



HEALTH

- THE ARTS
- CHILD POLICY
- CIVIL JUSTICE
- EDUCATION
- ENERGY AND ENVIRONMENT
- HEALTH AND HEALTH CARE
- INTERNATIONAL AFFAIRS
- NATIONAL SECURITY
- POPULATION AND AGING
- PUBLIC SAFETY
- SCIENCE AND TECHNOLOGY
- SUBSTANCE ABUSE
- TERRORISM AND HOMELAND SECURITY
- TRANSPORTATION AND INFRASTRUCTURE
- WORKFORCE AND WORKPLACE

This PDF document was made available from www.rand.org as a public service of the RAND Corporation.

[Jump down to document](#) ▼

The RAND Corporation is a nonprofit research organization providing objective analysis and effective solutions that address the challenges facing the public and private sectors around the world.

Support RAND

[Browse Books & Publications](#)

[Make a charitable contribution](#)

For More Information

Visit RAND at www.rand.org

Explore [RAND Health](#)

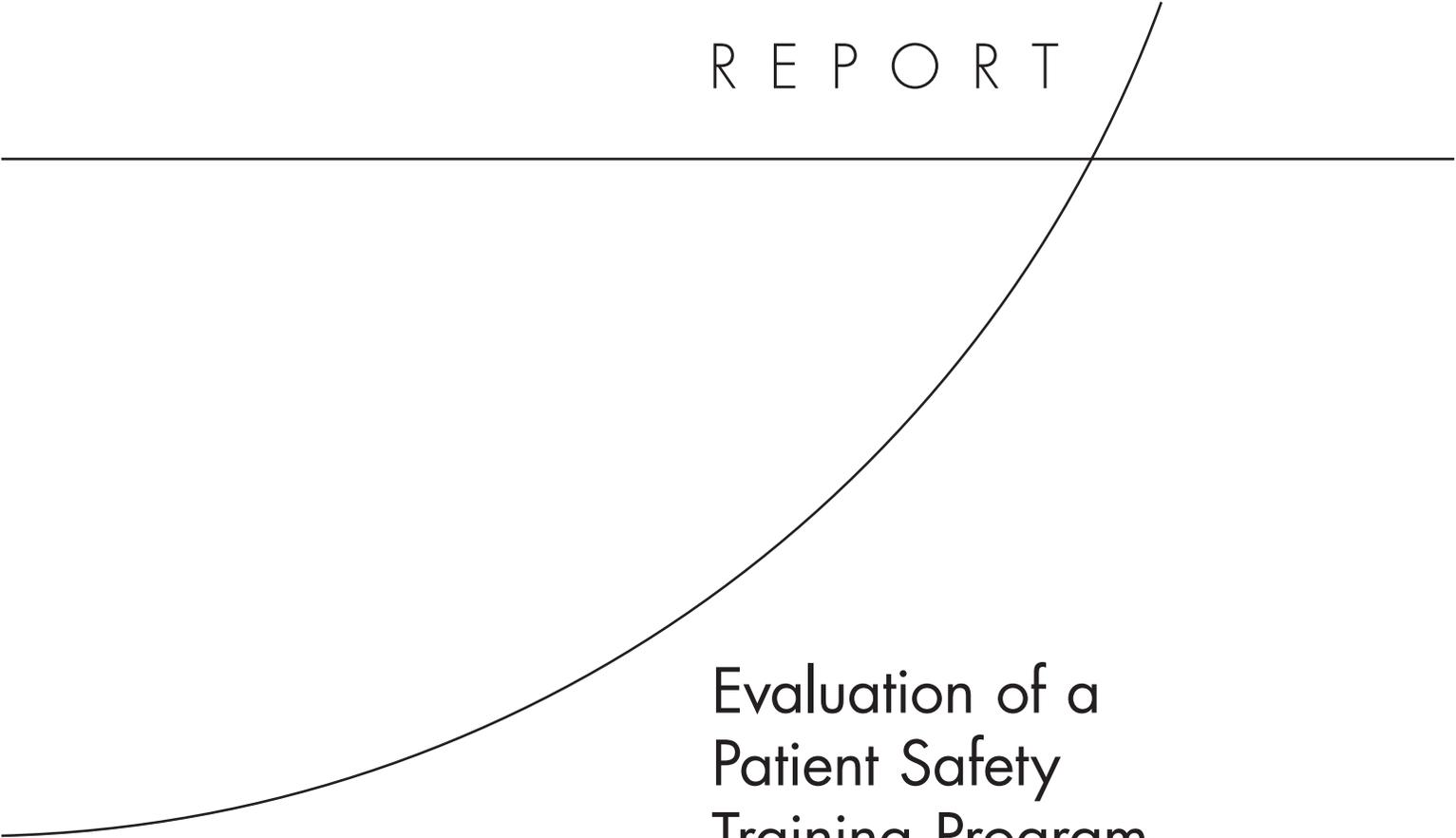
View [document details](#)

Limited Electronic Distribution Rights

This document and trademark(s) contained herein are protected by law as indicated in a notice appearing later in this work. This electronic representation of RAND intellectual property is provided for non-commercial use only. Permission is required from RAND to reproduce, or reuse in another form, any of our research documents for commercial use.

This product is part of the RAND Corporation technical report series. Reports may include research findings on a specific topic that is limited in scope; present discussions of the methodology employed in research; provide literature reviews, survey instruments, modeling exercises, guidelines for practitioners and research professionals, and supporting documentation; or deliver preliminary findings. All RAND reports undergo rigorous peer review to ensure that they meet high standards for research quality and objectivity.

R E P O R T



Evaluation of a Patient Safety Training Program

Christopher Nelson

Prepared for the Jewish Healthcare Foundation

The research described in the report was prepared for the Jewish Healthcare Foundation by RAND Health, a unit of the RAND Corporation.

The RAND Corporation is a nonprofit research organization providing objective analysis and effective solutions that address the challenges facing the public and private sectors around the world. RAND's publications do not necessarily reflect the opinions of its research clients and sponsors.

RAND® is a registered trademark.

A profile of RAND Health, abstracts of its publications, and ordering information can be found on the RAND Health home page at www.rand.org/health.

© Copyright 2005 RAND Corporation

All rights reserved. No part of this book may be reproduced in any form by any electronic or mechanical means (including photocopying, recording, or information storage and retrieval) without permission in writing from RAND.

Published 2005 by the RAND Corporation
1776 Main Street, P.O. Box 2138, Santa Monica, CA 90407-2138
1200 South Hayes Street, Arlington, VA 22202-5050
201 North Craig Street, Suite 202, Pittsburgh, PA 15213-1516
RAND URL: <http://www.rand.org/>
To order RAND documents or to obtain additional information, contact
Distribution Services: Telephone: (310) 451-7002;
Fax: (310) 451-6915; Email: order@rand.org

Preface

Medical errors account for somewhere between 44,000 and 98,000 preventable deaths annually in the United States, with some estimates as high as 195,000 preventable deaths a year. One widely discussed approach to addressing the problem of medical errors is improved professional training on safety science concepts for healthcare professionals. The Jewish Healthcare Foundation (JHF)—a Pittsburgh-based philanthropy dedicated to furthering the provision of high-quality healthcare—recently developed such a training curriculum. An initial pilot version of the curriculum was offered during the summer of 2004 under the auspices of the JHF/Coro Health Sciences Fellowship, a four-year-old program offered in partnership with the Coro Center for Civic Leadership.

In order to gain preliminary feedback on the effectiveness of this training curriculum, JHF contracted with RAND Health, a unit of the RAND Corporation to provide an in-process evaluation of the summer 2004 pilot, which, unlike earlier iterations of the Fellowship, included a primary focus on medical error and patient safety. The purpose of the evaluation was to assess the prospective merits of the new version of the Fellowship program as a mechanism for training students in the healthcare professions in the principles and practices of safety science.

The report should be of interest to policymakers, funders, medical professionals, medical educators, and others involved in efforts to reduce the prevalence of medical errors.

Table of Contents

Preface	iii
Table of Contents	v
List of Tables	viii
List of Figures	x
Summary	xii
Evaluation Questions and Methods	xii
Fellowship Design and Participant Characteristics	xiii
Implementation	xiv
Training Outputs.....	xiv
Recommendations	xv
Conclusions	xvii
1. Introduction	1
2. Course Design	3
2.1 Program Logic Model Used To Identify Intermediate Outputs.....	3
2.2 Core Content Included Safety Science Concepts and Teamwork Skills	5
2.3 Pedagogical Approach Emphasized Exposure to Local Clinical Settings	9
2.4 Summary and Conclusions	10
3. Evaluation Methods	11
3.1 Data Collection Sought Information on Outcomes and Mechanisms	11
3.2 Limitations	12
3.3 Summary	13
4. Participants	14
4.1 Most Fellows Were at an Early Stage in Their Professional Training	14
4.2 Fellows Came From Wide Range of Backgrounds	15
4.3 Differences in Perceived Motivation Across Fellowship Tracks	16
4.4 Most Participants Learned of Program Through Teachers	16
4.5 Summary and Conclusions	16
5. Implementation	17
5.1 Fellows' Perspectives on Course Design	17
5.2 Materials and Activities	19
5.3 Instructors and Course Management	21
5.4 Summary and Conclusions	23

6. Training Outputs	24
6.1 Discernible Increases in Awareness of Patient Safety Problems	24
6.2 Attitudes Changed in the Expected Direction	25
6.3 Discernible Increase in Knowledge and Skills, But Less Confidence in Ability to Apply Skills...27	
6.4 Fellows Generally Report That They're Likely to Act on Their New Skills	30
6.5 Generally Positive Overall Evaluations of the Fellowship	31
6.6 Summary and Conclusions	32
7. Conclusions and Recommendations	33
7.1 Key Findings	33
7.2 Recommendations	35
7.3 Conclusion	37
Appendix A: Survey Instrument	38
Appendix B: Focus Group Protocol	42
References	44

List of Tables

Table 2.1 Fellowship Topics	8
Table 4.1 Fellows' Professional Backgrounds	15
Table 4.2 Fellows' Intended Professions	15
Table 5.1 Fellows' Assessments of Readings (Means and Standard Deviations)	19
Table 5.2 Fellows' Assessments of Opportunities for Hands-On Experience (Means and Standard Deviations)....	20
Table 5.3 Fellows' Assessments of Instructors (Means and Standard Deviations)	22
Table 5.4 Fellows' Assessments of Session Management (Means and Standard Deviations)	23
Table 6.1 Changes in Fellows' Awareness of Patient Safety Issues (Means and Standard Deviations)	25
Table 6.2 Changes in Fellows' Attitudes About Patient Safety (Means and Standard Deviations)	26
Table 6.3 Self-Reported Changes in Fellows' Knowledge and Skills (Means and Standard Deviations)	28
Table 6.4 Self-Reported Ability to Apply Skills in Real-World Contexts	29
Table 6.5 Self-Reported Propensity to Act on Fellowship Skills	31
Table 6.6 Fellows' Overall Evaluation of the Fellowship	32

List of Figures

Figure 2.1 Program Logic Model	4
Figure 2.2 Content of the Index Card on “The Four Rules in Use”.....	7
Figure 2.3 The Ladder of Inference	9
Figure 3.1 Power Calculations	12

Summary

Medical errors account for somewhere between 44,000 and 98,000 preventable deaths annually in the United States (Institute of Medicine [IOM], 2000), with some estimates as high as 195,000 preventable deaths a year (Health Grades, 2004). Thus, more people die each year from medical errors in the United States than from motor vehicle accidents (IOM, 2000). Along with the obvious human costs, medical errors generate the need for otherwise unnecessary hospital admissions, lengthened hospital stays, and additional treatments to correct the initial error (IOM, 2001). Total national costs associated with medical errors are estimated at between \$17 billion and \$29 billion annually (IOM, 2000), including lost income, lost household production, disability and healthcare costs.¹ Other less tangible costs include a loss of patient trust and diminished job satisfaction among healthcare providers.

One widely discussed approach to addressing the problem of medical error is improved professional training on safety science concepts for healthcare professionals (IOM, 2000: 12). The Jewish Healthcare Foundation (JHF) a Pittsburgh-based philanthropy dedicated to furthering the provision of high-quality healthcare—recently developed such a training curriculum. An initial pilot version of the curriculum was offered during the summer of 2004 under the auspices of the JHF/Coro Health Sciences Fellowship, a four-year-old program offered in partnership with the Coro Center for Civic Leadership.

In order to gain preliminary feedback on the effectiveness of this training curriculum, JHF contracted with the RAND Corporation to provide an in-process evaluation of the summer 2004 pilot, which, unlike earlier iterations of the Fellowship, included a primary focus on medical error and patient safety.² The purpose of the evaluation was to assess the prospective merit of the new version of the Fellowship as a mechanism for training students in the healthcare professions in the principles and practices of safety science.

Evaluation Questions and Methods

The Fellowship's ultimate goal is to engender in participants the capacity to make system-improving changes in the healthcare settings in which they eventually work. Consequently, its impacts will become evident, if they become evident at all, long after the completion of this evaluation. Thus, the evaluation is not able to provide a rigorous estimate of the Fellowship's efficacy in achieving its *ultimate* goal. Instead, the report seeks to provide formative feedback on *intermediate* goals that might inform improvement, redesign, and decisions about scale-up of the summer 2004 Fellowship.

The evaluation questions that guide this report reflect the need for in-process feedback and fall into three categories:

- *Design and context.* What were the key elements of the Fellowship's design? What were the key characteristics of the program's Fellows?
- *Implementation.* How effective was the delivery of the curriculum during the summer 2004 pilot?
- *Training outcomes.* Is there evidence of program impacts on the Fellows' (a) knowledge of safety science concepts and skills, (b) willingness and ability to apply safety science concepts, and (c) commitment to error reduction and patient safety?

Data collection methods included document review, a participant survey, participant focus groups, observations, and an examination of work samples.

¹ Studies suggest that medication errors alone increase hospital stays by an average of 1.9 to 4.6 days, at an annual cost of \$2,262 to \$4,700 per admission (Bates, Spell, Cullen, et al., 1997; Classen, Pestotnik, Evans, et al., 1997).

² Hereafter, we use the term "Fellowship" to refer to the summer 2004 pilot version.

Fellowship Design and Participant Characteristics

Safety science concepts included in the Fellowship are derived primarily from the Toyota Production System (TPS) and a locally developed curriculum developed by a nonprofit alliance, the Pittsburgh Regional Healthcare Initiative (PRHI). The curriculum—Perfecting Patient Care—involves both *core principles* and a set of *methods of investigation*. The core principles include the following:

- Most medical errors involve systemic root causes.
- Process improvement is usually best undertaken by those working on the “front lines” of care and in real time.
- For front line workers to successfully address systemic root causes of medical errors, the support and leadership of management is required.

The method of investigation, in turn, includes:

- Identification of a patient need
- Careful observation of current work practices
- Identification of a problem with current work processes
- Investigation of the root causes of such problems
- Devising and testing a change in work processes
- Measuring results of the change.

The first three activities often involve selection of a “model line”—a productive process in the hospital or other clinical setting that might serve as an example for the rest of the organization—and a careful assessment of the “current condition” of that process. Root cause analysis, in turn, is described in the Fellowship as the process of asking “The 5 Whys.” Here, the analyst begins with an unsatisfactory condition (e.g., a high rate of central line infections) and then asks why the condition exists. Having identified a proximate cause through an answer to the first “why” question (e.g., use of equipment that has not been properly sterilized), the analyst continues to ask why each of the causes themselves exists. Having done this, the next step is to devise and implement a fix to the problem. TPS provides “four rules in use” that provide guidance for system redesign. The rules urge users to ensure that steps in the care chain (1) are related to outcomes, (2) are clear, (3) are highly specified, and (4) provide ongoing signals about system performance.

Along with the core safety science curriculum content, Fellows also received training in critical thinking and teamwork skills. The emphasis on critical thinking can be viewed as a complement to TPS’s emphasis on careful observation and the use of the scientific method to evaluate the impact of changes in healthcare processes.

Training on leadership and teamwork skills focused largely on a tool called OARRS, which stands for Outcomes, Agendas, Roles, and Rules. Fellows used the OARRS tool to structure and execute group activities. *Outcomes* refer to the intended goals of a group activity, while *agendas* refer to the sequence of events by which the activity is to proceed. Similarly, *roles* refer to the functions that each team member plays, while *rules* refer to the guidelines that govern interactions among team members. For, instance, participants used the tool to plan for sessions in which they interviewed guest speakers from local healthcare facilities. The Fellowship also sought to develop leadership and teamwork skills by including representatives from a wide range of healthcare disciplines.

Instruction was almost entirely built around a series of site visits to local clinical settings engaged in effective patient safety practices. This served to embed the content described above into real-world clinical contexts.

The summer 2004 Fellowship was offered in two tracks, “Getting to Zero” (GTZ) and “Getting to Excellence” (GTE), in an effort to experiment with minor variations in instructional techniques and content foci. While the focus of both tracks was on patient safety, the GTE track also included units on palliative care and public health. Similarly, while both tracks involved site visits, Fellows in the GTZ track also participated in hands-on exercises that typically involved interviewing practitioners, diagramming processes, and suggesting process and system redesigns.

