Users' Guides to the Medical Literature

XIX. Applying Clinical Trial Results
B. Guidelines for Determining Whether a Drug Is Exerting (More Than) a Class Effect

Finlay A. McAlister, MD, FRCPC
Andreas Laupacis, MD, MSc, FRCPC
George A. Wells, MSc, PhD
David L. Sackett, FRSC, MD, FRCP
for the Evidence-Based Medicine Working Group

MOST CLASSES OF DRUGS include multiple compounds. The opinions of clinicians, manufacturers, and purchasers may differ as to whether a particular drug is more efficacious, safer, or more cost-effective than others in its class.[1] In this article, we review the types of evidence commonly cited to support the prescribing of a particular drug rather than another of the same class and provide a hierarchy for grading studies that compare a drug with another of the same class, expanding on our discussion in part A of this Users' Guide.[2]

CLINICAL SCENARIOS

The Clinician

As a clinician, you care for many patients with elevated serum cholesterol levels. A speaker at a recent continuing medical education event reviewed the benefits of cholesterol-lowering therapy, particularly with 3-hydroxy-3-methylglutaryl coenzyme A reductase inhibitors (statins), in the primary and secondary prevention of ischemic heart disease but did not recommend a particular statin. You decide to consider statin therapy for all your patients with elevated cholesterol levels, but are uncertain which of the statins on the market is best. You ask a general internist, cardiologist, and endocrinologist for their opinions, and each suggests a different statin, citing different reasons. You contact pharmaceutical representatives to provide you with evidence that their statins are better than those of their competitors. Although you use the *JAMA* series on Users' Guides to the Medical Literature to assess the validity of published studies, faced with a variety of competing claims, you realize that you need a framework for grading the strength of these studies.

The Policymaker

Your colleague, a purchaser for a large health maintenance organization (HMO), is faced with a similar dilemma when she is asked to consider replacing the statin on her HMO's formulary with a newer one. She wonders whether there is enough evidence to support the contention that the new statin is as good as, or better than, the one currently on formulary. While the new statin is cheaper, it has been evaluated only in short-term trials, with cholesterol lowering as the solitary end point.

DRUG CLASSES

Although there is no uniformly accepted definition of a drug class—and some argue that it cannot be defined at all—drugs are generally said to belong to the same class for 1 of 3 reasons (TABLE 1). Herein, we define a drug class as those drugs that share a similar structure and mechanism of action. Most classes of drugs include multiple compounds, and because of their similar mechanisms of action, they are generally thought to confer similar pharmacologic effects and clinical outcomes (*class effects*). This assumption is a key medical heuristic[3] and underlies clinical practice guidelines in which evidence from studies involving 1 or more drugs within a class is extrapolated to other drugs of the same class. For example, it is recommended that β-blockers be prescribed for survivors of myocardial infarction or angiotensin-converting enzyme inhibitors to patients with heart failure. In this circumstance, clinicians are likely to be interested in the drug within each class

Author Affiliations: National Health Service Research and Development Centre for Evidence-Based Medicine, John Radcliffe Hospital, Oxford, England (Drs McAlister and Sackett); Division of General Internal Medicine, Ottawa Hospital (Dr Laupacis), and Clinical Epidemiology Unit, Loeb Health Research Institute (Drs Laupacis and Wells), Ottawa, Ontario.

Dr McAlister is currently at the Division of General Internal Medicine, University of Alberta Hospital, Edmonton.

The original list of members (with affiliations) appears in the first article of this series (*JAMA*. 1993; 270:2093-2095). A list of new members appears in the 10th article of the series (*JAMA*. 1996;275:1435-1439). The following members contributed to this article: Gordon Guyatt, MD, MSc, Virginia Moyer, MD, MPH, and David Naylor, MD, DPhil.

Corresponding Author: Finlay A. McAlister, MD, FRCPC, Division of General Internal Medicine, 2E3.24 WMC, University of Alberta, Canada Hospital, 8440 112 St, Edmonton, Alberta, Canada T5G 2R7.
Reprints: Gordon Guyatt, MD, MSc, McMaster University Health Sciences Centre, 1200 Main St W, Room 2C12, Hamilton, Ontario, Canada L8N 3Z5.
Users' Guides to the Medical Literature Section Editor: Drummond Rennie, MD, Deputy Editor (West), *JAMA*.

Table 1. Definitions of Drug Classes*

Definition	Example
Drugs with similar chemical structure	Dihydropyridine CCBs have dihydropyridine rings
Drugs with similar mechanism of action	CCBs block the voltage-dependent calcium channels on the surfaces of cell membranes
Drugs with similar pharmacologic effects	Antihypertensives (eg, CCBs, ACE inhibitors, β-blockers, thiazides, α-blockers) lower blood pressure

*CCB indicates calcium channel blocker; ACE, angiotensin-converting enzyme.

with the most attractive efficacy-to-safety ratio; purchasers, in the most cost-effective drug from a class; and manufacturers, in the most frequent prescribing of their drugs.

