Not All Soy Products Are Created Equal: Caution Needed in Interpretation of Research Results

John W. Erdman, Jr.,*  Thomas M. Badger,†  Johanna W. Lampe,**  Kenneth D.R. Setchell,‡  and Mark Messina+++  
*Department of Food Science and Human Nutrition, University of Illinois, Urbana, IL 61801; †Arkansas Children’s Nutrition Center, University of Arkansas for Medical Sciences, Little Rock, AR 72202; **Fred Hutchinson Cancer Research Center, Seattle, WA 98109; ‡Cincinnati Children’s Hospital Medical Center, Department of Pediatrics of the University of Cincinnati College of Medicine, Cincinnati, OH 45229; and +++Nutrition Matters, Inc., Port Townsend, WA 98368 and Loma Linda University, Loma Linda, CA 92350

ABSTRACT  Interest in the health benefits of soy foods has been intense among the research community, health professionals, and the public. At the same time, potential concerns associated with soy consumption, especially as related to soy isoflavones, have tempered the enthusiasm for making public health recommendations. On both accounts, the primary soybean isoflavone, genistein, has received the most attention. Because consumers are becoming increasingly confused by the often conflicting dietary messages, a balanced and accurate view of the risks and benefits of soy foods and soy food components is essential. Even among health professionals, confusion exists about proper nomenclature and about the precise composition of the agents under investigation. Levels of isoflavones are frequently assumed to be constant within categories of soy foods, and intakes are estimated rather than being directly analyzed. Furthermore, all too often research dealing singularly with genistein is interpreted by both health professionals and the media as equating directly with soy. Researchers often fail to fully understand the implications of their research outcomes and the context in which those outcomes should be placed. With the hundreds of publications yearly on soy and isoflavones, it is especially important to consider the literature in its entirety when making pronouncements about health effects. Efforts are needed by all to reduce the public confusion by adapting standardized approaches to the reporting of data. This paper provides a framework for both standardization of nomenclature and appropriate interpretation of data.  J. Nutr. 134: 1229S–1233S, 2004.

KEY WORDS:  • soy  • soy protein  • isoflavones  • genistein

A recent headline from an article typical of those that appear in popular health magazines was “Does Soy Have a Dark Side?” The author first reviewed the positive epidemiologic evidence linking soy food consumption to reduced risk of coronary heart disease and cancer. Then the reader was reminded about the recent approval by the U.S. Food and Drug Administration of a health claim for the cholesterol-lowering effects of soy protein and reduction of risk from heart disease (1). The remainder of the article raised possible negative issues of soy consumption, focusing on the potential increased risk of developing estrogen-sensitive cancers, impaired cognitive function, and reproductive problems. It is most appropriate to consider the possible adverse effects; indeed it is essential for journalists and the scientific community to scrutinize both sides of the risk-benefit continuum for any dietary substance. The author of the article was informed about the issues but was not able to put the different kinds of research results into perspective and was confused about the nomenclature used in different studies.

Confusion regarding the relative merits and potential public health significance of research reports using different research models plagues writers, scientists, and policy makers alike. Even among scientists debate exists about the pros and cons of various research models and approaches. Part of the confusion likely arises because of the lack of standardized nomenclature for referring to soy products and components in the scientific literature. In addition, it is difficult to compare results across a wide range of study designs that often lack biostatistical power and involve poorly defined test substances. Rather than addressing health and safety of soy foods per se, this paper will provide guidelines for those preparing or inter-
preparing research publications in regard to proper nomenclature for the different soy foods and extracts; fair and accurate interpretation of the variety of in vitro, cell culture, animal, and human studies; and accurate translation of research results for public health decision makers and the media. The paper will conclude with recommendations for manufacturers, researchers, and the media.

**Soy nomenclature and chemical composition**

Soy nomenclature has several confusing aspects. It is important to recognize that most epidemiologic studies that have examined the relationship between soy intake and health outcomes have involved Asian populations and therefore evaluated the intake of traditional soy foods, such as tofu or soy milk, that are derived from whole or dehulled soybeans. In contrast, few animal or human intervention studies have involved whole soy foods. Instead, soy concentrates, soy isolates, isolated isoflavone mixtures, supplements, or pure genistein are generally used. Some studies have evaluated full-fat or defatted soy flours or textured soy protein (usually mixtures of soy concentrates and soy flour). Thus, a major distinction usually occurs between the population-based association studies of soy food intake and health outcomes and most experimental research that is based on use of isolated fractions of the soybean.

