Data-driven prioritisation for optimise oncological screening impact.

08 December, 2024

Contents

Reflective writing

As the final assignment of this introductory course on Data Science for Health and Social Care, we were asked to compose a data-driven narrative around data from the OECD Health Statistics Repository, produce visualisations, and undertake exploratory analyses in R. I took this assignment as an opportunity to dig further into aspects related to my career, as a biologist researcher in Oncology.

I focused on analysing the relationship between cancer screening and patient survival at the international level and with a focus on Scotland. For the presentation, I attempted to portray it as a short introductory talk in a fictitious Research and Innovation Strategy Symposium at the University of Edinburgh. Such a Symposium would have been open to the University's community; hence, the presentation was intended for a broader audience aiming to enact an open discussion. The report, although it touches on the same topics, is more detailed in terms of the information delivered and a more explicit set of recommendations that would contribute to the fictitious Research and Innovation Strategy. Considering these two scenarios, the presentation and report, and their respective target audiences, I adjusted the information details delivered and the visuals. The course project itself was challenging due to the constraints of working with a specific dataset. However, it forced me to think more clearly on a data-driven project rather than a hypothesis-based approach, as I am used to. There are a few other elements I would want to highlight as a product of the course and the course's project.

Firstly, the analysis of this data enabled me to identify the temporal trends in screening across several countries and with more detail in Scotland. Analysing these trends together with survival (or mortality) trends allowed me to identify that although the trends seem to be weaker than expected, there is a positive relationship between screening and survival. A detailed analysis of Scotland allowed me to detect that Health Boards with high screening and high mortality could be related to specific genetic variants, according to research published last year, which results tremendously interesting for me because it allows me to link data-driven epidemiological patterns with mechanistic biological processes.

Secondly, although this was an exploratory analysis, I realised how complex it could be to intersect data analysis, visuals, and narrative when building data-driven storytelling. During the course, we learnt about different ways of data wrangling that enabled the development of this assignment. Additionally, some of the perspectives regarding a system approach to develop interventions in complex systems like healthcare are very useful when thinking about specific actions and how to optimise the use of resources and impacts.

Thirdly, reflecting on the whole process, I internalised the importance of data quality and access, wrangling and completion. For example, data from the OECD, although containing information from several countries, the temporal sequences are incomplete, and it is difficult to have a more integrated perspective. Luckily, countries like Scotland have invested in creating open repositories for their data that enable more local analysis that would help in the prioritisation of different components around a disease, from research to deployment and integration with other information systems.

Finally, putting it in perspective, with this assignment, I dove into an introduction to the complexities and the potential extent of the impact of a data-driven approach, what are the critical aspects of it and how to drive storytelling around it.

Background on early detection of cancer

Early detection of cancer increases the likelihood of survival Hankey et al. (1999). Therefore, the deployment of screening programs is crucial for successful patient care. Screening is a vital component of early detection (Smith et al. 2002), and many countries have extensive programs targeting about 80 of their population susceptible to having breast, cervical, colon or rectum cancers. These active surveillance programs are not widely adopted; hence, a variety of outcomes are expected to directly impact patient care. Here, it is presented the need to incorporating cancer screening as a priority into the Research and Innovation Strategy. the need is highlighted by analysing international data collected by the Organization for Economic Cooperation and Development (OECD) is analysed for the 21st century, including screening and survival rates for cervix and breast cancer. Furthermore, a focus on Scotland highlights the country's challenges and opportunities in the deployment of screening for improving patient care, integrating Data Science in the Health and Social care ecosystem to improve oncological patient care.

