

# Assessments of Pharmaceutical Advertisements: A Critical Analysis of the Criticism

J. HOWARD BEALES III, PH.D. \*  
WILLIAM C. MACLEOD, ESQ. \*\*

## I. INTRODUCTION

The impact and accuracy of pharmaceutical advertising has inspired decades of study and debate, as well as legislation and regulations. The importance of advertising to the medical profession is undeniable. Some scientific journals could not survive without advertisement revenues, and many physicians would not learn of new therapies without commercial messages from drug manufacturers. The utility of providing such information, however, has not ensured a complete acceptance of the practice. Drug advertising has been criticized for relying on questionable research, obscuring unfavorable evidence, failing to provide adequate information, and exaggerating the advantages of one product over another.

Such criticism, however, is vulnerable to the same charges. An examination of the last twenty years of literature on pharmaceutical advertising shows that prior to 1992, the discussion of the faults and virtues of advertising had been limited to personal opinions and isolated anecdotes.<sup>1</sup> This is not the kind of record on which a serious assessment of the subject could be based.

In June 1992, an article in the *Annals of Internal Medicine* attempted to remedy the lack of scientific rigor in the measure of pharmaceutical claims in advertising.<sup>2</sup> The authors of *Pharmaceutical Advertisements in Leading Medical Journals: Experts' Assessments* subjected over 100 pharmaceutical advertisements to a systematic form of peer review, and they reached an alarming conclusion that seemed to confirm the worst fears of advertising skeptics. The study concluded that "the public is at risk, because advertisements may be misleading doctors about the appropriateness of a drug, failing to inform doctors of the adverse effects of a drug, and inducing doctors to prescribe more expensive drugs than patients need."<sup>3</sup> One in four advertisements would have been rejected if the reviewers had possessed editorial authority to do so.<sup>4</sup> Another third of the advertisements would have required major revision before publication.<sup>5</sup>

---

\* Dr. Beales is an Associate Professor of Strategic Management and Public Policy, George Washington University School of Business and Public Management, Washington, D.C. He previously has served as Branch Chief, Office of Information and Regulatory Affairs, Office of Management and Budget, and Associate Director for Policy and Evaluation, Bureau of Consumer Protection, Federal Trade Commission.

\*\* Mr. MacLeod is a partner in the law firm of Collier, Shannon, Rill & Scott, Washington, D.C. He has also served as Director, Bureau of Consumer Protection, Federal Trade Commission (1986-1990), and Director, Chicago Regional office, Federal Trade Commission.

The original research and analysis upon which this article is based was supported by funding from Pfizer, Inc.

<sup>1</sup> See e.g., Jacob Jacoby & Constance Small, *The FDA's Approach to Defining Misleading Advertising*, 39 J. MARKETING 65 (1975).

<sup>2</sup> Michael S. Wilkes et al., *Pharmaceutical Advertising in Leading Medical Journals: Expert Assessments*, 116 ANNALS OF INTERNAL MED. 912 (1992).

<sup>3</sup>*Id.* at 918.

<sup>4</sup>*Id.* at 917.

<sup>5</sup>*Id.*

This article re-examines the conclusion that pharmaceutical advertising may mislead the doctors to whom it is directed. None of the significant articles published on the subject before 1992 provides a systematic basis for their conclusions and the analysis in *Annals* collapses under a careful review of the data. A critical examination of the methodology of the *Annals* study and the underlying data reveals that:

- If problems that the reviewers thought were minor are discounted, more than half the reported problems disappear.
- Most rejections of advertisements were not due to misleading claims, but to a desire by the reviewers for more information than the advertisement contained — information that has no bearing on the accuracy of the advertisements reviewed.
- The results were not significantly different from chance, but from the pattern of responses that would have been generated by reviewers who were accurate in identifying problems in the advertisements.

The methodology of the literature on pharmaceutical advertising has advanced beyond the reported impressions and prejudices of industry observers. Nevertheless, the one effort to lend scientific rigor to the literature failed to confirm or refute the assertions on either side of the debate. If there is a problem warranting regulatory reform, the case has not been made by the findings to date. Until it can be demonstrated that the general population of prescribing physicians is being misled by these advertisements, it is premature to consider policies that could restrict the supply of information to the medical community.

## II. THE EARLY STUDIES

For decades, commercial advertising has represented a significant source of information about the availability, indications, and advantages of prescription drugs. A 1977 editorial celebrating the fiftieth anniversary of *The Annals of Internal Medicine* claimed that “[d]rug advertisements in peer-review journals are a vital part of continuing pharmacologic education.”<sup>6</sup> This observation has been confirmed by studies of pharmaceutical advertising’s influence on physician prescription patterns.<sup>7</sup> Doctors respond to information in attractive packages with understandable terms and concise messages. Investigators who found that promotional activity played an important role in the practice of medicine concluded that “the predominance of nonscientific rather than scientific sources of drug information is consistent with what would be predicted from communications theory and marketing research data.”<sup>8</sup>

While there is general agreement as to advertising’s ability to inform doctors about drugs, “concerns have been expressed about inaccuracies or misleading claims” made in these advertisements.<sup>9</sup> If these concerns are valid, misleading claims could result in improper prescribing practices, inadequate treatment, and unnecessary side effects. Unfortunately, nothing in the literature from the 1970s and 1980s tested the validity of these concerns. Critics registered their personal disapproval of the idea that physicians

---

<sup>6</sup> Robert H. Moser, *Advertising and Our Journal*, 87 ANN INTERN MED. 114-115 (1977).

<sup>7</sup> E.g., Lawrence S. Linn & Milton S. Davis, *Physicians’ Orientation Toward the Legitimacy of Drug Use and Their Preferred Source of New Drug Information*, 6 SOC. SCI. MED. 199-203 (1972); Jerry Avorn, Milton Chen, & Robert Hartley, *Scientific versus Commercial Sources of Influence on the Prescribing Behavior of Physicians*, 73 AM. J. MED. 4-8 (1982).

<sup>8</sup> Avorn, et al., *supra* note 7, at 8.

<sup>9</sup> *Id.* at 912.

were influenced by noncommercial messages, while journal editors expressed their support. Neither critic nor supporter, however, came forward with evidence that might settle the issue. It was not until 1992 that anyone attempted to ascertain whether pharmaceutical advertising deserved a passing or failing grade for accuracy.

### III. THE WILKES STUDY

#### A. *The Methodology*

The Wilkes study set out to evaluate a broad spectrum of pharmaceutical advertising. Borrowing a methodology from behavioral science, the authors selected 109 advertisements from 10 medical journals and developed a survey with 36 questions, 28 of which related to the criteria the Food and Drug Administration (FDA) applies in reviewing advertising for accuracy.<sup>10</sup> The advertisements, references cited in the advertisements, and questionnaires were sent to physicians and pharmacists recognized as experts in their respective fields.<sup>11</sup> They were asked whether they agreed or disagreed (somewhat or strongly) that the advertising claims were in compliance with certain FDA standards of accuracy, and whether the advertisements met the standards that govern publication of articles in major medical journals.

#### B. *The Conclusions of the Study*

While some aspects of the advertisements received quite favorable grades, not all the reviews were good.<sup>12</sup> On the positive side, in only 8% of the 48 advertisements where reviewers found safety claims did two or more reviewers disagree with claims that the drug was safe.<sup>13</sup> In 86% of the advertisements, reviewers agreed with the safety claims they saw. For the remaining 6% of the advertisements, the reviewers could not agree one way or the other.<sup>14</sup> Of the 23 advertisements identified with tables and graphs, only 9% were thought to misrepresent the conclusions of clinical studies, while 82% of these advertisements received favorable reviews.<sup>15</sup>

On the other hand, two or more reviewers agreed at least somewhat that one-third of the 94 advertisements in which they saw headlines had headlines that misled the reader about efficacy or side effects.<sup>16</sup> Nearly one-quarter of the images in the 90 adver-

---

<sup>10</sup> *Id.* at 913. Specifically, the survey instruments asked respondents to record whether they were in strong agreement, somewhat in agreement, somewhat in disagreement, or in strong disagreement with questions such as, "This advertisement uses TABLES OR GRAPHS that are not adequately referenced." Michael S. Wilkes, et al., UCLA ADVERTISEMENTS REVIEW FORM 5 (1991) [*hereinafter* QUESTIONNAIRE].

