Errors in Operational Spreadsheets

Stephen G. Powell, Dartmouth College, USA

Kenneth R. Baker, Dartmouth College, USA

Barry Lawson, Dartmouth College, USA

ABSTRACT

Spreadsheets are used in almost all businesses, for applications ranging from the mundane to the mission-critical. Errors in the data, formulas, or manipulation of spreadsheets could be costly, even devastating. The received wisdom is that about 5% of all formulas in spreadsheets contain errors, and this rate is consistent across spreadsheets. However, this estimate is based on five studies, some of which are quite informal, and a total of only 43 spreadsheets. Our research was designed to deepen our understanding of spreadsheet errors. Specifically, we address three questions about errors in operational spreadsheets: what is the average cell error rate, how does it differ among spreadsheets, and what types of errors are most prevalent? We created a spreadsheet auditing protocol and applied it to 50 diverse operational spreadsheets. We found errors in 0.9% to 1.8% of all formula cells, depending on how errors are defined. We also found that the error rate differed widely from spreadsheet to spreadsheet. [Article copies are available for purchase from InfoSci-on-Demand.com]

Keywords: please provide

INTRODUCTION

Errors in software have been a problem since the beginning of the computer era. The discipline of software engineering (Sommerville, 2004) was created in part to develop methods to minimize errors in software code. Professionally developed software is now generally created and debugged using disciplined approaches from software engineering.

The advent of the personal computer in the 1980s increased by orders of magnitude the number of people who worked closely with computers. Just as professional programmers were learning to use disciplined approaches to programming, millions of so-called end-user programmers began to create their own programs in the form of spreadsheets. One of the appealing features of spreadsheet programming on a personal computer is that the end user can work independently of the constraints of a professional programmer or an MIS group. This allows the end user much more freedom and speed in completing the business task at hand.

But very few end users have any training for the programming tasks they attempt, and many have little appreciation for the risks inherent in a casual approach to programming.

As spreadsheets have spread throughout business, they have increasingly been used for mission-critical applications (Croll, 2005; Grossman, Mehrotra, & Özluk, 2005). At the same time, evidence has accumulated that many spreadsheets have errors (Panko, 2005) and that sometimes these errors cost the organizations that use them millions of dollars (EUSPRIG, n.d.). End users and organizations that rely on spreadsheets generally do not fully recognize the risks of spreadsheet errors. In fact, spreadsheets tend to be somewhat invisible, both as corporate assets and as sources of risk.

Although the received wisdom holds that errors are prevalent in spreadsheets, our review of the literature suggests that this conclusion is based more on casual empiricism than solid research. In fact, researchers have produced only limited information on the types of errors that occur, how frequently they occur on average, and how the error rate varies from one spreadsheet to another.

The current study was designed to improve our understanding of spreadsheet errors by testing a large sample of spreadsheets in actual use by organizations. We developed a detailed auditing protocol and trained a group of researchers to apply it consistently. Using an explicit protocol is important for two reasons: (1) it allows other researchers to replicate and improve on our work; and (2) it contributes to the development of improved auditing procedures, which is important in its own right. The auditing procedure is described in detail in Powell, Baker, and Lawson (2006b).

We begin this article with a summary of previous work on spreadsheet errors which comes from field audits and laboratory experiments. Then we describe the design of our study and the sample of spreadsheets we audited. Our results are summarized in terms of *error instances* (the occurrence of a single type of error) and *error cells* (the number of cells affected by a single error instance). We also categorize the errors as to whether they generate wrong numerical results. We report the total number of errors we identified, as well as how they were distributed across error types and across spreadsheets. This is the first data on errors and their distribution to appear in the research literature that can be replicated with parallel studies using similar methods.

PREVIOUS WORK ON SPREADSHEET ERRORS

Since our interest in this article is primarily with the classification and frequency of errors, we will discuss only those questions here. In a related article (Powell, et al., 2006a), we provide a review of the literature on all aspects of spreadsheet errors.

Types of Errors

Before we can productively study spreadsheet errors, we need to define and classify the types of errors we seek. Classification is useful because it helps us to understand what is common among errors. It also allows us to compare different studies of errors, and eventually to understand the causes of errors and to develop cures.

At the outset, we make a distinction between errors in the *use* of a spreadsheet and errors in the *cells* of the spreadsheet itself. A spreadsheet can be perfectly correct but used to solve the wrong problem, or its results could be misinterpreted. More concretely, a correct spreadsheet can lead to an error if sorting is done improperly, or if formulas are overwritten, or if incorrect data were entered temporarily. All of these errors would be difficult or impossible to detect purely from examination of the spreadsheet itself.