Global context of oncological screening of breast and cervical cancer

Data wrangling

```
overall_screening <- read.csv("data/0_screening_oecd_24112024.csv")
overall_inc_mor <- read.csv("data/1_survival_oecd_24112024.csv")

df_plt_screening <- overall_screening %>%
  filter(DATA_SRCE == "PRG",
         # using <2015 to match the availability of OECD cancer survival data
         TIME_PERIOD < 2015) %>%
  mutate(year = as.numeric(TIME_PERIOD),
         country = factor(Reference.area),
         value = OBS_VALUE) %>%
  group_by(Cancer.site, country) %>%
  # keep countries with > 12 records for the period 2000-2014
  filter(length(country) >= 12) %>%
  ungroup() %>%
  #calculate slide window with 5 years lag
  mutate(value_5y = slider::slide_dbl(value, mean, .before = 4, .after = 0),
         name_lab = if_else(year == 2014, country, NA_character_)) %>%
  #filter by years matching the cancer survival records
  filter(year %in% c(2004, 2009, 2014))

#generate a named color vector for all countries
colour_vec_all <- met.brewer("VanGogh2", length(unique(df_plt_screening$country)))
names(colour_vec_all)<- unique(df_plt_screening$country)


mortality_breast <- overall_inc_mor %>%
  filter(Measure == 'Breast cancer five-year net survival',
         Statistical.operation == 'Observed') %>%
  mutate(year = TIME_PERIOD,
         country = Reference.area,
         year_country = interaction(year, country, sep = '-'),
         mortality = OBS_VALUE) %>%
  select(year, country, mortality, year_country)
```

```
oecd_trends <- overall_screening %>%
  filter(DATA_SRCE == "PRG",
```

```
        Sex == 'Female',
        Reference.area %in% df_plt_screening$country,
        str_detect(Cancer.site, "cervix|breast"),
        TIME_PERIOD < 2015) %>%
  mutate(year = as.numeric(TIME_PERIOD),
        country = factor(Reference.area),
        value = OBS_VALUE,
        value_70 = case_when(value > 70 & year == 2005 ~"2000 above 70%",
                             value < 70 & year == 2005 ~"2000 below 70%")) %>%
  group_by(Cancer.site, country) %>%
  mutate( value_70_all = first(na.omit(value_70))) %>%
  ungroup() %>%
  mutate(value_70_all = ifelse(is.na(value_70_all), "no data 2000", value_70_all)) %>%
  ggplot(aes(year, value)) +
  geom_line(aes(group = country, color = country), alpha =0.6) +
  geom_smooth(aes(group = 1), method = "lm",
              size = 0.5, se = FALSE,
              linetype = 2, color = "black") +
  facet_grid(value_70_all~Cancer.site) +
  scale_colour_manual(values = colour_vec_all) +
  labs(title = '',
       x = 'Year',
       y = 'Screening (% target population)',
       colour = "Country") +
  theme_minimal() +
  theme(axis.title =  element_text(size = 10, hjust = 0.5,face = "bold"),
        legend.position = 'bottom',
        legend.title = element_text(size = 7, angle = 90, hjust = 0.5,face = "bold"),
        legend.text = element_text(size=6),
        legend.key.height = unit(0.15, 'cm'),
        legend.key.width = unit(0.5, 'cm'),
        strip.text = element_text(size = 8, hjust = 0,face = "bold"))
```

```
cancer_site = 'Malignant neoplasms of female breast'

c_names <- df_plt_screening %>%
  filter(Cancer.site == cancer_site) %>%
  pull(country) %>%
  unique() %>%
  as.character()

colour_vec = colour_vec_all[c_names]


screening_breast <- df_plt_screening %>%
  filter(Cancer.site == cancer_site) %>%
  select(year, country, value_5y) %>%
  # create a composed year_country to match with mortality records
  mutate(year_country = interaction(year, country, sep = '-'))
```

```
breast_join <- left_join(screening_breast,
                         mortality_breast,
                         by = 'year_country')

#set color vector for the countries represented
colour_vec <- colour_vec_all[as.character(unique(pull(breast_join,'country.x')))]
```

```r
oecd_densityplot <- screening_breast %>%
  mutate(year = case_when(year == 2004 ~"1999-2004",
                          year == 2009 ~ "2005-2009",
                          year == 2014 ~ "2010-2014"),
         year = factor(year)) %>%
  ggplot(aes(x = value_5y, color = year, fill = year)) +
