Designed Delays Versus Rigorous Pragmatic Trials

Lower Carat Gold Standards Can Produce Relevant Drug Evaluations

Malcolm Maclure, ScD, Bruce Carleton, PharmD,† and Sebastian Schneeweiss, MD, ScD‡§¶*

Background: Centralized administrative databases enable low-cost pragmatic randomized trials (PRTs) of drug effectiveness and safety. We simplified the PRT strategy by using designed delays (DD) to evaluate drug policies.

Objectives: To reassess our DD trial of a cost-saving nebulizer-to-inhaler conversion policy and a proposed DD trial of reduced restrictions on Cox-2 inhibitors.

Research Design: We randomized 52 pairs of communities and clusters of physician practices to the policy either on time or after a 6-month delay. Our 2-stage qualitative reassessment comprised: (1) applying criteria for reporting PRTs and (2) assessing DD trials in 3 domains of responsibility: policymakers' decisions, researchers' decisions, and joint decisions involving negotiation.

Measures: A draft checklist of 22 Consolidated Standards of Reporting Trials (CONSORT). Researchers' recollections of their degree of influence on decisions.

Results: DD trials deviated from ideal PRTs in the policymakers' domain: the policies affected mixtures of drugs, users, and illnesses, and implementation was not by strict protocol. Aspects negotiated by researchers and policymakers also deviated from ideal: length of delay; size and location of control group; unit of randomization; additional data collection; and communications to physicians. The DD trials complied better with CONSORT in the researchers' domain of analysis and interpretation.

Conclusions: DD trials can be negotiated with policymakers. Low cost and simplicity of DD trials partly compensate for some limitations for evaluating drug safety and effectiveness. The ethics question of whether a DD is routine evaluation or research depends on its purpose and generalizability.

Key Words: pharmaceutical care, policy, study design, randomized control trial

(*Med Care* 2007;45: S44–S49)

In October 1998, when the *British Medical Journal* marked the 50th anniversary of its publication of the first randomized clinical trial,[1] we were preparing to do our first randomized drug policy trial. In March 1999 in British Columbia (BC), the publicly-funded drug benefit program called PharmaCare, covering the elderly, the poor, and families with high drug expenditures, introduced a drug benefit restriction policy with a randomized delayed control group: 10% of general practices in BC were granted a 6-month optional delay in the policy. Encouraged by the success of its first trial, PharmaCare officials 2 years later explored the possibility of a randomized trial of expanding coverage to include Cox-2 inhibitors (coxibs), a new subclass of nonsteroidal antiinflammatory drugs (NSAIDs) that then included only celecoxib and rofecoxib.

The idea that PharmaCare should use a delayed control group to evaluate drugs and policy changes was inspired by an interview with Tom Chalmers, a pioneer of randomization in health care.[2] Reflecting on his 40 years of successes and failures to promote randomized trials, he said,

■ *When I first went to the Veterans Administration (VA) central office in 1968, coronary care units were just becoming popular. The VA was faced with a problem: they did not have the money to pay for coronary care units in all 150 general medical hospitals. I heard they had decided on the 10 biggest hospitals getting funded 1 year, and the next year the next 10, and the next year the next 10. And I tried to persuade the chief medical director to choose the top 20 hospitals and assign them at random, and the controls could get their coronary care unit a year later. They would have had 1 full year of functioning to compare the mortality in the 10*

From the *School of Health Information Science, University of Victoria, Victoria, British Columbia, Canada; †Pharmaceutical Sciences, University of British Columbia, Vancouver, British Columbia, Canada; ‡Department of Medicine, Harvard Medical School; §Department of Epidemiology, Harvard School of Public Health; and ¶Department of Medicine, Division of Pharmacoepidemiology and Pharmacoeconomics, Drug Evaluation and Outcomes Research, Brigham and Women's Hospital (BLI341), Boston, Massachusetts.

