REVIEW

Tobacco document research reporting

S M Carter

Tobacco Control 2005;**14**:368–376. doi: 10.1136/tc.2004.010132

Objective: To understand the use of internal tobacco industry documents in the peer reviewed health literature.
Design: Interpretive analysis of published research.
Sample: 173 papers indexed in Medline between 1995 and 2004 that cited tobacco industry documents.
Analysis: Information about year published, journal and author, and a set of codes relating to methods reporting, were managed in N*Vivo. This coding formed the basis of an interpretation of tobacco document research reporting.
Results: Two types of papers were identified. The first used tobacco documents as the primary data source (A-papers). The second was dedicated to another purpose but cited a small number of documents (B-papers). In B-papers documents were used either to provide a specific example or to support an expansive contention. A-papers contained information about purpose, sources, searching, analysis, and limitations that differed by author and journal and over time. A-papers had no clear methodological context, but used words from three major traditions—interpretive research, positivist research, and history—to describe analysis.
Interpretation: A descriptive mainstream form of tobacco document reporting is proposed, initially typical but decreasing, and a continuum of positioning of the researcher, from conduit to constructor. Reporting practices, particularly from experienced researchers, appeared to evolve towards researcher as constructor, with later papers showing more complex purposes, diverse sources, and detail of searching and analysis. Tobacco document research could learn from existing research traditions: a model for planning and evaluating tobacco document research is presented.

Correspondence to: Stacy M Carter, Room 129A Building A27, The University of Sydney, NSW 2006, Australia; carters@health.usyd.edu.au

Received 24 September 2004
Accepted 26 May 2005

This paper reflects on a decade of internal tobacco industry document research. Tobacco industry documents were released in two waves.[1] In 1994, whistleblowers and US Congress released 1384 documents from the British American Tobacco (BAT) group. The second wave commenced with the now-famous Minnesota lawsuit against US tobacco corporations.[1 2] The Master Settlement Agreement released tens of millions of pages from this second wave via the internet, excluding documents from the UK based BAT companies, continuing until 2010.[1–4]

A small number of papers have considered tobacco document research (TDR) method—that is, the specific procedures used to gather and analyse tobacco documents. They note that documents can be found by time consuming searching of online or physical archives: one archive is in Minnesota, USA, and the other in Guildford, UK.[4–12] Online industry sites enable wildcard and ''all fields'' searching; physical archives contain unique collections and provide indexing lists, but access is resource intensive; industry indexing is inconsistent.[2 4 9 12 13] Searching the Minnesota Depository versus industry websites produces comparable results.[13] The BAT depository in Guildford, UK, has been particularly inaccessible.[9–11] Electronic archiving projects in tobacco control are building more stable, accessible and comprehensively indexed alternatives to industry sources.[2 4 5 7 8 12 13]

Few papers have discussed TDR methodology—that is, why research is done a certain way, or the principles that determine how research procedures are used and interpreted. These most commonly observe that publicly available documents are not representative of the total document population.[3 4] This results from the documents' provenance—that is, whistle blowing and legal discovery—and from limitations imposed by the tobacco industry. Tobacco lawyers have destroyed documents, a subset of privileged documents is withheld, and, importantly, no business keeps every document it produces. It has also been observed that the relationship between industry intentions and actions is often unclear, and that TDR papers rarely discuss analysis, sometimes analyse documents out of context, and tend not to use interpretive social science methods.[2 3]

This paper examines use of the two waves of documents in the peer reviewed health literature indexed in Medline. It was beyond the scope of this work to include research based on other industry sources, or in other literatures such as legal journals, books, or journalism. I sought to answer several questions. First: What reporting traditions have been established in TDR published in the peer reviewed health literature? Then: What can the reporting traditions of TDR tell us about the nature of TDR? How do they relate to other research reporting traditions? What directions do they suggest for the future?

METHODS

To answer the first question I searched Medline, most recently on 1 February 2005, from 1995[3] to December 2004. I used the MESH term ''Tobacco Industry'', introduced in 1997, to search years 1997 to 2004, and combined the MESH terms ''industry'' and ''tobacco'' with the ''AND'' operator for 1995 and 1996, limiting the search to English language papers. I kept reports, reviews, and research papers that cited at least one industry document, excluding other formats, such as news, and secondary citations—for example, to newspaper articles reporting on documents. The original Medline search produced 1766 papers; 173 met the selection criteria.

Abbreviations: AJPH, *American Journal of Public Health*; BAT, British American Tobacco; DA, discourse analysis; NTR, *Nicotine and Tobacco Research*; TC, *Tobacco Control*; TDR, tobacco document research

I imported the 173 papers into N*Vivo as separate documents. They were divided into two sets, which I will refer to as A-papers and B-papers. A-papers were research papers primarily concerned with tobacco industry documents. There were 110 A-papers. B-papers were not primarily reports of tobacco document research. There were 63 B-papers. B-papers included, for example, literature reviews and survey research papers: all cited a small number of tobacco documents (generally one or two). The distinction between A-papers and B-papers was largely straightforward. Marginal decisions were based on the number of tobacco documents cited, whether they were acknowledged as a source and how they were used.

A-paper or B-paper status, author details, publishing journal and publication year information were encoded using N*Vivo's attributes function. Journals were divided into groups. Journals that had published eight papers or more were kept distinct in the analysis; other journals were divided into those that had published only one paper (1 paper journals) and those that had published between two and four papers (2–4 paper journals). Authors were divided into groups according to the number of papers they had published. Four groups of authors were identified: those that had published only one or two TDR papers (inexperienced authors); three to six papers (occasional authors); seven to 11 papers (frequent authors); and over 20 papers (prolific authors). No authors had published between 12 and 19 papers. Many inexperienced TDR authors were widely published in their own fields, including in tobacco control. Papers were divided into groups according to the number of papers published by their most published author. This ascribed reporting responsibility to the most experienced TDR author on a paper, although this may not always be the case in practice.