The absolute treatment effects seen with a drug (defined by the absolute risk reduction or number needed to treat) are influenced by the baseline risk or control event rate of those patients in whom it is used. Thus, the absolute risk reduction varies considerably among different groups of patients. On the other hand, the relative treatment effect of a drug (defined by the relative risk reduction [RRR]) is often (but not always[4]) similar, irrespective of the baseline risk of trial participants.[5,6] If 2 drugs are tested in separate placebo-controlled trials, only proportional effects such as the RRR resulting from each drug can be compared (and then only under the assumption of constant RRR over different control event rates). Although the point estimates of effect size vary, a class effect is considered to be present when drugs with similar mechanisms of action generate RRRs (or odds ratios [ORs]) that are similar in direction and magnitude. For example, the Collaborative Group on ACE Inhibitor Trials[7] suggested that there is a class effect for angiotensin-converting enzyme inhibitors in patients with symptomatic heart failure, despite the fact that the OR point estimates for effects on total mortality ranged from 0.14 (95% confidence interval [CI], 0-7.6) for perindopril (1 trial, 125 patients) to 0.78 (95% CI, 0.67-0.91) for enalapril (7 trials, 3381 patients). We are confident in this class effect, because the overall OR in 32 trials involving 7105 patients was 0.77 (95% CI, 0.67-0.88), the CIs for each of the angiotensin-converting enzyme inhibitors overlapped, and there was no statistical heterogeneity between trials of different agents.

Risks of Assuming a Class Effect

Although drugs of the same class typically exhibit similar pharmacological effects and clinical outcomes, this may not always be the case. Note the current controversy regarding the safety of sotalol hydrochloride in myocardial infarction survivors with congestive heart failure after the publication of the SWORD Trial,[8] which suggested an increase in mortality with sotalol, compared with the decrease in mortality with other β-blockers. It is useful to recall a previous controversy regarding the efficacy of β-blockers with intrinsic sympathetic activity (ISA) in patients with myocardial infarction. Although a meta-analysis[9] suggested that the treatment effect was greater with non-ISA β-blockers, subsequent trials[10] failed to confirm this, and the evidence[11] suggests there is little difference between β-blocker subgroups. It would seem reasonable to accept a priori that drugs within the same class exert similar effects, unless there is clear evidence of important differences.

However, this assumption can lead to 2 important errors of extrapolation with major clinical consequences. First, when agents in a class of drugs (such as the thiazide diuretics) all produce similar pharmacological effects (blood pressure lowering) and similar clinical effects (stroke reduction), a second class of drugs (for example, the calcium channel blockers) that produce the same pharmacological effects might be assumed to produce the same clinical benefits. In the absence of randomized trials verifying that final assumption, this type of extrapolation may be erroneous. For example, consider the issue raised in part A of this Guide—some calcium channel blockers have unfavorable effects on total mortality.[12] Second, even within the same class, individual drugs may have physiologic effects other than the mechanism of action that defined them as being from the same class. It therefore may be inaccurate to extrapolate the clinical outcomes shown in randomized trials of 1 drug in a class to another member of that class that has not been subjected to similar outcome-centered trials. For example, some authors have argued that, although all of the statins act on the 3-hydroxy-3-methylglutaryl coenzyme A reductase enzyme, they may have different nonlipid effects on the atherothrombotic process that may influence their clinical efficacy.[13]

To reduce the risk of faulty extrapolation and to maximize the optimal selection of treatments within a class of drugs, it may be useful to develop and apply a hierarchy of evidence when making decisions about the comparative clinical efficacy and safety of drugs within a class. As pointed out in part A of this Users' Guide, no matter how strong the pathophysiologic rationale or indirect evidence, the efficacy and safety of a new drug must be established in clinical outcome studies that test more than just biological plausibility.

Levels of Evidence

Levels of evidence are increasingly used by groups that make recommendations about patient care,[14-16] and we have used some of them to develop guidelines for comparing 1 drug with other drugs in the same class (TABLE 2). This comparison should occur as part of a systematic review of all the relevant evidence on the effects of a treatment, identified and assessed by thorough and clear methods such as those used in the Cochrane Collaboration [Update Software, Oxford, England; 1998]. We will describe each level in turn, using the choice of statin drugs as an example to illustrate their use (TABLE 3).

Level 1. Level 1 includes randomized clinical trials providing head-to-head comparisons of the drug of interest with other drugs of the same class for their effects on clinically important outcomes. This would generate the strongest evidence for the decision maker; however, there are potential

threats to validity (Table 2) and several methodologic issues unique to these trials. First, at least 1 of the drugs should have been shown to have a clinically important impact vs placebo in previous trials carried out in a population similar to that of the current trial. Second, the choice of appropriate dosage for each drug is a complicated issue, as this will affect the outcomes and safety profiles for both drugs. Finally, one must carefully consider the trial size and methods before concluding equivalence of 2 drugs—equivalence trials require much larger sample sizes than standard trials,[17] and any laxity in trial conduct or patient compliance will tend to mask any real differences between drugs.

The choice of clinically important outcomes for level 1 studies depends on the target intervention. In the case of therapies designed to prevent or arrest atherosclerosis (such as statins), this implies long-term efficacy data on events such as myocardial infarction, stroke, and all-cause mortality. On the other hand, for interventions designed to treat symptomatic diseases (such as gastroesophageal reflux disease), clinically important outcomes could include symptom scores and other quality-of-life measures.