  geom_density(alpha = 0.6) +
  geom_segment(data = screening_breast %>%
                 mutate(year = case_when(year == 2004 ~"1999-2004",
                                         year == 2009 ~ "2005-2009",
                                         year == 2014 ~ "2010-2014"),
                        year = factor(year)) %>%
                 group_by(year) %>%
                 summarise(median_screening = median(value_5y)) %>%
                 ungroup(),
               aes(x = median_screening,
                   xend = median_screening,
                   y = 0,
                   yend = Inf),
               colour= 'grey30')+
  geom_text(data = screening_breast %>%
              mutate(year = case_when(year == 2004 ~"1999-2004",
                                      year == 2009 ~ "2005-2009",
                                      year == 2014 ~ "2010-2014"),
                     year = factor(year)) %>%
              group_by(year) %>%
              summarise(median_screening = median(value_5y)) %>%
              ungroup(),
            aes(x = 30,
                y = 0.03,
                label = paste0('Median = ',median_screening, "%")),
            fontface = 'bold',
            size = 6,
            size.unit = "pt") +
  scale_fill_viridis_d(direction = -1, end = 0.9) +
  scale_color_viridis_d(direction = -1, end = 0.9) +
  theme(legend.position = "none") +
  labs(x = 'Screening (% target population)',
       y = 'Density') +
  guides(fill = 'none', colour= 'none') +
  facet_wrap(.~year) +
  theme_minimal() +
  theme(axis.title =  element_text(size = 10, hjust = 0.5, face = "bold"),
        legend.position = 'bottom',
        legend.title = element_text(size = 8, angle = 90, hjust = 0.5, face = "bold"),
        legend.text = element_text(size = 6),
        legend.key.height = unit(0.15, 'cm'),
        legend.key.width = unit(0.5, 'cm'),
        strip.text = element_text(size = 8, hjust = 0,face = "bold"))
```

```r
oecd_breast_mortalityscreening <- breast_join %>%
  mutate(year = as.numeric(year.x)) %>%
  rename(Country = country.x,
         Screening = value_5y,
         Survival = mortality, ) %>%
  mutate(name_lab = if_else(year == 2009, Country, NA_character_)) %>%
  ggplot(aes(x= Screening, y = Survival)) +
```

```
  geom_path(aes(group = Country, colour = Country),
            linewidth = 0.8,
            linejoin = 'round')+
  geom_text_repel(aes(color = Country, label = name_lab),
                  fontface = "bold",
                  size = 3,
                  force = 1,
                  direction = "y",
                  xlim = c(93, NA),
                  point.padding=unit(1,'lines'),
                  segment.alpha = 0,
                  hjust = 0,
                  box.padding = .45,
                  segment.curvature = 0)+
  geom_point(size = 0.5, shape = 16) +
  geom_text_repel(aes(label = year),
    size = 2,
    direction = "both",
    force_pull = 10,
    segment.size = 0.5,
    segment.alpha = 1,
    segment.linetype = "dotted",
    box.padding = 0.2,
    segment.curvature = -0.11,
    segment.ncp = 3,
    segment.angle = 20) +
  scale_colour_manual(values = colour_vec) +
  scale_x_continuous(limits = c(30,100),
                     breaks = seq(30,100, 10),
                     expand = expansion(mult=c(0.05,.12)))+
  scale_y_continuous(limits = c(65,95),
                     breaks = seq(65,95, 15))+
  labs(x = 'Screening (% target population)',
       y ='Net Survival (% of population with disease)') +
  guides(colour = 'none') +
  geom_text(data = data.frame(x = -Inf, y = Inf, label = 'D'),
            aes_string(x = 'x', y = 'y', label = "label"),
            hjust = 0,
            vjust = 1.5,
            fontface = 2,
            inherit.aes = FALSE) +
    theme_minimal() +
  theme(axis.title =  element_text(size = 10, hjust = 0.5,face = "bold"),
        legend.position = 'bottom',
        legend.title = element_text(size = 8, angle = 90, hjust = 0.5,face = "bold"),
        legend.text = element_text(size = 6),
        legend.key.height = unit(0.15, 'cm'),
        legend.key.width = unit(0.5, 'cm'))
```

