EMPIRICAL METHODS FOR BIOETHICS: A PRIMER

LIVA JACOBY
LAURA A. SIMINOFF
Editors
EMPIRICAL METHODS FOR BIOETHICS: A PRIMER
ADVANCES IN BIOETHICS

Series Editors: Robert Baker and Wayne Shelton

Recent Volumes:

Volume 3: Values, Ethics and Alcoholism – Edited by W. N. Shelton and R. B. Edwards
Volume 4: Critical Reflections Medical Ethics – Edited by M. Evans
Volume 5: Bioethics of Medical Education – Edited by R. B. Edwards
Volume 6: Postmodern Malpractice – Edited by Colleen Clements
Volume 7: The Ethics of Organ Transplantation – Edited by Wayne Shelton and John Balint
Volume 8: Taking Life and Death Seriously: Bioethics in Japan – Edited by Takao Takahashi
Volume 9: Ethics and Epidemics – Edited by John Balint, Sean Philpott, Robert Baker and Martin Strosberg
Volume 10: Lost Virtue: Professional Character Development in Medical Education – Edited by Nuala Kenny and Martin Strosberg
EMPIRICAL METHODS FOR BIOETHICS: A PRIMER

EDITED BY

LIVA JACOBY
Associate Professor of Medicine, Office of the Vice Dean for Academic Affairs and The Alden March Bioethics Institute, Albany Medical College, Albany, NY, USA

LAURA A. SIMINOFF
Professor and Chair, Department of Social and Behavioral Health, School of Medicine, Virginia Commonwealth University, VA, USA

ELSEVIER
JAI


JAI Press is an imprint of Elsevier
This book is dedicated to our husbands Bill Jacoby and Jacek Ghosh for their unfailing support
This page intentionally left blank
CONTENTS

LIST OF CONTRIBUTORS ix

INTRODUCTION
Liva Jacoby and Laura A. Siminoff 1

SECTION I: PERSPECTIVES ON EMPIRICAL BIOETHICS

THE ROLE OF EMPIRICAL DATA IN BIOETHICS: A PHILOSOPHER’S VIEW
Wayne Shelton 13

THE SIGNIFICANCE OF EMPIRICAL BIOETHICS FOR MEDICAL PRACTICE: A PHYSICIAN’S PERSPECTIVE
Joel Frader 21

SECTION II: QUALITATIVE METHODS

QUALITATIVE CONTENT ANALYSIS
Jane Forman and Laura Damschroder 39

ETHICAL DESIGN AND CONDUCT OF FOCUS GROUPS IN BIOETHICS RESEARCH
Christian M. Simon and Maghboeba Mosavel 63

CONTEXTUALIZING ETHICAL DILEMMAS: ETHNOGRAPHY FOR BIOETHICS
Elisa J. Gordon and Betty Worder Levin 83
SEMI-STRUCTURED INTERVIEWS IN BIOETHICS RESEARCH
Pamela Sankar and Nora L. Jones 117

SECTION III: QUANTITATIVE METHODS

SURVEY RESEARCH IN BIOETHICS
G. Caleb Alexander and Matthew K. Wynia 139

HYPOTHETICAL VIGNETTES IN EMPIRICAL BIOETHICS RESEARCH
Connie M. Ulrich and Sarah J. Ratcliffe 161

DELIBERATIVE PROCEDURES IN BIOETHICS
Susan Dorr Goold, Laura Damschroder and Nancy Baum 183

INTERVENTION RESEARCH IN BIOETHICS
Marion E. Broome 203

SUBJECT INDEX 219
LIST OF CONTRIBUTORS

G. Caleb Alexander  Section of General Internal Medicine, Department of Medicine, MacLean Center for Clinical Medical Ethics, The University of Chicago, Chicago, IL, USA

Nancy Baum  University of Michigan, School of Public Health, Ann Arbor, MI, USA

Marion E. Broome  Indiana University, School of Nursing, Indianapolis, IN, USA

Laura Damschroder  Ann Arbor VA HSR&D Center of Excellence, Ann Arbor, MI, USA

Jane Forman  Ann Arbor VA HSR&D Center of Excellence, Ann Arbor, MI, USA

Joel Frader  Feinberg School of Medicine, Northwestern University, Chicago, IL, USA

Susan Dorr Goold  Internal Medicine and Health Management and Policy, University of Michigan, Ann Arbor, MI, USA

Elisa J. Gordon  Alden March Bioethics Institute, Albany Medical College, Albany, NY, USA

Nora L. Jones  Center for Bioethics, University of Pennsylvania, Philadelphia, PA, USA

Maghboeba Mosavel  Center for Reducing Health Disparities, MetroHealth Medical Center, Case Western Reserve University, Cleveland, OH, USA

Sarah J. Ratcliffe  University of Pennsylvania, Department of Biostatistics, School of Medicine, Philadelphia, PA, USA
LIST OF CONTRIBUTORS

Pamela Sankar  Center for Bioethics, University of Pennsylvania, Philadelphia, PA, USA
Wayne Shelton  Program on Ethics and Health Outcomes, Alden March Bioethics Institute, Albany Medical College, Albany, NY, USA
Christian M. Simon  Department of Bioethics, School of Medicine, Case Western Reserve University, Cleveland, OH, USA
Connie M. Ulrich  University of Pennsylvania School of Nursing, Philadelphia, PA, USA
Betty Wolder Levin  Department of Health and Nutrition Sciences, Brooklyn College/Brooklyn City University of New York, Brooklyn, NY, USA
Matthew K. Wynia  The Institute for Ethics, American Medical Association, Chicago, IL, USA
INTRODUCTION

Liva Jacoby and Laura A. Siminoff

In recent years, concerns over how to use the results of scientific advances, changing expectations of how medical decisions are made, and questions about the implications of demographic changes have raised ethical challenges regarding allocation of resources, justice, and patient autonomy. Bioethics – no longer the singular purview of moral philosophy – is now accepted as a legitimate field in the academic health sciences and is helping to guide policy and clinical decision-making. To achieve its full potential, it must seamlessly integrate the methods of the humanities, social sciences and medical sciences.

This volume is intended to open a window to how empirically based social research helps illuminate and answer ethical questions in health care. Its primary aim is to examine the nature, scope and benefits of the relationship between empirical social science research and bioethics. Through a thorough examination of key research methods in sociology, anthropology and psychology and their applications, the book explores the study of bioethical phenomena and its impact on clinical and policy decision making, on scholarship and on the advancement of theory. The many and varied illustrations of research investigations presented in this book, allow readers to learn how different methodological approaches can address a wide range of ethical questions on both micro- and macro levels. In this vision of bioethics, fundamental questions are formulated using the tools of the social sciences, and then systematically studied with the thoughtful and methodical application of empirical methods. In this way, bioethics achieves the widest
lens allowing it to become a translational, as well as theoretical, area of inquiry.

The book provides a primer and a guide to those who are interested in learning how to collect and analyze empirical data that informs matters of bioethical concern. It provides readers an in-depth understanding about a range of qualitative and quantitative research methodologies and is designed to convey the breadth, depth and richness of such work. This book demonstrates how synergy between the social sciences and moral philosophy can integrate into a new vision of bioethics. It is our hope that this approach will expand not only our understanding of the complex bioethical environment surrounding the provision of health care, but also will encourage continued collaboration across disciplines relevant to bioethics.

This book builds on over 20 years of work in which individual researchers have attempted to shine an empirical spotlight onto bioethics. In 1984, Fox and Swazey (1984) characterized American bioethics as devoid of recognition of the social and cultural forces influencing ethical phenomena and being “sealed into itself” (p. 359). Five years later, Fox (1989) produced an eloquent analysis of the relationship between bioethics and the social sciences characterizing it as “… tentative, distant and susceptible to strain” (p. 237). In her analysis, she described how each field contributed to the tension – bioethics largely due to its focus on individualism and equating the social sciences with a quantitative and non-humanistic perspective, and the social sciences due to their limited interest in studying values and beliefs and favoring structural and organizational variables which, she contended, reduced their understanding of the importance of ethical and moral values in society. Her conclusion was that the ethos of both fields, with resultant “blind spots,” constituted barriers to collaboration and synergy.

Since this bleak picture was articulated two decades ago, the relationship between the two fields has evolved to the point where bioethics is a multidisciplinary field of study (as opposed to a singular discipline), where moral philosophy, the medical sciences, the humanities and the social sciences intersect. This evolution of bioethics into a truly multidisciplinary field makes the present book not only possible but also necessary. In a sense, the book represents a culmination of the coming together of these areas of inquiry. Thus, in order to situate it appropriately, we present a brief overview of some of the literature that has helped move this development forward over the past two decades.

Leading the path are Social Science Perspectives on Medical Ethics, edited by George Weisz (1990) and Bioethics and Society: Constructing the Ethical Enterprise, edited by DeVries and Subedi (1998), The SUPPORT study
(The SUPPORT Principal Investigators, 1995), and several articles by sociologists, philosophers/ ethicists and physician/ ethicists. George Weisz, in presenting the rationale for the book, describes three key areas where he believed social scientists could make important contributions to medical ethics: (1) provide data, (2) place ethical problems in their social contexts and (3) facilitate critical self-reflection on the part of medical ethicists. Of note is that although Weisz chronicles the important contributions of the social sciences to what he then termed “medical” ethics, he did not see the social sciences as an integral part of the field. In one of the chapters, and perhaps one of the most forceful deliberations on the topic at hand, Hoffmaster (1990) – a philosopher- ethicist – introduces the notion of ethical contextualism whose aim is to “explain the practice of morality” (p. 250). In order to understand and explain morality and the nature of ethical dilemmas, he posits that the focus has to be on practice and on the social and historical contexts in which these dilemmas are located. In the same volume, Fox expounds on the limited attention given by bioethicists to reciprocity, interconnectedness and community when analyzing ethical phenomena. Despite the often pessimistic perspectives presented on the differences in disciplinary approaches, Weisz concludes the volume with the hope that the exchange of ideas and the “stretching of disciplinary boundaries” (p. 15) will continue.

In 1992, Hoffmaster (1992) continued his vigorous arguments for incorporating contextualism into bioethical analysis by deliberating on the question “Can Ethnography Save the Life of Medical Ethics?” Presenting ethnography as a method to better understand the structural forces and individual particularities that influence values, behaviors and decisions in medical settings, he posits that moral theory must be tested in practice in order for theoretical development to occur. To justify his position, Hoffmaster uses several highly acclaimed ethnographic studies on decision making in neonatal intensive care units and genetic counseling.

One of the most comprehensive empirical studies in bioethics so far has been the SUPPORT study, published in 1995. With two physicians as its principal investigators, the study employed quantitative as well as qualitative methods to examine a range of end of life care issues, both descriptively and experimentally, and produced a vast amount of data showing not only troubling results but engendering new questions, debates and more significant research. Despite criticism from many quarters concerning the study’s methods and aspects of the intervention, this investigation was evidence of how bioethics and the social sciences informed each other and it constituted a vital example of the testing of ethical theories.
in practice. In fact, it demonstrated that thinking of bioethics simply as a branch of moral philosophy that could be “aided” by social scientific inquiry, was largely an outdated approach to the field.

In the first chapter of the volume Bioethics and Society: Constructing the Ethical Enterprise (1998), DeVries and Conrad call attention to the “bad habits” of analytic bioethics that have resulted in its asocial nature and practical irrelevance (1998). They also point out the shortcoming among social scientists of separating data from norms, arguing that, “a richly rendered social science ... must be normative” (p.6). Introducing the notion of “social bioethics,” they further posit that without empirical underpinnings, princip¬alism and analytic bioethics will never lead to workable solutions to moral problems. A recurring theme in this volume is a critique of how the weight given to autonomy and individualism in American bioethics limits bioethicists’ recognition of social and cultural contexts and respect for pluralism. In one chapter, bioethics is criticized for protecting the status quo through its lack of attention to justice and structural factors in the health care system.

In 2000, The Hastings Report moved the discourse about the relationship between the social sciences and medical ethics forward by publishing an issue entitled “What Can the Social Scientist Contribute to Medical Ethics?” In one of the articles, sociologist Zussman (2000) recognizes how medical ethicists have become more inclined to incorporate empirical data in their analyses and that social scientists studying medical ethics have shown more openness to the normative implications of their research. Importantly, he states that the classic “ought-is” distinction and other differences between the fields denote complementarity rather than incompatibility. He concludes by calling for a combination of empirical methods and an applied ethics model as an approach to pursuing scholarly work immersed in practice. In the same issue of the Hastings Report, Lindemann Nelson – a philosopher – states that the social sciences can and should help bioethics by enriching an understanding of prevailing ethical values and how these “come to be installed or resisted in patterns of practice” (2000; p. 15). Using the SUPPORT study as an illustration of his arguments, he recognizes the need to give attention to how structural and institutional factors impact human behaviors and practice patterns.

Bioethics in Social Context, edited by Hoffmaster, and Methods in Medical Ethics, edited by Sugarman and Sulmasy were published in 2001. Both constitute important work in this genre, and reflect what many of the scholars reviewed above have called for. The first volume consists of essays on qualitative research emphasizing the context within which ethical decisions take place. The second volume describes a wide range of empirical approaches
to studying bioethical questions – from religion and theology, history and legal methods to ethnography, survey research, experimental methods and economics. Setting the stage for their book with a discussion on the relationship between descriptive and normative research in medical ethics, Sugarman and Sulmasy posit that empirical research “can raise questions about the universalizability of normative claims” and “can identify areas of disagreement that are ripe for ethical inquiry” (p. 15). Claiming that “good ethics depends upon good facts” (p. 11), one of their premises is that good moral reasoning needs both moral and factual elements. Their conclusion is that descriptive and normative inquiries are mutually supportive. Including a number of methods outside as well as within the social sciences, the book sheds light on the wide range of disciplines that have contributed to the study of bioethical phenomena during the past couple of decades.

In the present book, we advance the field of empirical bioethical inquiry another step by focusing on empirical methods in bioethics and on their practical applications to investigating a wide spectrum of bioethical problems. One thing that is noteworthy, when compared to much of the work preceding this volume, is that we use the term “bioethics” rather than “medical ethics.” We believe the term “bioethics” denotes a broader meaning related to the study of the ethical, moral and social implications of the practice of medicine in all its aspects along with the social and ethical problems generated by new biotechnology and biomedical advances.

We have included eight basic research methodologies – and asked the authors to describe how they have employed “their” particular method to examine matters of bioethical concern. These range from informed consent, human subjects research, end of life care, decision making regarding organ donation, to the tension between privacy rights and the facilitation of medical research to community standards concerning health care spending priorities. Before the eight chapters on methodology, are two chapters that provide two different and, in many ways, complementary perspectives on contemporary empirical bioethics – one by a practicing physician/ethicist and the other by a clinical ethicist/philosopher. Their discussions of how empirical research has contributed to their areas of expertise, demonstrate how such research brings together moral philosophy, clinical practice and clinical ethics and serves as a valuable framework for the remainder of the book.

The methodology chapters are organized under the headings of methods that are generally classified as “qualitative” research methods and those that are “quantitative.” The particular methods were chosen that have been demonstrated to have a practical application in the study of bioethics, including those that are used frequently and have proven to yield valuable
results. The qualitative chapters include: Content analysis, Focus groups, Ethnography and Semi-structured interviews. Chapters that focus on quantitative methods are: Survey Research, Hypothetical Vignettes, Deliberative Procedures and Intervention Studies.

The section on qualitative methods begins with a chapter on content analysis by Forman and Damschroder. As the first chapter in this section, it introduces the reader to a method that constitutes the basis for much of analysis of qualitative research data and, as such, frames the following three chapters. The authors provide a detailed description of how qualitative content analysis can be used to analyze textual data of various kinds and is aimed at generating detail and depth. Their focus is on the examination of data gathered through open-ended interviews. By using specific examples, they illustrate how content analysis provides comprehensive descriptions of phenomena; illuminates processes; captures beliefs, motivations and experiences of individuals; and explains the meaning that individuals attach to their experiences. The chapter provides ample information on the many steps inherent in content analysis, from the framing of the research question, deciding on the unit(s) of analysis, to the various and specific forms of engaging with the data, and performing the actual analysis.

In the second chapter, Simon and Mosavel discuss focus groups as a useful method to stimulate discussion and gather data on multifaceted and complex bioethical issues. The use of focus groups in bioethical inquiry has increased in recent years, and drawing on their own research experiences in South Africa, the authors explore and highlight some of the uses of this method as an investigatory tool. Referring to focus groups as a method that is comparatively cost effective and easy to implement, Simon and Mosavel present the reader with practical information on the processes and procedures of designing and conducting focus groups in an ethical, culturally appropriate and scientifically rigorous way. They go on to present a novel and unique form of analysis of focus group data, referred to as “workshop-based summarizing and interpretation” and describe how this approach was used with members of communities in South Africa as part of their research. Finally, they provide insights into ways of disseminating findings from focus group research.

The chapter by Gordon and Levin illuminates how ethnography, as one of the most prominent empirical methods in early bioethics research, has and still does contribute significantly to our understanding of bioethical phenomena. In this chapter, Gordon and Levin start by giving a brief overview of seminal ethnographic work in the field, followed by a detailed description of participant observation as “the heart” of ethnography. Using examples from research of their own and that of others conducted in a
variety of health care settings, they continue by outlining the steps involved in preparing and implementing a participant observation study in the field. Their accounts give the reader valuable insights into the unique role of the participant observer, the significance of good note-taking, and common challenges encountered by ethnographers. The authors offer helpful ideas on precautions that researchers can take in order to maximize the rigor of their research and to generate valid and meaningful data. The section on the elements of data interpretation and analysis connects with Forman’s and Damschroder’s chapter on content analysis, providing the reader with a comprehensive guide to the collection and analysis of qualitative data. The authors end by reviewing ethical considerations in conducting ethnographic research as well as the strengths and weaknesses of such research.

The final chapter on qualitative methods describes semi-structured interviews. Along with surveys, interviews have long constituted one of the basic methods in social science research. In this chapter, Sankar and Jones begin by presenting the advantages of semi-structured interviews characterized by the combination of closed-ended questions with open-ended queries making possible both comparisons across subjects and the in-depth exploration of data. Comparing to quantitative research, the authors contend that the main strength of semi-structured interviews is in the richness of the data they generate and that the method is particularly useful in exploratory research. Paying a good deal of attention to considerations in designing an interview guide, Sankar and Jones discuss the importance of pilot testing, and using an example from their study on medical confidentiality, address ways of maximizing the validity of interview questions and steps involved in finalizing questions and queries. Important segments of the chapter are the discussion of sampling and the actual conducting of semi-structured interviews.

Focusing on audiotaping and the digital recording of interviews, Sankar and Jones provide a valuable complement to Gordon’s and Levin’s discussion of note taking in ethnography. Similarly, the review of coding procedures and the particular approach referred to as “multi level consensus coding,” add to the perspectives offered by Foreman and Damschroder in their chapter on content analysis.

The first chapter in the book’s section on quantitative methods presents the basics of survey research. As Alexander and Wynia point out, surveys have been the bedrock of much of the research conducted by social scientists, stating that, few researchers who conduct empirical research in bioethics do not use some survey research techniques. Alexander and Wynia further observe that surveys about ethically important topics have made important contributions to bioethics. However, it is not a simple task to
conduct a good survey. The authors make clear that in order to obtain meaningful information from a survey, the researcher needs to pay careful attention to the development of the survey’s design, including formulating the research question, how to draw the sample, developing the questionnaire and how the data will be managed and analyzed. The authors contend that using rigor throughout this process can be the difference between an important study that makes fundamental contributions and one that is irrelevant to ethical analysis, health policy or clinical practice. The chapter provides a primer to the reader in how to balance rigor with feasibility at all stages of survey development, fielding, analysis and presentation and helps the reader plan, develop and conduct a survey.

A related technique to survey research is the use of hypothetical vignettes. This technique is especially relevant to bioethics research where it can be difficult to directly observe certain ethical problems because of the intensely personal nature of the questions of interest (removal of ventilator support from a patient), the rarity of the occurrence (requests for organ donation in the hospital), or a sensitive question that may reside at the edge or over the edge of what is legally permitted (assisted suicide and euthanasia). As Ulrich and Ratcliffe point out, hypothetical vignettes provide a less personal and, therefore, less threatening presentation of such issues to research participants. The chapter provides an overview of hypothetical vignettes with examples of how this method has been used to examine and analyze critical ethical problems. It reviews ways to evaluate the reliability, validity, strengths and limitations of studies using vignettes. The chapter will take the reader through what constitutes a vignette, how to develop a vignette about a bioethics-relevant problem, how to evaluate the psychometric properties of vignettes, the determination of sample size and sampling considerations and examples of published vignettes used in empirical bioethics research.

The chapter by Goold, Damschroder and Baum will introduce many readers to a methodology unfamiliar to them – deliberative procedures. This methodology is based on theories of deliberative democracy with the idea of providing community members with a “voice” in community-wide decisions, for instance about health care spending priorities or research regulation. Deliberative procedures offer an opportunity for individuals to assess their own needs and preferences in light of the needs and desires of others. In bioethics research, deliberations involve individuals in a community decision-making process about bioethical issues with policy implications and may provide acceptance and legitimacy to a given issue within a population.
The authors describe how deliberative procedures entail gathering non-professional members of the public to discuss, deliberate and learn about a particular topic with the intention of forming a policy recommendation or casting an informed “vote.” For researchers involved in exploring bioethical issues, deliberative procedures can be a valuable tool for gathering information about public views, preferences and values. This chapter focuses on de novo deliberative procedures used for research purposes, or combined policy and research purposes, where sampling issues, and research aims are known and planned up front. The chapter offers a review of methodological considerations unique to, or particularly important for, deliberative methods include sampling, specifically substantive representation, what to measure and when, the use of group dialog in the data collection process, and the role that deliberative procedures can play in educating the public and informing policy.

The final chapter deals with intervention research. As Broome makes clear, the use of intervention designs, while a relatively recent phenomenon in bioethical inquiry, has a distinct and important role to play in advancing the field of bioethics. Its importance will grow as more empirical bioethics research provides data that not only informs policy and/or practice, but asks questions about what policies or practices work best. Although many ethical questions of interest are not appropriate for intervention research, the author contends that some questions can only be answered using experimental or quasi-experimental designs. The chapter provides the reader with a review of the application of experimental methods to bioethics research including randomized controlled trials and quasi-experimental designs ranging from the more rigorous two-group repeated measures or pre-test/post-test designs to the one-group post-test-only design. Strengths of each design, including the threats to internal and external validity, are presented. As Broome stresses, not all bioethical phenomena are appropriate for study using experimental or quasi-experimental designs. Examples of intervention research related to informed consent are provided.

With this book, we hope to create enthusiasm for empirical research that will continue to bring synergy between disciplines representing bioethics and, in so doing, further enhance our understanding of bioethical phenomena.

REFERENCES


SECTION I: PERSPECTIVES ON EMPIRICAL BIOETHICS
This page intentionally left blank
THE ROLE OF EMPIRICAL DATA IN BIOETHICS: A PHILOSOPHER’S VIEW

Wayne Shelton

THE CRISIS OF TRADITIONAL ETHICAL THEORY

How many textbooks or introductory articles in bioethics begin with a section on ethical theory? Of the many that do, the relevance of basic theories of utilitarianism, deontology, virtue ethics, feminist ethics, casuistry and so on, is assumed. These theories are also considered in light of the well-accepted principles of medical ethics: (1) respect for patient autonomy, (2) beneficence, (3) non-maleficence and (4) justice. Those of us trained in philosophy find these sections on theory terse summations of complex philosophical views. Physicians and nurses, and others not trained in philosophy, sometimes struggle to get their gist, and end up with an ability to make a basic analysis and formulate arguments about ethical problems from each of these perspectives, and to write and discuss the issues that arise with fellow ethicists. But how essential are these theoretical perspectives to the real work of clinical ethics consultants? It is important that we do not forget just how applied and practical that work is.

Regardless of one’s background perspective coming into bioethics, particularly clinical ethics, if he or she wishes to become a clinical ethics consultant and work in the field applied clinical ethics, it is essential to
grapple first hand with value conflicts in real life situations. Applied clinical ethics is the study of range of value-laden, ethical conflicts and dilemmas that arise in the clinical setting, and especially in the physician–patient relationship. The ethics consultant, a specialist in applied clinical ethics, combines both core clinical skills necessary to provide support and help people embroiled in value conflicts, as well as to employ advanced analytical skills and knowledge in offering a considered analysis of the opposing value positions in order to make recommendations consistent with the rights and obligations of those involved. Some ethics consultants are also expected to communicate directly with patients and families to facilitate an ethically acceptable outcome. For those ethics consultants directly involved in clinical value conflicts, both in terms of their practical resolution in individual situations and in terms of their import for academic reflection, what is the role of ethical theory in the work of applied ethics? And more specifically, can traditional ethical theory serve as a normative basis on which we judge one moral option better than another? These are indeed legitimate theoretical ethical questions. But, it is doubtful that our typical educational training prepares ethics consultants to answer them.

It is unfortunate that philosophical ethical theories still have an uncertain and, I would say, awkward fit into the practical realm of applied ethics. As someone who has taught a number of graduate courses in clinical ethics, the justification I use for providing a brief introduction to ethical theory, after the nature of real clinical value conflicts have been established, is to stimulate the student’s imagination in analyzing and seeing all the possible ways of viewing and justifying a case ethically. But again, there is the continuing sense of not knowing quite how to use ethical theory in bioethics in relation to problem solving, particularly in knowing the normative force of ethical theory. The purpose of this brief chapter is to provide an innovative alternative that was proposed by John Dewey in the early part of last century. I will argue that the emergence of an applied ethical field like bioethics provides the occasion to reconstruct our understanding of the relation of theory and practice, and creates a whole new appreciation of, and need for, empirical bioethics. From this perspective, ethical theory can be viewed in a different, and more constructive, light.

Dewey’s critique of western philosophy and ethical theory was the springboard for a new understanding of ethics. Dewey believed that western philosophy since Plato was largely based on false dualisms that created a bogus dichotomy between some ultimate, rational understanding of truth versus what can be known from ordinary experience. It is because of such dichotomies that the “Is-Ought” problem has appeared so impenetrable, i.e.
because reason and experience are assumed to be disconnected, empirical knowledge is of little or no help in deriving moral obligations. (Those interested in Dewey’s critique, please see *The Quest for Certainty: A Study of the Relation of Knowledge and Action* (1960)). Dewey was quite convinced, as are many of his followers, that western philosophy has failed to reach final understandings of rational truth, and that we are wasting our time to continue to search for them. He contended that there has never been and will never be final agreement about such quests. It is for this reason that Dewey believed philosophy, including ethics, is in need of reconstruction – a whole new pragmatic understanding that grounds ethical inquiry in human experience. In light of this critique of western philosophy, the problem of knowing just how to use traditional ethical theory can be fully appreciated, and why the pragmatic turn in bioethics marks a new beginning for philosophy and ethics.

**PRAGMATIC ETHICS**

Pragmatic ethics begins with the reality of lived experience of human beings who are connected biologically, socially and politically within a natural environment. Thus, the moral life is connected to the conditions that best foster human flourishing and reduce suffering. “Right” and “good” are the ends of moral inquiry, not assumptions grounded in normative religion or philosophical theory. Many have taken this approach as a step toward moral relativism and crisis of value. But pragmatists believe the turn toward naturalism is the occasion to fully grasp the vital role of human beings in shaping their own fate, and promoting a better society. Human beings’ highest aspirations – justice, peace and alleviation of suffering – lie in the enhancement of human intelligence and forms of inquiry that allows humans to better understand how to craft a better future for everyone. Questions of moral conflict, thus, become more problems of strategy for humans to resolve using principles that guide expedient resolution within a social circumstance or context. According to this perspective, since ethical reflection begins and ends in experience and not some pre-established, a priori patterns of reasoning, the tension between “is” versus “ought” begins to subside.

Because ethicists no longer work only in rarified academic settings but more and more function as applied ethicists alongside practitioners in clinical settings, there is a need to provide working solutions to medical ethical problems. Not many applied ethicists I know are looking for
ultimate answers; rather, they seek solutions that help people in conflict to make decisions and to cope better in their environments. Ethicists must themselves know the experiential landscape of a given situation in order to provide meaningful and helpful moral advice to practitioners. But in order to have viable interpretations of such situations ethicists need to be able to use imagination to creatively propose viable strategies for amelioration of the human condition – both at the micro and macro level. This requires knowledge and understanding of the empirical circumstances in which value conflicts arise. In order to generate empirical knowledge the field of bioethics must rely on the same empirical methodological basis as the social sciences. This is consistent with Dewey’s hope for philosophy: that it would be put to work along side the other sciences for the betterment of society.

**EMPIRICAL BIOETHICS**

In its reconstructed role, ethical theory is no longer isolated from experience in preexisting forms as it appears in traditional philosophical ethics, but is connected to experience and informed by it. The goals of “right” action is not to make a determination of a final moral duty, but to make provisional statements that must be continually monitored and, like scientific claims, revised when warranted by further empirical knowledge. The philosophical quest of seeking final rational answers is replaced by a commitment to improve at least certain aspects of human life. For the clinical ethics consultant, it is to improve the conditions for people with conflicting values and/or preferences in particular clinical settings.

Thus, empirical knowledge, and therefore, empirical bioethics are essential to the field of bioethics, and all ethical inquiries that seek solutions to value conflicts. The rise of empirical knowledge in bioethics also allows us to generate new knowledge about the empirical landscape of clinical ethical conflicts. This knowledge leads to greater insight into the associative and causal elements that generate ethical conflicts in the first place. With more empirical knowledge, more strategic mastery of the course of clinical events is possible. This can occur by providing more effective ethics consultations in individual cases. But empirical knowledge can also become a basis for developing broader, preemptive strategies for dealing with ethical conflicts. It is a sign of a maturing field, more aligned with other fields based in scientific methodology that many applied ethicists, health care practitioners, policy makers and others are becoming more focused on quality of care improvement by reducing the incidence of ethical conflicts. This is done by
testing, through scientific empirical studies, the efficacy of new strategies that focus on improved management of the contributing elements of ethical conflict. Most ethics consultants quickly realize that it is much more efficient to prevent major ethical conflicts from arising than to have to grapple with them in their often final and intractable manifestation as ethical dilemmas. This requires the ethicist to become proficient in empirical studies of clinical decision making and outcomes and related issues that affect patient and family care. A major empirical study in the Surgical ICU at an academic health care institution, led by a philosopher/ethicist, is an example.

AN ILLUSTRATION

The study grew out of extensive experience providing ethics consultations in the ICU setting. Based on recent ICU data accumulated from an internal study, we learned that up to 60% of patients where ethics consultations were done, died. In these cases, families typically go through extremely stressful and, sometimes, gut-wrenching experiences of decision making for their loved ones. Families in this setting are routinely required to make decisions about whether or not to continue life-sustaining treatments, and how best to follow the expressed wishes of the patient in situations where the patient lacks capacity. In the course of providing ethics consultations, it is often necessary to have extensive conversations with the family of an incapacitated patient allowing them to share their intimate knowledge of the patient’s values and how those values apply to medical decision making. Many times a consensus emerges about the right course of action, but unfortunately, sometimes there is deep disagreement between family members, and between family and care providers. At those times, the ethics consultant is there to help mediate what is often a value-laden conflict, which has reached an impasse. Attitudes and dispositions among those involved can become hardened and people’s positions entrenched. Conflicts that drag on frequently lead to a lack of clarity in defining goals of care and patients may stay in the ICU for an extended length of time, using costly resources. In situations where conflicts persist, it is possible that patients are receiving care that is not medically indicated. In many instances, the ethics consultant can serve as an outside mediator, clarifying facts and values, and help the parties in conflict reach mutually acceptable outcomes. But from the ethics consultant’s point of view the time of entry into the situation is late, and improvements in care are made, one case at a time. Commonly, much has happened prior to the point of the ethics consultation that impacts
the conflict, and often the key factor is how the information flow and communication has occurred between the health care team and the family. Over the years I have often drawn from my experience as an ethics consultant dealing with ethical conflicts in the ICU for teaching purposes to illustrate the ways how care providers can interact with families so as to prevent conflicts from arising in the first place. Based on these observations and insights, my hunch grew into a well-formulated research question about how a highly tailored plan for focused family support might reduce ethical conflicts, as well as increase family satisfaction and reduce length of stay for patients most at risk for extended length of stay. Thus, the inspiration for a study!!

As a philosopher trained in clinical ethics consultation, and also in social work and health care policy, my interests expanded to considering how one might do an empirical study of my, at that point, hunch. After an exhaustive literature review, numerous discussions with many clinicians and researcher, several pilot studies, two related publications (Gruenberg et al., 2006; Rose & Shelton, 2006) and extensive planning and proposal writing, funding was received to begin the “ICU study”. The study is designed to test the hypothesis that a focused, multidisciplinary model of family support, including the combined resources of ethics consultation, pastoral care, social work and palliative care, all led and coordinated by a nurse practitioner will lead to (1) increased family satisfaction with care, (2) decreased unnecessary and unwanted care, and (3) reduced cost and resource utilization.

The intervention will use a nurse practitioner to gather information about the family from the non-medical support services. Working directly with the physicians, he or she will ensure that this information is used meaningfully and robustly in the medical decision making and goal-setting process. Thus, the nurse practitioner will be the crucial link between the physicians in charge of directing medical care for the patient and the family. Thus, the intervention will resolve one of the most pervasive and well-known problems of hospital case: no one taking responsibility for the patient and family, with one central line of communication to manage the flow of medical and other essential information.

This perennial problem regarding the flow of information to families and patients in hospitals was identified by Michael Balint as “collusion of anonymity” in his book *The Doctor, His Patient and the Illness* first published in 1957. Since that time, the problem has grown much more complicated with the rise of highly complex, specialized medical fields, especially in the context of large teaching hospitals. It is common in such hospital settings for families to come into contact with numerous physicians
from many specialty services, each with their own medical perspective and sometimes at variance with one another in terms of the prognosis and goals of care for the patient. To say this situation can be confusing for a stressed family distraught over the illness of their loved one is an understatement. Such confusion can also fuel misunderstandings, leading to confusion and strong emotions of anger and resentment. Most experienced ethics consultants know, at least anecdotally from their own experiences, that such situations are a breeding ground for ethical conflicts and dilemmas, and for dissatisfaction with care among families.

The goal of the study intervention is to preclude clinical ethics dilemmas of these kinds, and to engage in what is sometimes referred to as “preemptive ethics” by preventing the problems from occurring. Instead of dealing with ethical problems one at a time, from one crisis to another, the study will provide the benefit of collecting empirical data that can illuminate and address underlying root causes of ethical conflicts at the bedside. Our premise is that by providing focused support to stressed families, they will gain an enhanced ability to make decisions with ease and understanding, according to their values and preferences and those of the patient. The emphasis of the study is on improving the quality of care for seriously ill patients, and testing whether such an improvement reduces cost of care.

Is this a clinical ethics study? I would say, yes! Most importantly, it shows how an emerging field like empirical bioethics is connected to other key areas of health care research, such as quality improvement, resource utilization and outcomes studies. Therefore, one significant point is that as bioethical inquiry evolves into empirical investigations it clearly becomes more multidisciplinary and gains more standing as an important area of health care research.

CONCLUSIONS

In light of this beginning trend toward empirical bioethics, where does this leave ethical theory, and what is its role in applied ethics? Applied ethicists who have taken the pragmatic turn and are now exploring empirical bioethics in terms of outcomes research, see the tradition of western ethical theory as a body of literature with no special moral authority. This is not to say ethical theory is not interesting or even important, but clearly the perspective of the naïve graduate student looking for foundational moral authority is gone. Instead, its use is more as a set of tools – they are handy to
have at one’s disposal. They provide ways of asking relevant questions, structuring arguments and formulating alternative resolutions to pressing problems. To the extent theory is now used for the empirically oriented, applied philosopher, they represent structured ways of conceiving alternative moral perspectives that stem from direct experience in the empirical details of ethical problem solving. The theories stem from the imagination and provide the ethical visions for forging a better state of affairs with respect to some human value problems.

Perhaps with more bioethical empirical research there will emerge a new type of fully articulated theory, as Dewey was hoping, that is better suited for practical problem solving. This also means much less of a focus, if any, on the questions that have flowed out of modern philosophy, based on dualisms that force us to separate theory and practice, mind and body, and fact and value. So at this point of a new beginning for applied philosophy, a field we call empirical bioethics, it is a good thing that philosophers are getting their hands dirty in the real world of experience and grounding their theoretical approaches toward improving the human condition.

REFERENCES


INTRODUCTION

While some of us enjoy engaging in many forms of bioethical activity, including philosophical analysis and debate, clinical ethics consultation, and empirical research, only the latter matters much to the practicing physician. Practically minded, most doctors have little concern with fine moral distinctions when faced with a patient’s request for assistance in dying or a pharmaceutical company’s offer to attend a product “consultation” session at a first class resort in addition to an attractive fee for participation. Physicians want to know what facts might bear on ethical questions they confront, how ethical conflicts that have an impact on patient care can be understood and resolved, and whether research reveals consistently clear, helpful findings. The following discussion offers some examples of how empirical research related to bioethical issues has provided evidence and guides for physicians at both individual-patient care and policy levels, and further reviews areas that warrant continued research attention.
AREAS OF RELATIVELY CLEAR EVIDENCE

Informed Consent

During the second-half of the twentieth century in the United States, medicine experienced enormous change. Scientific and technological advances made medical interventions vastly more effective and health care became a major economic engine. In accord with the political and cultural changes emphasizing the rights of individuals, ethical and legal thinking about the relationship between professional providers and researchers on the one hand and patients and subjects on the other hand, shifted from beneficent paternalism to consumerist autonomy. The doctrine of informed consent became the judicial and ethical cornerstone of decisions about medical treatment and/or research participation. Public policy, fashioned to prevent recurrences of Nazi medical atrocities, the kind of researcher arrogance documented by Beecher (1966), Barber, Lally, Makarushka, and Sullivan (1973) and Gray (1975), and medical over-treatment noted by the President’s Commission for the Study of Ethical Problems in Medicine and Biomedical and Behavioral Research (1983) and Weir (1989) focused on individual consent by the patient, subject, or surrogate authorization.

Unfortunately, research about the realities of informed consent shows that practice fails to live up to ethical theory and legal doctrine. In the context of medical treatment, evidence indicates that differences in knowledge, social status, and emotional states of practitioners and patients/family members undermine the value of information about the risks, benefits, and alternatives to proposed plans of care (Lidz et al., 1983; Appelbaum, Appelbaum, & Grisso, 1998; Siminoff & Drotar, 2004). Research also reveals that physicians often lack adequate training for and motivation to communicate clearly with patients and surrogates (Angelos, DaRosa, & Sherman, 2002; Sherman, McGaghie, Unti, & Thomas, 2005). Similar problems plague the implementation of the doctrine of informed consent in research. Potential subjects frequently suffer from the “therapeutic misconception” believing that proposed clinical research is designed to provide them with the best known therapy, despite randomization schemes ensuring distribution of subjects into treatment arms of competing value (Lidz & Appelbaum, 2002). In recent studies on research participation among children with cancer, Kodish, et al. (2004) have dramatically demonstrated that parents commonly do not understand that their children could receive treatment without enrolling in research nor do they typically comprehend that treatment assignment proceeds by chance,
rather than the doctor’s deliberative decision about which regimen would work best for their child. Finally, especially in the research context, studies have repeatedly shown that the obsessive, legalistic focus on written consent forms generally fails to produce truly informed consent (Baker & Taub, 1983; Ogloff & Otto, 1991; Waggoner & Mayo, 1995; Lawson & Adamson, 1995; Agre & Rapkin, 2003). Consent forms, despite all the attention lavished on them by investigators, research coordinators, institutional review board (IRB) members and staffs, commonly contain language far too complex and technical for the general population to comprehend. What seems astonishing, in the face of years and volumes of research about the failure of efforts to effectuate meaningful informed consent, is the continued reliance on incomprehensible consent forms, and the lack of adequate preparation of clinicians (Wu & Pearlman, 1988; Sherman et al., 2005) or research team members for communicating clearly to patients and their loved ones about illness, treatment, research, and about the maintenance of ethical and legal notions of the consent process. The disconnection between ethical theory and application suggests one of two things: the theory needs to change to better reflect social practicality and/or the need for clinicians and investigators to become more creative and diligent in the way they convey information and assess understanding.

Some recent developments, largely beyond the scope of this brief review, provide some hope that a much better job can be done with the process of informed consent, though not without substantial investment of time and resources. Researchers starting from the perspective of appropriate use of health care resources, particularly those noting wide regional variations in practice in the United States (Wennberg, Fisher, & Skinner, 2004), have become interested in standardized presentation of information to patients and surrogates. A recent review by O’Connor, Llewellyn-Thomas, and Flood (2004) concluded that use of various patient decision aids, such as interactive computer programs, up-to-date multimedia presentations, tools to help consumers identify and actualize their health-related values, and forms of personal coaching by health care professionals or experienced peers can improve the quality of decisions and reduce the use of invasive interventions without a decline in health outcomes. We will need additional efforts to assess whether such mechanisms enhance the informed consent process and if they have practical and cost effective use in everyday health care settings. This is particularly the case in situations involving emotionally charged decisions, such as testing for genetic susceptibility to disease, forgoing life support, or the choice among alternative therapies for cancer.
Advance Directives

With the dramatic increase in intensive care units (ICUs) in American hospitals during medical and technical advances in the past 50 years, patients and families became concerned with over-treatment leading to the “right to die” movement. A series of legal cases related to family members’ requests to discontinue unwanted prolongation of life for their loved ones led to the notion that patients could avoid undesired life-sustaining interventions through preparation of documents designed to communicate treatment preferences once they lost capacity to interact effectively with health care providers. The first case to receive national attention in the U.S., involved a young woman named Karen Ann Quinlan who sustained severe brain injury. After prolonged treatment, including mechanical ventilation, her parents requested discontinuation of life support, but the patient’s doctors and hospital administrators, fearing legal sanctions, refused. The New Jersey courts in this case (In re Quinlan, 1976) and another subsequent seminal case involving Claire Conroy (In re Conroy, 1985), an elderly woman with several chronic conditions and no longer able to speak for herself, helped established criteria doctors and surrogates might use to justify forgoing life-sustaining treatment, recognizing the importance of evidence of what the patient him- or herself would want under the circumstances. This became a matter of Supreme Court consideration in the Nancy Cruzan case (Cruzan v. Director, 1990), which came to center on the issue of the quality of evidence necessary to establish the patient’s wishes.

