Spreadsheet Errors and Decision Making: Evidence from Field Interviews[a]

Jonathan P. Caulkins,[1] Erica Layne Morrison,[2] and Timothy Weidemann[3]

Abstract

There is consensus in the literature that spreadsheets are both ubiquitous and error-prone, but little direct evidence concerning whether spreadsheet errors frequently lead to bad decision making. We interviewed 45 executives and senior managers/analysts in the private, public, and non-profit sectors about their experiences with spreadsheet errors and quality control procedures. Differences across sectors do not seem pronounced. Almost all respondents report that spreadsheet errors are common. Most can report instances in which the errors directly led to losses or bad decisions, but opinions differ as to whether the consequences of spreadsheet errors are severe. Error checking and quality control procedures are in most cases informal. A significant minority of respondents believe such ad hoc processes are sufficient because the "human in the loop" can detect any gross errors. Others thought more formal spreadsheet quality control processes could be beneficial.

Keywords: spreadsheets, errors, error control, managerial decision making, decision support systems, end user computing

[1] Corresponding Author: Jonathan P. Caulkins, Carnegie Mellon University Qatar Campus, Qatar Office SMC 1070, 5032 Forbes Ave., Pittsburgh, PA 15289. Caulkins@cmu.edu
[2] Business Consulting Services, IBM Global Services, 12902 Federal Systems Park Dr., Fairfax, VA 22033-4421, elaynemorrison@us.ibm.com. (412) 983 4797.
[3] Deloitte Consulting, (717) 526 0430, x5740, tweidemann@deloitte.com.

1. Introduction

The information revolution has provided leaders with many powerful decision support tools, but none is more frequently used than the familiar spreadsheet. Spreadsheets have made enormous contributions to decision making by democratizing analysis and speeding decision-cycles.

Field audits and laboratory studies consistently find, however, that a very large proportion of spreadsheets contain errors, although the precise frequency of errors may vary by spreadsheet type, task and development context. Given this, one might expect that (1) spreadsheet errors frequently lead to poor decisions that impose tangible costs and (2) concerned organizations would invest in more quality control procedures governing spreadsheet creation and use. However, there is little empirical evidence from practitioners to support either proposition.

We investigated both hypotheses through 45 semi-structured interviews with executives and senior managers / analysts in the public, nonprofit, and private sectors. Respondents were recruited from a convenience sample of people deemed likely to be cooperative and forthcoming.

Semi-structured field interviews have strengths that complement those of laboratory studies and spreadsheet audits. Laboratory studies inevitably raise questions of external validity; do their results pertain only to laboratory exercises or also to business practice? Auditing spreadsheets can overlook something that we discovered to be crucial. Spreadsheets inform decisions but rarely make them, so errors in spreadsheets do not inevitably lead to errors in decisions. To understand the impact of spreadsheet errors on decision making, it is important to think about the decision making process more generally, not just the spreadsheet artifact.

Our research approach has two limitations. First, the respondents are a convenience sample. Second, self-report data can be flawed, whether through imperfect memories or conscious deception. Given these limitations, we focus on broad qualitative conclusions.

The rest of this paper is organized as follows. Section 2 reviews some literature relevant to spreadsheet errors. Section 3 describes our data and methods. Section 4 discusses results pertaining to spreadsheet use, frequency of errors, reported effects on decisions, and error control procedures. Section 5 discusses implications for practice and future research.

2. Review of the Literature

The first major question motivating this research is, "How often do spreadsheet errors lead to major losses or bad decisions?" Two literatures are relevant to this question. One notes that spreadsheets are widely used (Gerson et al., 1992) in diverse domains, ranging from familiar manufacturing and supply chain applications (e.g., Buehlmann et al., 2000) to the sciences (e.g., Jain et al., 2000) and social sciences (Rydell and Everingham, 1994) and for many purposes, including support of high-level managerial decision making (Chan and Storey, 1996). Indeed, spreadsheets have become such a force in decision making that they generated a revolution in Management Science instruction in business schools (Plane, 1994; Grossman, 1997, 2001; Powell, 1997, 1998).

The second pertinent literature documents the frequency of spreadsheet errors. In a dozen studies reviewed by Kruck et al. (2003), the (simple) average proportion of spreadsheets with errors was 46%. A definitive synthesis of this literature by Panko (1996, 2000, 2005) suggested that even those numbers might understate the problem, since not all audits find all errors. Out of 8 studies auditing a total of 421 spreadsheets using better methodologies, only one study had an error rate below 21%. According to Panko, a weighted average of published spreadsheet audits since 1995 suggests that 94% of spreadsheets had errors (2005). Even if these stunning rates overstate the problem in practice, it is hard to avoid the conclusion that errors in spreadsheets are common. Spreadsheet end-users frequently do not follow design and development recommendations (Teo and Tan, 1999, Clermont et al., 2002) while organizational policies on spreadsheet creation and maintenance are rare (Cragg and King, 1993).

Given that spreadsheets are so frequently used and so frequently flawed, one might expect spreadsheets errors lead to a great many bad decisions. However, there is surprisingly little systematic study of that phenomenon. There are many case studies and anecdotes. The European Spreadsheet

Research Interest Group (EUSPIG) maintains a webpage listing literally scores of news stories reporting the consequences of spreadsheet errors (http://www.eusprig.org/stories.htm). However, spreadsheets are used by so many organizations that there could be scores of isolated examples even though the probability any given organization is impacted by errors is quite small. We started with a population of individuals and organizations for which there was no a priori reason to think spreadsheet errors were a particular problem. This approach has been taken by others (e.g., Cragg and King, 1993) for exploring the prevalence of defective spreadsheets, but we sought to shift the focus slightly to whether the organizations actually suffered losses or bad decisions because of such errors.

The second question motivating our research, "What organizations do and should do to control bad decisions stemming from spreadsheet errors?", has occasioned a considerable literature. Classic recommendations lean toward application of good software engineering principles (e.g., Mather, 1999; Janvrin and Morrison, 2000; Rajalingham et al., 2000) or formal theories generally (Isakowitz et al., 1995). Grossman and Ozluk (2004) recommend that researchers work to identify classes of spreadsheet applications for which specific engineering methodologies can be developed, while acknowledging the diversity of applications and organizational environments.

