These errors were then manually corrected. Since OCR software systems may not flag all incorrect entries, various statistical tests were used to correct these mistakes and analyze the error rate. In the case of this project, the error rate is defined as the ratio of incorrect entries produced by the OCR software to the total number of recognized entries.

The first test involved calculating the p-value of each row of data; a p-