Gareth Davies

2/2/2022

# WRRS (Welsh Results Reports Service) – supporting documentation - coverage report

## Datasets

The WRRS datasets are located in schema SAIL0911V, and exist in the specific tables:

| TABNAME |
| --- |
| WRRS\_OBSERVATION\_REQUEST |
| WRRS\_OBSERVATION\_RESULT |
| WRRS\_REPORT |

Total row count of SAIL0911V.WRRS\_OBSERVATION\_REQUEST = 389,275,283

Total row count of SAIL0911V.WRRS\_OBSERVATION\_RESULT = 2,131,639,361

Total row count of SAIL0911V.WRRS\_REPORT = 221,525,040

## Person counts (ALF\_E counts) and linkage of ALF\_E to C20 cohort

People are identified using the ALF\_E field, which is present in the report table (SAIL0911V.WRRS\_REPORT). In SAIL0911V.WRRS\_REPORT there are 4169403 distinct number of people.

In the C20 cohort (SAILW0911V.C19\_COHORT20) there are 3449059 distinct ALF\_E at the time of running this report.

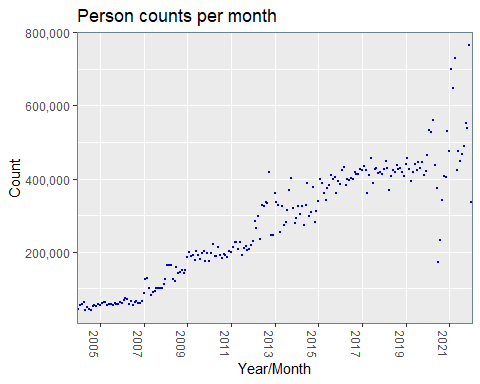
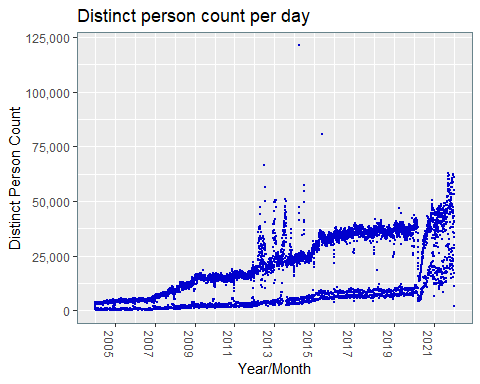
3174172 (76.13%) of the 4169403 distinct ALF\_E in SAIL0911V.WRRS\_REPORT were able to be linked to the C20 cohort.

The AUTHORISED\_DTTM field has been used to define dates. Person counts by (authorised) date are number of people having one or more events within each distinct day. Person counts by month are number of people having one or more events within each distinct day of the specified year/month. For example, if one person has multiple events on a particular day, that person is only counted once for that day.

The AUTHORISED\_DTTM field has been used to define dates. On examining the full date range of person and test counts over all health boards (see appendix), it is seen that there is minimal data prior to 2007. Therefore in order to focus on data rich years and provide a 3 year lag buffer, dates prior to 2004 have been filtered out.

Dates are filtered to years 2004 onwards in the following plots, as data coverage prior to 2004 is minimal. For the full date distribution and recent years distribution, see appendix.

Daily and monthly person counts for all health boards are displayed in the following 2 plots.

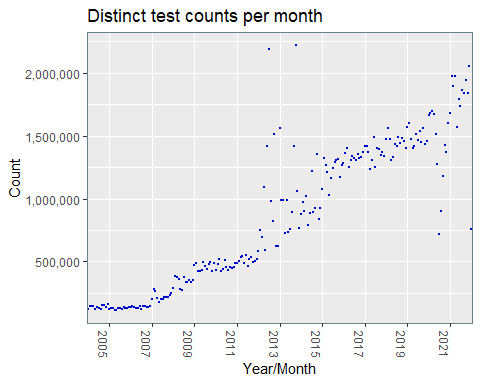
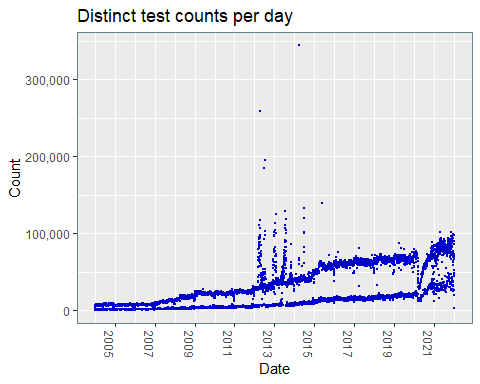
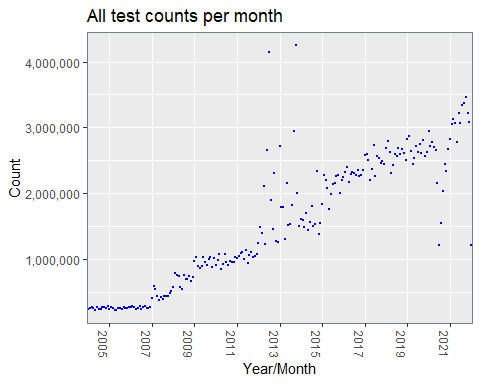
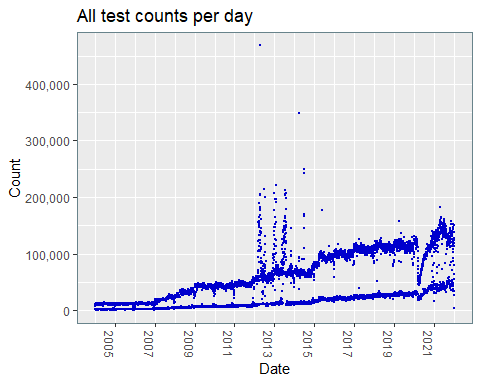


## Test counts (REPORTID\_E counts)

Daily and monthly test counts for all health boards are displayed in the following 4 plots. All test counts are the number of tests (reports) within each day/month. Distinct test counts are the number of dates for each test (report) - i.e. number of tests (report ids) within each day/month removing multiples - so if a report id has >1 occurrences on a single day/month, it is just counted once.

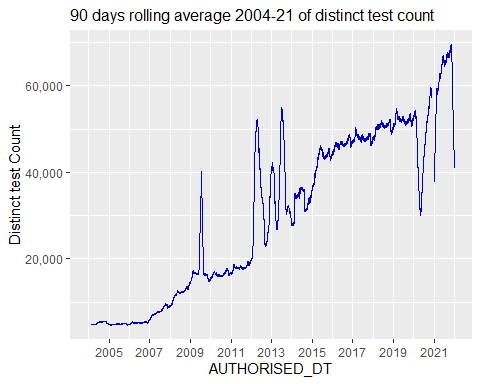
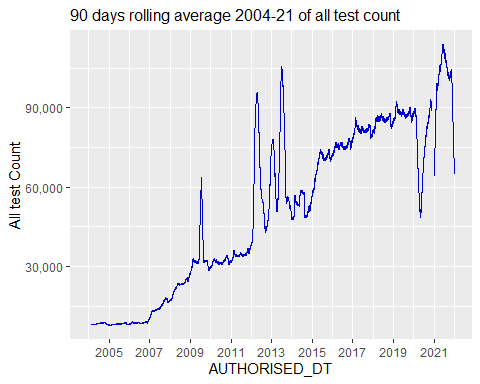
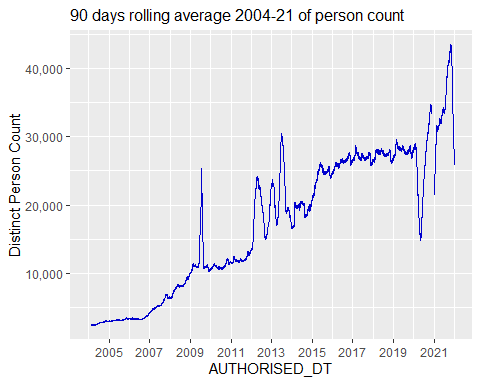
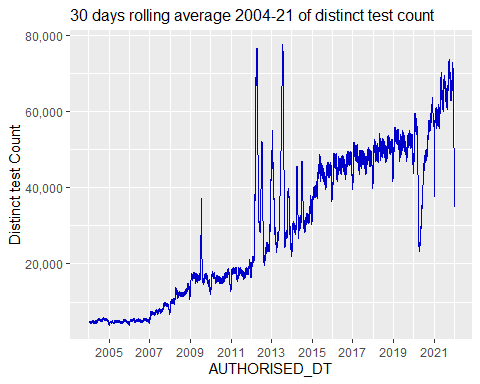
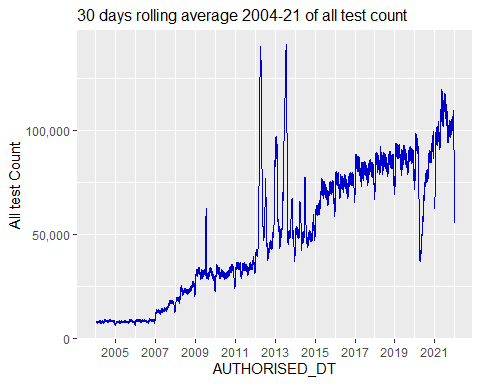
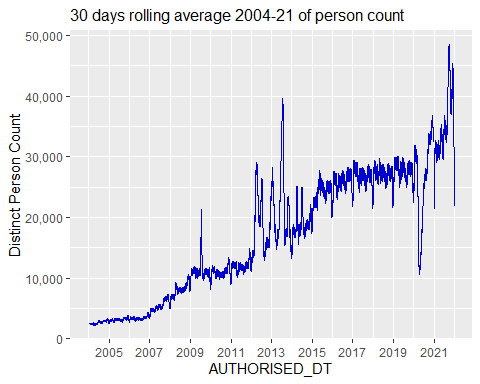
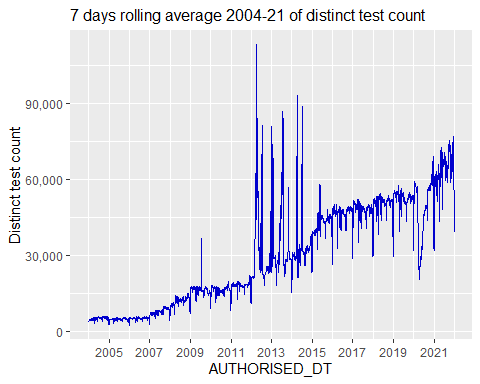
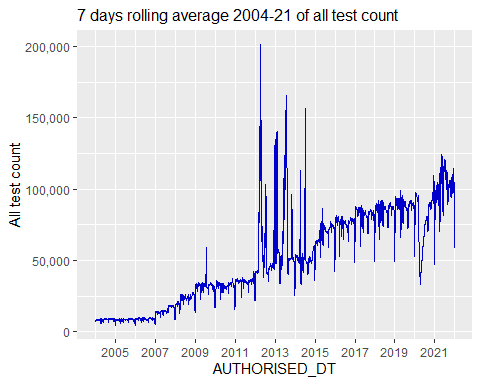
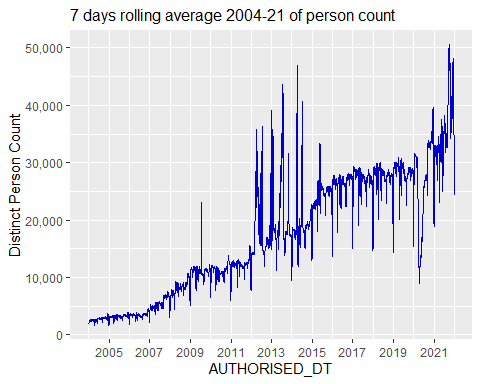
REPORTID\_E exists independently of ALF\_E. Test counts have not been combined with ALF\_E.

The AUTHORISED\_DTTM field has been used to define dates. Dates are filtered to years 2004 onwards in the following plots, as data coverage prior to 2004 is minimal. For the full date distribution and recent years distribution, see appendix.



## Rolling averages

Rolling averages for the period 2004 to present are shown below, for distinct person count, all test count. and distinct test count. All health boards are included. Rolling average timeperiods of 7,30 and 90 days have been explored.



## Coverage counts by health boards

The following abbreviations have been used for the 7 Welsh health boards:

SB: Swansea Bay University Health Board

AB: Aneurin Bevan University Health Board

BC: Betsi Cadwaladr University Health Board

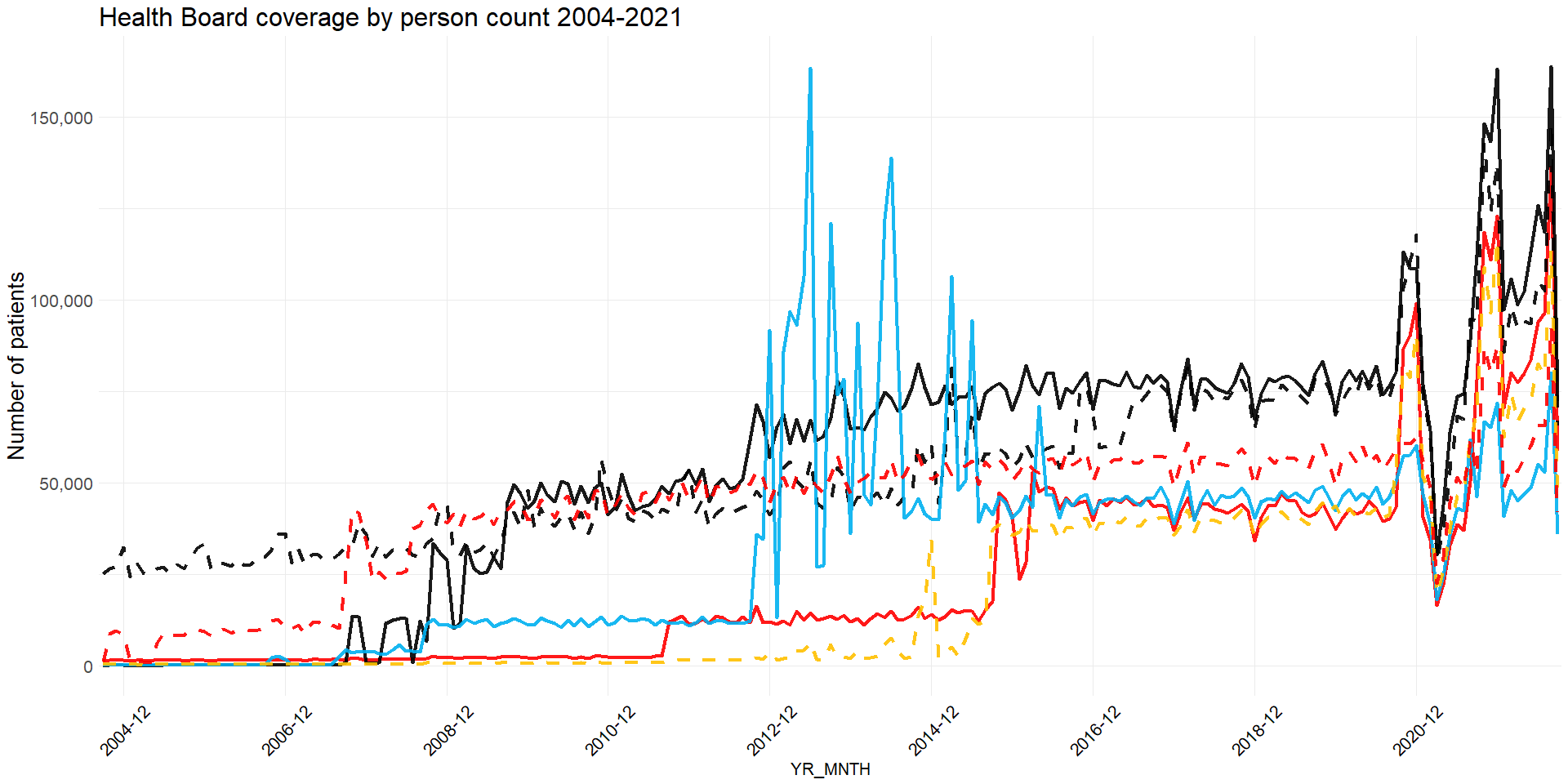
CV: Cardiff and Vale University Health Board

