Data preprocessing for analysis

Signal detection of spontaneous medical device reports over time

Ty Stanford and Curtis Murray

Table of contents

1 Set up

1.1 Packages

```
suppressPackageStartupMessages({
  library("readr")
  library("dplyr")
  library("tidyr")
  library("lubridate") # way to handle dates better than default R way
  library("ggplot2")
  library("ggthemes")
  library("purrr") # map(), map2() functions etc
  library("stringr")
  library("knitr")
  library("foreach")
  library("tictoc")
  library("arrow") # read/write parquet files
})
```

```
Warning: package 'ggthemes' was built under R version 4.3.2
```

```
# here are the functions written for this project
source("r/_funcs.R")
```

1.2 Constants

```
# arbitrarily, let's go with minimum cell count of 1
arbitrary_cell_min <- 1

# these are the thresholds for pain_topic to be pain == TRUE
# thresholds <- c(0.010, 0.025, 0.05, 0.075, 0.100, 0.150)
(thresholds <- sprintf("%1.3f", seq(0.010, 0.100, by = 0.005)))
```

```
 [1] "0.010" "0.015" "0.020" "0.025" "0.030" "0.035" "0.040" "0.045" "0.050"
[10] "0.055" "0.060" "0.065" "0.070" "0.075" "0.080" "0.085" "0.090" "0.095"
[19] "0.100"
```

```r
col_pal <- c("cyan4", "darkorange", "purple", "dodgerblue")


target_lst <-
  list(
    "pelvic_mesh",
    "pelvic_mesh",
    "pelvic_mesh",
    "hernia_mesh",
    c("hernia_mesh", "other_mesh")
  )

compar_lst <-
  list(
    "hernia_mesh",
    c("hernia_mesh", "other_mesh"),
    c("hernia_mesh", "other_mesh", "other_device"),
    "other_mesh",
    "other_device"
  )
```

1.3 Date function

```r
create_qtr_range <- function(start_qtr, end_qtr) {
  s_yr <- as.integer(substr(start_qtr, 1, 4))
  s_qr <- as.integer(substr(start_qtr, 7, 7))
  e_yr <- as.integer(substr(end_qtr, 1, 4))
  e_qr <- as.integer(substr(end_qtr, 7, 7))

  qtr_vec <- NULL
  if (s_yr > e_yr) {
    stop("End year must not be before start year")
  } else if ((s_yr == e_yr) & (s_qr > e_qr)) {
    stop("End quarter must not come before start quarter")
  } else if (s_yr == e_yr) {
    qtr_vec <- paste0(s_yr, "-Q", s_qr:e_qr)
  } else if (s_yr == (e_yr - 1)) {
    qtr_vec <-
      c(
        paste0(s_yr, "-Q", s_qr:4),
        paste0(e_yr, "-Q", 1:e_qr)
      )
  } else {
    yr_diff <- e_yr - s_yr - 1
    # print(yr_diff)
    qtr_vec <-
      c(
        paste0(s_yr, "-Q", s_qr:4),
        paste0(rep((s_yr + 1):(e_yr - 1), each = 4), "-Q", rep(1:4, yr_diff)),
        paste0(e_yr, "-Q", 1:e_qr)
      )
  }

  return(tibble(qtr = qtr_vec))

}
# create_qtr_range("2013-Q2", "2012-Q4") ### error tests
# create_qtr_range("2013-Q2", "2013-Q1")
create_qtr_range("2013-Q2", "2013-Q2")
```

```
# A tibble: 1 x 1
  qtr
  <chr>
1 2013-Q2
```

```r
create_qtr_range("2013-Q2", "2013-Q3")
```

```
# A tibble: 2 x 1
  qtr
  <chr>
1 2013-Q2
2 2013-Q3
```

```r
create_qtr_range("2013-Q2", "2014-Q1")
```

```
# A tibble: 4 x 1
  qtr
  <chr>
1 2013-Q2
```

```
2 2013-Q3
3 2013-Q4
4 2014-Q1
```

```r
create_qtr_range("2013-Q2", "2015-Q1")
```

```
# A tibble: 8 x 1
  qtr
  <chr>
1 2013-Q2
2 2013-Q3
3 2013-Q4
4 2014-Q1
5 2014-Q2
6 2014-Q3
7 2014-Q4
8 2015-Q1
```

```r
create_qtr_range("2013-Q4", "2015-Q1")
```

```
# A tibble: 6 x 1
  qtr
  <chr>
1 2013-Q4
2 2014-Q1
3 2014-Q2
4 2014-Q3
5 2014-Q4
6 2015-Q1
```

```r
create_qtr_range("2014-Q4", "2017-Q1")
```

```
# A tibble: 10 x 1
   qtr
   <chr>
 1 2014-Q4
 2 2015-Q1
 3 2015-Q2
 4 2015-Q3
 5 2015-Q4
 6 2016-Q1
 7 2016-Q2
 8 2016-Q3
 9 2016-Q4
10 2017-Q1
```

```r
create_qtr_range("2012-Q4", "2017-Q4") %>% print(., n = 21)
```

```
# A tibble: 21 x 1
   qtr
   <chr>
 1 2012-Q4
 2 2013-Q1
 3 2013-Q2
 4 2013-Q3
 5 2013-Q4
 6 2014-Q1
 7 2014-Q2
 8 2014-Q3
 9 2014-Q4
10 2015-Q1
11 2015-Q2
12 2015-Q3
13 2015-Q4
14 2016-Q1
15 2016-Q2
16 2016-Q3
17 2016-Q4
18 2017-Q1
19 2017-Q2
20 2017-Q3
21 2017-Q4
```

2 Data wrangling

2.1 Read data

```
clean_data_cols <-
  cols(
    Report_ID = col_double(),
    Date = col_date(format = ""),
    pain_word = col_logical(),
    pain_topic = col_double(),
    type = col_character()
  )

clean_data <- read_csv("dat/clean_data.csv", col_types = clean_data_cols)
```