Implementation

Having explicated the design of the Fellowship, the evaluation by RAND sought to assess the implementation of that design during the summer 2004 pilot. For the most part, respondents’ reactions to the execution of the course were positive. Most found that the material presented was new to them. However, there were mixed opinions about whether the Fellowship attempted to cover too much material. Many respondents noted that topics often were rushed. But few, if any, could identify material they thought might be dropped in the interest of time.

Some respondents also raised concerns about the overall coherence of the Fellowship, suggesting that sessions should be more focused around overarching themes and concepts. In a similar vein, many respondents noted that it was not always clear how course readings related to broader instructional goals. Respondents’ opinions were also mixed on the Coro skills (e.g., leadership, critical thinking, and group operations). While some respondents found presentation of the Coro skills to be too simplistic, others found them to be new and important, noting that they seldom receive such training in the course of their other studies.

Opportunities for hands-on experience and the use of concrete examples during the Fellowship were given high ratings by respondents. However, there was some concern that participants were not given enough time to prepare for presentations, and that project groups were not given enough time to complete their tasks and develop strong working relationships. Some respondents also pointed out that presentations by Fellows often came at the end of the session, with little or no opportunity for discussion. This, according to the respondents, diminished the presentations’ utility as learning tools.

Instructors and guest speakers generally received high marks. However, respondents generally wanted more opportunities to interact with guest speakers and with each other.

Training Outputs

As noted earlier, a key challenge lies in the fact that most of the Fellowship’s ultimate goals will happen—if they happen at all—long after the evaluation is over. The logic model discussed above identified the following goals: (1) enhanced *awareness* of patient safety issues, (2) increased prevalence of *attitudes* about safety that are congruent with current “systems” thinking on safety, and (3) the development of *knowledge and skills* that can be used to diagnose and respond to systemic causes of medical error as intermediate outputs that might be used to evaluate the Fellowship’s efficacy in the short term.

Data from the survey and focus groups suggest that the Fellowship largely succeeded in achieving each of these goals. First, respondents reported discernible increases in the extent to which they perceived patient safety as a significant problem. For instance, respondents were asked to rate their level of agreement with the statement “Medical error is a significant problem” on a five-point scale, with 1 representing “Strongly disagree” and 5 representing “Strongly agree.” Respondents were asked to rate their agreement with the statement both before the Fellowship and on the day they took the survey,³ with self-reported awareness levels increasing during the course of the Fellowship. Indeed, the average participant gained 1.1 points

³ Respondents took the survey near the end of the Fellowship. The logic of this “retrospective pre/post” design is discussed in Chapter 3.

on the five-point scale, a statistically discernible increase in their agreement with the statements presented to them.⁴

Second, respondents reported becoming more likely to hold attitudes about error causation that are congruent with current thinking in safety science. For instance, one item asked respondents to indicate the extent to which they agreed with the statement “Medical errors are primarily caused by individual carelessness” on a five-point scale, where 1 represented “Strongly disagree” and 5 represented “Strongly agree.” Given the Fellowship’s emphasis on the systemic causes of medical errors and the importance of blame-free organizational cultures, Fellowship developers clearly hoped that participation in the Fellowship would reduce the level of agreement with this statement. Indeed, the average respondent rating declined from 3.3 to 2.2, a statistically discernible decline of 1.1 points. Similarly, there were discernible increases in respondents’ self-reported knowledge of core safety science concepts and techniques. However, respondents were less certain of their readiness to apply these skills in real-world clinical settings.

Respondents also reported a willingness to act on their newly developed knowledge and skills. This included reading materials related to patient safety, and speaking to colleagues and supervisors about safety problems observed in the workplace.

Recommendations

While limitations in the design (e.g., the inability to observe Fellows applying their skills in actual clinical settings) prevented us from assessing the Fellowship’s success in meeting its ultimate goals, the evidence considered here suggests that, at the very least, the Fellowship succeeded in reaching its more immediate training goals. Nevertheless, the report identified a number of remaining issues that Fellowship developers would do well to address as they continue to refine the design. These include:

- The tradeoff between breadth and depth in the range of material covered
- The extent to which the Fellowship is well-organized around a set of clear and coherent instructional themes and goals
- The extent to which course readings and hands-on exercises are integrated into a larger instructional vision
- The appropriateness and use of training in critical thinking
- Concern about the ability to apply skills in practice
- Time for hands-on activities.

To address these and other issues, we offer a number of recommendations, some involving suggestions for further design and planning, others for increasing the program’s evaluability.

Recommendation 1: Use a logic modeling exercise to strengthen coherence of design. Fellows’ concerns about the program’s coherence suggest the need to reconsider the match between the Fellowship’s goals and elements of its design. Accordingly, JHF might consider a logic modeling exercise (one more elaborate and participatory than the logic model presented in this report) to ensure that the Fellowship’s specific elements are all well aligned with instructional goals. This exercise might include the following steps:

- Developing a list of typical use contexts and scenarios in which Fellows might eventually work
- Identifying capabilities that cut across the use contexts and scenarios

⁴ See Chapter 3 for a discussion of “statistical discernibility,” as the term is used in this report.

- Mapping training activities against Fellows’ capabilities
- Evaluating the whole design for completeness and coherence.

Recommendation 2: Reconsider the use of readings and exercises. The Fellowship has a strong orientation toward case-based learning—especially in the GTZ track. This orientation is appropriate given the complexity and variability in the skills it seeks to teach. Given the concerns raised in this report about tradeoffs between depth and breadth of coverage, it makes sense to review each of the exercises to ensure that they are sufficiently imbued with core concepts and skills, and that they are rich enough to provide Fellows with the experience of tackling complex problems without overwhelming them. Indeed, respondents to survey and focus-group questions often noted that they had too little time to learn from the exercises. Thus, Fellowship developers might look for opportunities to reduce the number of exercises, allowing more time for exercises that Fellows found most valuable.

Fellowship developers might also consider developing rich documentation for cases and exercises that might provide “scaffolding” for future instructors to support effective presentation of the cases. This will be particularly important should JHF consider scale-up of the Fellowship to other sites and locations.

It is important to emphasize that the findings presented here should be regarded as suggestive, but not conclusive. This caution stems from the fact that (1) we were unable to observe alumni as they seek to function in real-world healthcare settings; (2) practical considerations forced us to rely mainly on self-reported accounts of changes in awareness, attitudes, and skills related to patient safety; and (3) the fact that the evaluation design could not employ a no-treatment comparison group. Thus, the remaining recommendations we present provide suggestions to ensure that future implementations of the Fellowship are capable of yielding stronger evaluative inferences.

Recommendation 3: Track alumni over time and do follow-up surveys. First, JHF should consider developing a formal and rigorous system for tracking and surveying Fellowship alumni over time. Such a survey might include questions on the following:

- Career paths, including information about the institutional environments in which alumni work
- The extent to which alumni have been involved in activities to improve patient safety
- The extent to which the skills and concepts taught in the Fellowship have been adequate given the institutional contexts in which they work
- Which concepts and skills alumni have retained over time.

Tracking alumni can be challenging, especially in healthcare where significant numbers might move to other communities. For this reason, it would be useful for the tracking system to include names of family members and professional colleagues who might be able to assist JHF in locating alumni who have moved.

Recommendation 4: Develop more structured ways to evaluate hands-on exercises. The reliance on self-reported outcomes in this evaluation could be addressed in future evaluations if JHF were to develop more rigorous ways of evaluating participant projects and other work. Participants worked on a number of hands-on projects, culminating in a group project presented to other participants and members of the local healthcare community. These exercises would be more useful for evaluation purposes if they had been graded against a well-specified rating rubric. Ideally, such a process would be applied both to a work completed at the beginning *and* the end of the Fellowship. This would provide a clear baseline against which to assess growth over time.

Recommendation 5: Consider more careful structuring of differences between tracks. Finally, the existence of two Fellowship tracks provides an opportunity to determine how variations in content and

pedagogy affect outcomes. Such knowledge, in turn, can be useful in guiding improvements to the Fellowship. However, the fact that the GTE and GTZ tracks varied both in content *and* pedagogy made it difficult to assess the independent impact of either on implementation and outcomes. In the future, it would be desirable to plan such variations in a way that allows for clearer inferences.

Conclusions

The JHF-Coro Fellowship represents an early attempt to develop and implement a training curriculum in medical safety science for healthcare professionals in training. The results reported here suggest that the approach used by the Fellowship holds promise and is worthy of further effort. However, conclusions about its ultimate effect on patient outcomes await more sustained implementation and further evaluation research.

1. Introduction

Medical errors account for somewhere between 44,000 and 98,000 preventable deaths annually in the United States (Institute of Medicine [IOM], 2000). More recent estimates put the number as high as 195,000 preventable deaths a year (Health Grades, 2004). Thus, more people die each year from medical errors in the United States than from motor vehicle accidents (IOM, 2000). Along with the obvious human costs, medical errors generate the need for otherwise unnecessary hospital admissions, lengthened hospital stays, and additional treatments to correct the initial error (IOM, 2001). Total national costs associated with medical errors are estimated at between \$17 billion and \$29 billion annually (IOM, 2000), including lost income, lost household production, disability and healthcare costs.⁵ Other less tangible costs include a loss of patient trust and diminished job satisfaction and low morale among healthcare providers.

One widely discussed approach to addressing the problem of medical error is improved professional training on safety science concepts for healthcare professionals (IOM, 2000). The Jewish Healthcare Foundation (JHF)—a Pittsburgh-based philanthropy dedicated to furthering the provision of high-quality healthcare—recently developed such a training curriculum. An initial pilot version of the curriculum was offered during the summer of 2004 under the auspices of the JHF/Coro Health Sciences Fellowship, a four-year-old program offered in partnership with the Coro Center for Civic Leadership.

In order to gain preliminary feedback on the effectiveness of this training curriculum, JHF contracted with the RAND Corporation to provide an in-process evaluation of the summer 2004 pilot, which, unlike earlier iterations of the Fellowship, included a primary focus on medical error and patient safety.⁶ The purpose of the evaluation was to assess the prospective merit of the new version of the Fellowship as a mechanism for training healthcare professionals in the principles and practices of safety science.

The Fellowship's ultimate goal is to engender in participants the capacity to make system-improving changes in the healthcare settings in which they will eventually work. Consequently, its impact will be evident, if it is evident at all, long after the completion of this evaluation. Thus, the evaluation is not able to provide a rigorous estimate of the Fellowship's efficacy in achieving its *ultimate* goal. Instead, the report seeks to provide formative feedback on *intermediate* goals that might inform improvement, redesign, and decisions about scale-up of the summer 2004 Fellowship. Accordingly, the evaluation questions that underlie this report touch on not only the Fellowship's impact on participants' skills and knowledge but also the design and execution of the Fellowship program.⁷

- *Design and context.* What were the key elements of the Fellowship's design? What were the key characteristics of the Fellows?
- *Implementation.* How effective was the delivery of the curriculum during the summer 2004 pilot?
- *Training outcomes.* Is there evidence of program impacts on the Fellows' (1) knowledge of safety science concepts and skills, (2) willingness and ability to apply safety science concepts, and (3) commitment to error reduction and patient safety?

The remainder of this report is organized as follows. Chapter 2 provides a description of the key elements of the Fellowship and provides a logic model linking this design to key outputs and outcomes. Chapter 3 provides a brief overview of the data collection and evaluation methods employed in this study. Chapter

⁵ Studies suggest that medication errors alone increase hospital stays by an average of 1.9 to 4.6 days, at an annual cost of \$2,262 to \$4,700 per admission (Bates, Spell, Cullen, et al., 1997; Classen, Pestotnik, Evans, et al., 1997)

⁶ Hereafter, we use the term "Fellowship" to refer to the summer 2004 pilot.

⁷ Selection and conceptualization of the evaluation questions was guided by the CIPP model developed by Daniel Stufflebeam. CIPP stands for the four elements that the model says every evaluation should possess: context, inputs, process, and product (see, e.g., Stufflebeam, 2000, 2004).

4 describes the key characteristics of Fellowship participants, with particular attention paid to differences between the two tracks. Chapter 5 assesses how the curriculum was actually delivered during the summer 2004 pilot. Chapter 6 seeks to assess the Fellowship's success in improving participants' safety science skills and knowledge, as well as their awareness of, and motivation to become involved in, patient safety issues. Finally, Chapter 7 summarizes key findings from the evaluation and presents a number of recommendations.

2. Course Design

This chapter describes and analyzes the content covered and the instructional strategies used in the Fellowship.⁸ Here the focus is on the design of the course—how it was *intended* to function. Subsequent chapters examine the *actual implementation* of the Fellowship and evidence of its success in producing its intended training outcomes. The analysis relies on interviews with JHF and Coro staff, examination of program documents, and direct observations of Fellowship sessions.⁹

2.1 Program Logic Model Used to Identify Intermediate Outputs

The key methodological challenge in this evaluation is to identify a set of intermediate outputs that can be observed in the short run and used to determine whether the program is on the right track.

A commonly used strategy for identifying intermediate outputs is the development of program logic models. Logic models represent—usually in graphical format—the sequence of activities that link program inputs and processes to intermediate outputs and ultimate outcomes. Thus, they can aid evaluations of programs with “downstream” outcomes by identifying a set of “upstream” outputs that likely precede the downstream outcomes (see, e.g., Chen, 1990; Rossi & Chen, 1992).

The logic model in Figure 2.1 seeks to identify (1) the ultimate outcomes Fellowship developers hope participants will engender when they begin their work as practicing professionals, (2) a set of observable intermediate *training outputs*, and (3) how those intermediate outputs are related to key *Fellowship inputs*. While we could not observe ultimate outcomes, we could observe these intermediate outputs and the Fellowship inputs that are designed to engender them. The remainder of this chapter describes these intermediate outputs and the Fellowship’s content and instructional methods.

It is important to point out, however, that we could locate no empirically validated models of similar training programs. Thus, the linkages in the model are merely speculative.

Outcomes in the Ultimate Use Context

As noted above, the ultimate goal of the Fellowship is to develop in participants a set of capabilities that will enable them to act as change agents in clinical “use contexts.” Understanding the skills and other requirements of operating in these contexts helps identify the types of upstream outputs and Fellowship inputs that might leverage eventual change in clinical use contexts. These use contexts are represented in the far-right panel of Figure 2.1.

Current thinking on medical error emphasizes the importance of the systemic factors that drive individual performance in clinical contexts. Systems involve the ways in which individuals are organized in the pursuit of common goals. While inappropriate individual behaviors and choices are often the *proximate* cause of medical errors, their *root* causes often lie in latent conditions and organizational factors that precede them.¹⁰ For instance, while fatigue might be the proximate cause of a nurse’s medication error, the root cause might lie in scheduling practices that provide insufficient staffing levels and prevent nurses from getting enough rest.

In the logic model, Fellowship participants are shown as initiating attempts to remove systemic barriers to safe practice, and otherwise improving systems of care to reduce the likelihood of error. As described by the Institute of Medicine, a systems approach to medical error is similar to the approach of draining swamps instead of trying to kill mosquitoes one by one (IOM, 2000). Specific systemic issues widely

⁸ The training curriculum studied in this report represents a modification to a pre-existing health sciences fellowship for graduate and professional students in the health sciences. As used in this report, “Fellowship” refers to the modified version.

⁹ Chapter 3 describes the direct observations in detail.

¹⁰ This discussion is based on the work of James Reason (1997) and also draws upon Nelson (2004).

cited in the literature include reimbursement policies, workload pressures, inadequate training, unclear procedures (and thus increased cognitive burdens), culture, and communication (IOM, 2000; IOM, 2001; Nelson, 2004).