Most classifications of spreadsheet errors have focused primarily on the spreadsheet itself, not on errors in use. Because spreadsheets essentially consist of data and formulas, an error in the spreadsheet must arise from erroneous data or erroneous formulas.

There is no generally accepted taxonomy of spreadsheet errors, possibly because different classifications are appropriate for different purposes. The first researchers who studied spreadsheet use simply listed examples of errors without attempting to classify them. Subsequent authors offered distinctions (but not complete taxonomies), such as that between *domain* errors (lack of knowledge of the application area) and *device* errors (misuse of the spreadsheet itself).

Panko and Halverson (1996) were the first to offer a real taxonomy of errors. They distinguished *qualitative* and *quantitative* errors, and further decomposed quantitative errors into *mechanical, logic,* and *omission* errors. Quantitative errors lead to wrong numbers in the current version of the spreadsheet. The three types were defined as follows:

- Mechanical errors arise from typing or pointing errors.
- Logic errors arise from choosing the wrong function or creating the wrong formula.
- Omission errors arise from misinterpretation of the situation to be modeled.

Qualitative errors do not lead to wrong numbers in the current version of the spreadsheet, but are risky practices that could lead to errors in subsequent use. This category includes practices such as hard-coding numbers into formulas. It can also include long formulas, poor physical layout, or lack of sufficient documentation.

Rajalingham, Chadwick, and Knight (2000) and Purser and Chadwick (2006) have developed the most elaborate taxonomy of errors available to date (see Table 1). The first distinction in this hierarchy is between *application-identified* errors and *developer/user-identified* errors. Excel has eight categories of errors that it displays in a cell when it cannot resolve a formula; these are application-identified errors. The remaining errors are developer/user-identified errors.

Although this taxonomy may be interesting from a theoretical perspective, it has limitations as a practical tool for auditing. First, its categories overlap. For example, an error caused by lack of domain knowledge could occur during insertion. Second, it may not be possible in practice to determine the cause of an error and

Table 1. Error taxonomy (Rajalingham et al., 2000; Purser & Chadwick, 2006)

```
A. Application-Identified Errors

B. Developer/User-Identified Errors
    1. Qualitative Errors
        A. Structural
                a. Visible
                b. Hidden
        B. Temporal
    2. Quantitative Errors
        A. Reasoning
            a. Domain Knowledge
                    1. Real-World Knowledge
                    2. Mathematical Representation
            b. Implementation
                    1. Syntax
                    2. Logic
        B. Accidental
            a. Insertion
            b. Update
                    1. Modification
                    2. Deletion
```

therefore to properly categorize it. For example, a complex formula giving wrong results could be due to lack of domain knowledge, incorrect syntax, or an accidental typing mistake.

Powell et al. (2006a) summarized the literature on classification of errors as follows:

- Classifications are offered without specifying the contexts or purposes for which the classification is intended.
- The existing classifications do not include sufficient examples of specific errors that satisfy each category.
- Classifications are not rigorously tested to demonstrate that multiple users can consistently classify actual errors into the proper categories.
- The boundary between quantitative errors and qualitative errors remains vague.
- No studies compare the types of errors by development stage.

Frequency of Errors

Research on the frequency of errors in spreadsheets comes from a variety of sources, mostly experiments in the laboratory. One approach is to observe subjects as they build spreadsheets and record the number of errors they make. Another approach is to ask subjects to build a spreadsheet from a written description and determine the frequency of errors in the finished product. However, the approach of most relevance to our research is the field audit, in which operational spreadsheets are examined for errors.

Panko (2005) summarized the results of seven field audits in which operational spreadsheets were examined, typically by an outsider to the organization. His results show that 94% of spreadsheets have errors and that the average cell error rate (the ratio of cells with errors to all cells with formulas) is 5.2%.

Table 2 summarizes the data behind these estimates. In total, 88 spreadsheets are represented in the table. For all 88, the weighted average percentage of spreadsheets with errors is 94%. Data on cell error rates were available on 43 of these spreadsheets, and the weighted average for this sample is 5.2%.