2.2 Clean/remove duplicates

```
### all look like duplicates
inner_join(
  clean_data,
  clean_data %>%
    group_by(Report_ID) %>%
    summarise(n = n(), .groups = "drop") %>%
    dplyr::filter(n > 1),
  "Report_ID"
) %>%
  arrange(Report_ID) %>%
  print(., n = nrow(.))
```

```
# A tibble: 26 x 6
   Report_ID Date       pain_word pain_topic type             n
       <dbl> <date>     <lgl>          <dbl> <chr>        <int>
 1     29914 2014-07-03 TRUE          0.0270 other_device     2
 2     29914 2014-07-03 TRUE          0.0270 other_device     2
 3     31508 2014-07-03 TRUE          0.0882 other_device     2
 4     31508 2014-07-03 TRUE          0.0882 other_device     2
 5     32629 2014-07-03 FALSE         0      other_device     2
 6     32629 2014-07-03 FALSE         0      other_device     2
 7     36586 2015-03-25 FALSE         0      other_device     2
 8      36586  2015-03-25  FALSE         0       other_device      2
 9      36677  2015-06-26  FALSE         0       other_device      2
10      36677  2015-06-26  FALSE         0       other_device      2
11      36953  2015-06-06  FALSE         0       other_device      2
12      36953  2015-06-06  FALSE         0       other_device      2
13      41788  2016-12-08  FALSE         0       other_device      2
14      41788  2016-12-08  FALSE         0       other_device      2
15      43614  2016-12-13  FALSE         0       other_device      2
16      43614  2016-12-13  FALSE         0       other_device      2
17      45287  2017-06-04  FALSE         0       other_device      2
18      45287  2017-06-04  FALSE         0       other_device      2
19      45369  2017-05-20  FALSE         0       other_device      2
20      45369  2017-05-20  FALSE         0       other_device      2
21      45749  2017-10-06  FALSE         0       other_device      2
22      45749  2017-10-06  FALSE         0       other_device      2
23      46029  2017-10-06  FALSE         0       other_device      2
24      46029  2017-10-06  FALSE         0       other_device      2
25      46030  2017-09-06  FALSE         0       other_device      2
26      46030  2017-09-06  FALSE         0       other_device      2
```

```
# make dup free
clean_data <-
  clean_data %>%
  arrange(Report_ID, Date, desc(pain_word)) %>% # pain first in dups
  group_by(Report_ID) %>%
  dplyr::filter(row_number() == 1) %>%
  ungroup(.) %>%
  arrange(Date, Report_ID, desc(pain_word), desc(pain_topic))


clean_data %>%
  dplyr::filter(type == "other_mesh") %>%
  # select(Report_ID) %>%
  write_csv("out/other_mesh_ids.csv")
```

2.3 Inspect and summarise data

```
cat("First 10 rows of raw data\n")
```

```
First 10 rows of raw data
```

```
clean_data %>%
  arrange(Date) %>%
  dplyr::filter(row_number() < 11) %>%
  kable(.)
```

Report_ID	Date	pain_word	pain_topic	type
26696	2012-01-08	FALSE	0.0555556	other_device
27722	2012-01-08	FALSE	0.0000000	other_device
28827	2012-01-10	FALSE	0.0000000	other_device
28828	2012-01-10	TRUE	0.0500000	other_device
28452	2012-01-11	FALSE	0.0000000	other_device
28758	2012-01-11	FALSE	0.0000000	other_device
28826	2012-01-11	FALSE	0.0400000	other_device
29097	2012-01-11	FALSE	0.0000000	other_device
29100	2012-01-11	FALSE	0.0000000	other_device
29101	2012-01-11	FALSE	0.0000000	other_device

```
# clean_data <-
#   clean_data %>%
#   dplyr::filter(
#     type %in% c("pelvic_mesh", "hernia_mesh")
#   )


clean_data %>%
  with(., table(type, pain_word)) %>%
  knitr::kable(.)
```

	FALSE	TRUE
hernia_mesh	42	4
other_device	12741	1184
other_mesh	52	32
pelvic_mesh	32	70

```
clean_data %>%
  with(., table(type, pain_topic >= 0.05)) %>%
  knitr::kable(.)
```

	FALSE	TRUE
hernia_mesh	38	8
other_device	12386	1539
other_mesh	47	37
pelvic_mesh	25	77

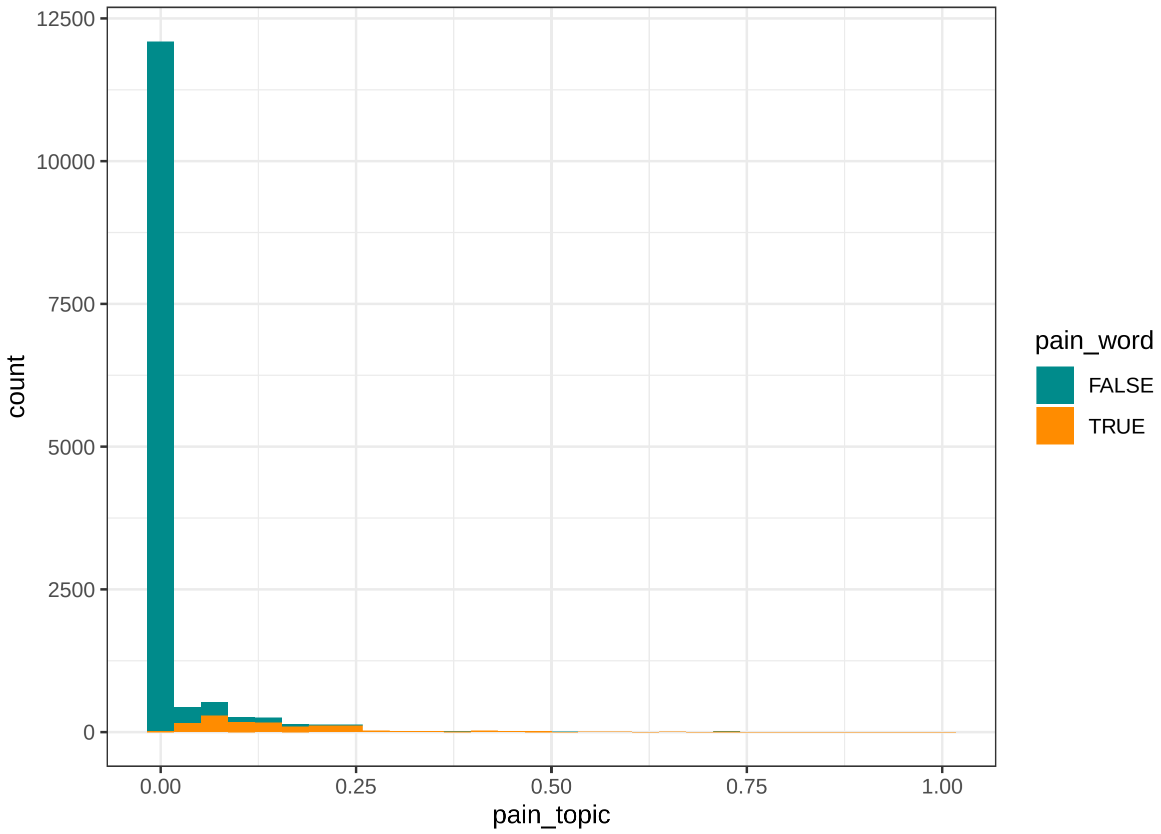
```
# These are the device groups and subgroups.
clean_data %>%
  group_by(type) %>%
  summarise(count = n()) %>%
  kable(.)
```

