

Cleaning of the data

Jono Tuke

21/03/2019

1 Nigel's cleaning

Nigel did some cleaning of the original dataset as follows

1.1 18th March 2019

- Copy Paper ID to all rows involving that paper.
- Copy Condition to all relevant rows.
- Create a new column containing "Continuous", "Quartile" etc.
- Delete blank rows within a paper, currently left blank rows separating papers to make it easier to read
- Highlighted any obviously dubious cells/entries in Red, left yellow ones from Stephen for clarification
- Tidied up "minus" entries - Note that the Confidence Interval notation is not standardised!
- Explicitly annotated "plus" with the prior categories
- Removed "adj" from conditions, where the covariates indicate a deeper analysis
- Inserted annotations, such as "- male", into the "Condition" description to uniquely identify each test.

2 Read in the data

```
TSH <- read_excel("../data/20190318_excel tsh_2019_data.xlsx")
```

```
## New names:  
## * `` -> ...19
```

```
TSH
```

```
## # A tibble: 509 x 19
##   Paper `Continuous V q... Condition FT4_N FT4_method FT4_Sig FT4_PV FT4_cov
##   <chr> <chr>             <chr>    <chr> <chr>      <chr>  <chr>  <chr>
## 1 Capp... Continuous      AF      2673 Cox PH      Y      < 0.0... Age, s...
## 2 Capp... Continuous      AF      2673 Cox PH      Y      0.02    Age, s...
## 3 Capp... Continuous      CHD      2215 Cox PH      N      0.13    Age, s...
## 4 Capp... Continuous      CHD      2215 Cox PH      N      0.280... Age, s...
## 5 Capp... Continuous      Incident... 2706 Cox PH      Y      4.000... Age, s...
## 6 Capp... Continuous      Incident... 2706 Cox PH      Y      0.03    Age, s...
## 7 Capp... Continuous      Composit... 2073 Cox PH      Y      8.000... Age, s...
## 8 Capp... Continuous      Composit... 2073 Cox PH      Y      0.02    Age, s...
## 9 Capp... Continuous      Hip Frac... 2803 Cox PH      N      0.79    Age, a...
## 10 Capp... Continuous      Hip Frac... 2803 Cox PH      N      0.560... Age, a...
## # ... with 499 more rows, and 11 more variables: TSH_N <chr>,
## #   TSH_method <chr>, TSH_Sig <chr>, TSH_PV <chr>, TSH_cov <chr>,
## #   T3_N <chr>, T3_method <chr>, T3_Sig <chr>, T3_PV <chr>, T3_cov <chr>,
## #   ...19 <chr>
```

3 Cleaning of the data

3.1 Missing rows

Nigel has mentioned that left blank rows between the papers, we will remove these

```
TSH <-
  TSH %>%
  filter(!is.na(Paper))
TSH
```

```
## # A tibble: 460 x 19
##   Paper `Continuous V q... Condition FT4_N FT4_method FT4_Sig FT4_PV FT4_cov
##   <chr> <chr>             <chr>    <chr> <chr>      <chr>  <chr>  <chr>
## 1 Capp... Continuous      AF      2673 Cox PH      Y      < 0.0... Age, s...
## 2 Capp... Continuous      AF      2673 Cox PH      Y      0.02    Age, s...
## 3 Capp... Continuous      CHD      2215 Cox PH      N      0.13    Age, s...
## 4 Capp... Continuous      CHD      2215 Cox PH      N      0.280... Age, s...
## 5 Capp... Continuous      Incident... 2706 Cox PH      Y      4.000... Age, s...
## 6 Capp... Continuous      Incident... 2706 Cox PH      Y      0.03    Age, s...
## 7 Capp... Continuous      Composit... 2073 Cox PH      Y      8.000... Age, s...
## 8 Capp... Continuous      Composit... 2073 Cox PH      Y      0.02    Age, s...
## 9 Capp... Continuous      Hip Frac... 2803 Cox PH      N      0.79    Age, a...
## 10 Capp... Continuous      Hip Frac... 2803 Cox PH      N      0.560... Age, a...
## # ... with 450 more rows, and 11 more variables: TSH_N <chr>,
## #   TSH_method <chr>, TSH_Sig <chr>, TSH_PV <chr>, TSH_cov <chr>,
## #   T3_N <chr>, T3_method <chr>, T3_Sig <chr>, T3_PV <chr>, T3_cov <chr>,
## #   ...19 <chr>
```

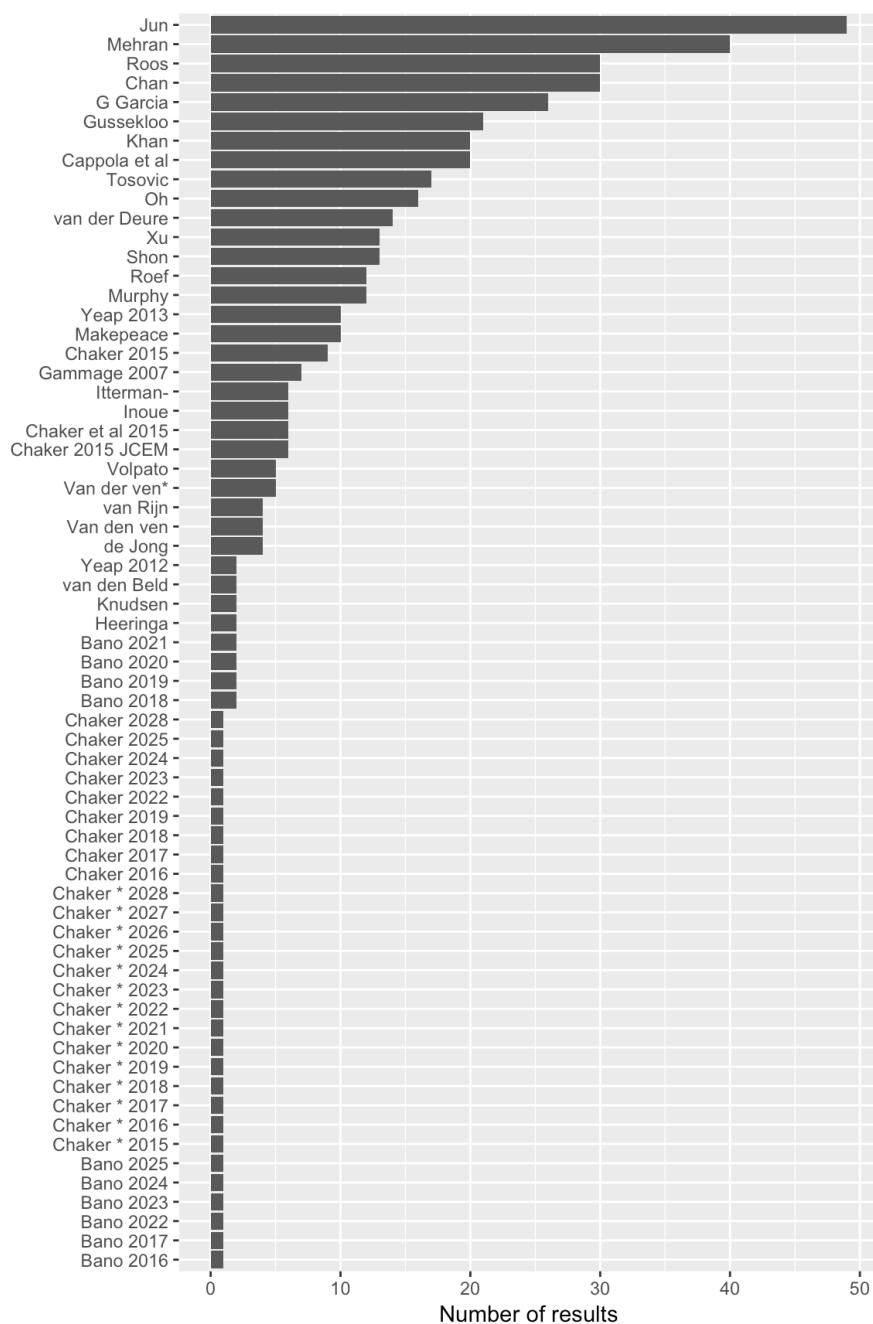
So drop from 509 rows to 460.

