

Table 5

Formatted: Left

## Addressing Commonly Encountered Problems in UI Wage Data Files

The Problem	How It Appears	What to Do
Duplicate records	Exact or partial duplicates by SSN, quarter, employer ID, and earnings. Could be two earnings amounts, or two employer IDs, in a quarter.	Verify that records are duplicates—could be two different jobs or <del>two stints at one job</del> <del>a data errors</del> . If <del>duplicates</del> <del>this is an error</del> , then delete one.
Outliers	Out-of-range values (extremely high or extremely low earnings). Could be shared SSN.	If possible, determine cause, then “treat” (impute) or delete. <del>See TDI blog post on dealing with outliers.</del>
Wrong identifiers collected at intake for some sample members	No data, when expected. Could also be a shared SSN issue.	Decide: (1) Can you live with the problem? (2) Can you improve the match (by submitting corrected identifiers)?
UI agency did not match all sample members	Record counts are too low. Could be a programming bug with the UI agency’s algorithm that incorrectly imported identifiers or incorrectly filtered out records.	Request a new file.

**Commented [JK1]:** Remove here and add as a NOTE at the bottom with a little more of an explain, and add a superscript here instead. We generally don’t link this way inside a table. Use this language in the NOTE: See more tips for dealing with outliers at [www.tanfdata.org/content/tdc-quick-tips-dealing-with-outliers-in-human-services-data?category%5B%5D=3](http://www.tanfdata.org/content/tdc-quick-tips-dealing-with-outliers-in-human-services-data?category%5B%5D=3).

NOTE: See more tips for dealing with outliers at [www.tanfdata.org/content/tdc-quick-tips-dealing-with-outliers-in-human-services-data?category%5B%5D=3](http://www.tanfdata.org/content/tdc-quick-tips-dealing-with-outliers-in-human-services-data?category%5B%5D=3).

**Commented [EY2]:** Will be posted on github resource page so need to update link in August.