

Data Description Sheet
“Whistleblowers and Outcomes of Financial Misrepresentation Enforcement Actions”

1. A description of which author(s) handled the data and conducted the analyses.

Gerald Martin and Jaron Wilde have both handled all data and jointly conducted the empirical analyses.

2. A detailed description of how the raw data were obtained or generated, including data sources, the date(s) on which data were downloaded or obtained, and the instrument used to generate the data (e.g., for surveys or experiments). We recommend that more than one author is able to vouch for the stated *source* of the raw data.

OSHA whistleblower data – Andy Call used Freedom of Information Act (FOIA) requests to obtain the OSHA whistleblower data from each of the national OSHA office and each of the 10 regional OSHA offices. Each of the ten OSHA offices responded with the relevant data from mid-2009 through early 2011.

Enforcement action data – Gerald Martin has hand collected the enforcement action data over the last 20+ years. Although the dataset itself is proprietary, the sources of the enforcement action data are publicly available, as discussed briefly in the paper and in other enforcement action research (e.g., Karpoff, Lee, and Martin 2008, *Journal of Financial Economics*; Karpoff, Koester, Lee, and Martin, *The Accounting Review*, *forthcoming*). Numerous studies in top accounting and finance studies have used the data. Gerald Martin is happy to provide any details about the dataset, including a comprehensive list of studies using the data.

Other data – Jaron Wilde collected other data from publicly available data sources (e.g., Compustat).

3. If the data are obtained from an organization on a proprietary basis, the authors should *privately* provide the editors with contact information for a representative of the organization who can confirm data were obtained by the authors. The editors would not make this information publicly available. The authors should also provide information to the editors about the data sharing agreement with the organization (e.g., non-disclosure agreement, any restrictions imposed by the organization on the authors with respect to publishing certain results).

Not applicable.

4. A complete description of the steps necessary to collect and process the data used in the final analyses reported in the paper. For experimental papers, we require information about subject eligibility and/or selection, as well as any exclusion criteria.

All relevant steps are detailed in the manuscript and/or outlined in the computer programs described below.

5. **Prior to final acceptance of the paper, the computer program used to convert the raw data into the dataset used in the analysis plus a brief description that enables other researchers to use this program. *Instead of the program*, researchers can provide a detailed step-by-step description that enables other researchers to arrive at the same dataset used in the analysis. The purpose of this requirement is to facilitate replication and to help other researchers understand in detail how the sample was formed, including the treatment of outliers, Winsorization, truncation, etc. This programming is in most circumstances not proprietary. However, we recognize that some parts of the data generation process may indeed be proprietary or otherwise cannot be made publicly available. In such cases, the authors should inform the editors upon submission, so that the editors can consider an exemption from this requirement.**

Three code and data files accompany the final paper:

1. A Stata dataset of the underlying data used in the paper, including variables, company name, and an SEC identifier (“recid”) that relates to the specific action in the SEC enforcement action dataset collected by Gerald S. Martin.
2. A Stata .do file that provides the code one can use to replicate the analyses in the paper with the data used in the paper (item #1 above).
3. A SAS program that provides code used to collect basic information on Compustat firms from the WRDS server, as used in the description statistics in Table 2, Panel C and Table A1, Panel C in the online appendix.

Instructions: After loading the dataset in Stata (item #1), run the code from the Stata .do file (item #2) to conduct the analyses presented in the paper.

6. **An assurance that the data and programs will be maintained by at least one author (usually the corresponding author) for at least six years, consistent with National Science Foundation guidelines.**

The data and code to estimate the primary analyses in the paper have been submitted with the paper for online access. The authors will also maintain a shared Dropbox account with the complete set of code and data for at least six years.