I then examined the content of the papers. I took an interpretive or constructivist approach—that is, I sought to make meaning from the papers and thus form an understanding of TDR, recognising that I would do this differently from other researchers.[14–19] To emphasise this I have expressed myself in the first person where possible, a common interpretive practice that acknowledges authors' active creation of research.[20]

My coding was guided by my research questions, my research experience (at time of writing I was one of the five most published TDR authors internationally), and my interest in epistemology—that is, the study of theories of knowledge or ways of knowing. Within those influences I kept codes concrete and as close as possible to the raw data.[21] I used the enumeration and cross tabulation functions of N*Vivo extensively to cross check completeness of coding, compare prevalent reporting practices with marginal practices, test emergent interpretations, and search for gaps and alternative explanations.[16 21–23] I kept a running record of my decision making. Although I could have enhanced the transparency of this paper by presenting sections of others' papers to illustrate, I avoided this because I did not wish to criticise particular authors.

Each paper was read repeatedly. The codes I developed from the methods reported in the papers were organised under five metathemes[23]:

- study purpose (only coded when explicitly stated)
- data sources
- search strategies
- analysis strategies
- limitations of the research.

To be clear: search strategies were the ways in which authors found documents, analogous to the data collection or survey administration phase of epidemiological research; analysis strategies were the ways in which authors made sense of the documents, analogous to the statistical analysis or modelling phase of an epidemiological study. All papers were re-coded iteratively as the coding system evolved.

RESULTS

TDR publishing patterns

The second wave of tobacco documents stimulated TDR publication: 93% of papers were published after 1998. Publication of B-papers, with one exception, commenced in 1999. Publication of A-papers peaked in 2002 and 2003; the number of B-papers increased annually from 2000 (fig 1).

Although 39 journals published TDR, a small subset dominated the field (table 1). *JAMA* published seven of the 12 papers to 1998 but little thereafter. After the second wave, *Tobacco Control* (TC) dominated (from 2000), publishing about half of all A-papers, followed by the *American Journal of Public Health* (AJPH, from 2001). Both published around five A-papers a year, except for 2002 and 2003, when TC published approximately one third of all A-papers (36 papers), many of them in supplements.

Almost a third of all B-papers were published in 2004. Up until 2003, most B-papers were published in TC (more than half) and 1 paper journals (a quarter). In 2004 this changed. The number of TDR papers in TC decreased. A quarter of the B-papers were published in 1 paper journals, as before. The rest were in published in journals that had previously rarely published B-papers, including *Nicotine and Tobacco Research* (NTR), *Lancet*, 2–4 paper journals, and AJPH.

Authors of TDR

A total of 222 authors were represented; 161 (almost three quarters) had worked on only one paper; only 17 had published five papers or more (table 1). Generally, A-papers were more likely to have highly published authors working on them, and B-papers, less published authors. The rush of A-papers in 2002 and 2003 was strongly dominated by frequent and prolific authors, mostly based at large centres funded specifically to do TDR. Otherwise each author group published about the same number of A-papers from 2000 to 2004. More than half of all B-papers were by inexperienced authors, half of these in 2004 alone; however, all author groups published B-papers (table 1).

Purpose of TDR

Of the 173 papers, 40 had no stated purpose, slightly more of those being B-papers. The simplest and most common purpose stated in the remaining papers was ''to describe what was in the documents''.

Five other more complex purposes occurred in fewer papers:

- to understand
- to argue, criticise, or assist with litigation
- to answer stated questions
- to analyse contributing factors or determine causation
- to compare the contents of document and non-document sources.

Data sources, searching, analysis, and limitations reported in TDR

The non-mutually exclusive sources and limitations reported are listed in table 2; searching and analysis practices reported are listed in table 3.

Patterns in reporting

Reporting was inconsistent. Most items in tables 2 and 3 occurred infrequently, combined differently by different

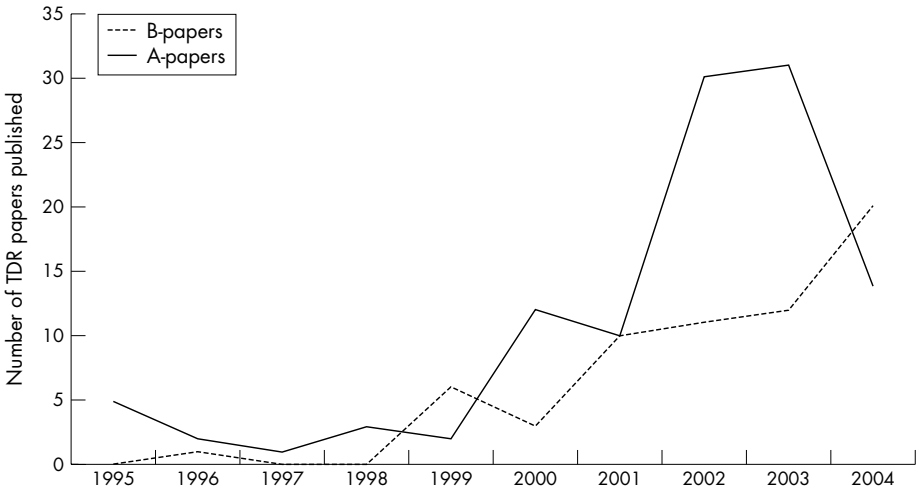

Figure 1 Number of tobacco document research (TDR) papers published over time.

authors. One third of all A-papers reported using multiple data sources, one third noted some limitations, about 80% provided some search information, and half some analysis information.

Patterns of reporting in A-papers

There was an apparent evolution in A-papers over time. These changes were complex, occurring more or less in some author groups or journals or even in some individuals. Generally, across time A-papers became more likely to:

- state a purpose
- state a purpose more complex than ''to describe what was in the documents''
- combine tobacco documents with other sources
- describe searching
- describe analysis.

Until 2001 most A-papers stated either no purpose or a descriptive purpose. From 2002 about half of A-papers had a more complex purpose, most often from frequent authors, least often from inexperienced authors. A-papers containing no source information were published until 2003, particularly in 1 paper journals, but became proportionally less prominent each year from 1998. There was a sharp decrease in A-papers containing no source information in 2001 and 2002, corresponding with a sharp proportional increase—to about 60%—in papers that reported using tobacco document sources only. In 2003 only, the majority of papers (just over half) reported combining document and non-document sources, again primarily from the frequent author group. Although this suggests evolution over time for these authors, this was true for only some individuals in the group. Papers with more complex purposes combined sources more often, as did papers in NTR and *Lancet*, and in AJPH.

