/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*CRC Analysis.sas\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Description: coding for infographs data

Date: Sept 1, 2021

Author: Nuha Wareg, Jamie Smith, Jane McElroy

Input: "C:\Users\smithjami\OneDrive - University of Missouri\Desktop\Nuha\sample\_dataset\_comc\_q2\_2022.XLSX" (COMC quarter 2 data)

Output:

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

\*\* Calling in excel data \*\*;

**PROC** **IMPORT** OUT = WORK.comc\_2022

DATAFILE= "C:\Users\smithjami\OneDrive - University of Missouri\Desktop\Nuha\sample\_dataset\_comc\_q2\_2022.XLSX"

DBMS=XLSX REPLACE;

GETNAMES=YES;

**RUN**;

**proc** **format**;

value fobt **0** = 'no FOBT'

 **1** = 'FOBT but not in quarter 2 or 1 year compliant'

 **2** = 'FOBT = quarter 2 only'

 **3** = 'FOBT = 1 year compliant only'

 **4** = 'FOBT = both quarter 2 and 1 year compliant';

value dna **0** = 'no DNA'

 **1** = 'DNA but not in quarter 2 or 3 year compliant'

 **2** = 'DNA = quarter 2 only'

 **3** = 'DNA = 3 year compliant only'

 **4** = 'DNA = both quarter 2 and 3 year compliant';

value col **0** = 'no colonoscopy'

 **1** = 'colonoscopy but not in quarter 2 or 10-year compliant'

 **2** = 'colonoscopy = quarter 2 only'

 **3** = 'colonoscopy = 10 year compliant only'

 **4** = 'colonoscopy = both quarter 2 and 10 year compliant';

value crc **1** = 'Colonoscopy'

 **2** = 'FIT-FOBT'

 **3** = 'FIT-DNA';

value age **0** = '< 50 yrs'

 **1** = '50-54 yrs'

 **2** = '55-59 yrs'

 **3** = '60-64 yrs'

 **4** = '65-69 yrs'

 **5** = '70-75 yrs';

**run**;

\*\* analysis file \*\*;

**data** comc2022\_q2;

 set comc\_2022;

 if exclusion = 'N';

 length provider $**35**;

\*\* New Variables \*\*;

year = **2022**;

quarter = **2**;

source = 'COMC';

\*\* collapsing positive/negative responses \*\*;

if fit\_dna\_result in ('Negative', 'NEG') then fit\_dna\_result = 'negative';

if fit\_fobt\_result in ('Negative', 'NEG') then fit\_fobt\_result = 'negative';

if fit\_fobt\_result in ('Positive', 'POSITIVE') then fit\_fobt\_result = 'positive';

\*\* quarterly/yearly compliance indicators \*\*;

\*\* Fit-FOBT quarter 2 and 1 year compliance range \*\*;

if (fit\_fobt\_date => **"01Apr2022"d** and fit\_fobt\_date <= **"30Jun2022"d**) then fobt\_q2\_2022 = **1**;

if (fit\_fobt\_date => **"01Jul2021"d** and fit\_fobt\_date <= **"31MAR2022"d**) then fobt\_1y\_2022 = **1**;

\*\* categorizing FOBT compliance \*\*;

if fit\_fobt\_date = **.** then fobt\_x = **0**; else

if fit\_fobt\_date ne **.** and fobt\_q2\_2022 = **.** and fobt\_1y\_2022 = **.** then fobt\_x = **1**; else

if fit\_fobt\_date ne **.** and fobt\_q2\_2022 = **1** and fobt\_1y\_2022 = **.** then fobt\_x = **2**; else

if fit\_fobt\_date ne **.** and fobt\_q2\_2022 = **.** and fobt\_1y\_2022 = **1** then fobt\_x = **3**; else

if fit\_fobt\_date ne **.** and fobt\_q2\_2022 = **1** and fobt\_1y\_2022 = **1** then fobt\_x = **4**;

\*\* Fit-DNA quarter 2 and 3 year compliance range \*\*;

if fit\_dna\_date ne **.** then dna = **1**;

if (fit\_dna\_date => **"01Apr2022"d** and fit\_dna\_date <= **"30Jun2022"d**) then dna\_q2\_2022 = **1**;

if (fit\_dna\_date => **"01Jul2019"d** and fit\_dna\_date <= **"31MAR2022"d**) then dna\_3y\_2022 = **1**;

