Scientific quality of original research articles on environmental tobacco smoke

Deborah E Barnes, Lisa A Bero

Abstract

***Objective*—To evaluate the scientific quality of original research articles on the health effects of environmental tobacco smoke; to determine whether poor article quality is associated with publication in non-peer-reviewed symposium proceedings or with other article characteristics.**

***Design*—Cross sectional study of original research articles on the health effects of environmental tobacco smoke published in peer reviewed journals and non-peer-reviewed symposium proceedings from 1980 to 1994. Article quality was assessed by two independent reviewers who used a valid and reliable instrument, were unaware of study hypotheses, were blinded to identifying characteristics of articles, and had no disclosed conflicts of interest.**

***Participants*—All symposium articles (n = 68) and a random sample of peer reviewed journal articles (n = 68) that satisfied inclusion/exclusion criteria.**

***Main outcome measure*—Mean quality scores, which could range from 0 (lowest quality) to 1 (highest quality).**

***Results*—Using multivariate regression analysis, symposium articles were of poorer scientific quality than peer reviewed journal articles when controlling simultaneously for the effects of study design, article conclusion, article topic, and source of funding acknowledged (P = 0.027). Article quality was not associated with either source of funding acknowledged or article conclusion in multivariate analyses.**

***Conclusions*—In published reports on environmental tobacco smoke, non-peer-reviewed symposium articles tend to be of poor quality. These articles should not be used in scientific, legal, or policy settings unless their quality has been independently assessed.**

(*Tobacco Control* 1997;6:19–26)

Keywords: environmental tobacco smoke; peer review; quality assessment; symposium proceedings

Institute for Health Policy Studies, School of Medicine, University of California, San Francisco, California, USA
DE Barnes

Department of Clinical Pharmacy, School of Pharmacy, University of California, San Francisco, California, USA
LA Bero

Correspondence to:
Dr Lisa A Bero, Institute for Health Policy Studies, 1388 Sutter Street, 11th floor, San Francisco, California 94109, USA; e-mail: bero@cardio.ucsf.edu

A symposium is a meeting or conference for discussion of a subject, or a collection of articles on a particular topic. A written symposium is usually based on a conference proceeding, and is typically published as either a supplement or special section of a scientific journal or as an individual book or monograph. Symposia are often industry sponsored and they are rarely peer reviewed, giving rise to concern that they may contain articles that are biased in content and poor in quality.[1 3]

The tobacco industry has sponsored at least four symposia related to environmental tobacco smoke.[4–7] Two additional symposia on this subject[8 9] did not explicitly acknowledge tobacco industry sponsorship but were organised by individuals with tobacco industry affiliations.[2] We have previously examined the content of symposia on environmental tobacco smoke,[2] finding that symposia are not balanced and that they tend to support the tobacco industry position that environmental tobacco smoke is not harmful to health. It is not known, however, whether symposium articles are also poor in quality.

It is important to examine the quality of symposium articles on environmental tobacco smoke because the tobacco industry has used data from symposia in scientific, legal, and policy settings. For example, the tobacco industry cited findings from symposium studies in its criticism of the US Environmental Protection Agency's risk assessment of environmental tobacco smoke.[10] In addition, R J Reynolds Tobacco Company has run advertisements that use quotes from symposia to suggest that exposure to environmental tobacco smoke is not hazardous.[11] Furthermore, several scientists who have published articles in tobacco industry sponsored symposia have testified as expert witnesses on the tobacco industry's behalf.[12] An examination of the quality of symposium articles may therefore prove useful to judges, policy makers, scientists, and journalists, who must decide whether to incorporate symposium articles on environmental tobacco smoke into their deliberations.

The primary objective of our study was to compare the overall scientific quality of peer reviewed journal articles and non-peer-reviewed symposium articles on environmental tobacco smoke. We defined scientific quality as the quality of design, reporting, analysis, and interpretation. We evaluated the overall scientific quality of publications, rather than simply the quality of methodology, because we are interested in guiding the use of scientific publications in policy and legal settings. A scientific publication should be well designed, accurately reported, properly analysed, and appropriately interpreted if it is used to guide policy or legal decisions.

In addition to comparing the quality of peer reviewed and symposium articles, we evaluated several other characteristics of articles that we hypothesised might be associated with quality.

These included research sponsorship (the sources of funding acknowledged), article conclusion (that environmental tobacco smoke is harmful or not harmful), article topic, and study design.

Previous studies have suggested that both research sponsorship and article conclusions have influenced the overall content of scientific reports. For example, several studies have shown that research sponsored by interested parties tends to produce results that are favourable to the sponsors.[13 17] Similarly, the existence of publication bias suggests that the reports may contain a preponderance of positive studies.[18 19] However, the relation between these factors and the quality of the reports has not been studied.

We hypothesised a priori that poor article quality would be associated with publication in symposium proceedings, with tobacco industry sponsorship, and with a conclusion that environmental tobacco smoke is not harmful to health. Article topic and study design (experimental versus observational) were evaluated as potential confounding variables.

