Herbal Medicines Today and the Roots of Modern Pharmacology

Peter Goldman, MD

The transformation of digitalis from a folk medicine, foxglove, to a modern drug, digoxin, illustrates principles of modern pharmacology that have helped make drugs safer and more effective. Digitalis was improved because its preparation was standardized, first by bioassay and then by chemical methods; however, few of today’s herbs are standardized by methods that can ensure a consistent product and, hence, consistent safety and efficacy profiles. Many herbs have been evaluated in randomized, controlled trials, and several—St. John’s wort and ginkgo, for example—are apparently effective. Yet, many trials of herbs have limited value because of poor design, small samples, and, above all, use of products of uncertain composition and consistency. The uncertain composition of many herbal products raises questions about their safety, as does evidence indicating that herbs may have harmful interactions with prescription drugs. Such adverse effects of herbs are probably underreported. Meanwhile, systematic studies, such as those identifying adverse reactions to drugs, are hindered because herbal preparations are not standardized—one brand of St. John’s wort, for example, will differ chemically from another—and, unlike for prescription drugs, there are no databases linking herb consumption to later medical problems. Since herbal medicines are regulated as dietary supplements, they are not subject to the premarketing regulatory clearance required for drugs. The burden of proof is on the U.S. Food and Drug Administration to show a dietary supplement is unsafe, unlike for drugs, which cannot be approved until the manufacturer has demonstrated safety and effectiveness.


For the author affiliation and current address, see end of text.

The desire to take medicines is one feature which distinguishes man, the animal, from his fellow creatures” (1). Thus did William Osler express skepticism about remedies available in the early 20th century and an avuncular indulgence toward patients who wanted them. His comment expresses the attitude of many physicians today toward consumers of herbal medicines, and indeed may be timeless: Medicinal herbs were found in the personal effects of an “ice man,” whose body was frozen in the Swiss Alps for more than 5000 years (2). Since these herbs appear to have treated the parasites found in his intestine (2), “the desire to take medicines” may signify a timeless quest for cures that flowers today in the form of widely acclaimed new drugs.

The effectiveness of a modern drug is ultimately judged by the results of clinical trials. Ordinarily, such trials are designed to test the assumption that a drug’s pharmacologic activity will favorably affect a disease process, which in turn is viewed in terms of a physiologic model. Clinical trials yield convincing results, however, only if they are conducted in accordance with principles that, for example, ensure elimination of bias and reduce the possibility that results occurred merely by chance. Trials must also use drug preparations with consistent pharmacologic properties. These principles apply to all drugs, whether they originate as traditional remedies or in precepts of molecular biology. Indeed, such principles have successfully guided digitalis from medicinal plant to modern drug; we might ask, therefore, how these principles apply to the evaluation of today’s herbal medicines.

**Digitalis: From Folk Remedy to Modern Drug**

Withering, who introduced foxglove to the medical profession in 1785 (3), took the first steps in transforming digitalis from a folk remedy to a modern drug when he simplified a “family receipt for dropsy” that contained more than 20 substances (3) by assuming that foxglove was the active ingredient. Careful clinical observations then enabled him to recognize the plant’s slim margin of safety and thus the importance of dose: just enough foxglove to cause diuresis, but not enough to cause vomiting or very slow pulse.

**Bioassays and Chemical Standardization**

By the early 20th century, it was understood that activities of medicines derived from foxglove were influenced by such factors as “the time when the leaves are gathered, and... climatic and soil conditions...” as well as the manner in which the drug is prepared for the market” (4). Clearly, plants have ingredients with therapeutic activity, but their preparations must be standardized to yield consistent products, which therefore can be given in doses that are maximally safe and effective. In 1906, the pharmacopeia contained a daunting
number of digitalis preparations—for example, “Digitalis,” “Extractum Digitalis,” and “Infusum Digitalis” (5)—whose potency had never been investigated. When these preparations were investigated by using a new bioassay based on the fact that digitalis causes asystole in the frog, the results were surprising: The potencies of 16 commercial digitalis preparations varied over a fourfold range (4). Fortunately, the bioassay also provided a way to control this problem, and the frog bioassay was soon officially adopted by the United States Pharmacopeia to standardize digitalis preparations.