type	count
hernia_mesh	46
other_device	13925
other_mesh	84
pelvic_mesh	102

```
cat("\n\n## Histogram of `pain_word` (boolean) v `pain_topic` (score)")
```

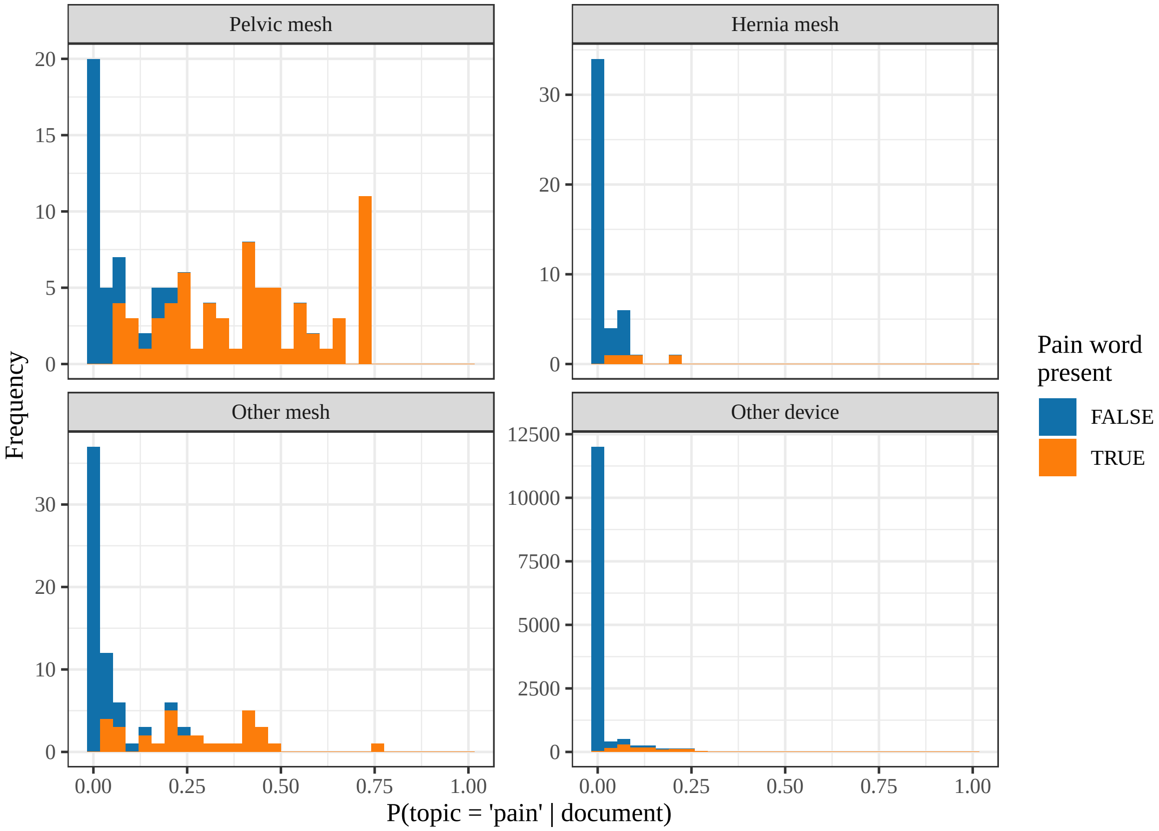
```
## Histogram of `pain_word` (boolean) v `pain_topic` (score)
```

```
clean_data %>%
  ggplot(., aes(pain_topic, fill = pain_word)) +
  geom_histogram(bins = 30) +
  scale_fill_manual(values = col_pal[1:2]) +
  theme_bw()
```



```r
type_lvls <- c("pelvic_mesh", "hernia_mesh", "other_mesh", "other_device")
type_lvls_edt <- str_to_sentence(str_replace_all(type_lvls, "_", " "))

clean_data %>%
  dplyr::filter(type %in% type_lvls) %>%
  mutate(
    type = str_to_sentence(str_replace_all(type, "_", " ")),
    type = factor(type, levels = type_lvls_edt)
  ) %>%
  ggplot(., aes(pain_topic, fill = pain_word)) +
  geom_histogram(bins = 30) +
  # scale_fill_manual(values = col_pal[2:1]) +
  scale_fill_tableau(palette = "Color Blind", direction = 1) +
  facet_wrap(~ type, scales = "free_y") +
  theme_bw() +
  theme(text = element_text(family = "serif")) +
  labs(
    x = "P(topic = 'pain' | document)",
    y = "Frequency",
    fill = "Pain word\npresent"
  )
```



```
ggsave(filename = "fig/pain_topic_dist.png", dpi = 900, width = 7, height = 4)
```