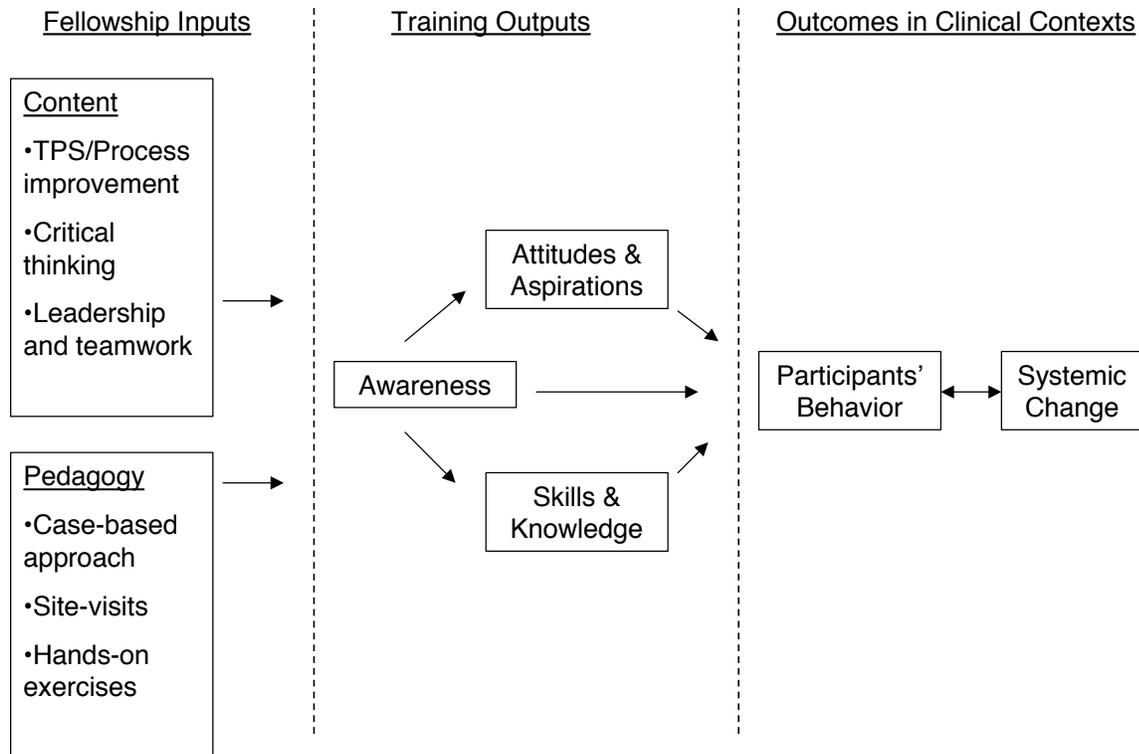


Figure 2.1 Program Logic Model

But the literature also suggests that there are formidable barriers to systemic change in healthcare systems. First, a decentralized network of actors delivers healthcare. Thus, appropriate execution of healthcare requires coordination among many individuals and groups, with the possibility of failure at each step in the care chain.¹¹ The problem is exacerbated by the increasing prevalence of chronic conditions requiring care from a number of largely autonomous and independent disciplines and specialties, which increases the likelihood of poor coordination, poor hand-offs, and other problems (IOM, 2000; 2001; Nelson, 2004). Thus, those who would initiate systemic reforms have the unhappy task of trying to motivate and coordinate a far-flung group of actors who are often not accustomed to working cooperatively to address problems.

Second, errors are likely to be under-reported due to a culture of blame in healthcare. Fault-based liability laws that discourage clinicians from confronting procedural problems openly sustain this culture. Pharmacists, for instance, often hesitate to call for clarification of physicians' prescriptions for fear of recriminations from physicians, resulting in increased risk of medication error. Hospital administrators, in turn, might fear that open discussion of medical errors could lead to increased legal liability (see, e.g., Runciman, Merry, & Tito, 2003). Thus, the logic model in Figure 2.1 contains a double-arrow running

¹¹ Pressman and Wildavsky (1984) offer a dramatic quantitative illustration of the impact of "the complexity of joint action" on the probability of program execution. Imagine that a given program task requires appropriate execution by ten separate actors. Even if the probability that each of those actors properly executes his or her part effectively is high (e.g., 0.9), the ultimate probability of success will still be low ($0.9^{10}=0.35$, assuming that the individual probabilities are independent of one another).

between participants' behavior and systemic change, in recognition of the fact that while individuals might seek to change systems, those same systems also present constraints on reformers.

Finally, while it might be possible to identify a core set of skills that can be used to effect change in organizations, the application of those skills will likely vary across organizations according to history, culture, individual personalities, and other factors. Thus, those seeking to effect changes in patient safety practices must be able to diagnose particular organizational contexts and adapt their skills to those contexts.

Intermediate Training Outputs

Although the ultimate goal of the Fellowship is to create in participants the capacity to act as change agents, it is necessary to identify more-immediate, short-term training outcomes that allow stakeholders to assess whether the program is “on the right track.” We know of no body of empirical research that identifies such leading indicators of future capacity in medical-error reduction and process improvement. Instead, we have patterned the logic model in Figure 2.1 on logic models created to evaluate programs in entrepreneurship education (see e.g., Krueger, Reilly, & Carsrud, 2000). As with the Fellowship, these programs typically seek to increase Fellows' skills levels, ability to leverage opportunities in organizational contexts, and their propensity to use their skills at some point in the future. Indeed, the ability to effect changes in patient safety practices represents a form of organizational entrepreneurship.

Based on this literature, we have identified three broad sets of variables that might serve as markers of future capacity. The first is *awareness* of medical error and patient safety issues among participants. Students and practitioners alike are subject to many competing demands on their time and attention. Thus, the first step is to ensure that trainees place a high priority on patient safety issues. Moreover, a challenge in many professional training contexts is to convince participants that the practices in which they are being instructed really do represent a clear departure from current practice. As Spillane, Reiser, & Reimer (2002) note, individuals tend to interpret new practices through the conceptual lenses of old practices, leading them to underestimate the extent to which changes in thinking and acting are required. Indeed, developers of the Fellowship often note a key challenge is making participants sufficiently “uncomfortable” with current practice and create a keen sense of “cognitive dissonance.”

Second, training must develop *knowledge* of patient safety concepts and approaches and the requisite *skills* in applying this knowledge in real-world settings. The distinction between knowledge and skills may also be described as the difference between “knowing that” and “knowing how” (see, e.g., Oakshott, 1962).

Finally, there is the need to develop the *motivation* to act on patient safety issues and to use skills and knowledge to effect change. A critical goal of the Fellowship, for instance, is to inculcate the expectation that the error rate in the healthcare system can, in fact, be driven to (or close to) zero.

In short, the literature on medical error suggests that attempts to improve patient safety must address systemic barriers to safety. Yet, these same healthcare systems possess features that hinder such changes. Thus, the Fellowship must inculcate a set of skills (1) that require considerable cooperation among a large set of actors, and (2) the execution of which will likely vary considerably from context to context. In the next sections, we explore the specific concepts and skills the Fellowship seeks to teach and how it seeks to inculcate them.

2.2 Core Content Included Safety Science Concepts and Teamwork Skills

The Fellowship seeks to inculcate both core concepts in patient safety and skills related to teamwork and critical thinking. Each is described in turn.

Systems Thinking: The Toyota Production System

Patient safety skills in the Fellowship are drawn largely from the Toyota Production System (TPS) and a locally developed curriculum developed by a nonprofit alliance, the Pittsburgh Regional Healthcare Initiative (PRHI). This curriculum—Perfecting Patient Care (PPC)¹²—involves both *core principles* and a set of *methods of investigation*.

Core principles. A number of core assumptions and value premises underlie the Toyota Production System and the Perfecting Patient Care system. The first of these is that most medical errors are rooted in systemic causes. Thus, improving patient safety requires careful root cause analysis, followed by interventions designed to address these systemic causes through redesign of core processes.

The second premise is that process improvement is usually most effective if undertaken by those at the “front lines” of care. This, in turn, requires a willingness on the part of management to give front-line workers the authority and discretion to make changes to core processes. Like other process-improvement approaches, however, TPS recognizes that systems change usually requires management support and action. Thus, management support—including encouragement, resources, and materials—is an essential element of the TPS approach.

One particularly important way in which management support is necessary for effective process improvement lies in the encouragement of a non-punitive culture that encourages open sharing of information about potential hazards. TPS require that management create and foster a climate in which workers do not fear reprisal for pointing out flaws in healthcare systems.

Finally, TPS is rooted in a set of core ethical values, including the importance of ensuring that healthcare workers’ work is meaningful and no more difficult than necessary. This requires attempts to eliminate wasted effort and maximize the amount of time workers can spend engaging in core patient care functions, as opposed to activities that merely “nurse the system.” Other values include ensuring the dignity of healthcare workers and patients, and appropriate recognition and reward for hard work.

Methods of investigation. The Toyota Production System also recommends methods for identifying and responding to systemic causes of medical errors, including the following:

- Identification of a patient need
- Careful observation of current work practices
- Identification of a problem with current work practices
- Investigation of the root causes of such problems
- Devising and testing a change in work processes
- Measuring results of the change.

The first three activities often involve selection of a “model line”—a productive process that might serve as an example for the rest of the organization—and a careful assessment of the “current condition” of that process. Identifying the current condition involves careful observation of supply sources and flows of information and materials—specifically, *who* is performing *what* activities in a productive process (e.g., delivery of medications to patients) along with *when* and *how* those activities are carried out (see, e.g., Harvard Business School, 2000). Such observations often generate a detailed diagram of steps for requesting material (e.g., medications) and the process by which those requests are responded to.

Root-cause analysis, in turn, is described in the Fellowship as the process of asking “the 5 Whys.” Here, the analyst begins with an unsatisfactory condition (e.g., a high rate of central line infections) and then asks why the condition exists. In a recent example from a Pittsburgh hospital, officials noticed an

¹² See www.prhi.org for a fuller description of the curriculum.

increase in the occurrence of pulmonary infections related to pseudomonas (Shannon, 2003). An initial examination by a problem-solving team suggested that the most likely cause was defective chemical sterilization. However, the team decided to inquire whether use of the defective chemical sterilization process was itself evidence of deeper problems. Thus, the team kept asking “why” questions in order to probe into deeper root causes. For instance, when the team asked why chemical sterilization was being used (which is less effective than sterilization using Ethylene Oxide gas), the team found that this quicker process was being used in response to time pressures. Further, the team determined that the increased time pressures were caused by the need for more bronchoscopies, which, in turn, were caused by a higher incidence of ventilator-associated pneumonias (VAP). Finally, the team linked the increase in VAP to the use of a new antibiotic regimen in the hospital.

After identifying a root cause (or set of root causes), the next step is to devise a fix for the problem. TPS provides “four rules in use” that provide guidance for system redesign. The rules (see Figure 2.2) are provided to Fellowship participants on a laminated index card. The rules urge users to ensure that steps in the care chain (1) are related to outcomes, (2) are clear, (3) are highly specified, and (4) provide ongoing signals about system performance.

<p><u>The Four Rules in Use</u></p> <p><u>Design in Use</u></p> <p><i>Rule 1:</i> All work must be highly specified as to content, sequence, timing, location, and expected outcomes.</p> <p><i>Rule 2:</i> Every customer-supplier connection must be highly specified and direct, and there must be an unambiguous yes-or-no way to send requests and receive responses.</p> <p><i>Rule 3:</i> The pathway for every product and service must be predefined, highly specified, simple, and direct, with no loops or forking.</p> <p><i>Rule 4:</i> Any improvement must be made in accordance with the scientific method, under the guidance of a teacher, at the lowest possible level in the organization.</p> <p><u>Test in Use</u></p> <p>All four Rules have built-in internal tests that let you know if the activities, connections, pathways, and improvements are being done as expected.</p> <p><i>Source: Steven Spear & H. Kent Bowen, “Decoding the DNA of the Toyota Production System”, Harvard Business Review, Sept-Oct, 1999, p. 96.</i></p>
--

Figure 2.2 Content of the Index Card on “The Four Rules in Use”

In the example above, the problem-solving team not only improved the sterilization process, but also developed a new pre-procedure checklist and clearer documentation of standardized procedures. In doing so, the team followed the TPS principles by clarifying and standardizing key processes in the care chain.

An essential element of the TPS and PPC systems is that these fixes are devised immediately after identification of the problem and root causes. The assumption is that root causes often become less visible, and the motivation to fix them less acute, as time passes. Having devised a fix, the analyst essentially executes a small field experiment to determine whether the fix is effective in addressing the problem. In most instances, this involves comparing system performance before and after introduction of the fix.

The summer 2004 Fellowship was offered in two tracks, “Getting to Zero” (GTZ) and “Getting to Excellence” (GTE), in an effort to experiment with minor variations in instructional techniques and

content foci. While the focus of both tracks was on patient safety, the GTE track also included units on palliative care and public health. Table 2.1 shows the schedule of topics used in both Fellowship tracks. The “core values” discussed above are evident in topics related to the “patient perspective” and “teamwork.” Similarly, the emphasis on the analytical processes of TPS and PPC are evident in topics related to observation, work redesign, and real-time problem solving.

Table 2.1. Fellowship Topics

Week	GTE Topics	GTZ Topics
1	Introduction and overview	Introduction and overview
2	Medical informatics; real-time data; team-based problem solving	Nature of the problem; basics of TPS and process improvement; teamwork
3	Simulation and teamwork	Reconnaissance and observation and patient perspective
4	Work redesign	Work redesign; clinical micro-systems
5	Leadership and hospital culture	Real-time data collection and problem solving
6	Palliative care	Patient safety from workers' perspective/worker safety
7	Public health/disease eradication	Care pathways and redesign
8	Team application	Team application

Source: JHF Documents

Leadership and Critical Thinking: The “Coro Skills”

Along with the core content on safety science, Fellowship participants also received ongoing training in critical thinking skills and teamwork. The emphasis on critical thinking can be viewed as a complement to TPS’s emphasis on careful observation and use of the scientific method to evaluate the impact of changes in healthcare processes. Fellowship instructors used a tool called the “Ladder of Inference” (see Figure 2.3) to emphasize the importance of clearly distinguishing facts from inferences, assumptions, and opinions.

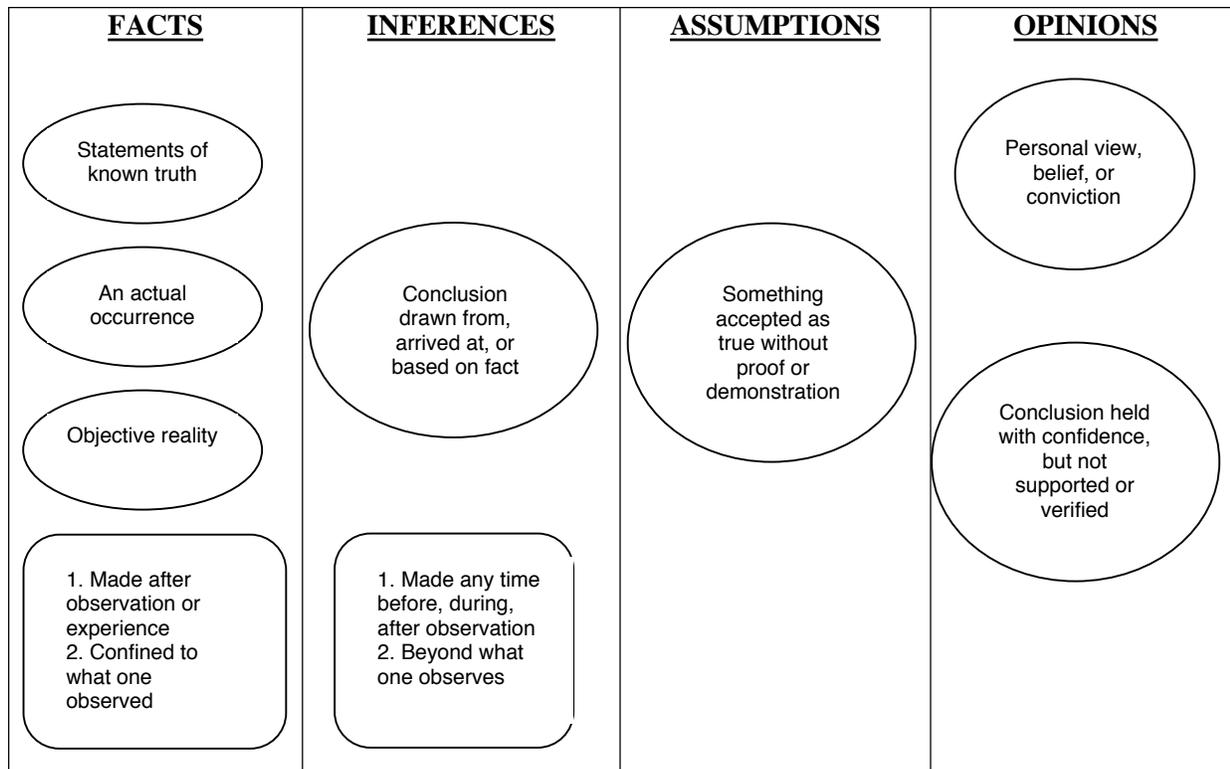


Figure 2.3 The Ladder of Inference

Training on leadership and teamwork skills focused largely on a tool called OARRS—Outcomes, Agendas, Roles, and Rules—which was used to train Fellows to structure and execute group activities. *Outcomes* refers to the intended goals of a collective activity, *agendas* refers to the sequence of events by which the activity is to proceed, *roles* refers to the functions that each team member performs, and *rules* refers to the boundaries or guidelines that govern interactions among team members. For, instance, participants used the tool to plan for sessions in which they interviewed—as a group—guest speakers from local healthcare facilities.

The Fellowship also sought to develop leadership and teamwork skills by exposing Fellows to experiences working in multi-disciplinary teams. As noted above, healthcare is jointly produced by clinicians from a wide variety of backgrounds, along with hospital administrators, policymakers, and others. Thus, the Fellowship’s ability to bring together participants from a wide variety of disciplinary backgrounds is itself an important element of the Fellowship.

2.3 Pedagogical Approach Emphasized Exposure to Local Clinical Settings

Training in the Fellowship was almost entirely built around a series of site visits to clinical settings in the Pittsburgh area. The site visit locations were chosen to exemplify best practices in patient safety and to feature innovative practitioners. During each session, an introductory lecture by JHF and Coro staff was followed by opportunities for the Fellows to listen to and interview (using OARRs) guest speakers.

Interviews were often followed by a tour, during which Fellows could directly observe patient safety–related practices. For instance, to encourage team functioning during a session on the use of medical simulators, Fellows actually used a simulator. Similarly, during a session at a local pathology lab, resident pathologists and lab technicians gave participants a tour of the lab.

