Journal of Business and Technical Communication
http://jbt.sagepub.com/

Missed Opportunities in the Review and Revision of Clinical Study Reports

Gregory P. Cuppan and Stephen A. Bernhardt

Journal of Business and Technical Communication 2012 26: 131 originally published online 30 January 2012
DOI: 10.1177/1050651911430624

The online version of this article can be found at:
http://jbt.sagepub.com/content/26/2/131

Published by:
SAGE
http://www.sagepublications.com

Articles

Journal of Business and Technical Communication
26(2) 131-170

DOI: 10.1177/1050651911430624
http://jbtc.sagepub.com

Missed Opportunities in the Review and Revision of Clinical Study Reports

Gregory P. Cuppan[1] and Stephen A. Bernhardt[2]

Abstract

Circulating written drafts and conducting roundtable reviews are two important document-development activities in many work sites. Previous studies suggest that review processes are frustrating for participants and have substantial inefficiencies caused by conflicting participant purposes. This article presents two case studies of the document-review practices for clinical study reports from a large pharmaceutical company, paying particular attention to whether review efforts contributed to improvements in document quality. Findings suggest that document review did not lead to demonstrable improvement in report quality. The authors offer recommendations for improving document-review practices.

Keywords

document review, medical writing, pharmaceutical development

[1] McCulley/Cuppan, LLC, Salt Lake City, UT, USA
[2] Department of English, University of Delaware, Newark, DE, USA

Corresponding Author:
Stephen A. Bernhardt, Department of English, University of Delaware, Memorial 132, Newark, DE 19716, USA
E-mail: sab@english.udel.edu

For 17 years, we have consulted regularly with various pharmaceutical companies to help them improve the quality of their documents. This work has taken various shapes: providing training workshops to develop effective individual and team practices, assessing document quality, facilitating document development in departmental and cross-functional teams, developing internal procedures and practice guidelines, coordinating uses of document technologies, and creating annotated model documents. An increasingly important focus of our work has been on review practices, the common activity of circulating drafts of reports for comments and revision. Much of our work has focused on clinical study reports (CSRs), the key filing documents for new drugs that report experimental studies of drug effectiveness and safety (Bernick, Bernhardt, & Cuppan, 2008; Fossati Wood & Foote, 2009). These reports are the focus of this article. Our title suggests we see room for improvement of review practices.

This article is a follow-up to an earlier study that detailed general issues with review practices for large documents in pharmaceutical companies (Bernhardt, 2003). Finding that review was frequently a frustrating and time-consuming practice, the study suggested that companies would benefit from review practices that were more intentional, with better articulated purposes, goals, and procedures. In our work, we have found that best practices in document review share some of the following characteristics:

- The reviews lead to improved document quality.
- The review practices respect the cost of participants' time, achieving some level of efficiency.
- Draft reports move relatively quickly toward final reports, with a reasonable number of review sessions.
- Review focus, method, and participation are staged according to the phase of document development (e.g., an early prototype or rough draft receives different sorts of attention than an almost-final draft).
- Reviews, especially those involving real-time group discussion, focus on important rhetorical concerns (purpose, audience, argument, logic, issues, conclusions, implications) as opposed to minor stylistic elements (word choice, format, order of details, consistency, style preferences).

Reviews should be used to consolidate the strategic knowledge that is often distributed unevenly across the organization. A truly strategic review of a CSR marshals pharmaceutical and clinical-development intelligence to identify critical issues, specific challenges, and arguments that must be addressed in order to produce a high-quality research report. A strategic

review as a work practice should help a team discover and reach consensus on the strongest arguments as warranted by the research data and scientific understanding. The review should test a document's usability insofar as it enables regulatory agents to perform their jobs. A high-quality report should be designed to meet the needs of the principal users of this document genre: the drug regulatory agents charged with approving drug products for specific therapeutic applications.

In this article, we seek to further the understanding of review practices as applied to large and complex technical documentation. This work is an extension of other studies of review practices in professional settings. Paradis, Dobrin, and Miller (1985) first established document review as a critical site of contentious interaction in a research-and-development environment (Exxon). These authors contrasted document review from the opposing perspectives of managers and employees, detailing the ways that individuals often worked without shared purpose or expectations, resulting in frustrating differences of perception, feelings of resentment, and the need to substantially rework documents during successive reviews. Van der Geest and van Gemert (1997) pinpointed a general sense of frustration with review processes in several Dutch companies, noting that "both writers and reviewers find reviewing the most cumbersome stage in the process of text production, given that many parties are involved in it and that it is loaded with different expectations" (p. 445). Henry (2000), working with data gathered by many interns at various professional sites, found that reviews were "fraught with second guessing" and required "interpretations of organizational culture to the ends of adequately and appropriately delivering discursive products" (p. 65). And several studies (Henry, 2000; Katz, 1998; Paradis, Dobrin, & Miller, 1985) discuss the activity of "document cycling," an activity in which documents pass through multiple and sometimes conflicting reviews, as various reviewers weigh in with commentary.

A theme throughout the literature is that document review frequently brings into play conflicting or competing purposes. In addition to improving a document, review sometimes functions to evaluate worker performance (Couture & Rymer, 1991) or to discipline individuals (Henry, 2000, p. 81). Thus, interpersonal dynamics are frequently in play. Katz (1998), in particular, highlighted how review processes can function positively as a way to socialize new workers, helping them learn how to both write and work successfully within the local culture. Henry (2000) shared this concern for how interns come to understand the ways that organizations perform work.

Some researchers contrast purposes of review in academic circles with the purposes and expectations of review in work settings. Shwom and Hirsch (1994) noted the problems caused by review of early drafts. They explained that the whole notion of *drafting*, viewed so positively in classrooms as an opportunity for feedback and improvement, is typically seen as inefficient in engineering organizations. The idea that writers need early feedback on rough drafts is not widely recognized in these professional environments. Shwom and Hirsch suggested that more carefully characterizing the purpose of various review stages would smooth the process.

To further our understanding of the complexities of document review, we present case studies of the review practices for two CSRs from different projects at one large company. In our assessment of these review practices, first we interviewed a range of participants about their experiences and expectations regarding document review. Then we tracked all comments and edits by all participants during the multiple review stages for each CSR. We analyzed these data for the kinds and frequencies of review comments and edits made individually by reviewers and collectively by each development team. We also assessed and compared the differences in document quality between early drafts and the final reports, paying particular attention to whether review efforts contributed to improvements in the quality of communication. While our focus is one company's practices, we will refer to our corroborating experiences at another large pharmaceutical company.

Like an earlier article (Bernhardt, 2003), this one is shaped by an action-research agenda (Clark, 2004). By *action research*, we mean activity in which partnering clients engage with consultant researchers to identify problems, collect data, test solutions, and refine approaches through workplace training and feedback. This project reflects our continuing work as consultants hired to study, analyze, and improve documentation practices. Over the past 17 years, we have worked with over 50 pharmaceutical companies to evaluate, coordinate, or improve CSRs, and we have worked in some fashion with more than 400 such reports. This consulting provides us an opportunity to see practices from inside organizations. In important ways, our consulting role limits our control over study design because we can do only what clients want us to do and are willing to pay for. But it also means that our clients are "in on the action," working collaboratively with us to define the problems they experience with document development and to improve outcomes. Our clients describe their frustrations, identify problems with reports, and confirm or reject our suggestions about what is going right or wrong and what ought to be done. As we interview participants about work practices, collect data, assess draft and final documents, develop

training for the organization, and evaluate outcomes, we develop emerging, shared understandings of the complexities of document development in scientific organizations.

The data in this article reflect our success in convincing two companies that, to prepare for document-review training, they would benefit by quantifying their review performances and measuring various attributes of performance. Unlike the earlier article (Bernhardt, 2003), which was quite general in its analysis and moved freely across multiple companies and projects, this article draws a tight focus on the review cycles for specific documents. Our goal here is to help reveal what transpires during document review and to relate those activities to desired outcomes, namely, improved document quality. This article should interest those in industry who coordinate or participate in review processes, industry trainers who work to implement best practices, and those in the teaching profession who prepare professional communicators.

Before we discuss our findings from these two case studies, we present a background on the genre of the CSR.

Background on CSRs

CSRs are a specific genre of research reports, an important component of the research-and-development dossier that presents the drug sponsor's case for approving a new drug or the extended therapeutic application of an existing drug. Depending on the stage of drug development, a CSR might aim to characterize toxicity or tolerability in healthy individuals, pharmacodynamic or kinetic profiles in humans, or efficacy or safety in one population or another. Besides showing that the new drug or extended application is safe and efficacious, as demonstrated by clinical studies and reported in CSRs, a drug sponsor must (in some markets) show a compelling therapeutic need or economic rationale for the drug. Additionally a drug sponsor must show, in a comprehensive compilation of research-and-development. reports of various sorts, that the drug product is a stable, well-characterized chemical or biologic entity, reproducible within narrow quality standards, environmentally acceptable, and characterized overall by a favorable risk/benefit profile. In pharmaceutical development, the CSR is thus situated within complex scientific, organizational, regulatory, business, and legal contexts. We describe these contexts and offer an analysis of the genre of the CSR elsewhere (Bernick et al., 2008). In brief, the CSR helps the pharmaceutical company that is developing a new drug to establish and build a convincing information base and argumentative positions for the

investigational drug's therapeutic activity and safety in various patient populations. Without strong, positive efficacy and safety studies, and without the corresponding CSRs, there is no case for approval.