The dozens of products used for research differ markedly in both macro (protein, fat, and carbohydrates) and micro (isoflavones, saponins, phytic acid, phytosterols, vitamins, and minerals) constituents (2). Table 1 shows the substantial compositional differences in macronutrient content among soy flours, concentrates, and isolates. Research articles often use terms such as soy, soy protein, or even soy supplements without specifying what is meant or has been used in a study. Using a soy flour, soy protein concentrate, or an isolate might mean a difference between 32%, 21%, or 3% carbohydrate content, for example. Unfortunately, researchers often neglect to provide sufficient details in the methods sections of papers to allow others to understand, interpret, or repeat their work.

Chemical composition of soy concentrates or isolates is affected by processing techniques, including both solubilization method and temperature. Both water and alcohol (ethanol) washing are used to concentrate the proteins by removing sugars and oligosaccharides from soy. Water, and to some degree ethanol, will also remove some of the low-molecular-weight peptides (those below 2000) that contain high amounts of sulfur amino acids. Alcohol washing also removes most of the isoflavones (and other lipid-soluble phytochemicals such as saponins) from the product (3,4). Table 2 lists the total and individual isoflavone contents of selective soy foods (5).

Recent research indicates alcohol washing may substantially alter the protein matrix and structure. For example, Gianazza et al. (6) using 2-dimensional electrophoresis demonstrated extensive degradation of the 7S globulin of protein in some soy isolates into proteins and peptides with smaller molecular weights. Peptides formed from the digestion of the α constituent subunit from 7S globulin may upregulate hepatic LDL cholesterol receptors and consequently contribute to the cholesterol-lowering effects of soy protein (7). Gianazza et al. (6) also found that some proteins accumulated more in ethanol-treated samples. Researchers often compare ethanol-washed with nonethanol-washed soy protein and then assume that any differences seen in health outcomes are due to the isoflavone component when saponins, other alcohol soluble material, and the protein and peptide composition could conceivably contribute or be responsible for the observations.

The composition of specific (standardized) soy foods varies substantially. Setchell and Cole (8) recently evaluated the variations in isoflavone levels in soy foods and soy protein isolates. They found a 2–3-fold variation in total isoflavones in soy protein isolates over 3 γ and fivefold differences in isoflavone levels in different commercial soy milks. Figure 1 shows the variations in total isoflavones in soy milks (8). The differences among products are due in part to the variety of soybeans, differences in growing and storage conditions, and differential food processing techniques. Clearly, it is important for the scientific community to accurately understand the chemical composition of study products.

Another major source of confusion is the nomenclature used for isoflavones. The soybean contains 6 major and 6 minor forms of isoflavones. As shown in Table 3, the majority of the isoflavones found in the soybean cotyledon are the β-glycosides, genistin and daidzin, and only small proportions of genistein and daidzein (unglycosylated, or aglycone forms) are present (9). In contrast the hypocotyl, or germ, contains high concentrations of glycitin and some glycitein, moderate amounts of daidzin, and relatively small amounts of genistin. Many commercial soy isoflavone supplements are made from the soy germ. Those soy isoflavone supplements do not contain the same isoflavone profile as that obtained from the consumption of soy foods that are produced primarily from the whole or the dehulled soybean. This difference leads to markedly different plasma isoflavone profiles (10). Other soy isoflavone products derived from whole soy or recovered from the ethanol wash during the preparation of soy concentrates and isolates

---

**TABLE 1**

<table>
<thead>
<tr>
<th>Percentage composition of soy protein products (moisture-free basis)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Constituent</strong></td>
</tr>
<tr>
<td>Protein (Nx6.25)</td>
</tr>
<tr>
<td>Fat</td>
</tr>
<tr>
<td>Crude fiber</td>
</tr>
<tr>
<td>Soluble fiber</td>
</tr>
<tr>
<td>Insoluble fiber</td>
</tr>
<tr>
<td>Ash</td>
</tr>
<tr>
<td>Carbohydrates (by difference)</td>
</tr>
</tbody>
</table>

1 Modified from Endres (2).

**TABLE 2**

| Estimated isoflavone contents of selected soy protein products | **Total isoflavones** |
|---|---|---|---|
| Daidzein | Genistein | Glycitein |
| Soy flour, full-fat, roasted | 99.27 | 98.75 | 16.40 | 198.95 |
| Soybean flakes defatted | 36.97 | 85.69 | 14.23 | 125.82 |
| Soy protein concentrate, aqueous washed | 43.04 | 55.59 | 5.16 | 102.07 |
| Soy protein concentrate, produced by alcohol extraction | 6.83 | 5.33 | 1.57 | 12.47 |
| Soy protein isolate | 33.59 | 59.62 | 9.47 | 97.43 |

1 Modified from Jackson and Gilani (5).
2 Values based on mg/100 g edible portion.
also contain various profiles and amounts of isoflavones and other constituents.