Kruck and Sheetz (2001) combed practitioner literature for axioms validated by empirical results, supporting aspects of the spreadsheet lifecycle theory (e.g., include planning / design and testing/debugging stages), as well as the recommendation to decrease formula complexity. Striving to understand what organizations actually do, as well as what academics believe they do, Finlay and Wilson (2000) surveyed 10 academics and 10 practitioners. Numerous factors influence the amount of validation, of which the most commonly mentioned by both groups were (a) aspects of the decision and (b) aspects of the spreadsheet underlying the decision context.

However, it would still be valuable to have greater knowledge of what error control methods are currently used and under what circumstances (Grossman, 2002). Furthermore, the range of methods relevant for preventing bad decisions is broader than just the range of methods relevant for preventing spreadsheet errors. That is, there are two ways of keeping a spreadsheet error from leading to a bad decision: (1) preventing the spreadsheet error from occurring and (2) preventing a spreadsheet error form translating into a bad decision. For our respondents, the latter was as important as the former. By shifting the focus slightly from spreadsheet errors to spreadsheet errors' effect on decision making, this paper contributes to the small but important organizational literature (e.g., Chan and Storey, 1996) that studies spreadsheet use in situ.

3. Data and Methods

3.1 Sampling

Interview subjects were identified primarily through personal and institutional contacts. Some were known directly to the authors, but most were reached via one-step referral. For three reasons this approach was preferred to cold-calling a random sample of some population of organizations. The first was to enhance response rates. Executives are busy and not always inclined to devote scarce time to academic projects, but only one person approached through these personal and institutional contacts declined to be interviewed. The second was to focus on respondents judged likely to be introspective and conscious about decision making processes and their strengths and weaknesses vis-à-vis analytical decision support. The final reason was the sensitive nature of the topic. We gave every assurance of anonymity, but respondents might still have been wary of admitting bad decisions to a stranger than to someone referred by a mutual acquaintance.

We originally intended to focus on differences across three sectors (for-profit, non-profit, and government) and two levels (executives vs. senior managers/analysts), with six subjects in each of the six resulting cells. We sought to interview twelve not just six non-profit executives in the expectation that non-profit executives would be a particularly interesting group. Three of the six cells produced one "extra" interview. We include those three extra interviews in the analysis because differences across sector and job classification were by and large not pronounced, yielding a total sample of 45 subjects.

Fifty-five people were interviewed, but ten were excluded from the analysis below. Seven were excluded because they were duplicate interviews within the same work group. In those cases we retained only the respondent with the most sophisticated perspectives concerning spreadsheet use within that work group. Two were eliminated because they had worked for more than one organization across multiple sectors, and it became ambiguous which of their comments concerned which organizations and, hence, which sectors. One was excluded because s/he used spreadsheets only for list-tracking and database functions.

3.2 Sample Characteristics

Since with one exception interviews were done in person, most of the interviewees represented organizations located in one region.[4] Eighty percent of the executives and sixty-five percent of the senior managers /analysts were from that region. Overall, 96% of our executive sample was male, and 45% of our manager/analyst sample was male. All of the executives and 90% of the manager/analysts were non-Hispanic White.

Table 1: Description of Sample

Group	N	Gender		Location	
		Women	**Men**	**Local**	**Other**
Public Executives	6	0%	100%	100%	0%
NonProfit Executives	13	8%	92%	85%	15%
Private Executives	6	0%	100%	50%	50%
Public Managers	7	57%	43%	57%	43%
NonProfit Managers	6	83%	17%	100%	0%
Private Managers	7	29%	71%	43%	57%
Total	45	27%	73%	73%	27%

Educational attainment ranged from several PhDs to a local government manager with a high school degree. None remembered any recent formal training in Excel. Most had learned spreadsheet techniques informally, in the office environment or with the help of practical manuals.

[4] The phone interview stemmed from meeting in person the CEO of a manufacturing firm who (rightly) suggested that his CFO, in another state, had more hands-on insights into the topic.

We classified interviewees by their highest spreadsheet use as basic calculation (n=6), simple modeling (n=20), or advanced modeling (n=19). This distinction was based on the mathematical complexity of formulas and models, as well as spreadsheet size (in number of cells and/or file size), links, functionality and features, such as optimization with Solver.

Our respondents used an average of 2.7 other types of software tools to support decision making, with database and accounting software mentioned most frequently. Advanced spreadsheet users used significantly more types of software tools in addition to spreadsheets (average of 3.3) than did simple (2.5) or basic (2.0) users (ANOVA, $F(2,40)= 3.58$, $p<0.037$).

3.3 Interview Protocol

The interview protocol was developed by conducting ten exploratory interviews with subjects who are not part of this analysis. Those interviews provided a wealth of information on error experiences and spreadsheet-based decision making in various forms, including personal stories, vignettes, intuition and beliefs based on accumulated experience.

Open-ended questions elicited richer responses than did tightly-scripted or multiple choice questions. Indeed, senior executives resisted highly structured questions. Furthermore, the most effective sequence in which to cover topics seemed to vary from interview to interview. Hence, we adopted a semi-structured interview protocol. The protocol helped ensure that a specific set of topics with associated probes were covered in every interview, but the conversation often moved nonlinearly through the topics.

Interviewees were sent a description of the research project and the interview protocol, presented in Appendix A. The final protocol addressed individual and organizational experience with spreadsheets, spreadsheet errors, error control, and effects on decision making.

Variables for quantitative analysis were coded by the primary interviewer based on audio recordings and detailed notes. Most of the responses were coded into categorical variables representing groupings of common responses. To test inter-rater reliability, a subset of the interviews were coded by both primary interviewers. Seventy-six percent of those items were coded consistently, and most discrepancies were instances in which one interviewer drew a conclusion and the other thought there was insufficient information to decide, not from the two interviewers drawing contradictory conclusions. The variables most vulnerable to discrepancies pertained to the advantages and disadvantages of spreadsheets for decision making. They were excluded from the analysis reported below.

