Checked for specificity of fuzzy-matching by comparing manual-found overlaps to fuzzy-matching overlaps.

| Manual-found overlap? | Fuzzy-match overlap? | Result: |
| --- | --- | --- |
| Yes | Yes | Correct |
| Yes | No | False Negative |
| No | Yes | False Positive |
| No | No | Correct |

**Our Findings(**[**link to spreadsheet we worked on**](https://docs.google.com/spreadsheets/d/1UX2H-2dG6H0o9Nz4oU0VdVPf7tZzHtpc4H48gOaoNn8/edit?gid=0#gid=0)**):**

For manual comparisons, we used the [FDA\_Approved\_Orphan\_Drug\_List](https://docs.google.com/spreadsheets/d/1SbpySVPpPDJM9E_sS7e9VD5BxOFEWq5E/edit?usp=sharing&ouid=111680076277879887268&rtpof=true&sd=true) file and the [FINALIZED\_CDSCO\_Dataset](https://docs.google.com/spreadsheets/d/1G9FWzetcS9YlcfKuvKTdKopPlCqYn-cyae9S3XzCTtw/edit?usp=sharing). We referred to the “validation\_report.csv” in Vikram’s GitHub for overlaps.

1st test Results: Tony manually checked entries 1-50, and found **0 false positives** and **0 false negatives**. All matches(2) were correct.

2nd test Results: Sahiti manually checked entries 51-100, and found **0 false positives** and **0 false negatives.** All matches(2) were correct.

3rd test Results: Tony manually checked 100 random entries from the remaining drugs(101-3725), and found **0 false positives** and **0 false negatives.** All matches(4) were correct.