3.2 Add an ID

```
TSH$ID <- seq_along(TSH$Paper)
```

3.3 Paper

```
TSH %>%  
  count(Paper) %>%  
  ggplot(aes(fct_reorder(Paper, n), n)) +  
  geom_bar(stat = "identity") +  
  coord_flip() +  
  labs(x = NULL, y = "Number of results")
```



TODO Not sure about Chaker and Chaker *

3.4 Continuous versus quantile

First rename to make easier

```
TSH <-  
  TSH %>%  
    rename(CvQ = `Continuous V quantile`)
```

```
TSH %>%  
  count(CvQ)
```

```
## # A tibble: 20 x 2  
##   CvQ                                n  
##   <chr>                          <int>  
## 1 <NA>                             3  
## 2 Continuous                      378  
## 3 dichotomised                     6  
## 4 not sure of differences here!)    2  
## 5 PERFORMANCE                     8  
## 6 PERFORMANCE CHANGE               8  
## 7 quartile analysis                 2  
## 8 Quartile analysis                 2  
## 9 Quartile analysis XS/1000         1  
## 10 Quartile analysis XS/1001        1  
## 11 Quartile analysis XS/1002        1  
## 12 Quartile analysis XS/1003        1  
## 13 Quartile analysis XS/1004        1  
## 14 Quartile analysis XS/1005        1  
## 15 quartiles                       9  
## 16 Quartiles                      10  
## 17 quintiles                       3  
## 18 Tertile analysis                 3  
## 19 tertiles                        18  
## 20 Tertiles                         2
```

I assume that `quartile analysis` and `Quartile analysis` are the same so collapse, similar process for others.

```
TSH <-  
  TSH %>% mutate(  
    CvQ = fct_recode(  
      CvQ,  
      `Quartile analysis` = "quartile analysis",  
      Quartiles = "quartiles",  
      Tertiles = "tertiles"  
    )  
  )
```

```
TSH %>%
  count(CvQ) %>%
  print(n = 20)
```

```
## Warning: Factor `CvQ` contains implicit NA, consider using
## `forcats::fct_explicit_na`
```

```
## # A tibble: 17 x 2
##   CvQ                                n
##   <fct>                          <int>
## 1 Continuous                        378
## 2 dichotomised                      6
## 3 not sure of differences here!)    2
## 4 PERFORMANCE                      8
## 5 PERFORMANCE CHANGE               8
## 6 Quartile analysis                 4
## 7 Quartile analysis XS/1000         1
## 8 Quartile analysis XS/1001         1
## 9 Quartile analysis XS/1002         1
## 10 Quartile analysis XS/1003        1
## 11 Quartile analysis XS/1004        1
## 12 Quartile analysis XS/1005        1
## 13 Quartiles                       19
## 14 quintiles                       3
## 15 Tertile analysis                3
## 16 Tertiles                        20
## 17 <NA>                            3
```

3.5 Condition

```
TSH %>%
  count(Condition)
```

```
## # A tibble: 237 x 2
##   Condition                                n
##   <chr>                                <int>
## 1 *delta thyroid diab                     6
## 2 abdo fat - female                      1
## 3 abdo fat - male                       1
## 4 ADL                                    2
## 5 AF                                      3
## 6 AF incident                           2
## 7 AF incident normal FT4/TSH, excluding thyroid medication 2
## 8 AF incident, normal FT4/TSH excluding thyroid medication 2
## 9 AF prevalent                          2
## 10 AF prevalent normal TSH/FT4, excluding thyroid medication users 4
## # ... with 227 more rows
```

Lots of mis-spelling etc. Will use the methods of key collision and ngram fingerprints to correct.

This is done by the `refinr` package

<https://cran.r-project.org/web/packages/refinr/vignettes/refinr-vignette.html> (<https://cran.r-project.org/web/packages/refinr/vignettes/refinr-vignette.html>).

These methods come from `openrefine` - see also

<https://github.com/OpenRefine/OpenRefine/wiki/Clustering-In-Depth> (<https://github.com/OpenRefine/OpenRefine/wiki/Clustering-In-Depth>)

```
clean_levels <- function(x, check = FALSE){  
  # Convert to character  
  x <- as.character(x)  
  # Apply key collision and ngram merge  
  x_refin <- x %>%  
    key_collision_merge() %>%  
    n_gram_merge()  
  if(check){  
    tab <-  
    tibble(  
      x, x_refin  
    ) %>%  
    filter(  
      x != x_refin  
    ) %>%  
    knitr::kable()  
    return(tab)  
  }  
  return(x_refin)  
}  
clean_levels(TSH$Condition, check = TRUE)
```

x	x_refin
Mortality	mortality
Hip fracture	Hip Fracture
AF incident, normal FT4/TSH excluding thyroid medication	AF incident normal FT4/TSH, excluding thyroid medication
AF incident, normal FT4/TSH excluding thyroid medication	AF incident normal FT4/TSH, excluding thyroid medication
prevalent AF	AF prevalent
prevalent AF	AF prevalent
incident AF	AF incident
incident AF	AF incident
Breast cancer pre-meno	Breast cancer premeno
Breast cancer post-meno	Breast cancer postmeno

x	x_refin
Breast cancer BMI<25	Breast cancer BMI <25
Breast cancer BMI>25	Breast cancer BMI <25
Breast cancer BMI >25	Breast cancer BMI <25
Breast cancer BMI >25	Breast cancer BMI <25
cognitive decline follow-up	cognitive decline follow up
cognitive decline followup	cognitive decline follow up
syst BP	BP syst
diastol BP	BP diastol
Triglycerides	triglycerides
chol-total	chol -total
chol-LDL	chol- LDL
fatty liver-men U/s+ ALT+	fatty liver-men U/s + ALT-
Fatty liver women U/s+ ALT+	Fatty liver women U/s + ALT-
Fatty liver	fatty liver
Fatty liver	fatty liver
chol-total	chol -total
chol-LDL	chol- LDL
Chol-HDL	chol-HDL
chol-total	chol -total
chol LDL	chol- LDL
delta thyroid diab	*delta thyroid diab
delta thyroid diab	*delta thyroid diab
delta thyroid diab	*delta thyroid diab
delta thyroid diab/female	delta thyroid diab female
delta thyroid diab/male	delta thyroid diab male
chol-LDL - male	chol LDL - male
chol-total - female	chol-total - female
chol-LDL - female	chol LDL - female
total chol	chol -total
chol -LDL	chol- LDL

x	x_refin
Triglycerides	triglycerides
prediabetes to diabetes	prediabetes to diab

So looks reasonable so change

```
TSH <-
  TSH %>% mutate(
    Condition = clean_levels(Condition)
  )
```

NEED A MAPPER

3.6 Condition mapper

First we read in the mapper:

```
condition_mapper <- read_excel("../data/20190324_condition_mapper.xlsx")
condition_mapper %>% kable()
```

condition	system
AF	cardiac
CHD	cardiac
heart failure	cardiac
CVD	cardiac
fracture	orthopaedic
dementia	neurological
mortality	mortality
death	mortality
bone	orthopaedic
lumbar	orthopaedic
neck	orthopaedic
femoral	orthopaedic
cortical	orthopaedic
buckling	orthopaedic
vertebral	orthopaedic
spine	orthopaedic
radius	orthopaedic

condition	system
tibia	orthopaedic
cancer	cancer
ca	cancer
cognitive	neurological
hippocampus	neurological
amygdala	neurological
liver	metabolic
BMI	metabolic
waist	metabolic
BP	cardiac
chol	metabolic
glucose	metabolic
triglycerides	metabolic
thyroid	endocrine
fat	metabolic
HDL	metabolic
diabetes	endocrine
hip	orthopaedic
memory	neurological
depression	neurological
triglyceride	metabolic
insulin	endocrine
cognition	neurological
attention	neurological
processing	neurological