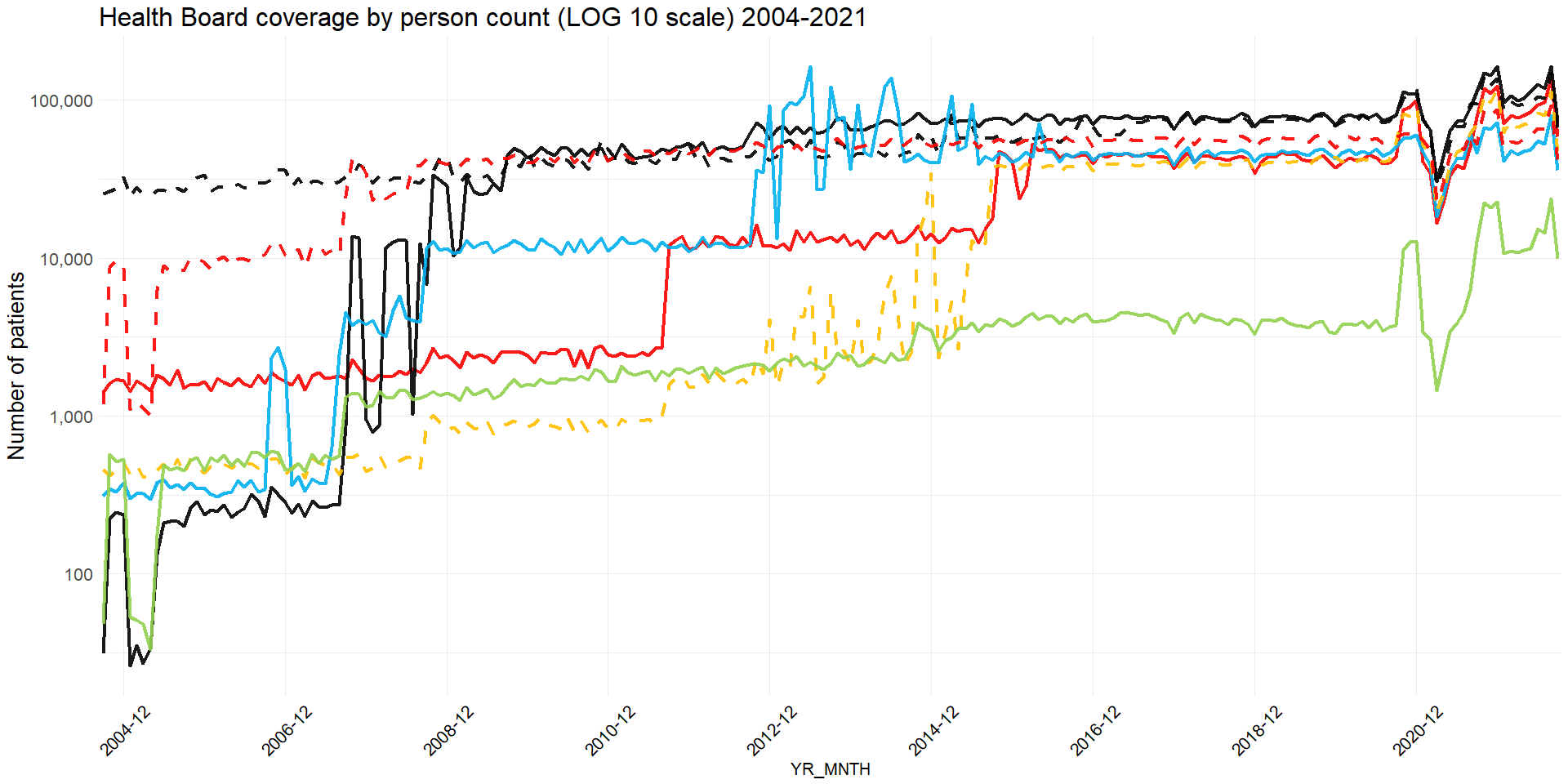
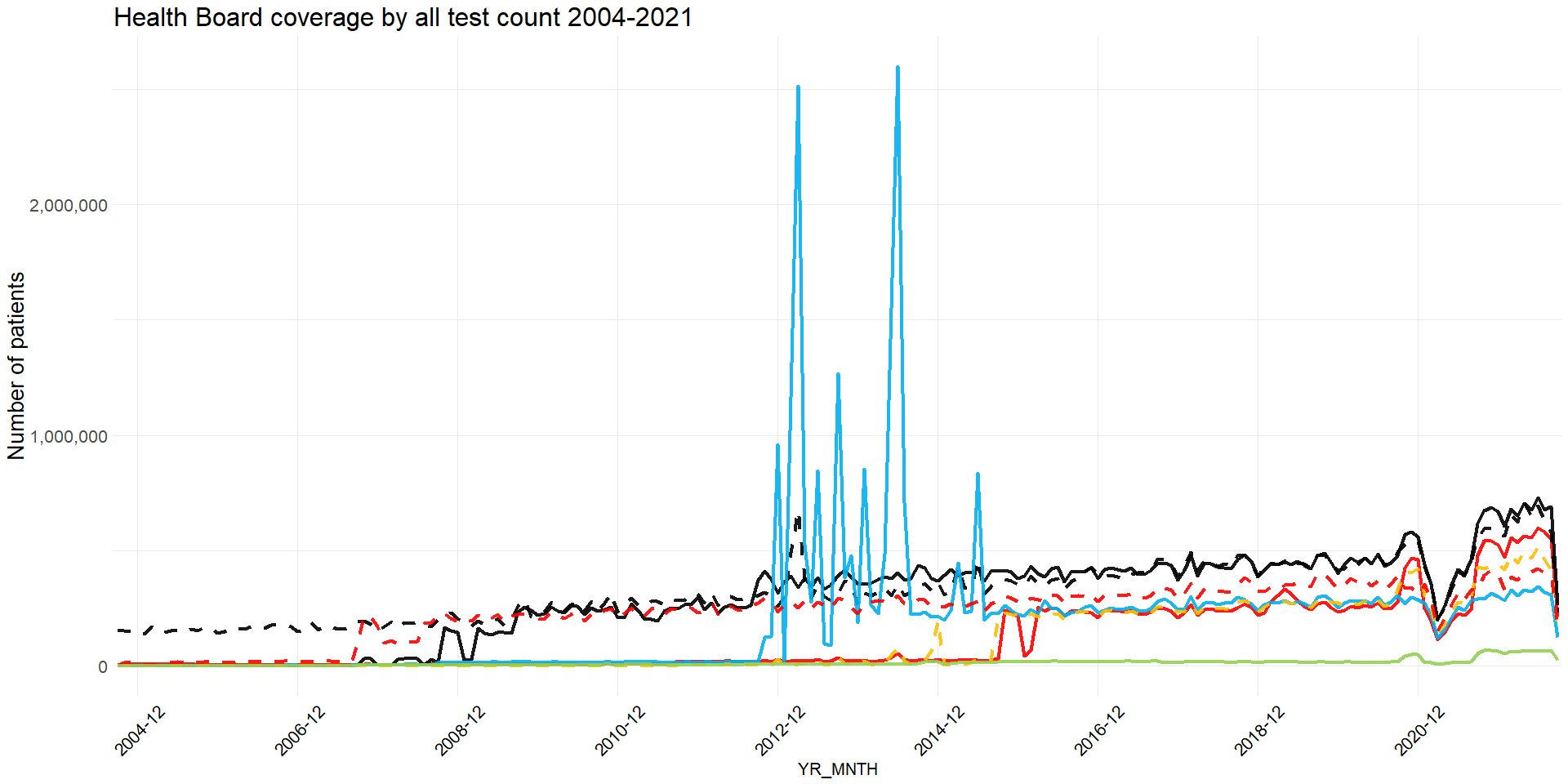
CT: Cwm Taf University Health Board

