Do Jobseekers Value Diversity Information? Evidence From a Field Experiment and Human Capital Disclosures

Data Description Sheet

Choi, Pacelli, Rennekamp, and Tomar

January 2023

This document provides information required to comply with the Data and Code Sharing Policy of the *Journal of Accounting Research*.

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

Professor Choi performed the primary analysis of the experimental data using the R Statistical Programming language with Sarah McDonald, Fiona Sequeira, and Jisang (Eddie) Yu, a research assistant at Stanford GSB. Professor Tomar performed the empirical analysis using the R Statistical Programming language. Professor Rennekamp conducted the survey analysis using JMP and SAS. All the authors except for Professor Pacelli have access to the experimental and archival data.

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).

Our experiment was conducted over an eleven-week period beginning in June of 2021 and ending in August of 2021. All data was captured and stored by Zippia over the course of the study. Zippia collects information on: 1) whether each email was opened by a recipient, 2) how many jobs were clicked on by a user within an open email, 3) which companies were associated with each job listing that was clicked on, and 4) the "position" of each job in the list that a participant receives (e.g., if a user clicks on the third job listed in an email message, this would be recorded as having a position of "3"). After collection, the data is anonymized by Zippia for participants' privacy, and provided to the author team for analysis.

To provide a better understanding of why jobseekers do or do not value diversity information, we conducted a novel follow-up survey that yielded responses from many of our original participants. On December 9th, 2021, we emailed the original participants and provided them with the opportunity to complete a follow-up survey in exchange for the opportunity to win a small

monetary award. The survey remained open through December 20th, 2021, and participants received one follow-up reminder email. The survey instrument is included in the internet appendix.

Regarding the archival analysis, Professor Tomar used the data that are downloaded from Compustat's Fundamentals Annual Data and the CRSP Stock File. Professor Choi and Tomar downloaded 10-Ks and parsed out human capital disclosures with support from the Data, Analytics, and Research Computing team at Stanford GSB. All archival data were downloaded March and November of 2021 and analysis was performed until December of 2022. Professor Choi and Tomar used the Armed Conflict Location & Event Data Project (ACLED) data, which were downloaded in August of 2021, to measure Black Lives Matter protests.

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.**

Chris Kolmar can be reached via e-mail at ckolmar@zippia.com. Chris Kolmar can confirm the data was collected on the dates that we specify above, and that the data was given directly to us from Zippia with the requirement that it remain confidential.

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 and survey papers, we require information about the instructions and instruments used to generate the data, subject eligibility and/or selection, as well as any exclusion criteria.**

This information is provided in the paper. Detailed instructions in collecting and processing the experimental data can be found in Sections III, IV, and V of the manuscript.

5. **The computer programs or code used to convert the raw data into the final dataset used in the analysis plus a brief description that enables other researchers to use this program. The purpose of this requirement is to facilitate replication and to help other researchers understand in detail how the raw data were processed, the final sample was formed, variables were defined, outliers were treated, etc. This code or programming is in most circumstances not proprietary. However, we recognize that some parts of the code or data generation process may be proprietary, including from the authors' perspective. Therefore, instead of the code or program, researchers can provide a detailed step-by-step description of the code or the relevant parts of the code such that it enables other researchers to arrive at the same final dataset used in the analysis. In such cases, the authors should inform the editors upon initial submission, so that the editors can consider an exemption from the code sharing requirement. Whenever feasible, authors should also provide the identifiers (e.g., CIK, CUSIP) for their final sample. Authors should consult our FAQ Sheet on the JAR website for further details.**

We describe the step-by-step description of the code provided in a separate document because we use proprietary datasets. We provide the list of GVKEYs for our final sample as a csv file.

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.**

Prof. Choi has stored all the datasets and code in the secured Stanford GSB server, Yen. He plans to maintain them for at least six years.