As a result of such cases, institutional policies and, eventually, federal and state laws, began to recognize “advance directives” as a valid means by which patients and surrogates could communicate their wishes in the face of lost decision-making capacity. One type of document, the “instructional” directive, in which the individual attempts to project what treatments to employ or avoid if she/he can no longer express him- or herself, was designed to provide the means for limiting – or, in some cases, continuing – medical interventions, especially in situations where treatment might involve marginal benefits.

Research has shown that instructional directives – whether oral or written – lead to little change in what happens to patients. In the SUPPORT study, research nurses with the responsibility for ascertaining and communicating treatment preferences of patients to physicians in ICUs, failed to affect physicians’ decisions when compared to decisions among matched control
patients without enhanced communication intervention (SUPPORT Principal Investigators, 1995). Various other studies have documented that even when one can find the properly executed document, no matter what form a prescriptive written instrument takes (i.e., using check boxes, blanks that patients fill in with their own words, prepared descriptions of therapies to avoid or use, etc.), clinicians frequently feel that their patients’ circumstances insufficiently match what the document says or anticipates. Thus, as suggested in an Institute of Medicine (IOM, 1997) study, *Approaching Death: Improving Care at the End of Life* and in a review by Lo and Steinbrook (2004), instructional directives may confuse, rather than clarify, end of life treatment decisions. Based on current research, and in the wake of the Schiavo case (Quill, 2005), “proxy” directives, or a combination of surrogate appointment, clarifying who should make decisions for the incapacitated patient, and instructions, rather than prescriptive documents alone, are needed.

**Health Care Disparities**

One of the tenets of modern bioethics involves the importance of social justice. With regard to medical care, most ethicists believe that everyone should have access to a “decent minimum” level of care, presumably including at least that which primary care physicians can deliver. In research settings, most agree and federal regulations require that the benefits and burdens of research and the advantages of access to research studies should be distributed equitably across social classes, ethnic groups, males and females, young and old, and other subgroups. Unfortunately, empirical research suggests that our health care and research systems have not lived up to these ideals.

Politicians like to proclaim the glories of the health care “system” in the U.S. Studies show that some miraculous things do occur, as long as patients or family members have the means to ensure payment for desired services. For example, kidney transplants for those with end stage renal disease occur in substantially greater proportion among white, middle-class patients than among those who are poor and African-American (Eggers, 1995, Epstein et al., 2000; Churak, 2005). This holds true even though kidney failure affects a significantly larger percentage of the African-American population (Martins, Tareen, & Norris, 2002). When eliminating the variable of direct payment, such as within the Veterans Administration system, research
shows that African-American patients with equivalent medical conditions receive less aggressive care for serious heart disease than do Caucasians, i.e., fewer referrals for cardiac catheterization and/or coronary artery bypass surgery (Whittle, Conigliaro, Good, & Lofgren, 1993; Peterson, Wright, Daley, & Thibault, 1994). Recent findings have indicated similar results when comparing male cardiac patients with females, with the latter receiving less intervention (Maynard, Every, Martin, Kudenchuk, & Weaver, 1997; Fang & Alderman, 2006). The lack of access of poor patients to primary care has meant greater, more expensive, and less efficient and effective care (often in emergency departments), leading to delayed or inadequate maintenance and preventive services, and thus more frequent and/or more severe exacerbations of chronic conditions such as asthma, arthritis, and diabetes (Forrest & Starfield, 1998; Stevens, Seid, Mistry, & Halfon, 2006).

Similar patterns can be seen in the research arena. For example, as only a small proportion of children require extensive medical intervention, drug manufacturers often decline to include children in clinical studies on new medications (Yoon, Davis, El-Essawi, & Cabana, 2006). As a result, those treating young patients lack systematic knowledge of proper dosing and differences in toxicities for immature human bodies. In the same vein, fear of liability, rather than patient benefit, seems to have affected pharmaceutical companies’ decisions not to study the effects of new drugs in women who are pregnant or at risk of becoming pregnant, even though these women may have or may acquire the same conditions in need of treatment as women who cannot bear children. Those favoring the inclusion of known-to-be or possibly pregnant women in clinical studies view their exclusion as discriminatory and feel that the women, not corporate legal counsel, can and should balance potential harms to themselves or their fetuses compared to the benefits that a clinical trial may offer (McCullough, Coverdale, & Chervenak, 2005). Only additional research focused on this issue can begin to clarify both the likelihood of harm and the adequacy of decision making in the face of possible or actual pregnancy.

Further research aimed at ameliorating injustice in the worlds of clinical care and clinical research seems imperative. As noted, while we know African-Americans receive disproportionately fewer kidney transplants than whites with end stage renal disease, it is poorly understood how much a lack of timely access to primary care, sub-specialty (nephrology) care, transplant center evaluation, or other factors contribute to the low transplantation rate. Without such knowledge, one cannot recommend ethically optimal interventions that can address the inequities.
AREAS OF CONFLICTING OR UNCLEAR EVIDENCE

Assessment of Decision-making Capacity

As reliance on the doctrine of informed consent depends on the ability of patients or subjects to understand and make use of information about proposed clinical or research interventions, many researchers and clinicians have sought accurate, reproducible methods to determine the adequacy of a patient’s or subject’s decision-making capacity. Putting aside efforts to decide if persons in the criminal justice system have sufficient capacity to stand trial, the search for an assessment instrument or process to adequately assess decisional capacity has had mixed success. In part, the problem seems to stem from the different questions researchers or clinicians believe ought to be asked. For example, does the individual have sufficient capacity to consent to medical care, to participate in research, to return to independent living, to make financial decisions, etc.? According to one review (Tunzi, 2001), the well known MacArthur Competence Assessment Tool (Grisso & Appelbaum, 1998) applies best to persons with known psychiatric or neurological disorders while others, such as the Capacity Assessment Tool (Carney, Neugroshi, Morrison, Marin, & Siu, 2001) or the Aid to Capacity Evaluation (Etchells et al., 1999) work in a general patient population. As Baergen (2002) and Breden and Vollman (2004) comment, the instruments may over- or under-estimate patients’ understanding of their situation, focus mostly on performance of cognitive tasks, inadequately assess the importance of a person’s values and feelings, and minimize the complexity of decision making in actual medical situations, fraught, as they often are, with several levels of uncertainty.

In one particular population, that of minors, confusion and conflict reign regarding decision-making capacity. The work of ethnographers and others, most notably Bluebond-Langner (1978), show that the experience of chronic illness brings knowledge and decision-making ability well beyond one’s years for many children with serious medical conditions. On the other hand, focusing on the need for broad and effective public policy, social psychologists and lawyers (Scott, Reppucci, & Woolard, 1995) rely on research findings that adolescents typically and excessively (1) attend to the short-term consequences of decisions and actions; (2) tolerate risks; and (3) bow to pressure from others, especially adults in authority or peers. From this perspective, one should delay adolescent decisional authority as long as possible, hoping for the onset of maturity. It seems that empirical research
has failed to provide information as to how to determine medical decision-making capacity among adolescents under particular circumstances. Additional research could clarify how to balance an adolescent’s experience and maturity against population-based concerns about psychological and social development. When might an individual teenager, for example one who has lived with severe cystic fibrosis for years, despite relatively young age, say 14 years, have sufficient judgment to decline another round of mechanical ventilation in the ICU, with or without support from her parents? Targeted clinical studies could provide valuable practical data for agonizing ethical decisions regarding such thorny issues.

Protection of Human Subjects of Research

Since the mid-1970s with the introduction of federal regulations governing how institutions must review and oversee research involving human subjects, distressed and disgruntled investigators have often wondered about the extent to which bureaucratic processes actually protect subjects from research-related risk(s). As noted above, the common failure to produce intelligible consent forms suggests the current regulatory structure might not be effective. On the other hand, in the face of enormous increases in biomedical and behavioral research with human subjects since World War II, commentators, e.g. Emanuel (2005) and Fost (2005), point to the apparent low frequency of coercive abuses of human subjects in the era of oversight by IRBs. From this perspective, the rare dramatic and tragic deaths of research subjects prove the rule that the system works well, despite federal rebuke to IRBs at institutions where the deaths occurred.

Only a few studies have systematically assessed the effectiveness of IRBs. One early study that used sham protocols submitted to IRBs around the U.S. found considerable differences in the quality and level of detail reviews, not to mention willingness to approve or reject questionably acceptable studies (Goldman & Katz, 1982). Several more recent publications regarding differing kinds of research (critical care, genetics, health services, and adult surgery) have indicated continuing variability in IRB procedures and responses to federal regulations aimed at protecting human subjects (Silverman, Hull, & Sugarman, 2001; McWilliams et al., 2003; Dziak et al., 2005). At least one ongoing study (Lidz, 2006) should begin to shed needed light on how IRBs routinely do their work.

With regard to research including children, a classic vulnerable group, one might expect greater concern for subject protection and therefore greater
consistency in the application of rules and regulations. However, some studies do not bear out that expectation. A review of published human subjects research with children led to a survey of authors whose articles failed to indicate whether their research had had IRB review and/or used required informed consent procedures (Weil, Nelson, & Ross, 2002). The survey found significant misclassification by IRBs of studies felt to be “exempt” from IRB review. Another survey of IRB chairs who frequently assessed studies involving children noted wide variations in the definitions the IRBs used for interventions constitution “minimal” or “minor increase over minimal” risk (Shah, Whittle, Wilfond, Gensler, & Wendler, 2004). A recent study showed large differences between IRBs in how they required investigators to respond to federal regulations regarding child assent to participation in research (Whittle, Shah, Wilfond, Gensler, & Wendler, 2004), while a review of IRB websites revealed some “incorrect advice” to investigators about meeting regulatory requirements concerning children (Wolf, Zandecki, & Lo, 2005). These studies begin to point to the areas where interventional research designed to determine how best to help IRB committee and staff members understand ethical considerations and regulations for pediatric research would be useful.

There have also been studies of IRB members that suggest potential problems for human subjects protections in general. Campbell et al. (2003) found that of the nearly 3000 US medical school faculty members responding to their survey, 11% had served on IRBs. Almost half of these individuals (47%) consulted for industry, raising concerns about conflicts of interest in judging research protocols. Van Luijn, Musschenga, Keus, Robinson, and Aaronson (2002) found that “a substantial minority” of the 53 IRB members they interviewed in the Netherlands about reviews of phase II clinical oncology trials “felt less than fully competent at evaluating” key aspects of the research protocols.

Several US federal government bodies have attempted to produce overviews of the adequacy of human subjects protections in the last several years. In 2000, The Office of the Inspector General of the Department of Health and Human Services (2002) issued a report entitled “Protecting Human Research Subjects: Status of Recommendations.” Citing an earlier report from 1998 from the same office warning of problems in the system, the follow-up report indicated that few of its previously recommended reforms had been put in place. They noted little progress in “continuing protections” of research subjects beyond initial IRB reviews, inadequate educational requirements for investigators or IRB members on protecting research subjects generally and preventing or minimizing conflicts of interest, especially.
In 2001, the National Bioethics Advisory Commission (NBAC, 2001), appointed by President Clinton, issued its report *Ethical and Policy Issues in Research Involving Human Participants*. The introduction, entitled “The need for change,” identified challenges faced by the research oversight system, highlighting the enormous workload faced by IRBs at research-intensive institutions and the high financial stakes involved. NBAC noted inadequate protections for potential subjects from vulnerable populations, inconsistency and rigidity in federal regulations, weaknesses in enforcement mechanisms available to agencies overseeing human subjects research, and inadequate resources for IRBs (administratively) and for IRB members in terms of time and education about research ethics.

In 2001 and 2002, the Institute of Medicine (IOM, 2001, 2002) published two volumes concerned with protection of human subjects. These reports point to various problems in participant protection, including the fact that federal regulations do not necessarily apply to non-federally funded research, depending on arrangements at the institution where the research takes place. Even if that issue were resolved, the IOM studies acknowledge a host of additional problems, some of them well-documented, others simply feared or hypothesized, such as the extent to which potential subjects are exposed to “coercive” efforts to secure their enrollment in research (Emanuel, 2005).

In summary, sufficient and clear data are lacking about the adequacy of protections of human subjects to assist physician-investigators considering referrals of patients to clinical studies and subjects considering participation in research. Some would claim that the combination of federal regulations, local IRB oversight, investigator education, and public good-will provide at least adequate protection for human subjects of biomedical and behavioral research in the US and Western Europe. (The controversies about research in the developing world fall outside the scope of this review.) Alternatively, some believe that the host of demonstrated and feared inadequacies in the system, especially when one considers the financial stakes involved, suggest widespread disregard of subject protection. The skeptics point to the few well-publicized deaths of research subjects in the last decade and suggest that we have only learned about the tip of the proverbially iceberg that is a poorly regulated and possibly corrupt system. We would all benefit from much more detailed studies of actual IRB function, including field observations of research team members as they interact with potential and actual subjects, and studies that monitor or audit compliance of institutions with existing rules for the conduct of human subjects research. For example, the latter research might attempt to generate generalizable results about
subject understanding of risks and benefits, the completeness of research records, including properly executed consent forms, and the appropriate notification of subjects when new information becomes available that might affect their willingness to continue in a study.

**IMPLICATIONS FOR MEDICAL PRACTICE**

Empirical research in the area of bioethics has helped clarify “best practices” in many areas. There is compelling evidence that the theory and practice of informed consent have failed to live up to expectations in both clinical and research arenas. Likewise, advance directives have provided insufficient guidelines to most patients and many clinicians for treatment decisions when individuals lose decision-making capacity. Reliance on directives available to clinicians has proved frustrating because the instructions frequently do not cover all possible circumstances that patients, surrogates, and clinicians may face. As a result, health care attorneys, hospital administrators, clinicians, and ethicists now tend to recommend that patients both clearly designate a proxy decision maker and engage in detailed discussions with the appointed surrogate regarding the values that should guide decisions when he/she no longer has decision-making capacity. Empirical studies have clearly shown that ethnic, economic, and gender disparities persist in health care and clinical research despite increases in civil liberties and social justice in the last century. Unfortunately, studies have not yet pointed to ways to reduce the inequities. Clinicians need to maintain a high level of awareness about the potential influences of their unconscious biases on treatment and research recommendations, emergency decision making, and interactions with patients and family members. Institutions may need to develop systems to monitor for inequitable patterns of care and, if discovered, system-wide methods to correct imbalances. Of course, to the extent that patterns reflect larger social problems regarding risk for disease and disability and inadequate health care insurance coverage and payment schemes, providers may face serious financial problems if they undertake to redress inequality on their own. Empirical evidence in other areas of bioethics still remains scant or presents an unclear picture. For example, with regard to the ability to assess the adequacy of patient or (potential) research subject decisional capacity, research results are somewhat mixed. Similarly, physician-investigators trying to decide whether sufficient protections exist for patients who might become research subjects – not least those who are part of vulnerable populations – will find a
confusing array of reassurance, scandal, and troublesome study results. Other bioethical issues that need research attention in efforts to optimize ethical standards and quality in medical care and research include a better understanding of the potential benefits and problems of clinical ethics consultation, the consequences – intended and unintended – of universal health care insurance and rationing schemes, and methods to reduce the administrative burdens on investigators and institutions of ethics reviews of human, animal, and embryonic stem cell research.

In conclusion, the above review and discussion indicate that there is fertile ground for more work at the intersection of bioethics and empirical research that can guide practitioners, investigators, patients, and their families in the quest to continue to advance medical practice and research consistent with ethical principles.

REFERENCES


In re Conroy, 486 A.2d 1209 (N.J. 1985).


SECTION II: QUALITATIVE METHODS
This page intentionally left blank
QUALITATIVE CONTENT ANALYSIS

Jane Forman and Laura Damschroder

INTRODUCTION

Content analysis is a family of systematic, rule-guided techniques used to analyze the informational contents of textual data (Mayring, 2000). It is used frequently in nursing research, and is rapidly becoming more prominent in the medical and bioethics literature. There are several types of content analysis including quantitative and qualitative methods all sharing the central feature of systematically categorizing textual data in order to make sense of it (Miles & Huberman, 1994). They differ, however, in the ways they generate categories and apply them to the data, and how they analyze the resulting data. In this chapter, we describe a type of qualitative content analysis in which categories are largely derived from the data, applied to the data through close reading, and analyzed solely qualitatively. The generation and application of categories that we describe can also be used in studies that include quantitative analysis.

QUANTITATIVE VERSUS QUALITATIVE CONTENT ANALYSIS

In quantitative content analysis, data are categorized using predetermined categories that are generated from a source other than the data to be
analyzed, applied automatically through an algorithmic search process (rather than through reading the data), and analyzed solely quantitatively (Morgan, 1993). The categorized data become largely decontextualized. For example, a researcher who wants to compare usage by physicians, patients, and family members of words such as die, dying, or death versus euphemisms such as pass away or demise, would make a list of words, use a computer to search for them in relevant documents (e.g., audio-recordings of oncology outpatient visits), and compare usage in each group using statistical measures (Hsieh & Shannon, 2005).

In qualitative content analysis, data are categorized using categories that are generated, at least in part, inductively (i.e., derived from the data), and in most cases applied to the data through close reading (Morgan, 1993). There is disagreement in the literature on the precise definition of qualitative content analysis; these differences are about how the data is analyzed once it has been sorted into categories. For some authors, qualitative content analysis always entails counting words or categories (or analyzing them statistically if there is sufficient sample size) to detect patterns in the data, then analyzing those patterns to understand what they mean (Morgan, 1993; Sandelowski, 2000). For example, in one study, researchers used qualitative data from semi-structured interviews to identify, count, and compare respondents who personalized the task they were asked to do versus those who did not (Damschroder, Roberts, Goldstein, Miklosovic, & Ubel, 2005). This study derived categorical data from the qualitative data in order to quantitatively analyze differences in the types of responses people gave to the different types of elicitations. Qualitative content analysis is defined more broadly by some researchers to also include techniques in which the data are analyzed solely qualitatively, without the use of counting or statistical techniques (Hsieh & Shannon, 2005; Mayring, 2000; Patton, 2002).

USES OF QUALITATIVE CONTENT ANALYSIS

Qualitative content analysis is one of many qualitative methods used to analyze textual data. It is a generic form of data analysis in that it is comprised of an atheoretical set of techniques which can be used in any qualitative inquiry in which the informational content of the data is relevant. Qualitative content analysis stands in contrast to methods that, rather than focusing on the informational content of the data, bring to bear theoretical perspectives. For example, narrative analysis uses a hermeneutical perspective that emphasizes interpretation and context, and focuses
“on the tellings [of stories] themselves and the devices individuals use to make meaning in stories” (Sandelowski, 1991, p. 162). There are some methods of qualitative inquiry (e.g., ethnography, grounded theory, and some types of phenomenology) that, though they bring a theoretical perspective to qualitative inquiry, use content analysis as a data analysis technique. Grounded theory, which has been used extensively in nursing research, uses a specific form of content analysis whose goal is to generate theory that is grounded in the data. Because the term “grounded theory” is often used generically to describe the technique of inductive analysis, it is often confused with a form of inquiry called qualitative description (Sandelowski, 2000) or pragmatism (Patton, 2002). The goal of these approaches to data analysis is to answer questions of practice and policy in everyday terms, rather than to generate theory. What we describe in this chapter is a generic form of content analysis commonly used in the health sciences to answer practical questions. The validity of an atheoretical approach is a source of controversy among qualitative researchers, but is widely accepted by researchers in the health sciences.

As compared to quantitative inquiry, the goal of all qualitative inquiry is to understand a phenomenon, rather than to make generalizations from the study sample to the population based on statistical inference. Examples include providing a comprehensive description of a phenomenon; understanding processes (e.g., decision-making, delivery of health care services); capturing the views, motivations, and experiences of participants; and explaining the meaning they make of those experiences. When used as part of a study or series of studies using a combination of qualitative and quantitative methods, qualitative methods can be employed to explain the quantitative results and/or to generate items for a closed-ended survey.

Qualitative content analysis examines data that is the product of open-ended data collection techniques aimed at detail and depth, rather than measurement. For example, a closed-ended survey can be used to measure the level of trust patients have in their physicians. As an alternative, open-ended interviews in which participant responses are not constrained by closed-ended categories can be used in order to explore the topic of trust more deeply. While a closed-ended survey may provide an assessment of patient trust, it fails to provide any information on the process through which patients come to trust or distrust their physicians, and what trust means to them.

Empirical bioethics studies take advantage of the open-ended nature of qualitative research to, for example, examine and challenge bioethical assumptions, inform clinical practice, policy-making or theory, or describe
and evaluate ethics-related processes or programs. For example, in a study to understand how housebound elderly patients think about and approach future illness and the end of life, interviewees “described a world view that does not easily accommodate advance care planning” (Carrese, Mullaney, Faden, & Finucane, 2002, p. 127). In another study, female patients talked about their views on medical confidentiality to inform clinical practice around confidentiality protections (Jenkins, Merz, & Sankar, 2005). Findings showed that some patients “might have expectations not met by current practice nor anticipated by doctors” (p. 499). As a first step toward developing benchmarks of clinical ethics practices, another study described and compared the structure, activity, and resources of clinical ethics services at several institutions (Godkin et al., 2005). Results indicated a high degree of variability across services and that increasing visibility was a challenge within organizations.

**DOING QUALITATIVE CONTENT ANALYSIS**

In the rest of this chapter, we will discuss the choice of qualitative content analysis as one of many decisions in designing a study. We will review the processes and procedures involved in this type of analysis, including data management, memoing developing a coding scheme, coding the data, using coding categories to facilitate further analysis, and interpretation. We will also discuss the use of software developed to aid qualitative content analysis.

**STUDY DESIGN**

As we discussed, qualitative content analysis is one of many techniques for performing analysis of textual data. Although it is beyond the scope of this chapter to discuss study design in any detail, we will briefly illustrate the types of design decisions required for qualitative studies and their effect on data analysis.

As with all empirical research, study design starts and flows from the research question(s). Thoughtfully and deliberately matching data sources, sampling strategy, data collection methods, and data analysis techniques matched to the research questions is fundamental to the quality and success of any study. After formulating the research question(s), Mason suggests that the researcher ask the following questions when designing a qualitative
study: “What data sources and methods of data generation are potentially available or appropriate? What can these methods and sources feasibly describe or explain? How or on what basis does the researcher think they could do this? Which elements of the background (literature, theory, research) do they relate to?” (Mason, 2002, p. 27) Examples of data sources are individuals and groups (e.g., persons with kidney failure, family members), sites (e.g., ICU, primary care clinic, governmental agency), naturally occurring interactions (e.g., primary care visit, congressional hearing), documents and records (e.g., policies and procedures, medical records, legislation), correspondence, diaries, and audio-visual materials. Data collection techniques include individual interviews, focus groups, observations, and sampling of written text(s).

Once it is clear that a qualitative approach is appropriate, a number of related issues need to be addressed. First, an exploration of what is already known about the topic of study is necessary. The more that is known about the topic, the more structured or deductive the data collection and analysis are likely to be. This is because previous empirical and theoretical work will provide a conceptual framework consisting of concepts and models that direct data collection and analysis (Marshall & Rossman, 2006).

The study’s research question will point to a particular unit or units of analysis. Units of analysis can be individual people, groups, programs, organizations, communities, etc. The unit of analysis is the object about which the researcher wants to say something at the end of the study. There may be more than one unit of analysis in a study (Patton, 2002). For example, in a multi-site study of communication in ICUs, ICUs, health provider groups (e.g., physicians and nurses) and individuals (e.g., individual nurse points of view) could all be units of analysis.

Having identified an appropriate unit of analysis, the researcher must decide how best to sample it. Sampling in qualitative studies is what is called purposeful. In purposeful sampling, the goal is to understand a phenomenon, rather than to enable generalizations from study samples to populations. In-depth study of a particular phenomenon involves an intensive look at a relatively small sample, rather than a surface look at a large sample. “Information-rich” cases are selected for in-depth study to provide the information needed to answer research questions. It is important to choose those cases that will be of most use analytically (Patton, 2002; Sandelowski, 1995b).

Finally, the study’s conceptual framework, unit(s) of analysis, sampling, and data collection technique(s) will affect the data analysis performed: the
conceptual framework will influence the categories used to code the data, as well as how deductive or inductive the analysis will be; the unit(s) of analysis will determine the entities around which the analysis is organized; the sampling strategy may create subgroups that can be compared; and the data collection technique selected will produce data of varying degrees of depth. On this last point, the more in-depth the data (e.g., a few extended in-depth interviews), the more challenging and time-intensive it is to analyze. Also, because more in-depth data provides more information about what participants mean by their statements, and may indicate causal linkages emerging from the data, it provides more evidence to support a higher inference analysis than data that focuses on breadth (e.g., a set of focus groups). By higher inference, we mean that the researcher can interpret the data at a higher level of abstraction. For example, in a focus group study, participants may name the quality of communication with various members of the health care team as a barrier to appropriate care in the ICU. Data collected through in-depth interviews may contain enough evidence to characterize types of communication barriers with more abstract concepts, such as manifest versus latent communication.

It is helpful to consider several additional points when designing a study. First, having several people analyzing the data demands a more structured approach. Second, resource constraints (e.g., time, money, personnel) force trade-offs between the richness of the data, the amount of data collected, and the quality of analysis. The researcher must make all these decisions must be made with the purpose of the study in mind by considering the resources available and the optimal way to expend the resources to obtain the study’s goals.

Finally, there is only so much that can be learned from hearing, reading, and thinking about content analysis. In our experience, knowing about qualitative methods is not the same as having experience using them. We suggest that when embarking on a qualitative study, the novice seek a mentor to learn how to use methods and techniques effectively. The most effective learning comes from being actively engaged in a project. Qualitative research requires a somewhat more hands-on approach than quantitative techniques.

**DATA MANAGEMENT**

As in any study, before starting data collection, it is important to develop a system for labeling the data (e.g., participant, site, etc.). Qualitative data can
take on more complex and varied forms than quantitative data that depend on its sources, for example, interviews, direct observation, complicated textual sources, and video-recordings. Therefore, the data sources must be clearly labeled in preparation for coding and electronic entry. For instance, audio-recordings of interviews should be labeled with the participant id number, the date, and the interviewer’s initials. After an interview, it is important to make a copy of the audio-recording so that there is a backup in case the original is lost or destroyed. In most qualitative content analyses, the next step is to transcribe the recording so that it appears as a written text. However, it is essential to recognize that any transcription of spoken words will be incomplete. Qualitative content analysts are interested in the informational content of the data. Therefore, the transcription focuses on a verbatim representation of the words on the audio-recording and may also include some indicators of emotion (e.g., exclamation point, notation of laughter). In contrast, transcription for analysis of discourse has highly detailed specifications that may include, for example, representation of interruptions, pauses, intonation, and simultaneous talk.

It is important to use clear rules for how recordings are to be transcribed, whether the transcription is done by the researcher or delegated to a transcriptionist. Having these rules avoids unnecessary work later on, and makes it easier to read and analyze multiple transcripts. Table 1 shows an example of rules that can be included. This is not meant to be an exhaustive list; items and rules for any specific project will vary.

### Table 1. Example Rules for Transcription.

*The finished product should be in word (no tables) – Version x; with the following specifications:*

- Times new roman 12 point font
- Create a separate file for each tape. The filename must include the tape, and site identifier (e.g. TapeB-SiteX.doc) and insert in the first line of the file
- Insert page numbers in footer
- Header 1: Person ID; place on a separate line from the text
- 1.5" margins
- Note interruptions and inaudible conversation by inserting [INTERRUPTION], [INAUDIBLE] in the text. (Only indicate [INAUDIBLE] when words cannot be distinguished because of the recording)
- Indicate significant deviations from normal conversational tone as [LAUGHTER], [ANGRY], [LOUD] after text
- Insert tape counter position periodically through the document
- If you are unsure of a word, put a [?] after the word(s)
After a transcript is completed, it will need to be compared to the recording to make sure it is accurate. It is best if the interviewer can verify the accuracy of the transcript because the interviewer is more likely to recall information that may be difficult for the transcriptionist to hear due to inaudible speech. If the interviewer is also the transcriptionist, additional editorial comments about the interpretation of what occurred during the interview at specific times can be notated. If editorial comments are added, it is vital to clearly label them as such so that they are not mistaken for raw data.

The final step in preparing the transcript for analysis is stripping it of information that identifies participants, such as names and places. It is important to have clear rules for how identifying information will be replaced. The aim is to include enough information about a particular word being replaced so that the informational contents are not lost.

**DATA IMMERSION, REDUCTION, AND INTERPRETATION**

Content analysis requires considerably more than just reading to see what’s there. (Patton, 2002, p. 5)

In qualitative content analysis, data collection and analysis should occur concurrently. One danger in qualitative research is the collection of large amounts of data with no clear way to manage or analyze it. To maximize the chances of success, one should “engage with” the data early on by starting to develop a coding scheme. By examining the data as it is collected, the researcher will become familiar with its informational content, and may identify new topics to be explored and develop analytic hunches and connections that can be tested as analysis progresses. These insights also inform subsequent data collection. For example, a question might be added to the interview guide to explore a new topic with a subsequent interviewee. Although this chapter presents content analysis as a series of sequential steps, it is important to note that it is an inherently iterative process.

One useful approach is to divide content analysis into three phases: immersion, reduction, and interpretation (Coffey & Atkinson, 1996; Miles & Huberman, 1994; Sandelowski, 1995a). Through each of these phases, the goal is to create new knowledge from raw, unordered data. Content analysis requires both looking at each case (e.g., participant, site, etc.) as a whole
and breaking up and reorganizing the data to examine individual cases systematically, and compare and contrast data across cases.

Immersion: Engagement with the Data

During immersion, the researcher engages with the data and obtains a sense of the whole before rearranging it into discrete units for analysis. There are several ways to accomplish this. First, the researcher can write what is called a “comment sheet” immediately after the data collection activity to record first impressions, new topics to be added in future data collection, comparisons to data collected previously (if this is not the first data collection activity), and analytic hunches. Second, the audio-recording is listened to. Third, when transcripts are available they can be read several times. The last process that can be employed is to write thoughts that are triggered while listening to and/or reading the data. This form of “free association” is often associated with meaningful insights that can be tested later on. While the science of qualitative content analysis – the process of developing and implementing a systematic approach to data analysis – is vital, much of the art of content analysis takes place when the analyst makes connections that occur once the data are considered holistically.

An essential part of the immersion phase is referred to as “memoing.” Memos are documents written by the researcher as he/she proceeds through the inspection of the data and can contain just about anything that will help make sense of it. Memos serve as a way to get the researcher engaged in the data by recording early thoughts and hunches. They also serve to initiate the data analysis by identifying and sharpening categories and themes (core topics or meanings) that begin to emerge. This diminishes the potential for losing ideas and thoughts in the process. Throughout the analysis, memos can describe themes and the connections among them that are developed through interactive inspection of the data. Memos serve as an audit trail of the researchers’ analytic processes and add credibility to the final analysis and conclusions. Memos can be coded along with the raw text data (e.g., transcripts) so that their contents appear in the reorganized data along with the raw data. The researcher can then read the raw data in a particular category along with the relevant category descriptions and analytic hunches recorded in memos during the immersion, preliminary coding, and code development phases of analysis. We describe the process of developing a coding scheme and coding the data in the next section.
Reduction: Developing a Consistent Approach to the Data

One of the most paralyzing moments in conducting qualitative research is beginning analysis, when researchers must first look at their data in order to see what they should look for in their data. (Sandelowski, 1995a, p. 371)

The reduction phase is when the researcher develops a systematic approach to the data. It constitutes the heart of the content analysis process and supplies rigor to the process. The goals of the reduction phase are to: (1) reduce the amount of raw data to that which is relevant to answering the research question(s); (2) break the data (both transcripts and memos) into more manageable themes and thematic segments; and (3) reorganize the data into categories in a way that addresses the research question(s).

Codes: What are They and Where do They Come from?

Codes provide the classification system for the analysis of qualitative data. Codes can represent topics, concepts, or categories of events, processes, attitudes or beliefs that represent human activity, and thought. Codes are used by the researcher to reorganize data in a way that facilitates interpretation and enables the researcher to organize and retrieve data by categories that are analytically useful to the study, thereby aiding interpretation. The thoughtful and deliberative development of codes provide rigor to the analytic process. Codes create a means by which to exhaustively identify and retrieve data out of a data set as well as enable the researcher to see a picture of the data that is not easily discernable in transcript form. As Coffey and Atkinson state, “attaching codes to data and generating concepts have important functions in enabling us rigorously to review what our data are saying” (Coffey & Atkinson, 1996, p. 27).

Codes can be either deductive or inductive. Deductive codes exist a priori and are identified or constructed from theoretical frameworks, relevant empirical work, research questions, data collection categories (e.g., interview questions or observation categories), or the unit of analysis (e.g., gender, rural versus urban, etc.). Inductive codes come from the data itself: analytical insights that emerge during immersion in the data and during what is called “preliminary coding” (see below). Although there are studies that use codes developed either deductively or inductively, content analysts most often employ a combination of both approaches. This means using a priori deductive codes as a way to “get into” the data and an inductive approach to identify new codes and to refine or even eliminate a priori codes.
Developing the Coding Scheme

We now turn to a description of how to develop a coding scheme and codebook. Coding the data allows the researcher to rearrange the data into analytically meaningful categories. Code definitions must be mutually exclusive; that is, they must have definitions that do not overlap in meaning. When coding categories are created it is important to consider how the coded data will look once retrieved, and how the rearranged data facilitate addressing the research question(s) during the next phase of the analysis.

Codebook development is an iterative process, and begins with what is called “preliminary coding.” This consists of reading through the text, highlighting or underlining passages that may be potentially important and relevant to the research questions, and writing notes in the margins. As noted above, because code development most often is based on inductive and deductive reasoning, it often starts with deductively developed codes but remains open to new topics suggested by the data (inductive codes).

To illustrate preliminary coding and subsequent codebook development (and, later in the chapter, the application of codes and data interpretation), we will use a short interview extract from the following hypothetical empirical bioethics study. The study explores the development of interpersonal trust in physicians from the point of view of patients with kidney disease who are at risk of losing enough function to need dialysis or a kidney transplant. Using semi-structured interviews, it seeks to understand which physician behaviors and qualities are important in the formation and maintenance of patient trust. This study conceptualizes trust as a phenomenon that arises from the patient’s experience of illness. Phenomenologists have shown that illness disrupts patients’ previously taken-for-granted ways of being in the world (Zaner, 1991). Patients’ physical, emotional, and existential need arise from these disruptions and attendant feelings of distress. Trust is a central moral issue in the physician–patient relationship because patients, in their vulnerable state, need to trust that physicians use their expertise and power in their patients’ best interest.

Table 2 shows an extract from an interview with a patient who has kidney disease and who is concerned about the progression of her disease. Based on the literature, which shows that physicians who more freely share complete information with patients engender more trust (Keating, Gandhi, Orav, Bates, & Ayanian, 2004), the researcher would begin with a deductive code, for example “clinical information.” The researcher would read the transcript, looking for statements related to “clinical information,” and highlight the text starting at line 4 in the transcript, in which the participant...
Table 2. Extract of Transcript from an Interview with a Patient with Kidney Disease Shortly after Visiting her Nephrologist.

I: How has your kidney disease been lately?

P: Well, Dr. [name of nephrologist] said that I’m pretty stable, but that eventually I’ll need to make a decision about what kind of treatment I want. You know, the thing I like most about him is that he always explains things really well. Like what my labs are compared to last time. And he tells me in regular language, so I can understand. Not like some doctors, who, um, well, you might as well be a number, for all they care. Like one doctor I went to, he barely looked at me, much less answered my questions. Dr. [name of nephrologist] is just the opposite. When he comes into the room, he looks you in the eye.

I: When you say that Dr. [name of nephrologist] explains things really well, what do you mean?

P: Well, you know, it’s scary having this disease. Like when I get a pain in my back, I’m thinking that my kidney is deteriorating or there’s too much poison in my blood or I have an infection or something. That actually happened today. I told Dr. [name of nephrologist] I was having these pains, and he examined me and looked at my lab tests and said that it was actually muscle pain; it didn’t have anything to do with my kidneys, just normal stuff. He said that if I had an infection, I’d have a fever and a real different kind of pain, described the difference. He also said that my kidneys getting worse wouldn’t cause pain like that. So I feel relieved. I mean I thought last week maybe I should come in and have him take a look at me, but I wasn’t sure, so I didn’t. So I just spent the week worrying. You know, I’m afraid I’m going to have to start dialysis or get put on the transplant list.

I: What has Dr. [name of nephrologist] told you about that?

P: Well, he was real clear about what I should expect. This sure isn’t going to get better, but I could stay the same for a while, or I could get worse more quickly. So he doesn’t know exactly what’s going to happen, but that’s ok, as long as he’s honest about it. It just helps to know. If I do get bad enough, I’ll have to start dialysis, and decide what to do about a transplant. He told me stuff about how that works, getting on the list and what I’ll need to consider. That puts my mind at ease; I know I won’t be hit with all this stuff that I don’t know about when the time comes.

says, “…eventually, I’ll need to make a decision about what kind of treatment I want … the thing I like most about him is that he always explains things really well.” The researcher would also make a notation in the margin about the concept “explaining,” and about how the patient may use the information, namely to make a treatment decision. After reading through several transcripts, the researcher may find descriptions of different uses of clinical, information and a deductive code called “clinical information” may evolve to “uses of clinical information.”

The researcher will also want to use inductive reasoning to develop new codes, specifically in vivo codes, which reflect the way informants make sense
of their world. For example, the text beginning at line 6, “… he tells me in regular language, so I can understand,” can be highlighted with a note about the patient’s desire for information from the physician in “regular language.” In the course of codebook development this statement may be used to support a code developed inductively, called “physician communication of clinical information.” Second, a preliminary code called, “relief from worry” might be created based on lines 18–26 in the transcript and other passages in this transcript and other transcripts. After reading through several transcripts, the code may evolve into a name that defines the concept more broadly, and frames it in terms of use of information, such as “reassurance.”

A codebook must be developed to organize codes and to help ensure they are used reliably. A codebook is especially important for projects using multiple coders. Table 3 shows an extract of the codebook from our hypothetical study, and contains a partial list of codes, definitions, and example quotes for each code. The example shows the fundamental elements that a codebook should contain: (1) name of the element; (2) an abbreviated label for that code (e.g., [REASSURE]); (3) the node type; (4) a description of the code that includes a clear definition, often with inclusion and exclusion criteria; and (5) example quotes that further illustrate the correct use of that code, along with a notation of the transcript and line numbers where the quote is located in the data set. The node type refers to the hierarchical position of that code in the coding framework. For example, uses of information [INFOUSES], a parent node, is a high-level category (code) that has four different types of uses: (1) what to expect on progression of the disease; (2) making a decision; (3) reassurance; and (4) monitoring symptoms. Fig. 1 shows how these codes relate to one another and help visualize how “parent” nodes relate to “child” nodes (see p. 43).

When working with a team, the code development process will proceed differently than for a solitary researcher. Team coding has its perils, but those are far outweighed by the benefits of having multiple perspectives to establish content validity and the ability to establish and test coding reliability. Multiple coders use much of the same procedures previously outlined in the chapter but their preliminary coding is done independently. The researchers come together to share their impressions of the data. It is important that all team members who might be involved in coding or later stages of the analysis (e.g., interpretation, writing manuscripts) be involved in these early meetings. The goal of this early development is not to review pages of transcripts but rather to engage in high-quality conceptualization through an iterative, negotiated process. Usually, to produce a revised
<table>
<thead>
<tr>
<th>Code</th>
<th>Node Type</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uses of clinical information (INFOUSES)</td>
<td>Parent</td>
<td>Discussion of the uses of information by the patient. EXCLUDE sub-codes</td>
<td>“… he was real clear about what I should expect. This sure isn’t going to get better, but I could stay the same for a while, or I could get worse more quickly. So he doesn’t know exactly what’s going to happen, but that’s ok, as long as he’s honest about it.” [30–33, 2035]</td>
</tr>
<tr>
<td>WHAT to expect on progression of the disease (EXPECT)</td>
<td>Child</td>
<td>Information about progression of the disease and what the patient can expect as it progresses, or the need for such information. Also, INCLUDE statements about the lack of information about disease progression and feeling in the dark about what to expect</td>
<td>“eventually I’ll need to make a decision about what kind of treatment I want” [4, 2035]</td>
</tr>
<tr>
<td>Making a decision (DECISION)</td>
<td>Child</td>
<td>Information that is useful to the patient to make decisions about medical treatment of the disease, or need for such information</td>
<td>“He also said that my kidneys getting worse wouldn’t cause pain like that. So I feel relieved.” [22–23, 2035]</td>
</tr>
<tr>
<td>Reassurance (REASSURE)</td>
<td>Child</td>
<td>Information to help relieve worry or fear, or the need for such information. Also, INCLUDE statements related to communication that causes worry or anxiety</td>
<td>“Like when I get a pain in my back, I’m thinking that my kidney is deteriorating or there’s too much poison in my blood or I have an infection or something.... He said that if I had an infection, I’d have a fever and a real different kind of pain, described the difference.” [15–22, 2035]</td>
</tr>
<tr>
<td>Monitoring symptoms (MONITOR)</td>
<td>Child</td>
<td>Information to help monitor symptoms, or the need for such information</td>
<td></td>
</tr>
<tr>
<td>Md communication of clinical information (INFOCOMMMD)</td>
<td>Parent</td>
<td>Descriptions of and/or judgments about the ability of the physician to communicate information in a way the patient can understand or apply</td>
<td>“he tells me in regular language, so I can understand.” [6–7, 2035]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“he barely looked at me, much less answered my questions.” [8, 2035]</td>
</tr>
</tbody>
</table>
(or initial) list of codes, the team will first review a few pages of one transcript. The team will then enter into an iterative process in which analysts apply the revised codes to a portion of the data and then meet again to add or delete codes, and further refine code definitions. After each meeting, decisions and definitions must be documented as codes are proposed, refined and finalized. This makes the process both transparent and systematic, thereby increasing the rigor of the analysis. The codebook is a good place to track changes over time by dating revisions made to code definitions.

