Creating and Synthesizing Evidence With Decision Makers in Mind

Integrating Evidence From Clinical Trials and Other Study Designs

David Atkins, MD, MPH

Background: Randomized controlled trials (RCTs) remain the accepted "gold standard" for determining the efficacy of new drugs or medical procedures. Randomized trials alone, however, cannot provide all the relevant information decision makers need to determine the relative risks and benefits when choosing the best treatment of individual patients or weighing the implications of particular policies affecting medical therapies.

Objectives: To demonstrate the limitations of RCTs in providing the information needed by medical decision makers, and to show how information from observational studies can supplement evidence from RCTs.

Methods: Qualitative description of the limitations of RCTs in providing the information needed by medical decision makers, and demonstration of how evidence from additional sources can aid in decision making, using the examples of deciding whether a 60-year-old woman with mildly elevated blood pressure should take daily low-dose aspirin, and whether a hospital network should implement carotid artery surgery for asymptomatic patients.

Conclusions: Even the most rigorously designed RCTs leave many questions central to medical decision making unanswered. Research using cohort and case-control designs, disease and intervention registries, and outcomes studies based on administrative data can all shed light on who is most likely to benefit from the treatment, and what the important tradeoffs are. This suggests the need to revise the traditional evidence hierarchy, whereby evidence progresses linearly from basic research to rigorous RCTs. This revised hierarchy recognizes that other research designs can provide important evidence to strengthen our understanding of how to apply research findings in practice.

Key Words: evidence-based medicine, health policy

(*Med Care* 2007;45: S16–S22)

From the Center for Outcomes and Evidence, Agency for Healthcare Research and Quality, Rockville, Maryland.

The opinions expressed herein are those of the authors alone and do not represent the offical positions or policies of the Agency for Healthcare Research and Quality or the US Department of Health and Human Services.

Reprints: David Atkins, MD, MPH, Center for Outcomes and Evidence, Agency for Healthcare Research and Quality, 540 Gaither Rd., Rockville, MD 20850. E-mail: david.atkins@ahrq.hhs.gov.

ISSN: 0025-7079/07/4500-0016

To a previous generation, a "medical decision maker" would have been synonymous with a physician. Today, however, the number of medical decision makers has proliferated. Regulatory agencies, health plans and insurance companies, professional societies, formulary committees, private industry, quality organizations and others are all involved in deciding what medical therapies and practices are approved, marketed, promoted, reimbursed, rewarded, or chosen by patients. Individual patients and family members often seek out new information on treatment alternatives, so that they can be more involved in decisions about their treatment. Unfortunately, we have not kept pace in our ability to produce the types of evidence that would help patients, their clinicians, and all the other decision makers make more informed decisions.[1]

Each new therapy triggers a range of clinical and policy decisions at different levels of the health care system, as illustrated in Table 1 for treatments for osteoporosis. The regulatory framework for the drug approval process ensures a relatively predictable body of evidence regarding pharmaceuticals. This evidence is centered on the protocol-driven randomized trials, which the FDA requires to establish safety and effectiveness before approving a new drug. Neither these trials nor the drug approval process itself is designed to address all the needs of the other decision makers noted in Table 1, however. The evidence needed for drug approval addresses only the first of 3 questions that Archie Cochrane (the forefather of the Cochrane Collaboration) noted must be answered for medical policy decision making: "Can it work? Will it work? Is it worth it?"[2] Knowing that an intervention can work is necessary but not sufficient for deciding whether to use it in an individual patient or to promote it for a broad population. Whether it will work in a specific patient, population or clinical setting, and whether the benefits will be worth any harms or costs are questions for which evidence from randomized trials is often lacking. As David Mant observed, "The clinical trial is the best way to assess whether an intervention works, but it is arguably the worst way to assess who will benefit from it."[3]

Answering these questions requires that we update the traditional evidence hierarchy and its emphasis on the randomized trial as the "gold standard."[4] Although such a hierarchy serves us well for answering questions of efficacy—"Can something work?"—we need to integrate evidence from trials with

TABLE 1. Decisions Relevant to Medical Therapies (Example: Therapy for Osteoporosis)

