SCREENING FOR ABDOMINAL AORTIC ANEURYSMS:
TIME-BASED MODELING FOR PUBLIC POLICY

Dennis G. Fryback
Department of Preventive Medicine
University of Wisconsin-Madison
Madision, Wisconsin 53796

Paul S. Frame
Tri-County Family Medicine Program
Park Avenue, Box 112
Cohocton, New York 14862

ABSTRACT

A deterministic, time-based model to evaluate cost effectiveness of programs to screen for abdominal aortic aneurysms (AAA) is presented. At its present complexity it has been used to inform debate by the US Preventive Services Task Force about recommended screening policies. Such models could be formulated in much more realistic (and complex) manner. However even in present form it pushes beyond the limits of data-based epidemiologic knowledge. Most simulation models in health care are based on poorly known parameters, often extracted secondarily from published studies. Given that these models are becoming influential in public policy, it is argued that the simulation and modeling community needs to take seriously the need for new thinking about validation, presentation, and review of complex models.

1 INTRODUCTION

Prevention of serious disease and its consequences is an important part of public health policy (Public Health Service, 1980; 1990). However controversy exists about which preventive interventions are in fact effective (having demonstrable benefits outweighing possible harms). Professional organizations (*e.g.*, American College of Physicians) and private interest organizations (*e.g.*, American Cancer Society, American Heart Association) have issued sometimes conflicting recommendations and guidelines.

Recommendations about screening, one particular type of prevention activity, can be particularly controversial -- and difficult to evaluate. Screening is defined as "the examination of asymptomatic people in order to classify them as likely, or unlikely, to have the disease that is the object of screening" (Morrison, 1985, p. 3). Because they can affect large segments of the population, policy decisions to recommend or fund screening programs can involve considerable expense as well as potential benefit.

There are two main pre-requisites for screening to be deemed effective. First is that treatment of the disease in an early phase (before it would normally come to attention) is more beneficial than later treatment. Second is that diagnostic evaluation of people with the disease and of people without the disease, but whom might be erroneously identified by the screening test, does not have side effects outweighing the benefits of early treatment. Because of many inherent uncertainties about disease processes, and testing and treatment for diseases, the only secure experimental evidence of benefit from screening must be derived from prospective, randomized controlled trials. But conduct of such experiments may be expensive (10's or even 100's of millions of dollars), time-consuming (10 years or more is not unusual), and administratively not feasible. Thus definitive data most often does not exist and evaluation is based on bits and pieces of evidence from many places in the literature and anecdotal experience of practitioners.

The US Preventive Services Task Force (USPSTF) was commissioned by the Office of Disease Prevention and Health Promotion of the Assistant Secretary for Health, US Department of Health and Human Services, to assess the evidence for effectiveness of preventive interventions and to make recommendations about use of specific preventive services by primary care physicians in the US. The USPSTF published a book collecting their assessments of the effectiveness of 169 interventions (USPSTF, 1989). Although many individual members have changed, the USPSTF continues work and anticipates publishing a revised edition in 1995. The work reported here was undertaken in support of ongoing USPSTF deliberations.

Although the USPSTF does not explicitly consider the costs of screening in making its recommendations, costs must ultimately be weighed at least implicitly by

public policy makers (e.g., in Health Care Financing Administration) and others (e.g., large insurance companies) who must make decisions about payment for screening programs.

1.1 Abdominal Aortic Aneurysm

Abdominal aortic aneurysm (AAA) is a potentially lethal problem which can be detected in an early phase, and for which a curative treatment is available; therefore it might be a suitable subject for a screening program (Lederle, 1990). Because a screening program for AAA might represent a considerable investment in health care we constructed a model to estimate potential benefits and costs of screening for AAA.

The abdominal aorta is the major vessel supplying arterial blood to the entire body below the lungs. An AAA is a localized abnormal dilation of this artery usually occurring below the takeoff of the renal arteries and around or above the bifurcation of the abdominal aorta into the femoral arteries. The natural time course of AAA is characterized by intermittent slow, asymptomatic expansion. Symptoms, if they occur, are generally related to pressure of the enlarged AAA on other structures or due to embolization from blood clot formed within the lumen of the AAA.