3 Create (monthly) data for analysis from raw data

3.1 Creation of analysis data

```r
### testing: example 1
# Use pelvic mesh as group 1 and hernia_mesh mesh devices as group 2.
# The value of interest is the pain topic, being above the threshold of 0.05.
# (i.e. 5% of the document contains words from the pain topic)
# You can adjust the topic threshold if you want to balance the groups more.
# A higher topic_threshold will look for documents that discuss "pain" more, and
# hence find less pain documents.

# get_signal_dat(
#   g1 = "pelvic_mesh",
#   g2 = "hernia_mesh",
#   pain_type = "pain_topic",
#   thresh = 0.05,
#   cell_min = 1,
#   cumul = TRUE,
#   verbose = FALSE
# ) %>%
#   bind_cols(., thresh = 0.05)



# takes ~ 20 sec
tic()
cumul_dat <-
  foreach(i = 1:length(target_lst), .combine = bind_rows, .packages = "dplyr") %do% {
    foreach(th_j = thresholds, .combine = bind_rows, .packages = "dplyr") %do% {

      get_signal_dat(
        g1 = target_lst[[i]],
        g2 = compar_lst[[i]],
        pain_type = "pain_topic",
        thresh = as.numeric(th_j),
        cell_min = 1,
        cumul = TRUE,
        verbose = FALSE
      ) %>%
        mutate(
          grps =
            paste0(
              "(", letters[i], ") ",
              paste(target_lst[[i]], collapse = "/"),
              " v ",
              paste(compar_lst[[i]], collapse = "/")
            ),
          dat_type = "cumulative",
          thresh = th_j
        ) %>%
          select(grps, dat_type, thresh, everything())

    }
  }
toc()
```

```
4.17 sec elapsed
```

```
cumul_dat
```

```
# A tibble: 4,523 x 8
   grps                          dat_type   thresh mnth     nA    nB    nC    nD
   <chr>                         <chr>      <chr>  <chr> <dbl> <dbl> <dbl> <dbl>
 1 (a) pelvic_mesh v hernia_mesh cumulative 0.010  2013~     3     7     1     2
 2 (a) pelvic_mesh v hernia_mesh cumulative 0.010  2013~     3     7     1     4
 3 (a) pelvic_mesh v hernia_mesh cumulative 0.010  2013~     3     7     1     5
 4 (a) pelvic_mesh v hernia_mesh cumulative 0.010  2013~     4    10     1     5
 5 (a) pelvic_mesh v hernia_mesh cumulative 0.010  2013~     4    11     1     7
 6 (a) pelvic_mesh v hernia_mesh cumulative 0.010  2013~     5    11     1     7
 7 (a) pelvic_mesh v hernia_mesh cumulative 0.010  2013~     5    11     2     9
 8 (a) pelvic_mesh v hernia_mesh cumulative 0.010  2013~     8    11     2     9
 9 (a) pelvic_mesh v hernia_mesh cumulative 0.010  2013~     9    11     2     9
10 (a) pelvic_mesh v hernia_mesh cumulative 0.010  2014~     9    11     2    10
# i 4,513 more rows
```

```
# takes ~ 20 sec
tic()
snpsh_dat <-
  foreach(i = 1:length(target_lst), .combine = bind_rows, .packages = "dplyr") %do% {
    foreach(th_j = thresholds, .combine = bind_rows, .packages = "dplyr") %do% {
      get_signal_dat(
        g1 = target_lst[[i]],
        g2 = compar_lst[[i]],
        pain_type = "pain_topic",
        thresh = as.numeric(th_j),
        cell_min = 1,
        cumul = FALSE,
        verbose = FALSE
      ) %>%
        mutate(
          grps =
            paste0(
              "(", letters[i], ") ",
              paste(target_lst[[i]], collapse = "/"),
              " v ",
              paste(compar_lst[[i]], collapse = "/")
            ),
          dat_type = "snapshot",
          thresh = th_j
        ) %>%
        select(grps, dat_type, thresh, everything())

    }
  }
toc()
```

```
4.28 sec elapsed
```

```
snpsh_dat
```

```
# A tibble: 4,523 x 8
   grps                            dat_type thresh mnth       nA    nB    nC    nD
   <chr>                           <chr>    <chr>  <chr>   <dbl> <dbl> <dbl> <dbl>
 1 (a) pelvic_mesh v hernia_mesh snapshot 0.010  2013-01     3     7     1     2
 2 (a) pelvic_mesh v hernia_mesh snapshot 0.010  2013-02     0     0     0     2
 3 (a) pelvic_mesh v hernia_mesh snapshot 0.010  2013-04     0     0     0     1
 4 (a) pelvic_mesh v hernia_mesh snapshot 0.010  2013-05     1     3     0     0
 5 (a) pelvic_mesh v hernia_mesh snapshot 0.010  2013-07     0     1     0     2
 6 (a) pelvic_mesh v hernia_mesh snapshot 0.010  2013-08     1     0     0     0
 7 (a) pelvic_mesh v hernia_mesh snapshot 0.010  2013-09     0     0     1     2
 8 (a) pelvic_mesh v hernia_mesh snapshot 0.010  2013-11     3     0     0     0
 9 (a) pelvic_mesh v hernia_mesh snapshot 0.010  2013-12     1     0     0     0
10 (a) pelvic_mesh v hernia_mesh snapshot 0.010  2014-03     0     0     0     1
# i 4,513 more rows
```

