

Economic
Models of **COLORECTAL**
CANCER SCREENING
in Average-Risk Adults

Workshop Summary

INSTITUTE OF MEDICINE
OF THE NATIONAL ACADEMIES

Economic Models of Colorectal Cancer Screening in Average-Risk Adults

Workshop Summary

National Cancer Policy Board

Board on Science, Technology, and Economic Policy
Policy and Global Affairs Division

Michael Pignone, Louise Russell and Judith Wagner, *Editors*

INSTITUTE OF MEDICINE *AND*
NATIONAL RESEARCH COUNCIL
OF THE NATIONAL ACADEMIES

THE NATIONAL ACADEMIES PRESS
Washington, D.C.
www.nap.edu

THE NATIONAL ACADEMIES PRESS 500 Fifth Street, N.W. Washington, DC 20001

NOTICE: The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine.

This study was supported by Contract No. N01-OD-4-2139 between the National Academy of Sciences and the National Cancer Institute. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the organizations or agencies that provided support for the project.

International Standard Book Number 0-309-09539-5

Additional copies of this report are available from the National Academies Press, 500 Fifth Street, N.W., Lockbox 285, Washington, DC 20055; (800) 624-6242 or (202) 334-3313 (in the Washington metropolitan area); Internet, <http://www.nap.edu>.

For more information about the Institute of Medicine, visit the IOM home page at: **www.iom.edu**.

Copyright 2005 by the National Academy of Sciences. All rights reserved.

Printed in the United States of America.

THE NATIONAL ACADEMIES

Advisers to the Nation on Science, Engineering, and Medicine

The **National Academy of Sciences** is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Bruce M. Alberts is president of the National Academy of Sciences.

The **National Academy of Engineering** was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. Wm. A. Wulf is president of the National Academy of Engineering.

The **Institute of Medicine** was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Harvey V. Fineberg is president of the Institute of Medicine.

The **National Research Council** was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Bruce M. Alberts and Dr. Wm. A. Wulf are chair and vice chair, respectively, of the National Research Council.

www.national-academies.org

NATIONAL CANCER POLICY BOARD

JOSEPH SIMONE (*Chair*), Simone Consulting, Dunwoody, GA
ELLEN STOVALL (*Vice Chair*), Executive Director, National Coalition for Cancer Survivorship, Silver Spring, MD
BRUCE W. STILLMAN (*Vice Chair*), Director, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
JILL BARGONETTI, Associate Professor, Department of Biological Sciences, Hunter College, New York, NY
TIMOTHY EBERLEIN, Bixby Professor and Chairman, Department of Surgery, Washington University School of Medicine, St. Louis, MO
KATHY GIUSTI, President, Multiple Myeloma Research Foundation, New Canaan, CT
KAREN HERSEY, Senior Counsel, Massachusetts Institute of Technology, Cambridge, MA
JIMMIE C. HOLLAND, Chair, Department of Psychiatry and Behavioral Sciences, Memorial Sloan-Kettering Cancer Center, New York, NY
WILLIAM KAELIN, Professor, Harvard Medical School, Boston, MA
WILLIAM MCGUIRE, Chief Executive Officer, UnitedHealth Group, Minnetonka, Minnesota
JOHN MENDELSON, President, M. D. Anderson Cancer Center, University of Texas, Houston, TX
KATHLEEN HARDIN MOONEY, Professor, University of Utah College of Nursing, Salt Lake City, UT
PATRICIA NOLAN, Director, Rhode Island Department of Health, Providence, RI
DAVID PARKINSON, Oncology Therapeutic Area Head, Amgen, Inc. Thousand Oaks, CA
JOHN POTTER, Senior Vice President and Director, Cancer Prevention Research Program, Fred Hutchinson Cancer Research Center, Seattle, WA
LOUISE RUSSELL, Professor, Rutgers University, New Brunswick, NJ
THOMAS J. SMITH, Professor, Virginia Commonwealth University, Richmond, VA
ROBERT C. YOUNG, Past-President, American Cancer Society and President, the Fox Chase Cancer Center, Philadelphia, PA

Study Staff

JUDITH L. WAGNER, Scholar in Residence, Institute of Medicine
ELIZABETH J. BROWN, Research Assistant
ANIKE L. JOHNSON, Administrative Assistant
MARY ANN F. PRYOR, Project Assistant

LOUISE B. RUSSELL, Research Professor, Institute for Health, Health Care Policy and Aging Research, Rutgers University, Workshop Chair
MICHAEL P. PIGNONE, Assistant Professor of Medicine, University of North Carolina

REVIEWERS

All presenters at the workshop have reviewed and approved their respective sections of this report for accuracy. In addition, this workshop summary has been reviewed in draft form by independent reviewers chosen for their diverse perspectives and technical expertise, in accordance with procedures approved by the National Research Council's Report Review Committee. The purpose of this independent review is to provide candid and critical comments that will assist the Institute of Medicine (IOM) in making the published workshop summary as sound as possible and to ensure that the workshop summary meets institutional standards. The review comments and draft manuscript remain confidential to protect the integrity of the deliberative process.

The National Cancer Policy Board and IOM thank the following individuals for their participation in the review process:

John Inadomi, University of Michigan Medical Center and VA Ann Arbor Health Care System

David Lieberman, Oregon Health Sciences University

Diana Petitti, Kaiser Permanente of Southern California

David Ransohof, University of North Carolina at Chapel Hill

The review of the report was overseen by **Clyde J. Behney, Deputy Executive Director, IOM**, who was responsible for making certain that an independent examination of this report was carried out in accordance with institutional procedures and that all review comments were carefully considered. Responsibility for the final content of this report rests entirely with the editors and the institution.

TABLE OF CONTENTS

INTRODUCTION AND BACKGROUND	1
THE COLLABORATIVE MODELING EXERCISE	2
Origin of the Exercise	2
General Approach	3
Results	12
Lessons Learned from the Exercise	19
MAJOR CHALLENGES TO MODELING THE COST-EFFECTIVENESS OF CRC SCREENING	22
Uncertainty	23
Modeling Reality or an Ideal World?	27
How Complex Should Models Be?	28
NEXT STEPS	29
REFERENCES	31
APPENDIXES	
A Workshop Agenda: Economic Models of Colorectal Cancer Screening in Average-Risk Adults	35
B Workshop Participants	37
C Workshop Speaker and Staff Biographies	39
D Overview Of Harvard Model: Karen M. Kuntz, ScD.	47
E Description of the Laudabaum Colorectal Cancer Screening Model: Uri Ladabaum, M.D., M.S.	61
F MISCAN-colon: An Overview: Marjolein van Ballegooijen, Iris Vogelaar, Rob Boer, Franka Loeve, Ann Zauber, Gerrit van Oortmarssen, and Dik Habbema	73
G The Vanderbilt Colorectal Cancer Model R.M. Ness, R.W. Klein, R.S. Dittus	84
H Overview of the Vijan: Colorectal Cancer Screening Model: Sandeep Vijan, M.D., M.S.	105
I Cost-Effectiveness Analyses of Colorectal Cancer Screening: Results from a Pre-conference Modeling Exercise: Michael Pignone, M.D., M.P.H.	118
J Recent Findings on Test Performance: Brian P. Mulhall, M.D., M.P.H.	154
K Cost Issues in Cost Effectiveness Modeling of Colorectal Cancer Screening: Martin L. Brown, Ph.D.	189
L Current Evidence on Compliance: Sally W. Vernon, M.A., Ph.D.	208
M Recent Trends in Follow-up Surveillance in Medicare Beneficiaries: Todd Anderson, M.S	220
N Preliminary Results from CDC's Estimate of the National Capacity for Colorectal Cancer Screening and Follow-Up: Laura C. Seeff, M.D.	232
O Colorectal Cancer Surveillance Testing After Polypectomy: Deborah Schrag, M.D., M.P.H.	241

P	Natural History of Colorectal Adenomas and Cancer: T. R. Levin, M.D.	263
Q	CISNET—Cancer Intervention and Surveillance Eric J. (Rocky) Feuer, Ph.D.	285

TABLES AND FIGURES

TABLES

1	Summary of Standardized Assumptions: Pre-workshop Exercise	
2	Basic Assumptions: Pre-workshop Exercise	
3	Summary of Original Assumptions	
4	Settings for Experimental Runs: Pre-workshop Exercise	
5	Predicted Years of Life Lived and Lifetime CRC Costs per Capita, No Screening: Original Model Assumptions	
6	Incremental Cost-Effectiveness Ratios of Five CRC Screening Strategies: Original Assumptions	
7	Incremental Cost-Effectiveness Ratios of Five CRC Screening Strategies: Standardized Assumptions	
8	Effect of Standardizing Individual Assumption Groups on Variation Across Models: Ratio of Highest Estimate to Lowest Estimate	
9	Effect of Standardizing Specific Assumption Groups on Variation in Cost-Effectiveness Ratios Across Models-Ratio of Highest Estimate to Lowest Estimate	Ef-
10	Impact of Excluding Non-Adenomatous Polyps from the Vanderbilt Model	
11	Impact of Detecting Non-Adenomas on Incremental Cost-Effectiveness Ratios	

FIGURES

1a	Years of life gained from screening: original assumptions
1b	Lifetime costs of screening: original assumptions
2	Cost-effectiveness of screening: original assumptions
3a	Years of life gained from screening: standardized assumptions
3b	Lifetime costs of screening: standardized assumptions
4	Cost-effectiveness of screening: standardized assumptions

Workshop Summary

INTRODUCTION AND BACKGROUND

Colorectal cancer (CRC) is the second leading cause of death from cancer in the United States (Edwards et al., 2002). Research has shown that screening adults for early cancers or their precursor lesions, followed by appropriate therapy and continued surveillance, can reduce CRC incidence and mortality (Curry, 2003). A general consensus has emerged that periodic screening of adults over age 50 is a valuable preventive intervention and today most health plans cover CRC screening (United States General Accounting Office, 2004). Yet, there is continued uncertainty about the specific screening strategies that should be offered to individuals who are at average risk for CRC.

There are two reasons for the prevailing uncertainty about what screening strategies make sense for these average-risk adults. First, the number of potential screening strategies is large, encompassing not only the choice of technology (or technologies) but also decisions about the age at which screening should begin, the frequency with which it should occur, and the age at which routine screening should end. Several medical technologies are available to detect early cancers or benign adenomas, the polyps that precede most colorectal cancers. Those technologies vary widely both in cost and detection capabilities. The list includes flexible sigmoidoscopy, colonoscopy, barium enema x-ray, and fecal occult blood tests. The choices are growing, too. New technologies, including imaging and molecular markers, are currently under development. Their entry will expand the range of alternative screening strategies even further.

A second factor that makes it difficult to settle on a specific strategy is that much is unknown about the natural history of colorectal cancer—how fast or slowly it develops, how frequently it arises from pre-existing benign adenomas, and how long those adenomas remain in a benign but detectable state before they convert to cancer. Although new information about these questions has emerged in recent years, it is indirect because, once they are detected, cancers or adenomas are virtually never left behind to grow and be observed. The effectiveness and cost of any screening strategy depend on the details of natural history and as long as those details remain unknown, it is impossible to be sure that one strategy is unequivocally better than another in the absence of a head-to-head trial comparing different strategies. Such a trial is unlikely to be performed because the cost and duration would be prohibitive.

Economic models of CRC screening offer a means for addressing questions about how to screen for CRC. Beginning with the work of David Eddy in the late 1970s (Eddy, 1980), many academic and government researchers have built computer models to describe the natural history of CRC and analyze the costs and effects of altering that history with selected screening strategies (Eddy et al., 1987; Frazier et al., 2000; Glick et al., 1998; Joseph et al., 1988; Khandker et al., 2000; Ladabaum et al., 2004b; Lieberman, 1995; Loeve et al., 1999, 2000; Neilson and Whynes, 1995; Ness et al., 2000; Sonnenberg and Delco, 2002; Sonnenberg et al., 2000; Vijan et al., 2001; Wagner et al., 1991, 1996; Whynes, 2004). The purpose of such models is to help decision makers evaluate which strategies to pay for, recommend, adopt, or use. As the field of cost-effectiveness analysis (CEA) in medicine advances (Gold et al., 1996), and as new evidence on the natural history of CRC emerges, the models have improved. But they have not been able to resolve the uncertainty about the comparative performance of different CRC screening strategies. Rather, they continue to disagree about how alternative strategies stack up against one another in their health effects and costs (Curry, 2003; Pignone et al., 2002).