The threat of AAA is rupture, which is almost uniformly fatal without emergency surgery. Rupture of AAA causes 1.2% of male deaths and 0.6% of female deaths in persons over 65 years of age in the United States.

AAA is an age-related disease, with higher incidence in older persons, and in males. Other risk factors appear to be tobacco smoking, coexisting vascular disease, and familial predisposition. The prevalence of AAA in older men in one of the larger studies to report prevalence is 5.4% (Collin et al. 1988). The rate of rupture for AAAs over 4 cm. in diameter is directly related to size; overall 3-6% of AAAs over 4 cm. will rupture annually (Nevitt, Ballard, and Hallett, 1989; Cronenwett et al 1985). Generally a patient in whom an AAA > 4 cm. is found will be offered elective surgery to prevent rupture; although this threshold has been debated in more recent literature, existing data are reported only for the 4 cm threshold. Three to five percent of initially small AAAs (< 4 cm.) will progress to potentially operable size each year (Bengtsson et al. 1989; Collin, Heather, and Walton, 1991).

Two tests have been proposed as screening tests. Abdominal ultrasound (US) has reported sensitivity (probability of positive result given AAA present)

from 82-99%, with specificity (probability of negative result given no AAA) approaching 100% (Canadian Task Force on Periodic Health Examination, 1991). The current charge for US is about $150. Physical examination (PE) is less accurate; Lederle et al. (1988) report a sensitivity of 50% and specificity of 96%. Detection rate varied with size of the AAA; 80% of AAAs larger than 5 cm. diameter were detected by PE. Charges for PE that could be apportioned to palpation for AAA might reasonably range from $5 to $30.

Finally, the cost of surgery for AAA varies depending on its urgency. The charge for elective surgery averages $27,000 for the episode of care (Breckwoldt, Mackey, and O'Donnell, 1991). Emergency surgery is estimated at $52,000.

Modelers not familiar with the medical literature should realize that all the numbers above are ill-known at best. They have been collected for many different purposes, under varying conditions on two continents, in three different health care systems, and in heterogeneous population subgroups. Further, they are often reported in highly summary fashion leaving much to the imagination. Contrary to many industrial simulations, it is rarely possible to collect new data with which to instantiate a model. Still, the public policy questions are real and need to be addressed.

1.2 Screening recommendations

The current recommendation of the Canadian Task Force on Periodic Health Examination (CTF) is: "There is poor evidence to include screening through physical examination or ultrasonography for [AAAs] in or exclude it from the periodic health examination of asymptomatic people." (CTF, 1991) Oboler and La Force (1989) recommend "abdominal examination ... be done yearly in all men older than 60 years." Bengtsson et al. (1989) recommend screening men at ages 60, 67, and 74 years of age by abdominal ultrasound.

2 THE MODEL

We used a computer "spreadsheet" program (Quattro Pro 4.0, ©1991 Borland International, Inc.) to simulate possible AAA screening programs for a cohort of 10,000 men aged 60-79 years at the initiation of screening and followed over a period of 21 years. (The size of the cohort was picked for convenience only.) Although the model is computationally deterministic in its present form, it is designed as state-transition model to represent time-based events occurring to the cohort over time, and thus may be

deemed a simulation model. The purpose of this model was to integrate the diverse information presented above into a unitary computational "laboratory" with which to explore ramifications of differing assumptions and program structures.

Table 1 shows parameters needed for the model.

Table 1
Variables in the Model

Variable	base case	range
• AAA prevalence	5.4%	2-7.8%
• annual incidence of new AAA	0.16%	0.1%-0.16%
• % patients who choose elective surg.	70%	60-90%
• emergency surg. mortality	50%	21-66%
• elective surg. mortality	5%	1-12.5%
• AAA ≥4cm rupturing annually	4%	3-6%
• AAA < 4cm rupturing annually	0%	0%
• % AAA < 4cm expanding to ≥ 4cm annually	6%	3-10%
• % AAA diagnosed annually in absence of screening	33%	20-50%
• sensitivity of US	100%	100%
• specificity of US	100%	100%
• sensitivity of PE	50%	22-96%
• specificity of PE	96%	81-96%
• cost of US	$150	$100-175
• cost of PE	$20	$5-30
• cost of elective surgery	$27000	$12000-$27000
• cost of emergency surgery	$52000	$17500-52000
• annual, all-cause age-specific mortality rate	US Vital statistics tables	-
• annual discount rate for present value calculations	5%	-