Drug approval—Is slow release sodium fluoride safe and effective for preventing fractures?
Drug coverage—Which bisphosphonate drugs should be included on a drug fomulary?
Clinical practice guidelines—When should therapy for low bone density be initiated?
Patient decisions—Should I take raloxifene, alendronate, or calcium and vitamin D to prevent osteoporosis?
Health plans and insurers: Should we pay for follow-up assessment of bone density for women on treatment, and how often?
Health system policies—Should we institute primary care-based ultrasound screening for osteoporosis?
Quality measurement—What is an appropriate measure of high-quality care in the treatment of osteoporosis?

the best evidence from other study designs if we hope to answer the questions "will it work, for whom, and with what balance of benefits and risks?"

Limitations of Randomized Trials

The strength of randomized trials lies in the protection that randomization provides against sources of bias and confounding that often affect observational studies, where treatments are chosen in the course of usual care rather than randomly assigned by an investigator. Although randomization and other features of well-designed trials (eg, careful follow-up and blinded assessment of outcomes) help ensure the "internal validity" of a study, they can reduce its relevance or "external validity" (also known as "applicability" or "generalizability"). Rothwell[5] has enumerated in detail how factors such as design, recruitment, execution, and follow-up of trials may affect their external validity. The end result is that the question answered by a trial may differ in important ways from the most important question for patients, clinicians or policymakers—differences arising from the population and the intervention studied, the comparison chosen, or the outcomes measured (Table 2).

For example, trials apply careful selection criteria and frequently use run-in periods to ensure adherence and to detect side effects. These maximize the effect of the intervention, minimize risk to patients, and reduce sources of underlying variation. At the same time, they produce a trial population that often differs from the target population in demographics, clinical status, and underlying risk for both benefits and harms. Similarly, clinicians and clinical sites are usually selected to maximize the quality of the interventions, which may be further enhanced by careful training, treatment protocols, and cointerventions. Follow-up and monitoring is designed to promote adherence to the drugs under study and to detect any complications. As a result, the benefits of the intervention observed in the trial may exceed those that can be expected when applied in practice.

Regulatory standards in the United States encourage placebo comparisons to isolate the effect of the therapy under study. In clinical practice, however, the important question is how a new treatment compares to available alternatives.

TABLE 2. Factors Affecting External Validity of Randomized Controlled Trials

Patient factors
Underlying risk (high or low)
Demographics: age, race, gender
Comorbid disease
Adherence to therapies
Disease stage and severity
Risk of complications or side-effects of therapy
Intervention factors
Representative or selected settings
Level of training, quality of intervention
Timing of intervention
Cointerventions, quality of care
Adherence (for drugs)
Comparison
Placebo or usual care
Usual care or optimal care
Intention to treat or "on treatment"
Outcomes
Surrogate or clinical outcomes
Individual outcomes or composite
Patient-centered or disease-focused
Length of follow-up
Completeness of harms data

Adapted from Rothwell.[5]

Usual care control groups better estimate the effectiveness of a new intervention relative to current practice, but they can be problematic when current practice is variable and suboptimal.[6] For example, trials of osteoporosis drugs have not regularly ensured that control women were getting adequate calcium and vitamin D.[7]

Intention-to-treat analysis is widely accepted as minimizing bias in randomized trials. When there is substantial dropout or crossover, however, intention-to-treat can underestimate both benefits and possible harms for the individual patient who adheres to intended therapy. In the Women's Health Initiative, the excess risk of thromboembolism associated with estrogen/progestin therapy more than doubled when adherence to actual therapy was considered,[8] whereas benefits of aspirin in randomized controlled trials (RCTs) are up to 50% higher among those patients who are most adherent to therapy.[9]

Finally, the choice of outcomes measured in trials often does not provide a complete picture of the benefits and harms that are important to patients. Surrogate outcomes such as change in a risk factor (eg, bone density) or disease marker (eg, tumor progression) may correlate poorly with the clinical outcomes of greatest interest (eg, hip fracture, cancer survival).[10] Often, events that may vary considerably in their severity (eg, minor and major gastrointestinal bleeding) are lumped together. Aggregate outcomes such as cardiovascular events may conceal important differences in effects on individual components such as stroke and myocardial infarction.