<sup>11</sup> The 109 advertisements typically were reviewed two or three times each, although not all the questions applied to all the advertisements. For example, two or more reviewers found tables or graphs in only 23 advertisements. Wilkes et al., *supra* note 2 at 915. Thus, evaluations of the use of tables and graphs were not based on the entire sample of 109 advertisements, but on less than one quarter of the sample.

<sup>12</sup> For purposes of analysis, the article combined those who "strongly agreed" and those who "somewhat agreed" into a single category, and combined disagreements in the same fashion.

<sup>13</sup> Wilkes et al., *supra* note 2, at 914. Further analysis of the underlying data reported in the Wilkes study revealed a number of discrepancies, most of which were minor. *See infra* part IV. Only one difference was significant. The analysis determined that 59 advertisements were found to include safety claims, not 48 as reported by Wilkes and his colleagues. For the remainder of this article, numbers derived from the further analysis are employed where different from numbers included in the Wilkes study.

<sup>14</sup> *Id.*

<sup>15</sup> *Id.* at 915.

<sup>16</sup> *Id.*

tisements with images were held to mislead by two or more reviewers.<sup>17</sup> Although only 9% of the 23 advertisements with graphs and tables were found to misrepresent conclusions of studies, 30% of the graphs and tables (those in 7 advertisements) were found likely to mislead the reader.<sup>18</sup> The reviewers concluded that nearly one-fifth of all 109 advertisements would have been rejected if judged by journalistic peer-review standards.<sup>19</sup>

These findings led the authors to conclude that FDA rules “appear to go unheeded,”<sup>20</sup> and to suggest that the FDA was “unable or unwilling to enforce adequately its rules relating to drug advertising.”<sup>21</sup> The Wilkes study’s authors warned that “[t]he public is at risk, because advertisements may be misleading doctors about the appropriateness of a drug, failing to inform doctors of the adverse effects of a drug, and inducing doctors to prescribe more expensive drugs than patients need.”<sup>22</sup> The authors advocated the establishment of a peer review system for pharmaceutical advertisements.<sup>23</sup>

The conclusions presented in the Wilkes study are disturbing and its recommendations for reform are far-reaching. Before adopting a new and untested procedure for restricting pharmaceutical advertising, it would be wise to review the basis of the study’s conclusions. Do the findings withstand analysis? A re-examination of the data shows that the evidence falls short of credible support for the authors’ conclusions.

#### IV. METHODOLOGY

Copies of the completed survey forms used in the Wilkes study were obtained from the FDA by filing a Freedom of Information Act (FOIA)<sup>24</sup> request. To confirm its authenticity, the data from these forms were then compared with the Wilkes study’s reported findings. A replication of the authors’ analysis revealed only a few discrepancies with the authors’ description of the data. For example, the Wilkes study reported 309 completed questionnaires, while 311 were obtained from the FOIA request. Two advertisements had four reviewers (which was not reported), and a few tallies on specific issues varied by one or two votes from the results reported by the authors. No significant discrepancies, however, emerged. The ability to duplicate the authors’ results verified that the data obtained and recorded was the same data used in the original article.

Additional analyses was performed to assess the validity of the authors’ conclusions. Aspects of the survey that Wilkes and his colleagues did not address — such as how often the physicians and pharmacists agreed on their interpretation of the advertisements and their perception of what messages the claims conveyed — also were examined. An assessment of advertising claims requires not only knowledge of the science on which a claim is based, but also expertise in discerning what an advertisement is communicating to its audience (in this case, an audience of practicing physicians). The original results were tested for statistical significance, freedom from bias,

---

<sup>17</sup> *Id.*

<sup>18</sup> *Id.*

<sup>19</sup> *Id.* at 916. Two reviewers agreed on the rejection of 21 advertisements, or 19.3% of all advertisements. The percentage Wilkes and his colleagues report is based only on the 76 advertisements where at least 2 reviewers agreed on some recommendation, and excludes the 33 advertisements where reviewers were not in accord.

<sup>20</sup> *Id.* at 918.

<sup>21</sup> *Id.* at 912.

<sup>22</sup> *Id.* at 918.

<sup>23</sup> *Id.*

<sup>24</sup> Pub. L. No. 89-487, 80 Stat. 250 (1966), as amended 5 U.S.C. § 552 (1988).

and relevance to the issues being addressed.

### V. INTERPRETING AND ANALYZING AN ADVERTISING CLAIM

Assessing the veracity of an advertising claim requires two distinct analytical steps. First, the message that an advertisement conveys must be determined, and, second, a judgment must be made whether that message is consistent with the facts that are known to be true. Because each of these steps depends on the complicated process of comprehending and evaluating communications, it is difficult to assess with confidence the likelihood that an advertisement would mislead the typical viewer.

The accepted procedure for determining whether an advertisement is misleading is to expose it to representatives of the audience for whom it is intended. In considering charges of false and deceptive advertising, "the public's impression is the only true measure of deceptiveness."<sup>25</sup> The potential for a claim to mislead then can be measured by asking viewers what messages the advertisement conveyed to them.

Wilkes and his colleagues, however, did not survey a sample of doctors from the audience targeted by these advertisements. Instead, the authors selected reviewers who were specialists in their fields and probably conversant with scientific literature, but who only could provide indirect evidence of the advertisements' effects on the target audience. It is safe to assume that after studying the advertisements and comparing them to the references supplied, the reviewers were not deceived by the advertisements. Thus, the survey was not intended to determine whether the advertisements misled the reviewers. The premise of the authors' methodology is that reviewers who presumably were not misled would be able to determine whether other viewers of the advertisements were misled.

There is no reason to expect reviewers untrained in the science of consumer perception to have extra insight or expertise on what the target audience might have thought after seeing the advertisements. Wilkes' reviewers were asked to do what reputable scientists in the field of consumer perception are reluctant to do — speculate in the absence of evidence. This practice has been discredited by government agencies<sup>26</sup> and courts<sup>27</sup> that pass judgment on misleading advertising. If there is any question as to the meaning of a claim, little weight is credited to opinions derived without the benefit of evidence, even when those opinions are held by experts qualified to assess the target

---

<sup>25</sup> *FTC v. Brown & Williamson Tobacco Corp.*, 778 F.2d 35, 39 (D.C. Cir. 1985) (quoting LOUIS ALTMAN & RUDOLF CALLMAN, *THE LAW OF UNFAIR COMPETITION, TRADEMARKS AND MONOPOLIES* § 5.04, at 5-32 (4th ed. 1981)).

<sup>26</sup> The Federal Trade Commission (FTC) has recognized that survey evidence is the best means of judging what messages an advertisement may convey. In *Thompson Medical Co., Inc.*, the Commission stated that unless it could "conclude with confidence" that a particular reading of an advertisement was shared by consumers, "we will not find the ad to make the implied claim unless extrinsic evidence allows us to conclude that such a reading of the ad is reasonable." 104 F.T.C. 648, 789 (1984), *aff'd* 791 F.2d 189 (D.C. Cir. 1986). Moreover, the Commission noted, "[t]he extrinsic evidence we prefer to use and to which we give great weight is direct evidence of what consumers actually thought upon reading the advertisement in question." *Id.* Although courts have continued to allow the FTC to rely on its "expertise" in appropriate cases, they also have encouraged the Commission to use extrinsic evidence whenever possible. *Kraft, Inc. v. FTC*, 970 F.2d 311, 321 (7th Cir. 1992), *cert. denied*, 113 S. Ct. 1254 (1993).

<sup>27</sup> In Lanham Act cases, courts generally have required survey evidence to establish the "representative reactions of the trade and consuming public," *Stiffel Co. v. Westwood Lighting Group*, 658 F. Supp. 1103, 1112 (D.N.J. 1987), unless the advertising statement is "false on its face." *Coca-Cola v. Tropicana Prods., Inc.*, 690 F.2d 312, 317 (2d Cir. 1982). See Charles J. Walsh & Marc S. Klein, *From Dog Food to Prescription Drug Advertising: Litigating False Scientific Establishment Claims Under the Lanham Act*, 22 SETON HALL L. REV. 389, 414 (1992).

audience's reactions to a communication.<sup>28</sup>

It is apparent that the reviewers encountered difficulties in interpreting the advertisements from their responses to the assignment of judging if a particular claim appeared. An examination of how often the reviewers agreed on this first step of advertising analysis revealed extensive disagreement, an observation that would not surprise practitioners in perception testing. The reviewers frequently disagreed on the interpretation of the basic claims in the advertisements.