Supported by Distinguished Scholar Award from the Michael Smith Foundation for Health Research, Vancouver, British Columbia (to M.M.); research grants from the National Institutes of Health (AG19463), Department of Health and Human Services, Rockville, Maryland (to S.S.); and Canadian Health Services Research Foundation, plus in-kind contributions to the policy trial from the British Columbia Ministry of Health, Victoria, British Columbia (to B.C.); and contracts from PharmaCare to advise on drug policy (to B.C.); and Agency for Healthcare Research and Quality, Gaithersburg, MD.

M.M. is an employee of the Ministry of Health on extended leave for the duration of his award, 2002–2006.

Presented at the DEcIDE Program Symposium *Comparative Effectiveness and Safety: Emerging Methods,* Agency for Healthcare Research and Quality, June 19–20, 2006, Gaithersburg, MD.

Reprints: Malcolm Maclure, ScD, School of Health Information Science, University of Victoria, P.O. Box 3050, Station CSC, Victoria, B.C. V8W 3P5, Canada. E-mail: malcolmmaclure@shaw.ca.

ISSN: 0025-7079/07/4500-0044

hospitals that got units with the mortality in the 10 that would have got them but did not. A very logical way to test right off the bat whether coronary care units were life saving or not, and to compare costs. [*But they did not agree to do it.*] *Then, after I left the VA, I heard they had the same problem with CAT scans. So I wrote a long letter to the medical director reciting all the arguments in favor of randomly assigning CAT scans to hospitals, and measuring various outcomes and mortality . . . They would not do it. Then I read about MRIs and I tried again a third time . . . they turned it down.*[3]

In 1994, a year after interviewing Chalmers, one of us (M.M.) was asked by PharmaCare how to evaluate an impending restriction on drug coverage—payment of no more than the lowest market price among chemically identical drugs, ie, reimbursement for only the cheapest generic agent. He suggested a randomized delay, and Chalmers agreed to speak in favor of that design. The opportunity did not arise, however; PharmaCare's policy implementation committee initially rejected the idea because of concerns about how the media and public might view such a trial.

We regarded PharmaCare's initial reactions as hypotheses to be tested. When we conducted focus groups and interviews with patients and clinicians, we found greater receptiveness to randomized delays than the PharmaCare staff imagined. This finding paved the way for the policy trial in 1999. Its smooth execution led to PharmaCare's participation with us in planning other policy trials.

In January 2005, we reported on our experience at a workshop on pragmatic randomized trials (PRTs) in health care at the Institute for Clinical Evaluative Sciences in Toronto. Participants were trialists, epidemiologists, journal editors, and policymakers who had conducted PRTs or coauthored papers on Consolidated Standards on Reporting Trials (CONSORT). CONSORT is a 22-item checklist of recommendations, from the title and abstract to the overall conclusions.[4] Adaptations of CONSORTs for special types of trials have been developed.[5,6] The purpose of the 2005 workshop was to adapt CONSORT to PRTs.

METHODS

At the workshop, we presented our completed and planned trials of educational interventions to improve prescribing,[7–9] as well as the 2 trials of policies that we discuss below. We called them designed delay (DD) trials, a special type of PRT with less generalizability because of greater dependence on the local administrative context. The workshop organizers asked us to assess draft modifications to the CONSORT by applying them to specific studies. When we applied the checklist of 22 recommendations on *reporting* to our past *conduct* of the 2 DD trials of policies, we noticed a pattern of deviation from ideal practice that warranted a critical reassessment. In June 2006, we presented a preliminary reassessment at a symposium sponsored by the Agency for Healthcare Research and Quality in Gaithersburg, Maryland. Two questions raised were, "What is special about a DD trial compared with other PRTs?" and "When is a DD classified as research rather than routine program evaluation?"

In addressing the first question, the most distinctive feature of DD trials emerged—the large influence of policymakers on how the trial unfolds. This led us to restructure our reassessment so that it covered 3 domains of responsibility: (1) policymakers, (2) researchers, and (3) joint decisions involving negotiation.