Assessing harms can be particularly problematic in efficacy trails. Harms are variably and often poorly reported;

the exclusion criteria in trials systematically exclude subjects at highest risk of harms (eg, the elderly, those on multiple medications or with multiple comorbidities), and other features of trial protocols (careful monitoring and dose adjustment) further minimize risks of harm to patients. Additionally, trials are often too short or too small to detect harms that are rare or which emerge slowly. Finally, information from industry-funded trials may remain unpublished, limiting the ability to examine potential safety issues completely.

Heterogeneity of Treatment Effect

Trials face a second set of limitations for informing individual patient decision-making. Because of heterogeneity of treatment effect, the average benefits observed within a trial may differ substantially from those that might be expected for a given individual.[11] A major source of apparent heterogeneity is variation in baseline risk—the risk of the outcome in the absence of treatment. Benefits on an absolute scale (eg, number of hip fractures prevented) typically increase along with the underlying risk.[12] For example, although lipid-lowering treatment is equally effective in patients with diabetes and nondiabetics on a relative scale (roughly 25% reduction in risk of myocardial infarction), treatment confers more than double the absolute benefit for diabetic patients due to their higher risk of heart disease. Patients may also differ in their responsiveness to treatment, due to differences in drug metabolism, receptor affinity, or other genetic factors that influence the mechanisms of disease progression and treatment response.[13] The role of sophisticated new markers in selecting therapy remains unresolved, however.[14]

A Model for Moving Forward

A number of proposals have been offered for improving the relevance of clinical trials to clinicians and policy makers. Chief among these is greater use of "pragmatic" trials (also known as practical or effectiveness trials). Sean Tunis, who experienced the deficiencies of the evidence to inform coverage decisions as the former Chief Medical Officer for Medicare, has advocated for the need for more relevant trial evidence by designing trials that enroll representative patients, replicate "real-world" rather than ideal practice, and address comparisons relevant for policy makers.[1] Such trials remain more common outside of the United States, but the growth of practice-based research networks and electronic health records will make it increasingly feasible to conduct large research studies in community-based practice settings here in the United States. Nonetheless, such studies can be time consuming and expensive, and it remains doubtful that the major funders of clinical research will shift their priorities sufficiently to develop a robust body of effectiveness evidence on all the questions of interest.

Subgroup analysis within trials can help explore important issues of heterogeneity of treatment effect[15] but must be used with caution to distinguish true differences from those arising by chance.[16] Systematic reviews and meta-analyses of RCTs provide another avenue for improving our understanding of trial evidence. Reviews can probe sources of heterogeneity among trials, examine rare effects by pooling smaller trials, provide a more complete picture of outcomes, and examine subgroup differences more effectively than an individual trial.[17] Meta-analysis of group level data, however, often lacks power to explore important patient predictors of outcomes. Individual patient meta-analysis is more powerful for this purpose, and has been used with success in several areas with large numbers of trials.[18] These efforts are the exception, however, and remain difficult given the variety of industry and public-funded trials.

Glasziou and Irwig[19] proposed individualizing treatment decisions by combining estimates of relative benefits and harms from trials with information to predict baseline risk of an individual patient, to derive estimates that may be more applicable to individual patients. This concept can be expanded by carefully applying a range of evidence from prospective cohort studies, registries, and other nonexperimental data to assess applicability of trials and to generate information more relevant to the specific clinical or policy decision at hand. Nonrandomized studies will never supplant the need for rigorously conducted trials but they can: enrich our understanding of how patients treated in practice differ from those in the trials; provide tools for estimating baseline risk and potential benefits for individual patients; examine whether trial results are replicable in community settings; examine outcomes not carefully studied in trials (especially rare or slow emerging harms); explore sources of heterogeneity in safety or effectiveness arising from variation among patients, clinicians, and settings; and together produce a more complete picture of the potential benefits and harms of a clinical decision for individual patients or health systems (Table 3).