Although there are examples of level 1 evidence in other branches of medicine,[18,19] they are rare in the cardiovascular literature. Our literature search failed to find any level 1 evidence for statins.

Level 2. Level 2 includes randomized clinical trials providing head-to-head comparisons of the drug of interest with other drugs of the same class for their effects on validated surrogate outcomes or comparisons across 2 or more placebo-controlled trials for effects on clinically important outcomes or validated surrogate outcomes. Part A of this Users' Guide discussed criteria for deciding whether to accept results of trials based on surrogate outcomes. Ecologic studies, cohort studies, and randomized clinical trials with prestatin lipid-lowering agents were supportive of the lipid-lowering hypothesis[20] (that lowering low-density lipoprotein [LDL] cholesterol levels lowers the risk of atherosclerotic heart disease); however, it was not until the publication of the large-scale statin trials[21-25] (Table 3) consistently linking reductions in LDL cholesterol to reductions in morbidity and mortality that we agreed to accept the surrogate end point of LDL cholesterol lowering as a proxy for clinically important outcomes. Thus, to accept head-to-head comparisons for surrogate outcomes as level 2 evidence, at least 1 of the comparators must have demonstrated efficacy in long-term trials with clinically important outcomes.

Whereas a randomized trial[26] comparing 4 statins for their effects on LDL cholesterol, high-density lipoprotein cholesterol, and triglycerides during an 8-week period would be an example of level 2 evidence, it also is important to incorporate considerations of the size and duration of trials in the decision-making process.

On the other hand, a number of level 2 comparisons can be made between various statins—for example, one can compare the treatment effects seen with simvastatin vs pravastatin in secondary prevention trials (such as the 4S[21] and LIPID[25] studies [Table 3]). Although consistency of effects in such comparisons would be strong evidence for the presence of a class effect, these comparisons are less useful in determining whether a drug is more efficacious than another,

Table 2. Levels of Evidence for Comparing the Efficacy of Drugs Within the Same Class*

Level	Comparison	Study Patients	Outcomes	Threats to Validity
1	Within a head-to-head RCT	Identical (by definition)	Clinically important	Failure to conceal randomization scheme Failure to achieve complete follow-up Failure to achieve double-blinding Soundness of outcome assessment
2	Within a head-to-head RCT	Identical (by definition)	Validated surrogate	Those of level 1 *plus* validity of surrogate outcome for clinically important outcomes
2	Across RCTs of different drugs vs placebo	Similar or different (in disease and risk factor status)	Clinically important or validated surrogate	Those of level 1 *plus* differences between trials in: Methodologic quality (adequacy of blinding, allocation concealment, etc) End point definitions Compliance rates Baseline risk of outcomes
3	Across subgroup analyses from RCTs of different drugs vs placebo	Similar or different	Clinically important or surrogate	Those of level 1 (*plus or minus* those of level 2) *plus:* Multiple comparisons, posthoc data dredging Underpowered subgroups Misclassification into subgroups
3	Across RCTs of different drugs vs placebo	Similar or different	Unvalidated surrogate	Surrogate outcomes may not capture all of the effects (beneficial or hazardous) of a therapeutic agent
4	Between nonrandomized studies (observational studies and administrative database research)	Similar or different	Clinically important	Confounding by indication, compliance, and/or calendar time Unknown/unmeasured confounders Measurement error For outcomes research: limited databases, coding systems not suitable for research

*Clinically important outcomes refer to long-term efficacy data, and the particular end points depend on the condition being treated. For statins used to prevent or treat atherosclerotic disease, clinically important outcomes would include all-cause mortality, myocardial infarction, and stroke. Surrogate outcomes are considered validated only when the relationship between the surrogate outcome and clinically important outcomes has been established in long-term randomized clinical trials (RCTs).

because the advantages of randomization are lost, and the comparison is essentially that between 2 or more cohorts. In addition to the potential biases outlined in Table 2, there is also the possibility of confounding a subject's risk or responsiveness with exposure to a particular treatment in those situations in which subjects from different trials have different risk statuses. For example, if one were to compare the statin used in a primary prevention trial (such as lovastatin in AFCAPS/TexCAPS[24]) with another statin tested in a secondary prevention trial (such as simvastatin in 4S[21]), such a comparison would only be valid if the drug efficacy is known to be independent of baseline risk, an assumption that appears valid to make in some situations (such as antiplatelet[6] or antihypertensive[5] therapy) but has been questioned for the statins.[27-32]

It is theoretically possible to compare the efficacy of 2 drugs tested in separate placebo-controlled trials. As outlined by Bucher et al,[33] an indirect estimate of the association between drugs A and B can be obtained by comparing the OR (or relative risk) from studies of drug A vs placebo (p) and the OR from studies comparing drug B vs placebo: $OR_{A\ vs\ B} = OR_{A\ vs\ p} / OR_{B\ vs\ p}$. However, this assumes that none of the potential biases outlined in Table 2 are operative and that an intervention's treatment effect is consistent across different patient subgroups. Furthermore, these indirect estimates may provide substantially different effect-size estimates than direct comparisons of drug A against drug B. For example, a systematic overview of strategies to prevent *Pneumocystis cari-*

