This describes what happens in /CHRP/NEOCARE/Corrected Data-MH Jan 2019/dedupication.Rmd, which is essentially the MH side of the deduplication process.

* A list of study-qualifying MH patients was constructed as follows
  + The raw MH demographics and encounters data sets were loaded from their corresponding SAS files and merged together (inner join by PAT\_ID).
  + Patients’ [truncated] age in years at the time of encounter were calculated, and rows were removed if the patient was purportedly under 18 or over 115 when the encounter occurred.
  + If a patient then only had one unique encounter date (“CONTACT\_DATE”) associated with them, their rows were removed.
* The MH patient demographics data were then filtered to only include patients in the study-qualifying MH patients list, and that was further filtered to only include patients with a CCF Care-Everywhere ID (CEID). **This was sent to CCF for stage 1 matching (i.e, matching on CCF CEID).**
* Stage 1 matching was performed on the CCF side, and the results were sent back to MH. Stage 1 matches were extracted from the MH patient demographics table and set aside. **The MRN, patient ID, last name, first name, last four digits of social security number, and birth year of the remaining MH patients (i.e, the stage 1 non-matches) were sent back to CCF for stage 2 matching.**
* On the CCF side, a study ID was randomly assigned to all patients (CCF, MH, or both). Study id key tables were constructed for the MH patients (one for stage-1 matches and one for stage-1 nonmatches) and sent back to MH. These study id key tables were merged with the MH patient data to create a single MH study id key table at /NEOCARE/Corrected Data-MH Jan 2019/mh\_study\_id\_key.rds. This key table was subsequently merged with all patient data in order to limit the amount of explicit PHI that was subsequently sent from MH to CCF.