The CSR must describe in sufficient detail the purpose of the work, the relationships studied, and the specifics of study design, conduct, and analysis. These details enable the drug regulatory agents (the primary audience for CSRs) to determine whether the characteristics of an adequate and well-controlled study are present. (These terms, *adequate* and *well-controlled*, are charged with meaning—they broadly signify whether a study is a success or failure.) The CSRs produced by pharmaceutical companies and their contract agents routinely follow the guidance on content and structure issued by the International Conference on Harmonisation and accepted by regulatory authorities such as the United States Food and Drug Administration. The underlying generic structure of the CSR is the familiar scientific journal research report, but with many adaptations, additions, and requirements that dramatically alter the scale of the reports. The typical CSR submitted to regulatory audiences will include 50 to over 100 pages of main text, plus appendixes that often comprise thousands of pages of patient case records and data sets.

In terms of argument, the CSR must report and interpret for the reader both the statistical and the clinical significance—in terms of drug efficacy and safety—of the findings collected on all subjects enrolled in a study. Piantadosi (1997), as well as Lang and Secic (1997), suggested that for a successful study, statistical and clinical reasoning must converge within the body of the CSR. Explanatory research on pharmaceutical drugs therefore requires two interdependent logical or argumentative tasks:

- generalizing observations from few to many (statistical reasoning)
- integrating empirical data with theory-based and practice-based knowledge (clinical reasoning)

This demand for the convergence of statistical and clinical significance establishes two primary argumentative topoi. The study must be designed to produce results that offer statistical confidence in the findings—that observed differences in the data (or lack of differences) are determined reliably and suggest a low probability of alternative results upon further study. Furthermore, the study must make a convincing argument that any observed differences would actually be relevant in clinical settings, where people with diseases or conditions are seeking relief.

The warrants for such arguments are explicitly framed by the study's statistical model. While studies may demonstrate small, statistically significant differences between large groups, these differences might not be viewed as clinically meaningful. Successful arguments need to establish warranted claims that respond to questions such as these:

- Is the study well designed and conducted?
- Does the study provide sufficiently strong evidence for efficacy and safety?
- Does the drug actually affect the disease state or does it merely alter or suppress certain markers associated with the disease (e.g., reduce heart attack or stroke incidence vs. reduce cholesterol, improve cancer survival rates vs. shrink tumors)?
- What are the clinical implications of the adverse-event profile? Do drug side effects or patient intolerability outweigh the benefits of the drug?
- Will certain patient subgroups be limited in the use of this product?

The need to demonstrate both statistical and clinical significance puts CSR authoring and reviewing teams into a difficult argumentative forum—a forum where they must explain that a finding is the result of a biological process shared by a group of patients that is indeed subject to explanation, measurement, prediction, and ultimately, control.

A study unfolds over time within an emerging framework of data and understanding, and although hypotheses, methods, populations, and measures are represented prospectively in the study protocol, a CSR must account for everything that did and did not go as expected. In any clinical program, and in any particular study, issues emerge that might be either expected with a given class of compounds or surprising. Data are often equivocal, unexpected, or mystifying. Trials are complicated, often spanning years and involving hundreds or thousands of subjects at multiple research sites, often across countries. The studies frequently have significant turnover in key personnel.

A key challenge is that a CSR must detail and explain the meaning of what happened during a study, but it also must avoid constraining the company if a future study provides results that are inconsistent with prior data, arguments, or conclusions. The tendency, thus, is to be conservative, especially in earlier studies, in offering generalizations and conclusions. Once all studies are completed, the data can be read retrospectively with greater insight and confidence, allowing fewer opportunities for misinterpretation. But companies cannot wait to write up earlier studies until later ones unfold

because timely completion of study reports (complete with logical arguments, discussions, and conclusions of significance) is generally required by regulatory agencies.

A pharmaceutical company does have an opportunity to draw study findings together because they must provide to regulatory agencies integrated summaries of all findings on safety, efficacy, and risk/benefit based on the whole, multiyear program of clinical studies (each with its attendant CSR). In these integrated summaries, a company can attempt to reconcile inconsistencies or shortcomings in the data and prior interpretations. Yet, the rhetorical bind persists: The individual report must be well argued and conclusive, addressing any complexities raised by the study. At the same time, it must leave room for future interpretive positions so that the whole development program can be integrated within a coherent summary document that favorably situates the company for drug approval with as few prescribing restrictions as possible on disease conditions and patient populations.

The organizational context of the CSR also poses multiple rhetorical challenges to the development team. A report requires the contributions of experts of various specializations. Different people routinely write sections that must come together into a coherent whole. The process might best be characterized as cooperative authorship, in which individuals contribute sections according to individual expertise, as opposed to collaborative authorship, in which individuals or teams actually plan and compose together. Common practice is to use review as a trigger to foster collaboration and to use review roundtables as a forum for reaching consensus on the final shape of each section of the document.

The number of individuals closely involved in developing and reviewing a particular report will vary across organizations and even project teams. Typically, and true for the cases reported here, the review team will consist of eight to ten people. The primary author is usually a medical writer, either employed by the company within a medical writing group or contracted as an outside agent. Invariably, there is a group of contributing authors representing various specializations associated with the conduct of human medicine research. Most often the group involves a clinician, a pharmacologist, a drug-safety expert, a biostatistician, and perhaps a geneticist or pharmacoeconomist. The rest of the review team tends to consist of the head of the drug-development program and specialists from clinical operations, regulatory affairs, or quality control. External medical experts are commonly involved in the review process, especially if they contributed to study design or conduct. In our experience, senior managers (associated with the therapeutic specialty, pharmacology, clinical operations, or regulatory

affairs) routinely enter the review process late and often engage only in reviews of pivotal studies or studies with troublesome issues.

Once a drug-filing dossier is submitted to health authorities (the Food and Drug Administration in the United States, the European Medicines Agency in the European Union, and various other regional or national authorities), the filing undergoes rigorous review by teams of internal and external scientists and regulatory experts. This review process is characterized only in general terms because the regulatory authorities are circumspect about their internal work practices due to public and political scrutiny and a desire to maintain an arm's-length relationship with drug sponsors. Thus, regulatory bodies are careful about the kinds of communication their agents have with drug sponsors so that the process appears unbiased, rigorous, and objective. This rather distant author–audience relationship makes it difficult for companies to have a good sense of whether individual reports in the filing are well written and rhetorically successful, that is, whether they achieve their purposes. Typically the feedback from the health authority agent is limited to scientific challenges or requests for clarification. Rarely is there commentary on document quality beyond what can be inferred when a reviewer misinterprets information or misses a point of fact. Only in worst-case scenarios do health agencies provide explicit feedback to drug sponsors about their submission documents. Because authors and regulatory reviewers are necessarily kept at some distance, companies must imagine how these reviewers will respond to their documents, anticipate their reactions, and develop documents that are responsive to a wide range of reviewers.

We offer this context to underscore the rhetorical and organizational challenge of constructing CSRs. The reports must do difficult work, and the companies we work with continually strive to improve the quality of the reports. They invest immense resources in conducting the studies, and many scientists devote countless hours to drafting, revising, and finalizing the reports. They have a vested interest in offering reports that are convincing in the case for approval, that generate few challenges from regulatory authorities, and that answer the core questions: Is this study adequate and well controlled? Is this drug convincingly safe and reasonably effective in the targeted patient population? Do the benefits outweigh the risks?

Method

The assessments we describe here characterize the review efforts made by two teams, each engaged in the process of drafting and revising a CSR. Our

goal was to help the client, a large pharmaceutical company, gain greater understanding of whether the review efforts of each team were strategic and effective. We started the assessment by interviewing 21 individuals across the project teams in order to gain insight into how people in the organization approached the task of review. Then we examined and categorized the review comments made by the two document-development teams during each round of review of two CSRs, each reporting different drug-development efforts. These two reports were identified by the company as being at an appropriate stage of development, with review processes just getting under way. We also assessed document quality to look for improvement between initial and final draft versions of the CSRs. We collected the data for these studies during 2005–2006. We present our findings as case studies, organized by project.

Reviewer Interviews

As part of a consulting intervention to improve review practices, we developed a common set of questions and interview prompts. We interviewed 21 individuals working across the various functional areas associated with clinical drug development at various operational and management levels. The client selected all interviewees, the majority of whom were involved with one of the two document-review projects described here. We did not interview the report writers (both medical writers) of the two CSRs. We conducted the interviews by telephone over a 5-week period; each interview was scheduled for 50 minutes, though some lasted longer. We digitally recorded the majority of interviews and relied on interviewer notes for those few that were not recorded. We briefed interviewees on the nature of the project and the corporate rationale for pursuing the project, and we asked them to speak openly about their concerns and perspectives. The interview questions focused on these topics:

- What is the principal focus or main purpose of your review?
- At what document-development stage do you believe your input is most valuable? Please describe what value you specifically added to the review of a CSR.
- Does the focus of your review change with the advancement of the document?
- What works really well in the document-review cycle in your group?
- What does not work so well?
- In general, is a sufficient amount of time allocated for doing reviews?