Expressing isoflavone content

Considerable confusion exists about how the isoflavone content of a product is expressed. As previously noted, isoflavones are naturally present in the soybean primarily as glycosides. However, as a result of processing, and especially fermentation, some of the glycosides in soy foods are converted to aglycones. Researchers often express the isoflavone content of the product fed to subjects without indicating whether the stated amount refers to the aglycone or glycoside value. This distinction is critical because the sugar moiety does not contribute to the biological activity, and the amount of the aglycone present is ~60% of the total weight of the glycoside. Thus, 100 mg isoflavones may represent anywhere from 60 to 100 mg active isoflavones depending on the type of soy product. Simply indicating the form of the isoflavones in a product is not sufficient. Ideally, all values should be expressed as the aglycone equivalent. Of course, as noted by Murphy et al. (11), confusion surrounding isoflavone content could be avoided if amounts were expressed in molar values, but this has not been the convention within the scientific community for food labeling. In addition, summing of isoflavone intakes may not be useful because the biological activities of daidzein and genistein are quite different and the activities of other isoflavones are not well characterized.

Interpretation of research publications

Even scientists specializing in soy nutrition find it difficult to synthesize the findings from the hundreds of papers published yearly on soy foods, soy isoflavones, and genistein. These publications range from epidemiologic investigations to human intervention trials, animal studies, and cell culture work.

Although several epidemiologic studies involving Western populations have noted positive associations between soy intake and favorable health outcome, some questions remain about the plausibility of there being a soy-related biological basis for the reported results because of the very limited soy intakes noted in these studies. As is always the case with epidemiologic studies, there is the issue of uncontrolled confounding variables. Because soy foods, which are generally perceived as healthful, are not a traditional part of Western diets, higher soy consumption may reflect an overall healthful lifestyle that is not so easily identified or controlled for (12). Similarly, in Asian countries, soy intake may be more a marker of a traditional diet and lifestyle (low fat, low red meat, more green tea, more exercise, etc.) so it is not known, for example, whether it is the soy or the other components of the traditional Asian diet and lifestyle that are important.

In Japan and some parts of China, soy foods contribute ~5% of total energy and provide ~8–12 g/d of protein, which represents ~10% of adult total protein intake (13–16). Not unexpectedly, recent surveys suggest that daily soy protein and isoflavone intake among adults of non-Asian descent in Western countries is quite small, likely being ≤1–2 g and ≤3 mg, respectively (12,17–20). Precise quantification is difficult because small amounts of soy protein containing variable amounts of isoflavones are added to literally hundreds of commonly consumed processed foods. In contrast to the modern Asian soy intake, most clinical trials involve feeding between 20 and 50 g/d of soy protein. Although the goals of these trials are important, we should be mindful that provision of high amounts of soy protein and isoflavones exceeds habitual intakes in soy-consuming populations.

Unquestionably, the population whose soy protein and isoflavone exposure is greatest is infants consuming 100% of their diets as soy formula. On a kilogram body weight basis, isoflavone intake is ~4–8 times greater in infants exclusively fed soy infant formulas when compared with intake of adults. Not unexpectedly, the concentration of isoflavones in blood and tissues of infants is 10 times greater than in adults with comparable daily intakes of soy products, especially because infants are continually consuming isoflavones throughout the day (21). Despite these very high levels with their potential for biological effects, it is worth noting that soy infant formula has been generally shown to be nutritionally safe and adequate for normal growth and development.

In animal trials soy protein is typically provided at levels of 10–20% of the diet. Moreover, acute species differences in colonic metabolism and biotransformation of isoflavones may influence response to an intervention (22). Although experiments in animals are usually designed to show a difference

---

**TABLE 3**

Isoflavone distribution within the soybean

<table>
<thead>
<tr>
<th>Isoflavone</th>
<th>Hypocotyl</th>
<th>Cotyledon</th>
<th>Seed coat</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total isoflavones, mg/g</td>
<td>21.54</td>
<td>3.80</td>
<td>0.05</td>
</tr>
<tr>
<td>Daidzin, mg/g</td>
<td>8.38</td>
<td>1.45</td>
<td>0.02</td>
</tr>
<tr>
<td>Daidzein, mg/g</td>
<td>0.35</td>
<td>0.11</td>
<td>—</td>
</tr>
<tr>
<td>Genistin, mg/g</td>
<td>2.46</td>
<td>2.10</td>
<td>0.02</td>
</tr>
<tr>
<td>Genistein, mg/g</td>
<td>0.16</td>
<td>0.14</td>
<td>0.01</td>
</tr>
<tr>
<td>Glycitin, mg/g</td>
<td>10.04</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Glycitein, mg/g</td>
<td>0.15</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Total soybean isoflavone content, %</td>
<td>11.98</td>
<td>87.94</td>
<td>0.08</td>
</tr>
</tbody>
</table>