Although methods sections ranged from one or two sentences up to many detailed paragraphs, the proportion of A-papers containing some search or analysis information increased over time, more consistently for searching than for analysis (fig 2). Inexperienced authors and 1 paper journals were least likely to provide search or analysis information. Frequent authors and prolific authors were most likely to provide search information; analysis information was most commonly provided by occasional authors and frequent authors, and in AJPH, 2–4 paper journals, and TC. AJPH and TC published the widest range of analysis information. Papers that stated a complex purpose, and those that combined sources, were more likely to describe searching or analysis.

Most innovations in reporting searching were introduced between 2000 and 2002, after which authors recombined established elements. In 2000 the first papers based on depository and online searching were published. Online searching is more common, and this imbalance is reflected in the search information observed. In 2000 authors most

Table 1 Publication patterns and authorship

	Total number of papers	Number of A-papers	Percentage of all A-papers	Number of B-papers	Percentage of all B-papers
Total	173	110	100	63	100
Number of TDR papers published by journal					
1 paper journals (27 journals)	27	12	11	15	24
2–4 paper journals (*Addiction, BMJ, Central European Journal of Public Health, Health Promotion Practice, Journal of Epidemiology and Community Health, Medical Journal of Australia, New England Journal of Medicine*)	19	9	8	10	16
Nicotine and Tobacco Research and *Lancet* (exclusively second wave)	16	7	6	9	14
JAMA (predominantly first wave)	9	9	8	0	0
American Journal of Public Health (exclusively second wave)	25	20	18	5	8
Tobacco Control (almost exclusively second wave)	77	53	48	24	38
TDR author group of most senior author on paper					
Most senior author is an inexperienced author (1 or 2 papers)	53	20	18	33	52
Most senior author is an occasional author (3–6 papers)	27	13	12	14	22
Most senior author is a frequent author (7–11 papers)	43	38	35	5	8
Most senior author is a prolific author (>20 papers)	50	39	35	11	17

Table 2 Sources and limitations reported in order of frequency of use

Tobacco document sources reported	Limitations reported
MSA tobacco industry web sites	Returned document set not complete
Tobacco Documents Online	Searching inefficient, returned document set voluminous
Legacy/UCSF	Searching problems
Other non-tobacco industry websites	Depositories hard to access
Tobacco documents non-MSA or BAT/B&W	Nature of document evidence creates limitations
Minnesota Depository	Documents returned by search redundant
BAT Guildford Depository	Not clear from documents whether plans were implemented
Early BAT and B&W doc set	Documents stop in 1990s
Non-tobacco document sources	Selection bias in total document population
General or specialist press, internet	No triangulation of documents with other sources
Legislation, lawsuits, hearings	
Published literature	
Interviews and surveys	
Government or state information	
Resources from individuals or organisations	
Quantitative data	
Observation	
Agencies working for the tobacco industry	
Patents	
Filings from stock exchanges, financial records	
Advertising campaigns	

BAT, British American Tobacco; B&W, Brown and Williamson; MSA, Master Settlement Agreement; UCSF, University of California, San Francisco.

often listed the keywords used and number of documents returned; less often, search dates, search strategy, imbalance in sources, and searching by people or groups. In 2001, authors began describing their searching as systematic, and reported combining search terms and searching in more fields including title, source file or location, dates, and document type. In 2002, authors began using Malone and Balbach's 2000 methods paper[12] to state that they had used normative techniques, and to report using optical character recognition (available at Tobacco Documents Online). Authors also began reporting searching for documents in context, including searching for Bates numbers consecutive to found documents, searching for all documents arising from particular events, projects or accounts, or searching by request for production (RFP) codes (for explanation see Cummings *et al*[24]).

Reporting of analysis was slower to start and less prevalent than reporting of searching, and is still evolving. The two least specific and earliest analysis practices reported—summarising documents and culling according to ''relevance''—appeared most often and across years. From 2000, authors talked about analysis using three kinds of terms that resonated with three different research traditions, sometimes coexisting in one paper. (Note that it was rare for the terms to be referenced to literature or the papers to be structured so as to allow the reader to understand and verify the authors' relationship to a particular tradition. Comparisons of different research traditions can be found in Denzin and Lincoln.[25])

The earliest terminology used to discuss TDR analysis borrowed loosely from interpretive or constructivist research. Researchers described identifying themes, used adjectives such as narrative, descriptive, and qualitative in reference to their analysis, and wrote that they coded inductively—that they created codes from reading the documents rather predetermining codes. A second way of writing about analysis borrowed from the research traditions generally referred to as positivist, post-positivist, or reductionist: first, formal selection criteria for culling and multiple researchers for coding; later pre-defined exhaustive coding systems that researchers were trained to use, and in a small number of papers, quantitative analyses. A third group of terms resonated with historical research traditions. The earliest and most common, from 2000, was the idea of chronology, a

Table 3 Searching and analysis practices reported in order of frequency of use

Searching practices	Searching practices continued	Analysis practices	Analysis practices continued
Search strings or terms listed	Notes more documents returned from some companies than others	Authors stated they kept ''relevant'' documents	Quantitative techniques used
Searching by topic or keyword	Date range of documents returned listed	Abstracting or summarising	Coders were trained
Number of documents found listed	Searching labelled ''exhaustive''	Selection criteria described	Analysis labelled ''qualitative''
Describes use of snowballing	Searching by title	Documents ordered chronologically	Analysis labelled ''mixed methods''
Searching for people or names	Searching labelled ''systematic''	Documents coded or indexed	Authors described considering quality of document information in analysis
Dates of search listed	Searching for projects or accounts	Specific analysis methods named (e.g. case study methods, content analysis)	Analysis labelled ''analytical''
Searching for organisations committees or companies	Searching labelled ''normative''	More than one coder used	Authors described interpreting the documents against other material
Search terms combined	Physical searching in depositories	Themes identified	Authors described emphasising documents for which a context was available
Searching by dates	Searching by request for production codes	Documents coded inductively	Authors described emphasising repeated rather than single instances
Overall search strategy described	Searching by document type	Analysis labelled ''narrative''	Analysis labelled ''historical''
Explains why search strategy used	Uses optical character recognition	Analysis labelled ''descriptive''	
Searching by source, file or location	Searching for meetings or events	Statistical tests used	
Searching by Bates numbers	Used keyword variation (e.g. misspellings)	Software used to manage data	