4. Results

4.1. Types of Spreadsheet Errors Reported

With one exception, respondents reported encountering errors in spreadsheets. The most commonly mentioned types were *inaccurate data* (76%), *errors inherited from reuse* of spreadsheets (49%), *model errors*, including structural errors and errors of omission (33%), and *errors in the use of functions* (also 33%). While we recorded the presence or absence of a type of error, we have no information on the frequency of occurrence. Error experiences are not scaled to the complexity of the spreadsheet so this may create a masking effect, where error rates and types are constant as the level of spreadsheet complexity changes.

Table 2: Commonly mentioned errors, by sophistication of highest spreadsheet use and by sector

Error Type	All	By Sophistication of SS Use			By Sector		
		Advanced	Simple	Basic	Public	Nonprofit	Private
Inaccurate data	76%	74%	80%	67%	77%	95%	46%
Errors inherited from reusing spreadsheets	49%	63%	40%	33%	31%	63%	46%
Model error	33%	42%	25%	33%	23%	16%	69%
Error in use of functions	33%	21%	45%	33%	46%	26%	31%
Misinterpretation of output/report	27%	47%	15%	0%	15%	32%	31%
Link broken/failed to update	22%	37%	15%	0%	8%	26%	31%
Copy/Paste	22%	21%	30%	0%	23%	26%	15%
Other	11%	11%	15%	0%	8%	5%	23%
Lost file/saved over file	7%	5%	10%	0%	0%	5%	15%
No errors	2%	5%	0%	0%	8%	0%	0%
N	45	19	20	6	13	19	13

Inaccurate Data

Inaccurate data was the most commonly cited source of errors, particularly for nonprofits. Ninety-five percent of respondents in nonprofit organizations mentioned inaccurate data, versus 77% of those in public organizations and 46% of those in private organizations (ANOVA $F(2,42)= 5.91$, $p<0.005$).

Originally we thought this category would primarily reflect miskeyed data and other "mechanical errors" to use Panko's (1999) term. Indeed, such "typos" were mentioned by all types of users. However, there were several other sources of data value errors.

One was bad data piped into a spreadsheet automatically, e.g., from a database query or a web-based reporting service. One health insurance firm reported that a change in prescription pill units triggered automatic reimbursements that were too high because of a units inconsistency between the revised database and the spreadsheet. At one level this is just an example of "Garbage In, Garbage Out." However, it illustrates an issue raised by quite a few respondents and the literature (Panko, 2005). Numbers reported in a spreadsheet can sometimes be attributed with an inordinate aura of inerrancy, lulling users into not reviewing them as critically as they might have in another context.

Respondents mentioned human bias as another source of inaccurate data, varying in culpability from "wishful thinking" to outright fraud. Wishful thinking included analysts who had a (perhaps unrecognized) bias concerning the results and who played around with uncertain parameters, ultimately choosing "base case" values that "made sense" to them because they gave the (perhaps unconsciously) preferred conclusion. The more extreme version was willful self-serving manipulation of input assumptions.

At a mathematical level there is no distinction among these sources of error. However, the implications for quality control are very different. Asking analysts to check parameter values a second time might help catch typos. It would do little or nothing to address a fraudulent self-serving bias.

Errors Inherited from Reusing Spreadsheets

Almost all respondents reused the majority of their spreadsheets, and 49% had experienced errors from reuse of their own or colleagues' spreadsheets. One respondent inherited a model containing a vestigial ‘assumptions page’ that did not link to the model itself. Several described spreadsheet errors that endured for an extended period of time. One large nonprofit had relied on a faulty ROI worksheet for several years, affecting contracts worth ~$10 million. Reuse error reports increased with the complexity of spreadsheet applications.

Opinions differed on the value of reuse. For many respondents, the ability to reuse templates was a key advantage of spreadsheets. Some echoed Nardi et al.’s (1990) observation that templates enable sharing of domain expertise as well as spreadsheet skills. However, a sizable minority felt reused spreadsheets were difficult to control, since when users update worksheets, more errors can be introduced. Furthermore, respondents noted that small errors replicated many times can lead to substantial losses over time. On the other hand, several analysts reported being extremely confident in their spreadsheets because they had used them successfully many times and were now certain that they were error-free.

Errors in the Use of Functions

Thirty-three percent of respondents had experienced errors in functions, ranging from inappropriate use of built-in functions to mistaken operators and cell addressing problems. Careful inspection of formulas was relatively rare unless motivated by an observed discrepancy. In the words of one respondent, “Formulas are only examined in depth if there’s a reason.” The most frequently mentioned reason was the difficulty of review. One senior analyst observed: “Spreadsheets are not easy to debug or audit…It’s a very tedious process to check someone else’s cells, especially two or three levels [of cell references] down.”

Model Error

One-third of respondents reported model errors, including errors in assumptions, overall structure, errors of omission, and other major distortions of the modeled situation (as identified by the respondent). These errors were more frequently reported in private organizations (69%) than in public

(23%) or nonprofit organization (16%), differences that were statistically significant (ANOVA $F(2,42)=6.62$, $p= 0.003$).

Model errors are not programming mistakes, but rather erroneous judgments in modeling real world situations. Because identifying model errors requires domain and business context, they cannot be detected solely by a review that focuses on the spreadsheet's internal logic. The same can be said for correction of the fifth most commonly cited type of error, namely misinterpretation of spreadsheet output or results.

There were some statistically significant differences in error types reported by sophistication of spreadsheet use. Advanced modelers were more likely to report misinterpretation of an output (ANOVA $F(2,42)= 4.36$, $p<0.019$). There is some indication that advanced modelers may also experience link and update errors more frequently (ANOVA $F(2,42)= 4.36$, $p<0.100$). On the other hand, advanced users experience fewer copy and paste errors. Lost and saved over files were more common among basic users, perhaps reflecting poor computing quality control skills, not skill deficits specifically related to spreadsheets.