The importance of the second question was underscored by a subsequent commentary on the ethics of a new policy called coverage with evidence development (CED) at the Centers for Medicare and Medicaid Services (CMS). "There is wide variation in attitudes on the boundary between quality improvement and research activities. Regulatory implications give this boundary great importance. . .. The questions regarding which elements of CED are research and which are not have plagued the policy discussions at CMS."[10] According to the Code of Federal Regulations, 45 CFR 46_102 (d), research is any systematic investigation designed to develop or contribute to *generalizable* knowledge.[11] Whether a DD trial produces *generalizable* knowledge is sometimes beyond both the interests of policymakers and the influence of researchers. Adding to the ambiguous status of our first DD trial, one of us (M.M.) was then a PharmaCare employee with a simultaneous academic position, who believed the DD methodology, but not the findings of the policy evaluation, would be generalizable to other jurisdictions.

RESULTS

Comparison With Standards for Pragmatic Randomized Trials

Table 1, showing results of our reassessment, made us aware of a strong association between the early stages of a DD trial and the degree of influence of policymakers. The CONSORT checklist of recommendations for *reporting* trials (column 4) roughly corresponds to the chronology of a trial. When we used the checklist, we observed that DD trials deviated from ideal trials more during the planning and targeting of the intervention, the domain of policymakers—at the top of the Table 1—than during later steps.

The policymakers' crucial role partly accounts for our using the phrase *designed delays* rather than an alternative research term already in use. *Pragmatic randomized trials* are done in real-world settings with everyday patients and practices,[12] as distinct from *explanatory* trials done with selected patients in purer settings. The adjective "pragmatic" recalls the American philosophy of pragmatism with its focus on outcomes in normal life.[13] The phrase *practical trial*,[14] although synonymous, is apt to be misinterpreted to mean "feasible." Policymakers like the terms pragmatic and practical, but we have witnessed their discomfort with the word "randomized." Randomization suggests an element of chaos and loss of control. After many years, we settled on the phrase *designed delay* because policymakers and managers seem comfortable with it. They know almost every policy or program involves natural delays. Adding an element of design to those delays feels like an increase in control rather

TABLE 1. Key Elements in the Conduct of Two Designed Delay Trials in British Columbia

Three Domains of Decisions	Decisions on Conduct of DD Trials in BC	Deviations of DD Trials From Ideal Trials	CONSORT Items for Reporting Trials
Policymaker	The policy: its rationale, context, communication, timing, implementation, people affected, settings, monitoring, enforcement	Mixtures of drugs, users, illnesses, and settings; letters to control group; flexible implementation	Title, background, participants, interventions, objectives
Joint decisions, negotiated	Duration of delay, size and location of controls, whether to randomize, units of randomization, subset of population randomized, baseline data availability, outcome measures, extra outcome data if needed, recruitment if needed	Arguably: delay could have been longer, control group larger, or randomization units smaller; there were no rules for stopping the delays in the controls; just one of DD trials had to collect data by survey of patients	Outcomes, sample size and stopping rules, participant flow, recruitment, baseline data
Researcher	Grant proposals with statistical methods plan, run focus groups, run scientific advisory panel, matching and blocking, randomization methods, blinding of analysis, statistical analyses, assess generalizability, overall interpretation, publication of findings	Hypotheses and statistical analysis plan were in grant proposal, but policy evolved after that; blinding of analyses was hard to maintain because of monitoring; outcome analyses adjusted for nonadherence, plus used historical controls	Hypotheses, random sequence generation, concealment of random allocation, implementation of randomization, blinding, statistical analyses, interpretation

than a loss. We explain that randomization is good method of design, but it is not essential.