3.2 Check analysis data

```
nrow(cumul_dat)
```

```
[1] 4523
```

```
if (nrow(cumul_dat) != nrow(snpsh_dat)) {
  stop("logic of creating analysis data producing different # rows in data")
}


chk_start_vals <-
  inner_join(
    cumul_dat %>%
      group_by(grps, dat_type, thresh) %>%
      dplyr::filter(row_number() == 1) %>%
      ungroup(.),
    snpsh_dat %>%
      group_by(grps, dat_type, thresh) %>%
      dplyr::filter(row_number() == 1) %>%
      ungroup(.),
    c("grps", "thresh")
  ) %>%
    mutate(
      mnth_same = (mnth.x == mnth.y),
      counts_same = (nA.x = nA.y) & (nB.x = nB.y) & (nC.x = nC.y) & (nD.x = nD.y)
    )

chk_start_vals %>%
  select(grps, thresh, dat_type.x, dat_type.y, mnth_same, counts_same)
```

```
# A tibble: 95 x 6
   grps                        thresh dat_type.x dat_type.y mnth_same counts_same
   <chr>                       <chr>  <chr>      <chr>      <lgl>     <lgl>
```

```
 1 (a) pelvic_mesh v hernia_~ 0.010  cumulative snapshot   TRUE      TRUE
 2 (a) pelvic_mesh v hernia_~ 0.015  cumulative snapshot   TRUE      TRUE
 3 (a) pelvic_mesh v hernia_~ 0.020  cumulative snapshot   TRUE      TRUE
 4 (a) pelvic_mesh v hernia_~ 0.025  cumulative snapshot   TRUE      TRUE
 5 (a) pelvic_mesh v hernia_~ 0.030  cumulative snapshot   TRUE      TRUE
 6 (a) pelvic_mesh v hernia_~ 0.035  cumulative snapshot   TRUE      TRUE
 7 (a) pelvic_mesh v hernia_~ 0.040  cumulative snapshot   TRUE      TRUE
 8 (a) pelvic_mesh v hernia_~ 0.045  cumulative snapshot   TRUE      TRUE
 9 (a) pelvic_mesh v hernia_~ 0.050  cumulative snapshot   TRUE      TRUE
10 (a) pelvic_mesh v hernia_~ 0.055  cumulative snapshot   TRUE      TRUE
# i 85 more rows
```

```
with(chk_start_vals, table(mnth_same, counts_same, useNA = "ifany"))
```

```
         counts_same
mnth_same TRUE
     TRUE   95
```

```
# check the first + second row in snapshot == second row in cumulative data
inner_join(
  cumul_dat %>%
    group_by(grps, thresh) %>%
    dplyr::filter(row_number() %in% 1:2) %>%
    ungroup(.),
  snpsh_dat %>%
    group_by(grps, thresh) %>%
    dplyr::filter(row_number() %in% 1:2) %>%
    ungroup(.),
  c("grps", "thresh", "mnth")
)
```

```
# A tibble: 190 x 13
   grps  dat_type.x thresh mnth   nA.x  nB.x  nC.x  nD.x dat_type.y  nA.y  nB.y
   <chr>  <chr>      <chr>  <chr> <dbl> <dbl> <dbl> <dbl> <chr>      <dbl> <dbl>
 1 (a) p~ cumulative 0.010  2013~      3     7     1     2 snapshot        3     7
 2 (a) p~ cumulative 0.010  2013~      3     7     1     4 snapshot        0     0
 3 (a) p~ cumulative 0.015  2013~      3     7     1     2 snapshot        3     7
 4 (a) p~ cumulative 0.015  2013~      3     7     1     4 snapshot        0     0
 5 (a) p~ cumulative 0.020  2013~      5    11     1    10 snapshot        5    11
 6 (a) p~ cumulative 0.020  2013~      8    11     1    10 snapshot        3     0
 7 (a) p~ cumulative 0.025  2013~     5     11      1     10 snapshot        5    11
 8 (a) p~ cumulative 0.025  2013~     8     11      1     10 snapshot        3     0
 9 (a) p~ cumulative 0.030  2013~     5     11      1     10 snapshot        5    11
10 (a) p~ cumulative 0.030  2013~     8     11      1     10 snapshot        3     0
# i 180 more rows
# i 2 more variables: nC.y <dbl>, nD.y <dbl>
```

3.3 Export analysis data

```
# all spontaneous report analysis data
sra_dat <-
  bind_rows(
    cumul_dat,
    snpsh_dat
  )

sra_dat %>%
  write_parquet(., sink = "dat/sra_dat.parquet")
```