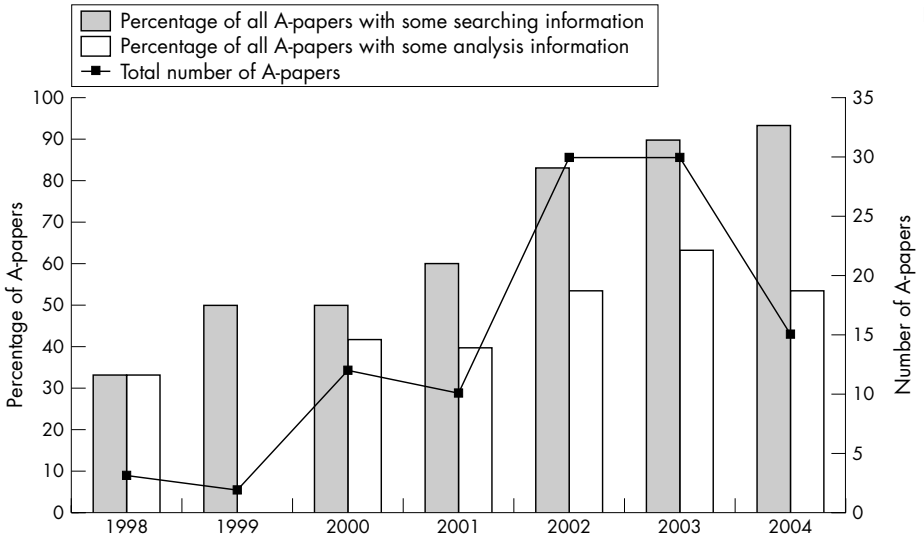

Figure 2 Searching and analysis information by year.

problematic but central concept in history. By 2003–4 a few TDR authors used language suggesting the quality of document evidence could not be taken at face value, a core issue in historiography. These authors described emphasising documents that were consistent with other documents or for which a meaningful context was available (such as a file), and making judgements about the trustworthiness of documents—for example, evaluating documents against other sources as opposed to simply treating different sources as stores of additional facts.

This late shift towards questioning the documents as evidence was also reflected in the limitations noted by TDR authors. Until 2003 limitations noted used language more common in positivist, post-positivist, or reductionist approaches, mostly describing technical barriers such as searching problems and selection bias, and focusing on whether documents examined by a researcher were representative of industry documents generally. In 2003 and 2004 a few authors raised limitations that seemed fundamentally different, suggesting that regardless of the representativeness of a document sample, TDR was limited because it was based on documentary evidence from the past. These authors argued the past could not be evaluated by the standards of the present, framed their work as an interpretation, suggested that the documents could not always tell researchers why things were done, and interrogated the quality of the information in the documents (for example, including only industry research that met certain criteria in their document analysis).

Patterns of reporting in B-papers

About three quarters of all B-papers fell into one of two types, both written mostly by inexperienced authors: reviews or essays that provided no source information; or research papers that provided detailed source information about non-document data, most commonly from surveys or interviews (almost 30% of B-papers), but referenced one or two documents with no related source information. A smaller subset of B-papers drew on a wide variety of sources including documents.

Document provenance was rarely discussed in B-papers. In about three quarters of B-papers documents were used in the introduction or discussion as evidence for a point—for example, that tobacco companies market to teenagers. More than half used quotes, demonstrating the perceived power of using the voice of the industry. A quarter of B-papers cited documents as though they constituted the same kind of evidence as published research—for example, authors would make a statement and support it with both an industry document and a piece of peer reviewed research, or would cite a single industry document instead of the published, peer reviewed TDR that existed on that subject.

Most B-papers used only one of two contrasting practices that suggested different views of the documents as evidence. In just over half of B-papers, a small number of documents were used to make an expansive statement about ''the industry'', a single quote representing all industry conduct. In contrast, in just under half of B-papers, documents provided a specific example or instance, located in place and time. The latter practice became more common over time and was most common in papers published in NTR, *Lancet*, and AJPH and by frequent authors: an example is Proctor's use of two tobacco documents along with other sources to detail the operations of ''Project Cosmic''.[26]

DISCUSSION

This section will address the second set of questions raised in the introduction.

Insights into the nature of TDR from its reporting traditions: my interpretations of the field

The TDR I examined was diverse: no clear standard had been established for it. Although most A-papers were produced by frequent or prolific authors, they often worked with inexperienced authors. Across the study period inexperienced authors independently published A-papers and, more often, B-papers. Journals began publishing TDR for the first time in recent years, frequently B-papers. The inexperience, newness, diversity, and lack of methodological context in TDR suggest that a clearer consideration of why we do things the way we do could be beneficial for the field.

I observed a ''descriptive mainstream'' in A-papers, becoming slowly less prevalent over time. ''Descriptive mainstream'' papers nominated no purpose or a descriptive purpose, no sources or only document sources, and had a typical structure: an introduction summarising an issue and promising to describe the contents of industry documents on that issue; methods; a long report of events, ordered chronologically and/or by issue, sometimes with extensive quotation; a short conclusion calling for action. Within this descriptive mainstream there were variations, such as early papers with no introduction or methods. A reasonable

example (with an atypical introduction) is my first TDR paper.[27]

In A-papers and B-papers, treatment of industry documents seemed to fall along a continuum, which I have labelled the continuum from researcher as conduit to researcher as constructor. When the researcher was positioned as a passive conduit, documents were used as straightforward nuggets of general truth (B-papers with one quote purporting to prove an expansive contention; A-papers with no source, searching, analysis, or researcher information, telling a story without qualification). When the researcher was positioned as an active constructor, documents were treated as problematic, complex sources of specific information that needed a context to be understood, and readers were made aware of the way in which the researcher had constructed their account of the past (B-papers giving precise, qualified examples located in place, person and time; A-papers with complex purposes, multiple sources, details of searching and analysis, and the limitations inherent in the documents). Most papers fell between these two positions and many contained elements of both, but in a general sense there was progression away from conduit and towards constructor. The methods and the discussion sections of Le Cook *et al*'s 2003 paper on cigarette design, for example, position the researchers more as constructors than conduits.[28]

The descriptive mainstream and the researcher as conduit became less prevalent over time and seemed connected. These patterns may have arisen partly from the initial ''forbidden fruit'' nature of the documents. The emphasis on secrecy, urgent tone, and dramatic language used in some early TDR seemed to telegraph inherent importance and trustworthiness, implicitly defining TDR as a special kind of research, with unique reporting standards.