Mason (2002) provides guidance as to the number of codes and precision of definitions to use when creating codes. Codes allow the researchers to index the data so that they can easily find salient text segments that relate to particular topics or concepts in the next stage of analysis. Discovering, even after coding, that it is difficult to find meaning in the data because the reorganized text is not focused sufficiently on analytically useful topics or concepts is an indication that the codes have been defined too broadly. Codes can also be defined too narrowly. When this occurs coders will have difficulty discerning which of the two closely defined codes should be applied. Too narrow coding can also obscure the ability to see larger patterns and themes. Study goals, resources and the amount of time available will dictate the depth and breadth of coding done and the extent to which multiple, independent coders can be used.

The framework for codes refers to the way the codes are arranged in relation to each other to form a conceptual map. It must be carefully designed in a way that best fits the data and that meets the goals of the study. Although each study must be approached individually, the development of 20–40 codes is the norm. As we saw in the example above, it is helpful conceptually to create coding “trees” in which there is a primary or parent code, with all related sub-codes or child codes listed under the parent (e.g., uses of clinical information and its children). Codes should parsimoniously categorize text and yet thoroughly cover the richness of information contained in that text. The framework chosen can make the

![Diagram Showing Relationship Between Types of Coding Nodes.](image-url)
difference between juggling hundreds of unrelated codes versus a fraction of that number of codes organized conceptually; the difference between creating a coding quagmire and providing a launching point for the next phase of analysis.

_Coding Agreement_

When code definitions have become substantially stable, and prior to applying the codes to the entire data set, coding agreement must be established. Agreement is when two or more coders who code text data independently, using the same codebook, can consistently apply the same codes to the same text segments. Although differences in how codes are applied are almost guaranteed to occur, regardless of how detailed codebook definitions may be, a sound conceptualization process, along with a well-constructed codebook with well-defined codes will help guide all coders to apply codes consistently. These constitute key methods in ensuring rigor in content analysis. When working in teams, the codebook is especially important for facilitating agreement because several different people may code different portions of the data. When using these methods it is common for solo researchers to assess agreement by having a second coder code a portion of the data and compare the results.

The issue of coding agreement exposes a basic tension between the positivist view that bias introduced by human involvement in research must be minimized to increase the validity of research results, and the constructivist view that validity is derived from community consensus, through the social process of negotiation (Lincoln & Guba, 2003; Sandelowski & Barroso, 2003). A tenet of qualitative research is that the researcher is the primary instrument of the research, and brings with her particular experiences, assumptions, and points of view that will affect interpretation of the data (Mason, 2002). Also, some coders may be more familiar with study aims or the data set than others. Thus, multiple coders mean multiple research instruments. Those with a positivist orientation label this as bias and aim to minimize it, while those with a constructivist orientation see it as an inherent feature of the interpretive process.

These differing orientations lead to two basic approaches as to how the agreement process should be structured in order to increase the validity of study findings. The first is measuring inter-coder agreement: using quantitative measures of agreement of the coding of two or more independent coders to establish coding reliability. Agreement is measured toward the end of coding scheme development; when it reaches a particular level, the codes are deemed reliable, and coding of the whole data set
proceeds. There are many ways to quantitatively measure agreement (Lombard, Snyder-Duch, & Bracken, 2002) and some qualitative researchers, following a positivist philosophy, believe that using quantitative measures are essential to establish reliability, especially when working in teams (Krippendorff, 2004; Neuendorf, 2002).

The second basic approach to the agreement process is using a consensus process in which two or more coders independently code the data, compare their coding, and discuss and resolve discrepancies when they arise, rather than measuring them. Qualitative researchers who follow a constructivist philosophy do not believe that quantitative measures of reliability are appropriate in content analysis, largely because of their view that unanimity among coders often leads to over-simplification that compromises validity, and that reflexivity and reason-giving are more important aspects of an agreement process than achieving a pre-specified level of agreement independently (Harris, Pryor, & Adams, 2006; Sandelowski & Barroso, 2003). Mason (2002) defines reflexivity as “thinking critically about what you are doing and why, confronting your own assumptions, and recognizing the extent to which your thoughts, actions and decisions shape how you research and what you see” (p. 5). A negotiated agreement process happens when coders meet to discuss the rationale they used to apply particular codes to the data. Through discussion, team members are able to explain their perspectives and justifications, how and why it differs from other team members’ perspectives, and reach consensus on how the data ultimately should be coded. It is important to understand the strengths and weaknesses of each approach and develop a process that best fits the study. The preferred approach will depend on study aims, the coding process used, the type of codes that are being applied (e.g., low versus high inference), the richness of the text being analyzed, the degree of interpretation required for the final product, and the targeted venue for publication and dissemination of the study results.

*Coding and Reorganizing the Data*

After a codebook is developed, and the codes can be used reliably or a consensus process is established, the codes can be applied to all the text in the data set. Once accomplished, the text is rearranged into code reports, which list all of the text to which each particular code has been applied.

When applying a code to a segment of text, the coder must be sure to include text that will provide sufficient context so that its meaning can be discerned. For example, the entire section spanning lines 15–26 in Table 2 could be coded as REASSURANCE to provide full context for how the
physician was able to “put [her] mind at ease.” To understand why this is necessary, imagine if only lines 22–23 were coded, “He also said that my kidneys getting worse wouldn’t cause pain like that. So I feel relieved.” When the text segment is read in a report that contains all of the text in the data set coded with REASSURANCE, instead of read in the context of a transcript, the reader loses information as to what led the patient to worry, including the connection to her experience of illness as “scary” because it could result in the need for dialysis or a kidney transplant.

Even after the codebook can be used reliably or a consensus process has been established, there may be changes in code definitions. New codes may be added as existing codes are applied to new text and as conceptualization progresses. It is a challenge to manage the tension between the desire for a predictable, sequential, and efficient process and allowing the process to be guided by intuitions, concepts, and theories arising from the data. Especially for large data sets, however, there is a point when the codebook, including code definitions, should be considered final, unless it is deemed critical to add a new code. Depending on resources, it may be possible to recode smaller data sets, say less than 20 transcripts, when new code definitions and codes arise.

Interpretation

Data to be interpreted include the code reports and memos that can contain anything from interpretive notes to preliminary conclusions, as mentioned earlier. These products need to be further analyzed, interpreted, and synthesized in order to formulate results. This phase of the analysis involves using the codes to help re-assemble data in ways that promote a coherent and revised understanding or explanation of it. Through this process the researcher can identify patterns, test preliminary conclusions, attach significance to particular results, and place them within an analytic framework (Sandelowski, 1995a). There is no clear line between data analysis and interpretation; ordering and interpretation of data occurs throughout the analysis process. However, by the interpretation phase, the groundwork has been laid to produce a finished product that communicates what the data mean. There are many ways to go about interpreting data, but almost all will include re-organizing it, writing descriptive and interpretive summaries, displaying key results, and drawing and verifying conclusions (Miles & Huberman, 1994).

To reorganize data in a way that facilitates interpretation, the researcher chooses to produce particular code reports, and organize these reports by
cases, i.e., subsets of the data that represent the unit(s) of analysis, for example, site or health provider type. (If, as in our example study, the unit of analysis is the individual, each participant counts as a case.) Code reports can represent all of the data in the data set coded with a single code (e.g., REASSURANCE), or a combination of codes (e.g., INFOCOMMMD and DECISION). Code reports and how they are summarized are determined by the research questions, what has been learned from the data analysis to date, and the specifics of what the researcher wants to examine. Code reports can enable case-by-case analysis or can help the researcher delve more deeply into a particular topic. The process is dynamic and iterative and guided by what is learned from the data, so a number of code reports may be produced.

After choosing and organizing the code reports, the researcher writes descriptive and interpretive summaries of the data contained in each report. The structure of these code report summaries will depend on the project, but usually includes the main points obtained from reading the report, quotations selected to provide evidence for those points, and an interpretive narrative at the code and/or case level. As discussed earlier, it is vital to draw a distinction between the raw data and the interpretation of the data. Summaries for each case should be grouped together so that each can be examined before making cross-case comparisons.

Data displays (e.g., matrices, models, charts, networks) can be helpful for exploring a single case, but are particularly helpful in looking across cases. Miles and Huberman (1994) define a display as “an organized, compressed assembly of information that permits conclusion drawing and action” (p.11). Seeing the data in a compressed form, organized in a systematic way, makes it easier to recognize patterns. It facilitates comparisons, which are important in drawing conclusions from the data. For example, a matrix with categories found to be analytically meaningful arrayed horizontally across the top and cases arrayed vertically can be created. These categories can be codes that were used to break up the data and/or themes which reflect a higher level of interpretive understanding that are developed as interpretation progresses. Each cell in the matrix is filled in with text, numbers, or ordinal group (e.g., text excerpts; main points; 1, 2, 3; high, medium, low) that summarize the characteristics of that category in each case. Fig. 2 is an extract from a data display matrix from our example study (an actual display would include more participants). It assumes that the researcher identified a new theme early in the interpretive phase: “Attributing physician motives,” defined as what motivations the participant attributes to their physician to account for the way the physician communicates with
The researcher creates a data display that arrays attributed motivations along with a brief summary of the contents of each of the codes listed in the example codebook. The researcher would then look at the display to discern patterns in the data and draw preliminary conclusions. Data displays are powerful tools and often are used throughout the interpretation process.

### Drawing Conclusions

Drawing and verifying conclusions involve developing preliminary conclusions and testing them by going back into the data. The researcher may develop the conclusion that one of the ways patients develop trust in their physicians is when physicians explain clinical information in a way that they can understand and that these behaviors denote physician competence to the patient. It is important to look for alternative themes and conclusions that

<table>
<thead>
<tr>
<th>ID</th>
<th>Attribution of MD motivation</th>
<th>INFOCOMM MD</th>
<th>EXPECT</th>
<th>DECISION</th>
<th>REASSURE</th>
<th>MONITOR</th>
</tr>
</thead>
</table>
| 101 | “doesn’t know what he’s doing”
“doesn’t think I’d understand” | Didn’t answer my questions | Wants more info from MD on potential progression to dialysis | Not mentioned | Worry re: acute symptoms (pain): Does it indicate disease progression? MD did not address satisfactorily when asked. | Wants to know whether should call the physician when she has pain or fever; are kidneys infected? |
| 102 | “a brilliant physician”
“cares about me” | Can understand what MD says (no jargon) Gives detailed info | Describes what MD told her about expected treatment progression; was detailed and included uncertainty | Needs to make decision about a transplant; MD gave useful info, including rationale for each treatment option | Was worried re: fatigue. MD told her it was typical and why it was occurring. | Not mentioned |

Fig. 2. Data Display Matrix.
may “fit” the data better throughout the interpretation process and not settle on premature analytic closure.

Conclusion verification is derived from going back into the data to find evidence that supports or refutes a particular conclusion. The result of verification can be finding that the conclusion holds in most cases, or is refuted, or that an alternative or refined conclusion is supported. If it does hold in most cases, it is not enough to report the theme and show supporting evidence for it. The researcher also must examine “negative cases” – cases for which the conclusion does not hold. In the example, patients for whom information is not useful and who instead judge physician competence and develop trust based on non-verbal and social cues is one such instance. In the final product, discussion of these “negative” cases as they relate to the conclusions adds credibility to findings by showing that the researcher searched for what made most sense rather than simply using data to support one conclusion (Patton, 2002). What about them or their situation is different and what does that say about the phenomenon under study? Finally, the act of writing a report or manuscript presenting study findings refines and clarifies the interpretation and should not be minimized as an important step in the interpretive process.

Using Software

Qualitative researchers are increasingly using software to manage data and facilitate data analysis and interpretation. Commonly used software includes ATLAS.ti (http://www.atlasti.com/index.php), MaxQDA (http://www.maxqda.com/), and NVivo (www.qsrinternational.com). It should be emphasized that software is a tool to help manage, retrieve, and connect data, but cannot perform data analysis. Too often, researchers unfamiliar with qualitative coding invest in a software purchase in the mistaken belief that the software will produce and code the data. Correctly used, and with the appropriate data set, such programs can enhance the efforts of the qualitative researcher.

Nearly any kind of data source (e.g., text, pictures, video clips) can be imported into the software and coded or linked using tools within that software. Some researchers will use software primarily to enter codes and rearrange their data into coding reports. Others will use it more comprehensively as they work through code development, coding, creating code reports, interpretation, and final manuscript writing.
The software allows researchers to link source documents with notes, memos, summaries and even theoretical models. For example, notes in the margin of a transcript can be created within NVivo by adding “annotations,” memos and summaries of code reports can be created and coded or linked to other documents in the data set. These kinds of software programs are especially helpful when working in teams because they facilitate sharing annotations, code reports, and summaries. For example, after the interview shown in Table 2 is coded, code reports can be generated to include text related to “uses of clinical information,” grouped by each of the sub-codes (progress, monitor, reassure, and decision). The software also allows one to do special queries, such as reporting all text coded with a designated union or intersection of codes. Analyses can be performed on subsets of transcripts (e.g., patient groups, sites) so that a variety of focused comparisons can be made.

**SUMMARY**

In this chapter, we have defined qualitative content analysis and discussed the choice of this method in qualitative research in the field of bioethics. As compared to quantitative inquiry, the major goal of qualitative inquiry is to understand a phenomenon, rather than to make generalizations from study samples to populations based on statistical inference. Qualitative content analysis is one of the many ways to analyze textual data, and focuses on reducing it into manageable segments through application of inductive and/or deductive codes, and reorganizing data to allow for the drawing and verification of conclusions (Miles & Huberman, 1994). The product of this process is an interpretation of the meaning of the data in a particular context. Qualitative content analysis that can be used by itself or in combination with other empirical methods can be employed to examine textual data derived from several sources and constitutes a versatile strategy to explore and understand complex bioethical phenomena.

**NOTE**

1. The data and analysis presented here are based on an unpublished project on “The Physician-patient Relationship in Function-threatening Illness” funded by the Niarchos Foundation, on which Dr. Forman was a co-investigator with Dr. Daniel Finkelstein and Dr. Ruth Faden.
ACKNOWLEDGMENT

The authors wish to thank Dr. Holly A. Taylor for her insightful comments on this chapter.

REFERENCES


ETHICAL DESIGN AND CONDUCT OF FOCUS GROUPS IN BIOETHICS RESEARCH

Christian M. Simon and Maghboeba Mosavel

ABSTRACT

Focus groups can provide a rich and meaningful context in which to explore diverse bioethics topics. They are particularly useful for describing people’s experiences of and/or attitudes toward specific ethical conundrums, but can also be used to identify ethics training needs among medical professionals, evaluate ethics programs and consent processes, and stimulate patient advocacy. This chapter discusses these and other applications of focus group methodology. It examines how to ethically and practically plan and recruit for, conduct, and analyze the results of focus groups. The place of focus groups among other qualitative research methods is also discussed.

INTRODUCTION

Focus groups are a versatile and useful tool for bioethical inquiry. Successful focus groups shed light on the diversity of views, opinions, and experiences of individuals and groups. Their group-based, participatory
nature is ideal for stimulating discussion of the kinds of multifaceted and contentious issues that bioethicists wrestle with daily. Apart from their role as a source of information on people’s perceptions of topical issues, focus groups can also be used to determine the ethics-related needs of health care institutions and professionals, to evaluate the effectiveness of interventions, and to generate rapport and trust among research subjects and communities. Fontana and Frey (MacDougall & Fudge, 2001) have summed up the advantages of focus groups as, “being inexpensive, data rich, flexible, stimulating to respondents, recall aiding and cumulative and elaborative” (p. 118).

This chapter explores some of the multiple characteristics and uses of focus groups with the purpose of highlighting their potential as a useful investigatory tool in bioethics. We consider a number of issues that we anticipate will be of special interest to bioethics researchers. These issues include the question of how one designs and conducts focus groups in an ethical, culturally appropriate, and scientifically rigorous way, and how researchers can use focus groups to stimulate critical reflection and generate new knowledge on key ethical and social issues. We draw on our experiences conducting focus groups in South Africa to illustrate both the challenges and rewards of using this methodology. Further information about using focus groups in empirical research can be found in a variety of sources, including our own work (Mosavel, Simon, Stade, & Buchbinder, 2005) and in several general guides to focus group methodology (Krueger & Casey, 2000; Morgan, 1997).

BACKGROUND

A focus group typically is composed of one or two moderators and six to ten individuals who may or may not share a common interest in the issue or topic under investigation. Focus group research originated in the 1930s among social scientists seeking a more open-ended and non-directive alternative to the one-on-one interview (Krueger & Casey, 2000). They later became popular in the marketing world as a tool for establishing consumer preferences for or opinions about different products, brands, and services. This commercial use of focus groups made some social scientists mistrustful of the methodology; however, focus groups are widely used and reported on today in the literature of the social sciences and other disciplines.

Focus groups are especially popular among researchers of health and patient care issues, in part because they are comparatively cost effective,
easy to implement, and less intimidating to some patients or individuals than interviews, questionnaires, or other methods of inquiry. In bioethics, focus groups have been used in a variety of ways, including exploration of public perceptions of the continued influence of the Tuskegee experiments on the disinclination of some groups from participating in biomedical research (Bates & Harris, 2004); how to improve end-of-life care (Ekblad, Marttila, & Emilsson, 2000; McGraw, Dobihal, Baggish, & Bradley, 2002); genetic testing and its medical, social, cultural, and other implications (Bates, 2005; Catz et al., 2005); environmental justice and environmental health issues (Savoie et al., 2005); and the appropriateness and effectiveness of medical informed consent procedures (Barata, Gucciardi, Ahmad, & Stewart, 2005). They have also been used as a tool for evaluating the effectiveness of medical ethics education and training initiatives (Goldie, Schwartz, & Morrison, 2000). In community-based health research, focus groups have been used to explore community health needs and concerns, build rapport and trust, and to empower community members to work toward constructive change (Mosavel et al., 2005; Clements-Nolle & Bachrach, 2003). Other uses of focus groups are possible and likely to emerge in the future.

THE FOCUS GROUP PROCESS

Some initial considerations: Empirical researchers often face the question of when it is appropriate to use focus groups, as opposed to other empirical methods such as individual interviews or surveys, in order to explore a particular issue, problem, or phenomenon. In making this decision, the researcher will want to bear in mind several factors. Focus group research typically involves far fewer research subjects than interview or survey research, where sample sizes tend to be significantly larger. This may make focus group research less costly and time consuming to conduct than individual interviews or surveys. However, the generally small sample sizes in focus group research also mean that it is harder to generalize findings to the larger group, community, or population from which the focus group participants were sampled.

For these reasons, among others, focus groups are often used to obtain preliminary or formative data that can be used to gain an initial impression of participant opinions and attitudes, and to inform the development of individual interviews, surveys, vignettes, or other instruments to be administered at a later date and with larger samples.
Focus groups are often used in combination with methods such as individual interviews, surveys, or other methods of inquiry. Findings from focus groups can help improve the design of a larger study, including the content, language, and sequence of items, and other essential elements of an interview or survey instrument. An example of this approach is provided by Fernandez et al. who conducted focus groups as a first step in determining the kinds of interview questions to ask pregnant women with regard to the ethical and other issues surrounding the collection, testing, and banking of cord blood stem cells (Fernandez, Gordon, Hof, Taweel, & Baylis, 2003). However, focus groups need not always be used as a first step in empirical exploration. In some cases, researchers have reversed the order of methods, for example, by using surveys to identify salient issues that would benefit from further, in-depth, exploration through focus group discussions (Weston et al., 2005).

Selecting Participants for Focus Groups

The proper sampling of research participants is essential for all types of empirical research. Different sampling techniques can be used in focus group research (MacDougall & Fudge, 2001); however, a primary objective of all these techniques is to bring together individuals who are generally representative of the larger group, community, or population that the research is interested in. One relatively simple way to achieve this representation is through intentional or purposive sampling, that is, by purposefully selecting specific individuals representative of the age, gender, racial and ethnic characteristics, professional training and skills, and other characteristics evident in the group or community of interest. This sampling approach has the advantage of being flexible and can evolve as the study develops (MacDougall & Fudge, 2001, p. 120). Researchers can draw on informal networks of colleagues, community organizations, advocacy groups, or other sources to help identify potential participants to invite to the focus groups.

A second approach to focus group sampling is to randomly select potential subjects. Random sampling typically has more scientific cachet than other approaches do; however, it does present unique challenges. For example, a focus group study aimed at better understanding how pediatric oncologists view the merits and problems of assent with children is unlikely to result in diverse and rich focus groups if participants simply have been randomly selected from, say, a pediatric hospital’s directory of oncologists.
The resulting sample is likely to be overwhelmingly English-speaking, male, and Caucasian. This level of homogeneity may be perfectly acceptable if the research question at hand is limited to exploring the attitudes and perceptions of individuals who share only these characteristics. However, if the attitudes of individuals of different genders and linguistic and ethnic backgrounds are also of interest to the research, stratified random sampling needs to be employed. In this case, the researcher aiming to explore the issue of assent will need to sample a cross section of pediatric oncologists, sorted into separate lists according to gender, ethnicity, among other possible characteristics. Thus, in addition to male Caucasian oncologists, the researcher may want to invite a number of female and minority oncologists to join his or her focus groups. Of course, if there happens to be only one or two female and/or minority oncologists at the institution the researcher has selected for study, stratified random sampling will not be possible. In the event that there is no diversity of gender or racial background among the oncologists at the institution, the researcher may want to convene a focus group at a second, more demographically diverse institution, or at a number of institutions. Alternatively, the sample can be broadened beyond physicians to include nurses or nurse practitioners, residents, and other oncology staff. However, the decision to take this step would depend on the particular research question at hand.

A stratified random sample is therefore one way in which focus group research can be made more representative and rigorous. A variety of sources have discussed sampling procedures for focus groups in more detail. Interested readers are referred to, among other sources, MacDougall and Fudge (2001) on the purposive approach to sampling and Krueger and Casey (2000) on random sampling.

Preparing for Focus Groups

Focus groups typically involve a series of questions that the moderator poses in order to generate discussion and get feedback on a particular topic. Depending on the length of the focus group (typically between 60 and 90 min), between 8 and 10 core questions are usually posed. Focus group questions need to be carefully developed so that they address the research question(s) at hand, are relevant, comprehensible and interesting to participants, and can be covered in the time allotted. They also need to be appropriately sequenced, for example, by asking questions of a more complex or controversial nature after participants have had a chance to
develop confidence and rapport among themselves, and with the moderator. Questions also need to flow logically and be guided by participants’ particular responses and by the moderator’s comments as facilitator.

Researchers may need to consider the sensitive and complex nature of many of their bioethics-related research topics when deciding on the questions they want to pose in a focus group. Inquiring about people’s views on stem cell or genetics research, informed consent, end-of-life care, and other similar topics has the potential to intimidate focus group participants due to being politically and morally loaded. Similarly, questions that require conceptual or technical knowledge may be beyond participants’ level of knowledge or experience. One way of avoiding the dreaded silences that questions about such issues can introduce into a focus group is to first consult with group or community leaders on how best to approach the issue at hand. In fact, this community consulting process should be considered for a range of reasons, including to facilitate access to potential focus group participants, to develop linguistically and culturally appropriate questions, and to enrich the analysis of focus group data (see description later). By consulting with key stakeholders, researchers can quickly establish what kinds of topics and questions will likely be viewed as acceptable and engaging, or which ones ought better be avoided. Different levels of community or group engagement can be sought. For example, a researcher may want simply to submit a list of focus group questions to selected community members or leaders to gain their feedback on how comprehensible, engaging, and appropriate the questions are, or, he or she may involve the community from the beginning in the formulation of the research questions so that they are reflective of the interests and concerns of the wider community in which the research is taking place. Cost, logistics, and other considerations will likely determine which of these options the researcher can feasibly take. Regardless, focus groups are far more likely to attract a good turnout, lively discussion, and rich data if preceded by efforts to engage the target group or community in developing and validating the questions to be asked.

Informed Consent for Focus Groups

The consent process for potential participants of focus groups should be responsive to all the elements of good informed consent, plus a number of additional considerations. The group-based nature of focus groups makes it harder to ensure confidentiality when compared to individual interviews or
surveys. Focus group participants may disclose personal and sensitive information about themselves to the moderators, the researchers, and other focus group participants. Some researchers attempt to discourage this kind of disclosure by asking participants to comment generally on the topic and not to share personal information about themselves. This may be partly effective, however, the interactive and intimate nature of focus groups can get the better of participants and prompt them to share sensitive personal information before the moderator can intervene and stop them. While the researcher can take steps to keep focus group recordings and transcripts containing personal information confidential, he or she has little to no control over whether or not the information will be more widely shared by group participants once the focus group is over.

Researchers can take a number of steps to help reduce, if not eliminate, concerns about the confidentiality and privacy of information that is shared in focus groups. One such step is to ask participants to use only their first names while engaged in focus group discussions. This will help protect participants’ identities if they are not already known to one another. Another step is to balance the need for diversity in focus groups against the need for confidentiality and respect among their participants, which may mean not including in the same focus group individuals who may be antagonistic toward one another on ideological, religious, or other grounds. However, such focus groups are unlikely to stimulate constructive discussion or yield important data. Finally, focus group moderators can help dispel some anxiety about confidentiality. This should not involve a formal review of the confidentiality statement that would have been included in the consent document for the research study, but a brief verbal reminder that what is said during the focus group must remain in the group. The usual practice is to provide this reminder before the focus group discussion begins and once again when it ends.

Selecting and Training Moderators

Selecting and training effective moderators are critical steps in the successful conduct of focus groups. A poorly selected or trained focus group moderator will not be able to promote optimal interaction among participants, keep their discussion from straying, and ask pertinent follow-up questions. Many focus groups use two moderators. This has the advantage of allowing one moderator to concentrate fully on introducing the topic, asking questions, and stimulating discussion while the other keeps
track of time, operates the audio recorder, and helps in asking follow-up questions. Obviously, if two moderators are used, they need to clearly understand their own as well as one another’s roles and responsibilities, and spend time training together.

Training is particularly critical for first-time moderators, and should focus, among other things, on the development of appropriate responses to classic focus group problems. For example, training could involve a few participant-actors who respond to the moderator’s questions in any number of predetermined and realistic ways. In actual focus group encounters, for example, it is not unusual for some participants to gravitate toward dominating discussions, while others grow increasingly passive. Training sessions that simulate this dynamic can be used to help moderators identify and negotiate this potential problem. Actors and moderators can debrief afterwards to discuss what sorts of moderator-initiated interventions work and do not work in the effort to address issues of passivity and dominance, among others.

It is also important to consider how well moderators are matched to the focus group participants they will be interacting with. Moderators who are matched to focus group participants in terms of age, gender, social and ethnic background, dress code, and so forth will engender greater rapport and openness. It may be appropriate for the researcher him- or herself to facilitate the focus groups, for example, if the focus groups include researcher colleagues who may expect a certain level of sophistication from their interactions with their moderator. On the other hand, some focus group participants may feel intimidated if the researcher serves as moderator. This may be the case particularly if the focus groups are led by the researcher–moderator is older or more experienced than most of the participants. These and other potential advantages and drawbacks should be carefully considered before the researcher decides whether or not to serve as a moderator.

Where to Conduct Focus Groups and How

Focus groups, particularly if they are addressing sensitive or controversial issues, will need to be conducted in as private and comfortable and accessible facilities as possible. Such settings may be difficult to secure given that many researchers may want or need to conduct their focus groups in hospitals, clinics, medical schools, and other health-related facilities. Space constraints, noise, unplanned interruptions, and other typical features of
many medical and clinical settings need to be negotiated as a result. Furthermore, if participants are patients and/or family members, they may not feel comfortable openly sharing their opinions in a medical or clinical setting, even if their health care providers are not immediately present. Health care facilities may also be hard for participants to access in terms of location, parking, or finding the room where the focus group is being held. Alternative locations for focus groups might include local libraries, community centers, or other public facilities that can offer quiet and comfortable environments.

The focus groups themselves should be flexible, exploratory, and not overly controlled. At the same time, too little structure or direction in a focus group can result in a lack of focus, confusion, argument, and, ultimately, highly disconnected data. Hence, focus group researchers have used a variety of strategies in an effort to balance the need for flexibility and informality against the need for direction and focus. These strategies include the use of question-and-answer formats, discussion guides, vignettes, “show cards,” video or audiotape, and Internet. For example, in a study designed to identify the key issues associated with the use of human-genes in other organisms, the Bioethics Council in New Zealand used show cards depicting various scientific claims associated with gene research to stimulate discussion on the topic (retrieved October 6, 2005 from http://www.bioethics.org.nz/publications/human-genes). The well-established “case study” in bioethics also potentially lends itself well to stimulating focus group discussion on any number of ethics topics.

These and other techniques can be used in combination with a question-asking approach. A well-developed series of questions has the capacity to generate lively discussion and valuable feedback on the topic of interest to the researcher. Many sources caution against the temptation to ask focus group participants too many questions; typically, participants in an hour-long focus group should not be asked to consider more than five core or primary questions.

Moderators should promote mutual respect among focus group participants by, for example, stating at strategic moments throughout the focus group that, “there are no right or wrong answers.” Focus groups dealing with sensitive topics can also be introduced through an “icebreaker question.” Here, again, prior consultation with community or group members can be helpful. For example, as part of our focus group research on a key social justice issue in South Africa, namely access to women’s health care resources, we asked community members to suggest an appropriate way of starting off the focus groups. Because some of our
groups included youth, the suggestion was made that each participant should provide their first name and the name of a country whose first letter matched that of their name. This simple strategy helped to get the focus groups off on a lighthearted note, without asking participants to grapple with too difficult a topic or to divulge anything too intimate about themselves.

**Monitoring the Focus Group Process**

Monitoring focus group processes is essential to the quality and success of a focus group study. There are many different ways in which researchers can accomplish this; here, we mention two possible steps: (1) the use of debriefing reports that are put together by moderators after each focus group, and (2) ongoing review of the focus group audiotapes and/or transcripts by research staff, including the principal investigator (PI). Debriefing reports are completed by the moderator usually within 24 h of a focus group discussion and summarize the group dynamics, the quality of responses to questions, the main themes, any peculiarities in the group that may have affected its responses, and suggestions for conducting future groups. Data for debriefing could also be based on observations made or notes taken by moderators of participants’ nonverbal behaviors and of their own responses. Researchers can also meet and verbally debrief with moderators immediately following a focus group. However, a written debriefing report is useful as a record that can later help in the evaluation and analysis phase of the research process (see description later). The second step, reviewing the audiotapes or transcripts from a focus group, allows researchers to evaluate for themselves the quality of interaction and response in a given focus group. Both these steps help ensure not just that quality data are being collected, but that the focus group experience is mutually rewarding for both the researcher and the participants.

**DATA ANALYSIS**

**Preliminary Analysis**

The data analysis process for focus groups is largely driven by the research question or aims of the research. Data analysis can begin immediately after
the first group ends or once all focus groups have been conducted. It is a good idea to start informally analyzing the data from the outset of the study so that the moderator and researcher have the opportunity to identify any challenges in the process or with the content. Materials for this informal analysis can include the audiotapes or transcripts (if already available) of the discussions and any debriefing notes that the moderator(s) may have taken about nonverbal and other behaviors among participants. Preliminary data analysis can also be used to identify issues or themes that the researcher may want to take up in subsequent focus groups. However, this strategy can be problematic if the study design or its anticipated outcomes depend on consistency in the kinds of questions being asked from one focus group to the next.

Full Analysis

Analysis of focus group data can be and often is quite complex, especially given that the researcher may need to process large amounts of narrative. Despite the qualitative nature of the data, its analysis must be systematic, verifiable, and context driven. The analysis process should be guided by the aims, overall philosophy, and anticipated outcomes of the research. Often the main outcome will take the form of a report of the pertinent issues discussed in the focus groups. As noted above, in other cases, the researcher may use the focus group data as part of formative research which will inform a larger, more representative study.

Researchers have found that engaging multiple participants in the data analysis process can greatly facilitate rich analysis and interpretation of focus group data. For example, the authors employed and trained South African community members as well as US-based research assistants to help in analyzing their focus group data. This approach helped address the researchers’ concern that their data might be distorted if analyzed through the social and cultural lens of only South African or American research assistants. In this approach, the South African research assistants brought their intimate knowledge of the local community and its wider social and cultural context to the data, while the US. research assistants added a useful critical distance, along with their prior training and experience in data analysis. By involving South African community members in the analysis phase, the researchers were also being consistent with their participatory philosophy, which emphasized the need for active community involvement in all phase of their research.
Below, we describe one method that is increasingly used in focus group data analysis: workshop-based summarizing and interpretation of data. However, it should be noted that focus group researchers have described many different kinds of methods for analyzing data. Different software programs also exist to facilitate qualitative data analysis, including focus group data. These programs offer ways of innovatively and rapidly sorting and organizing qualitative data, moving between data sets, and streamlining their analysis. Online tutorials can help users learn how to use these programs; however, the upfront time and effort required to master these programs are still significant. The choice of analytic method may be affected by many factors, including the size of the focus group study. Data analysis for large studies involving 10 or more focus groups, for example, may best be conducted using a computer program rather than through a workshop-based or manual cut and paste method (Krueger & Casey, 2000). The research objectives and anticipated outcomes should also play a key role in deciding what method to use for data analysis.

Workshop-Based Summaries and Interpretation

One practical and useful way to analyze focus group data is to begin by reviewing the audiotapes and transcripts, and then creating summaries of participants’ responses to each question asked. Since the summaries contain only a synopsis of what was said, the original audiotapes and transcripts may need to be repeatedly consulted to place the summaries back into their contexts. Research assistants can help with this process by writing summaries and offering, in a separate section, their initial interpretations of what was said and not said in the focus group, and what appeared to be the most compelling theme or themes for each question asked. Having these summaries and interpretations independently generated by at least two people will allow them to be placed side by side for comparison and validation.

Data analysis workshops can be used to review and verify summaries and interpretations. Often, what to include or leave out of a summary or interpretation will need to be decided through a process of discussion and negotiation among the researchers and research assistants. This process can be difficult, in part because people can be overzealous in their efforts to highlight certain themes in the data or to interpret the information in one way or another. Nonetheless, this workshop-based negotiation of the data and its interpretation may be one of the most effective ways of minimizing
the intrusion of individual bias into the data analysis and interpretation phase of the research.

In our South African work, we held workshops both in the US and South Africa, aimed at analyzing and interpreting our focus group findings. Research Associates (RAs) in both countries were trained to create comprehensive, substantive response summaries for each focus group question. Each summary had three components: a quantitative list of responses, showing how many times a particular behavior, issue, or theme was mentioned or talked about over the course of the focus group; a narrative synopsis of the themes and issues that emerged from participants’ responses; and, the RAs’ personal interpretations of the responses to each focus group question. An example of each of these components is provided below:

Example of analyzed focus group data
Focus group question: What do young people in this community do for fun?

1. Quantitative list of responses:
   - Listen to music – 8
   - Singing – 1
   - Go out to malls – 1
   - Watch movies – 2
   - Washing and cleaning my place at home – 1
   - Hanging out with friends – 2

2. Qualitative summary

The majority of focus group participants (8 girls) reported that they liked to listen to music, particularly R&B. Other types of music the participants reported liking are hip-hop, gospel, Kwaito, jazz, and Indian music. The participants reported a wide variety of activities they do for fun. The girls reported that friends’ houses, the library, taverns, game shops, malls, and the ice rink are places where they spend time with their friends. Game shops are places where alcohol is not served and children go to play games. One participant explained that taverns are places where grown people drink alcohol: “Sorry, there’s a difference between a game shop and the taverns, they don’t put games in the taverns, grown people go there, and at the game shops, children go there” (p. 3). On the other hand, several girls said that young girls, including some of their friends, from their school go to taverns.
A few participants reported that crime makes it difficult for youth to have fun. When asked, one participant said that crime means: “Some people will rape you, take you away from your home, and kill you,” (p. 4).

3. Interpretations

**Interpretation (RA 1)**

These girls implicitly communicate the importance of friends and the peer group, particularly in relation to how they spend their time. The prevalence of sexual violence and other types of violence in the daily lives of these girls is evident. The fact that they feel they have nowhere safe to go indicates a dangerous environment that must be navigated. These girls have identified a few safe places for themselves, however, such as the library, though from their laughter it seems that may not necessarily be a fun place to spend time.

**Interpretation (RA 2)**

While participants list a wide variety of fun activities that they do in their spare time, they also depict their community as a dangerous place where girls must face potential violence at many public places, even the supermarket. It seems that a lot of bad activity centers at the taverns, which several attribute to the alcohol there. The girls portray their town with a matter-of-fact attitude and seem to speak optimistically, despite all the crime.

These summaries were created using a workshop approach, both by RAs based in the United States and in South Africa. Prior to each workshop, each member of the research team read the full transcripts for each particular focus group. One RA prepared a detailed, three-part summary for each question, consisting of the list of categories, narrative synopsis, and subjective interpretations. We used the same process both in the US and in South Africa, with slight variations due to the local context and resources.

Process in United States: Prior to each workshop, the research team, consisting of the two principal investigators (PIs) and two RAs, read the full transcripts for each particular focus group. One RA would prepare a detailed, three-part summary for each question. For the first two focus groups we analyzed, each of the two RAs summarized the same transcript and then merged their summaries. It was then determined that the summaries were similar enough that this verification process was unnecessary. In the workshops, we read the summaries of each focus group question for accuracy, and discussed modifications. In general, the
summaries did not require substantive changes. Nonetheless, the subjective interpretation portions of each summary were especially illuminating in that they often revealed the contextual biases of the RAs. Interpretations were discussed, but not corrected. After we reached agreement on the accuracy of the summaries for all focus groups, we analyzed the narratives for common themes across all groups.

Process in South Africa: Similar to the US data analysis procedure, two RAs were assigned to each transcript. The summaries were written in English. To establish coder reliability, and to replicate the role of the study investigators in the US as much as possible, a senior RA was assigned to verify the accuracy and completeness of each summary. In addition, the PIs received and delivered regular feedback via telephone and email to the analysis team in South Africa.

This analytic approach had several distinctive advantages. It allowed for the creation of a manageable and dynamic data set, consisting of short summaries of otherwise long and potentially unwieldy transcripts; a combination of qualitative and quantitative data; and, interpretations reflective of the social and cultural context in which they were done. It was also a creative and lively way to analyze focus group data, and, in South Africa, a way of keeping the community members whom we hired to help us with the research involved and interested in a phase of research that can be all too easily desegregated from the community setting. One potential drawback of this approach is that key information can be lost as a result of generating short narrative summaries of the much longer transcripts. We addressed this limitation by moving back and forth between the summaries and the original transcripts in order to contextualize quotations, to identify what may have been said both before and after a summarized segment, and to retain the original tone or texture of the focus group discussion. It is possible and in some cases more appropriate to use other methods to analyze focus group data. It is also possible to construct only one or two components of our approach, for example, only the quantitative listings or only the narrative summaries. However, using them in combination will make for a richer, more multifaceted analysis of focus group data.

DATA DISSEMINATION

After the analysis, the nature and the scope of data dissemination is largely determined by the purpose of the focus groups and the overarching
philosophy guiding the research. In cases where the researcher was contracted to do the research, the goal in this phase is usually to provide the client with a report. Reports have different formats including narrative, report memo, top-line, and bulleted reports (Krueger & Casey, 2000). The narrative report is the lengthier document, which is framed by the main focus group questions or the issues that has emerged from the data. This report usually includes recommendations for the client. The report memo is typically geared toward focus group participants and its purpose is to assure the focus group participants that they were heard. It commonly focuses on progress that has been made since the groups met or includes future goals that will further address the concerns of participants. Similarly, the top-line report is a much more concise report which includes a combination of bulleted points and narrative about the focus group. In fact, this report may be somewhat similar to the debriefing report in that it is usually presented to the client within a day or two after the group. The top-line report is usually prepared without careful data analysis but is based on a more immediate evaluation of the focus group. This report is a standard in market research but it may have value to the researchers attempting to disseminate findings about sensitive or controversial issues.

When focus groups are conducted for academic research purposes, their findings are usually shared with an academic audience through journals, other publications, or at conference proceedings. However, other audiences can also be reached with focus group findings. In fact, because focus group research often marks the researcher’s first entree into a community, it provides a unique opportunity to help build support and trust through data sharing and dissemination. Researchers may find it particularly useful to share data with individuals, groups, or communities who participated in the research in order to stimulate constructive discussion of sensitive or controversial issues. Sharing data in this way can boost trust, support, and accountability between researchers and research participants.

The format and scope of data dissemination needs to be determined by the research question. In their South African study, the authors returned to the community and met with various local stakeholders to share their findings (Mosavel et al., 2005). One of their goals in doing this was to demonstrate accountability to their initial contacts and to the community. The researchers used a variety of reporting methods including written reports, structured and unstructured conversational reports, as well as informal and formal briefing sessions. They met with representatives of local government, with school principals, and with potentially interested health care agencies and provided them with an executive summary of the focus
group findings. They also discussed the findings in informal briefings with community teachers, clinic staff, and library personnel who had daily contact with young people in the community. Other members of the community were invited to informal presentations at which the results of the focus groups were presented for commentary and discussion.

In sharing data with research participants or communities, the focus group researcher should anticipate that some people might ask, “how is this information relevant to us?” or “why are you telling us this?” It is also important for the presenter to emphasize that the focus group does not generate generalizable data. These and other factors may lead audiences to justifiably question the relevance or significance of focus group data. Krueger and Casey call this the, “ho-hum syndrome,” which tends to accompany the presentation of focus group data that has not been appropriately or clearly presented (Krueger & Casey, 2000). Researchers need to anticipate this reaction and provide clear information that indicates to the audience the relevance of the data and how it might affect their lives or the life of their communities.

**CONCLUSION**

Focus groups can provide a rich and meaningful context for exploring many different kinds of bioethics issues. They are an excellent tool for formative exploration of sensitive or controversial topics and for partnership building with research participants and communities. The strengths of focus group methodology lie in its participatory and interactive nature. By generating interaction and discussion, focus groups can explore issues to a degree not possible in interviews, surveys, or other empirical tools. The tradeoff for this depth and richness lies in the limited generalizability of focus group findings, the large amount of qualitative data that focus groups generate, and the challenge of analyzing these data. Researchers interested in using focus groups in their studies need to evaluate these strengths and limitations against their research goals, available resources, and other factors.