4 Create (quarterly, complete) data for analysis from raw data

4.1 Creation of analysis data

```
cumul_qtrly_dat <-
  cumul_dat %>%
  mutate(
    mnth_qtr =
      quarter(
        as_date(paste0(mnth, "-01")),
        type = "quarter"
      ),
    mnth_qtr = paste0(substr(mnth, 1, 5), "Q", as.character(mnth_qtr))
  )

cumul_qtrly_dat <-
  cumul_qtrly_dat %>%
  group_by(grps, dat_type, thresh, mnth_qtr) %>%
  dplyr::filter(row_number() == n()) %>%
  ungroup()

cumul_qtrly_dat
```

```
# A tibble: 1,691 x 9
   grps                    dat_type thresh mnth      nA    nB    nC    nD mnth_qtr
   <chr>                   <chr>    <chr>  <chr> <dbl> <dbl> <dbl> <dbl> <chr>
 1 (a) pelvic_mesh v her~ cumulat~ 0.010  2013~     3     7     1     4 2013-Q1
 2 (a) pelvic_mesh v her~ cumulat~ 0.010  2013~     4    10     1     5 2013-Q2
 3 (a) pelvic_mesh v her~ cumulat~ 0.010  2013~     5    11     2     9 2013-Q3
 4 (a) pelvic_mesh v her~ cumulat~ 0.010  2013~     9    11     2     9 2013-Q4
 5 (a) pelvic_mesh v her~ cumulat~ 0.010  2014~     9    11     2    10 2014-Q1
 6 (a) pelvic_mesh v her~ cumulat~ 0.010  2014~    10    12     3    12 2014-Q2
 7 (a) pelvic_mesh v her~ cumulat~ 0.010  2014~    12    14     5    19 2014-Q3
 8 (a) pelvic_mesh v her~ cumulat~ 0.010  2014~    30    15     7    24 2014-Q4
 9 (a) pelvic_mesh v her~ cumulat~ 0.010  2015~    31    15     7    25 2015-Q1
10 (a) pelvic_mesh v her~ cumulat~ 0.010  2015~    31    16     7    25 2015-Q3
# i 1,681 more rows
```

```
cumul_qtrly_dat <-
  cumul_qtrly_dat %>%
  mutate(mnth = mnth_qtr) %>%
  select(-mnth_qtr)

cumul_qtrly_dat_summ <-
  cumul_qtrly_dat %>%
  group_by(grps, dat_type, thresh) %>%
  summarise(
    min_dte = min(mnth),
    max_dte = max(mnth),
    n_row = n(),
    .groups = "drop"
  )



cumul_qtrly_dat_summ
```

```
# A tibble: 95 x 6
   grps                          dat_type   thresh min_dte max_dte n_row
   <chr>                         <chr>      <chr>  <chr>   <chr>   <int>
 1 (a) pelvic_mesh v hernia_mesh cumulative 0.010  2013-Q1 2017-Q4    19
 2 (a) pelvic_mesh v hernia_mesh cumulative 0.015  2013-Q1 2017-Q4    19
 3 (a) pelvic_mesh v hernia_mesh cumulative 0.020  2013-Q3 2017-Q4    17
 4 (a) pelvic_mesh v hernia_mesh cumulative 0.025  2013-Q3 2017-Q4    17
 5 (a) pelvic_mesh v hernia_mesh cumulative 0.030  2013-Q3 2017-Q4    17
 6 (a) pelvic_mesh v hernia_mesh cumulative 0.035  2013-Q3 2017-Q4    17
 7 (a) pelvic_mesh v hernia_mesh cumulative 0.040  2013-Q3 2017-Q4    17
 8 (a) pelvic_mesh v hernia_mesh cumulative 0.045  2013-Q3 2017-Q4    17
 9 (a) pelvic_mesh v hernia_mesh cumulative 0.050  2013-Q3 2017-Q4    17
10 (a) pelvic_mesh v hernia_mesh cumulative 0.055  2013-Q3 2017-Q4    17
# i 85 more rows
```

```
cumul_qtrly_dat_summ <-
  cumul_qtrly_dat_summ %>%
  mutate(
    range = map2(.x = min_dte, .y = max_dte, .f = create_qtr_range)
  ) %>%
  unnest(cols = range)

cumul_qtrly_dat_summ %>%
  print(., n = 22)
```

```
# A tibble: 1,707 x 7
   grps                          dat_type   thresh min_dte max_dte n_row qtr
   <chr>                         <chr>      <chr>  <chr>   <chr>   <int> <chr>
 1 (a) pelvic_mesh v hernia_mesh cumulative 0.010  2013-Q1 2017-Q4    19 2013-Q1
 2 (a) pelvic_mesh v hernia_mesh cumulative 0.010  2013-Q1 2017-Q4    19 2013-Q2
 3 (a) pelvic_mesh v hernia_mesh cumulative 0.010  2013-Q1 2017-Q4    19 2013-Q3
 4 (a) pelvic_mesh v hernia_mesh cumulative 0.010  2013-Q1 2017-Q4    19 2013-Q4
 5 (a) pelvic_mesh v hernia_mesh cumulative 0.010  2013-Q1 2017-Q4    19 2014-Q1
 6 (a) pelvic_mesh v hernia_mesh cumulative 0.010  2013-Q1 2017-Q4    19 2014-Q2
 7 (a) pelvic_mesh v hernia_mesh cumulative 0.010  2013-Q1 2017-Q4    19 2014-Q3
 8 (a) pelvic_mesh v hernia_mesh cumulative 0.010  2013-Q1 2017-Q4    19 2014-Q4
 9 (a) pelvic_mesh v hernia_mesh cumulative 0.010  2013-Q1 2017-Q4    19 2015-Q1
10 (a) pelvic_mesh v hernia_mesh cumulative 0.010  2013-Q1 2017-Q4    19 2015-Q2
11 (a) pelvic_mesh v hernia_mesh cumulative 0.010  2013-Q1 2017-Q4    19 2015-Q3
12 (a) pelvic_mesh v hernia_mesh cumulative 0.010  2013-Q1 2017-Q4    19 2015-Q4
13 (a) pelvic_mesh v hernia_mesh cumulative 0.010  2013-Q1 2017-Q4    19 2016-Q1
14 (a) pelvic_mesh v hernia_mesh cumulative 0.010  2013-Q1 2017-Q4    19 2016-Q2
15 (a) pelvic_mesh v hernia_mesh cumulative 0.010  2013-Q1 2017-Q4    19 2016-Q3
16 (a) pelvic_mesh v hernia_mesh cumulative 0.010  2013-Q1 2017-Q4    19 2016-Q4
17 (a) pelvic_mesh v hernia_mesh cumulative 0.010  2013-Q1 2017-Q4    19 2017-Q1
18 (a) pelvic_mesh v hernia_mesh cumulative 0.010  2013-Q1 2017-Q4    19 2017-Q2
19 (a) pelvic_mesh v hernia_mesh cumulative 0.010  2013-Q1 2017-Q4    19 2017-Q3
20 (a) pelvic_mesh v hernia_mesh cumulative 0.010  2013-Q1 2017-Q4    19 2017-Q4
21 (a) pelvic_mesh v hernia_mesh cumulative 0.015  2013-Q1 2017-Q4    19 2013-Q1
22 (a) pelvic_mesh v hernia_mesh cumulative 0.015  2013-Q1 2017-Q4    19 2013-Q2
# i 1,685 more rows
```

```r
nrow(cumul_qtrly_dat)
```

```
[1] 1691
```

```r
nrow(cumul_qtrly_dat_summ)
```

```
[1] 1707
```

```r
cumul_qtrly_dat <-
  left_join(
    cumul_qtrly_dat_summ %>% select(grps, dat_type, thresh, mnth = qtr),
    cumul_qtrly_dat,
    c("grps", "dat_type", "thresh", "mnth")
  )
nrow(cumul_qtrly_dat)
```

```
[1] 1707
```

```
cumul_qtrly_dat <-
  cumul_qtrly_dat %>%
  arrange(grps, dat_type, thresh, mnth)

cumul_qtrly_dat
```

```
# A tibble: 1,707 x 8
   grps                           dat_type   thresh mnth      nA    nB    nC    nD
   <chr>                          <chr>      <chr>  <chr> <dbl> <dbl> <dbl> <dbl>
 1 (a) pelvic_mesh v hernia_mesh cumulative 0.010  2013~      3     7     1     4
 2 (a) pelvic_mesh v hernia_mesh cumulative 0.010  2013~      4    10     1     5
 3 (a) pelvic_mesh v hernia_mesh cumulative 0.010  2013~      5    11     2     9
 4 (a) pelvic_mesh v hernia_mesh cumulative 0.010  2013~      9    11     2     9