Cuppan (the first author) and another senior consultant in our firm reviewed and independently analyzed the interview recordings. Both consultants developed a written description that represented current organizational review practices based on their individual interpretations of the interviews. We folded together these individual analyses to generate the overall findings regarding interviewee concerns and perspectives about their individual and their organization's approach to reviewing CSRs.

Data Collection for Review Remarks

The second phase of the project required tracking review practices as the documents were circulated. The company used an electronic document-management system to track and control versions and to collect review remarks. We received the marked-up versions of the CSR files for each reviewer for each round of review. Some reviewers marked up paper versions of the files. We received copies of these documents as well. In one instance, a reviewer submitted commentary via a series of e-mails. We received printouts of this e-mail correspondence. We sat in during roundtable review meetings, keeping notes of our observations during these roundtables. We did not attempt to record the suggested changes or comments on the drafts during these roundtables; rather, we observed for more general behaviors.

Categorization of Review Remarks

A single analyst (not either of us) examined the summary files to count and categorize the review remarks by report section for each draft version of the CSR. We logged the tallies to an Excel® spreadsheet. During the early phases of analysis, a second analyst (not either of us) examined 20% of the work to verify that the identification and categorization of the review remarks were accurate and consistent. During a conference call that Cuppan moderated, the primary and secondary analysts discussed points in which they disagreed on the identification and categorization in order to clarify distinctions and make the analysis more reliable. The primary analyst then completed the analysis. Review remarks were categorized as either edits or comments (see Figure 1). *Edits* make specific changes to words, punctuation, sentences, paragraphs, or sections, including changes to numbers in the text or in tables, without any accompanying commentary. *Comments* suggest the need for changes to a sentence, paragraph, tabled data, or section

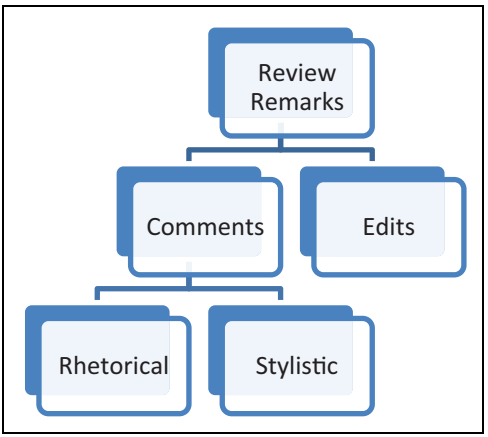

Figure 1. Categorization of review remarks.

or question the perceived meaning of a data value or a set of data, providing some level of explanation of the desired change or why the change is needed.

Review remarks made in the track changes mode of MS Word® were all treated as edits unless the reviewer included a statement regarding why the specific change was made; in such cases, the track change and comment were counted collectively as one comment. MS Word formatting edits of headers and footers, bulleted and numbered lists, and tables were not counted nor were global formatting changes of font styles. A reviewer's highlighting of a text block to indicate the scope of a remark was also ignored as an edit. If, in one comment post, a reviewer asked multiple questions or covered multiple topics, then each separate topic was counted as a comment. Comments left by the author of a report or section were all counted as review comments (e.g., in-text remarks in a different font color, notes about edits to be made in future drafts, notes placed in a document to inform reviewers about data needing to be confirmed, or answers to reviewer questions).

In a close reading of only the review comments, we established whether the review comments addressed stylistic versus rhetorical aspects of the document. *Stylistic* aspects included language (wording, grammar, style), organization (local ordering of information), and presentation (format, layout, labeling). We characterized comments as stylistic if they were low-level or local as opposed to high-level or global; stylistic comments

addressed formal features of the document, invoking the traditional form–content dichotomy. In contrast, we characterized comments as *rhetorical* if they addressed purpose, logic, interpretation, evidence, and arguments made in the report. Rhetorical comments suggested adding, deleting, or changing content in order to correct or align intended meaning, support arguments, make the logic compelling, or deliver stronger claims, conclusions, or generalizations.

We recognized during the project that some review comments were editorial in nature and that certain edits (e.g., the addition or deletion of a specific word such as *generally* in front of the term *well tolerated*) significantly altered the meaning of a particular passage—any edit results in some change to meaning, and some of those changes are substantial. Our arguments that follow are predicated on the proportions of remarks that are edits versus comments and on the proportions of comments that are stylistic versus rhetorical. We think that these proportions make a compelling argument that reviewers often misplaced their attention by focusing on low-level stylistic features at the expense of attending to rhetorical concerns.

Changes to and Improvements in Draft Quality

We assessed the draft and final versions of each report based on the document quality standards that our consulting group had defined (see Appendix for the assessment rubric). We have used these quality standards for over 15 years, and we periodically revisit the standards and the criteria associated with each standard to refine the wording and to inform our shared understanding as a consulting group. We try to strike a balance in the wording between our client's language—the ways they talk about writing—and our own vocabulary, which is more informed by the research on professional writing.

We use this rubric to report the results of the document assessments that we do for various companies. Over time, we have compiled a database of assessments by document genre (e.g., a chemical synthesis report vs. a preclinical animal study vs. a CSR, which reports on studies in humans) so that we can offer clients a comparison of the quality of their reports versus the industry standard. We report numeric results on each of the seven standards (purpose, logic, context, content, organization, presentation, and language), and we organize our reports to the clients with generalizations and comments under each standard. These comments, which establish what is well done in a report and what should be improved, illustrate and give meaning

to the standards for the client. If we score a report low on the standard of purpose, for example, we produce a set of criticisms to demonstrate that the purpose of the report is not explicitly stated, is not carried through to the conclusions and recommendations, or fails to inform the content or organization of certain subsections of the report.

We frequently identify illogical or inconsistent interpretations, missing data, or unresolved issues, describing and exemplifying the problems using the language of the criteria for each standard. Thus, we do not simply judge a report to be unclear; we point to specific places where a reviewer would probably not be able to sort out just what was done or observed. If we say the logic is faulty, we point to places in the report where a warrant must be offered to support an interpretation of the data, or we identify gaps between what was planned for the study and what was actually carried out and reported. We understand that simply admonishing a writer to be clear or accurate or complete does little good, so in assessing a report's clarity of language or adequacy of content, we refer to specific places in the report, showing what we mean by providing examples that include suggestions for improvement. And because CSRs are regulated by health-authority guidelines, we are able to use these guidelines to support our judgment that content is not complete, accurate, or relevant.

Whenever possible, to increase reliability, we use two independent reviewers to assess a report. Each reviewer spends about 8 hours reviewing a document and preparing commentary to present to the client. In the cases reported here, we were only able to have one reviewer do the assessment and prepare commentary because the company was paying us to improve its review practices not to assess its documents. The assessments for the two CSRs reported here were conducted by a single consultant (neither of us) with extensive document, scientific, and regulatory experience. We draw on the quality scores and comments offered by this assessor in both case studies as our primary evidence for whether review processes improved drafts substantially and whether problems we identified in the early drafts were addressed in final reports.

We have received repeated confirmations from many companies (we have evaluated multiple reports for over 50 companies) of the usefulness of these assessment reports and the accompanying rubric. Typically, clients will comment that they do not understand how report weaknesses escaped the attention of their in-house reviewers. They are surprised at the weaknesses we find in the report's arguments and the inconsistencies of interpretation. And they are often surprised to learn that the report did not explicitly discuss results for all of the study's end points. They agree with us that the

conclusions offered often merely restate the results, and they recognize that the report does not convincingly respond to the regulatory issues that it raises. With many companies, we then go to the next phase of client engagement: defining review conduct and training scientists to review documents according to the quality standards.

Findings From Interviews

Our interview findings suggest that knowledge about good review practices was poorly distributed across the organization. We did not uncover effective plans for teams or function areas to inform reviewers sufficiently of their roles at the outset of the review process. CSRs were typically distributed with only an advisement regarding the deadline for returning review comments and a date for the team roundtable review session.

Most interviewees were aware of the company's standard operating procedure (SOP) on review practices. But the document describing the SOP for CSR review provides little useful guidance into effective review practices because the document largely represents only the who and when of review. The SOP does indicate that once a document is distributed, reviewers should identify and help resolve content-oriented (what we term *rhetorical*) issues rather than focus on stylistic or editorial issues. But the SOP does not offer examples to clarify this distinction. It stipulates that reviewer feedback should be submitted to the author prior to roundtable reviews, a way of working, we discovered, that is routinely ignored.

Reviewers did not typically distinguish strategic review (perspective-based reading in order to improve arguments) from quality-control review (inspection-based reading in order to fix errors). Interviewees suggested that they saw their role as identifying problems in documents but not necessarily offering solutions that would improve document quality. Their comments suggested that a lack of common expectations and uncertainty characterized authorship and review at every stage. Interviewees often stated that only through tenure with the company do people eventually learn about organizational expectations and adjust their practices to conform to these expectations. Interviewees described learning "what to do" during review principally through "trial and error" rather than via mentoring, guidance documentation, or training.