Analysis of cancer screening globally

Country-based screening data for malignant neoplasms of cervix uteri, female breast, colon, rectum, and anus was retrieved from the OECD Healthcare Utilisation data set (OECD 2024b). Survival data for breast, cervical, colon and rectal cancers was retrieved from the OECD Healthcare Quality and Outcomes data set (OECD 2024a) providing indicators for 5-year net survival per country. Survival is available as aggregated measures for non-overlapping intervals of five years between 2000 and 2014. Screening data is available for the range 2000-2023

at yearly intervals, therefore, to match the survival data format and assess the association with survival data, screening data was summarised for 2004, 2009, and 2014, averaged for the previous 5 years inclusive. Screening data was filtered considering countries with screening data available between 2000 and 2014. For the time window analysed, one could split the countries into different groups based on a 70 cutoff screening rate at the beginning of the study period (year 2000). The first group, > 70, shows a steady screening rate, without major changes (e.g., Sweden, UK, New Zealand). The second group, with screening below 70 at 200, shows an increase in their efforts of screening over time for breast and cervical cancers. A third group of countries without data reported for 2000 shows a modest trend to increase screening of cervical cancers. These findings have been adapted for this report including higher granularity in the temporal dynamics.

```
tag_facet2(oecd_trends)
```

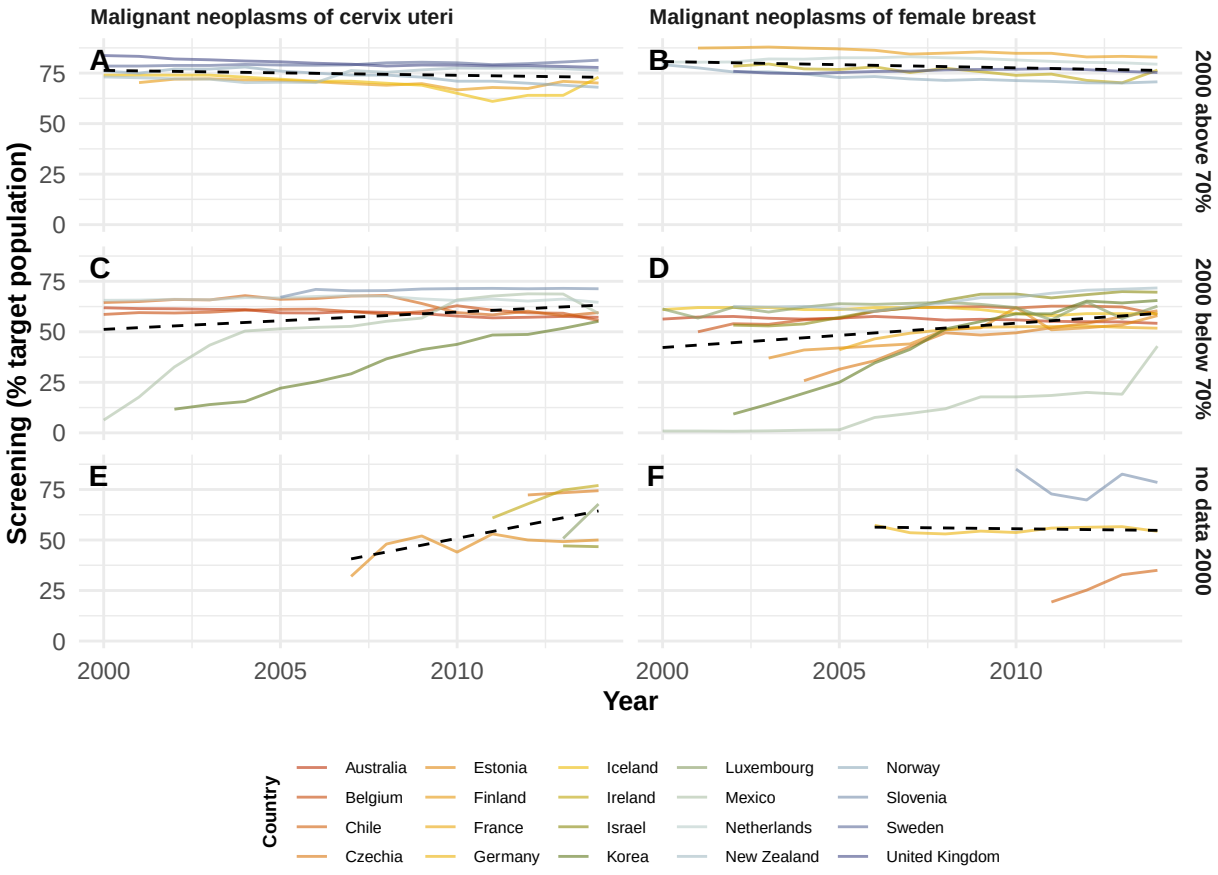

Figure 1: Trends in cancer screening worldwide. Temporal dynamics in the screening rate of cervical and breast malignant neplasms per country. Countries are straified according the data availability by 2000, if their screening rate was above 70% (A-B), below 70% (C-D), or data was not available (E-F). Segmented line represent a linear regression between time and screening percentage for each strata and neoplasm type. Data source: OECD.

For the countries that have enough records during the 2000-2014 period, there is a modest increase in the breast cancer screening rate globally. For some countries below the 70 cutoff for the range 2000-2004, such as Korea or Estonia, there is an overall gain of 5-years net survival over time, such pattern is not present in the countries above 70 for the range 2000-2004, such as the United Kingdom and Finland.

```
tag_facet2(oecd_densityplot)/oecd_breast_mortalityscreening + plot_layout(heights = c(1,4))
```

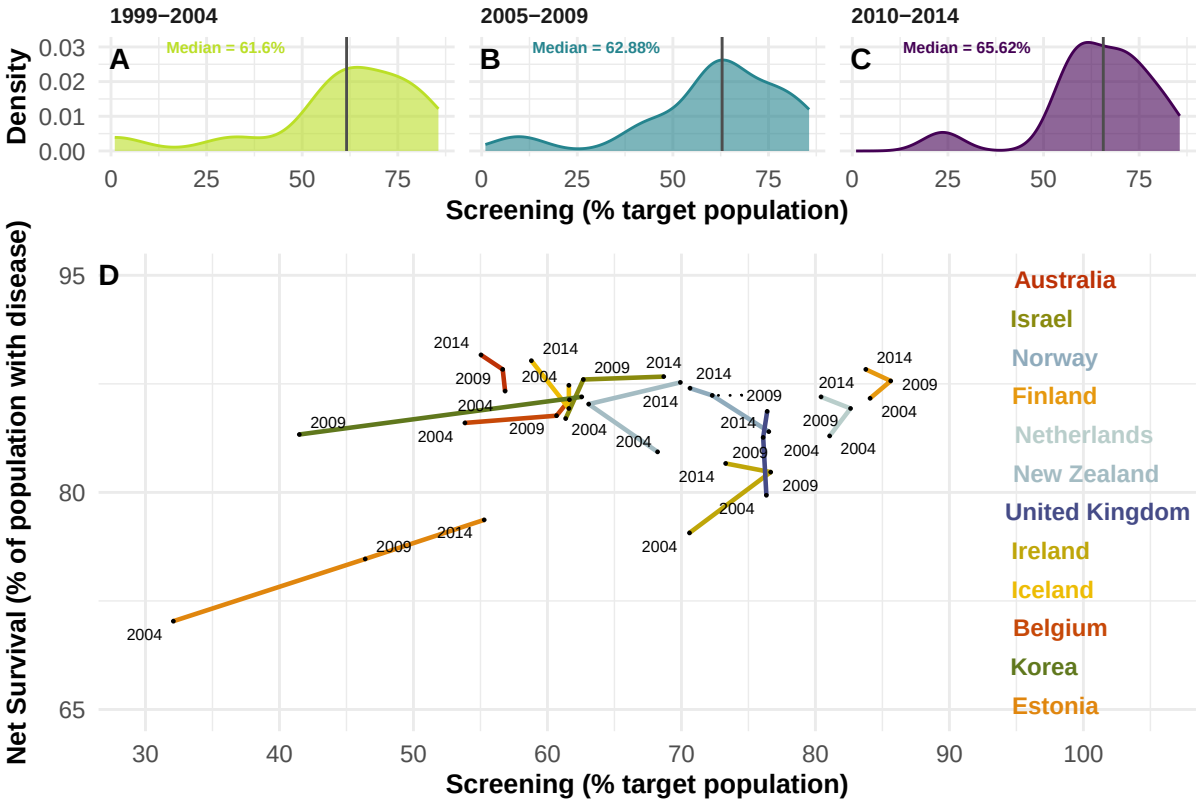

Figure 2: Screening of malignant neoplasms of the breast. Target population: Female, 50-69 years old. (A-C) Temporal changes in the distribution of screening uptake for each 5-years period. (D) Association between screenning rate uptake and 5-years net survival. Source: OECD (https://data-explorer.oecd.org/).