In addition, participants in the GTZ track were given a series of hands-on exercises that typically involved interviewing practitioners, diagramming processes, and suggesting process and system redesigns.¹³ The case-based approach is used by the Coro Center in its own programs on leadership development. There, teams of participants from varied backgrounds are put into situations in which they have to solve problems with little preparation. In the words of one Coro official, the situations are designed to “force people to synthesize.”

2.4 Summary and Conclusions

This chapter sought to provide an overview of the design and content of the Fellowship. While subsequent chapters examine how the course was actually implemented, the goal in this chapter was to describe how the Fellowship was *intended* to function.

For the purposes of an evaluation, the Fellowship’s most salient feature is that, like many training programs, it seeks to engender in participants the capacity to make changes in the future. Consequently, the Fellowship’s outcomes will happen, if at all, long after the end of this evaluation project. Thus, we used a logic model to identify intermediate outputs that can be observed in the near term, including (1) enhanced awareness of patient safety issues, (2) increased prevalence of attitudes about safety that are congruent with current “systems” thinking on safety, and (3) the development of knowledge and skills that can be used to diagnose and respond to the systemic causes of medical error.

The Fellowship emphasized not just core concepts in safety science (e.g., process improvement) but also critical thinking, leadership, and teamwork skills. This can be viewed as a supplement to the process-improvement methods’ emphasis on careful observation and the scientific method. Teamwork and leadership skills, in turn, reflect the fact that most healthcare systems involve multiple—and often autonomous—actors from a wide variety of professional and disciplinary backgrounds. The Fellowship’s pedagogical approach featured exposure to a variety of practitioners in local clinical settings.

The Fellowship was offered in two “tracks”—GTE and GTZ. While the focus of both was on patient safety, the GTE track also included sessions on palliative care and public health. In addition to these minor content differences, each track used different instructional methods, with Fellows in GTZ getting more opportunities for hands-on problem solving. Accordingly, where possible, the analyses provided in the remainder of the report include comparisons across the two tracks.

¹³ The case-based approach might address all three of the intermediate outcomes described in the logic model. The case might be expected to build *awareness* through direct exposure to patient safety problems in clinical settings. Similarly, the case-based approach might build *attitudes and motivation* through exposure to exemplary individuals and institutions. In particular, this approach seeks to expose participants to those who really believe that one can “get to zero” medical errors. Finally, the approach develops *skills* by allowing participants to talk with, and observe, practitioners who are putting patient safety skills into practice.

3. Evaluation Methods

Given that the Fellowship design process is in its early stages, the work reported on here is a *formative* evaluation. Formative evaluations are typically conducted during the development of a program and provide useful feedback on program design, redesign, and scale-up. *Summative* evaluations, by contrast, are designed to inform “thumbs-up, thumbs-down” decisions about whether programs deserve funding (or continued funding) (Scriven, 1991). Formative evaluations are especially appropriate when outcomes are distal and decisionmakers seek to make mid-course corrections along the way.¹⁴

3.1 Data Collection Sought Information on Outcomes and Mechanisms

Data collection strategies were designed to measure not only outcomes but also mechanisms and processes. The RAND Human Subjects Protection Committee approved all data collection instruments.

Document Review

Course syllabi, readings, and other artifacts were examined in order to determine the content focus of the sessions. In addition, the evaluation team reviewed recent literature on patient safety and medical error in order to understand how aspects of the curriculum relate to what is currently known about the drivers of medical errors and error reduction strategies.

Participant Survey

A survey was administered near the end of the course to solicit information on participants’ experiences. Some survey items were designed to elicit information relevant to attitudinal outcomes and perceptions of the relevance of course content to professional practice. The survey employed a retrospective pre/post design (see, e.g., Pratt, McGuigan, & Katzev, 2000), in which respondents are asked to judge changes after the fact. Unlike the traditional pre/post approach, which compares responses given before and after the intervention, the retrospective pre/post method asks respondents to recall their pre-intervention attitudes retrospectively. The advantage of the design is that it avoids contamination of pre/post assessments by changes in respondents’ evaluative standards. Commonly, individuals come away from training with a greater appreciation for what they do not know, which can bias conventional pre/post impact assessments. JHF staff reviewed the survey instrument prior to administration. A copy of the survey instrument is provided in Appendix A.

Surveys were distributed during the final session of the Fellowship along with a notification to Fellows of their rights as human subjects and a postage-paid return envelopment. The final response rate was 79 percent, with 31 of the 39 participants returning the survey. The response rate was slightly higher for the GTE track than for the GTZ track, with 19 of 22 (86 percent) GTE and 12 of 17 (71 percent) GTZ participants returning surveys. Statistical techniques used to analyze the data are discussed as they arise throughout the report.¹⁵

Participant Focus Groups

Focus groups with participants sought to gather more-nuanced data about the design, execution, and outcomes of the Fellowship. Scheduling considerations compelled us to run three separate focus groups, two for participants in the GTZ track and one for participants in the GTE track. Participation in each was low, with three in one group, two in the second group, and one in the third. The focus groups were tape-

¹⁴ As Stake summarizes, “When the cook tastes the soup, it is formative evaluation; when the customer tastes the soup, it is summative evaluation” (Stake, 2000).

¹⁵ Analysis of items in the survey covering similar content shows moderate correlations in the expected directions. For instance, items on root cause analysis had Spearman’s rho values of 0.4 to 0.6. The relationships were statistically discernible at the 0.05-level.

recorded and used a semi-structured protocol (see Appendix B). All Fellows were invited to participate in focus groups. The fact that only five did participate suggests considerable risk of selection bias.

Observations

Three sessions were observed for each of the two Fellowship tracks. Accompanying teaching materials and artifacts were also collected. Observations were guided by the research questions but did not employ a formal protocol.

Work Samples

De-identified work samples were examined to supplement the assessment of learning outcomes.

3.2 Limitations

Given its formative orientation and data limitations, this evaluation cannot provide a definitive estimate of program impact on training outcomes. Impact estimates require comparing outcomes in a treatment group with outcomes in a comparison or control group. No such comparisons were available for this study. However, Chapter 7 provides some recommendations about how such a contrasted-group study might be designed and executed.

Given content and pedagogy differences across the two Fellowship tracks, the analysis presented below seeks to detect differences in responses across the two tracks. Given the small sample sizes, however, the statistical analysis of survey data suffers from low statistical power, which reduces the probability of finding statistically discernible differences across the Fellowship tracks. To illustrate the problem, Figure 3.3. plots required sample sizes (per track) by the minimum detectable effect we might wish to find. As is indicated by the plot line in the figure, given the sample sizes achieved, the minimum detectable difference for this study is approximately 0.6 points on a five-point scale.¹⁶ Detection of effects as small as 0.2 points on a five-point scale would require approximately 200 subjects from each track.

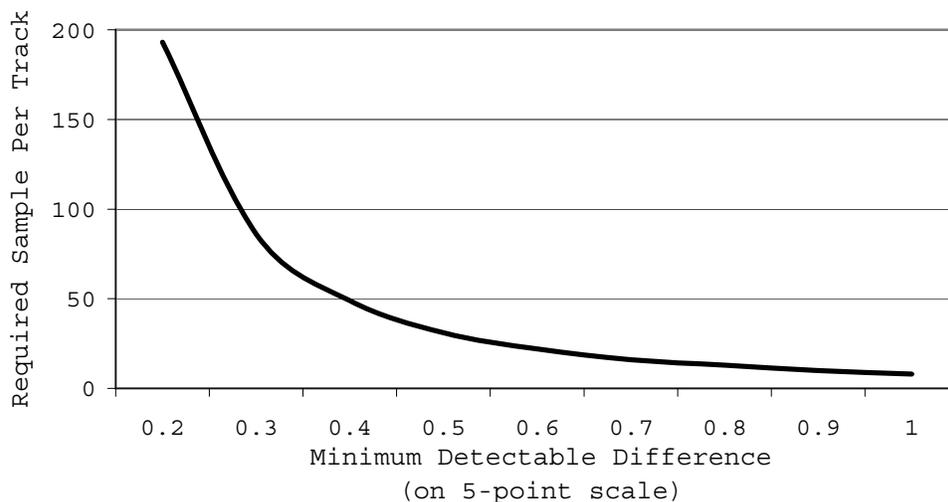


Figure 3.1 Power Calculations

¹⁶ Calculations assume power = 0.8, $p = 0.05$, and a standard deviation of 0.7 points on a 5-point scale.

Perhaps more importantly, attempts to compare the causal efficacy of the GTE and GTE approaches are limited by self-selection of participants into these two groups. Conversations with JHF staff suggest that participants' decisions on which track to sign up for were driven largely by scheduling consideration. However, it is likely that in at least some cases participants selected a track on the basis of factors correlated with their propensity to benefit from one of the two tracks, which would confound any attempt to draw causal inferences from observed differences between the two groups.

In spite of these limitations, we provide information in this report that will be useful in considering improvements to the Fellowship, and might inform the design of similar instructional curricula.

3.3 Summary

The evaluation methods used in this study were designed to elicit information about inputs, processes, and outcomes. Information about inputs and processes is necessary in order to derive actionable recommendations for improving program outcomes. Data collection methods included document review, a participant survey, participant focus groups, observations, and examination of work samples. Given data limitations, however, this report is unable to provide rigorous assessments of program impact. Moreover, small sample sizes limit the power of statistical comparisons.

4. Participants

This chapter provides an overview of Fellowship participants, including their educational and professional backgrounds, prior exposure to patient safety issues, and how they became aware of the Fellowship. Such knowledge can be useful in determining whether the Fellowship reached its target audience, and will also inform sections of the report on curriculum, pedagogy, course implementation, and outcomes. Throughout, the report notes significant differences between the two Fellowship tracks – Getting to Excellence and Getting to Zero.

Data for this chapter came from the participant surveys. The need to ensure respondent confidentiality, along with the small number of participants, limited the range of questions we could ask about background characteristics.¹⁷

Throughout, readers should bear in mind that the survey response rate, while reasonably high, was not perfect (see Chapter 2). Thus, survey findings might be biased if those choosing to respond to the survey differed in systematic ways from those who did not respond. Moreover, the response rate was higher for participants in the GTE track than for those the GTZ track.

4.1 Most Fellows Were at an Early Stage in Their Professional Training

The survey data suggest that the Fellowship succeeded in its goal of reaching Fellows at an early stage in their professional training. Some 27 of 31 survey respondents (87 percent) reported having been students immediately prior to enrolling in the course (see Table 4.1). Most of the remaining respondents reported working as some kind of practicing healthcare professional. Six respondents reported being both students and practicing healthcare professionals. There was little difference between the tracks, with 89 percent of GTE and 83 percent of GTZ participants reporting that they were students.

Most respondents (71 percent) reported a post-baccalaureate degree as their highest degree level, more so for GTE than GTZ Fellows (83 versus 63 percent). However, the difference in Fellowship tracks was not statistically discernible.¹⁸

All respondents reported having had at least some prior exposure to patient safety issues. Some 81 percent reported having read or heard about patient safety issues in the news. Nearly half had either read a book, report, or journal article on the subject, or had found material on the Internet. One respondent reported being exposed to patient safety issues through prior coursework in health policy and through work experiences. Overall, respondents from the GTE track reported more sources of exposure (an average of 3.5 sources) than did respondents from the GTZ track (an average of 2.3 sources)—a statistically discernible difference ($p = 0.04$).¹⁹ Fellows in the GTE track were also more likely to have read books or reports on patient safety. Finally, more GTE than GTZ respondents had themselves observed troubling patient safety practices (68 versus 42 percent).

¹⁷ As noted in Chapter 2, respondents were asked to not identify themselves on the survey. However, including detailed questions about one's background might lead to identification through inference. As with all data-collection instruments, the survey was approved by RAND's Human Subjects Protection Committee.

¹⁸ We use the term “statistically discernible” instead of the more conventional “statistically significant” to distinguish statistical from practical significance. Given the low statistical power of the sample, we consider differences to be “discernible” if $p < 0.10$, rather than the conventional $p < 0.05$. Generally, findings that do not meet this threshold should be interpreted as holding for the population included in this evaluation, but not necessarily for other populations.

Unless otherwise noted, the nonparametric Mann-Whitney test was used to assess the statistical discernibility of cross-track (GTZ versus GTE) differences. The Wilcoxon sign rank test, which allows for non-independence between groups, was used to assess the statistical discernibility of retrospective pre/post differences.

¹⁹ The statistical test was the Kruskal-Wallis test, a nonparametric analogue of analysis of variance (ANOVA).

Table 4.1 Fellows' Professional Backgrounds

Occupation	Number of Respondents		
	Getting to Excellence	Getting to Zero	All Respondents
Student	17	10	27
Practicing physician/nurse	3	1	4
Other practicing healthcare professional	1	1	2
Other	3	1	4
Number of Respondents* ^a	19	12	31

^a Total number of respondents does not add up to total number in occupation categories because some respondents checked more than one item.

Source: RAND research

4.2 Fellows Came from a Wide Range of Backgrounds

As noted above, the Fellowship sought to recruit from a wide range of disciplines in order to provide a view of healthcare as an interconnected system. Fragmentation in the healthcare system has been identified as a condition that encourages medical errors, whether through poor hand-offs of patients and materials or poor coordination and communication (IOM, 2000; IOM, 2001). Fellows represented a wide variety of professional orientations. As Table 4.2 shows, more than half of respondents expected to go into administration, research, or public policy.²⁰ Just one-quarter of respondents expected to be practicing physicians or nurses. The proportion of those intending to work in administration, research, or policy was slightly higher in the Getting to Zero track, but the difference was small and not statistically discernible ($p = 0.31$).

Table 4.2 Fellows' Intended Professions

Occupation	Number of Respondents		
	Getting to Excellence	Getting to Zero	All Respondents
Student	2	3	5
Practicing physician/nurse	3	4	7
Other practicing healthcare professional	9	8	17
Other	7	5	12
Number of Respondents ^a	19	12	31

^{a*} Total number of respondents does not add up to total number in occupation categories because some respondents checked more than one item.

Source: RAND research

²⁰ Once again, our ability to collect fine-grained data on respondents' backgrounds was limited by the need to prevent the possibility of identification by inference.

4.3 Differences in Perceived Motivation Across Fellowship Tracks

Most respondents reported that other Fellows were well motivated. The average response to the item “Other students were well motivated” was 4.0 on a scale of 1 to 5, with 5 indicating “Strongly Agree.” However, there was a notable difference between the two tracks, with GTZ Fellows giving higher marks than GTE Fellows to their peers’ motivation (4.3 versus 3.8). While the difference is not quite statistically discernible ($p = 0.14$), it is worthy of mention.

4.4 Most Fellows Learned of Program Through Teachers

Most Fellows (77 percent) learned about the Fellowship through a teacher, professor, or other instructor. The next most common way of learning about the Fellowship was through a JHF or Coro recruiter (42 percent), followed by contacts with past JHF/Coro Health Sciences Fellows (26 percent), professional colleagues/peers (9 percent), and the JHF and Coro Web sites (6 percent). Half of the respondents learned about the Fellowship from more than one source.

4.5 Summary and Conclusions

In summary, analysis of the surveys suggests that Fellows were generally in the early stages of their professional training. Thus, the summer 2004 Fellowship appears to have achieved its goal of targeting developing healthcare professionals. However, the surveys also suggest that Fellows in the GTE track had more prior training, professional experience, and first-hand experience with patient safety issues than those in the GTZ track. Fellows also came from a wide variety of disciplinary backgrounds, suggesting that the Fellowship also met its goal of attracting a professionally diverse group. Finally, this chapter also provided information on how Fellows learned of the Fellowship, which might prove useful in planning for future recruitment efforts. Given that most Fellows were students before joining the Fellowship, it is not surprising to find that most learned about the Fellowship through teachers.

5. Implementation

This chapter examines the actual implementation of the Fellowship curriculum during summer 2004. The chapter begins by examining Fellows' perspectives on course design, with a discussion of the overall pedagogical approach, course materials, activities, and interactions among Fellows and between Fellows and instructors. Data for this chapter come from a survey, focus groups, and direct observation of a sample of sessions (see Chapter 3 for details).

5.1 Fellows' Perspectives on Course Design

The surveys and focus groups sought to elicit Fellows' perspectives on the Fellowship's content and pedagogy. Generally, reactions were very positive. However, many respondents provided suggestions for improvement. For the most part, there were few differences between the GTE and GTZ tracks. However, GTZ respondents provided fewer responses to open-ended questions on the survey instrument.²¹ We begin with a discussion of course content and turn next to pedagogy.²²

Respondents Found the Material to be New and Useful

Focus group participants were asked to discuss their opinions about the content covered in the Fellowship. For the most part, the response was very positive, with most reporting that they found the topics relevant and interesting. The participant survey asked respondents to rate the extent to which the "Fellowship presented material that was new to me." The average response was 4.5 on a scale of 1 to 5, with 1 indicating strong disagreement and 5 indicating strong agreement. The responses were very consistent, with all but two of the 31 respondents responding with a 4 or 5.²³

An open-ended question on the survey asked respondents to list specific concepts and skills that were new to them. The most frequent responses included:

- Holding effective meetings
- Ladder of Inference (see Chapter 2) and evaluating factual claims
- Toyota Production System (see Chapter 2)
- Observation
- Assessing systems, including "diagramming work flows and designing processes."