Methods

SELECTION OF PEER REVIEWED AND SYMPOSIUM ARTICLES

Peer reviewed articles on the health effects of environmental tobacco smoke were identified by searching Medline from 1 January 1980 to 15 March 1994 using the key words "environmental tobacco smoke," "tobacco smoke pollution," "passive smok#," "involuntary smok#," or "sidestream smoke," retrieving a total of 2189 citations. The search results were narrowed electronically by excluding non-English language publications (n = 393); by excluding studies not related to human subjects (n = 215); and by excluding articles classified by Medline as reviews, news items, editorials, commentaries, or letters to the editor (n = 504). The remaining 1077 articles were evaluated for inclusion.

Articles were included if they described the study of the health effects of environmental tobacco smoke in humans, were original research studies, and were published in peer reviewed journals, as defined below. Of the 1077 articles considered, 386 (36%) satisfied our inclusion/exclusion criteria.

Studies of environmental tobacco smoke

Articles were included if environmental tobacco smoke was the focus of the study, or if it was one of several risk factors examined. Studies of the effects of maternal smoking on the fetus were excluded because the fetus is exposed to mainstream smoke rather than environmental tobacco smoke.

Studies of health outcomes

Articles were included if they studied any health outcome related to environmental tobacco smoke, ranging from lung cancer to eye irritation. Articles were excluded if they discussed policy issues related to environmental tobacco smoke, measured exposure without assessing the effects, or discussed theoretical issues related to study design or analysis.

Studies conducted in human subjects

Only studies conducted in human subjects were included. In vitro, cellular, and animal studies were excluded.

Studies that presented original research findings

Original research articles were defined as articles that presented data based on research conducted by the authors. Articles were included if they incorporated a short review of published reports followed by a presentation of original research findings. Articles were also included if they presented reanalyses of another investigator's data, because the reanalysis was considered to be original work. Meta-analyses, which would have required a different instrument for quality assessment, were excluded but are being assessed in a different study.

Studies published in peer reviewed journals

Articles were included only if they were published in peer reviewed journals. A journal was considered peer reviewed if it explicitly stated that its articles were peer reviewed, if it published a list of peer reviewers, or if it required submission of multiple copies of manuscripts for review before publication. Articles published in symposium proceedings within a peer reviewed journal were considered for inclusion as symposium articles, as described below.

Symposium articles on environmental tobacco smoke were identified through three sources. (1) For a previous study,[2] we identified 11 symposia that primarily contain articles on environmental tobacco smoke by searching Medline, Catalog, Dialog, Conference Papers Index, Toxline, and International Guide to Periodicals using the key words "environmental tobacco smoke," tobacco smoke pollution," "passive smoking," or "involuntary smoking" *and* "symposia," "proceedings," or "conference paper." (2) Additional articles were identified by searching Medline as described above for peer reviewed articles, except that only symposium articles were considered for inclusion. (3) Finally, we identified symposia on indoor air by searching Catalog using the title words "indoor air" and either "symposi#," "proceedings," "conference," "meeting," or "congress", and we then examined their contents for articles on environmental tobacco smoke; 301 articles were identified through these three sources and were considered for inclusion. Articles were included if they studied the health effects of environmental tobacco smoke in humans and were original research studies, as defined above for peer reviewed articles, and if they were published in a symposium. "Symposium" was defined as a publication that contains a collection of articles based on a conference or proceeding.

Of the 301 symposium articles identified, 90 (30%) satisfied our inclusion/exclusion criteria. However, 22 (24%) of these were excluded because they were duplicate publications; that

is, they discussed the same outcome data from the same group of research subjects as another symposium article in the study. In one case, the same article had been published in four different symposia.[20-23] We identified duplicate publications within the symposium article database by comparing articles that had at least one author with the same last name. The article that appeared to present the most complete discussion of the data was selected for inclusion. (A similar technique was used to identify duplicate publications within the peer reviewed article database; none were identified.)

Our final sample consisted of 68 articles from 22 symposia: 31 articles had been included in our previous study of symposia on environmental tobacco smoke, while 37 articles were from new symposia identified through Medline or Catalog. The number of articles per symposium ranged from 1 to 11, with a median of 2 (interquartile range: 1 to 4). None of the symposia were peer reviewed.

Sixty eight peer reviewed journal articles were randomly selected for comparison with the 68 symposium articles. Because the quality of the scientific literature has improved over time,[24-26] we wanted to ensure that the publication dates of the peer reviewed and symposium articles did not differ significantly. We therefore determined the number of symposium articles published from 1980 to 1984 (n = 17), 1985 to 1989 (n = 32), and 1990 to 1994 (n = 19) and used a random number generator to select the same number of peer reviewed journal articles within each time block. The 68 peer reviewed articles selected had been published in 41 journals. The number of articles per journal ranged from 1 to 6, with a median of 1 (interquartile range: 1 to 2).