Public health policy makers increasingly rely on CEAs to help them sift through the many choices confronting them. When different CEA models give different answers to the same question, confidence in their usefulness may suffer, since it is unclear to what extent the disagreement arises from uncertainty about the underlying evidence, which affects all decision making approaches, or from the modeling methods used by different modelers. Understanding the reasons for differences among models is therefore an important first step in building the public's confidence that CEA can provide objective and informative insights into the consequences of health policy choices.

The Institute of Medicine's (IOM's) National Cancer Policy Board (NCPB) convened the workshop, "*Economic Models of Colorectal Cancer Screening in Average-Risk Adults*" on January 26-27, 2004, to explore the reasons for differences among leading CEA models of CRC screening. Participants discussed the results of a collaborative pre-workshop exercise undertaken by five research teams that have developed and maintained comprehensive models of CRC screening in average-risk adults. The purpose of the exercise was to provide workshop participants with insights into each model's structure and assumptions and possible explanations for differences in their published analyses. Workshop participants also examined the current state of knowledge on key inputs to the models with a view toward identifying areas where further research may be warranted.

In keeping with the purpose of IOM workshops, this summary of its proceedings presents the individual perspectives and research of people who made presentations at the workshop and of many other experts who participated. This summary does not contain consensus recommendations, nor does it represent a consensus opinion of the IOM's NCPB. Nor is it intended as a guide for conducting or using cost-effectiveness analyses in CRC screening decisions.

It is particularly important to recognize that the purpose of the workshop was *not* to consider the relative merits of different strategies for CRC screening, or to suggest which CRC screening strategy is best. It was solely to consider the commonalities and differences among the CEA models bearing on the subject. The demand for more certain guidance from models by those who recommend or pay for screening strategies, while clearly a motivating force behind the workshop, was not its focus. More certain guidance may result in the future as modelers continue to grapple with and explain the differences in their findings.

THE COLLABORATIVE MODELING EXERCISE

Origin of the Exercise

The idea for collaboration among research teams that maintain published models of CRC screening grew out of a recent review by Michael Pignone and colleagues for the U.S. Preventive Health Services Task Force (Pignone et al., 2002). They systematically reviewed seven published CEAs of periodic CRC screening in average-risk adults. That review identified several aspects of model structure and underlying assumptions which, taken together, might account for most of the differences in cost-effectiveness rankings of CRC screening strategies. However, each model involves dozens of assumptions, and the reviewers concluded that the published reports provided insufficient information to determine which assumptions or aspects of model design were most important in explaining differences in conclusions across models.

The goal of the collaborative pre-workshop exercise was to shed light on the degree to which difference across models could be reduced by standardizing the values of key input parameters, or assumptions, across models. Any residual variation in model outcomes would be the result of differences either in parameters that remained unstandardized or in the structure of the models themselves. Secondary objectives were to demonstrate the benefit of collaborative interactions among modelers and to ascertain the research resources (time and money) required to mount such exercises.

General Approach

Five research teams with published CEAs of colorectal cancer screening agreed to participate in a comparative modeling exercise to further explore the reasons for disparate cost-effectiveness findings. Each of the models can track (via computer) a hypothetical cohort of average-risk Americans, beginning at age 50, over their remaining lifetimes and can estimate the number of years of life lived and the medical costs incurred by the members of that cohort.¹ The participating research teams were:

- The Harvard Model (Frazier et al., 2000), led by Karen Kuntz, Ph.D.;
- The Ladabaum Model (Ladabaum et al., 2004a; Song et al., 2004), led by Uri Ladabaum, M.D.;
- The Miscan Model (Loeve et al., 1999, 2000), led by Marjolein van Ballegooijen, M.D.;
- The Vanderbilt Model (Ness et al., 2000) led by Reid Ness, M.D.; and
- The Vijan Model (Vijan et al., 2001), led by Sandeep Vijan, M.D.

At the workshop, each team leader described essential features of the model's structure and assumptions. (See the appendixes with speakers' presentations.) The teams further agreed to provide cost-effectiveness results for a set of five specific screening strategies across 10 different combinations of assumptions, starting with the assumptions in their original models.

The Screening Strategies

All the strategies included in the pre-workshop exercise envisioned periodic screening of all average-risk Americans beginning at age 50 and ending at age 80. The five selected strategies were:

- 1) **F/S**: Annual fecal occult blood testing in combination with a flexible sigmoidoscopy every five years;
- 2) **S**: Sigmoidoscopy every five years;
- 3) **R**: A prototype radiology procedure every five years, with specific test characteristics and costs;
- 4) **C**: Colonoscopy every 10 years; and
- 5) **F**: Annual fecal occult blood testing.

These strategies were selected not for any posited superiority over other CRC screening approaches, but for the frequency with which they are advocated by practitioners today. Some of them represent strategies that have been recommended by professional groups (Smith et al., 2004; U.S. Preventive Services Task Force, 2002; Winawer et al., 2003). They also represent a wide range of procedure cost and test accuracy.

¹ Some of models can track all age cohorts of adults over a long period of time as well as specific age cohorts.

The prototype radiology strategy differed from the others by virtue of being defined by specific assumptions about costs and test performance. That route was necessary because some research teams had not investigated CRC screening with radiological technologies and therefore had no original assumptions at the ready. Moreover, an emerging imaging technique—virtual colonoscopy—may eventually join a much older radiology procedure—double-contrast barium enema (DCBE)—as an entry in the mix of available screening technologies. (Cotton et al., 2004; Pickhardt et al., 2003; Ransohoff, 2004). The assumptions specified for the prototype strategy represent an optimistic mix of cost and test performance characteristics based on the old and new radiology procedures.

The Standard Assumptions

The pre-workshop exercise specified standard assumptions in each of four groups listed below:

- 1) follow-up and periodic surveillance regimens—the assumptions that modelers make about how the health care system responds to a positive screening test, both in the short term (diagnostic follow-up) and after removal of a pre-cancerous adenoma (surveillance);
- 2) test performance characteristics—the sensitivity, specificity, and medical risk of tests for screening, follow-up, and periodic surveillance after treatment;
- 3) medical costs—the costs of screening, follow-up, and surveillance, as well as the costs of treating colorectal cancer at various stages; and
- 4) compliance—expected levels of adherence to the screening, follow-up, and surveillance strategies under evaluation.

The standardized assumptions in each of these groups are shown in Table 1.

TABLE 1. Standardized Assumptions for Pre-workshop Collaborative Exercise

COSTS	
Fecal occult blood est	\$10
Colonoscopy-diagnostic	\$625
Colonoscopy with polypectomy	\$900
Pathology per polyp	\$65
Sigmoidoscopy-screening	\$200
Sigmoidoscopy with polypectomy	N/A
Sigmoidoscopy with biopsy	\$375
Prototype radiology procedure	\$200
Lifetime CRC treatment cost	
Local	\$24,000
Regional	\$31,000
Distant	\$40,000
Cost of treating perforation	\$24,000

TABLE 1 Continued

TEST PERFORMANCE

Sigmoidoscopy	
Reach (percent of polyps)	50 percent
Sensitivity for polyps	85 percent
Sensitivity for cancer	95 percent
Specificity	100 percent
Fecal occult blood test-not rehydrated	
Sensitivity for polyps	10 percent
Sensitivity for cancer	40 percent
Specificity	97 percent
Colonoscopy	
Sensitivity for polyps	85 percent
Sensitivity for cancer	95 percent
Specificity	100 percent
Prototype radiology procedure	
Sensitivity for polyps	70 percent
Sensitivity for cancer	80 percent
Specificity	90 percent
Complications	
Colonoscopy major complications (perforation)	0.10 percent
Colonooscopy mortality rate	0.01percent
Sigmoidoscopy major complication	0 percent
Sigmoidoscopy mortality rate	0 percent
Prototype radiology	0 percent

FOLLOW-UP

Fecal occult blood test	Assume all positive fecal occult blood tests are followed by colonoscopy with polypectomy if true positive, or diagnostic colonoscopy if false positive
Sigmoidoscopy	a) Assume all positive screens are followed by colonoscopy with polypectomy if true positive, or diagnostic colonoscopy if false positive b) Assume positive screen involves no biopsy or polypectomy—cost is for screening sigmoidoscopy
Colonoscopy	Assume positive screen involves polypectomy with biopsy
Prototype radiology	Assume all positive screens are followed by colonoscopy with polypectomy if confirmed as true positive, or diagnostic colonoscopy if false positive

continued

TABLE 1 Continued

SURVEILLANCE	All individuals with adenomatous polyps get surveillance with colonoscopy every 5 years, beginning with fifth year post-polypectomy. Continued until 80 years of age or death
COMPLIANCE	100 percent with all aspects of strategy (screen, follow-up and surveillance)

A small number of basic assumptions, such as the discount rate, were also specified to remove possible sources of variation among models deriving from technical details (see Table 2).

Each research team first produced results with its own original assumptions, as shown in Table 3.² Then they produced results in successive runs when assumptions in one group at a time were assigned standardized values, leaving the rest at their original values. They generated a third set of results for a series of runs when one group of assumptions was left at its original values while the rest of the groups were standardized. A final run produced estimates when all assumptions in the exercise were standardized.

The standardized assumptions were *not* selected with the goal of specifying “correct” values. For the most part they were selected to strike a compromise among the five research teams’ original assumptions. However, some values were set to accommodate the least specific model in order to avoid the need for extensive reprogramming. For example, standardized compliance was set at 100 percent. Although an abundance of evidence suggests compliance is far less than perfect, it would have been time-consuming or impossible for all of the research teams to reconfigure their models to accommodate more realistic assumptions. This somewhat opportunistic standardization process underscores the danger of interpreting the standardized results as endorsing any specific colorectal cancer screening strategy, especially because the effectiveness of some strategies is bound to be more heavily dependent on high rates of compliance than others.

TABLE 2. Basic Assumptions

Population size (at 50 years of age):	100,000
Population demographics:	Average-risk individuals, U.S. population, both sexes and all races
Discount rate:	3 percent per year
Quality adjustments for all health states short of death:	None
Type of output:	Cohort model, followed from age 50 through age 85

² Researchers were given the choice of using the same assumptions as those made in their published papers or using other assumptions if more recent work had led them to new assumptions in the current versions of their models.

