Taking Stock
A Bibliometric Analysis of the Focus of Tobacco Research from the 1980s to the 2000s

Joanna E. Cohen, PhD, Michael O. Chaiton, PhD, Lynn C. Planinac, MHSc

Background: Little is known about the body of tobacco research as a whole.

Purpose: This paper examines the changes in literature focus (1980s to 2000s) and identifies areas in need of increased attention.

Methods: Tobacco articles randomly selected from searches of the MEDLINE and Web of Science databases were coded according to (1) epidemiologic framework component; (2) study focus; and (3) form of tobacco. Frequencies, cross-tabulations, and tests of proportions were conducted. The analysis was conducted in 2009.

Results: From the 1980s to the 2000s, there was a significant decrease in tobacco-related articles focusing on the "agent" and an increase in articles focusing on the "host." Few articles in either decade focused on the "environment" or on the "vector" (<10%). The percentage of study foci addressing health effects decreased, whereas prevalence/use and cessation foci increased. Approximately two thirds of articles focused on the cigarette.

Conclusions: The nature of tobacco research has shifted from examining the links between cigarettes and disease to understanding why people smoke and how to help them quit. Proportionately more research could focus on the environment and vector components of the epidemiologic framework, to expand strategies for reducing tobacco-related disease.
(Am J Prev Med 2010;39(4):352–356)

Introduction

Tobacco research is a mature field of study, with at least three academic journals now dedicated to this topic: *Tobacco Control*, *Nicotine & Tobacco Research*, *Tobacco-Induced Diseases*. Although the tobacco literature is replete with reviews, there have been few studies examining the focus of published tobacco research as a whole.

Bibliometrics is a research method used in library and information science that often uses quantitative analysis to describe publications within a given field or body of literature. Five studies were identified that have used bibliometric analysis to examine components of the tobacco literature. Two of these studies focused exclusively on research originating from Spain: (1) Garcia-Lopez[1] reported on collaboration among authors, research centers, and institutions; productivity by Spanish communities; journals where the research was published; and the number of authors per paper; and (2) de Granda-Orive et al.[2] described international collaborations, number of citations, and journals where the research was published. Qui and Chen[3] described nicotine research originating from China, reporting on citations, impact factors, international and institutional collaboration, and subject disciplines within the broader area of nicotine research. de Granda-Orive and colleagues[4] examined scientific collaboration in the published literature on smoking over a 5-year period (1999 to 2003). They found that the United Kingdom published the highest number of articles with inter-institutional collaboration, followed by the U.S. and Germany, whereas the U.S. published the highest number of articles with international collaboration, followed by the United Kingdom and France. Articles resulting from collaborations received a higher number of citations than those with no collaborations. Finally, Kusma et al.[5] analyzed research self-identified as "tobacco control"; they visually describe country and institutional collaboration, the number of subject disciplines re-

From the Ontario Tobacco Research Unit (Cohen, Chaiton, Planinac); Dalla Lana School of Public Health, University of Toronto (Cohen, Chaiton); and Centre for Addiction and Mental Health (Cohen), Toronto, Ontario, Canada

Address correspondence to: Joanna E. Cohen, PhD, Director of Research and Training, Ontario Tobacco Research Unit, 33 Russell Street T5, Toronto, Ontario, Canada M5S 2S1. E-mail: joanna_cohen@camh.net.

0749-3797/$17.00
doi: 10.1016/j.amepre.2010.06.003

ported in the database, and the journals where this research was published.

Although these studies provide some information about the body of tobacco research, and especially about collaboration, very little is known about what this body of research is actually studying. By understanding how the areas of study have changed over time, and which areas of study currently receive more attention and which receive less, one might identify areas that are in need of further investigation.

Moreover, although not without limitations, the epidemiologic framework is a useful framework for gaining comprehensive understanding of disease and its prevention. Traditionally, the framework has been composed of agent, host, and environment; the addition of vector has sometimes been described as the "public health model."[6,7] Slade[8] is one of the first to use these terms in discussing the prevention and control of cigarette use.