4.2. Spreadsheet Errors and Decision Making

We are not interested here primarily in spreadsheet errors per se, but rather in whether spreadsheet errors lead to organizational losses or bad decisions. A slim majority of interview subjects whose responses could be coded (25 of 44) expressed strong concern about the consequences of spreadsheet errors in their organization.

Did the remaining 43% of respondents have outstanding error control procedures? Not necessarily. Nine of the nineteen who were not strongly concerned did not use spreadsheets in ways that were integral to strategic or high-stakes decision making. Almost by definition, spreadsheet errors could not cause those organizations grave harm. Looking solely among organizations that used spreadsheets explicitly to inform decisions, the slim majority of concerned subjects (57%) becomes a substantial majority (71%).

Still, ten organizations that use spreadsheets to inform important decision and which experienced spreadsheet errors reported no major concern about adverse impact on the ultimate decisions made. We discuss why errors in spreadsheets that inform decisions do not inexorably translate into erroneous decisions in a later section.

We looked for relationships between membership in these three categories and other variables, ranging from whether the organization was a for-profit, nonprofit, or governmental organization to the types of errors described, but surprisingly few correlations emerged. A few of the descriptive statistics are reported in Table 3.

Table 3: Descriptive Statistics by whether respondent (a) thought spreadsheet errors were a significant threat to decisions, (b) used spreadsheets to inform decisions but was not particularly concerned about errors, or (c) did not use spreadsheets very directly to inform decisions.

Decision Making Group	**N[a]**	**Average Spreadsheet Sophistication** (1=basic, 2 = simple, 3=advanced)	**Decision making Responsibility** Percentage of group that is Executive	**Organization Type** Public	Non-profit	Private
Spreadsheet errors were a significant threat to decisions	25	2.4	72%	38%	56%	77%
Used spreadsheets to inform decisions but was not particularly concerned about errors	10	2.2	60%	38%	17%	15%
Did not use spreadsheets very directly to inform decisions	9	2	22%	23%	28%	8%
N[a]				13	18	13

[a]Number of observations totals to 44 not 45 because for one respondent there was insufficient information to ascertain group membership.

Opinions about whether spreadsheet errors led to bad decisions could also be categorized into three groups in a slightly different way: (1) a small minority who thought spreadsheet errors were always caught before decisions were made, (2) a larger group who acknowledged that not all errors are detected but who nonetheless thought any errors of consequence would be detected before they misinformed a decision, and (3) the plurality who thought spreadsheet errors could have a significant adverse impact on decisions.

In our sample there were more people in the third group (worried that spreadsheet errors could lead to major errors) than in the other two groups, but precise numbers are hard to determine because of incomplete information for some respondents. Still, it appears that these three categories all exist in significant numbers.

We draw two fundamental conclusions from this section. First, it is clear that spreadsheet errors sometimes lead to major losses and/or bad decisions in practice. Indeed, we heard about managers losing their jobs because of inadequate spreadsheet quality control. Second, many senior decision makers whose organizations produce erroneous spreadsheets do not report serious losses or bad decisions stemming from those flawed spreadsheets. Hence, it seems in no way inevitable that errors in spreadsheets that inform decisions automatically lead to bad decisions.

4.3. Quality Control Methods Employed

In our interviews, 12 error reduction techniques were mentioned repeatedly. Most organizations described 4-5 practices that they believed would detect and reduce errors. Note: these are techniques respondents reported and presumably *believe* reduce errors. However, we have no direct evidence concerning whether they actually work.

As mentioned above, the academic literature stresses the benefits of applying fairly formal software engineering practices to spreadsheet development. A typical theme is the benefit of explicitly adopting a lifecycle perspective to spreadsheet development, with phases of design, development, and review. However, essentially none of our respondents thought in those terms. Instead, the methods actually employed might usefully be divided into three categories: (1) informal quality control, (2) formal but organizational methods such as peer review, and (3) technical tools that are specific to spreadsheets. (See Table 4.)

The most frequently applied quality control step was the ‘gut check’ or ‘sniff test,’ a cursory examination of bottom line figures for reasonableness. Organizational methods, documentation and review, were more widely applied than technical error correction tools.

Overall, the quality control methods respondents described might be characterized as stemming primarily from the application of common sense or general managerial acumen. They generally do not differ in obvious ways from quality control steps that would be employed to review any other form of analysis. Hence, what distinguished conscientious from lackadaisical organizations was not necessarily technical sophistication. Rather it was the formality with which general purpose quality control procedures such as peer review were employed.

Table 4: Proportions of respondents reporting use of various quality control methods

			Spreadsheet Sophistication (Highest Level)			
Quality Control Method	Method Type	N[a]	Avg of All Groups	Advanced Modeling	Simple Modeling	Basic Calculation
Gut check against the bottom line	I	45	96%	100%	90%	100%
Review by developer	2	44	86%	84%	85%	100%
Review by someone other than developer	2	44	73%	67%	70%	100%
Crossfooting	3	45	60%	63%	55%	67%
Review by multiple reviewers other than developer	2	44	45%	56%	45%	17%
Documentation	2	45	42%	53%	40%	17%
Keep it simple	1	43	33%	33%	30%	40%
Input controls	3	45	22%	37%	15%	0%
Prevent deletion of calculation cells (protection)	3	45	20%	21%	25%	0%
Other		45	16%	11%	15%	33%
Test cases	3	45	13%	26%	5%	0%
Separate page for audit/change tracking	3	45	13%	21%	10%	0%
Audit tools	3	45	7%	16%	0%	0%

[a]Number of observations is less than 45 for items for which insufficient information was gathered to ascertain whether that particular quality control method was employed. Percentages reported are for respondents who gave complete information concerning this topic.

Informal Quality Control: The ‘Sniff Test’ and ‘Keeping It Simple’

The most frequently cited quality control procedure was the ‘sniff test,’ which almost all respondents employed (96%, 43 out of 45). Many respondents acknowledged the limitations of ‘sniff tests,’ at least in complex models. One analyst reported finding a major modeling error worth several million dollars by “dumb, blind luck” the night of the presentation to the board. Pointing to a blank cell accidentally excised a significant portion of the analysis, but even that obvious error passed many sniff tests.