This bioassay, which indicated the importance of laboratory studies for the emerging science of pharmacology, provided the means to standardize the potency of a chemically complex herbal medicine, even when its active ingredients were uncertain. Soon the quest for even better methods of standardizing digitalis yielded several dozen bioassays in more than six different animal species (6). Thus, the cat heart assay replaced the frog heart assay, which in turn was replaced by the pigeon assay. The ultimate bioassay, however, was done in humans; it was based on the digitalis-induced changes in a patient’s electrocardiogram (7). Although digoxin, now the preferred form of digitalis, can be standardized chemically, a bioassay of sorts is still required to establish its bioavailability (8) and, hence, the pharmaceutical standardization needed to carry out the clinical trials that shape our current perspective on the drug (9).

HERBAL REMEDIES IN THE UNITED STATES TODAY

Challenges in Standardizing Herbal Medicines

Unfortunately, standardization methods such as those described for digitalis are not suitable for many herbs. Bioassays must be based on biological models, which are not available for the health claims made for many popular herbs, and chemical analysis has limited value when the ingredients responsible for a plant’s activity have not been identified. In addition, if the active ingredient of an herb were known, it would remain unclear whether the crude herb would be preferable to its purified active principle. In the absence of definitive information in this regard, such traditional herbal preparations as digitalis leaf and opium have been replaced by such drugs as digoxin and codeine, respectively.

How can an herb be standardized if its active ingredients are not known and there is no suitable bioassay? EGB 761, a patented extract of *Ginkgo biloba*, is a commendable attempt to solve this problem and to achieve a consistent formulation of ginkgo. Thus, EGB 761 sets feasible standards for how and where ginkgo is grown and harvested, how the leaves are extracted, and the target values for several chemical constituents of the medicinal product (10). EGB 761, which aims for chemical consistency and, presumably, therapeutic consistency, was used in three of four studies that, on the basis of a meta-analysis, concluded that ginkgo conferred a small but significant benefit in patients with Alzheimer disease (11). In the absence of evidence to the contrary, those who hope to replicate these trial results would justifiably select this ginkgo product in preference to others with less well-specified standards of botanical and chemical consistency.

Recent studies with St. John’s wort, however, remind us of the potential pitfalls of standardizing a medicinal herb to constituents that may not be responsible for therapeutic activity. For years, St. John’s wort, which meta-analysis finds superior to placebo for treatment of mild to moderate depression (12), has been standardized by its content of hypericin. Hypericin, however, has never been confirmed as the herb’s active ingredient and may be no more than a characteristic ingredient of the plant, useful for botanical verification but not necessarily for therapeutic standardization.

Another constituent of St. John’s wort, hyperforin, now appears to be a more potent antidepressant than hypericin. Thus, the potency of various St. John’s wort extracts for inhibiting the neuronal uptake of serotonin, a characteristic of conventional antidepressants such as fluoxetine, increases with increasing hyperforin content. Studies in animal models of depression (13) and patients with mild to moderate depression (14) suggest that antidepressant activity is related to content of hyperforin, not hypericin. For example, a three-arm clinical trial of 147 patients that compared two St. John’s wort extracts of equal hypericin content with placebo found antidepressant activity to be higher for the extract that had a 10-fold higher hyperforin content (14).

Although this trial was relatively small and therefore of limited statistical significance, its results suggest that antidepressant activity demonstrated in a meta-analysis of past studies (12) may have resulted from the fortuitous inclusion of hyperforin in many of the St. John’s wort formulations included. If the active ingredient of
St. John’s wort products used in these studies was not optimized, the studies as a group would undoubtedly underestimate the potential antidepressant activity of St. John’s wort.

Additional evidence suggests that the consumer is not receiving the full possible benefit of St. John’s wort. On a recent visit to a local food store, I found St. John’s wort preparations that were reminiscent of digitalis formulations at the beginning of the 20th century. Some were said to contain “0.3% (300 mg)” hypericin, another was a liquid formulation containing 180 mg of “hypericins,” and a third contained “0.3% (450 mg)” hypericin. The highest content of “0.3% hypericin” was 530 mg. Yet another product carried the label “St. John’s wort,” but its contents were not quantified. Hyperforin content was listed only for some products, whereas other products indicated that St. John’s wort had been combined with such ingredients as kava, Echinacea, licorice root, or coconut. The parts of the plant used in the preparations were described as “leaf, flowers and stem,” “aerial parts,” or simply “flowers and leaf.” Although labels on some St. John’s wort products indicated an awareness of recent studies on hyperforin, other labels confirmed that there is no barrier to selling herbal preparations of doubtful scientific rationale and uncertain potency.