According to this conceptualization of agent, host, environment and vector, a change in any of the four components will alter the existing equilibrium to increase or decrease the frequency of disease. The implication is that all components of this epidemiologic framework need to be understood and addressed if efforts to reduce death and disease are to be maximally effective.

The purpose of this study is to take stock of the recent body of tobacco research by comparing its focus from the most recent decade to that from the 1980s. In addition to examining changes in the nature of the tobacco literature over time, areas that might be in need of increased focus are also identified. It was hypothesized that research on the vector was least represented in the tobacco literature and that the proportion of the tobacco literature focused on health effects would show a decline over time.

Methods

Article Selection

The MEDLINE and Web of Science databases were searched from 1960 to the end of 2008 to determine the number of tobacco articles published during these 5 decades. The following terms were used to search within the title field: smok* or cigar* or nicotine or tobacco or narghile or shisha or hookah or snus or snuff. Articles were limited to those published in English. Results from the searches of the two databases were combined and duplicates were removed. A detailed analysis was conducted for the 1980s (1980–1989) and the 2000s (2000–2008). For these years, a 10% random sample was selected for coding (Figure 1). Articles were excluded that clearly did not refer to tobacco (e.g., smoke inhalation among fire victims) as well as those that focused exclusively on the tobacco plant (e.g., transgenic tobacco). Based on these criteria, the number of articles eligible for coding was 2,050 for the 2000s and 780 for the 1980s. Additionally, a search was conducted in MEDLINE to determine the total number of articles *on all topics* published from 1960 to 2008, to compare with the trend of number of tobacco articles for these decades. The search for total number of published articles was conducted in the "search fields" option of MEDLINE by entering the applicable years as search terms under "publication year."

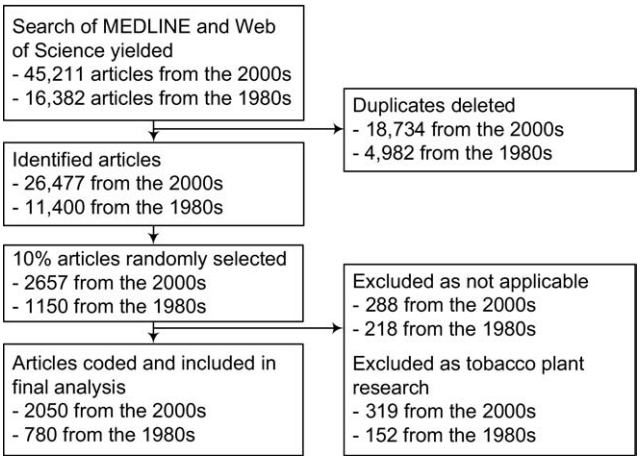

Figure 1. Literature search results

Coding

Articles were coded based on their title and, if necessary, their abstract. Each article was coded according to three dimensions:

1. Epidemiologic framework component: agent (e.g., the cigarette); host (e.g., smoking behaviors, characteristics of smokers, treatment for smoking); environment (e.g., regulations, community interventions); and vector (e.g., the tobacco industry). For example, genetics and health effects studies were coded as "agent" because it is the agent that causes the different phenotypes and health effects; preclinical studies were coded as "agent" because those studies examined the effects of the tobacco/nicotine product. Traditional epidemiologic studies on the characteristics of smokers were coded as "host."
2. Study focus: protection (from secondhand smoke); prevention; cessation; prevalence/use; preclinical (i.e., studies on animals or in vitro); health effects; harm reduction; chemistry (studies looking at the chemical composition of the product); economics; methodology; and tobacco control. "Focus" was not necessarily a subset of the epidemiologic framework classification; a particular study focus could fall under different components depending on the intervention being studied. Categorizing articles by "study focus" provided an additional approach to understanding the subject being examined.
3. Form of tobacco (including pharmaceuticals): cigarette; nicotine; secondhand smoke; pharmaceuticals; smokeless tobacco; cigar; pipe; waterpipe; and tobacco (generic). Nicotine replacement therapy was classified as a pharmaceutical.

The coding scheme was developed by the three authors with a goal of capturing the key focus of each article while keeping the number of categories as small as possible. The coding scheme was refined during the coder training sessions (see below).