However, the contrary view was also expressed by a significant minority. Recall the definition of the second category in terms of perceived impacts on decision making: people who acknowledged that not all errors are detected but who nonetheless thought all errors of consequence would be detected before they misinformed a decision. To an important degree that group's position was based on confidence that 'sniff tests could catch the larger errors.'

The other informal method respondents mentioned was to 'keep it simple.' This included simplifying the analysis as well as the spreadsheet itself. Executives were slightly more likely to recommend 'keeping it simple' (36%) than were analysts and managers (27%). Advanced developers used advanced methods in addition to, not instead of these informal methods. For example, four out of five decision makers engaged in advanced modeling recommended 'keeping it simple.'

Formal Quality Control Methods that Are Organizational in Nature

Respondents mentioned two tools that were organizational in nature: documentation and spreadsheet review. We expected that advanced users would be the most likely to formally document spreadsheets. Yet documentation was by no means limited to advanced modeling. Approximately 40% of our respondents doing simple modeling used documentation, as did one basic user, so these differences were not statistically significant. However, documentation was highly related to use of the technical tools described below. Those who document use an average of 2.0 technical tools, while those who do not document use an average of only 0.8 technical tools (ANOVA $F(1,43)=10.36$, $p<0.002$). Respondents who had experienced misinterpretation of a report were also more likely to report using documentation (ANOVA $F(1,43)= 4.20$, $p<0.046$). Documentation concerns were mentioned more frequently in respect to assumptions and parameter values, rather than formulas, which is consistent with Galetta et al. (1996).

Most (73%) respondents reported that spreadsheets were reviewed by someone besides the developer. For the forty-five percent using multiple outside reviewers, this ranged from two colleagues in the office to review by several executives in teleconference. One organization mentioned inadvertent outside review. A public manager confessed that their annual budget, which invariably contained errors in the first release, would be scrutinized by unions and other groups for their own interests, helping to correct the document.

Although review was common, one-quarter of respondents did not mention any kind of spreadsheet review by others. Furthermore, almost no respondents spent even the minimum time on validation that is suggested by Olphert and Wilson (2004).

Technical Tools

We define technical tools as quality control methods that are specific to spreadsheets, such as protecting portions of the spreadsheet as a form of input control. Advanced modelers were the most likely to use these technical tools (ANOVA $F(2,42)=2.47$, $p<0.097$), particularly those beyond crossfooting (redundant calculations) (ANOVA $F(2,42)= 3.91$, $p<0.028$). Presumably, knowledge of advanced functionality and technical tools for quality control go together.

Test cases were rarely used. Twenty six percent of advanced modelers used *test cases*, compared to 5% of simple modelers and no basic users (ANOVA $F (2,42)=2.56$, $p<0.089$). A similar pattern emerged for Separate Pages for Audit / Change Tracking, although the differences were not statistically significant.

Least frequently mentioned were automatic audit tools (16% of advanced modelers and no one else). Few used formal software tools (e.g., Morrison et al., 2002) even though there are commercial packages available. Few even knew about or used Excel's own (limited) built in Audit feature, and those who did reported having used it on an occasional basis only.

Other Methods

The other error control methods were diverse, and each was mentioned by only one respondent. A number were interesting inasmuch as they represented "outside the box" thinking relative to methods typically discussed in the spreadsheet quality control literature. Two were personnel-related: firing people who produced flawed spreadsheets and, on the positive side, hiring talented people in the first place, where talent referred to general analytical competence not spreadsheet-specific experience or skills. Another method cited was to not use spreadsheets at all or as often, e.g., by converting the organization to some enterprise resource planning (ERP) system or by encouraging use of other analytical software. Another created duplicate models to verify results.

4.4. Organizational Policies Concerning Spreadsheet Quality Control

Most respondents reported that their organizations had no formal policies intended to ensure spreadsheet quality. Financial services sector organizations were the exception. Common practices included the use of 'standardized' spreadsheet models, created by IT personnel at corporate offices and distributed to branches, along with selected data protection. The only example outside the financial sector was a health care organization whose guidelines for managers included a reference to commonly used spreadsheet formulas, but no specification of how to build the spreadsheet.

Without formal policies, several workgroups still engendered high quality environments where review and documentation were routine. It was easy to understand why some respondents had stressed good quality control in their organization. One was a PhD computer scientist familiar with the risks of spreadsheet programming. Two others were extensively trained in accounting and taught spreadsheet modeling to graduate students.

For most, the absences of policies were not the result of any thoughtful decision balancing the benefits of improved quality against the overhead associated with rigorous quality control. Many responded by saying, in essence, "Never thought about it, but it sounds like a great idea." Others contended that despite errors, the risk was just not great enough to justify the resources needed for thorough review. As Grossman points out, that is not necessarily unreasonable: "Tolerance of spreadsheet errors is not necessarily foolish or even irrational; it is a matter of degree and perceived risk" (Grossman, 2002). One government analyst reported that she often had to create or modify spreadsheets with as little as half an hour preparation. Errors were sometimes discovered later in those spreadsheets, but timeliness was paramount. Another public manager resisted an auditor's recommendations to hire additional staff to check spreadsheet accuracy, citing budget constraints. The perception among several managers that checking spreadsheets more carefully would require hiring another employee may be inconsistent with the conventional wisdom in software development that time invested in quality control upfront is more than recouped in reduced rework down the road.

The literature suggests that formal policies governing spreadsheet use often encounter resistance because the main advantages of spreadsheets include empowering end users to complete analysis independently (Cragg and King, 1993 and Kruck and Sheetz, 2001). Some of our respondents expressed a related yet distinct concern. These respondents all worked in small, non-corporate environments and suggested that guidance should be informal and/or implicit in these close-knit workplaces. Their arguments centered on unintended consequences of formality on workplace culture, such as concern that spreadsheet policies would betray a lack of trust in staff competence, rather than effects on spreadsheet productivity per se.