Function to convert condition to system

```

cond_to_system <- function(x, mapper){
  x <- str_to_lower(x)
  mapper$condition <- str_to_lower(mapper$condition)
  systems <- unique(mapper$system)
  n_systems <- length(systems)
  N <- length(x)
  M <- matrix(rep(0, n_systems * N), nc = n_systems, nr = N)
  colnames(M) <- systems
  for(i in 1:n_systems){
    conditions <-
      mapper %>% filter(system == systems[i]) %>%
      pull(condition)
    for(j in conditions){
      M[,i] <- M[,i] + str_detect(x, j)
    }
  }
  M <- pmin(M, 1)
  get_system <- function(row){
    if(sum(row) == 0){
      return("other")
    } else if(sum(row) == 1){
      return(names(row[row == 1]))
    } else {
      return("mixed")
    }
  }
  return(apply(M, 1, get_system))
}

TSH$system <- cond_to_system(TSH$Condition, condition_mapper)
TSH %>%
  filter(system == "other") %>%
  pull(Condition)

```

```
## [1] "cross sectional area"
## [2] "cross sectional area"
## [3] "body BMD"
## [4] "body BMC"
## [5] "ADL"
## [6] "instrumental ADL"
## [7] "ADL"
## [8] "instrumental ADL"
## [9] "frailty"
## [10] "frailty"
## [11] "baseline frailty"
## [12] "follow-up frailty"
## [13] "change in frailty"
## [14] "baseline restricted to participants c followup"
## [15] "uric acid"
## [16] "weight gain"
## [17] "obesity"
## [18] "HbA1c - male"
## [19] "HOMA-IR - male"
## [20] "HbA1c - female"
## [21] "HOMA-IR - female"
## [22] "number of metabolic risk factors"
## [23] "low metabolic risk no worse."
## [24] "low metabolic risk worse"
## [25] "high metabolic risk better"
## [26] "high metabolic risk stable"
## [27] "high metabolic risk worse"
## [28] "Met S - incident"
## [29] "Met S - incident"
## [30] "Met S - incident"
## [31] "HOMA-IR"
## [32] "HOMA-IR"
## [33] "HOMA-IR FT4/low TSH or vice versa"
## [34] "Apo A1"
## [35] "Apo A1"
## [36] "Apo A1"
## [37] "Apo B"
## [38] "Apo B"
## [39] "Apo B"
## [40] "SCD excluding possible SCDs"
## [41] "SCD excluding possible SCDs"
```

3.7 FT4_N

```
table(TSH$FT4_N)
```

##			
##	1007	10163	10206
##	1	2	2
##	10225	1025	1047
##	6	2	1
##	1151	1177	1257
##	14	2	1
##	1278	1338	1375
##	6	3	2
##	1477	1572	1581
##	4	13	30
##	1623	168/185	1741
##	6	1	3
##	1754	1824	1853
##	6	3	7
##	1920	2026	2073
##	3	6	2
##	2136	2139	2215
##	8	8	3
##	2223	2393	2616
##	1	40	15
##	2673	2700 approx	2706
##	3	10	3
##	2803	2843	302/315
##	3	3	1
##	3033	317/329	3376
##	12	1	2
##	3442	3547	3615
##	5	2	12
##	3619	3649	378
##	3	18	1
##	3875	3885	403
##	2	5	2
##	4082	4255	4360
##	2	2	2
##	441	461/469	464
##	2	1	5
##	472	4762	489
##	1	2	2
##	5257	5285	5321
##	6	1	2
##	5345	5365	5403
##	2	3	1
##	5519	558	5698
##	2	20	2
##	578	5812	5860
##	1	2	2
##	6126	6235	629/644
##	1	14	12
##	6416	676cases/680 controls	677
##	3	1	12
##	701	7114	7131

##	1	2	2
##	7409	7740	7746
##	4	4	2
##	8038	8519	8642
##	1	2	4
##	8734	878	882
##	2	13	1
##	8881	9160	9199
##	7	2	1
##	9439	9640	9704
##	2	1	2
##	9742	9882	not stated
##	2	2	8

So some seem to have cases and controls. I will use the total as N, also not stated I will see to NA. Finally approx will be set to the value.

```
clean_n <- function(x, check = FALSE){
  origin_x <- x
  # Get rid of not stated
  x[str_detect(x,"not stated")] <- NA
  # Get rid of approx
  x <- str_replace(x, " approx", "")
  # Get rid of cases and controls
  x <- str_replace(x, "cases", "")
  x <- str_replace(x, "controls", "")
  # Grab enteries of type case/control and add
  case_control <- str_match(x, "(\\d+)/((\\d+))")
  total <- as.numeric(case_control[, 2]) + as.numeric(case_control[, 3])
  index <- which(!is.na(total))
  x[index] <- total[index]
  x <- as.numeric(x)
  # Compare
  if(check){
    tab <-
      tibble(
        origin_x, x
      ) %>%
        filter(origin_x != x) %>%
        knitr::kable()
    return(tab)
  }
  return(x)
}
clean_n(TSH$FT4_N, check = TRUE)
```

origin_x	x
676cases/680 controls	1356
168/185	353
461/469	930

origin_x	x
317/329	646
302/315	617
629/644	1273
629/644	1273
629/644	1273
629/644	1273
629/644	1273
629/644	1273
629/644	1273
629/644	1273
629/644	1273
629/644	1273
629/644	1273
629/644	1273
2700 approx	2700
2700 approx	2700
2700 approx	2700
2700 approx	2700
2700 approx	2700
2700 approx	2700
2700 approx	2700
2700 approx	2700
2700 approx	2700
2700 approx	2700

```
TSH <-
  TSH %>%
  mutate(
    FT4_N = clean_n(FT4_N)
  )
```

```
summary(TSH$FT4_N)
```

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
##	353	1477	2393	3146	3649	10225	8

3.8 FT4 method

```
TSH %>%
  count(FT4_method)
```

```
## # A tibble: 32 x 2
##   FT4_method      n
##   <chr>          <int>
## 1 <NA>            2
## 2 ?fixed effects    6
## 3 adjusted regression 12
## 4 Anova/Tukey        6
## 5 Anova/Tukey/multiple log reg 13
## 6 Anova/X sq          4
## 7 correlation analysis 24
## 8 Cox PH           121
## 9 Cox PH logFT4      10
## 10 Cox PH/log reg     6
## # ... with 22 more rows
```

```
clean_levels(TSH$FT4_method, check = TRUE)
```

[illegible]

x	x_refin
CoxPH	Cox PH
CoxPH	Cox PH
CoxPH	Cox PH
CoxPH*	Cox PH
CoxPH	Cox PH
CoxPH	Cox PH
CoxPH	Cox PH
CoxPH	Cox PH
CoxPH	Cox PH
CoxPH	Cox PH

```
TSH <-
  TSH %>%
  mutate(
    FT4_method = clean_levels(FT4_method)
  )
```

3.9 FT4 sig

```
TSH %>% count(FT4_Sig)
```

```
## # A tibble: 4 x 2
##   FT4_Sig      n
##   <chr>    <int>
## 1 <NA>         5
## 2 N         194
## 3 N*          8
## 4 Y        253
```

Not sure what N* is.