 5 (a) pelvic_mesh v hernia_mesh cumulative 0.010  2014~      9    11     2    10
 6 (a) pelvic_mesh v hernia_mesh cumulative 0.010  2014~     10    12     3    12
 7 (a) pelvic_mesh v hernia_mesh cumulative 0.010  2014~     12    14     5    19
 8 (a) pelvic_mesh v hernia_mesh cumulative 0.010  2014~     30    15     7    24
 9 (a) pelvic_mesh v hernia_mesh cumulative 0.010  2015~     31    15     7    25
10 (a) pelvic_mesh v hernia_mesh cumulative 0.010  2015~     NA    NA    NA    NA
# i 1,697 more rows
```

```
which_nas <- which(with(cumul_qtrly_dat, is.na(nA)))
# problem children
cumul_qtrly_dat %>% dplyr::filter(row_number() %in% which_nas)
```

```
# A tibble: 16 x 8
   grps                           dat_type   thresh mnth      nA    nB    nC    nD
   <chr>                          <chr>      <chr>  <chr> <dbl> <dbl> <dbl> <dbl>
 1 (a) pelvic_mesh v hernia_mesh cumulative 0.010  2015~     NA    NA    NA    NA
 2 (a) pelvic_mesh v hernia_mesh cumulative 0.015  2015~     NA    NA    NA    NA
 3 (a) pelvic_mesh v hernia_mesh cumulative 0.020  2015~     NA    NA    NA    NA
 4 (a) pelvic_mesh v hernia_mesh cumulative 0.025  2015~     NA    NA    NA    NA
 5 (a) pelvic_mesh v hernia_mesh cumulative 0.030  2015~     NA    NA    NA    NA
 6 (a) pelvic_mesh v hernia_mesh cumulative 0.035  2015~    NA    NA    NA    NA
 7 (a) pelvic_mesh v hernia_mesh cumulative 0.040  2015~    NA    NA    NA    NA
 8 (a) pelvic_mesh v hernia_mesh cumulative 0.045  2015~    NA    NA    NA    NA
 9 (a) pelvic_mesh v hernia_mesh cumulative 0.050  2015~    NA    NA    NA    NA
10 (a) pelvic_mesh v hernia_mesh cumulative 0.055  2015~    NA    NA    NA    NA
11 (a) pelvic_mesh v hernia_mesh cumulative 0.060  2015~    NA    NA    NA    NA
12 (a) pelvic_mesh v hernia_mesh cumulative 0.065  2015~    NA    NA    NA    NA
13 (a) pelvic_mesh v hernia_mesh cumulative 0.070  2015~    NA    NA    NA    NA
14 (a) pelvic_mesh v hernia_mesh cumulative 0.075  2015~    NA    NA    NA    NA
15 (a) pelvic_mesh v hernia_mesh cumulative 0.080  2015~    NA    NA    NA    NA
16 (a) pelvic_mesh v hernia_mesh cumulative 0.085  2015~    NA    NA    NA    NA
```

```
# rows prior to problem children
cumul_qtrly_dat %>% dplyr::filter(row_number() %in% (which_nas - 1))
```

```
# A tibble: 16 x 8
   grps                          dat_type   thresh mnth     nA    nB    nC    nD
   <chr>                         <chr>      <chr>  <chr> <dbl> <dbl> <dbl> <dbl>
 1 (a) pelvic_mesh v hernia_mesh cumulative 0.010  2015~    31    15     7    25
 2 (a) pelvic_mesh v hernia_mesh cumulative 0.015  2015~    31    15     7    25
 3 (a) pelvic_mesh v hernia_mesh cumulative 0.020  2015~    31    15     6    26
 4 (a) pelvic_mesh v hernia_mesh cumulative 0.025  2015~    31    15     6    26
 5 (a) pelvic_mesh v hernia_mesh cumulative 0.030  2015~    31    15     6    26
 6 (a) pelvic_mesh v hernia_mesh cumulative 0.035  2015~    31    15     6    26
 7 (a) pelvic_mesh v hernia_mesh cumulative 0.040  2015~    30    16     5    27
 8 (a) pelvic_mesh v hernia_mesh cumulative 0.045  2015~    28    18     4    28
 9 (a) pelvic_mesh v hernia_mesh cumulative 0.050  2015~    27    19     4    28
10 (a) pelvic_mesh v hernia_mesh cumulative 0.055  2015~    27    19     4    28
11 (a) pelvic_mesh v hernia_mesh cumulative 0.060  2015~    26    20     4    28
12 (a) pelvic_mesh v hernia_mesh cumulative 0.065  2015~    26    20     3    29
13 (a) pelvic_mesh v hernia_mesh cumulative 0.070  2015~    25    21     2    30
14 (a) pelvic_mesh v hernia_mesh cumulative 0.075  2015~    24    22     2    30
15 (a) pelvic_mesh v hernia_mesh cumulative 0.080  2015~    24    22     2    30
16 (a) pelvic_mesh v hernia_mesh cumulative 0.085  2015~    24    22     1    31
```

```
cumul_qtrly_dat$nA[which_nas] <- cumul_qtrly_dat$nA[which_nas - 1]
# cumul_qtrly_dat %>% dplyr::filter(row_number() %in% which_nas)
cumul_qtrly_dat$nB[which_nas] <- cumul_qtrly_dat$nB[which_nas - 1]
cumul_qtrly_dat$nC[which_nas] <- cumul_qtrly_dat$nC[which_nas - 1]
cumul_qtrly_dat$nD[which_nas] <- cumul_qtrly_dat$nD[which_nas - 1]
```

```
# fixed? (yes)
cumul_qtrly_dat %>% dplyr::filter(row_number() %in% which_nas)
```

```
# A tibble: 16 x 8
   grps                          dat_type   thresh mnth     nA    nB    nC    nD
   <chr>                         <chr>      <chr>  <chr> <dbl> <dbl> <dbl> <dbl>
 1 (a) pelvic_mesh v hernia_mesh cumulative 0.010  2015~    31    15     7    25
 2 (a) pelvic_mesh v hernia_mesh cumulative 0.015  2015~    31    15     7    25
 3 (a) pelvic_mesh v hernia_mesh cumulative 0.020  2015~    31    15     6    26
 4 (a) pelvic_mesh v hernia_mesh cumulative 0.025  2015~    31    15     6    26
 5 (a) pelvic_mesh v hernia_mesh cumulative 0.030  2015~    31    15     6    26
 6 (a) pelvic_mesh v hernia_mesh cumulative 0.035  2015~    31    15     6    26
 7 (a) pelvic_mesh v hernia_mesh cumulative 0.040  2015~    30    16     5    27
 8 (a) pelvic_mesh v hernia_mesh cumulative 0.045  2015~    28    18     4    28
 9 (a) pelvic_mesh v hernia_mesh cumulative 0.050  2015~    27    19     4    28
10 (a) pelvic_mesh v hernia_mesh cumulative 0.055  2015~    27    19     4    28
11 (a) pelvic_mesh v hernia_mesh cumulative 0.060  2015~    26    20     4    28
12 (a) pelvic_mesh v hernia_mesh cumulative 0.065  2015~    26    20     3    29
13 (a) pelvic_mesh v hernia_mesh cumulative 0.070  2015~    25    21     2    30
14 (a) pelvic_mesh v hernia_mesh cumulative 0.075  2015~    24    22     2    30
15 (a) pelvic_mesh v hernia_mesh cumulative 0.080  2015~    24    22     2    30
16 (a) pelvic_mesh v hernia_mesh cumulative 0.085  2015~    24    22     1    31
```

```
cumul_dat %>% distinct(grps)
```

```
# A tibble: 5 x 1
  grps
  <chr>
1 (a) pelvic_mesh v hernia_mesh
2 (b) pelvic_mesh v hernia_mesh/other_mesh
3 (c) pelvic_mesh v hernia_mesh/other_mesh/other_device
4 (d) hernia_mesh v other_mesh
5 (e) hernia_mesh/other_mesh v other_device
```

