Following is a response from DHS Research & Data Analysis Director, Tom Pullum:

When I run the program on your data, after "tab _merge" I get 70,099 cases with _merge=1. These are people in the PR file who are never named as caregivers. There are 21 cases with _merge=2. These are children for whom the caregiver is not in the household (there really should not be any such cases). There are 4267 cases with _merge=3. These are children who are successfully matched with their caregiver.

I suggested you keep the cases with _merge =2 or =3. You could just keep the cases with _merge=3, because the 21 cases with _merge=2 really don't contain any information about the caregiver. I suggested keeping them because then you have a complete file of children who have caregivers.

I don't know why you have so many variables in the "keep hhid...." line. But if you want to keep variables like that you will have to distinguish between whether they refer to the child or to the caregiver. You can just construct more "rename...child" and "rename....caregiver" commands, for those variables.

Hope this helps.