Terms for similar designs include *wait-listed design*,[15] *phased implementation trial*,[16] *randomized start or randomized withdrawal trial*,[17] *staggered start trial*,[18] and *firms trial*.[19] We prefer DD trials because delay covers both starting and stopping, as well as minor changes such as delayed mailings[7] that are too simple to call implementation. "Wait-listed" suggests that earlier is better. "Phased" can be confusing when the program or policy is already multistaged. "Staggered" suggests more than 1 delay. "Firms" are ongoing administrative structures.

Nebulizer Policy Trial

The Policymakers' Domain: The Policy Itself

In 1995, PharmaCare introduced a "maximum allowable cost" policy, called Reference-Based Pricing (a misnomer, as it was not a price control policy). The first waves of this policy were evaluated retrospectively, initially by PharmaCare. PharmaCare's internal evaluations, not considered research, were not reviewed by a research ethics committee. Later, researchers conducted more thorough evaluations using the same data,[20] after grant agencies and university ethics committees had reviewed their protocols.

In early 1997, PharmaCare decided to include respiratory drugs within the umbrella of Reference-Based Pricing. We persuaded PharmaCare to evaluate this step prospectively with a randomized delay. At the time, no principal investigator and no research grant existed. We assumed PharmaCare analysts would analyze the data. We did not discuss whether the trial would be classified as research or routine evaluation.

When the policy change occurred on March 1, 1999, it was no longer part of Reference-Based Pricing. It was a prior authorization policy called the Nebulizer-to-Inhaler Conversion Program. Reimbursement for medications that were delivered to the lungs by wet nebulizer machines would be available only if physicians faxed an appropriate clinical justification to PharmaCare. The rationale of the policy was that drugs for nebulizers were more expensive than the equivalent metered-dose inhalers, yet the 1996 Canadian Asthma Consensus Conference stated that "nebulized medication is rarely, if ever, indicated in the management of asthma in older children or adults."[21] The final wording of the policy allowed for more exemptions by prior authorization than some of the clinical advisors anticipated: 29% of patients continued to get coverage for nebulizer medications rather than a predicted 20%.

The Researchers' Domain: Grant Funding, Ethics Approval, Statistical Analyses

Funding and Ethics Review. The question of whether or not the trial was research was settled during the 2-year delay of the policy. We obtained a grant from the Canadian Health Services Research Foundation to pay for mailed asthma-related quality-of-life questionnaires, data entry, and statistical analyses. This necessitated approval by the university's research ethics committee. The committee agreed that the government policy did not require written informed consent from either the individuals immediately affected by the policy or those in the control group given a 6-month delay; analysis of anonymous central administrative data would also not require patient consent. However, the committee regarded the questionnaires sent to patients as research materials subject to their standard requirements for invitations to subjects to participate in research.

Impact Analyses. Elsewhere we have published our methods of analysis, using randomized concurrent and historical controls.[22] Briefly, we did controlled time-series analyses using central administrative databases, including hospitalizations, medical services, and PharmaNet, an online pharmacy database and network capturing all dispensings (with rare exceptions) of prescription drugs in all BC community pharmacies.

Domain of Negotiation Between Researchers and Policymakers

Designed Delay. In 1997, when we persuaded PharmaCare to evaluate the forthcoming respiratory drug policy prospectively, we did not know what the policy would finally be nor how generalizable its evaluation would be. We initially proposed to PharmaCare a delay of 6 or 12 months, and this was left open.

Scientific Advisory Panel. PharmaCare requested an international panel of scientific advisors. We chose 4 distinguished pharmacoepidemiologists from Canada and the United States and included PharmaCare's top external advisor, a clinical pharmacologist. They met by telephone 3 times and added advice by e-mail.

Duration of Delay. The scientific panel recommended a 1-year delay. Our focus groups with clinicians revealed some discomfort with a delay of 1 year. Almost all clinicians were comfortable with a 3-month delay. In BC, 3 months is a typical lag time between a policy change and the date it affects those patients who luckily (or deliberately) refilled their medications on the last day before the policy became effective, because physicians normally prescribe a 100-day supply of medications for chronic users. By the time of the policy, PharmaCare was most comfortable with a 6-month delay.