Limitations of the OECD dataset

There are a few limitations to the use of OECD data and potential inference. For instance, there was no evaluation of lag between the screening and survival metrics; this could be critical for a more accurate inference of the effects of increasing the efforts, which would be unlikely to have observable results in patient survival within five years. Another caveat is related to coverage. Out of the 30 countries that have data up to 2014 for the breast cancer screening program, only 14 have records that could be used for generating reliable metrics screening rates to match the 5-year aggregate data from the survival records. Another limitation regarding the OECD Healthcare Quality and Outcomes data set is the resolution since they do not provide a yearly report accounting for oncological patient survival. These two metrics reduce the overall data coverage and limit the inference that can be made out of it. Nevertheless, international comparative analysis of screening programs has shown to be a useful tool when rates are calculated across a target population with uniform criteria, as shown for colorectal cancer (Klabunde et al. 2015) and cervical cancer screening (Bruni et al. 2022). Alternatively, other international databases have emerged and could be incorporated (e.g., CanScreen5 (Zhang et al. 2023)), which resources can be explored complementary to this repository.

Oncological screening in Scotland

Data wrangling

```
#read file dictionary of health boards names and codes
hb21 <- read.csv("data/2_hbscot_codes_name.csv")
```

```r
#read a simple geojson to generate the map
map_HB_scot <- st_read("data/3_hbscot_map.geojson", quiet=TRUE)

# read mortality data Scotland per Health Board
mort <- read.csv("data/4_mortality_hbScot_24112024.csv")

# filter mortality for breast cases
mort_filter <- mort %>%
  filter(Year == '2016-2020',
         CancerSite == 'Breast',
         Sex == 'Female')

#merge mortality dataset with dictionary
mort_filter_hb21 <- merge(hb21, mort_filter, by.x = 'HB21CD', by.y = 'HB') %>%
  # adding \n so long HB names with & get split
  mutate(HB21NM = str_replace_all(HB21NM," and ", "\n & " ))

map_HB_scot <- map_HB_scot %>%
  mutate(h06_name = str_replace_all(h06_name," & ", "\n & " ))

# merge mortality and the dataframe from the map by Health Board name
map_HBmortality_scot_breast <- merge(map_HB_scot,
                                     mort_filter_hb21,
                                     by.x = 'h06_name',
                                     by.y = 'HB21NM')

mort_filter <- mort %>%
  filter(Year == '2016-2020',
         CancerSite == 'All cancer types incl NMSC',
         Sex == 'All')

#merge mortality dataset with dictionary
mort_filter_hb21 <- merge(hb21, mort_filter, by.x = 'HB21CD', by.y = 'HB') %>%
  # adding \n so long HB names with & get split
  mutate(HB21NM = str_replace_all(HB21NM," and ", "\n & " ))

map_HB_scot <- map_HB_scot %>%
  mutate(h06_name = str_replace_all(h06_name," & ", "\n & " ))

# merge mortality and the dataframe from the map by Health Board name
map_HBmortality_scot_all <- merge(map_HB_scot,
                                  mort_filter_hb21,
                                  by.x = 'h06_name',
                                  by.y = 'HB21NM')
```

```r
# data load and wrangling
#data source:
screening_scotland <- read_excel("data/5_screening_hbscot_2223.xlsx")
hb_names <- str_replace_all(map_HB_scot$h06_name, '\n', '')

colour_vec_hb <- met.brewer("VanGogh2", length(hb_names))
names(colour_vec_hb) <- hb_names

#add color black for national level screening rates
colour_vec_hb_scot = c(colour_vec_hb,  setNames("black", "Scotland"))
#define line type vector, segmented for national level
lty_hb_scot <- c(rep(1, length(hb_names)), 2)
```

```r
names(lty_hb_scot) <- names(colour_vec_hb_scot)

#plot production
scot_screen_breast <- screening_scotland %>%
  select(`Scottish Breast Screening Programme`:...12) %>%
  filter(`Scottish Breast Screening Programme` %in% c(hb_names,
                                                      "Scotland",
                                                      "NHS Board of Residence")) %>%
  row_to_names(row_number = 1) %>%
  select(where( ~!all(is.na(.x)))) %>%
  pivot_longer(names_to = "Period",
               cols = `2011/14`:`2020/23`,
               values_to = 'Uptake') %>%
  rename(hb_name = `NHS Board of Residence`) %>%
  separate_wider_delim(Period,"/", names = c(NA, "Period")) %>%
mutate(Uptake = as.numeric(Uptake),
       Period = as.numeric(Period),
       hb_name = factor(hb_name),
       hb_name_label = ifelse(Period == 23, as.character(hb_name), NA_character_),
       ) %>%
  ggplot(aes(Period, Uptake, group = hb_name,
             linetype = hb_name,
             label = hb_name_label)) +
  geom_line(aes(colour = hb_name))+
  geom_text_repel(aes(label = str_replace_all(hb_name_label, "&", "\n&"),
                      colour = hb_name_label),
                  fontface = "bold",
                  size = 3.8,
                  direction = "y",
                  segment.alpha = 0,
                  xlim = c(23.5, NA),
                  hjust = 0,
                  box.padding = .45) +
  scale_colour_manual(values = colour_vec_hb_scot)+
  scale_linetype_manual(values = lty_hb_scot) +
  #show target screening rate at 80%
  geom_segment(aes(x = 14, xend = 23, y = 80, yend = 80), colour= 'black') +
  geom_label(aes(x = 19, y=80, label="Target"), fill='white') +
  scale_x_continuous(expand=expand_scale(mult=c(0.05,.6)),
                     labels = paste0("20",seq(11,20),"-20",seq(14,23)),
                     breaks = seq(14,23)) +
  guides(colour = 'none', linetype = 'none')+
  labs(y = 'Screening (% target population)',
       x = '') +
    theme_minimal() +
  theme(axis.title =  element_text(size = 10, hjust = 0.5,face = "bold"),
        legend.position = 'bottom',
        legend.title = element_text(size = 8, angle = 90, hjust = 0.5,face = "bold"),
        legend.text = element_text(size=6),
        legend.key.height = unit(0.15, 'cm'),
        legend.key.width = unit(0.5, 'cm'),
        axis.text.x = element_text(angle = 45))
```

```r
# Plot with ggplot2's sf geom
all_map <- ggplot(data = map_HBmortality_scot_breast,aes(label=h06_name)) +
  geom_sf(aes(fill = WASR), color='white') +
  scale_fill_distiller(palette = "RdPu", direction = 1,
```

```
                       name = "Age-standardised mortality rate (WASR)") +
  geom_sf_text(fontface = "bold",size = 2)+
  theme_map()+
  theme(
    legend.position = "bottom",
    legend.justification = 0.5,
    legend.key.size = unit(0.5, "cm"),
    legend.key.width = unit(1.0, "cm"),
    legend.title.position = 'top',
    legend.text = element_text(size = 8),
    legend.margin = margin())+
  annotate(geom = 'text',
           label='A',
           x = -Inf,
           y = Inf,
           hjust = -0.5,
           vjust = 1.5,
           fontface = 2,
           family = "")

#plot breast neoplasms
breast_map <- ggplot(data = map_HBmortality_scot_breast, aes(label=h06_name)) +
  geom_sf(aes(fill = WASR), color='white') +
  scale_fill_distiller(direction = 1,
                       name = "Age-standardised mortality rate (WASR)")+
  geom_sf_text(fontface = "bold",size = 2)+
  theme_map()+
  theme(
    legend.position = "bottom",
    legend.justification = 0.5,
    legend.key.size = unit(0.5, "cm"),
    legend.key.width = unit(1.0, "cm"),
    legend.title.position = 'top',
    legend.text = element_text(size = 8),
    legend.margin = margin()) +
  annotate(geom = 'text',
           label='B',
           x = -Inf,
           y = Inf,
           hjust = -0.5,
           vjust = 1.5,
           fontface = 2,
           family = "")
```