Clinical Trials of Herbs

Randomized clinical trials have become the gold standard for evaluating the efficacy of a drug and have assumed a similar status for evaluating an herbal remedy. Although the methodology of herbal trials is improving, some studies cited in herbal compendia have shortcomings. One problem is that results of herbal trials often do not reach statistical significance because they enroll fewer participants than trials of a conventional drug, and the role of chance may be overlooked in interpreting such trials. For example, the results of clinical studies were recently examined to determine whether parthenolide, a characteristic component of feverfew, was necessary for feverfew’s apparent role in prevention of migraine. It was reasoned (15) that parthenolide could not be the sole active ingredient of feverfew because the parthenolide content of the feverfew preparation used in one negative trial (16) was twice that of another trial with positive results (17). However, because there were only about 50 participants in each trial, the observed difference cannot be considered a statistically significant argument against the activity of parthenolide.

Yet, when the active principle of an herb is not known and there is no accepted method of standardization, a clinical trial offers an attractive approach to evaluating the activity of an herbal preparation. In other words, instead of standardizing a medicinal herb before it is tested in a clinical trial, the results of a clinical trial might be used to identify an effective herbal formulation. This strategy, however, requires both a consistent formulation and a large study sample.

Statistical considerations also apply to a controversy over the significance of a recent randomized, placebo-controlled trial study of 25 participants in which “garlic oil” was found to have no effect in decreasing serum cholesterol levels (18). This negative result, which seemed to contradict results obtained with other garlic products, was criticized because of the processing of this garlic oil and the apparent lack of bioavailability of the ingredient thought to be responsible for garlic’s cholesterol-lowering effect (19). Differences in the composition of garlic preparations offer a reasonable explanation for different study results. The possibility remains, however, that the apparently disparate results arise simply from the inherent variability of small clinical trials. Two independent meta-analyses, which included this garlic oil trial and at least a dozen other garlic products (20, 21), concluded that garlic decreased serum cholesterol levels and confirmed that chance alone readily explains the results of this garlic oil trial.

Sometimes, poorly designed clinical trials or incomplete reports make it difficult to evaluate published studies. For example, when feverfew was considered for meta-analysis to determine its effectiveness in prophylaxis against migraine, four randomized trials were found, but the three that reported positive results had more methodologic deficiencies than did the one with a negative conclusion (22). Furthermore, the three trials with positive results were flawed by possible selection bias, since each had recruited participants who had previously benefited from feverfew (23). The negative study (16), on the other hand, was conducted in patients with no previous exposure to feverfew.

Pooling data from individual trials by using meta-analysis is one way to reach an interpretation of the
results of a group of inconclusive trials. The examples of feverfew and garlic raise the issue of how to decide which products should be eligible for inclusion in the analysis. Being too inclusive risks including products that are ineffective, thus decreasing the power of the meta-analysis to detect effectiveness of the products as a group. But on what basis can products be excluded if an herb’s active ingredients are in doubt and there is no consensus on methods for standardization?

Fortunately, this question does not always require an answer. A meta-analysis of the effectiveness of feverfew as prophylaxis against migraine cannot be done because data are insufficient (22). In contrast, there are enough studies to conclude that garlic decreases serum cholesterol levels, even when the controversial small garlic oil study is included (20, 21). As for St. John’s wort, it is apparently effective even with the inclusion of studies of products that may not have optimized the herb’s active ingredients. Of course, such favorable conclusions apply to the group of garlic and St. John’s wort products tested and do not warrant that each garlic or St. John’s wort product on the market has the claimed activity. In other words, these herbs appear to have activity, which can be assured only if the products derived from them are adequately standardized.

Safety

Until recently, the safety of herbal preparations was considered in medical journals only when toxicity was detected from a contaminated herbal product, usually because of careless or unscrupulous manufacturing practices. A toxic herb might replace the traditional one (24), a conventional drug might be added covertly (25), or a harmful contaminant might appear without the manufacturer’s knowledge (26). True herbal toxicity, on the other hand, is almost certainly underreported. Users of herbal remedies are generally convinced of their safety and are therefore biased against reporting an adverse clinical event of possible herbal origin. Furthermore, physicians are often unaware of the herbs their patients are taking, either because they do not ask about them or the patient does not tell them (27). Indeed, half of the herbs taken by patients are not reported to their physicians (28). Adverse reactions to herbs, however, are now receiving attention formerly accorded only to adverse reactions to drugs. The medical literature now contains both case reports and periodic updates of adverse reactions to herbs (29).