The initial distribution of ages in the cohort was presumed to be that of the male US population in this age range:

60-64 years	--	35%
65-69 years	--	29%
70-74 years	--	22%
75-79 years	--	14%

The cohort is presumed to be unscreened until the first year of the simulation. The model simulates several processes affecting this cohort simultaneously over time: onset and growth of AAAs, clinical surfacing of AAAs in the absence of screening, rupture and death from AAAs, aging and death from other causes than AAA rupture or surgical complication, and case finding by screening, and repair by elective or emergency surgery.

Members of the cohort are considered to be in one of seven states at any given time: (1) Alive without AAA, (2) Having an AAA <4cm in diameter that has not been discovered, (3) Having an AAA ≥4 cm in diameter that has not been discovered, (4) Having an AAA <4cm which has been discovered, (5) Having an AAA ≥4 cm, previously discovered when <4cm and not re-examined since, (6) Having an AAA ≥4 cm, known to the medical system, (7) Dead.

Because we used deterministic calculations in a simplified model, these processes were represented as sequential events occurring each year in the cohort. The status of men in the cohort is represented as transitions among the seven states. The sequence in each year is presumed as follows:

Day 1. If this is a year in which screening is done, all men without a previously discovered AAA will be screened on this day; elective surgery is offered to each in which an aneurysm of 4 cm or larger is found. Screening moves men from states 2 and 3 to states 4 and 6. Elective surgery is presumed to return the patient to state 1 (no AAA) if successful, or to state 7 (dead) if not. Not all men will elect to undergo surgical repair; the proportion declining is a parameter in the model.

Day 2. All previously discovered AAA's (except those found on screening yesterday) are examined in an office visit and with ultrasound. Any men who are found to now have an AAA 4 cm or larger are offered elective surgery. Those that elect repair may be returned to the normal state or die of operative complications.

Day 3. On this day all interval cases that would surface during the year by means other than screening do in fact surface. (We will term this process "interval case finding" to differentiate it from cases found by

screening.) Elective surgery is offered to those with AAAs 4 cm or larger, as above.

Day 4. Any AAAs that will rupture during the year do so today. Some proportion of these arrive at a hospital in time for emergency surgery, the remainder die. Those undergoing emergency surgery may be cured or die.

Day 364. All mortality from other causes happens on this penultimate day of the year. The fraction

dying is computed using age-specific mortality rates for US males.

Day 365. All new AAAs occur on this day. Some fraction of existing AAAs that were less than 4 cm grow to 4 cm or larger (i.e., these men transition from state 2 to 3 or state 4 to 5).

