Comparative Effectiveness and Personalized Medicine Research Using Real-World Data

Thomas Debray

2023-05-03

Table of contents

1 Preface

Thomas Debray (Smart Data Analysis and Statistics B.V.)

About this book

This book provides practical guidance for estimating the effectiveness of treatments in real-world populations. It explains how real-world data can directly be used or combined with other data sources to derive overall and individualized estimates of treatment effect. The book explains statistical methods for implementing bias adjustments, conducting evidence synthesis and individualizing treatment effect, whilst also providing illustrative examples and supporting software. The chapters and contents of the book are written by leading experts, with a track record in the generation and/or evaluation of real-world evidence.

This book is intended as a pivotal textbook for statisticians, epidemiologists, methodologists, regulators and/or regulatory scientists considering, undertaking or appraising the real-world evidence of treatment effectiveness. It covers key concepts and stages to derive and evaluate treatment effect estimates for entire populations and specific individuals. The book offers a conceptual framework towards estimating treatment effects at both the population and individualized level, where modelling methods may include traditional regression-based and machine learning methods.

Motivation

Although randomized clinical trials traditionally form the cornerstone of comparative effectiveness research, there is a growing demand to consider evidence from "real-world data" (RWD) in clinical decision-making. These data are often available from observational cohort studies, administrative databases, and patient registries, and may offer additional insights into the comparative effectiveness and safety of treatments. Yet, the analysis of RWD and the evaluation of real-world evidence face many operational and methodological challenges.

In this book, we aim to address three current needs. First, this book will offer the guidance that is currently lacking on assessing the quality of RWD and on implementing appropriate

statistical methods to reduce bias of single study estimates of treatment effects. Second, this book will provide researchers with advanced approaches to pooling estimates from multiple non-randomized studies for which traditional evidence synthesis methods are not suitable. Finally, to answer the growing need to translate average estimates of treatment effects to individualized clinical decision-making, this book will present recent methods for more tailored approaches where patient characteristics are used to derive their individualized prognosis and treatment benefit.

This book aims to explain key principles and state-of-the-art methods for deriving treatment effects in entire populations and specific individuals using RWD. It will not only discuss statistical theory by key experts in the field; it will also provide illustrative examples and practical guidance for implementation in R. In short, the book aims to prepare a new generation of researchers who wish to generate and integrate evidence from both randomized and non-randomized data sources to investigate the real-world effectiveness of treatments in populations and individual patients.

Contents

The book is divided into six sections:

1. **Introduction**. This section introduces the relevance of real-world data for conducting comparative effectiveness research, and discusses various concerns regarding their use.
2. **Principles of treatment effect estimation using real-world data**. In this section, we discuss key principles of treatment effect estimation in non-randomized data sources. We explain methods to adjust for confounding (including propensity score analysis and disease risk score analysis) and missing data when estimating the treatment effect for a specific (sub)population.
3. **Principles of evidence synthesis**. In this section, we discuss statistical methods for estimating the treatment effect using (individual participant and/or aggregate) data from multiple studies. To this purpose, key principles of meta-analysis are introduced and explained, including the standard fixed effect and random effects meta-analysis models, methods for individual patient data (IPD) meta-analysis, methods for network meta-analysis, and methods for data-driven and tailored bias adjustment.
4. **Advanced modelling issues for dealing with additional bias in both randomized and non-randomized data sources**. In this section, we discuss advanced statistical and machine learning methods for dealing with time-varying confounding, informative visit schedules, and measurement error.
5. **Individualizing treatment effects for personalized medicine**. In this section, we discuss statistical methods to estimate and evaluate individualized treatment effects.
6. Closing

2 Validity control and quality assessment of real-world data and real-world evidence

Christina ReadThomas Debray (Smart Data Analysis and Statistics B.V.)

```
library(readxl)
library(robvis)
```

The quality of real-world data is often suboptimal and can therefore lead to bias when generating real-world evidence (RWE). In this chapter, we will introduce key quality concerns of RWD, including their accuracy, completeness, and timeliness. Subsequently, we will discuss which steps can be taken to assess the quality of RWD, and determine their fitness for use. The chapter will also introduce directed acyclic graphs to explain how the analysis of RWD may be affected by different types of bias. We will put particular focus on confounding bias, selection bias, and information bias, and explain how these biases can be addressed by referring to specific chapters from the book. Finally, the chapter presents common quality appraisal tools that can be used to assess the quality of real-world evidence (for instance when conducting a systematic review).