Most organizations did not have formal policies governing how to respond when errors were detected, but respondents described an interesting range of responses. A common response was to fix the specific error but do nothing else. Forty percent went further to investigate the associated processes to detect other, related errors. Other responses were less obviously effective. The extreme example was an individual who threw out the computer on which a perpetually buggy spreadsheet was being run, in the belief that the hardware was somehow at fault.

A more subtle issue pertains to a practice mentioned by thirty percent of respondents: possibly rebuilding a spreadsheet from scratch when errors are detected. This practice could be eminently sensible, particularly if the original design was not well thought out. Often a cumbersome exploratory spreadsheet can be replaced by a more elegant and less error-prone second version. On the other hand, this may also reflect undue confidence that rebuilding the spreadsheet might not introduce new, more serious errors. Some respondents seemed to view errors as aberrations that can be eliminated if the work is redone carefully, rather than a predictable outcome, as in the software engineering perspective of a predictable number of errors per thousand lines of code (cf., Panko, 1999).

4.5. Other Factors Mediating the Application of Spreadsheet Quality Control

An important question distinct from what methods are ever used is how *consistently* quality control methods are applied. Our protocol did not address this topic, so we do not have systematic data from all 45 respondents, but it came up spontaneously in many interviews.

One ideal espoused in the literature is *not* to apply all methods to all spreadsheets. Rather, the risk analysis philosophy suggests investing more in quality control methods for certain high-risk spreadsheets. This approach has also been recommended in the popular literature (Whittaker, 1999) and makes good sense for organizations with limited resources. However, at least eight respondents mentioned situations in which, perversely, the level of review for important spreadsheets was less, not more, rigorous:

- Highly independent executives often completed one-off, ad hoc and first-time analysis for an important decision without the benefit of review.
- When the spreadsheet was highly confidential, few people in the organization had access to it, making effective review difficult.
- Important decisions were often associated with time pressures that made formal review inconvenient.

On the other hand, the respondents suggested some points that might usefully extend and improve the prioritization schemes common in the literature. In particular, the literature suggests focusing on spreadsheets that are the most likely to contain errors. Our respondents stressed that it is at least as important to consider the decision context, such as the stakes of the decision. It might make sense to review a simple spreadsheet underpinning a $1B decision as carefully as a complex and error-prone decision underpinning a million dollar decision.

One contextual issue raised in our interviews has not typically been stressed in the literature (Finlay and Wilson, 2000, being a partial exception). That issue is whether the decision was internal as opposed to being part of a public or adversarial proceeding such as labor negotiations or law suits. One example cited stemmed from a highly partisan political budgeting battle. The respondent noted that the opposition's budget literally did not add up. The sum of itemized subcategories did not (quite) match the category total, a minor discrepancy both in absolute dollars and as a percentage of the total budget. However, the respondent was able to exploit that small but incontrovertible error to cast doubt on the credibility of all of the other party's analysis, leading fairly directly to a dramatic political victory before the legislative body.

4.6. Social Elements that Insulate Decisions from Spreadsheet Errors

We began our research with a rather narrow vision of spreadsheets and decision making that was shaped by our experience teaching decision modeling courses and might be caricatured as follows. "Leaders sometimes analyze decisions quantitatively with spreadsheets that explicitly project the bottom-line consequences of different courses of action. The spreadsheet thus effectively if not literally encodes choices as "decision variables" whose values can be selected to achieve managerial goals. Decision makers combine that analysis with expert judgment and factors outside the model to select a particular course of action."

That naïve view suggests an almost one-to-one or automatic connection between spreadsheet errors and decision errors. However, while our respondents almost universally reported spreadsheet errors, significantly fewer reported that they led to bad decisions. From an organizational perspective, apparently there were decision making processes mitigatingd the consequences of spreadsheets errors.

We note three of the biggest disjunctions between our original "academic" view of organizational decision making and what our respondents described, since they bear directly on the present topic and also on avenues for further research.

First, even when there was a well-defined decision to be made, most quantitative analysis in spreadsheets provided descriptive, contextual information. For example, the spreadsheet might be a budget model identifying the scale of investment required by a proposed project, but with no "objective function" cell to be optimized or even necessarily a bottom line performance metric, such as impact on profit. The decision makers might greatly appreciate the budget model's ability to clarify what it is the organization would be committing to if it made a “go” decision, but they were not looking to the spreadsheet to forecast bottom line results (particularly for government and non-profit organizations that usually grapple with multiple objectives, not all of which are easily quantified).

Second, respondents thought that people, not spreadsheets, make decisions. When bad decisions were made, even when the underlying spreadsheet analysis was flawed, they did not scapegoat the spreadsheets. As one executive in the financial sector said, quantitative analysis is only “the first 75% of a decision. … this [spreadsheet] is a decision making tool, not an arbiter.” Another senior financial manager emphasized that it was “his job” to interpret the spreadsheet analysis in light of other critical qualitative factors. Because human decision making based on experience and domain knowledge was critical, these interviewees were less concerned about the impact of spreadsheet errors.

By definition, in decision support applications, there is a human in the loop. Willemain et al. (2003) argued based on experimental evidence that there is a "robust human ability to overcome flawed decision support." Our field findings are consistent with that view. There were instances in which the spreadsheets essentially made the decision, such as when spreadsheet-based programs automatically generated bills. However, in those applications, the spreadsheet was being used more for routine data processing than for (strategic) decision support.

Third, spreadsheets often informed decision making that was much more diffuse than the narrow image of a leader (or committee) sitting down at a particular point in time to make a specific, discrete, well defined choice. For example, the decision to invest in an area might not be a one-shot decision to write a check of a certain size on a particular day, but rather a commitment to nurture an organizational effort by exhibiting leadership and oversight in a variety of large and small ways over an extended period. In such cases, the leader might consume with great interest spreadsheets describing the financial performance of the subsidiary or product line in question. However, even if those spreadsheets contained errors it would be hard to point to specific decisions where wrong choices could be traced to those spreadsheets.

