Data Alert for RAND HRS Family Data 2014 (V1) (September 2018)

We released the RAND HRS Family Data 2014 (V1) in May 2018. A Family Data user has since encountered potential duplicate Respondent-kid records. As noted in our data codebook, the technique of matching records by OPN to track children longitudinally is limited by a number of factors in the HRS raw data, including assignment of more than one OPN to children in some households across waves. Consequently, the RAND HRS Family Data includes some Respondent-kid records with different OPNs and KIDIDs which are in fact the same child. This issue is also present in earlier versions of Family Data.

Looking at households containing children with different OPNs but identical ages (KwAGEBG) and genders (KwGENDER), as well as comparing number of in-contact children (HwNKID) to number of child records, we have estimated that up to 3% of children may have more than one Respondent-kid record.

During preparation for our next data release, we will further develop our algorithms linking children across waves, sorting through records not matched on OPNs to identify potential double records based on date of birth, gender and other information from HRS restricted data. It may not be possible to identify these cases as duplicates with complete accuracy. Confounding scenarios include twins, blended families and varying reports across years of child birthdates and names. Due to inconsistencies in the raw data, the links for such cases will remain a "best guess," and we will maintain the LINK variable which flags cases for which the cross-wave links are questionable.

It is worth noting that, despite this discovery, over 95% of Respondent-kid records are matched with a high degree of certainty after applying numerous algorithms developed by the RAND HRS Team to link Family Data records across HRS waves even given the limitations described here and outlined in the Family Data codebook (See Section 2, Linking Across Waves).