We conclude this chapter with a note on possible future innovations in focus group research. Researchers have recently started exploring the benefits of conducting ethics-focused and other kinds of focus group research utilizing the Internet or World Wide Web. The benefits of this technology include the ability to “bring together” participants who may live or work far apart, even in different countries. Online focus group research may also provide certain social groups with an appreciated sense of
anonymity, social distance, and safety. One possible drawback of internet-based focus groups is that they may be able to recruit only individuals who have access to and are able to use online computers. Economically disadvantaged, elderly, and other individuals who typically have limited access to the Internet therefore may be excluded from important research. Adequate informed consent may also be difficult to obtain over the Internet, and discussions may be hard to keep private if they are conducted, recorded, and/or stored online. Unless they are televised in some way, online focus groups may also lack the visual and proximal intimacy that enriches face-to-face interaction. These and other potential advantages and drawbacks still need to be fully explored before the usefulness of conducting online bioethics focus group research can be determined.

REFERENCES


CONTEXTUALIZING ETHICAL DILEMMAS: ETHNOGRAPHY FOR BIOETHICS

Elisa J. Gordon and Betty Wolder Levin

ABSTRACT

Ethnography is a qualitative, naturalistic research method derived from the anthropological tradition. Ethnography uses participant observation supplemented by other research methods to gain holistic understandings of cultural groups’ beliefs and behaviors. Ethnography contributes to bioethics by: (1) locating bioethical dilemmas in their social, political, economic, and ideological contexts; (2) explicating the beliefs and behaviors of involved individuals; (3) making tacit knowledge explicit; (4) highlighting differences between ideal norms and actual behaviors; (5) identifying previously unrecognized phenomena; and (6) generating new questions for research. More comparative and longitudinal ethnographic research can contribute to better understanding of and responses to bioethical dilemmas.

INTRODUCTION

Ethnography aims to understand the meanings that individuals attach to situations or events under study and the myriad of factors that affect beliefs
and behavior. Because of this, ethnography is well suited to the study of bioethics. Bioethical issues and dilemmas are morally charged, laden with meaning, and unfold through social interaction. Ethnographic research is therefore ideal for opening the door to the world of meanings attributed to health-related events and moral decisions, and for understanding the broader socioeconomic and political factors shaping how cultures and cultural members frame, interpret, and respond to such phenomena.

Ethnography was one of the first methods used to conduct empirical research on bioethical issues (Fox, 1959; Glaser & Strauss, 1965). Ethnographic studies relating to bioethics can be categorized into 3 groups. The first group is specifically about the work of bioethics itself, such as research on the role and functioning of Institutional Review Boards (IRBs) or hospital ethics committees (chapters in Weisz, 1990; DeVries & Subedi, 1998; Hoffmaster, 2001). The second group aims to elucidate bioethical issues with a focus on how bioethical dilemmas and conflicts develop and/or are addressed (see Guillemin & Holmstrom, 1986; Levin, 1986; Anspach, 1993; Zussman, 1992; DeVries, Bosk, Orfali, & Turner, 2007). The third set of ethnographic studies is not framed primarily as research in bioethics, but is relevant to the field, such as Fox’s and Swazey’s (1978, 1992) classic studies of dialysis and organ transplantation and Bosk’s (1979) examination of the socialization of surgeons in the context of surgical mistakes. Other seminal studies in this genre include Bluebond-Langer (1978) on children with leukemia; Estroff (1981) on people living with mental impairments; Ginsburg (1989) on the abortion debates in an American community; Rapp (1999) and Bosk (1992) on prenatal genetic testing; and Farmer (1999) on the social context of AIDS and other infectious diseases.

Bioethicists may see ethnography as a good source of illustrative cases or dramatic stories gathered simply through observation. But ethnography involves more than just observation – it relies on understanding the social processes underlying phenomena that are observed, building on prior research and analytic methods developed by social scientists, and on the systematic analysis of data that goes beyond simple description.

The art of ethnography relies on the skills, knowledge, and sensitivity of the researcher. Interpretation based on good ethnography may provide all the information one needs in many circumstances. Or, it may be only the starting point, raising questions to be further investigated with the use of other methods such as a survey instrument that can more systematically collect information from larger numbers of respondents than can be observed through ethnography.
In this chapter, we introduce the qualitative method of ethnography, provide practical guidance on how to conduct ethnographic research, particularly as it relates to bioethical issues, and in doing so, highlight the value of ethnography to bioethics. Although some ethnographic research has been undertaken in the area of public health ethics and related topics (Brugge & Cole, 2003; Marshall & Rotimi, 2001; Lamphere, 2005), most ethnographic research cited in the field of bioethics has been clinically based. Thus, we primarily discuss ethnographic research done in clinical settings and will include examples drawn from the authors’ own experiences conducting ethnographic research on decision-making within neonatal intensive care units (NICUs), and on kidney transplantation, to offer insights into this research endeavor.

**Defining Ethnography and Culture**

There are many ways to define ethnography, grounded in different theoretical schools of thought about social research. However, all agree that ethnography has the following characteristics: it is a qualitative, naturalistic research method that derives from the anthropological tradition. An ethnographer typically goes into the field with a research question developed to build on existing social theory and/or previous substantive research with the aim of better understanding a cultural group or groups. The ethnographer aims to explicate the behaviors as well as the meaning of those behaviors of the people observed within a holistic context. In other words, ethnographers aim to describe a culture – whether it be the culture of the Navajo, intensive care, or people involved in organ donation and transplantation – by examining the worldviews, beliefs, values, and behaviors among its members. Often ethnography seeks to explore the historical, social, economic, political, and/or ideological factors which may account for cultural phenomena.

The work is inductive, rather than deductive. It does not test a pre-established, fixed hypothesis and does not collect data only using previously defined variables. Instead, concepts and variables emerge through the ethnographic process. As we detail below, the prime method of ethnography is participant observation. This entails immersion in the field situation, establishing rapport with individuals, and gaining knowledge through first-hand observation of social behaviors. This is complemented by direct interaction with people in the field, and participation in their activities. Ethnographers may also use other techniques such as semi-structured
interviews, surveys, focus groups, and do textual analysis of relevant documents. Ethnography is not only the process just described, but also a product – a written account or ethnography – derived from the process (Roper & Shapira, 2000).

The concept of culture is essential to ethnography. Some fundamental points to understand about culture are: (1) culture is shared among a group of people (i.e., members of a nation, religion, profession, or institution); (2) culture entails patterns of behavior, values, beliefs, language, thoughts, customs, rituals, morals, and material objects made by people; (3) culture provides a framework for interpreting and modeling social behavior; (4) culture is learned through social interaction; (5) structural factors determine social positions which affect people’s worldviews and behaviors; (6) culture interacts with gender, class, ethnicity/race, age, (dis)ability, and other social characteristics; (7) culture is fundamental to a person’s self-identity; and (8) cultures change over time in response to changes in social, political, economic, and physical environments.

There are many definitions of culture, and ethnographers vary in the approaches they use to describe it. Here, we present two definitions – the first is a classic definition by Tylor (1958[1871]) that provides a broad sense of culture as “... that complex whole which includes knowledge, belief, art, morals, law, custom, and any other capabilities and habits acquired by man as a member of society” (p. 1). A second and often-quoted conception of culture is provided by Geertz who stated: “man is an animal suspended in webs of significance he himself has spun ... I take culture to be those webs, and the analysis of it to be ... not an experimental science in search of law but an interpretive one in search of meaning” (Geertz, 1973, p. 5). According to this definition, the meaning of all actions and things are socially constructed and shared. Culture is comprised of the symbols and meanings attached to actions and other phenomena that help people communicate, interpret, and understand their world.

Both definitions can be helpful for analyzing the culture of medicine, and specifically the culture of bioethics. Bioethics is situated at the confluence of complex legal and moral systems. Given their moral content, bioethical issues are laden with multiple meanings, and the actions agents take to resolve ethical problems or dilemmas are symbolically charged. For example, many clinicians perceive withdrawing life-sustaining therapy as different from not initiating life-sustaining therapy. So, by custom, they try to avoid withdrawing therapy when their goals can be met by not initiating a new therapy. However, for most bioethicists, these practices are conceptually and ethically synonymous.
Most ethnographies conducted in the area of bioethics have been done in complex, pluralistic societies where biomedicine is the dominant professional medical system. In such societies, there may be significant cultural variations between members from different backgrounds, social statuses, and other groups (e.g., health professionals, patients, ethnic/religious groups, nationalities, genders, and age groups). People from different groups often vary in worldview and behaviors. When people interact they may assume that all the social “players” in a given situation share the same assumptions, knowledge, and beliefs that they hold, even when this is not the case. Moreover, people may believe their way of seeing things is the most valid or only way of interpreting reality. Conversely, people may assume that other groups hold different beliefs from their own, even when they do not. These assumptions can be a source of confusion, and contribute to bioethical dilemmas and value conflicts. Ethnography is an excellent method for examining cultural assumptions and studying their effects on behaviors, social interactions, and decisions in the health care environment.

Objectives of Ethnography

Researchers conduct ethnography to accomplish one or more of the following objectives: (1) understanding a phenomenon from the “native’s” or “participant’s” (i.e., member of the culture’s) point of view; (2) describing a given culture by making culturally embedded norms or tacit assumptions shared by members of a cultural group explicit; (3) discerning differences between ideal and actual behavior; (4) explaining behavior, social structure, interactions within and/or between groups, or the effects of economic, institutional, global, or ecological factors; (5) examining social processes in-depth; and (6) revealing unanticipated findings that can generate new research questions. Each of these objectives is described below.

(1) **Understanding human behaviors from the “insider’s” point of view.** To describe and analyze the insider’s point of view, ethnographers distinguish between “emic” and “etic.” Emic refers to words and concepts that the people who are observed in a given setting use themselves (and are therefore significant in their culture), while etic refers to the abstracted concepts that scholars or researchers use for analysis. This distinction is illustrated well in a study of terms used to describe infections: whereas the emic “folk” term “flu” was often used by lay people, the emic term “viral syndrome” was used by physicians.
documenting patient’s problems (McCombie, 1987). However, neither term was adequate for epidemiologists conducting enteric disease surveillance. Epidemiological investigations and disease control required the use of etic or analytic terms to more precisely categorize diseases by the specific causal agent (McCombie, 1987).

(2) Revealing culturally embedded norms or tacit assumptions shared among members of a cultural group. Because people understand their physical and cultural worlds through the lens of their cultural understandings, their values, beliefs, norms, and facts are usually assumed to be naturally given or taken for granted and are sometimes subconscious. Accordingly, members often cannot explain or even articulate many aspects of their culture. As Jenkins and Karno (1992) state, “In everyday life, culture is something people come to take for granted—their way of feeling, thinking and being in the world—the unselfconscious medium of experience, interpretation, and action” (p. 10).

Traditionally, ethnographers have conducted research on cultural groups to which they do not belong. However, conducting ethnographic research about one’s own culture, which is common in current bioethics research, can be very difficult. Ethnographers may find it easier to understand the perspectives of people from their own culture, but more difficult to identify subconscious assumptions. The ability to identify tacit assumptions depends on not having been socialized as a member of the group under examination, and/or by using techniques designed to reveal such assumptions.

For example, in a 1977 study of decision-making in the NICU, the researcher initially knew little of the culture of biomedicine or bioethics and was naïve about the distinction between CPAP (a device supplying continuous positive airway pressure to keep lungs inflated) and a respirator. She was confused when a nurse told her that an infant on CPAP might not be put back on a respirator if his respiratory condition deteriorated because they believed he probably had a cerebral bleed and therefore would have a poor quality of life (Levin, 1985, 1986). The researcher knew that decisions were sometimes made not to treat a baby on the basis of the future quality of life. However, when she asked what difference there would be between keeping him on CPAP or putting him on the respirator, the clinicians just described the technical differences in the two technologies. Through participant observation of the care of infants in the unit, talking to doctors, nurses, social workers, and parents, attending rounds, reading charts, and reading the medical and bioethics literature, the ethnographer realized that clinicians made emic
distinctions between treatments according to the level of “aggressiveness” as well as between “withholding” and “withdrawing” treatment. Since taking into consideration these aspects of the treatment seemed “natural” to them, they could not articulate clearly the reasons for their decision.

(3) **Discerning differences between ideal behavior and what actually happens.** For example, physicians may say that it is important for patients or family members to be involved in important decisions about care. Yet in many cases, informed consent has become a ritual where clinicians follow the letter, but not the spirit, of the law in the process of obtaining consent. Physicians frequently influence patients’ treatment decisions in subtle and direct ways, through body language or by emphasizing the risks of one treatment and the benefits of another treatment to obtain the decision physicians prefer (Zussman, 1992).

(4) **Viewing phenomena in their economic, political, social, ideological, and historical contexts.** In the NICU study mentioned above, the ethnographer endeavored to understand the clinicians’ views and decisions that were made in the context of the history of the care of newborns, the development of life support technology and of the care of people with disabilities, as well as those who were critically and terminally ill. In the middle of the study, when the “Baby Doe Controversy” over the treatment of “handicapped newborns” occurred (Caplan & Cohen, 1987), the investigation was expanded to include examination of the ways NICU physicians’ understanding of the regulations and the controversy affected decisions about care (Levin, 1985, 1988).

Examining how political, economic and social forces impact cultural understandings, cultural values, and social structures, and how these in turn shape both behavior and the discourse surrounding an issue highlights the cultural construction and “situatedness” of given phenomena. This kind of information is valuable because it demonstrates that matters of bioethical concern do not have to be framed in only one way, and it illustrates alternative routes to construing issues or addressing them. For example, Margaret Lock’s (2002) work on views of death and organ transplantation in Japan is an excellent illustration of the ways a non-Western cultural system leads to different ethical perspectives than a Western culture even when the biotechnology is similar in both cultures.

(5) **Examining social processes and social phenomena in greater depth.** Immersion in the research setting enables researchers to understand the subtleties and nuances of phenomena under study and how components of phenomena are related. For example, one can conduct a series of case
studies to examine the experiences, attitudes, and behaviors of patients, family members, and health care professionals encountering ethical dilemmas. Such research aims to elucidate how social interactions and power dynamics in the clinical setting affect the resolution of bioethical problems. Additionally, immersion helps ethnographers to become sensitive to the political implications of what they observe as well as how they represent cultural perspectives in their reports (Clifford & Marcus, 1986).

(6) Uncovering factors and processes unanticipated at the beginning of the research process. Quite often, serendipitous events lead to new and revealing observations and insights. Accordingly, ethnography can be helpful in generating new analytic frameworks, research questions, and hypotheses that can be tested using other research methods. For example, during the course of a study of disparities in gaining access to kidney transplantation (Gordon, 2001b), emerging evidence indicated that patients faced difficulties with maintaining the transplant. This concern led to development of a new research question concerning long-term graft survival.

The Application of Ethnography to Bioethics

In the ethnographic study of bioethical issues, three main, albeit overlapping, sets of problems are generally examined. First is the examination of everyday ethics in clinical settings. However, as Powers (2001) states, “The challenge of recognizing everyday ethical issues lies in their ordinariness” (p. 339). In other words, it may be difficult to problematize or consider as cultural those practices and beliefs that members of the group being studied treat as “normal” or “natural.” Ironically, it is precisely when cultural members construe issues as “normal” or “natural” that ethnographers can identify a phenomenon with important cultural dimensions. Second, much research focuses on examining whether bioethical principles and assumptions derived from philosophy are actually applied in reality, and if not, why not. For example, a study by Drought and Koenig (2002) examining the principle of respect for autonomy and the concept of patient “choice” in decision-making of dying patients, their families, and their health care providers, found that patients did not perceive that they had choices when discussing treatment options, contrary to bioethics scholars’ expectations that patient autonomy would be respected. A study by Dill (1995) also illustrates a challenge to assumptions about the principle of respect for autonomy in hospital discharge planning decisions for older adults.
Third, ethnographic research in bioethics aims to elucidate how cultural context(s) shape ethical reasoning. Researchers taking this approach seek to highlight Western assumptions that pervade bioethics, and/or investigate ethical reasoning in other cultures for a comparative approach. For example, Jecker and Berg (1992) examined how the face-to-face dynamics of living in a small, rural American setting shaped the way scarce medical resources were allocated on an individual patient level in a primary care setting. This contrasts with philosophical expectations about justice as a blinded, impersonal process. A related approach seeks to examine the diversity of experiences within a culture or among subgroups regarding particular bioethical phenomena. For example, numerous studies have shown that African Americans prefer more aggressive life-sustaining treatments compared to European Americans (Blackhall et al., 1999). Other research suggests that Koreans and Mexican Americans appear to favor physician disclosure of grave diagnoses or terminal prognoses to family members instead of to the patient in order to protect the patient and enable familial decision-making rather than patient autonomy (Blackhall Murphy, Frank, Michel, & Azen, 1995; Orona, Koenig, & Davis, 1994; Hern, Koenig, Moore, & Marshall, 1998).

ETHNOGRAPHIC DATA COLLECTION TECHNIQUES

Ethnography is best learned through experience, reading ethnographies, engaging in discussions with people who have done them, and doing smaller or pilot studies, rather than through purely didactic training. Accordingly, ethnography is a difficult research method to teach. We emphasize that ethnography is more than just observing a situation. One must know what to look for, how to make observations through the lens of the complex concept of culture, and how to interpret and analyze data in light of the social, historical, and cultural contexts in which data are collected. Ethnographers generally use multiple data sources and data collection techniques to obtain rich and overlapping data. Their primary source of data is participant observation. They may also use other techniques including interviews and case studies. These techniques are discussed below along with the skills necessary to use each one.

Participant Observation

Participant observation is the heart of ethnography; it is a strategy enabling the ethnographer to “listen to and observe people in their natural
environments” (Spradley, 1979, p. 32). Participant observation allows for the examination of several dimensions of a social situation simultaneously – the physical, behavioral, verbal, nonverbal, and interactional – in the context of the broader social and physical environment. Doing so “give[s] the researcher a grasp of the way things are organized and prioritized, how people relate to one another, and the ways in which social and physical boundaries are defined” (Schensul & LeCompte, 1999, p. 91). A strength of participant observation is that researchers are the “instrument of both data collection and analysis through [their] own experience” (Bernard, 1988, p. 152).

Ethnographic techniques vary depending on the extent to which ethnographers identify themselves as insiders or outsiders; how involved the people who are observed are in the data collection effort, and the kinds of activities that researchers engage in as part of fieldwork (Atkinson & Hammersley, 1994). Although the technique is commonly referred to as participant observation, not all ethnographers are actual “participants.” For example, in bioethics research, ethnographers are often participant observers and join providers during rounds and team meetings, or share lunchtime conversations with staff. However, they do not express their opinions about bioethical or other issues discussed. In some settings, however, the ethnographer may be only an observer. Examples of this are attending a committee meeting and listening to and observing interactions of its members. Ethnographers who are studying problems in bioethics may find it enlightening to “observe” phenomena outside the clinical setting such as advocacy groups or representations of issues in the media. For example, members of Not Dead Yet, which advocates for disability rights, constitute a valuable source for understanding non-clinical perspectives on end of life practices (see, for example, http://www.notdeadyet.org/docs/about.html). All data collection approaches can be used simultaneously. In one study examining access to transplantation, the ethnographer observed transplant team meetings; the formal interactions between transplant coordinators and transplant candidates and their families; clinical encounters between nephrologists and dialysis patients; shadowed transplant surgeons on their medical rounds; and observed monthly social support group meetings run by the transplant center (Gordon & Sehgal, 2000; Gordon, 2000).

**Skills Necessary for Effective Participant Observation**

Ethnography, specifically the strategy of participant observation, requires that researchers develop many skills, including self-reflexivity, having a good memory, attending to details, flexibility, interpersonal skills, the ability to exert discretion, building rapport, and appreciating cultural differences.
Self-reflexivity is critical to an investigator’s success in using her or himself as a research tool. Self-reflexivity can be defined as working to assess one’s own biases and their potential influence on perceptions of phenomena (Frank, 1997; Ahern, 1999). (For an excellent example of how a medical anthropologist engaged in self-reflexivity to understand her informant’s disability experience, see Gelya Frank’s (2000) monograph, *Venus on Wheels*).

An important skill required in ethnography – whether observing or interviewing – is having a good memory. During unstructured interviews, it is essential not only to have the next question ready at hand, but also several possible leads for additional questions or comments in mind, based on earlier parts of the conversation. In addition, it can be useful to keep a list of important points to cover during an unstructured interview. Because the interviewer cannot always take notes or audiotape interviews, useful strategies are: writing everything down immediately after making observations; avoiding speaking to people about the observations before writing them down; recalling things chronologically as they were witnessed; and drawing a map of the physical space in which events occurred (Bernard, 1988, p. 157).

Developing “explicit awareness” of details of ordinary life is another skill (Spradley, 1979, p. 55). People go about life aware, but not attending to many details, e.g., what people are wearing, music playing in public places, the process involved in deciding which products to buy at supermarkets, etc. (Bernard, 1988). In fact, this occurs because many aspects of cultural life are tacit. Since one definition of culture refers to the knowledge necessary for a person to get by in his/her culture, generating information on such ordinary details provides insight into the daily lives of members in a given culture. Attending to details helps to keep biases in check and often leads to insights about assumed realities. For example, ethnographers in medical settings can gain important knowledge by noticing which cases are talked about the longest during rounds or noticing who regularly attends or skips rounds or meetings.

The researcher must exert a fair degree of flexibility when conducting ethnographic research. For example, flexibility to alter the course of inquiry if needed is a hallmark of the ethnographic approach. Drought and Koenig (2002) exercised flexibility as the data emerging during their data collection indicated that the AIDS and cancer patients in their study did not conform to the categories anticipated by the normative assumptions in bioethics regarding the existence of discrete decision points at the end of life.

Conducting ethnography requires that the researcher have interpersonal skills and exercise discretion. Building rapport with people is necessary to
establish trust and open the door to communication. Ethnographers need to speak to people of diverse backgrounds to understand how power dynamics and social status shape attitudes and behaviors. In the context of bioethics research in the clinical setting, this may entail going on medical rounds daily, shadowing clinicians, talking to providers and staff, sitting at the computer station with them, drinking coffee together, staying at the unit on night shift hours, and participating in social events on the unit. The ethnographer’s goal is to “fade” into the social fabric of the group under observations with as little impact as possible on the phenomena under study.

Participant observers must also engage key informants to help guide them about the culture under study and provide insider information about aspects of a culture. One needs at least one or two key informants in each setting who have necessary competence to provide in-depth information about a particular domain of culture. Researchers can ask key informants about how things work, factual information, or about things that are typically perceived by members of the culture to be essential to understand their culture. For example, to understand treatment decisions about critically ill neonates, one might ask, for what kinds of conditions do NICU babies tend to have an ethics consultation requested (Orfali & Gordon, 2004). However, informants are not selected for their representativeness and they may not be able to report accurately about opinions or attitudes that vary in the culture (Bernard, 1988). Key informants can also be extraordinarily helpful in guiding observations, enlisting the involvement of others, and partaking in ad hoc interviews. In the study of dialysis patients’ choices about transplantation, one key informant was an administrative secretary who provided behind-the-scenes information about the structure and functioning of the transplant center and background of the health care professionals working there (Gordon, 2000).

It is important to be mindful of how one selects key informants in terms of the political alignments among people in the particular social context. One should avoid aligning with key informants who are too marginal to the group under study so that one does not lose access to people and information. Ethnographers must also be careful that alignment with a key informant does not alienate other members of the group who are important to the study. Informants may emerge through establishing friendships based on trust or luck. The best informants tend to be articulate people who are somewhat cynical about their own culture and, even though they are insiders, feel somewhat marginal to it (Bernard, 1988).

Ethnography requires an ability to adopt a culturally relative perspective when doing research. By this we mean endeavoring to understand and
respect the views of others and events within cultural, social, and historical contexts rather than making judgments about what is observed based on one’s own cultural perspective. Ethnographers must be able to appreciate that people are “rational” and systematic in their thinking (Horton, 1967), and that the beliefs and values underlying people’s thought processes and behaviors may differ from their own. As a result, when confronted with the same choices, people may vary in the kinds of conclusions they reach. Using a culturally relative perspective when doing research does not require ethnographers to accept or adopt a different value system than their own, but to strive to be as nonjudgmental and open-minded as possible, in order to understand, and explicate alternative worldviews when collecting and interpreting data.

**Interviews**

Interviews constitute another major technique for collecting ethnographic data. There are three main types of interviews: unstructured, semi-structured, and structured, of which unstructured interviews are the most commonly used. Unstructured and semi-structured interviews can occur spontaneously in the course of participant observation research, or researchers may seek out people to ask them about specific issues. Informal and unstructured interviews are ideally conducted in the midst of participant observation when ethnographers can get immediate input on the meaning of events as they occur, especially those that are unexpected. The ethnographer may have a general idea about the topic they want to learn about or they may let the respondent drive the course of the interview. This can allow respondents to raise issues unanticipated by the researcher.

When conducting semi-structured interviews, researchers prepare a written interview schedule with a set of questions or discussion points. Whereas the bulk of unstructured interviewing is open-ended, semi-structured and structured interviews commonly include both open- and closed-ended questions (for a detailed discussion on this method, see chapter on Semi-Structured Interviews).

**Case Studies**

Case studies are another ethnographic technique that entails “examining most or all aspects of a particular distinctly bounded unit or case (or series of
cases)” (Crabtree & Miller, 1999, p. 5; Stake, 1994). Cases can be individual patients, sets of interactions, programs, institutions, nations, etc. The goal of this data collection method is to describe cases in great detail and context, which may generate hypotheses or explain relationships of cause and effect (Aita & McIlvain, 1999). Cases may be selected based on whether they are representative of a phenomenon, setting, or demographics, and whether they “offer an opportunity to learn” (Stake, 1994, p. 243). Alternatively, one may choose atypical cases to explore the limits of what is the norm, and to set limits to generalizability (Stake, 1994, p. 245).

Rapid Assessment Process

An adaptation of ethnography has been developed called “rapid ethnography” or “rapid assessment process” (RAP) (Scrimshaw & Hurtado, 1988; Bloor, 2001; Hahn, 1999). RAP is used as a faster approach to data collection in various applied settings, including public health. RAP was developed by the World Health Organization (WHO) to accommodate shorter time frames for conducting research and tends to be more problem-oriented than traditional ethnography. Generally, the RAP data collection period lasts from 3 days to 6 weeks, depending on time, resources, and previous data collected; and typically RAP uses small sample sizes (Trotter & Schensul, 1998). The RAP approach commonly uses several observers, a narrower research focus, and multiple collaborators and emphasizes the use of a number of methods including direct observation, informal conversation, and key informant interviews.

Other Methods in Ethnography

Ethnographers can draw upon a wealth of additional data sources, such as administrative records kept at hospitals or public records regarding morbidity and mortality in a population. Other relevant data sources constitute media reports, newspaper clippings, television programs, and other forms of popular culture. The Internet offers the opportunity to engage in different kinds of observations. For example, online support groups, such as a listserv that provides a venue for dialysis patients to exchange views and give each other advice, served as a useful counterpart to observations of the in-person support group, Transplant Recipients International Organization, which was observed by one of the authors.
Ethnographers may also draw upon results of surveys previously conducted by other researchers as Long (2002) did for her work on euthanasia in Japan to understand how Japanese people construed the issue.

Ethnographers also collect quantitative data (Bernard, 1988). For example, one can collect data on the number of cases with certain characteristics, the number of days in the hospital for people with different medical or demographic characteristics, ask individuals to list the number of people they could call on to care for them in their homes after leaving the hospital, etc. Although ethnographers conducting statistical tests should not assume that the quantitative data represent a random sample from a defined population, they can use descriptive statistics to identify rates and correlations using such data in combination with qualitative data.

DEVELOPING ETHNOGRAPHIC RESEARCH

Conducting ethnographic research requires considerable preparation in advance. Key preparatory steps include: (1) conceptualizing a research question, (2) establishing a research plan, (3) obtaining permission to gain access to the research site, and (4) determining the unit(s) of analysis and sampling frame(s). Depending on the nature of the study, it may involve additional elements.

1. Conceptualizing a research question. The first step in ethnographic research is conceptualizing a research question. An important characteristic of ethnographic research is that it does not usually aim to test an a priori hypothesis and is not geared toward achieving generalizability, characteristic of statistical hypothesis testing. Rather, as discussed above, the ethnographer is seeking to gain an in-depth understanding of a problem, group, or social setting within its broader context. Conceptualizing a research question requires sufficient familiarity with a topic or human group to find a gap in the scholarly literature. Accordingly, ethnographers develop a research topic derived from social theory and/or from substantive gaps in knowledge. A topic may not have been investigated at all, or phenomena may have been studied with a different theoretical framework, methodological approach, or in a different group or setting or timeframe. As with all research, a comprehensive review of the existing literature is essential in this phase.

The central research question is generally broad and subsumes multiple other questions. For example, the investigation of social and cultural factors shaping treatment decision-making regarding renal transplantation (Gordon, 2001a,b) included inquiries into the doctor–patient
communication, transplant evaluation, and patients’ decisions about
donation (Gordon & Sehgal, 2000; Gordon, 2001a). In addition, researchers
must consider the potential significance of their proposed research, i.e.,
the extent to which the knowledge gained will: (a) advance theory or
methodology in bioethics and/or in the social sciences, i.e., help to
re-conceptualize the doctor–patient relationship or bioethical principles,
(b) change clinical practice, (c) inform health interventions, or (d) inform
health policy.

2. Establishing a research plan. An important step after conceptualizing a
research question is devising a preliminary plan for undertaking the
research. This can be done by a comprehensive review of the existing
substantive and methodological literature and by identifying and, if
possible, preparing the methodological technique(s) to use during fieldwork.
The next step entails distilling the research design logistics: identifying who
or what situations to observe, which people to interview, when to do these
steps, and how. Although ethnographers usually enter the field with a fairly
open plan to develop a holistic view of the situation, ethnographers may draft
interview guides, have experts in the field to review them, or conduct pilot
studies with specific methods before initiating the main phase of research.
Collecting background data i.e., population statistics, medical data, or
administrative information, is important for gaining an appreciation for the
broader context. However, ethnographers cannot plan all of their fieldwork
in advance. Serendipitous occasions and unanticipated informal conversa-
tions occur during participant observation, which may lead to the most
important insights – this constitutes the heart of ethnography.

3. Obtaining permission to gain access to the research site. The next step
entails gaining access to the field site. Essential components of this
process illustrated below are based upon experiences by one of the
authors in conducting research with kidney transplant and dialysis
patients (Gordon 2001a,b). In this study, the investigator planned to
conduct the research in dialysis units. This necessitated obtaining
permission from the nephrologist directing all the dialysis centers that
served as observation settings. The next group from whom the ethno-
grapher needed permission was the individual nephrologists who ran each
dialysis center, followed by the other health care providers who staffed
the dialysis centers and provided the ethnographer with direct entrée to
the dialysis patients. Finally, the IRB approval was required and consent
forms were needed in order to speak with patients.

Negotiating access with chairs of clinical units or with the IRB can also be
challenging given that many chairs and IRB members lack knowledge about
qualitative research and often perceive it as a “fishing expedition” or “not really science” (Koenig, Black, & Crawley, 2003). Nonetheless, anticipating in advance potential critiques of social science and qualitative methods can better enable ethnographers to address them when they arise. For example, it is common for biomedical practitioners to comment that social science research – particularly qualitative research – is “subjective” rather than “objective.” By highlighting the strengths of qualitative research in terms of validity rather than generalizability, ethnographers can advance the understanding of their methods, foster acceptance and support for their research, facilitate IRB approval, and improve communication between ethnographer and study subjects (Anspach & Mizrachi, 2006; Koenig et al., 2003; Cassell, 1998, 2000; Lederman, 2007).

Gaining access to the research site is an informal as well as formal process. In addition to the formal permission from health care institutions and IRBs, it is crucial to obtain “buy-in” from relevant administrative and/or clinical staff and other individuals who can effectively facilitate or hinder data collection. Such informal gatekeepers can help the ethnographer gain access to electronic data, unfolding situations, potential research participants, and answer questions, all of which can help ethnographers to realize their research goals.

Moreover, gaining access to the research site is not a one-time matter but rather a process that must be negotiated repeatedly. The researcher often needs to re-introduce him/herself and re-explain his or her role to staff in order to ensure that the research itself and the researcher’s presence are understood and accepted. This is especially true in clinical settings with high staff turnover and multiple shifts. Staff may feel that ethnographers are interfering with their work space, time, and patients. To minimize this perception, ethnographers must make clear to staff that they are in the setting to learn about the people there, that they will keep information confidential, and will not spread gossip. The ethnographer can let cultural members know that his/her process of learning entails asking basic questions, the answers to which might be perceived as obvious to cultural members but help the ethnographer uncover tacit and fundamental assumptions. After they have been in the field for a long time, the ethnographer must be careful to not be seen as a regular member of the group under study as this will diminish their effectiveness (Bosk, 2001).

Finally, ethnographers may encounter challenges gaining access to research sites because (a) respondents/subjects hold powerful positions and (b) ethical issues are sensitive in nature. Nevertheless, “studying-up” or studying the powerful – such as health care professionals, administrators, or
ethicists – is important because they affect the well-being of many other members of society (Nader, 1972; Sieber, 1989, p. 1). The powerful are typically difficult to study. Anthropologist Laura Nader explains just why this is the case: “The powerful are out of reach on a number of different planes: they don’t want to be studied; it is dangerous to study the powerful; they are busy people; they are not all in one place, and so on” (Nader, 1972, p. 302). Nader provides further insights into the concerns held by the powerful about being studied: “Telling it like it is may be perceived as muckraking by the subjects of study … or by fellow professionals who feel more comfortable if data is [sic] presented in social science jargon which protect the work from common consumption” (Nader, 1972, p. 305).

Indeed, the effort by one of the co-authors to examine how ethics consultants discuss cases during ethics committee meetings was thwarted by some of the committee leaders and members because of fears about the uses of data to be obtained. Finally, the sensitive nature of many medical ethical issues requires a great deal of tact, strong communication skills, discretion, and often empathy on the part of the ethnographer.

4. Determining the units of analysis and sampling frame. Because ethnographic research operates at a number of different levels of inquiry at once, it is essential to consider the particular unit(s) of analysis before implementing a study. The choice of unit(s) is driven by the study’s aims and focus and may be at the micro level of patients or clinicians, or can be broader and include an entire medical floor or hospital or a society’s response to an ethical issue. In the study of ethics consultations, the unit of analysis was the aggregate of the patient, family, and health care professionals treating one patient (Kelly, Marshall, Sanders, Raffin, & Koenig, 1997). This study examined the interactions between individuals within these small groups and with ethics consultants to obtain a rich understanding of how consultants influence decision-making and of each party’s perspectives about the value of ethics consultations.

Determining the unit of analysis is related to the issue of sampling. In ethnographic research, sampling is not geared toward statistical generalization. Instead, sampling is performed to enable ethnographers to study a sufficient scope and depth of processes and interactions to gain understanding of situations in all their complexity. An in-depth appreciation often comes at the expense of generalizability. Data collection ideally proceeds until saturation has been achieved. Saturation is the point where no new insights, themes, or patterns are being generated. However, the ability to reach saturation may depend on the unit(s) of analysis. A tradeoff often exists between the number of situations observed and the richness of data
collected (Morse, 2000). To enhance validity of a study’s findings and obtain a better sense of the range of phenomena under study, ethnographers may make visits to other comparable programs, medical units, or institutions, albeit with less time than in the primary site.

**Informed Consent**

Ethnographic research presents challenges regarding informed consent not usually encountered in other ethnographic or bioethical research. Traditionally, much of what anthropologists observe has been “public” behavior or has occurred in settings which they have been invited to attend as participants. Often, an invitation may not even be required after anthropologists have immersed themselves and become accepted by the group or community under study. IRBs consider much ethnographic research to be exempt or subject to expedited review; oral consent is often considered adequate. Yet in clinical settings, because of the sensitive nature of the situations observed, written informed consent is expected from patients, family members, clinicians, and others who participate in formal interviews or whose cases or behavior are studied in depth. However, it may not be practical or even possible to obtain prospective informed consent from everyone who will be observed during an ethnographic study (e.g., from the dozens of people who will pass through a unit or clinic on a given day, and who are not the main focus of research). Although some IRBs will exempt ethnographers from obtaining written consent, most require verbal informed consent for observations and interviews that states the purpose of the research, guarantees confidentiality, and enables people who would be studied the choice not to participate or to withdraw from the study at any time. Of note is also the fact that the act of requesting consent to be an observer of group events (e.g., clinical meetings and consultations) sets the researcher off from the group being studied which may undermine efforts to establish rapport (Miller, 2001).

The American Anthropological Association (AAA) (2004) has published a statement on Ethnography and Institutional Review Boards ([URL: http://www.aaanet.org/stmts/irb.htm] accessed 6-28-07). Briefly, the AAAs position is that anthropological research often falls under expedited and exempt review, yet investigators may have difficulty getting the IRB to use such mechanisms. Some IRB members are unaware of how conducting ethnographic research diverges from the biomedical model leading to difficulties for ethnographers in obtaining IRB approval for verbal consent. The AAA points out that the common rule does allow for waivers of written
consent, according to certain criteria. The tension between the requirements of good ethnography and the strictures of modern informed consent regulations has been further intensified by the Health Information Portability and Accountability Act (HIPAA) of 2003 that aims to protect identifying information and prevent its transmission outside of an institution. Local variations of how informed consent for ethnography is handled mean that there are few clear guidelines for ethnographers seeking to do research in clinical settings.

Other Groundwork

Throughout the preparation process ethnographers need to be involved in other groundwork. For instance, researchers may need to learn a foreign language, in this context, that of medicine, bioethics, and the local clinic. Whereas much technical medical language can be learned in advance, learning how it is used on an everyday basis occurs through observation after entering the field. It is essential for the ethnographer to become familiar with the vocabulary, jargon, and lore of those under study; study subjects will likely teach ethnographers if they indicate a willingness to learn (Bernard, 1988). For this purpose attending rounds and case conferences can be invaluable.

Other groundwork involves learning the lay of the land, which is essential for understanding the social and symbolic meanings associated with physical spaces. For example, for the NICU study it was helpful to know that babies in the beds on one side of the unit were in the most critical condition. This awareness helped make sense of their treatment regimens and clarified that movement from room to room denoted improvement and preparation for discharge. Early stages of ethnographic research may also entail a macro-level form of site survey (Anderson & McFarlane, 2004), where one assesses the broader environmental setting in which the study takes place, i.e., the neighborhood.

DATA MANAGEMENT AND ANALYSIS

Recording Data

Ethnographic research generates large quantities of data such as field notes, responses to interviews and surveys, photographs and video-recordings, and copies of documents.
Field notes are the main method of recording data. These start as handwritten notes which are then entered into a word processing program and/or database. Field notes record what ethnographers observe about the unfolding events, the environment or context, and/or the people in it, and information provided by participants (Dewalt, Dewalt, & Wayland, 1998; Pelto & Pelto, 1978). Much has been written on how to write good field notes (Emerson, Fretz, & Shaw, 1995). Some brief, general guidelines are as follows: First, and most importantly, ethnographers must write down notes promptly, and not rely on memory. Without the note, there is no data point (Dewalt et al., 1998; Bernard, 1988; Pelto & Pelto, 1978). Field notes should be written as close to the time that the data are collected as possible to prevent memory loss. Notes should be taken throughout the day or course of observation, and not only at the end of the day. Second, it is important that ethnographers exercise discretion about note-taking while undertaking participant observation. Not infrequently, observations of emotionally laden interactions common in sensitive situations may preclude taking notes as an event unfolds. Study participants and others in the field may perceive the researcher’s note-taking as intrusive. Note-taking may also disrupt the natural flow of events. When taking notes is not appropriate, writing brief quotes, key words, or short descriptions immediately after observing events, interactions, and discussions is important to optimize proper recall and documentation.

Third, ethnographers should record at a low level of abstraction. This means that researchers should note concrete details regarding the actions, interactions, communication, and nonverbal communication that they observed. Furthermore, they need to document the specific observations that led to their impression about a cultural process or pattern, such as emotional events transpiring during the observation. For example, researchers studying communication between health professionals and patients would benefit by noting nonverbal expressions, the physical proxemics, and relative positions between social actors, such as facial expressions or crying, in order to later interpret and make sense of these expressions in light of cultural meanings and expectations in the clinical setting. Otherwise, taking note that a patient expressed “agreement” (an abstraction based on one’s own cultural lens) may be grossly inaccurate since, for example, for many Asian patients, nodding is a form of respect rather than agreement (McLaughlin & Braun, 1998).

Fourth, writing field notes does not necessarily mean writing down every bit of minutia observed, but should be directed toward a certain focus, depending on one’s theoretical orientation and study aim (Pelto & Pelto, 1978). For instance, noting the color and length of physicians’ jackets/coats may be important for a study of the symbolic power of physicians and
interactions based on their status, but would be moot if the focus is on language used during doctor–patient communication during telephone consults. Further, it is common for researchers to jot down key phrases heard, rather than noting entire sentences word for word, to stay abreast of observations without getting bogged down with the logistics of recording. At the end of a day of data collection, ethnographers should fill in the details of their field notes and write fuller notes by reviewing their jottings, reflecting on the events observed. Notes should also record steps in the development of the ethnographer’s analysis. If an ethnographer is collaborating with others, then debriefing with mentors or colleagues can trigger memories of observations and should also be recorded.

Fifth, it is important for ethnographers to recognize that their notes are not objective representations of events, but rather constructions infused with their own interpretation and analysis (Dewalt et al., 1998). Accordingly, there is a debate within anthropology about the use of one or more sets of field notes to record: (a) log of field activities, (b) observations and information, (c) analytic interpretations, and (d) personal perceptions and experiences to aid in self-reflexivity. Some anthropologists recommend that ethnographers keep a separate journal or diary to record personal experiences in line with the notion of self-reflexivity. One way to keep check on biases in field notes is to write reflexively, staying as fully attuned as possible and documenting how the events observed make the ethnographer feel and respond. However, others contend that such experiences are themselves data which reveal much about the difference between the culture observed and the ethnographer’s own cultural framework. Thus, such notes yield a form of cross-cultural comparison that should be kept with notes on observations of the setting.

In addition to field notes, ethnographers may collect data from interviews and surveys which may or may not be audio-recorded. Even if they are audio-recorded, it is important to take notes during interviews as a backup in case recordings fail, and to write notes after the interview to provide greater context with details. In many clinical settings, health care professionals have expressed great reluctance to have their interactions and meetings tape recorded, likely owing to the American cultural context of litigation as well as to confidentiality considerations.