At each step any costs that are generated are computed and discounted to present value using a discount rate set as a parameter of the program (5%
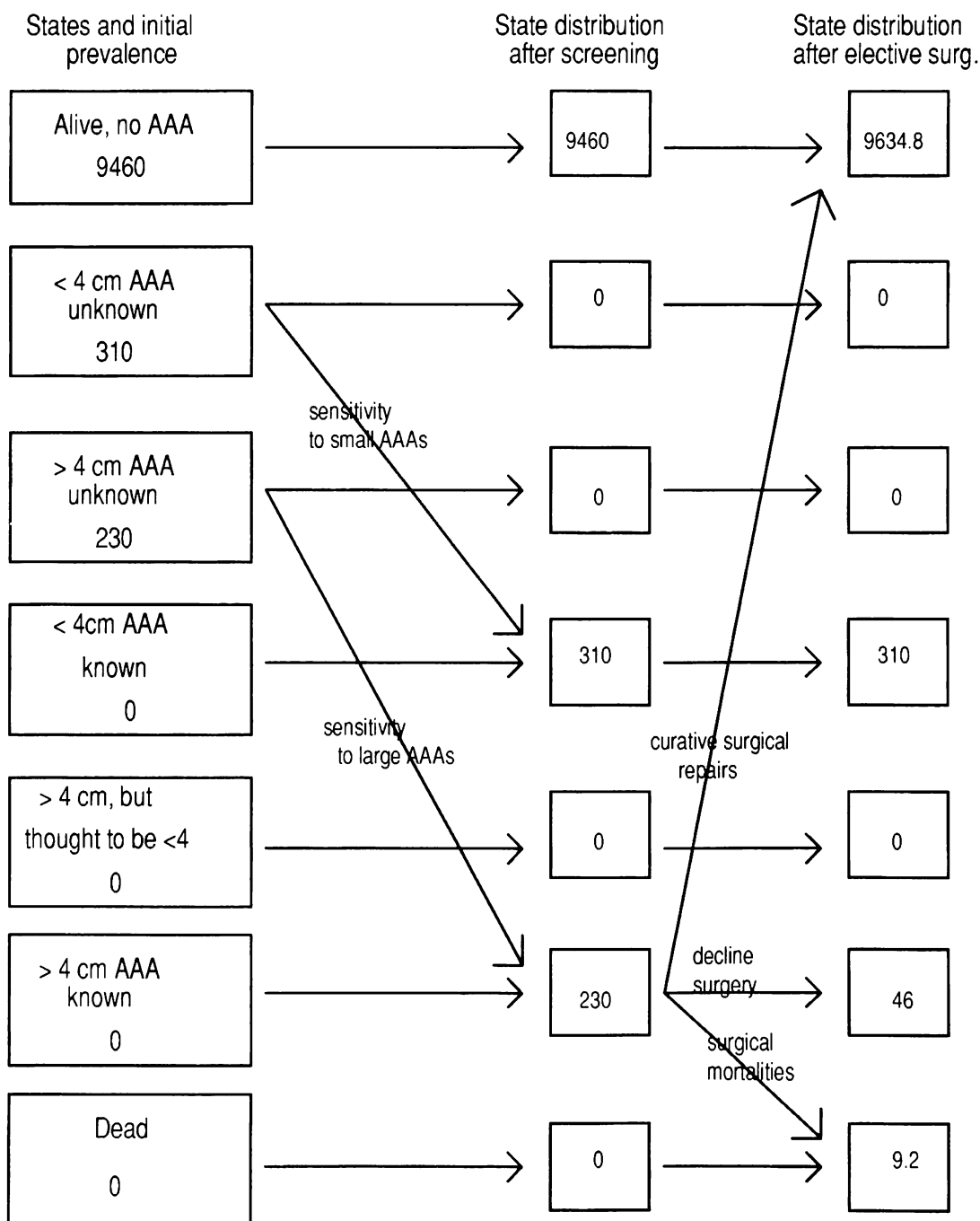

Figure 1: Initial States and Example Transitions in the AAA Model

Table 2
Incremental Cost per LY Gained for Various Protocols

Row	Protocol	All model variables set at base case values	All model variables at best for screening	All variables set at least favorable for screening
1.	"do-nothing" baseline	-	-	-
2.	emergency repair only (compared to Row 1.*)	$15,537	$32,070	$3,319
3.	emergency repair for ruptures + elective surg. for incidentally discovered cases (compared to Row 2.)	$27,035	$4,704	loses life years
4.	One-time US screen (compared to Row 3.)	$41,197	$10,776	loses life years
5.	US first year, then at 5th year (compared to Row 4.)	$1,596,069	$100,803	loses life years
6.	One-time PE→US (compared to Row 3.)	$25,062	$8,724	loses life years
7.	PE→US in year 1 and at 5th year (Compared to Row 6.)	$142,782	$22,802	loses life years

*Incremental $/LY were computed for this table by taking the difference between total dollar expenditures for the program in a given row and the total dollar expenditure in the comparison row, then dividing this difference by the analogous difference for total LYs accumulated by the cohort.

for results reported here). At the end of each simulated year survivors and those dying of competing causes at year's end each accumulate one life-year (LY). These are discounted to present value at the same discount rate. The program can weight LYs by a quality factor to account for living with a known AAA; however differential weighting was not used for the results reported here.