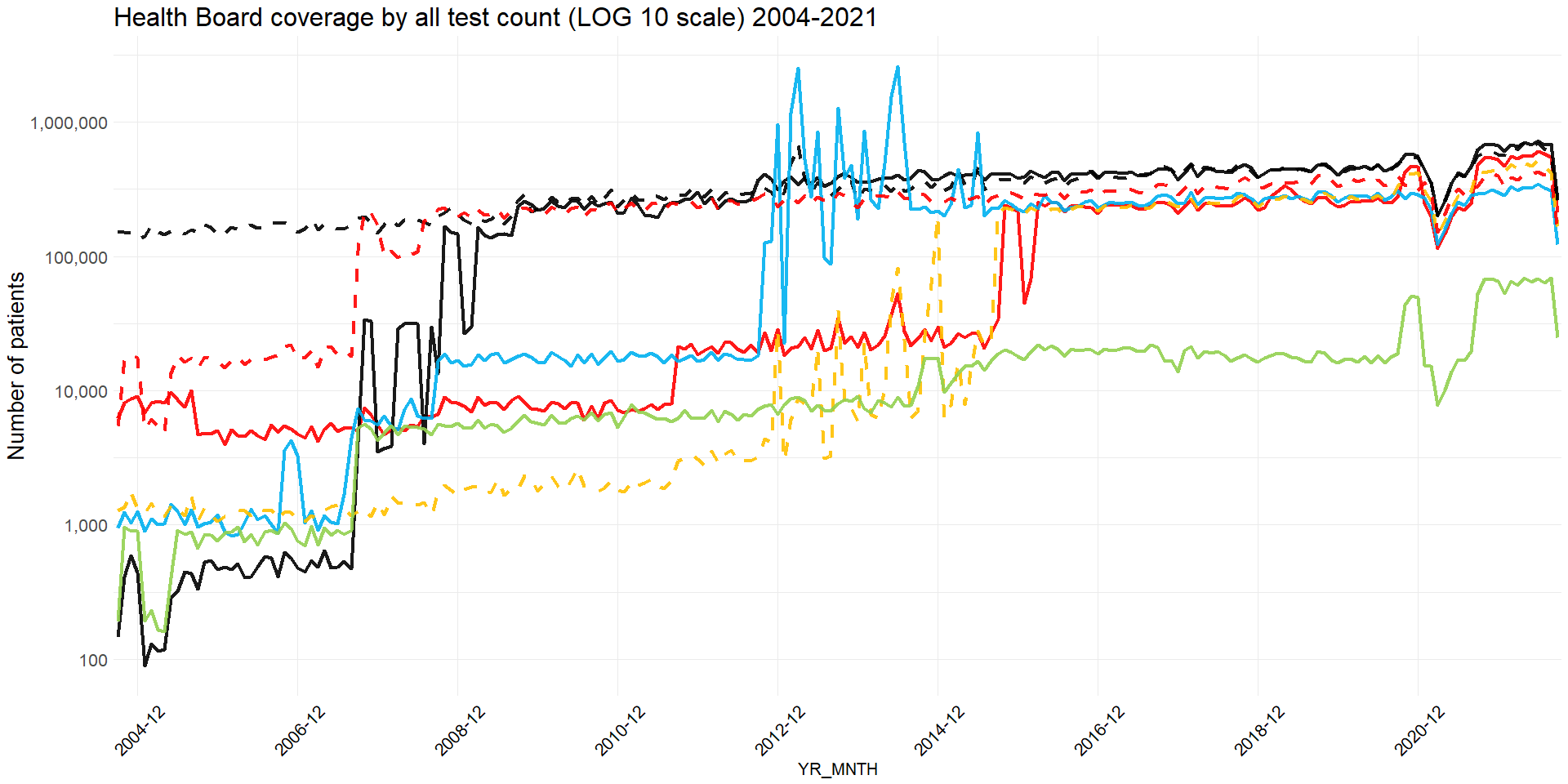
HD: Hywel Dda University Health Board

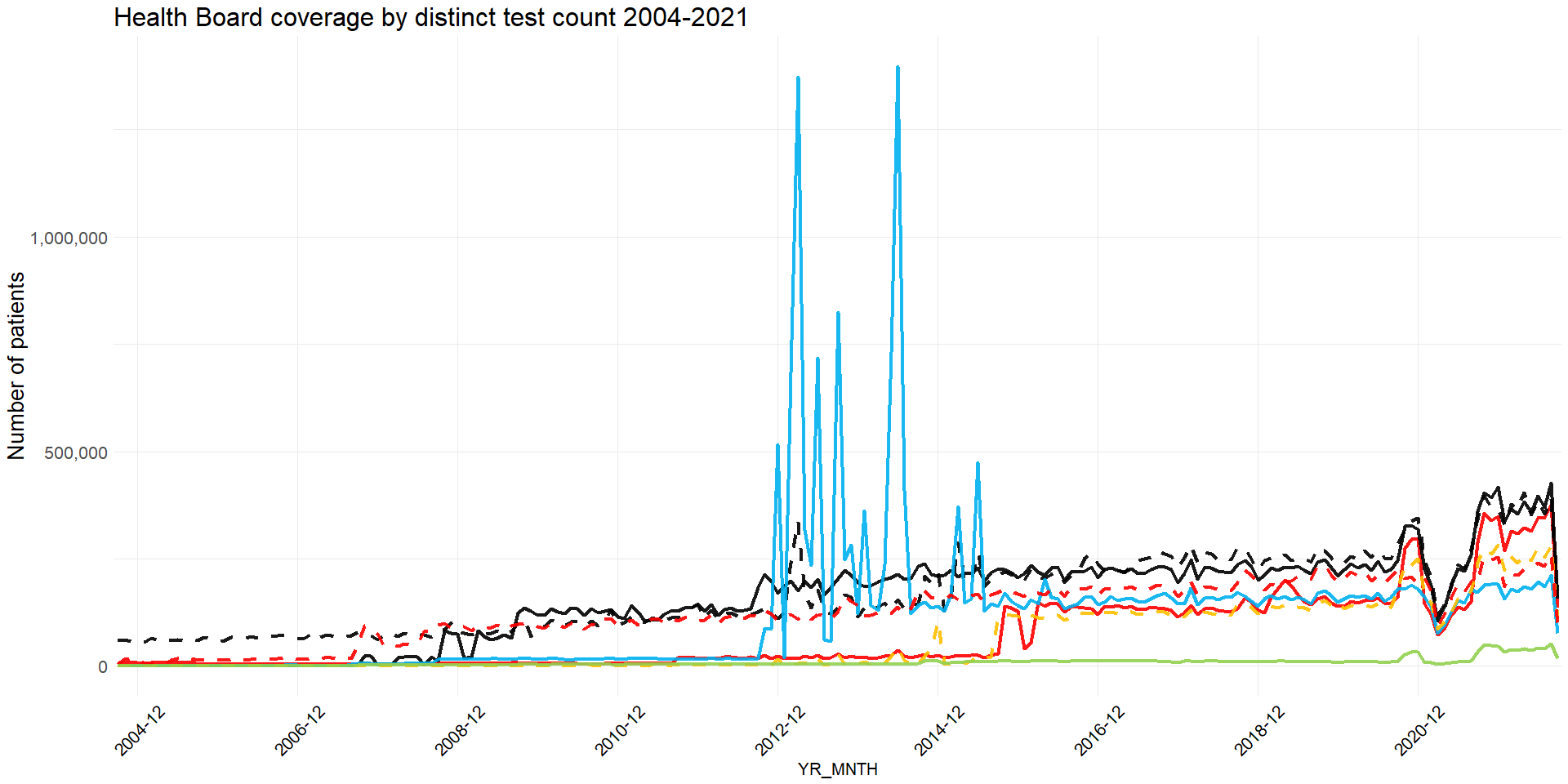
PT: Powys Teaching Health Board

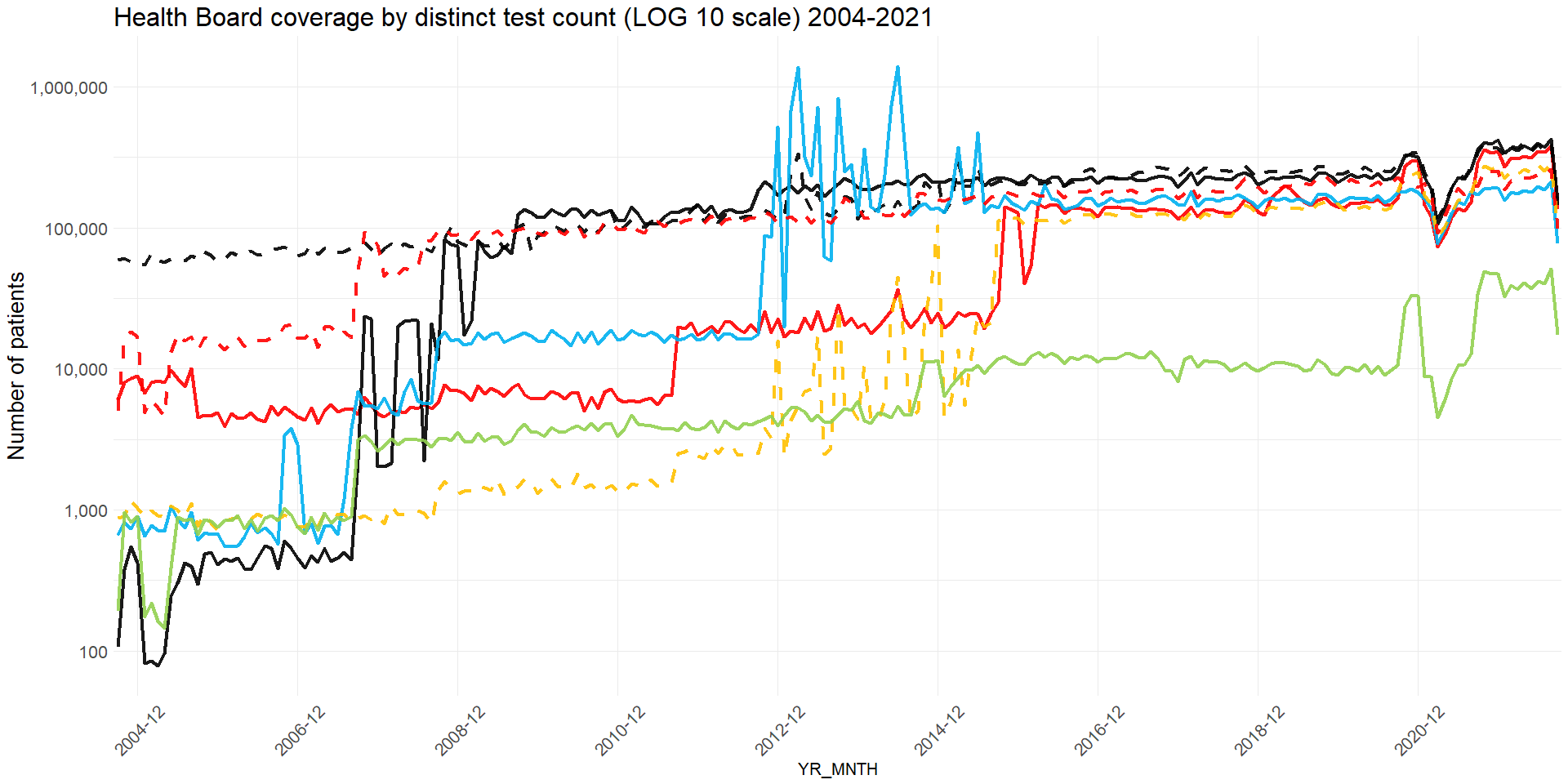
The STATS\_CURR\_CENSUS\_LSOA\_CD field in SAIL0911V.WRRS\_OBSERVATION\_REQUEST was used to differentiate people into the 7 different Welsh health boards.