Interviews with many of the senior (in terms of tenure) reviewers and approvers suggested that judgments of quality were often predicated on what was done in previous documents or projects that received senior-management or external approval. In such cases, arguments were based

on "how documents have been done" as opposed to how content and arguments might best be managed in a particular document under review. Interviewees suggested that a common bit of advice proffered to new writers was to route them to previous project documents with the instruction: "Write your section of the report just like this one." The contributing authors were then left to their own devices to surmise what aspects of the report made it a model document worthy of emulation. A lack of shared, articulated standards of document quality thus hampered both authorship and reviews. We view this situation as a missed opportunity to mentor and socialize new workers, an important purpose of review (Katz, 1998).

Interviews suggested that the purposes and audiences for CSRs were neither well understood nor made salient during reviews. None of the interviewees suggested that a helpful role in the review process would be to act as a surrogate for the ultimate reader of the CSR, the external health-authority regulator. We found little evidence to suggest that issues were considered from the perspective of the primary audience. Often, issues were addressed to explain what an individual or part of the team did as a kind of accountability maneuver, to make sure someone or some department was "covered."

The interviewees generally estimated that the vast majority (over 80%) of their own review comments addressed matters of rhetoric or argument in the CSR—the intellectual content of the report. When explaining what they actually did during review, however, many interviewees described their actions as focused on assessing the accuracy of the data and the ways in which the data were presented and described in the text rather than assessing whether and how the data supported the arguments or messages being made in the document. We thus sense a lack of fit between their stated intent and their actual performance.

The Two Cases

The company operated from two principal locations, one in North America and one in Europe. Each clinical study was conducted by a team drawn up with people working at each of the sites. These teams were responsible for generating the CSR. Each team member represented a specific scientific or professional discipline. Team members were likely involved with multiple research studies, aligned within a specific product-development program or therapeutic area.

Case 1

This CSR described an early phase of research. Two draft versions were taken through review. The team, consisting of 10 members, reviewed only the body of the report, not the appendixes, which numbered in the hundreds of pages. In the final version of the report, the body consisted of 52 pages. After the medical writer released draft 1, the team had 5 days to review it; for the second draft, the team had 3 days. A roundtable review meeting followed each review period. Of the 10 invited reviewers, 6 participated in the roundtable: Four reviewers attended the first session, and six attended the second.

Both roundtable reviews were technology-facilitated meetings, mediated via an Internet-based conferencing tool to view the document and a telephone link to offer and listen to review comments. Reviews advanced on a page-by-page basis, with the team viewing the projected document in real time as the author, a medical writer, either made the changes endorsed by the meeting attendees in the displayed file or left a remark to be resolved later. The elapsed time to move from the first draft to the published version of the document was 8 weeks.

Findings on Categorization of Review Remarks. The overwhelming majority of the review remarks were edits as opposed to comments, and the majority of these comments concerned stylistic issues versus rhetorical issues. No comments centered on whether the document sufficiently addressed issues, followed an effective argumentative strategy, or achieved its purpose. A few comments directly addressed interpretation of study results. In total, we identified 352 remarks, 237 made on the first draft and 115 on the final draft of the report (see Figure 2). Of those remarks, more than 97% addressed style concerns: word choice, table design, or low-level edits. The few remaining comments addressed rhetorical considerations: added content, deletions, and changes in wording that affected interpretation of findings.

Observations on Review Practices. When circulating the drafts for review, the medical writer advised reviewers to review the document according to their areas of expertise. But the writer did not further advise reviewers regarding how to engage in the task of review, what parts of the document to examine, or who might have overlapping expertise.

During the roundtable review, the medical writer sought the team's endorsement for each page in turn in the body of the report (a total of 52 pages), asking reviewers if they had any concerns or questions related to the

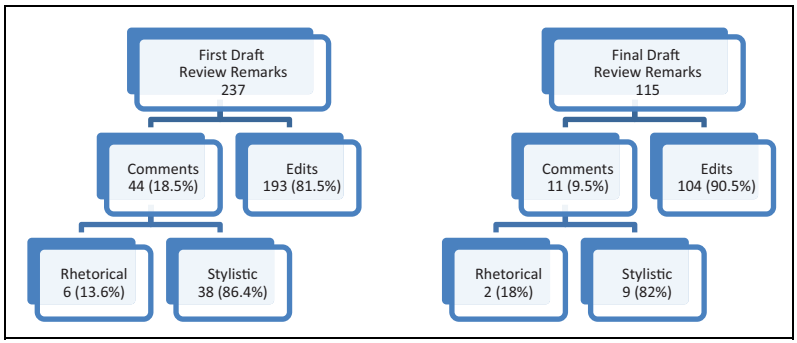

Figure 2. Categorization of review remarks for Case 1, first and final drafts, *n* (%).

currently projected page of the draft document. The reviewers principally narrowed their focus to the projected page, making little attempt at a more global evaluation. This approach appeared to encourage reactive versus reflective review remarks. Reviewer comments focused on inspecting stylistic and grammatical deficiencies at the page level instead of considering how messages and issues were handled within and across sections of the report. Reviewers appeared to be socialized to offer comments exclusively on what was wrong with a document because none of the comments addressed what a reviewer found as appropriate or well done in a draft.

The company's SOP states that to facilitate roundtable review, reviewers should either be familiar with the whole report or have their copy of the document and the consolidated review comments in front of them. Comments made by reviewers during both roundtables suggest that some attendees did not have the most recent version of the document in front of them and that some had not read the document completely before the start of the roundtable review session. The medical writer admonished the reviewers during both meetings in such terms as these: "If you had read it from the beginning, then you'd see that...." Lack of reviewer preparation likely contributed to the prevalence of stylistic comments and disagreements as opposed to intellectual, rhetorical, or content discussion.

Our analysis of interview commentary indicates that many felt that roundtable reviews failed to leverage their particular expertise because meetings often became bogged down in attending to minor stylistic matters. During the two roundtable reviews, we observed senior reviewers engaged in discussing mechanics and minor stylistic adjustments. For example, an extensive discussion focused on whether to describe data as clinically

"relevant" versus "meaningful" and whether to use the phrase "statistically significantly reduced" versus "statistically significant reduction." In spite of the directive in the SOP to focus roundtable reviews on content as opposed to style, the reviewers frequently digressed or engaged in long discussions about minor matters of language and format that greatly extended the length of the review meetings. The result was that some individuals attending the meeting solely to address one or two particular points in the document left to attend to other matters and then could not be found when their points of interest finally entered the discussion.

Additionally, we observed another disruptive behavior pattern: extensive and often lengthy side conversations that were largely unavailable to reviewers on the other end of the teleconference. At times the side discussions, combined with the ambient room noise, made it difficult to hear and follow the primary discussion topic. Our observations suggest that reviewers were not aware of how negatively such behavior affects efficiency in technology-assisted collaboration. The amount of time (more than 4 hours) required for each roundtable was complicated by cross-continental logistics, which made the need for efficiency and clarity a top priority.

Issues were rarely discussed from the perspective of the ultimate audience of the CSR, the regulatory agent. Principally, individual reviewers addressed issues from the perspective of what had been done in other teams in the past. The team spent considerable time clarifying for individual team members what the data represented (how it was collected, categorized, and measured). Little time was left for discussions of interpretation (i.e., what the data might mean or what their argumentative force might be). During roundtables, the teams did not reach any consensus on how they actually preferred to interpret the clinical significance of the research data.

More important, the team did not discuss how a skeptical regulatory agent might interpret findings and the consequences of such interpretations. For example, though discussed in both roundtable sessions, the relevance of the frequency of particularly important adverse events in patients receiving the test drug was left unresolved. The reviewers did not commit to an interpretive position. Instead, they decided simply to restate results in the conclusion section, repeating the data as opposed to offering an interpretation or conclusion.

Reviewers resolved conflict in one of three ways during discussion: (a) by retreating to precedents established in previous documents, for example, by saying "This is the way we did it in previous reports" and moving on; (b) by subordinating and minimizing the representation of the issue in the text, for example, by minimizing description of troublesome adverse events; or

(c) by ignoring the issue because they did not have a strategy or were uncomfortable with committing to a strategy for resolving the issue. Only rarely within the document or during the roundtable review did a reviewer raise an argument for how work might best be represented in the report, given the existing rhetorical situation. We rarely witnessed any appeal to audience, to how the regulatory reviewer would react to an unresolved issue, to the lack of an interpretive position on a study objective, or to the lack of supporting data on a stated conclusion. We also rarely heard comments framed within a rhetorical strategy, that a finding or complication should be represented in a certain way because doing so would build a stronger case for approval.

Quality Assessment of Draft and Final Reports. Our document assessment suggested that changes from the first to final draft led to only marginal improvement in quality, despite the number of changes requested during review and the extended roundtable discussions. Our assessor identified several major faults in the final report that also were present in the first draft and were not subsequently addressed by the review team in either round of review:

- The purpose of the research study was left unclear beyond simply reporting the results of the trial. While the study objectives—to evaluate the pharmacokinetics and evaluate the bioavailability of two formulations—were perfectly clear, readers were offered no insight into why the two formulations were being evaluated (i.e., what was the company trying to learn from this study? Did they want the formulations to perform in a similar manner or were they looking for the best formula to advance in further studies?). Without understanding why the study was performed, regulatory agents reading the report cannot possibly ascribe significance to the study findings. Without an explicitly stated purpose, readers cannot follow the logic from objectives to results to interpretive conclusions.
- The safety narrative in the CSR merely restated study data without drawing any conclusion. As is typical in most CSRs, this report included the evaluation of safety parameters as a major study objective. But to fulfill the study objectives, a report should offer high-level conclusions and a discussion of safety and tolerability. The narrative also failed to advance any argument about the relationship between drug safety and patient tolerability for the drug. In the Conclusion section of the report, only a single bulleted point addressed the study objective of safety.

