\* SWAB\_POS\_ELIFE.DO

\* analysis for paper on Positives and Ct Values

\* - include all positives

\* - add false-positives out of all tests at end

\* updated for ELife revision March and April 2021

adopath + "P:\Working\ado/"

adopath + "P:\Working\user\_asw\myado/"

\* date of files and for analysis

global DATE "20210315"

global TODAY "15mar2021"

global WindowDate =d(13mar2021)

noi di \_n "Swabs back from " %d $WindowDate

\* UseDate is the Sunday AFTER WindowDate

global UseDate=d(14mar2021)

assert dow($UseDate)==0

assert inrange($UseDate-$WindowDate,0,6)

\* directories

global MainData "P:/Working/user\_statistical\_processing/raw$DATE/Data"

global Data "P:/Working/user\_asw/analysis/$DATE"

global Logs "P:/Working/user\_asw/analysis/docs/2020 positives/analysis/eLife$DATE/Logs"

global Graphs "P:/Working/user\_asw/analysis/docs/2020 positives/analysis/eLife$DATE/Graphs"

cap log close

log using "$Logs/swab\_pos\_paper\_$DATE.log", replace

noi di "Run by ASW on $S\_DATE $S\_TIME"

noi version

noi di "Dataset: $DATE"

noi di "WindowDate: " %d $WindowDate

noi di "UseDate: " %d $UseDate

set more off

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\* A: positives out of all tests

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

noi di \_n \_dup(80) "\*" \_n \_dup(80) "\*" \_n "A: All linked test results" \_n \_dup(80) "\*" \_n \_dup(80) "\*"

noi use participant\_id result\_mk visit\_date visit\_id lab\_id country ageg\_sy hh\_id result\_\* assay\_\* n\_participants\_i swab\_barcode\_cleaned if result\_mk<=1|result\_tdi<=1 using $MainData/data\_participant\_clean, clear

assert visit\_date<=$UseDate

\* drop one duplicate positive from same person on same day - not in main dataset

sort participant visit\_date result\_mk swab\_barcode\_cleaned

noi by participant visit\_date: drop if \_n>1 & result\_mk==1 & result\_mk[1]==1 & inlist(participant,"DHR-200928237909")

by participant visit\_date: assert \_N==1|(result\_mk[1]==0 & result\_mk[2]==1 & \_N==2 & inlist(participant,"DHR-200925221500")) if result\_mk==1

\* save as temporary file, as used below

tempfile main

save "`main'", replace

noi di \_n "ALL POS/NEG SWABS to " %d $UseDate

\* drop void results

drop if result\_mk>=9

noi tab result\_mk

qui summ result\_mk if result\_mk<=1

noi cii prop r(N) r(sum), exact

noi tab country result\_mk

\* first or subsequent test in study

preserve

gsort participant visit\_date -result\_mk

by participant: gen byte first=1 if \_n==1

noi tab result\_mk first if result\_mk==1, miss

noi di \_n "All participants"

noi tab result\_mk first if first==1, col nokey

gen byte flag=(result\_mk==1)

egen byte max=max(flag), by(participant)

egen int count=group(hh\_id)

qui summ count

noi di \_n "Number of households: " r(max)

drop if max==0

noi di \_n "Participants with any positive"

noi tab result\_mk first, col nokey

noi tab country first, col nokey

drop count

egen int count=group(hh\_id)

qui summ count

noi di \_n "Number of households with a positive: " r(max)

restore

\* distribution of first test result, swabs per participant

sort participant visit\_date

by participant: gen long first\_test\_month=mdy(month(visit\_date),1,year(visit\_date)) if \_n==1

format first\_test\_month %d

noi tab first\_test\_month

by participant: gen byte n\_swabs=\_N if \_n==1

noi tabstat n\_swabs, by(first\_test\_month) c(s) s(median p25 p75 min max)

noi tab n\_swabs

\* number of prior negative tests

use $Data\swabpositives\_evidence, clear

format ct\_mean %7.1f

assert lab\_id==2 if visit\_date<d(16aug2020)

assert lab\_id==1 if visit\_date>d(9feb2021)

confirm var imd\_pc\_new imdg ageg\_sy

\* household classification - anyone else in the household positive EVER

bysort participant: gen byte flag=(\_n==1)

egen byte hh\_pos=sum(flag), by(hh\_id)

drop flag

noi tab hh\_pos

noi tab hh\_pos n\_participants

gen byte hh\_pos\_gt1=(hh\_pos>1)

noi di \_n "Number of prior negatives for first positive in the study"

noi tabstat count\_prior\_neg if any\_prior\_pos==0, by(first\_test) c(s) s(n median p25 p75 min max)

noi datacheck count\_prior\_neg>0 if any\_prior\_pos==0 & first\_test~=1, mess(Check numbers tie up with first test above) vars(participant first\_test count\_prior\_neg count\_prior\_pos) noobs nod

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\* B: Ct patterns and values overall

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

noi di \_n \_dup(80) "\*" \_n \_dup(80) "\*" \_n "B: Ct patterns, positive targets and values overall: Table 1" \_n \_dup(80) "\*" \_n \_dup(80) "\*"

noi tab count\_pos\_target, miss

noi di \_n "Drop if missing Ct (non-standard lab)"

noi tab lab\_id ct\_mean if ct\_mean>=., miss

noi drop if ct\_mean>=.

noi tab count\_pos\_target, miss

noi di \_n "Split count\_pos\_target by before or after 16nov (PHE technical briefing 2)"

noi replace count\_pos\_target=4 if count\_pos\_target==2 & ctpattern==4 & visit\_date>=d(16nov2020)

label define count\_pos\_target 4 "OR/N 16nov-"

label values count\_pos\_target count\_pos\_target

noi tab count\_pos\_target

noi tab ctpattern count\_pos\_target

noi tab count\_pos\_target any\_prior\_pos, col nokey chi

gen byte CTPATTERN=ctpattern

replace CTPATTERN=8 if ctpattern==4 & visit\_date>=d(16nov2020)

label define ctpattern 8 "OR+N 16nov-", add

label values CTPATTERN ctpattern

noi tab CTPATTERN any\_prior\_pos, col nokey chi

noi di \_n "S only"

gen long visit\_month=mdy(month(visit\_date),1,year(visit\_date))

format visit\_month %d

noi tab visit\_month if ctpattern==3

sort visit\_date

noi datacheck ctpattern~=3 if visit\_date>d(17may2020), mess(S only after 17 May 2020) vars(participant visit\_date ctpattern ct\_mean lab) noobs nod