Unit of Randomization. Initially, we had proposed to PharmaCare that we randomize by patient. The trial could be done as an easy-to-describe additional exemption: people with a Personal Health Number ending with a particular digit (eg, "lucky number 7") could be exempt for a year. (In BC, such numbers are assigned sequentially and the last digit, a verification digit from 1 to 9, is known to be virtually random.) PharmaCare accepted this, pending advice from the scientific panel. At first, the panel split on whether to randomize by patient, physician, or community. In our focus groups, physicians in group practices said the policy should not differ within their group. We also noticed that people are more comfortable with financial inequities if they are geographically separated. Therefore, we proposed, and PharmaCare agreed, that the trial be limited to 52 pairs of remote communities and clusters of general practices in more sparsely populated urban areas.

Size of Control Group. The decision to limit the control group to about 10% of the province was determined largely by the cost to PharmaCare in delayed savings. Six months of delayed savings in 10% of nebulized medication users would equal 2.5 weeks of delayed savings for the whole province. After many months of delays, a further delay in savings equal to 2.5 weeks was tolerable to PharmaCare.

Communications. PharmaCare welcomed researchers' suggestions on wording of communications. We recommended that the letter to all physicians describing the policy should say that, among patients who would be exempt (eg, children younger than18 years old) were "patients in a random sample of practices who are participating in an independent evaluation." PharmaCare found the simplicity of the description very helpful. Against our preference, PharmaCare sent the letter announcing the policy 6 weeks ahead to all physicians, including the *delayed* group. Controls were told of their 6-month exemption 2 weeks later in a separate letter. Perhaps because they did not see the second letter, 52% of the *delayed* group did not take advantage of the optional 6-month delay. Subtracting the 29% exempt in the *early* group, this meant the prevalence difference of nebulized medication use between groups was only 19%.

Reuse of Control Group for Next Policy Trial. In April 1999, PharmaCare agreed to use the same design for prospective evaluation of the next policy, a maximum allowable cost for statins. During the committee meeting, there was no discussion of whether this would be analyzed by external researchers or internal analysts. However, the policy was not implemented because of its complexity and controversy.

Coxib Policy Trial

The Policymakers' Domain

In 2001, PharmaCare was receiving conflicting advice from experts on whether to cover coxibs. Rheumatologists from the BC Arthritis Society strongly urged PharmaCare to cover them. By contrast, PharmaCare's official evidence-review agency, the Therapeutics Initiative in the Department of Pharmacology and Therapeutics at University of British Columbia, concluded that neither the long-term effectiveness nor the safety of coxibs had been demonstrated. We proposed to PharmaCare that a policy trial, analogous to the nebulizer policy trial, could resolve this dispute. PharmaCare tentatively agreed to a DD trial of expanded coverage.

Eligibility and Representativeness. Many patients were already using coxibs and paying out of pocket or by private insurance. Therefore, for many higher income patients, the policy would not influence their drug use, only their pocketbooks. If we researchers could have designed the policy, we would have restricted it to low-income patients who were not getting the drug already. This aspect of policy design, however, was not in our purview. We expected the policy would not be targeted by income, so we prepared instead to do the targeting in our statistical analysis.

The Researchers' Domain: Grant Funding and Ethics Approval

Funding. We obtained a 2-year federal government grant to study the feasibility of a coxib policy trial and related issues. This step would have settled the question of whether the policy trial was research, had the policy gone ahead. If we had not been awarded the grant, however, we are unsure whether the DD trial would have been considered research. PharmaCare was mainly interested in knowing the outcomes of the policy, not in producing generalizable knowledge.