A minimum sample size of 38 articles in each comparison group (symposium *v* peer reviewed) was necessary to detect a difference in quality scores of 0.10 (on a scale of 0 to 1) at a two tailed $\alpha = 0.05$, $\beta = 0.05$ (95% power), and standard deviation = 0.12.[27]

QUALITY ASSESSMENT

We hired two reviewers to assess the quality of the articles independently, using an instrument that has previously been tested for validity and reliability.[28] Both reviewers were postdoctoral fellows who had received doctorate degrees in social psychology as well as MPH (Masters in Public Health) degrees in epidemiology. Reviewers were trained to use the instrument and were given a detailed set of instructions for referral during the study. In order to minimise potential biases,[29 30] reviewers were unaware of our study hypotheses, did not have extensive knowledge of the literature on environmental tobacco smoke, stated that they had never been involved with either pro-tobacco or tobacco control groups, and stated that they did not have any financial conflicts of interest.

Articles were masked so that identifying characteristics—such as author names and affiliations, journal and book titles, dates of publication, sources of funding, and acknowledgments—had been completely removed. Articles were sent to reviewers in a random order using a random number generator on a computer.

Our quality assessment instrument is designed to measure the quality of design, reporting, analysis, and interpretation of original research articles.[28] This instrument was specifically designed to be applicable to a wide range of study designs, regardless of article topic. Compared to other published quality assessment instruments, ours has high reliability.[31] The instrument consists of a 22 item questionnaire that asks reviewers to determine whether certain items associated with high quality are present in an article (items listed in table 2). Reviewers may answer that the item is "present", "partially present", "not present", or "not applicable" for the article. Reviewers are also asked to classify the study design of the article into one of 15 categories. An overall quality score is calculated based on the study design and the number of items present using a scoring system has been previously described.[28] Quality scores may range from 0 (lowest quality) to 1 (highest quality).

Our analyses were based on the mean quality score for each article (that is, the average of the two reviewers' scores). If the reviewers' scores differed by more than one standard deviation, the masked article was discussed by both reviewers until consensus was achieved, and the consensus score was used. The interrater reliability between the two reviewers was assessed using Kendall's coefficient of concordance with adjustment for tied ranks (W) and the intraclass correlation coefficient (R). Initial interrater reliability was good (W = 0.85, R = 0.64). Reviewer scores differed by more than one standard deviation for 22 articles (six peer reviewed, 16 symposia). After reviewers had reached consensus on these articles, interrater reliability was high (W = 0.92, R = 0.87).

CLASSIFICATION OF OTHER VARIABLES

Sponsorship

Research sponsorship was determined based on statements in the articles and was classified as either "tobacco industry", "government", "other", or "none acknowledged." Articles were classified as tobacco industry sponsored if a tobacco company, the Tobacco Institute, the Council for Tobacco Research, the Center for Indoor Air Research, the Smokeless Tobacco Research Council, or a combination of these, were the sole sources of funding acknowledged. Articles were classified as government sponsored if one or more government agencies, such as the National Institutes of Health, were the sole source of funding acknowledged. Articles that acknowledged other sources of funding, such as private foundations, or which acknowledged funding from multiple sources (for example, government and private foundations) were classified in the "other" category.

Article conclusion

Article conclusions were coded as "positive" (environmental tobacco smoke harmful) if the article concluded that environmental tobacco

smoke had an adverse effect on one or more of the outcomes measured; "negative" (environmental tobacco smoke not harmful) if the article concluded that environmental tobacco smoke did not have an adverse effect on any of the outcomes measured; or "no conclusion" if the article controlled for environmental tobacco smoke as a confounding variable and did not draw any conclusions about its health effects.

Study design

Our scoring system weights experimental studies higher than observational studies.[13] To control for study design as a potential confounding variable, we classified each article as either *observational* (for example, the effects of environmental tobacco smoke were determined by asking people about their history of exposure) or *experimental* (for example, the effects of environmental tobacco smoke were determined by exposing subjects to it in an experimental chamber).

Article topic

A preliminary examination of the data indicated that peer reviewed and symposium articles tended to study different topics. To evaluate an article topic as a potential confounding variable, the topic of each article was coded as either "lung cancer", "respiratory effects", "other chronic disease", "biochemical effects"(for example, urinary mutagenicity), "sick building syndrome" (for example, headache and eye irritation in the workplace), or "other."

DUPLICATE PUBLICATIONS

As described above in the section on selection of peer reviewed and symposium articles, we eliminated duplicate publications that occurred within the symposium article database (symposium/symposium duplicates) and within the peer reviewed article database (peer/peer duplicates). However, we were interested in comparing the quality of duplicate publications that were published in both a symposium and a peer reviewed journal (peer/symposium duplicates). We identified these duplicate publications by searching our entire database of 386 peer reviewed articles using the last names of the authors from the 68 symposium articles included in the study. Articles that had at least one author with the same last name were compared. Articles were considered to be duplicates if they discussed the same outcome data from the same subject population. Fifteen pairs of peer/symposium duplicates were identified. In four cases, there was only partial overlap between the articles (for example, one article discussed lung cancer, respiratory disease, and heart disease in a group of subjects, whereas the other article only discussed the data related to lung cancer). All of the symposium articles and five of the peer reviewed articles had been included in the main study. The quality of the 10 remaining peer reviewed articles was assessed by mixing them in randomly with the other articles.