noi di \_n "Correlation between Ct values when called positive for both genes"

noi spearman ctNgene ctORF1ab if inlist(ctpattern,7,4)

noi spearman ctSgene ctORF1ab if inlist(ctpattern,7,5)

noi spearman ctNgene ctSgene if inlist(ctpattern,7,6)

noi di \_n "Ct per sample: all positives"

noi tabstat ct\_mean, by(count\_pos\_target) c(s) s(n median p25 p75 min max) format(%6.1f)

noi kwallis ct\_mean, by(count\_pos\_target)

noi tabstat ct\_mean, by(CTPATTERN) c(s) s(n median p25 p75 min max) format(%6.1f)

qreg ct\_mean i.count\_pos\_target i.CTPATTERN

noi qreg

noi testparm i.CTPATTERN

noi di \_n "Ct per sample: first positives"

noi tabstat ct\_mean if any\_prior\_pos==0, by(count\_pos\_target) c(s) s(n median p25 p75 min max) format(%6.1f)

noi tabstat ct\_mean if any\_prior\_pos==0, by(CTPATTERN) c(s) s(n median p25 p75 min max) format(%6.1f)

noi di \_n "Ct over 37"

noi tab ct\_mean if ct\_mean>37

noi tab visit\_month if ct\_mean>37

noi tab visit\_month lab if ct\_mean>37

noi tab CTPATTERN lab if ct\_mean>37

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\* D: Ct & symptoms over calendar time

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

noi di \_n \_dup(80) "\*" \_n \_dup(80) "\*" \_n "D: Ct by symptoms and over time" \_n \_dup(80) "\*" \_n \_dup(80) "\*"

\* use symptoms around test as this is what is most predictive

gen byte evidence\_sympt=any\_evidence\_sympt\_now

replace evidence\_sympt=any\_evidence\_sympt\_around

sort visit\_date

noi datacheck evidence\_sympt<., mess(Missing all symptoms - set to none for plots as all old) vars(participant visit\_date count\_pos ct\_mean \*evidence\*) noobs nod

noi replace evidence\_sympt=0 if evidence\_sympt>=.

\*\*\*\*\*\*\*\*

\* violin plot of Ct and symptoms: Figure 1

\*\*\*\*\*\*\*\*

gen byte temp=2\*CTPATTERN+evidence\_sympt-1

label define temp 1 "N only: no sympt" 2 "N only: sympt" 3 "OR only: no sympt" 4 "OR only: sympt" 5 "S only: no sympt" 6 "S only: sympt" /\*

\*/ 7 "OR+N -16nov: no sympt" 8 "OR+N -16nov: sympt" 9 "OR+S: no sympt" 10 "OR+S: sympt" 11 "N+S: no sympt" 12 "N+S: sympt" 13 "OR+N+S: no sympt" /\*

\*/ 14 "OR+N+S: sympt" 15 "OR+N 16nov-: no sympt" 16 "OR+N 16nov-: sympt"

label values temp temp

vioplot ct\_mean, over(temp) graphregion(color(white)) ylabel(10(5)30 34 37, angle(0) format(%2.0f) labsize(small)) ytitle(Ct value) /\*

\*/ xtitle(" ") xlabel(,labsize(small) valuelabel angle(45)) xline(2.5(2)14.5,lpattern(dash) lcolor(gray) lw(\*0.5))

graph export "$Graphs/violin\_plot\_fig\_1.emf", replace

preserve

table CTPATTERN evidence\_sympt, c(median ct\_mean p25 ct\_mean p75 ct\_mean min ct\_mean max ct\_mean) replace

for num 1/5 \ any median p25 p75 min max: rename tableX Y

export excel using "$Logs/Figure1.xlsx", replace firstrow(var)

restore

\*\*\*\*\*\*\*\*

\* symptoms by Ct: Figure 2

\*\*\*\*\*\*\*\*

preserve

replace ct\_mean=round(ct\_mean,1)

recode ct\_mean 36/39=36 9/11=11

assert inrange(ct\_mean,11,36)

compress

collapse (count) n=result\_mk n\_now=any\_evidence\_sympt\_now n\_around=any\_evidence\_sympt\_around (sum) \*cgh\* any\_evidence\_sympt\_now any\_evidence\_sympt\_around, by(ct\_mean)

local max=\_N

foreach var of varlist any\_evidence\_sympt\_now sympt\_now\_cgh {

 gen `var'\_p=100\*`var'/n\_now

 for any l u: gen byte `var'\_X=.

 forval i=1(1)`max' {

 local N=n\_now[`i']

 local n=`var'[`i']

 qui cii prop `N' `n', exact

 qui replace `var'\_l=r(lb)\*100 if \_n==`i'

 qui replace `var'\_u=r(ub)\*100 if \_n==`i'

 }

}

foreach var of varlist any\_evidence\_sympt\_around sympt\_around\_cgh {

 gen `var'\_p=100\*`var'/n\_around

 for any l u: gen byte `var'\_X=.

 forval i=1(1)`max' {

 local N=n\_around[`i']

 local n=`var'[`i']

 qui cii prop `N' `n', exact

 qui replace `var'\_l=r(lb)\*100 if \_n==`i'

 qui replace `var'\_u=r(ub)\*100 if \_n==`i'

 }

}

rename any\_evidence\_sympt\_now\* now\*

rename any\_evidence\_sympt\_around\* aro\*

rename sympt\_now\_cghfevamn\* now\_cgh\*

rename sympt\_around\_cghfevamn\* aro\_cgh\*

gen plot1=ct\_mean-0.15

gen plot2=ct\_mean-0.05

gen plot3=ct\_mean+0.05

gen plot4=ct\_mean+0.15

twoway rcapsym now\_u now\_l plot2, color(black) lcolor(black) s(i) lwidth(thin) || /\*

\*/ scatter now\_p plot2, color(black) lcolor(black) s(o) c(l) || /\*

\*/ rcapsym aro\_u aro\_l ct\_mean, color(gs8) lcolor(gs8) s(i) lwidth(thin) || /\*

\*/ scatter aro\_p ct\_mean, color(gs8) lcolor(gs8) s(s) c(l) || /\*

\*/ rcapsym now\_cgh\_u now\_cgh\_l ct\_mean, color(black) lcolor(black) s(i) lwidth(thin) || /\*

\*/ scatter now\_cgh\_p ct\_mean, color(black) lcolor(black) s(o) c(l) mfcolor(white) || /\*