Ethics. The rheumatologists and the pharmaceutical company representatives concluded that *withholding* funding for this

class of drugs was unethical. In contrast, we researchers regarded preliminary evidence of adverse cardiovascular outcomes in the clinical trials meant that the effects of prescribing and covering these drugs was uncertain. Under our reasoning, the policy trial was ethical. However, we do not know what an ethics review committee might have decided (before the withdrawal of rofecoxib for safety concerns 3 years later) because PharmaCare decided not to expand coverage of coxibs and stopped planning the trial.

The Domain of Negotiation Between Researchers and Policymakers

Reuse of the Same Controls. Our initial idea was to use the same cluster-randomized design as in the nebulizer policy trial, except in reverse. The PharmaNet computer, at the center of the network linking computers in all BC community pharmacies, would be programmed to approve coxib reimbursement claims from patients in the group of practices that had been randomized to 6-month delays of the nebulizer policy. PharmaCare agreed to this as a starting point for planning.

Additional Data Collection. The rheumatologists believed that, if we proceeded, the trial must include collection of data from patients about their symptoms. However, the Therapeutics Initiative noted that manufacturers had presented no evidence that coxibs relieved symptoms or reduced inflammation any better than other NSAIDs. Manufacturers claimed only that coxibs presented lower risks of major adverse events than other NSAIDs, particularly hospitalizations for gastrointestinal bleeding. Centralized hospitalization data were available to PharmaCare at no cost. Therefore, we researchers and PharmaCare agreed that no additional data collection was needed. For that reason, the Board of the BC Arthritis Society rejected the trial as having insufficient scientific merit.

CONCLUSIONS

We have shown that conducting a DD trial is possible and that it can be described in 1 sentence, requires little extra work from policymakers, and need not entail any extra time and effort from clinicians and patients. The main barriers are attitudes about ethics, the wisdom of the policies themselves, and timing. The main limitation is potentially low generalizability of findings to other jurisdictions.

Ethics and Equipoise

As for any clinical trial, a policy trial should have "equipoise"—equity of position—meaning no known advantage for a patient to be in 1 arm or another. In the nebulizer policy trial, we knew from previous clinical trials that some patients would do better on metered-dose inhalers. People who could not physically use inhalers by themselves might do better if they continued with nebulizers, and many patients might experience no difference. The relative numbers of these types of patients in the population were not known. However, we assumed that physicians would request exemptions for patients who would do better on a nebulizer and would already have prescribed inhalers if they were better for certain patients. Therefore, the patients who switched medications because of the policy would be in the third category: their physicians could not predict whether the change would make a difference. Thus, equipoise was a reasonable assumption.

Controversial Policies

Equipoise is more likely when a policy is controversial. Accordingly, a threat to a planned DD trial is that the policy proposal will be shelved or modified beyond recognition. Indeed, both the statin policy trial and the coxib policy trial did not proceed because of controversies.

Unequal Timing

A common concern about policy trials is that a new policy should apply immediately to all patients equally. This overlooks the fact that, often, much inequality in timing already occurs. For example, the nebulizer policy affected institutionalized patients sooner because they had prescriptions for shorter durations, often no longer than 1 week. Also, patients who heard in advance about the coming policy had an opportunity to refill their prescriptions just before the restriction applied. Indeed, we observed a spike in claims for nebulizer medications just before the policy started. In other words, when agencies implement policies of this sort, they may well induce undesigned delays. If these points are kept in mind when judging ethics, policymakers, clinicians, and others may have fewer concerns about adding an element of design to otherwise haphazard delays.

Duration of Delays

Fortunately, the policy concerned respiratory drugs with short-term effects. A 6-month trial would have been more problematic for drugs, such as statins, with cumulative effects that might take a year or more to appear. Whether a drug's effects are manifested within days or months, influences the perceived appropriateness of PRTs, as is discussed elsewhere in this issue of *Medical Care*.[23] For the coxib trial, we were aiming for a 12-month delay. This could have been extended if we had observed no statistically significant differences. In retrospect, after the withdrawal of rofecoxib, we believe the most ethical policy trial for coxibs would have been a self-designing trial.[24] In such trials, the final size and duration (and possibly the number of allocation arms) are unknown in advance but are determined by regular statistical comparisons of outcomes among the groups.