The Management of Data

The collection of extensive data requires that researchers manage them systematically. This means establishing proper storage and retrieval systems
(Huberman & Miles, 1994). Most data can be entered into computer databases. Software for storing and analyzing qualitative data are readily available e.g., QSR NUD*IST, THE ETHNOGRAPH, ASKASAM, NVIVO, and ATLAS-TI. Some of these programs even include statistical capacities. Additional ways to manage and protect data entail using a physical filing system, dubbing audiotapes and transcribing tape recordings, making photocopies of all documents, and backing up electronic data on a regular basis. An important facet of data management involves the development of a codebook ensuring that data are managed in a systematic manner. This fosters clarity and enhances consistency in data collection, interpretation, and coding procedures. Codebook development, data collection, coding, and analysis are not linear but inform each other and are refined over the course of research. (For more information about these aspects of qualitative data management, see chapters on Content Analysis and Semi-Structured Interviews).

THE ANALYSIS AND INTERPRETATION OF DATA

Social scientists have developed many methods of inductive analysis. Excellent sourcebooks on how to conduct qualitative data analysis are available (Miles & Huberman, 1994; Strauss, 1987), and some of these techniques are discussed further in other chapters of this book. Here, we briefly define some analytic approaches that are commonly used in or suited to ethnographic research in bioethics: grounded social theory, componential analysis, and content analysis.

Grounded Theory

Iterative procedures have always been the norm in the practice of ethnography. While ethnographers are in the field they begin to interpret their data. Subsequent data collection is guided by earlier observations and endeavors to develop more nuanced understandings of the phenomena under study. Efforts are taken to reject or elaborate earlier interpretations and develop better explanations. In the 1960s, sociologists Glaser and Strauss (1967) formally developed a rationale for this approach and showed how such interpretations grounded in data could be used to develop theory. They labeled their approach “Grounded Theory” and called for rigorous attention to the process used by researchers to move between data and
analysis. In particular, they advocated the use of a technique they called the "constant comparative method" in which the ethnographer would use later observations to verify or modify previously suggested theories (Strauss & Corbin, 1994). According to Strauss and Corbin, an important feature that distinguishes grounded theory methodology from other more descriptive inductive approaches is that sampling, questioning, and coding are all explicitly theoretically informed and chosen with the intention of explaining the relationships between concepts and patterns of social behavior. They wrote:

“In doing our analyses, we conceptualize and classify events, acts and outcomes. The categories that emerge, along with their relationships, are the foundations for our developing theory. This abstracting, reducing, and relating is what makes the difference between theoretical and descriptive coding (or theory building and doing description)” (Strauss & Corbin, 1998, p. 66).

It should be made clear that grounded theory describes a broader approach for handling the analysis of data, but does not provide guidance on specific cognitive elements to focus on for analysis, as does componential analysis, described below.

Componential Analysis

Although componential analysis has not been employed much in bioethics research, it could be especially useful. Componential analysis, which is also referred to as semantic analysis, entails two goals: (1) describing how members of a cultural group categorize a given meaningful, culturally valid, behavioral issue from an emic perspective and (2) delineating the cognitive processes, components of meaning, or criteria that cultural members use to distinguish between cultural categories (Bernard, 1988; Pelto & Pelto, 1978). This process entails charting out, classifying, and contrasting semantic networks of emic constructs among group members to understand how they view an issue under study. Anthropologist Edward Sapir hypothesized that language is a reflection of a cultural group’s conception of the world, which is often tacitly embedded in how people classify the things in their world (Mandelbaum, 1949). Accordingly, identifying how people classify, for example, kinds of personhood, the meanings attached to the concepts of person, and the rationale for the classification system, may yield insight into the moral underpinnings for decisions made about medical treatment.

Research in bioethics has used techniques derived from sociolinguistics. These entail the study of discourse as it naturally occurs in terms of its content and structure. For example, bioethics researchers have inquired into
the multiple meanings of constructs like “fairness” in allocation of primary health care services (Jecker & Berg, 1992) and of “do-not-resuscitate orders” for different social and cultural groups (Muller, 1992).

Content Analysis
A commonly used technique for analyzing data is content analysis. Content analysis is a process of analyzing data by systematically searching for themes and repetitions emergent from the data (Luborsky, 1994; Huberman & Miles, 1994). The first step is to prepare data for analysis – whether that means pulling out descriptions of cases, organizing hand-written and word-processed field notes systematically, transcribing audiotapes verbatim, or coding data for use with a computer program. Content analysis is an iterative process whereby themes are developed by grouping coded segments into schema of larger domains, followed by a review of the categorization schema for appropriate thematic fit, and adjusting and reviewing the schema again until the researcher(s) reach consensus. Traditionally, ethnographers coded data by hand, e.g., by highlighting or marking portions of handwritten or typed text, or by sorting index cards. Although many researchers continue to code by hand, there are now many qualitative data analysis software programs available, as noted above. (For a detailed discussion, see the chapter on Content Analysis in this volume).

Validity Checking
Ethnographers use numerous techniques to enhance the internal validity of their research, that is, “the extent to which it gives the correct answer” (Kirk & Miller, 1986, p. 19; Patton, 1999). To increase the validity of the data, ethnographers use multiple techniques including thick description, triangulation, member checking, using paradigm cases, and maintaining a detailed paper trail. These will be discussed further below.

Thick description entails depicting phenomena in rich detail (Geertz, 1973). Geertz likened culture to “an assemblage of texts” (1973, p. 448) with socially shared and generated meanings that can be interpreted. The ethnographer’s analytic goal is to uncover and understand the meanings woven throughout his/her field notes. Detailed accounts in field notes help to increase the likelihood that interpretations remain data-driven and are truly inductive accounts of the cultural group or other phenomena under study.

Triangulation involves the use of several kinds of methods or sources of information to obtain overlapping data which helps to support reliability and
validity of findings. There are different kinds of triangulation: (1) investigator triangulation which involves several researchers collecting data; (2) theory triangulation which entails using multiple perspectives to interpret data; (3) methodological triangulation which uses multiple data collection methods and strategies; and (4) interdisciplinary triangulation involving the inclusion of other disciplines to inform the research process (Crano, 1981).

**Member checking** involves asking informants to evaluate aspects of the researcher’s analysis to find out if interpretations are analytically sound. It can take several forms such as asking informants how and why cases are categorized in a certain way, or asking informants to review a written description of an aspect of their culture. For example, after having observed meetings during which transplant professionals made decisions about placing patients on the waiting list, I used member checking by having a transplant surgeon review a draft of my manuscript for accuracy, correct technical details, and to obtain clarification (Gordon, 2000).

Looking for “paradigm” cases generally involves researchers identifying “typical” or representative cases to illustrate situations in which certain norms apply. Ethnographers may also select anomalous or contradictory cases which do not conform to the dominant pattern. Study of these exceptional cases can help the ethnographer to differentiate between those conditions in which the cultural or ethical norms apply while simultaneously help to explain limits of application; this may reveal the “rules for breaking the rules.” In a study of perceptions about truth-telling by members of different ethnic groups, Blackhall, Frank, and Murphy (2001) deliberately used this technique by seeking further interviews with at least two respondents whose responses to an initial survey were atypical in order to obtain insight into the diversity within groups.

Finally, maintaining a detailed and accurate **paper trail** is another useful form of validity checking (Pelto & Pelto, 1978). This can be fostered by ensuring that dates and decision points are written on all data materials to provide a chronological account of the development of interpretations while data are being collected.

**ETHICAL CONSIDERATIONS**

Conducting ethnographic research raises several ethical concerns and challenges. Earlier in this chapter, we discussed informed consent and the lack of familiarity IRBs frequently have regarding ethnographic methods. Detailed accounts of the problems and proposed solutions to working with
IRBs regarding ethnography can be found in Gordon (2003), Marshall (2003), and Koenig et al. (2003).

Other ethical challenges in ethnographic research pertain to “studying up,” studying one’s colleagues, confidentiality, and the questions about whether to intervene in ethically troublesome situations. Confidentiality is a major concern for ethnographers who frequently live or work intimately with those they are studying. Informants may see their roles and lives described in detail or in unanticipated ways that they feel can harm their reputations (Bosk, 2001). To protect confidentiality, most ethnographers try to conceal or mask the exact location of the work and identities of the participants. Part of the problem with good ethnography is that it “penetrate[s] deeply enough into the social world being described,” and “make[s] the latent manifest” (Bosk, 2001, p. 209). Fieldworkers can thereby make subjects uncomfortable (Bosk, 2001). Moreover, informants may feel betrayed when ethnographers leave the field after they have developed friendships or paid concerted attention to them over time. Respect for study participants and their culture can be demonstrated by sharing results with them and/or having them review interim findings during the study.

Care must be taken so that individuals in the group being studied cannot be identified in written reports. One way to circumvent this problem is to write composite cases that incorporate aspects from multiple cases into one. Additionally, ethnographers commonly use pseudonyms for participants and anonymize the research site. It is important to convey to participants that all data will be kept confidential and that their names will not be used in publications or presentations.

Ethnographers in their research can become privy to many kinds of behaviors and discussions, including illicit and unethical behavior. This raises the question of whether the ethnographer should intervene. Intervening could help those in need. At the same time, it can potentially jeopardize the quality and reliability of data, as well as impact relationships with key informants and those providing access to the research setting. Anthropologists have examined this issue as part of the ethical standards for professional practice (URL: [http://www.aaanet.org/stmts/irb.htm](http://www.aaanet.org/stmts/irb.htm) accessed 2-28-07).

**SUMMARY**

Ethnography is a qualitative research method based predominantly on participant observation; this is usually supplemented by the use of interviews, surveys, and document analysis. The concept of culture is
central – the ethnographer seeks to articulate how the members of the culture(s) being studied themselves understand phenomena (emic perspective), and to analyze (etic perspective) the complex set of historical and contemporary forces that shape beliefs and behavior. Ethnography is best learned through experience, reading ethnographies, engaging in discussions with people who have done them, and doing smaller or pilot studies, rather than through purely didactic training. Accordingly, ethnography is a difficult research method to teach. We emphasize that ethnography is more than just observing a situation. One must know what to look for, how to make observations through the lens of the complex concept of culture, and how to interpret and analyze data in light of the social, historical, and cultural contexts in which data are collected. In the following, we discuss a number of strengths and weaknesses of ethnography.

Strengths

Ethnography offers much strength as a research method in bioethics. Foremost, ethnography enables researchers to contextualize bioethical issues in their broader social, historic, economic, political, ideological, and cultural contexts. Ethnography provides insight by uncovering elements of a culture that other methods cannot provide because of the tacit nature of culture. It also provides unique insights because it focuses on natural behaviors occurring in specific settings. It collects data on both what people say they do and what the ethnographer observes them doing. Ethnography does not depend on a rigid research plan; therefore, it can detect unanticipated phenomena. The sensitivity ethnography brings to examining subtlety and meaning renders it “an ideal vehicle for examining normative language or decision making” (DeVries & Conrad, 1998, p. 248). Ethnography is valuable for showing “how flexible values are, how the same values are used to justify a wide range of seemingly incompatible behaviors” (Bosk, 2001, p. 200). In addition, ethnography enables researchers to understand the complexity of a phenomenon. Since one of the hallmarks of ethnography is its flexibility, investigators are able to follow up leads that arise unexpectedly during the research, adjust techniques, and modify research questions as unforeseen information emerges to more thoroughly investigate a topic (Drought & Koenig, 2002; Briggs, 1970). Moreover, ethnography lends itself to informing other research methods. Specifically, it can help generate hypotheses that can be tested using other qualitative or quantitative methods, and interpret the findings of such research.
Weaknesses

A major weakness of ethnography is its time consuming nature. Most studies last at least one year, frequently longer. Thus, ethnography is expensive and resource intensive. Another weakness that critics are quick to point out is that it is subjective in data collection and analysis. By using the ethnographer as a primary tool for data collection, personal biases may enter unchecked into the data collection and analysis processes. Another weakness is the limited ability to generalize study findings.

Other weaknesses pertain to the process of conducting ethnography itself. Ethnography relies, to a great extent, on luck. It is unknown what kinds of cases or events will emerge over the course of the research period, and ethnographers are essentially dependent on the luck of the draw. Moreover, levels of cooperation by members of the group under study with the researcher may vary unexpectedly. Research in clinical settings may also be affected by turnover in medical and administrative staff each month and by changes in nursing staff requiring re-negotiation throughout a study.

ACKNOWLEDGMENTS

This work was supported by grant DK063953 from the National Institute of Diabetes and Digestive and Kidney Diseases (EJG). We thank Becky Codner and Elizabeth Schilling for their research assistance, and Laura Siminoff Ph.D. and Liva Jacoby, Ph.D. for their helpful suggestions with this manuscript.

REFERENCES


Contextualizing Ethical Dilemmas: Ethnography for Bioethics


SEMI-STRUCTURED INTERVIEWS IN BIOETHICS RESEARCH

Pamela Sankar and Nora L. Jones

ABSTRACT

In this chapter, we present semi-structured interviewing as an adaptable method useful in bioethics research to gather data for issues of concern to researchers in the field. We discuss the theory and practice behind developing the interview guide, the logistics of managing a semi-structured interview-based research project, developing and applying a codebook, and data analysis. Throughout the chapter we use examples from empirical bioethics literature.

INTRODUCTION

Interviewing provides an adaptable and reliable means to gather the kind of data needed to conduct empirical bioethics research. There are many kinds of interviews to choose among, and a primary feature that distinguishes them is the degree of standardization imposed on the exchange between interviewer and respondent. Semi-structured interviews, as their name suggests, integrate structured and unstructured exchanges. They rely on a fixed set of questions but ask respondents to answer in their own words, and they allow the interviewer to prompt for a more detailed answer or for...
clarification. Such interviews are more structured than in-depth or ethnographic interviews that explore topics by following and probing a respondent’s largely self-directed account. But they are less structured than typical surveys that administer the same questions to all subjects and confine responses to a fixed set of choices. Semi-structured interviews combine the advantage of closed-ended questions, which allow for comparison across subjects, and the advantage of open-ended queries, which facilitate exploring individual stories in-depth. Similarly, semi-structured interview data is more flexible than data generated from strictly quantitative studies. Analysis of semi-structured interview data can run the spectrum from the statistical to the descriptive and in-depth. This combination of types of analyses contributes to the strength and wide-ranging applicability of semi-structured interviewing.

Researchers can use semi-structured interviews for many types of inquiry, including confirming existing research results or opening up new fields of inquiry, but the method is particularly well suited to exploratory research or to very focused theory-testing. This is the case because the effort entailed in conducting good semi-structured interviews and then analyzing the considerable data that they produce work against the large sample sizes required to examine certain kinds of population-based questions characteristic of epidemiological and some sociological studies. While semi-structured interview results can be analyzed quantitatively, this is not their primary strength. Their main contribution lies in the richness of the data they provide, a strength that is sacrificed when data are reduced to numerical values.

In bioethics, semi-structured interviews have been used effectively to examine several topics, including genetic testing, end of life care, medical confidentiality, and informed consent. Genetic testing studies demonstrate the suitability of semi-structured interviews to explore new areas of ethical concern. For example, studies relying on semi-structured interviews have been instrumental in elucidating the challenge of successfully educating patients about the complexities of BRCA1/2 testing (Press, Yasui, Reynolds, Durfy, & Burke, 2001), examining the special circumstances of some patient groups, such as men who test positive for the gene (Hallowell et al., 2005; Liede et al., 2000), and studying how families communicate about test results (Claes et al., 2003; Forrest et al., 2003). Research on end of life care demonstrates the advantages of semi-structured interviews when examining sensitive topics such as how patients and families would like clinicians to answer questions about physician-assisted suicide (PAS) (Back et al., 2002) or what factors influence a terminally ill patient’s wish to hasten
death (Kelly et al., 2002). Another example of exploring sensitive topics with semi-structured interviews is the authors’ study of patients’ deliberations to tell or not tell their physicians about medical problems they consider sensitive. Using semi-structured interviews, we found that patients withhold information about a wider range of topics than practitioners might expect (Sankar & Jones, 2005). Semi-structured interviews also have been productively used to study situations distinguished by complex or confusing communication (Featherstone & Donovan, 1998; Stevens & Ahmedzai, 2004; Pang, 1994).

**THE INTERVIEW PROCESS**

*Developing the Interview Guide*

Semi-structured interviews rely on an interview guide that includes questions, possible prompts, and notes to the interviewer about how to handle certain responses. Constructing the interview guide derives from the study’s research question(s) and begins with reviewing the literature on a chosen topic to see how others have approached similar inquiries. Working back and forth between the previous research – what has been done – and the draft interview guide – what could be done – helps to clarify the aims and hypotheses of a study and brings into focus possible questions to include in the interview guide. The guide should be designed to balance the need to collect roughly similar kinds of information from all subjects while capturing each subject’s unique perspective on that information. Novices to interview guide development might consider reading existing interview guides available from colleagues or those posted online associated with published articles (see Box 1 for an excerpt of the interview guide we developed for our study on medical confidentiality). Several excellent articles also offer suggestions on how to formulate interview questions (Britten, 1995; Fowler, 1992; Mishler, 1986a, 1986b; Sanchez, 1992; Weiss, 1994). Here we offer some pointers for crafting and asking semi-structured interview questions, as well as suggestions for helping subjects answer those questions (Box 2). Reading draft questions aloud or trying them on colleagues can be useful. Very often what makes sense to the author eludes the listener.

The order of questions in an interview guide merits careful consideration. Topics lend themselves to different schemes, such as chronological, or by degrees of complexity or intrusiveness. The interview guide for our medical
confidentiality study, for example, started with questions about how subjects thought medical confidentiality operated in a clinic setting, and then asked them how they thought it should operate. Only after 30 or 40 min of exchange did we ask respondents to recount their own personal experiences related to medical confidentiality (Jenkins, Merz, & Sankar, 2005; Sankar & Jones, 2005).

Once a rough draft of the interview guide is complete, it is essential to try it out, start to finish, with a colleague to see if the questions make sense and flow sensibly from one to the next. Subsequent to this, it is advisable to revise the guide and try it on members of the study population while taping these pilot interviews, and if possible take notes while conducting them. Asking subjects to explain if and why a question seems unclear is important. Transcribing these interviews is not necessary, although if working in a large research team, transcribing allows everyone to review the results and facilitates revising the guide. The goal is to create an interview guide that is comprehensible to respondents and that evokes responses which answer the questions the research asks. While this goal might seem self-evident, achieving it can be far more difficult than anticipated. Absent careful review of pilot interview transcripts, the insight that the interview failed to ask the

---

**Box 1.** The Interview Guide: A Brief Example.

7. **DEFINE “CONF”:** If you had to explain to a friend what “confidentiality” means, what would you tell her it means?
   7a. So then, what part of the information that you talk about with your doctor or nurse is “use respondent’s wording, e.g., ‘just between a person and her doctor’”?
   7b. Does this confidential information go into your medical record?

[SKIP 8 & 9 IF ALREADY ANSWERED IN 7]

8. **RELEASE?**: Do you think there are any situations where your doctor or nurse would decide NOT to keep confidential information use respondent’s wording, e.g., ‘just between a person and her doctor’ – an give out or release your confidential medical information?

   a. NO  b. YES

9. Could you give me examples of this kind of situation?
Box 2. Asking Questions: Some Pointers.

- Tense and Specificity. The verb tense of a question influences its likely responses
  - Questions asked in the present tense, for example, “What happens when you call the doctor?” tend to elicit generalized accounts.
  - Conversely, questions phrased in the past tense, such as “What happened the last time you called the doctor?” will tend to elicit specific actions and experiences.

- Phrasing
  - Avoid phrasing that allows for yes/no answers. Questions beginning with what or how are better at eliciting fuller answers than those starting with did, as in “Did you ask the doctor any questions?”
  - Ask only one question at a time. If an interviewer asks, “What did you do after your diagnosis and how did you feel about that?” respondents can get caught up in answering the second part of the question and be unable or unwilling to go back and provide a complete answer to the first part.
  - Often the best follow-up question consists of repeating back the last phrase of the subject’s previous statement, as in, “…you called her after dinner …” when the subject has stated, “I couldn’t reach the doctor all day so I had to call her after dinner.” This response indicates that you are listening and encourages the subject to continue talking without directing him or her toward any particular response.

- Helping respondents answer your questions
  - Extending: Can you tell me more about how you met her?
  - Filling in detail: Can you walk me through that? So were talking to him after the phone call?
  - Making indications explicit: when subjects nod or sigh, try to confirm verbally what they indicated. “So would you say that was good news?” or “Tell me what you’re thinking when you sigh like that.” Or, most directly, “Can you say what you’re thinking for the tape recorder.”
“right” question may not emerge until all the data are collected and the opportunity to revise the interview guide long past. Reviewing the interview guide through extensive pilot testing should also mitigate the need to revise questions during the data collection phase of a project, a common remedy for poorly phrased questions that complicates data analysis and reduces internal validity.

Formulating, testing, reviewing, and revising questions so that they successfully communicate to the subject what interests the researcher is roughly equivalent to the steps that quantitative researchers go through in scale development when they “validate” an instrument. Validity refers generally to the relationship between measurements (such as scales or questions) and the phenomenon they claim to capture or measure. The closer that relationship is the more “valid” a measure.

There are several types of validity, including internal, external, and construct validity. For a useful review of these concepts written for qualitative researchers, see Maxwell (1996). A major difference between the import of these ideas for qualitative and quantitative research is the degree of specificity or formalism in assessing validity. The latter is standard in quantitative research. The logical problems posed by validity concerns, however, are equally salient in qualitative research. For example, in question development for semi-structured interviews, construct validity focuses attention on assessing whether the question asked is the question answered. In our medical confidentiality study for example, we struggled with devising a question to determine how subjects understood medical confidentiality. Asking them directly, “What does medical confidentiality mean?” seemed to suggest a test question with a right/wrong answer. Subjects interpreted the question more like “What is the definition of medical confidentiality?” and responded with their approximation of a dictionary definition. The question they answered was not the question we meant to ask. We went through several iterations, including versions that added “to you” at the end of the original question, and another that asked subjects what it meant if a doctor or nurse told them they would keep something confidential (which elicited responses about the doctor or nurse’s character, such as “It would mean they cared about me”). Finally we arrived at the solution of asking the subject to tell us how she would explain to a friend what medical confidentiality meant. These questions elicited answers such as, “I would tell her that it means keeping a secret. That whatever I tell a doctor goes to no one else.” These were the kind of responses we were seeking, that is, what patients think it means in a medical setting when a health care practitioner characterizes an exchange as “confidential.” Formulating, testing, and
reformulating questions brought the query we posed and the phenomenon we were trying to measure closer and closer into alignment.

Piloting the interview guide will require finalizing most of the bureaucratic and practical steps that actual interviews require. This includes completing all necessary human subjects research review, and deciding who will be interviewed, how they will be contacted, and where the interviews will be conducted. While such complete preparation for piloting might seem onerous and unnecessary, every moment spent testing the interview guide and refining recruitment and other logistics before the study’s full implementation will pay off by revealing obstacles that eventually might have delayed the project.

**Sampling: Choosing Whom to Interview**

The research question will determine the population to be sampled for interviews. How to approach the population and solicit its members for participation needs to be worked out carefully in advance based on familiarity with the targeted group and discussion with some of its members. The best designed study can founder on unanticipated obstacles in subject recruitment. If the target group is large and loosely defined, such as women who have had mammograms, the issues are quite different than if research calls for talking with a small or more isolated group, such as people who have undergone genetic testing for hearing loss. Abundance might pose a sampling problem in the former – which women who have had mammograms are of interest? Old, young, insured, uninsured, rural, or urban? Scarcity presents a challenge in the latter. How does one find people who have had genetic testing, who are few in number, geographically dispersed and, if themselves deaf or hard of hearing, who might not be reachable by methods typically used by the hearing research community?

A method often relied on in qualitative research is snowball sampling, in which potential participants are identified by asking initial subjects to suggest names of other people who might be interested in participating in the study. The resulting sample is likely to consist of people who know one another, which also means it might represent a fairly narrowly defined group. Inferences from such a study cannot be generalized much beyond the subjects interviewed or individuals deemed highly similar to them. While this might be acceptable for studies examining very specific issues confined to a well-defined group, it is less effective for more broadly relevant issues. Another common strategy for identifying potential participants is to use
clinic staff or other gatekeepers to estimate the flow of people through a recruitment site. Relying solely on this information can be a mistake. Overestimating is common, as is failure to recall seasonal fluctuations. Researchers would do well to review any sort of available records that might document actual numbers of individuals or groups pertinent to the research.

Estimating how many and what kinds subjects to enroll requires knowing what kind of inferences you plan to make from findings. The way one estimates and selects a sample will directly influence what inferences one can make from the data, and these sampling decisions will have to be explained in any publication that results from one’s research. There are several excellent articles on sampling for semi-structured interviews (Creswell, 1994; Curtis, Gesler, Smith, & Washburn, 2000; Marshall, 1996; Morse, 1991; Schensul, Schensul, & LeCompte, 1999). In general, sampling should control as much as possible for subject bias. Biased sampling occurs when researchers assume that the perspective of one group of potential respondents is representative of a more inclusive group, for example, using race or ethnicity as proxy for a wider range of socioeconomic and sociodemographic variables. A study focused on minority perspectives on genetic testing that only sampled northern urban African Americans would suffer from such sample bias, if it purported to speak for a general US minority population that is more diverse than the group sampled.

Finally, in studies focused on the beliefs, attitudes, and behaviors of a narrowly defined group of people, such as oncology clinic patients, or about a specific topic, such as attitudes about hospice care, the question of the exact number of subjects may be left open until after a few rounds of preliminary coding. Some researchers decide to keep recruiting subjects until a point of theoretical saturation, meaning that new interviews begin to be redundant and are not adding new substantive issues (Eliott & Olver, 2003).

Conducting the Interview

The setting for the interview should be comfortable and foster an open exchange, which often, but not necessarily, means a private room. Regardless of where the interview is conducted, it is important to make sure the space is available for the entire time needed to conduct the interview. Completing the interview in a timely manner shows respect to the participant and minimizes scheduling problems.

Regardless of what is explained when arranging to meet the respondent, he/she needs to be re-oriented of the study objectives. It is important to
explain what in his/her background makes him or her able to contribute to the research. Another essential strategy is to briefly review the aim(s) of the study, topics of the interview and the order in which they will appear, and to underscore that unlike other interviews or surveys that ask for simple yes or no answers, this type of interview emphasizes learning what respondents consider important about the research topic and relies on them to provide answers in their own words.

Semi-structured interviews allow a high degree of flexibility not only to the respondent but to the interviewer as well. A question mistakenly asked so that it can be answered with “yes” or “no” when the objective was to obtain a detailed response, can simply be re-stated. Having gone to the trouble to start the interview, respondents usually share the researcher’s interest in successfully completing it.

Technicalities of Data Collection and Transcription

There are two primary reasons why audiotaping or digitally recording interviews is preferable to note taking: memory and foresight. As semi-structured interviews can be very lengthy, it is highly unlikely that the interviewer can remember the full story behind the abbreviated notes jotted down during an interesting story early in the interview. Also, as research progresses, new patterns in the interviews emerge and new questions can be asked about the data. At the time of the interview, these new issues will not catch the interviewer’s attention and will not be documented, or not be documented fully enough to permit subsequent detailed analysis.

To assure high quality recordings and to avoid contributing to the overflowing stock of “interviews that got away” stories, one should obtain reliable equipment, bring several extra batteries and cassettes, and test the recorder before each interview. After obtaining informed consent, it is useful to establish the respondent’s permission to record the interview and have the respondent and the interviewer both speak while the recorder is in the exact position it will be in during the interview. It is also important to play back this recording and check for levels and intelligibility. It is not advisable to use the “voice activation” option as words or whole sentences can be cut-off. A good idea is to glance at the recorder every 10 min or so to make sure it is still functioning properly. Moreover, it is advisable to use one tape for each interview and not taping the next interview on the same cassette, as it is too easy to mix up the sides and tape over the previous interview. Cassettes can be reused after transcribing a set of interviews, so this practice need not
be wasteful. More and more, researchers in the field are switching to digital recording to eliminate tapes altogether, but it is important to practice a few times with any new technology before relying on it in an interview.

Finally, it is essential to carefully label interview cassettes with an appropriate identification code and arrange for transcription. With respect to transcription, every hour of talk takes upwards of 3–4 h to transcribe. The task is tedious and best left to a professional transcriptionist if the project can afford one, although in smaller studies transcribing interviews can provide a good way to review data. Whether analysis will rely on a software program or pen and paper coding, it will proceed more smoothly with established transcription guidelines to help assure consistency across transcripts and to facilitate a review of the documents for errors or to remove any identifying personal information.

Creating a data preparation chart to track the progress of individual interviews helps to organize interview tasks. It is important to keep any information that might identify a subject out of a data preparation chart and to create instead a code that links the data from this chart to other data sets containing subject contact information. A data preparation chart might contain fields that track an interview from its occurrence (including the date, time, place, and initials of the person who conducted an interview) to its completion as a fully prepared document ready for analysis (including the date the cassette was sent and received back for transcription, checks on transcript accuracy, reviews to remove personally identifying information, formatting text for database importing, and finally, the cassette’s erasure).

CODING AND DATA ANALYSIS

Researchers opt for semi-structured interviews because they provide more and more nuanced data than close-ended questions. However, the amount and kind of data that such interviews can produce will also greatly increase data management and analysis tasks. Conducting fewer interviews might cut down on these burdens, but analyzing any amount of unstructured text inevitably increases the work required in the post-data collection phase of research.

Much analysis of semi-structured interview data is the same irrespective of the number of interviews conducted or the means of analysis (note cards or qualitative data analysis software programs). However, here we assume that study data will be entered into a qualitative data analysis program and that somewhere between 30 and 100 interviews have been conducted. This
range is somewhat arbitrary, but reflects our experience that studies with fewer than 30 subjects are less likely to require the team coding process described here, and those that exceed 100 are likely to benefit from primarily quantitative analysis, which we address only in passing. There are examples of studies with larger sample sizes using semi-structured interviews (Kass et al., 2004; Mechanic & Meyer, 2000), but these studies are atypical, either in their lengthy duration (extending over many years) or in their reliance on a high proportion of close-ended questions, which minimizes coding burden.

Coding is the process of mapping interview transcripts so that patterns in the data can be identified, retrieved, and analyzed. In other words, coding provides a means to gain access to interview passages or data, rather than becoming the data itself. Unlike coding survey responses for quantitative analysis, which requires reducing responses to numeric values, the goal of coding semi-structured interview transcripts is to index the data to facilitate its retrieval, while retaining the context in which data was originally identified.

Researchers should investigate different qualitative software packages before beginning the codebook development process, as each program has different systems for coding. Common programs worth exploring include N6 (2002), ATLAS.ti (Muhr, 2004) and HyperResearch (2003). See the “Computer Assisted Qualitative Data Analysis (CAQDAS)” website for software comparisons and tips for choosing a qualitative software program (www.caqdas.soc.surrey.ac.uk/).

**Codes**

Codes classify and represent relevant passages in the interview. Devising codes that accomplish these tasks requires identifying the central themes in the research, developing concepts to represent them, and assessing how these concepts relate to each other and to overall research questions that motivate the study. Developing and defining codes thus constitutes the first stage in data analysis, which means that developing the coding manual requires careful attention. Coding itself consists of reading transcripts and deciding how and where to apply codes. Ideally the coding manual contains codes that are sufficiently well suited to the data and well enough defined that the process of coding, while unlikely to be automatic, does not require continually rethinking the meaning or limits of codes. There are several types of codes, which can be thought of as roughly paralleling the stages of coding, or at a minimum, the stages of developing a coding manual.
The first stage of developing a coding manual consists of listing codes that cover information that the interview explicitly sought. For example, in our study on confidentiality, we asked women how they would explain medical confidentiality to a friend. All of the passages containing answers to this question were assigned the code “DEF CONFI,” which stood for “definition of confidentiality.” Many codes of this type need to be divided into sub-codes. Sometimes sub-codes are logically suggested by the original code, as one might expect a code such as “family member” to have the following sub-codes: “spouse,” “child,” and “parent.” Other times sub-codes are needed to represent distinctions that emerge from the data.

Subdividing codes based on emergent data moves the focus from codes that are implied in the interview guide to codes that capture themes that the interviews have elicited. Thematic coding identifies recurrent ideas that are implicit in the data. Some of these codes might be envisioned in advance of the interviews. More typically, however, thematic coding is held out as the means for capturing what one learns or discovers when examining the interviews together as a completed set. In our confidentiality study, repeated transcript readings and examinations of the primary code “DEF CONFI” revealed that respondents often ground their definitions either in the personal relationships between patient and practitioner or in bureaucratic procedures required to safely store and accurately transmit sensitive information. Two thematic sub-codes of “DEF CONFI,” “personal” and “bureaucratic,” captured these recurrent ideas (Jenkins et al., 2005). Also, further analysis determined that these themes connected with many other features emblematic of different ways of understanding and using medical confidentiality.

A similar staged coding process was used by researchers studying patients’ interactions with their physicians regarding PAS (Back et al., 2002). The first level of primary codes included “interactions with health care providers,” “reasons for pursuing PAS,” and “planning for death.” Further examination of the text coded to the primary code “interactions with health care providers” revealed three distinctive themes, which became three individual sub-codes: “explicit PAS discussion,” “clinician willingness to discuss dying,” and “clinician empathy.” Through this process, researchers were able to see what their subjects valued in their clinician’s responses, or what they wanted but did not get.

The provisional coding manual provides draft definitions of the codes, which are tested by applying them to a set of completed interviews.
The experience of trying to fit draft codes to actual interviews clarifies which codes effectively capture responses and which do not, which might with revision, and what new codes should be tested. This process is analogous to the work required to develop interview questions that accurately express the researcher’s interests in that the task is one of proposing, matching, and testing representation to concepts. Testing draft codes also helps to identify example passages from the interviews that researchers can incorporate in the coding manual to demonstrate a code’s definition or limits.

After a test-round of coding, a revised coding manual is generated and applied to a second set of interviews that preferably includes some from the prior coding round as well as new ones. The second round continues largely in the manner of the first. Codes are evaluated and revised, and their definitions revised in parallel. This process continues until a codebook emerges that contains a set of codes judged to have sufficient range to effectively represent the material of interest to researchers and sufficient clarity to be consistently applied by coders (see Box 3 for an excerpted section of the codebook used in our confidentiality study).

Multi-Level Consensus Coding (MLCC)

With studies based on 30 or more interviews, coding typically requires, and nearly always benefits from, more than one coder. We have developed an approach to coding large semi-structured interview data sets called multi-level consensus coding (MLCC) (Jenkins et al., 2005). MLCC meets the challenge of analyzing a large number of open-ended qualitative interview transcripts by generating a high degree of inter-coder reliability and comparability among interviews, while maintaining the singular nature of interviewee’s experiences. This coding model depends on the use of a code and retrieval software program, in this case, N6, that aids in the analysis of qualitative data through flexible coding index systems and searching protocols that identify thematic patterns, essential in the mapping of qualitative knowledge.

MLCC differs from other coding strategies described in most standard methods texts in two ways. First, it makes explicit the connections between training to code, coding, and analysis by fore-fronting the vital connection between the codes, the coders, and the process of coding. The codes do not exist separately from the coders in the sense that assigning a code (that is, coding) requires coders to have a shared understanding of the relationship between code and text. In turn, having to reach consensus about a code
reproduces and, overtime, changes that understanding. The change in understanding should not be taken to mean the original relationship between the code and its definition and examples was flawed (although certainly that is sometimes the case). Rather, the change in understanding refers to evolving understanding of the phenomenon to which the code was meant to apply and to the capacity of the code to represent it. Second, MLCC formalizes and makes a virtue of the need for several stages of coding. While similar to the types of coding described in grounded theory methods (open, axial, and so on), the levels or stages laid out in MLCC are

**Box 3. Sample Codebook.**

This codebook example with numbering follows the system used in Nud*ist v4 (non-numerical unstructured data indexing searching and theorizing) (see Box 1 to compare the codebook with the interview guide).

(7) Q7 – Def confi
(7 1) Between 2 people/personal
(7 2) Need to know
(7 3) Continuity of care model/bureaucratic model
(7 4) Other
(7 5) Don’t know
(7 6) NA

(8) Q8-release?
(8 1) Yes
(8 2) No

(9) Q9 – Desc Release
(9 1) Happens – w/consent
(9 2) Happens – but not supposed to
(9 3) Happens – continuity of care model
(9 4) Happens – infectious disease
(9 5) Happens – emergency
(9 6) Happens – research
(9 7) Happens – other
(9 8) Happens – doesn’t specify
(9 9) Don’t know
(9 10) NA
less tied to a specific interpretive framework, and are intended more to serve a broader range of projects. Further, it highlights the steps required to code semi-structured interview data as a particular method because as much as the basic steps described here differ little from those that appear in many qualitative methods textbooks, a review of articles claiming to follow qualitative methods suggests that many studies omit or abbreviate the process in ways that seriously undermine the validity of the data and the legitimacy of claims that accepted qualitative data analysis produced it.

MLCC responds to concerns about inter-coder reliability and comparability between interviews through extensive training of coders and through consensus coding, or group decision-making. At the beginning of training coders jointly code two to three transcripts with PI supervision. When the coders have gone through a sufficient number of transcripts so that they are familiar with the content of the interviews and the coding rules, they then move onto “round-robin” coding. A round-robin coding session is a group of four coders (at least one PI among them) who each code three interviews, each interview with a different partner. The discussion session for round-robin coding is an occasion to test individual perceptions of the meaning of coding categories and rules, and an occasion where irregular coding is brought into conformity with the rules. The round-robin training sessions are similar to consensus sessions that coders will participate in during regular coding. Coders then code several more interviews individually and in pairs, mirroring the regular coding method.

Following training the coding teams are divided into pairs, and two coders independently read and code each level of each interview. Coding pairs are rotated so that each coder codes at least a subset of interviews with every other coder, although often certain pairs seem to work better together if only because of scheduling issues. Rotating members of the pair helps to prevent “coding drift,” which occurs over the course of a long coding effort when one or more coders starts to apply codes in an increasingly idiosyncratic manner. Occasional general data checks of all coded interviews can also help catch the wayward evolution of codes. Alternatively, the wayward codes these checks identify sometimes prove to be better than the approved ones. If this happens, the new code can either be added to the coding manual or substituted for an existing code. In either case, changing the coding manual should be done only after careful consideration, as it requires reviewing and amending all completed coding to bring it in line with the new codes.

After each member of the coding pair has coded an interview, they meet to compare individual coding choices and come to agreement on the final
coding for the interview. Printing line and page numbers on the interview transcripts facilitates such crosschecking. Disagreements are discussed in these pairs, and any that the pair is unable to resolve are brought to a weekly consensus meeting, which all pairs of coders, as well as the project manager, attend. Data that cannot be coded after being discussed in these consensus meetings are omitted from analysis.

Using the MLCC method, interviews are coded in three levels or stages. The objective of first-level coding is to generate standard, comparable answers to basic questions asked during each interview. The objective of second-level coding is to generate a series of vignettes characterizing the central experiences, ideas, and issues informing each interviewee’s perceptions. When examined together through N6, first- and second-level coding blend a structured and comparative analytic lens with a nuanced representation of each research participant. Third-level coding uses the first two layers of coding to theorize about the themes that emerge from the interviews.

While requiring considerable organizational effort, the method has several advantages, foremost among them creating a high level of consistency, or validity, across interviews. This process also creates a collective memory of how codes have been applied, a regular forum that catches problem codes earlier than might be the case if coders interact less frequently, and a handful of people intimately acquainted with the data and interested in proposing ideas for its analysis.

Analysis

Once coding has been completed, the codes need to be entered into the qualitative data analysis computer program. This can even be done during the preliminary coding phase if the program will be used for codebook development.

Most software programs allow for online coding, but pair coding and consensus meetings require that paper versions of coded documents be available. Our typical strategy is to start on paper and then enter coding into the program only after if has been finalized. Each program has different logistics for entering the codebook and the coding interviews, but in general, what these programs do is create a “filing cabinet” that stores the interviews and the coding. This filing cabinet organizes material in a myriad of ways – by interview, by the codes, or by demographic information. In response to different search commands the program retrieves any combination of
"files," or sections of coded text. Some programs can also use the existence of particular coding to create matrices that can be exported to various statistical software packages.

It is the automated data retrieval function of these programs that so benefits qualitative analysis by allowing researchers to retrieve all data coded to a particular code and to examine relationships between different segments of coded data. For example, a search of our medical confidentiality interviews for all passages coded to "sexuality concerns" might return primarily responses from young interviewees. A second search on these passages for the context in which these concerns were expressed might suggest they are most common among college age women attending a university health service. These data are useful and could be reported as a quantitative statement, such as: "Three fourths of women who reported having hesitated to discuss sexuality concerns with health care practitioners were college students between the ages of 18 and 22 who used the college health service for their care." However, ending there would miss the point of semi-structured interview data analysis. The software has simply located the respondents and passages where sexuality was characterized as a sensitive topic.

Complete analysis entails going back to the transcripts and examining the selected passages in the context of the interview to discern whether the comments relate to one of the thematic codes, or whether they suggest ideas for additional analyses. Given the initial finding that younger women in university health service settings had a high number of "sexuality concerns," one might, for example, query the status or gender of the health care practitioner seen by these women and what women expressed as their concerns in this setting (if these factors have been coded), and how those concerns related to other themes in the interview. As patterns start to emerge linking concerns, setting, age, and gender, hypotheses about the relationships between age and gender relations in health care might be generated that can be tested on broader segments of the sample or on the whole sample. The point, however, is always to reach back into the interview, at least to the coded passages if not to lengthier exchanges, with the goal of situating patterns in the context of the respondent's story as a whole. These forays into the data also provide the opportunity to collect exemplary quotes that can be used to illustrate the resulting presentation or article.

Before beginning analysis, it is important to devise an overall plan that includes steps needed to examine a set of relationships between thematic codes. Researchers can and typically do deviate from such a plan. Still,
having a written version of a strategy helps to highlight when one has departed from it, which encourages examining the reasons for doing so and keeping track of why old ideas were discarded and new ones proposed. Also, most qualitative data analysis programs lack an easy or effective way to keep track of the order in which analyses were conducted so it is the responsibility of the analyst to do so.