TABLE 3. Summary of Assumptions in Five Economic Models of CRC Screening

	Harvard	Laudabaum	Miscan	Vanderbilt	Vijan
TEST PERFORMANCE					
FOBT					
Rehydrated?	No	No	NA	No	No
Sensitivity-small adenoma	0.1	0.08	0.02	0.05 (<=10 mm)	0.05
Sensitivity-large adenoma	0.1	0.1	0.05	0.11	0.05
Sensitivity-high-risk adenoma	0.1	0.1	-	0.11	0.05
Sensitivity-early CA	0.33	0.4	0.6	0.13	0.3
Sensitivity-regional CA	0.33	0.4	0.6	0.13	0.5
Sensitivity-late stage CA	0.33	0.4 ^a	0.6	0.13	Not modeled ^f
Specificity	0.97	0.92	0.98	0.95	0.975
Test performance on each screen independent?	Yes	Yes	Yes	Yes	Yes
Medical risk of procedure	None	None	None	None	None
Each screen Independent?	Yes	Yes	Yes	Yes	Yes
FSIG					
Average percentage of lesions reached	About 50 ^d	50	64		55 percent
Sensitivity-small adenoma	0.85	0.7	0.75 for <=5 mm, 0.85 for 6-10 mm	0.75 (<5 mm); 0.80 (5-10mm)	0.85
Sensitivity-large adenoma	0.95	0.8	0.95	0.85	0.95
Sensitivity-high-risk adenoma	0.95	0.8	-	0.85	0.95
Sensitivity-early CA	0.95	0.9	0.95	0.95	0.95
Sensitivity-regional CA	0.95	0.9	0.95	0.95	0.95
Sensitivity-late stage CA	0.95	0.9 ^a	0.95	0.95	0.95
Specificity	1	0.95	0.92-hyperplastic polyp = FP	1	1
Risk of perforation	Not modeled explicitly—cost of complications embedded in procedure cost	0.0001 ^b	0	0.0001	0
Risk of bleeding			0	0	0
Risk of death—complications	0.0000014	0.00001 (0.0001×0.10)	0	0.00002	0

continued

TABLE 3 Continued

	Harvard	Laudabaum	Miscan	Vanderbilt	Vijan
Each screen independent?	Yes	Yes	Yes	Yes	Yes
Other?				Non-adenoma found in random 20 percent of exams—removed with additional risk and cost	
CSCPYP					
Sensitivity for small adenomas	0.85	0.85	0.80 for <=5mm, 0.85 for 6-10mm	0.75 (<5 mm); 0.80 (5-10mm)	0.85
Sensitivity for large adenomas	0.95	0.9	0.95	0.85	0.95
Sensitivity for high-risk adenomas	0.95	0.9	-	0.85	0.95
Sensitivity for early CA	0.95	0.95	0.95	0.95	0.95
Sensitivity for regional CA	0.95	0.95	0.95	0.95	0.95
Sensitivity for late stage CA	0.95	0.95 ^a	0.95	0.95	0.95
Specificity	1	1	1	1	1
Risk of perforation	Not modeled explicitly—cost of complications embedded in cost of procedure	0.001 ^b	0.002	0.001	0.001
Risk of bleeding			-	0.009	0
Risk of death from complications	0.00005	0.0001(0.001 × 0.10)	0.0001	0.0002	0.075
Each screen independent?	Yes	Yes	Yes	Yes	Yes
Other?				Non-adenoma found in random 26 percent of exams- removed with additional risk and cost	
FOLLOW-UP ASSUMPTIONS					
FOBT					
Does positive FOBT result in colonoscopy?	Yes	Yes	Yes	Yes	Yes
FSIG					
Is polyp found on screen removed in same procedure?	Yes	No	No	No	No
Is polyp found on screen biopsied?	NA	No, but CRC biopsied	No	No	Yes
Is a polyp found on screen referred to colonoscopy w/o biopsy or removal?	NA	Yes	Yes	Yes	No

continued

TABLE 3 Continued

	Harvard	Laudabaum	Miscan	Vanderbilt	Vijan
Is a polyp found on screen referred to colonoscopy after biopsy or removal?	Yes	No	No	No	Yes
SURVEILLANCE ASSUMPTIONS					
Is surveillance contingent on size or type of polyp? (if yes, explain)	Yes High-risk polyps only Every 3 years	No Every 5 years until age 80	Yes Depends on size and number of polyps 5 years if <3 small adenomas 3 years if ≥3 small adenomas 3 years if ≥1 large adenomas 5 years after negative surveillance No age bounds	Yes Adenomas only Colonoscopy every 5 years	Yes Only for those with ≥1 cm or multiple polyps Colonoscopy every 5 years
Describe colonoscopic surveillance schedule					
COST ASSUMPTIONS					
FOBT (unit cost)	\$38	\$20	\$4.50	\$4	\$18.40
FSIG (unit cost)	\$279	\$290	\$100	\$401	\$389
FSIG with biopsy (unit cost)	Embedded in polypectomy cost	\$440		NA	\$397
FSIG with polyp removal	\$528(279+249)	NA		NA	Not modeled
CSCPYP-diagnostic	\$1,012	\$820	\$650	\$681	\$653
CSCPYP with polyp removal	\$1,065(1012+553)	\$1,200	\$750	\$808	\$737
Pathology	Embedded in polypectomy cost	Included in above	Included in CSCPYP with polypectomy	\$181	\$95
Cost of treating complications	Embedded in colonoscopy cost	\$26,000	\$30,000	Hemorrhage:\$4,360; perforation/death:\$13,000	\$13,361
Cost of treating CRC-local (lifetime to age 85—discounted 3 percent)	Initial cost Continuing cost Final cost	\$21,941 \$15,000 \$392/year \$18,014	\$46,000 \$26,800 \$2,100/year	\$17,843 (first 6 months) \$473/yr (prorated) \$18,589 (last 6 months)	\$34,252 \$47,227
Cost of treating CRC-regional (lifetime to age 85—discounted 3 percent)	Initial cost Continuing cost	\$68,000 \$17,920 \$1793/year	\$26,800 \$2,100/year	\$20,517 (first 6 months) \$2,161/year (prorated)	\$41,602
Cost of treating CRC-distant (lifetime to age 85—discounted 3 percent)		\$71,000			

continued

TABLE 3 Continued

	Harvard	Laudabaum	Miscan	Vanderbilt	Vijan
Initial cost	\$29,235		\$26,800	\$23,449 (first 6 months)	
Continuing cost	\$19,562/year		\$2,100/year	\$23,578/year (prorated)	
Final cost	\$15,531			\$18,589 (last 6 months)	
COMPLIANCE ASSUMPTIONS					
Does compliance with screening vary by type of patient? (if so explain)	No	No	No	Yes	No
Does compliance with screening change for individual patient over time?	No	No	No	No	No
Is patient's compliance at any time dependent on prior compliance?	No	No	Depends on past compliance	Yes	Yes
Does compliance with screening vary with screening technology?	Yes	No	Yes	Yes	No
Compliance with FOBT	0.6				
Compliance with FSIG	0.6				
Compliance with CSCP	0.6				
Compliance with CSCP used in surveillance or as a F/U test	0.8				
Describe approach to modeling compliance	Persons have a "p" percent chance of complying in a screening year	Compliance is all-or-nothing	Colonoscopy—first: 30 percent; subsequent: if previously Attended, 63.2 percent; if previously not attended, 15.8 percent Radiology: 40 percent, 72.7 percent, 18.2 percent FSIG: 40 percent, 72.7 percent, 18.2 percent FSIG/FOBT: 35 percent, 68.3 percent, 17.1 percent FOBT: 60 percent 85.7 percent, 21.4 percent	Compliance varies with frequency of screening and screening modality. FOBT: every 1 years—35 percent every 2 years—14 percent every 5 years—11 percent one time—10 percent never—30 percent Endoscopy/Radiology every 5 years—53 percent one time—17 percent never—30 percent Surveillance Endoscopy 80 percent (80-100)	

NOTES: All assumptions are based on current versions (as of December 20, 2003) of researchers' models, except for Harvard model, which followed assumptions published in Frazier et al. (2000). *a.* not relevant-distant cancers assumed to present with symptoms within the year; *b.* includes all major complications; *c.* assumed all late-stage cancers are symptomatic; *d.* distal and rectal lesions, age- and sex-specific

Note also that assumptions about the natural history of colorectal cancer screening differ across models, but standardizing those assumptions is especially difficult to do and was not attempted. Natural history assumptions—the prevalence and incidence of adenomas and other benign polyps, how fast adenomas progress to cancer, what proportion of cancers are preceded by benign adenomas, and how fast cancers progress from early to late stages, and life-expectancy of the population with and without colorectal cancer—are interrelated with one another. They can be specified at various levels of detail, by age, sex and race, or other risk factors, as well as by location of the lesion in the colon and by the existence of past or concurrent adenomas. Some models can incorporate very detailed natural history assumptions, whereas others cannot. Additionally, model structures vary in the kind of natural history inputs required. For example, some models require data on the monthly or annual probability that an adenoma will progress to early cancer, whereas others require estimates of the number of years of growth required before an adenoma makes the transition to colorectal cancer. Because of these difficulties, the research teams agreed that the comparative modeling exercise should not attempt to standardize assumptions regarding natural history. Instead, they agreed to provide some intermediate results: the number of adenomas or polyps detected, deaths from CRC, and total mortality at each age between 50 and 85 in the absence of screening. Those results would allow an indirect comparison of natural history assumptions across the models.³

Specification of Model Outputs

For every model run, the research teams provided the coordinators of the exercise⁴ with estimates of the total number of years of life lived and total medical costs incurred by a population of 100,000 average-risk 50-year-old adults from age 50 until death or age 85, whichever comes first.⁵ These outputs were reported both as simple totals and in terms of their net present value (NPV) at the starting age (age 50).⁶

The cost-effectiveness of any screening strategy compared with any other strategy or with no screening, may be calculated from those outputs. For example, the cost-effectiveness of a strategy compared with no screening at all is as follows:

$$CE = \frac{(\text{Lifetime cost with strategy} - \text{Lifetime cost with no screening})}{(\text{Years lived with strategy} - \text{Years lived with no screening})}$$

If both numerator and denominator are positive, then the C/E ratio represents the extra costs required to achieve each extra year of life. If the numerator of the ratio is negative, while the denominator is positive, then the strategy saves both costs and lives and is unequivocally superior to doing nothing.

³ The intermediate results were not presented at the workshop and are therefore not discussed in this summary.

⁴ Four workshop participants, Martin Brown, Louise Russell, Michael Pignone, and Judith Wagner, led the development of the pre-workshop exercise and coordinated the analysis of its results. Michael Pignone presented the analysis at the Workshop (See his presentation in Appendix I.)

⁵ Screening programs lasted 30 years, but the reporting period continued for 35 years.

⁶ All comparisons using NPV applied an annual discount rate of 3 percent (Gold et al., 1996).

TABLE 4. Assumption Settings for Pre-workshop Exercise

Run Number	Cost	Test Performance	Follow-Up and Surveillance	Compliance
1	Orig	Orig	Orig	Orig
2	Std	Orig	Orig	Orig
3	Orig	Std	Orig	Orig
4	Orig	Orig	Std	Orig
5	Orig	Orig	Orig	Std
6	Std	Std	Std	Std
7	Std	Std	Std	Orig
8	Std	Std	Orig	Std
9	Std	Orig	Std	Std
10	Orig	Std	Std	Std

NOTES: “Std” = all assumptions in the group are standardized (see Table 2); “Orig” = all assumptions in the group are set at their original values (see Table 3).

The Comparisons

The five research teams were asked to report results for the baseline—no screening—as well as for 10 runs for each of the five screening strategies, 50 runs in all, as noted in Table 4. Each team ran its model 52 times (twice for the no-screening strategy,⁷ and 10 times for each of the 5 screening strategies). Thus, the research teams submitted a total of 260 separate computer runs for analysis by the coordinators.

Two runs represent the extremes of the standardization spectrum. Run number 1 produced results for the model’s original assumptions in all four areas—follow-up, test performance, cost, and compliance. Run number 6 showed the results when all assumptions were set to their standardized values. All other model runs involved combinations of original and standardized assumptions.

Results

Baseline Estimates (No Screening)

The research teams estimated the number of years of life lived (life expectancy) by an average 50-year-old and lifetime CRC-related costs per person, when no screening program was in effect and all assumptions were set to each team’s original values (Table 5). Any differences among models in those estimates would reflect variations either in model structure or in assumptions about age-specific mortality in the U.S. population, age- and stage-specific incidence of colorectal cancer, and costs of treating colorectal cancer by age and stage.

The research teams reported a range of estimates of years of life lived. The average life expectancy in the model with the highest predicted value was about 2.25 years or 1.1 times longer than in the model with the lowest value. Two models predicted almost identical life expectancies of 25 years; three predicted identical life expectancies of 27 years. In reviewing these results at the workshop, several researchers suggested that the

⁷ The no-screening strategy required Runs #1 and #2, because standardizing costs of treating colorectal cancer (as in Run # 2) would change model outcomes even without screening. All other runs involve changes in assumptions that would occur only under a screening regimen.