TABLE 3. Roles of Different Types of Data in Medical Decision Making

Study Design	Advantages
Efficacy trials	Least biased estimate of the effect of specific intervention under ideal conditions. Careful protocols maximize ability to detect effect of intervention.
Effectiveness trials	More representative estimates of benefits and harms in typical patients.
Systematic reviews and meta-analyses	Pooled results may allow estimates of less frequent events and more stable estimates of treatment effect; explore heterogeneity across different settings, demographic and risk groups.
Cohort study	Can examine longer-term outcomes and populations excluded from trials. Can identify more specific estimates of baseline risk to help target treatment.
Registry	More representative data on range of outcomes, including harms; can explore risk factors for good and bad outcomes.
Administrative database	Estimates of major harms across large populations; detection of rare events; estimate adherence.
Case-control studies	Examine risk for uncommon harms and factors that modify risk.
Audit or survey	Examine appropriateness of practice patterns and patient selection.

We provide 2 examples, one involving decision making about a medical therapy for an individual patient and one involving policy decisions regarding implementation of a surgical intervention, to illustrate how this approach can add to the information from a single large and seemingly definitive trial.

Assembling Evidence for Individual Patient Decision-Making: Should a 60-Year-Old Woman With Mildly Elevated Blood Pressure (Systolic Blood Pressure of 140 mm Hg) Take Daily Low-Dose Aspirin?

Individual patients need to know more than that a therapy works on average. They want to know about the benefits and risks of a given treatment relative to available alternatives for "someone like me"—ie, someone with the same age, gender, and collection of risk factors as themselves. Table 4 illustrates a series of steps and the sources of data that can help take the results from a single trial and make them more relevant for an individual patient.

Although aspirin's benefits have been long established for people with underlying cardiovascular disease, its role for primary prevention in women has been less clear. The Women's Health Study (WHS), specifically designed to address the benefits of low dose aspirin in initially healthy middle-aged women, found that aspirin significantly lowered the risk of stroke but not the risk of myocardial infarction, and increased the risk of gastrointestinal (GI) bleeding.[9] Its applicability to our 60-year-old hypertensive patient may be reduced, because WHS enrolled healthy female health professionals, 60% of whom were under age 55. A meta-analysis that included 2 additional studies enrolling older and higher-risk women, however, confirms the WHS findings that aspirin use reduces risk of stroke (pooled estimate 17% reduction) but increases risk of serious GI bleeding (pooled estimate 68% increase—see Table 5 for individual trial results).[20]

Answering whether daily aspirin is "worth it" for this patient requires careful consideration of how benefits and harms for the individual might differ from that represented for the total population of trial participants. Exploring subgroups within the WHS yields conflicting findings. Although aspirin had no effect on cardiovascular risk in women in her age category (age 55–64), it significantly reduced risk in women who were nonsmokers and in women with systolic blood pressure of 140 mm Hg or higher. A second approach is to explore variations in benefits and harms across trials, which enrolled patients of different underlying risk. An analysis of 7 primary prevention trials suggests that the

TABLE 4. Evidence to Answer Patient-Specific Questions About a Medical Therapy: Should a 60-Year-Old Woman Take Aspirin for Primary Prevention of Cardiovascular Disease