\*\* categorizing DNA compliance \*\*;

if fit\_dna\_date = **.** then dna\_x = **0**; else

if fit\_dna\_date ne **.** and dna\_q2\_2022 = **.** and dna\_3y\_2022 = **.** then dna\_x = **1**; else

if fit\_dna\_date ne **.** and dna\_q2\_2022 = **1** and dna\_3y\_2022 = **.** then dna\_x = **2**; else

if fit\_dna\_date ne **.** and dna\_q2\_2022 = **.** and dna\_3y\_2022 = **1** then dna\_x = **3**; else

if fit\_dna\_date ne **.** and dna\_q2\_2022 = **1** and dna\_3y\_2022 = **1** then dna\_x = **4**;

\*\* colonoscopy quarter 2 and 10 year compliance range \*\*;

if colonoscopy\_date ne **.** then col = **1**;

if (colonoscopy\_date => **"01Apr2022"d** and colonoscopy\_date <= **"30Jun2022"d**) then col\_q2\_2022 = **1**;

if (colonoscopy\_date => **"01Jul2012"d** and colonoscopy\_date <= **"31MAR2022"d**) then col\_10y\_2022 = **1**;

\*\* categorizing colonoscopy compliance \*\*;

if colonoscopy\_date = **.** then col\_x = **0**; else

if colonoscopy\_date ne **.** and col\_q2\_2022 = **.** and col\_10y\_2022 = **.** then col\_x = **1**; else

if colonoscopy\_date ne **.** and col\_q2\_2022 = **1** and col\_10y\_2022 = **.** then col\_x = **2**; else

if colonoscopy\_date ne **.** and col\_q2\_2022 = **.** and col\_10y\_2022 = **1** then col\_x = **3**; else

if colonoscopy\_date ne **.** and col\_q2\_2022 = **1** and col\_10y\_2022 = **1** then col\_x = **4**;

\*\* Primary codes (colonoscopy\_date = A and Sigmoidoscopy = C and fit\_fobt\_date = D and fit\_dna\_date = F) \*\*;

if colonoscopy\_date = **.** and sigmoidoscopy = **.** and fit\_fobt\_date = **.** and fit\_dna\_date = **.** then primary\_code = **0**;

if col\_q2\_2022 = **1** and (sigmoidoscopy = **.** and fit\_fobt\_date = **.** and fit\_dna\_date = **.**) then primary\_code = **1**;

if fobt\_q2\_2022 = **1** and (colonoscopy\_date = **.** and sigmoidoscopy = **.** and fit\_dna\_date = **.**) then primary\_code = **2**;

if dna\_q2\_2022 = **1** and (colonoscopy\_date = **.** and sigmoidoscopy = **.** and fit\_fobt\_date = **.**) then primary\_code = **3**;

if (col\_10y\_2022 = **1** and col\_q2\_2022 = **.**) and (sigmoidoscopy = **.** and fit\_fobt\_date = **.** and fit\_dna\_date = **.**) then primary\_code = **4**;

if (fobt\_1y\_2022 = **1** and fobt\_q2\_2022 = **.**) and (colonoscopy\_date = **.** and sigmoidoscopy = **.** and fit\_dna\_date = **.**) then primary\_code = **5**;

if (dna\_3y\_2022 = **1** and dna\_q2\_2022 = **.**) and (colonoscopy\_date = **.** and sigmoidoscopy = **.** and fit\_fobt\_date = **.**) then primary\_code = **6**;

if (colonoscopy\_date ne **.** and col\_q2\_2022 = **.** and col\_10y\_2022 = **.**) and (sigmoidoscopy = **.** and fit\_fobt\_date = **.** and fit\_dna\_date = **.**) then primary\_code = **7**;

if (fit\_fobt\_date ne **.** and fobt\_q2\_2022 = **.** and fobt\_1y\_2022 = **.**) and (colonoscopy\_date = **.** and sigmoidoscopy = **.** and fit\_dna\_date = **.**) then primary\_code = **8**;

