Subject: Re: duplicate caseid
Posted by Janet-DHS on Fri, 13 May 2022 14:59:54 GMT
View Forum Message <> Reply to Message

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

What survey and what file are you using? In the KR file, for example, there is a record for every child born in the past five years. caseid is the mother's ID code. Because many women had more than one child in the past five years, there will be several records with the same value of caseid, but children of the same mother will have different values of bidx (1, 2, etc.). In the IR file, there should never be a repeat of the same caseid, although very rarely we will find a duplicate. To check for duplicates in the IR file, in Stata, enter "gen ncases=1", then "collapse (sum) ncases, by(caseid)", then "tab ncases" and "list if ncases>1, table clean".