December 2002

A FOLLOW-UP STUDY OF ACTUARIAL DETECTION OF DATA ERRORS

Barbara Klein
University of Michigan-Dearborn

Follow this and additional works at: http://aisel.aisnet.org/amcis2002

A FOLLOW-UP STUDY OF ACTUARIAL DETECTION OF DATA ERRORS

Barbara D. Klein
University of Michigan-Dearborn
bdklein@umd.umich.edu

Abstract

As organizations become increasingly dependent on computerized databases, the impact of errors in these databases is likely to increase. If users are able to detect these errors (and the errors can be corrected), the problems associated with data errors may be lessened. Prior research shows that some users of data can find and correct data errors. A pilot study of practicing actuaries showed that the degree to which data were inspected for errors was influenced by the perceived likelihood of errors and by perceived costs and benefits of a more accurate dataset. The study proposed here reports the findings of a second set of interviews with practicing actuaries. The interview protocol used in this study is based on the theoretical insights gained from the earlier pilot study.

Introduction

In general, organizational databases are not free of data errors (e.g., Laudon 1986). The presence of data errors poses potential problems for organizations that are increasingly dependent on databases for support of critical organizational functions. Additionally, efforts to improve organizational performance may be hampered by errors in organizational databases (Redman 1998).

While efforts to automatically validate data as they are input to systems and efforts such as automated data cleansing are certainly valuable, few organizations can afford to ignore the potential for human involvement in the detection and correction of data errors (Orr 1998). The best approach to data quality improvement in most organizations is likely to include both automated methods and human efforts. This study examines the human side of error detection by focusing on practicing actuaries, a group of end users who have been found to be particularly good detectors of data errors in prior research. Theoretical insights gained from these users may be useful in efforts to improve the effectiveness of end users working in other professional domains.

Background

Early research on data quality proposed definitions of the concept. No single definition of data quality has been accepted by everyone working in the area. However, accuracy, currency, and completeness are generally recognized as important aspects of data quality. The research proposed in this paper focuses primarily on the accuracy dimension of data quality. A second area of concern has been the measurement of error rates. In general, studies have found that organizational databases are not free of data errors, and in some cases error rates as high as ten percent have been found (e.g., Laudon 1986).

Research conducted in the late 1990s provides evidence that end users are effective detectors of data errors under some conditions. A pilot study initiating this research stream found that end users working as practicing actuaries were able to detect errors in data (Klein 1997). These end users were more likely to spend time reviewing data for errors when they believed the likelihood of errors being present was high and when they believed the benefits of a more accurate dataset outweighed the costs of finding and correcting data errors. Base rate expectations developed through direct experience with data, incentive structures, and error detection goals have been shown to affect performance in the detection of data errors in laboratory experiments (Klein et al. 1997; Klein 2001). A field study was conducted to link the findings of these laboratory experiments to practice in organizations. The findings of the field study show that strong informal incentives, perceptions about the materiality of data errors, and perceptions

about the base rate of errors in data influence the detection of data errors. These conclusions are based on interviews with four types of end users (actuaries, municipal bond analysts, consumer product managers, and inventory managers). In this study, the interviews with practicing actuaries conducted as part of the field study will be analyzed. It is important to note that none of these actuaries were interviewed in the initial pilot study of practicing actuaries.

A Theory of Error Detection

A theory of error detection has been developed using Campbell's (1990) theory of individual task performance and theories of effort and accuracy in decision making (Payne 1982).

Campbell's (1990) theory of individual task performance argues that individual task performance is affected by experience, knowledge, and effort. In this study, task performance is viewed as the successful or unsuccessful detection of a data error. Both declarative knowledge and procedural knowledge affect error detection performance. Differences in expectations about the base rate of errors in data and assessments of the payoffs of error detection affect error detection through the choices related to effort.

Experience and Knowledge in Error Detection. Expertise rests on a foundation of significant amounts of experience (e.g., Ericsson and Chase 1982). In the task of detecting data errors, each error encountered by an end user provides an opportunity for experience in this task. With enough experience, an end user may develop expertise in error detection. Users working with data with a high rate of errors have more opportunities for these types of experience and hence for the development of expertise.

The Role of Effort in Error Detection. Theories of effort and accuracy in decision making argue that specific task requirements affect the mental resources or effort that humans devote to a task. For example, Payne (1982) showed that task requirements affect the selection of information processing strategies. In the task of detecting data errors, expectations about the base rate of errors in data and user assessments of the payoffs of error detection may affect users' effort expended to detect data errors. Choices about the degree of effort to expend in the detection of errors will in turn influence performance. Several factors influencing these choices are suggested by an analysis of data collected in an earlier pilot study on the use of imperfect data by actuaries (Klein 1997). These factors are discussed below.

Expectations about the Base Rate of Errors in Data. Research shows that decision makers are sensitive to base rates when asked to generate hypotheses in diagnostic tasks (Weber et al. 1993). There is similar evidence from the pilot study of the actuaries showing that expectations about the base rate of errors in a source of data influence effort expended in error detection. Greater effort may be devoted to error detection when users expect more errors in data.