**Summary**

Empirical research is increasingly gaining an equal footing with philosophical analysis in bioethics inquiry. Among the myriad of methods brought to bear in this work, semi-structured interviewing is a reliable and flexible means to gather data. The semi-structured interview allows for comparison across subjects as well as the freedom to explore what distinguishes them.

Semi-structured interviews are useful for examining the complex moral issues that bioethics confronts. The give and take of this method allows the interviewer to follow the subject’s lead, within the parameters of the study. The volume of the data acquired through semi-structured interviews entails, however, a trade off, usually in sample size. Given the demands of the method, including its time-consuming coding, and the difficulty of adapting a single interview guide to widely variant populations, the generalizability of findings will always be somewhat limited. Additionally, it can be difficult for interviewers to repeatedly hear lengthy personal stories that are distressing, such as those related by women about a breast cancer diagnosis or confidentiality breaches.

Research findings from studies relying on semi-structured interviews can be considered problematic for policymakers because results are not easily generalizable, given the uniqueness of the real-world situations from which the data are drawn (Koenig, Back, & Crawley, 2003). Similarly, for clinicians looking for strategies to improve practice, the absence of easily digested bullet points in many qualitative studies may lead readers to ignore such research. However, these limitations are offset by the distinct strengths of this method, including detailed, in-depth, and unanticipated responses. Such data can often help identify novel factors or explain complex relationships, which can contribute substantially to the understanding of complex ethical issues, policy assessment, decision-making, and development of interest in bioethics.
REFERENCES


SECTION III:
QUANTITATIVE METHODS
This page intentionally left blank
SURVEY RESEARCH IN BIOETHICS

G. Caleb Alexander and Matthew K. Wynia

ABSTRACT

Surveys about ethically important topics, when successfully conducted and analyzed, can offer important contributions to bioethics and, more broadly, to health policy and clinical care. But there is a dynamic interplay between the quantitative nature of surveys and the normative theories that survey data challenge and inform. Careful attention to the development of an appropriate research question and survey design can be the difference between an important study that makes fundamental contributions and one that is perceived as irrelevant to ethical analysis, health policy, or clinical practice. This chapter presents ways to enhance the rigor and relevance of surveys in bioethics through careful planning and attentiveness in survey development, fielding, and analysis and presentation of data.

INTRODUCTION

Surveys are a common method in empirical bioethics. They cannot say what is right or wrong, but they can reflect, within bounds, what people are actually doing or thinking. They can also provide information concerning whether consensus exists about a given issue. As a result, they can both inform ethical analysis and be useful in policy making and clinical practice.
Consider three examples of surveys examining ethically important topics. The first, a survey of healthy volunteers and firefighters regarding their attitudes towards quantity and quality of life, indicates that approximately one-fifth would choose radiation instead of surgery to manage laryngeal cancer in an effort to preserve voice, even though doing so would lead to a lower likelihood of survival (McNeil, Weichselbaum, & Pauker, 1981). The second, a survey of family members, demonstrates that their ability to accurately predict a loved one's preference for life-sustaining treatment is quite limited (Seckler, Meier, Mulvihill, & Paris, 1991). The third, a survey of physicians, suggests that nearly two-fifths of respondents report having used one of three tactics to deceive insurance companies in order to help their patients obtain coverage for restricted services (Wynia, Cummins, VanGeest, & Wilson, 2000).

These examples illustrate the diversity of ethically important topics that can be explored by surveys, as well as the various purposes they can serve (Table 1). The first survey, examining the tension that may exist between treatments that impact both quality and length of life, demonstrates that length of life is not always paramount, and some people are willing to make trade-offs in "cure-free survival" for improved quality of life such as preservation of voice. Many surveys of ethically important topics are of this type; they are useful in the process of developing, supporting or refuting normative empirical claims (e.g. the claim that most people value quantity over quality of life).

The second survey provides important evidence of limitations in the ability of surrogates to predict what family members without decision-making capacity would want in situations of life-threatening illness. This has profound and direct clinical implications: physicians should strive to ensure that patients’ wishes are known, work with family members to distinguish between their own wishes and those of the patient whose best interests they are trying to serve, and be cautious of the reliability of "substituted judgment" to truly reflect patient’s wishes. Many other noteworthy surveys examine similarly important topics in clinical ethics, such as those exploring the accuracy of physician’s prognoses for their terminally ill patients (Christakis & Lamont, 2000).

The final example examines insurance company deception. It demonstrates regulatory failures or the inoperability of common normative assumptions that physicians should not deceive on behalf of their patients, whether to protect a third party (Novack et al., 1989) or to secure reimbursement for services (Wynia et al., 2000). Such findings may lead to policy changes to make the adjudication of appeals to managed care
<table>
<thead>
<tr>
<th>Group</th>
<th>Objectives</th>
<th>Sample</th>
<th>Main Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Novack et al. (1989)</td>
<td>To assess physicians’ attitudes toward the use of deception in medicine</td>
<td>211 practicing physicians</td>
<td>Most physicians were willing to misrepresent a screening test as diagnostic test to get insurance payment, 1/3 indicated that they would give incomplete or misleading information to patient’s family if a mistake caused patient’s death</td>
</tr>
<tr>
<td>Wynia et al. (2000)</td>
<td>To examine physicians attitudes toward and frequency of manipulation of reimbursement rules to obtain insurance coverage for services</td>
<td>720 practicing physicians</td>
<td>Thirty-nine percent of physicians used one of the three techniques to manipulate reimbursement rules during the previous year</td>
</tr>
<tr>
<td>Cohen, Fihn, Boyko, Jonsen, and Wood (1994)</td>
<td>To assess physicians’ attitudes toward physician-assisted suicide and euthanasia</td>
<td>938 physicians in one state</td>
<td>Forty-eight percent of physicians thought euthanasia is never ethically justified and 39% believed physician-assisted suicide is never ethically justified. Fifty-four percent believed euthanasia and 53% believed physician-assisted suicide should be legal in some situations</td>
</tr>
<tr>
<td>Patients or family</td>
<td>Seckler et al. (1991)</td>
<td>To determine accuracy of family members’ and physicians’ predictions of patients’ wishes to be resuscitated</td>
<td>70 patients and their family members and physicians</td>
</tr>
<tr>
<td>Group</td>
<td>Objectives</td>
<td>Sample</td>
<td>Main Findings</td>
</tr>
<tr>
<td>-----------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Ganzini et al. (1998)</td>
<td>To determine attitudes toward assisted suicide among patients diagnosed with amyotrophic lateral sclerosis (ALS) and their caregivers</td>
<td>100 patients with ALS and 91 family caregivers</td>
<td>Fifty-six percent of patients would consider assisted suicide and 73% of caregivers and patients have the same attitude toward assisted suicide</td>
</tr>
<tr>
<td>General public</td>
<td>McNeil et al. (1981) To explore preference for laryngectomy (high survival rate, high loss of speech) versus radiation (low survival rate, low loss of speech) for throat cancer</td>
<td>37 healthy volunteers, 12 firefighters, 25 middle and upper management executives</td>
<td>Twenty percent of volunteers would choose radiation to preserve quality of life over quantity</td>
</tr>
<tr>
<td>Multiple groups</td>
<td>Ubel et al. (1996) To determine whether individuals make cost-effective decisions about medical care given budget constraints</td>
<td>568 prospective jurors, 74 medical ethicists, 73 decision-making experts</td>
<td>A little over half of the jurors and medical ethicists made non-cost-effective medical decisions when the non-cost-effective choices offered more equity among patients</td>
</tr>
<tr>
<td>Bachman et al. (1996)</td>
<td>To assess physician and citizen attitudes toward physician-assisted suicide</td>
<td>1119 physicians and 998 members of the general public</td>
<td>Most physicians and members of the public preferred legalizing physician-assisted suicide over banning it</td>
</tr>
<tr>
<td>Degner and Sloan</td>
<td>To determine how involved cancer patients and members of the general public would want to be in their treatment decisions</td>
<td>436 newly diagnosed cancer patients and 482 members of the public</td>
<td>Most patients wanted physicians to make treatment decisions on their behalf; most members of the public wanted to decide on their own treatment if they developed cancer</td>
</tr>
</tbody>
</table>
organizations as transparent and fair as possible and, equally important, these conclusions tell us something about how physicians view the equity of the current health insurance system.

Despite the many purposes of surveys in bioethics, there are important limitations to the data they produce, including the authenticity of the responses and how far they can be generalized. In addition, the number of people doing X does not necessarily mean that X is desirable or undesirable. Heterogeneity of an attitude or behavior is almost always present, warranting consideration at the outset of how this will affect data analysis, conclusions, and presentation of an empirical inquiry (Lantos, 2004).

**FINDING A QUESTION**

A successful survey begins with identifying a good research question (Table 2). Identifying such a question is not trivial. The literature is vast and certain areas of inquiry in ethics, such as end-of-life care, have been extensively explored. However, the tremendous effort to examine ethical

<table>
<thead>
<tr>
<th>Domain</th>
<th>Examples of Survey Topics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Principles of medical ethics</td>
<td>Cancer patient end-of-life care</td>
</tr>
<tr>
<td></td>
<td>Resource allocation; duty to treat</td>
</tr>
<tr>
<td></td>
<td>Physician involvement in capital punishment</td>
</tr>
<tr>
<td>Truth-telling</td>
<td></td>
</tr>
<tr>
<td>Justice</td>
<td></td>
</tr>
<tr>
<td>Non-malfeasance</td>
<td></td>
</tr>
<tr>
<td>Ethically concerning behaviors</td>
<td>Physician support for deception of third-party payers</td>
</tr>
<tr>
<td>Deception</td>
<td></td>
</tr>
<tr>
<td>Poor quality care</td>
<td>Non-adherence to appropriate guidelines</td>
</tr>
<tr>
<td>Unprofessional behavior</td>
<td>Sexual relations with patients</td>
</tr>
<tr>
<td>Prognostication</td>
<td>Accuracy of physicians’ prognoses</td>
</tr>
<tr>
<td>Euthanasia</td>
<td>Endorsement of euthanasia to treat suffering</td>
</tr>
<tr>
<td>Surrogate judgment</td>
<td>Predictive fidelity of surrogate judgments</td>
</tr>
<tr>
<td>Informed consent</td>
<td>Adequacy of process of informed consent</td>
</tr>
<tr>
<td>Trade-offs</td>
<td>Willingness to forgo length to enhance quality of life</td>
</tr>
<tr>
<td>Quantity versus quality of life</td>
<td>Endorsement of insurance company deception</td>
</tr>
<tr>
<td>Advocacy versus honesty</td>
<td></td>
</tr>
</tbody>
</table>
dimensions of end-of-life care speaks to the perceived importance of the topic, and an investigator with a strong interest in a well-explored area should not be dissuaded simply because others have examined the topic already. Big problems benefit from numerous approaches, and none are solved by one study alone. On the other hand, the best research questions are not only those that are feasible, but also novel. Whether or not a topic has previously been studied, any survey topic should also be relevant, either contributing to the development and understanding of normative theories or to more directly inform health policy or clinical care. Steps to identifying a good research question include a comprehensive review of the literature and existing theories, findings from qualitative studies, and discussion with content experts. Furthermore, the researcher must maintain a careful balance of creative intellectual meandering, considering multiple questions, and focusing on a specific topic. The importance of working from a theoretical framework and hypothesis varies somewhat depending upon the type of question being examined; detailed conceptual models and hypotheses are especially crucial for studies examining predictors or determinants of behaviors or attitudes. On the other hand, simple descriptive analyses of the frequency of an important outcome may demand a less well-developed theoretical framework to make the results interesting and useful. The question to ask, as in any research, is: “So what?” Who might care, and why, if you find what you expect to find, or if you do not? What ethical, policy, and/or clinical implications would follow from having answers to the questions you hope to pose?

**STUDY DESIGN**

There are a host of issues to consider in survey design and analysis, many of which this chapter will explore only briefly. There are several excellent references for further reading on general survey design, conduct, and analysis (Aday, 1996; Dillman, 2000; Fink, 1995). Here, we will focus on a few aspects of survey research that are most likely to arise during surveys on ethical issues.

*Who is to be Surveyed?*

Most bioethics surveys are of patients, family members, clinicians, others involved in health care (chaplains, social workers, and so on), or the general
public. The population selected should be driven by the research question and can inform survey development. For example, health professionals may have high literacy levels and the capacity to respond to complex survey designs due to their familiarity with standardized testing. On the other hand, health professionals also have a low threshold of tolerance for any time-consuming activity, suggesting that short surveys will be better received than longer ones. Certain populations of patients might have lower literacy or might not speak English – obviously, both would have a profound effect on survey design and administration. Although methodologically more complex, there are times when comparing or contrasting two different populations, rather than examining a single group, may be advantageous. Such an effort allows a broader and more rigorous approach to describing and reaching conclusions about attitudes or behaviors and may highlight important similarities and differences between groups of decision-makers, such as patients and families (Ganzini et al., 1998) or patients and clinicians (Alexander, Casalino, & Meltzer, 2003).

**Sampling**

Once a target group to be surveyed has been selected, it is important to consider how the sample within this group will be selected; this is called “sampling” or “sample design.” The degree to which those who receive the survey reflect the larger group from which they are drawn will affect the generalizability of survey findings, and hence the survey’s relevance and usefulness. At the same time however, important subgroups might be overlooked if special care is not taken in the survey’s sample design. Ethical issues might be rare, or might be most relevant to certain subsets of survey recipients (e.g. those recently testing positive for a genetic screening test).

There are two main types of sampling designs, probability and non-probability designs. The most common probability sampling design is a simple random sample, in which the probability of a sub-group’s selection into the survey sample is proportional to the frequency of that sub-group within the universe from which the sample is derived. In some cases, more sophisticated probability sampling methods may be applied to enhance the rigor of the survey protocol. For example, a stratified random sample allows for sampling of different sub-groups or strata within the sample universe, which is particularly helpful when different strata of interest are not equally represented within the sample universe. For instance, a survey of patients in the intensive care unit (ICU) might be designed to over-sample (i.e., include
more than would be chosen by chance) patients who are younger than 55 years, so that it can provide more information about this particular population, which is relatively uncommon in the ICU. Although the use of stratified probability sampling may improve efficiency and allow for greater representativeness and accuracy for selected populations, it introduces increased expense and complexity into the project and also depends upon a priori information about the populations (or strata) of interest. For analysis and reporting, most surveys using probability designs will benefit from developing sample weights. How to develop weights is beyond the scope of this chapter, but suffice it to say that weighting data allows one to report the results as if the whole sample had been drawn at random and was exactly representative of the universe of potential respondents. Weights allow for this by accounting for stratified sampling designs, as well as by taking into account potential sources of bias among survey respondents, such as differential response rates among different subsets of participants (Korn & Graubard, 1999).

There are many types of non-probability sample designs. For example, convenience samples comprise populations that are surveyed because of ease and availability, such as a sample derived of patients and family members passing through a hospital cafeteria. Purposive samples and quota samples consist of subjects surveyed based on pre-specified criteria; in the latter case subjects are sampled until certain quota are fulfilled, such as a certain number of subjects of a given age range or race. Snowball samples use survey participants to recruit other potential participants, such as by asking physicians in retainer medical practices to name other physicians in similar type practices that might be contacted to participate in the survey (Alexander, Kurlander, & Wynia, 2005).

Non-probability survey designs are neither inferior nor superior to probability designs overall; rather, each design has strengths and limitations and should be chosen based on the research question, available resources, and the degree to which external validity of the findings is important. External validity refers to how generalizable the data are to a larger (external) population. In some cases, survey findings only need to apply to a limited population of special interest. For example, a survey of physicians who practice in pediatric ICU’s was conducted to assess whether family presence during resuscitation attempts was more or less acceptable to them, and desired by patient families (Kuzin, et al., 2007). Whether these physicians’ views were representative of the general pediatrician population was not especially important, since most pediatricians outside the ICU have very little experience with acute resuscitations (Tibbals & Kinney, 2006).
On the other hand, samples that are too small or poorly reflective of the population from which they were drawn should not be used to derive conclusions about the broader universe of subjects. For example, researchers studying the coping skills of parents of children hospitalized for chronic diseases might err in generalizing the results of their study to parents of children hospitalized for acute illness; these two groups of parents may differ in important ways with respect to the research question at hand. Or a study of family physicians’ views regarding home childbirth might be inappropriately generalized to obstetricians and general practitioners, but again, important differences may exist among these specialties with regard to the topic of interest.

**Survey Mode**

Most surveys are administered in-person or else conducted by U.S. mail, telephone, or increasingly, the Internet. Each survey modality has strengths and limitations (Table 3). A particular survey mode (e.g. telephone) should be selected with care, since it may influence both the response rate and patterns of responses. In addition, some populations may be more or less responsive to certain types of surveys. For example, those with low literacy will be less likely to respond to a written survey, and telephone surveys only reach people owning telephones. Recent trends in ownership of cellular versus landline phones also should be taken into account, such as the younger ages of people who own a cell phone but not a home phone (Tuckel & O’Neill, 2005).

**Survey Design**

Good survey questions are simple, clear, and often very easy to answer. However, survey development is a time-consuming task, and formulating specific survey questions to be “just right” may be quite difficult. Nevertheless, the time and effort required to create a good survey are well spent. Poor wording and design of surveys do not just frustrate respondents, it also leads to difficulties in item interpretation, higher rates of non-response and, as a result, more difficulty getting the results analyzed, published, and used.

Survey questions (often referred to as “items”) consist of three parts: (1) the introduction to the question, (2) the question itself (*question stem*),
Table 3. Pros and Cons of Different Survey Modes.

<table>
<thead>
<tr>
<th>Mode</th>
<th>Commonly Targeted Respondents</th>
<th>Benefits</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phone</td>
<td>General public</td>
<td>- Inexpensive</td>
<td>- High non-response rates</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Random-digit dial allows broad sampling of general public</td>
<td>- Misses those without phones</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Avoids problems of illiteracy</td>
<td>- Prevents use of visual aids as well as visual cues from respondent</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Rapid data entry with Computer Assisted Telephone Interviewing (CATI) software</td>
<td></td>
</tr>
<tr>
<td>In-person</td>
<td>Patients, General public</td>
<td>- Allows for sampling of patients or family members in clinical setting or public space</td>
<td>- Very expensive</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Offers use of visual aids and visual cues from respondents</td>
<td>- Use in clinical settings over-samples frequent users of care</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Allows means of assisting with survey completion</td>
<td>- Use in public settings generally based on convenience sample of respondents who may differ from non-respondents</td>
</tr>
<tr>
<td>U.S. Mail</td>
<td>Physicians</td>
<td>- Common method to reach physicians</td>
<td>- Expensive</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Well-established protocols and mailing lists</td>
<td>- Use of survey waves complicates using anonymous design</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Some ability to track and compare respondents with non-respondents</td>
<td></td>
</tr>
<tr>
<td>Internet</td>
<td>Physicians, Internet users</td>
<td>- Very inexpensive</td>
<td>- Limited ability to generalize findings to non-Internet users</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Automatically entry of data while collected</td>
<td>- Email address lists often contain wrong addresses</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Fast data collection</td>
<td>- High non-response rates</td>
</tr>
</tbody>
</table>
and (3) the responses to the question (response frame). This structure is helpful to understand, both for conceptual clarity in communicating with others, as well as during the process of survey design.

The introduction to a question or set of questions is crucial because it orients the respondent as to what is expected. Framing the issue is often critically important in ethics surveys (see section on Precautions with Sensitive Topics). For example, respondents can be put at ease through reassurance (e.g. "there is no right or wrong answer") or by acknowledging the challenge of the question (e.g. "please balance both the patients' quality and length of life").

Survey questions are either forced-choice or, less frequently, open-ended. The benefits of an open-ended question, where a respondent writes in a response (or provides the verbal equivalent during an interview), are that it is less leading and elicits a broader range of responses than a forced-choice question. Drawbacks are that the coding of such questions is technically complicated due to illegible (on self-administered surveys) or long-winded responses, and is conceptually complicated due to responses that are unclear in meaning or intent (regardless of survey mode). In addition, response rates tend to be lower for open-ended questions.

A good survey question uses simple words, presents a simple idea, and uses as few words as possible to do so. Each survey question should contain a single idea and ask a single question. So-called "double-barrelled" questions, where two questions are asked at once, are generally to be avoided because answers to these questions are very difficult to interpret (for example, "How often have you been sued or been afraid you were going to be sued?"). The challenge with question design is to navigate certain tensions that are inherent in this process. For example, sometimes more words are needed to precisely explain a question or complicated idea, yet more words may make a survey item more cumbersome to read and understand. There is a large literature devoted to the psychology of survey response, and readers interested in advanced study in this area should refer to one of several excellent references (Aday, 1996; Fink, 1995; Tourangeau, Rips, & Rasinski, 2000).

Finally, the response options available for answering a survey question are important to consider. Response frames can be simple (e.g. yes/no or agree/disagree) or more complex (e.g. excellent, very good, good, fair, or poor) and sometimes it is necessary to create unique response frames that are specific to certain questions. The response frame of an item is crucial to making the question easy for the respondent to answer accurately. It should be developed based on an iterative process of piloting and pre-testing to
develop a response frame that is comprehensive, comprehensible, and, in the case of best choice answers, mutually exclusive.

In addition to giving careful attention to design of the introductions, question stems, and response frames, care should be given to other survey design factors that may influence data quality. These include the survey aesthetics, framing (e.g. survey title), item grouping, and item order. For example, although selecting the title of a survey may seem a mundane or unimportant task, even this can make a considerable difference in response rates. Consider creating a title that is engaging and that will draw the recipient in (e.g. Managed care: What do physicians think?), but avoid titles that might alienate recipients or that suggest a bias in the research (e.g. Are doctors aware of widespread inequities in American health care?). Similarly, it is important to consider the ordering of items. Generally, it is advisable to avoid placing overly sensitive or conceptually difficult items early in the survey or first in a series of questions, use open-ended prior to forced-choice questions if they are about similar topics, and try to group items that have the same response frame together so as to maximize the flow and readability of the survey. In addition, it is helpful to number items within each response frame to maximize the accuracy and ease of data entry.

**SURVEY PLANNING**

A considerable amount of work on survey analysis, sampling, and projections of costs should be conducted prior to fielding the survey instrument. This helps to minimize the collection of data that inevitably are not used, and the inadvertent omission of data that would have been useful. Questions to consider in planning the survey include: What is (are) the main outcome(s) of interest? Will simple descriptive statistics of the frequency of an attitude or behavior suffice, or is the goal to conduct more detailed analyses of associations between different variables? If the latter, what is the dependent variable and what are the independent (predictor) variables of interest? What potential variables may confound the associations of interest? There is considerable value to developing “mock-up” tables of the expected results or lists of potential correlates prior to fielding the survey. This can help to locate gaps in the survey and allow careful consideration for how responses will be analyzed.

The costs of surveys should also be considered during planning. Methods can be tailored to better fit the available budget, for example, by modifying the survey mode, sample size, financial incentives, or rigor of survey
development and statistical analyses. Many surveys of ethically important topics have been conducted on small budgets. However, if the budget allows, there are many organizations, often university-affiliated, that can be contracted with to develop, test, or administer surveys. Similarly, firms exist that will provide survey support for telephone-based or Internet samples.

Ethics surveys demand special attention to response rates. Since they are often about sensitive topics, ethics surveys may suffer from poor response rates, and there may be important differences between respondents and non-respondents (i.e., non-response bias, see below). These issues can stimulate questions about a survey’s relevance. General efforts to improve response rates are of several types: (1) financial or non-financial incentives; (2) endorsements from opinion leaders/people of influence; (3) minimizing the burden of the survey; (4) personalizing the survey through efforts such as hand-written notes, adhesive stamps, or human signature; and (5) persuasion about the importance of the topic and the respondents’ views. This last point can be accomplished, for example, through a particular way of framing the survey in the cover letter and through its title, or through the use of special delivery methods (e.g. Federal Express for mailed survey). Of these five methods, the use of financial incentives has been studied most extensively. In general, studies of financial incentives suggest that the marginal benefit of larger financial incentives may be relatively small compared with the impact of smaller incentives (VanGeest, Wynia, Cummins, & Wilson, 2001). In addition, financial incentives, when used, are more effective when a small amount is offered upfront to everyone, rather than the use of a lottery or incentives upon survey completion (Dillman, 2000).

SURVEY DEVELOPMENT AND PRE-TESTING

After developing a general research question and theoretical framework, efforts should turn to identifying specific conceptual domains and factors to be explored within these domains. Qualitative data are often helpful to inform this process, and there are various ways to gather such data, including key informant interviews, focus groups, and ethnography (see chapters on qualitative methods). Such efforts are often invaluable in helping to identify important areas of inquiry, and may provide sufficient material for analyses that can take place in parallel with the quantitative
focus of the survey. Finally, most surveys need to be piloted and pre-tested to enhance and ensure brevity, clarity, and measurement accuracy.

Accuracy consists of both the validity and reliability of the survey instrument. Validity is the degree to which a survey actually measures the phenomenon it purports to measure. In this regard, it is important to note that ethics surveys often attempt to measure phenomena (or "constructs") that are complex and on which clear consensus may not exist. For instance, ethics surveys may explore the meaning of "consent" or the importance of "privacy" or "fairness" in health care. When assessing such challenging and obscure constructs, particular attention must be given to the validity of the survey to ensure that it measures what it claims to measure. There are three main types of validity relevant to survey research: face validity, content validity, and construct validity.

Face validity refers to whether the survey domains and items appear reasonable at face value. A common way to assess face validity is to share the instrument with relevant parties, such as patients, family members, and clinicians, and ask if they agree that the items measure what you intend them to measure.

Content validity refers to how well the items examined represent the important content of the domains of interest. To assess content validity one must ask, does the survey address all the facets of the issue in question or, on the other hand, does it include aspects that are unrelated to the issue? For example, in the development of their survey instruments for assessing privacy in hospitals, researchers asked a group of experts to evaluate the survey items by rating each item on a 1–10 scale where (1) represented items necessary to the issue and (10) represented items extraneous to the issue. The experts were also asked for ideas on any areas that might have been missed. As this example shows, ensuring adequate content validity is usually achieved by the researchers' comprehensive knowledge of the subject matter the survey is examining and is supported by careful review of the survey domains and items with experts in the field of inquiry. As "soft" as simple reliance on experts may seem, content validity is crucial to a good survey on an ethically important topic. Criticisms of ethics surveys are frequently focused on items that the survey team did not ask, or items that while the team asserts are related to the central question, there is not a strong outside consensus from experts that this is so.

Finally, construct validity refers to the degree to which one can safely make inferences from a survey to broader theoretical constructs operationalized in the survey. Construct validity is very difficult to prove in many ethics surveys, because the constructs in question are often quite complex. How
does one prove that constructs such as “informed consent,” “concern for privacy,” “fear,” or a sense of “disrespect” or “mistrust” are accurately measured? Construct validity is assessed by determining that the measure in use is related in expected ways to other known measures of the construct in question. For instance, the results of a survey that purports to measure “pain” would be expected to correlate with other measures that are known to be associated with pain, such as sweating, rapid pulse, and asking for pain medication. The challenge in ethics surveys is often to determine, a priori, what are the expected correlates of the relevant constructs.

In addition to being valid, surveys should be reliable as well: they should get the same results each time they are used in the same circumstances. The reliability of a survey can be assessed through repeated administrations to one individual (intra-observer, or test-retest reliability) and by assessments of a given event or practice across multiple individuals (inter-observer reliability) (also see Chapter on The Use of Vignettes). Statistical tests of reliability are well described in most basic biostatistics or clinical epidemiology textbooks.

**SURVEY FIELDING AND DATA ENTRY**

In any ethics survey, especially one using new items, there is considerable benefit in examining early responses as they come in. For instance, it may be helpful to perform some analyses on the first wave of survey responses. This may allow for the identification of potentially serious systematic flaws, such as if respondents are unintentionally skipping questions printed on the back of a page.

As data are collected, several methods can be used to ensure systematic initiation of data entry and analysis. Errors in data entry are almost impossible to avoid, but the rigor of data entry may be enhanced by double entry or randomly checking a subset of respondents for the frequency of incorrectly entered data, or both. Questions about how to code unclear responses, such as when a response is somewhat illegible or does not clearly fit the response frames offered, are inevitable. The researcher should anticipate these and work to treat them in a systematic and fair way that does not introduce unnecessary bias into the measurements. Data entry can be enhanced by the use of limited data fields and specialized software that may simplify tasks such as providing easier database management of complicated survey skip patterns. Data entry can also be simplified by minimizing the number of steps between survey query and response in the
case of telephone or in-person surveys through the use of Computer Assisted Telephone Interviewing (CATI) software. Even with these efforts, data cleaning is important to ensure that the data are of high quality; that is, data for given variables are within expected ranges (e.g. a survey item with a response frame from 1–5 should not have any 8s entered as responses) and the distributions of data make sense (e.g. an item where all respondents answered the same would raise suspicions of a data entry error).

BIASES AND RESPONDENT BURDEN

Respondent bias and burden are two of the most important considerations to guide survey development. Bias refers to any systematic tendency to over- or underestimate whatever is being measured. There are numerous types of bias that are important to consider.

Socially desirable response bias, sometimes referred to as “yeah-saying,” is a special threat to the rigor of ethics surveys. It can be addressed in several ways, including sensitively wording questions, carefully designing the survey framing, paying attention to item ordering, ensuring protection of confidentiality (or, rarely, even ensuring anonymity), and by using statistical methods that help to adjust for the likelihood of this bias.

Recall bias may be present when asking about past events, and can be minimized by carefully framing the time period in question, such as by limiting the length of the retrospective period (e.g. “In the last week …”) or by using discrete events that are more likely to be accurately recalled (e.g. “At the last Grand Rounds you attended, ...”).

Non-response bias refers to bias introduced by systematic differences between respondents and non-respondents. That is, those who return the survey may be different in some relevant way from those who do not. This type of bias is a perennial challenge to survey development, fielding, and interpretation, and it is especially relevant to ethics surveys, which often touch on very sensitive or controversial topics. For instance, survey recipients who are especially affected by the survey topic (e.g. malpractice) may be more, or less, likely to respond. In addition, the association between respondent burden and poor response rates should not be underestimated. Although low response rates do not mean that non-response bias is present, longer and more cumbersome surveys are less likely to be completed, and lower response rates, all other things being equal, will raise concerns of possible non-response bias. In addition to seeking to improve response rates, there are four common methods used to address non-response bias.
First, one can compare respondents with non-respondents on all known variables (the absence of differences suggests that non-response bias is less likely). Second, one can look for "response-wave bias" in surveys conducted by Internet or U.S. mail by exploring whether there is any association between the length of time until survey response and the primary outcome(s) of interest. Response-wave bias is based on an assumption that respondents who took a long time to respond to the survey, such as those responding to a third survey wave, somewhat resemble non-respondents in that they were less motivated to respond than their counterparts. The absence of any association between length till response and the primary outcome(s) of interest suggest that non-response bias is less likely. Third, the active pursuit of a subset of non-respondents may be helpful to ascertain the response frequencies to one or two key questions among this group. For example, in a survey of physicians' support for capital punishment, researchers might be concerned of a significant non-response bias among the 45% of subjects who were non-respondents. To examine for this bias, the researchers might select a random 10% sample of non-respondents and call them by phone with one short question from the survey to find out whether their global beliefs about capital punishment are similar to survey respondents. Finally, advanced statistical methods can be used, such as weighting of survey responses, in an effort to account for non-response bias. Other biases are important to consider depending upon the unique circumstances of the project (e.g. interviewer bias, which introduces variation in survey response based on the characteristics of the interviewer administering the survey), but are less ubiquitous threats to survey validity than those discussed above.

**PRECAUTIONS WITH SENSITIVE TOPICS**

Many ethics surveys examine the prevalence of specific behaviors. One-way to identify the frequency of any behavior, sensitive or not, is to directly question the respondent (e.g. "In the last week, how often did you ...?"). However, the benefit of being able to report direct prevalence must be balanced with an acknowledgment that such reports are especially prone to socially desirable response bias and therefore may over- or underestimate the actual frequency of the behavior in question. Positive behaviors are likely to be over-reported, while negative behaviors are likely to be under-reported. In the case of negative behaviors, direct questions can also alienate survey recipients, because direct questions about negative behaviors are often perceived as leading (e.g. "How often, if ever, do you kick your...")
dog?”). As a result, direct questions may be most useful for non-sensitive topics or under circumstances in which modest misestimation of prevalence might not reduce the likely ethical, policy, or clinical impact of the survey results. For instance, even if only a few respondents directly report a very concerning behavior it might be worth investigation (e.g. admission of illicit drug use among physicians).

An alternative to direct questions is to use indirect, or third party, questions, which can allow respondents to discuss sensitive topics without having to personally admit to socially undesirable or stigmatized behavior. For example, instead of asking individual patients whether they have ever faked an illness to skip work, inquire as to whether they know of any colleagues that have done so. The trade-off is that such questions do not allow for direct assessment of the frequency of the behavior and hence, these questions too might not produce good estimates of prevalence. For instance, perhaps only one employee has skipped work on a medical excuse to go fishing, but many employees completing the survey know about this situation – in this case, the frequency of this behavior might be overestimated.

There are several methods to help maximize respondent honesty and comfort when completing both direct and indirect questions about sensitive topics. First, the way that each question is framed is crucial, and calming stems that acknowledge the legitimacy of different or controversial viewpoints and actions are helpful to allow respondents to answer honestly (e.g. “Patients often find health care frightening and stressful, and they handle this stress in many ways …” or “There are no right answers to the following questions …”). Question order can be used to one’s advantage as well – generally, questions that are more sensitive should be introduced later in the survey while more sensitive responses are better earlier in the response frame. The risk of socially desirable response bias in response to direct questions can also be diminished with any interventions that help to protect respondents’ confidentiality or anonymity.

Another option, useful for both sensitive and non-sensitive topics, is hypothetical vignettes. For a detailed discussion of this method, please see chapter on hypothetical vignettes.

**DATA ANALYSIS**

After data entry and cleaning, examination of univariate (single item) distributions is helpful. A blank survey instrument can be used as a template upon which to write these distributions. In some, but not all, cases, bivariate
and multivariate distributions may be of interest, in order to see how the primary measure(s) of interest, such as patient’s preference for end-of-life care, may be associated with other variables of interest, such as illness chronicity, hospice availability, and the specialty of the treating clinician. Multivariate regression analyses and other advanced statistical analytic techniques are possible; however, they require additional skills that may be beyond most bioethicists and they may not be relevant to the research question at hand. While it is sometimes critical to evaluate associations while holding other factors constant, as multivariate regression allows one to do, in many important ethics-related surveys, simple descriptive statistics are sufficient to examine the question of interest (e.g. what proportion of surgeons would override a patient’s “Do Not Resuscitate” order in the immediate post-operative period?). In this regard, it is important to return to the initial survey question or hypothesis and the conceptual model that one is using to frame the question. Where advanced statistical methods are required, it may be helpful to collaborate with statisticians and others with advanced training in health services research. Despite the relative ease with which statistical programs can generate multivariate analyses, creating appropriate models, using appropriate tests, and interpreting the results of these models demands special expertise.

CONCLUSIONS

Surveys have great promise as a method to inform bioethics, clinical practice, and health policy. To achieve this promise, the researcher must balance rigor with feasibility at all stages of survey development, fielding, analysis, and presentation. Table 4 provides a case study of an empiric ethics survey and illustrates examples of some key elements of survey design, administration, and analysis. Identifying a good research question is crucial, yet the difficulty of this stage of survey research may be easily overlooked. When conducting ethics surveys, it is particularly important to guard against constant threats to survey validity, such as unclear wording of survey questions or biased responses to questions, because the issues under study are often complex, conceptually inchoate, and/or sensitive or controversial. Finally, in bioethics the relationship between the quantitative data gathered by surveys and the qualitative nature of normative theories is a dynamic one, making survey interpretation a challenge. Nevertheless, a well-constructed, carefully analyzed survey in ethics can have a meaningful impact on policy and practice. Surveys are well suited for examining areas of

<table>
<thead>
<tr>
<th>Aspect of Survey</th>
<th>Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goal of survey</td>
<td>• To explore physicians’ bioterrorism preparedness, willingness to treat patients despite personal risk, and beliefs in a professional duty to treat during epidemics</td>
</tr>
<tr>
<td>Finding a question</td>
<td>• Question motivated by: (1) significant topical interest in duty to treat since September 11th, 2001, (2) prior debates regarding duty to treat during outset of HIV epidemic, (3) long and varied historical tradition of physicians’ response to epidemics, and (4) salient ethical dilemmas regarding balance of physician self-interest with beneficence</td>
</tr>
<tr>
<td>Who is to be surveyed?</td>
<td>• Decision to survey patient-care physicians</td>
</tr>
<tr>
<td>Sampling</td>
<td>• Simple random sample taken from universe of all practicing patient-care physicians in the United States; use of representative sample allowed for inferences to be made regarding broader physician population</td>
</tr>
<tr>
<td>Survey mode</td>
<td>• Survey conducted by US mail given impracticality and costs of in-person or phone surveys, and given generally low response rates to Internet surveys and the absence of reliable email addresses for physicians</td>
</tr>
<tr>
<td>Survey design</td>
<td>• Single page survey to maximize response rate; most response frames similar to reduce respondent burden</td>
</tr>
<tr>
<td></td>
<td>• Survey title emphasized importance of respondents’ experiences; survey layout facilitated completion in less than 5 min</td>
</tr>
<tr>
<td>Survey planning</td>
<td>• Decision that main outcomes of analysis would be simple descriptive statistics; awareness a priori that causal direction of any associations would be unclear</td>
</tr>
<tr>
<td></td>
<td>• Length and rigor of pretesting balanced with need for timely data collection given potential shifting interest in the topic among policy makers, providers, and the general public</td>
</tr>
<tr>
<td></td>
<td>• Efforts to obtain maximal response rates included minimizing burden by limiting survey to one sheet of paper, using persuasion that important topic, and using a $2 financial incentive</td>
</tr>
<tr>
<td>Development and pretesting</td>
<td>• Face validity maximized through piloting and pretesting with practicing clinicians</td>
</tr>
<tr>
<td></td>
<td>• Content validity maximized by expert review of survey by clinicians involved in disaster response planning</td>
</tr>
<tr>
<td>Fielding and data entry</td>
<td>• Construct validity maximized by examining expected correlations between items on training and preparedness</td>
</tr>
<tr>
<td></td>
<td>• Analysis of early survey respondents allowed us to observe an association between survey response time and duty to treat; it was unclear if this was due to response-wave bias or temporal trends</td>
</tr>
</tbody>
</table>
health care that raise vexing ethical issues for patients, clinicians, and policymakers alike.

**REFERENCES**


HYPOTHETICAL VIGNETTES IN EMPIRICAL BIOETHICS RESEARCH

Connie M. Ulrich and Sarah J. Ratcliffe

ABSTRACT

Hypothetical vignettes have been used as a research method in the social sciences for many years and are useful for examining and understanding ethical problems in clinical practice, research, and policy. This chapter provides an overview of the value of vignettes in empirical bioethics research, discusses how to develop and utilize vignettes when considering ethics-related research questions, and reviews strategies for evaluating psychometric properties. We provide examples of vignettes and how they have been used in bioethics research, and examine their relevance to advancing bioethics. The chapter concludes with the general strengths and limitations of hypothetical vignettes and how these should be considered.

INTRODUCTION

The Significance and Value of Vignettes in Empirical Bioethics Research

The value and significance of empirical bioethics research lies in its ability to advance our knowledge and understanding of a variety of ethical issues to promote opportunities for dialog among clinicians, researchers,
policy-makers, and members of other disciplines. Additionally, this type of research generates new lines of descriptive and normative inquiry.

Hypothetical vignettes represent one type of empirical approaches to examining bioethical issues. This method can be used to better understand attitudes, beliefs, and behaviors related to bioethical considerations in clinical practice, health service delivery and financing, and health policy (Finch 1987; Flakerud, 1979; Gould, 1996; Hughes & Huby, 2002; Veloski, Tai, Evans, & Nash, 2005). Issues range from the macro level, such as access to care, costs of care, and allocation of resources to micro level concerns of bedside rationing, provider–patient relationships, and beginning of life and end-of-life care problems. Given the potentially sensitive nature of many bioethic problems, hypothetical vignettes provide a less personal and, therefore, less threatening presentation of issues to research participants. Furthermore, some bioethics-related events are relatively rare and vignettes provide a mechanism to explore the attitudes and extrapolated behaviors concerning such events using larger numbers of participants than would be otherwise possible.

This chapter will give an overview of hypothetical vignettes in research with examples of how this method has been used to examine and analyze critical ethical problems. We will also review ways to evaluate the reliability, validity, strengths, and limitations of studies using vignettes.

**WHAT IS A VIGNETTE?**

Vignettes have been used in social science research since the 1950s (Gould, 1996; Schoenberg & Ravdal, 2000) and have been described as “short stories about hypothetical characters in specified circumstances, to whose situation the interviewee is invited to respond” (Finch, 1987, p. 105). Stories are designed to represent an issue of importance that simulates real life and require a focused response from participants. Depending on the research question of a study, vignettes are appropriate for both qualitative and quantitative methodological designs and studies using mixed methods. They can be used in isolation or as adjuncts to other data collection methods (Hughes & Huby, 2002), for example, self-administered survey questionnaires, focus groups, and face-to-face semi-structured interviews. Vignettes can be presented in a variety of ways such as verbal administration, written surveys, audiotape, videotape, and computers. The latter allows for a flexible approach to reaching groups across various settings.
Unlike attitudinal scales that ask direct questions about values and beliefs, vignettes offer an approach that assesses individuals’ attitudes or values in a contextualized scenario or situation (Alexander & Becker, 1978; Finch, 1987; Flaskerud, 1979; Gould, 1996; Hughes, 1998; Hughes & Huby, 2002; Schoenberg & Ravdal, 2000; Veloski et al., 2005). Various disciplines have used vignettes in health care and social science research among nurses, physicians, and in the general population to examine and ascertain attitudes, values, behaviors, and norms (Alexander, Werner, Fagerlin, & Ubel, 2003; Asai, Akiguchi, Miura, Tanabe, & Fukuhara, 1999; Barter & Renold, 2000; Christakis & Asch, 1995; Denk, Fletcher, & Reigel, 1997; Emanuel, Fairclough, Daniels, & Claridge, 1996; Gump, Baker, & Roll, 2000; Kodadek & Feeg, 2002; McAlpine, Kristjanson, & Poroch, 1997; Nolan & Smith, 1995; Rahman, 1996; Wolfe, Fairclough, Clarridge, Daniels, & Emanuel, 1999). The interpretation of standardized vignettes by different groups of people within the same study can also be examined and compared (Barter & Renold, 1999).