if (fit\_dna\_date ne **.** and dna\_q2\_2022 = **.** and dna\_3y\_2022 = **.**) and (colonoscopy\_date = **.** and sigmoidoscopy = **.** and fit\_fobt\_date = **.**) then primary\_code = **9**;

if (fobt\_q2\_2022 = **1** and col\_10y\_2022 = **1** and col\_q2\_2022 = **.**) and (sigmoidoscopy = **.** and fit\_dna\_date = **.**) then primary\_code = **21**;

if (dna\_q2\_2022 = **1** and col\_10y\_2022 = **1** and col\_q2\_2022 = **.**) and (sigmoidoscopy = **.** and fit\_fobt\_date = **.**) then primary\_code = **22**;

if (col\_q2\_2022 = **1** and fobt\_1y\_2022 = **1** and fobt\_q2\_2022 = **.**) and (sigmoidoscopy = **.** and fit\_dna\_date = **.**) then primary\_code = **23**;

if (fobt\_q2\_2022 = **1** and dna\_q2\_2022 = **1** and col\_10y\_2022 = **1** and col\_q2\_2022 = **.**) and (fit\_dna\_result in ('positive','negative')) and (source in ('COMC','CHCCMO','Access')) then primary\_code = **31**;

if (col\_q2\_2022 = **1** and fobt\_q2\_2022 = **1** and dna\_3y\_2022 = **1** and dna\_q2\_2022 = **.**) and (fit\_dna\_result in ('positive','negative')) and (source in ('COMC','CHCCMO','Access')) then primary\_code = **32**;

if (fobt\_q2\_2022 = **1** and dna\_q2\_2022 = **1** and col\_10y\_2022 = **1** and col\_q2\_2022 = **.**) and (fit\_dna\_result in ('positive','negative')) and (source not in ('COMC','CHCCMO','Access')) then primary\_code = **33**;

if (col\_q2\_2022 = **1** and fobt\_q2\_2022 = **1** and dna\_3y\_2022 = **1** and dna\_q2\_2022 = **.**) and (source not in ('COMC','CHCCMO','Access')) then primary\_code = **34**;

\*\* adjustments to primary\_code above to account for odd cases \*\*;

if primary\_code = **.** and col\_x = **2** then primary\_code = **1**; \*\* this is assigning obs where there was a colonoscopy in quarter 2, but other criteria did not fit other code definitions defined above (e.g., FOBT done but not in compliance or within 2 years) \*\*;

if primary\_code = **.** and col\_x = **3** then primary\_code = **4**; \*\* this is assigning obs where there was a colonoscopy within 10 years, but other criteria did not fit other code definitions defined above (e.g., FOBT done but not in compliance or within 2 years) \*\*;

\*\* need to create counts here for primary codes 24/25 below \*\*;

compliance\_tot = sum(of col\_10y\_2022 fobt\_1y\_2022 dna\_3y\_2022);

quarter\_tot = sum(of col\_q2\_2022 fobt\_q2\_2022 dna\_q2\_2022);

if (compliance\_tot > **1** and quarter\_tot = **.**) and (source in ('COMC','CHCCMO','Access')) then primary\_code = **24**;

if (compliance\_tot > **1** and quarter\_tot = **.**) and (source not in ('COMC','CHCCMO','Access')) then primary\_code = **25**;

\*\* calculating some date differences (in weeks) \*\*;

fobt\_col\_diff = INTCK('Week',fit\_fobt\_date, colonoscopy\_date);

if fobt\_col\_diff < **0** then fobt\_col\_diff = **.**;

dna\_col\_diff = INTCK('Week',fit\_dna\_date, colonoscopy\_date);

if dna\_col\_diff < **0** then dna\_col\_diff = **.**;

\*\*colonoscopy came after fobt or dna \*\*;

if colonoscopy\_date > fit\_fobt\_date then col\_after\_fobt = **1**;

if (fit\_fobt\_date = **.** or colonoscopy\_date = **.**) then col\_after\_fobt = **.**;

if colonoscopy\_date > fit\_dna\_date then col\_after\_dna = **1**;

if (fit\_dna\_date = **.** or colonoscopy\_date = **.**) then col\_after\_dna = **.**;