There are several reasons to question the reliability of these estimates. First, three of the seven sources are unpublished (Hicks, 1995; Lukasic, 1998; Butler, 1992). Second, the majority of the sources gave little or no information on their definition of errors or on the methods used to find errors. Third, Lawrence and Lee (2004), whose observations account for 70% of the sample used to estimate the cell error

Table 2. Spreadsheet error rates (Panko, 2005)

Source	Number Audited	Percent with Errors	Cell Error Rate (% of cells)
1. Coopers and Lybrand (1997)	23	91	N/A
2. KPMG (1998)	22	91	N/A
3. Hicks (1995)	1	100	1.2
4. Lukasic (1998)	2	100	2.2
5. Butler (2000)	7	86	0.4
6. Clermont (2002)	3	100	3.0
7. Lawrence and Lee (2004)	30	100	6.9
Average		**94***	**5.2****

Weighted average of 88 spreadsheets in sources 1-7.
**Weighted average of 43 spreadsheets in sources 3-7.*

rate, did not actually report a cell error rate of 6.9% on *completed* spreadsheets. Rather, they reported that auditors had "issues" concerning 6.9% of the cells on the *initial review* of models that they subsequently discussed with the developers. The authors provided no details on the definition of an "issue" or its relation to an actual error. Moreover, the initial model version that was reviewed was revised an average of six times before the model was complete. Thus the estimate of a 6.9% error rate applies only to issues raised by auditors on initial model versions, not to errors in completed spreadsheets. If, as seems warranted, we exclude this study from the calculation of the cell error rate, the average cell error rate decreases from 5.2% to 1.3% on a base of only 13 audited spreadsheets.

Our review of the literature on spreadsheet errors draws attention to several shortcomings. First, no generally accepted classification of errors exists. Second, the classifications that do exist are of more theoretical than practical value. Third, existing estimates of error rates are based on extremely limited data. Fourth, no studies of errors have fully revealed the sources of spreadsheets tested, how errors were defined, and the auditing methods that were used.

RESEARCH DESIGN

Our research into spreadsheet errors is predicated on several guiding principles or constraints. First, our interest is in errors in *completed, operational* spreadsheets, not errors made in a laboratory setting or errors made during the development of a spreadsheet. A second principle is that we conduct the audit using only the information in the spreadsheet itself, without relying on the developer. A third principle is that we use an explicit auditing protocol that any moderately experienced user of Excel can master.

These principles limit our study in certain ways. When we work without access to the spreadsheet developers, we are not able to check our understanding of a model with an expert. In practice, this means that at times we will accept a suspicious formula as correct because we cannot be *sure* that it is incorrect. It also means that we cannot hope to uncover errors in formulating the underlying problem or errors in interpreting spreadsheet results. Detecting these types of errors requires a much more time-intensive longitudinal study of how spreadsheets are used in the broader context of problem solving. Offsetting these limitations is the fact that we can audit a much larger volume of spreadsheets using our protocol than otherwise.

Our procedures almost certainly lead to an *underestimate* of the actual error rates in the spreadsheets we audit. We may occasionally classify a cell incorrectly as an error, but we have been conservative in limiting our definition of errors to cells for which we can have a high degree of confidence that our judgment is correct. On the other hand, there are entire classes of errors that we cannot hope to identify with the procedures used in this study. For example, most (but not all) errors in input data are beyond our scope.

Sample Spreadsheets

Our sample of 50 spreadsheets came from a wide variety of sources. Some were obtained during site visits to companies. We carried out site visits at two consulting companies, a bank, a college, a large energy company, and a state government agency. We also obtained spreadsheets from a variety of organizations through the alumni and faculty networks at the Tuck School of Business. No single source contributed more than five spreadsheets to the sample of 50 analyzed here.

We also obtained spreadsheets from various Web sites. Several software companies post spreadsheets on their Web site, either to illustrate how to use their software or to showcase the results that practitioners have had with their software. The Web site of Decisioneering Inc. (http://www.decisioneering.com), makers of the *Crystal Ball* add-in for Excel, is typical. This site lists hundreds of sample models used in industries ranging from aerospace to utilities. Each of these spreadsheets has been contributed

by a practitioner, and it can be assumed that these spreadsheets are carefully engineered and tested by the time they appear on the Web site.

Our search for sample spreadsheets was not strictly random. We asked potential contributors to give us important spreadsheets that were completed and had been in use for some time. We eliminated any spreadsheet that contained only data and no formulas. We also eliminated spreadsheets that contained a very small number of cells or that were too simple to warrant auditing (using only a couple of SUM functions, for example). We eliminated some spreadsheets that involved broken links to other spreadsheets. We also eliminated some that used complex macros. Thus our sample is by design not representative of the class of very small spreadsheets, which we believe are very common in practice. It is also not representative of the most elaborate spreadsheets in use. (Grossman et al. [2005] reported on some very large Excel models used in critical applications.)

While our sample is not strictly *random*, it is certainly *representative* of the general population of spreadsheets (with the caveats cited above). The sample includes spreadsheets from many different types of organizations, spreadsheets created by both novice and expert developers, and spreadsheets that span a broad range from small and simple to large and complex.