Analysis of Scottish screening programs

For a more focused analysis in response to the unmet oncological needs in the country, we can look at Scotland's trends in screening and survival. Scotland has seven screening programs, of which three are cancer-related: bowel, breast, and cervical cancer. We leveraged the public availability of the screening data for the breast screening program collected and published by Public Health Scotland (PHS) (Scotland 2023). Cancer mortality data was also collected from PHS (Scotland 2023). We analysed the data at the Health Board (HB) level since it may represent better units for further focus and future deployment of interventions. There has been a notorious consistent increase in Scotland's Breast Screening Programme efforts during the last decade. During the last six years, all NHS HBs have exceeded the acceptable 70% screening rate, aiming for an 80% screening of the target population (females between 50-70 years old). Shetland and Orkney HBs already met the desired target. The NHS Greater Glasgow and Grampian HB just surpassed the 70% rate of acceptable screening, making a significant effort since

it is the largest HB, covering about 1.3 million people.

```
scot_screen_breast
```

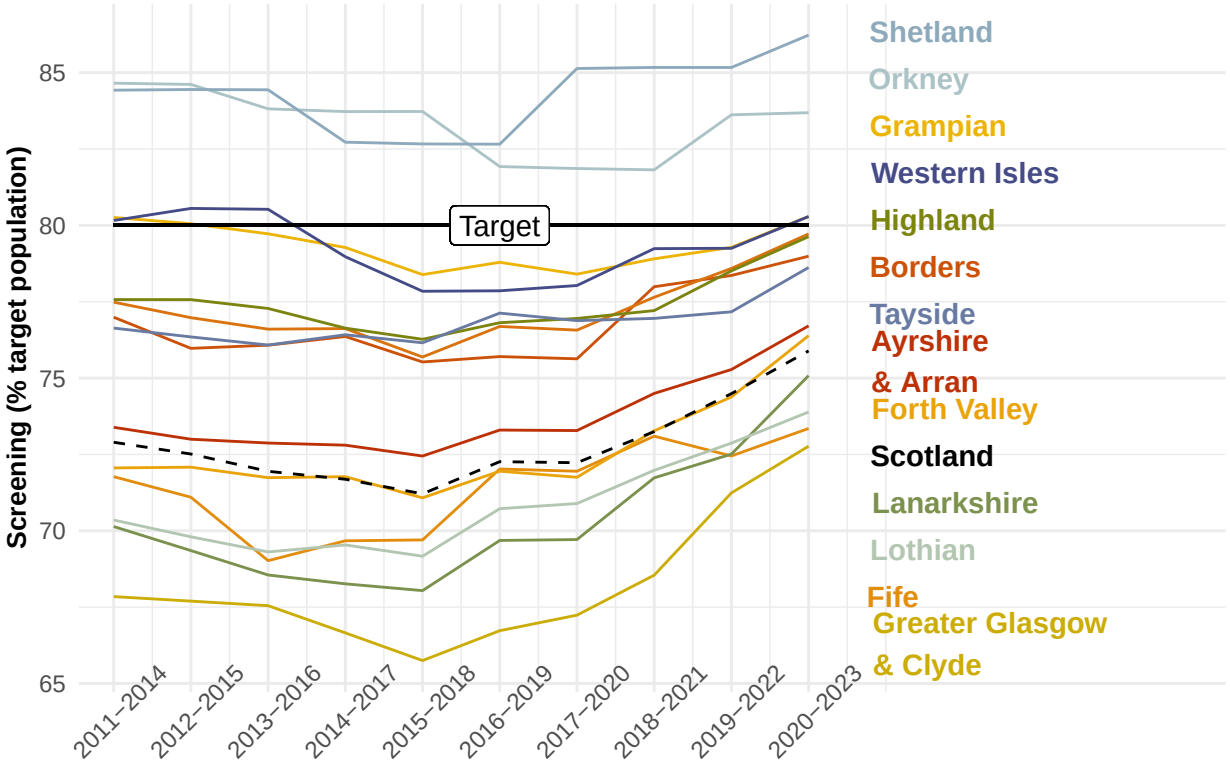

Figure 3: Screening of malignant neoplasms of the breast in Scotland. Health Board screening uptakes over three-year rolling period. Target population: Female, 50-70 years old. Horizontal line indicates target screening of 80%. Data source: Public Health Scotland (https://publichealthscotland.scot/)

Regarding cancer mortality, across all cancer types, it shows considerable heterogeneity across the country, which makes drawing specific conclusions rather challenging since there is a multi-dimensional nature of the drives of such a pattern, social and biological. Specifically for breast cancer, the trend is heterogeneous across HBs too. Some HBs have a high rate of mortality, such as the NHS Shetland and Orkney HBs, which contrast with the screening efforts. Other NHS HBs in the south, Boarders and Dumfries & Galloway, also experience challenges related to breast cancer mortality rates above the national median. The cases of the northern NHS HBs of Shetland and Orkney are interesting since although there are incredible efforts on screening; the mortalities are the highest in the country. Recent genetic evidence, however, has shed light on these patterns (Kerr et al. 2024), indicating that quantifying BRCA1 and BRCA2 genetic variants could explain the high mortality of breast cancer in the north. Furthermore, this entails an important biomarker for people with Orkney ancestry.

```
all_map | breast_map
```