Most studies that use vignettes rely on the constant-variable vignette method (CVVM) where identical scenarios are presented to respondents with multiple forced-choice questionnaires or rating scales (Cavanagh & Fritzsche, 1985; Wason, Polonsky, & Hymans, 2002). For example, in a survey using vignettes developed by Christakis and Asch (1995) to measure physician characteristics associated with decisions to withdraw life support, respondents were asked to rate on a 5-point Likert scale how likely they would be to withdraw life support from a hypothetical character presented in the scenarios (Box 1).

In another study to assess whether informed consent should be required for biological samples derived clinically and/or from research, Wendler and Emanuel (2002) surveyed two cohorts of subjects via telephone using three vignettes. Instead of a Likert response set, however, participants were simply asked to respond by using the following categories: “yes,” “no,” “don’t know,” and “it depends.”

Different versions of the same vignette can also be constructed in a single study using a factorial design. In this design, researchers can test multiple hypotheses by systematically manipulating two or more independent variables (for example, age and gender) within the vignette and randomly allocate subjects to the vignette as a means to evaluate both main effects, of each independent variable on the outcome variable, and interaction effects of two or more independent variables on the outcome variable (Polit & Hungler, 1999).
Qualitative researchers use vignettes to explore meanings of particular issues by asking participants to discuss the situation presented in their own words in response to open-ended questions. For example, to better understand the ethical awareness among first year medical, dental, and nursing students, Nolan and Smith (1995) asked students to respond to different vignettes that contained ethical dilemmas. In response to the vignettes, subjects were asked: “What course of action would you suggest, giving reasons?” Other authors (Kodadek & Feeg, 2002) have used similar open-ended questions based on vignettes to explore how parents approach end-of-life decisions for terminally ill infants by asking respondents (1) “What is your first reaction as you begin to think about this problem?”; (2) “What specific questions will you ask the physician when you have a chance to discuss this problem?”; or (3) “Name at least 5 aspects of the problem that you will consider in making your decision.” Open-ended response sets allow for in-depth probing and identification of issues deemed salient and emerging themes. Grady et al. (2006) presented four different hypothetical scenarios about financial disclosure to active research participants in NIH intramural sponsored protocols. Subjects were asked to openly discuss their views on each scenario, whether or what they would want to know about the financial interests of investigators, and how knowledge of financial disclosures would influence their research participation. Lastly,

**Box 1.** Example of a Vignette Used to Assess Physicians’ Perceptions on Withdrawal of Life Support.

EL is a 66-year-old patient of yours with a 15-year history of severe chronic pulmonary disease. One week ago, he was admitted to the ICU with pneumonia, hypotension, and respiratory failure. He required antibiotics, intravenous vasopressors, and mechanical ventilation to survive. He has now lapsed into a coma and shows no signs of clinical improvement. Consultant pulmonologists assert that his lung function is such that he will never be independent of the ventilator.

After his most recent hospitalization, the patient had clearly expressed to his family and to you that he would never want to live by artificial means. In view of these wishes and his poor prognosis, the family asks you to withdraw life support. You are deciding whether to stop the intravenous vasopressors or the mechanical ventilation. (Christakis & Asch, 1995)
Berney et al. (2005) developed two clinical vignettes from semi-structured interviews related to allocation of scarce resources and followed by discussions of the vignettes in focus groups with general practitioners. Participants were asked, through open-ended queries, to describe the ethical issues they perceived when presented with the vignettes. Usually, both qualitative and quantitative research ask respondents to rank, rate, or sort particular aspects of vignettes into categories, or to choose what they think the hypothetical characters should or ought to do in the presented scenario(s) (Martin, 2004).

HOW TO DEVELOP A VIGNETTE

The complexity of many bioethical problems often necessitates constructing vignettes to meet a study’s particular purpose. If possible, however, using established valid and reliable vignettes is always preferable. Vignettes can be constructed from the literature as well as from other sources, and must appear realistic, relevant, and be easily understood (Hughes, 1998; Wason et al., 2002). Several approaches to the development and evaluation of vignettes can be used as described below (Box 2).

1. Focus groups
   Focus groups provide an opportunity to gather information for vignettes from a population of interest related to the specific bioethical issue being

---

**Box 2. How Can Vignettes be Constructed and Presented in Bioethics Research?**

**Constructed:**
- Previous research findings/literature review
- Real life experiences with clinical and/or research cases
- Focus groups
- Cognitive interviewing

**Presented:**
- Narrative story
- Computer based, music videos
- Comic book style; flip book; cards; surveys; audio tapes (Hughes, 1998)
studied, especially if limited information on the topic exists (Krueger & Casey, 2000). Focus groups generally range from 6 to 8 individuals who participate in an interview for a specified time period with the object of obtaining “high-quality data in a social context where people can consider their own views in the context of the views of others” (Patton, 2002, p. 386).

Using focus groups to develop vignettes allows for the identification of the test variables (i.e., age, gender, level of education), how to structure and present content of interest in a vignette, and the number of vignettes needed. Schigelone and Fitzgerald (2004) convened a focus group of experts in geriatrics, nursing, and social science to identify key variables for geriatric vignettes to assess the treatment of older and younger patients by first year medical students. Based on the responses, the authors drafted initial versions of vignettes that examined age of patients in relation to levels of aggressive treatment (for more information, please see chapter on focus groups).

Focus groups help to:

- A priori refine the objectives of the research utilizing vignettes.
- Clarify and provide in-depth understanding of how subjects think and interpret the topic of interest.
- Design and construct vignettes for a larger quantitative study.

2. Cognitive Interviewing

Cognitive interviewing (CI) is an important technique used to evaluate survey questionnaires and/or identify difficulties in vignettes. CI is used to “understand the thought processes used to answer survey questions and to use this knowledge to find better ways of constructing, formulating, and asking survey questions” (Forsyth & Lessler, 1991). Because vignettes often are presented in written surveys, CI explores respondents’ abilities to interpret vignettes presented within a survey format, assesses whether the wording of questionnaires accurately conveys the objective of the vignette(s) and subjects’ techniques for retrieving information from memory, as well as their judgment formation on the material presented (Willis, 1994, 2006). CI is usually undertaken “between initial drafting of the survey questionnaire and administration in the field” (Willis, 2006, p. 6). Subjects are often asked to “think aloud,” that is, to verbalize their thoughts as they respond to each vignette (self administered or interviewer administered). One researcher usually conducts the interview while a second researcher will observe the interview, tape record, take notes, and transcribe responses. By asking respondents if the proposed vignettes are measuring what they claim
to measure (content validity), an important means of pre-survey evaluation can be established (Polit & Hungler, 1999).

Example of ‘‘Think Aloud’’ exercise and verbal probes: ‘‘While we are going through the questionnaire, I’m going to ask you to think aloud so that I can understand if there are problems with the questionnaire. By ‘‘think aloud’’ I mean repeating all the questions aloud and telling me what you are thinking as you hear the questions and as you pick the answers.’’ Verbal probing techniques can be used as an alternative to ‘‘think aloud exercises’’ allowing the interviewer to immediately ask follow-up questions based on the subject’s response to the vignette. Probes can be both spontaneous and scripted (prepared prior to the interview and used by all interviewers) to test subjects’ comprehension of questionnaire items, clarity, interpretation, and intent of responses and recall. Examples of verbal probes are given below (Grady et al., 2006; Willis, 2006).

- **General probe:**
  - “Do you have any questions or want anything clarified about this study?”
  - “What are your thoughts about this study?”
  - “How did you arrive at that answer?”

- **Paraphrasing probe:**
  - “Can you repeat the question in your own words?”

- **Recall probe:**
  - “What do you need to remember or think about in order to answer this question?’’

- **Comprehension probe:**
  - “What does the term ‘‘research subject’’ mean to you?”

- **Confidence judgment:**
  - “How sure are you that your health insurance covers your medications?”

- **Specific probe:**
  - “Would you consider individuals who consent to participate in this study vulnerable in any way?’’ ‘‘Vulnerable to what?’’ ‘‘Why?’’

3. Field Pre-testing

The purpose of a pretest is to simulate the actual data collection procedures that will be used in a study and to identify problem areas associated with the vignettes prior to administration with a sample (Fowler, 2002; Platek, Pierre-Pierre, & Stevens, 1985; Presser et al., 2004). Pre-testing is carried out with a small convenience sample of subjects, similar to the characteristics of the population planned for in the actual study. It is important to assess if the vignettes and related questions are consistently understood and believable so that researchers can improve on any reported
practical problem(s). Confusing and complex wording, misunderstandings, or misreading of vignettes can lead to missing data that ultimately biases sample estimates, underestimates correlation coefficients, and decreases statistical power (Platek et al., 1985; Kalton & Kaspryzk, 1986; Ulrich & Karlawish, 2006). Problems may also be related to length of completion time and burden or simply typographical errors. Therefore, evaluating response distributions through pre-testing can help the researcher to revise questions. For example, if a vignette with open-ended questions is planned, the pre-test can identify redundant participant responses and help to narrow the range of answers needed to respond. As a result, the researcher might deem fixed response categories as being more appropriate (Fowler, 1995). Revising and refining the instrument based on pre-testing improves the data collection procedures and data quality.

Areas to assess in pre-testing:

- Completion time: How long is the questionnaire? Are all items completed? If items were skipped, why?
- Clarity of wording: Are any items or terms used in the vignettes and questionnaire confusing to respondents? Are questions unidimensional?
- Complication of the instructions: Are the instructions easy to follow, understandable, and comprehensive? Does the questionnaire flow well; does the question order appear logical?
- Is the questionnaire user-friendly? Are the vignettes easy to understand?
- Is the information in the vignettes accurate and unambiguous?
- Are important terms in the vignette defined?
- Do respondents perceive any of the items as too sensitive to answer?
- What was the response rate?

Illustrations of Pre-Survey Testing

Siminoff, Burant, and Younger (2004) conducted 12 focus groups of different ethnicities, including Hispanics, Muslims, and African Americans to guide their research on public attitudes surrounding death and organ procurement. The focus groups were the basis for the development of vignettes that presented varying patient conditions (i.e., brain death, severe neurological damage, and persistent vegetative state). In doing so, these groups provided clarification of the ethical terms in lay language with diverse perspectives, addressed concerns not readily apparent in the literature, and provided input into the final random digit dial version of a survey that included vignettes and was to be conducted with citizens in
Ohio. Following the focus groups, a pre-test of the survey was randomly administered to 51 individuals to further assess the questionnaire and vignettes for clarity, completion time, and reliability of respondents’ classification of when death occurs. Curbow, Fogarty, McDonnell, Chill, and Scott (2006) developed eight video vignettes to measure the effects of three physician-related experimental characteristics (physician enthusiasm, physician affiliation, and type of patient–physician relationship) on clinical trial knowledge and acceptance, and beliefs, attitudes, information processing, and video knowledge. To evaluate the video vignettes, pre-testing was conducted with eight focus groups of former breast cancer patients and patients without cancer. Alterations to the vignettes were made based on participants’ responses.

**EVALUATING PSYCHOMETRIC PROPERTIES OF VIGNETTES**

Before administering vignettes to a sample and analyzing the results, it is important that the vignettes are valid and reliable. Only reliable and valid vignettes should be used to describe phenomena or test hypotheses of interest in a study. A comprehensive review of reliability and validity issues can be found in Litwin (1995) and Streiner and Norman (2003).

*Validity of Vignettes*

Internal validity is important for quantitative and qualitative measures and expresses the extent to which an instrument adequately reflects the concept(s) under study. Because vignettes in bioethics research are often constructed solely for the particular topic under study, validity and reliability are essential to achieve meaningful analyses and interpretation of data. In other words, empirical bioethics researchers often have fewer pre-existing instruments to draw from and need to develop measurement tools de novo (Ulrich & Karlawish, 2006). Thus, vignettes must be internally consistent. Two issues related to the internal validity of vignettes are important: (1) the extent to which the vignette(s) adequately depict the phenomenon of interest and (2) the degree to which each question in response to the vignette(s) measures the same phenomenon (Flaskerud, 1979; Gould, 1996).
Content Validity

Content validity, constituting one type of internal validity, addresses the degree to which an instrument represents the domain of content being measured and is a function of how it was developed and/or constructed (Waltz, Strickland, & Lenz, 1991). Ways to assess content validity of vignettes include using a panel of experts, focus group interviews, and/or CI techniques. Using a panel of experts, at least two (to a maximum of 10) experts in the field are asked to quantify or judge the relevance of vignettes and corresponding questions based on the following criteria: (1) does the vignette adequately reflect the domain of interest?; (2) is the vignette plausible and easily understood?; (3) are the corresponding questions representative of the vignette’s content?; and (4) do the objectives that guided the construction of the vignette correspond with its content and its response set? (Lanza & Cariaeo, 1992; Lynn, 1986; Waltz et al., 1991). Relevancy is generally rated on a four-point scale, from totally irrelevant (1) to extremely relevant (4). A formal content validity index (CVI) can be calculated based on the proportion of experts who rate a vignette 3 or 4. CVI hence indicates the extent of agreement by expert raters on the relevancy of the vignettes. Generally, an index of 0.80 or higher represents good content validity. Haynes, Richard, and Kubany (1995) provide a thorough guide to assessing content validity (Table 1).

**Illustration of Content Validity**

To compare adolescent and parental willingness to participate in minimal risk and above-minimal risk pediatric asthma research protocols, Brody, Annett, Scherer, Perryman, and Cofrin (2005) asked an expert panel of ethicists and pediatric pulmonary investigators to review 40 pediatric asthma protocol consent forms and choose those protocols that represented minimal risk and above-minimal risk. The researchers then

---

**Table 1.** Types of Reliability Important for Quantitative Vignettes.

*Test-retest:* Repeated measurements of the vignettes can determine the stability of the measure’s performance at two distinct time periods with the same group of subjects (generally within two weeks time). A Pearson’s product moment correlation coefficient is calculated; a coefficient closer to 1.00 generally represents stability of the measure. (Non-parametric measures of association are used for nominal and/or ordinal data).

*Internal Consistency:* A measure of an instrument’s reliability is how consistently the items in the scale measure the designated attribute. Cronbach’s alpha is the most widely used measure of reliability and an alpha of 0.70 is considered acceptable.

developed standardized vignettes using key information from each of the selected protocols.

Reliability of Vignettes

A reliable instrument is one that is consistent, dependable, and stable on repeated measurements. Thus, it is free of measurement error (Waltz et al., 1991). Test-retest, inter- and intra-rater, and internal consistency reliability are three different types of reliability that can be reported for quantitative instruments. Test-retest reliability measures how stable study results are over time. Thus, if vignettes are given to respondents on two occasions using a time interval in which conditions for subjects have not changed, vignette results should be comparable. For continuous or ordinal responses (e.g. ratings or rankings), Pearson or Spearman correlation is often used to measure reliability. The kappa coefficient (Cohen, 1960) or a weighted kappa (Cohen, 1968) can be used to measure the agreement between dichotomous response categories.

Intra-rater reliability is similar to test-retest but measures one rater’s variation as a result of multiple exposures to the same stimulus. Inter-rater reliability refers to the stability of responses when rated by multiple raters. When multiple experts rate the same vignette, the intra-class correlation coefficient (ICC) is often used to measure the reliability.

Internal consistency refers to the homogeneity of the items used to measure an underlying trait or attribute via a scale. For a scale to be internally consistent, items in the scale should be moderately correlated with each other and with the total scale score. One option is to use Cronbach’s alpha (Cronbach, 1951) to assess the homogeneity of the scale, with a value of 0.70 generally being considered acceptable. Qualitative vignettes can also be measured for rigor and reliability by addressing the following questions (Lincoln & Guba, 1985):

- How credible are the vignettes?
- Are the findings transferable? How applicable are the findings to other areas of inquiry?
- Are the findings dependable? Was an audit trail or process of verification used to clarify each step of the research process?

Illustration of Reliability

Gump et al. (2000) developed vignettes depicting six ethical dilemmas (two justice-oriented situations, two care-oriented situations, and two mixed orientations), each with 8 response items to test a measure of moral
justification skills in college students. Eight expert judges determined the
degree of representation of the justice and care constructs by the vignettes
and their corresponding items. Test-retest and internal consistency reliability
scores were established for each subscale.

ADDITIONAL EXAMPLES OF
PUBLISHED VIGNETTES USED IN EMPIRICAL
BIOETHICS RESEARCH

In bioethics research, vignettes have been used to address several issues at
the end-of-life, such as withdrawing life support, euthanasia, physician
assisted suicide, neonatal ethics, and other sensitive topics (Alexander et al.,
2003; Asai et al., 1999; Christakis & Asch, 1995; Emanuel et al., 1996;
Freeman, Rathore, Weinfurt, Schulman, & Sulmasy, 1999; Kodadek &
Feeg, 2002; McAlpine et al., 1997; Wolfe et al., 1999). Using a sophisticated
factorial vignette design, Denk et al. (1997) conducted a computer assisted
telephone interview to assess Americans’ attitudes about treatment decisions
in end-of-life cases. To avoid a limited content domain and maturation bias,
several vignettes were randomly presented to participants. Attitudes were
solicited on continuation or termination of costly medical care of critically
ill patients. Manipulated variables in vignettes included patients’ age,
contribution to medical condition, quality of life, type of insurance, and
patients’ right to decide about treatment.

Several authors have used vignettes in self-administered mailed ques-
tionnaires. Mazor et al. (2005) surveyed 115 primary care preceptors who
were attending a faculty development conference to examine the factors that
influence their responses to medical errors. The researchers developed two
medical error vignettes and randomly varied nine trainee-related factors,
including gender, trainee status, error history, and trainee response to error.
In another study, Alexander et al. (2003) developed two clinical vignettes to
study public support for physician deception of insurance companies by
surveying 700 prospective jurors in Philadelphia. The vignettes depicted
clinical situations in which a 55 year old individual (gender was changed for
each vignette) with a known condition required further invasive and/or
noninvasive procedures, determined by their physician, for which the
insurance company would not pay. Respondents were asked to either accept
or appeal the restriction or misrepresent the patient’s condition to receive
desired service.
Similarly, Freeman et al. (1999) developed six clinical vignettes to study perceptions of physician deception in a cross-sectional random sample of internists using a self-administered mailed questionnaire. The vignettes varied in terms of clinical severity and risks ranging from a life-threatening illness to the need for a psychiatric referral to a patient in need of cosmetic surgery (rhinoplasty). Based on the vignettes, respondents were asked to indicate whether a colleague should deceive third party payers and how they, their colleagues, and society would judge such behavior. Although these clinical vignettes were less threatening to respondents than if presented in interviews, they are limited in capturing actual misrepresentation and/or deception in clinical practice.

A few authors have used vignettes within a multi-method framework. Arguing that the use of vignettes with both qualitative and quantitative methods is a powerful tool, Rahman (1996) used long and complex case vignettes with both open-ended and fixed choice responses to understand coping and conflict in caring relationships of elderly individuals. Using an innovative internet survey design, Kim et al. (2005) used qualitative and quantitative approaches to understand the views of clinical researchers on the science and ethics of sham surgery in novel gene transfer interventions for Parkinson Disease patients. Researchers were asked to quantitatively estimate a number of issues, as well as to provide open commentary on their responses (Kim et al., 2005).

**METHODOLOGICAL CONSIDERATIONS**

*Sample Size Estimates for Studies using Vignettes*

The required number of respondents (sample size) for a quantitative study using vignettes varies depending on the study aims and design. Factors that influence the design include the research questions; the type of measurements (e.g. dichotomous forced choice, Likert scale) for respondents to rate, rank, or sort vignettes; the number of vignettes given to each respondent; and the number of respondent and situational characteristic effects to be examined. Once these factors have been determined, sample size estimates can be calculated based upon the statistical analysis planned.

For example, in the simplest case, suppose a study is designed to examine only the effect of respondents’ race (Caucasian and African American) on organ donation. Each respondent would be given a single vignette and asked to rate, on a Likert scale, how willing they are for their organs to be...
donated in the particular situation. The differences in the ratings by race would then be analyzed using a Student’s \( t \)-test, assuming normally distributed responses. Sample size calculations are based on a meaningful minimum difference (or effect size = difference/standard deviation \( \sigma \)) between the mean response rates in the two groups, that is, the smallest difference that will be found to be statistically significant in the analyses (Table 2).

When multiple vignettes are given to each respondent, there is an inherent correlation between responses from the same individual. That is, the response to one vignette is assumed to be related to or affected by the responses given to the other vignettes. When only two vignettes are given to each respondent, analyses and, hence, sample size calculations, can be based on the differences between each individual’s responses. However, as the number of vignettes increases, the within-respondent variance needs to be explicitly taken into account. This is particularly important as the within-respondent variance can potentially be large, and may “swamp” the situational effects of interest. Larger sample sizes are needed in order to detect any situational effects. For example, to test the hypothesis of no differences between two groups or four situations when four vignettes are given to each respondent and comparisons are to be made between two respondent groups, 16 respondents per group may be required when the within-respondent variation is 2, but 60 respondents per group would be required if the within-respondent variation was doubled. For the interested reader, sample size tables can be found in Rochon (1991).

### Table 2. Number of Subject per Group for Two-sided \( t \)-test with \( \alpha = 0.05 \) and Power of at least 80%, Assuming Equal Group Sizes.

<table>
<thead>
<tr>
<th>Difference in means</th>
<th>0.10( \sigma )</th>
<th>0.20( \sigma )</th>
<th>0.30( \sigma )</th>
<th>0.50( \sigma )</th>
<th>0.75( \sigma )</th>
<th>1.00( \sigma )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of subjects needed per group</td>
<td>1571</td>
<td>394</td>
<td>176</td>
<td>64</td>
<td>29</td>
<td>17</td>
</tr>
</tbody>
</table>

**Factorial Designs**

When characteristics in vignettes are manipulated, a factorial design is often employed. If responders are given every possible vignette, the study would be considered a full factorial design. When only a couple of situational characteristics are being studied, a full factorial design may be burdensome to the respondents. For example, if only the gender (male/female) and race (Caucasian/African American) of the hypothetical person in the
vignette were changed, each subject is given $2 \times 2 = 4$ vignettes to respond to. However, when there are a number of characteristics to be changed with multiple levels or categories, the number of vignettes can become excessive. For example, changing 5 characteristics with 2 options for each would result in $2^5 = 32$ vignettes being given to each respondent. In these cases, a fractional factorial design would be appropriate.

In fractional factorial designs, each respondent is given a fraction of all the possible situational combinations being studied. While this design does not affect the sample size, it does limit the hypotheses that can be tested. Generally, a study is designed so that main effects of characteristics can be estimated (e.g. differences between races) but higher order interactions cannot (e.g. age × race × gender), as they are assumed to be zero. If the effect of one situational change does vary with the level of another change and is not taken into account in the design, effect estimates may become biased or confounded. Thus, the fraction to be used is based upon the hypotheses or inferences that are most important in the study as well as hypothesized or known relationships between the situational characteristics. For example, Battaglia, Ash, Prout, and Freund (2006) used a fractional factorial design to explore primary care providers' willingness to recommend breast cancer chemoprevention trials to at-risk women. Five different dichotomous characteristics, including age, race, socioeconomic status, co-morbidity, and mobility, were manipulated in clinical vignettes to assess physician decision-making. Using all five characteristics would yield 32 possible vignettes in a complete factorial design ($2^5$). To reduce this number, a balanced fraction approach of half of all possible vignette combinations were used and participants were asked to respond to one of the 16 versions of the vignette.

**Hypothesis Testing**

In quantitative hypothesis-testing studies, and when each respondent is only given one vignette, standard statistical methods such as $t$-tests, analysis of variance (ANOVA), and nonparametric tests can be used to analyze the data. In studies when multiple vignettes are considered by each respondent, statistical methods need to account for the inherent correlation between measurements from the same respondent, as noted above. The data are considered balanced if every pair of situational characteristics is presented to respondents an equal number of times, and complete if there is no missing data. For balanced and complete data, repeated measures ANOVA or
analysis of covariance (ANCOVA) can test the hypothesis of interest. Issues of multiple comparisons need to be addressed before using any of these models. If the data are not balanced or complete, more advanced statistical methods can be used, such as linear mixed effect models (Laird & Ware, 1982) or generalized estimating equations (GEE) (Zeger & Liang, 1986). These methods also allow for testing complex correlation structures that may emerge in some studies.

VALIDITY OF CONCLUSIONS

External validity, or the extent to which generalizations can be made from the study sample to the population, is limited in studies employing hypothetical vignettes since vignettes may not reflect the clinical nuances of real-life situations. Therefore, caution must be used in interpreting predictive relationships between what participants report “ought to be done” in constructed vignettes and actual behaviors. However, vignettes provide a means for understanding attitudes, opinions, and beliefs about moral dilemmas. Rahman (1996) notes that findings will be more generalizable when a vignette is closer to real-life situations.

STRENGTHS AND LIMITATIONS OF HYPOTHETICAL VIGNETTES

Vignettes in bioethics research pose many practical advantages. They are economical; gather large amounts of data at a single time; provide a means of assessing attitudes, beliefs, and practices on sensitive subject areas; are less personal and threatening than other methods; and avoid observer influences (Alexander & Becker, 1978; Finch, 1987; Flasketud, 1979; Gould, 1996; Hughes, 1998; Hughes & Huby, 2002; Rahman, 1996; Schoenberg & Ravdal, 2000; Wilson & While, 1998) (see Table 3).

Vignettes have been criticized, however, for their limited applicability to “real life.” Hughes (1998) argues that the characters, social context, and situation of vignettes must be presented in an authentic, relevant, and meaningful way for participants. Vignettes are also subject to measurement error, which may lead to “satisficing.” Stolte (1994) described satisficing “as a tendency for participants to process the vignettes less carefully than under real conditions” (p. 727). This may occur because of difficulties completing
the interpretation of and/or response to a vignette, or because of insufficient motivation of participants to perform these tasks. In turn, participants’ responses may be biased or incomplete by simply choosing the first presented response option that seems reasonable, acquiescing to common assertions, randomly choosing among the offered responses, failing to differentiate responses on a particular measure, or reporting a “don’t know” answer (Krosnick, 1991). Attention to contextual factors, such as interview setting, participant compensation, instrument attributes, and mode of administration may help to control satisficing (Stolte, 1994).

**CONCLUSION**

Although the use of vignettes in bioethics research has largely focused on end-of-life care issues, this type of data collection method provides a
practical and economical means to understanding complex, challenging, and burgeoning ethical concerns. Vignettes can be constructed from a variety of sources and can be presented in several formats. The method is flexible in allowing the researcher to manipulate experimental variables of interest for a sophisticated analytic design. It can be incorporated in mailed questionnaires (paper-and-pencil method or web-based approach) or be administered face-to-face. Caution must be given in generalizing research findings based on the use of vignettes since responses regarding hypothetical behaviors are not necessarily indicative of actual behaviors. Overall, however, reliable and internally valid vignettes provide an important means to empirically advance our understanding of bioethics in key arenas such as clinical practice, research, and policy.

REFERENCES


Wendler, D., & Emanuel, E. J. (2002). The debate over research on stored biological samples. *Archives of Internal Medicine, 162*, 1457–1462.


DELiberative PROCEDURES
IN BIOETHICS

Susan Dorr Goold, Laura Damschroder and Nancy Baum

abstract

Deliberative procedures can be useful when researchers need (a) an informed opinion that is difficult to obtain using other methods, (b) individual opinions that will benefit from group discussion and insight, and/or (c) group judgments because the issue at hand affects groups, communities, or citizens qua citizens. Deliberations generally gather non-professional members of the public to discuss, deliberate, and learn about a topic, often forming a policy recommendation or casting an informed vote. Researchers can collect data on these recommendations, and/or individuals’ preexisting or post hoc knowledge or opinions. This chapter presents examples of deliberative methods and how they may inform bioethical perspectives and reviews methodological issues deserving special attention.

In the face of scarcity, deliberation can help those who do not get what they want or even what they need come to accept legitimacy of a collective decision.

–Gutmann & Thompson, 1997
INTRODUCTION

Bioethical issues are, by definition, morally challenging. Topics in bioethics of interest to empirical researchers frequently carry enormous policy relevance, and would benefit from informed, reflective public input. Unfortunately, such topics as stem cell research, “rationing,” cloning, and organ transplantation tend to be technically and conceptually complex, intimidating and sometimes even frightening, making public input difficult. Deliberative procedures, based on deliberative democratic theory, may be an appropriate choice when either: (a) an informed opinion is needed but difficult to obtain; (b) individual opinions will benefit from group discussion and insight; and/or (c) group judgments are relevant, usually because the issue affects groups, communities, or citizens. The development of health policy (including many bioethics issues) can benefit from several of these characteristics, and potentially from deliberative public input as well.

Theories of deliberative democracy, despite important differences, share an emphasis on political decision making that relies on a process in which political actors listen to each other with openness and respect, provide reasons and justifications for their opinion, and remain open to changing their point of view through a process of discourse and deliberation. Deliberation has been justified by appeals to develop a more informed public (Fishkin, 1995), create decisional legitimacy (Cohen, 1997), and/or claim that participants in deliberations and their constituents have consented to informed decisions (Fleck, 1994). Just as traditional bioethical principles have been invoked to ensure that individual patients have a voice in their own medical decision making, so might deliberation provide community members with a “voice” in community-wide decisions, for instance about health spending priorities (Goold & Baum, 2006) or research regulation. Deliberative procedures offer an opportunity for individuals to assess their own needs and preferences in light of the needs and desires of others. Morally complex decisions may enjoy public legitimacy if they are the result of such fair and public processes. Beyond legitimacy, individuals involved in a community decision-making process about bioethical issues with policy implications may be more likely to accept even intensely difficult decisions if they feel they have had an opportunity to fully understand and consider the issues and to contribute to the final resolution.

WHAT ARE DELIBERATIVE PROCEDURES?

In general, deliberative procedures call for gathering non-professional (non-elite and lay) members of the public to discuss, deliberate, and learn
about a particular topic with the intention of forming a policy recommendation or casting an informed vote. Some deliberative procedures aim primarily to inform policy. Occasionally, deliberative procedures include both research and policy aims. We will focus this chapter on de novo deliberative procedures used for research purposes, or combined policy and research purposes, where research aims are known and planned up front.

For researchers, deliberative procedures can be a valuable tool for gathering information about public views, preferences, and values, and provide some advantages over other methods. Although it is common practice to conduct public opinion polls or nationally representative surveys to gauge public views on particular policy questions, opinions measured in this way can be unstable, subject to manipulation, poorly informed, or individuals may not have formed opinions (Bartels, 2003). This may especially be the case when issues are morally and/or technically complex. For example, a public survey about a rare genetic test is likely to encounter ignorance about genetics, genetic testing, and how a particular condition may impact health. Respondents may refuse to answer, respond despite their lack of knowledge, or respond based on flawed information. Also, using a survey, one may encounter a great deal of influence from the way the questions are framed. Furthermore, surveys often fail to capture the “public” aspect of public input, since the information collected is aggregated individual opinion. Deliberative methods prompt a discussion about what we should do as a political community; participants in deliberation are encouraged to reconsider their opinions in light of the interests of others. Like national polls, deliberative efforts can engage a representative (though smaller) sample from a constituency. Unlike national polls, deliberative procedures emphasize reasons and rationales for and against an issue or policy as a natural part of the deliberative process.

Individual and group interviews, while providing opportunities for reflection, generally do not aim to inform participants. Individual interviews would be poorly suited to discovering what people think as a community (rather than as individuals). “Town hall meetings” have been used to gather public input on policies, although they can also include time spent informing the audience. However, they may suffer from heavily biased attendance and often are structured more for political than research objectives.

**OPERATIONALIZING DELIBERATIVE PROCEDURES**

In this section, we emphasize methodological issues unique to, or particularly important for, deliberative methods. Similar issues and concerns can be
found in other methods, for instance the role of group dynamics that arises in focus group research. We emphasize a limited number of methodological questions, and recommend that the reader turn to the many other excellent chapters in this volume and other resources related to specific portions of empirical research (e.g., survey question wording). To illustrate some of the issues, we provide two examples of deliberative projects.

**Representation, Recruitment and Sampling**

Representation, recruitment, and sampling are key aspects of research using deliberative procedures. Randomized selection methods can be used to recruit participants into a deliberative study, however, since deliberation nearly always involves gathering participants into groups, assembling them inherently introduces bias, which has both research and policy implications. In deliberative projects, even more so than, for instance, focus group projects, equal participation and equal opportunity for participation on the part of citizens is vital. A sample that omits important voices of those affected by the issue at hand, or a deliberative group that stifles certain points of view because of a dominant point of view or experience, an imbalance in perceived power, or stigma or sensitivity, undermines the legitimacy of the process and can lead to dissatisfaction, distrust, and/or recommendations that are not truly reflective of the public. There is some evidence that heterogeneous groups deliberate more effectively than homogenous groups (Schulz-Hardt, Jochims, & Frey, 2002). However, it may be important to have homogenous groups when some participants might not otherwise speak freely about sensitive issues, for example, mental illness or racism.

Substantive representation presents an alternative to random (proportional) sampling, and entails selecting those most affected by a policy (Goold, 1996). It resembles the practice of convening “stakeholders,” although distinguishes those most affected from those with the greatest interests at stake. For example, medical researchers have an interest in policies for human subjects research but the policies disproportionately affect the human subjects themselves. In addition, the opinions of researchers and laypersons on the topic are likely to be quite different, researchers often already have a forum for input on the topic, and, finally, researchers’ expertise could stifle laypersons’ comfort and active participation in deliberations. In this example, one can justify the disproportionate inclusion of laypersons in a deliberative project about policies for human subjects research, and should ensure that researchers (if they are involved in the project) are in
groups separate from laypersons. Importantly, community-based research (of which deliberative methods represent one type) includes input from the community at the beginning of the project to guide priorities and needs; this input can identify groups that would be most important to include and to identify types of groups where homogeneity can be especially advantageous.

Early community involvement in the research process can help with recruitment of participants and add legitimacy to recruitment methods. Most projects with policy-making implications should include a component of public advertising in recruitment to ensure that important voices are identified and heard and that the project adheres to standards of openness. Careful screening of volunteers, including questions about motivation for participation, can help minimize the potential for bias.

Methods and Structure of Deliberation

Deliberative sessions can last as long as 4 days (Rawlins, 2005; Lenaghan, 1999), a weekend (Fishkin, 1995), or just a single day (Damschroder et al., 2007; Ackerman & Fishkin, 2002). A deliberative project may also consist of a series of discussions over time (Goold, Biddle, Klipp, Hall, & Danis, 2005). Sometimes a large group (several hundred individuals) breaks into smaller groups and then reconvenes.

Deliberative methods typically include: educating and informing, discussion and deliberation, and describing and/or measuring group (and often individual) views. Balanced education is one approach commonly used to ensure that the information provided to participants meets their needs and enhances credibility. In one arrangement, experts who represent a variety of perspectives on the topic present information from their respective perspectives, and then respond to questions from participants. The opportunity for participants to construct their own questions helps guard against undue influence by the research team in what or how information is provided. The use of “competing experts” allows participants, like a jury in a court case, to judge for themselves the credibility of particular experts based on responses to questions raised by participants. However this type of approach can present a problem when an issue is of an adversarial nature. Participants may take sides rather than listen to experts and fellow deliberators with an open mind. It can also be frustrating if participants sense there are no “right answers.” Alternatively, printed or other materials rather than (or in addition to) experts can be provided to participants.
Discussion and deliberation should be led by professional facilitators, trained specifically for the deliberation at hand, whenever feasible. Trained facilitators avoid influencing participants (“leading”) and ensure that participation in the deliberation is as equally distributed as possible. Even more important in deliberations than in focus group research, dominant personalities need to be diffused and the opinions and perspectives of quieter participants actively sought. A round robin method, nominal group technique, or other approaches should ensure that all participants have an opportunity to speak. Equally important, participants must feel comfortable speaking; besides group composition, facilitator characteristics (men leading a group of women discussing sex, for instance), may be important considerations.

The structure of deliberative methods ranges from highly unstructured protocols, starting with an open-ended general question about a topic, to highly structured sequences of tasks. The structure of deliberations can profoundly influence the credibility and legitimacy of the method for public input on policy as well as the rigor of the research. Less structure in the outline for deliberations is advantageous because participants have more freedom to frame or emphasize issues from their own perspective. It is possible, however, that participants will not stay on task or obtain needed information, resulting in an unfocused discussion of less important aspects of the issue at hand. Maintaining the appropriate balance between greater structure and more openness depends on the topic, time available, resources, and other factors. Deliberative groups need enough structure to remain on task and cover important information domains, but participants also need enough flexibility to raise questions and process responses.

WHAT TO MEASURE AND WHEN?

Research aims determine what data to collect, as well as how and when to collect it. The following includes examples of data that might be collected in a deliberative project:

1. Individual participant viewpoints *ex ante* relevant to the topic
2. Relevant characteristics or experiences (e.g., participation in research or out-of-pocket health spending)
3. Political engagement, self-efficacy, judgment of social capital (before and/or after deliberation)
4. Individual participant viewpoints *ex post* relevant to the topic
5. Views on the deliberative process (e.g., others’ sincerity, chance to present views, group decision)
6. Group dialog and/or behavior
7. Group decisions or recommendations (or lack thereof)
8. Impact on policy

Typically, data collected in deliberative projects include a combination of survey responses, group dialog, and group recommendations, decisions, and statements. Researchers may want to know if individuals change with respect to knowledge or opinion as a result of the deliberative process, and so propose to measure a given variable before and after deliberation. Data collected about individuals, besides the usual demographic information, can include pre-deliberation opinions, knowledge of the topic, and measures of characteristics that are likely to influence views. For example, if you are measuring opinions about mental health parity, personal or family experience with mental illness would be a relevant variable to include. Other data that is often valuable to collect, particularly after group deliberation, includes judgments and views of the group process or the group’s final decision. In the project described below, for instance, we used measures of perceived fair processes and fair outcomes (Goold et al., 2005).

Group dialog can be audiotaped and, if needed, transcribed. Observation of group dialog and group behavior can include structured or open-ended options to document the group dynamic or process. For example, observation can document the distribution of participation in conversation, decision-making style, dominance of particular individuals in the discussion, judgments about the group’s cooperative or adversarial decision-making processes, and the like. Group dialog lends itself to a number of analytic possibilities. One can analyze dialog for the reasons, rationalizations, arguments, or experiences used to justify points of view. One can analyze the quality of reasoning, the persuasiveness of arguments, or characteristics of group dialog that influenced individual or group viewpoints.

EXAMPLES OF STUDIES USING DELIBERATIVE METHODS

Veterans, Privacy, Trust, and Research

The Federal Privacy Rule was implemented in the United States in 2003, as part of the Health Insurance Portability and Accountability Act of 1996
HIPAA), with the hope that it would address growing concerns people had about how personal medical information was being used in contexts outside of medical treatment. However, the Rule has affected research in unanticipated ways. Researchers generally can only access medical records if they have permission from each individual patient or they obtain a waiver of this consent requirement from an oversight board (an Institutional Review Board (IRB) or a privacy board). For researchers who need to review thousands of medical records going back a long period of time, obtaining individual authorization is difficult if not impossible (US HHS (United States Department of Health and Human Services), 2005). Requiring patient permission for each study can add significant monetary costs and result in selection biases that threaten the validity of findings (Armstrong et al., 2005; Ingelfinger & Drazen, 2004; Tu et al., 2004).

**Study Aim**

In the study described below, the investigators wanted to learn what patients thought about researchers’ access to medical records and what influenced those opinions.

**Representation and Sampling**

A sample of 217 patients from 4 Veteran Affairs (VA) health care facilities deliberated in small groups at each of 4 locations. They had the opportunity to question experts and inform themselves about privacy issues related to medical records research and patient privacy. Participants were recruited from a randomized sample of patients (stratified by age, race, and visit frequency) from four geographically diverse VA facilities to participate in baseline and follow-up phone surveys. Ensuring balanced numbers of older and heavier users of the health care system was important in order to gain insight into whether these patients were more sensitive about researchers using their medical records or, conversely, whether they were more supportive of the need for research compared to those who have a lower burden of illness. Additionally, the sampling approach used in this study sought to ensure that the voices of people with racial minority status, particularly African Americans (since they represent the vast majority of non-white patients for the four sites) were adequately represented. In past studies, African-American patients have expressed a reluctance to participate in clinical research (Corbie-Smith, Thomas, Williams, & Moody-Ayers, 1999; Shavers, Lynch, & Burmeister, 2002) and have exhibited lower levels of trust in researchers than white patients (Corbie-Smith, Thomas, & St George, 2002).
Deliberation Procedure and Data Collection

Participants who completed a baseline survey were invited to an all-day deliberative session. Each participant was randomly assigned to a group of 4–6 individuals and each deliberative session included 9 or 10 of these smaller deliberative groups. A non-facilitated deliberative process was chosen to minimize researcher bias and to encourage a fresh approach to the complex issues at hand. Small groups used a detailed, written protocol starting with a review of background information about medical records, minimal risk research, and the HIPAA Privacy Rule. Participants were asked to imagine that they were acting as an advisory committee for a “research review board … [that] judges whether a research study will pose minimal risk and whether the study will adequately protect private information.” The protocol and background information explained that researchers cannot use personally identifiable medical records in a research study unless the IRB agrees that three waiver criteria have been met. Small group deliberations were interspersed with larger, plenary sessions led by presentations from experts in medical records research and privacy advocacy. Participants had the opportunity to pose questions to the experts and hear the answers as a plenary group.

Baseline and follow-up surveys (including some completed on-site the day of the session) elicited the level of trust in various health care entities, attitudes about privacy, prior knowledge about research and privacy, and general demographic information. Analysis followed a mixed-methods approach, combining qualitative data from deliberations with quantitative data from baseline surveys.

Results

Baseline opinions of participants confirmed that many aspects of the topic for deliberation were unknown to participants; 75% did not know that sometimes researchers could access their medical records without their explicit consent and 39% had never heard of the HIPAA Privacy Rule. The issue was value-laden on the heels of much publicity surrounding implementation of the Rule; 75% of participants said they were very or somewhat concerned about privacy but 89% also said that conducting medical records research in the VA was critically or very important.

