Subject: Re: duplicate caseid
Posted by Janet-DHS on Fri, 20 May 2022 19:45:29 GMT
View Forum Message <> Reply to Message

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

I have to tell you that a comment such as "I appreciate swift responses" will not accelerate our response to a forum question.

You are looking for a way to extract the different columns of caseid (or hhid) and convert them from strings to numeric. The response to a recent forum question (#24358) describes how to do this. "destring caseid, gen(Ind_ID)" will not work because embedded blanks should not be interpreted as zeroes.

Usually, caseid just combines v001 and v002 and v003, and hhid combines hv001 and hv002. You can identify cases just as easily with those components, which are numeric, as with caseid or hhid.

The usual variables in the KR file for having received the basic vaccines are h0, h2, h3, h4, h5, h6, h7, h8, h9, h9a (you should check your survey). These variables are coded 0 if the child did not receive a specific vaccine. You could do something like "drop if (h0+h2+h3+h4+h5+h6+h7+h8+h9a+h9b)==0". I recommend caution with dropping cases from the file. An alternative would be "gen condition=0" and "replace condition=1 if (h0+h2+h3+h4+h5+h6+h7+h8+h9a+h9b)==0". Then you can exclude those cases from a specific command with something like "tab A B if condition==0"