Safety conclusions should have been presented and justified, especially in light of the fact that major, drug-related adverse events occurred in 40% of the subjects. Additional contextual information could have been included in the Safety Evaluation section in order to build support for the assertion in the Discussion section that the major adverse reactions seen in the study are common with this class of drug agents.

- The Conclusion section essentially repeated study findings without offering interpretive positions on clinical significance.
- The results reported in the Conclusion section presented values that were slightly different than the values in the Results section. Precision is a hallmark of a well-written research report, and such inconsistencies may, in the mind of the regulatory agent, cast doubt on the representation of other data in the CSR.
- Key research issues were left unresolved. The report introduction mentioned that the study would closely follow three adverse events of special interest, yet no further analysis or discussion was provided on the three events.
- The Results section of the report lacked deductive organization. Many of the subsections began, "Data are summarized in Table X and Table Y," or "Protocol deviations are listed by subject in . . . " Main messages were not positioned prominently but routinely appeared some three to four paragraphs into the subsection.

The assessor made many more comments under each of the quality standards, identifying problems with the report that should have been corrected between the first and final drafts. The preceding bulleted points underscore that the report was not responsive to the purpose standard—any of these shortcomings would undermine the report's authority and likely raise troublesome challenges from health authority reviewers. To allow such problematic issues with the report to persist through two review cycles without being identified, addressed, and resolved is a shortcoming of the review process.

Case 2

The contextual details regarding the corporate SOP for review apply to Case 2 as well. The CSR in Case 2 had two draft versions plus an additional review that covered only the rewrite of the Discussion section in the CSR. We collected formal review comments from 10 members of the review team and observed two roundtable review sessions.

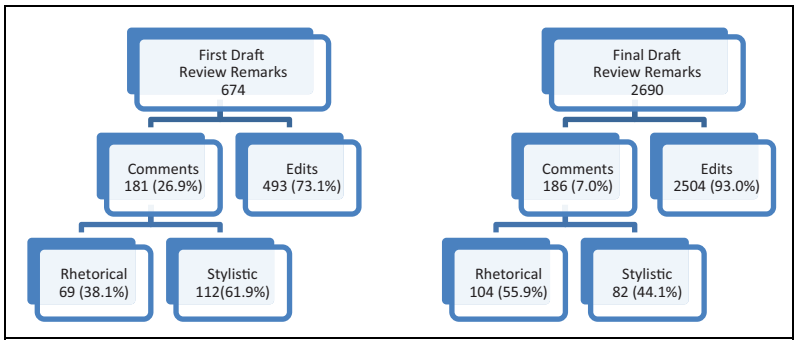

Figure 3. Categorization of review remarks for Case 2, first and final drafts, *n* (%).

As in Case 1, the review team focused on the body of the report, not the appendixes. The first and second drafts were 198 and 226 pages in length, respectively. A separate review of the Discussion section, a document that was four pages in length, addressed the substantial rewrite of this section by a senior team member. The elapsed time to move from the medical writer's release of the first draft to the published version of the document was 13 weeks.

As in Case 1, the review team in Case 2 used roundtable reviews to adjudicate outstanding review remarks and to provide reviewers who did not submit marked-up documents to the medical writer a forum to share their review remarks with the team. Here again, the method for review was to move through each page in serial progression, with the medical writer asking whether anybody in the convened group had comments. From time to time, the medical writer would identify a passage that had conflicting reviewer comments and suggest an approach to reconcile the comments.

Findings on Categorization of Review Remarks. For the first draft, there were six reviewers providing commentary. Four of the reviewers made their review remarks in the MS Word version of the document while the other two reviewers provided remarks as marginalia on hard copy. One reviewer also provided additional review remarks via an e-mail message. The review team made 674 review remarks on the first draft (see Figure 3). Of this total, 181 were either comments made in MS Word comment boxes or as marginalia on hard copy. Of these comments, 69 attended to rhetorical matters in the report.

Most of the review comments (112 of the 181) attended to elements of the document: completeness and accuracy of data, language or word choice,

and data presentation. Such concerns are important and addressed under organization, presentation, and language, the more formal standards in our document-assessment rubric. Here are examples of review comments that attended to stylistic elements:

> Write this statement like we did in the Study 754 shell.
> Put this sentence in a different paragraph.
> Should we call ourselves Sponsor or use [Company Name]?
> Make the columns in the table narrower.
> I'd break this up into two paragraphs.

In contrast to the many stylistic comments, no review comments addressed the overall argument, integrity of key messages, resolution of issues, or completeness of the logic trail. In terms of our document standards, the review comments tended not to address the large rhetorical issues of purpose, logic, context, and content.

The majority of review comments (108 of the 181) on this first draft were made by the clinical program director, who reviewed on hard copy. Many of his comments were queries to the medical writer about data and details on individual subjects profiled in the report. The medical writer provided handwritten responses to all 108 comments. As the following examples demonstrate, the program director's comments were challenging in tone and provoked a defensive reaction from the medical writer:

> Program director comment: Why do we have this table in the report?
>
> Writer response: Since it was part of the data programming, the team made the decision to keep it in the report.
>
> Program director comment: How was this related? How low was the platelet count in this subject? What was the dose?
>
> Writer response: Why do you keep asking these questions? I told you that details of this type are presented in the individual patient narratives.
>
> Program director comment: This table is missing details. Why do you not have descriptions in table?
>
> Writer response: If you had gone to last page of the table, you'd have seen these details appear as footnotes.

In the roundtable review, many of the comments focused on word choice, phrasing for table titles, or formatting (how to display a table on the page). Here are some examples the reviewers discussed:

- whether to use the term *application* or *administration* with regard to a therapy
- whether to say correlation coefficients were generally *negative* or to select a different word
- whether to change phrasing to accommodate this reading: "From a stylistic standpoint, I would take out the phrase *The onset of* . . . as you are talking about time to onset."
- whether to accommodate this review comment: "Can you insert a line between SOCs to make it more pleasing to the eye?"
- how to respond to a comment that tables need better "headers"

Discussions of word choice are clearly connected to meaning, as in the following example, which raised discussion of terms: "Adverse Events Leading to Deaths (AEDs) in subjects experiencing difficult to control partial seizure." One reviewer did not want to state *difficult to control* and suggested using the term *pharmacoresistant.* Others felt that using the term *pharmacoresistant* would alter the intended meaning of the passage and rejected the change. Such stylistic changes are clearly important, but they are also at a fairly low level in terms of the overall arguments in the document—their rhetorical force is local, not global. Occasionally, a reviewer would attempt to keep simple stylistic changes separate from the roundtable review: "Biostats requests some wording changes—will give them to medical writer after roundtable." Typically, however, low-level suggestions tended to occupy the attention and time of everyone at the roundtable.

The second, or final, draft, had eight reviewers; four of the eight had also participated in the first-draft review. Some major rhetorical issues raised by reviewers in the first draft regarding how to interpret findings remained unresolved in the second draft. The team was apparently struggling with how to characterize study findings on the lower of two drug doses used in the study because the efficacy results for this dose did not achieve significance.

The review team made extensive edits on this second draft (2,504 of the 2,690 remarks were edits; see Figure 3). The majority of these edits were made by three reviewers. The following example is typical of these edits:

> Subject 116407, a 24-year-old male randomized to the ABC 400mg/day treatment group, was discontinued during the Titration Phase due to an increase in QTc greater than 60ms between Visit 3 and Visit 4. This QTc result was not confirmed by the central reader.

The edit changed the passage to read as follows (change underlined):

> Subject 116407, a 24-year-old male randomized to the ABC 400mg/day treatment group, was discontinued during the Titration Phase due to an increase in QTc greater than 60ms between Visit 3 and Visit 4. This protocol directed action was taken prior to confirmation by the central reader.

This edit is important because it makes clear that the discontinuation was in accord with the study design's requirements—that it was per protocol—so that the reader would not infer that a deviation in study conduct occurred.

A second example of a typical edit leads to minor improvement:

> There were 6 subjects in the ABC 200mg/day treatment group . . . and 1 subject in the ABC 400mg/day treatment group . . . that had normal alkaline phosphatase values at Baseline that shifted to high at end of Treatment.

The edit changed the passage to read as follows (change underlined):

> There were 6 subjects in the ABC 200mg/day treatment group . . . and 1 subject in the ABC 400mg/day treatment group . . . that had normal alkaline phosphatase values at Baseline that shifted to above normal (high) at end of Treatment.

More than half of the review team's comments on this final draft were rhetorical (104 of the 186). The following example represents a rhetorical issue that was appropriately raised and addressed during the roundtable. A reviewer wondered about the validity of a claim in light of other data:

> There does not appear to be a withdrawal effect associated with reducing the dose of ABC by 200mg/day/week.

Ultimately, the claim is recast and rewritten as follows (change underlined):

> Based on analyses for increased seizure frequency and AEs reported during the Taper Phase, there does not appear to be a withdrawal effect associated with reducing the dose of ABC by 200mg/day/week.