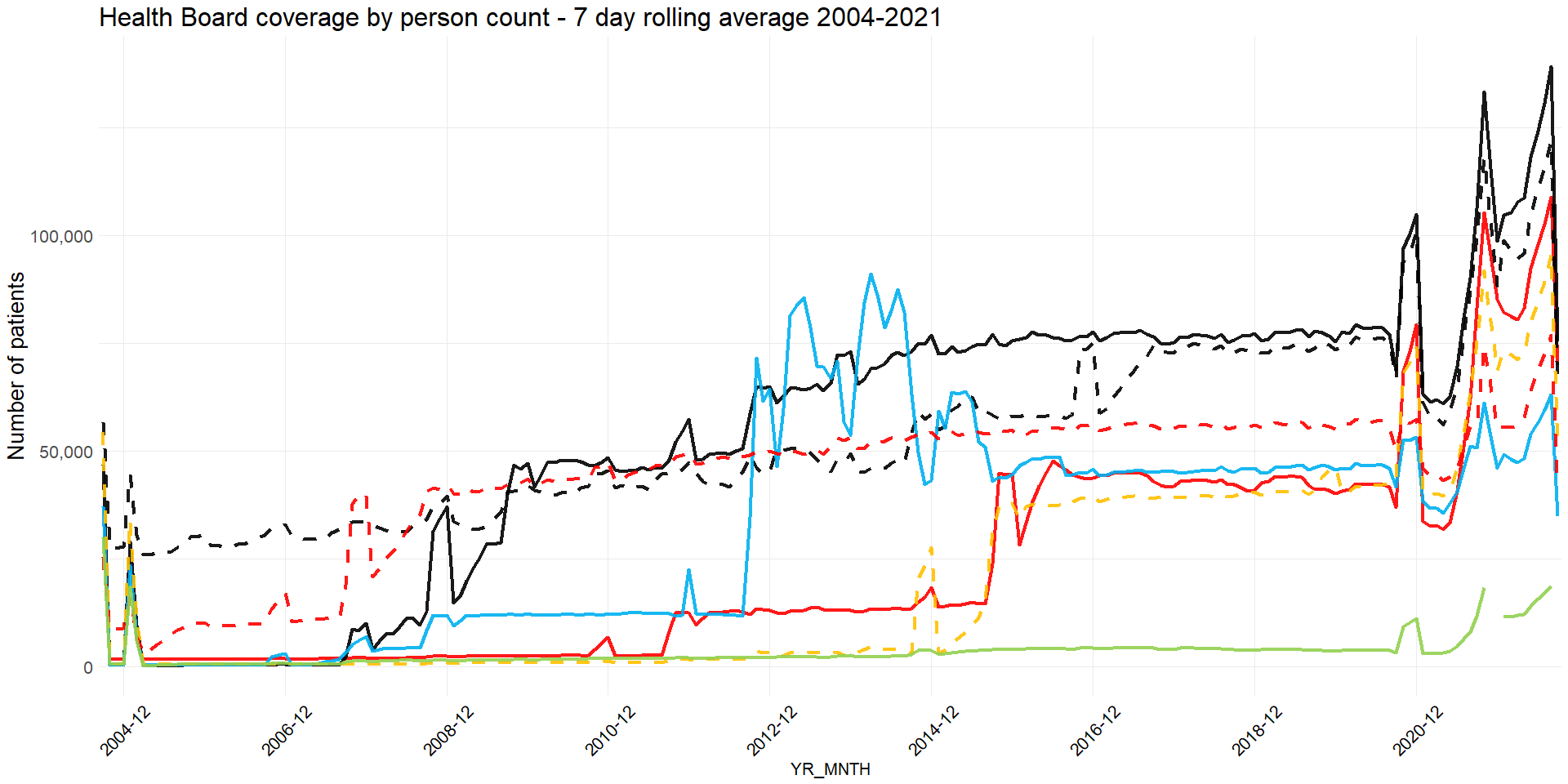
 

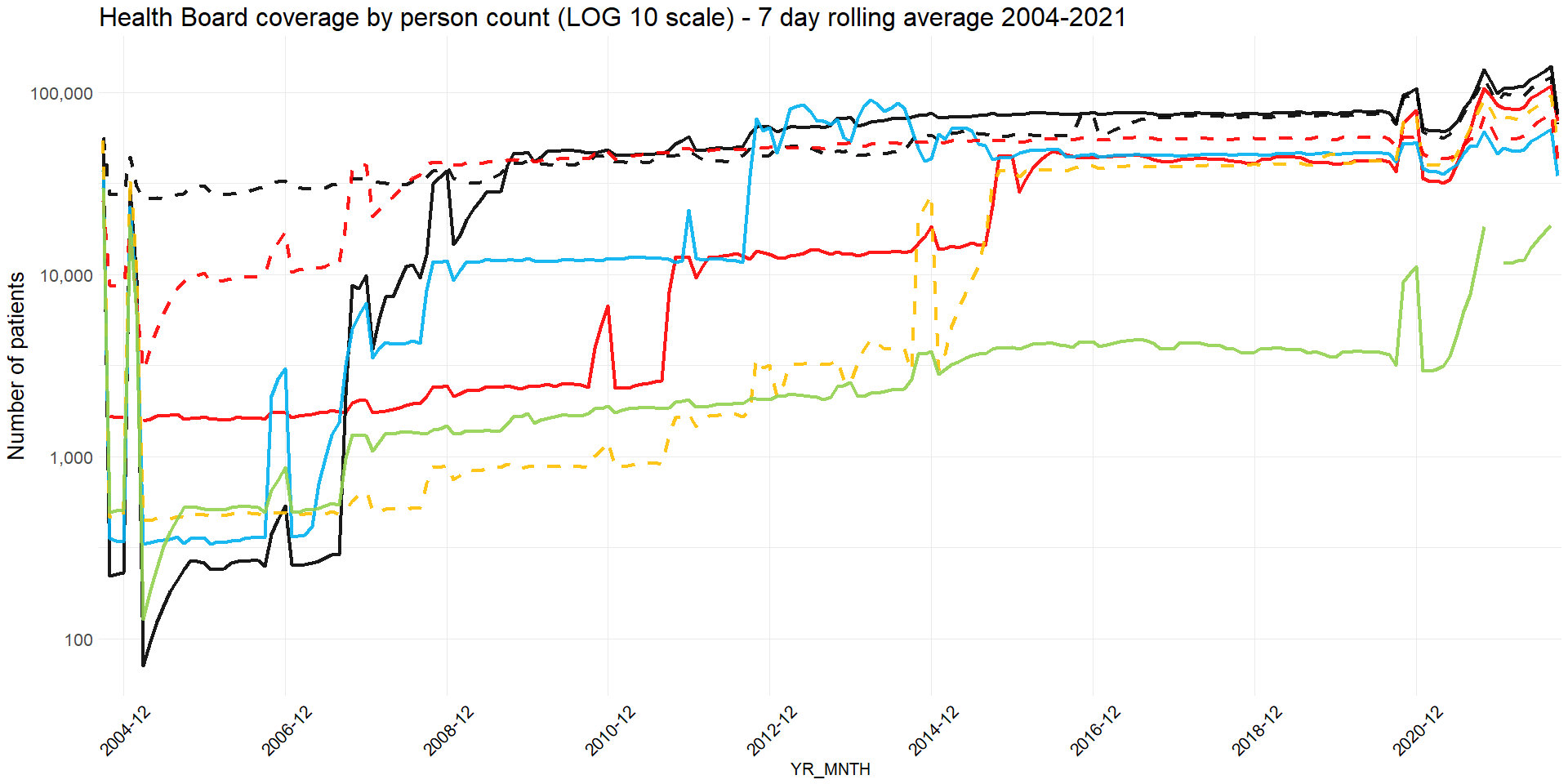
## 

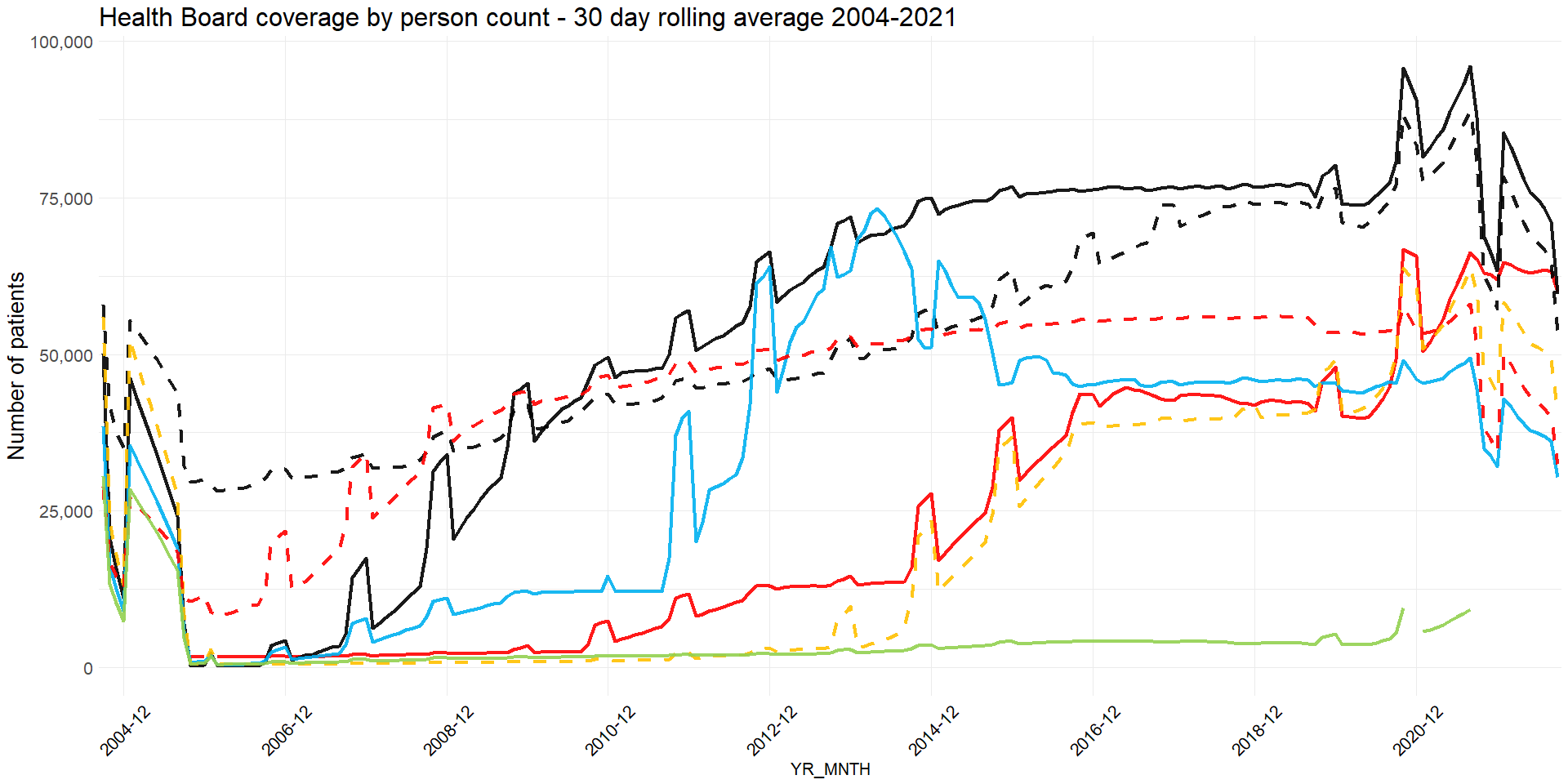
## Rolling averages by health boards

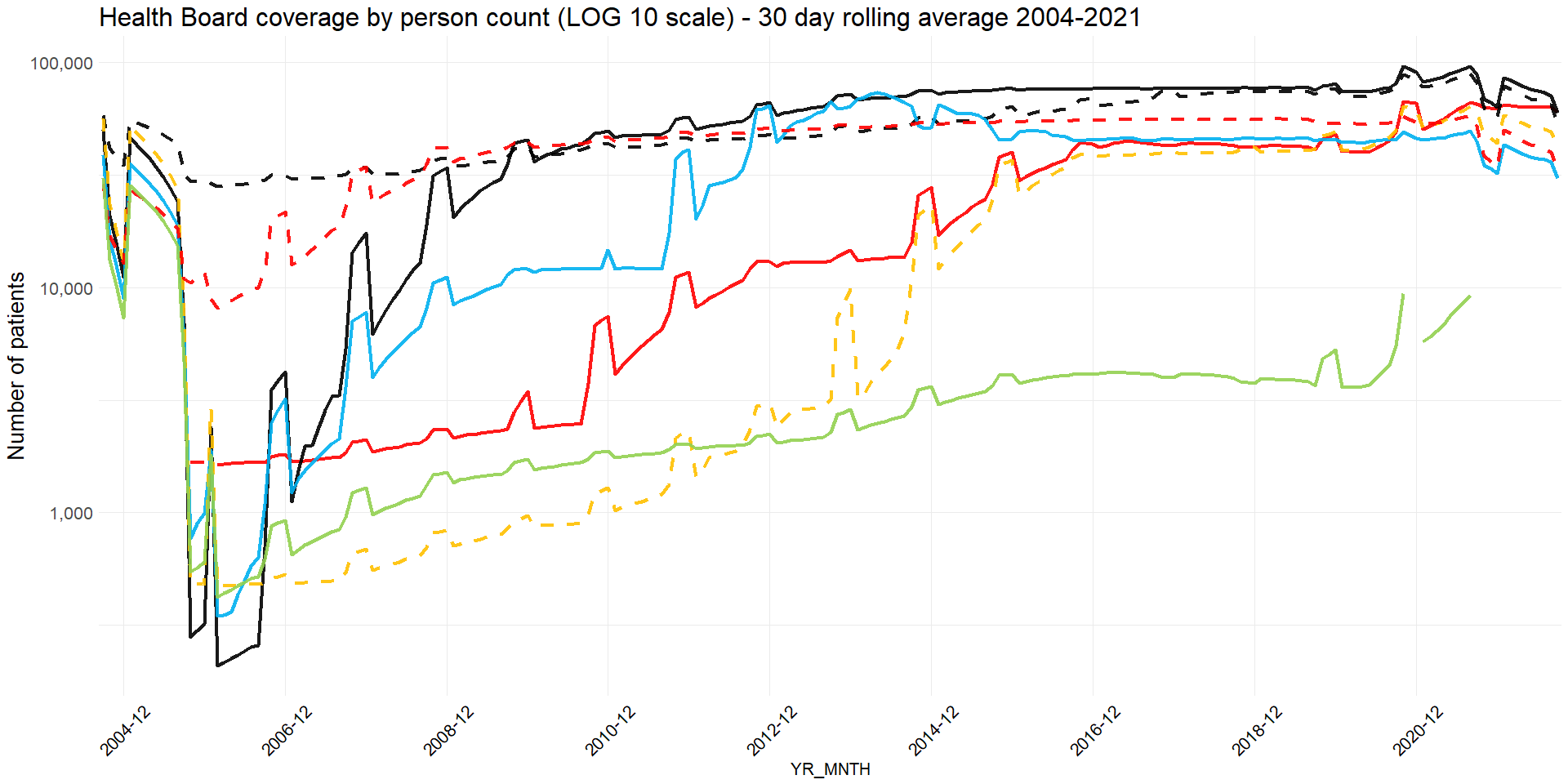
The STATS\_CURR\_CENSUS\_LSOA\_CD field in SAIL0911V.WRRS\_OBSERVATION\_REQUEST was used to differentiate people into the 7 different Welsh health boards, as defined by the LOCAL\_HEALTH\_BOARD\_NAME\_ENGLISH field in the reference table SAILW0911V.JH\_REGION\_LOOKUP.

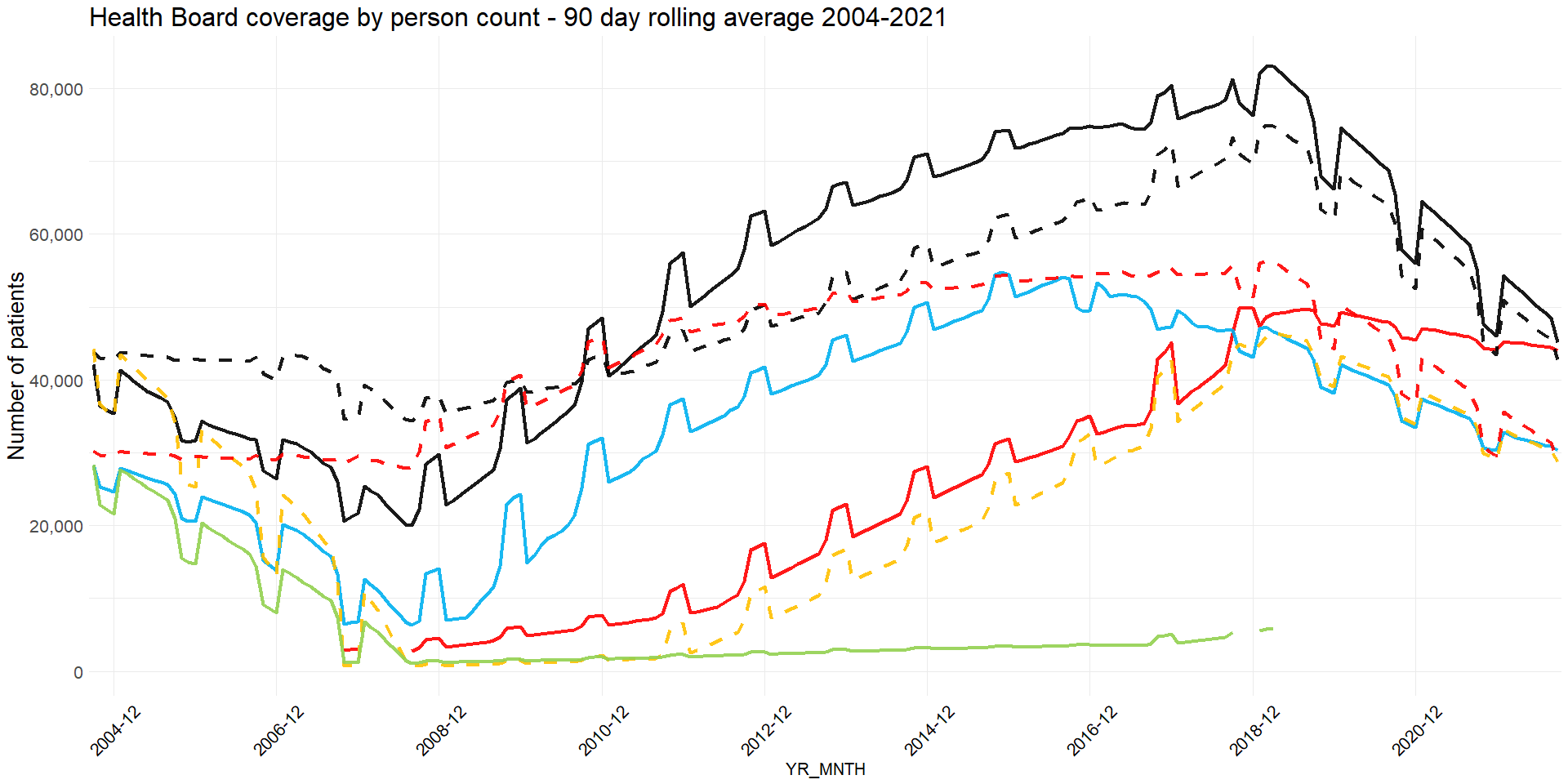
Rolling averages for 7, 30 and 90 days are explored, for person count, all test count and distinct test count.