Payoffs of Error Detection. A task described by a subject in the pilot study of actuaries will be used to illustrate the impact of assessments of payoffs on error detection performance. The task was the determination of whether an organization's financial reserves for its pension fund were sufficient. This judgment depends in part on the pay rate and the number of years of organizational service of each employee in the organization. Data provided by a client included this information along with other personnel information for each employee as of the end of the year. Imagine a specific case in which this data (as of the end of 2002) contains a record holding information about an accountant in a position requiring a CPA certificate in which the value of the Date of Birth field is "December 31, 1977" and the value of the Number of Years of Service field is "10".

An actuary using this data might or might not suspect that the data in one of these fields is wrong (i.e., it is unlikely that a firm would hire an accountant at the age of 15). An actuary analyzing a pension fund might be likely to detect this error because it is material to the judgment about the sufficiency of the firm's pension reserves. On the other hand, a payroll manager reviewing the same dataset might be unlikely to find the error because errors in the Date of Birth and Number of Years of Service fields are not material to a firm's payroll.

Materiality. Thus, beliefs about the materiality of a potential error may influence the degree of effort expended to detect the error. Users may expend more effort to detect errors that they believe will have a significant impact on their calculations or decisions. There is evidence from the pilot study of actuaries that the impact of data errors on the work being performed using the data is considered in the determination of the level of effort to expend in error detection. For example, one actuary stated that there are some types of errors that he does not try to detect when pricing insurance because the errors would not have a significant impact on his calculations.

Incentives. Organizational incentives may also play an important role in users' assessments of the payoffs of error detection. For example, an incentive system that discourages the use of time to investigate and correct errors may create an environment in which many errors in data go unnoticed.

Ease of Verification and Correction. The degree of effort expended to detect an error may also be affected by the ease with which an error can be corrected. Users may not expend much effort to detect errors if it is difficult to confirm that a suspected error is actually an error or if a confirmed error cannot be corrected.

Methodology

Interviews have been conducted with five practicing actuaries. None of the actuaries interviewed in the earlier pilot study were interviewed for this study. Instead, new participants were recruited. Potential participants were asked to participate in a study of the use of data in their work. To control for selection bias, the terms "data quality" and "error detection" were not used when recruiting participants.

Data were collected using a semi-structured interview. The interview protocol was developed using the theoretical framework discussed above. The interviews have been recorded and transcribed. An analysis of the interview transcripts using methodologies outlined by Miles and Huberman (1994) and King (1994) will be conducted. A coding scheme based on the theoretical framework has been developed. Transcripts were coded for expectations about the error rate, materiality, incentives, ease of verification and correction, and successful and unsuccessful error detection episodes. The transcripts have been coded by two independent coders using this scheme. Analysis of the coded transcripts will be presented at the conference.

Conclusion

The results of this study will generate fresh theoretical insights to explain human use of data containing errors. The results may be beneficial to researchers in suggesting new theoretical constructs for study in future empirical studies. Practitioners may benefit from the research if methods employed by the actuaries to detect data errors suggest error detection strategies that can be taught to end users in other professional domains.

References

Campbell, J. P. "Modeling the Performance Prediction Problem in Industrial and Organizational Psychology," In M. D. Dunnette and L. M. Hough (Eds.), *Handbook of Industrial and Organizational Psychology* (2nd ed., Vol. 1, pp. 687-732). Palo Alto, CA: Consulting Psychologists Press, Inc., 1990.

Ericsson, K. A., and Chase, W. G. "Exceptional Memory," *American Scientist* (70), 1982, pp. 607-614.

King, N. "The Qualitative Research Interview," In C. Cassell and G. Symon (Eds.), *Qualitative Methods in Organizational Research* (p. 14-36). Thousand Oaks, CA: Sage Publications, 1994.

Klein, B. D. "How Do Actuaries Use Data Containing Errors? Models of Error Detection and Error Correction," *Information Resources Management Journal* (10:4), 1997, pp. 27-36.

Klein, B. D. "Detecting Errors in Data: Clarification of the Impact of Base Rate Expectations and Incentives*," Omega: International Journal of Management Science* (29), 2001, pp. 391-404.

Klein, B. D., Goodhue, D. L., and Davis, G. B. "Can Humans Detect Errors in Data? Impact of Base Rates, Incentives, and Goals," *MIS Quarterly* (21:2), 1997, pp. 169-194.

Laudon, K. C. "Data Quality and Due Process in Large Interorganizational Record Systems," *Communications of the ACM* (29) 1986, pp. 4-11.

Miles, M. B., and Huberman, A. M. *Qualitative Data Analysis*, Thousand Oaks, CA: Sage Publications, 1994.

Orr, K. "Data Quality and Systems Theory," *Communications of the ACM* (41:2) 1998, pp. 66-71.

Payne, J. W. "Contingent Decision Behavior," *Psychological Bulletin* (92) 1982, pp. 382-402.

Redman, T.C. "The Impact of Poor Data Quality on the Typical Enterprise," *Communications of the ACM* (41:2) 1998, pp. 79-82.

Weber, E. U., Bockenholt, U., Hilton, D. J., and Wallace, B. "Determinants of Diagnostic Hypothesis Generation: Effects of Information, Base Rates, and Experience," *Journal of Experimental Psychology: Learning, Memory, and Cognition* (19), 1993, pp. 1151-1164.