One respondent was pleased to learn more about "identifying a need and identifying an answer to that need." The same respondent also valued the opportunity to learn more about "prioritizing needs" and "the value of observation." However, most respondents were less enthusiastic about a Fellowship session on computational simulations. Here, respondents were concerned about the technical nature of the material and questioned whether the material was sufficiently related to other course content to merit inclusion.

Tension Between Breadth and Depth in the Range of Topics Covered

There was considerable disagreement among respondents about the relative virtues of breadth versus depth in content coverage. A few focus group participants cited the range of topics covered as a strength

²¹ As noted in Chapter 2, the survey response rate also was slightly lower for GTZ participants.

²² Where findings are based on the participant survey, precise numbers are offered to support the interpretations. Otherwise, readers may assume that findings are based on findings from the focus groups. As noted in Chapter 3, five Fellows participated in the focus groups.

²³ The standard deviation for the item was 0.6.

of the Fellowship. One respondent stated, *“One of the strengths was the fact that so much material and so many topics were covered and presented.”*

However, a larger number of focus group respondents expressed some concern that breadth of topics covered came at the expense of depth. One respondent wrote that there was *“not enough time to digest all of the information presented at each session.”* Other respondents wished that there could have been greater depth on a number of key points. One, for instance, believed that Fellows would have benefited from a deeper and more nuanced examination of TPS. This respondent pointed to a site visit in which Fellows observed a nursing unit that had fully embraced TPS. However, quite accidentally, some Fellows came across another nursing group in which TPS was *“practically unheard of.”* To this respondent, the occurrence pointed out the need to explore more fully the limitations of, and barriers to, the TPS approach. Other respondents countered that a certain degree of simplification was necessary when teaching new concepts. However, even these respondents agreed that there had not been enough attention to *“what can be done when things don’t work out as planned.”*

A smaller group of respondents believed that, if anything, there could have been more breadth in course content. These respondents noted that nearly all of the site visits and concepts presented in the course involved clinical micro-systems, and suggested replacing one of the sessions with a topic that involves inter-clinical relationship or policy-level topics such as the incentives that reimbursement policies create (or fail to create) for quality and error reduction. Another respondent reported being *“surprised”* by the *“specificity”* of the topics, but found that each of the topics and site visits contained enough generality to be interesting to a broad range of Fellows. *“I was usually able to find how these topics applied to my own work,”* noted this respondent.

Mixed Views on Leadership and Critical Thinking Skills

The tension between breadth and depth was perhaps most evident in respondents’ opinions about attempts to include leadership and critical thinking skills (the “Coro skills”) in the curriculum. Many respondents found the Coro-skill topics to be useful and interesting; as noted above, Coro skills (especially OARRs and the Ladder of Inference) were among those skills most often cited by respondents as being new to them. However, opinion was deeply divided the importance of Coro-skill instruction. One respondent believed that, if anything, teaching the Coro skills is more important than teaching some other topics. This respondent stated that while JHF and Coro can assume that everyone knows something about healthcare, they should make no assumptions about Fellows’ backgrounds in leadership skills. *“None of us are trained in that,”* according to this participant. *“This is our opportunity [to learn these skills].”* Other respondents noted that the Coro skills *“got brushed over,”* especially as the end of the Fellowship neared. *“I think the skills are really important,”* stated one respondent. *“But I’m not sure I necessarily got it.”* Another respondent was more succinct, stating that the Coro skills were *“useful, but not used.”*

A smaller number of respondents thought that the Coro skills were too simplistic. Another speculated that training in leadership skills might be more appropriate for younger, less-experienced Fellows than for those with considerable work experience.²⁴

Several respondents wondered whether it is feasible to introduce both patient safety concepts and leadership/critical thinking skills in an eight-week session. One respondent noted that the Coro approach of *“throwing students into the deep end to see if they can swim”* requires more time than is possible in an eight-week session that also covers patient safety. According to this participant, *“People tend to shut down and not ask all of the questions they might have”* when confronted with such time pressures. Nevertheless, most respondents were reluctant to see the Coro skills dropped in the interest of time.

²⁴ Unfortunately, the data collected did not permit a straightforward examination of whether opinions about the Coro skills were related to level of experience.

Some Concern About Overall Coherence

One participant suggested that the eight-week course could, in fact, accommodate both the patient safety and leadership/critical-thinking content if it were more clearly focused around “*a centralized vision.*” Similarly, other respondents noted that Fellows would often “*walk into sessions blind*” and would have benefited from a clearer overarching analytical framework for the course. One respondent noted that often the most important questions occurred to Fellows only after they had participated in a session, and that a stronger overarching framework would have allowed Fellows to get more out of the sessions. Another noted the need for more focused “*big questions*” before sending students out to do assignments. As noted later in this chapter, however, most Fellows seemed content with the degree of overall course coherence.

5.2 Materials and Activities

Having described the Fellowship’s content and overall design, we turn next to Fellows’ views on materials and activities.

Readings Perceived As Useful, But Not Integrated Well Enough with Other Materials and Activities

Each week, Fellows were asked to read approximately three to five articles related to the week’s topic and site visit. Reactions to the readings were mixed.

For the most part, respondents found the readings to be appropriate to the sessions. One respondent, for instance, said that the readings provided “*good background*” and “*helped to answer the question, ‘Why are we even bothering to cover this?’*” This same respondent also noted that the readings “*save time for the actual get-together*” by laying out important concepts prior to the session.

For most respondents, the readings were at the appropriate level of difficulty. Indeed, the average response to the survey item, “Course readings were too difficult”, was 2 on a scale of 1 to 5, with 1 representing “Strongly disagree” (see Table 5.1). Interestingly, GTZ Fellows found the readings a little more difficult than GTE Fellows, with an average rating of 2.3 for GTZ and 1.7 for GTE—a statistically discernible difference ($p = 0.03$). Data from the focus groups suggest that this difference might reflect the fact that Fellows in the GTZ track had less prior experience, and that this track had more students seeking to juggle Fellowship readings with readings from other courses. Some Fellows thought that the readings were often repetitive, and that the reading list could be pared back.

**Table 5.1. Fellows’ Assessments of Readings
(Means and standard deviations)**

Survey Item	GTE Fellows	GTZ Fellows	All Fellows
Course readings were too difficult.	1.7 (1.0)	2.3 ^a (0.8)	2.0 (1.0)
Course readings were clearly integrated into the sessions.	3.8 (0.9)	3.5 (1.1)	3.7 (0.9)

^a Differences between tracks are statistically discernible at $p < 0.05$ (Mann-Whitney Rank Sum test).

Source: RAND research.

For the most part, however, respondents believed that the reading load was reasonable. One noted that the reading burdens associated with the Fellowship were similar to that in other courses that students typically take during their training.

There was, however, some concern about the extent to which the readings were integrated with other instructional activities. The average response to the item “Course readings were clearly integrated into

the sessions” was 3.7 on a scale of 1 to 5, with 5 representing “Strongly Agree.” This average is somewhat lower than the typical average for similarly scaled items, which was between 4 and 5. While the rating was slightly higher for the GTE track, the difference was not statistically discernible ($p = 0.51$). One respondent noted that, “*Sometimes the readings weren’t mentioned, so I wasn’t sure how exactly they pertained to the topic of interest.*” Another respondent believed it would have been useful to have a stronger sense of priorities among the readings. This respondent said, “*I really wish I would have known what was considered by them [to be] more relevant than others.*”

Some respondents suggested reducing the number of readings but increasing the amount of class time spent discussing the remaining readings, which, they believed, might promote greater depth in understanding. Another respondent suggested providing study questions with the readings in order to focus Fellows’ attention on core issues while they read. One respondent acknowledged that instructors did attempt to introduce the next week’s readings during each session, but suggested that the introductions were not always sufficient.

Opportunities for Hands-On Experience Were Well-Received

Generally, respondents cited the site visits and guest speakers as a particular strength of the Fellowship. One focus group participant noted the importance of actually meeting and seeing practitioners who have accepted “zero errors” as a realistic goal. For the most part, this positive view was borne out in the participant surveys. For instance, the average response to the item “Sessions presented enough concrete examples to provide a thorough understanding of key concepts” was 4 on a scale of 1 to 5, with 5 indicating “Strongly Agree” (see Table 5.2). However, respondents were less sanguine about their opportunities to *apply* Fellowship skills during sessions. The average response to the item “Sessions provided adequate opportunity for me to learn how to apply the skills we were expected to master” was 3.5 on a scale of 1 to 5, with 5 indicating “Strongly Agree.”

**Table 5.2. Fellows’ Assessments of Opportunities for Hands-On Experience
(Means and standard deviations)**

Survey Item	GTE Fellows	GTZ Fellows	All Fellows
Sessions presented enough concrete examples to provide a thorough understanding of key concepts.	3.8 (0.8)	4.2 (0.7)	4.0 (0.8)
Sessions provided adequate opportunity for me to learn how to apply the skills we were expected to master.	3.5 (1.0)	3.5 (1.1)	3.5 (1.0)

Source: RAND research.

Respondents in the GTZ track were more likely than those in the GTE track to agree that the Fellowship provided enough concrete examples (an average score of 4.2 compared with an average score 3.8). The difference is notable, if not statistically discernible ($p = 0.26$).²⁵ However, there were no such differences in responses to the item about opportunities to apply skills. The failure to find differences between the tracks on this latter item is surprising given that the GTE track, by design, relied less on practical exercises than did the GTZ track. However, the differences did come through in responses to open-ended survey questions. As one GTE respondent put it, “*Less talk, more shop.*”

Some Concern About Student Presentations

The Fellowship made extensive use of student presentations. First, each week, a student would provide a brief (usually three- to five-minute) presentation on readings relevant to the week’s topic, site visit, and

²⁵ Given the small samples sizes, the comparisons presented in this report have low statistical power. Thus, it is appropriate to consider findings that fail to meet conventional thresholds of statistical discernibility. See Chapter 2 for further discussion of statistical power.

speaker. However, several respondents (including those responding to open-ended survey questions and those in focus groups) noted that they were not sufficiently integrated into the remainder of the sessions. Others noted that presentations often came *after* interviews with the guest speakers, reducing their utility as tools for preparing for the interviews. Other respondents suggested that there was not enough time provided for the presentations. A few concluded that the weekly presentations did not add enough value to be worth including in the Fellowship.

Second, group break-out sessions were often followed by brief presentations by students on the groups' findings. Once again, several respondents believed that there was not enough time given for these presentations. Others doubted whether they added enough value to warrant the use of scarce session time. One respondent noted that group presentations often were very similar to one another, which this respondent took as evidence that the group sessions did not add enough value or were not structured well enough to elicit valuable responses.

Finally, both Fellowship tracks culminated in a public presentation to other Fellows and a group of local practitioners. Here Fellows were given a scenario in which a medical error had occurred and were asked to formulate—within a period of about an hour—a strategy for addressing the error (and any appurtenant systemic causes). Many respondents said that they had been given too little time with their groups throughout the Fellowship. As with the weekly presentations, some respondents expressed concern that the assignment was too vague, and that such vagueness was “*driving down the quality*” of the presentations.

Concerns about vague directions might reflect the design of the Fellowship, particularly the elements informed by the Coro Center. The Coro approach, as noted above, features placing students into situations and forcing them to “figure it out” on their own. Thus, it is not clear to what extent these findings reflect (1) a failure to adequately communicate this pedagogical approach, (2) inadequate execution of the approach, or (3) participant resistance to the approach.

5.3 Instructors and Course Management

The surveys and interview protocols also included items on instructors, guest speakers, and course management.

Instructors and Guest Speakers Generally Received Favorable Ratings

Fellowship sessions were lead by a core team from JHF and Coro and supplemented by a series of guest speakers affiliated with the site visit locales. For the most part, respondents reported being impressed by the instructors' grasp of the material. The average response to the item “Fellowship staff were knowledgeable about patient safety” was 4.5 on a scale of 1 to 5, with 1 indicating “Strongly Disagree” and 5 “Strongly Agree.” Responses to a similar question about guest speakers were largely the same, with an average score of 4.3 on the same scale. Moreover, responses to both items were very consistent, with all respondents marking a 4 or 5 on the item about Fellowship staff, and all but one respondent marking a 4 or 5 on the item about guest speakers.

For the item on Fellowship staff, responses from the GTE and GTZ tracks were indistinguishable. However, respondents' assessments of the guest speakers' knowledge were discernibly higher in GTZ (a mean of 4.8) than in GTE (a mean of 4.4). Descriptive statistics are provided in Table 5.3. The slightly lower scores from the GTE track were reflected in responses to open-ended questions on the survey. Whereas GTZ respondents offered little, if any, elaboration, a few respondents from the GTE group questioned whether some of the guest speakers had been adequately prepped for the sessions. One noted that, “*Some speakers did not clearly understand our expectations or had unclear expectations of us.*” In a similar vein, another respondent noted that, “*I think some of the speakers could have used more direction as to how the time would be spent.*” As noted above, however, the Fellowship is designed to

require the Fellows themselves to structure and run interactions with guest speakers. Thus, it is possible that these respondents either did not fully understand, were not equipped for, or did not welcome this role.

Respondents' assessments of instructors' ability to present the material clearly and effectively was positive. Here average ratings on the same five-point scale were 4.3 and 4.2 for Fellowship staff and guest speakers, respectively (see Table 5.3). Interestingly, there was slightly more variation in respondents' evaluations of guest speakers' presentation skills, as indicated by the higher standard deviations. Nonetheless, no more than a handful of respondents gave ratings lower than 3.

**Table 5.3. Fellows' Assessments of Instructors
(Means and standard deviations)**

Survey Item	GTE Fellows	GTZ Fellows	All Fellows
Fellowship staff were knowledgeable about patient safety.	4.5 (0.5)	4.4 (0.5)	4.5 (0.5)
Guest presenters were knowledgeable about patient safety.	4.4 (0.5)	4.8 ^a (0.5)	4.5 (0.6)
Fellowship staff effectively and clearly presented the material.	4.4 (0.6)	4.3 (0.9)	4.3 (0.7)
Guest speakers effectively and clearly presented the material.	4.2 (0.8)	4.3 (0.9)	4.2 (0.8)

^a Differences between tracks are statistically discernible at $p < 0.10$. (Mann-Whitney Rank Sum test).
Source: RAND research

Respondents were generally quite sanguine about the overall coherence of the material presented. The average response to the item "I could understand how the material presented was linked to overall Fellowship goals" was 4.2 on a scale of 1 to 5, with 5 representing "Strongly Agree" (see Table 5.4). However, responses were more positive among GTZ respondents, perhaps reflecting the tighter focus on patient safety issues in this track. This difference approached, but did not attain, conventional standards of statistical significance (Mann-Whitney $p = 0.15$). As noted above, a few respondents spoke of the need for "*a more centralized vision*" to focus Fellowship activities.

Respondents Wanted More Opportunities to Interact with Speakers and Other Fellows

Respondents also expressed some concerns about opportunities to interact with their peers. While respondent ratings for most of the items in the survey were 4 or higher on a five-point "Strongly Disagree"–"Strongly Agree" scale, responses to items about opportunities for interaction with other Fellows were lower. Indeed, the average response to the item "I had enough opportunities to interact with other students" was 3.5 (see Table 5.4). Interestingly, there was more variability in responses to this item than most others, with a standard deviation of 1.4. Unlike most other items, for which it was exceedingly rare to see ratings of 3 or lower, 7 of the 31 respondents (23 percent) responded to this item with a 1 or 2. While there was a notable difference in responses from GTE and GTZ respondents, the difference was not statistically discernible ($p = 0.33$).²⁶ Many respondents also wanted more opportunities to interact with course instructors and guest speakers. The average response to a survey item on the subject was 3.6 on a scale of 1 to 5 (see Table 5.4). There were no discernible differences between tracks on this item.

²⁶ In part, the failure to reach statistical discernibility reflects the relatively higher variance in responses to the item, as well as the small sample sizes (see Chapter 2).

**Table 5.4. Fellows' Assessments of Session Management
(Means and standard deviations)**

Survey Item	GTE Fellows	GTZ Fellows	All Fellows
I had enough opportunities to interact with course instructors and guest speakers.	3.5 (1.1)	3.8 (1.0)	3.6 (1.1)
I could understand how the material presented was linked to overall Fellowship goals.	4.1 (0.7)	4.4 (1.0)	4.2 (0.8)
I had enough opportunities to interact with other students.	3.4 (1.1)	3.8 (1.6)	3.5 (1.4)

Source: RAND research

Some respondents considered the fact that there were too few opportunities for peer interaction as a lost opportunity, given the diversity of backgrounds, disciplines, and viewpoints reflected in the Fellowship. One respondent noted that Fellows from each discipline “*have different thought processes.*” Another noted that developing awareness of healthcare as a system requires “*walking a mile in others’ shoes.*” But another respondent was surprised that there was not more opportunity for “*spontaneous interaction*” among Fellows given the Fellowship’s emphasis on peer communication in process improvement. As this respondent put it, the Fellowship “*didn’t always mirror its own philosophy.*”

5.4 Summary and Conclusions

For the most part, respondents’ reactions to the execution of the course were positive. Most found the material presented to be new. However, there were mixed opinions about whether the Fellowship attempted to cover too much material. Many respondents noted that topics were often rushed, given time constraints. But few, if any, could identify material they thought might be dropped in the interest of time. Some respondents also raised concerns about the overall coherence of the Fellowship, suggesting that sessions should be more clearly focused around overarching themes and concepts. In a similar vein, many respondents noted that it was not always clear how course readings related to broader instructional goals. Respondent opinions were also mixed on the Coro skills (e.g., leadership, critical thinking, and group operations). While some respondents found them too simplistic, others found them to be new and important, noting that they seldom receive such training in the course of their other studies.