\*\* Secondary Codes (note: uses some variables derived above) \*\*;

if primary\_code in (**21**,**22**,**23**,**24**,**25**,**31**,**32**,**33**,**34**) then secondary\_code = **40**;

if fit\_fobt\_result = 'negative' and fobt\_col\_diff <= **52** and fobt\_col\_diff ne **.** then secondary\_code = **41**;

if fit\_dna\_result = 'negative' and dna\_col\_diff <= **156** and dna\_col\_diff ne **.** then secondary\_code = **42**;

if (fit\_fobt\_result = 'positive' or fit\_dna\_result = 'positive') and (col\_after\_fobt = **1** or col\_after\_dna = **1**) then secondary\_code = **43**;

if primary\_code in (**21**,**22**,**23**,**24**,**25**) and ((fobt\_col\_diff ne **.** and fobt\_col\_diff <= **1**) or (dna\_col\_diff ne **.** and dna\_col\_diff <= **1**)) then secondary\_code = **44**;

\*\* assigning current CRC codes \*\*;

if primary\_code in (**1**,**4**) then current\_crc = **1**; else

if primary\_code in (**2**,**5**) then current\_crc = **2**; else

if primary\_code in (**3**,**6**) then current\_crc = **3**;

\*\* calculating patient age based on last test date in system \*\*;

last\_date = max (of colonoscopy\_date fit\_fobt\_date fit\_dna\_date);

age = round((last\_date-date\_of\_birth)/**365.25**,**.1**);

\*\* fixing provider names (note: this may need done each time) \*\*;

if usual\_provider = 'ATKINS, KRISTI' then provider = 'Atkins, Kristi';

if usual\_provider = 'BAGBY, MARK' then provider = 'Bagby, Mark';

if usual\_provider = 'BAILEY, KELVIN' then provider = 'Bailey, Kelvin';

if usual\_provider = 'BOWMAN, ELIZABETH' then provider = 'Bowman, Elizabeth';

if usual\_provider = 'CLARY, MICHELE' then provider = 'Clary, Michele';

if usual\_provider = 'GULLEY, VICKY' then provider = 'Gulley, Vicky';

if usual\_provider = 'Galloway, John' then provider = 'Galloway, John';

if usual\_provider = 'JONES, JUANITA' then provider = 'Jones, Juanita';

if usual\_provider = 'Jones, Jeffrey' then provider = 'Jones, Jeffrey';

if usual\_provider = 'KING, L' then provider = 'King, Lisa';

if usual\_provider = 'KING, LISA' then provider = 'King, Lisa';

if usual\_provider = 'KRANTZ, T' then provider = 'Krantz, Terry';

if usual\_provider = 'KRANTZ, TERRY' then provider = 'Krantz, Terry';

if usual\_provider = 'Kirkover, Michael' then provider = 'Kirkover, Michael';

if usual\_provider = 'MACIAS, MARIO' then provider = 'Macias, Mario';

if usual\_provider = 'Nagy, Virginia' then provider = 'Nagy, Virginia';

if usual\_provider = 'Neis, Miles' then provider = 'Neis, Miles';

if usual\_provider = 'Powell, Jennifer' then provider = 'Powell, Jennifer';

if usual\_provider = 'RADEMAN ALONSO, ERIN' then provider = 'Rademan Alonso, Erin';

if usual\_provider = 'ROBERTSON, CHANDLER' then provider = 'Robertson, Chandler';

if usual\_provider = 'SCHMIDT, DANIEL' then provider = 'Schmidt, Daniel';

if usual\_provider = 'SLUYTER, ERIN' then provider = 'Sluyter, Erin';

if usual\_provider = 'THOMAS, JAMIE' then provider = 'Thomas, Jamie';

if usual\_provider = 'Thomas, Thomas' then provider = 'Thomas, Thomas';

if usual\_provider = 'VANDIVER, AMBER' then provider = 'Vandiver, Amber';

if usual\_provider = 'VOGL, DANIEL' then provider = 'Vogl, Daniel';

if usual\_provider = 'WARREN, BARTON' then provider = 'Warren, Barton';