Figure 1 illustrates initial steps in the simulation. In the left column is shown the seven states. The initial prevalence of men with AAAs of the two sizes is determined by input parameters for the model. As shown in fig. 1, the cohort is screened at the outset of the first year. Because ultrasound is assumed here to be 100% sensitive and 100% specific no cases of AAA are left in the "undiscovered" states. Among the 230 men discovered with AAAs of 4 cm or larger these calculations show 80% (184) have elected surgery, leaving 46 in the "known large AAA" state. Assuming a 5% elective surgery mortality rate, on average 9.2 of the 184 die, leaving 174.8 who are returned to the "no AAA" state.

To evaluate repeated screening, we used the program to compare a program consisting of an initial

screen, then another screening 5 years later for all men in the cohort who were alive and did not have previously diagnosed and untreated AAAs. This was compared incrementally to the program described above. It gains slightly more than one-half LY in the entire cohort compared to the one-time screen and accrues additional costs of slightly over $900,000. The additional screen at 5 years thus gains LYs at a cost of $1,596,069 per LY (Table 2).

Table 2 displays similar calculations using the parameter values most favorable to screening. If all the parameters in Table 1 most favorable to screening were to be simultaneously true, then the one-time screen with US gains LYs at a rate of $10,776 per LY and the additional screen at five years gains a few more LYs at a rate of $100,803 each. If all the parameter values least favorable to screening were to occur simultaneously a screening program loses slightly more LYs due to elective surgical mortality than it gains by avoiding AAA deaths.

The spreadsheet model is not programmed directly to evaluate screening tests with less than 100% specificity. However we can approximate the value of a screening protocol using physical examination (PE)

by palpation, followed by ultrasound if PE is positive. The combined protocol will have the sensitivity of PE, but the specificity of ultrasound (here presumed to be 100%). The average cost per screened individual will be a function of the costs of PE, the specificity of PE, the prevalence of AAA, and the cost of ultrasound and can be computed directly. Using these values in the simulation we can approximate the incremental cost-effectiveness ratios for screening with this protocol. Table 2 shows these results. As with the case of screening by US, the additional screen at the fifth year yields very little in LYs for the cohort because of the low prevalence of AAA in the cohort after the first screen. Again under worst-case assumptions there is no net gain of LYs for the cohort.

Our calculations are for a cohort sampled from the 60-79 year old age group. Changing the age of the group to the low or high end of this range affects the calculations somewhat. If the cohort were men 60-64 the base case for a one-time screen with US is $31,603/LY. If the cohort were all 75-79 at the first screen this figure is $66,313.

3 DISCUSSION

3.1 Implications of the Model

How much is too much cost in terms of spending dollars to save LYs? A recent review of cost-effectiveness of cholesterol lowering programs notes that there are health intervention programs with public support costing less that $40,000/LY, and that programs costing more than $60,000/LY appear controversial due to their expense. (Goldman et al. 1992) Using our base case assumptions a one-time screening program of PE followed by US if positive falls into the favorable range. One-time screening with US of men 60-79 from the general population has a less favorable cost-effectiveness ratio, bordering the high end of this range. Both ratios become more favorable if screening is targeted to a higher prevalence population such as men with peripheral vascular aneurysms.

A general internist might consider screening men 60-79 with physical examination. Screening with US one time around ages 60-64 might be cost-effective. It appears quite unlikely that a second screen after five years would be cost-effective unless further research indicates a much higher annual incidence rate than suspected to date.

With current data in the literature we are unable to evaluate protocols based on a 5 cm threshold for elective surgery. As better epidemiologic data become

available about the natural history of AAA the model can be reprogrammed to evaluate alternative protocols.

The model appears secure enough to inform decisions about screening beyond an initial screen. Within reasonable assumptions this seems to be very expensive for the return and thus we recommend against. But the issue of whether to do an initial screen falls uncomfortably close to borderline. Under one set of assumptions it appears expensive for the return and under another it appears not unreasonable.