\*/ rcapsym aro\_cgh\_u aro\_cgh\_l plot3, color(gs8) lcolor(gs8) s(i) lwidth(thin) || /\*

\*/ scatter aro\_cgh\_p plot3, color(gs8) lcolor(gs8) s(s) c(l) mfcolor(white) || /\*

\*/ , graphregion(color(white)) ytitle(Percentage of positive tests (95% CI)) ylabel(0(10)100, angle(0)) /\*

\*/ xlabel(11(1)36, valuelabels labsize(vsmall) angle(45)) xtitle("Ct value (rounded)") /\*

\*/ legend(order(2 "Any evidence of symptoms at test" 6 "Cough, fever, anosmia at test" 4 "Any evidence of symptoms around test" 8 "Cough, fever, anosmia around test" ) size(vsmall) rows(2))

graph export "$Graphs\symptom\_over\_ct\_mean\_fig\_2.emf", replace

export excel ct\_mean now\_p now\_l now\_u aro\_p aro\_l aro\_u now\_cgh\_p now\_cgh\_l now\_cgh\_u aro\_cgh\_p aro\_cgh\_l aro\_cgh\_u using "$Logs\Figure2.xlsx", replace firstrow(var)

restore

\*\*\*\*\*\*\*\*\*

\* Evidence: Table 2

\*\*\*\*\*\*\*\*\*

noi di \_n "Level of evidence"

noi tab evidence

noi tab strength

noi tab strength if evidence==1

noi tab evidence count\_pos\_target, row nokey

noi tabstat ct\_mean, by(evidence) c(s) s(n median p75 p90 p95 p99) format(%6.1f)

\_pctile ct\_mean if evidence==1, percentiles(90(1)97 97.5 98 99)

noi return list

noi tab evidence any\_ct\_under\_34, row nokey

sort ct\_mean

noi di \_n "Factors determining evidence classification - symptoms and patient-facing/care work"

noi tab evidence any\_evidence\_sympt\_around, row nokey chi

noi tab evidence any\_evidence\_sympt\_around if evidence<=2, row nokey chi

gen byte occupation=max(ever\_patient,ever\_care\_home)

noi tab evidence occupation, row nokey

noi tab evidence occupation if evidence<=2, row nokey chi

noi di \_n "Other factors"

noi tab evidence sympt\_now\_cgh, row nokey chi

noi tab evidence sympt\_around\_cgh, row nokey chi

noi tab evidence sympt\_around\_cgh if evidence<=2, row nokey chi

noi tab evidence any\_prior\_pos, row nokey chi

noi tab evidence first\_test, row nokey chi

noi tab evidence wgs\_status if wgs\_status<=2, row nokey chi

noi di \_n "Ct in those positive one target with WGS"

noi tabstat ct\_mean if streng==3, c(s) s(n median p75 p90 p95 p99) format(%6.1f)

noi tab evidence hh\_pos\_gt1, row nokey chi

noi di \_n "- where >1 participant"

noi tab evidence hh\_pos\_gt1 if n\_participants>1, row nokey chi

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\* C: Predictors of Ct values

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

noi di \_n \_dup(80) "\*" \_n \_dup(80) "\*" \_n "C: Predictors of Ct values" \_n \_dup(80) "\*" \_n \_dup(80) "\*"

\* use quantile regression

noi summ ct\_mean, detail

\* spline for date

noi spearman ct\_mean visit\_date

summ visit\_date, detail

local min=r(min)

local max=r(max)

local p50=r(p50)

local p10=r(p10)-`p50'

local p90=r(p90)-`p50'

local p25=r(p25)-`p50'

local p75=r(p75)-`p50'

noi di \_n "EQUALLY SPACED KNOTS for visit\_date"

local p50=(`max'+`min')\*0.5

local p10=`min'+(`max'-`min')\*0.10-`p50'

local p90=`min'+(`max'-`min')\*0.90-`p50'

local p25=`min'+(`max'-`min')\*0.25-`p50'

local p75=`min'+(`max'-`min')\*0.75-`p50'

gen long DATE=visit\_date-`p50'

noi mkspline DATE\_=DATE, cubic nknots(5) knots(`p10' `p25' 0 `p75' `p90') displayknots

qui qreg ct\_mean DATE\_\*

noi qreg

noi testparm DATE\_\*

noi testparm DATE\_2 DATE\_3 DATE\_4

\* reset to zero on reference date

for num 2/4: summ DATE\_X if DATE\_1==0 \ replace DATE\_X=DATE\_X-r(mean)

noi di \_n "Reference date " %d `p50'

\* age

summ age\_at\_visit, detail

\* truncate at 80

replace age\_at=80 if age\_at>80 & age\_at<.

local min=r(min)

local max=r(max)

local p50=r(p50)

local p50=(`max'+`min')\*0.5

local p10=`min'+(`max'-`min')\*0.10-`p50'

local p90=`min'+(`max'-`min')\*0.90-`p50'

local p25=`min'+(`max'-`min')\*0.25-`p50'

local p75=`min'+(`max'-`min')\*0.75-`p50'

gen long AGE=age\_at\_visit-`p50'

noi mkspline AGE\_=AGE, cubic nknots(5) knots(`p10' `p25' 0 `p75' `p90') displayknots

drop AGE

\* imd

summ imd\_pc\_new, detail

local min=r(min)

local max=r(max)

local p50=r(p50)

local p50=(`max'+`min')\*0.5

local p10=`min'+(`max'-`min')\*0.10-`p50'

local p90=`min'+(`max'-`min')\*0.90-`p50'

local p25=`min'+(`max'-`min')\*0.25-`p50'

local p75=`min'+(`max'-`min')\*0.75-`p50'

gen long IMD=imd\_pc\_new-`p50'

noi mkspline IMD\_=IMD, cubic nknots(5) knots(`p10' `p25' 0 `p75' `p90') displayknots

drop IMD

fvset base 3 ageg\_sy

fvset base 5 imdg

foreach var of varlist count\_pos first\_test any\_prior any\_evidence\* \*split\* sex ethnicityg ethnicity\_wo ageg\_sy imdg hh\_pos\_gt1 ever\_patientfacing occupation ever\_lthc smoke {

 local mult=inlist("`var'","ageg\_sy","ethnicityg","count\_pos\_target")

 if strpos("`var'","split")>0 local mult=1

 noi di \_n \_dup(80) "=" \_n "`var' (UNIVARIABLE and BIVARIABLE WITH DATE)" \_n \_dup(80) "="

 noi tabstat ct\_mean, by(`var') c(s) s(n median p25 p75 min max) format(%6.1f)

 cap assert `var'<.