ANALYSES

We used χ^2 tests to compare peer reviewed and symposium articles in terms of qualitative characteristics. To examine the relation between our outcome variable, mean quality score, and the various predictor variables, we first conducted univariate analyses using one way analysis of variance (ANOVA). We then conducted multivariate regression analyses in which two models were developed. In the first model (full model), all of the variables were included regardless of whether they were significantly associated with quality scores. Categorical variables were represented using dummy variables. To develop our second, simpler model, we began by including all variables that were significantly associated ($P < 0.05$) with quality scores in the full model. We then added additional variables one at a time, focusing on those that were associated with poor article quality. Variables were retained if they were associated with poor quality scores, or if they changed the association between quality scores and the other variables in the model. Duplicate publications were compared using the paired t test. All tests were conducted using a two tailed $\alpha = 0.05$. Confidence intervals were calculated based on the t distribution.

Assumption checking

Both ANOVA and multivariate regression analyses use the F test, which is based on the assumption that the outcome variable (quality scores) is approximately normally distributed, with constant variance at all values of the independent variables. In our sample, peer reviewed articles had normally distributed scores, but the distribution of symposium articles was slightly skewed (Shapiro-Wilk test, $P = 0.03$). However, the median symposium score (0.375) did not differ greatly from the mean symposium score (0.365). In addition, the standard deviations for peer reviewed articles (0.08) and symposium articles (0.13) did not differ greatly. Furthermore, the F test is reasonably robust when sample sizes are greater than 30 and roughly balanced in each group. A plot of residual values from the full multivariate regression model did not reveal any obvious violations of the assumptions. We therefore believe that our analyses provide valid estimates of the relations between article quality, peer review status, and the other predictor variables examined.

Results

CHARACTERISTICS OF PEER REVIEWED AND SYMPOSIUM ARTICLES

Table 1 shows that peer reviewed journal articles and non-peer-reviewed symposium articles differed from each other in terms of funding sources acknowledged, article conclusion, study design, and topic. Symposium articles were more likely than peer reviewed articles to fail to acknowledge their sources of funding, to conclude that environmental tobacco smoke is not harmful, to use experimental study designs, and to study topics related to sick building syndrome. Peer reviewed articles, on the other hand, were more

Table 1 Characteristics of peer reviewed and symposium articles

	Per cent of articles	
	Peer reviewed articles (n=68)	*Symposium articles (n=68)*
Source of funding[a]		
Tobacco industry	2.9	4.4
Government	27.9	7.4
Other	23.5	14.7
None acknowledged	45.6	73.5
Article conclusion[b]		
ETS harmful	76.5	45.6
ETS not harmful	20.1	41.2
No conclusion	2.9	13.2
Study design[c]		
Experimental	8.8	25.0
Observational	91.2	75.0
Topic[d]		
Lung cancer	10.3	13.2
Respiratory disease	51.5	35.3
Other chronic disease	16.2	1.5
Biochemical effects	10.3	10.3
Sick building syndrome	4.4	27.9
Miscellaneous	7.4	11.8

ETS, environmental tobacco smoke.
[a] χ^2=14.2; df=3; P=0.003.
[b] χ^2=14.4; df=2; P=0.001.
[c] χ^2=6.3; df=1; P=0.012.
[d] χ^2=23.0; df=5; P<0.001.

likely than symposium articles to acknowledge government sponsorship, to conclude that environmental tobacco smoke is harmful, to use observational study designs, and to study respiratory disorders and other chronic diseases. Given the underlying differences between peer reviewed and symposium articles, it is particularly important to conduct multivariate analyses in order to ensure that any differences observed in quality scores are not actually due to these other factors.

UNIVARIATE ANALYSES

In the univariate analyses, article quality was highly associated with peer review status. Peer reviewed journal articles had mean quality scores of 0.45 (95% confidence interval (CI): 0.43 to 0.47), whereas non-peer-reviewed symposium articles had mean quality scores of 0.36 (95% CI: 0.33 to 0.40) ($F = 21.15$, $df = 1$, $P < 0.0001$). The scores of peer reviewed articles ranged from 0.24 to 0.61, while the scores of symposium articles ranged from 0.07 to 0.71

Table 2 presents an item by item analysis of the differences between peer reviewed and symposium articles. Peer reviewed articles scored higher than symposium articles on most of the items in our quality assessment instrument. For example, peer reviewed articles were more likely than symposium articles to have a well described study question, to use an appropriate study design, to specify inclusion and exclusion criteria, to study appropriate subject populations, to report findings completely, and to have conclusions that were consistent with their results. The categories in which peer reviewed and symposium articles did not differ tended to be those in which both groups scored poorly. For example, neither symposium nor peer reviewed articles routinely reported blinding of investigators or subjects. Symposium articles did not receive significantly higher scores than peer reviewed articles for any of the criteria measured.