3.10 FT4 PV


```
clean_pv <- function(x, check = FALSE){  
  origin_x <- x  
  # Get rid of not stated  
  x[str_detect(x,"not stated")] <- NA  
  # Get rid of <  
  x <- str_replace(x, "<", "")  
  x <- as.numeric(x)  
  # Compare  
  if(check){  
    cat("PVs removed\n\n")  
    print(origin_x[is.na(x)])  
    tab <-  
    tibble(  
      origin_x, x  
    ) %>%  
    filter(origin_x != x) %>%  
    knitr::kable()  
    return(tab)  
  }  
  return(x)  
}  
clean_pv(TSH$FT4_PV, check = TRUE)
```

```
## Warning in clean_pv(TSH$FT4_PV, check = TRUE): NAs introduced by coercion
```

```

## PVs removed
##
## [1] "11/1000 (5-16.9)" "8/1000 (1.1-15)"
## [3] "7.8/1000 (2-13.6)" "14.3/1000 (6.1-22.5)"
## [5] "0.5/1000 (-2.8-3.8)" "7/1000 -1.4-15)"
## [7] "1.16 (1.11-1.20)" "1.16 (1.11-1.20)"
## [9] "1.20 (1.14-1.27)" "1.21 (1.15-1.28)"
## [11] "1.06 (1.02-1.06)" "1.07 (1.03-1.12)"
## [13] "1.09 (1.03-1.15)" "1.10 (1.04-1.15)"
## [15] "1.08 (1.03-1.14)" "1.08(1.03-1.14)"
## [17] "1.09(1.03-1.15)" "1.87 (1.18-2.96)"
## [19] "1.76 (1.10-2.86)" "1.77 (1.09-2.86)"
## [21] "2.54 (1.48-4.40)" "2.24 (1.31-4.40)"
## [23] "2.26 91.30-3.94)" "not stated"
## [25] "not stated" "not stated"
## [27] "not stated" "not stated"
## [29] "1.32(1.06-1.65)" "1.40(1.10-1.77)"
## [31] "1.47(0.91-2.38)" "1.30(0.99-1.71)"
## [33] "1.24(0.88-1.75)" "1.51(1.07-2.12)"
## [35] "1.88(0.96-1.69)" "2.48(1.12-5.50)"
## [37] "3.99(0.71-22.48)" "2.17(0.88-5.32)"
## [39] "2.82(0.82-9.76)" "2.30(0.80-6.59)"
## [41] "1.11(0.59-2.07)" "1.41(0.88-2.26)"
## [43] "1.34 (0.84-2.15)" "1.72(0.93-317)"
## [45] "1.76(0.95-3.25)" "1.15(0.60-2.21)"
## [47] "1.22(1.00-1.49)" "1.47(0.86-2.51)"
## [49] "1.36(0.80-2.34)" "0.99(0.69-1.43)"
## [51] "0.97(0.63-1.50)" "-0.11 (-0.18;-0.04)"
## [53] "-0.11 ( -0.16;-0.05)" "1.46(1.25-1.69)"
## [55] "1.34(1.15-1.56)" "1.26(1.08-1.47)"
## [57] "1.24(1.06-1.45)" "1.19(1.02-1.39)"
## [59] "1.51(1.29-1.77)" "1.36(1.16--1.59)"
## [61] "1.27(1.08-1.50)" "1.25(1.06-1.47)"
## [63] "1.19(1.02-1.41)" "1.72(1.20-2.48)"
## [65] "1.38(0.98-1.94)" "1.36(0.93-2.00)"
## [67] "0.90(0.60-1.36)" "1.00(0.65-1.55)"
## [69] "1.01(0.66-1.55)" "1.6(0.9-1.3)"
## [71] "1.1(0.8-1.4)" "1.0(0.6-1.8)"
## [73] "1.7(1.0-2.9)" "1.27(1.01-1.60)"
## [75] "1.26(1.05-1.52)" "1.10(0.93-1.32)"
## [77] "0.33(0.22;0.48)" "0.42(0.28;0.63)"
## [79] "0.42(0.24;0.74)" "0.52(0.29;0.92)"
## [81] "0.41(0.09;1.73)" "0.59(0.13;2.59)"
## [83] "not stated" "not stated"
## [85] "not stated" "not stated"
## [87] "not stated" "not stated"
## [89] "not stated" "not stated"
## [91] "not stated" "not stated"
## [93] "not stated" "not stated"
## [95] "not stated" "not stated"
## [97] "not stated" "not stated"
## [99] "not stated" "not stated"

```

## [101]	"not stated"	"not stated"
## [103]	"0.59(0.39-0.9) "	"0.78(0.5-1.2) "
## [105]	"0.82(0.6-1.2) "	"0.44(0.3-0.6) "
## [107]	"0.49(0.3-0.6) "	"1.24(1.02-1.51) "
## [109]	"1.34(1.05-1.7) "	"1.35(1.05-1.7) "
## [111]	"1.05(0.88-1.2) "	"1.09(0.9-1.3) "
## [113]	"1.1(0.88-1.3) "	"0.39(0.28-0.50) "
## [115]	"0.56(0.4-0.7) "	"0.57(0.41-0.7) "
## [117]	"0.85(0.68-1.060) "	"0.84(0.65-1.07) "
## [119]	"0.84(0.66-1.07) "	"not stated"
## [121]	"not stated"	"not stated"
## [123]	"not stated"	"not stated"
## [125]	"not stated"	"not stated"
## [127]	"not stated"	"not stated"
## [129]	"not stated"	"0.96(0.93-0.99) "
## [131]	"0.96(0.93-0.99) "	"0.92(0.89-0.97) "
## [133]	"0.93(0.89-0.98) "	"0.96(0.92-0.99) "
## [135]	"0.94(0.92-0.99) "	"0.90(0.85-0.95) "
## [137]	"0.91(0.86-0.97) "	"0.70(0.53-0.92) "
## [139]	"3.36(1.69-6.66) "	"3.39(1.68-6.81) "
## [141]	"2.20(1.17-4.15) "	"2.02(1.07-3.82) "
## [143]	"3.00(1.44-6.18) "	"2.79(1.34-5.71) "
## [145]	"2.99(1.45-6.17) "	"2.80(1.34-5.82) "
## [147]	"2.42(1.14-5.16) "	"2.25(1.05-4.82) "
## [149]	"2.48(1.00-6.11) "	"2.22(0.95-5.52) "
## [151]	"3.38(1.41-8.14) "	"3.04(1.27-7.03) "
## [153]	"1.1(0.8-1.4) "	"1.0(0.1-7.2) "
## [155]	"1.0(0.8-1.4) "	"1.6(0.9-3.1) "
## [157]	"1.1(0.8-1.6) "	

origin_x	x
< 0.001	0.0010
0.280000000000000003	0.2800
4.00000000000000001E-3	0.0040
8.00000000000000002E-3	0.0080
0.560000000000000005	0.5600
7.00000000000000001E-3	0.0070
<0.001	0.0010
<0.001	0.0010
8.00000000000000002E-3	0.0080
5.00000000000000001E-3	0.0050
<0.001	0.0010
4.00000000000000001E-3	0.0040

[illegible]

origin_x	x
<0.05	0.0500
<0.05	0.0500
<0.05	0.0500
<0.05	0.0500
<0.05	0.0500
0.19800000000000001	0.1980
0.13700000000000001	0.1370
8.3000000000000004E-2	0.0830
2.599999999999999E-2	0.0260
3.5000000000000003E-2	0.0350
8.899999999999996E-2	0.0890
8.999999999999993E-3	0.0090
1.4E-2	0.0140
4.599999999999999E-2	0.0460
0.22500000000000001	0.2250
0.4849999999999999	0.4850
0.32200000000000001	0.3220
0.8259999999999996	0.8260
0.8249999999999996	0.8250
0.55200000000000005	0.5520
0.8289999999999996	0.8290
0.8149999999999995	0.8150
0.52900000000000003	0.5290
0.38300000000000001	0.3830
0.38200000000000001	0.3820
0.4789999999999998	0.4790
0.4749999999999998	0.4750
0.26200000000000001	0.2620
<0.001	0.0010
6.0000000000000001E-3	0.0060

origin_x	x
1E-3	0.0010
8.999999999999993E-3	0.0090
1E-3	0.0010
0.14000000000000001	0.1400
<0.001	0.0010
<0.001	0.0010
<0.0001	0.0001
<0.0001	0.0001
<0.0001	0.0001
<0.001	0.0010
1E-3	0.0010
1.299999999999999E-2	0.0130
3.0000000000000001E-3	0.0030
0.20200000000000001	0.2020
0.2949999999999998	0.2950
<0.001	0.0010
7.0000000000000007E-2	0.0700
0.1429999999999999	0.1430
0.2929999999999998	0.2930
0.55900000000000005	0.5590
1.799999999999999E-2	0.0180
8.999999999999993E-3	0.0090
<0.001	0.0010
0.64400000000000002	0.6440
1E-3	0.0010
4.0000000000000001E-3	0.0040
<0.001	0.0010
4.0000000000000001E-3	0.0040
<0.01	0.0100
4.0000000000000001E-3	0.0040