 if \_rc~=0 {

 noi di \_n "In missings"

 noi tabstat ct\_mean if `var'>=., c(s) s(n median p25 p75 min max) format(%6.1f)

 }

 noi kwallis ct\_mean, by(`var')

 qui qreg ct\_mean i.`var'

 noi qreg

 noi testparm i.`var'

 if `mult'>0 noi testparm i.`var', equal

 \* adjusted for DATE

 qui qreg ct\_mean i.`var' DATE\_\*

 noi qreg

 noi testparm i.`var'

 if `mult'>0 noi testparm i.`var', equal

 if "`var'"=="ageg\_sy" {

 noi spearman ct\_mean age\_at\_visit

 qui qreg ct\_mean age\_at\_visit

 noi qreg

 qui qreg ct\_mean age\_at\_visit DATE\_\*

 noi qreg

 qui qreg ct\_mean AGE\_\*

 noi qreg

 noi testparm AGE\*

 noi testparm AGE\_2 AGE\_3 AGE\_4

 qui qreg ct\_mean AGE\_\* DATE\_\*

 noi qreg

 noi testparm AGE\*

 noi testparm AGE\_2 AGE\_3 AGE\_4

 }

 else if "`var'"=="imdg" {

 noi spearman ct\_mean imd\_pc\_new

 qui qreg ct\_mean imd\_pc\_new

 noi qreg

 qui qreg ct\_mean imd\_pc\_new DATE\_\*

 noi qreg

 qui qreg ct\_mean IMD\_\*

 noi qreg

 noi testparm IMD\_\*

 noi testparm IMD\_2 IMD\_3 IMD\_4

 qui qreg ct\_mean IMD\_\* DATE\_\*

 noi qreg

 noi testparm IMD\_\*

 noi testparm IMD\_2 IMD\_3 IMD\_4

 }

 else if "`var'"=="hh\_pos\_gt1" {

 noi spearman ct\_mean hh\_pos

 qui qreg ct\_mean hh\_pos

 noi qreg

 }

}

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

noi di \_n \_dup(80) "\*" \_n \_dup(80) "\*" \_n "FULL MULTIVARIABLE MODEL with SYMPTOMS AT TEST" \_n \_dup(80) "\*" \_n \_dup(80) "\*"

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

qui qreg ct\_mean i.count\_pos\_target i.any\_evidence\_sympt\_now i.any\_prior\_pos i.first\_test i.sex i.ethnicity\_wo

noi qreg

noi testparm i.count\_pos\_target

noi testparm 2.count\_pos\_target

noi testparm 3.count\_pos\_target

noi testparm 4.count\_pos\_target

noi testparm i.count\_pos\_target, equal

noi testparm 4.count\_pos\_target 3.count\_pos\_target, equal

noi margins i.count\_pos\_target, atmeans

foreach var of varlist any\_evidence\_sympt\_now any\_prior\_pos first\_test sex ethnicity\_wo {

 noi testparm i.`var'

 noi margins i.`var'

}

noi di \_n \_dup(80) "=" \_n "WITH DATE" \_n \_dup(80) "="

qui qreg ct\_mean i.count\_pos\_target i.any\_evidence\_sympt\_now i.any\_prior\_pos i.first\_test i.sex i.ethnicity\_wo DATE\_\*

noi qreg

noi testparm DATE\_\*

noi di \_n \_dup(80) "=" \_n "SPLITTING SYMPTOMS" \_n \_dup(80) "="

qui qreg ct\_mean i.count\_pos\_target i.sympt\_now\_split i.any\_prior\_pos i.first\_test i.sex i.ethnicity\_wo

noi qreg

noi testparm i.sympt\_now\_split

noi testparm i.sympt\_now\_split, equal

noi testparm 1.sympt\_now\_split

noi testparm 2.sympt\_now\_split

noi margins i.sympt\_now\_split

noi di \_n \_dup(80) "=" \_n "Effect of other variables" \_n \_dup(80) "="

foreach var of varlist hh\_pos\_gt1 ethnicityg ageg\_sy imdg ever\_patientfacing occupation ever\_lthc smoke {

 qui qreg ct\_mean i.count\_pos\_target i.any\_evidence\_sympt\_now i.any\_prior\_pos i.first\_test i.sex i.ethnicity\_wo i.`var'

 noi qreg

 noi testparm i.`var'

 \*noi margins i.`var'

}

foreach thing in age\_at\_visit AGE\_\* imd\_pc\_new IMD\_\* {

 qui qreg ct\_mean i.count\_pos\_target i.any\_evidence\_sympt\_now i.any\_prior\_pos i.first\_test i.sex i.ethnicity\_wo `thing'

 noi qreg

 noi testparm `thing'

}

noi di \_n \_dup(80) "=" \_n "Mixed model with RE for hh\_id - check similar" \_n \_dup(80) "="

qui mixed ct\_mean i.count\_pos\_target i.any\_evidence\_sympt\_now i.any\_prior\_pos i.first\_test i.sex i.ethnicity\_wo || hh\_id: || participant\_id:

noi mixed

qui mixed ct\_mean i.count\_pos\_target i.any\_evidence\_sympt\_now i.any\_prior\_pos i.first\_test i.sex i.ethnicity\_wo || hh\_id:

noi mixed

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

noi di \_n \_dup(80) "=" \_n \_dup(80) "=" \_n "FULL MULTIVARIABLE MODEL with SYMPTOMS AROUND TEST (more people)" \_n \_dup(80) "=" \_n \_dup(80) "="

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

qui qreg ct\_mean i.count\_pos\_target i.any\_evidence\_sympt\_around i.any\_prior\_pos i.first\_test i.sex i.ethnicity\_wo

noi qreg

noi testparm i.count\_pos\_target

noi testparm 2.count\_pos\_target

noi testparm 3.count\_pos\_target

noi testparm 4.count\_pos\_target

noi testparm i.count\_pos\_target, equal

noi testparm 4.count\_pos\_target 3.count\_pos\_target, equal

noi margins i.count\_pos\_target, atmeans

foreach var of varlist any\_evidence\_sympt\_around any\_prior\_pos first\_test sex ethnicity\_wo {

 noi testparm i.`var'

 noi margins i.`var'