TABLE 5. Predicted Years of Life Lived and Lifetime per-Capita CRC Costs No Screening: Original Model Assumptions

Model	Years of Life Lived	Rank	Lifetime Costs	Rank
Harvard	25.12	2	\$1,811	3
Ladabaum	27.23	3	\$2,994	4
Miscan	27.29	4	\$2,283	5
Vanderbilt	27.39	5	\$1,792	2
Vijan	25.07	1	\$1,665	1

NOTE: Results are not discounted to net present value.

differences were due to the use of mortality statistics from different years. Sandeep Vijan and Karen Kuntz remarked that assumptions about life expectancy in their models were based on older life tables. Mortality rates have decreased substantially in the last decade, especially in older age groups.

The variation among models reported by the research teams in estimated lifetime costs was larger than the variation in effects, with the highest estimate about 1.8 times higher than the lowest. Those disparities reflect the models' very different assumptions about and approaches to estimating the cost of treating colorectal cancer. When treatment costs were standardized to the values shown in Table 1, the range of estimated costs diminished substantially to a ratio of 1.2 between the highest and lowest values. Some participants posited that differences in assumptions about cancer incidence probably account for the remaining variation in colorectal cancer costs.

Screening Estimates Under Original Assumptions

Differences among the five models in estimates of the effect of screening under each team's original assumptions were presented by Michael Pignone and discussed by the research teams and other participants.

Comparing screening with no screening. Figure 1 shows the net increases in years of life lived and lifetime costs (discounted to their NPV), compared with no screening under the full set of original assumptions adopted by each research team. The research teams reported wide variation in ratio terms for each of these two components of cost-effectiveness. For example, the NPV of lifetime cost reported for a screening program of flexible sigmoidoscopy every 5 years ranged from \$224 per person (Miscan) to \$1,159 per person (Vanderbilt), a five-fold difference between the two. The predicted gains in life expectancy from screening are less varied than for costs, but still high. For example, the net present value of life-years gained from flexible sigmoidoscopy ranged from 2,723 per 100,000 50-year-olds (Miscan) to 4,265 (Vanderbilt), a ratio between the highest and lowest of about 1.6.

The research teams reported that the most effective strategy differed across the models. Two models predicted that F/S gains the most years of life for the population, two models predicted that R would be most effective, and one model predicted C is the most effective. The least costly strategy also differed across models. Two models predicted that S is the least costly strategy, two that F is least costly, and one that R is least costly⁸

⁸ In discussion following the presentation of the results, Louise Russell and others pointed out that the analysis focused on differences across models in their single best estimates of effects and costs, whereas the research teams have acknowledged and reported on the range of uncertainty surrounding their estimates in their research papers. Had uncertainty been modeled in this exercise, the range of reported results might have overlapped.

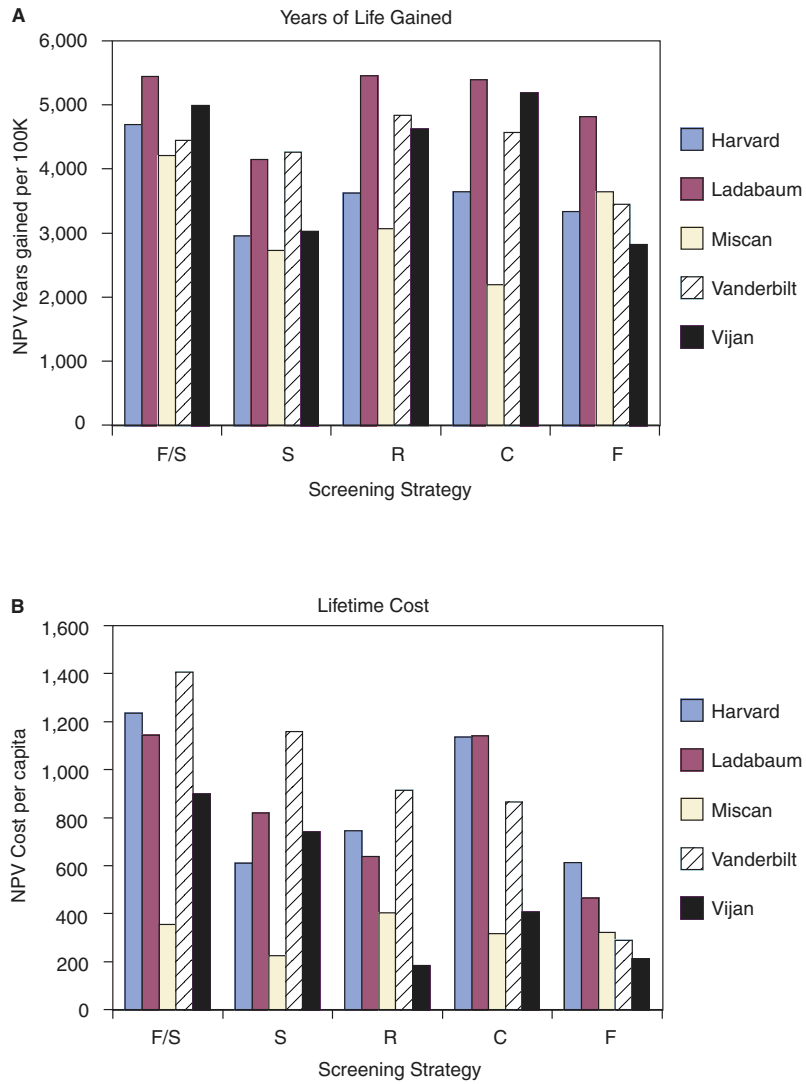


FIGURE 1. Years of life gained and lifetime costs of screening: original assumptions.

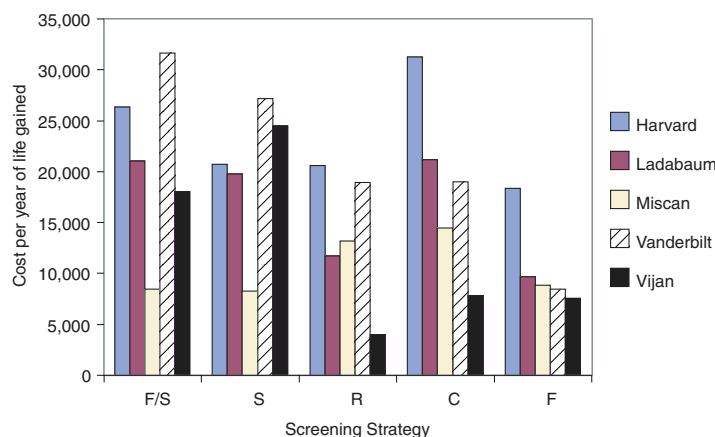


FIGURE 2. Cost effectiveness of screening: original assumptions.

As a result, estimates of the cost-effectiveness ratio also varied across the five models, in some cases by a five-fold difference between the highest cost-effectiveness ratio and the lowest (Figure 2). Despite that variation, Michael Pignone pointed out, all the models show that all of the strategies meet common benchmarks of cost-effectiveness. Every research team estimated that, when compared with no screening, colorectal cancer screening could deliver an additional year of life for a cost of less than \$40,000, regardless of which strategy is adopted.

Comparing strategies with one another. The goal of cost-effectiveness analysis is to compare alternative strategies with one another (Gold et al., 1996). The disparities among CRC models in such comparisons prompted the Workshop to begin with. So, participants reviewed the performance of strategies with each other as reported by the research teams.

The first step in making such comparisons is to rule out any screening strategy that is both less effective and more costly than at least one other. Strategies ruled out at this stage are referred to as “strongly dominated.” The second step is more subtle. It requires ruling out any strategy whose gains in life expectancy, compared with the next most effective strategy, come at an incremental cost that is higher than the incremental cost of achieving gains at least as great through still another strategy. Strategies ruled out at this stage are referred to as “weakly dominated.” Any strategies surviving this two-step elimination process present a true trade-off between successively higher costs and greater health benefits. Louise Russell reminded participants, however, that the process is based on point estimates, which are subject to uncertainty. All research groups have routinely assessed the effect of uncertainty on those estimates. Had the exercise included such analyses, it might have found that some strategies that were ruled out were essentially equivalent to those ruled in.

Once the strategies surviving the two rule-out tests are identified, their incremental cost-effectiveness ratios can be calculated by sorting them into ascending order of effectiveness, measuring the differences in both cost and years of life gained compared with the next most effective strategy (or with no screening for the least effective strategy), and calculating the cost-effectiveness ratio. Michael Pignone summarized the results.

TABLE 6. Incremental Cost-Effectiveness Ratios of Five CRC Screening Strategies: Original Assumptions

	Harvard	Ladabaum	Miscan	Vanderbilt	Vijan
F/S	\$45,976	SD	\$8,848	SD	SD
S	WD	SD	\$8,230	SD	SD
R	WD	\$27,069	SD	\$44,936	\$3,980
C	WD	SD	SD	WD	\$38,854
F	\$18,347	\$9,631	WD	\$8,409	SD

NOTES: F/S = annual fecal occult blood test; sigmoidoscopy every 5 years; S = sigmoidoscopy every 5 years; R = prototype radiology procedure every 5 years; C = colonoscopy every 10 years; F = annual fecal occult blood test; WD = strategy is weakly dominated by at least one other strategy; SD = strategy is strongly dominated by at least one other strategy.

SOURCE: M. Pignone Workshop Presentation (Appendix I).

Across the five models, the surviving strategies differed substantially and the incremental cost-effectiveness ratios of those strategies also differed (Table 6). Thus, according to Pignone, under their original assumptions, the five research teams would present very different options to policy makers.

Estimates of Screening Under Standardized Assumptions

Michael Pignone presented the effect of standardizing all of the assumptions in the four groups together on differences among models.

Comparing screening with no screening. Under the full set of standardized assumptions, the two components of the cost-effectiveness ratio still varied across models. Sometimes, but not always, by less in ratio terms than when the models used their original assumptions (Figure 3). Differences across models in predicted per capita lifetime costs were greatest for strategy S, where they ranged from \$718 per person (Miscan) to \$1,436 per person (Vanderbilt), a two-fold difference between the two. (Recall that the difference was five-fold under the original assumptions.)

Differences across models in years of life gained from screening did not change in a systematic way after standardization. The range of variation grew modestly for two strategies and declined for the other three. The NPV of life-years gained from strategy S ranged from 3,470 per 100,000 50-year-olds (Vijan) to 6,954 (Vanderbilt), a ratio of 2.0 between the highest and lowest, compared with a ratio of 1.6 under the teams' original assumptions. The two strategies involving sigmoidoscopy seemed to resist convergence in predicted years of life gained more than other strategies.

Standardization of assumptions did result in agreement across models on the most effective and least costly strategies. All of the research teams estimated that F/S gains the most years of life and all found F to be the least costly strategy.

The cost-effectiveness ratio for each strategy continued to vary across the five models, but the range of difference as measured by the ratio of the highest to lowest narrowed with full standardization (Figure 4). With all tested assumptions standardized, the cost-effectiveness ratio varied across models by a factor of 1.5 to 2.0 for every strategy.