origin_x	x
5.0000000000000001E-3	0.0050
<0.001	0.0010
<0.001	0.0010
<0.001	0.0010
<0.001	0.0010
<0.001	0.0010
<0.001	0.0010
<0.001	0.0010
<0.01	0.0100
0.7159999999999997	0.7160
0.4889999999999999	0.4890
0.5889999999999997	0.5890
0.8379999999999997	0.8380
<0.001	0.0010
<0.001	0.0010
<0.001	0.0010
4.1000000000000002E-2	0.0410
<0.001	0.0010
1.9E-2	0.0190
<0.001	0.0010
<0.01	0.0100
<0.05	0.0500
<0.001	0.0010
<0.05	0.0500
<0.001	0.0010
<0.001	0.0010
<0.001	0.0010
<0.05	0.0500
<0.05	0.0500
<0.001	0.0010

origin_x	x
<0.05	0.0500
<0.05	0.0500
<0.05	0.0500
<0.001	0.0010
8.0000000000000002E-3	0.0080
<0.001	0.0010
<0.001	0.0010
<0.001	0.0010
<0.001	0.0010
<0.001	0.0010
<0.001	0.0010
<0.001	0.0010
<0.001	0.0010
<0.001	0.0010
<0.001	0.0010
<0.001	0.0010
2.4E-2	0.0240
1.4E-2	0.0140
1.0999999999999999E-2	0.0110
7.0000000000000001E-3	0.0070
6.0000000000000001E-3	0.0060
6.0000000000000001E-3	0.0060
8.6999999999999994E-2	0.0870
9.299999999999999E-2	0.0930
3.4000000000000002E-2	0.0340
1.9E-2	0.0190
<0.001	0.0010
<0.001	0.0010
<0.001	0.0010
8.1000000000000003E-2	0.0810
8.2000000000000003E-2	0.0820

origin_x	x
6.0000000000000001E-3	0.0060
2.9000000000000001E-2	0.0290
<0.001	0.0010
5.299999999999999E-2	0.0530
<0.001	0.0010
8.999999999999993E-3	0.0090
0.1789999999999999	0.1790
<0.001	0.0010
<0.001	0.0010
<0.001	0.0010
<0.001	0.0010
<0.001	0.0010
<0.001	0.0010
<0.0001	0.0001
<0.0001	0.0001
0.7129999999999997	0.7130
0.79300000000000004	0.7930
0.44800000000000001	0.4480
6.0000000000000001E-3	0.0060
5.7000000000000002E-2	0.0570
9.4E-2	0.0940
0.22700000000000001	0.2270
<0.001	0.0010
0.16800000000000001	0.1680
4.0000000000000001E-3	0.0040
0.88300000000000001	0.8830
<0.001	0.0010
3.0000000000000001E-3	0.0030
0.68100000000000005	0.6810
1.7000000000000001E-2	0.0170

origin_x	x
1.4E-2	0.0140
3.699999999999998E-2	0.0370
4.0000000000000001E-3	0.0040
1.299999999999999E-2	0.0130
<0.001	0.0010
7.0000000000000001E-3	0.0070
3.0000000000000001E-3	0.0030
<0.001	0.0010
2.3E-2	0.0230
0.14000000000000001	0.1400
0.6909999999999995	0.6910
0.5869999999999997	0.5870
9.299999999999999E-2	0.0930
<0.001	0.0010
3.799999999999999E-2	0.0380
0.46200000000000002	0.4620
1.9E-2	0.0190
2.3E-2	0.0230
0.3659999999999999	0.3660
3.5000000000000003E-2	0.0350
0.2049999999999999	0.2050
4.599999999999999E-2	0.0460
0.13500000000000001	0.1350

```
TSH <-
  TSH %>%
  mutate(FT4_PV = clean_pv(FT4_PV))
```

```
## Warning in clean_pv(FT4_PV): NAs introduced by coercion
```

Have removed CI, and converted \$< p \$ to p.

```
## Warning: Removed 157 rows containing non-finite values (stat_bin).
```

x	x_refin
Crude	crude
age,sex,smoking	age,sex, smoking
age,sex,smoking	age,sex, smoking
age,sex,smoking	age,sex, smoking
age,sex,smoking	age,sex, smoking
age,sex,smoking	age,sex, smoking
age,sex,smoking	age,sex, smoking
age,sex,smoking	age,sex, smoking
age,sex,smoking	age,sex, smoking
age,sex,smoking	age,sex, smoking
age,sex,smoking	age,sex, smoking
age,sex,smoking	age,sex, smoking
age,sex,interaction	age,sex and interaction
age,sex BMI	age,sex,BMI
age,sex BMI	age,sex,BMI
not stated?	not stated
not stated?	not stated
age,sex BMI	age,sex,BMI
age,sex, smokingplus BMI	age,sex, smoking plus BMI

```
TSH <-
  TSH %>%
  mutate(
    FT4_cov = clean_levels(FT4_cov)
  )
```

3.12 TSH_N

```
clean_n(TSH$TSH_N, check = TRUE)
```

origin_x	x
676/680	1356
169/175	344

origin_x	x
459/469	928
326/328	654
302/316	618
630/646	1276
630/646	1276
630/646	1276
630/646	1276
630/646	1276
630/646	1276
630/646	1276
630/646	1276
630/646	1276
630/646	1276
630/646	1276
630/646	1276
2700 approx	2700
2700 approx	2700
2700 approx	2700
2700 approx	2700
2700 approx	2700
2700 approx	2700
2700 approx	2700
2700 approx	2700
2700 approx	2700
2700 approx	2700

```
TSH <- TSH %>%
  mutate(TSH_N = clean_n(TSH_N))
```

3.13 TSH_method

```
clean_levels(TSH$TSH_method, check = TRUE)
```

[illegible]

```
TSH <-  
  TSH %>%  
  mutate(TSH_method = clean_levels(TSH_method))
```

3.14 TSH_sig

```
TSH %>%  
  count(TSH_Sig)
```

```
## # A tibble: 5 x 2  
##   TSH_Sig      n  
##   <chr>    <int>  
## 1 <NA>         4  
## 2 N          336  
## 3 N*          20  
## 4 Y           99  
## 5 Y*           1
```