Much of what I counted as reporting on analysis would not be considered such in other disciplines. I believe my definitions were extremely generous. Nonetheless, although reporting on analysis increased over time, analysis was discussed less often than searching, suggesting that TDR authors perceive analysis to be less salient than searching. To me this practice has a subtext: that once a researcher finds the documents, and justifies the way in which they were found, making sense of them is straightforward. This is consistent with TDR's lack of methodological context. When TDR authors provided background references they were generally to the seminal TDR methods papers,[12 13] which provide guidance on what to do. Authors rarely provided methodological or theoretical references that would explain why things were done a certain way.

TDR authors used vocabulary from positivist or reductionist research traditions, interpretive research traditions, and history, but rarely demonstrated that they were working explicitly or purposefully within these traditions. This is unsurprising, as many TDR authors, including some of the most careful and widely published, come from research disciplines that do not traditionally work with text, or think about methodology. Many of us, for example, have come from fields, such as clinical medicine, statistics or epidemiology, that take the positivist/post-positivist approach for granted.

In the next section I will discuss possibilities for analysis in TDR under three headings, reflecting the vocabulary used in TDR: interpretive textual research, positivist or reductionist textual research, and history. (There are both positivist and interpretive historians: I discuss history separately because historians' use of archives and focus on the past are uniquely relevant to TDR.) I am not suggesting that a single TDR paper or author should try to adopt all three traditions at once (to try to do so would be absurd and unhelpful). I am also not suggesting that current TDR is not useful because it has not adopted these traditions more fully. Rather I am exploring different directions that TDR authors could take to enrich their current practice.

TDR reporting traditions and other research reporting traditions

Interpretive or constructivist approaches and TDR

Many researchers working interpretively use an informal hermeneutic thematic analysis, much like the approach taken in this paper.[23] Other more formal interpretive approaches that could be adopted in TDR include frame, discourse, case study, or ethnographic analysis.

Frame analysis is used in media studies, linguistics, and policy studies.[29] Although highly contested,[29–33] in essence frame analysis seeks to elucidate the conflicting, invisible perspectives through which a single event or issue is presented in texts. Different frames highlight or suppress aspects of the same situation, frequently for ideological reasons, and serve functions such as laying blame, suggesting solutions, or calling to arms.[29 34] Frame theory is compatible with TDR: it transcends literal meanings, highlights conflict, is used to study social movements, and lends itself to advocacy.[29] A frame analysis could, for example, use documents and other sources to examine a government inquiry, contrasting the frames through which the issues were represented by different players, and emphasising the functions these frames played.

Discourse analysis (DA) is another fractured field, crossing many disciplines including psychology, linguistics, organisational studies, and politics.[35–38] Most DA approaches focus on language in detail and in context, proscribing summarisation or decontextualisation of documents, and limiting researchers to examining small bodies of text. DA presumes that language does not simply reflect, but creates, our social world, providing ways of asking ''how was this issue constructed?''. DA could be used to analyse a set of letters between a tobacco CEO and a politician on an issue, for example, not just for what was said, but how it was said. What was made important or unimportant? Were ideas recycled for emphasis? Were actions presented as obligations or as options? Did the letters tell stories, argue, instruct, negotiate? Did they include or exclude others' points of view? Whose voices were vicariously expressed: tobacco control, children, parents, retailers? Some DA traditions would relate these linguistic elements back to the larger social and political context.

The case study approach focuses on a highly specific and ''bounded'' case,[39] such as a decision or a programme, narrowly situated in place and time, and seems readily applicable to TDR. (A small number of TDR papers stated that they used case study methods, but rarely provided the detail needed to allow other TDR researchers to understand or replicate these methods.) A case study approach would not attempt to ''describe what the documents say about economics'' but could, for example, study the recent infamous Philip Morris commissioned economic report to the Czech government,[40] defining the study period tightly—perhaps six months before and six months after the report. The researcher would interpret (not just describe) the case, emphasising and detailing the context for the reports' production.[39 41 42] The case study approach emphasises diversity of sources: documents would be combined with, for example, archives, news reports, interviews, and observation where possible.[39 41 42] This approach is best suited to studying relatively recent or current events in which the researcher can directly immerse themselves: this would be its major limitation for TDR.[42]