Table 3. Features of Randomized, Placebo-Controlled Statin Trials Designed to Detect Differences in Clinically Important End Points*

	Trial				
	4S[21]	WOSCOPS[22]	CARE[23]	AFCAPS/TexCAPS[24]	LIPID[25]
Study design	Secondary prevention, multicenter	Primary prevention, single center	Secondary prevention, multicenter	Primary prevention, multicenter	Secondary prevention, multicenter
Treatment (dose once daily)	Simvastatin (20 mg)	Pravastatin (40 mg)	Pravastatin (40 mg)	Lovastatin (40 mg)	Pravastatin (40 mg)
Patient inclusion criteria†	Age 35-70 y, prior angina or AMI, fasting total cholesterol 5.5-8.0 mmol/L	Age 45-64 y, no prior AMI, fasting LDL cholesterol 4.0-6.0 mmol/L	Age 21-75 y, prior AMI, fasting LDL cholesterol 3.0-4.5 mmol/L	Age 45-73 y (males) or 55-73 y (females), no prior AMI, fasting LDL cholesterol 3.4-4.9 mmol/L	Age 31-75 y, prior AMI or unstable angina, fasting total cholesterol 4.0-7.0 mmol/L
Cointerventions, %					
Aspirin	37	None	83	None	82
β-Blockers	57	None	40	None	47
Duration of follow-up, y	5.4 (Median)	4.9 (Mean)	5.0 (Median)	5.2 (Mean)	6.1 (Mean)
Patients					
No.	4444	6595	4159	6605	9014
Mean age, y	58.6	55.2	59	58	62
Males, %	81	100	86	85	83
Smokers, %	26	44	21	12	10
Diabetes mellitus, %	5	1	15	2	9
Baseline cholesterol, mean mmol/L†					
Total	6.8	7.0	5.4	5.7	5.6
LDL	4.9	5.0	3.6	3.9	3.9
Control event rates, %					
Death	11.5	4.1	9.4	0.44	14.1
AMI	22.6	7.9	10	0.56	10.3
Treatment effects					
Change in lipids (active treatment vs placebo), %	−25 (Total cholesterol) −35 (LDL cholesterol) +8 (HDL cholesterol) −10 (Triglycerides)	−20 (Total cholesterol) −26 (LDL cholesterol) +5 (HDL cholesterol) −12 (Triglycerides)	−20 (Total cholesterol) −28 (LDL cholesterol) +5 (HDL cholesterol) −14 (Triglycerides)	−18 (Total cholesterol) −25 (LDL cholesterol) +6 (HDL cholesterol) −15 (Triglycerides)	−18 (Total cholesterol) −25 (LDL cholesterol) +5 (HDL cholesterol) −11 (Triglycerides)
Relative risk reductions, % (95% CI)					
Death	30 (15 to 42)	22 (0 to 40)	9 (−12 to 26)	−4 (Not given)	22 (13 to 31)
AMI	27 (20 to 34)	31 (17 to 43)	25 (8 to 39)	40 (17 to 57)	29 (18 to 38)
Number needed to treat‡					
To prevent 1 death	27 (5 y)	111 (5 y)	125 (5 y)	5000 to harm§	32 (6 y)
To prevent 1 AMI	10 (5 y)	42 (5 y)	40 (5 y)	435 (5 y)	34 (6 y)

*4S indicates Scandinavian Simvastatin Survival Study; WOSCOPS, West of Scotland Coronary Prevention Study; CARE, Cholesterol and Recurrent Events Trial; AFCAPS/TexCAPS, Air Force/Texas Coronary Atherosclerosis Prevention Study; LIPID, Long-term Intervention with Pravastatin in Ischaemic Disease Study; AMI, acute myocardial infarction; LDL, low-density lipoprotein; HDL, high-density lipoprotein; and CI, confidence interval.
†To convert cholesterol levels to milligrams per deciliter, divide by 0.02586.
‡Point estimates only. Years in parentheses indicate number of years needed to treat that number of patients to prevent 1 event.
§Since all-cause mortality was nonsignificantly increased in the active treatment arm, results are presented as number needed to treat to cause 1 death.

nii pneumonia in human immunodeficiency virus–positive patients documented that the indirect comparison of trimethoprim-sulfamethoxazole vs a combination of dapsone and pyrimethamine suggested a much larger effect size from trimethoprim-sulfamethoxazole (OR, 0.37; 95% CI, 0.21-0.65) than was seen in the direct comparisons (overall OR, 0.64 in the 9 trials of trimethoprim-sulfamethoxazole vs dapsone and pyrimethamine; 95% CI, 0.45-0.90).[33] Thus, the strength of inference from indirect comparisons is limited.