Question	Evidence	Findings
Can it work?		
Can aspirin reduce CVD in women?	Individual RCT[9]	NS 9% reduction in CVD; 17% reduction in stroke; 40% increase in serious GI bleeding.
	Meta-analysis[20]	12% reduction in CVD; 17% reduction in stroke; 68% increase in major bleeding.
Will it work?		
Will I have trouble adhering to aspirin?	Trials; registries[25]	Adherence high in RCTs and community patients with heart disease (>80%); little data on primary prevention.
Is it worth it?		
Do benefits exceed harms in trials?	Individual RCT[9]	Over 10 yr, 45 fewer strokes but 36 more serious GI bleeds in 20,000 women taking aspirin.
	Meta-analysis[20]	Prevent 3 CVD events, cause 2.5 bleeds per 1000 women over 6.4 yr.
	Overviews of trials[21]	CVD benefits exceed major bleeds only when CVD risk >1% per yr.
Will my benefits differ from those of trial participants?	Individual RCT subgroup[9]	No CVD benefit in women 55-65; greater benefit in nonsmokers and women with elevated blood pressure.
	Risk prediction from cohort studies[22]	Patient's stroke risk comparable to that of WHS women (1.3% over 10 yr).
Will my harms differ from those of trial participants?	RCTs and cohort studies[24]	Annual GI risk from aspirin up to 10 times higher in community (1-3 per 1000) than in WHS (0.2 per 1000).
	Case-control studies on risk factors	Age, concomitant NSAID or steroid use, past history of ulcers increase risk.
Will expected benefits exceed harms for me?	Modeling	Aspirin might cause 2-4 times more serious GI bleeds as strokes prevented in this 60-yr-old woman.
	Qualitative studies of impact of stroke, aspirin on patient experience	Apparent unfavorable balance may depend on severity of strokes, patient preferences.

CVD indicates cardiovascular disease; MI, myocardial infarction; NS, nonsignificant result; NSAID, nonsteroidal anti-inflammatory drug.

TABLE 5. Stroke and Major Bleeding Risk in Primary Prevention Trials Containing Women

Trial	Population	No. and Mean Age of Women	Average Follow-up (yr)	Stroke (Women Only)			Major Gastrointestinal Bleed (Women Only)		
				Control Rate*	RRR (%)	Strokes Prevented†	Control Rate*	RRI (%)	Bleeds Caused†
WHS	Healthy female healthcare workers	39,876 (54.6)	10	1.3	17	0.2	0.5	40	0.2
HOT	Men and women with hypertension	8,883 (62.3)	4	3.7	19	0.7	1.4	89	1.2
PPP	Men and women with ≥1 cardiovascular risk factor	2,583 (64.7)	3.6	2	44	0.9	0.4	363	1.4
Pooled	All				17			68	

*Per 1000 untreated patients per year.
†Per 1000 patients per year taking aspirin.
HOT indicates Hypertension Optimal Treatment Study; PPP, Primary Prevention Project; RRR, relative risk reduction; RRI, relative risk increase.
Data adapted from Berger et al.[20]

vascular benefits of aspirin exceed major bleeding risks only when the 10-year cardiovascular risk exceeds 10%.[21] Most of these data come from trials of men, however, and assume the risk of bleeding is relatively constant.

A final approach is to examine how baseline risks, for both benefits and harms, of the individual patient compare with trial participants. Where they differ, we can create better estimates of benefits and harms by applying the pooled trial estimates of relative benefits (17% reduction in stroke risk) and harms (68% increase in major bleeding) to baseline risk estimates that are more applicable to our patient.[19] Data derived from careful prospective population-based cohort studies such as the Atherosclerotic Risk in Communities study have produced tools to calculate stroke or heart disease risk based on individual risk factors.[22] Based on age and risk factors, our patient's 10-year risk of stroke is estimated at 1.2%, comparable to that of the women in the WHS (1.3%),[9] suggesting that the benefits of aspirin in WHS seem to be applicable to our patient.

Can we similarly apply harms estimates from WHS? The women in WHS were relatively healthy, carefully screened, and underwent a run-in period, all of which may have contributed to their low risk of major GI bleeding. As seen in Table 5, the baseline risk of bleeding observed in the control group in WHS (0.5 per 1000 per year) is similar to that in the Primary Prevention Project, but less than half that in the Hypertension Optimal Treatment trial. Moreover, the excess number of major bleeds attributable to aspirin (0.2 per 1000 per year) is substantially lower that that observed in the 2 other trials (1.2–1.4 per 1000 annually). Moreover, in population-based studies, the annual risk of serious GI bleeding in adults taking aspirin is 5- to 10-fold higher.[23,24] Putting the data from trials and other studies together, we can estimate that aspirin might modestly lower the risk of stroke in our 60-year-old patient over 10 years (from 1.2% to 1%, or a number needed to treat 500 to prevent 1 stroke). At the same time, her excess risk over 10 years of a major GI bleed (serious enough to require hospitalization) is likely to be substantially higher than the 2 per 1000 observed in the WHS. The other trials enrolling less healthy women suggest a risk as much as 5 times higher, and cohort studies, which may better represent risk in unselected patients, suggest it could be even higher.