A recommendation here is not trivial. If even half of the approximately 9 million white males between the ages of 65 and 80 (a prime risk group for AAA) were to be screened with ultrasound at $150/screen, the aggregate bill just for the initial test is $675 million. This does not include the some 63,000 elective surgeries (2% prevalence of operable AAAs and 70% compliance with recommended surgery) that could result (another $1.7 billion) and nearly 1000 premature deaths resulting from complications of the elective surgery.

3.2 Implications for Medical Simulation Modeling

Models such as ours are going to be increasingly used to inform public policy decisions of this magnitude. No doubt there may be substantial errors in many of these models in spite of careful work by their creators. Recently the National Research Council has identified two "major deficiencies that demand attention if policy models, of whatever type, are to provide cost-effective information to the legislative debates of the future. The first...is lack of systematic model validation. ... The second...is under investment and consequent deterioration in the scope and quality of needed input data for policy models." (Citro and Hanushek, 1991, pp 2-3.)

Similar problems exist for medical simulation models. We wish to briefly recount some of these here.

Validation. Models are made to predict quantities that we do not know directly. Without making a $2 billion investment and waiting 20 years for the data, we cannot directly validate the main predicted quantities in the AAA model. Can a small experiment be run? Because of the low prevalence rates of AAA, and even lower incidence rates, and because of many contingent uncertainties determining AAA-related events, huge sample sizes would be required to validate even intermediate results of the model. This is not feasible.

The only feasible "validation" is then by exploratory sensitivity analysis, pushing the model to extremes to discover counterintuitive behavior, and by peer review. We have employed the former method quite extensively and indeed have discovered "bugs" in the model programming and formulation. Over time these seemed to be decreasing in the potential magnitude of impact on model results, and we have not found any for a long time. Other than this we can make no absolute guarantee however.

Reviewing models. Can the model be adequately validated by peer review? Any publication of the model is necessarily a brief written abstraction such as presented in section 2 of this paper. We have tried to give sufficient details, but any reader -- or reviewer -- will have to take much on faith. Even if we were to present reviewers with a disk containing the entire model (we are very willing to do so!), the time investment in confirming someone else's code is substantial burden on a reviewer. This is more than is traditionally expected of reviewers for journal articles.

If models were reported in some cannonical form the review task might be simplified. It appears to us hat every problem's nuances may defeat standardization. The development of "stochastic trees", recently reported in the medical literature by Hazen (1992) may assist here however.

Data for medical models. A central problem with medical simulation modeling was noted earlier: the data to instantiate a model are fragmentary, derived from many secondary sources, and often of dubious quality for modeling purposes. Unlike many industrial simulations, where data may be collected specifically to support the simulation, medical simulations seem to be almost wholly dependent on secondary data. Processes are often modeled on a different time scale than the original data were collected on; *e.g.*, screening intervals of 1, 3, and 5 years might be investigated using data originally collected at a 4-year interval (Dasbach et al. 1991). Next to nothing is known about higher order joint distributions of parameters in medical simulation models, so deterministic univariate and bivariate sensitivity analyses tend to be the rule as simulation experiments are impossible.

Model Abstraction. Finally, we note that not all modelers would instantiate a given model in the same way. It is a secure bet that were we to deliver the AAA model as an empty shell to another research team, they may well come to different conclusions about base case parameter values (not to mention ranges) from their review of the literature. If another research team were told to make a deterministic state-transition model for the problem, not only would the parameter values likely differ, but the model may well be quite different structurally. We would hope the functional implications of their model and ours would be similar, but there is no guarantee. Finally, were the form of the model not dictated, the other AAA model could be radically different. Any modeling effort is a long series of judgments made by the modelers in their process of developing a computational abstraction of the problem. We are not aware of any studies of variability in models due to these judgments.

3.1 A Recommendation

Important public health care decisions will depend on medical simulation models. There are substantial questions about the validity of such models. Although we believe it important to use modeling techniques to draw together what knowledge exists to support policy decisions, it seems important to know more about the performance of these techniques.