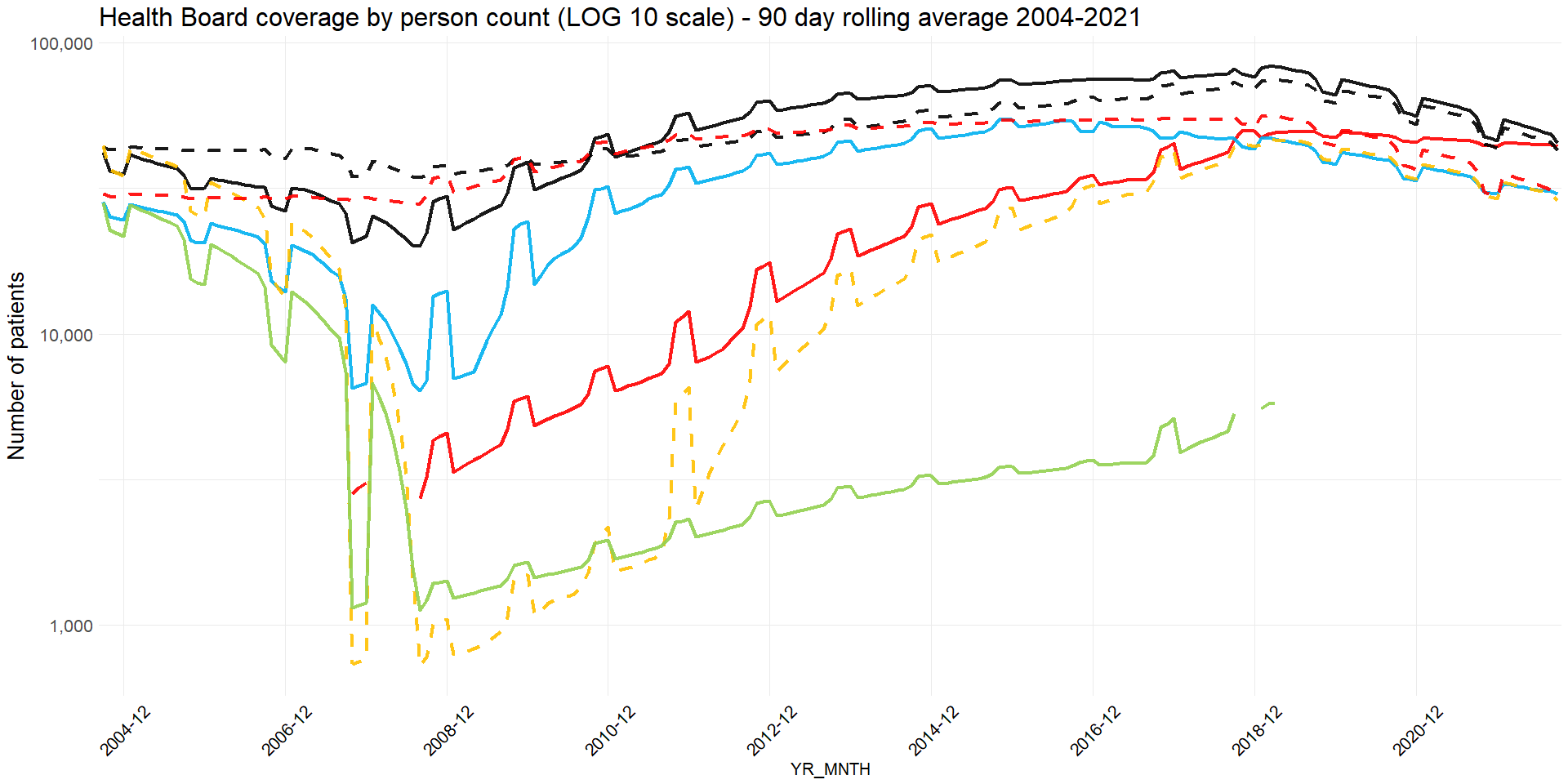
 

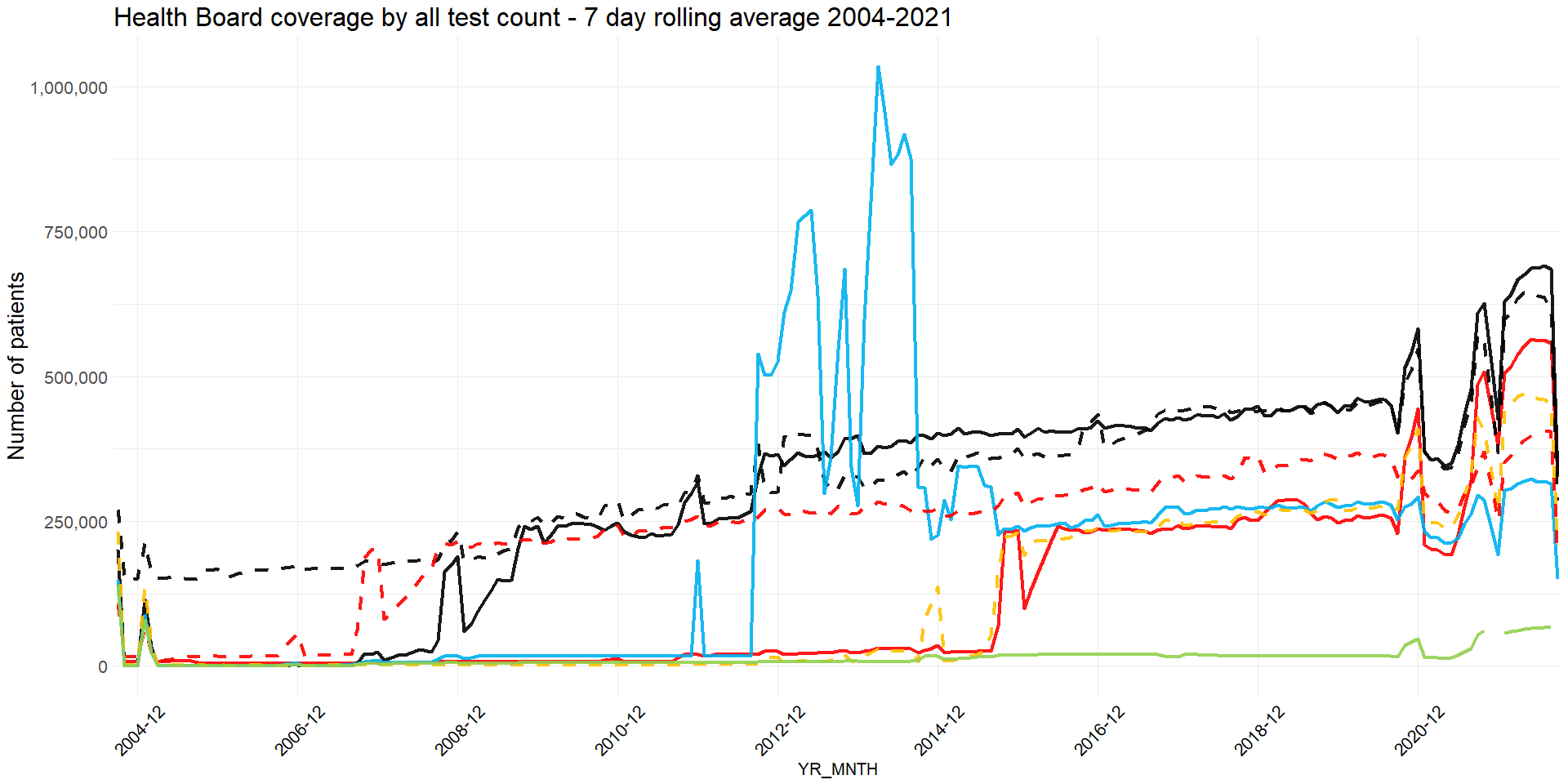
 

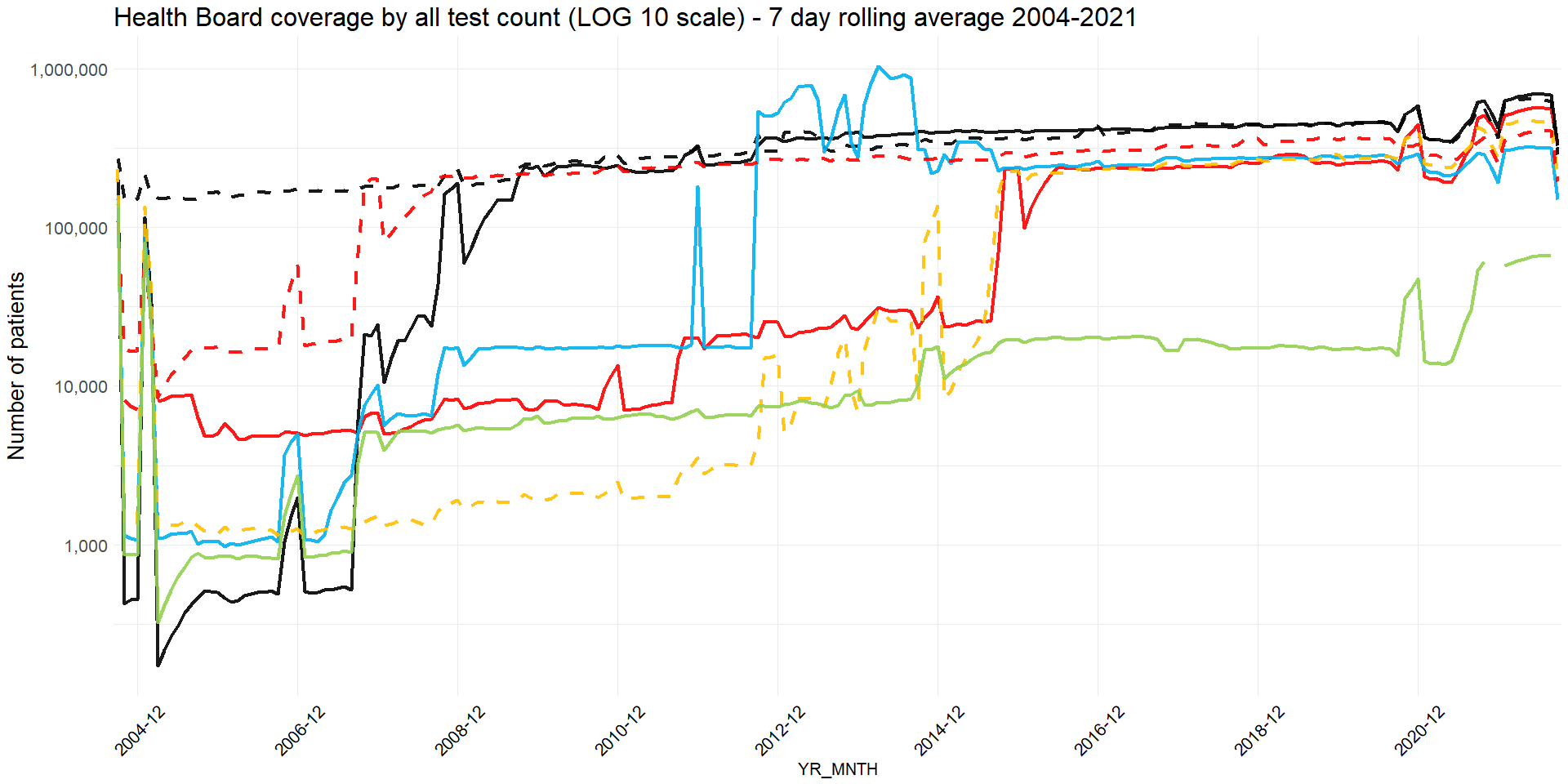
 

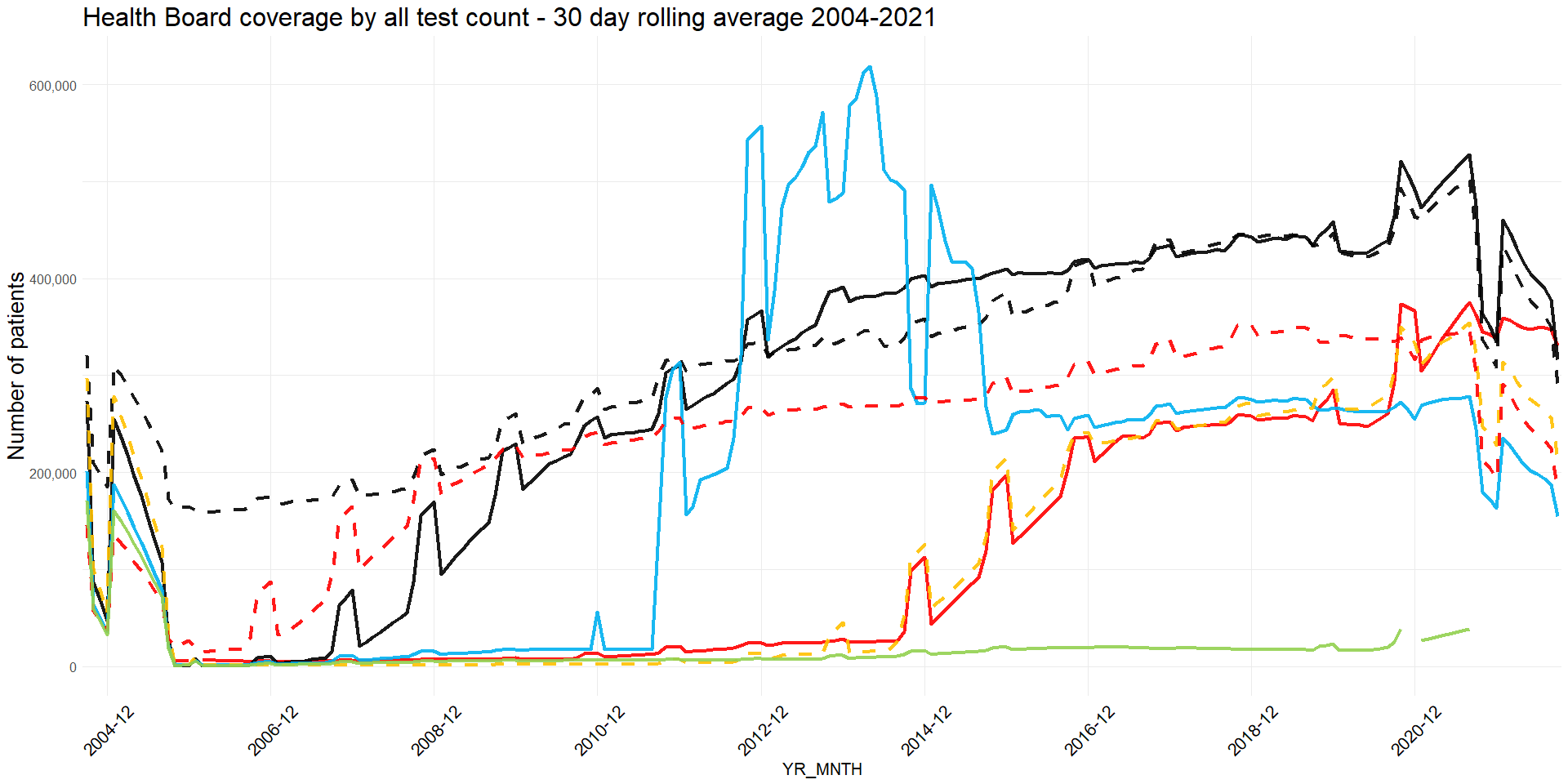
 

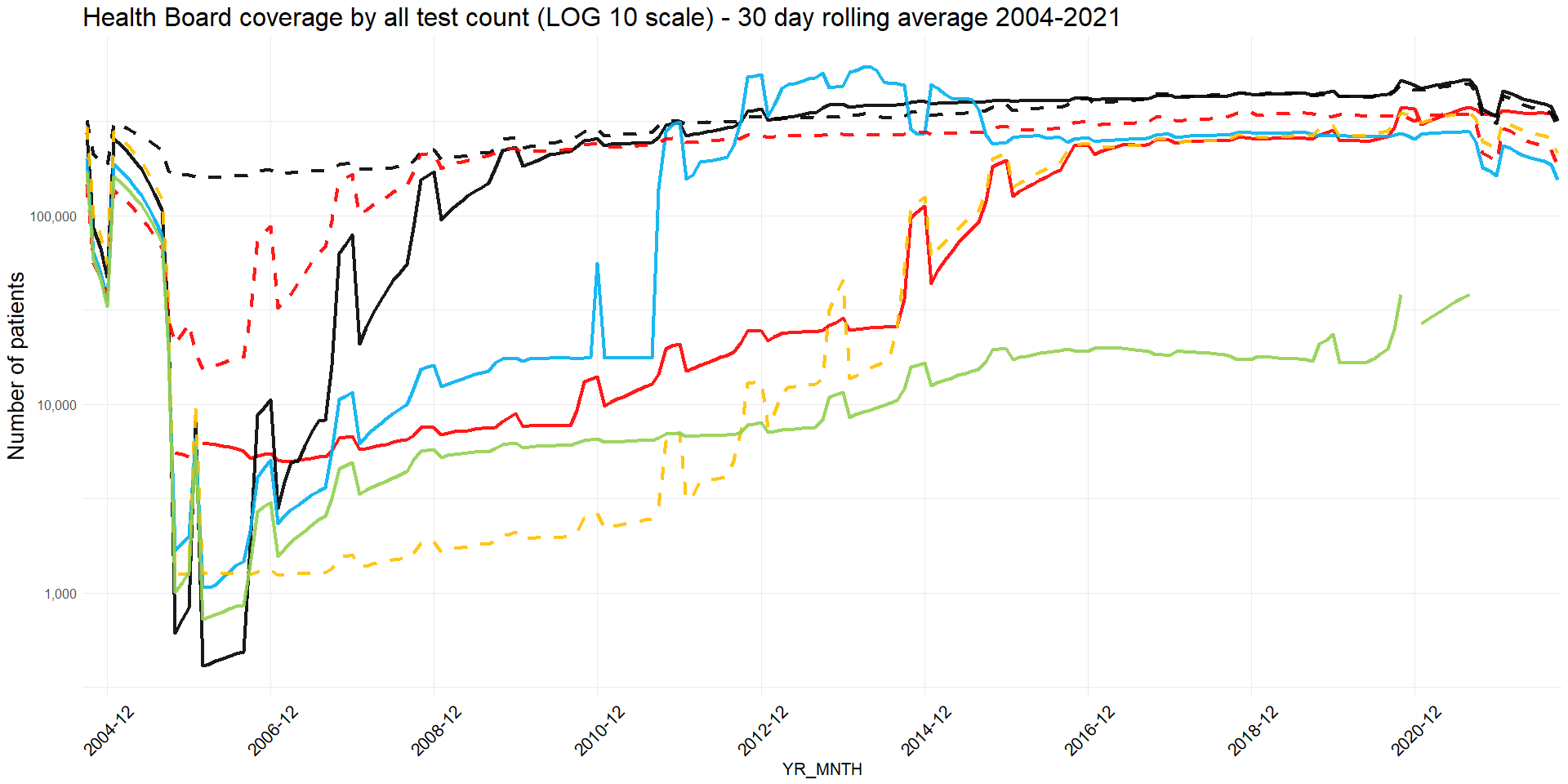
 

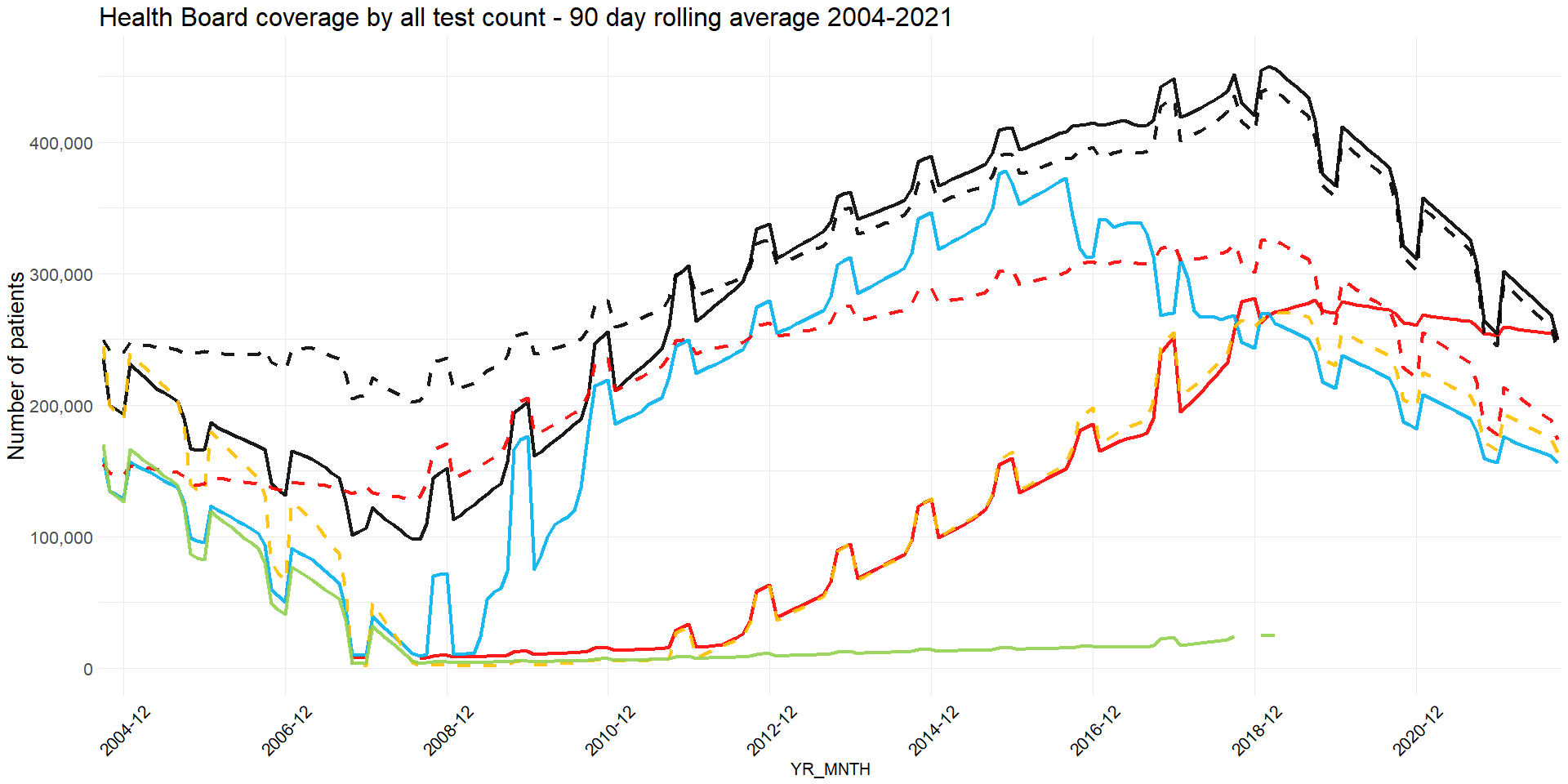
 

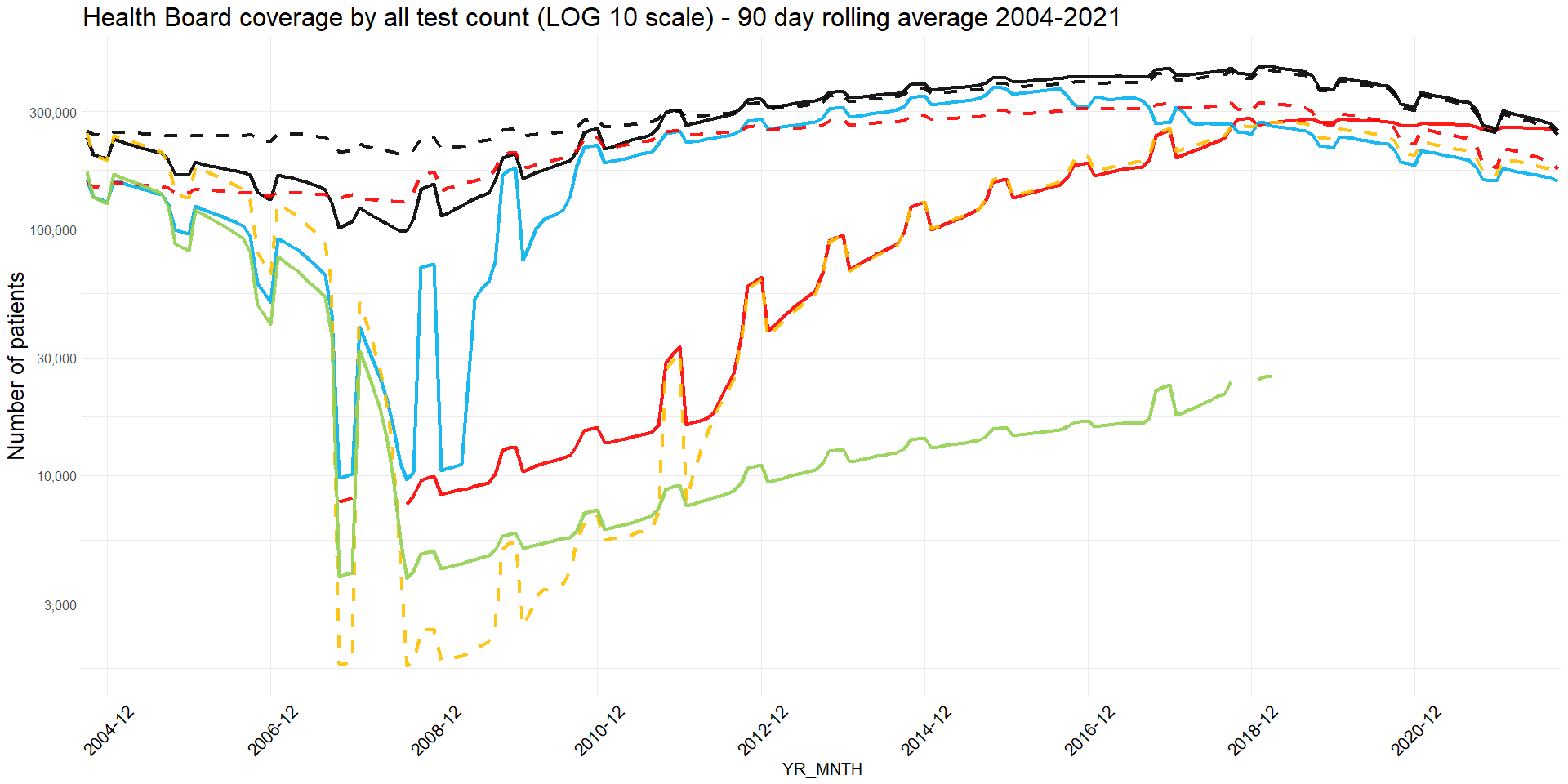
 

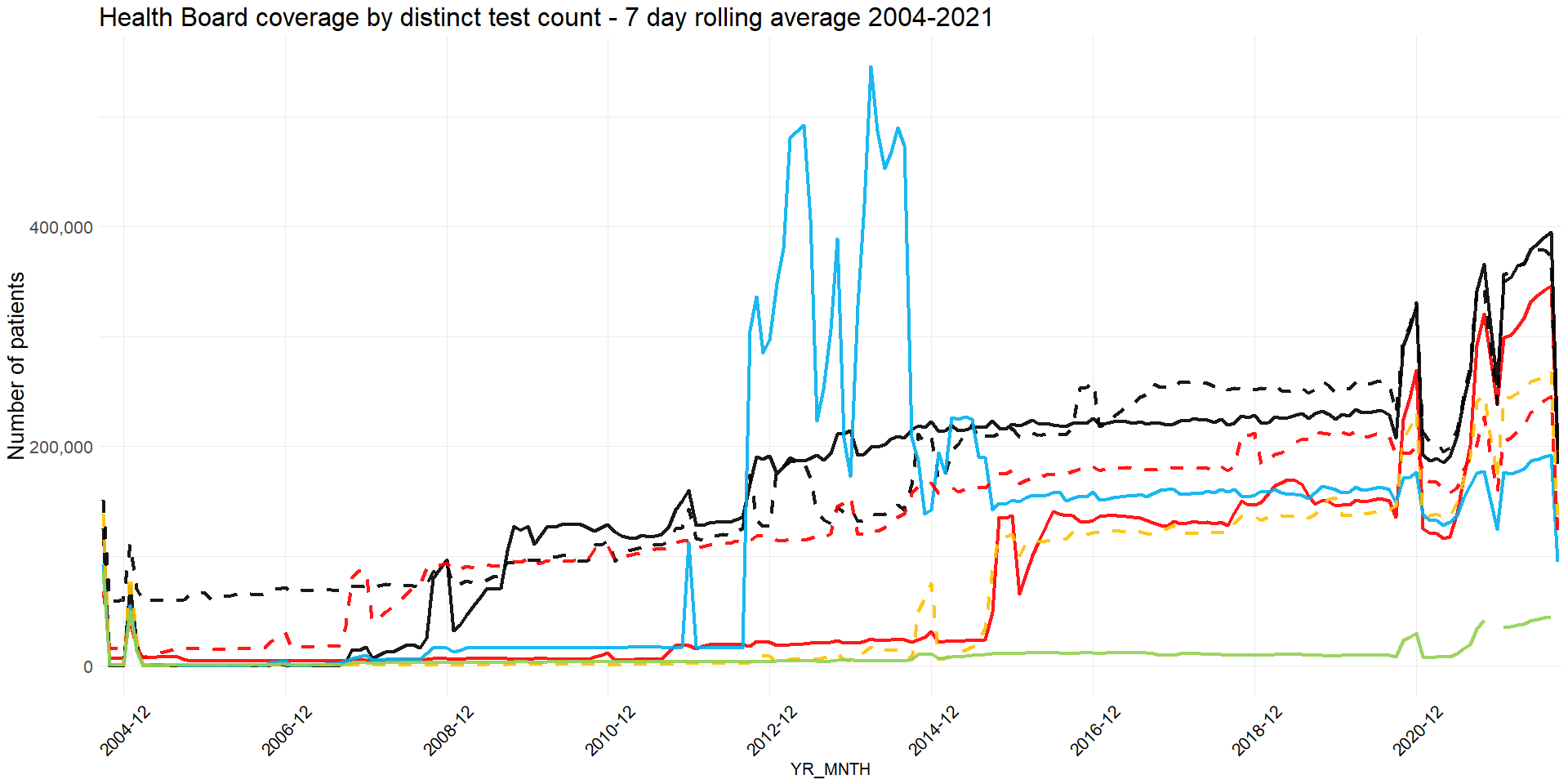
 

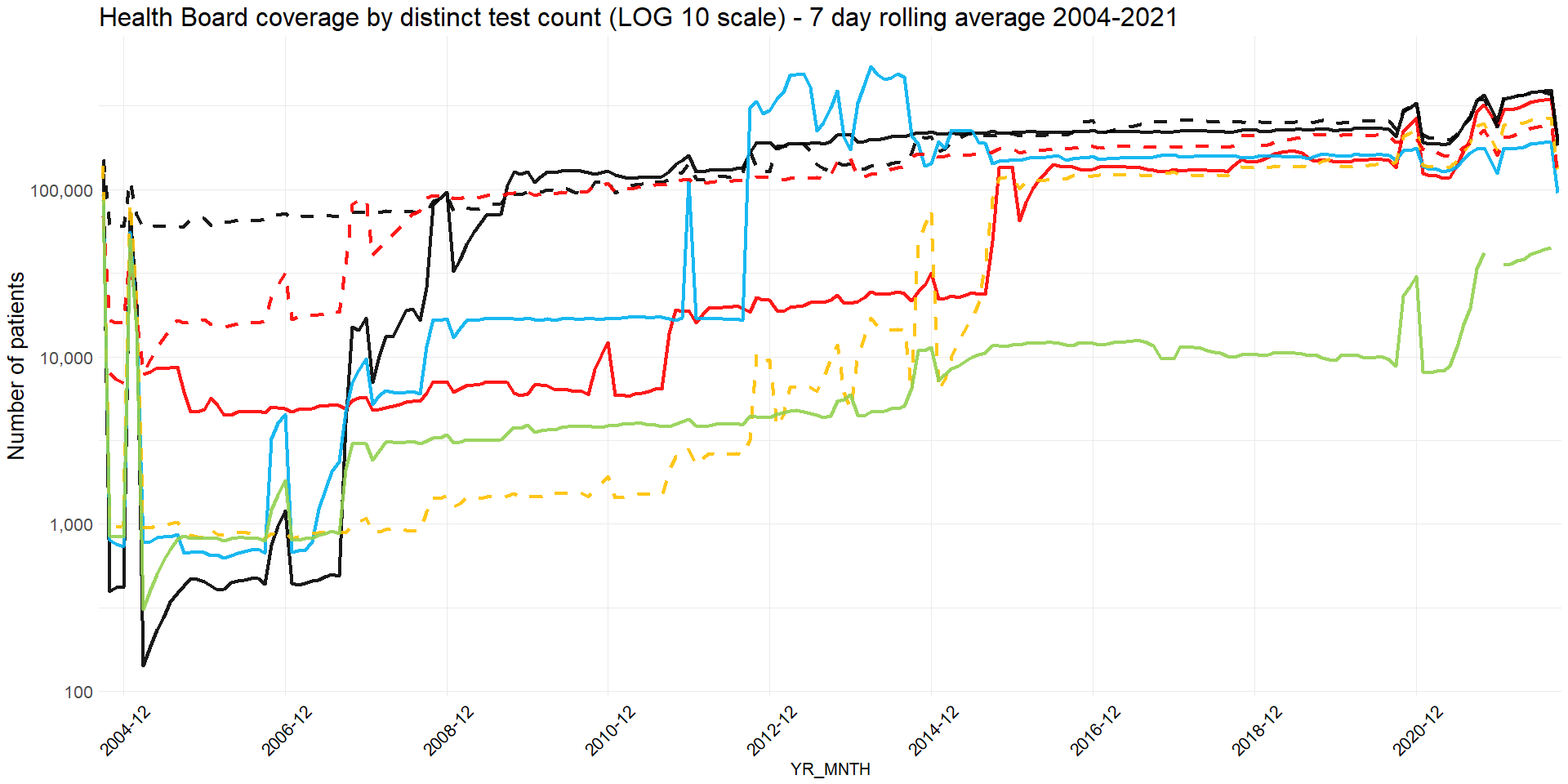
 

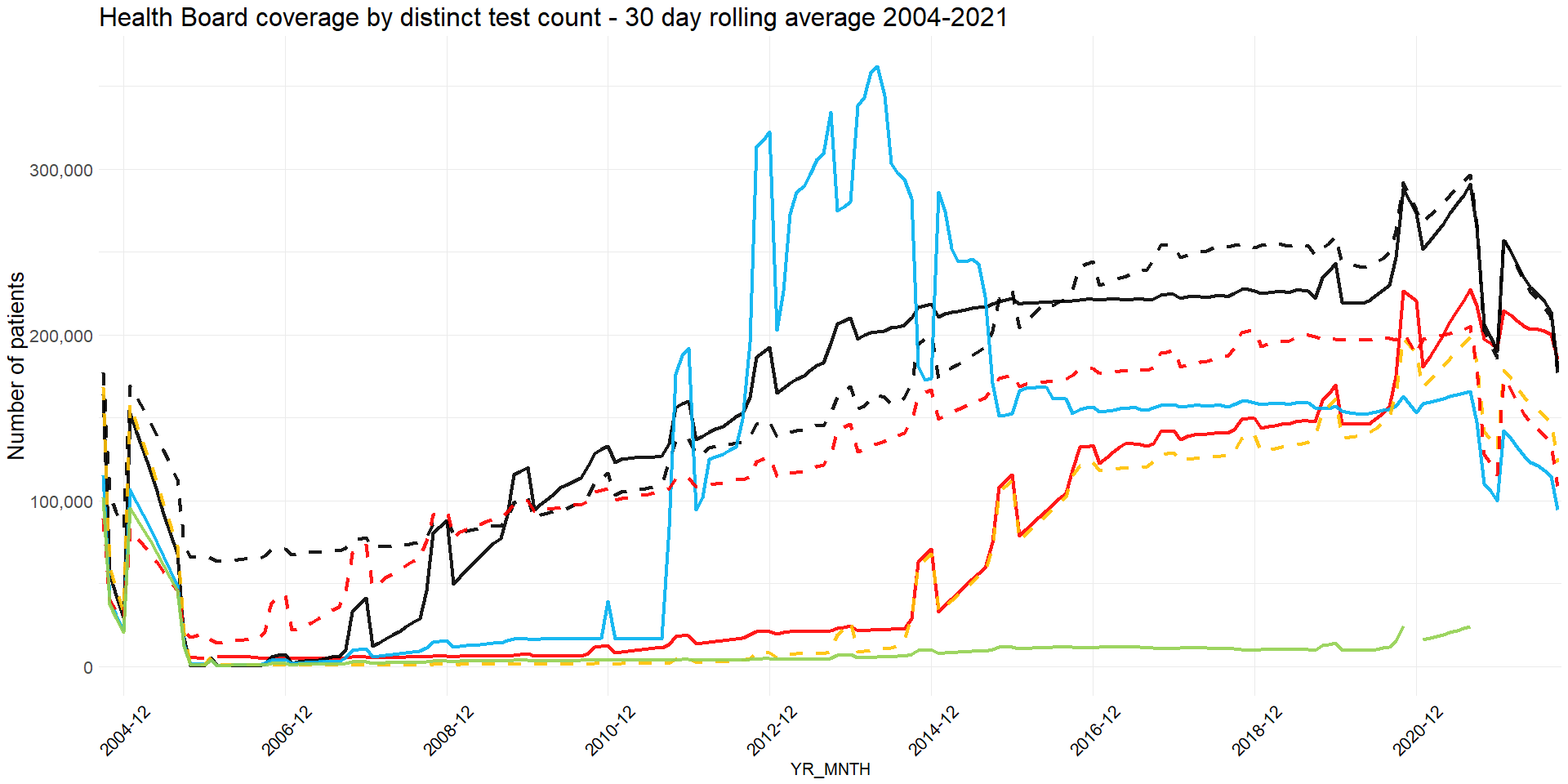
 

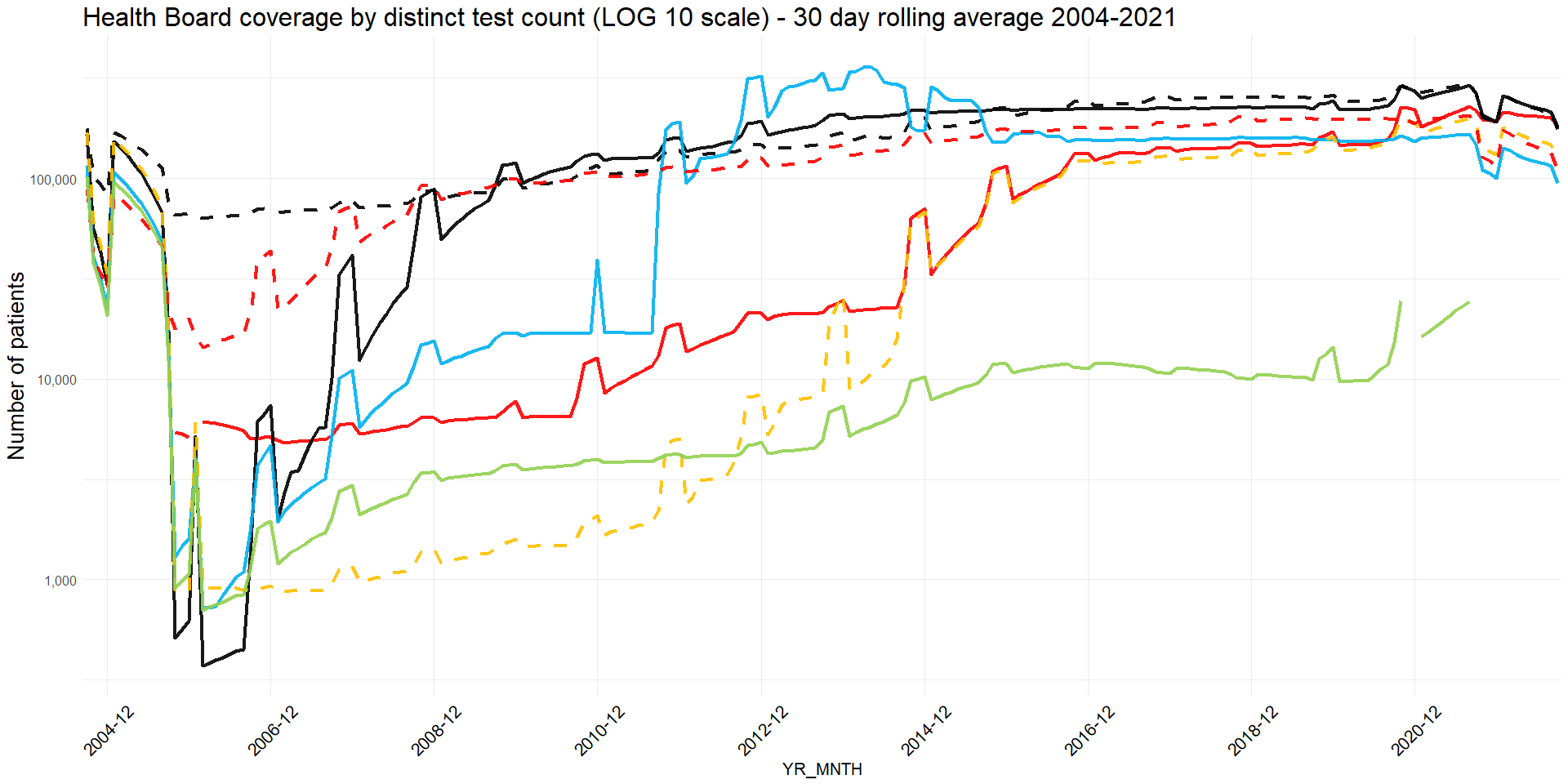
 

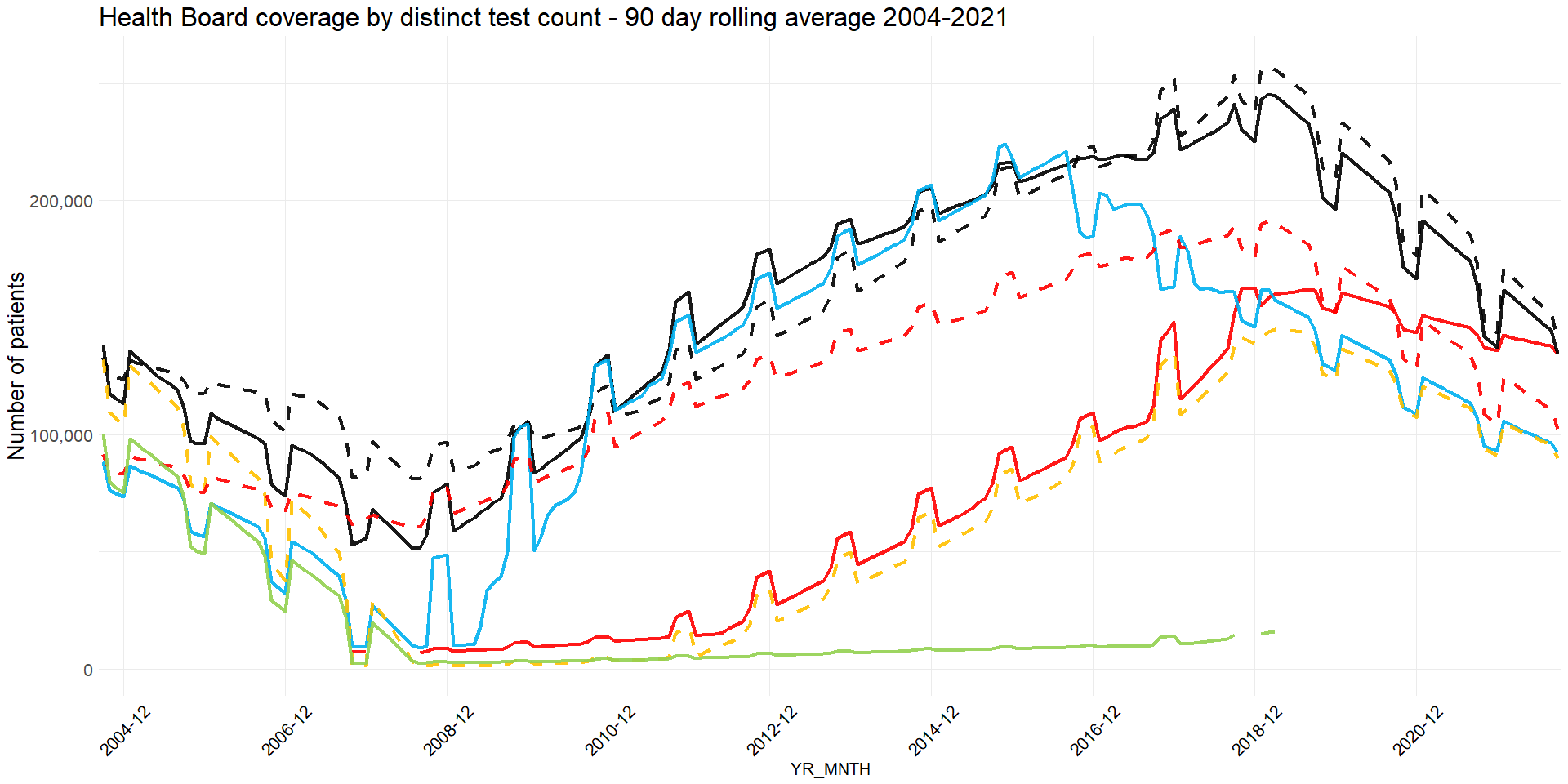
 

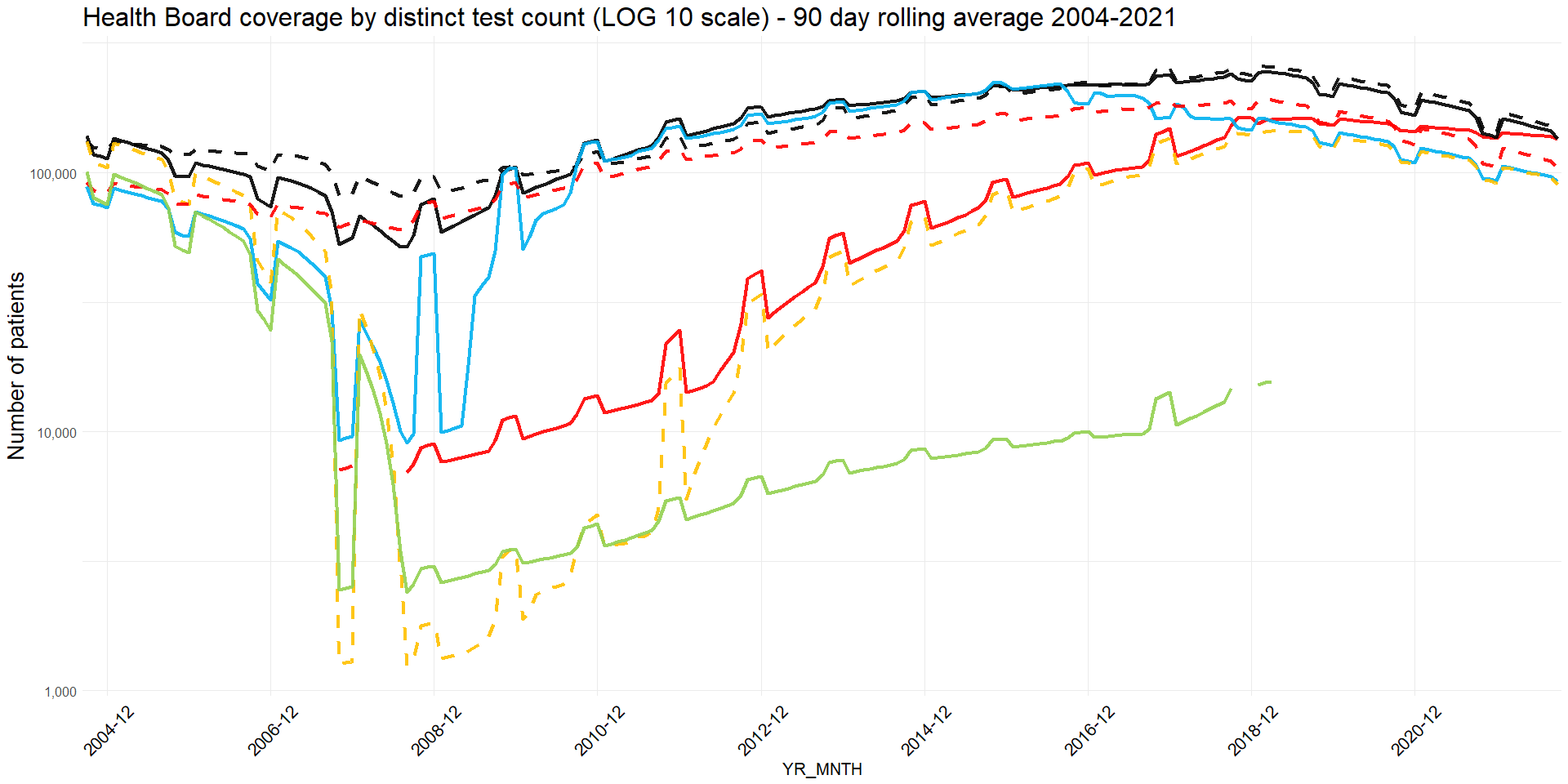
 

## Good coverage assessment

An SQL algorithm was developed to assess good data coverage periods. This was based on input parameters of rolling average timeperiod and count ratio to overall average. The algorithm starts at the latest date, and looks backwards. When two consecutive timeperiods have a count to average ratio that drops below the threshold, the coverage ends.

The overall timeperiod examined was 2004 to present day. Results for 7,30 and 90 day timeperiods; and ratio to average of 0.7, 0.5, 0.3 and 0.1 were explored for person counts and all (non distinct) test counts. The results from the algorithm were compared to a manual visual assessment to the log plots (which are easier to visually assess), and strong agreement was coloured green, mild agreement colored pale green. On this basis, an initial choice of parameters was ratio = 0.3, timeperiod = 30days.

The automated good coverage periods (using 30days, 0.3 ratio, strict method) for each health board are as follows:

Person count coverage:

ALL Health boards: start = 2008-02-14 end = 2021-12-31   
SB : start = 2012-05-23 end = 2021-12-31   
AB : start = 2004-02-05 end = 2021-12-31   
BC : start = 2008-04-14 end = 2021-12-31   
CV : start = 2014-07-12 end = 2021-12-31   
CT : start = 2012-03-24 end = 2021-12-31   
HD : start = 2007-01-20 end = 2021-12-31   