The univariate analyses also revealed that mean quality scores were associated with funding source acknowledged ($F = 2.87$, $df = 3$, $P = 0.039$) and with article topic ($F = 9.33$, $df = 5$, $P < 0.0001$). Post hoc testing using the Scheffe test suggested that articles that failed to acknowledge their funding sources had poorer scores than those that acknowledged any funding source, although the differences were not statistically significant. In addition, articles related to sick building syndrome had significantly lower scores than articles on all other topics, based on the post hoc Scheffe

Table 2 Item by item analysis of quality scores

Item	Per cent (number) of articles containing item		
	Peer reviewed articles	*Symposium articles*	*Fischer's exact P value*
1 Study question well described	96% (65/68)	75% (50/67)	0.0006*
2 Study design appropriate	84% (57/68)	55% (36/65)	0.0006*
3 Inclusion/exclusion criteria specified	56% (37/66)	15% (9/59)	<0.0001*
4 For case report only: patient characteristics reported	0% (0/0)	100% (1/1)	n/a
5 Subject population appropriate	88% (60/68)	59% (36/61)	0.0002*
6 Control population appropriate	64% (14/22)	28% (7/25)	0.02
7 Subjects randomly selected	14% (9/66)	9% (5/58)	0.41
8 Method of random selection described	0% (0/9)	0% (0/6)	n/a
9 Method of random allocation described	0% (0/0)	0% (0/2)	n/a
10 Investigators blinded	4% (2/53)	7% (3/46)	0.66
11 Subjects blinded	7% (3/45)	3% (1/39)	0.62
12 Measurement bias accounted for	46% (31/68)	27% (17/62)	0.04
13 Confounders accounted for by study design	43% (18/42)	16% (8/50)	0.006
14 Confounders accounted for by analysis	70% (42/60)	50% (29/58)	0.04
15 A priori sample size justification	3% (2/67)	0% (0/61)	0.50
16 Post-hoc power calculations for non-significant results	36% (22/61)	18% (9/51)	0.04
17 Statistical analyses appropriate	83% (54/65)	68% (36/53)	0.08
18 Statistical tests stated	95% (62/65)	80% (40/50)	0.02
19 Exact P values or confidence intervals reported	47% (30/64)	32% (16/50)	0.13
20 Attrition of subjects discussed	67% (6/9)	23% (3/13)	0.08
21 Results completely reported	97% (66/68)	75% (46/61)	0.0004*
22 Findings support conclusions	96% (64/67)	73% (48/66)	0.0003*

The data are presented as the percentage of articles for which one or both reviewers rated the criterion as "present". The denominator in each cell varies because the article was excluded if one or both reviewers felt that the criterion was "not applicable" (n/a) for the article.
*Using a Bonferroni adjustment for multiple comparisons, only $P < 0.0023$ should be considered statistically significant.

Table 3 Multivariate regression analysis: full model

	Coefficient	95% CI	t	P value
Peer review status				
Peer reviewed *v* symposium	0.046	0.005 to 0.086	2.243	0.027
Source of funding				
Tobacco industry *v* other	−0.063	−0.168 to 0.042	−1.182	0.24
Government *v* other	0.016	−0.040 to 0.072	0.581	0.562
None *v* other	−0.029	−0.074 to 0.016	−1.291	0.199
Article conclusion				
ETS harmful *v* no conclusion	0.022	−0.042 to 0.087	0.689	0.49
ETS not harmful *v* no conclusion	0.023	−0.043 to 0.090	0.688	0.49
Study design				
Experimental *v* observational	0.088	0.027 to 0.149	2.856	0.005
Topic				
Lung cancer *v* miscellaneous	0.076	−0.011 to 0.162	1.731	0.086
Respiratory diseases *v* miscellaneous	0.068	−0.005 to 0.141	1.853	0.066
Other chronic diseases *v* miscellaneous	0.086	−0.004 to 0.176	1.897	0.060
Biochemical studies *v* miscellaneous	0.011	−0.069 to 0.090	0.269	0.788
Sick building syndrome *v* miscellaneous	−0.073	−0.152 to 0.005	−1.845	0.067

CI, confidence interval.

test. Article conclusion was marginally associated with quality scores ($F = 2.99$, $df = 2$, $P = 0.054$), but study design was not ($F = 0.75$, $df = 1$, $P = 0.39$).