Several questions in the survey allowed for a direct test of the reviewers' ability to identify claims. The three different questions about use of statistics in the advertisement, for example, gave reviewers the opportunity to respond "no statistics cited in advertisement."<sup>29</sup> This question of whether an advertisement did or did not contain statistics should have elicited widespread agreement. It did not.

Ultimately, the reviewers disagreed more often than they agreed. Depending on which of the three questions concerning statistics is examined, reviewers agreed unanimously that the advertisement did or did not cite statistics either 46%, 56%, or 65% of the time. In only 29% of the 108 advertisements, where two or more reviewers saw statistical analysis, did all the reviewers consistently agree that the advertisements contained statistics.

Whether an advertisement cites a statistic or not is a straightforward question. As a result it is difficult, however, to place much confidence in the reviewers' judgments on more subtle questions, such as whether the advertisement included the suggestion that a drug is safe, the drug of choice, or useful in a broad range of applications. Moreover, disagreements were not limited to the questions on statistics. In all of the advertisements, there was disagreement among the reviewers as to whether an advertisement made one or more claims.

Disagreement about the meaning of an advertisement is not unusual. It is because such disagreements are commonplace that the courts require evidence of the message actually conveyed to the target audience. Wilkes and his colleagues skipped this step, assuming that their substantive experts could judge audience reaction.

## VI. EXAMINING THE WILKES STUDY'S RESULTS

The reviewers also had difficulty deciding whether the advertisements were inaccurate or misleading, and often were unsure of criticizing an advertisement. For example, one reviewer somewhat agreed that a table in an advertisement was likely to mislead the reader, because the reviewer could not tell whether the advertisement was peer reviewed.<sup>30</sup> Another reviewer somewhat agreed that a headline was misleading because no data were provided to support a claim of efficacy in the headline.<sup>31</sup> One advertisement was judged misleading because it failed to disclose that other medications were available for the indication.<sup>32</sup> One reviewer stated, "The image is interesting — the subliminal bold color . . . may cause some M.D.'s to 'change to one and only one' antibiotic . . . I would prefer to see imageless ads . . ."<sup>33</sup> Another reviewer con-

---

<sup>28</sup> See, e.g., *Kraft, Inc.*, *supra* note 26. (FTC rejected claim interpretations based on unsupported opinion of a consumer perception expert).

<sup>29</sup> QUESTIONNAIRE, *supra* note 10, at 6.

<sup>30</sup> Reviewer #105, *Advertisement #022*, QUESTIONNAIRE, *supra* note 10, at 5.

<sup>31</sup> Reviewer #050, *Advertisement #078*, *id.*, at 4.

<sup>32</sup> Reviewer #1035, *Advertisement #093*, *id.*, at 10.

<sup>33</sup> Reviewer #1019, *Advertisement #059*, *id.*, at 10.

cluded that an advertisement did not claim superiority over other products, but found it misleading, because “[the drug] is effective . . . , but no more so than any other [drug].”<sup>34</sup> What each of these reviewers found objectionable was not the factual accuracy of the advertisement, but the claim that the advertisement did not make, or the failure of the advertisement to make a claim the reviewer wanted it to make.

The survey needs to be viewed as a whole. The data that the authors collected but did not report yield some interesting clues as to what the reviewers thought about the advertising they saw and the problems they found. A basic testing of the survey design and the analysis of the data reveals that many of the authors’ conclusions cannot withstand scrutiny.

### A. *The Reviewers Rate the Importance of the Problems*

A principal conclusion of the Wilkes study was that many advertisements contain deficiencies in areas relating to FDA standards. Wilkes and his colleagues, however, noted that the reviewers were not trained to apply FDA regulations to pharmaceutical advertising.<sup>35</sup> To elicit the reviewers’ opinions regarding compliance with FDA standards, the authors adapted the language used in the regulations to the questions in the survey instrument. Reviewers were asked, for example, whether an advertisement presented “a fair balance between information relating to efficacy and information relating to side effects/contraindications.”<sup>36</sup> This language comes directly from the Code of Federal Regulations.<sup>37</sup> Numerous other questions also were drawn from the regulations, such as the sections asking whether headlines or illustrations were used in a misleading way.<sup>38</sup> The reviewers, however, were not told how to interpret these criteria.

Wilkes and his colleagues acknowledged that some of the FDA-related problems identified by the reviewers should be regarded as less serious than others. For example, deficiencies in claims that a drug is the drug of choice should not arouse much concern, compared to deficiencies in safety claims.<sup>39</sup> In fact, problems were less frequently found in safety claims (where only 8% of the advertisements were faulted) than in drug of choice claims (where 30% of the advertisements received unfavorable reviews).<sup>40</sup> How serious some of the other problems may be is less obvious, but there is a way to test the gravity of the reviewers’ concern.

The importance of a problem can be measured by examining how it affected the reviewers’ conclusions about the advertisements. Every reviewer was asked whether an advertisement would lead to improper prescribing and whether the advertisement would have been accepted for publication if professional peer review standards had been applied to it. The reviewers’ judgment of the seriousness of any identified problem can be assessed by comparing their findings of problems with their recommendations regarding acceptance and their opinion of whether the advertisement would lead to proper prescribing. If a reviewer would have accepted an advertisement without change or with only minor revisions, or agreed that the advertisement alone, with no other information, would lead to appropriate prescribing behavior, any problem he or she saw cannot be regarded as serious.

---

<sup>34</sup> Reviewer #1042, *Advertisement #045, id.*, at 4.

<sup>35</sup> Wilkes et al., *supra* note 2, at 918.

<sup>36</sup> QUESTIONNAIRE, *supra* note 10, at 6.

<sup>37</sup> 21 C.F.R. §202.1(e)(5) (1994).

<sup>38</sup> QUESTIONNAIRE, *supra* note 10, at 4, 8.

<sup>39</sup> Wilkes et al., *supra* note 2, at 918.

Using the authors' majority rule for determining whether an advertisement had a problem,<sup>41</sup> the average rate of problems was 28% over all the FDA-related criteria. In other words, combining the 1512 claims related to FDA standards that the reviewers identified in the advertisements, they agreed with approximately 75% and disagreed with approximately 25% of the messages. Most of those disagreements, however, did not affect the reviewers' ultimate judgment of the advertisement. If only the problems that affected the recommendation to publish an advertisement or that aroused concerns about proper prescribing are counted, the average rate of disagreements with FDA-related claims drops from 28% to 12% across all categories. Over half the alleged FDA violations did not cause the reviewers to recommend major revision/ rejection of an advertisements, or to believe that the advertisement would lead to inappropriate prescribing behavior. Thus, more than half of the problems that the authors report were not regarded as serious by their reviewers.<sup>42</sup>

### B. *Accepting or Rejecting for Publication*

One of the most surprising conclusions of the study related to the application of peer-review standards to the advertisements. The article reported that when reviewers were asked to judge the advertisements in the same manner they would judge an article submitted for publication in a journal, they recommended major revision or rejection for 61% of the advertisements.<sup>43</sup> While the authors reported the reasons for the rejection of some advertisements, they did not report how often two or more reviewers agreed on the reasons for rejection or revision. This article performs that missing level of analysis, first for major revisions and then for rejections.

When deficiencies are identified by the voting rule Wilkes and his colleagues established (a majority vote of reviewers being sufficient to find a deficiency) the major revisions that the reviewer most often wanted to see were more information or references in the advertisement. The most commonly cited reason for recommending revision was the desire to see more information on efficacy. The second most frequently mentioned deficiency was a failure to cite enough references (*See* table 1). Seldom did reviewers agree that a deficiency in need of correction was misleading. For example, only 4 of the 109 advertisements would have needed major revisions because they contained a misleading statement. Only 2 were found to have misleading graphs or tables that merited major changes. Consequently, by a margin of more than 4 to 1, the reviewers seeking significant changes wanted the advertisements expanded, rather than corrected.