Level 3. Level 3 includes comparisons across subgroups from different placebo-controlled trials or comparisons across placebo-controlled trials in which outcomes are restricted to unvalidated surrogate markers. In addition to the biases that affect higher-level studies, comparisons based on subgroup analysis are potentially flawed (Table 2). Both simple statistics and experience have taught us that many initial subgroup conclusions (especially those that result from data-dredging) are subsequently disproven.[34,35] An example of such a comparison would be looking at the efficacy of simvastatin in the 4S subgroup with the lowest lipid levels (241 patients with total cholesterol levels of 5.5-6.24 mmol/L [213-241 mg/dL])[28] vs the efficacy of pravastatin in the CARE subgroup with comparable lipid profiles (2087 patients with total cholesterol levels of 5.4-6.21 mmol/L [209-240 mg/dL]).[23]

Level 3 evidence may also include the use of surrogate markers that, although they may lie along a recognized pathogenetic pathway from mechanisms of action to important clinical outcomes, have not been validated in long-term randomized clinical trials. To return to an example cited in part A of this Users' Guide, this would involve making inferences about reductions in fractures from the effects on bone density of 2 different bisphosphonates in 2 independent randomized trials.

Level 4. Level 4 includes comparisons involving or confined to nonrandomized evidence. This type of evidence is only possible for conditions in which there are a large number of potential treatments commonly used by practitioners. Nonrandomized evidence can include cohort or case-control studies, modeling studies (using risk-prediction equations such as those derived from the Framingham data[36]), and/or outcomes research using administrative databases. Although these types of analyses can provide useful insights (particularly with respect to dose-response relationships),[37] they are best viewed as exercises in hypothesis-generation. In particular, outcomes research studies, originally developed to determine whether the efficacy of interventions proven in randomized trials have their anticipated impacts at a population level, have sometimes been used to pursue the primary determination of efficacy—a purpose for which they were not intended. When used to establish efficacy, they present, in addition to other limitations (Table 2), unique problems in interpretation that restrict the validity of inferences drawn from them about the relative efficacy of medications from the same class.[38]

An example of level 4 evidence is a recent reanalysis of the WOSCOPS database, designed to infer whether pravastatin's efficacy exceeds that expected of other statins.[29] Using the constellation of risk factors and mean on-treatment cholesterol levels seen in the trial, the observed coronary event rates in pravastatin-treated patients were compared with those predicted from the Framingham coronary risk equation to determine whether the treatment benefit with pravastatin exceeded that expected from the degree of cholesterol lowering achieved.

Level 3 and 4 studies have numerous flaws as outlined above and are best viewed as exercises in hypothesis generation.

Other Considerations

Amount of Efficacy Evidence. While we have thus far focused on the validity of the evidence, the number, size, and duration of studies are essential factors to be considered in the decision-making process. Certainly, the superiority of 1 drug within a class can only be definitively established with level 1 evidence. However, while level 1 evidence would be ideal for establishing that a group of drugs exert a class effect (by showing narrow confidence limits around the difference between drugs), we recognize that it is rarely available and is unlikely to ever be available for many classes of drugs because of difficulties in funding and conducting trials so large that they are unlikely to appeal to researchers, manufacturers, or funders. In this situation, the amount of level 2 evidence becomes important. For instance, one would feel more comfortable in concluding that a drug produced a class effect if there were a number of placebo-controlled trials demonstrating that various drugs from the same class had similar treatment effects. However, our goal is not to set a level that must be achieved before a drug can be claimed to be superior to others in its class or before a class effect can be established. Those are decisions that individual clinicians or policymakers must make, taking into account their local circumstances and individual comfort levels.

Safety. In the past decade, there have been numerous examples of drugs within the same class that have been shown to have different safety profiles. Although not our primary focus, considerations of drug safety are part of any treatment or purchasing decision, so we offer a set of levels of evidence for determining drug safety in TABLE 4. Phase 1 drug studies in humans are designed to determine the maximally tolerated dose, and clinical trials are generally designed to determine the efficacy of the drug. As such, the sample sizes of neither are adequate to detect uncommon adverse effects. The inverse rule of 3 states that to be 95% sure of seeing at least 1 adverse drug reaction that occurs once in every given number of patients, you need to follow up 3 times that many patients.[39] Given the size and duration of most clinical trials, adverse effects that occur in fewer than 1 in 1000 participants or that take more than 6 months to appear will generally remain undetected.[3] However, randomized clinical trials are still the strongest design for detecting real differences in adverse effects (such as the different rates of intracranial bleeding with different thrombolytic agents[40,41]), and meta-analyses of such

trials can give unbiased estimates of excess hazards. In the absence of clinical trials, premarketing safety data must be considered preliminary, and large, phase 4 studies or systematic postmarketing surveillance data are necessary to confirm the safety of new drugs.

Convenience/Compliance. While once-a-day medications are more convenient and usually have higher compliance rates, evidence about drug compliance derived from trials may translate poorly in clinical practice. For instance, while compliance with the various statins described in Table 3 ranged from 90% to 94% during the course of the trials, analyses of administrative databases in Canada and the United States[42] revealed that only half of statin-treated patients were still taking their medication 1 year after it was prescribed.

