SUMMARY. Taking a careful look at each of the outcomes measured in randomized, controlled trials of digoxin suggest that discrepancies in results may be more apparent than real. Digoxin does work, but clinically important benefit is restricted to a relatively small proportion of congestive heart failure (CHF) patients. The play of chance, the dose of digoxin used, and the severity of heart failure in patients enrolled in the studies are other factors that may explain the variability in results that were observed. A systematic examination of the sort undertaken here is likely to help resolve apparent differences in outcomes of clinical trials of new (and old) therapies in CHF patients.

KEY WORDS. digoxin, digitalis, congestive heart failure, randomized controlled trials

Digitalis has been the mainstay of therapy for congestive heart failure (CHF) for more than two centuries. Nevertheless, use of the drug in patients in sinus rhythm remains controversial. The controversy has been fueled by the apparent conflicting results of trials of digoxin in CHF patients. In this article we shall present a case suggesting that the results of digoxin trials, in fact, provide a clear and consistent message regarding the effectiveness of digoxin and the patients in whom the drug is of benefit. We shall begin by briefly reviewing the reasons why trials of drug efficacy might differ, using examples from the literature concerning digoxin in CHF patients in sinus rhythm, before systematically reviewing the major trials of digoxin. We have framed the possible explanations for discrepant results as a series of five questions, which are summarized in Table 1.

Possible Explanations for Discrepant Results in Drug Trials

Who Were the Patients Included in the Study?

While a number of definitions of heart failure have been suggested, none of them are very specific. While there would be no disagreement about patients with completely normal cardiac function, or about those with pulmonary edema, it is possible that classification of patients with mild to moderate impairment of cardiac function would differ. A recent review of inclusion criteria in clinical trials of CHF stressed the heterogeneity of diagnostic criteria [1]. Of 51 randomized control trials (RCTs) reviewed, only 23 (45%) specified criteria for the diagnosis of CHF. Only four of 23 trials used the same criteria, and these four were done by the same research team. Under these circumstances, it is possible that very different patients are being included in the clinical trials.

Table 1. Questions to ask when seeking explanations for discrepant results in drug trials

<table>
<thead>
<tr>
<th>Question</th>
<th>1. Who were the patients in the study?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2. What was the intervention?</td>
</tr>
<tr>
<td></td>
<td>3. What was the study design?</td>
</tr>
<tr>
<td></td>
<td>4. What was the sample size of the trial?</td>
</tr>
<tr>
<td></td>
<td>5. What were the measures of outcome?</td>
</tr>
</tbody>
</table>

Even assuming that inclusion criteria were similar, it is still possible for a different spectrum of patients to be included in different studies. Patients differ with respect to etiology (hypertension, coronary artery disease and myocardial infarction, valvular heart disease, etc.); physiology (systolic versus diastolic dysfunction); and severity. It is quite possible that digoxin could have different effects on patients whose etiology, physiology, and severity of disease differ. We will argue that digoxin is of symptomatic benefit only in more severely affected patients.

A final issue in patient selection is the bias that could intrude when a subpopulation of eligible patients are enrolled. Let us assume that most members of the eligible population have been exposed to digoxin. Those who had perceived a benefit from the drug might be reluctant to participate, while those who thought...
the drug wasn’t doing any good would be more willing. If the patients’ perceptions are correct, a subgroup of nonresponders could be enrolled and result in a false negative trial [2]. Results of trials could differ depending on the extent to which this bias was avoided. This mechanism was certainly operating in one parallel group trial of digoxin, captopril, and placebo in CHF: Patients who could not tolerate digoxin withdrawal were excluded from the study [3]. While it is impossible to judge the extent of bias from this mechanism without a detailed account of the patients who refused participation and the reasons for refusal (an account that only two trials [4,5] have provided), this mechanism may explain some of the variation remaining after consideration of the other factors we will invoke.

What Was the Intervention?

While all the trials we will consider tested the efficacy of digoxin, the dose and the methods of drug monitoring varied. If the effect of digitalis is dose related, studies using lower doses could underestimate drug efficacy (or side effects). We will suggest that one explanation for discrepant results in digoxin trials may be the administration of suboptimal doses in “negative” trials.

What was the Study Design?

The problems with inadequate study design in drug evaluation are now generally recognized. The potent placebo effect, an expectation of drug effectiveness on the part of both patient and investigator, the natural history of the condition (spontaneous improvement may occur independent of the intervention), and bias in allocation (patients given active drug may be destined to do better than patients not given the drug, irrespective of therapeutic effect) are all well known. These problems can be minimized only by random allocation of patients to alternative therapies and by keeping patients and care-givers unaware of patient allocation. Thus, we will consider only double-blind, randomized, controlled trials in our analysis.

Randomization and double-blinding is not sufficient to guarantee validity; a series of lesser pitfalls remain in carrying out randomized, controlled trials. One such problem is cointervention, the administration of interventions other than those under consideration. If cointervention is administered unequally to the treatment groups, bias can result. This occurred in a controlled trial of digoxin, captopril, and placebo [3] in which modification of diuretic therapy was allowed. Such modification occurred more frequently in the placebo than the digoxin group, and may have contributed to the negative findings in some major outcomes, such as exercise capacity, in the comparison of digoxin and placebo. Allowing diuretic changes may have been appropriate if the investigators were interested in the management question—Does digoxin benefit patients in the “natural” clinical setting in which diuretic adjustment will inevitably be made—rather than the explanatory question—When all else is held equal, can digoxin provide benefit? Nevertheless, cointervention such as alterations in diuretic therapy may be an important explanation of differing results of randomized, controlled trials.

What was the Sample Size of the Trial?

One way to guarantee a “negative” result for a trial is to choose an insufficient sample size. Using a crossover design, Taggart et al. found that 4 of 22 patients deteriorated while on placebo, whereas 2 of 22 deteriorated on digoxin. This is a negative trial insofar as the results could easily be explained by chance (i.e., conventional statistical significance was not achieved) [5]. On the other hand, the results suggest a 50% reduction in the rate of treatment failure on digoxin, finding that would represent an important benefit were it true. Whenever one has positive and negative randomized, controlled trials, one of the first explanations that should be considered is inadequate sample size in the negative trials.

What Were the Measures of Outcome?

The results of a study may differ depending on the measures of outcome one chooses. One of the most dramatic examples of this phenomenon comes from the heart failure literature. In a randomized, controlled trial of minoxidil versus placebo, Franciosa et al. found that left ventricular ejection fraction improved in actively treated patients, while the probability of clinically worsening CHF and death was increased [6].

Ideally, outcome measures chosen should be both clinically relevant to both patients and physicians, and responsive to small but still clinically important differences [7,8]. The latter issue is illustrated in a comparison of two crossover randomized, controlled trials of digoxin. In a study already mentioned, Taggart found that two patients deteriorated while on active drug, while four deteriorated while on placebo [5]. Lee reported that 6 of 25 patients deteriorated while on active drug, while 9 deteriorated while on placebo [9]. These findings appear very similar; nevertheless, Taggart and colleagues reported their study as essen-