if usual\_provider = 'Wang, Michael' then provider = 'Wang, Michael';

if usual\_provider = 'Whitman, Mary' then provider = 'Whitman, Mary';

if usual\_provider = 'coon, errica' then provider = 'Coon, Errica';

if usual\_provider = 'z-BUCHANAN, SARAH' then provider = 'Buchanan, Sarah';

if usual\_provider = 'z-SPARR, LAURIE' then provider = 'Sparr, Laurie';

if usual\_provider = 'z-WARREN, SARA' then provider = 'Warren, Sara';

if usual\_provider = 'zREYNOLDS, EVAN' then provider = 'Reynolds, Evan';

if usual\_provider = 'zzzMATTOX, SHAYNA' then provider = 'Mattox, Shayna';

\*\* age category \*\*;

if age < **50** then agegrp = **0**; else

if **50** le age lt **55** then agegrp = **1**; else

if **55** le age lt **60** then agegrp = **2**; else

if **60** le age lt **65** then agegrp = **3**; else

if **65** le age lt **70** then agegrp = **4**; else

if age ge **70** then agegrp = **5**;

format fobt\_x fobt. dna\_x dna. col\_x col. current\_crc crc. last\_date mmddyy10. agegrp age.;

**run**;

\*\* Creating a count of patients seen by provider \*\*;

**proc** **sort** data=comc2022\_q2; by provider; **run**;

**data** provider\_count (keep=provider pat\_total);

 length provider $**35**;

 set comc2022\_q2;

 by provider;

if first.provider then pcount = **1**; else pcount+**1**; \*\* creating a count of records \*\*;

if last.provider then pat\_total = pcount; \*\* getting total # of patients \*\*;

if pat\_total = **.** then delete; \*\* reducing file to one obs per provider \*\*;

**run**;

\*\* merging total patient counts back into file \*\*;

**proc** **sort** data=provider\_count; by provider; **run**;

**data** comc2022\_final;

 length provider $**35**;

 merge comc2022\_q2(in=keeper) provider\_count;

 by provider;

 if keeper;

\*\* creates a variable that denotes providers who have less than 10 patients in file \*\*;

if pat\_total < **10** then pt\_less10 = **1**;

**run**;

\*\* Earlier - Requested Tables (note: these include all providers and do not restrict by age) \*\*;

**proc** **tabulate** data=comc2022\_final missing;

class primary\_code location;

table all location, n colpctn primary\_code\*(n rowpctn colpctn);

where primary\_code in (**0**,**1**,**2**,**3**);

**run**;

**proc** **tabulate** data=comc2022\_final missing;

class current\_crc crc\_status;

table all current\_crc, n colpctn crc\_status\*(n rowpctn colpctn);

**run**;

**proc** **freq** data=comc2022\_final;

table fit\_dna\_result\*primary\_code/list missing;

where primary\_code in (**3**,**6**,**9**,**22**,**33**); \*\* note: there were two additional DNA tests (both neg), but the primary code was a 24 \*\*;

**run**;

**proc** **freq** data=comc2022\_final;

tables age;

where secondary\_code = **40**;

**run**;

**proc** **tabulate** data=comc2022\_final missing out=provider\_output; \*\* this code outputs the data needed for the provider data below \*\*;

class provider crc\_status;

tables all provider, n crc\_status\*(n rowpctn);

where pt\_less10 ne **1**;

**run**;

\*\* Creating datasets for provider output/database \*\*;

\*\* total \*\*;

**data** CRC\_T(keep=provider crc\_n);

 set provider\_output;

 where CRC\_STATUS = '';

\*\* renaming variables \*\*;

crc\_n = N;

if provider = '' then provider = 'Total';

label

crc\_n = 'Total # of Patients';

**run**;

\*\* CRC = No \*\*;

**data** CRC\_N(keep=provider crc\_n\_n crc\_n\_pct);

 set provider\_output;

 where CRC\_STATUS = 'N';