PT : start = 2007-01-20 end = 2021-12-31

Event date count coverage:

ALL Health boards: start = 2008-02-14 end = 2021-12-31   
SB : start = 2015-02-07 end = 2021-12-31   
AB : start = 2004-02-05 end = 2021-12-31   
BC : start = 2008-05-14 end = 2021-12-31   
CV : start = 2014-12-09 end = 2021-12-31   
CT : start = 2012-03-24 end = 2021-12-31   
HD : start = 2007-01-20 end = 2021-12-31   
PT : start = 2007-02-19 end = 2021-12-31

## 

## Anomalies or spikes in coverage

It is seen that there are a number of dramatic spikes in coverage between 2010-2015. The health board breakdown helps to reveal some further information on this, in that they seem to originate from Cardiff and Vale and Cwm Taf health boards. Further investigation is needed here, as to which codes are responsible, with a view to explaining these anomalies.

## Location fields

Locations are present in the fields PROV\_DEPT\_SITEID\_E, SUBJECTSITEID\_E, SUBJECT\_LOC\_ID\_E and STATS\_CURR\_CENSUS\_LSOA\_CD. These location IDs do not have associated descriptions, unlike the previous SAIL0911V.PATH\_WRRS\_PATH dataset. Apart from the LSOA field it is unclear without further input where these IDs map to. Currrently, it is only possible to map the STATS\_CURR\_CENSUS\_LSOA\_CD field. A lookup table for the other location descriptions is currently in progress, but was not completed at the time of running this report.

# Appendix

## Fields

The fields present in SAIL0911V.WRRS\_OBSERVATION\_REQUEST are:

| COLNAME | TYPENAME | LENGTH |
| --- | --- | --- |
| ALF\_E | INTEGER | 4 |
| ALF\_STS\_CD | INTEGER | 4 |
| ALF\_MTCH\_PCT | DOUBLE | 8 |
| SPCM\_COLLECTED\_DT | DATE | 4 |
| ID\_E | BIGINT | 8 |
| REPORTID\_E | BIGINT | 8 |
| REPORT\_SEQ | INTEGER | 4 |
| REQUEST\_SEQ | INTEGER | 4 |
| CD | VARCHAR | 50 |
| NAME | VARCHAR | 150 |
| PROV\_SYSTEM\_CD | VARCHAR | 50 |
| READ\_CD | VARCHAR | 5 |