Opportunities for hands-on experience and the use of concrete examples during the Fellowship were given high ratings by respondents. However, there was some concern that Fellows were not given enough time to prepare for presentations, and that participant groups were not given enough time to complete their tasks and develop strong working relationships. Some respondents also pointed out that student presentations often came out of sequence during sessions, which diminished their utility.

Finally, instructors and guest speakers generally received high marks. However, respondents generally wanted more opportunities to interact with guest speakers and with other Fellows.

6. Training Outputs

Chapter 2 proposed a logic model to identify a set of intermediate outputs that we might observe to determine whether the Fellowship is “on track” toward achieving its ultimate goal of training Fellows to be effective agents of systemic improvement in healthcare organizations. Those intermediate outputs are:

- *Awareness.* Increasing awareness of medical error and patient safety issues.
- *Attitudes.* The development of attitudes relevant to patient safety, including those concerning the balance between individual- and system-level responsibility for errors and the aspiration of “getting to zero.”
- *Acquisition of knowledge and skills.* The development of knowledge and skills that might be used to ameliorate patient safety problems.
- *Propensity to act.* The propensity to act on such knowledge and skills.

The remainder of this chapter examines the extent to which the Fellowship was successful in engendering these intermediate outputs. Evidence comes from the participant surveys, which included both quantitative scaled items and open-ended responses. The analysis was also supplemented by participant focus groups. Details on the survey are provided in Chapter 2 and Appendix A.

6.1 Discernible Increases in Awareness of Patient Safety Problems

The first intermediate output examined was awareness of problems with patient safety. To assess this output, respondents were asked to rate their level of agreement with the statement “Medical error is a significant problem” on a five-point scale, with 1 representing “Strongly disagree” and 5 representing “Strongly agree.” Respondents were asked to rate their agreement with the statement both before the Fellowship and on the day they took the survey.²⁷ As is evident in Table 6.1, self-reported awareness levels were moderate before participation in the Fellowship, but considerably higher afterward. Indeed, the average participant gained 1.1 points on the five-point scale, a statistically discernible increase. Gains in the GTZ track (1.6 points) were discernibly greater than those in the GTE track (0.7 points), a statistically discernible difference ($p = 0.01$).²⁸ However, this difference is explained by the fact that respondents in the GTZ group reported their pre-Fellowship awareness as being lower than that of the GTE group ($p = 0.008$); indeed, both groups reported identical levels of awareness after the Fellowship.

Not surprisingly, respondents reporting more sources of prior exposure to patient safety issues (see Chapter 4) also reported greater prior awareness, although the relationship failed to reach statistical significance. Moreover, respondents with greater prior exposure and awareness reported less dramatic gains in awareness, a statistically discernible relationship.²⁹ The Fellowship’s impact on awareness was also evident in the qualitative data. One respondent noted that learning about the prevalence of medical error was “*a real eye-opener*,” even though the respondent had considerable prior exposure to the issue.

²⁷ Readers should recall that respondents made these judgments retrospectively. The logic of this “retrospective pre/post” design is discussed in Chapter 3.

²⁸ Differences in growth across the tracks were tested by comparing the distribution of post-minus-pre scores for GTZ and GTE using the Mann-Whitney rank sum test. Differences in pre-Fellowship scores were assessed using the same test.

²⁹ Relationships between exposure and prior awareness and between prior awareness and pre/post gains were assessed using contingency tables. The gamma statistic for the relationship between prior exposure and prior awareness was 0.70 ($\chi^2 = 18.5, p = 0.24$). The same statistic for the relationship between prior awareness and pre/post gains was -1.0 ($\chi^2 = 54.4, p < 0.001$).

**Table 6.1. Changes in Fellows' Awareness of Patient Safety Issues
(Means and standard deviations)**

Survey Item	Fellows	Pre-Fellowship	Post-Fellowship	Difference (Post – Pre)
Medical error is a significant problem	All	3.7 (0.9)	4.8 (0.4)	1.1*** (1.0)
	GTE	4.1 (0.8)	4.8 (0.4)	0.7*** (0.9)
	GTZ	3.2 (0.9)	4.8 (0.4)	1.6*** (0.9)
<i>Difference (GTZ – GTE)</i>		-0.9***	0.0	0.9***

*** $p < 0.01$

** $p < 0.01$

* $p < 0.01$

NOTE: p -values are based on Wilcoxon sign rank tests.

Source: RAND research.

6.2 Attitudes Changed in the Expected Direction

The second intermediate output examined was changes in attitudes and values related to patient safety. As noted in Chapter 2, the Fellowship sought to convince Fellows that (1) medical errors usually have systemic causes, (2) a just safety culture precludes reflexive blame of individuals for those errors, and (3) “zero errors” is an attainable goal.

Results from this section of the survey are presented in Table 6.2. Items were scaled so that declines in the level of agreement with the statement are desirable. That is, the list of attitudes or values presented to respondents was contrary to the values and attitudes espoused by the Fellowship. For instance, the first item asked respondents to indicate the extent to which they agreed with the statement “Medical errors are primarily caused by individual carelessness” on a five-point scale, where 1 represented “Strongly disagree” and 5 “Strongly agree.” Given the Fellowship’s emphasis on the systemic causes of medical errors and the importance of blame-free organizational cultures, Fellowship developers clearly hoped that participation in the Fellowship would reduce the level of agreement with this statement. Indeed, the average respondent reported a 1.1-point decline, from 3.3 to 2.2 on a five-point scale ($p < 0.05$).

Analysis of the survey data indicated similar declines in all other items. As with the items discussed above, all declines were statistically discernible. However, the magnitude of the declines was not as high on the remainder of items, ranging from a half-point decline on the item “A team of highly vigilant individuals will generally not produce medical errors” to the 1.1-point decline on the item discussed in the previous paragraph.

In two instances, there were notable differences between tracks. For instance, respondents from GTE showed a larger decline in agreement with the item “Medical errors can be prevented by retraining or firing a few individuals.” However, the difference was not statistically discernible. By contrast, there was a nearly statistically discernible difference by track on the item “A certain number of medical errors is inevitable.” Here, the average decline among GTZ respondents was 1.5, compared with 0.3 for GTE respondents ($p = 0.11$).³⁰ The fact that changes in attitude on this item were greater among GTZ respondents is perhaps not surprising, given that the title of the track was “Getting to Zero.” This change in attitudes about the inevitability of medical errors was also reflected in open-ended survey items and focus groups. Indeed, some of the most commonly cited effects of the Fellowship were to increase (1) awareness of the systemic causes of medical error and (2) acceptance of “zero errors” as an attainable goal.

³⁰ The difference was assessed using a Mann-Whitney rank sum test, where the pre/post difference was the outcome variable and “track” the grouping variable.

**Table 6.2. Changes in Fellows' Attitudes About Patient Safety
(Means and standard deviations)**

Survey Item	Fellows	Pre-Fellowship	Post-Fellowship	Difference (Post – Pre)
Medical errors are primarily caused by individual carelessness.	All	3.3 (1.2)	2.2 (1.2)	-1.1*** (1.4)
	GTE	3.2 (1.2)	2.1 (1.1)	-1.1** (1.3)
	GTZ	3.4 (1.2)	2.3 (1.3)	-1.1* (1.6)
Medical errors can be prevented by retraining or firing a few individuals.	All	2.9 (1.1)	2 (1.2)	-0.9** (1.5)
	GTE	3.0 (1.2)	1.8 (1.0)	-1.2** (1.2)
	GTZ	2.8 (0.9)	2.3 (1.4)	-0.5 (2.0)
A certain number of medical errors is inevitable.	All	3.6 (1.2)	2.8 (1.2)	-0.8** (1.6)
	GTE	3.5 (1.2)	3.2 (1.2)	-0.3 (1.4)
	GTZ	3.8 (1.3)	2.3 (1.2)	-1.5** (1.6)
An atmosphere in which individuals are publicly reprimanded for errors is most likely to reduce error rates.	All	2.3 (1.2)	1.5 (0.9)	-0.8** (1.6)
	GTE	2.4 (1.2)	1.5 (0.8)	-0.9** (1.4)
	GTZ	2.3 (1.4)	1.6 (1.0)	-0.7 (1.6)
It is not worthwhile for hospitals to try to achieve error rates any lower than those seen at similar hospitals.	All	2.3 (1.3)	1.3 (0.6)	-1.0** (1.3)
	GTE	2.3 (1.1)	1.2 (0.4)	-1.1** (1.0)
	GTZ	2.3 (1.5)	1.5 (0.8)	-0.8 (1.7)
A team of highly vigilant individuals will generally not produce medical errors.	All	3.2 (0.9)	2.7 (1.2)	-0.5** (1.3)
	GTE	2.4 (1.2)	1.5 (0.8)	-0.9** (1.0)
	GTZ	2.3 (1.4)	1.5 (1.0)	-0.8 (1.7)

***p < 0.01

** p < 0.05 Wilcoxon sign rank tests

Source: RAND research.

Interestingly, respondents with more sources of prior exposure to patient safety issues (see Chapter 4) reported larger shifts in attitudes. However, the relationship between prior exposure and attitudinal change was statistically discernible only for the item “It is not worthwhile for hospitals to try to achieve error rates any lower than those seen at similar hospitals.” Surprisingly, there was a negative relationship between level of prior education and degree of attitudinal change. However, the relationship was

statistically discernible only for the item “Medical errors are primarily caused by individual carelessness” ($p = 0.03$).³¹ While these findings are difficult to interpret, it does suggest that learning outcomes might be systematically different for various participant subgroups. This issue deserves further exploration in the future, ideally in studies with larger sample sizes.

6.3 Discernible Increase in Knowledge and Skills, but Less Confidence in Ability to Apply Skills

The third intermediate outcome explored in this chapter is changes in self-reported knowledge and skills. First and foremost among the skills the Fellowship sought to teach was “systems” thinking, particularly as manifested in the Toyota Production System.³²

Results from this section of the survey are presented in Table 6.3. Unlike the items discussed in the previous section, these items were scaled so that *increases* were desirable. For instance, one item asked respondents to indicate the extent to which they were familiar with root-cause analysis, both before and after the Fellowship. Here, the average respondent reported an increase of 1.8 points on a five-point scale, a statistically discernible increase ($p < 0.01$).

Respondents reported statistically discernible gains on all other skills included in the survey (see Table 6.3). The largest self-reported gain was in the TPS. This is not surprising given the amount of time and attention devoted to the topic.³³ Generally, the largest increases in self-reported skills were in specialized skills, such as TPS’s “systems” thinking, the “5 Whys,” and root-cause analysis. Smaller gains were reported in skill areas such as observation. In part, this output might represent a failure of the survey instrument to reveal what respondents think of as “observation” as opposed to the term’s general meaning. Relatively small gains were also reported in team-based problem solving and in providing critical feedback to others. In part, this might be due to the fact that respondents rated their pre-Fellowship skills in these areas more favorably than their pre-Fellowship skills in other areas, leaving less room for improvement. Given the qualitative findings reported in Chapter 5, however, it is also possible that these less impressive findings represent difficulties in integrating leadership and critical thinking skills into the patient safety curriculum.

On many skill items, gains reported by GTZ respondents were greater than those reported by GTE respondents. Yet, GTZ respondents often reported lower pre-Fellowship skill levels and, perhaps, had more room for improvement. However, these GTZ-GTE differences were statistically discernible only on the items pertaining to root-cause analysis and careful observation. One possible explanation for the differences in these skill areas is that learning these skills might be more dependent on practical, hands-on experience than are the other skills. As noted earlier, opportunities for hands-on learning were more abundant in the GTZ track.

As noted in Chapter 2, an important outcome of patient safety training should be the ability to apply skills in a variety of practical contexts. While RAND could not directly observe Fellows in such activities, the survey did ask them to rate their own preparation for real-world applications. The results of this analysis are provided in Table 6.4.

³¹ Respondents with graduate degrees—a small proportion of all respondents—actually reported increases in the extent to which they agreed with this item. The relationship between attitudinal change and prior educational attainment was assessed using the Kruskal-Wallis test, with change in the agreement scale as the response variable and a 0/1 dummy variable for graduate degree as the grouping variable.

³² Given that the curriculum was still being finalized as the survey instrument was being designed, there might be skills and knowledge areas that were covered in the Fellowship that are not adequately represented on the survey instrument and in this analysis.

³³ As noted earlier, TPS forms the heart of the Perfecting Patient Care system of the Pittsburgh Regional Healthcare Initiative, from which much of the Fellowship curriculum was derived.

**Table 6.3. Self-Reported Changes in Fellows' Knowledge and Skills
(Means and standard deviations)**

Survey Item	Fellows	Pre-Fellowship	Post-Fellowship	Difference (Post – Pre)
Team-based problem solving	All	3.7 (1.1)	4.3 (0.7)	0.6* (0.8)
	GTE	3.7 (1.2)	4.3 (0.8)	0.6* (0.8)
	GTZ	3.7 (1.1)	4.4 (0.7)	0.7* (0.8)
The “5 Whys”	All	2.1 (1.0)	4.0 (0.8)	1.9* (1.1)
	GTE	2.0 (0.9)	4.1 (0.8)	2.1* (1.0)
	GTZ	2.2 (1.1)	3.0 (0.9)	1.8* (1.2)
Root-cause analysis	All	2.5 (1.1)	4.3 (0.7)	1.8* (1.3)
	GTE	2.6 (1.1)	4.2 (0.8)	1.6* (1.3)
	GTZ	2.3 (1.0)	4.5 (0.7)	2.2* (1.1)
Work design and redesign	All	2.4 (1.1)	4.2 (0.8)	1.8* (1.1)
	GTE	2.6 (1.1)	4.3 (0.8)	1.7* (1.0)
	GTZ	2.2 (1.0)	4.0 (1.0)	1.8* (1.2)
Careful observation	All	3.6 (1.1)	4.4 (0.7)	0.8* (1.0)
	GTE	3.9 (1.0)	4.3 (0.8)	0.4* (0.6)
	GTZ	3.2 (1.1)	4.6 (0.7)	1.4* (1.2)
Giving critical performance feedback to others	All	3.4 (0.9)	4.2 (0.7)	0.8* (0.7)
	GTE	3.7 (1.0)	4.3 (0.7)	0.6* (0.7)
	GTZ	3.1 (0.7)	3.9 (0.7)	0.8* (0.7)
Understanding how systems work	All	2.7 (1.1)	4.2 (0.6)	1.5* (1.0)
	GTE	2.9 (1.2)	4.3 (0.7)	1.4* (1.1)
	GTZ	2.4 (0.8)	4.1 (0.5)	1.2* (0.9)
Toyota Production System	All	1.5 (0.8)	4.2 (0.6)	2.7* (0.9)
	GTE	1.7 (0.9)	4.3 (0.6)	2.6* (1.1)
	GTZ	1.2 (0.6)	4.1 (0.7)	2.9* (0.7)

* p < 0.01 Wilcoxon sign rank tests
Source: RAND research.

Table 6.4. Self-Reported Ability to Apply Skills in Real-World Contexts

Survey Item	Fellows	Mean (Standard deviation)
There are likely to be significant barriers to my ability to use the skills I learned in my professional work.	All	3.0 (1.3)
	GTE	2.8 (1.0)
	GTZ	3.3 (1.6)
The Fellowship provided the skills I will need to overcome barriers I might encounter in putting medical safety skills into practice.	All	3.8 (1.3)
	GTE	3.6 (0.9)
	GTZ	4.1 (0.9)
The Fellowship helped me to identify resources in my (future) workplace that might allow me to address barriers to use of skills.	All	3.9 (1.0)
	GTE	4.1 (0.9)
	GTZ	3.7 (1.1)
I am confident that I could use team-based problem solving skills in my (future) workplace without any further training.	All	3.8 (1.0)
	GTE	4.0 (0.9)
	GTZ	3.4 (1.1)
I am confident that I could use systems thinking (e.g., the “5 Whys”) in my (future) workplace without any further training.	All	3.6 (0.9)
	GTE	3.7 (0.9)
	GTZ	3.4 (1.0)
I am confident that I could use root-cause analysis in my (future) workplace without any further training.	All	3.8 (1.0)
	GTE	3.9 (0.9)
	GTZ	3.6 (1.0)
I would feel comfortable approaching a professional peer in order to discuss an unsafe practice he or she was engaging in.	All	3.8 (1.0)
	GTE	4.0 (0.8)
	GTZ	3.6 (1.3)
I would feel comfortable approaching a manager, administrator, or supervisor in order to discuss an unsafe practice.	All	4.2 (0.7)
	GTE	4.2 (0.6)
	GTZ	4.1 (0.9)

Source: RAND research.