Cost. Faced with a decision as to whether a new drug from a class should be offered to eligible patients within the population, clinicians and policymakers have different perspectives. For clinicians, this decision usually hinges on the efficacy, safety, convenience or compliance, cost of the new drug vs the old, and the applicability of the trial evidence to their patients.[43] However, for policymakers, these issues form only 1 piece of the puzzle. They also must evaluate the efficiency, affordability, and opportunity costs of any new drugs. The efficiency of any intervention is determined by formal economic analyses, and the Users' Guides series offers criteria for evaluating methodological quality.[44] Although cost-minimization analysis is the simplest and least controversial of the economic analysis techniques, it requires proof that the outcomes resulting from both alternatives are the same. As this rarely exists, the policymaker must rely on other types of analyses (cost-effectiveness, cost-benefit, or cost-utility analyses) that involve varying degrees of assumption and guesswork. As pointed out by Naylor and colleagues,[45] economic analyses should be viewed as "promising, clearly helpful, still in need of refinement and open, like any new technology, to both wise use and well-intentioned abuse."

The decision as to whether a new drug is efficient enough to warrant its adoption depends critically on the social, political, and economic realities of the particular health care setting, complicating the policymaker's task. Thus, attempts to establish universal cut-points (using cost or quality-adjusted life-year ratios) have been largely unsuccessful.[46] Although there are occasions for which there is compelling evidence for a new drug's adoption (the new drug is as effective or more effective than others of its class and is less costly) or rejection (the new drug is less effective than others of its class and is more costly), the policymaker operates most often in a cost-utility gray zone between these 2 extremes.[45]

Table 4. Levels of Evidence for Comparing the Safety of Drugs Within the Same Class

Level	Type of Study	Advantages	Threats to Validity
1	Randomized clinical trial(s)	Only design that permits the detection of adverse effects when the adverse effect is similar to the event that treatment is trying to prevent	Underpowered for detecting adverse effects
2	Cohort	Prospective data collection, defined cohort	Critically depends on follow-up, classification, and measurement accuracy
3	Case-control	Cheap and fast to perform	Selection and recall bias; temporal relationship may not be clear
4	Phase 4	If sufficiently large, can detect rare but important adverse effects	No, or unmatched, control group; critically depends on follow-up, classification, and measurement accuracy
5	Case series	Cheap and fast to perform	Small sample size; selection bias; no control group
6	Case report(s)	Cheap and fast to perform	Small sample size; selection bias; no control group

RESOLUTION OF SCENARIOS

The Clinician

Given the qualitative consistency of the RRR for acute myocardial infarction in patients treated with 3 of the statins in large trials with clinically important outcomes (Table 3) and the convincing nature of LDL cholesterol lowering as a surrogate outcome,[20,30,47-49] our clinician concludes that there is a class effect of statin drugs on the occurrence of ischemic heart disease. In the apparent absence of differences in safety or compliance profile between the various statins, he decides to pursue a cost-minimization strategy. While the newer statin has been evaluated only for cholesterol-lowering efficacy in a short-term trial (<6 months), he decides to prescribe it because it is the cheapest statin in his local setting.

The Policymaker

The policymaker agrees with the clinician that the statins appear to exert a class effect in terms of efficacy. However, she is concerned that the efficacy of the newer statin has not been evaluated in long-term trials with clinically important outcomes or validated surrogate outcomes. Thus, she decides to keep the older (and more expensive) statin on her formulary until level 1 or long-term level 2 evidence is available that proves that the newer statin is as good as or better than the currently provided statin.

CONCLUSION

While it would be preferable that every drug in each class (and indeed every dose and every formulation) be evaluated in randomized clinical trials with active comparators from the same class for its effects on clinically important outcomes, this has not been accomplished for several important classes of drugs. We believe that advocates of newer drugs within a class must provide evidence of equivalence (or superiority) to the older agents and "randomized comparative trials . . . remain the preferred evidentiary standard."[50] Recognizing that this criterion standard is not always attainable (in the case of

the statins, such randomized clinical trials would require very large sample sizes and long follow-up to detect significant differences in myocardial infarction or death between 2 different statins), we suggest that discussions about class effects will benefit from citing the levels of evidence behind the arguments and recognizing the strengths and weaknesses inherent in each study design.

Funding/Support: Dr McAlister is supported by the Medical Research Council of Canada and the Alberta Heritage Foundation for Medical Research, Dr Laupacis is supported by the Medical Research Council of Canada, and Dr Sackett is supported by the Research and Development Programme of the National Health Service, United Kingdom.

Acknowledgment: We gratefully acknowledge the input of various members of the University of Ottawa Clinical Epidemiology Unit and attendees of the EQUINOX symposium for earlier discussions of this topic, J. Glennie, PharmD, for providing information on the approaches of regulators to these issues, and Ian Chalmers, MB, ChB, and F. Lawson, MB, ChB, for reviewing earlier versions of this article.