}

noi di \_n \_dup(80) "=" \_n "WITH DATE" \_n \_dup(80) "="

qui qreg ct\_mean i.count\_pos\_target i.any\_evidence\_sympt\_around i.any\_prior\_pos i.first\_test i.sex i.ethnicity\_wo DATE\_\*

noi qreg

noi testparm DATE\_\*

noi di \_n \_dup(80) "=" \_n "SPLITTING SYMPTOMS" \_n \_dup(80) "="

qui qreg ct\_mean i.count\_pos\_target i.sympt\_around\_split i.any\_prior\_pos i.first\_test i.sex i.ethnicity\_wo

noi qreg

noi testparm i.sympt\_around\_split

noi testparm i.sympt\_around\_split, equal

noi testparm 1.sympt\_around\_split

noi testparm 2.sympt\_around\_split

noi margins i.sympt\_around\_split

noi di \_n \_dup(80) "=" \_n "Effect of other variables" \_n \_dup(80) "="

foreach var of varlist hh\_pos\_gt1 ethnicityg ageg\_sy imdg ever\_patientfacing occupation ever\_lthc smoke {

 qui qreg ct\_mean i.count\_pos\_target i.any\_evidence\_sympt\_around i.any\_prior\_pos i.first\_test i.sex i.ethnicity\_wo i.`var'

 noi qreg

 noi testparm i.`var'

 \*noi margins i.`var'

}

foreach thing in age\_at\_visit AGE\_\* imd\_pc\_new IMD\_\* {

 qui qreg ct\_mean i.count\_pos\_target i.any\_evidence\_sympt\_around i.any\_prior\_pos i.first\_test i.sex i.ethnicity\_wo `thing'

 noi qreg

 noi testparm `thing'

}

noi di \_n \_dup(80) "=" \_n "Mixed model with RE for hh\_id - check similar" \_n \_dup(80) "="

qui mixed ct\_mean i.count\_pos\_target i.any\_evidence\_sympt\_around i.any\_prior\_pos i.first\_test i.sex i.ethnicity\_wo || hh\_id: || participant\_id:

noi mixed

qui mixed ct\_mean i.count\_pos\_target i.any\_evidence\_sympt\_around i.any\_prior\_pos i.first\_test i.sex i.ethnicity\_wo || hh\_id:

noi mixed

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

noi di \_n \_dup(80) "\*" \_n \_dup(80) "\*" \_n "EXCL SYMPTOMS, COUNT: FULL MULTIVARIABLE MODEL with SYMPTOMS AT TEST" \_n \_dup(80) "\*" \_n \_dup(80) "\*"

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

qui qreg ct\_mean i.any\_prior\_pos i.first\_test i.sex i.ethnicity\_wo i.ageg\_sy i.imdg

noi qreg

foreach var of varlist any\_prior\_pos first\_test sex ethnicity\_wo ageg\_sy imdg {

 noi testparm i.`var'

 noi margins i.`var'

}

qui qreg ct\_mean i.any\_prior\_pos i.first\_test i.sex i.ethnicity\_wo i.ageg\_sy i.imdg DATE\_\*

noi qreg

qui qreg ct\_mean i.any\_prior\_pos i.first\_test i.sex i.ethnicity\_wo AGE\_\* imd\_pc\_new

noi qreg

foreach var of varlist any\_prior\_pos first\_test sex ethnicity\_wo {

 noi testparm i.`var'

 noi margins i.`var'

}

noi testparm imd\_pc\_new

noi lincom 20\*imd\_pc\_

\* marginal effect of age

summ any\_prior\_pos

local b\_prior=r(mean)

summ first\_test

local b\_first=r(mean)

summ sex

local b\_sex=r(mean)-1

summ ethnicity\_wo

local b\_eth=r(mean)-1

summ imd\_pc

local b\_imd=r(mean)

gen plot\_age=\_b[\_cons]+`b\_prior'\*\_b[1.any\_prior\_pos]+`b\_first'\*\_b[1.first\_test]+`b\_sex'\*\_b[1.sex]+`b\_eth'\*\_b[1.ethnicity\_wo]+`b\_imd'\*\_b[imd\_pc\_new ]

for num 1/4: replace plot\_age=plot\_age+AGE\_X\*\_b[AGE\_X]

noi di \_n \_dup(80) "=" "Effect of other variables" \_n \_dup(80) "="

foreach var of varlist hh\_pos\_gt1 ethnicityg ever\_patientfacing occupation ever\_lthc smoke {

 qui qreg ct\_mean i.any\_prior\_pos i.first\_test i.sex i.ethnicity\_wo AGE\_\* imd\_pc\_new i.`var'

 noi qreg

 noi testparm i.`var'

}

foreach thing in IMD\_2-IMD\_4 DATE\_\* {

 qui qreg ct\_mean i.any\_prior\_pos i.first\_test i.sex i.ethnicity\_wo AGE\_\* imd\_pc\_new `thing'

 noi qreg

 noi testparm `thing'

}

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\* G: CT, symptoms, evidence over time

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

noi di \_n \_dup(80) "=" \_n "G: CT, symptoms, evidence over time" \_n \_dup(80) "="

noi tab wofy

preserve

collapse (p10) p10=ct\_mean (p25) p25=ct\_mean (p50) p50=ct\_mean (p75) p75=ct\_mean (p90) p90=ct\_mean (mean) mean=ct\_mean (min) visit\_date (count) n=ct\_mean, by(wofy WOFY)

summ wofy

local min=r(min)

local max=r(max)

gen byte y=40 if mod(wofy,2)==0

replace y=39 if mod(wofy,2)==1

format n %4.0f

scatter mean p10 p25 p50 p75 p90 wofy, c(l..) s(i..) lp(dash solid..) color(red navy blue orange purple gray) || /\*

\*/ scatter y wofy, s(i) c(i) mlabel(n) mlabcolor(black) mlabpos(0) mlabsize(\*0.6) || /\*

\*/ , graphregion(color(white)) ylabel(10(5)35 37 40 "N", angle(0) format(%2.0f) labsize(small)) ytitle(Ct value) /\*

\*/ xtitle("Week starting (2020/1)") xlabel(`min'(2)`max', valuelabels labsize(vsmall) angle(45)) /\*

\*/ legend(order(1 "Mean" - "Percentiles" 2 "10th" 3 "25th" 4 "50th" 5 "75th" 6 "90th") rows(1) size(vsmall) keygap(\*0.1) symxsize(\*0.5))

graph export "$Graphs/ct\_monitoring\_fig\_3A\_UK.emf", replace

export excel wofy mean p10 p25 p50 p75 p90 using "$Logs/Figure3A.xlsx", replace firstrow(var)

restore

preserve

keep if country==0

collapse (p10) p10=ct\_mean (p25) p25=ct\_mean (p50) p50=ct\_mean (p75) p75=ct\_mean (p90) p90=ct\_mean (mean) mean=ct\_mean (min) visit\_date (count) n=ct\_mean, by(wofy WOFY)

gen byte y=40 if mod(wofy,2)==0

replace y=39 if mod(wofy,2)==1

format n %4.0f

summ wofy

local min=r(min)

local max=r(max)