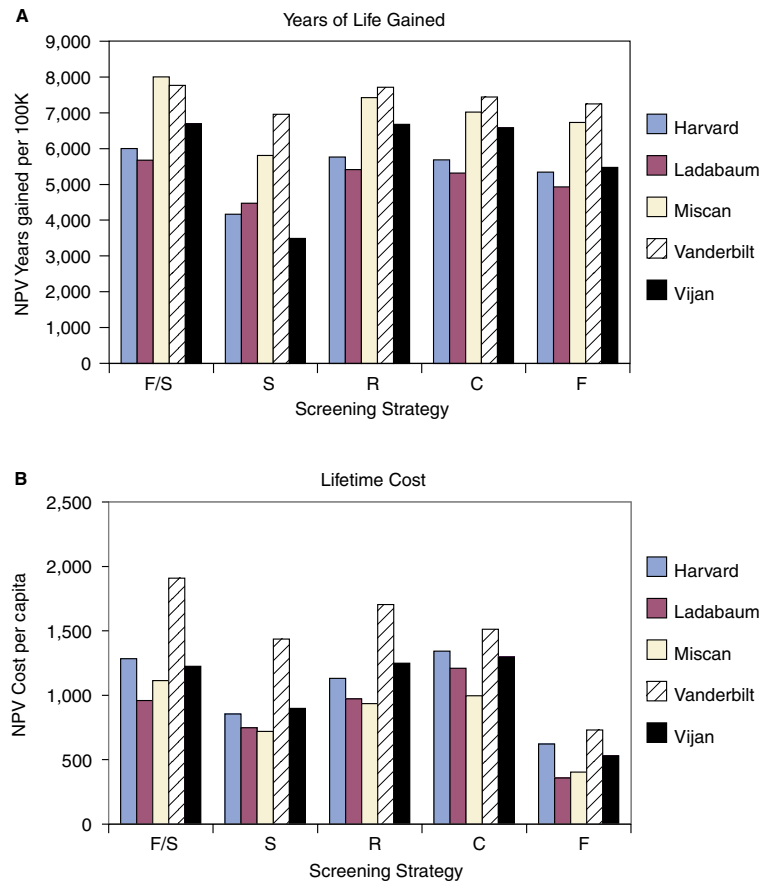


FIGURE 3. Years of life gained and lifetime costs of screening: standardized assumptions.

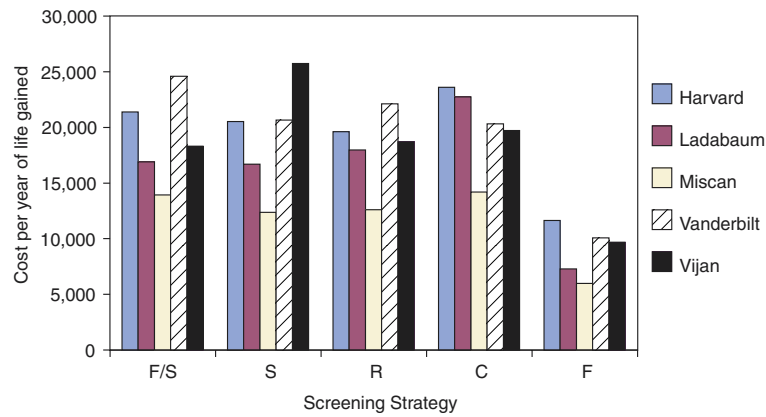


FIGURE 4. Cost-effectiveness of screening: standardized assumptions.

TABLE 7. Incremental Cost-Effectiveness Ratios of Five CRC Screening Strategies: Standardized Assumptions

	Harvard	Ladabaum	Miscan	Vanderbilt	Vijan
F/S	\$99,997	\$79,920	\$55,828	\$355,647	\$56,969
S	SD	SD	SD	SD	SD
R	WD	SD	WD	\$209,906	SD
C	SD	SD	SD	SD	SD
F	\$11,632	\$7,232	\$5,980	\$10,073	\$9,676

NOTES: F/S = annual fecal occult blood test; sigmoidoscopy every 5 years; S = sigmoidoscopy every 5 years; R = prototype radiology procedure every 5 years; C = colonoscopy every 10 years; F = annual fecal occult blood test; WD = strategy is weakly dominated by at least one other strategy; SD = strategy is strongly dominated by at least one other strategy. SOURCE: M. Pignone Workshop Presentation (Appendix I).

Comparing strategies with one another. Under standardized assumptions, all models agreed about which strategies survived the dominance test (Table 7), but the incremental cost per year of life gained for F/S, versus S, still varied widely.

Effect of Specific Assumption Groups on Variations across Models

The research teams examined the separate effect of each group of assumptions on the estimates for four strategies (versus no screening).⁹ They compared model results when each of the four assumption groups was standardized while the rest were set to their original values. Estimated years of life gained did not show any general pattern of convergence (Table 8). The ratio between the highest and lowest estimate of years of life gained actually increased for some strategies when some assumption groups were standardized. The range of estimates for lifetime costs associated with a particular strategy declined substantially for two strategies but increased slightly for two others.¹⁰

TABLE 8. Effect of Standardizing Individual Assumption Groups on Variation Across Models: Ratio of Highest Estimate to Lowest Estimate

	F/S	S	C	F
Net Present Value of Years of Life Gained Due to Screening				
Original assumptions	1.3	1.6	2.4	1.7
Costs std	1.3	1.6	2.4	1.7
Test performance std	1.2	1.8	1.4	1.4
FU/Surveillance std	1.3	1.7	1.5	1.7
Compliance std	1.4	1.8	1.3	1.3
All groups std	1.4	2.0	1.4	1.5
Net Present Value of Lifetime Costs Incurred Due to Screening				
Original assumptions	3.9	5.2	3.6	2.9
Costs std	2.8	3.2	4.1	NA
Test performance std	4.7	7.8	3.8	6.0
FU/Surveillance std	4.0	5.0	4.2	2.5
Compliance std	3.0	3.9	2.5	5.6
All standard	1.9	2.0	1.5	2.0

NOTES: Net present value computed at 3 percent per annum; FU/Surveillance = Follow up and Surveillance; std = standardized; NA = Not available. One model predicted net cost savings, thereby making calculation of a ratio impossible.

⁹ Recall that strategy R was conceived with standard test performance and test costs at the outset because at the time not all of the research teams had studied a radiologic technology for screening. Therefore, the analysis of the impact of individual strategies did not include R.

¹⁰ One model estimated negative net lifetime costs for strategy F under standardized cost assumptions. The absolute difference in costs between the highest-cost and lowest-cost estimates increased compared with the difference among the models when original assumptions were used.

TABLE 9. Effect of Standardizing Specific Assumption Groups on Variation in Cost-Effectiveness Ratios across Models: Ratio of Highest Estimate to Lowest Estimate

	Strategy				
	F/S	S	R	C	F
Original assumptions	3.7	3.3	5.2	4.0	2.4
Costs standard	3.1	2.3	5.6	2.5	NA
Test Performance standard	4.4	6.2	5.5	3.7	5.9
FU/Surveillance standard	3.8	3.6	1.8	2.4	2.4
Compliance standard	3.5	3.8	5.7	3.2	5.3
All groups standard	1.8	2.1	1.8	1.7	1.9

NOTES: Cost-effectiveness compared with no screening. NA = Not available. One model predicted net cost savings, thereby making calculation of a ratio impossible.

The cost-effectiveness ratios for each strategy did converge across models as a result of standardizing costs (Table 9). That result led Michael Pignone to conclude that standardizing cost assumptions seemed to have the biggest effect on convergence among models. However, he also warned that standardizing other groups of assumptions individually did not lead to systematic convergence across models in the estimated cost-effectiveness of any strategy. Because the cost-effectiveness ratios converged when all four assumption groups were standardized, Pignone observed, it is probable that the assumption groups interact in their effects on model outcomes.

Lessons Learned from the Exercise

The results of the pre-workshop exercise prompted substantial discussion among the workshop participants. Comments focused both on the strengths and limitations of the exercise itself and on the implications of the collaborative exercise for further model development.

The Impact of Subtle Differences in Model Structure

Workshop participants identified some subtle differences in structure across models that affected the results of the exercise itself. One is how the different models account for polyps that are not adenomas. As described by T.R. Levin, most experts believe that the vast majority of colorectal cancers arise from pre-cancerous adenomas. These lesions come in a variety of morphologic and histological forms and they grow and progress to cancer at varying speeds. They are not, however, the only polyps that appear in the colon or rectum—other kinds of benign lesions, notably hyperplastic polyps, are quite common in older people (Lieberman et al., 2003). Although hyperplastic lesions are thought to present a low risk for progression to cancer (Imperiale et al., 2003; Lieberman et al., 2004), some screening technologies may detect them with higher frequency than others. In particular, endoscopy and radiology would be more likely to detect non-adenomatous polyps than would fecal occult blood testing, because non-adenomatous polyps rarely bleed.¹¹ Once detected, however, such lesions are typically removed and sent for biopsy because they cannot be differentiated from adenomas by any other method. Martin Brown observed that the cost of follow-up procedures triggered by detection of a non-adenomatous lesion may have a major effect on the incremental cost of screening.

¹¹ Even when a non-adenomatous polyp is not detectable by the screening test, it could be found serendipitously through diagnostic follow up of a test result that was positive for reasons unrelated to its presence.

TABLE 10. Impact of Excluding Non-Adenomatous Polyps from the Vanderbilt Model: Percent Change in Outcomes Resulting from Exclusion

Strategy	Change in Lifetime Cost Percent	Change in Years of Life Gained Percent
F/S	- 34	- 1
S	- 45	- 9
R	- 38	- 1
C	- 14	- 0
F	-14	- 0

NOTES: F/S = annual fecal occult blood test, sigmoidoscopy every 5 years; S = sigmoidoscopy every 5 years; R = prototype radiology procedure every 5 years; C = colonoscopy every 10 years; F = annual fecal occult blood test; WD = strategy is weakly dominated by at least one other strategy; and SD = strategy is strongly dominated by at least one other strategy.

SOURCE: R. Ness Workshop Presentation (Appendix G).

The research teams reported that not all of the models account for the implications of detecting non-adenomatous lesions. Karen Kuntz noted that the Harvard model did not include such lesions at all. Some recognize them implicitly rather than explicitly by making a downward adjustment in the assumed specificity (i.e., increasing the false positive rate) of the screening test, or an upward adjustment in the average cost of diagnostic follow-up of adenomas detected through screening. Reid Ness observed that standardizing assumptions in the groups involving test performance (i.e., test specificity) and costs (i.e., follow-up costs) masked these subtle differences in model structure.

The Vanderbilt team was the first to recognize the impact of non-adenomas on the standardized results of the pre-workshop exercise. Vanderbilt's estimates of the lifetime costs of all screening strategies were much higher than those reported to the workshop by the other research teams (see Figure 3B). The Vanderbilt model explicitly recognizes the prevalence of non-adenomatous polyps and independently records the costs of diagnostic follow-up of those lesions. Because other models either excluded those costs or considered them implicitly through adjustments in other assumptions, they effectively ignored them when test specificity and unit costs were standardized. The Vanderbilt team assessed the importance of this difference in model structure by reanalyzing the five strategies under fully standardized assumptions after setting the prevalence of non-adenomas to zero in their model. Reid Ness reported that the lifetime costs of all screening strategies declined (Table 10). Those with the highest relative decline were the screening strategies most likely to detect non-adenomas, namely those that involve direct visualization of the colon and diagnostic follow-up of all polyps with colonoscopy.¹² Ignoring non-adenomas also had a small negative impact on life years gained, because doing so would imply fewer referrals to colonoscopy. Such referrals generated by a screening test that was positive because of a non-adenomatous polyp would sometimes result in serendipitous discovery on follow-up of an adenoma or cancer, with consequent life-extending benefits.

¹² Note that under the standardized assumptions\ all polyps found on sigmoidoscopy are referred to full colonoscopy for removal and biopsy.

TABLE 11. Impact of Detecting Non-Adenomas on Incremental Cost-Effectiveness Ratios Study

Strategy	Harvard	Ladabaum	Miscan	Vijan	Vanderbilt (old)	Vanderbilt (new)
F/S	\$99,977	\$79,920	\$55,878	\$56,969	\$355,647	\$355,608
S	SD	SD	SD	SD	SD	SD
R	WD	SD	WD	SD	\$209,906	\$114,510
C	SD	SD	SD	SD	WD	SD
F	\$11,632	\$7,272	\$5,980	\$9,676	\$10,073	\$8,659

NOTES: F/S = annual fecal occult blood test, sigmoidoscopy every 5 years; S = sigmoidoscopy every 5 years; R = prototype radiology procedure every 5 years; C = colonoscopy every 10 years; F = annual fecal occult blood test; WD = strategy is weakly dominated by at least one other strategy; and SD = strategy is strongly dominated by at least one other strategy.
SOURCE: R. Ness Workshop Presentation (Appendix G).