Even when the reviewers voted to reject an advertisement, more often than not their reason was because the advertisements lacked information or references. A desire for more information explained more than twice as many rejections as did a finding of an inaccuracy. FDA regulations, however, do not require advertisements to contain the kinds of references and details the reviewers were accustomed to finding in articles.

---

<sup>40</sup> *Id.* at 914.

<sup>41</sup> At least two reviewers had to identify a claim and then at least two had to agree (either somewhat or strongly) that there was a problem with the claim.

<sup>42</sup> Even this assessment of the importance of problems may overstate the extent to which the reviewers thought problems appeared. A reviewer might decide, for example, to reject an advertisement with five problems, each of which he or she regarded as minor. Under the revised test, each of those minor problems would be counted among the 12% classified as significant.

<sup>43</sup> Wilkes et al., *supra* note 2, at 917.

Physicians are unlikely to be misled by an advertisement because it lacks such references. Nevertheless, many of the reviewers justified rejection of an advertisement for such reasons. Corrections of misleading references, statements, tables, or images explained few of the rejections. Of the 21 rejected advertisements (out of a sample of 109), only 7 would have been rejected because of a statement perceived to be misleading. Three would have been rejected because of a misleading graph or table. (*Table 1 appears on the next page*).

### C. The Survey Design

The foregoing discussion assumes that the survey used by Wilkes and his colleagues accurately reflected the opinions of the reviewers. This assumption is subject to challenge. The interpretation of survey results is impossible without first assessing the design of the survey, because a defective design can yield inaccurate results. In the Wilkes survey, several basic errors in design and analysis compounded one another to impart potential bias.

First, the survey failed to account for a flaw that is commonly encountered in survey construction known as the “yea-saying” bias. Survey respondents tend to agree with a question, and research has shown that, whether they mean to or not, respondents agree in larger numbers than they disagree.<sup>44</sup> Consequently, the form of the questions can exaggerate findings of misleading advertising if affirmative replies indicate deception. In the Wilkes survey, every question pertaining to whether an advertisement was misleading required an affirmative answer for the reviewer to find it misleading, thus this source of potential bias was significant.

Compounding this difficulty was the failure of the questions to give reviewers the option of abstaining — neither agreeing nor disagreeing. If a question regarding FDA compliance was applicable to an advertisement reviewers had only four choices: strongly agree, somewhat agree, somewhat disagree, strongly disagree. For example, one question asked, “[i]f an image is present, it is used in a manner which misleads the reader about the efficacy of the drug.”<sup>45</sup> In response, each reviewer had to state a preference — whether or not he or she had one. A reviewer could not answer, “none of the above” or “no opinion” even though subjective judgments on some of the questions and uncertainty about science on others may have made agreement or disagreement impossible.

When surveys do not give respondents the option of neutrality, it is difficult to know how many respondents wanted to abstain, but instead agreed or disagreed. While some with no opinion may have declined to answer a particular question, most probably selected the option to “somewhat” agree or disagree. Given that the yea-saying bias can affect surveys with neutrality options, this forced choice makes it more likely that the ambivalent subjects will “agree” with the questions.<sup>46</sup> Many reviewers who partially agree and partially disagree with a statement will suppress their disagreement and give the agreeable response. In the Wilkes survey, one reviewer scratched out “somewhat disagree” and substituted “somewhat agree.”<sup>47</sup> Such confusion is understandable, because the word “somewhat” could be interpreted as “both agree and disagree.”

Finally, the problems inherent in yea-saying bias and forced choice were com

---

<sup>44</sup> See e.g. Jacob Jacoby, Remarks to Coalition of Healthcare Communicators, New York, N.Y. (Oct. 20, 1992).

<sup>45</sup> QUESTIONNAIRE, *supra* note 10, at 8.

<sup>46</sup> Jacoby, *supra* note 44.

<sup>47</sup> Reviewer #1026, Advertisement #4, QUESTIONNAIRE, *supra* note 10, at 6.



pounded by the way Wilkes and his colleagues analyzed and reported the data. Equivocal positions were combined with strong positions, and then characterized only as positions. If one reviewer "somewhat agreed" with the statement that an image misled the reader, and another reviewer "strongly agreed," the two responses would be grouped as an unfavorable review. The fact that one voter may have had no opinion or been ambivalent did not prevent his or her vote from providing a verdict against an advertisement.

#### *D. The Impact of the Flaws*

To determine whether forced agreement and yea-saying bias affected the survey results, the opinions that were strongly held were examined separately from those where the reviewer only somewhat agreed or disagreed. In the absence of any forced choice or yea-saying bias, it would be expected that reviewers who somewhat agree or disagree would give similar responses to those who strongly agree or disagree. If one-third of the respondents with strong views agreed with a question, one-third of the reviewers with weaker views also should agree. If the structure of the survey questions biases the results, however, should be more agreement among reviewers who only somewhat agree or disagree.

Considering all twenty-eight questions together, the results indicate the effects of bias. Strongly held opinions agreed approximately 33% of the time: 37% of the answers agreed with the question, and 63% disagreed. This pattern, however, does not extend to the equivocal positions. Among the "somewhat" opinions, the results are reversed: 60% of the "somewhat" answers are agreements, and only 40% are disagreements. These differences, illustrated in Figure 1, are significant.

The contradiction between strong opinions and equivocal opinions is evident in many responses important questions. As shown in Table 2, weaker opinions contradicted strong opinions in all questions about side effects and contraindications, as well as in those questions for which the Wilkes survey reported high percentages of advertisements containing misleading information. This reversal of opinion that occurs in the move from strong to equivocal answers suggests that many of the unfavorable reviews reported by Wilkes and his colleagues resulted from “somewhat” agreements, even though it is difficult to ascertain whether a reviewer who “somewhat” agreed actually agreed, disagreed, or had no opinion at all.

Some of the effects of the yea-saying bias and most of the effects of the forced choice can be reduced by discounting weaker opinions and examining only strong opinions. If a majority rule is applied only to the strong opinions (such that two strong findings are necessary to establish, for example, a tendency to mislead or a lack of fair balance) the unfavorable reviews virtually disappear. A unanimity rule applied to strong opinions eliminates almost every disagreement with the advertisements (*see* Table 3).

An examination of the most important of the FDA-related questions relating to whether an aspect of an advertisement was misleading highlights the extent to which the authors’ results depend on equivocal opinions. When “somewhat agreements” are eliminated from the responses to these questions, and Wilkes’ majority rule is applied, the changes are striking (*see* Figure 2 on page 000).

- The percentage of headlines found to mislead about efficacy drops from 30% to 4%.
- The percentage of headlines found to mislead about adverse reactions drops from 17% to 1%.
- The percentage of graphs and tables found to misrepresent clinical studies drops from 9% to 4%.
- The percentage of graphs and tables found to mislead drops from 26% to 0%.
- The percentage of images found to mislead about efficacy drops from 27% to 1%.

These results do not support the conclusions of Wilkes and his colleagues that a significant number of pharmaceutical advertisements are misleading physicians or violating FDA regulations. Because few reviewers had clear opinions when the authors’







majority rule is applied, their votes contradict the article's conclusions. Instead, the findings are consistent with the responses of the equivocal reviewers. Reviewers who were confident and held strong opinions rarely agreed that there were problems in the advertisements.

#### VII. THE STATISTICAL SIGNIFICANCE OF THE RESULTS<sup>48</sup>

---

<sup>48</sup> In discussing statistical significance, the Wilkes study notes that, for reasons of simplicity, they did not report confidence intervals. Their data was not intended to allow the reader to make inferences about all

The question of confidence applies not only to the individual reviewers, but also to the study as a whole. The projectability of an opinion sample is uncertain. The typical means of limiting the uncertainty in survey research is to collect a large enough sample of observations to provide reasonable confidence that the results actually are representative of the population being studied. In the Wilkes study, the authors attempted to diminish the influence of the mistaken reviewer by taking multiple observations (typically three per ad) and accepting the reviewers' judgments only when two or more agreed on an assessment of an advertisement.<sup>49</sup> The authors, however, did not report how successful they were in discounting the role of chance in their findings.