MULTIVARIATE REGRESSION ANALYSIS

The results of our multivariate analysis using the full model are presented in table 3. The most important predictors of article quality were peer review status ($P = 0.027$) and study design ($P = 0.005$). The coefficients may be interpreted as the change in mean quality score associated with a given variable while controlling for all other variables in the model. Thus when controlling for all other factors measured, peer reviewed articles had mean quality scores that were 0.046 points higher (on a scale of 0 to 1) than symposium articles. Similarly, experimental studies had mean quality scores that were 0.088 points higher than observational studies when controlling for all other variables. Neither funding source nor article conclusion was significantly associated with article quality in the multivariate model. Roughly 37% of the variability in quality scores was explained when all of the variables were included ($R^2 = 0.372$).

The results of the final, simpler model are given in table 4. We included both peer review status and study design, because they were significant in the full model. In addition, we obtained the most significant effects when the funding variable was dichotomised as any funding acknowledged versus no funding acknowledged and the topic variable was dichotomised as sick building syndrome versus all other topics. In the simpler model, the primary predictors of article quality were peer review status ($P = 0.005$) and sick building syndrome ($P < 0.0001$). Roughly 33% of the variability in quality scores was explained using the simpler model ($R^2 = 0.33$).

DUPLICATE PUBLICATIONS

Our analysis of duplicate publications revealed that the articles published in peer reviewed journals had a mean quality score of 0.49 (95% CI: 0.44 to 0.54), whereas the same studies published in symposia had a mean quality score of 0.43 (95% CI: 0.40 to 0.46) (paired *t* test, $P = 0.01$). Quality scores were higher for peer reviewed articles than for symposium articles in 13 of the 15 pairs.

Discussion

Our findings confirm our primary hypothesis that articles published in symposium proceedings are associated with poor quality. We found that symposium articles had significantly poorer mean quality scores than peer reviewed articles even after controlling for the effects of funding source acknowledged, article conclusion regarding the health effects of environmental tobacco smoke, type of study design used, and article topic.

The criterion by criterion analysis showed that peer reviewed articles were superior to symposium articles in terms of study design, reporting, and interpretation. For example, peer reviewed articles were more likely to use appropriate study designs and subject populations; to report their study objectives, inclusion/exclusion process, and results

Table 4 Multivariate regression analyses: final model

	Coefficient	95% CI	t	P value
Peer review status				
Peer reviewed *v* symposium	0.053	0.016 to 0.089	2.868	0.005
Source of funding				
None acknowledged *v* any acknowledged	−0.032	−0.067 to 0.003	−1.824	0.070
Study design				
Experimental *v* observational	0.042	−0.002 to 0.087	1.871	0.064
Topic				
Sick building syndrome *v* all other topics	−0.127	−0.174 to −0.081	−5.391	<0.0001

CI, confidence interval.

adequately; and to draw conclusions that were consistent with their findings.

Our two other a priori hypotheses were not confirmed. We had originally proposed that tobacco industry sponsorship would be associated with poor quality. However, because only five articles in our study (two peer reviewed, three symposium) acknowledged tobacco industry sponsorship, we did not have enough statistical power to evaluate this hypothesis. Interestingly, our findings suggest that the failure to acknowledge funding sources is marginally associated with poor quality. It is possible that articles that fail to acknowledge funding sources may tend to have poor reporting in general, and this may explain why their quality scores tend to be lower. This finding should be investigated in future studies.

Our third a priori hypothesis was that articles with negative conclusions (environmental tobacco smoke is not harmful to health) would be associated with poor quality. This hypothesis was not confirmed. Article conclusion was not associated with article quality when controlling for factors such as peer review. We have previously found that peer reviewed articles are more likely than symposium articles to conclude that environmental tobacco smoke is harmful.[32] The tobacco industry has suggested that this may be due to a bias in the peer reviewed journals against publishing negative studies on environmental tobacco smoke.[33] Our findings suggest, however, that negative studies on environmental tobacco smoke are published in the peer reviewed journals when they are of high quality.

Our findings also suggest that articles related to sick building syndrome may be of poorer quality than articles on other topics. We did not establish this hypothesis a priori, and this finding should be further evaluated by other studies.

Our analysis of duplicate publications showed that peer reviewed articles are higher in quality even when they discuss the same data as symposium articles. Because these pairs of duplicates, by definition, had the same study design characteristics, the differences observed between peer reviewed and symposium articles were most likely to have been due to factors such as poor reporting and poor interpretation of findings.

The findings from this study complement our previous work on the content of symposia on environmental tobacco smoke. We have previously found that symposium articles on environmental tobacco smoke are associated with a lack of balance and that they tend to support the tobacco industry position that environmental tobacco smoke is not harmful.[2] Our findings reported in this article suggest that symposium articles on environmental tobacco smoke are also associated with poor quality. Taken together, these findings suggest that symposium articles are not reliable sources of information about environmental tobacco smoke.

Our findings are also consistent with previous research on the quality of symposium articles in the pharmaceutical literature. For example, Rochon[3] has found that randomised controlled drug trials are of poorer quality when published in journal supplements than in parent journals. On the other hand, Cho and Bero[13] did not find a difference in the quality of articles on drugs published in symposia and peer reviewed journals; however, they noted that their power to detect a difference when controlling for the effect of study design was low.