Auditing Software

When we began this research, we were aware of the auditing and error checking tools available in Excel, but we knew little about third-party auditing software. As we learned more, we realized that many auditing tools are available, and some research suggests that appropriate software can make auditing easier. Accordingly, we decided to include a small number of auditing tools in our protocol. From a list of roughly 50 such tools supplied by Roger Grinde of the University of New Hampshire, we selected 10 to review in some detail. Eventually we settled on two to use in our auditing protocol: *XL Analyst* (http://www.codematic.net/default.

htm) and *Spreadsheet Professional* (http://www.spreadsheetinnovations.com/).

XL Analyst is an Excel add-in that evaluates 28 aspects of a spreadsheet, from "Formulas evaluating to an error" to "Use of SUMIF." *Spreadsheet Professional* is a collection of tools for building, testing, analyzing, and using spreadsheets. In our auditing protocol we made use of two features of this tool: maps and calculation tests. The mapping tool creates a coded version of each worksheet in a workbook. Each non-blank cell is coded as a label, a number, or a formula. It also shows which formulas have been copied from an original formula. The calculation test tool checks a workbook for 25 conditions such as "Blank cell referenced." For each of these categories it gives the number of cells involved and their cell addresses.

Selection and Training of Auditors

Our spreadsheet auditors were undergraduates, graduate students, or recent alumni from business or engineering schools. All had several years of experience with Excel, usually in a business setting. None were professional programmers or professional spreadsheet developers.

For an auditor, the first step in training was to study the auditing protocol, which describes in detail the stages an audit goes through and the data to be gathered at each stage. Then an auditor was given two or three spreadsheets to audit. The finished audited workbooks were then reviewed by the researchers for adherence to the auditing protocol and for the quality of the audit. In total, auditor training took an average of 10 hours.

Auditing Protocol

The auditing protocol we used in this research is a highly detailed document that specifies the steps to take in auditing a spreadsheet for size, complexity, several types of qualitative features, and errors. (A complete description of the protocol is available at http://mba.tuck.dartmouth.edu/spreadsheet/index.html.)

The auditing protocol involves the following 11 steps:

1. Run the two software tools *XLAnalyst* and *Spreadsheet Professional.*
2. Transfer selected results from the software tools to a data record sheet.
3. Record the purpose of the workbook and each worksheet.
4. Examine workbook for use of Excel functions.
5. Review the results of *XL Analyst* and use them to locate errors.
6. Review the *Spreadsheet Professional* maps and use them to locate errors.
7. Review the *Spreadsheet Professional* calculation tests and use them to locate errors.
8. Review all formulas not already reviewed for errors.
9. Conduct sensitivity analyses to uncover errors.
10. Rate the workbook on various aspects of spreadsheet design (e.g., use of modules).
11. Record the total time taken by the audit and record comments on any special situations encountered.

This particular sequence of steps evolved over several months of testing. During this time we trained auditors and tested the protocol ourselves and through our assistants on dozens of operational spreadsheets. More details on the auditing protocol are available in Powell et al. (2006b).

Error Taxonomy

One of the challenges of spreadsheet error research is how to categorize errors. As we pointed out earlier, many different error classifications have been offered. Most of these suffer from the same flaw: errors that arise from different causes cannot be distinguished by an auditor. For example, when we encounter an error in a formula, we rarely can determine whether the error was due to sloppy typing, lack of domain knowledge, lack of Excel knowledge, a subsequent user changing the formula, or some other cause. We can, however, easily detect some formulas that give the wrong result. With somewhat more effort we can sometimes identify formulas that would give a wrong result with a different set of numerical inputs. An example of the latter error is an IF formula whose True condition is active in the current version and is correct, but whose False condition can be activated with a different set of inputs and is incorrect. We can also identify many practices that are likely to cause errors as the spreadsheet is used or that simply will make it harder than necessary to use the spreadsheet productively. The most common of these poor practices is placing numerical inputs directly into formulas. Other poor practices include limited or nonexistent documentation, duplication of inputs, illogical physical layout, and so on. A fundamental question in spreadsheet error research is where to draw the line between errors and poor practices.

Initial Error Taxonomy

Recall that our study required our auditors to detect errors in the spreadsheets themselves, without access to external information or to the spreadsheet developer. Thus the types of errors we sought had to be *observable* by our auditors. A second requirement for our taxonomy was *consistency*: a single error must with high probability be placed in the same category by different auditors.