The first test an analyst performs after collecting evidence in a scientific inquiry is to determine whether the evidence supports a hypothesis different from random chance. Wilkes and his colleagues did not report the statistical significance of their findings, stating that the study was not generalized across all pharmaceutical advertisements.<sup>50</sup> It would not have been appropriate to claim that the results of the Wilkes study revealed some general characteristics about pharmaceutical advertising or practicing physicians. To do so would have required the sample of advertisements and reviewers to be representative of the populations from which they were drawn. Wilkes and his colleagues did not contend that their samples met these criteria. Nevertheless, it is appropriate to ask whether the conclusions drawn about the specific reviewers and these advertisements were established with the confidence that is necessary in scientific research reports. It also is appropriate to compare the results reported to the outcomes that could have resulted from pure chance, and to measure the likelihood that there is a difference between the two.

### A. *Significance of Votes on Individual Advertisements*

Most of the results reported in the Wilkes study were founded on grades given by reviewers to individual advertisements. Each aspect of an advertisement was judged based on the majority vote of two or three reviewers.<sup>51</sup> The aggregate results (for example, the proportions of advertisements that received favorable or unfavorable reviews) were accumulations of the individual reviews. Thus, the validity of the conclusions, such as the percentage of advertisements in compliance with FDA criteria, depended on the reliability of the individual reviews. One way to determine reliability is to assess whether the responses of the reviewers diverged from random chance.<sup>52</sup>

For the twenty-eight questions relating to FDA criteria, the Wilkes study combined votes into two groups. All reviews that "strongly agreed" that there was a problem were combined with answers that "somewhat agreed" in order to count as "unfavorable reviews." "Favorable reviews" similarly were derived by combining "somewhat" agreements with "strong" agreements.<sup>53</sup> Aggregating the results in this manner allowed each

---

advertisements but rather to evaluate a select population of advertisements published in several leading medical journals. Wilkes et al., *supra* note 2, at 915.

<sup>49</sup> *Id.* at 914.

<sup>50</sup> *Id.* at 915.

<sup>51</sup> *Id.* at 914.

<sup>52</sup> On purely technical questions the reviewers could be expected to lend their expertise to the answers. In fact, technical accuracy, to which the reviewers' expertise is most relevant, was also the aspect of the advertisements least often judged as misleading. *Id.* at 914-15. Reviewers also were required, however, to determine what message the advertisement would convey to physicians before they could determine whether that message was accurate, and then to assess whether the claim, accurate or not, would mislead reviewers. On these questions, the reviewers were not experts.

<sup>53</sup> *Id.* at 914.

reviewer to express one of two views — either favorable or unfavorable — about an advertisement.<sup>54</sup> If no distinction is made between strong and weaker votes, greater or lesser weight cannot be accorded to any individual agreement or disagreement, and the results are comparable to a model of chance.

Expanding the scope of the survey from one reviewer to two or three does not improve confidence in the results to generally acceptable levels. With only two possible outcomes, it is not uncommon for a completely random process to generate the same outcome three times in a row. In such a scenario, 12.5% of the time the results of three out of three random trials will be the same; two out of three results will be the same 37.5% of the time. At most, researchers can be only 87.5% confident that a unanimous unfavorable vote from three reviewers was different from chance. The level of confidence generally accepted in scientific inquiry is 95%.<sup>55</sup> Most researchers insist that they be 95% certain that their results were not generated by random chance before they claim to have found something significant. By this standard, there is not a single favorable or unfavorable review on the twenty-eight FDA-related questions reported in the *Annals* article that could be represented as statistically significant, because the number of reviewers was so small.

Although the study's analysis was based on grouping the reviews in two categories, favorable and unfavorable, reviewers actually were given four choices: somewhat agree, strongly agree, somewhat disagree, and strongly disagree.<sup>56</sup> With four choices, there is a better opportunity of distinguishing the results from chance. In this case, the odds that chance would produce three strong findings that an advertisement is misleading is less than 5%. Thus, there is a basis for concluding with 95% confidence that when three reviewers strongly agreed that there was a problem, their votes were different from chance. In 18 of the 28 FDA-related questions, however, no advertisements were unfavorably reviewed by a unanimous, strong vote.<sup>57</sup> None of the questions that asked whether the advertisements were misleading was answered affirmatively by three strong votes. The study offers no evidence that any single advertisement can be judged as misleading by traditional scientific standards; there is no statistically significant evidence that FDA regulations "appear to go unheeded."<sup>58</sup>

### B. Significance of the Vote Pattern

In subsequent statements, Wilkes moderated the conclusion that he and his colleagues had recorded in their article. He stated that they were looking for trends that the reviewers might have discovered in the advertising.<sup>59</sup> This proposition can be tested. The pattern of results that Wilkes and his colleagues obtained may be different from

---

<sup>54</sup>An additional effect of combining strong and weaker opinions is to increase the likelihood that at least two reviewers will agree. With three reviewers and only two choices, at least two reviewers necessarily will agree on each question. With four possibilities, there is no such necessity; it may well be that all three reviewers express different opinions. Thus, the Wilkes study's procedure artificially increases the likelihood that at least two reviewers will agree on a particular question.

<sup>55</sup>LOTHAR SACHS, *APPLIED STATISTICS: A HANDBOOK OF TECHNIQUES* 247 (1984).

<sup>56</sup>Wilkes et al., *supra* note 2, at 914.

<sup>57</sup>The questions that did elicit such reactions typically related to whether sufficient information had been provided, and an average of less than 4% of the advertisements fell short on this basis. *See infra* Table 4.

<sup>58</sup>Wilkes et al., *supra* note 2, at 918.

<sup>59</sup>Michael S. Wilkes, Remarks to Coalition of Healthcare Communicators, New York, N.Y. (Oct. 20, 1992).

chance, even if valid conclusions cannot be drawn about the individual advertisements. No conclusions can be reached about whether an analysis is fair from a single experiment with only three reviewers. If the experiment is repeated 100 times, however, it may be possible to draw conclusions from the pattern of results across the 100 experiments. By asking the same 28 questions for the 109 advertisements reviewed, Wilkes repeated 28 experiments 109 times.

Wilkes and his colleagues did not use the same advertisement for each set of three reviewers. Instead, they used 109 separate advertisements. Testing whether the answers pattern differs from chance first requires an analysis of the results that chance alone produces. Two possible models can be used to explain a distribution of votes that differs from chance.<sup>60</sup> In the first of these, the "problem advertisement" model, votes are presumed to reflect actual problems with the advertisements. This model should yield the same results that Wilkes found. Poor reviews would cluster around certain advertisements. An alternative model is the "biased reviewer" model, in which the votes are assumed to be driven only by differences between individual reviewers (some atypically harsh, others excessively lenient). Poor review in this model would cluster around certain reviewers. The biased reviewer model should yield different results than those reported in the Wilkes study, which assumed that the accuracy of the advertisements, and not the predilection of the reviewers, explained the votes.

### 1. *The Problem Advertisement Model*

The methodology of the *Annals* study assumed that there was a fraction of advertisements with problems, and that reviewers were reasonably accurate in identifying those problems. This can be called the problem advertisement model because it assumed that there were real problems with the advertisements. An independent calculation of the expected number of advertisements for which a given number of reviewers would find problems should be consistent with the results that Wilkes and his colleagues actually obtained, if the "problem advertisement" model accounted for the voting behavior of the reviewers. Such a calculation would be inconsistent with the results of the study if something other than problems with the advertisements accounted for the negative votes.

This article assumes the reviewers were highly accurate in identifying claims and finding them deceptive. The level of accuracy used in the "problem advertisement" model is 90%; it is assumed that if a problem was present, there was a 90% chance that the reviewer reported it. (If there were no problems, there is an assumed 10% chance that the reviewers erroneously identified a problem.)<sup>61</sup> The high rate of disagreement among the reviewers suggests that this assumption is too generous, but it still shall be used for the purpose of this analysis.

To calculate the expected number of advertisements for which reviewers would

---

<sup>60</sup> In applying these models, this article uses Wilkes' procedure of combining strong and weaker opinions into a single vote that a problem exists. The effect of this approach is to increase the amount of apparent agreement among reviewers. Because the statistical tests examine whether there is more or less agreement than would be expected due to chance, combining the categories tends to bias the results in favor of the Wilkes study's hypothesis.