Uncovering adverse reactions to herbs, however, is more challenging than uncovering adverse reactions to drugs. First, for herbs, there is no equivalent to prescription records, which document exposure to drugs and therefore permit associations to be made with later clinical events (30, 31). In addition, such associations become directly meaningful because standardization of drugs by the United States Pharmacopeia and the U.S. Food and Drug Administration (FDA) links the drug by name to its pharmacologic effect. Comparable associations for herbs are hampered by product variations such as those just described for St. John’s wort and documented by several analytical comparisons. Among 10 ginseng products, for example, the ginsenoside content varied over a 10-fold range (32), and such indications of variability do not include analysis of other components in these products that might trigger an adverse reaction. An added difficulty in monitoring the safety of herbal products in the United States is the surfeit of products on the market—an estimate from 1997 was more than 1500 (33). If an adverse reaction to an herb is suspected, it should be reported to the FDA’s MedWatch program (www.fda.gov/medwatch) (34), even though getting to the root of the problem may require considerable detective work (26) that is not always within the scope of the FDA’s resources.

Attention was recently called to the potentially serious clinical implications of possible interactions between herbs and prescription drugs. For example, the administration of St. John’s wort for 10 days to a group of normal volunteers reduced the absorption of digoxin by an average of 25% (35). St. John’s wort taken for 2 weeks also reduced the total absorption of indinavir by 50%, which would have been large enough to cause treatment failure (36). The effects of St. John’s wort appear to be pervasive, possibly because of its induction of P-glycoprotein (35); case reports indicate significant increases in the metabolism of other drugs, including cyclosporine (37), warfarin (38), and oral contraceptives (39). It is not clear whether these effects are all attributable to the same ingredients of St. John’s wort or whether these interactions can be attributed to the ingredients responsible for the herb’s antidepressive effects.
Regulation

Starting in 1906, the U.S. Congress passed legislation that, when implemented by the FDA (or its predecessors), provides assurance that drugs are accurately labeled (1906), safe (1938), and effective for the labeled indications of use (1962) (40). Thus, the United States has developed a rigorous system for evaluation and approval of new drugs that is widely emulated. The United States, however, never emulated such countries as Japan and Germany, which accommodated national traditions by developing special regulations for herbal medicines. The FDA maintains that a drug is “any substance or mixture of substances intended for the cure, mitigation, diagnosis, or prevention of disease . . .” (40), a definition derived from the Pure Food Act of 1906 (except that “diagnosis [of disease]” was omitted from the original Act). The results are regulations that make it almost impossible for medicinal herbs to fulfill the FDA’s existing standards for drug approval.

The marketing of medicinal herbs received a boost in 1994 when Congress passed the Dietary Supplement and Health Education Act (DSHEA), which declared that herbal medicines are not drugs. The Congressional solution was simple: Herbal medicines would be called “botanicals” and classified along with vitamins, minerals, and other health products in a new category called “dietary supplements.” Because they were not drugs, botanicals and other dietary supplements could not claim to cure, treat, prevent, or diagnose disease; neither, however, were they subject to the expert scientific evaluation that had helped ensure the safety and effectiveness of drugs. Although DSHEA prevents manufacturers of botanicals from making “disease claims,” they are allowed to make “health” or “structure/function” claims (41).

The distinction between disease claims, which are allowable only for drugs, and health claims, which are allowable for dietary supplements, may be difficult and even paradoxical, as when an herb’s effectiveness is convincingly documented only for treating a disease. A case in point is St. John’s wort, which a meta-analysis concludes is “more effective than placebo for the short-term treatment of patients with mild to moderately severe depressive disorders” (12). This conclusion is based on an analysis of 27 trials, all conducted in depressed patients in medical facilities. Nine of these trials included only patients who met formal diagnostic criteria for major depression, and 21 evaluated treatment success in terms of the Hamilton Rating Scale for Depression, an instrument “developed for use in assessing the symptoms of patients diagnosed as suffering from depressive states” (42). Nevertheless, suggested claims for St. John’s wort under DSHEA are clearly nonmedical, for example, to “help support healthy emotional balance” or to “help maintain a positive attitude” (41).

A similar paradox affects Ginkgo biloba, which a meta-analysis concluded has “a small but significant effect” on objective measures of cognition in patients with Alzheimer disease. The diagnosis in these patients was made according to accepted psychiatric criteria: those in the Diagnostic and Statistical Manual of Mental Disorders or those of the National Institute of Neurological Disorders and Stroke (11). Yet, suggested acceptable claims for ginkgo under DSHEA are so vague as to be misleading, for example, to “promote mental sharpness” or to “improve memory and concentration” (41). The search for consensus on the definition of disease and, hence, the basis for distinguishing between drug and permissible botanical claims has occupied lawyers representing the dietary supplement industry and their FDA counterparts since DSHEA became law.