The Vanderbilt team reported that their reanalysis had a limited effect on incremental cost-effectiveness ratios under standardized assumptions. Their results were in closer agreement with those of the other models, but their estimate of the incremental cost-effectiveness of moving from F to F/S was still much higher (Table 11). Ness concluded that different approaches to non-adenomas may have been responsible for some of the variation among models. However, other factors recognized but not fully understood by participants continued to support a high level of variation in the incremental cost-effectiveness estimates for alternative screening strategies.

Other Limitations of the Exercise

Several participants noted that standardizing to a single set of values in each assumption group is insufficient if the goal is to determine unequivocally the extent to which variation across models can be explained by different values in the four groups of assumptions.¹³ Other standardized values for the same group of assumptions might have generated more, or less, agreement among models than did the values chosen for the pre-workshop exercise. In the extreme, it might be possible to force a measure of agreement among models by selecting standardized assumptions that strongly favor certain strategies. Judith Wagner noted that the standardized assumptions selected in the pre-workshop exercise may have differentially favored the two strategies involving fecal occult blood testing—F and F/S—since the effectiveness of fecal occult blood testing is especially sensitive to assumptions about compliance. A more robust exercise would have tested multiple values for standardized assumptions, perhaps selected probabilistically from a range of possible values. Such an exercise—involving hundreds or thousands of model runs—would have required time and resources that none of the research teams could afford without external funding.

Several participants noted that convergence among models does not necessarily imply that the models are valid representations of the true cost and effectiveness of any given CRC screening strategy. To paraphrase Marjolein van Ballegoijen, if the models merge when we standardize, should we believe the merged results? The ultimate test of any model is how well it predicts what occurs in the real world. If all models share flawed designs or assumptions, agreement does not constitute validity.

¹³ Recall that natural history assumptions were not standardized in the exercise.

Michael Pignone observed that a key structural aspect shared by all the models is that the sensitivity of each subsequent screening test performed in a periodic screening program is independent of the results of earlier tests. That assumption may be questioned most strongly in the case of annual fecal occult blood testing. Researchers have posited that some adenomas may never bleed, while others may bleed regularly. If more were known about whether such patterns actually exist, and the frequency with which they do, models could be constructed that would adjust the assumed probability that people with adenomas receive positive fecal occult blood testing results in the second and subsequent years of a screening program, based on their test results in previous years. In Pignone's view, adjustments such as these could have profound effects on the estimated effectiveness of periodic fecal occult blood testing. At present, however, data simply do not exist to provide reasonable estimates of such contingent probabilities, and modeling them would be a complex undertaking.

MAJOR CHALLENGES TO MODELING THE COST-EFFECTIVENESS OF CRC SCREENING

The workshop benefited from presentations by leading researchers on the current state of knowledge about the natural history of CRC and the effects of screening, follow-up and surveillance. Those presentations took place on the afternoon of the first day and are published in the appendixes to this workshop report. They covered the following topics:

- Evidence on test performance of current and experimental screening technologies—Brian Mulhall, M.D.;
- Issues in measuring the costs of CRC screening and treatment—Martin Brown, Ph.D.;
- Evidence on rates of compliance with CRC screening—Sally Vernon, Ph.D.;
- Evidence on endoscopy utilization and capacity—Laura Seefe, M.D.;
- Evidence on compliance with follow up and surveillance—Todd Anderson, M.S.;
- Evidence on efficacy of follow-up and surveillance protocols—Deborah Schrag, M.D.;
- Evidence on the natural history of CRC—T.R. Levin, M.D.; and
- The National Cancer Institute's CISNET Program and its approach to model validation—Eric (Rocky) Feuer, Ph.D.

The presentations were intended to identify the best evidence both to improve models and to identify gaps in knowledge. Together with the collaborative modeling exercise, they stimulated workshop participants to confront three major issues that challenge the ability of models to provide useful information to health policy makers.

Uncertainty

Workshop participants spent a good deal of time addressing the uncertainty underlying the costs and effects of colorectal cancer screening. Louise Russell and Michael Pignone argued that pervasive uncertainty makes bottom-line conclusions about the comparative performance of different screening strategies, although not about the overall cost-effectiveness of screening itself, potentially inappropriate because they presume a degree of precision that the current state of knowledge cannot support and may never be able to support. Ironically, it is exactly that kind of precise, bottom-line guidance that decision makers seek.

The Extent of the Problem

The workshop presentations underscored how little is known about many aspects of screening or its consequences. Brian Mulhall's review of the wide range of estimates of fecal occult blood test sensitivity and specificity for adenomas in people who are recommended for screening suggested that uncertainty about test performance is not limited to new, emerging, or uncommon technologies. Fecal occult blood testing, one of the oldest technologies available for CRC screening, has been the focus of several large-scale randomized screening trials, all of which have demonstrated that it can reduce mortality from CRC (Jorgensen et al., 2002; Mandel et al., 1999; Scholefield et al., 2002). However, according to Mulhall, none of those trials has provided definitive evidence on its sensitivity and specificity for adenomas.

The range of uncertainty about the performance of other common screening technologies, such as sigmoidoscopy and colonoscopy, may be lower, but there is still substantial variation in findings across studies, according to Mulhall. Long considered the gold standard for detection of adenomas, with a sensitivity that was believed to be close to 100 percent, colonoscopy was found in a recent head-to-head comparison with virtual colonoscopy, a new radiological procedure, to miss about 11 percent of advanced adenomas (Pickhardt et al., 2003).

Martin Brown emphasized the uncertainty surrounding estimates of CRC treatment cost, which is a major component of the lifetime cost of a screening program. Variation in estimates of this cost, and its distribution over the years following diagnosis, can make for very large differences among models in estimates of the net cost of screening. A strategy that prevents a large number of cancers is far less costly when treatment costs, and thus savings from early detection, are high rather than when they are low. Although it might seem easy to make accurate estimates of such costs because billing and claims data are available from health care providers or insurers, prices can vary widely from provider to provider and across different payers. Moreover, estimates vary depending on whether they are based on prices charged or on audits of the amount and value of the labor and other inputs required to produce each service. Thus, Brown concluded, a seemingly straightforward element—the cost of treating colorectal cancer—is in practice subject to considerable uncertainty.

In her discussion of the current evidence on compliance with CRC screening, Sally Vernon emphasized the uncertainty about compliance in a screening program that continues over a patient's lifetime. Though much is known about factors that affect patients' adherence to screening, notably insurance coverage, education, and physician recommendations, survey evidence is insufficient to provide accurate estimates of the levels of adherence to periodic screening that could be expected over the long term. William Lawrence observed that we do not know whether compliance rates estimated from one-time

consumer surveys represent a combination of some patients who fully adhere to a lifelong screening program and others who never receive any screening, or whether they represent a more homogenous population all of whom adhere to a screening strategy intermittently. In Lawrence's view, these distinctions have important implications for the effectiveness and cost of screening programs.

Uncertainty about compliance is also high because surveys define compliance differently. Sally Vernon observed that surveys that measure the number of patients who receive fecal occult blood test kits from their physicians typically report high compliance, whereas those that measure the number of test kits returned for analysis show much lower rates.

Uncertainty about the natural history of colorectal cancer in average-risk individuals was also a topic of discussion throughout the workshop. T.R. Levin summarized the evidence on several important aspects of that history. His review of one aspect—the proportion of CRCs that arise *de novo*, without spending time as a pre-cancerous adenoma—illustrates how difficult it can be to resolve uncertainty. Cancers arising from fast-progressing adenomas or from adenomas that are difficult to detect with even the most sensitive screening tests could be mislabeled as *de novo*. The difference could be important for comparing strategies because, for example, underestimating the proportion of cancers that arise *de novo* could favor screening technologies that have high sensitivity for adenomas over those that are better at diagnosing early cancer. Brian Mulhall observed that the emergence of molecular assays in the near future may offer new opportunities for definitive research on the question of *de novo* cancer.

How to Think About Uncertainty

A prerequisite to dealing with the effects of uncertainty is to recognize that it comes in different forms. Rocky Feuer offered a three-level classification. The first type, which he referred to as stochastic variation, arises from inherent randomness in disease processes and human behavior across the members of a population. Put simply, not everyone's disease follows the same course and screening and treatment do not have the same effects from person to person. Dealing with stochastic variation by itself is straightforward. Modelers would simply specify the known distribution of values for an input parameter. Statistical confidence intervals for model outcomes can be generated through analytic or simulation techniques. Feuer pointed out that the uncertainty discussed by workshop speakers and participants does not fall into this category.

The second level—parameter uncertainty—refers to the far more common situation in which the true values of the parameter, e.g., test sensitivity or the cost of treating CRC, are not well understood. Estimates about the population distributions of model inputs, drawn from medical and epidemiological research, are the “assumptions” on which models are built. Feuer explained that sensitivity analysis is the most appropriate approach to dealing with this kind of uncertainty. In sensitivity analysis, modelers let assumptions vary across a range of likely values and the resulting range of costs and effects is reported. When many parameters are uncertain, as they are in the case of colorectal cancer screening, experts recommend the use of probabilistic cost-effectiveness analysis to generate a bottom-line “confidence interval” or “credible interval” of cost-effectiveness ratios (Briggs et al., 2002; Gold et al., 1996). Such an interval allows users to understand the simultaneous effect of uncertainty about many parameters on the range of cost-effectiveness ratios that result.

The third level of uncertainty is structural, according to Feuer. In that case, researchers may have little to go on about the relationships and interactions among key parameters. They may therefore choose to model those relationships in different, perhaps even arbitrary ways. The debate over whether some cancers arise *de novo* or from pre-existing adenomas is an example of structural uncertainty. To deal with this unknown, some modelers have assumed that such lesions are simply very fast-moving adenomas, while others have assumed that they can never be detected as adenomas. Other examples of structural uncertainty are the effects of including or excluding the consequences of detecting non-adenomatous polyps, the cost of a patient's time engaged in screening, or the impacts on individuals' quality of life from both screening and colorectal cancer. In Feuer's view, the five models highlighted in this workshop represent five different approaches to resolving structural uncertainty.

Strategies for Managing Uncertainty

As workshop participants grappled with how best to deal with the effects of parameter and structural uncertainty on the ability of CEA models to produce the answers policy makers want, many ideas surfaced on how to reduce, or at least manage, those effects.

Research strategies. To reduce the uncertainty about important assumptions, many participants called for more primary research, particularly on those factors that account for the greatest variation among models. The areas most frequently cited were costs and compliance. Laura Seefe outlined work that the CDC is conducting with several states to enhance the utilization of CRC screening. That research should provide more evidence on the degree of compliance that can be expected from different screening program designs. Alan Gerling endorsed more studies of the impact of public awareness programs and other recruitment strategies on adherence, along the lines of those currently underway in pilot studies in the United Kingdom. Michael Pignone suggested that research aimed at getting better estimates of the lifetime cost of treating colorectal cancer might do more to resolve differences among models than would research on other parameters.

Another approach mentioned by several workshop participants to help understand the effect of uncertainty is to evaluate model predictions against independent results from well-designed trials of screening programs. The presentations by leaders of the five research teams showed that several have used data from large fecal occult blood testing screening trials to evaluate the extent to which their models' predictions of cancer incidence and mortality over time agree with the results found in the trials. The ongoing PLCO trial (Schoen et al., 2003; Gohagan et al., 1995), which is testing sigmoidoscopy screening, will soon provide a new dataset to support validation, according to Robert Schoen. NCI's CISNET program acts as a catalyst for sharing useful databases from NCI-sponsored studies among member research teams, said Rocky Feuer.

Karen Kuntz reminded the workshop participants that the paucity of data on important assumptions often leads model builders to use data from trials to inform their choice of values for critical assumptions, such as the sensitivity and specificity of screening tests. She warned that validating a model with data that were also used in part to build model assumptions does not provide a true test of the validity of model predictions.