The final considerations for patients are their own preferences, including their ability to adhere with daily aspirin therapy. Some patients might place a higher value on preventing a stroke, but they would also need to consider the small risk of fatal bleeding and other less serious risks associated with aspirin use, including ulcer disease, hematuria, and minor GI bleeding. Finally, although adherence was good in prevention trials and in community registries of patients with heart disease,[25] data from volunteers and symptomatic patients may not apply to healthy subjects in the community. In conclusion, a treatment that looked potentially beneficial in a landmark trial now seems, on consideration of a greater array of evidence, to be likely to do more harm than good for the specific patient in question.

Answering Questions for Clinical Policy Makers: How Should Carotid Artery Surgery for Asymptomatic Patients be Implemented in a Hospital Network?

The potential benefits of carotid endarterectomy to prevent stroke in patients with asymptomatic carotid stenosis was addressed in a large, National Institutes of Health-funded clinical trial published in 1995.[26] The Asymptomatic Carotid Artery Surgery (ACAS) study, indicating that surgery reduced combined risk of stroke and death over 5 years, led to an immediate and substantial increase in use of endarterectomy.[27] To a policy maker, the trial raises the question of how a specific health system can appropriately implement the findings of this trial (Table 6). Having established that the surgery can work, the important question is whether it will work in the specific patients and hospitals within the system in question (eg, a network of hospitals and group practices). Trials of surgical procedures, as with studies of other complex interventions that depend highly on the training and technical skill of the team delivering care, raise important questions of generalizability to wider practice. The ACAS trial achieved a 53% reduction in death or ipsilateral stroke

TABLE 6. Evidence to Answer Population Policy: How Should a Health Plan Implement Findings of a Randomized Trial on Carotid Endarterectomy (CEA)?

Question	Evidence	Findings
Can it work?		
Can CEA reduce stroke and death in persons with asymptomatic carotid stenosis?	Individual RCT[26]	Over 5 yr, aggregate risk of ipsilateral stroke or death reduced 53%; benefit lower in women.
	Meta-analysis[28]	29% reduction in risk over 3–4 yr.
Will it work?		
Will CEA be as safe and effective in the community as in the trial?	Trials and registries[29]	All trials carefully selected surgeons. Operative morbidity and mortality in community double that achieved in trials.
Is it worth it?		
Do benefits of CEA outweigh harms in trials?	Individual RCT and meta-analysis[26,28]	Net benefits emerge at 2 yr but are small (100 operations to prevent 1 event per year). No effect on total mortality/stroke rates.
Will benefits of CEA outweigh harms in our patients and our hospitals?	Cohorts and administrative databases[29,30]	State outcomes variable but morbidity usually exceeds that of trials. Outcomes best in high-volume hospitals.
Which patients are most likely to benefit?	Overviews of trial subgroups; cohort studies of operative outcomes[15]	Benefit significantly greater in men than women (51% vs. 4% reduction in risk) and in younger vs. older persons. Surgical risk higher in women and elderly.
Which settings are most likely to have good outcomes?	Hospital performance audits; state comparisons[29,30]	Substantial state variation in complication rates; high volume hospitals have lower complication rates; variation in process measures.
What policies can increase the appropriate use of carotid endarterectomy to achieve the best outcomes?	Quality improvement studies[31]	Modest improvements in outcomes with feedback of data to hospitals.