We recommend the medical simulation community discuss goals, processes, and perhaps standards for validation, presentation and review, and for use of simulation models for use in health policy. We further recommend that there be a concerted empirical assessment of medical simulation modeling as a process. This might reasonably include a federally funded, centrally administered experiment involving multiple research teams modeling the same problems. With billion dollar policy decisions in the balance an investment in knowing more about the validity and variability of the modeling process seems prudent.

REFERENCES

Bengtsson H., D. Bergqvist, S. Jendteg, et al. 1989. *World Journal of Surgery* 13:266-271.

Breckwoldt W.L., W.C. Mackey, and T.F. O'Donnell. 1991. The economic implications of high-risk abdominal aortic aneurysms. *Journal of Vascular Surgery* 13:798-804.

Canadian Task Force on the Periodic Health Examination. 1991. *Canadian Medical Association Journal* 145:783-789.

Citro, C.F. , and E.A. Hanushek (Ed.) 1991. *Improving Information for Social Policy Decisions: The Uses of Microsimulation Modeling, Volume I.* Washington, D.C.: National Academy Press.

Collin J., L. Araujo, J. Walton, et al. 1988. Oxford screening programme for abdominal aortic aneurysm in men aged 65 to 74 years. *Lancet* 2:613-615.

Collin J., B. Heather, J. Walton. 1991. Growth rates of subclinical abdominal aortic aneurysms -- implications for review and rescreening programmes. *European Journal of Vascular Surgery* 5:141-144.

Cronenwett J.L., T.F. Murphy, G.B. Zelenock, et al. 1985. Actuarial analysis of variables associated with rupture of small abdominal aortic aneurysms. *Surgery* 98:472-483.

Dasbach, E.J., D.G. Fryback, P.A. Newcomb, et al. 1991. Cost effectiveness of strategies for detecting diabetic retinopathy. *Medical Care* 29:20-39.

Goldman L., D.J. Gordon, B.M. Rifking, et al. 1992. Cost and health implications of cholesterol lowering. *Circulation* 85:1960-1968.

Hazen G.B. 1992. Factored stochastic trees: a tool for solving complex temporal medical decision models. *Medical Decision Making* 12:163-178.

Lederle, F.A. 1990. Screening for snipers: the burden of proof. *Journal of Clinical Epidemiology* 43:101-104.

Lederle F.A., J.M. Walker, D.B. Reinke. 1988. Selective screening for abdominal aortic aneurysms with physical examination and ultrasound. *Archives of Internal Medicine* 148:1753-1756.

Morrison, A. S. 1985. *Screening in Chronic Disease.* New York: Oxford University Press.

Nevitt, M.P., D.J. Ballard, J.W. Hallett. 1989. Prognosis of abdominal aortic aneurysm: a population-based study. *New England Journal of Medicine* 321:1009-14.

Public Health Service. 1980. *Promoting Health/Preventing Disease: Objectives for the Nation.* Washington, DC: U.S. Department of Health and Human Services, 1980.

Public Health Service. 1990. *Healthy People 2000.* Washington, DC: U.S. Department of Health and Human Services.

U.S. Preventive Services Task Force. 1989. *Guide to Clinical Preventive Services.* Baltimore: Williams & Wilkins.

Weinstein, M.C., and W.B. Stason. 1977. Foundations of cost-effectiveness analysis for health and medical practices. *New England Journal of Medicine* 296:1977.

AUTHOR BIOGRAPHIES

DENNIS G. FRYBACK, Ph.D., is Professor of Preventive Medicine and of Industrial Engineering at the University of Wisconsin-Madison. His principal research interests are decision analysis and technology assessment in health care. He is a past president of the Society for Medical Decision Making and edited the journal *Medical Decision Making* for three years. He chairs the Health Care Technology Study Section for the Agency for Health Care Policy and Research, and is a member of the recently re-formed U.S. Preventive Services Task Force.

PAUL S. FRAME, M.D., is a family practice physician in private practice with the Tri-County Family Medicine Program in Dansville, NY. He was Science Advisor to the original U.S. Preventive Services Task Force, and appointed to the re-formed USPSTF. He has a long-time interest in preventive medicine and determining what actions should and should not be included in periodic health examinations. He also has active federally supported research programs to develop computer-assisted reminder systems to promote better adherence to prevention protocols by physicians and patients.