2.1 Example code

A risk of bias assessment was conducted in the COVID-NMA review. We can create a summary table of risk of bias assessment and produce a traffic light plot as follows:

```
Risk_of_Bias <- read_excel("resources/RoB-covid.xlsx")

#creation of traffic light plot
trafficlight_rob <- rob_traffic_light(data = Risk_of_Bias, tool = "ROB2")
trafficlight_rob
```


Risk of bias domains
D1
D2
D3
D4
D5
Overall
Study
Abd−Elsalam et al
Ader
Beigel (ACTT−1)
Cao et al
Cavalcanti et al
Criner*
Davoudi−Monfared
Horby (RECOVERY; DEXA)
Horby (RECOVERY; LOPI)
Horby (RECOVERY; TOCI)
Islam^
Lyngbakken et al
Mahajan
Pan (WHO Solidarity; HCQ)
Pan (WHO Solidarity; INTB)
Pan (WHO Solidarity; LOPI)
Pan (WHO Solidarity; REM)
Spinner
Wang
Broman
Declerq
Gordon (REMAP−CAP)
Hermine
Rohmani
Rosas (COVACTA)
Rutgers
Salama(EMPACTA)
Salvarani*
Soin (COVINTOC)
Stone*
Tomazini
Talaschian
Ulrich et al
Veiga
Domains:
D1: Bias arising from the randomization process.
D2: Bias due to deviations from intended intervention.
D3: Bias due to missing outcome data.
D4: Bias in measurement of the outcome.
D5: Bias in selection of the reported result.
Judgement
High
Some concerns
Low

Version info

This chapter was rendered using the following version of R and its packages:

```
R version 4.2.3 (2023-03-15 ucrt)
Platform: x86_64-w64-mingw32/x64 (64-bit)
Running under: Windows 10 x64 (build 19045)

Matrix products: default

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

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

other attached packages:
[1] robvis_0.3.0.900 readxl_1.4.2

loaded via a namespace (and not attached):
 [1] cellranger_1.1.0 pillar_1.9.0     compiler_4.2.3   tools_4.2.3
 [5] digest_0.6.31    jsonlite_1.8.5   evaluate_0.21    lifecycle_1.0.3
 [9] tibble_3.2.1     gtable_0.3.3     pkgconfig_2.0.3  rlang_1.1.1
[13] cli_3.6.1        rstudioapi_0.14  yaml_2.3.7       xfun_0.39
[17] fastmap_1.1.1    withr_2.5.0      stringr_1.5.0    dplyr_1.1.2
[21] knitr_1.43       generics_0.1.3   vctrs_0.6.3      grid_4.2.3
[25] tidyselect_1.2.0 glue_1.6.2       R6_2.5.1         fansi_1.0.4
[29] rmarkdown_2.22   farver_2.1.1     tidyr_1.3.0      purrr_1.0.1
[33] ggplot2_3.4.2    magrittr_2.0.3   scales_1.2.1     codetools_0.2-19
[37] htmltools_0.5.5  colorspace_2.1-0 utf8_1.2.3       stringi_1.7.12
[41] munsell_0.5.0
```

3 Confounding adjustment using propensity score methods

Tammy Jiang (Biogen)
Thomas Debray (Smart Data Analysis and Statistics B.V.)

3.1 Introduction

The purpose of this document is to provide example R code that demonstrates how to estimate the propensity score and implement matching, stratification, weighting, and regression adjustment for the continuous propensity score. In this example using simulated data, we have two disease modifying therapies (DMT1 and DMT0) and the outcome is the number of post-treatment multiple sclerosis relapses during follow-up. We will estimate the average treatment effect in the treated (ATT) using propensity score matching, stratification, and weighting. We will estimate the average treatment effect in the population (ATE) using regression adjustment for the continuous propensity score. The treatment effects can be interpreted as annualized relapse rate ratios (ARR).

We consider an example dataset with the following characteristics:

```
head(dat)
```

```
   age female prevDMTefficacy premedicalcost numSymptoms prerelapse_num
1:  50      1            None        3899.61           1              1
2:  51      0            None        9580.51           1              0
3:  56      0            None        4785.89           1              0
4:  44      1            None        8696.80           1              1
5:  63      0            None        2588.03           1              0
6:  28      1            None        5435.57           1              0
   treatment y      years      Iscore
1:      DMT1 0 1.78507871 Moderate A1
2:      DMT1 0 0.01368925     High A1
```

```
3:      DMT1 2 3.25530459     High A1
4:      DMT1 2 5.73853525     Neutral
5:      DMT1 0 1.31143053     High A1
6:      DMT1 0 0.59137577 Moderate A0
```