Rapid Evaluation

We demonstrated the feasibility of rapid, rigorous impact evaluation. Anonymous PharmaNet data on dispensed drugs in the *early* and *delayed* groups were available each week. Medical services data from both groups were available each month. Just 6 weeks after the end of the 6-month delay, we presented preliminary findings at a scientific conference.[25]

Generalizability

A large explanatory clinical trial or, better, a meta-analysis of such trials is a "high-carat" gold standard of

evidence. But generalizability of randomized clinical trials is often limited because they are restricted to patients with few comorbidities or concomitant medications. Results of PRTs are more generalizable because the samples of patients are statistically representative of *risk modifiers* (comorbidities, medications and other exposures) in real-world patients. A DD trial is a special type of PRT with further *administrative* modifiers (variations in policy context, implementation, communication, and administrative databases) that *reduce* generalizability. For example, in the nebulizer policy trial, the use of nebulizers in the *early* and *delayed* groups differed by only 19%, meaning the findings applied to only a fraction of the population. This is analogous to limitations on generalizability of analyses using instrumental variables, discussed elsewhere in this issue of *Medical Care*.[26] Therefore, a DD trial may be done solely to assess a local policy for local reasons. In that sense, DD trials are "low-carat" gold standards that can produce locally more relevant evaluations.

Designed Delays Without Randomization

Once we reached agreement with PharmaCare on how to design delays into the policy implementation, we found that randomization was a relatively easy step. Other groups, however, may find that randomization causes a program evaluation or quality improvement activity to be classified as research. We do not believe the line between routine program evaluations and research should be drawn simply when delays, or starting times, are randomized. For example, in 2004, the CMS offered early enrollment into the new Medicare drug plan to up to 50,000 seniors *by lottery*,[27] although it failed to attract many applicants. Had the lottery succeeded, a routine evaluation comparing health services utilization data by the early applicants and the delayed applicants would have constituted a DD trial. Such an evaluation should not be classified automatically as research just because it involves randomization.

Our decade of slowly gained experience with DD trials has confirmed the prudence of Chalmers's advice, "I have suggested to payers that when there is a new procedure, they should say, 'Sure, we will pay for that . . . if the patient is part of a randomized control trial to determine whether we should pay for it.'"[3] He would have been delighted with CMS's policy of coverage with evidence development.