This comment and change align specific data with a generalization, resulting in an improved argument; admittedly, the change is at a local level in terms of argument strategy within the CSR.

An additional review of the Discussion section was made by five of the team members. Their objective was to address issues and rhetorical positions left unaddressed at the end of the final round of review. A senior project team member made a number of substantial edits intended to satisfy these communication weaknesses (see Figure 4).

Our analysis showed that these reviewers engaged in a meaningful attempt to address some of the communication weaknesses found in the

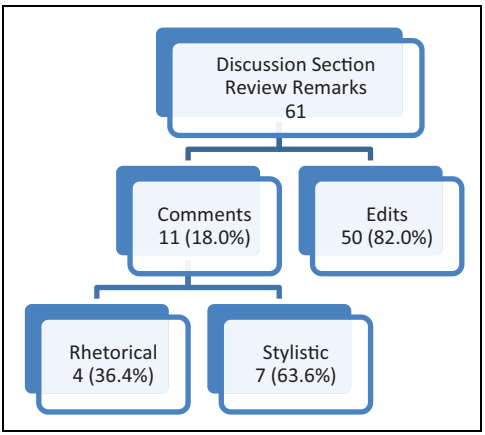

Figure 4. Categorization of review remarks for Case 2, review of report discussion section.

discussion section, but they also placed extensive effort into editing passages previously reviewed and deemed acceptable by the full team. Of their 61 review remarks, 50 were edits placed in sections that had been previously reviewed by the full team. In our opinion, many of these edits did little to alter the communication quality of the particular passage. Many edits involved deletion and insertion of words:

- The phrase *change in sample size* was edited to *sample size increase*.
- The phrase *seizure frequency reduction* was edited to *reduction of seizure frequency*.
- The phrase *may be due to an increase* in was edited to *may be from an increase in*.

The majority of review comments addressed stylistic elements, typified by these comments:

> Switching these sentences flows better.
> Add the word generally here since next paragraph uses this modifier.
> Remove since this wording is not used in the previous paragraph.

In contrast, there were several comments that attempted to consider communication quality:

I think the comment re: reduced effect size is confusing because . . . I suggest to drop and just describe the data.

I think this is a more neutral comment re. results than the word *variability*.

Using *significant* implies statistical testing. I would reword . . . .

Observations on Review Practices. During these roundtable reviews, organizational goals, document purpose, and audience expectations were never explicitly described or discussed. Reviewers were clearly working with an eye to the future—wanting to hold off on committing to an interpretation of the study data until data from other studies had been finalized and compared with these data. The reviewers raised some meaningful topics, but these topics were not effectively addressed during the roundtable reviews.

Our observations of both roundtable review sessions suggest that some reviewers were reading the document for the first time during the course of the roundtable meeting. This unfamiliarity with the document raised the noise-to-substance ratio because it created a need for lengthy side discussions (often involving only two individuals) to either discuss study results or inform an individual about reporting expectations.

Both first and final drafts went into review with substantial holes: missing data, absent analyses, and incomplete interpretations. The ongoing data analysis led to many comments during the second roundtable review that, for other projects, would normally have been addressed and settled at a much earlier time point. During roundtable discussions, reviewers tended to raise issues of content accuracy and completeness (which required specific expertise in the statistical conduct of clinical studies available to only two or three people in the review) rather than address substantive rhetorical issues. Attempting to resolve these study conduct issues distracted reviewers from attending to the rhetorical demands of the draft. Queries about the accuracy of data are important, but such queries should occur before the roundtable review session so that time can be spent on strategy as opposed to data verification.

A principal driver in the reviewers' decision-making process was precedent—how certain information was represented in previous documents. For example, a reviewer commented that a paragraph narrative following a table merely repeated what was already represented in the table; therefore, it was redundant and should be deleted. But this recommendation was refuted by a senior reviewer with the response, "No, do not change this as all our CSRs do this." Another reviewer commented, "This introduction does not read *exactly* like the 786 study. It is better if it reads the same." As

in Case 1, then, this team tended to appeal to prior practice to justify revisions rather than raise questions of rhetorical strategy.

Quality Assessment of Draft and Final Reports. Our document assessment of the first and final drafts found substantial quality improvement in the final draft. But most of these improvements came from the addition of content and could not be attributed to review remarks and suggestions made during the roundtable review meetings. In other words, the writer (likely with input from key team members) improved the draft, but not in ways we could track to reviewers' remarks on the document and during roundtable review sessions.

The assessor found some major communication flaws in the final report that were not addressed at all during the review process. The greatest weaknesses of the report were (a) unaddressed issues related to study conduct and (b) inadequate explanations of negative findings on a key statistical end point. The CSR stated that 21% of subjects had a deviation from study conduct and design criteria defined in the protocol, and nearly 11% of subjects had a major deviation from these criteria. The study enrolled some subjects who were not defined per protocol as acceptable, others who failed to receive proper test-drug kits, and still others who violated one or more of the study conduct rules. These types of study deviations are widely recognized as points of concern to regulatory agencies because the enrolled study population may not truly represent the population targeted in the protocol. The report also identified a large number of study subjects with implausible laboratory test results for blood levels of the test drug. Some subjects on the placebo arm were identified as having the test drug in their blood plasma samples; correspondingly, some subjects on the active-drug arm presented plasma samples that showed no trace of the drug. Additionally, some subjects on the active-drug arm were identified as having the drug in their blood plasma even before the initial doses of the drug were distributed. Such findings call into question the conduct of the study. The report goes into considerable detail in representing the incorrect results and suggests that the implausible results were due to an error in the lab's sampling process. Such an error may be of considerable concern to a regulatory agent because it can call into question the overall study controls and ultimately the interpretation of the data.

In terms of report quality, the team did a good job in descriptively representing the deviations and the implausible lab values. But the report fell short on explaining how such deviations and implausible values could occur and why, despite these events, the study was indeed well controlled. During

the document-review process, none of the reviewers commented on how these issues in the CSR should be addressed. The team's only response to these problems seems to be the reiteration of the following statement in the Discussion and Conclusion sections of the report: "The study was determined to be adequate and well-controlled." This statement appears four times over four pages. The team may indeed have been trapped by the poor study conduct, but the claim of control is neither supported by the data nor backed by an explicit warrant. Bluntly repeating an unsupported generalization does not make an effective argument. We reexamined the review comments for both drafts of the CSR and found that no team member wrote a comment regarding how a regulatory agent may react to the issues regarding study conduct.

The report also did not adequately address its negative findings for a key statistical end point. The Discussion section describes the statistical outcomes for all end points except one. Instead of openly recognizing the failed statistical test, the report describes the outcome as follows: "Drug ABC 200 mg/day showed a numerically improved treatment difference over placebo." It then goes on to conclude, in what is essentially boilerplate text used in many reports, the following: "This adequate and well-controlled trial supports that Drug ABC 200 mg and 400 mg are each effective." This statement is difficult to support, given both the implausible lab data and the negative statistical result for a key study end point. We did not observe any reviewer express concern with how the company would defend this claim if challenged by the regulatory reader, which would be highly likely. We were surprised that multiple rounds of review did not uncover the need for an interpretive position that would satisfy a regulatory audience.

As in Case 1, the review team made decisions about how to address study issues or communication weaknesses by either following the precedent of what had been stated (or not stated) in previous reports, minimizing the representation of the issue, or ignoring the issue. Our document assessment found that the language in the report was of good quality, which is an expected outcome given the team's extensive focus on fine-tuning the report's phrasing and sentences. Unfortunately, the review team neglected the deeper issues of interpretation and argument that persisted throughout all drafts.

Discussion

In this study, we examined document-review practices within a single pharmaceutical company. In addition to interviewing participants, we observed review sessions and collected and analyzed review remarks on the first and

final drafts for two CSRs. We also evaluated the first and final drafts to assess and compare document quality.

To summarize our findings, we found that far too much time, attention, and commentary were directed toward low-level or local features of the reports and that far too little time and attention were given to constructing compelling arguments and addressing issues associated with study conduct or results. In spite of voluminous commentary during roundtable reviews, the quality of the reports did not seem to improve as a result of the review activities. Certainly, the reviewers corrected statements, adjusted data, and rearranged information in ways that were important and made sense to them. But what was lacking was focused attention on developing clear claims and supporting arguments while attending to important issues of interpretation.

Also lacking was a sense of audience that would have helped the reviewers think through the kind of work these reports must accomplish. These reports face a highly skeptical audience of health authority agents; these scientists read against the report, seeking inconsistencies, unresolved issues, and unsupported conclusions. Overall, the company reviewers were inclined to see their roles as inspecting drafts to identify mistakes or problems of data representation and narrative style rather than improving the overall effectiveness of the reports by considering the purpose, audience, and particular situation of the given study.

The data we collected describe review processes that are dominated by attention to detail, with the overwhelming preponderance of the remarks on the documents and the discussion during review sessions being concerned with very local decisions about wording and correctness as opposed to rhetorical strategy. Obviously, all classification categories leak to some extent, and whether a comment was stylistic or rhetorical was not always a clear matter. In all likelihood, we mischaracterized some comments, and some comments probably do double duty, such as a stylistic edit that has rhetorical implications. But the sheer volume of remarks across multiple reviews makes the patterns in the data quite clear because edits and stylistic comments so greatly outweighed any remarks that we could comfortably classify as rhetorical.