LIMITATIONS

There are several limitations to our findings. First, as mentioned above, we did not have enough statistical power to assess the relation between funding source and article quality. Therefore our finding that article quality was not associated with sponsorship should be interpreted cautiously.

Another potential source of criticism lies in our selection of peer reviewed and symposium articles. Our symposium articles were drawn from 22 symposia while our peer reviewed articles were drawn from 41 journals. If a single journal or symposium contained extremely high or low quality articles, then it could theoretically have shifted the mean quality score of the entire group. We do not believe that this occurred because no single journal or symposium dominated our samples. The maximum number of articles per symposium was 11 (16%), while the maximum number of articles per peer reviewed journal was six (9%). The quality of articles from these sources did not differ from the overall quality of articles in the groups.

Another potential source of selection bias is that the peer reviewed articles in our study were identified exclusively through Medline, whereas the symposium articles were identified through a variety of electronic databases. We used Medline to identify peer reviewed journal articles on environmental tobacco smoke because it is the database most commonly used in the United States to gather information about health related research. However, several peer reviewed journals that publish articles related to environmental tobacco smoke, including *Tobacco Control*, were not indexed by Medline during the period of our study.[34 35] Future studies on the quality of the scientific literature should therefore consider drawing their samples from multiple sources, such as searching a variety of electronic databases, hand searching journals, and checking references.

Another potential criticism of our study is that the difference in mean quality scores observed between peer reviewed and symposium articles, although statistically significant, was not particularly large. We feel that table 2 provides qualitative insight into the quantitative differences between the groups. This table shows that peer reviewed articles are more likely than symposium articles to be well designed, to report results completely and accurate, and to draw appropriate conclusions. When these qualitative and quantitative findings are taken together, they provide strong

evidence that peer reviewed articles tend to be of higher quality than symposium articles.

Finally, it is not clear whether our findings may be generalised to symposia on other topics. Our study primarily contained articles from symposia on environmental tobacco smoke and symposia on indoor air, and we found that article quality scores were similarly poor in both groups (data not shown). Other studies have found that the quality of symposium articles in the pharmaceutical literature tends to be poor.[3] Taken together, these findings suggest that symposia in general may be poor in quality. However, this should be confirmed by future studies.

CONCLUSIONS

In summary, our study provides strong evidence that peer reviewed journal articles on the health effects of environmental tobacco smoke are superior to symposium articles. These findings support the decision of the US Supreme Court[36] that expert scientific testimony should be "scientifically valid" and that "[a] pertinent consideration is whether the theory or technique has been subjected to peer review and publication."

However, our findings also suggest that peer review is not a guarantee of high quality. The peer reviewed articles in our sample had a mean quality score of 0.45 on a scale of 0 to 1, suggesting that there is a great deal of room for improvement. Because decisions should be based on the highest quality evidence available, we propose that researchers, policy makers, and judges should evaluate the quality of *all* scientific research before using the findings in research, policy, or legal settings. A lack of peer review should be seen as a flag that an article has a higher probability of being poor in quality.

On the basis of our findings, we believe that journal and book editors should develop standards for labelling scientific material. It is often difficult to determine whether an article has been published in a peer reviewed journal or a non-peer-reviewed symposium because this information is not printed directly on the articles. We therefore recommend that the first page of all scientific research articles should be clearly labelled to indicate whether the article was peer reviewed and whether it is being published as part of a symposium. This will allow policy makers, researchers, the press, and the public to identify those articles that are most likely to be associated with high quality.

This study was funded by the Robert Wood Johnson Foundation (award No 024783), and by the Cigarette and Tobacco Surtax Fund of the State of California, through the tobacco related disease research program of the University of California (award No 4RT0035).

We wish to thank our quality assessors, Valerie Hoffman PhD, MPH and Judith Moskowitz PhD, MPH; the Writing Seminar at the Institute for Health Policy Studies, which provided invaluable insight during the preparation of this manuscript; Phillip Lollar for administrative assistance; and Steve Selvin PhD for statistical advice.

The results of this study were presented in part at the 1995 meeting of the American Public Health Association in San Diego, California.