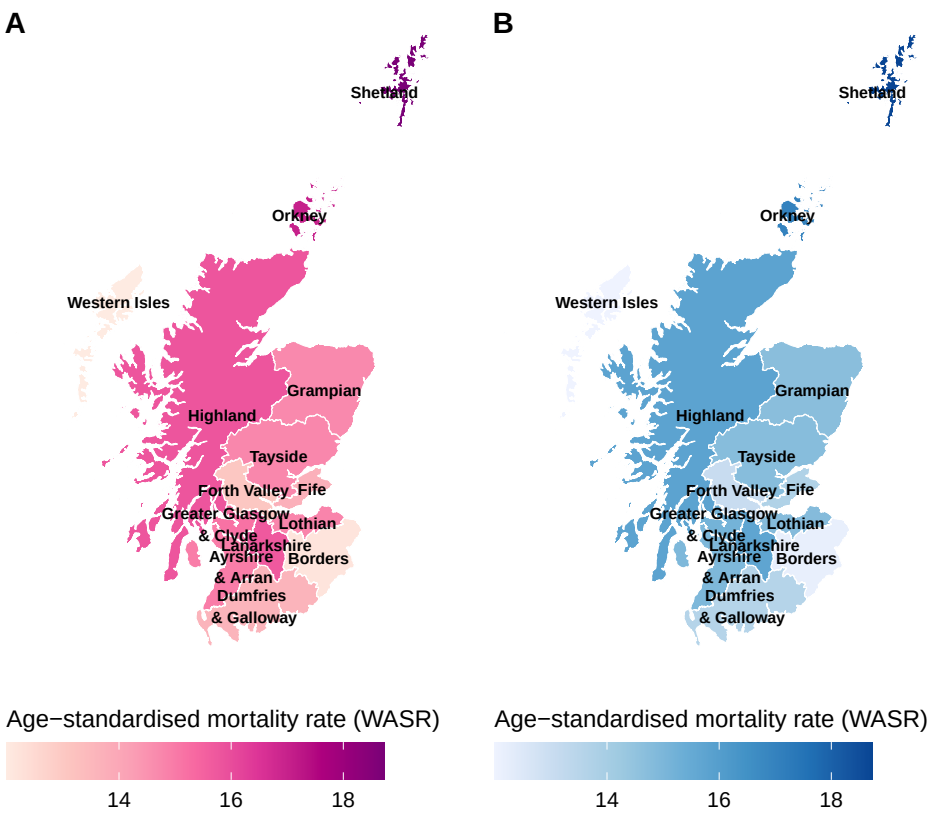

Figure 4: Spatial heterogeneity of cancer mortality in Scotland. World age-standardised mortality rate (WASR) per 100,000 person-years at risk for period 2016-2020 across (A) all cancer types and (B) breast cancer by Health Board. Data source: Public Health Scotland (https://www.opendata.nhs.scot).

Recommendations ahead in the context of the Research and Innovation Strategy

Following a more detailed discussion and data analysis in this report, in comparison to the presentation, there are a series of recommendations for incorporating cancer screening as an interdisciplinary program into the Research and Innovation Strategy. Specifically:

- Identify social determinants of health across health boards and how they shape the likelihood of the population accessing primary care for cancer screening.
- Deploy a pilot molecular screening to assess the occurrence of specific genetic variants that could drive pathogenicity reflected in high mortality rates at the country level.
- Collect evidence from other countries to further optimise the use of resources on patient screening and increase the specificity and sensitivity of current biomarkers.
- Identify how automated systems such as those powered by artificial intelligence can drive faster screenings and what is the digital infrastructure required to succeed.

References

Bruni, Laia, Beatriz Serrano, Esther Roura, Laia Alemany, Melanie Cowan, Rolando Herrero, Mario Poljak, et al. 2022. “Cervical Cancer Screening Programmes and Age-Specific Coverage Estimates for 202 Countries and Territories Worldwide: A Review and Synthetic Analysis.” *The Lancet Global Health* 10 (8): e1115–27.

Cedolini, Carla, Serena Bertozzi, Ambrogio P Londero, Sergio Bernardi, Luca Seriau, Serena Concina, Federico Cattin, and Andrea Risaliti. 2014. “Type of Breast Cancer Diagnosis, Screening, and Survival.” *Clinical Breast Cancer* 14 (4): 235–40.

Crosby, David, Sangeeta Bhatia, Kevin M Brindle, Lisa M Coussens, Caroline Dive, Mark Emberton, Sadik Esener, et al. 2022. “Early Detection of Cancer.” *Science* 375 (6586): eaay9040.

Hankey, Benjamin F, Eric J Feuer, Limin X Clegg, Richard B Hayes, Julie M Legler, Phillip C Prorok, Lynn A Ries, Ray M Merrill, and Richard S Kaplan. 1999. “Cancer Surveillance Series: Interpreting Trends in Prostate Cancer—Part i: Evidence of the Effects of Screening in Recent Prostate Cancer Incidence, Mortality, and Survival Rates.” *Journal of the National Cancer Institute* 91 (12): 1017–24.

Investigators, International Early Lung Cancer Action Program. 2006. “Survival of Patients with Stage i Lung Cancer Detected on CT Screening.” *New England Journal of Medicine* 355 (17): 1763–71.

Kerr, Shona M, Lucija Klaric, Marisa D Muckian, Emma Cowan, Lesley Snadden, Gannie Tzoneva, Alan R Shuldiner, Zosia Miedzybrodzka, and James F Wilson. 2024. “Two Founder Variants Account for over 90% of Pathogenic BRCA Alleles in the Orkney and Shetland Isles in Scotland.” *European Journal of Human Genetics*, 1–8.

Klabunde, Carrie, Johannes Blom, Jean-Luc Bulliard, Montse Garcia, Lea Hagoel, Verna Mai, Julietta Patnick, Heather Rozjabek, Carlo Senore, and Sven Törnberg. 2015. “Participation Rates for Organized Colorectal Cancer Screening Programmes: An International Comparison.” *Journal of Medical Screening* 22 (3): 119–26.

OECD. 2024a. *Healthcare Quality and Outcomes*. https://data-explorer.oecd.org/.

———. 2024b. *Healthcare Utilisation*. https://data-explorer.oecd.org/.

Scotland, Public Health. 2023. *Scottish Breast Screening Programme Statistics*. https://data-explorer.oecd.org/.

Smith, Robert A, Vilma Cokkinides, Andrew C von Eschenbach, Bernard Levin, Carmel Cohen, Carolyn D Runowicz, Stephen Sener, Debbie Saslow, and Harmon J Eyre. 2002. “American Cancer Society Guidelines for the Early Detection of Cancer.” *CA: A Cancer Journal for Clinicians* 52 (1): 8–22.

Zhang, Li, Isabel Mosquera, Eric Lucas, Mary Luz Rol, Andre L Carvalho, and Partha Basu. 2023. “CanScreen5, a Global Repository for Breast, Cervical and Colorectal Cancer Screening Programs.” *Nature Medicine* 29 (5): 1135–45.