3.15 TSH PV

```
clean_pv(TSH$TSH_PV, check = TRUE)
```

```
## Warning in clean_pv(TSH$TSH_PV, check = TRUE): NAs introduced by coercion
```

```

## PVs removed
##
## [1] "not stated" "not stated"
## [3] "not stated" "not stated"
## [5] "not stated" "minus 9.6 (-17.9 to -1.2)"
## [7] "0.93 (0.83-1.05)" "0.94(0.84-1.05)"
## [9] "0.94 (0.74-1.19)" "0.94 (0.74-1.19)"
## [11] "0.94(0.84-1.06)" "0.91 (0.81-1.03)"
## [13] "0.91 (0.073-1.09)" "0.91 (0.73-.09)"
## [15] "not stated" "not stated"
## [17] "not stated" "not stated"
## [19] "not stated" "0.91 (0.80-1.03)"
## [21] "0.91 (0.80-1.03)" "0.92 (0.80-1.04)"
## [23] "0.81 (0.63-1.04)" "0.80 (0.62-1.03)"
## [25] "0.80 (0.62-1.04)" "not stated"
## [27] "not stated" "not stated"
## [29] "not stated" "not stated"
## [31] "not stated" "0.86(0.69-1.08)"
## [33] "0.85(0.67-1.07)" "0.90(0.56-1.43)"
## [35] "0.88(0.67-1.15)" "0.95(0.68-1.37)"
## [37] "0.77(0.55-1.08)" "0.90(0.77-1.06)"
## [39] "0.91(0.76-1.09)" "0.86(0.61-1.22)"
## [41] "0.95(0.77-1.17)" "0.95(0.74-1.25)"
## [43] "0.83(0.64-1.07)" "0.96(0.91-1.02)"
## [45] "0.97(0.94-1.00)" "0.89(0.77-1.03)"
## [47] "0.97(0.85-1.11)" "0.95(0.83-1.09)"
## [49] "1.00(0.86-1.16)" "0.99(0.92-1.07)"
## [51] "1.00(0.89-1.13)" "1.01(0.89-1.14)"
## [53] "0.95(0.89-1.10)" "0.96(0.91-1.01)"
## [55] "0.88(0.75-1.02)" "0.95(0.83-1.10)"
## [57] "0.94(0.81-1.08)" "0.96(0.82-1.13)"
## [59] "0.95(0.88-1.03)" "1.09(0.89-1.15)"
## [61] "1.01 (0.89-1.15)" "1.13(0.51-2.52)"
## [63] "0.77(0.27-2.19)" "0.07(-0.19;0.33)"
## [65] "-0.06 (-0.27;0.16)" "0.96(0.82-1.13)"
## [67] "0.99(0.84-1.16)" "0.97(0.83-1.14)"
## [69] "0.99(0.84-1.16)" "0.85(0.85-1.17)"
## [71] "0.98(0.84-1.16)" "1.00(0.85-1.17)"
## [73] "0.98(0.83-1.16)" "0.99(0.84-1.17)"
## [75] "1.00(0.85-1.18)" "0.94(0.61-1.43)"
## [77] "1.03(0.68-1.57)" "0.99(0.64-1.53)"
## [79] "1.86(1.16-2.98)" "1.39(0.92-2.12)"
## [81] "1.34(0.89-2.01)" "1.1(0.9-1.3)"
## [83] "0.9(0.6-1.3)" "1.8(1.0-3.1)"
## [85] "1.1(0.9--1.5)" "not stated"
## [87] "not stated" "0.66(0.48-0.98)"
## [89] "0.84(0.66-1.07)" "not stated"
## [91] "1.09(1.10;1.19)" "1.07(0.98;1.17)"
## [93] "1.13(1.00;1.27)" "1.08(0.95;1.23)"
## [95] "1.55(1.09;2.20)" "1.49(1.04;2.15)"
## [97] "not stated" NA
## [99] "not stated" "not stated"

```


## [101]	"not stated"	"not stated"
## [103]	"not stated"	"not stated"
## [105]	"not stated"	"not stated"
## [107]	"not stated"	"not stated"
## [109]	"not stated"	NA
## [111]	"not stated"	"not stated"
## [113]	"not stated"	"not stated"
## [115]	"not stated"	"not stated"
## [117]	"not stated"	"not stated"
## [119]	"not stated"	"not stated"
## [121]	"not stated"	"not stated"
## [123]	"not stated"	"not stated"
## [125]	"not stated"	"not stated"
## [127]	"not stated"	"not stated"
## [129]	"0.99(0.99-1.0002) "	"0.99(0.99-0.999) "
## [131]	"0.99(0.99-0.9999) "	"1(0.99-1.02) "
## [133]	"1.01(0.99-1.02) "	"0.99(0.96-1.03) "
## [135]	"0.99(0.99-1) "	"0.99(0.99-1) "
## [137]	"0.97(0.94-1.01) "	"0.97(0.93-1.01) "
## [139]	"0.97(0.93-1.01) "	"1(0.99-1.01) "
## [141]	"1(0.99-1.03) "	"1(0.99-1) "
## [143]	"1(0.99-1) "	"1(0.99-1.01) "
## [145]	"1(0.99-1.01) "	"not stated"
## [147]	"not stated"	"not stated"
## [149]	"not stated"	"not stated"
## [151]	"not stated"	"not stated"
## [153]	"not stated"	"not stated"
## [155]	"not stated"	"not stated"
## [157]	"1.09(1.06-1.12) "	"1.06(1.00-1.13) "
## [159]	"1.17(1.07-1.27) "	"1.13(1.03-1.24) "
## [161]	"1.16(1.04-1.30) "	"1.14(1.02-1.27) "
## [163]	"1.26(1.08-1.47) "	"1.21(1.04-1.41) "
## [165]	"1.37(1.05-1.78) "	"0.74(0.53-1.04) "
## [167]	"0.74(0.53-1.03) "	"0.85(0.65-1.21) "
## [169]	"0.84(0.64-1.11) "	"0.83(0.64-1.080" "
## [171]	"0.83(0.64-1.08) "	"0.83((0.64-1.08) "
## [173]	"0.83(0.64-1.08) "	"0.81(0.62--1.06) "
## [175]	"0.81(0.62-1.06) "	"0.80(0.58-1.09) "
## [177]	"0.80(0.58-1.10) "	"0.91(0.65-1.28) "
## [179]	"0.91(0.65-1.28) "	"0.9(0.6-1.3) "
## [181]	"0.9(0.5-1.4) "	"1.4(0.6-3.3) "
## [183]	"1.1(0.9-1.4) "	"0.2(0.1-0.7) "

origin_x

x

7.0000000000000007E-2

0.0700

7.0000000000000007E-2

0.0700

7.0000000000000007E-2

0.0700

0.28599999999999998

0.2860

0.55600000000000005

0.5560

origin_x	x
0.91600000000000004	0.9160
6.5000000000000002E-2	0.0650
0.28599999999999998	0.2860
0.77500000000000002	0.7750
0.28599999999999998	0.2860
0.16700000000000001	0.1670
0.16700000000000001	0.1670
2.8000000000000001E-2	0.0280
0.28999999999999998	0.2900
0.55000000000000004	0.5500
0.28000000000000003	0.2800
<0.05	0.0500
<0.05	0.0500
0.34899999999999998	0.3490
0.57599999999999996	0.5760
0.39200000000000002	0.3920
0.54300000000000004	0.5430
1E-3	0.0010
5.0000000000000001E-3	0.0050
1E-3	0.0010
0.28599999999999998	0.2860
0.52600000000000002	0.5260
0.61899999999999999	0.6190
0.29799999999999999	0.2980
0.53200000000000003	0.5320
0.52800000000000002	0.5280
0.70199999999999996	0.7020
0.54600000000000004	0.5460
0.33500000000000002	0.3350
0.71199999999999997	0.7120

origin_x	x
0.544000000000000004	0.5440
0.29099999999999998	0.2910
0.61299999999999999	0.6130
0.898000000000000002	0.8980
0.643000000000000002	0.6430
0.74399999999999999	0.7440
0.914000000000000003	0.9140
0.70499999999999996	0.7050
5.00000000000000001E-3	0.0050
8.9999999999999993E-3	0.0090
0.550000000000000004	0.5500
0.533000000000000003	0.5330
3.00000000000000001E-3	0.0030
<0.0001	0.0001
1.0999999999999999E-2	0.0110
2.5999999999999999E-2	0.0260
6.3E-2	0.0630
0.468000000000000003	0.4680
0.802000000000000005	0.8020
<0.001	0.0010
0.72499999999999998	0.7250
4.2999999999999997E-2	0.0430
0.98699999999999999	0.9870
2.10000000000000001E-2	0.0210
0.384000000000000001	0.3840
0.928000000000000005	0.9280
6.90000000000000006E-2	0.0690
0.509000000000000001	0.5090
0.59399999999999997	0.5940
<0.001	0.0010