As shown at the top of Table 6.4, respondents were neutral with respect to the statement “There are likely to be significant barriers to my ability to use the skills I learned in my professional work,” with an average rating of 3.0 on a five-point scale, with 1 representing “Strongly disagree” and 5 “Strongly agree.” Fellows in the GTZ track were more likely to anticipate such barriers, perhaps because of the more sustained exposure to nuanced and complex real-life examples in that track. However, this difference was not statistically discernible ($p = 0.32$). Interestingly, Fellows in the GTZ track were also more likely to believe that the Fellowship had provided the skills needed to overcome such barriers. This difference was statistically discernible ($p = 0.08$).

The remainder of the survey items asked respondents to rate their level of comfort in applying specific skills and knowledge areas covered by the Fellowship. For the most part, responses were modest to strong, with the average rating on most items ranging from 3.3 to 4.2 on a five-point scale. The most positive response came to the statement “I would feel comfortable approaching a manager, administrator, or supervisor in order to discuss an unsafe practice,” with an average rating of 4.2 on a five-point scale. With the exception of the item mentioned above, there were no discernible differences between the GTE and GTZ tracks on this set of items. Nor was there a discernible relationship between perceived ability to apply skills and prior exposure to patient safety issues.

Interestingly, respondents’ ratings of their ability to apply skills (see Table 6.4) were generally lower than their ratings of their familiarity with those skills (see Table 6.3) after the Fellowship. For instance, while the average rating for familiarity with root-cause analysis after the Fellowship was 4.3, the average rating for ability to apply it was just 3.8. From this, it appears that Fellows believed that they learned a lot (i.e., experienced gains) but still wondered about adequacy (i.e., the level) of their skills.³⁴ The qualitative data yield similar conclusions. When one focus group participant was asked whether he/she would be ready to implement TPS in a clinical setting, the participant said, *“I feel comfortable that I understand TPS and could be a good advocate. I could suggest things that people should be ready to think about when implementing TPS. But I couldn’t necessarily run the training myself.”*

This same respondent, however, said that he/she *“would know who to bring in”* to help with the training. Another participant echoed this response, stating, *“My greatest strength would be my excitement. They would already have my buy-in, and I would already have contacts. I could at least ready those around for what they were about to experience.”*

6.4 Fellows Generally Report That They Are Likely to Act on Their New Skills

The last intermediate outcome examined was propensity to act on the skills and knowledge presented in the Fellowship. We considered both future training activities and the use of Fellowship skills in real-world contexts. The results of this analysis are summarized in Table 6.5.

All items were placed on a five-point scale, with 1 representing “Not likely at all” and 5 representing “Very likely.” First, we found that respondents were not particularly inclined to engage in further training activities, with an average score of 2.9 on the five-point scale. Interestingly, GTE respondents were more likely to report a willingness to engage in further training than were GTZ respondents, with a statistically discernible difference ($p = 0.01$). However, it is not clear whether this result reflects respondents’ evaluation of the course or the possibility that GTZ respondents perceived their time to be more constrained than did GTE respondents. Regarding this point, GTE respondents said they were more likely than GTZ respondents to “Read a book, report, or other publication on safety science not covered during the Fellowship.” In this case, however, the difference was not statistically discernible.

³⁴ However, such an inference is limited by the fact that the scales used on the two sets of items were different.

Table 6.5. Self-Reported Propensity to Act on Fellowship Skills

Survey Item	Fellows	Mean (Standard deviation)
Take another course in medical safety or safety science	All	2.9 (1.1)
	GTE	3.3 (0.9)
	GTZ	2.3 (1.2)
Read a book, report, or other publication on safety science not covered during the Fellowship	All	4.2 (1.0)
	GTE	4.4 (0.8)
	GTZ	3.8 (1.3)
Talk to a professional colleague about an unsafe practice	All	4.0 (1.0)
	GTE	4.4 (0.7)
	GTZ	3.3 (1.1)
Talk to a supervisor about an unsafe practice	All	4.1 (1.0)
	GTE	4.3 (0.7)
	GTZ	3.7 (1.2)

Source: RAND research.

Responses were relatively high for the propensity to speak with colleagues or to speak with supervisors about unsafe practices, with average ratings for all respondents of 4.0 and 4.1, respectively. Interestingly, GTZ respondents reported being less likely to speak with peers and to speak with supervisors—although the difference for only the former was statistically discernible ($p = 0.007$). As noted in Chapter 3, Fellows in the GTZ track were more likely than GTE Fellows to be students. Thus, it is possible that GTZ respondents are less likely to have a professional peer or supervisor with whom to speak about these matters.

6.5 Generally Positive Overall Evaluations of the Fellowship

Finally, we asked respondents to provide an overall evaluation of the Fellowship. Specifically, we asked whether (1) respondents would recommend the Fellowship to a colleague, (2) the Fellowship was a good use of time, and (3) the Fellowship met expectations. As shown in Table 6.6, these evaluations were quite favorable, with average ratings on each ranging between 4.0 and 4.3 on a five-point scale. There were no statistically discernible differences between satisfaction of respondents in the GTE track and those in the GTZ track.

Table 6.6. Fellows' Overall Evaluation of the Fellowship

Survey Item	Fellows	Mean (Standard deviation)
I would recommend the Fellowship or other medical safety training to a peer, colleague, or supervisor.	All	4.3 (1.0)
	GTE	4.5 (0.6)
	GTZ	4.1 (1.4)
The Fellowship was a good use of my time.	All	4.2 (0.8)
	GTE	4.3 (1.0)
	GTZ	4.2 (1.0)
The Fellowship met my expectations.	All	4.0 (1.0)
	GTE	4.1 (1.0)
	GTZ	3.9 (1.2)

Source: RAND research.

6.6 Summary and Conclusions

This chapter sought to assess the Fellowship's success in engendering a number of intermediate training outcomes, including (1) increased awareness of patient safety issues, (2) changes in attitudes consistent with current thinking in safety science, (3) the acquisition of knowledge and skills, and (4) the propensity to act on those skills. For the most part, the Fellowship succeeded in achieving each of these goals. First, respondents reported discernible increases in the extent to which they perceive patient safety as a significant problem. Second, respondents were more likely to believe that errors stem from system-level factors and to express other attitudes congruent with current thinking in safety science about error causation. Similarly, there were discernible increases in respondents' self-reported knowledge of core safety science concepts and techniques. However, respondents were less certain of their readiness to apply these skills in a real-world context. Finally, with the exception of their not wanting to take more coursework, respondents reported a willingness to act on their newly developed knowledge and skills.

7. Conclusions and Recommendations

The Jewish Healthcare Foundation contracted with the RAND Corporation to provide an in-process evaluation of the summer 2004 pilot training curriculum on patient safety. The purpose of the evaluation was to assess the prospective merit of the pilot training curriculum as a mechanism for training healthcare professionals in the principles and practices of safety science. The evaluation

- described the elements of the program
- described key characteristics of the Fellows
- evaluated the implementation and execution of the curriculum design
- assessed measurable progress toward learning outcomes.

This final chapter summarizes key findings from the evaluation and provides recommendations designed to assist in the ongoing development of the Fellowship.

7.1 Key Findings

The purpose of the evaluation was to inform decisionmaking about program redesign and scale-up. As such, it was important to assess not only the extent to which the Fellowship had its intended impacts but also its design and implementation. An important feature of the Fellowship was that each of two “tracks”—Getting to Excellence and “Getting to Zero”—used slightly different approaches. Thus, wherever possible this report examined differences across the tracks.

Course Design

The analysis of the Fellowship design discussed in Chapter 2 covered both the range of safety science *content* presented in the Fellowship and the *pedagogical approach* used to deliver that content. The core focus of the Fellowship was on systems concepts related to patient safety. As noted in Chapter 2, current thinking on medical error and safety science emphasizes that medical errors are often rooted in the characteristics of healthcare systems, including the flow of materiel, scheduling, and organizational culture. Accordingly, the Fellowship focused on the Toyota Production System, which advocates real-time problem solving to address flaws in healthcare processes. Having identified a patient need that is not currently being met, users of TPS (1) carefully observe current work practices in order to establish the “current condition,” (2) identify problems in current processes, (3) identify root causes of these problems, and (4) devise and test the consequences of a fix designed to address the problems and the root causes.

The Fellowship also sought to supplement core safety science and process improvement concepts with training in critical thinking and leadership/teamwork skills. Training in these skills was derived from materials developed by the Coro Center for Civic Leadership. Here, the training focused on a number of conceptual tools designed to help Fellows to think more clearly about the logical and factual bases of inferences about healthcare processes. The Fellowship also sought to provide tools designed to help Fellows to more effectively structure group processes. The focus on critical-thinking skills can be viewed as an attempt to supplement process-improvement methods’ emphasis on careful observation and the scientific method. Teamwork and leadership skills, in turn, may be viewed as a response to the fact that most healthcare systems involve multiple—and often autonomous—actors from a wide variety of professional and disciplinary backgrounds.

Given the Fellowship’s focus on practical skills, its *pedagogical approach* featured exposure to a variety of practitioners in local clinical settings. The GTZ track focused exclusively on patient safety issues, while the GTE track also included sessions on palliative care and public health. Also, GTZ Fellows performed a series of hands-on tasks designed to give them practical experience in process improvement.

Fellowship Participants

Fellows were generally inexperienced and still in the midst of their professional training. Thus, the summer 2004 program appears to have achieved its goal of targeting developing healthcare professionals. However, Fellows in GTE had more prior training, professional experience, and first-hand experience with patient safety issues than those in GTZ. Moreover, Fellows came from a wide variety of disciplinary backgrounds, suggesting that the summer 2004 program also met its goal of attracting a professionally diverse group. This report also provided information on how Fellows learned about the Fellowship, which might prove useful in planning for future recruitment efforts. Given that most Fellows were students before joining the Fellowship, it is not surprising to find that most learned about the Fellowship through their teachers.

Course Implementation

For the most part, respondents' reactions to the execution of the course were positive. Most found that the material presented to them was new. However, there were mixed opinions about whether the Fellowship attempted to cover too much material. Many respondents noted that topics were often rushed, while few could identify material that they thought might be dropped in the interest of time. Some respondents also raised concerns about the overall coherence of the Fellowship, suggesting that sessions should be more focused around overarching themes and concepts. In the same vein, many respondents noted that it was not always clear how course readings related to broader instructional goals. Respondents' opinions were also mixed on the Coro skills (e.g., leadership, critical thinking, and group operations). While some respondents found them to be too simplistic, others found them to be new and important, noting that they seldom receive such training in their other studies.

Opportunities for hands-on experience and the use of concrete examples during the Fellowship were given high ratings by respondents. However, there was some concern that Fellows were not given enough time to prepare for presentations, and that participant groups were not given enough time to complete their tasks and develop strong working relationships. Some respondents also pointed out that presentations often came at the end of the session, at which point the presentations' utility was diminished.

Finally, instructors and guest speakers generally received high marks. However, respondents generally wanted more opportunities to interact with guest speakers and with other Fellows.

Learning Outcomes

As noted in Chapter 2, the ultimate goal of the Fellowship is to engender in its participants the capacities and skills to lead changes in healthcare organizations that enhance patient safety. Given that most Fellows are still in training, however, it is likely that these impacts will occur well beyond the period included in this evaluation. Thus, the evaluation focused on a number of intermediate training outcomes, including enhanced awareness of patient safety issues, the development of attitudes consistent with the system-related nature of medical error that is common in current safety science thinking, and the development of specific knowledge and skills related to patient safety.

For the most part, the Fellowship succeeded in raising awareness of patient safety issues, engendering attitudes consistent with systems-oriented conceptions of medical error, and inculcating useful knowledge and skills. First, respondents reported discernible increases in the extent to which they perceive patient safety as a significant problem. Second, post-Fellowship, respondents were more likely to believe that errors stem from system-level factors and expressed other attitudes congruent with current thinking in safety science about error causation. Similarly, there were discernible increases in respondents' self-reported knowledge of, and skills in, core safety science concepts and techniques. However respondents were less certain of their readiness to apply these skills in real-world clinical settings. Finally, with the

exception of their not wanting to take more coursework, respondents reported a willingness to act on their newly developed knowledge and skills.

Given that we were unable to directly observe the Fellows' ability to put these skills into practice, we were unable to assess whether the Fellowship was successful in its ultimate goal of generating a cadre of professionals capable of effecting systemic change in healthcare institutions. Thus, the findings presented here must be regarded as suggestive, but not conclusive.

7.2 Recommendations

Overall, this evaluation report suggests that the Fellowship represents a promising approach to training in medical safety science concepts and practices. Nevertheless, the report identified a number of remaining issues that program developers would do well to address as they continue to refine the Fellowship's design:

- The tradeoff between breadth and depth in the range of material covered
- The extent to which the Fellowship is well-organized around a set of clear and coherent instructional themes and goals
- The extent to which course readings and hands-on exercises are integrated into a larger instructional vision
- The appropriateness and use of training in critical thinking
- Concern about ability to apply skills in practice
- Time for hands-on activities
- Time for interaction with each other and guest speakers.

Further Design and Planning Activities

This section presents two recommendations designed to facilitate an assessment of the current design of the Fellowship and to guide a process for identifying strategies for redesign.

Recommendation 1: Use a logic modeling exercise to strengthen coherence of program design. The concerns raised by Fellows about the overall coherence of the program suggest the need to reconsider the goals of the Fellowship and how its specific elements relate to those goals. Accordingly, JHF might consider engaging in a logic modeling exercise. Chapter 2 provided a draft logic model developed by RAND to guide this evaluation. However, JHF and its stakeholders might consider either elaborating upon this logic model or starting afresh with a new one. The logic modeling exercise might include some or all of the following steps:

- *Develop a list of typical use contexts and scenarios in which Fellows might eventually work.* As noted in Chapter 2, the Fellowship seeks to develop in Fellows the capability to induce system changes in real-world healthcare contexts. Thus, the logic modeling exercise might begin by identifying the range of particular healthcare contexts in which alumni might reasonably be expected to practice, along with the key reform tasks that they might be expected to perform. For instance, while diagnosing problems in current care processes is probably a common feature in all such contexts, it is likely that the ways in which "symptoms" present themselves might vary considerably across contexts. For instance, chronic-care contexts are more likely than acute-care contexts to involve a multitude of organizations and professions (see, e.g., Institute of Medicine, 2001). The list of typical use contexts/scenarios might be developed by (1) consulting with practitioners and/or (2) reviewing the literature on healthcare process improvement and organizational change.

- *Identify capabilities that cut across the use contexts/scenarios.* Next, JHF might seek to identify the capabilities required to function effectively in each of these contexts.³⁵ Here, it would be important to look for capabilities that apply across a wide range of scenarios. Once again, this process could be informed by consultation with practitioners and review of the literature.
- *Map training activities against capabilities.* The next step in the process might be to consider the extent to which elements of the current Fellowship help generate the identified capabilities and determine whether some capabilities require more attention in the curriculum. The relationships between training activities and desired capabilities might be represented by an arrow diagram similar to that in Figure 2.1. This step might be guided by consultation with practitioners, as well as review of the literature.
- *Evaluate the whole design for completeness and coherence.* Having mapped the Fellowship’s elements and its goals, the next step might be to assess the design for completeness—does it address all of the capacities and goals it should? Similarly, the design could be assessed for coherence. This assessment might include (1) the extent to which the design possesses a reasonably small number of core ideas and focusing concepts, (2) the degree to which the various elements of the program support each other, and (3) the degree of duplication in the design. It is important to note that duplication is sometimes desirable, especially when core concepts are revisited several times from different perspectives.

Recommendation 2: Reconsider the use of readings and exercises. As noted in Chapter 2, the Fellowship has a strong orientation toward case-based learning, especially in the GTZ track. This is appropriate given the complexity and variability in the skills it seeks to teach. Given the concerns raised in this report about tradeoffs between depth and breadth of coverage, it makes sense to review each of the exercises to ensure that they are sufficiently imbued with core concepts and skills and that they are rich enough to provide students with the experience of tackling complex problems without overwhelming them. Indeed, survey and focus-group respondents often noted that they had too little time to learn from many of the exercises. Thus, Fellowship developers might look for opportunities to reduce the number of exercises, retaining only the most valuable ones.

Fellowship developers might also consider developing richer documentation for cases and exercises that might provide “scaffolding” for future instructors. This will be particularly important should JHF consider scale-up of the Fellowship to other sites and locations.

Improving the Evaluability of the Fellowship

As noted at several points in this report, the findings of this evaluation should be regarded as suggestive, but not conclusive. This caution stems from the fact that (1) we were unable to directly observe the ability of Fellowship alumni to function in a real-world healthcare context; (2) practical considerations forced us to rely mainly on self-reported accounts of changes in awareness, attitudes, and skills related to patient safety; and (c) the evaluation design could not make use of a no-treatment comparison group. Thus, the remaining recommendations provide suggestions to ensure that future implementations of the Fellowship are capable of yielding stronger evaluative inferences.