\* England lockdowns 4nov-1dec; 20dec-8 march; first one schools open 1 June

gen byte upper=38 if inrange(wofy,45,49)

gen byte upper2=38 if inrange(wofy,52,62)

gen byte upper3=38 if inrange(wofy,18,23)

gen byte upper4=38 if inrange(wofy,62,65)

gen byte lower=10 if inrange(wofy,45,49)

gen byte lower2=10 if inrange(wofy,52,62)

gen byte lower3=10 if inrange(wofy,18,23)

gen byte lower4=10 if inrange(wofy,62,65)

twoway rarea upper lower wofy, color(gs15) || /\*

\*/ rarea upper4 lower4 wofy, color(gs15) || /\*

\*/ rarea upper2 lower2 wofy, color(gs13) || /\*

\*/ rarea upper3 lower3 wofy, color(gs13) || /\*

\*/ scatter mean p10 p25 p50 p75 p90 wofy, c(l..) s(i..) lp(dash solid..) color(red navy blue orange purple gray) || /\*

\*/ scatter y wofy, s(i) c(i) mlabel(n) mlabcolor(black) mlabpos(0) mlabsize(\*0.6)|| /\*

\*/ , graphregion(color(white)) ylabel(10(5)35 37 40 "N", angle(0) format(%2.0f) labsize(small)) ytitle(Ct value (England only)) /\*

\*/ xtitle("Week starting (2020/1) (England only)") xlabel(`min'(2)`max', valuelabels labsize(vsmall) angle(45)) /\*

\*/ legend(order(5 "Mean" - "Percentiles" 6 "10th" 7 "25th" 8 "50th" 9 "75th" 10 "90th") rows(1) size(vsmall) keygap(\*0.1) symxsize(\*0.5))

graph export "$Graphs/ct\_monitoring\_fig\_3B\_Eng.emf", replace

export excel wofy mean p10 p25 p50 p75 p90 using "$Logs/Figure3B\_Eng.xlsx", replace firstrow(var)

restore

\* symptoms & evidence by WEEK

gen byte high=(evidence==1)

gen byte mod=(evidence==2)

gen byte low=(evidence==3)

\* NUMBERS TOO FEW INITIALLY - have to group

preserve

recode wofy 19=18 21=20 23=22 25=24 27=26 29=28 31=30 33=32 35=34

collapse (count) n=result\_mk n\_now=any\_evidence\_sympt\_now n\_around=any\_evidence\_sympt\_around (sum) \*evidence\* \*cgh\* high mod low, by(wofy)

local max=\_N

foreach var of varlist high mod low {

 gen `var'\_p=100\*`var'/n

 for any l u: gen byte `var'\_X=.

 forval i=1(1)`max' {

 local N=n[`i']

 local n=`var'[`i']

 qui cii prop `N' `n', exact

 qui replace `var'\_l=r(lb)\*100 if \_n==`i'

 qui replace `var'\_u=r(ub)\*100 if \_n==`i'

 }

}

foreach var of varlist any\_evidence\_sympt\_now sympt\_now\_cgh {

 gen `var'\_p=100\*`var'/n\_now

 for any l u: gen byte `var'\_X=.

 forval i=1(1)`max' {

 local N=n\_now[`i']

 local n=`var'[`i']

 qui cii prop `N' `n', exact

 qui replace `var'\_l=r(lb)\*100 if \_n==`i'

 qui replace `var'\_u=r(ub)\*100 if \_n==`i'

 }

}

foreach var of varlist any\_evidence\_sympt\_around sympt\_around\_cgh {

 gen `var'\_p=100\*`var'/n\_around

 for any l u: gen byte `var'\_X=.

 forval i=1(1)`max' {

 local N=n\_around[`i']

 local n=`var'[`i']

 qui cii prop `N' `n', exact

 qui replace `var'\_l=r(lb)\*100 if \_n==`i'

 qui replace `var'\_u=r(ub)\*100 if \_n==`i'

 }

}

rename any\_evidence\_sympt\_now\* now\*

rename any\_evidence\_sympt\_around\* aro\*

rename sympt\_now\_cghfevamn\* now\_cgh\*

rename sympt\_around\_cghfevamn\* aro\_cgh\*

gen byte y=80 if mod(wofy,2)==0

replace y=78 if mod(wofy,2)==1

gen int y2=105 if mod(wofy,2)==0

replace y2=103 if mod(wofy,2)==1

gen plot1=wofy-0.2

gen plot2=wofy-0.2/3

gen plot3=wofy+0.2/3

gen plot4=wofy+0.2

for var plot\*: label values X wofy

summ wofy

local min=r(min)

local max=r(max)

twoway rcapsym now\_u now\_l plot1, color(black) lcolor(black) s(i) lwidth(thin) || /\*

\*/ scatter now\_p plot1, color(black) lcolor(black) s(o) c(l) || /\*

\*/ rcapsym aro\_u aro\_l plot2, color(gs8) lcolor(gs8) s(i) lwidth(thin) || /\*

\*/ scatter aro\_p plot2, color(gs8) lcolor(gs8) s(s) c(l) || /\*

\*/ rcapsym now\_cgh\_u now\_cgh\_l plot3, color(black) lcolor(black) s(i) lwidth(thin) || /\*

\*/ scatter now\_cgh\_p plot3, color(black) lcolor(black) s(o) c(l) mfcolor(white) || /\*

\*/ rcapsym aro\_cgh\_u aro\_cgh\_l plot4, color(gs8) lcolor(gs8) s(i) lwidth(thin) || /\*

\*/ scatter aro\_cgh\_p plot4, color(gs8) lcolor(gs8) s(s) c(l) mfcolor(white) || /\*

\*/ scatter y wofy, s(i) c(i) mlabel(n) mlabcolor(black) mlabpos(0) mlabsize(\*0.6)|| /\*

\*/ , graphregion(color(white)) ytitle(Percentage of positive tests (95% CI)) ylabel(0(10)70 80 "N", angle(0) labsize(small)) /\*