To summarize, when thinking about how spreadsheet errors contribute to bad decisions, it may be useful to think of that not primarily in terms of spreadsheets recommending the wrong course of action but rather as misinforming deliberations. Furthermore, the overall information context underpinning decisions is almost always murky, ill structured, and incomplete. So it is not as if decision makers would have had a crystal clear understanding of all relevant factors if only the spreadsheet had not had an error. Rather, the spreadsheet error might just thicken the fog surrounding the decision. Indeed, sometimes the flawed spreadsheet might even help dispel some of that fog, just less effectively than it would have if it had not contained an error (cf., Hodges, 1991).

None of these observations are meant in any way to minimize the problem of spreadsheet errors or to question the premise that reducing spreadsheet errors can improve organizational performance and decision making. Rather, they are offered to explain the results above and to suggest that the unit of analysis in some future research on the topic might usefully be the organizational decision processes, not just the individual spreadsheet.

5. Discussion

Our interviews affirmed two common findings: (1) Spreadsheets are frequently used to inform decisions and (2) spreadsheets frequently have errors. Given this, one might expect these respondents to be able to recount many instances of spreadsheet errors leading to bad decisions or other losses. Indeed, the majority of our respondents could cite such instances and viewed them as a serious problem. However, a significant minority did not view this as a serious problem and even among those who did, the sky was not falling. No respondent suggested that the proportion of flawed decisions in any way approached the proportion of spreadsheets the literature finds to be flawed.

Disaster was not being avoided because of systematic application of formal, spreadsheet-specific quality control policies and procedures. Indeed, few organizations outside the financial sector had formal policies, and the actual practices reflected general, common sense concern for the quality of analysis more than they did technical or spreadsheet-specific tools or procedures.

Three alternative but not mutually exclusive explanations emerged as to why spreadsheet errors lead to some, perhaps even many, but still not an overwhelming number of flawed decisions. The first view, espoused by a significant minority of respondents, was that informal quality control methods work reliably for precisely those errors that could be most problematic. When the spreadsheet analysis is wildly off, experienced decision makers can sniff that out. Small spreadsheet errors might not be noticed, but small errors either won't tip the scales in a decision or, even if they do, will have minor consequences.

The second explanation is that for some organizations (nine of forty-five in our sample), the spreadsheets are used to crunch numbers, but not in ways that are tied to specific decisions. The spreadsheets might be used for various types of information processing, ranging from database-like functions to synthesizing and graphing data drawn from another system, but the spreadsheets are not being used in ways that inform specific strategic decisions.

The third explanation from our respondents is that even when spreadsheets are used to inform specific decisions and large errors might go undetected, the spreadsheet analysis is merely informing but not driving or determining the decisions. The image one should have is not that an analyst enters all relevant considerations into a spreadsheet, analyzes that spreadsheet, and the organization implements whatever course of action that spreadsheet suggests. Instead, there is some organizational decision process, involving at least one but often multiple people. Those people bring to the table a great deal of judgment and wisdom, as well as a range of data, mental models, forecasts, etc. Spreadsheets may well have been used to inform or even generate some of those data, mental models, forecasts, etc., but other sources of information are also drawn upon. At the end of the day, it is humans exercising human judgment that make the decision.

Usually that judgment is exercised in the face of terribly incomplete and imperfect information. A good spreadsheet analysis might fill in some but not all of that incomplete information. A bad spreadsheet analysis might increase the amount of imperfect information. The murkier the information, the greater the risk of bad decisions, so spreadsheet errors can lead to bad decisions, but the overall organizational decision processes do not necessarily break down in the face of some bad information.

To the extent that results from our sample generalize, they have three implications for research on spreadsheet errors and decision making. First, this is a topic of importance to which field interviews contribute interesting insights, so further research would be of value. Second, spreadsheets are used for diverse purposes, and not all spreadsheet analysis is closely connected to decision making. Third, it might be productive for some research to make the unit of analysis be the decision, or the collectivity of all analysis in support of that decision, not the spreadsheet.

References

Buehlmann, U., Ragsdale, C.T., & Gfeller, B. (2000). A Spreadsheet-Based Decision Support System for Wood Panel Manufacturing. Decision Support Systems, 29(3), 207-227.

Chan, Y.E., & Storey, V.C. (1996). The Use of Spreadsheets in Organizations: Determinants and Consequences. Information and Management, 31(3), 119-134.

Cragg, P.B., & King, M. (1993). Spreadsheet Modeling Abuse: An Opportunity for OR. J, 44(8), 743-752.

Clermont, M., Hanin, C., & Mittermier, R. (2002). A Spreadsheet Auditing Tool Evaluated in an Industrial Context. https://143.205.180.128/Publications/pubfiles/pdffiles/2002-0125-MCCH.pdf.

Finlay, P.N., & Wilson, J.M. (2000). A Survey of Contingency Factors Affecting the Validation of End-use Spreadsheet-based Decision Support Systems. JORS, 51, 949-958.

Gerson, M., Chien, I.S., & Raval, V. (1992). Computer Assisted Decision Support Systems: Their Use in Strategic Decision Making. Proceedings of the 1992 ACM SIGCPR Conference on Computer Personnel Research held in Cincinnati, Ohio. New York: ACM Press 152-160.

Grossman, T.A. (1997). End-User Modeling. OR/MS Today, 24(5), 10.

Grossman, T.A. (2001). Causes of the Decline of the Business School Management Science Course. INFORMS Transactions on Education 1(2).

Grossman, T.A. (2002). Spreadsheet Engineering: A Research Framework. European Spreadsheet Risks Interest Group 3rd Annual Symposium, Cardiff . Available at: http://www.usfca.edu/sobam/faculty/grossman_t.html

Grossman, T.A., & Ozluk, O. (2004). A Paradigm for Spreadsheet Engineering Methodologies. Personal correspondence. European Spreadsheet Risks Interest Group 5th Annual Symposium, Klagenfurt, Austria. Available at: http://www.usfca.edu/sobam/publications/ AParadigmforSpreadsheetEngineeringMethodologies2004.pdf.

Hodges, J. (1991). Six (or so) Things You Can Do With a Bad Model. Operations Research, 39, 355-365.