Short of evaluating the predictive validity of models with independent data sources, research teams can assess other measures of validity,¹⁴ such as whether a model contains all of the components of cost and effect that one would expect to be important. For ex-

¹⁴ For a description of different kinds of validity, see the research methods web page maintained by William Trochim at <http://www.socialresearchmethods.net/kb/index.htm> (Trochim, 2004).

ample, Reid Ness's presentation of his team's reanalysis of the pre-workshop exercise revealed that leaving out the cost of working up non-adenomatous polyps can have a large effect on model outcomes. Louise Russell mentioned that an important component of cost omitted in all of the CRC models presented at the workshop is the value of patients' time lost in screening. That omission biases model results toward screening technologies that are time intensive for the patient and against technologies that are fast and convenient. Martin Brown explained that with no published empirical studies of this cost component modelers typically exclude it.

Several participants expressed skepticism that either research approach—primary research on uncertain assumptions or excluded components or greater availability of independent data sets for model validation—will fully resolve uncertainty, in part because of the cost of generating new information but also because technological advances in screening and treatment continually create new unknowns. In technical terms, the confidence or credible interval for the cost-effectiveness of one strategy is likely to overlap that of others. Recognizing this reality, they suggested steps that might help decision makers make more appropriate use of the information that CEAs can generate.

User strategies. One line of thinking expressed at the workshop was that policy makers have little choice but to accept the discrepant results from models because those results simply reflect the lack of medical and epidemiological evidence. Policy makers could adopt a message that focuses on the value of CRC screening in general, leaving specific choices of strategies to physicians and patients. Robert Dittus suggested that an appropriate message for providers might be, "here is a collection of approaches that we think are good and they all fit within the general realm of 'it's a whole lot better than nothing.'" Robert Smith, on the other hand, argued for opinion leaders to advocate a practical screening strategy that offers the greatest protection and best outcomes for patients given what we know today.

Others noted that colorectal cancer screening involves high cost as well as great medical benefits, so choosing one strategy over another can mean differences of billions of dollars and hundreds of thousands of years of life when summed over the entire U.S. population 50 years of age and older. Although Richard Lilford observed that "the complexity of choice is so great in medicine that the questions are unanswerable," he also recognized that decisions must nevertheless be made based on the best information available. In his view, models offer one type of information to assist in those decisions. The ultimate choice of screening strategy, according to Lilford, will be influenced by political pressures and preferences as well as by models laying out costs and effects as best they can.

Some participants addressed ways to help policy makers better understand the levels of uncertainty represented in models. Judith Wagner commented that editors of medical journals can play a useful role in this regard. The pre-workshop exercise revealed, for example, how uneven and sometimes vague the descriptions in published papers were of models' assumptions about compliance and diagnostic follow-up protocols. Clear descriptions of the assumptions in these and other important areas should be a priority. Wagner also observed that published CEAs of CRC screening have often evaluated a single screening strategy not examined in published work by other modeling teams. That practice makes it difficult for readers to assess the level of agreement across models. Authors seeking to evaluate the cost-effectiveness of a new screening technology might be asked to report on the model's outcomes for a common set of well studied screening strategies, such as the five strategies used in the pre-workshop exercise. Decision makers

could then assess whether the cost-effectiveness results for the new technology might differ if assessed by other models.

Some participants called for research teams to provide more access to their models, even suggesting that models be placed in the public domain on the Internet to allow decision makers to test the impact of different assumptions or strategies. Others pointed out, however, that models entail a substantial investment in researchers' intellectual capital, which could be compromised by open access. Rocky Feuer suggested as a middle ground the Model Profiler currently under development as part of the CISNET program. The Profiler, described in Feuer's presentation, is expected to provide open web-based access to detailed information on model structure, assumptions and outputs, but not to the models themselves.

Modeling Reality or an Ideal World?

Another issue that threaded its way through the discussion concerned which of the following two questions CRC screening models should seek to answer: "What can be expected to happen under a given strategy?" Or "What could happen if the strategy is implemented under ideal conditions?"

The modeling of compliance is an obvious example of this issue. Perfect adherence to a strategy, including the periodic screening examinations, the specified follow-up, and surveillance protocols is the ideal, but it is not achieved in practice, as Sally Vernon showed in her presentation.

Another example of the tension between modeling the ideal and modeling reality is how models handle diagnostic follow-up of a positive sigmoidoscopy. Some experts recommend that polyps found on sigmoidoscopy be biopsied or removed during that procedure, with referral for a full colonoscopy only if the polyp is found to be a high-risk or advanced adenoma. But Todd Anderson's presentation on the frequency of follow-up and surveillance procedures suggests that the vast majority of polyps removed from patients undergoing sigmoidoscopy are removed in a subsequent colonoscopy. Whether models are based on recommended or actual practice in this regard could affect the costs and effects of sigmoidoscopy.

Several participants noted that limits on the supply of screening procedures represent an area in which reality may force departures from ideal conditions. Certain procedures— notably colonoscopy, but in the future, perhaps, virtual colonoscopy—require trained specialists to perform or interpret them. Like most people, medical specialists respond to economic incentives and higher reimbursements for screening or surveillance procedures would induce them to do more. However, some strategies may require so many colonoscopies that the supply of gastroenterologists would be completely inadequate.¹⁵ The same may be true of radiologists, should virtual colonoscopy become a routine screening procedure (Herdman and Norton, 2005). The supply of specialists cannot be expanded quickly, so real constraints on capacity may have to be taken into account. Sandeep Vijan observed that assuming low compliance for certain screening or surveillance procedures is one way models could implicitly account for such constraints.

Seth Glick emphasized the divergence between test performance under ideal quality assurance programs and test performance in current practice. The quality of many screening processes may be poor today, in Glick's view. Estimates of test sensitivity, specificity, and medical risk, usually taken from studies where good quality assurance existed,

¹⁵ Recent news accounts suggest that waiting times for colonoscopy are growing in certain areas of the country (Kowalczyk, 2004). New evidence also suggests that some surveillance colonoscopies are being done more frequently than is recommended by professional guidelines (Mysliwiec et al., 2004).

therefore overestimate the performance of screening in the absence of strong quality assurance programs.

The basic problem, in the view of Martin Brown, is not the choice between the real and the ideal *per se*, but the failure of researchers to make explicit their choices about it. Moreover, their implicit choices may vary within a model, with optimal assumptions in one area and realistic assumptions in another and no explicit rationale for the differences. He and several other participants argued that both kinds of analysis are useful. For example, analysts could tell patients and physicians what can be expected if consumers fully comply with a strategy, and then show the decrements in effects and costs resulting from less complete compliance. Michael Pignone cautioned that researchers who model ideal conditions need to include the full costs of achieving those conditions. Quality assurance and high rates of compliance do not come free. They are usually the result of intensive programs of behavior change that must continue for the duration of a screening program.

How Complex Should Models Be?

Participants returned repeatedly throughout the workshop to the question of whether models should be capable of evaluating complex screening strategies. This question surfaced often because the workshop participants, including the modelers, are fundamentally interested in the health policy question—what screening strategy is best?—not in modeling for its own sake. When the high lifetime costs of some very effective screening strategies become apparent in all models, a natural next step is to explore how those costs could be reduced, without compromising effectiveness, by fine-tuning strategies. Such fine-tuning drives modelers to add more branches to their strategies, which places even greater demands on the clinical and epidemiological evidence available to support such modeling.

Many ideas for complex screening strategies were put forward at the workshop. Participants suggested strategies such as changing the screening, follow-up, or surveillance schedule as a person ages, offering different screening strategies to different demographic groups with different relative risks, or changing a schedule contingent on the results of a previous screening, follow-up, or surveillance test. For example, Ann Zauber observed that men and women have different profiles of adenoma and CRC incidence, with women developing CRC an average of 10 years later than men do (Chu et al., 1994; Cooper et al., 1995; Devesa and Chow, 1993). Different screening strategies for men and women might make sense and could be explored by models. Donatus Ekwueme raised similar possibilities for tailoring strategies by race in recognition of the systematic racial differences in incidence, prevalence, and location in the colon of adenomas and polyps (Devesa and Chow, 1993; Theuer et al., 2001; Walker Jr et al., 1995). Michael Pignone and Marjolein van Ballegooijen suggested that it might make sense to alter the type of screening test as a person ages, saving more sensitive but more expensive tests until the individual is at higher risk of advanced adenomas or CRC.

To John Inadomi, the most important clinical question is whether surveillance following polypectomy is cost-effective and how often it is needed. He and Reid Ness argued that selective post-polypectomy surveillance strategies—where high-risk individuals are monitored more often than those at low-risk of future polyps or cancer—have the greatest potential to reduce costs. But, to make such judgments without imperiling outcomes, accurate data on the factors that matter are needed. Deborah Schrag summarized the results of the National Polyp Study (Winawer et al., 1993), which found that intensive surveillance strategies offer little additional benefit compared with protocols that condi-

tion future surveillance on the nature of the polyp removed and on the result of the first surveillance test.

Mark Fendrick raised the possibility of adjusting screening strategies to account for individuals who had already received a colonoscopy for symptoms or non-screening reasons. David Lieberman commented that in reviewing a large endoscopic database he found 40 to 50 percent of all colonoscopies were performed either for vague symptoms or for rectal bleeding. The results of most of those procedures either are negative or show benign polyps. If models were adjusted to assume that 40 to 50 percent of individuals have already been screened before a formal screening program starts, rather than assuming that no one has been screened, as they currently do, the predicted cost of screening would be lower.

Despite the enthusiasm for complex strategies (and for estimating their potential to save costs or increase effectiveness), several participants sounded notes of caution. Mark Fendrick emphasized barriers to making complex screening strategies operational. He described the difficulty one major medical center had in providing same-day follow-up colonoscopy for people with positive screening sigmoidoscopic examinations, even when those patients had been clinically prepped beforehand for a possible colonoscopy. He and Sandeep Vijan also warned that presenting too many options could overwhelm patients and ultimately reduce their willingness to participate in screening or surveillance. Robert Dittus held out hope that new medical information systems, such as automated test ordering and electronic medical records with built-in guidelines, will make it easier for physicians to implement complex strategies.

Several participants held that if complex strategies offer substantial hope for moderating costs without reducing the benefits of screening, then models should stand ready and be capable of assessing them. But, argued Amnon Sonnenberg, given the information requirements of complex models, we may be expecting models to do too much. Sometimes if models become too complex, they go off the mark simply because they must make too many assumptions based on too little evidence. Thus, the discussion of complexity ended with a reprise of the first problem for modeling, as for decision making in general: uncertainty.

NEXT STEPS

The workshop was not intended to evaluate the effectiveness or cost-effectiveness of colorectal cancer screening in general. Virtually all economic models, drawing from a wealth of clinical trials and epidemiological studies, have found that colorectal cancer screening decreases mortality from the disease. The message to physicians, payers, and patients that periodic screening for colorectal cancer is an effective preventive measure continues to have urgency.

The focus of the workshop was not on “whether to screen” but on “how to screen.” Nevertheless, it was NOT intended to evaluate alternative CRC screening strategies. No evidence was presented to recommend one strategy over others. Instead, the purpose was to explore and enhance the usefulness of cost-effectiveness models in helping medical policy makers make such judgments. An obvious next step would be for modelers and clinicians to continue to explore together how the factors affecting model outcomes that were identified in this workshop—both parameter assumptions and structural assumptions—can be resolved through better information or better modeling. Such explorations take time and resources beyond those available to a one-time workshop. The workshop did show, however, that collaboration can identify critical sources of variation—such as

assumptions about the cost of treating disease—that lead to conflicting findings. Those differences confuse decision makers, who must grapple with the underlying question of “which is the best strategy?”