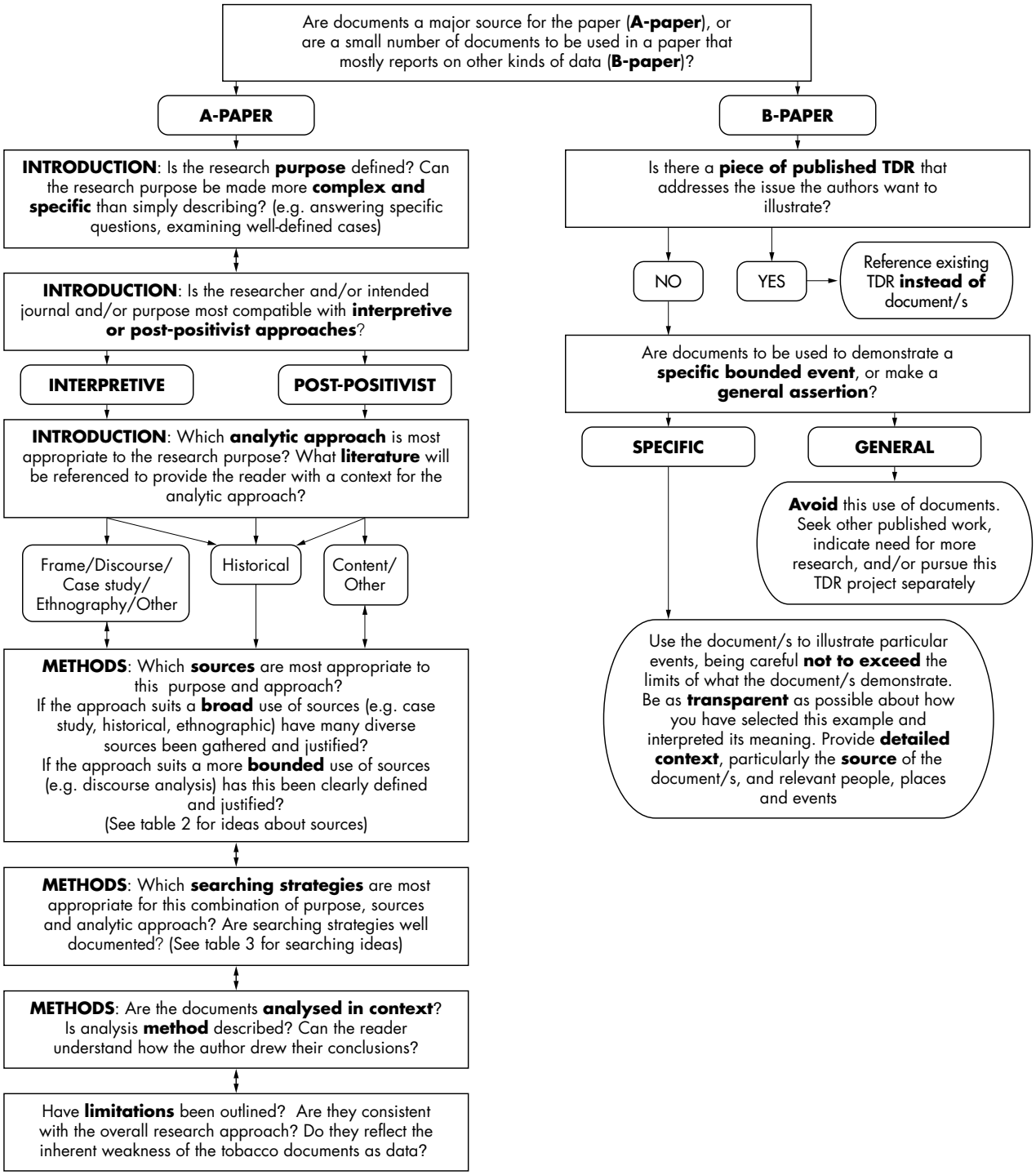

Figure 3 Flowchart for designing and evaluating TDR.

Ethnography[41 43–45] also uses diverse data sources, but focuses on culture. I can imagine two ways in which cultural questions could be asked in TDR. First, questions about specific groups in the population. Some TDR papers have studied particular countries or subcultures, but have mostly asked: "what do the documents say about this group?" In contrast, an ethnographic approach would ask a question such as "what does smoking mean in lesbian culture in Sydney?" using the documents, other archives, participant observation, and in-depth interviews with group members. Ethnographic TDR would move away from being "about" particular groups and towards including the perspectives of group members. The second kind of ethnographic TDR would study a tobacco company. Anecdotally, for example, BAT documents appear to be more explicit about document destruction. A researcher could seek to understand this by understanding the culture of the BAT group, or its document destruction operation. There are many examples of ethnographic and case study organisational research, although these models would require modification for TDR because of lack of meaningful access to the tobacco industry.

Positivist, post-positivist, or reductionist approaches and TDR

As public health is founded on positivist and post-positivist research traditions, it is unsurprising that some of the most scrupulous TDR researchers have adopted standards from these traditions, although mostly in relation to searching rather than analysis. For an excellent example of a carefully reported reductionist analysis of text (albeit mostly of sources other than tobacco documents) see Bryan-Jones and Bero 2003.[46] Few TDR papers in the health literature have reported using content analysis, the primary reductionist method used to analyse texts in many disciplines.[47 48] Content analysis involves counting concrete textual elements and statistically testing the patterns in which they occur: a good example is Balbach *et al*'s 2003 study, although content analysis is applied to magazine advertising rather than the documents themselves.[49] In situations where statistical demonstration of repeatability is important, content analysis may prove useful.

Possible lessons from historical research for the analysis of tobacco industry documents

Although TDR rarely acknowledges its historical character, and has been criticised as ahistorical,[50] discussions in historiography are particularly relevant to TDR because they address the challenges of researching the past. A major limitation on TDR researchers learning from historians is that history papers generally do not have methods sections as public health papers do, although they do have extensive footnotes. If TDR researchers are to learn to ''think historically'', they will probably need to consult historiography texts.

Positivism/post-positivism and interpretivism/constructivism conflict over their claims to objectivity and subjectivity respectively, and historiography contains similar tensions. One compromise position taken by many in the debate,[51–56] is that histories are competing readings of past events. Because the past is massive, surviving artefacts and accounts are fragmented, and we always view the past through the eyes of a very different present, we can never comprehend the past exactly or totally, difficulties shared but rarely acknowledged by TDR.[52 54 55 57–60] Sources, the only window to the past, are a key concern in history; many historiographers advocate amassing diverse sources to counterbalance their intrinsic flaws.[55 60] In contrast, while TDR amasses sources, they are rarely diverse.

Many historiographers argue that while the accuracy of facts is fundamentally important, what matters more is the interpretation of facts,[51 52 54 57 61] a distinction that seems not to have informed TDR writing. Historians disagree over the notion that their identities and ideologies affect their interpretations.[51 52 55 57 59 62] Traditional history has been criticised for serving powerful groups,[51 54 55 62] and recently histories have been written of previously excluded groups—for example, women, indigenous people, or workers—to energise political action.[51 53–55 59 62] Like ethnography and case study research, these marginal histories engage in detail with people's lives. TDR similarly criticises powerful groups, and tobacco documents could be included in marginal histories—for example, an historical study of tobacco workers. TDR could also learn from interpretive historians' conscious positioning of themselves in the research—for example, as feminists or indigenous people or unionists. Although many TDR researchers positioned themselves as a neutral conduit, TDR is often implicitly activist, and could adopt from interpretive history a more overt acknowledgment of this position.

Finally, historians, like discourse analysts, case study researchers, or ethnographers, emphasise the need for context and complexity.[26] This includes the need to consider institutional records as part of organisational processes and in the series in which they were originally produced,[55] consistent with some experienced TDR authors' recent prioritisation of contextualised documents.

What this paper adds

Authors have published research based on the internal documents of the tobacco industry since 1995. A small number of important papers have proposed methods that can be used in such research, but there has been little discussion of why tobacco document research should be done a certain way—that is, methodology—or of how tobacco document research is reported.

This study details the purposes, sources, searching, analysis, and limitations reported in peer reviewed research in the health literature based on tobacco industry documents. A ''descriptive mainstream'' form of reporting was observed to be gradually declining. Tobacco document research appeared to be slowly evolving away from positioning the researcher as a passive conduit and towards positioning the researcher as an active interpreter. A model for planning, writing, and evaluating tobacco document research is proposed to encourage this evolution and increase transparency.