The random sample of articles was distributed across three coders. The coders underwent two training sessions and participated in three rounds of practice coding, discussing any discrepancies and fine-tuning the coding definitions.

A number of articles were deemed as having legitimate double coding (i.e., discussed both smokeless tobacco use and cigarette use). Twenty-nine articles had double epidemiologic framework coding, 37 had double study focus coding, and 45 had double tobacco coding. To resolve this issue, the most unique aspect of the study was coded (i.e., articles addressing cigarette and smokeless coded as smokeless).

A total of 6% (n=175) of the total article sample was randomly selected for the assessment of inter-rater reliability. Inter-rater reliability was assessed by comparing two raters' classifications for each of the individual categories that were coded (epidemiologic framework, study focus, and form of tobacco), as well as for overall classifications.

Data Analysis

Frequencies and cross-tabulations were performed using SAS version 9.1.3. Differences between proportions were conducted by performing a standard two-tailed, two-sample test of proportion at a significance level of 0.05. Kappa statistics were calculated to assess inter-rater reliability. The analysis was conducted in 2009.

Results

The number of tobacco-related articles has increased substantially from the 1960s, when there were 3752 relevant articles published compared to 26,477 (and counting) for the 2000s, representing more than a 600% increase. The total number of articles published on all topics during this time frame also increased substantially from 1,533,069 (1960s) to 5,801,739 (2000s), representing a 278% increase.

Inter-rater reliability coding was very good. The kappa statistic was 0.81 for the epidemiologic framework coding, 0.79 for the study focus coding, and 0.81 for the tobacco form coding. The overall kappa of all classifications between two raters was 0.83.

From the 1980s to the 2000s, there was a significant reduction in the proportion of papers focusing on the agent ($p<0.001$) and significant increases in the proportion of papers focusing on the host ($p<0.001$) and the environment ($p<0.001$). More than half of the tobacco-related articles in the 1980s (71%) and in the 2000s (54%) dealt with the agent (focusing on the product itself, including design, components including constituents of the product and of the smoke, health effects, studies dealing with genetic polymorphisms, preclinical studies; Table 1). The host (prevalence, individual-level risk factors for smoking and cessation, individual treatment, efficacy/effectiveness of a pharmacologic agent for smoking prevention/cessation) was the second most frequent focus for tobacco articles in the 1980s (22%) and the 2000s (34%). Fewer than 10% of articles (5% in 1980s, 9% in 2000s) focused on the environment (tax increases, smoking bylaws, advertising/promotion bans, community-level risk factors for smoking and cessation, population-based tobacco control programs and strategies, schools and worksites, comparison of use across countries/jurisdictions/schools, societal costs). The vector (focusing on the tobacco industry, such as strategies used to maximize sales, business issues including supply chain from farmer to retailer, tobacco in movies) represented 2% of articles in the 1980s and 3% of articles in the 2000s.

Table 1. Tobacco articles[a] according to epidemiologic framework concept, study focus, and form of tobacco, 1980s and 2000s, % (n)

Article characteristic	1980s[b]	2000s[b]
Epidemiologic framework concept		
Agent	71 (557)	54 (1110)
Host	22 (169)	34 (698)
Environment	5 (39)	9 (191)
Vector	2 (15)	3 (51)
Study focus		
Health effects	45 (351)	37 (751)
Prevalence/use	14 (109)	21 (432)
Cessation	6 (48)	14 (283)
Preclinical	19 (146)	13 (260)
Methodology	4 (29)	4 (78)
Chemistry	6 (43)	3 (61)
Tobacco control	2 (14)	3 (55)
Protection	2 (13)	2 (43)
Economics	1 (11)	2 (40)
Prevention	1 (9)	1 (29)
Harm reduction	1 (7)	1 (18)
Form of tobacco		
Cigarette	68 (530)	63 (1301)
Nicotine	13 (99)	11 (230)
Secondhand smoke	6 (44)	6 (131)
Pharmaceuticals	1 (8)	5 (102)
Smokeless	2 (19)	2 (35)
Other[c]	10 (80)	12 (251)
Total n	780	2050