<sup>61</sup> Reasonably high levels of reviewer accuracy are necessary if we are to believe Wilkes and his colleagues actually identified some characteristic of the advertisements. If reviewers are extremely accurate, however, most advertisements should have unanimous judgments, which they do not. The 90% accuracy level was chosen as a reasonable balance between these considerations. As discussed in note 65, the results are not sensitive to the assumed level of accuracy.

find problems, the fraction of advertisements with problems first needs to be determined. Wilkes' assumption that this fraction can be estimated based on the number of advertisements where two reviewers agreed that there was a problem is used herein, because this article seeks to ascertain whether the data are consistent with the Wilkes model.<sup>62</sup> Taking the fraction of advertisements that Wilkes found deceptive and assuming that reviewers are accurate 90% of the time, it is easy to calculate the number of advertisements for which a given number of reviewers would find problems.<sup>63</sup>

Focusing on the twenty-eight questions relating to FDA criteria, for most questions the actual number of advertisements where three reviewers found a problem is significantly different from the expected number that would have occurred if the reviewers had been 90% accurate at identifying real problems.<sup>64</sup> This comparison of the reviewers' actual performance to the performance predicted by the problem advertisement confirms that the actual votes reveal far less agreement about problems than the hypothetical votes of the model. As shown on Table 4, actual responses were consistent with hypothetical responses for only 8 of the 28 questions relating to FDA criteria.<sup>65</sup> Only on these questions is it possible to ascertain that the reviews formed a pattern that may be consistent with problems in the advertisements.<sup>66</sup> (The pattern, however, does not indicate anything of significance about any individual advertisement.)

Accurate reviewer assessments of real problems should produce relatively high levels of agreement among the reviewers that particular advertisements either satisfied or failed to satisfy a standard applied by the FDA. In reality, however, there are much higher levels of reviewer disagreement than the problem advertisement assumption can explain. Thus, for the vast majority of the criteria Wilkes and his colleagues examined, the data are not consistent with the implicit assumptions that reviewers are accurate and that some advertisements have problems.

## 2. *The Biased Reviewer Model*

Before attaching significance to the eight criteria for which the reviewers' votes were consistent with the "problem advertisement" model, a second question needs to be

---

<sup>62</sup> This article's analysis was confined to advertisements that were reviewed by three reviewers. Thus, the sample of advertisements ranges from 45 to 90, depending on the number of reviewers who answered a particular question. Across all 28 questions, the average number of advertisements reviewed by three reviewers was 82. A similar analysis is possible with advertisements reviewed by only two reviewers, but the relatively small number of such advertisements would make this a weak test.

<sup>63</sup> For example, if reviewers are 90% accurate in identifying problems, then for an advertisement with a problem, all three reviewers should identify a problem about 73% of the time ( $.9 \times .9 \times .9$ ). If 50% of the advertisements have problems, then in the Wilkes study sample, three reviewers should agree that a problem is present in about 37% of the advertisements ( $73\% / 2$ ). Similarly, it is possible to calculate how many advertisements should have two reviewers who find a problem, only one reviewer who finds a problem, or no reviewers who find a problem. *See infra* statistical appendix.

<sup>64</sup> This methodology could also be used to compare the number of advertisements where two, one or none of the reviewers found problems.

<sup>65</sup> This article's results are not sensitive to the assumed rate of reviewer accuracy. In fact, if the level of reviewer accuracy is assumed to be 50% (closer to the 52% to 68% level of agreement that Wilkes and his colleagues report for the reviewers), there are only two FDA-related criteria for which reviewer results are consistent with the problem advertisement model. (On both of those criteria, the hypothesis that the "problem" was identified by a random selection of minority reviewers whose answers had nothing to do with an individual advertisement cannot be rejected.)

<sup>66</sup> This test is particularly likely to accept the possible validity of the problem advertisement model when the estimated proportion of advertisements with problems is low. In fact, for all of the questions for which the problem advertisement model cannot be rejected, the estimated incidence of problems is below 8%, and in 5 of the 8 questions, the estimated incidence of problems is less than 4%.





asked: can the results be distinguished from a random scattering of predetermined opinions? Suppose that the advertisements have nothing to do with the pattern of reviewers' votes, and that an individual reviewer was equally likely to find problems no matter what advertisement he or she saw. Could this explain the pattern of results reported in the Wilkes study? This also can be tested.

The starting point for the "biased reviewer" model is the assumption that the reviewers' predilections will influence how they interpret an advertising claim, whether they agree with it, and whether they believe it is stated properly. Advertising interpretation can challenge respondents who have no expertise in communications.<sup>67</sup> Even in their area of expertise reviewers may hold different opinions, because the scientific community is unlikely to agree unanimously on many propositions put before it. For example, the question of whether advertisement headlines were "used in a way that will promote appropriate use of the drug" likely will engender differing responses from different physicians and pharmacists, who have different opinions about the "appropriate use" of a given drug.<sup>68</sup> Even if the vast majority of experts — and the FDA — agree that a particular use is appropriate, there will be some individual reviewers who sincerely disagree with that majority opinion and will find a problem with advertisements discussing that use. Through the operation of chance, some advertisements will be assigned to two reviewers who both disagree with the prevailing opinion among experts.<sup>69</sup> Unless the reviewers agree about both the claims in the advertisements and the scientific basis for the claims, a voting procedure such as that used by Wilkes and his colleagues will identify problems in some advertisements because they are assigned to two reviewers who disagree with the consensus of the scientific community.

To determine whether reviewers' responses can be explained by the "biased-reviewer" model, the chance that a particular reviewer will disagree with a particular claim, regardless of the advertisement being reviewed, must be calculated. The best estimate of that chance is obtained by pooling all the responses to each question over all the advertisements, and assuming that each reviewer is as likely as the group to agree or disagree with the question. In response to the question about whether headlines encourage appropriate use, for example, 30.5% of the reviewers, over all the advertisements under review, said that it did not.<sup>70</sup> If that same 30.5% chance of disagreement applies to every reviewer, regardless of what advertisement he or she reviews, it is straightforward to calculate the number of advertisements in which chance assignment of reviewers would lead to three reviewers saying there is a problem.<sup>71</sup> Whether the reviewers' responses differed from those that would be expected due to chance alone can be tested. If a difference exists, there may be some significance in the pattern.

---

<sup>67</sup> See *supra* notes 26-28 and accompanying text.

<sup>68</sup> See QUESTIONNAIRE, *supra* note 10, at 4. A recent *Journal of the American Medical Association* article found that several years lapsed between findings of the benefits of new therapies and the time a majority of experts in the field recommended them for specific indications. Eliot M. Antman et al., *A Comparison of Results of Meta-analyses of Randomized Control Trials and Recommendations of Clinical Experts*, 268 JAMA 240 (1992).

<sup>69</sup> Even if there were unanimity about the facts, disagreements among reviewers about what the advertisement said would produce the same problem.

<sup>70</sup> This calculation is based on asking whether a reviewer thought there was a problem, defined as strongly or somewhat disagreeing with the statement that the headline would promote appropriate use. Reviewers who thought that the headline made no claim about efficacy are counted as finding "no problem."

<sup>71</sup> Although the same analysis with advertisements that are assigned to only two reviewers could be conducted, the relatively small number of advertisements would mean that differences between the results expected and chance would almost never be significant.

For the question about whether headlines encourage appropriate use of the drug, the pattern of responses cannot be distinguished from a random assignment of biased reviewers. Nor can chance be ruled out for the questions about whether readers are misled about efficacy or side effects/contraindications. Indeed, as shown in Table 5, for 15 of the 28 problem criteria, the hypothesis that the results are due to chance alone cannot be rejected. Consequently, for a majority of the criteria that Wilkes and his colleagues examined, the pattern of results could be due to chance. More than half the time, the advertisements examined could have had nothing to do with the reviewers' conclusions of whether or not there was a problem.

### 3. *Combining the Two Models*

Considered together, the two methods of analyzing the distribution of reviewers' votes indicate that few of the reported results can be distinguished from chance. The problem advertisement model identified only eight criteria for which the reviewers voted as if they were accurately identifying problems in the advertisements, and the biased reviewer model identified only twelve criteria in which it could confidently be said that the reviewers' votes were distributed differently from chance. As shown in the final column of Table 5, the number of criteria that meet both standards is four. Of the twenty-eight FDA-related areas where the reviewers were asked whether problems existed, in only four did they answer with a degree of consistency that is both different from chance and consistent with the notion that there may be problems in some advertisements.