When asked whether they were satisfied with provisions of the HIPAA Privacy Rule, 66% wanted a procedure in place that would cede more control to patients over who sees their medical records. However, no consensus was reached as to how much control should be in the hands of patients. After deliberation, nearly everyone (96%) said they would be
willing to share their medical records with VA researchers conducting a study about a serious medical condition.

Participants’ trust in VA researchers was the most powerful determinant of the degree of control they recommended for patients who overuses their medical records. Mechanic and Schlesinger’s (1996) work on trust inspired a framework to describe trust between patients and a medical research enterprise: (1) Are medical records kept confidential? (2) Does the research being conducted demonstrate high priority on patient welfare? (3) Are researchers held accountable and responsible for protecting privacy? (4) Are systems to protect medical records sufficiently secure? (5) Do researchers fully disclose the research being conducted and how medical records are used to conduct that research? (Mechanic, 1998). Participants reported the need to see and understand how their records are kept private, that violators will be punished consistently and relatively severely (e.g., job loss, fines), and assurance that computerized systems are truly secure. Participants expressed the need to see that the institution’s research actually benefits patients and is not subject to conflicts of interest. They wanted transparency in the research process in terms of what research is being done and how their medical information may have contributed to new findings. Further analyses will explore reasons for the apparent contradiction that participants are willing to share their own information, yet united in their call for more control over how their medical records are used.

Deliberation changed opinions in that the overwhelming majority of participants who said they would be willing to share their medical records with VA researchers actually increased compared to baseline willingness (89%). When asked how important it was for researchers to obtain permission for each and every research study before deliberation, 74% said it was critically or very important to do so at the time of the baseline survey, whereas immediately and 4–6 weeks after deliberation, only 48% and 50%, respectively, held this position.

Participants were satisfied with the deliberation process. When asked anonymously, 98% of participants thought the deliberation process was fair, 94% felt others in their group listened to what they had to say, and 89% said they would make the same recommendation as the one made by their group as a whole (through consensus or voting). One participant summarized his experience by saying, “...with more exposure and thought, my decisions are more in line with my moral values.”

In this example, a single, one-day deliberative session generated informed, and possibly more public-spirited, points of view about an important, complex, and morally challenging health policy issue.
Health Spending Priorities

How to allocate and prioritize limited resources fairly and openly is perhaps the most pressing moral and practical concern in the health care arena today; certainly the issue is of importance to nearly all citizens. Setting health care priorities for the use of scarce public resources requires attention to both economics and justice; justice, in turn, is enhanced by the participation of those most affected by the decisions. According to this participatory conception, rationing decisions should incorporate the preferences and values of those affected (Eddy, 1990; Menzel, 1990; Fleck, 1992, 1994; Goold, 1996; Emanuel, 1997; Daniels & Sabin, 1998). Engaging and involving citizens in health care priority setting, however, confronts obstacles. Because health concerns lack salience for most healthy citizens, decisions tend to be influenced heavily by health care experts or those, like the disabled or senior citizens, with specific interests (Kapiriri & Norheim, 2002; Goold, 1996; Jacobs, 1996). Discussions about health services and financing can be complex, making citizens frustrated or intimidated and reliant on others to decide for them, trusting that the services they need will be provided to them in the event of illness (Goold & Klipp, 2002). Talking about future health care needs requires that people think about illness and death; the emotionally laden tension when money appears to be pitted against health in rationing decisions has been a thorny problem for proponents of the “citizen involvement in rationing” model (Ham & Coulter, 2001).

Besides the complexity and value-laden nature of the problem, health care trade-offs involve pooled (often public) resources. Individual health and health care priorities must be balanced by the current or future needs of others in a community. The need for interpersonal trade-offs, and the balancing of individual with social or group needs requires either procedural justice (i.e., fair processes for decision making) or distributive justice (i.e., fair distribution of benefits and burdens), or, ideally, both. The topic of health care spending priorities lends itself well to deliberative procedures.

Accordingly, Drs. Goold and Danis designed CHAT “Choosing Healthplans All Together,” a simulation exercise based on deliberative democratic procedures in which laypersons in groups design health insurance benefit packages for themselves and for their communities. Participants make trade-offs between competing needs for limited resources. To test the choices they have made as individuals and as a group, participants encounter
hypothetical “health events” that illustrate the consequences of their priorities, which inform subsequent group deliberations.

**Study Aims**
The project described below (for full details see Goold et al., 2005) aimed to describe public priorities for health benefits and evaluate the CHAT exercise as a deliberative procedure. We assessed aspects of feasibility, structure (including the content of the exercise), process, and outcomes.

**Representation and Sampling**
Participants were North Carolina residents without health care expertise, recruited from ambulatory care and community settings. Groups were homogeneous with regard to type of health insurance coverage (Medicare, Medicaid, private, uninsured) and heterogeneous with regard to other characteristics (gender, age, race). Low-income participants were over-sampled to assess whether the exercise was accessible and acceptable since typically, these groups are less well represented in policy decisions.

**Deliberation Procedure and Data Collection**
The CHAT simulation exercise is a highly structured, iterative process, led by a trained facilitator using a script, and progresses from individual to group decision making. Data collection included pre-exercise questionnaires about demographics, health and health insurance status, health services utilization, out-of-pocket costs, and the importance of health insurance features (Mechanic, Ettel, & Davis, 1990). Post-exercise questionnaires rated participant enjoyment of CHAT, understanding, ease of use, and informativeness (Danis, Biddle, Henderson, Garrett, & DeVellis, 1997; Biddle, DeVellis, Henderson, Fasick, & Danis, 1998; Goold et al., 2005). Other items asked participants to rate their affective response to the exercise, perceptions of the group process, outcome of decision making, informational adequacy, and range of available choices. Half the group discussions were tape-recorded to analyze the values, justifications, and reasons expressed by participants during deliberation.

**Results**
Five hundred sixty-two individuals took part in 50 sessions of CHAT. Transcripts of group deliberations were analyzed to understand the reasoning, values, and justifications participants emphasized during
deliberation. Over 60 themes were identified and organized into four major
categories: (1) Insurance as Protection against Loss or Harm; (2) Preferences
for the Process of Care; (3) Economics or Efficiency; and (4) Equity or
Fairness.

Insurance as Protection against Loss or Harm described the greatest
proportion of the dialog (55% of coded text) and had the largest number of
sub-themes. Dialog was coded under this theme when participants justified
coverage selections on the basis of planning for future health care needs and
avoiding future harms or losses through adequate insurance coverage. For
example, participants discussed the likelihood that they or others would
suffer health related losses/harms in the future:

*I know in our group we had chosen medium coverage because depending on [heredity] and
your age, you may have a lot more dental problems and it becomes certainly more
important as you age. So that would be my choice.*

Participants also frequently discussed the types (financial, physical, social,
and emotional) and the magnitude of the harms or losses that might be
avoided, either because care would be prohibitively expensive without
coverage, because *not* having coverage could lead to serious harm, or
because coverage was essential for large numbers of people.

The second major category, Economics or Efficiency, represented
participants’ concerns about costs and resources (20% of coded text). Such
concerns included the use/abuse of insurance, the perceived value of care,
and issues of supply and demand of health care services. Participants were
concerned with waste in the health care system, and recognized the potential
for increased use related to insurance coverage – the concept of “moral
hazard”:

*Looking in the long run, if you, specialty just staying at the basic is kind of like a gate
keeper. If you’re thinking of group coverage, some people may, I don’t, but some people
may jump to a specialist quicker than they need to if they have the free choice of going
there.*

Dialog coded in the third theme, Preferences for the Process of Care,
included concerns for quality of care, the ability to choose physicians and
treatment facilities, system “hassles” such as the wait time for appointments
and, in some instances, one’s own beliefs or attitudes about health care
services. This example illustrates the value placed on timely access:

*This is very important. I think …. I mean, just the thought of waiting four weeks for a
routine appointment. What do they consider routine?*
The fourth major theme, Equity or Fairness, (8% of coded dialog) included discussions of justice, including dialog about equality, personal responsibility, and social responsibility, especially a responsibility to care for the worse off:

...because there are people of course that just simply can’t afford $50 a day. It would be an extreme hardship on them....

Several policy and research projects have used CHAT to involve citizens in health benefit design. Evaluation data from the project reported above and several other CHAT projects shows that participants, including low-income and poorly educated participants (Goold et al., 2005):

1. Find CHAT understandable, informative, and easy to do
2. Judge the fairness of the group process and decision favorably
3. Would be willing to abide by the decisions made by their groups
4. Gain an understanding of the reality of limited resources and trade-offs between competing needs
5. Alter the choices they make for themselves and their families after the exercise, for instance by more frequently choosing coverage for mental health services.

As is true for many deliberative methods, however, much more critical and rigorous evaluations of processes and outcomes are needed.

**DATA MANAGEMENT, ANALYSIS, AND INTERPRETATION: SPECIAL ISSUES**

Since deliberative procedures always involve gathering individuals into large or small groups for discussion, any data analysis using individuals as the unit of analysis needs to adjust for clustering effects. Owing to the fact that membership in a particular deliberative group discussion and other events will vary from group to group, individual responses on any post-deliberation measures may be affected. Another important issue for deliberative procedures is the need, for many researchers, to compare individual responses before and after deliberations. Since merely measuring something once can work as an intervention (a well documented effect in the educational research literature), it is better to either (a) use a control group that does not participate in deliberations, or (b) ensure the sample size is large enough to randomize a portion of participants to complete the measures only after group deliberations, and analyze the impact of
completing the measure prior to deliberations on post-deliberation responses to the same measure. A related concern inherent in deliberative methods relates to the effect of the group on the individual. Do individuals tend to change their point of view depending on the group’s overall point of view? What about particular events (knowledge statements, e.g.,) that occur in some groups but not others? The recognition or mention of a particular issue (e.g., discrimination) could sway whole groups and many individuals to a different opinion. Finally, while group dialog is a valuable and rich source of information, it can be a challenge to manage. Transcription can be error-prone and should be checked in the usual manner. Transcribers who have not been present at the group discussion are unlikely to be able to recognize the gender of speakers, much less consistently identify individual speakers.

Policy makers and other users of data from deliberative procedures often raise concerns about the representativeness of non-random sampling, inevitable whenever recruitment takes place into groups. When this occurs, it is important to acknowledge the limitations of non-random sampling but compare that limitation with the limitations of other methods, including comparisons with proportional representative sampling for surveys where generalization may be easier but the responses may be less informed, less insightful, more vulnerable to framing, and have other profound limitations. It is also important to talk about alternatives to proportional representative sampling, such as substantive sampling or, often more familiar to policy makers, the inclusion of important stakeholders. Compare, for example, a randomly selected public opinion survey related to high-risk brain surgery for Parkinson’s Disease to information gained from deliberative groups that purposefully includes family members of those with Parkinson’s, those with early Parkinson’s Disease, or those otherwise at high risk of having the condition.

A variety of analytic techniques may be used in deliberative studies. Analyzing individual responses to survey items, before and/or after deliberations, will typically include descriptive statistics, but also should, when possible, include analyses of the responses of important subgroups. For example, in a project about health care resource allocation, subgroup analyses might include low-income or chronically ill participants. The specific approach to analysis of group dialog will be guided by the study’s specific research aims and how much is known about public views on the issue. For example, although group discussions about resource allocation might include many valuable insights, research might focus on dialog about the relative importance of mental health services. Some analysis will be
straightforward (e.g., labeling comments about job discrimination in a
discussion of privacy), while some may require more interpretation (e.g.,
expectations of beneficence as an element of trust in medical researchers).

APPLICATIONS AND IMPLICATIONS
OF THE METHOD

Deliberative procedures have been designed for public input into policy
making, and hence have special applications for that arena. For example,
where policies and/or regulations are relatively silent about surrogate
decision making for research participation, deliberative methods can help
learn how the public feels about degrees of risk, informed consent, and other
issues. As such, these methods can be useful for the examination of the
public’s views on bioethical issues that are often complex and benefit from
public reflection, discourse and understanding.

SUMMARY

In this chapter, we have tried to review briefly the use and application of
deliberative procedures to empirical research questions in bioethics,
illustrated with two projects. Like many other methods, the use of
deliberative procedures is varied and consequently the strengths, weak-
nesses, and implications of research results vary as well. Statistical
(proportional) generalization is not a strength of deliberative procedures
since convening groups will eliminate the ability to consider a sample truly
random. However, deliberative procedures can gain public opinion that is
informed, reflective, and more focused on the common good than on
individual interests.

Deliberative procedures hold a great deal of promise for research on
relevant bioethical policy questions. However, like any other method,
there is a need for conceptual and empirical research to better define
when deliberative procedures are most appropriate, describe the impact of
particular methodological choices, and improve our ability to draw
conclusions. It may be tempting to regard deliberative procedures as simply
the means to a better end – the end being “better” decisions and outcomes.
Proponents of this outcome-oriented view evaluate only the products of
deliberations (Abelson, Forest et al., 2003; Rowe & Frewer, 2000).
Evaluating only outcomes, however, misses the normative argument that “good” deliberative democratic processes can be valued in and of themselves, and that the procedures can be justifiably criticized if they fail to meet normative procedural standards, for example, fair representation or transparency.

Research on deliberative methods should examine, in particular, the impact of choices of sampling methods, group composition (relatively heterogeneous or homogeneous), and deliberations’ structure. Researchers using deliberative procedures should be encouraged to include these and other sorts of “methods” questions, as survey researchers have included research aims that address issues of framing, question ordering, and the like. Recently, a number of scholars (Abelson, Eyles et al., 2003; Fishkin & Luskin, 2005; Neblo, 2005; Steenberger, Bächtigerb, Spöndlib, & Steinerab, 2003) have begun to examine and evaluate deliberative procedures, and a few (e.g., Fishkin) have used deliberative procedures directly to answer research as well as policy questions. Researchers should use and interpret the results of deliberations carefully until more is known about the influence of particular aspects of the method.

There are sound theoretical and philosophical reasons for involving the public directly in policy decisions, including policy decisions in bioethics. Deliberative methods present a promising research approach to addressing morally, technically, and politically challenging policy questions, with the additional advantage that research and policy aims can, at times, be fruitfully combined.

REFERENCES


This page intentionally left blank
INTERVENTION RESEARCH IN BIOETHICS

Marion E. Broome

ABSTRACT

This chapter discusses the role of intervention research in bioethical inquiry. Although many ethical questions of interest are not appropriate for intervention research, some questions can only be answered using experimental or quasi-experimental designs. The critical characteristics of intervention research are identified and strengths of this method are described. Threats to internal validity and external validity are discussed and applied to a case example in bioethical research. Several recent intervention studies that were federally funded in the area of informed consent are discussed, and recommendations for future intervention research are presented.

INTRODUCTION

Empirical research in bioethics has been defined as “the application of research methods in the social sciences (i.e., anthropology, epidemiology, psychology, and sociology) to the direct examination of issues in medical ethics” (Sugarman & Sulmasy, 2001, p. 20). Empirical research methods
have been successfully applied to several areas of bioethical inquiry, such as informed consent, education of health professionals in ethical reasoning, assessment of patient–provider communication preferences, end-of-life decision making, and assessment of the effectiveness of various interventions.

The purpose of this chapter is to discuss intervention studies applied to the study of bioethical phenomena. The chapter will review topic areas to which intervention research has been successfully applied, describe critical design elements of such research, discuss the strengths and challenges of this approach, and provide an in-depth analysis from a specific intervention study on informed consent. Finally, recommendations will be made for future research in bioethics that may benefit from intervention methods.

**INTERVENTION STUDIES IN BIOETHICS**

The use of empirical methods in the study of bioethical inquiry has increased over the past two decades. Sugarman, Faden & Weinstein (2001) conducted an analysis of empirical studies posted in BIOETHICSLINE during the decade of the 1980s. At that time 3.4 percent or 663 of the postings were reports of empirical research. In a subsequent analysis of the reports of empirical studies in bioethics in MEDLINE in 1999 (which by then had subsumed the BIOETHICSLINE database), the number of postings had doubled from 0.6 percent of all MEDLINE postings in 1980–1984 to 1.2 in 1995–1999 (Sugarman, 2004). The most frequently studied topic, regardless of type of empirical approach, was informed consent with physician–patient relationship and ethics education among the top 30 of 50 topics overall. Since 2000, studies have reported on the effectiveness of different types of materials (written, tailored, videotapes, etc.) in increasing individuals’ understanding of a health condition they have or are at risk for (Skinner et al., 2002; Rimer, et al., 2002). Sugarman identified eight types of empirical research ranging from “purely descriptive studies” to “case reports.” Only one type is based on interventional research which Sugarman calls “demonstration projects.” This multitude of types of empirical approaches is very appropriate in the field of bioethics as, the overwhelming majority of the topics in bioethics would not lend themselves to intervention studies, or manipulation of the independent variables, such as medical care at end-of-life, confidentiality, or euthanasia (Sugarman & Sulmasy, 2001).
EXPERIMENTAL DESIGNS

Many descriptive studies in bioethics have examined issues surrounding clinical trials, including therapeutic misconception, informed consent, placebo groups, randomization, physician–scientist conflict, and the provision of drugs for individuals who are addicted (Friedman, Furberg, & DeMets, 1998; Sugarman, 2004). Such studies are designed to build a descriptive body of knowledge generating hypotheses that can be tested in intervention studies (Sugarman, 2004) and by now some topics have been sufficiently well described to suggest that interventions be developed and tested. The purpose of a well-designed clinical trial is to prospectively compare the efficacy of an intervention in one-group of human participants to the effects in another group of individuals to which no treatment was intentionally administered. The application of experimental designs, such as the randomized controlled trial (RCT), has been limited in bioethical inquiry for several reasons, including difficulties applying the stringent requirements for random selection, as well as difficulties with assignment, blinding, and achieving a sufficient sample size (Friedman et al., 1998). Therefore, instead of using the RCT, much of intervention research conducted in bioethics often follows the precepts of quasi-experimental designs, in which randomization is limited to random assignment to conditions and control groups are referred to as comparison groups (Rossi, Freeman, & Lipsey, 1999). Quasi-experimental designs range from the more rigorous two-group repeated measures or pre-/post-test designs to the one-group post-test-only design. The latter provides the least amount of control over extraneous factors. The nature of these designs allows for varying degrees of control for group differences in pre-existing characteristics (e.g., education, age, etc.) and events that may occur during the study and that may influence outcomes (Cozby, 2007). The degree of control is determined in part by the topic under study, and as noted above, there are topics in bioethics where not all conditions can be met for a purely experimental design. In quasi-experimental designs, the degree to which findings are generalizable (which depends on controls such as random selection and assignment), and the degree to which cause and effect conclusions (which depends on manipulation of the independent variable and use of control groups) can be drawn will be limited.

Essential Attributes of Experimental Designs

Experimental designs use a variety of procedures to distribute equally pre-existing differences among participants across conditions (i.e., experimental
and control), in order to maximize the similarity of circumstances under which participants receive an intervention and study groups are observed. These procedures can be categorized into three primary components: (1) randomization of control and experimental groups to the intervention (random assignment), (2) manipulation and systematization of the intervention, and (3) exposing the control group to contextual experiences as similar as possible as the experimental group during the study (Shadish, Cook, & Campbell, 2002).

Randomization to intervention and control groups is essential in order for the investigator to assure that any differences that individuals bring to the study, such as previous experiences and personal or socio-demographic characteristics that could interact with the intervention, will be equally distributed across groups and thus not influence the outcome (Shadish et al., 2002; Lipsey, 1990). Manipulation of the independent variable, constituting an intervention in the experimental group only, is an essential attribute of experimental studies (Lipsey, 1990) and in this respect differs from field studies in which naturally occurring phenomena that affect a group are observed as the study unfolds. For instance, in a hypothetical observational field study of informed consent in a research trial, the investigator would observe conversations between the researcher and study participants under different conditions to draw conclusions about how factors may influence participants’ understanding about risk level. In this study the investigator does not manipulate an intervention. In an experimental intervention study, the investigator would randomly assign researchers to different scripts, use vignettes with varied characteristics, or manipulate other variables in the experimental group in order to ascertain how different types of information delivery affect a participant’s understanding of risk.

Another important aspect of intervention studies is standardization of the intervention. This means that the investigator must develop and adhere to a protocol so that all participants in the experimental group are exposed to the intervention in the same way and for the same amount of time (Cozby, 2007). This will enable the investigator to interpret results with more confidence related to the effect of the intervention. Additionally, others can then replicate the study by following the same protocol.

Finally, the researcher must make efforts to ensure that external events or experiences that may have a bearing on the study outcome are not significantly different for participants in experimental and control conditions. The RCT is the most rigorous and best-controlled experimental design. Adhering to proscribed guidelines (CONSORT, 2004), this design always employs a control group, randomization, often random selection, and an
intervention that follows a strict, standardized protocol. The RCT is well established, highly reliable, and valid, allowing for the use of multivariate statistical procedures to make causal inferences about the effect of an intervention on an outcome. It requires as much systematic control over extraneous variables as is feasible. When this is not possible, the extent to which an effect can be attributed to the intervention is less certain.

**Strengths of the Experimental Design**

The overall aim of an experimental design is to control as many threats to internal and external validity as is possible, with RTC being the design that provides the most rigorous application of control. When a researcher examines the relationships between two or more variables he/she must be concerned about minimizing threats to internal validity and external validity. Internal validity is defined as the extent to which the effects detected are a true reflection of reality rather than being the result of extraneous variables (Burns & Grove, 1997, p. 230). That is, the researcher wants to be assured that the relationship between variables of interest is not influenced by unmeasured variables. External validity refers to the extent to which study findings are considered generalizable to other persons, settings or time (Shadish et al., 2002). The significance of a study is judged, at least in part, by whether the findings can be applied to individuals and groups separate from those in the sample studied (Burns & Grove, 1997).

**THREATS TO INTERNAL VALIDITY**

There are 12 commonly acknowledged threats to internal validity, or the ability to be confident that a proposed causal relationship reflects known, rather than unknown or unmeasured variables. These threats are: history; maturation; testing; instrumentation; statistical regression; selection; mortality; ambiguity about direction of a causal relationship; interactions with selection; diffusion or imitation of intervention; compensatory equalization of treatments; and compensatory rivalry by or resentful demoralization of respondents receiving less desirable treatments (Cook & Campbell, 1979; Shadish et al., 2002). Of these, history, selection, testing, instrumentation, and diffusion of intervention are especially relevant in bioethical intervention studies. Each of these threats will be illustrated by an intervention study whose purpose was to examine the effectiveness of education about research
integrity delivered to graduate students with the use of different teaching techniques. Outcome variables were knowledge and attitudes about ethical approaches to research and scientific misconduct. Content was delivered via standard in-person classes compared to self-paced, interactive modules via web-based platform and both were compared to a group of students not taking the course.

Table 1. Case Example: Illustration of Threats to Internal and External Validity.

| Purpose: The purpose of this study was to examine the effectiveness of education about research integrity delivered using different teaching techniques. The outcome variables included knowledge and attitudes about ethical approaches to research and scientific misconduct. The content was delivered via standard in-person classes compared to self-paced, interactive modules via web-based platform and both were compared to a group of students not taking the course. |
| Methods: Ninety-six Ph.D. students in psychology from two different universities were randomly selected from a group of 200 volunteers to participate in the study. These 96 were then randomly assigned to one of two interventions or one control group. The first intervention group consists of a series of 4 on-line modules to be completed over a four-week period, the second intervention consists of six in class two hour sessions and the control group did not receive any instruction. The Scientific Misconduct Questionnaire-Revised (Broome, Pryor, Habermann, Pulley, & Kincaid, 2005) was used to assess knowledge, attitudes, and experiences with scientific misconduct after completion of the intervention and one year later for all three groups. All participants were surveyed using the SMQ-R pre-intervention, eight weeks and one year after the start of the study. |

The threat of history refers to the possibility that participants in the experimental and the control group have different experiences while under observation in a study and that such experiences (extraneous variables) influence the outcomes differently in the groups. In the case study described in Table 1, the psychology students at one of the universities were all mandated to take a research ethics workshop provided by an official from the Office of Research Integrity at NIH. That is, the experience of taking a mandated research ethics course had the potential of affecting attitudes and knowledge about research ethics in that university but not in the other. This meant that the control group at one of the sites was no longer in the control condition after being exposed to the mandated class.
Selection

The threat of selection refers to bias that occurs when recruiting and assigning study subjects does not give every potential participant the same opportunity to become enrolled in the study (random selection) or the same opportunity to be randomly assigned to the treatment or control condition (random assignment). In the case example, the investigators advertised the study widely to all psychology graduate students on both campuses. As one would expect, only those who were interested in participating volunteered, and those who volunteer to participate in studies may represent a specific subset of the population. The investigators addressed this bias by randomly selecting 96 of the 200 students who volunteered. Random selection involves selecting study subjects by chance (e.g., using a table of random numbers) to represent the population from which they are chosen (Shadish et al., 2002) and gives each volunteer an equal chance of being chosen. The investigators then randomly assigned each individual to one of the three-groups: one which completed four on-line modules over a four week period; another which took part in six didactic two-hour sessions; and a control group that received no instruction. Thus, the threat of selection bias was minimized.

Testing

Testing is a threat that occurs when participants are asked to respond to the same measure on several occasions and, as a result, may remember some of the specific items. In the case study, after the first administration, some of the students may have become sensitized to the items on the instrument used to evaluate knowledge and attitudes (Broome et al., 2005) and remembered how they answered the questions when being administered the same instrument second time. Hence, any change in responses may not have been explained by being exposed to the intervention. Some investigators handle this problem by using alternative but parallel forms of a measure, so that the questionnaire tests the same concepts but uses different wording for the items. Others measure both control and intervention groups pre- and post with the same instrument so that patterns of change can be statistically tested to control for initial differences between groups. That is, investigators can test the assumption that due to random assignment, one would expect no differences in pretest scores and the potential for testing problems would be the same for both groups. Thus, differences in outcome would be attributed to the intervention.
Instrumentation

This threat is related to a change in the method of measurement from pre- to post intervention. An example from the case study is the use of a quantitative survey to measure the students’ knowledge about research integrity before the intervention and the use of open-ended interviews assessing this knowledge after the intervention. Any change in knowledge (either via a score on the survey or a coded analysis of open-ended responses) cannot, with a reasonable level of certainty, be attributed to the intervention, but could likely be affected by the manner in which knowledge was measured at the two points in time.

Another problem related to instrumentation can occur as a result of the response option format. For instance, when intervals on a scale are narrower on the ends than in the middle (e.g., extremely positive, very positive, positive, somewhat positive, somewhat negative, negative, very negative, extremely negative), responses on a second administration of an instrument may often tend to cluster around the middle rather than reflecting the full-range of options (Cook & Campbell, 1979). This may be due to individuals becoming frustrated at attempting to differentiate between the outer options of “very” and “extremely,” rendering the measure less a reflection of subjects’ actual responses and more a result produced by the format of the instrument itself.

Diffusion of the Intervention

When an intervention study is designed to assess the acquisition of knowledge or skills, and when individuals in the intervention and control groups can interact about an intervention (e.g., discuss it outside the study), the control group may gain access to information that may reduce differences between the two-groups on outcome measures. In the case example, this threat is especially salient as psychology graduate students often take the same classes or participate in similar activities within a university and may thus discuss with each other particulars of the educational intervention. A method to control for this is to randomize the intervention at the university level (across settings) rather than at the individual level (within a given setting).

THREATS TO EXTERNAL VALIDITY

Threats to external validity can limit the ability of the investigator to generalize the results of a study beyond the current sample, which, in turn,
limits the usefulness of the study's findings. The three major threats fall into the following categories: (1) interaction of selection and treatment; (2) interaction of setting and treatment; and (3) interaction of history and treatment. In the first situation, study participants possess a specific characteristic as a group that interacts with the intervention in such a way that change in the outcome is not generalizable to other groups of individuals. For example, a significant effect of an intervention to increase medical students' skills in clinical decision making in morally ambivalent situations may not be generalizable to a group of nursing assistants. In this case, the significant difference in educational level between groups is such that it interacts with the intervention to produce different outcomes. In the second and third situations (i.e., setting and history) contextual events (e.g., pay raises, new institutional leadership climate, additional bioethics workshops) that may occur during an intervention study are not replicable in a subsequent study or in other settings and will therefore restrict generalizability of findings.

In summary, it is important that an investigator select a research design to test an intervention that will control for as many threats to internal and external validity as possible. Some threats (e.g., selection, testing, and history) can be controlled for by using randomization. Other threats (diffusion of treatment and instrumentation) can be planned for by randomizing across sites rather than within one site and using the same instrument for all assessments. Choosing well-established measures that have been tested in other studies and which have demonstrated adequate reliability and validity are crucial to obtaining credible responses. Maintaining strict protocols regarding instrumentation and data collection help to ensure that data are collected under as similar conditions as possible and that any differences in responses between groups is due to the intervention. Plans that include adequate time and rigor in cleaning and managing data and applying statistical tests best fit for the type of data collected will decrease threats to both types of validity and enhance the reliability and credibility of the findings (Cook & Campbell, 1977).

ADVANTAGES OF INTERVENTION RESEARCH IN BIOETHICAL INQUIRY

Many questions asked by bioethicists can be answered by any one of the many non-experimental designs available to researchers. However, there are
several important questions that require the investigator to systematically and rigorously compare, in a controlled context, the effects of one or more treatments on selected outcomes. The experimental or quasi-experimental designs used to test the effectiveness of an intervention are most useful when assessing changes in knowledge, understanding, attitudes, or behaviors related to some aspect of ethical phenomena (Danis, Hanson, & Garrett, 2001). Examples of some questions that can only be answered by intervention designs are included in Table 2.

There are at least five distinct advantages to intervention research in bioethical inquiry: (1) the ability to examine whether a certain action (intervention) can be safely used with a selected group of individuals (efficacy) under relatively controlled conditions; (2) the ability to examine how useful an intervention is when used in the real world with a variety of people (effectiveness); (3) the ability to use multiple measurement methods (e.g., behavioral observation and questionnaires) to assess the impact of an intervention; (4) the possibility to maximize confidence in relationships found between variables (i.e., cause and effect); and (5) the ability to reach greater acceptance of findings by the larger scientific research community.

### CHALLENGES IN INTERVENTIONAL RESEARCH IN BIOETHICAL INQUIRY

Not all bioethical phenomena are appropriate for study using experimental or quasi-experimental designs. In general, these include naturally occurring phenomena that cannot be manipulated, the presence of conditions to which

---

**Table 2.** Selected Research Questions for Intervention Research.

1. Does the timing of information (72 h, 24 h, and immediately before an elective procedure) about a research study influence the understanding of the purpose of the study, the risks and benefits, and the refusal rate of participation?
2. Are nurses who are assigned to work with patient actors who are labeled as hospice patients more likely to discuss organ donation during a clinic visit than those assigned to patient actors who have cancer but are not enrolled in hospice?
3. Are tailored instructional materials for individuals at-risk for genetic conditions more effective than standardized materials in increasing knowledge and satisfaction?
4. Do school-age children who are shown a DVD depicting a child engaged in a study demonstrate greater comprehension and a more positive affect in regards to study participation?
one cannot randomly assign individuals, and circumstances that cannot be scripted or protocolized. For instance, one cannot manipulate or randomize individuals to different conditions at the end-of-life to study how they cope. Another challenge is the ethics of studying ethical issues. For example, in some situations the imposition of a research study, no matter how well intended, can burden participants during difficult or stressful times (Sachs et al., 2003). One example of this is to study how parents make decisions about whether or not to enroll their child in an end-of-life research study soon after the death of the child. Therefore, it is critical to carefully evaluate the study’s potential for benefit (in relation to risk or harm) and for expanding knowledge prior to approval or implementation.

Another challenge in intervention research is the ability to recruit an adequate number of participants so that statistically significant differences between groups can be detected. Given the nature of bioethical phenomena, it can be difficult to attract a large enough sample that meets requirements for rigorous statistical analyses. Multi-site studies that facilitate obtaining large samples can address this challenge.

ILLUSTRATIONS OF INTERVENTION RESEARCH – APPROACHES AND FINDINGS

In 1998, the National Institutes of Health funded 18 studies on the topic of informed consent. The purpose of this initiative was to produce (1) new and improved methods for the informed consent process; (2) methods that would address the challenges in obtaining consent from vulnerable populations; and (3) data to inform public policy (Sachs et al., 2003). The studies tested interventions designed to improve two different but related areas: (1) knowledge among potential research participants, and (2) decision-making abilities in vulnerable individuals (Agre et al., 2003). Six studies were RCTs and one used a quasi-experimental design where patients were not randomized to groups. The studies had in common the testing of a variety of media interventions such as videotapes, decision aids, and computer software to convey information to potential research subjects. A selection of the seven funded projects that were intervention studies will be reviewed below to illustrate the range of questions, designs, and analyses used in such research.

One of the studies was conducted with patients and families in a hospital waiting room who were going to make a decision about participating in a
clinical trial (Agre et al., 2003). Subjects were randomly assigned to one of the four modalities: standard verbal consent, a video, a computer program, or a booklet. A brief quiz was used to measure knowledge as the outcome. Findings revealed that the more complex the research protocol described to the potential participant, the higher the knowledge score. It also showed that the better-educated patients had higher scores, while more distressed individuals and those who were minorities scored lower. There were no primary effects for the four media suggesting none were superior to the others. However, the video was more effective for those deciding about complex protocols and for minority participants, while the booklet was more effective for those in poor health.

In a second study on informed consent also focusing on knowledge outcomes, Campbell, Goldman, Boccio, and Skinner (2004) conducted a simulated recruitment for two pediatric studies, one high risk (e.g., insertion of device in patients awaiting heart transplant) and one low risk (e.g., longitudinal assessment of low birth weight infants), with parents of children enrolled in Head Start. Four different interventions were tested: (1) a standard consent form; (2) a consent form with enlarged type and more white space; (3) a videotape; and (4) a PowerPoint presentation. None of the four methods of conveying information was superior. However, parents were significantly less likely to enroll their child in a high-risk protocol regardless of the nature of the method tested.

In one of the other studies (an Early Phase Research Trials – EPRT – with oncology patients), researchers first conducted a descriptive study in which interactions of patients and their physicians were audiotaped when the study was described and participation was offered (Agre et al., 2003). Based on this, an intervention was designed to increase the patients’ understanding of EPRT consisting of a 20 minute, self-paced, touch screen computer-based educational program. Patients were randomly assigned to the intervention or control group, with the latter receiving a pamphlet that explained the EPRT. Results showed that the intervention had minimal impact on agreement to participate, with equal numbers in both groups deciding to join the trial, although patients in the intervention group were more likely to say the intervention changed the way they made their decision.

Mintz, Wirshing and colleagues (Agre et al., 2003) developed two videotapes preparing potential participants to consider enrollment in a medical study or in a psychiatric study. In addition to content and information, the intervention videotape encouraged individuals to be active participants during the informed consent process. The control video presented historical information and federal regulations about informed
consent. Participants for both studies were divided into intervention and control conditions. Knowledge about consent processes among participants in both the medical and psychiatric studies improved as a result of the intervention video compared to those viewing the control tape.

In another study, Merz and Sankar recruited participants from a General Clinical Research Center (GCRC) to test the effectiveness of a standard consent form compared to a series of vignettes (Agre et al., 2003) on participants’ knowledge. At post-test, participants in both groups demonstrated relatively good comprehension and no significant difference emerged between control and intervention groups on knowledge.

Overall, these studies show that the medium used to deliver the message did not consistently make a difference in the outcome variables of interest. What several of the studies did show were some important subgroup differences as results of the interventions. The strength of these efforts is that for the first time several studies with similar designs and interventions (albeit different populations) could be compared and some preliminary conclusions made that validate previous thinking about informed consent. These include assumptions that younger age, higher education, higher literacy, and stronger medical knowledge influence outcomes in positive directions. This suggests that investigators must give more attention to how consents are presented, the characteristics of subjects, and how comprehension is evaluated (Agre et al., 2003).

Limitations of these informed consent studies include a lack of diversity in the samples, use of patient surrogates, participants who were well educated, and the use of complex designs with multiple variables. The majority of the subjects were white and well educated, reflecting the ethnic and socio-demographic settings in which most clinical trials are undertaken and, thus, restricting the generalizability of findings. The use of patient vignettes as a method to inform potential research subjects, while not unusual given the sensitive nature of many of the clinical trials (e.g., blood donation for DNA banking), also limits the generalizability of results to researchers, patients, and families (for a detailed discussion of the use of vignettes, see chapter on hypothetical vignettes). The testing of multiple variables in some of the studies reviewed, although realistic, also presents a challenge related to how external variables or the particular medium used may have influenced outcomes. Yet, findings from these studies provide preliminary evidence on which to build future research toward expanding our knowledge and offering ways to maximize the protection of human subjects through optimizing their comprehension and informed decision making related to consent to participate in research investigations.
FUTURE RESEARCH

The use of intervention designs, while a relatively recent phenomenon in bioethical inquiry, has a distinct and important role to play in advancing the field of bioethics. It is especially important to examine ethical practices that have widespread implications for patients, families, and health care professionals. This is particularly the case with practices that have been proposed by policy makers, such as advanced directives, organ donation, and do not resuscitate orders. Although not all, or even most, bioethical questions are appropriate to study using empirically driven intervention designs, some are and, in fact, some questions must be addressed using intervention models. Without a systematic, controlled approach to examining the effectiveness of interventions designed to change and test various outcomes, we will never know which actions work and which do not. Clinicians depend on rigorously designed intervention studies that provide evidence for the establishment of guidelines for conflict resolution and decision making in the delivery of quality health care. Furthermore, findings generated by intervention studies in bioethics makes available data for policy makers to formulate policies, fund programs, and enact legislation that may assist clinicians and ethicists to resolve value conflicts and other ethical problems. Ultimately, this will improve the lives of patients and families who experience suffering, not only from their illnesses, but also from vexing questions related to these illnesses.

REFERENCES


This page intentionally left blank
SUBJECT INDEX

Advance directives 24, 31
applied clinical ethics 14, 15, 16, 19
ATLAS.ti 127
audio-recordings 45, 47, 125
automated data retrieval 133
Bias 54, 75, 154, 186, 191
Categories 47, 48
clarity of wording 168
clinical ethics 14
clinical ethics consultation 32
closed-ended survey 41
clustering effects 196
codebook 49, 51
code development 59
code reports 55, 56, 57
codes 48, 53, 127
coding agreement 54
coding drift 131
coding manual 128
coding reliability 51, 54
coding scheme 49
cognitive interviewing 166
community-based research 187
community decision-making 184
community involvement 68, 73
computer assisted telephone interviewing 154
conceptual map 53
confidentiality 42, 68, 156
conflict of interest 29
consensus coding 131
consensus process 55
CONSORT 206
constant-variable vignette method 163
construct validity 122, 152
content validity 51, 152, 170
control groups 205, 206
convenience samples 146
cost of surveys 150
Data immersion 47
data preparation chart 126
debriefing notes 73
decision-making capacity 27, 31
decisional capacity 31
deductive codes 48, 49
descriptive studies 205
diffusion of intervention 207
digitally recording interviews 125
distributive justice 193
doctrine of informed consent 22, 27
Ethical theories 13, 14, 15, 16, 19, 23
ethics consultants 13, 14, 16, 17, 18, 19
ethics consultations 16, 17, 18
experimental designs 205
exploratory research 118
external validity 176, 203, 207
extraneous variables 207
Face validity 152
factorial design 174
financial incentives 150, 151
focus group questions 67
focused theory-testing 118
follow-up surveys 191
forced-choice responses 149
forgoing treatment 24
formative data 65
factorial designs 175

Generalizability 79, 211
goals of care 19
grounded theory 41

Health care disparities 25
health care priorities 184, 193
HIPAA Privacy Rule 191
human subjects research 186
hypothesis-testing 175

Inductive codes 48
inductive reasoning 50
informed consent 31, 68, 125, 198, 213, 214, 215
Institute of Medicine 25, 30
instrumentation 207
inter-coder reliability 129
internal consistency 169, 171
intervention studies 204
interview guide 46, 119
intra-rater reliability 171
IRBs 28, 29, 30

Justice 15, 193

Karen Ann Quinlan 24
key informant interviews 151

Labeling 44
legitimacy 184, 186, 188
life-sustaining interventions 24
Likert scale, 173
literacy levels 145

Memoing 47
memos 56, 60

multi-level consensus coding (MLCC)
129

Nancy Cruzan 24
National Bioethics Advisory Commission 30
node type 51
nominal group technique 188
non-probability sampling 145
non-random sampling 197
non-response bias 154

Online coding 132
open-ended interviews 41
open-ended queries 118, 168
ordering of survey items 150

Pilot testing 120, 122
positivist view 54
pragmatic ethics 15
pre-/post-test designs 205
preliminary coding 49
pretest 167
privacy of information 69
procedural justice 193
protecting human subjects 28, 30
provisional coding manual 128
proxy directives, 25, 31
public opinion 198
public opinion polls 185
purposeful sampling 43, 146

Qualitative content analysis 40
qualitative data analysis 132
quantitative content analysis 39
quasi-experimental designs 205, 212, 213
question order 156
quota samples 146

Randomization 186, 190, 205, 206, 213
rates of non-response 147
recall bias 154
Subject Index

recruitment of subjects 123, 186
reliability 152, 169, 171
representative 185, 186, 194
research questions 18, 43, 57, 151
respondent burden 154
response frames 149
response rates 151
response summaries 75
response-wave bias 155
right to die 24
round robin method 188

Sampling 43, 66, 123, 124, 145, 146, 150, 173, 186, 190, 194, 199
selection bias 207
self-administered surveys 149
semi-structured interviews 40, 117
simple random sample 145
simulation exercise 193, 194
snowball sampling 123, 146
social justice 25, 31
socially desirable response bias, 154, 156
stratified random sampling 67, 145
sub-codes 128

substantive representation 186
SUPPORT study 24
survey development 145
survey fielding 153
survey question wording 186
survey responses 189

Team coding 51
test-retest reliability 171
textual data 39
thematic coding 128
themes 47, 48, 128
theoretical saturation 124
therapeutic misconception 22
town hall meetings 185
transcription 45, 47, 120, 126

Units of analysis 43, 196
univariate (single item) distributions 156

Validity 54, 122, 152, 169
value conflicts 14
vulnerable populations 28, 30
This page intentionally left blank