REFERENCES

1. Skolnick AA. Drug firm suit fails to halt publication of Canadian Health Technology Report. *JAMA*. 1998;280:683-684.
2. Bucher HC, Guyatt GH, Cook DJ, Holbrook A, McAlister FA, for the Evidence-Based Medicine Working Group. Users' guides to the medical literature, XIX: applying clinical trial results, A: how to use an article measuring the effect of an intervention on surrogate end points. *JAMA*. 1999;282:771-778.
3. McDonald CJ. Medical heuristics. *Ann Intern Med*. 1996;124:56-62.
4. Rothwell PM. Can overall results of clinical trials be applied to all patients? *Lancet*. 1995;345:1616-1619.
5. Collins R, Peto R, MacMahon S, et al. Blood pressure, stroke, and coronary heart disease, part 2: short-term reductions in blood pressure. *Lancet*. 1990;335:827-838.
6. Antiplatelet Trialists' Collaboration. Collaborative overview of randomised trials of antiplatelet therapy—I [published correction appears in *BMJ*. 1994;308:1540]. *BMJ*. 1994;308:81-106.
7. Garg R, Yusuf S, for the Collaborative Group on ACE Inhibitor Trials. Overview of randomized trials of angiotensin-converting enzyme inhibitors on mortality and morbidity in patients with heart failure [published correction appears in *JAMA*. 1995;274:462]. *JAMA*. 1995;273:1450-1456.
8. Waldo AL, Camm AJ, de Ruyter H, et al, for the SWORD Investigators. Effect of d-sotalol on mortality in patients with left ventricular dysfunction after recent and remote myocardial infarction [published correction appears in *Lancet*. 1996;348:416]. *Lancet*. 1996;348:7-12.
9. Yusuf S, Peto R, Lewis J, Collins R, Sleight P. Beta blockade during and after myocardial infarction. *Prog Cardiovasc Dis*. 1985;27:335-371.
10. Boissel JP, Leizorovicz A, Picolet H, Peyrieux JC. Secondary prevention after high-risk acute myocardial infarction with low-dose acebutolol. *Am J Cardiol*. 1990;66:251-260.
11. McAlister FA, Teo KK. Antiarrhythmic therapies for the prevention of sudden cardiac death. *Drugs*. 1997;54:235-252.
12. Furberg CD, Psaty BM. Calcium antagonists. *Am J Hypertens*. 1996;9:122-125.
13. Rosenson RS, Tangney CC. Antiatherothrombotic properties of statins. *JAMA*. 1998;279:1643-1650.
14. Guyatt GH, Sackett DL, Sinclair JC, Hayward RC, Cook DJ, Cook RJ, for the Evidence-Based Medicine Working Group. Users' guides to the medical literature, IX: a method for grading health care recommendations [published correction appears in *JAMA*. 1995;275:1232]. *JAMA*. 1995;274:1800-1804.
15. Canadian Task Force on the Periodic Health Examination. The Periodic Health Examination, 2: 1987 update. *CMAJ*. 1988;138:618-626.
16. Cook DJ, Guyatt GH, Laupacis A, Sackett DL, Goldberg RJ. Clinical recommendations using levels of evidence for antithrombotic agents. *Chest*. 1995;108(4 suppl):227S-230S.
17. Donner A. Approaches to sample size estimation in the design of clinical trials. *Stat Med*. 1984;3:199-214.
18. Naguib M, el Bakry AK, Khoshim MH, et al. Prophylactic antiemetic therapy with ondansetron, tropisetron, granisetron and metoclopramide in patients undergoing laparoscopic cholecystectomy. *Can J Anaesth*. 1996;43:226-231.
19. Korttila K, Clergue F, Leeser J, et al. Intravenous dolasetron and ondansetron in prevention of postoperative nausea and vomiting. *Acta Anaesthesiol Scand*. 1997;41:914-922.
20. Law MR, Wald NJ, Thompson SG. By how much and how quickly does reduction in serum cholesterol concentration lower risk of ischaemic heart disease? *BMJ*. 1994;308:367-373.
21. Scandinavian Simvastatin Survival Study Group. Randomised trial of cholesterol lowering in 4444 patients with coronary heart disease. *Lancet*. 1994;344:1383-1389.
22. Shepherd J, Cobbe SM, Ford I, et al, for the West of Scotland Coronary Prevention Study Group. Prevention of coronary heart disease with pravastatin in men with hypercholesterolemia. *N Engl J Med*. 1995;333:1301-1307.
23. Sacks FM, Pfeffer MA, Moye LA, et al, for the Cholesterol and Recurrent Events Trial Investigators. The effect of pravastatin on coronary events after myocardial infarction in patients with average cholesterol levels. *N Engl J Med*. 1996;335:1001-1009.
24. Downs JR, Clearfield M, Weis S, et al, for the AFCAPS/TexCAPS Research Group. Primary prevention of acute coronary events with lovastatin in men and women with average cholesterol levels: results of AFCAPS/TexCAPS. *JAMA*. 1998;279:1615-1622.
25. The Long-term Intervention with Pravastatin in Ischaemic Disease (LIPID) Study Group. Prevention of cardiovascular events and death with pravastatin in patients with coronary heart disease and a broad range of initial cholesterol levels. *N Engl J Med*. 1998;339:1349-1357.