In these two cases, the practices that informed review activities tended to direct attention to low-level inspection as opposed to high-level strategy. Although the company had an SOP with some good recommendations for conducting reviews, this guidance document did not seem to affect practice. The coordinating medical writers tended to underplay their opportunity to focus the review, establish roles, and predetermine where they could best place their energy. Some reviewers arrived unprepared, apparently looking

at the draft for the first time during the review session. The common practice of going through the report page by page encouraged reviewers to focus on local edits as opposed to global revisions, often resulting in their spending excessive time on low-level corrections as opposed to strategic or rhetorical thinking. Uses of technology, both Web-meeting software and review mark-up tools, did not improve the process and often caused their own difficulties. Marshaling technologies to support review practices is possible, as demonstrated by Swarts (2004), but in the cases we observed, the technologies did not lead to efficient or effective reviews. Possibly, the mark-up tools themselves tend to favor minor corrections and local comments rather than global comments on the effectiveness of a document for a given purpose, audience, and situation.

We recognize that a document may need more editing comments than rhetorical comments over the same number of pages and that editorial best practice is to note every instance of a problem (e.g., inconsistent use of a term or a missing comma) whereas a recurrent rhetorical issue might be addressed only once. This difference between editing and commenting perhaps accounts for some of the large discrepancy in the number of edits versus rhetorical comments. But we believe the real problem stems from confusion regarding the proper roles of the subject-matter expert reviewer and the document author or editor. Certainly, documents should be correct and consistent in all respects, but devoting the attention of a full team of experts to minor details is inefficient and quite costly in several respects. We believe that reviewers need a more a refined mental model for how to review drafts of documents—how to concentrate on the big-picture rhetorical concerns and set aside edits and stylistic corrections. Through observation and interviews, we have found that most reviewers work toward one mental model of a CSR—the final published version—with the primary goal of inspecting for correctness. Assessing the details of a document to ensure their correctness is not the same mental activity as assessing the argumentative or rhetorical effectiveness of a document.

We have not emphasized here the general sense of frustration that review processes engender, though this point has been made in the literature. One of our reviewers referred to the process as "torturous" while another noted that roundtable reviews frequently stretched on for more than 5 hours. While European participants excused themselves to go home late at night, U.S. reviewers continued to plow through the document page by page.

Pointing to the general sense of frustration identified in the literature on review practice, van der Geest and van Gemert (1997) reported that subjects in various Dutch companies across three studies found reviewing to be

"cumbersome" (p. 445). We certainly saw similar patterns in our study. We also saw many of the problems that Couture and Rymer (1991) identified in their study of review in an architectural firm. They suggested that managers and writers may have "vastly different perceptions about the function of discourse interaction" (p. 106). In their study, managers saw the writing-review session as an opportunity to correct a writer's mistakes while writers expected reviews to be informative. In our study, everyone seemed to see review as corrective rather than as an opportunity to inform and realize the potential rhetorical force of a draft report.

It would be difficult to measure specifically how the CSR documents discussed here functioned in the hands of regulatory health agents, given that the individual reports are compiled and submitted in a dossier with other clinical, chemical, and preclinical research reports. We do know that the two drug products studied and reported in these CSRs were approved by health agencies for marketing. We do not have access to information concerning whether the rhetorical shortcomings that we identified in our document assessment were raised by the health authority agents.

We do, however, have the reactions of a number of scientists at the company who have participated in document-review training. We recently used the report from Case 2 for training purposes. The participants agreed with our characterization of the problems in this report, particularly its shortcomings in demonstrating that the study was well controlled, and they understood the seriousness of such shortcomings. The training participants concurred that the report left open the question of why such a high percentage of subjects were misenrolled, and they agreed that the implausible lab values would likely cloud a reader's judgment of study outcomes unless those values could be explained convincingly. Many were certain that a report with such deficiencies would indeed invite inquiry from the health authorities. Sharing with clients our analysis in such a manner confirms our own interpretation, a distinct advantage of this sort of action research. Recognizing the faults in their own recent reports and gaining perspective on the limits of their own review practices were also strong incentives for them to attend to the training and pursue improved practices.

In our training, we encourage teams to focus during review on troublesome issues that are likely to raise questions or challenges from health authorities. Once these issues are articulated, we suggest that teams should make certain that their documents make the best possible case, being careful to link conclusions to supporting data and provide warrants that connect data to claims. This process requires some imaginative projection: The researchers inside the company must imagine the

reviewing behavior of a health authority agent. Although taking such an audience-focused approach is evidently not an easy or typical behavior for review teams, it can be an outcome of training once teams gain a realistic perspective on their current practices and are given some ideas and practice in new ways of working.

We do not believe that our findings for this one company are exceptional. Consider the data shown in Table 1, which we collected recently from another large pharmaceutical company. In this case, a massive effort took a report through 10 separate reviews, generating 5,485 remarks in total. Even in the 10th round of review, 550 remarks were offered on the near-final draft. As Figure 5 shows, a low percentage of these remarks focused on rhetorical changes (always <10% and often <2%). The majority of review remarks suggested low-level stylistic changes, either marked directly on the text or offered with comments to justify the edits.

Revising Our View of Document Review. Companies need to become more evaluative and methodical in their own work practices, in part because the costs associated with planning, authoring, and reviewing individual research reports are substantial. If we consider the time and costs of authoring, reviewing, and publishing, we estimate that a final report might range in cost from $50,000 to well over $200,000, in addition to the direct costs of running a clinical study. Although the costs vary widely according to study complexity and size, we estimate a cost range between $700,000 and $3.5 million for a clinical study. Thus, these companies have much at stake in the review process.

Because we have witnessed and documented what we believe are ineffective and inefficient review practices, in our training and consulting practice, we recommend specific actions to encourage effective review practices:

- Articulate and rely on defined document quality standards and SOPs as guidance for executing effective reviews. Privilege shared standards over individual preferences.
- Define the scope, purpose, audience, data displays, and argumentative strategy for the document before drafting. In turn, use this early document planning to guide reviews.
- Define reviewer roles and responsibilities, acknowledging unique and strategic expertise. Involve specific reviewers for specific purposes. Explain to all reviewers what their roles are, where they should focus their energies during review, and what the differences are between being a reviewer versus being an editor.

Table 1. Remarks (Comments and Edits) per Draft of a Report at Another Pharmaceutical Company

Document Draft Versions											
Section	1	2	3	4	5	6	7	8	QC	Reg.	Total by Section
Company signature page	0	11	0	0	0	0	0	0	2	0	13
Investigator signature page	0	1	0	0	0	2	2	0	0	0	5
Report title page	2	5	18	10	16	9	11	0	0	1	72
Report synopsis	5	20	62	57	58	86	73	2	2	6	371
Glossary of abbreviations	0	1	0	1	0	3	4	0	47	30	86
Trademark information	0	0	0	0	0	1	1	0	1	1	4
1 Study admin	0	0	0	0	0	5	4	1	1	0	11
2 Ethics statement	0	1	0	0	0	0	0	0	0	0	1
3 Study objectives	0	5	0	1	0	2	2	0	0	0	10
4 Study rationale	0	8	9	6	0	8	7	0	0	0	38
5 Diagnosis & criteria	0	0	2	0	3	6	9	0	0	0	20
6 Treatment admin	0	2	0	1	0	3	2	1	1	1	11
7 Study procedures	0	7	2	11	7	20	20	5	1	3	76
8 Methods	12	4	2	14	11	21	17	2	1	0	84
9 Study population	5	29	88	38	25	16	37	0	41	33	312
10 Safety results	14	119	765	578	423	98	153	3	272	370	2,795
11 Pharmacokinetic results	1	59	531	186	47	12	23	0	21	15	895
12 Pharmacogenetic	0	10	0	0	0	0	0	0	0	0	10
13 Efficacy results	3	13	26	41	32	9	20	0	1	21	166
14 Discussion/conclusions	0	7	9	52	62	60	81	0	1	23	295
15 References	1	6	0	1	3	0	5	0	0	1	17
16 Patient narratives	2	3	3	64	41	0	16	0	3	42	174
17 Tables and listings	0	0	0	3	3	0	4	1	1	2	14
Modular appendixes	0	0	0	3	0	0	1	0	0	1	5
Total by draft	45	311	1,517	1,067	731	361	492	15	396	550	5,485

Note. QC = quality control review; Reg = regulatory affairs review Individuals from these departments reviewed the draft following the eight team reviews.

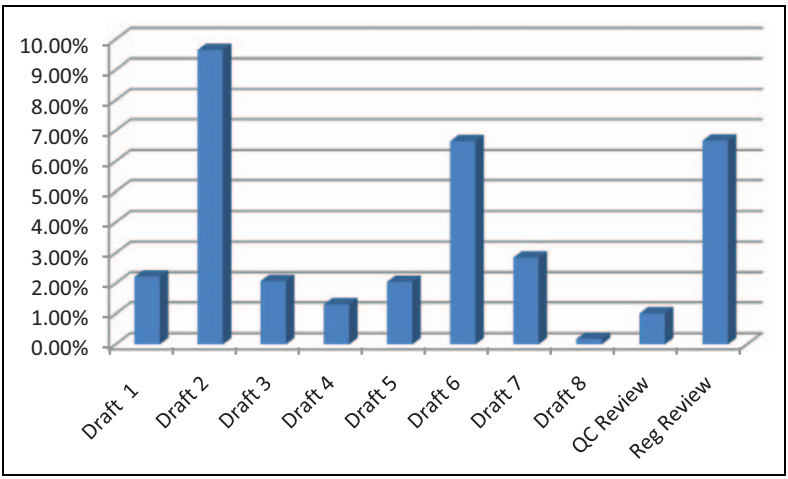

Figure 5. Percentage of rhetorical comments out of total review remarks per draft of report of another pharmaceutical company. *Note.* QC = quality control review; Reg = regulatory affairs review. Individuals from these departments reviewed the draft following the eight team reviews.