REFERENCES

1. Chalmers I. Unbiased, relevant, and reliable assessments in health care: important progress during the past century, but plenty of scope for doing better. *BMJ*. 1998;317:1167–1168.
2. Thomas Chalmers [obituary]. *BMJ*. 1996;312:118.
3. Maclure M. Tom Chalmers 1917–1995, Part 2: The tribulations of a trialist. *Can Med Assoc J*. 1996;155:986–988.
4. Begg C, Cho M, Eastwood S, et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA*. 1996; 276:637–639.
5. Campbell MK, Elbourne DR, Altman DG; The CONSORT Group. CONSORT statement: extension to cluster randomised trials. *BMJ*. 2004;328:702–708.
6. Moher D, Schulz KF, Altman DG; The CONSORT Group. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *JAMA*. 2001;285:1987–1991.
7. Dormuth CR, Maclure M, Bassett K, et al. Periodic letters on evidence-based drug therapy improves prescribing: a randomized trial. *CMAJ*. 2004;171:1057–1061.
8. Herbert CP, Wright JM, Maclure M, et al. Better prescribing project: a randomized controlled trial of the impact on prescribing of case-based educational modules and personal prescribing feedback. *Fam Pract*. 2004;21:575–581.
9. Maclure M, Nguyen A, Carney G, et al. Measuring prescribing improvements in pragmatic trials of educational tools for general practitioners. *Basic Clin Pharmacol Toxicol*. 2006;98:243–252.
10. Pearson SD, Miller FG, Emanuel EJ. Medicare's requirement for research participation as a condition of coverage: is it ethical? *JAMA*. 2006;296:988–991.
11. Code of Federal Regulations. Title 45, Volume 1. Public Welfare: Department of Health and Human Services. Part 46: Protection of Human Subjects. Section 102. Definitions. December 13, 2001:72. Available at http://www.research.utah.edu/irb/board/45%20CFR%20Part%2046.pdf.
12. Godwin M, Ruhland L, Casson I, et al. Pragmatic controlled clinical trials in primary care: the struggle between external and internal validity. *BMC Med Res Methodol*. 2003;3:28.
13. Wiener PP. Pragmatism. In: Wiener PP, ed. *The Dictionary of the History of Ideas: Studies of Selected Pivotal Ideas*. Vol 3. New York, NY: Charles Scribner's Sons;1973:551–570.
14. Glasgow RE, Magid DJ, Beck A, et al. Practical clinical trials for translating research to practice: design and measurement recommendations. *Med Care*. 2005;43:551–557.
15. Brown CH, Wyman PA, Guo J, et al. Dynamic wait-listed designs for randomized trials: new designs for prevention of youth suicide. *Clin Trials*. 2006;3:259–271.
16. Godfrey-Faussett P. District-randomized phased implementation: strengthening the evidence base for cotrimoxazole for HIV-positive tuberculosis patients. *AIDS*. 2003;17:1079–1081.
17. Bodick N, Forette F, Hadler D, et al. Protocols to demonstrate slowing of Alzheimer disease progression. Position paper from the International Working Group on Harmonization of Dementia Drug Guidelines. The Disease Progression Sub-Group. *Alzheimer Dis Assoc Disord*. 1997; 11(Suppl 3):50–53.
18. Whitehouse PJ, Kittner B, Roessner M, et al. Clinical trial designs for demonstrating disease-course-altering effects in dementia. *Alzheimer Dis Assoc Disord*. 1998;12:281–94.
19. Cebul RD. Randomized, controlled trials using the Metro Firm System. *Med Care.* 1991;29(7 Suppl):JS9–JS18.
20. Schneeweiss S, Maclure M, Dormuth C, et al. Pharmaceutical cost containment with reference-based pricing: time for refinements. *CMAJ*. 2002;167:1250–1251.
21. Canadian Asthma Consensus Conference. Summary of recommendations. *Can Respir J*. 1996;3:89–100.
22. Schneeweiss S, Maclure M, Carleton B, et al. Clinical and economic consequences of a reimbursement restriction of nebulised respiratory therapy in adults: direct comparison of randomised and observational evaluations. *BMJ*. 2004;328:560–567.
23. Mazor KM, Sabin JE, Boudreau D, et al. Cluster randomized trials: opportunities and barriers identified by leaders of eight health plans. *Med Care*. 2007;45(Suppl 2):S29–S37.
24. Fisher LD. Self-designing clinical trials. *Stat Med*. 1998;17:1551–1562.
25. Carleton BC, Maclure M, Dormuth C, et al. Randomized controlled trial of a drug policy change by British Columbia PharmaCare. In: *Proceedings of the Third International Conference on the Scientific Basis of Health Services in Toronto*, Ontario, Canada; October 1999.
26. Brookhart MA, Rassen JA, Wang PS, et al. Evaluating the validity of an instrumental variable study of neuroleptics: can between-physician differences in prescribing patterns be used to estimate treatment effects? *Med Care*. 2007;45(Suppl 2):S116–S122.
27. Young D. Medicare holds lottery for access to expensive drugs. *Am J Health Syst Pharm*. 2004;61:1632,1636.