Our first attempt to define an operational taxonomy used three categories:

1. wrong result with current inputs,
2. wrong result with different but allowable inputs, and
3. poor practice.

After some experience with this classification, we discovered that the category of poor practices was so ill-defined that no two auditors could agree in specific instances. As a result we eliminated this category from our

protocol. We also discovered that the second category, wrong result with different inputs, was ambiguous because it forced us to decide which numerical inputs were allowable or plausible. For example, most spreadsheets can be made to generate errors with unlikely inputs, such as text instead of numbers.

Interim Error Taxonomy

Eventually we settled on the following six error types that our experience with auditing suggested were well defined in theory and could be identified with high reliability in practice:

1. *Logic Error*—a formula is used incorrectly, leading to an incorrect result.
2. *Reference Error*—a formula contains one or more incorrect references to other cells.
3. *Placing Numbers in a Formula*—one or more numbers appear in formulas.
4. *Copy/Paste Error*—a formula is wrong due to inaccurate use of copy/paste.
5. *Data Input Error*—an incorrect data input is used.
6. *Omission Error*—a formula is wrong because one or more of its input cells is blank.

More details on this taxonomy, including examples of situations in which a cell would or would not be classified as an error, are available in Powell et al. (2006b). Not surprisingly, we found that we had to test our categorization extensively before we could be confident that our auditors could implement it effectively. Even after auditing dozens of spreadsheets, we found it necessary for the senior researchers to review each audited workbook and to make judgments as to which of the errors cited would be included. (In the end, our overall inter-rater reliability was on the order of 90%).

Our experience has been that errors come in hundreds of subtle variations, many of which do not fit neatly into any classification scheme. If nothing else, this suggests that we should be skeptical of drawing firm conclusions from previous studies of errors in which both the errors themselves and the procedures used to find errors are not clearly spelled out.

Final Error Taxonomy

After completing our audits of 50 spreadsheets using the taxonomy described above, we returned to the question that motivated our initial taxonomy: which of these errors actually give wrong results and which are merely poor practices?

In order to investigate this distinction, we reviewed all the instances of errors we found and categorized each one as either a *wrong result* or a *poor practice*. While this distinction may appear to be objective, there are many subtle cases that require interpretation. For example, simply embedding a number in a formula is usually a poor practice, not a wrong result. However, we encountered situations in which one number was embedded in a formula and a related but *different* number was included elsewhere among inputs. For example, an input module might contain the number 4% next to the heading "Inflation," but the actual inflation rate used in a formula was 3.5%. We categorized this case as a wrong result, even though the actual spreadsheet may have been numerically correct. Again we stress that determining whether a particular cell is an error (or gives a wrong result) is not a purely objective task.

ERRORS FOUND DURING AUDITS

We audited 50 spreadsheet workbooks and in three of them were unable to find any errors of the types included in our auditing protocol. This implies a spreadsheet error rate of 94%, which is identical to the spreadsheet error rate estimated by Panko (2005).

In the remaining 47 spreadsheets, we found a total of 483 instances of errors involving a total of 4,855 error cells. (An error instance is a single occurrence of one of the six errors in our taxonomy. Typically a single instance involves

more than one cell.) Thus, on average each error instance involved 10.05 error cells. The average cell error rate over all 270,722 formulas audited was 1.79%. This is considerably lower than the 5.2% cell error rate cited by Panko (2005), although somewhat higher than the corrected estimate of 1.3% we derived earlier from Panko's data.

Table 3 shows how the error instances and error cells were distributed by error type.

Hard-coding errors were the most common (37.7% of instances and 43.5% of cells), followed by reference errors (32.9% of instances and 22.1% of cells) and logic errors (21.9% of instances and 28.6% of cells). The remaining three categories—copy/paste, omission, and data input errors—together accounted for less than 5% of instances and cells.

As explained earlier, not all of the 483 error instances we identified involved wrong results. In fact, roughly 40% of these were hard-coding errors, in which a number was included in a formula rather than a cell reference; most of these did not produce wrong results. In order to determine how many errors actually gave rise to wrong results, we categorized each instance as a wrong result or a poor practice. With this more restrictive definition, we found that 7 of the 50 spreadsheet workbooks we audited were error-free, for a spreadsheet error rate of 86%.

In the remaining 43 spreadsheets, we found a total of 281 instances of wrong results involving a total of 2,353 cells. Using this definition, the average cell error rate over all 270,722 formulas audited was 0.87%. This is, of course, substantially lower than the 1.79% cell error rate we estimated using our more inclusive definition of error, and much less than Panko's estimate of 5.2%.