- Identify objectives to focus each review. Reviews might focus on the logic of an argument and its supporting data, on critical issues of interpretation, on key data displays, or on the limits of generalization with regard to statistical and clinical significance. Problems with word choice, style preferences, transcription accuracy, and format should be directed to the writer or editor but not made the focus of review.
- Train review roundtable facilitators to establish and follow a meeting agenda that encourages strategic review as opposed to a page-by-page protocol that largely encourages editing.

As a result of the work reported here, as well as similar work in other companies, we have changed our consulting and training approach to make use of pooled information from several organizations. We can extend our characterization of review practice by adding quantifiable information about where people spend time in a document and how much energy is given to editing as opposed to high-level rhetorical commentary. The use of quantitative data helps support an authorial voice in training sessions. We talk about the impact of those conditioned behaviors (i.e., to inspect and correct) associated with typical review performance, and we stress the mental discipline it takes to review a document at a strategic level. We also

describe the author's challenge of having to wade through a large volume of edits and comments to create a next draft version. Based on the data we have collected, we now characterize what people do when examining documents as *editing*, *inspecting*, and *reviewing*, and we describe various tactics and techniques that can be applied to support efficient practices. Specifically related to CSRs, we now deploy review checklists with heuristic prompts related to each section of the report in order to guide reviewers and encourage them to look at global features of the document rather than pursue a low-level, page-focused analysis. As we have always done, we encourage clients to adopt document standards and to shape their thinking and commentary to this rubric.

Sharing results from these investigations with clients has helped. The most noteworthy change as a result of our review assessments is that people now appreciate that there is a significant difference between their perceived efforts and their actual efforts. Although conditioned behaviors change slowly, reviewer discipline has improved. Performance varies in relation to the adaptability of project teams. Teams with strong leaders who embrace the concepts associated with improved review practices often show dramatic improvement. Document-development teams now recognize that a critical performance factor is to shape group expectations regarding the review process before starting the review. At the company described in this study, teams often use the roundtable reviews as the first point in time in which the team truly collaborates on strategy.

The other company we studied (see Table 1 and Figure 5) has now deployed various metrics to track review performance. The most obvious metric is the number of drafts required to reach a final document. Before training, the company averaged six drafts and 18 months to document finalization; following training of 300 employees, the company averages three drafts, with some projects actually reaching a finished document with two rounds of review, the stated organizational target.

We do not believe that changing nonproductive, conditioned, inefficient practices is an easy matter, or companies would have already done so. We do believe that recognizing nonproductive review practices and understanding the causes for such practices should be an object of focus for more organizations. We understand that collaborating to develop complex documents with sound arguments involves difficult cognitive and social practices. But if a company establishes the goal of producing quality documentation through efficient and effective review practices, it will find that it must do a lot of work to counter the ingrained tendencies of review teams to focus on low-level stylistic edits as opposed to high-level rhetorical concerns.

Appendix

Document-Assessment Rubric

Standard	Criteria
Purpose	The document is purposeful at all levels: • The purpose of the document (and major sections) is clear. Explicit statements of purpose early in the document or section help establish a clear sense of purpose. • The document clearly achieves its purpose through adherence to all of the document standards.
Logic	The document presents a clear logic trail that makes the report coherent: • The goal of the reported work is clear. • Objectives (specific and measurable tasks) are explicitly stated, preferably early in the document. • Variables are defined, along with the standards against which they were assessed. • Results lead through strong arguments to clearly stated conclusions. • Conclusions are explicitly related to objectives. The practical, medical, and scientific significance of the work is clear and confirms achievement of the goal of the work. Arguments are formed with claims that are supported by evidence and justified by reasons based on appropriate warrants and qualifiers. Throughout the document, decisions are justified with explicit rationales.
Context	Background information (important medical, scientific, or technical context) provides a framework so the reader understands the work: • The *specific* expertise that shaped the work is conveyed (the local knowledge, developed within the company, that helps an external reader make sense of the development work). • The place of the work within the overall development project is clear and information is linked to other development studies or functions (i.e., clinical to preclinical, CMC to clinical) as appropriate. Information is linked to the scientific literature as appropriate.

(continued)

Appendix (continued)

Standard	Criteria
Content	The information in the document is • complete—includes everything necessary and sufficient • relevant—connects logically to the purpose • accurate—contains no factual errors • precise—uses appropriate and consistent degrees of exactness Issues are identified (preferably early in the document) and resolved with clear messages in appropriate locations. There is a good balance between general "big-picture" information and detail. The document conforms with relevant regulatory and corporate guidelines.
Organization	Main messages, summary statements, or conclusions appear at the beginning of sections and paragraphs. Document sections, especially Introduction, Results, or Discussion, are organized deductively, from general to specific, from most important to least important. Other sections are well organized, for example: • Procedural steps are sequential. • Safety analyses may have a conventional order (as dictated by the World Health Organization). Organizational devices guide the reader: • Table of Contents • Usable headers, footers, and pagination • Numbered, informative headings • Explicit internal and external references
Presentation	Information is presented in the most appropriate form (text, table, or figure) to help reader access and comprehension: • Page layout is inviting and makes information easily accessible. • Format (bullets, boldface, etc.) emphasizes important information. • Visuals (tables and figures) display, emphasize, or clarify important points. Visuals are introduced in the text with comments that focus readers on key elements. Each visual should do the following: • convey a clear message • use effective design, free of visual clutter • have a clear, informative title with helpful numbering • contain clear and useful labels, legends, and interpretive comments • stand alone

(continued)

Appendix (continued)

Standard	Criteria
Language	Sentences use strong subject–verb–object patterning, with effective subordinate and parallel structures, to deliver clear, unambiguous meaning. The language is correct, reflecting standard conventions of contemporary English (grammar, punctuation, and spelling). Terms are used consistently. Specialized terms and acronyms are defined, either in text or, preferably, in a glossary table.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

References

Bernhardt, S. A. (2003). Improving document review practices in pharmaceutical companies. *Journal of Business and Technical Communication*, *17*, 439-473.

Bernick, P., Bernhardt, S. A., & Cuppan, G. P. (2008). The genre of the clinical study report in drug development. In B. Heifferon, & S. Brown (Eds.), *Rhetoric of healthcare: Essays toward a new disciplinary inquiry* (pp. 115-132). Cresskill, NJ: Hampton Press.

Clark, D. (2004). Is professional writing relevant? A model for action research. *Technical Communication Quarterly*, *13*, 307-323.

Couture, B., & Rymer, J. (1991). Discourse interaction between writer and supervisor: A primary collaboration in workplace writing. In M. M. Lay, & W. M. Karis (Eds.), *Collaborative writing in industry: Investigations in theory and practice* (pp. 87-108). Amityville, NY: Baywood.

Fossati Wood, L., & Foote, M. A. (2009). *Targeted regulatory writing techniques: Clinical documents for drugs and biologics.* Basel, Switzerland: Birkhauser Verlag.

Henry, J. (2000). *Writing workplace cultures: An archaeology of professional writing.* Carbondale: Southern Illinois University Press.

Katz, S. (1998). *The dynamics of writing review: Opportunities for growth and change in the workplace.* Vol. 5 in ATTW Contemporary Studies in Technical Communication. Stamford, CT: Ablex.

Lang, T. A., & Secic, M. (1997). *How to report statistics in medicine: Annotated guidelines for authors, editors, and reviewers.* Philadelphia, PA: American College of Physicians.

Paradis, J., Dobrin, D., & Miller, R. (1985). Writing at Exxon ITD: Notes on the writing environment of an R&D organization. In L. Odell, & D. Goswami (Eds.), *Writing in nonacademic settings* (pp. 281-308). New York, NY: Guilford.

Piantadosi, S. (1997). *Clinical trials: A methodologic perspective.* New York, NY: Wiley.

Shwom, B. L., & Hirsch, P. L. (1994). Managing the drafting process: Creating a new model for the workplace. *Bulletin of the Association for Business Communication*, *57*, 1-10.

Swarts, J. (2004). Technological mediation of document review: The use of textual replay in two organizations. *Journal of Business and Technical Communication*, *18*, 328-360.

Van der Geest, T., & van Gemert, L. (1997). Review as a method for improving professional texts. *Journal of Business and Technical Communication*, *11*, 433-450.

Bios

Gregory P. Cuppan is the owner of McCulley/Cuppan, LLC, a firm that provides consulting and training services to over 30 companies in the pharmaceutical industry.

Stephen A. Bernhardt is a professor of English and the Andrew B. Kirkpatrick. Jr., Chair in Writing at the University of Delaware.