| OBSERVATION\_STS\_CD | VARCHAR | 2 |
| SENSITIVITY\_STS\_CD | VARCHAR | 4 |
| AUTHORISEDBYID | INTEGER | 4 |
| REPORTDISPLAYORDER | VARCHAR | 3 |
| AUTHORISED\_DTTM | TIMESTAMP | 10 |
| AVAIL\_FROM\_DT | DATE | 4 |

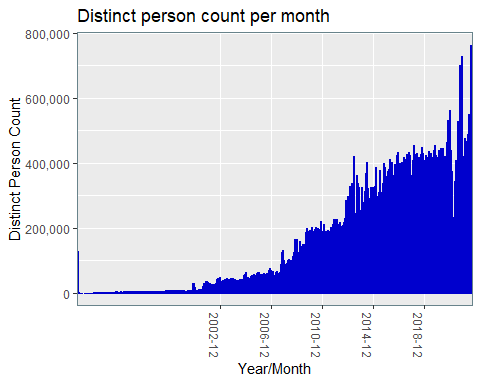
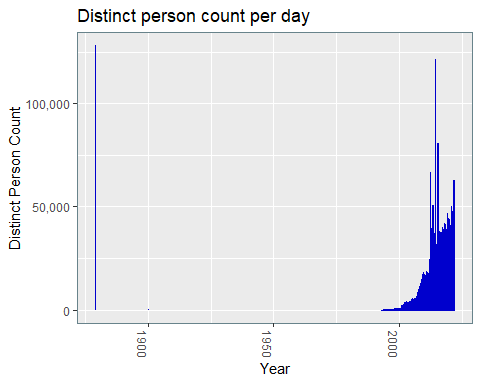
The fields present in SAIL0911V.WRRS\_OBSERVATION\_RESULT are:

| COLNAME | TYPENAME | LENGTH |
| --- | --- | --- |
| ALF\_E | INTEGER | 4 |
| ALF\_STS\_CD | INTEGER | 4 |
| ALF\_MTCH\_PCT | DOUBLE | 8 |
| SPCM\_COLLECTED\_DT | DATE | 4 |
| OBSERVATION\_REQST\_ID\_E | BIGINT | 8 |
| REQUEST\_SEQ | INTEGER | 4 |
| REPORTID\_E | BIGINT | 8 |
| REPORT\_SEQ | INTEGER | 4 |
| VAL\_TYPE | VARCHAR | 3 |
| CODE | VARCHAR | 30 |
| NAME | VARCHAR | 90 |
| PROV\_SYSTEM\_CD | VARCHAR | 7 |
| READ\_CD | VARCHAR | 9 |
| VAL | VARCHAR | 100 |
| UNITOFMEASUREMENT | VARCHAR | 25 |
| REFERENCERANGE | VARCHAR | 60 |
| ABNORMAL\_STS\_CD | VARCHAR | 30 |
| AVAIL\_FROM\_DT | DATE | 4 |

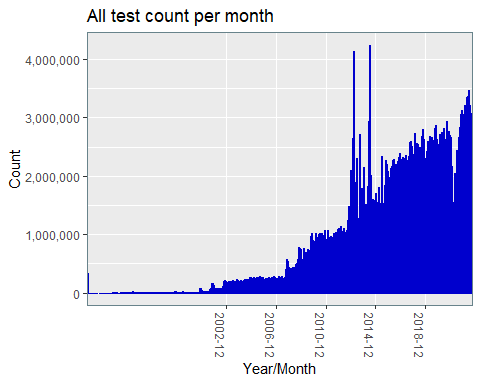
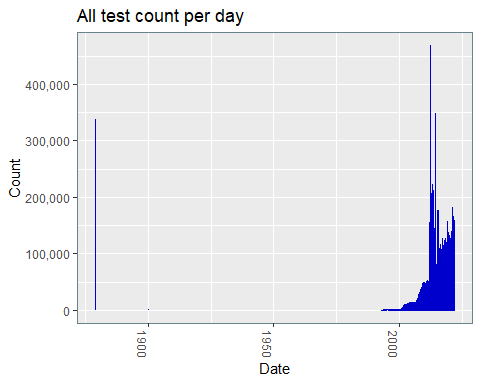
The fields present in SAIL0911V.WRRS\_REPORT are:

| COLNAME | TYPENAME | LENGTH |
| --- | --- | --- |
| ID\_E | BIGINT | 8 |
| REPORT\_SEQ | INTEGER | 4 |
| MASTERREPORTID\_E | BIGINT | 8 |
| REPORTVERSION | INTEGER | 4 |
| REPORTTITLE | VARCHAR | 200 |
| PROV\_TYPE\_CD | VARCHAR | 10 |
| REPORTSENSITIVITY\_STS\_CD | VARCHAR | 10 |
| REPORTSENSITIVE | VARCHAR | 6 |
| ABNORMALRESULTS | VARCHAR | 3 |
| PLACER\_REQST\_NUM | INTEGER | 4 |
| PATHWAY\_ID | INTEGER | 4 |
| PROV\_SYSTEMID\_E | BIGINT | 8 |
| PROV\_DEPT\_ID\_E | BIGINT | 8 |
| PROV\_DEPT\_SITEID | INTEGER | 4 |
| PROV\_DEPT\_ORG\_ID\_E | BIGINT | 8 |
| SUBJECT\_LOC\_ID | INTEGER | 4 |
| SUBJECTSITEID | INTEGER | 4 |
| SUBJECT\_ORG\_ID\_E | BIGINT | 8 |
| REQUESTORID\_E | BIGINT | 8 |
| REQUESTOR\_SPEC\_ID\_E | BIGINT | 8 |
| PLACERID\_E | BIGINT | 8 |