Table 4 shows how wrong results break down by type of error. Most, but not all, hard-coding errors were categorized as poor practice. Most, but not all, of the other error categories were labeled wrong result. In total, 58% of error instances and 48% of error cells were categorized as wrong result.

Our results suggest that the average cell error rate across all spreadsheets is in the range of 1% when we use a restrictive definition and 2% when we use an inclusive definition. But the average error rate does not tell us how the error rate varies across spreadsheets. The received wisdom is that errors are unavoidable in spreadsheets, as they are in most facets of life. In fact, Panko (2005) cites an error rate of 5% as the norm for many different activities. This suggests that we should expect most spreadsheets to have a cell error rate close to the average.

Figure 1 shows the distribution of cell error rates using our original, inclusive definition.

Using that definition, 46% of our sample spreadsheets had error rates below 2%; 70% had error rates below 5%. However, several spreadsheets had error rates above 10%; in fact one had a cell error rate of 28%. Figure 2 shows

Table 3. Errors categorized by type

Error Type	Instances	Cells
Hard-Coding	182 (37.7% of total)	2,111 (43.5% of total)
Reference	159 (32.9%)	1,074 (22.1%)
Logic	106 (21.9%)	1,389 (28.6%)
Copy/Paste	17 (3.4%)	206 (4.2%)
Omission	13 (2.7%)	65 (1.3%)
Data Input	6 (1.2%)	10 (0.2%)
Total	**483**	**4,855**

Table 4. Wrong results and poor practices

Instances		
Error Type	**Wrong Results**	**Poor Practices**
Hard-Coding	31 (11.0% of total)	151 (74.8% of total)
Reference	137 (48.8%)	22 (10.9%)
Logic	89 (31.7%)	17 (8.4%)
Copy/Paste	14 (5.0%)	3 (1.5%)
Omission	7 (2.5%)	6 (3.0%)
Data Input	3 (1.1%)	3 (1.5%)
Totals	**281**	**202**
Cells		
Error Type	**Wrong Results**	**Poor Practices**
Hard-Coding	191 (8.1% of total)	1,920 (76.7% of total)
Reference	857 (36.4%)	217 (8.7%)
Logic	1,073 (45.6%)	316 (12.6%)
Copy/Paste	203 (8.6%)	3 (0.1%)
Omission	26 (1.1%)	39 (1.6%)
Data Input	3 (0.1%)	7 (0.3%)
Totals	**2,353**	**2,502**

the distribution of cell error rates using the more restrictive definition. In this case 68% had cell error rates below 2% and 90% were below 5%. The highest single error rate was 17%.

We began this article by noting that the received wisdom on spreadsheet errors maintains that errors average 5% of cells and this rate is consistent across spreadsheets. Our results give a very different impression. The average cell error rate appears to be closer to 1% or 2%, depending on how errors are defined. Perhaps even more important, roughly half of the spreadsheets we tested had rates below these levels, although a few had astonishingly high rates. These results suggest that errors are not a constant factor, but presumably depend on the situation being modeled and the developer's skills. Finally, we document for the first time in the published literature the sources of errors. Thirty to forty percent of errors are due to embedding numbers in formulas. The next most common error type is reference errors, followed by logic errors. The remaining error types are far less common.

SUMMARY AND DIRECTIONS FOR FUTURE RESEARCH

We have audited 50 operational spreadsheet workbooks and discovered 483 error instances involving 1.79% of all formulas. This study is unique in several ways. First, this is the largest published sample of spreadsheets that has been audited for errors. In fact, our sample of 50 operational spreadsheets exceeds the *total* number of spreadsheets (43) cited in Panko's (2005) meta-study to estimate cell error rates. Second, we have been explicit about our definition of errors. Our work in this area has shown that errors manifest themselves in a wide variety of guises, so estimated cell error rates may be highly sensitive to how one defines errors. In