Future directions for TDR reporting

Asserting a standard for TDR reporting will be difficult, not least because of inexperience in the field. However, certain reporting practices became more prevalent over time and ''clumped'' together—complex purposes, diverse sources, detail of searching and analysis—suggesting that experience was producing an evolution in some TDR reporting towards positioning the researcher as constructor. I believe this trend from more senior TDR authors is valuable because it increases transparency. Figure 3 proposes a process for planning and evaluating TDR that positions the researcher as constructor. There will certainly be traditions that I have neglected in the model. I hope it will be improved by adjustments from other TDR authors.

A standardised structure for TDR methods sections would encourage continuation of the trend towards researcher as constructor. Methods could include dedicated sections on sources (see examples in table 2), searching (see examples in table 3), and analysis (presently poorly defined in the literature: see discussion in previous section). Sources, searching, and analysis interact with purpose, which interacts with methodology (fig 3). In my view methodology should be interpretive/constructivist *or* positivist/post-positivist, *or* adopt another approach (one possible analytic frame not discussed here is the law).

Although word limits are problematic, few disciplines would consider it appropriate to have a one sentence methods section to marginally lengthen a results section. The model in fig 3 should produce more compact results by encouraging synthesis, rather than description. Some qualitative researchers use audit trails to help address word limits[63 64]; tables of raw data linked to one of my published papers provide one possible model for TDR audit trails.[65]

Although I believe TDR has much to learn from the rich analytic traditions described above, TDR researchers may have developed unique analyses specifically suited to the documents. The problem at the moment is that we rarely describe our analysis. TDR practice is clearly evolving. If this evolution continues, greater consistency and more detailed and complex documentation of analysis, purpose, sources, searching, and limitations should result.

Limitations of this research

This work was dependent on Medline indexing, and is an interpretive analysis of TDR reporting, not practice. The model and discussion are not intended to be prescriptive, but to offer possibilities that other researchers can build on and modify.

ACKNOWLEDGEMENTS

Thank you to the reviewers, Edith Balbach, Jeff Collin and Stan Glantz, and to Claire Hooker, Simon Chapman, Mary Assunta, Michelle Scollo, Fiona Byrne and Jenny Knight for their time, thoughtful feedback, encouragement during the writing of earlier drafts, and specific suggestions for changes (thanks especially to Jeff and Claire for their comments leading to the section on ethnography of the organisation and Michelle for her observations on the lack of a legal perspective in TDR in the health literature). Responsibility for this final version lies entirely with me.

Sponsor details: this work was supported by an Australian Postgraduate Award.