\*/ xlabel(`min'(2)`max', valuelabels labsize(vsmall) angle(45)) xtitle("") /\*

\*/ legend(order(2 "Any evidence of symptoms at test" 6 "Cough, fever, anosmia at test" 4 "Any evidence of symptoms around test" 8 "Cough, fever, anosmia around test" ) size(vsmall) rows(2))

graph export "$Graphs\symptom\_over\_wofy\_fig\_3C.emf", replace

export excel wofy now\_p now\_l now\_u aro\_p aro\_l aro\_u now\_cgh\_p now\_cgh\_l now\_cgh\_u aro\_cgh\_p aro\_cgh\_l aro\_cgh\_u using "$Logs\Figure3C.xlsx", replace firstrow(var)

replace plot1=wofy-0.2

replace plot3=wofy+0.2

twoway rcapsym high\_u high\_l plot1, color(black) lcolor(black) s(i) lwidth(thin) || /\*

\*/ scatter high\_p plot1, color(black) lcolor(black) s(s) c(l) || /\*

\*/ rcapsym mod\_u mod\_l wofy, color(gs8) lcolor(gs8) s(i) lwidth(thin) || /\*

\*/ scatter mod\_p wofy, color(gs8) lcolor(gs8) s(d) c(l) mfcolor(white) || /\*

\*/ rcapsym low\_u low\_l plot3, color(gs13) lcolor(gs13) s(i) lwidth(thin) || /\*

\*/ scatter low\_p plot3, color(gs13) lcolor(gs13) s(t) c(l) mfcolor(white) || /\*

\*/ scatter y2 wofy, s(i) c(i) mlabel(n) mlabcolor(black) mlabpos(0) mlabsize(\*0.6) || /\*

\*/ , graphregion(color(white)) ytitle(Percentage of positive tests (95% CI)) ylabel(0(10)100 105 "N", angle(0) labsize(small)) /\*

\*/ xlabel(`min'(2)`max', valuelabels labsize(vsmall) angle(45)) xtitle("") /\*

\*/ legend(order(2 "Higher evidence" 4 "Moderate evidence" 6 "Lower evidence" ) size(vsmall) rows(1))

graph export "$Graphs\evidence\_over\_wofy\_fig\_3D.emf", replace

export excel wofy high\_p high\_l high\_u mod\_p mod\_l mod\_u low\_p low\_l low\_u using "$Logs\Figure3D.xlsx", replace firstrow(var)

restore

\* Ct distribution by evidence over time

\* NUMBERS TOO FEW INITIALLY - have to group

preserve

recode wofy 19=18 21=20 23=22 25=24 27=26 29=28 31=30 33=32 35=34

collapse (p25) p25=ct\_mean (p50) p50=ct\_mean (p75) p75=ct\_mean (count) n=ct\_mean, by(wofy evidence)

reshape wide p25 p50 p75 n, i(wofy) j(evidence)

summ wofy

local min=r(min)

local max=r(max)

scatter p501 p502 p503 wofy, c(l..) s(i..) lp(dash solid..) color(black gs8 gs13) || /\*

\*/ , graphregion(color(white)) ylabel(10(5)35 37, angle(0) format(%2.0f) labsize(small)) ytitle(Median Ct value) /\*

\*/ xtitle("Week starting (2020/1)") xlabel(`min'(2)`max', valuelabels labsize(vsmall) angle(45)) /\*

\*/ legend(order(1 "Higher evidence" 2 "Moderate evidence" 3 "Lower evidence") rows(1) size(vsmall) keygap(\*0.1) symxsize(\*0.5))

graph export "$Graphs/ct\_by\_evidence\_fig\_4A.emf", replace

for num 1/3 \ any high mod low: rename p50X median\_Y

export excel wofy median\_high median\_mod median\_low using "$Logs\Figure4A.xlsx", replace firstrow(var)

restore

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\* F: Antibody response

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

noi di \_n \_dup(80) "=" \_n \_dup(80) "=" \_n "F: Antibody response - first positive only" \_n \_dup(80) "=" \_n \_dup(80) "="

keep participant visit\_date

sort participant visit\_date

by participant: keep if \_n==1

rename visit\_date first\_pos\_date

merge 1:m participant using "`main'", update

drop if \_m==2

assert \_m==3

drop \_m

\* remove failed results

replace result\_tdi=. if result\_tdi>1

gen byte flag=(result\_tdi<=1)

egen byte any\_ab=max(flag), by(participant)

egen byte count\_ab=sum(flag), by(participant)

drop flag

bysort participant (visit\_date): gen byte first=1 if \_n==1

noi tab any\_ab if first==1

noi tab count\_ab if first==1

noi tab count\_ab if first==1 & count\_ab>1

keep if result\_tdi<.

\* check no duplicates ab on same day

gsort participant visit\_date -result\_tdi

by participant visit\_date: assert \_N==1

\* plots of antibody values

\* one participant has non-usual lab positive test (no Ct) BEFORE this test

noi datacheck visit\_date>=first\_pos\_date if result\_mk==1, mess(Swab positive BEFORE first swab positive date - dropped positive from Loughborough) vars(participant visit\_date result\_mk first\_pos\_date) noobs nod

noi replace first\_pos\_date=d(16dec2020) if first\_pos\_date==d(22dec2020) & participant=="DHR-200913152768"

gen int day=visit\_date-first\_pos\_date

noi summ day, det

\* truncate at 5th and 95th percentiles closest to multiples of 30

local lower=-120

local upper=150

sort participant day

\* keep at most one at the end and beginning to stop vertical lines

by participant: drop if \_n<\_N & day<`lower' & day[\_n+1]<=`lower'

by participant: drop if \_n>1 & day>`upper' & day[\_n-1]>=`upper'

replace day=`lower' if day<`lower'

replace day=`upper' if day>`upper'

\* count numbers before and after first pos

for any min max: egen int X=X(day), by(participant)

drop first

bysort participant (visit\_date): gen byte first=1 if \_n==1

\* bin responses

gen int DAY=day

for num `lower'(30)`upper': replace DAY=X if inrange(day,X,X+29)

summ DAY

replace DAY=DAY-30 if DAY==r(max)

replace DAY=14 if inrange(day,14,29)

label define DAY -120 "[-120,-91]" -90 "[-90,-61]" -60 "[-60,-31]" -30 "[-30,-1]" 0 "[0,13]" 14 "[14,29]" 30 "[30,59]" 60 "[60,89]" 90 "[90,119]" 120 "[120,150]"

label values DAY DAY

label var DAY "Days from first positive swab"

noi di \_n "All results"

noi tab DAY result\_tdi, row nokey

\* any positive on or after swab positive

gen byte flag=result\_tdi\*(visit\_date>=first\_pos\_date)

egen byte any\_pos\_ab=max(flag), by(participant)

label var any\_pos\_ab "Positive Ab at or after 1st pos swab"