[illegible]

origin_x	x
<0.001	0.0010
<0.001	0.0010
<0.001	0.0010
<0.001	0.0010
3.0000000000000001E-3	0.0030
8.0000000000000002E-3	0.0080
5.0000000000000001E-3	0.0050
<0.001	0.0010
<0.001	0.0010
0.5799999999999996	0.5800
<0.001	0.0010
3.0000000000000001E-3	0.0030
<0.001	0.0010
<0.001	0.0010
1.499999999999999E-2	0.0150
0.1739999999999999	0.1740
0.7750000000000002	0.7750
1.6E-2	0.0160
4.599999999999999E-2	0.0460
<0.001	0.0010
<0.0001	0.0001
<0.0001	0.0001
0.7439999999999999	0.7440
0.9320000000000005	0.9320
0.6969999999999995	0.6970
0.5210000000000002	0.5210
1.4E-2	0.0140
0.6490000000000002	0.6490
1E-3	0.0010
1E-3	0.0010

origin_x	x
<0.001	0.0010
<0.001	0.0010
6.0000000000000001E-3	0.0060
0.65800000000000003	0.6580
0.6989999999999995	0.6990
0.5939999999999997	0.5940
0.40200000000000002	0.4020
1.6E-2	0.0160
0.4109999999999998	0.4110
9.8000000000000004E-2	0.0980
8.0000000000000002E-3	0.0080
<0.001	0.0010
2E-3	0.0020
1E-3	0.0010
2.599999999999999E-2	0.0260
8.999999999999993E-3	0.0090
0.7139999999999997	0.7140
0.75800000000000001	0.7580
0.9429999999999995	0.9430
7.9000000000000001E-2	0.0790
0.9929999999999999	0.9930
0.16200000000000001	0.1620
0.39600000000000002	0.3960
0.80300000000000005	0.8030
0.3509999999999998	0.3510
0.63400000000000001	0.6340
7.499999999999997E-2	0.0750
0.5809999999999996	0.5810

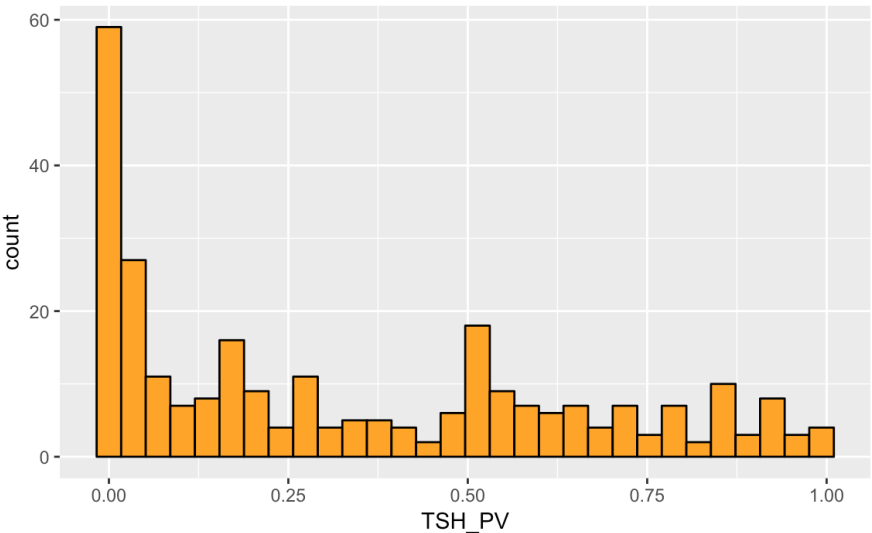
```
TSH <-
  TSH %>%
  mutate(
    TSH_PV = clean_pv(TSH_PV)
  )
```

```
## Warning in clean_pv(TSH_PV): NAs introduced by coercion
```

```
TSH %>%
  ggplot(aes(TSH_PV)) + geom_histogram(col = "black", fill = "orange")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 184 rows containing non-finite values (stat_bin).
```



```
## TSH cov
```

```
clean_levels(TSH$TSH_cov, check = TRUE)
```

x	x_refin
age, sex	age,sex
age, sex	age,sex
age, sex	age,sex
age, sex	age,sex
age, sex	age,sex
age, sex	age,sex
age BMI smoking	age,smoking,BMI
age BMI smoking	age,smoking,BMI

x	x_refin
age BMI smoking	age,smoking,BMI
age BMI smoking	age,smoking,BMI
age BMI smoking	age,smoking,BMI
age BMI smoking	age,smoking,BMI
Crude	crude
age,sex,smoking	age,sex, smoking
age,sex,smoking	age,sex, smoking
age,sex,smoking	age,sex, smoking
age,sex,smoking	age,sex, smoking
age,sex,smoking	age,sex, smoking
age,sex,smoking	age,sex, smoking
age,sex,smoking	age,sex, smoking
age,sex,smoking	age,sex, smoking
age,sex,smoking	age,sex, smoking
age,sex,smoking	age,sex, smoking
age,sex,smoking	age,sex, smoking
age,sex,smoking	age,sex, smoking
age,sex,interaction	age,sex and interaction
age,sex BMI	age,sex,BMI
age,sex BMI	age,sex,BMI
not stated?	not stated
not stated?	not stated
age,sex BMI	age,sex,BMI
age,sex, smokingplus BMI	age,sex, smoking plus BMI

```
TSH <-
  TSH %>%
  mutate(
    TSH_cov = clean_levels(TSH_cov)
  )
```

3.16 T3 N

```
clean_n(TSH$T3_N, check = TRUE)
```


origin_x	x
676cases/680controls	1356
168/175	343
461/469	930
327/329	656
302/315	617
640/644	1284
640/644	1284
640/644	1284
640/644	1284
640/644	1284
640/644	1284
640/644	1284
640/644	1284
640/644	1284
640/644	1284
640/644	1284
640/644	1284

```
TSH <-
  TSH %>%
  mutate(
    T3_N = clean_n(T3_N)
  )
```

3.17 T3 method

```
clean_levels(TSH$T3_method, check = TRUE)
```

x	x_refin
Cox PH*	Cox PH
CoxPH	Cox PH
CoxPH	Cox PH
CoxPH	Cox PH

x	x_refin
CoxPH	Cox PH
CoxPH	Cox PH
CoxPH	Cox PH

```
TSH <-
  TSH %>%
  mutate(
    T3_method = clean_levels(T3_method)
  )
```

3.18 T3 sig

```
TSH %>%
  count(T3_Sig)
```

```
## # A tibble: 5 x 2
##   T3_Sig      n
##   <chr>    <int>
## 1 <NA>    287
## 2 N       81
## 3 N*       3
## 4 not stated 1
## 5 Y      88
```

Convert not stated to NA

```
TSH$T3_Sig[TSH$T3_Sig == "not stated"] <- NA
```

```
TSH %>%
  count(T3_Sig)
```

```
## # A tibble: 4 x 2
##   T3_Sig      n
##   <chr>    <int>
## 1 <NA>    288
## 2 N       81
## 3 N*       3
## 4 Y      88
```

3.19 T3 PV

```
clean_pv(TSH$T3_PV, check = TRUE)
```

```
## Warning in clean_pv(TSH$T3_PV, check = TRUE): NAs introduced by coercion
```

PVs removed

##

##	[1] NA	NA	NA
##	[4] NA	NA	NA
##	[7] NA	NA	NA
##	[10] NA	NA	NA
##	[13] NA	NA	NA
##	[16] NA	NA	NA
##	[19] NA	NA	NA
##	[22] NA	NA	NA
##	[25] NA	NA	NA
##	[28] NA	NA	NA
##	[31] NA	NA	NA
##	[34] NA	NA	NA
##	[37] NA	NA	NA
##	[40] NA	NA	NA
##	[43] NA	NA	NA
##	[46] NA	NA	NA
##	[49] NA	NA	NA
##	[52] NA	NA	NA
##	[55] "not stated"	"not stated"	"not stated"
##	[58] "not stated"	"not stated"	"0.95(0.76-1.26)"
##	[61] "1.00(0.791.26)"	"1.07(0.67-1.71)"	"0.91(0.69-1.19)"
##	[64] "0.77(.55-1.08)"	"1.10(.79-1.54)"	"0.61(.20-1.88)"
##	[67] "0.68(.19-2.60)"	"0.62(0.95--1.86)"	"0.45(0.99-2.02)???"
##	[70] "0.75(.46-1.24)"	"0.87(.15-4.91)"	NA
##	[73] NA	NA	NA
##	[76] NA	NA	NA
##	[79] NA	NA	NA
##	[82] NA	NA	NA
##	[85] NA	NA	NA
##	[88] NA	NA	NA
##	[91] NA	NA	NA
##	[94] NA	NA	NA
##	[97] NA	NA	NA
##	[100] NA	NA	NA
##	[103] NA	NA	NA
##	[106] NA	NA	NA
##	[109] NA	NA	NA
##	[112] NA	NA	NA
##	[115] NA	NA	NA
##	[118] NA	NA	NA
##	[121] NA	NA	NA
##	[124] NA	NA	NA
##	[127] NA	NA	NA
##	[130] NA	NA	NA
##	[133] NA	NA	NA
##	[136] NA	NA	NA
##	[139] NA	NA	NA
##	[142] NA	NA	NA
##	[145] NA	NA	NA
##	[148] NA	NA	NA