| PAT\_TYPE\_ID | INTEGER | 4 |
| PAT\_CAT\_ID | INTEGER | 4 |
| REPORT\_DTTM | TIMESTAMP | 10 |
| SPCM\_COLLECTED\_DT | DATE | 4 |
| SPCM\_COLLECTED\_TM | VARCHAR | 19 |
| SPCM\_RECEIVED\_DT | DATE | 4 |
| SPCM\_RECEIVED\_TM | VARCHAR | 10 |
| UPDATED\_DTTM | TIMESTAMP | 10 |
| ISARCHIVED | INTEGER | 4 |
| TRANSFORMATIONPROCESSID | INTEGER | 4 |
| ISBINARY | VARCHAR | 6 |
| ALF\_E | INTEGER | 4 |
| ALF\_STS\_CD | INTEGER | 4 |
| ALF\_MTCH\_PCT | DOUBLE | 8 |
| STATS\_CURR\_CENSUS\_LSOA\_CD | VARCHAR | 9 |
| AVAIL\_FROM\_DT | DATE | 4 |

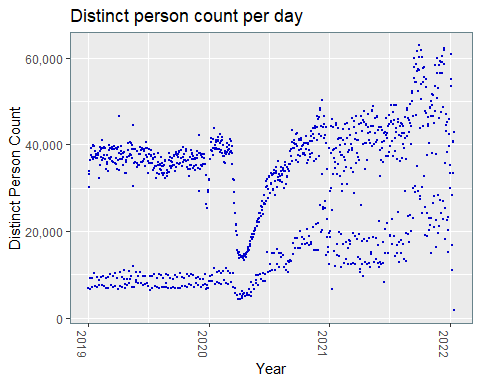
## Person count coverage unfiltered



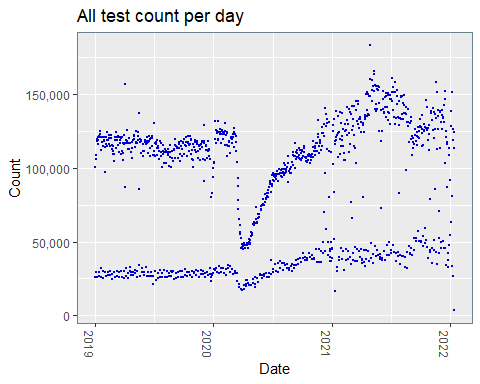
## All test count coverage unfiltered



## Person count coverage post 2019



## All test count coverage post 2019



### Default (false) date entries

The >‘1900’ filter is applied as some dates are set to a defaults of 1900-01-01 and 1879-01-01.

The number of dates set to 1900-01-01 or 1879-01-01 in SAIL0911V.WRRS\_OBSERVATION\_REQUEST.AUTHORISED\_DTTM = 338088.

The number of dates set to 1900-01-01 or 1879-01-01 in SAIL0911V.WRRS\_REPORT.REPORT\_DTTM = 95.

The number of dates set to 1900-01-01 or 1879-01-01 in SAIL0911V.WRRS\_REPORT.UPDATED\_DTTM = 0.

### Location lookup

LSOAs were filtered into their respective health boards as defined by the LOCAL\_HEALTH\_BOARD\_NAME\_ENGLISH field in the reference table SAILW0911V.JH\_REGION\_LOOKUP.

Note that some dates might pre-date health board name changes, and dates correspond to LSOAs as categorized at the time SAILW0911V.JH\_REGION\_LOOKUP was created (2020-04-09). Abertawe Bro Morgannwg Health Board in SAILW0911V.JH\_REGION\_LOOKUP has been renamed Swansea Bay University Health Board in this report, to bring the health board names up to date. Boundary changes have not been adjusted however, and categorisation into health boards is as in SAILW0911V.JH\_REGION\_LOOKUP.