Jain, A.K., Banerjee, A., Gupta, A.K., Thomas, K.N., & Chandra, K. (2000). A Spreadsheet Program to Optimize Kinetic Parameters of Organic Matter from Laboratory Pyrolysis Data. Computers & Geosciences, 26(9-10), 1023-1028.

Kruck, S.E., Maher, J.J., & Barkhi, R. (2003). Framework for Cognitive Skill Acquisition and Spreadsheet Training. Journal of End User Computing, 15(1), 20-37.

Kruck, S.E., & Sheetz, S.D. (2001). Spreadsheet Accuracy Theory. Journal of Information Systems Education, 12(2), 93-107.

Isakowitz, T., Schocken, S., & Lucas, Jr., H.C. (1995). Toward a Logical/Physical Theory of Spreadsheet Modeling. ACM Transactions on Information Systems, 13(1), 1-37.

Janvrin, D., & Morrison, J. (2000). Using a Structured Design Approach to Reduce Risks in End User Spreadsheet Development. Information and Management, 37, 1-12.

Mather, D. (1999). A Framework for Building Spreadsheet Based Decision Models. Journal of the Operational Research Society, 50, 70-74.

McGill, T. (2002). User-Developed Applications: Can End Users Assess Quality? Journal of Organizational and End User Computing, 14(3), 1-15.

McGill, T., & Klobas, J.E. (2005). The Role of Spreadsheet Knowledge in User-developed Application Success. Decision Support Systems, 39, 355-369.

Morrison, M., Morrison, J., Melrose, J., & Wilson E.V. (2002). A Visual Code Inspection Approach to Reduce Spreadsheet Linking Errors. Journal of End User Computing, 14(3), 51.

Olphert, C.W., & Wilson, J.M. (2004). Validation of Decision-Aiding Spreadsheets: The Influence of Contingency Factors. Journal of the Operational Research Society, 55, 12-22.

Panko, R. (2000). Two Corpuses of Spreadsheet Error. IEEE: Proceedings of the 33rd Hawaii International Conference on Systems Sciences. Available at: http://panko.cba.hawaii.edu/ssr/Mypapers/HICSS33-Panko-Corpuses.pdf

Panko, R.R. (1999). Applying Code Inspection to Spreadsheet Testing. Journal of Information Management Systems, 16(2), 159-176.

Panko, R.R. (1998). What We Know About Spreadsheet Errors. Available at: http://panko.cba.hawaii.edu/ssr/Mypapers/whatknow.htm, updated 2005. Previously published in Journal of End User Computing 10(2).

Plane, D.R. (1994). Spreadsheet Power. OR/MS Today. 21(6), 32-38.

Powell, S.G. (1997). Leading the Spreadsheet Revolution. OR/MS Today 24(6), 50-54.

Powell, S.G. (1998). Requiem for the Management Science Course. Interfaces, 28(2), 111-117.

Rajalingham, K., Chadwick, D., & Knight, B. (2000). Classification of Spreadsheet Errors. British Computer Society (BCS) Computer Audit Specialist Group (CASG) Journal, 10(4), 5-10.

Rajalingham, K., Chadwick, D., Knight, B., Edwards, D. (2000). Quality Control in Spreadsheets: A Software Engineering-Based Approach to Spreadsheet Development. Proceedings of the Thirty-Third Hawaii International Conference on System Sciences, Maui, Hawaii, January 2000.

Rydell, C.P., & Everingham, S.S. (1994). Controlling Cocaine. Supply Versus Demand Programs. Santa Monica: RAND.

Teo, T.S. H., & Tan, M. (1999). Spreadsheet Development and What if Analysis: Quantitative Versus Qualitative Errors. Accounting, Management and Information Technology, 9, 141-160.

Whittaker, D. (1999). Spreadsheet Errors and Techniques for Finding Them. Management Accounting, 77(9), 50-51.

Willemain, T.R., Wallace, W.A., Fleischmann, K.R., Waisel, L.B., & Ganaway, S.N. (2003). Bad Numbers: Coping with Flawed Decision Support. Journal of the Operational Research Society, 54, 949-957.

Appendix A – Interview Protocol (SS = spreadsheet)

1. Introduction of researchers and project topic
    - Managers frequently use SS to analyze and inform decisions; research has shown that many of these SS contain errors.
    - This project will investigate how these errors affect the quality of decision making, and propose recommendations on the best ways of reducing these errors.

2. How often do you build SS decision making tools?
    - Do you personally create SS to support decisions?
    - How complex are these SS (in terms of the calculations performed in them, or the amount of data contained in them)?

3. How often do you use SS for making decisions?
    - Does your staff present you with SS or SS-based analysis on a regular basis?
    - How complicated are the SS you encounter (in terms of the calculations performed in them, or the amount of data contained in them)?
    - What decisions are these SS being used to support?
    - What makes SS useful for this decision making?
    - Do you use other quantitative tools for decision making?

4. What is your level of expertise with SS modeling and/or other development environments?
    - Have you had formal training in Excel or software programming/development?
    - Have you ever held a position that required daily creation and manipulation of SS?
    - Which features of Excel are familiar to you?

5. Please describe your experiences with SS that were known to contain errors
    - Do SS errors have the potential to cause significant impact?
    - Was the source of the error(s) ever determined?
    - Were the errors caught before damage was done? If not, what was the extent of damage?
    - Describe the errors and what you think caused them.
    - How were the errors fixed?

6. What are the advantages and disadvantages of using SS for decision making?
    - Are there particular features or tools that you have found most useful in your SS?
    - What are the limitations of SS?
    - Is the quality and reliability of your SS a concern for you?
    - Is there anything that might reduce your concerns?

7. Please describe any processes or tools you have used to ensure the integrity of SS
    - Have you or your staff used Excel's built in tools for error-checking?
    - Have you or your staff used Add-ins provided by another vendor?
    - Does your organization follow a particular development process for creating SS models?
    - What other methods are used to detect errors?

8. Are there any other issues related to the topic that you would like to talk about?
    - Do you have advice for other decision-makers?
    - Any stories/anecdotes about particularly helpful solutions to SS problems, or horror-stories about the impact of errors?
    - Recommended readings, web sites, other resources?