Models are like maps. Maps are useful when they serve as guides to underlying territories. A map that is too vague is useless; one that is completely accurate merges with the territory itself and is also useless as a guide. The participants spent considerable time discussing the optimal balance for models along the continuum from rough guide to complete accuracy. They struggled with questions of how detailed CRC models should be if they are to be useful to decision makers and how detailed they can be, given the available information.

Richard Lilford provided valuable perspective with his observation that “modeling is a way of having a conversation.” That is precisely what occurred during the day and a half when modeling teams and experts came together to compare assumptions, results, and the underlying evidence base for modeling. Many participants commented on the value of the conversation for further refinement of their models (in the case of the research teams) and for research ideas (in the case of clinical and epidemiological researchers). The pre-workshop modeling collaboration demonstrated that too many lives and dollars are at stake not to continue to work on understanding and communicating both the strengths and weaknesses of cost-effectiveness models.

REFERENCES

- Briggs AH, O'Brien BJ, Blackhouse G. 2002. Thinking outside the box: Recent advances in the analysis and presentation of uncertainty in cost-effectiveness studies. *Annu Rev Public Health* 23: 377-401.
- Chu KC, Tarone RE, Chow WH, Hankey BF, Ries L. 1994. Temporal patterns in colorectal cancer incidence, survival, and mortality from 1950 through 1990. *J Natl Cancer Inst* 86(13): 997-1006.
- Cooper GS, Yuan Z, Landefeld CS, Johanson JF, Rimm AA. 1995. A national population-based study of incidence of colorectal cancer and age: Implications for screening in older Americans. *Cancer*. 75(3): 775-781.
- Cotton PB, Durkalski VL, Pineau BC, Palesch YY, Mauldin PD, Hoffman B, Vining DJ, Small WC, Affronti J, Rex D, Kopecky KK, Ackerman S, Burdick JS, Brewington C, Turner MA, Zfass A, Wright AR, Iyer RB, Lynch P, Sivak MV, Butler H. 2004. Computed tomographic colonography (virtual colonoscopy): A multicenter comparison with standard colonoscopy for detection of colorectal neoplasia. *JAMA* 291(14): 1713-1719.
- Curry S, Byers T, Hewitt M (eds). 2003. *Fulfilling the Potential of Cancer Prevention and Early Detection*. National Cancer Policy Board Institute of Medicine. Washington DC: The National Academies Press.
- Devesa SS, Chow WH. 1993. Variation in colorectal cancer incidence in the United States by subsite of origin. *Cancer* 71(12): 3819-3826.
- Eddy DM. 1980. *Screening for Cancer: Theory, Analysis and Design*. Englewood Cliffs, NJ: Prentice-Hall.
- Eddy DM, Nugent FW, Eddy JF, Collier J, Gilbertsen V, Gottlieb LS, Rice R, Sherlock P, Winawer S. 1987. Screening for colorectal cancer in a high-risk population. Results of a mathematical model. *Gastroenterol* 92(3): 682-692.
- Edwards BK, Howe HL, Ries LAG, Thun MJ, Rosenberg HM, Yancik R, Wingo PA, Jemal A, Feigal EG. 2002. Annual report to the nation on the status of cancer, 1973-1999, featuring implications of age and aging on U.S. cancer burden. *Cancer* 94(10): 2766-2792.
- Frazier AL, Colditz GA, Fuchs CS, Kuntz KM. 2000. Cost-effectiveness of screening for colorectal cancer in the general population. *JAMA* 84(15): 1954-1961.
- Glick S, Wagner JL, Johnson CD. 1998. Cost-effectiveness of double-contrast barium enema in screening for colorectal cancer. *AJR AM JRoentgeno* 170(3): 629-636.
- Gohagan JK, Prorok PC, Kramer BS, Hayes RB, Cornett JE. 1995. The prostate, lung, colorectal, and ovarian cancer screening trial of the National Cancer Institute. *Cancer* 75(7 Suppl.): 1869-1873.
- Gold MR, Siegel J, Russell L, Weinstein, M. 1996. *Cost-Effectiveness in Health and Medicine* New York: Oxford University Press.
- Herdman R, Norton L (eds). 2005. *Saving Women's Lives: Strategies for Improving Breast Cancer Detection and Diagnosis*. National Cancer Policy Board, Institute of Medicine. Washington DC: The National Academies Press.
- Imperiale TF, Wagner DR, Lin CY, Larkin GN, Rogge JD, Ransohoff DF. 2003. Using risk for advanced proximal colonic neoplasia to tailor endoscopic screening for colorectal cancer. *Ann Intern Med* 139(12): 959-965.
- Jorgensen OD, Kronborg O, Fenger C. 2002. A randomised study of screening for colorectal cancer using faecal occult blood testing: Results after 13 years and seven biennial screening rounds. *Gu*. 50(1): 29-32.
- Joseph AM, Crowson TW, Rich EC. 1988. Cost effectiveness of HemoQuant versus Hemoccult for colorectal cancer screening. *J Gen Intern Med* 3(2): 132-138.
- Khandker RK, Dulski JD, Kilpatrick JB, Ellis RP, Mitchell JB, Baine WB. 2000. A decision model and cost-effectiveness analysis of colorectal cancer screening and surveillance guidelines for average-risk adults. *Int J Technol Assess Health Care* 16(3): 799-810.

- Kowalczyk L. 2004, August 1. Colonoscopies tax health industry. *The Boston Globe* A-1.
- Ladabaum U, Song K, Fendrick AM. 2004. Colorectal neoplasia screening with virtual colonoscopy: When, at what cost, and with what national impact? *Clinical Gastroenterology & Hepatology* 2(7): 554-563.
- Lieberman DA. 1995. Cost-effectiveness model for colon cancer screening. *Gastroenterol* 109(6): 1781-90.
- Lieberman DA, Prindiville S, Weiss DG, Willett W. 2003. Risk factors for advanced colonic neoplasia and hyperplastic polyps in asymptomatic individuals. *JAMA* 290(22): 2959-2967.
- Lieberman DA, Prindiville S, Weiss DG, Willett W, Imperiale T. 2004. Risk factors for advanced colonic neoplasia: From evidence to application. *Evidence-Based Gastroenterology* 5(3): 86-87.
- Loeve F, Boer R, Van Oortmarssen GJ, Van Ballegooijen M, Habbema JDF. 1999. The MISCAN-COLON simulation model for the evaluation of colorectal cancer screening. *Comput Biomed Res* 32(1): 13-33.
- Loeve F, Brown ML, Boer R, Van Ballegooijen M, Van Oortmarssen GJ, Habbema JDF. 2000. Endoscopic colorectal cancer screening: A cost-saving analysis. *J Natl Cancer Inst* 92(7): 557-563.
- Mandel JS, Church TR, Ederer F, Bond JH. 1999. Colorectal cancer mortality: Effectiveness of biennial screening for fecal occult blood. *J Natl Cancer Inst* 91(5): 434-437.
- Mysliwiec PA, Brown ML, Klabunde CN, Ransohoff DF. 2004. Are physicians doing too much colonoscopy? A national survey of colorectal surveillance after polypectomy. *Ann Intern Med* 141(4).
- Neilson AR, Whynes DK. 1995. Cost-effectiveness of screening for colorectal cancer: A simulation model. *IMA J Math Appl Med Biol* 12(3-4): 355-367.
- Ness RM, Holmes AM, Klein R, Dittus R. 2000. Cost-utility of one-time colonoscopic screening for colorectal cancer at various ages. *Am J Gastroenterol* 95(7): 1800-1811.
- Pickhardt PJ, Choi JR, Hwang I, Butler JA, Puckett ML, Hildebrandt HA, Wong RK, Nugent PA, Mysliwiec PA, Schindler WR. 2003. Computed tomographic virtual colonoscopy to screen for colorectal neoplasia in asymptomatic adults. *N Engl J Med* 349(23): 2191-2200.
- Pignone M, Saha S, Hoerger T, Mandelblatt J. 2002. Cost-effectiveness analyses of colorectal cancer screening: A systematic review for the U.S. Preventive Services Task Force. *Ann Intern Med* 137(2): 96-104.
- Ransohoff DF. 2004. Virtual colonoscopy: What it can do vs what it will do. *JAMA* 291(14): 1772-1774.
- Schoen RE, Pinsky PF, Weissfeld JL, Bresalier RS, Church T, Prorok P, Gohagan JK. 2003. Results of repeat sigmoidoscopy 3 years after a negative examination. *JAMA* 290(1): 41-48.
- Scholefield JH, Moss S, Sufi F, Mangham CM, Hardcastle JD. 2002. Effect of faecal occult blood screening on mortality from colorectal cancer: Results from a randomised controlled trial. *Gut* 50(6): 840-844.
- Smith RA, Cokkinides V, Eyre HJ. 2004. American Cancer Society guidelines for the early detection of cancer, 2004. *CA Cancer J Clin* 54(1): 41-52.
- Song K, Fendrick AM, Ladabaum U. 2004. Fecal DNA testing compared with conventional colorectal cancer screening methods: A decision analysis. *Gastroenterol* 126(5): 1270-1279.
- Sonnenberg A, Delco F. 2002. Cost-effectiveness of a single colonoscopy in screening for colorectal cancer. *Arch Intern Med* 162(2): 163-168.
- Sonnenberg A, Delco F, Inadomi JM. 2000. Cost-effectiveness of colonoscopy in screening for colorectal cancer. *Ann Intern Med* 133(8): 573-584.
- Theuer CP, Wagner JL, Taylor TH, Brewster WR, Tran D, McLaren CE, Anton-Culver H. 2001. Racial and ethnic colorectal cancer patterns affect the cost-effectiveness of colorectal cancer screening in the United States. *Gastroenterol* 120(4): 848-856.
- Trochim WM. 2004. The Research Methods Knowledge Base. 2nd ed. [online] <http://www.socialresearchmethods.net/kb/index.htm>.
- United States General Accounting Office. 2004. Coverage of Common Colorectal Cancer Screening Tests Is Common. Washington DC: United States General Accounting Office.

- U.S. Preventive Services Task Force. 2002. Screening for colorectal cancer: Recommendation and rationale. *Ann Intern Med* 137(2): 132-141.
- Vijan S, Hwang EW, Hofer TP, Hayward RA. 2001. Which colon cancer screening test? A comparison of costs, effectiveness, and compliance. *Am J Med* 111(8): 593-601.
- Wagner JL, Herdman RC, Wadhwa S. 1991. Cost effectiveness of colorectal cancer screening in the elderly. *Ann Intern Med* 115(10): 807-817.
- Wagner J, Tunis S, Brown M, Ching A, Almeida R. 1996. Cost-Effectiveness of Colorectal Cancer Screening in Average-Risk Adults. In: Young G, Rosen P, Levin B (eds). *Prevention and Early Detection of Colorectal Cancer*. Philadelphia: Saunders. Pp. 321-356.
- Walker Jr B, Figgs LW, Zahm SH. 1995. Differences in cancer incidence, mortality, and survival between African Americans and whites. *Environ Health Perspect* 103(Suppl. 8): 275-281.
- Whynes DK. 2004. Cost-effectiveness of screening for colorectal cancer: Evidence from the Nottingham faecal occult blood trial. *J Med Screen* 11(1): 11-15.
- Winawer S, Fletcher R, Rex D, Bond J, Burt R, Ferrucci J, Ganiats T, Levin T, Woolf S, Johnson D, Kirk L, Litin S, Simmang C. 2003. Colorectal cancer screening and surveillance: Clinical guidelines and rationale: Update based on new evidence. *Gastroenterol* 124(2): 544-560.
- Winawer SJ, Zauber AG, May Nah Ho, O'Brien MJ, Gottlieb LS, Sternberg SS, Wayne JD, Schapiro M, Bond JH, Panish JF, Ackroyd F, Shike M, Kurtz RC, Hornsby-Lewis L, Gerdes H, Stewart ET, Lightdale CJ, Edelman M, Fleisher M. 1993. Prevention of colorectal cancer by colonoscopic polypectomy. *N Engl J Med* 329(27): 1977-1981.