**List of questions about the Swiss MMG data**

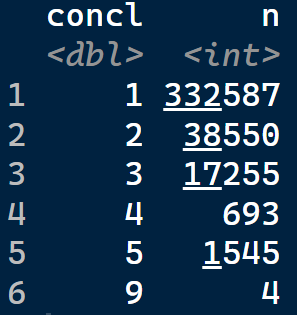
Diagram, text

Description automatically generated with medium confidence

1) A total of 632+756=1388 women with breast cancer (endpoint=1) have a detection mode that is either **unknown**(detec=5) or "**other**" (detec=3). What is the difference between unknown and other in this context?

2) Ideally each woman with breast cancer should either be clinical (detec=0) or screen detected (detec=1,2,4). It is a key variable for overdiagnosis estimation, so if there is no direct assignment, we must either impute the mode of detection, or remove the participant from the analysis – both can of course lead to serious bias. Can you think of any way to gain more information on the detec field from the cancer registry?

3) Among the 4545 cancers (adding up column 'n' in the above table), 906 (20%) are coded as clinical (detec=0). Since we don't expect the true ratio of clinical cancers among all cancers to be any higher than that (in the US it's about 10%), we thus expect that the vast majority of the 1388 women who are coded as "unknown" (detec=5) or "other" (detec=3) are in fact screen-detected. Do you have reason to believe that this is true?

4) In Swiss practice, is it common for women with final BI-RADS finding of **concl**=3 (probably benign finding) to get a biopsy? We are asking because there are (i) 4545 cancers (of which only 906 are confirmed clinical), yet (ii) there are only 693+1545=2238 screens with BI-RADS of 4 (Suspiciously abnormal finding) or 5 (Probably malignant finding). It is possible that many of the concl=3 screens led to screen-detected cancers, but we would like to make sure this fits the clinical practice patterns in Switzerland.

5. We found that a non-trivial number of clinical cancers were diagnosed shortly after an in-program screen (we can prepare some statistics for Thursday). It would be very helpful to better understand the registrar coding rules used to determine whether a cancer was clinical vs screen-detected.

6) We found some apparent inconsistencies when comparing the variables BI-RADS (concl) and "reason for (not) attending" (nopartreas): shouldn't BI-RADS be missing (concl=9) when the participant did not attend the screen (nopartreas=1,2,3)? See in particular row 6 in the following table:  
A picture containing table

Description automatically generated

7) We found several instances where a cancer was coded as “screen-detected in program" (detect=1), yet in the actual screening history there was no screen that could be matched to the event. Below is an example of a woman (nodoss = 1003347125) with three screens, two of which were attended (nopartreas=9), but none seems to have triggered the screen detection on 2016-12-19.

Text

Description automatically generatedWe are trying to understand whether this is "noise" in the data (to be expected in any large registry, of course), or whether there is a plausible explanation for such patterns.

8) Some screens (examdt) take place *after* the endpointdt. Is this normal?

Text

Description automatically generated9) We found specific dates with a large number of censor dates (endpoint=5)? For instance, 1612 women were censored on May 14, 2018 (see picture below). Is this normal?