CEA indicates carotid endarterectomy; CVD, cardiovascular disease.

over 5 years by carefully selecting surgeons based on their operative complication rate, excluding the majority of potential participating surgeons.[26] A second trial, with slightly higher complication rates, found similar benefits of surgery.[27] A pooled analysis of trial data suggests a 29% reduction in risk over 4 years.[28] Examination of Medicare data in the follow-up to the ACAS, however, casts doubt on the question of whether carotid endarterectomy will work as well in more representative settings. Mortality rates from endarterectomy were 8-fold higher in the community than in the ACAS trial.[29] Similar audits of state-by-state performance indicated wide variation in complication rates between states.[30] These differences may arise from less careful selection of surgeons and patients and differences in quality of pre- and postoperative care delivered in the community. Outcomes seem best in hospitals performing a high volume of surgeries.[29]

This variation highlights concerns that endarterectomy, if not carefully applied in the right settings to the right patients, won't be "worth it"—that is, it could result in operative deaths and complications that outweigh any long-term benefits. Subgroup analysis of trials indicates that the benefits are much greater in asymptomatic men than in asymptomatic women (51% vs. 4% reduction in risk, respectively); cohort studies indicate that this is a result both of higher surgical risk in women and lower risk of stroke on medical therapy.[15] In addition, benefits may go down with age; elderly people have higher risks with surgery and are more likely to die of other causes.

With knowledge of specific complication rates for a state, hospital, or individual surgeon, a hospital network can steer patients to better performing centers and feed back data to promote best practices regarding perioperative care and surgical interventions. Finally, education of providers and patients can attempt to influence careful selection of appropriate patients; audits could examine the rates at which women and patients over age 80 are undergoing surgery, because the balance of harms and benefits is especially problematic in these groups. Quality improvement studies at the state level indicate some of these interventions can change processes and outcomes of care.

CONCLUSIONS

As these cases illustrate, even landmark trials can raise as many questions as they answer. If we are to implement the findings of these studies effectively, we have to pay close attention to the applicability of trial findings to the settings and patients of interest. Research using cohort and case-control designs, disease and intervention registries, outcomes studies using administrative databases, and quality improvement methods can all shed important light on who is most likely to benefit, what the important tradeoffs are, and how policy makers might promote the safe, effective, and appropriate use of new interventions. This use of evidence suggests a different model from that suggested by the traditional evidence hierarchy, with a linear progression from basic

research to observational studies, to pilot intervention studies, and finally to definitive controlled clinical trials. For many of the decisions we face, trials generate questions and highlight gaps in evidence that must be examined and bridged with other study designs. Instead of being a narrow pillar where each study rests on the preceding one, robust evidence is better likened to a web. Trials often provide the strong strands that create the central structure, but the strength of the completed web relies on a variety of supporting cross strands made up of evidence from a more diverse array of studies.