From a historical perspective, DSHEA regulations merely reflect the reemergence of the philosophy said to guide the original Pure Food Act. According to Temin (40), the Pure Food Act of 1906 intended simply to provide consumers with accurate information that would enable them to make informed decisions about purchases of foods and drugs. In Temin’s view, the Acts of 1938 and 1962 changed the emphasis to consumer protection; decisions about drugs would be made not by consumers themselves, but on their behalf by “experts.” Thus, the 1906 Act was intended to protect fair competition and leave decisions about drug use to the consumer; the intent of later laws was to protect the consumer (40).

In this context, DSHEA represents a return to the philosophy of consumer choice that Temin attributes to the 1906 Act. Both Acts emphasize product labeling as a means of providing consumers with product information. The 1906 Act, however, recognized the importance of standards, although it left the setting of these standards to traditional, nongovernmental, compendial organizations, such as the United States Pharmacopoeia or the National Formulary. The DSHEA does not mention standards directly, but it stresses consumer educa-
tion by regulating dietary supplement labeling and such matters as the relationship between herbal product literature and the products themselves when both are sold in the same store.

To emphasize that botanicals are no longer considered drugs, their regulation is implemented from the Office of Special Nutrition within the FDA’s Center for Food Safety and Nutrition. Furthermore, “there is no FDA pre-market review of ingredients or finished products considered to be dietary supplements” (34), and the FDA is empowered to remove a botanical from the market only if it can be proved unsafe. Placing the burden of proof on the FDA to show that a product is unsafe, rather than on the manufacturer to show that it is safe, marks a return to the weak regulations that had applied to drugs until 1938. The law mandating premarketing clearance for a drug’s safety grew out of an incident in 1937 when more than 100 deaths resulted from a sulfanilamide syrup formulated with ethylene glycol (40).

Clearly, safety is jeopardized if herbs are adulterated, but the long-term safety of some plants, even those used for medicinal purposes (29, 43–47), can be questionable. Recent reports on the delayed clinical consequences of a mislabeled herbal weight-reduction product remind us that some herbs may be carcinogenic (48) as well as toxic (49).

**Practical Considerations**

Current knowledge about the safety and effectiveness of herbs now on the market is clearly inadequate and not likely to improve under current regulations. A fundamental problem is the lack of funds to support research on medicinal herbs, in part because herb manufacturers, lacking product exclusivity, cannot profit from their research as do drug manufacturers. Thus, physicians, who consider medicinal herbs to be drugs and wish to evaluate them by comparable standards of safety and efficacy, are not satisfied with the information now available. Since manufacturers’ claims for botanicals are not regulated by the FDA, a botanical preparation has no document like the FDA-approved package inserts reproduced in the Physician’s Desk Reference (50). The United States Pharmacopeia plans to fill the information gap by preparing monographs on many popular botanicals that should be useful to both consumers and health professionals (51). In the meantime, general information may be obtained from a recent English translation of the German Commission E Monographs (44) and the series “Adverse Effects of Herbal Drugs” (45–47).

To many consumers, DSHEA has provided a welcome opportunity to gain access to remedies that apparently provide simple solutions to their health concerns. To physicians and other health care providers, however, the new remedies are not so simple because reliable information about them is lacking. Thus, the risk–benefit assessment that physicians ordinarily make with their patients tends to become dichotomous: The patient is more enthusiastic about the claimed benefits of a botanical, and the physician is more concerned about unknown risks. Clearly, consumers and their health care providers must reach a consensus with vendors of herbal medicines and the FDA about how medicinal herbs will be further investigated and adequately regulated.

From Harvard Medical School, Harvard School of Public Health, and the Center for Alternative Medicine Research and Education, Beth Israel Deaconess Medical Center, Boston, Massachusetts.

**Grant Support:** In part by grant 1 RO1 AT00607-01 from the National Center for Complementary and Alternative Medicine.

**Requests for Single Reprints:** Peter Goldman, Center for Alternative Medicine Research and Education, Beth Israel Deaconess Medical Center, 330 Brookline Avenue, Boston, MA 02215.

**References**

10. DeFeudis FV. *Ginkgo biloba* Extract (EGb761). Wiesbaden, Germany.