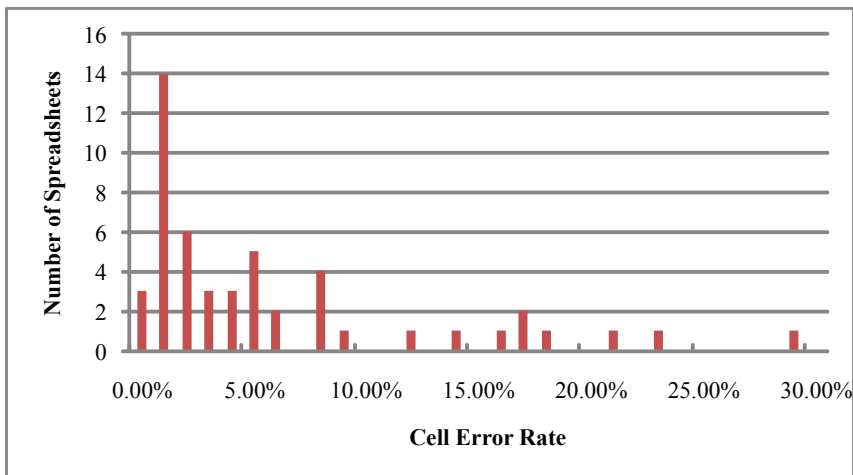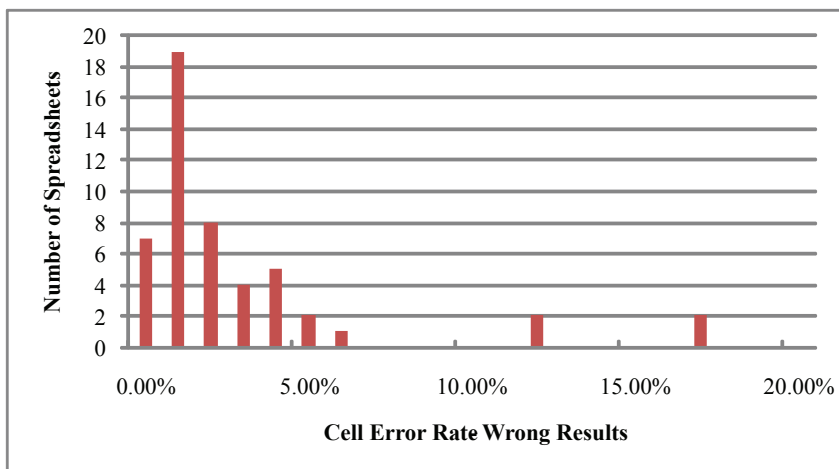*Figure 1. Distribution of cell error rates across spreadsheets*

Figure 2. Distribution of cell error rates across spreadsheets (wrong results)

fact, our own estimates range from 0.87 to 1.79% of cells, depending on the definition of error. Finally, the method we used to detect errors is fully documented and available to other researchers. In most previous studies the methods used to search for errors were not documented at all; in others only a brief summary of the methods is offered. Our study involves an explicit, detailed auditing protocol. It also used two types of auditing software to assist in the search for errors.

Our results confirm the general belief among those who have studied spreadsheets that errors are commonplace. The specific estimates we have made are somewhat lower than what has become received wisdom, but we note that previous studies of spreadsheet errors cannot easily be compared since they use different (and generally undocumented) auditing methods and definitions of errors. We now have sound evidence for the hypothesis that around 1% of all formulas in operational spreadsheets are in error.

Obtaining reliable estimates of cell error rates is only a first step toward understanding the fundamental question of the impact that spreadsheets have on the quality of decision making in organizations. Errors in formulas should be avoided, of course, but by themselves they do not necessarily translate into significant errors in the outcomes of spreadsheet use. The next step we envision in this line of research is to estimate the *numerical impact* of errors in spreadsheets. This investigation would require working closely with the developers of the spreadsheet to determine which outputs are most important and to determine the impact on those outputs of correcting errors.

Errors in spreadsheet data and formulas are not the only possible causes of errors in spreadsheet use. In fact, many of the press accounts reported by EUSPRIG involve *misuse* of spreadsheets. For example, spreadsheet results can be sorted incorrectly, or an out-of-date version of a spreadsheet may be used. More generally, poor decisions based on spreadsheet models can arise by modeling the wrong problem or by misinterpreting the results. These types of errors cannot be identified by auditing the data and formulas of the spreadsheet itself. To determine how frequent and damaging these types of errors are to decision making would involve a kind of ethnographic research, in which spreadsheet users were closely observed in their natural surroundings over months or years.

Finally, very little research (and none outside the laboratory) has been conducted to identify methods for avoiding or mitigating the impacts of errors in operational spreadsheets. One promising approach would be to audit a sample of spreadsheets from different organizations and attempt to correlate error rates with corporate or individual practices around spreadsheet use. Many practitioners have offered advice on what constitutes good spreadsheet practice, but there is no scientific evidence from the field yet to back up these claims.

ACNOWLEDGMENT

REFERENCES

Butler, R.J. (1992). Personal communication to Ray Panko.

Clermont, J.M. (2002). A spreadsheet auditing tool evaluated in an industrial context. *Proceedings of the European Spreadsheet Risk Interest Group Annual Conference,* Cardiff, Wales.