\* any negative within 120 days before negative to 21 days post negative

replace flag=(result\_tdi==0)\*inrange(visit\_date,first\_pos\_date-120,first\_pos\_date+21)

egen byte any\_neg\_ab=max(flag), by(participant)

label var any\_neg\_ab "Negative Ab in[-120,+21] of 1st pos swab"

drop flag

noi di \_n "Latest positive per window"

sort participant DAY result\_tdi day

noi by participant DAY: keep if \_n==\_N

noi tab DAY result\_tdi, row nokey

drop first

bysort participant (visit\_date): gen byte first=1 if \_n==1

egen byte postonly=min(DAY), by(participant)

replace postonly=(postonly>=0)

label define postonly 0 "Antibody result(s) before 1st swab-positive (day 0)" 1 "No antibody result(s) before 1st swab-positive (day 0)"

label values postonly postonly

noi tab postonly first

noi tab postonly any\_pos\_ab if first==1

noi tab postonly any\_neg\_ab if first==1, cell nokey

noi bysort postonly: tab DAY result\_tdi, row nokey

noi bysort DAY: tab postonly result\_tdi, row nokey chi

gen byte i=1

graph export "$Graphs/swab\_positive\_ab\_over\_time\_perc\_fig\_S4A.emf", replace

graph bar (count) i, subtitle(, size(small)) by(postonly, graphregion(color(white)) note("")) over(result\_tdi) asyvars over(DAY, label(angle(45) labsize(small))) stack percentage bar(1, color(white) lcolor(black)) bar(2, color(black)) ytitle(Percentage of participants with antibody tests) ylabel(0(20)100, angle(0)) legend(order(1 "Negative" 2 "Positive") rows(1))

graph export "$Graphs/fig\_5.emf", replace

collapse (sum) i, by(postonly DAY result\_tdi)

reshape wide i, i(postonly DAY) j(result\_tdi)

gen perc=100\*i1/(i1+i0)

export excel postonly DAY perc using "$Logs\Figure5.xlsx", replace firstrow(var)

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\* H: All tests

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

noi di \_n \_dup(80) "=" \_n "H: all tests by wofy" \_n \_dup(80) "="

use if result\_mk<=1 using "`main'", clear

merge 1:1 participant visit\_id using $Data\swabpositives\_evidence, keepusing(evidence)

assert inlist(\_m,1,3)

drop \_m

assert inlist(result\_mk,0,1)

assert dow(d(1jan2020))==3

gen int dofy=visit\_date-d(1jan2020)+2 if year(visit\_date)==2020

gen int wofy=1+int(dofy/7)

label values wofy wofy

\* 2021

assert dow(d(1jan2021))==5

assert wofy==53 if visit\_date==d(31dec2020)

replace wofy=53 if inrange(visit\_date,d(1jan2021),d(3jan2021))

replace dofy=visit\_date-d(4jan2021) if visit\_date>d(3jan2021)

replace wofy=54+int(dofy/7) if visit\_date>d(3jan2021)

\* 26 April is only 1 person - reset

summ wofy

assert r(min)==17

summ result\_mk if wofy==17 & result\_mk==1

assert r(N)==1

recode wofy 17=18

gen int WOFY=wofy

assert WOFY<.

drop dofy

noi tab wofy

noi tab wofy result\_mk

\* evidence by wofy

gen byte high=(evidence==1)

gen byte mod=(evidence==2)

gen byte low=(evidence==3)

\* NUMBERS TOO FEW INITIALLY - have to group even more

recode wofy 18/20=18 21/23=21 24/26=24 27/29=27 30/32=30 33/35=33

collapse (count) n=result\_mk (sum) high mod low, by(wofy)

local max=\_N

foreach var of varlist high mod low {

 gen `var'\_p=100\*`var'/n

 for any l u: gen byte `var'\_X=.

 forval i=1(1)`max' {

 local N=n[`i']

 local n=`var'[`i']

 qui cii prop `N' `n', exact

 qui replace `var'\_l=r(lb)\*100 if \_n==`i'

 qui replace `var'\_u=r(ub)\*100 if \_n==`i'

 }

}

for any high\_p mod\_p low\_p: gen str6 Xstr=string(X,"%4.3f")+"%"

\*\ replace Xstr=subinstr(Xstr,"0.",".",.)

gen plot1=wofy-0.3

gen plot2=wofy+0.3

for var plot\*: label values X wofy

\* zero low - plot at lowest other point

summ low\_l if low\_l>0

local min=r(min)

assert `min'>0.0005

for var low\_l low\_p: replace X=`min' if X<=0.0005

summ wofy

local min=r(min)

local max=r(max)

\* line plot rather than bar chart

twoway rcapsym high\_u high\_l plot1, color(black) lcolor(black) s(i) lwidth(thin) || /\*

\*/ scatter high\_p plot1, color(black) lcolor(black) s(s) c(l) || /\*

\*/ rcapsym mod\_u mod\_l wofy, color(gs8) lcolor(gs8) s(i) lwidth(thin) || /\*

\*/ scatter mod\_p wofy, color(gs8) lcolor(gs8) s(d) c(l) mfcolor(white) || /\*

\*/ rcapsym low\_u low\_l plot2, color(gs13) lcolor(gs13) s(i) lwidth(thin) || /\*

\*/ scatter low\_p plot2, color(gs13) lcolor(gs13) s(t) c(l) mfcolor(white) || /\*

\*/ , graphregion(color(white)) ytitle(Percentage of all tests (95% CI) [log scale]) ylabel(0.001 0.003 0.005 0.01 0.03 0.05 0.1 0.3 0.5 1.0 0.0005 "0", angle(0) labsize(small)) yscale(log) /\*

\*/ xlabel(`min'(2)`max', valuelabels labsize(vsmall) angle(45)) xtitle("") /\*

\*/ legend(order(2 "Higher evidence" 4 "Moderate evidence" 6 "Lower evidence" ) size(vsmall) rows(1))

graph export "$Graphs\positives\_over\_wofy\_fig4b.emf", replace

export excel wofy high\_p high\_l high\_u mod\_p mod\_l mod\_u low\_p low\_l low\_u using "$Logs\Figure4B.xlsx", replace firstrow(var)

noi list wofy n \*\_p\*, noobs nod

log close