Of the four questions for which the reviewers' responses pass the statistical tests in both the problem advertisement and biased reviewer models, only one relates to misrepresentation. That question addressed whether graphs and tables misrepresented conclusions of clinical studies.<sup>72</sup> Only 23 advertisements were identified as having made the claim, and only 9% of these (or 2 of the 109 advertisements) were found by the reviewers to have misled the reader. The fact that the pattern of answers to this question passed these tests of statistical significance does not indicate that the reviews of these two advertisements or any other individual advertisements were statistically significant.

Neither the reviews of the individual advertisements, nor the patterns of answers to the twenty-eight FDA-related questions, reveal any statistically significant conclusions about the accuracy or misleading nature of the advertisements in the Wilkes study.

## VIII. THE CALL FOR MORE REGULATION

A careful analysis of the data provides no support for the judgment Wilkes and his colleagues rendered on pharmaceutical advertising.

The finding of problems in a large proportion of pharmaceutical advertisements is troublesome, given research suggesting that drug advertising serves as a major source of information for practicing physicians.

---

<sup>72</sup> QUESTIONNAIRE, *supra* note 10, at 4. The other three questions that passed both tests asked the reviewers for their opinions on: (1) the adequacy of references for a claim, which has nothing to do with whether the reviewer was misled by the claim; (2) whether one drug was more effective than another, where only nine advertisements were reviewed unfavorably (none with unanimous and strong opinions); and (3) whether the advertised drug was useful in a broad range of patients, where only six advertisements received unfavorable reviews (none with unanimous and strong opinions). *Id.* at 2-4.





[S]tandards for honesty, accuracy and balance in pharmaceutical advertising currently exist in the form of FDA regulations, but these regulations appear to go unheeded . . . .<sup>73</sup>

It is therefore appropriate to reassess the authors' recommendation for reform.

Existing regulations may be adequate if combined with stricter federal enforcement (perhaps through requiring federal approval of an advertisement before publication), larger penalties for improprieties, and the involvement of journals and professional organizations in the monitoring of drug advertising . . . .

[To] assure such compliance, a meaningful program of surveillance must be developed by the FDA, the journals, or both.<sup>74</sup>

The authors suggested a new program of surveillance, the presumed effect of which would be to eliminate or substantially alter much of the advertising that appears in medical journals today. Wilkes and his colleagues acknowledged that any such program would impose costs,<sup>75</sup> but they did not explore the implications of increasing the costs on the quantity of promotion. The immediate impact of an increase in pharmaceutical advertising costs would be a decrease in the use of advertising resources.

Any assessment of reform that only focuses on the financial cost of a new advertising review mechanism will miss the far greater cost that patients will bear because of the reduction of pharmaceutical information reaching physicians. The benefits of regulation must be weighed against this principal cost.

For most doctors, the first formal source of systematic information about drug therapy is medical school, but that experience cannot provide all of the information that a physician will need to become an effective practitioner. Moreover, the store of knowledge about specific drug products that comes with a medical degree quickly becomes obsolete unless it is updated continually.<sup>76</sup> The challenge confronting the practicing physician is how best to remain current with advancing science. Patients hope that their doctors continue their education by reading relevant articles reporting new findings about old and new drugs. There would be no time left to practice medicine, however, if a practitioner tried to absorb the thousands of articles that are published each year in the hundreds of scientific journals. Every hour a physician spends on his or her education is an hour away from administering to patients. It is not surprising, therefore, that studies of physician-prescribing behavior indicate that doctors tend to prescribe drugs that were state of the art when they were in medical school.<sup>77</sup>

Information that can reduce sickness, shorten hospital stays, avoid expensive procedures, and save lives is not immediately converted into the practice of medicine because learning takes time and time (especially a doctor's time) is expensive. A recent

---

<sup>73</sup> Wilkes et al., *supra* note 2, at 918.

<sup>74</sup> *Id.*

<sup>75</sup> *Id.* at 918-19.

<sup>76</sup> In fact, medical students do not have to wait until graduation before technology outruns their knowledge of pharmaceuticals. One study found that many medical school text books were obsolete at the time they were being used. Antman et al., *supra* note 68, at 245.

<sup>77</sup> "[P]rescribing patterns are well predicted by the age of physicians (and even better by the year of entry into medical school); that is there is a differential reliance on the products that were state-of-the-art while the physician was training." Keith Leffler, *Assessing Prescription Drug Advertising: Is Information the Proper Criteria?* Comments for UCLA Public Health-Economics Seminar 4 (Dec. 10, 1992).

article published in the *Journal of the American Medical Association* found that several years elapsed between the findings that five new therapies significantly reduced deaths from heart attacks and the time a majority of experts in the field recommended them.<sup>78</sup> Another survey found that a minimum of five years elapsed between the introduction of a medical innovation and the appreciable diffusion of that innovation into practice.<sup>79</sup>

Delays are expensive. The benefits of rapid diffusion were measured in a 1975 study that compared the diffusion of the anti-psychotic drug chlorpromazine to the diffusion of the polio vaccine.<sup>80</sup> The study found that if this drug, which reduced the average hospitalization of mental patients by one-third, had been recognized by physicians as rapidly as they had adopted the polio vaccine, the nation's health care system could have saved 645,000,000 patient days between 1961 and 1964.<sup>81</sup>

By helping speed the diffusion of knowledge about new pharmaceuticals, advertising can be a significant force for reducing unnecessary expenditures on health care. Researchers have found that pharmaceutical advertising substantially increases knowledge. One study showed that physicians' awareness of the association between certain clinical disorders and specific products more than tripled, from 18% to 57%, after seven months of journal advertising.<sup>82</sup> Diffusion increased further when the advertising was supplemented by other methods of promotion.<sup>83</sup> The informational value of pharmaceutical advertising and promotion was confirmed in an empirical analysis that found promotional expenditures were associated with the rapid entry and adoption of new therapies.<sup>84</sup>

These studies of pharmaceutical advertising's value are consistent with research on other industries. Studies have found that advertising lowers prices,<sup>85</sup> accelerates innovation,<sup>86</sup> and diffuses knowledge more broadly among the population.<sup>87</sup> Similarly, when pharmaceutical products are approved for new uses, promotional expenditures are associated with a more rapid diffusion of the new therapy among physicians, and with lower prices for competitive products.<sup>88</sup> There is widespread recognition of the essential function that information performs in virtually every market where its effect has been examined, and the pharmaceutical market is no different.

Advertising that distorts the facts or misleads the reader generates little support, but there is much support for government agencies that protect buyers from misleading promotional claims. An unsettled question, however, is how to regulate pharmaceutical advertising without depriving doctors and patients of important information. Just as FDA regulations recognize that the failure to provide certain information in advertising

---

<sup>78</sup> Antman et al., *supra* note 68, at 243.

<sup>79</sup> MARY L. FENNELL & RICHARD B. WARNECKE, *THE DIFFUSION OF MEDICAL INNOVATIONS: AN APPLIED NETWORK ANALYSIS* (1988).

<sup>80</sup> Sam Peltzman, *The Diffusion of Pharmaceutical Innovation*, in *DRUG DEVELOPMENTS AND MARKETING* (Robert Helms ed., 1975).

<sup>81</sup> *Id.* at 22.

<sup>82</sup> PHARMACEUTICAL MFRS. ASS'N, *MARKETING AND PROMOTION OF PHARMACEUTICALS: ADDING VALUE TO THE AMERICAN HEALTH CARE SYSTEM* 39 (1992).

<sup>83</sup> *Id.*

<sup>84</sup> Keith Leffler, *Persuasion or Information? The Economics of Prescription Drug Advertising*, 24 *J.L. & ECON.* 45 (1981).

<sup>85</sup> Lee Benham, *The Effect of Advertising on the Price of Eyeglasses*, 15 *J.L. & ECON.* 337 (1972); Robert L. Steiner, *Does Advertising Lower Consumer Prices?* 37 *J. MKTG.* 19 (1973).

<sup>86</sup> Pauline M. Ippolito & Alan D. Mathios, *Health Claims in Advertising and Labeling/A Study of the Cereal Market* (Aug. 1989).