\*\* renaming variables \*\*;

crc\_n\_n = N;

crc\_n\_pct = pctn\_10;

if pctn\_10 = **.** then crc\_n\_pct = pctn\_00;

crc\_n\_pct = round((crc\_n\_pct),**.1**);

if provider = '' then provider = 'Total';

label

crc\_n\_n = 'CRC Status No (N)'

crc\_n\_pct = 'CRC Status No (%)';

**run**;

\*\* CRC = Yes \*\*;

**data** CRC\_Y(keep=provider crc\_y\_n crc\_y\_pct crc\_y\_plus10 id);

 set provider\_output;

 where CRC\_STATUS = 'Y';

\*\* renaming variables \*\*;

crc\_y\_n = N;

crc\_y\_pct = pctn\_10;

if pctn\_10 = **.** then crc\_y\_pct = pctn\_00;

crc\_y\_pct = round((crc\_y\_pct),**.1**);

crc\_y\_plus10 = round((crc\_y\_n\***.10**),**.1**); \*\* # patients needed if increased by 10% \*\*;

if provider = '' then provider = 'Total';

label

crc\_y\_n = 'CRC Status Yes (N)'

crc\_y\_pct = 'CRC Status Yes (%)'

crc\_y\_plus10 = '# of patients needed for 10% increase'

provider = 'Provider Name';

if provider = 'Total' then id = **1**; else id = **2**;

**run**;

\*\* sorting files prior to merge \*\*;

**proc** **sort** data = crc\_t; by provider; **run**;

**proc** **sort** data = crc\_n; by provider; **run**;

**proc** **sort** data = crc\_y; by provider; **run**;

**data** crc\_providers;

 merge crc\_t crc\_n crc\_y;

 by provider;

**run**;

**proc** **sort** data=crc\_providers; by id; **run**;

\*\* final file for providers \*\*;

**data** crc\_provider\_info (drop=id);

 set crc\_providers;

**run**;

**proc** **sort** data=crc\_provider\_info out=plot\_sort; by descending crc\_y\_pct; **run**; \*\* sorting so that highest percentage first \*\*;

\*\* exporting new file \*\*;

**proc** **export** data=crc\_provider\_info

 outfile='C:\Users\smithjami\OneDrive - University of Missouri\Desktop\Nuha\crc\_provider\_2022\_q2.xlsx'

 dbms=xlsx replace;

**run**;;

\*\* Scatter plot for Q2 2022 CRC Provider Screening Rates \*\*;

ods rtf file = 'C:\Users\smithjami\OneDrive - University of Missouri\Desktop\CRC preliminary results for infographics(09-01-22).doc'; \*\* location of word file, adjust as needed \*\*;

**proc** **print** data=crc\_provider\_info label noobs;

**run**;

ods graphics / reset=all border=off width=**600**px height=**400**px;

**proc** **sgplot** data=plot\_sort noborder;

 scatter y=crc\_y\_pct x=provider/markerattrs=(color=black size=**10** symbol=squarefilled);

 yaxis display=(nolabel) min=**0** max=**80**;

 xaxis label = 'Q2 2022 CRC Screening Rates of Providers' display=(novalues noticks);

 where provider ne 'Total';

 title1 justify=RIGHT "National Goal - 80%" ;

**run**;

\*\* Infosheet - 2022 (Q2) CRC Screening Data for Patients Aged 50-75 \*\*;

**proc** **tabulate** data=comc2022\_final missing;

class crc\_status current\_crc agegrp race sex\_at\_birth location;

tables all crc\_status current\_crc agegrp race sex\_at\_birth, n colpctn location\*(n colpctn);

tables all current\_crc agegrp race sex\_at\_birth location, n colpctn crc\_status\*(n colpctn);

where agegrp ne **0**;

title '2022 (Q2) CRC Screening Data for Patients Aged 50-75';

**run**;

**proc** **logistic** descending data=comc2022\_final ;

class sex\_at\_birth(ref='F') agegrp(ref='50-54 yrs') location(ref='Camdenton'); \*\* assign reference group here instead of dummy coding \*\*;

model CRC\_status = sex\_at\_birth agegrp location/risklimits ;

where agegrp ne **0** and location in ('Camdenton','Osage Beach','Richland');

title 'Logistic Model Results - CRC Status 2002(Q2)';

**run**;

ods rtf close;