1 Bero LA, Galbraith A, Rennie D. The publication of sponsored symposiums in medical journals. *N Engl J Med* 1992;327:1135–40.
2 Bero LA, Galbraith A, Rennie D. Sponsored symposia on environmental tobacco smoke. *JAMA* 1994;271:612–7.
3 Rochon P. Evaluating the quality of articles published in journal supplements compared with the quality of those published in the parent journal. *JAMA* 1994;272:108–13.
4 Ecobichon DJ, Wu JM. *Environmental tobacco smoke: proceedings of the International Symposium at McGill University.* Lexington, Massachusetts: Lexington Books, 1990.
5 Symposium proceedings. Assessing low-risk agents for lung cancer: methodological aspects. (International Symposium held in Clearwater, FL, USA, October 10–14, 1989.) *Int J Epidemiol.* 1990;19 (suppl 1).
6 Workshop proceedings. ETS—Environmental tobacco smoke. (Workshop on effects and exposure levels, held at the University of Geneva, Switzerland, March 15–17, 1983.) *Eur J Respir Dis* 1984;65(suppl 133).
7 Workshop proceedings. Environmental tobacco smoke effects of the non-smoker. (Workshop, Bermuda, 1974.) *Scand J Respir Dis* 1974; suppl 91.
8 Perry R, Kirk P. *Indoor and ambient air quality.* London: Selper, 1988.
9 Symposium proceedings. International Symposium on Medical Perspectives on Passive Smoking. *Prev Med* 1984; 13.
10 Bero LA, Glantz SG. Tobacco industry response to a risk assessment of environmental tobacco smoke. *Tobacco Control* 1993;2:103–13.
11 R J Reynolds Tobacco Company advertisement. "Passive smoking: an active controversy." [Available from authors upon request.]
12 Everingham R, Woodward S. *Tobacco litigation.* Sydney: Legal Books, 1991.
13 Cho MK, Bero LA. The quality of drug studies published in symposium proceedings. *Ann Intern Med* 1996;124: 485–9.
14 Barnes DE, Bero LA. Industry-funded research and conflict of interest: an analysis of research sponsored by the tobacco industry through the Center for Indoor Air Research. *J Health Politics Policy Law* 1996;21:515–42.
15 Rochon PA, Gurwitz JH, Simms RW, *et al.* A study of manufacturer-supported trials of nonsteroidal anti-inflammatory drugs in the treatment of arthritis. *Arch Intern Med* 1994;154:157–63.
16 Swaen G, Meijers J. Influence of design characteristics on the outcome of retrospective cohort studies. *Br J Indust Med* 1988;45:624–9.
17 Davidson R. Source of funding and outcome of clinical trials. *J Gen Intern Med* 1986;1:155–8.
18 Dickersin K. The existence of publication bias and risk factors for its occurrence. *JAMA* 1990;263:1385–9.
19 Dickersin K, Min Y-I, Meinert CL. Factors influencing publication of research results: Follow-up of applications submitted to two institutional review boards. *JAMA* 1992; 267:374–8.
20 Weber A. Acute effects of environmental tobacco smoke. *Eur J Respir Dis* 1984;65:98–108.
21 Weber A. Annoyance and irritation by passive smoking. *Prev Med* 1984;13:618–25.
22 Weber A. Irritating and annoying effects of passive smoking. *Tokai J Exp Clin Med* 1985;10:341–5.
23 Weber A. Environmental tobacco smoke exposure: acute effects—acceptance level—protective measures. In: Berglund B, Lindvall T, Sundell J, eds. *Indoor air.* Volume 2. *Radon, passive smoking, particulate and housing epidemiology.* Stockholm: Lyber Tryck AB, 1985: 297–301.
24 Hemminki E. Quality of clinical trials—a concern for three decades. *Meth Inform Med* 1982;21:81–5.
25 Andrew E, Eide H, Fuglerud P, *et al.* Publications on clinical trials with X-ray contrast media: differences in quality between journals and decades. *Eur J Radiol* 1990;10: 92–7.
26 Sonis J, Joines J. The quality of clinical trials published in *The Journal of Family Practice,* 1974-1991. *J Fam Pract* 1994;39:225–35.
27 Hulley SB, Cummings SR. *Designing clinical research.* Baltimore, Maryland: Williams & Wilkins, 1988.
28 Cho MK, Bero LA. Instrument for assessing the quality of drug studies published in the medical literature. *JAMA* 1994;272:101–4.
29 McNutt RA, Evans AT, Fletcher RH, Fletcher SW. The effects of blinding on the quality of peer review. *JAMA* 1990;263:1371–6.
30 Ernst E, Resch K-L. Reviewer bias: a blinded experimental study. *J Lab Clin Med* 1994;124:178–82.
31 Moher D, Jadad AR, Nichol G, Penman M, Tugwell P, Walsh S. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. *Controlled Clin Trials* 1995;16:62–73.
32 Bero LA, Glantz SA, Rennie D. Publication bias and public health policy on environmental tobacco smoke. *JAMA* 1994;272:133–6.
33 Flue-Cured Tobacco Cooperative Stabilization Corporation and The Council for Burley Tobacco, Inc and Universal Leaf Tobacco Company, Inc and Philip Morris Incorporated and R J Reynolds Tobacco Company and Gallins Vending Company *vs* United States Environmental Protection Agency and Carol Browner, 1993.
34 Liguori A, Hughes JR. Where is smoking research published? *Tobacco Control* 1996;5:37–8.
35 Davis RM. *Tobacco Control* joins *Index Medicus* and Medline. *Tobacco Control* 1996;5:99.
36 Supreme Court of the United States. *William Daubert v Merrell Dow Pharmaceuticals, Inc* (No 92–102). June 28, 1993.