<sup>87</sup> *Id.*

<sup>88</sup> J. Howard Beales, *New Uses for Old Drugs*, in *COMPETITIVE STRATEGIES IN THE PHARMACEUTICAL INDUSTRY* (Robert Helms ed., 1995).

may lead to poor prescribing decisions by individual physicians, the regulatory mechanism also must recognize that the inability to provide any information can produce precisely the same result. One of the means of protecting buyers contemplated in the Wilkes study is a ban of all pharmaceutical advertising.<sup>89</sup> While this certainly would rid the market of deceptive claims it also would silence claims that help heal patients. Neither the Wilkes report nor the regulatory authorities have given this approach serious consideration. Most agencies charged with regulating information appreciate the value of information that is truthful and not misleading. Considering the mandate of the First Amendment, they have no choice.

The free flow of commercial information is valuable enough to justify imposing on would-be regulators the costs of distinguishing the truthful from the false, the helpful from the misleading, and the harmless from the harmful.<sup>90</sup>

Although categorical bans of truthful statements are easily recognized and rejected, random bans of truthful speech are more difficult to detect and more tempting. Wilkes and his colleagues advocated a peer review mechanism for determining which pharmaceutical advertisements to ban.<sup>91</sup> While such a review system seems attractive and reasonable, this article's analysis of the Wilkes study's data indicates that it would mistakenly find truthful messages to be misleading and prohibit the communication of those claims. If this process were adopted as the new clearance procedure, some advertising that does not mislead would be rejected (or revised to the point that an advertiser would lose the incentive to run it), physicians would lose valuable information, and patients could not obtain some important treatments.

## IX. CONCLUSION

There has been no demonstration that there is a serious problem with pharmaceutical advertising or a flaw in the FDA's performance in protecting against misleading advertising. The most significant survey of pharmaceutical advertising to date, the Wilkes study, failed to report results with generally accepted levels of significance. Given the current evidence, it cannot be concluded that the nation's physicians are being misled by pharmaceutical advertisements in medical journals.

On one issue, there is widespread agreement, which the Wilkes study's authors share: pharmaceutical advertising delivers important information to physicians, who can put this information to work healing patients. Unless and until it can be demonstrated that an advertisement misleads physicians, the information it communicates should not be suppressed by reviewers.

### STATISTICAL APPENDIX: THE PROBLEM ADVERTISEMENT MODEL

The problem advertisement model is the model that Wilkes and his colleagues implicitly assumed to have generated the data. This appendix explains the logic of that model, works through an example of its application to one of the criteria that they examined, and presents the test statistics for both the problem advertisement and the

---

<sup>89</sup> Wilkes et al., *supra* note 2, at 918.

<sup>90</sup> *Shapiro v. Kentucky Bar Ass'n*, 108 S. Ct. 1916, 1924 (1988) (quoting *Zauderer v. Office of Disciplinary Counsel*, 471 U.S. 626, 646 (1985)).

<sup>91</sup> Wilkes et al., *supra* note 2, at 918.

problem reviewer models.

Figure 3 illustrates the logic of the problem advertisement model in the form of a decision tree. A particular advertisement is selected, which in fact is either deceptive or not deceptive.<sup>92</sup> The top of the figure represents deceptive advertising, while the bottom branch represents nondeceptive advertising. In either case, three reviewers are selected who evaluate the advertisement to determine whether it is deceptive or not.<sup>93</sup> On either branch, the last column of the figure depicts possible outcomes: three reviewers may agree that the advertisement is deceptive, only two reviewers may agree, only one reviewer may find the advertisement deceptive, or no reviewers may find the advertisement deceptive.

The probability of reaching any given branch in the tree must be determined to apply the model. Wilkes and his colleagues assumed that the probability that an advertisement is deceptive can be calculated by determining the number of advertisements whose deceptiveness is agreed on by at least two reviewers. Thus, the number of advertisements where at least two reviewers say the advertisement is deceptive divided by the number of advertisements with three reviewers provides the probability that the advertisement is deceptive.

If the advertisement is deceptive, the likelihood that three reviewers will agree it is deceptive is determined by the assumed reliability of the reviewers. It is assumed that

---

<sup>92</sup> Wilkes and his colleagues assume a two step process. The selected advertisement either makes a claim about a particular criterion or does not make a claim, and only if it makes a claim is that claim evaluated as deceptive or not deceptive. The probability that a selected advertisement is deceptive in this one-step model is the product of the probability that the advertisement makes the claim and the probability that the claim is deceptive given that it is made.

<sup>93</sup> This analysis is confined to the advertisements that had three reviewers. Although the tests could be applied to advertisements with two reviewers, the number of such advertisements is too small for statistically reliable results.

reviewers are 90% reliable, in the sense that if the advertisement is in fact deceptive, there is a 90% chance that each reviewer will identify it as deceptive, and a 10% chance that he or she will say it is not deceptive. Thus, the probability that all three reviewers will find the advertisement deceptive is  $.9 \times .9 \times .9$ , or 72.9%. There are three possible ways that two reviewers could find the advertisement deceptive: reviewers 1 and 2 could agree, reviewers 1 and 3 could agree, or reviewers 2 and 3 could agree. The probability of each outcome is  $.9 \times .9 \times .1$ , or 8.1%, so the total probability that exactly two reviewers will agree that a deceptive advertisement is deceptive is  $3 \times .081$ , or 24.3%. Similarly, there are three ways that only one reviewer will find the advertisement deceptive, and the probability of each is  $.1 \times .1 \times .9$ , or 0.9%, for a total probability that exactly one reviewer will find the advertisement deceptive of 2.7%. Finally, the probability that all three reviewers will say the advertisement is not deceptive is  $.1 \times .1 \times .1$ , or 0.1%. The probability that three, two, one, or no reviewers will find the advertisement deceptive when it is not deceptive can be computed in an analogous fashion.

The expected number of advertisements in the sample for which three reviewers say the advertisement is deceptive now can be calculated. There are two possibilities: either the advertisement is deceptive, and three reviewers agree, with a probability equal to .729 multiplied by the probability the advertisement is deceptive, or the advertisement is not deceptive and three reviewers are mistaken, with a probability equal to 0.001 multiplied by the probability the advertisement is not deceptive. The expected number of advertisements for which two reviewers will find deception similarly can be calculated.

A statistical test compares this expected distribution if the problem advertisement model is correct to the actual distribution found in the sample. If the two distributions are close, we cannot reject the problem advertisement model. If, however, the two distributions differ substantially, the problem advertisement model does not accurately

describe the data. The test statistic is a chi-square statistic, based on the comparison of the expected and the actual distributions.

The application of this model to claims that a drug is the "drug of choice" for at least one condition is illustrated in Figure 4. There were 89 advertisements for which three reviewers answered the question. The estimated probability that an advertisement is deceptive if the model is correct is 14.61%. This probability multiplied by the 72.9% chance that three reviewers who are 90% accurate would find the advertisement deceptive implies that in 10.65% of the cases, the advertisement is deceptive and three reviewers will find it deceptive. On the bottom branch, if the advertisement is not deceptive, there is only a 0.1% chance that three reviewers will label it deceptive. Multiplying this probability by the probability that an advertisement is not deceptive, 0.09% of the advertisements in the sample should be nondeceptive but identified as deceptive by three reviewers. The other probabilities in Figure 4 are calculated in the same fashion.

The expected number of advertisements in the entire sample where a particular number of reviewers find the advertisement deceptive can now be calculated. In 10.65% of the advertisements, the advertisement is deceptive and three reviewers find it deceptive. In an additional 0.09% of the advertisements, the advertisement is not deceptive but three reviewers erroneously label it deceptive. Thus, it is expected that three reviewers would label 10.74% of the 89 advertisements, or about 10 advertisements, deceptive.

Table A1 presents the expected number of advertisements for each possible outcome, along with the actual number. The expected and actual numbers of advertisements differ substantially. If the problem advertisement model were correct, substantially more advertisements where three reviewers identify a problem than actually found would exist. The value of the chi-square test statistic is 21.97, with 3 degrees of freedom. This value is too large to be due to chance if the problem advertisement model is correct. Consequently, the problem advertisement model does not fit the data for this question.

The statistical tests for each question for the problem advertisement model, as well as for the random reviewer model, are presented in Table A2.