4.2 Export analysis data

```
cumul_qtrly_dat %>%
  write_parquet(., sink = "dat/cumul_qtrly_dat.parquet")
```

5 Session information

```r
# Sys.info()[!(names(Sys.info()) %in% c("login", "nodename"))] %>%
#   as.data.frame(.)
format(Sys.time(), '%d %b %Y')
```

```
[1] "04 Dec 2023"
```

```r
sessionInfo()
```

```
R version 4.3.1 (2023-06-16 ucrt)
Platform: x86_64-w64-mingw32/x64 (64-bit)
Running under: Windows 10 x64 (build 19045)

Matrix products: default


locale:
[1] LC_COLLATE=English_Australia.utf8  LC_CTYPE=English_Australia.utf8
[3] LC_MONETARY=English_Australia.utf8 LC_NUMERIC=C
[5] LC_TIME=English_Australia.utf8

time zone: Australia/Adelaide
tzcode source: internal

attached base packages:
[1] stats     graphics  grDevices utils     datasets  methods   base

other attached packages:
 [1] arrow_12.0.1.1  tictoc_1.2      foreach_1.5.2   knitr_1.43
 [5] stringr_1.5.0   purrr_1.0.1     ggthemes_5.0.0  ggplot2_3.4.2
 [9] lubridate_1.9.2 tidyr_1.3.0     dplyr_1.1.2     readr_2.1.4

loaded via a namespace (and not attached):
 [1] utf8_1.2.3        generics_0.1.3    stringi_1.7.12    hms_1.1.3
 [5] digest_0.6.33     magrittr_2.0.3    evaluate_0.21     grid_4.3.1
 [9] timechange_0.2.0  iterators_1.0.14  fastmap_1.1.1     jsonlite_1.8.7
[13] fansi_1.0.4       scales_1.2.1      textshaping_0.3.6 codetools_0.2-19
[17] cli_3.6.1         rlang_1.1.1       crayon_1.5.2      bit64_4.0.5
[21] munsell_0.5.0     withr_2.5.0       yaml_2.3.7        tools_4.3.1
```

```
[25] parallel_4.3.1    tzdb_0.4.0        colorspace_2.1-0  assertthat_0.2.1
[29] vctrs_0.6.3       R6_2.5.1          lifecycle_1.0.3   bit_4.0.5
[33] vroom_1.6.3       ragg_1.2.5        pkgconfig_2.0.3   pillar_1.9.0
[37] gtable_0.3.3      glue_1.6.2        systemfonts_1.0.4 xfun_0.39
[41] tibble_3.2.1      tidyselect_1.2.0  rstudioapi_0.15.0 farver_2.1.1
[45] htmltools_0.5.5   rmarkdown_2.23    labeling_0.4.2    compiler_4.3.1
```