26. Jones P, Kafonek S, Laurora I, Hunninghake D. Comparative dose efficacy study of atorvastatin versus simvastatin, pravastatin, lovastatin, and fluvastatin in patients with hypercholesterolemia [published correction appears in *Am J Cardiol*. 1998;82:128]. *Am J Cardiol*. 1998;81:582-587.
27. Sacks FM, Moye LA, Davis BR, et al. Relationship between plasma LDL concentrations during treatment with pravastatin and recurrent coronary events in the Cholesterol and Recurrent Events Trial. *Circulation*. 1998;97:1446-1452.
28. Scandinavian Simvastatin Survival Study Group. Baseline serum cholesterol and treatment effect in the Scandinavian Simvastatin Survival Study (4S). *Lancet*. 1995;345:1274-1275.
29. West of Scotland Coronary Prevention Study Group. Influence of pravastatin and plasma lipids on clinical events in the West of Scotland Coronary Prevention Study (WOSCOPS). *Circulation*. 1998;97:1440-1445.
30. Fager G, Wiklund O. Cholesterol reduction and clinical benefit: are there limits to our expectations? *Arterioscler Thromb Vasc Biol*. 1997;17:3527-3533.
31. Smith GD, Song F, Sheldon TA. Cholesterol lowering and mortality [published correction appears in *BMJ*. 1993;306:1648]. *BMJ*. 1993;306:1367-1373.
32. Sacks FM, Gibson CM, Rosner B, Pasternak RC, Stone PH, for the Harvard Atherosclerosis Reversibility Project Research Group. The influence of pretreatment low density lipoprotein cholesterol concentrations on the effect of hypocholesterolemic therapy on coronary atherosclerosis in angiographic trials. *Am J Cardiol*. 1995;76:78C-85C.
33. Bucher HC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *J Clin Epidemiol*. 1997;50:683-691.
34. Oxman AD, Guyatt GH. A consumer's guide to subgroup analyses. *Ann Intern Med*. 1992;116:78-84.
35. Yusuf S, Wittes J, Probstfield J, Tyroler HA. Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials. *JAMA*. 1991;266:93-98.
36. Anderson KM, Wilson PW, Odell PM, Kannel WB. An updated coronary risk profile: a statement for health professionals. *Circulation*. 1991;83:356-362.
37. Psaty BM, Siscovick DS, Weiss NS, et al. Hypertension and outcomes research: from clinical trials to clinical epidemiology. *Am J Hypertens*. 1996;9:178-183.
38. Marshall WJ. Administrative databases: fact or fiction? *CMAJ*. 1998;158:489-490.
39. Sackett DL, Haynes RB, Gent M, et al. Compliance. In: Inman WH, ed. *Monitoring for Drug Safety*. Philadelphia, Pa: Lippincott; 1980:427-438.
40. Gruppo Italiano per lo Studio della Sopravvivenza nell' Infarto Miocardico. GISSI-2: a factorial randomised trial of alteplase versus streptokinase and heparin versus no heparin among 12,490 patients with acute myocardial infarction. *Lancet*. 1990;336:65-71.
41. Third International Study of Infarct Survival Collaborative Group. ISIS-3: a randomised comparison of streptokinase vs tissue plasminogen activator vs anistreplase and of aspirin plus heparin vs aspirin alone among 41,299 cases of suspected acute myocardial infarction. *Lancet*. 1992;339:753-770.
42. Avorn J, Monette J, Lacour A, et al. Persistence of use of lipid-lowering medications: a cross-national study. *JAMA*. 1998;279:1458-1462.
43. Dans AL, Dans LF, Guyatt GH, Richardson S, for the Evidence-Based Medicine Working Group. Users' guides to the medical literature, XIV: how to decide on the applicability of clinical trial results to your patient. *JAMA*. 1998;279:545-549.
44. Drummond MF, Richardson WS, O'Brien BJ, Levine M, Heyland D, for the Evidence-Based Medicine Working Group. Users' guides to the medical literature, XIII: how to use an article on economic analysis of clinical practice, A: are the results of the study valid? *JAMA*. 1997;277:1552-1557.
45. Naylor CD, Williams JI, Basinski A, Goel V. Technology assessment and cost-effectiveness analysis: misguided guidelines? *CMAJ*. 1993;148:921-924.
46. Laupacis A, Feeny D, Detsky AS, Tugwell PX. How attractive does a new technology have to be to warrant adoption and utilization? *CMAJ*. 1992;146:473-481.
47. Gotto AM Jr. Cholesterol management in theory and practice. *Circulation*. 1997;96:4424-4430.
48. Grover SA, Paquet S, Levinton C, Coupal L, Zowall H. Estimating the benefits of modifying risk factors of cardiovascular disease [published correction appears in *Arch Intern Med*. 1998;158:1228]. *Arch Intern Med*. 1998;158:655-662.
49. Lacour A, Derderian F, LeLorier J. Comparison of efficacy and cost among lipid-lowering agents in patients with primary hypercholesterolemia. *Can J Cardiol*. 1998;14:355-361.
50. Tu JV, Naylor CD. Choosing among drugs of different price for similar indications [published correction appears in *Can J Cardiol*. 1998;14:662]. *Can J Cardiol*. 1998;14:349-351.