Coopers and Lybrand. (1997). Retrieved from http://www.planningobjects.com/jungle1.htm

Croll, G.J. (2005). The importance and criticality of spreadsheets in the city of London. *Proceedings of the European Spreadsheet Risks Interest Group Conference,* Greenwich, UK.

EUSPRIG. (n.d.). European Spreadsheet Risks Interest Group stories. Retrieved from http://www.eusprig.org/stories.htm

Grossman, T.G., Mehrotra, V., & Özluk, Ö. (2005). *Spreadsheet information systems are essential to business.* Working paper, University of San Francisco, USA.

Hicks, L. (1995). Personal communication to Ray Panko.

KPMG. (1998). *Supporting the decision maker—a guide to the value of business modeling.* Retrieved from http://www.kpmg.co.uk/uk/services/manage/press/970605.html

Lawrence, R.J., & Lee, J. (2004). Financial modelling of project financing transactions. *Institute of Actuaries of Australia Financial Services Forum, 19.*

Lukasik, T. (1998). Personal communication to Ray Panko.

Panko, R. (2005). *What we know about spreadsheet errors.* Retrieved September 2, 2006, from http://panko.cba.hawaii.edu/ssr/Mypapers/whatknow.htm

Panko, R., & Halverson, R. (1996). Spreadsheets on trial: A survey of research on spreadsheet risks. *Proceedings of the 29th Annual Hawaii International Conference on Systems Sciences* (pp. 326-335).

Powell, S., Baker, K., & Lawson, B. (2006a). *A critical review of the literature on spreadsheet errors.* Spreadsheet Engineering Research Project working paper, Dartmouth College, USA.

Powell, S., Baker, K., & Lawson, B. (2006b). *An auditing protocol for spreadsheet models.* Spreadsheet Engineering Research Project working paper, Dartmouth College, USA.

Purser, M., & Chadwick, D. (2006). Does an awareness of differing types of spreadsheet errors aid end-users in identifying spreadsheet errors? *Proceedings of the European Spreadsheet Risk Interest Group Annual Conference* (pp. 185-204), Cambridge, UK.

Rajalingham, K., Chadwick, D., & Knight, B. (2000). Classification of spreadsheet errors. *Proceedings of the European Spreadsheet Risks Interest Group Annual Conference* (pp. 23-34), Greenwich, England.

Sommerville, K. (2004). *Software engineering* (7th ed.). Boston: Addison-Wesley.

Steve Powell is a Professor at the Tuck School of Business at Dartmouth. His primary research interest lies in modeling production and services processes, but he has also been active in research in energy economics, marketing, and operations. At Tuck, he has developed a variety of courses in management science, including the core Decision Science course and electives in the Art of Modeling, Business Process Redesign, and Applications of Simulation. He originated the Teacher's Forum column in Interfaces, and has written a number of articles on teaching modeling to practitioners. He is the academic director of the INFORMS Annual Teaching of Management Science Workshop. In 2001 he was awarded the INFORMS Prize for the Teaching of Operations Research/Management Science Practice. He is the co-author with Kenneth Baker of The Art of Modeling with Spreadsheets (Wiley, 2004).

Ken Baker is a faculty member at Dartmouth College. He is currently Nathaniel Leverone Professor of Management at the Tuck School of Business and also adjunct professor at the Thayer School of Engineering. At Dartmouth, he has taught courses relating to decision science, manufacturing management, and environmental management. Over the years, much of his teaching and research has dealt with production planning and control, and he is widely known for his textbook Elements of Sequencing and Scheduling, in addition to a variety of technical articles. He has served as the Tuck School's associate dean and directed the Tuck School's management development programs in the manufacturing area. In 2001 he was named a Fellow of INFORMS's Manufacturing and Service Operations Management (MSOM) Society, and in 2004 a Fellow of INFORMS. He is the co-author with Stephen Powell of The Art of Modeling with Spreadsheets (Wiley, 2004).

Barry Lawson is a research associate at the Tuck School of Business at Dartmouth and is also a visiting scholar in the geography department of the college. He founded and has served as president of Barry Lawson Associates, a consulting firm, since 1978. As visiting scholar, he coordinates the development of an atlas of the upper Connecticut River Watershed in New Hampshire and Vermont. As research associate at Tuck he serves as the program manager for the Tuck Spreadsheet Engineering Research Project. Lawson has taught in graduate programs at Boston University and Wayne State University as well as in short courses at Bentley College. He has moderated a host of public hearings for local, state and federal governments on controversial environmental and energy- and waste-related projects, and has considerable experience in group facilitation, conflict resolution and simulation design.