Recommendation 3: Track alumni over time and do follow-up surveys. First, JHF should consider developing a formal and rigorous system for tracking and surveying Fellowship alumni over time. Such a survey might include questions on the following:

- Career paths, including information about the institutional environments in which alumni work
- The extent to which alumni have been involved in activities to improve patient safety

³⁵ This discussion borrows from the concept of capabilities-based planning. See, e.g., Davis (2002) for an introduction to this concept.

- The extent to which the skills and concepts taught in the Fellowship have been adequate given the institutional contexts in which they have worked
- Which concepts and skills they have retained over time and have found most useful in practice.

Tracking alumni can be challenging, especially in healthcare where significant numbers might move to other communities. For this reason, it would be useful for the tracking system to include names of family members and professional colleagues who might be able to assist JHF in locating alumni who have moved.

Recommendation 4: Develop more-structured ways to evaluate hands-on exercises. The reliance on self-reported outcomes in this evaluation could be reduced if JHF were to develop more-rigorous ways of evaluating participants' projects and other work. As noted in Chapter 2, Fellows worked on a number of hands-on projects, culminating in a group project presented to other Fellows and members of the local healthcare community. These exercises would be more useful for evaluation purposes if they were graded by a jury of instructors and local experts against a well-specified rating rubric. Ideally, such a process would be applied to a piece of work completed by Fellows at the beginning and end of the Fellowship to provide a clear baseline against which to assess growth over time.

Recommendation 5: Consider more-careful structuring of differences between tracks. The existence of two Fellowship tracks provides an opportunity to determine how variations in content and pedagogy affect outcomes. Such knowledge, in turn, can be useful in guiding improvements to the Fellowship. Cross-track comparisons, however, are most useful if variations between the tracks are planned to support rigorous comparisons. During the summer 2004 program, the GTE and GTZ tracks varied both in content and pedagogy, making it difficult to assess the independent impact of either. In the future, it would be desirable to plan such variations in a way that allows for clearer inferences.

Finally, after the Fellowship design has been honed and finalized, JHF should plan for a rigorous impact assessment designed to yield a stronger estimate of program impact. Generally, the strong impact assessments involve random assignment to a treatment and a control condition to ensure that any observed differences were not actually caused by non-program factors. Randomization, as is well known, can be difficult to implement. First, those selected into the control group sometimes seek out comparable training from other sources. Second, it is often difficult to track control group participants over time. A delayed treatment design—where those selected into the control group are promised the intervention after a reasonable wait—can help address some of these difficulties. Similarly, there are a number of quasi-experimental designs capable of yielding reasonably strong causal inferences (see, e.g., Shadish, Cook, & Campbell, 2002). Nonetheless, it would be wise to hold off on such a rigorous—and likely expensive—evaluation until the program has been developed and refined further and until it can recruit enough subjects to yield inferences with a considerable degree of statistical power.

7.3 Conclusion

Preventable medical errors kill somewhere between 44,000 and 98,000 people in the United States each year, with some studies placing the number as high as 195,000. The JHF-Coro Fellowship represents an early attempt to develop and implement a training curriculum in medical safety science for developing healthcare professionals. Given the newness of the program and the evaluation's inability to directly observe linkages to patient outcomes, this report was designed to assist JHF in developing and refining the program. The results reported here suggest that the approach holds promise and is worthy of further effort. However, conclusions about its ultimate effect on patient outcomes await more sustained implementation and further evaluation research.

Appendix A

SURVEY INSTRUMENT

This survey is part of an independent evaluation of the Health Sciences Fellowship conducted by the RAND Corporation. The survey, given to all members of the program, will be used to help RAND advise JHF and Coro on how the program might be improved. Your participation is completely voluntary, and your answers will remain confidential. In order to protect your confidentiality, PLEASE DO NOT WRITE YOUR NAME ON THIS SHEET. As further protection to confidentiality, responses will be reported at the aggregate level only.

About You

We'd like to learn a little bit about your prior exposure to medical safety issues and skills and about your future career plans.

- 1) Before taking this course, what was your primary vocation? (*Check all that apply*)
 - Student
 - Practicing nurse or physician (including interns, residents, etc)
 - Other practicing medical professional
 - Other

- 2) What do you expect your primary vocation to be after completing your education? (*Check all that apply*)
 - Student
 - Practicing nurse or physician (including interns, residents, etc.)
 - Administration, research, or public policy
 - Other

- 3) Prior to being a JHF/Coro Fellow, what was your highest degree level?
 - Baccalaureate degree
 - Post-baccalaureate degree (including professional degrees)
 - Other

- 4) Prior to being a JHF/Coro Fellow, what exposure did you have to medical safety/error issues? (*Check all that apply*)
 - Read or heard about it in the news
 - Read a book, government report, or journal article
 - Found material on the internet
 - Took a course that covered medical safety
 - I or someone I know has been the victim of a medical error
 - Observed troubling practices by healthcare workers in medical settings

5) How did you learn about the Fellowship program? (*Check all that apply*)

- Web site
- Heard about it from a teacher, professor, or other instructor
- Heard about it from a past JHF/Coro Fellow
- Heard about it from a JHF or Coro recruiter
- Heard about it from a peer or professional colleague
- Other (*please elaborate below*)

What You Learned From This Course

6) Please indicate the extent to which you disagree or agree with the following statements. Please note that we are interested in your answers for both *before* this course and for *today*.

	<i>Before this course</i>					<i>Today</i>				
	Strongly disagree		↔		Strongly agree	Strongly disagree		↔		Strongly agree
a) Medical error is a significant problem.	1	2	3	4	5	1	2	3	4	5
b) Medical errors are primarily caused by individual carelessness.	1	2	3	4	5	1	2	3	4	5
c) Medical errors can generally be prevented by retraining or firing a few individuals.	1	2	3	4	5	1	2	3	4	5
d) A certain number of medical errors are simply inevitable.	1	2	3	4	5	1	2	3	4	5
e) An atmosphere in which individuals are publicly reprimanded for errors is most likely to reduce error rates.	1	2	3	4	5	1	2	3	4	5
f) It is generally not worthwhile for hospitals to try to achieve error rates any lower than those seen at similar hospitals.	1	2	3	4	5	1	2	3	4	5
g) A team of highly vigilant, error-free individuals will generally not produce medical errors.	1	2	3	4	5	1	2	3	4	5

7) Please describe any other ways in which the JHF/Coro Fellowship program changed your understandings of and attitudes about medical safety.

8) Please answer the following questions about the degree to which you were/are *familiar* with the following skills and concepts.

	<i>Before this course</i>					<i>Now</i>				
	Not at all familiar		↔		Very familiar	Not at all familiar		↔		Very familiar
1) Team-based problem-solving	1	2	3	4	5	1	2	3	4	5
2) The "5 Whys"	1	2	3	4	5	1	2	3	4	5
3) Root cause analysis	1	2	3	4	5	1	2	3	4	5
4) Work design and redesign	1	2	3	4	5	1	2	3	4	5
5) Careful observation	1	2	3	4	5	1	2	3	4	5
6) Giving critical performance feedback to others	1	2	3	4	5	1	2	3	4	5
6) Understanding how systems work	1	2	3	4	5	1	2	3	4	5

7) The Toyota Production System 1 2 3 4 5 1 2 3 4 5

9) Please describe any other new skills or concepts you learned during the Fellowship.

10) Please indicate the extent to which you agree or disagree with the following statements.

	Strongly disagree		←→		Strongly agree
a) There are likely to be significant barriers to my ability to use the skills I learned in my professional work.	1	2	3	4	5
b) The Fellowship provided the skills I will need to overcome barriers I might encounter in putting medical safety skills into practice.	1	2	3	4	5
c) The Fellowship helped me to identify resources in my (future) workplace that might allow me to address barriers to use of skills.	1	2	3	4	5
d) I am confident that I could use team-based problem-solving skills in my (future) workplace without any further training.	1	2	3	4	5
e) I am confident that I could use systems thinking (e.g., the “5 Whys”) in my (future) workplace without any further training.	1	2	3	4	5
f) I am confident that I could use root cause analysis in my (future) workplace without any further training.	1	2	3	4	5
g) I would feel comfortable approaching a professional peer in order to discuss an unsafe practice he or she was engaging in.	1	2	3	4	5
h) I would feel comfortable approaching a manager, administrator, or supervisor in order to discuss an unsafe practice.	1	2	3	4	5
i) The Fellowship presented material that was new to me.	1	2	3	4	5

11) Please describe any specific barriers to use of the skills you learned in your (future) workplace you expect to find.

12) Please describe any specific resources in your (future) workplace that might support you in the use of the skills developed during the Fellowship.

13) Please rate the likelihood that you will engage in the following activities during the next year.

	Not at all likely		←→		Very likely
a) Take another course in medical safety or safety science.	1	2	3	4	5
b) Recommend the Fellowship or other medical safety training to a peer, colleague, or supervisor.	1	2	3	4	5
c) Read a book, report or other publication on safety science not covered during the Fellowship.	1	2	3	4	5
d) Talk to a professional colleague about an unsafe practice.	1	2	3	4	5
e) Talk to a supervisor about an unsafe practice.	1	2	3	4	5

How This Course Was Taught

14) Please indicate the extent to which you agree or disagree with the following statements about how Fellowship sessions were conducted.

	Strongly disagree		←→		Strongly agree
a) Sessions presented enough concrete examples to provide a thorough understanding of key concepts.	1	2	3	4	5
b) I could understand how the material presented was linked to overall Fellowship goals.	1	2	3	4	5
c) Fellowship staff were knowledgeable about patient safety.	1	2	3	4	5
d) Guest presenters were knowledgeable about patient safety.	1	2	3	4	5
e) Fellowship staff effectively and clearly presented the material.	1	2	3	4	5
f) Guest speakers effectively and clearly presented the material.	1	2	3	4	5
g) I had enough opportunities to interact with other students.	1	2	3	4	5
h) Other students were well motivated.	1	2	3	4	5
i) Course readings were clearly integrated into the sessions.	1	2	3	4	5
j) Course readings were too difficult.	1	2	3	4	5
k) Sessions provided adequate opportunity for me to learn how to apply the skills we were expected to master.	1	2	3	4	5
l) I had enough opportunities to interact with course instructors and guest speakers.	1	2	3	4	5
m) I had enough opportunities to interact other JHF/Coro Fellows.					
n) The locations in which the sessions were held were conducive to learning.	1	2	3	4	5

15) Do you have any suggestions for how the material might be better presented next time? Please elaborate.

Overall Evaluation

16) Please indicate the extent to which you agree or disagree with the following statement:

	Strongly disagree		←→		Strongly agree
a) The Fellowship was a good use of my time.	1	2	3	4	5
b) The Fellowship met my expectations.	1	2	3	4	5

17) What were the primary strengths of the Fellowship?

18) What aspects/elements of the Fellowship are in most need of improvement?

Appendix B

FOCUS GROUP PROTOCOL

Introduction

This focus group is part of an independent evaluation of the JHF Health Sciences Fellowship conducted by the RAND Corporation. Your participation is completely voluntary, and your answers will remain confidential. At any time you may ask to stop the focus group or choose not to answer any questions. This focus group is part of an effort to improve the program. Accordingly, your candor is appreciated.

You should know that your name will not be used in any communications about this data collection with JHF or anyone outside of the research team. References in briefings and reports will be to “students” in general, not any specific individual student. Moreover, any identifying characteristics (e.g., age, educational background) will be omitted from all such communications. Also, I ask that each of you also protect the confidentiality of others in the group. Please do not repeat anything that is said here in a way that is attributable to particular people. And during the discussion, do not use the name or other identifying information of anyone as you talk about them. And please do not provide details about a specific medical error you or a colleague was involved in.

And finally, we're tape-recording the session because we don't want to miss any of your comments. If anyone is uncomfortable with the idea of being taped, just say so and we won't use it. Again, remember that we will not be able to connect the taped information with your name or anything that identifies you.

Questions

- 1) How aware of patient safety issues were you prior to being a JHF/Coro Fellow? How did you know about the issue?

- 2) What motivated you to join the Fellowship program? What did you hope to learn?

- 3) How do you expect the Fellowship to relate to your career/intended career?

- 4) Has the Fellowship changed the way you view medical errors? If so, how?

- 5) What are the most useful concepts and tools you have learned/acquired in the Fellowship program? Explain.

- 6) Do you expect any challenges in applying these concepts and tools to your professional practice/future education? If so, has the Fellowship provided you with strategies for addressing those challenges?
- 7) After the Fellowship, do you expect to take any other steps to increase your awareness of medical safety? Your skills set?
- 8) What are the primary strengths and weaknesses in the way the material has been presented?
- 9) Are there any skills, concepts, or topics missing from the Fellowship curriculum? Explain.
- 10) What is your overall assessment of the Fellowship program?
- 11) Aside from what you have already said, are there things that might be done to improve the Fellowship for future cohorts?

REFERENCES

- Bates, D. W., N. Spell, D. J. Cullen, et al. (1997). "The costs of adverse drug events in hospitalized patients. Adverse Drug Events Prevention Study Group," *JAMA*, Vol. 277, No. 4, Jan 22-29, pp. 307-311.
- Chen, Heuy-tsyh. (1990). *Theory-Driven Evaluations*. Newbury Park, CA: Sage.
- Classen, D. C., S. L. Pestotnik, R. S. Evans, et al. (1997). "Adverse drug events in hospitalized patients. Excess length of stay, extra costs, and attributable mortality," *JAMA*, Vol. 277, No. 4, Jan 22-29, pp. 301-306.
- Davis, P. (2002). *Analytical Architecture for Capabilities Based Planning, Mission-System Analysis, and Transformation*. MR-1513-OSD. Santa Monica, CA: RAND Corporation.
- Harvard Business School. (2000). *Deaconess-Glover Hospital*. Harvard Business School Case 9-601-022.
- Health Grades, *Health Grades Quality Study: Patient Safety in American Hospitals*, 2004.
- Institute of Medicine. (2001). *Crossing the Quality Chasm: A New Health System for the 21st Century*, Washington, D.C.: National Academy Press.
- Institute of Medicine. (2000). *To Err Is Human: Building a Safer Health System*, Washington, D.C.: National Academy Press.
- Krueger, N., M. Reilly, and A. Carsrud. (2000). "Competing models of entrepreneurial intentions," *Journal of Business Venturing*, Vol. 15, pp. 411-432.
- Nelson, C. (2004). *Cleaning Up Care: Eliminating Errors and Waste*. Technical paper prepared for the Highmark Healthcare Cost Summit, Pittsburgh, September 23, 2004.
- Oakeshott, M. (1962). *Rationalism in Politics and Other Essays*. New York: Basic Books.
- Pratt, C., W. McGuigan, A. Katzev. (2000). "Measurement program outcomes: Using retrospective pretest methodology," *American Journal of Evaluation*, Vol. 21: 341-349.
- Pressman, J., and A. Wildavsky. (1984). *Implementation: How Great Expectations in Washington are Dashed in Oakland: Or, Why It's Amazing that Federal Programs Work At All, This Being a Saga of the Economic Development Administration as Told by Two Sympathetic Observers Who Seek to Build Morals on a Foundation of Ruined Hopes*, Berkeley: University of California Press.
- Reason, J. (1997). *Managing the Risks of Organizational Accidents*, Burlington, VT: Ashgate.
- Rossi, P., and Heuy-tsyh Chen, eds. (1992). *Using Theory to Improve Program and Policy Evaluations*. New York: Greenwood Press.
- Runciman, W. B., A. F. Merry, and F. Tito. (2003). "Error, blame, and the law in health care—an antipodean perspective," *Ann Intern Med*, Vol. 138, No. 12, Jun 17, pp. 974-979.
- Scriven, M. S. (1991). *Evaluation Thesaurus*. Thousand Oaks, CA: Sage.
- Shadish, W. T. Cook, and D. Campbell. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. New York: Houghton-Mifflin.
- Shannon, R. (2003). "The anatomy of a medical error," Pittsburgh Regional Health Initiative White Paper, Pittsburgh: PRHI.
- Steven Spear & H. Kent Bowen, "Decoding the DNA of the Toyota Production System", *Harvard Business Review*, Sept-Oct, 1999, p. 97-111

- Spillane, J. B., Reiser, B.J.,and T. Reimer. (2002). "Policy implementation and cognition: reframing and refocusing implementation research," *Review of Educational Research* Vol. 72 No. 3, Fall, pp. 387-431.
- Stake, R. (2000). "Program evaluation, particularly responsive evaluation," in D. L. Stufflebeam, G. F. Madaus, and T. Kellaghan, eds., *Evaluation Models: Viewpoints on Educational and Human Services Evaluation*. Boston: Kluwer.
- Stufflebeam, D. (2004). "CIPP checklist," Online at <http://www.evaluation.wmich.edu>.
- Stufflebeam, D. (2000). "The methodology of metaevaluation," in D. L. Stufflebeam, G. F. Madaus, and T. Kellaghan, eds., *Evaluation Models: Viewpoints on Educational and Human Services Evaluation*. Boston: Kluwer.
- Stufflebeam, D., and Webster, W. (1988). "Evaluation as an administrative function," in N. Boyan, ed., *Handbook of Educational Administration*, White Plains, NY: Longman, pp. 569-601.