[a]Non-empirical articles such as editorials and letters were excluded as part of the search strategy.
[b]Percentages do not add up to 100% as a result of rounding.
[c]"Other" includes cigar, pipe, waterpipe/narghile/hookah/shisha, and tobacco (generic, all kinds, plant components, tobacco company/industry)

From the 1980s to the 2000s there was a significant reduction in the proportion of papers examining health effects ($p<0.001$); preclinical topics ($p<0.001$); and chemistry ($p=0.002$), while the proportion of papers examining prevalence/use and cessation increased ($p<0.001$). More than one third of articles dealt with health effects (45% in the 1980s and 37% in the 2000s; Table 1). The other major study foci in the 2000s were prevalence/use (21%; includes research on smoking behaviours); cessation (14%); and preclinical (13%).

From the 1980s to the 2000s, there was a significant decrease in the proportion of papers studying the cigarette ($p<0.029$), and a significant increase in the proportion of papers studying pharmaceutical products ($p<0.001$). Approximately two thirds of tobacco-related articles focused on the cigarette (68% in the 1980s and 63% in the 2000s; Table 1). Nicotine was the next most frequent form of tobacco studied (13% in the 1980s and 11% in the 2000s). Secondhand smoke, pharmaceuticals, smokeless tobacco, cigars, pipes, and waterpipe each represented less than 10% of the tobacco-related articles in both the 1980s and the 2000s.

Discussion

This is the most extensive description of the tobacco literature to date. As expected, research on the vector was least represented in the tobacco literature, and there was a decline over time in the proportion of tobacco literature focused on health effects. The decline in the proportion of studies examining health effects is expected as our understanding of the harms caused by tobacco products has solidified.

The paucity of research on the environment and vector is concerning given the international consensus on types of strategies needed to reduce tobacco-caused disease and death. The WHO's "MPOWER" vision[9] to reverse the global tobacco epidemic calls for a focus on interventions that address the environment and the vector (e.g., smoke-free environments, stringent limits on tobacco companies' ability to advertise and promote their products, population-based counter-tobacco mass media campaigns, and taxing tobacco products). Similarly, three of Frieden and Bloomberg's[10] five approaches to prevent 100 million deaths from tobacco focus on the environment and the vector. The recently proposed Health Impact Pyramid[11] postulates that interventions that change the environmental context have a large potential health impact, second only to addressing socioeconomic factors (such as reduction in poverty and improvement in education). Research needs to be conducted to expand the evidence base for the development and evaluation of interventions aimed at the environment and vector.

Our findings regarding study focus are consistent with other literature. In their analysis of the public health literature on infectious diseases, Durando et al.[12] found that research on "epidemiology and surveillance" (study focus of prevalence/use in this analysis) was better represented than research on "prevention and control" (study focus of tobacco control, protection, and prevention in this analysis). Hughes and Oliveto[13] found that treatment research in drug and alcohol disorders grew rapidly over a 20-year period—similar to the finding of a doubling of research with a study focus of cessation in this analysis.

It is not surprising that the majority of the tobacco literature studied cigarettes, as this is the form of tobacco most commonly used in countries that make the largest contribution to the research enterprise. The significant increase in the proportion of papers studying pharmaceutical products is expected, given that many of these products were developed and introduced during the period under study; despite this important increase, the overall proportion of papers studying pharmaceutical products should be increased beyond the 5% found in the 2000s. We expect that research on waterpipes will increase over time with continued investment in research in low- and middle-income countries.

This study does have some limitations. First, articles written in languages other than English were excluded. It is possible that the distribution of papers across the epidemiologic framework concepts differ in English versus non-English papers. However, the vast majority of tobacco articles indexed by MEDLINE and Web of Science were published in English, particularly in the most recent decade (89% published in English in 2000s; 75% in 1980s). Second, choices did have to be made about what would be included in each of the coding categories. One key choice made was to code papers assessing the health effects of tobacco products as "agent" rather than as "host" (because the cigarette, or agent, itself causes the health effects). If these health effects papers had been coded as "host," a dramatically different picture of the epidemiologic framework would have emerged (particularly because 39% of the sample of papers focused on health effects), although the trend toward increased "host" papers and the finding regarding a paucity of papers dealing with the environment and the vector would still stand. Third, it is important to note that just because a smaller proportion of papers deal with one epidemiologic framework apex versus another, this does not necessarily mean that the neglected areas of research are poorly developed; it could be that those papers have more direct implications for strategies to address tobacco use.