Competing interests: none declared

Ethics approval: not required

REFERENCES

1 **Cummings KM**, Pollay RW. Exposing Mr Butts' tricks of the trade. Introduction. *Tobacco Control* 2002;**11**(suppl I):i1–4.
2 **Bero L**. Implications of the tobacco industry documents for public health and policy. *Annu Rev Public Health* 2003;**24**:267–88.
3 **Glantz S**, Barnes D, Bero L, *et al*. Looking through a keyhole at the tobacco industry. *JAMA* 1995;**274**:219–24.
4 **MacKenzie R**, Collin J, Lee K. *Handbook on the use of tobacco industry documents for research*, London School of Hygiene and Tropical Medicine and Rockefeller Foundation, 2003. http://www.lshtm.ac.uk/cgch/tobacco/Handbook%2008.07.03.pdf.
5 **Legacy Tobacco Documents Library**. American Legacy Foundation and University of California San Fransisco Library Center for Knowledge Management. http://legacy.library.ucsf.edu/ (accessed 17 Feb 2005).
6 **Tobacco Archives**. http://www.tobaccoarchives.com/, (Accessed 15 Feb 2005).
7 **Tobacco Documents Online**. http://tobaccodocuments.org/, (Accessed 15 Feb 2005).
8 **British American Tobacco Documents Archive**. http://bat.library.ucsf.edu, (Accessed 15 Feb 2005).
9 **Collin J**, Lee K, Gilmore AB. Unlocking the corporate documents of British American Tobacco: an invaluable global resource needs radically improved access. *Lancet* 2004;**363**:1746–7.
10 **Lee K**, Gilmore AB, Collin J. Looking inside the tobacco industry: revealing insights from the Guildford Depository. *Addiction* 2004;**99**:394–7.
11 **Muggli ME**, LeGresley EM, Hurt RD. Big tobacco is watching: British American Tobacco's surveillance and information concealment at the Guildford depository. *Lancet* 2004;**363**:1812–9.
12 **Malone RE**, Balbach, eds. Tobacco industry documents: treasure trove or quagmire? *Tobacco Control* 2000;**9**:334–8.
13 **Balbach ED**, Gasior RJ, Barbeau EM. Tobacco industry documents: comparing the Minnesota Depository and internet access. *Tobacco Control* 2002;**11**:68–72.
14 **Angen MJ**. Evaluating interpretive inquiry: reviewing the validity debate and opening the dialogue. *Qualitative Health Research* 2000;**10**:378–95.
15 **Lacity MC**, Janson MA. Understanding qualitative data: a framework of text analysis methods. *Journal of Management Information Systems* 1994;**11**:137–156.
16 **Marshall C**. *Designing qualitative research*. Thousand Oaks, California: SAGE Publications, 1999.
17 **Poland BD**. Learning to ''walk our talk'': the implications of sociological theory for research methodologies in health promotion. *Can J Public Health* 1992;**83**:S31–46.
18 **Guba EG**, Lincoln YS. Competing paradigms in qualitative research. In: Denzin NK, Lincoln YS, eds. *Handbook of qualitative research*. Thousand Oaks, California: SAGE Publications, 1994.
19 **Denzin NK**, Lincoln YS. Entering the field of qualitative research. In: Denzin NK, Lincoln YS, eds. *Handbook of qualitative research*. Thousand Oaks, California: SAGE Publications, 1994.
20 **Gilgun JF**. ''Grab'' and good science: writing up the results of qualitative research. *Qualitative Health Research* 2005;**15**:256–62.
21 **Seale C**. *The quality of qualitative research*. London: SAGE, 1999.
22 **Silverman D**. *Doing qualitative research: a practical handbook*. London: SAGE publications, 2000.
23 **Ryan GW**, Bernard HR. Techniques to identify themes. *Field Methods* 2003;**15**:85–109.
24 **Cummings KM**, Morley CP, Horan JK, *et al*. Marketing to America's youth: evidence from corporate documents. *Tobacco Control* 2002;**11**(suppl I):i15–17.
25 In: **Denzin NK**, Lincoln YS, eds. *Handbook of qualitative research*. Thousand Oaks, California: SAGE Publications, 1994.
26 **Proctor RN**. Should medical historians be working for the tobacco industry? *Lancet* 2004;**363**:1174–5.
27 **Carter SM**. Mongoven, Biscoe & Duchin: destroying tobacco control activism from the inside. *Tobacco Control* 2002;**11**:112–8.
28 **Le Cook B**, Wayne GF, Keithly L, *et al*. One size does not fit all: how the tobacco industry has altered cigarette design to target consumer groups with specific psychological and psychosocial needs. *Addiction* 2003;**98**:1547–61.
29 **Benford RD**, Snow DA. Framing processes and social movements: an overview and assessment. *Annual Review of Sociology* 2000;**26**:11–39.
30 **Goffman E**. *Frame analysis: an essay on the organization of experience*. New York: Harper & Row, 1974.
31 **D'Angelo P**. News framing as a multi-paradigmatic research program: a response to Entman. *Journal of Communication* 2002;**52**:870–88.
32 **Scheufele DA**. Framing as a theory of media effects. *Journal of Communication* 1999;**49**:103–22.
33 **Entman RM**. Framing: toward clarification of a fractured paradigm. *Journal of Communication* 1993;**43**:51–8.
34 **Ryan C**. *Prime time activism*. Boston, Massachusetts: South End Press, 1991.
35 **Gee J**. *An introduction to discourse analysis*. London: Routledge, 1999.
36 **Schiffrin D**, Tannen D, Hamilton H. *The handbook of discourse analysis*. Oxford: Blackwell, 2003.
37 **Antaki C**, Billig M, Edwards D, *et al*. Discourse analysis means doing analysis: a critique of six analytic shortcomings (also see online responses). Discourse analysis online. www.shu.ac.uk/daol/articles/v1/n1/a1/antaki2002002-paper.htm (Accessed 25 Feb 2005).
38 **Blommaert J**, Bulcaen C. Critical discourse analysis. *Annual Review of Anthropology* 2000;**29**:447–66.
39 **Stake R**. *The art of case study research*. Thousand Oaks, California: SAGE Publications, 1995.
40 **Arthur D**. Little International Inc. Public Finance Balance of Smoking in the Czech Republic: executive summary, http://www.mindfully.org/Industry/Philip-Morris-Czech-Study.htm (Accessed 10 March 2005).
41 **Cresswell JW**. *Qualitative inquiry and research design: choosing among five traditions*. London: SAGE Publications, 1998.
42 **Yin RK**. *Case study research, design and methods*, 3rd ed. Newbury Park, California: SAGE Publication, 2002.
43 **Fetterman DM**. *Ethnography: step by step*. Newbury Park, California: SAGE Publication, 1989.
44 **Hammersley M**, Atkinson P. *Ethnography: principles in practice*, 2nd ed. New York: Routledge, 1995.
45 **Wolcott HF**. *Transforming qualitative data: description, analysis and interpretation*. Thousand Oaks, California: SAGE Publications, 1994.
46 **Bryan-Jones K**, Bero LA. Tobacco industry efforts to defeat the occupational safety and health administration indoor air quality rule. *Am J Public Health* 2003;**93**:585–92.
47 **Krippendorff K**. *Content analysis: an introduction to its methodology*, 2nd ed. Thousand Oaks, California: SAGE Publications, 2004.
48 **Neuendorf KA**. *The content analysis guidebook*. Thousand Oaks, California: SAGE Publications, 2002.
49 **Balbach ED**, Gasior RJ, Barbeau EM. R.J. Reynolds' targeting of African Americans: 1988–2000, *Am J Public Health* 2003;**93**:822–7.
50 **Talley C**, Kushner HI, Sterk CE. Lung cancer, chronic disease epidemiology, and medicine, 1948–1964. *Journal of the History of Medicine & Allied Sciences* 2004;**59**:329–74.
51 **Appleby J**, Hunt L, Jacob M. *Telling the truth about history*. New York and London: WW Norton & Company, 1994.
52 **Carr E**. *What is history?*. Houndmills, UK: Palgrave, 1961.
53 **Iggers G**. *Historiography in the twentieth century*. Hanover and London: Wesleyan University Press, 1997.
54 **Jenkins K**. *Re-thinking history*. London and New York: Routledge, 1991.
55 **Tosh J**. *The pursuit of history: aims, methods and new directions in the study of modern history*. Harlow, Essex, UK: Longman Group Ltd, 1984.
56 **Evans R**. *In defense of history*. London: Granta Publications, 1997.
57 **Becker CL**. What is evidence? The relativist view - ''Everyman his own historian''. In, Winks RW, eds. The historian as detective, *essays on evidence*, 2nd ed. New York, Evanstone and London: Harper and Row, 1969.
58 **Winks RW**. Introduction. In: Winks RW, eds. *The historian as detective: essays on evidence*, 2nd ed. New York, Evanstone and London: Harper and Row, 1969.
59 **Fernandez-Armesto F**. Epilogue. In: Cannadine D, eds. *What is history - now?* Houndmills, UK: Palgrave MacMillan, 2002.
60 **Collingwood RG**. The pleasures of doubt: re-enacting the crime. In: Winks RW, eds. *The historian as detective: essays on evidence*, 2nd ed. New York, Evanstone and London: Harper and Row, 1969.
61 **Pedersen S**. What is political history now? In: Cannadine D, eds. *What is history - now?* Houndmills, UK: Palgrave MacMillan, 2002.
62 **Rubin M**. What is cultural history now? In: Cannadine D, eds. *What is history - now?* Houndmills, UK: Palgrave MacMillan, 2002.
63 **Wolf ZR**. Exploring the audit trail for qualitative investigations. *Nurse Educator* 2003;**28**:175–8.
64 **Koch T**. Commentary: expert researchers and audit trails. *J Adv Nurs* 2004;**45**:134–5.
65 **Carter SM**, Chapman S. Supplementary material for the paper ''Smoking disease and obdurate denial''. http://tobacco.health.usyd.edu.au/site/gateway/docs/pdf/Carter_Health_1980s_Supplementarydata.pdf (Accessed 30 March 2005).