## [151] "not stated"	"not stated"	NA
## [154] NA	NA	NA
## [157] NA	NA	NA
## [160] NA	NA	NA
## [163] NA	NA	NA
## [166] NA	NA	"not stated"
## [169] "not stated"	NA	NA
## [172] NA	NA	NA
## [175] NA	"not stated"	"not stated"
## [178] NA	NA	NA
## [181] NA	NA	NA
## [184] NA	NA	NA
## [187] NA	NA	NA
## [190] NA	NA	NA
## [193] NA	NA	NA
## [196] NA	NA	NA
## [199] NA	NA	"not stated"
## [202] NA	NA	NA
## [205] NA	NA	NA
## [208] NA	NA	NA
## [211] NA	NA	NA
## [214] NA	NA	NA
## [217] NA	NA	NA
## [220] NA	NA	NA
## [223] NA	NA	NA
## [226] NA	NA	NA
## [229] NA	NA	NA
## [232] NA	NA	NA
## [235] NA	NA	NA
## [238] NA	NA	NA
## [241] NA	NA	NA
## [244] NA	NA	NA
## [247] NA	NA	NA
## [250] NA	NA	NA
## [253] NA	NA	NA
## [256] NA	NA	NA
## [259] NA	NA	NA
## [262] NA	NA	NA
## [265] NA	NA	NA
## [268] NA	NA	NA
## [271] NA	NA	NA
## [274] NA	NA	NA
## [277] NA	NA	NA
## [280] NA	NA	NA
## [283] NA	NA	NA
## [286] NA	NA	NA
## [289] NA	NA	NA
## [292] NA	NA	NA
## [295] NA	NA	NA
## [298] NA	NA	NA
## [301] NA	NA	NA
## [304] NA	NA	NA

[307] NA

NA

NA

[310] NA

NA

origin_x	x
0.14299999999999999	0.143
0.48899999999999999	0.489
8.8999999999999996E-2	0.089
5.0000000000000001E-3	0.005
1E-3	0.001
0.14199999999999999	0.142
0.40699999999999997	0.407
0.79400000000000004	0.794
0.68500000000000005	0.685
2E-3	0.002
6.0000000000000001E-3	0.006
8.0000000000000002E-3	0.008
7.0000000000000001E-3	0.007
6.0000000000000001E-3	0.006
<0.01	0.010
<0.001	0.001
<0.001	0.001
<0.001	0.001
5.0000000000000001E-3	0.005
3.0000000000000001E-3	0.003
<0.001	0.001
0.57999999999999996	0.580
8.0000000000000002E-3	0.008
0.55900000000000005	0.559
4.1000000000000002E-2	0.041
8.6999999999999994E-2	0.087
0.40799999999999997	0.408
4.599999999999999E-2	0.046

origin_x	x
5.0000000000000001E-3	0.005
0.18099999999999999	0.181
5.299999999999999E-2	0.053
4.0000000000000001E-3	0.004
0.52700000000000002	0.527
0.90300000000000002	0.903
0.8269999999999996	0.827
4.299999999999997E-2	0.043
0.36799999999999999	0.368
0.25800000000000001	0.258
0.79100000000000004	0.791
4.8000000000000001E-2	0.048
<0.001	0.001
6.0000000000000001E-3	0.006
0.28599999999999998	0.286
<0.001	0.001
<0.001	0.001
2.7E-2	0.027
<0.001	0.001
2.3E-2	0.023
1.9E-2	0.019
<0.01	0.010
<0.05	0.050
<0.05	0.050
<0.001	0.001
<0.05	0.050
<0.001	0.001
<0.001	0.001
<0.001	0.001
<0.001	0.001

origin_x	x
<0.001	0.001
<0.001	0.001
<0.001	0.001
<0.001	0.001
<0.001	0.001
<0.001	0.001
<0.001	0.001
<0.001	0.001
<0.001	0.001
<0.001	0.001
<0.001	0.001
<0.001	0.001
<0.001	0.001
<0.001	0.001
<0.001	0.001
<0.001	0.001
<0.001	0.001
<0.001	0.001
2.1000000000000001E-2	0.021
1.7000000000000001E-2	0.017
1.4999999999999999E-2	0.015
4.0000000000000001E-3	0.004
<0.001	0.001
2E-3	0.002
<0.001	0.001
<0.001	0.001
<0.01	0.010
<0.001	0.001
<0.001	0.001
1E-3	0.001
1E-3	0.001

origin_x	x
<0.001	0.001
<0.001	0.001
<0.001	0.001
0.5929999999999997	0.593
7.8E-2	0.078
8.599999999999993E-2	0.086
3.599999999999997E-2	0.036
8.999999999999993E-3	0.009
8.0000000000000002E-3	0.008
5.899999999999997E-2	0.059
6.7000000000000004E-2	0.067
<0.001	0.001
<0.001	0.001
<0.001	0.001
0.865999999999999	0.866
0.26900000000000002	0.269
4.299999999999997E-2	0.043
2.5000000000000001E-2	0.025
7.199999999999995E-2	0.072
8.599999999999993E-2	0.086
0.23300000000000001	0.233
0.32600000000000001	0.326

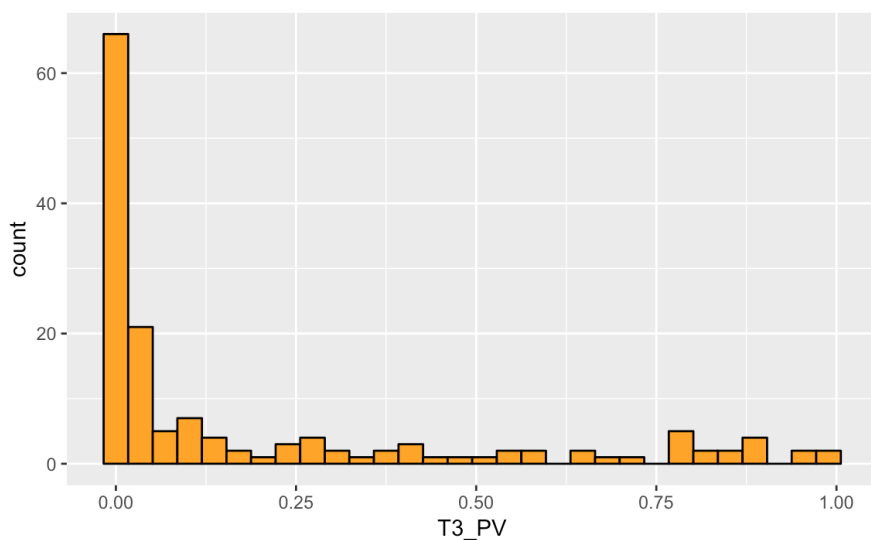
```
TSH <-
  TSH %>%
  mutate(
    T3_PV = clean_pv(T3_PV)
  )
```

```
## Warning in clean_pv(T3_PV): NAs introduced by coercion
```

```
TSH %>%
  ggplot(aes(T3_PV)) + geom_histogram(col = "black", fill = "orange")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 311 rows containing non-finite values (stat_bin).
```



3.20 T3 cov

```
clean_levels(TSH$T3_cov, check = TRUE)
```

x	x_refin
Crude	crude
not stated?	not stated

```
TSH <-  
  TSH %>%  
  mutate(  
    T3_cov = clean_levels(T3_cov)  
  )
```

3.21 Notes

The final column seems to be a comments section, will rename as such.

```
TSH <-  
  TSH %>%  
  rename(  
    comments = ...19  
  )
```

4 Save clean data

```
write_rds(TSH, glue("../data/{get_label()}"))
```