REFERENCES

1. Tunis SR, Stryer DB, Clancy CM. Practical clinical trials, increasing the value of clinical research for decision making in clinical and health policy. *JAMA*. 2003;290:1624–1632.
2. Haynes B. Can it work? Does it work? Is it worth it. *BMJ*. 1999;319: 652–653.
3. Mant D. Can randomized trials inform clinical decisions about individual patients. *Lancet*. 1999;353:743–746.
4. Harris RP, Helfand M, Woolf SH, et al; for the Methods Word Group, third US Preventive Services Task Force. Current methods of the US Preventive Services Task Force: a review of the process. *Am J Prev Med*. 2001;20(3 Suppl):21–35.
5. Rothwell PM. External validity of randomized controlled trials: "To whom do the results of this trial apply? " *Lancet*. 2005;365:82–93.
6. NIH Program on Clinical Research Policy and Analysis and Coordination. Considering usual medical care in clinical trial design: scientific and ethical issues. Proceedings of a meeting in Bethesda, MD, November 14–15, 2005. Available at: http://crpac.od.nih.gov/Draft_UsualCareProc_06062006_cvr.pdf.
7. Nelson HD, Helfand M, Woolf SW, et al. Screening for postmenopausal osteoporosis: a review of the evidence for the US Preventive Services Task Force. *Ann Intern Med*. 2002;137:529–541.
8. Rossouw JE, Anderson GL, Prentice RL, et al. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's Health Initiative randomized controlled trial. *JAMA*. 2002;288:321–333.
9. Ridker PM, Cook NR, Lee IM, et al. A randomized trial of low-dose aspirin in the primary prevention of cardiovascular disease in women. *N Engl J Med*. 2005;352:1293–1303.
10. Ridker PM, Torres J. Reported outcomes in major cardiovascular clinical trials funded by for-profit and not-for-profit organizations: 2000–2005. *JAMA*. 2006;295:2270–2274.
11. Kravitz RL, Duan N, Braslow J. Evidence-based medicine, heterogeneity of treatment effects, and the trouble with averages. *Milbank Q*. 2004;82:661–687.
12. Jackson R, Lawes CM, Bennett DA, et al. Treatment with drugs to lower blood pressure and blood cholesterol based on an individual's absolute cardiovascular risk. *Lancet*. 2005;365:434–441.
13. Ginsburg GS, Konstance RP, Allsbrook JS, et al. Implications of pharmacogenomics for drug development and clinical practice. *Arch Intern Med*. 2005;165:2331–2336.
14. Cooper RS, Psaty BM. Genomics and medicine: distraction, incremental progress, or the dawn of a new age. *Ann Intern Med*. 2003;138:576–580.
15. Rothwell PM. Subgroup analysis in randomized controlled trials: importance, indications, and interpretation. *Lancet*. 2005;365:176–186.
16. Oxman AD, Guyatt GH. A consumer's guide to subgroup analyses. *Ann Intern Med*. 1992;116:78–84.
17. Egger M, Davey Smith G, Altman DG. *Systematic Reviews in Health Care*. London, England: BMJ Books; 2001.
18. Leonardi-Bee J, Bath PM, Bousser MG, et al. Dipyridamole for preventing recurrent ischemic stroke and other vascular events: a meta-analysis of individual patient data from randomized controlled trials. *Stroke*. 2005;36:162–168.
19. Glasziou PP, Irwig LM. An evidence based approach to individualizing treatment. *BMJ*. 1995;311:1356–1359.
20. Berger JS, Roncaglioni MC, Avanzini F, et al. Aspirin for the primary prevention of cardiovascular events in women and men. *JAMA*. 2006; 295:306–313.
21. Patrono C, Rodriguez LAG, Ladolfi R, et al. Low-dose aspirin for the prevention of atherothrombosis. *N Engl J Med*. 2005;353:2373–2383.
22. Chambliss LE, Heiss G, Shahar E, et al. Ischemic stroke risk prediction in the Atherosclerosis Risk In Communities (ARIC) Study. *Am J Epidemiol.* 2004;106:259–269. Available at: http://aricnews.net/strokecalc/html/s_RC1.html. Accessed November 2, 2006.
23. Petty GW, Brown RD Jr, Whisnant JP, et al. Frequency of major complications of aspirin, warfarin, and intravenous heparin for secondary stroke prevention. *Ann Intern Med*. 1999;130:14–22.
24. Laine L. Gastrointestinal bleeding with low-dose aspirin—what's the risk. *Aliment Pharmacol Ther*. 2006;24:897–908.
25. Newby LK, Lapointe NM, Chen AY, et al. Long-term adherence to evidence-based secondary prevention therapies in coronary artery disease. *Circulation*. 2006;113:203–212.
26. Executive Committee for the Asymptomatic Carotid Atherosclerosis Study. Endarterectomy for asymptomatic carotid artery stenosis. *JAMA*. 1995;273:1421–1428.
27. Gross GP, Steiner CA, Bass EB, et al. Relation between prepublication release of clinical trial results and the practice of carotid endarterectomy. *JAMA*. 2000;284:2886–2893.
28. Chambers BR, Donnan GA. Carotid endarterectomy for asymptomatic carotid stenosis [review]. *Cochrane Database Syst Rev.* 2005;4: CD001923.
29. Wennberg D, Lucas F, Birkmeyer J, et al. Variation in carotid endarterectomy mortality in the Medicare population. *JAMA*. 1998;279:1278–1281.
30. Sheikh K, Bullock C. Variation and changes in state-specific carotid endarterectomy and 30-day mortality rates, United States, 1991–2000. *J Vasc Surg*. 2003;38:779–784.
31. Kresowik TF, Bratzler DW, Kresowil RA, et al. Multistate improvement in process and outcomes of carotid endarterectomy. *J Vasc Surg*. 2004;39:372–380.