Citation analyses could be conducted to assess the use of these literatures.[14–16] Fourth, some relevant literature on nicotinic receptors may have been missed because the trunk nicotin* was not used. If this literature was indeed underrepresented (which would have been coded as agent), then it further reinforces the conclusion that research on the environment and the vector are underrepresented in the tobacco literature.

Fifth, source of funding could not be examined given the limited data available in the MEDLINE and Web of Science databases. It is possible that source of funding is associated with study focus, and this could be examined in future studies. In addition to source of funding, amount of funding is also important to assess, as it may portray an alternate representation of the focus of the literature beyond number of articles. Finally, publication bias, and the time lag between obtaining study results and publication, would likely be similar across the epidemiologic framework categories.

Conclusion

The field of tobacco research is vast, prolific, and increasing more rapidly than all scientific literature combined; however, research to date has focussed predominantly on tobacco products themselves and the people who use them. Proportionately more research could focus on the environment and vector apexes of the epidemiologic framework, to expand strategies for reducing tobacco-related disease.

We thank Ms. Diane van Abbe, Information Specialist at the Ontario Tobacco Research Unit, who conducted the literature searches.

This work was supported by the Ontario Tobacco Research Unit, which receives funding from the Ontario Ministry of Health Promotion.

No financial disclosures were reported by the authors of this paper.

References

1. Garcia-Lopez JA. Bibliometric analysis of Spanish scientific publications on tobacco use during the period 1970–1996. Eur J Epidemiol 1999;15:23–8.
2. de Granda-Orive JI, Garcia Rio F, Aleixandre Benavent R, et al. [Spanish productivity in smoking research relative to world and European union productivity from 1999 through 2003, analyzed with the science citation index.] Arch Bronconeumol 2007;43:212–8.
3. Qui H, Chen Y. Bibliometric analysis of nicotine research in China during the period 1991 to 2007. Inhal Toxicol 2009;21:965–71.
4. de Granda-Orive JI, Villanueva Serrano S, Aleixandre Benavent R, et al. [Network of international scientific collaboration on smoking: analysis of coauthorship through the Science Citation Index.] Gac Sanit 2009;23:222.e34–43.
5. Kusma B, Scutaru C, Quarcoo D, Welte T, Fischer TC, Gronebergh-Kloft B. Tobacco control: visualization of research activity using density-equalizing mapping and scientometric benchmarking procedures. Int J Environ Res Public Health 2009;6:1856–69.
6. Orleans CT, Slade J. Nicotine addiction: principles and management. New York: Oxford University Press, 1993:ix.
7. Giovino G. Epidemiology of tobacco use in the U.S. Oncogene 2002; 21:7326–40.
8. Slade J. A disease model of cigarette use. NY State J Med 1985;85:294–7.
9. WHO. MPOWER: a policy package to reverse the tobacco epidemic. Switzerland, 2008.
10. Frieden TR, Bloomberg MR. How to prevent 100 million deaths from tobacco. Lancet 2007;369:1758–61.
11. Frieden TR. A framework for public health action: the health impact pyramid. Am J Public Health 2010;100:590–5.
12. Durando P, Stichhi L, Sasso L, Gasparini R. Public health research literature on infectious diseases: coverage and gaps in Europe. Eur J Public Health 2007;17(S 1):19–23.
13. Hughes JR, Oliveto A. The growth of treatment research in drug and alcohol use disorders. Drug Alcohol Depend 1990;26:81–4.
14. Small HG. A co-citation model of a scientific specialty: a longitudinal study of collagen research. Soc Stud Sci 1977;7:139–66.
15. Chen C. CiteSpace II: detecting and visualizing emerging trends and transient patterns in scientific literature. J Am Soc Inf Sci Technol 2006;57:359–77.
16. Howard MO, Howard DA. Citation analysis of 541 articles published in drug and alcohol journals: 1984–1988. J Stud Alcohol 1992;53:427–34.