1 Modified from Gugger (8).
2 On a dry weight basis hypocotyl, cotyledon, and seed coat are represented by 2.2%, 91.5%, and 6.3% of weight, respectively.
3 Not detected.
between diet groups, interpretation or translation of those results to humans should be done with the knowledge that these effects may not occur at the lower intake levels typically found in Western or perhaps even in Asian cultures. Studies that use the addition of genistein either by injection or diet provide valuable information about its pharmacological actions, but caution is needed when extrapolating findings to humans consuming soy protein foods, especially because genistein is rarely present in substantial amounts in most human diets.

Specialized animal models are frequently used to study specific health aspects of food components. There are a variety of cancer, diabetes, and coronary heart disease models, usually involving rodents (rats and mice) bred especially to express these diseases. Because no individual animal model is totally comparable with humans with regard to disease progression or metabolism, researchers should consider the strengths and weaknesses of an animal model system when relating the results to humans. This is especially important when investigating the effects of in utero and prepubescent soy exposure.

The microflora in the intestinal tract of rodents and monkeys are more efficient than microflora in humans in producing the isoflavone metabolite equol; in fact, rodents and monkeys are referred to as to equol machines (22). Only 30–50% of humans can produce equol from daidzein in the gastrointestinal tract and even in those who do, equol levels in serum are relatively modest compared with genistein and daidzein (22–24). Because equol has been suggested to be a particularly potent biologically active isoflavone (22), outcomes of studies with rats and mice may not predict the effect of soy isoflavone consumption by humans.

Perhaps the source of the most confusion and conflicting results is the use of supraphysiological levels of genistein or isoflavone mixtures in cell culture studies. Cell culture studies often use isoflavone concentrations of 10–1000 times higher than physiological levels and provide isoflavones without other soy components. Numerous studies have shown that the half-life of isoflavones in blood after a soy-containing meal is not long (10,25–30). In healthy adult men and women the terminal elimination half-life of genistein and daidzein ranges from 3 to 9 h after consumption of a soy protein isolate drink (30). Although cell culture studies are aimed at identifying possible isoflavone effects and defining mechanisms, the relevance of these studies to in vivo situations is uncertain. For example, many independent estrogen-receptor-binding effects of genistein that are noted in vitro, such as the inhibition of DNA topoisomerases, tyrosine protein kinases, and cancer cell growth, occur at concentrations that exceed 25 μmol (31). However, peak serum genistein concentrations typically are no higher than 5 μmol and 95% are conjugated with glucuronic acid and are less biologically active. The extent to which in vitro observations involving high micromole levels of isoflavones are relevant to humans is unclear.

In vitro studies have an important role, particularly for studying structure-activity relationships and mechanisms whereby soy components may affect cellular or molecular events. However, there is concern that in vitro work may either overestimate or underestimate the in vivo potency of isoflavones. An underestimation may occur because in vitro systems may lack cellular factors that potentiate the effects of isoflavones in vivo or because isoflavone metabolites are produced in vivo that are more potent than the parent compound. In any event, it is quite clear that in vitro data should be interpreted with caution and should be used primarily to help determine the biological actions of individual phytochemicals and to identify the mechanisms underlying those effects.

Conclusions
Confusion regarding soy and health research is not surprising considering the complexity of this field. The confusion often starts when researchers do not provide sufficient details regarding the products and compounds that they use or do not relate the amounts or concentrations used to real-life consumption patterns. Researchers often fail to translate appropriately their results in risk-benefits terms and to compare their findings produced from 1 experimental model with results from relevant research from other types of experimental models. Consequently, the media and health professionals have a difficult time translating the results for consumers and public health messages. Health professionals should place new research into perspective of the entire wealth of published literature when communicating to the public or making public health recommendations.

Recommendations
To reduce the confusion regarding soy and health research, we make the following recommendations:

- Manufacturers who provide information about the isoflavone content on their product labels should express contents as an aglycone amount. In addition, they should provide clear content information on test products to researchers.
- Researchers should provide a detailed and unambiguous description of products, concentrations, and amounts used in their studies.
- Researchers should clearly relate their study conditions to usual soy and isoflavone intakes and/or tissue levels of isoflavones.
- Researchers should consider the risks and benefits of their findings for human health.
- All researchers, but especially those conducting cell culture or animal studies, should outline the benefits and limitations of the model system used.
- The media and public health officials should look at the totality of the evidence when commenting on the potential significance of new findings and demand that researchers be clear on what soy product or extract was used and how any new findings compare with the existing literature.

LITERATURE CITED