3.2 Comparing baseline characteristics

- `DMT1` is the treatment group and `DMT0` is the control group
- `prevDMTefficacy` is previous DMT efficacy (none, low efficacy, and medium/high efficacy)
- `prerelapse_num` is the number of previous MS relapses

	DMT0	DMT1
n	2300	7700
age (mean (SD))	51.39 (8.32)	44.25 (9.79)
female = 1 (%)	1671 (72.65)	5915 (76.82)
prevDMTefficacy (%)		
None	1247 (54.22)	3171 (41.18)
Low_efficacy	261 (11.35)	858 (11.14)
Medium_high_efficacy	792 (34.43)	3671 (47.68)
prerelapse_num (mean (SD))	0.39 (0.62)	0.46 (0.68)

3.3 Estimating the propensity score

3.3.1 Logistic regression

We sought to restore balance in the distribution of baseline covariates in patients treated with DMT1 (index treatment) and DMT0 (control tratment). We fit a multivariable logistic regression model in which treatment was regressed on baseline characteristics including age, sex, previous DMT efficacy, and previous number of relapses.

```r
# Fit logistic regression model
ps.model <- glm(treatment ~ age + female + prevDMTefficacy + prerelapse_num,
                data = dat, family = binomial())

# Summary of logistic regression model
summary(ps.model)
```

```
Call:
glm(formula = treatment ~ age + female + prevDMTefficacy + prerelapse_num,
    family = binomial(), data = dat)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.7949   0.2585   0.5220   0.7478   1.5033

Coefficients:
                                   Estimate Std. Error z value Pr(>|z|)
(Intercept)                        4.809473   0.157127  30.609  < 2e-16 ***
age                               -0.086708   0.002996 -28.939  < 2e-16 ***
female1                            0.253611   0.057664   4.398 1.09e-05 ***
prevDMTefficacyLow_efficacy        0.310394   0.083022   3.739 0.000185 ***
prevDMTefficacyMedium_high_efficacy 0.660266   0.054393  12.139  < 2e-16 ***
prerelapse_num                     0.156318   0.039288   3.979 6.93e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 10786  on 9999  degrees of freedom
Residual deviance:  9597  on 9994  degrees of freedom
AIC: 9609

Number of Fisher Scoring iterations: 5
```

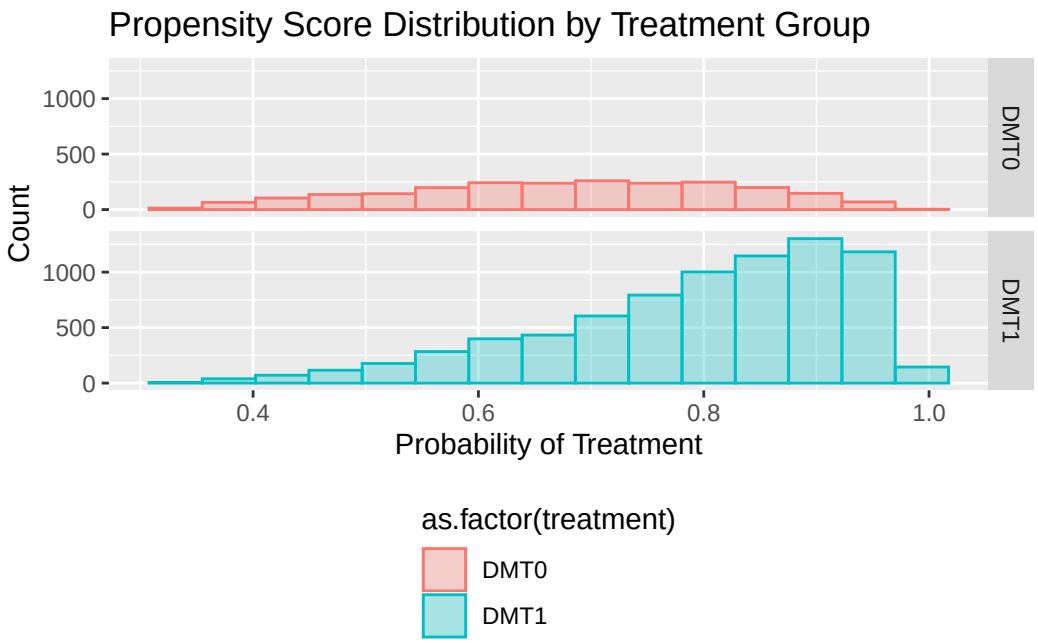
```r
# Extract propensity scores
dat$ps <- predict(ps.model, data = dat, type = "response")
```

3.3.2 Assessing overlap

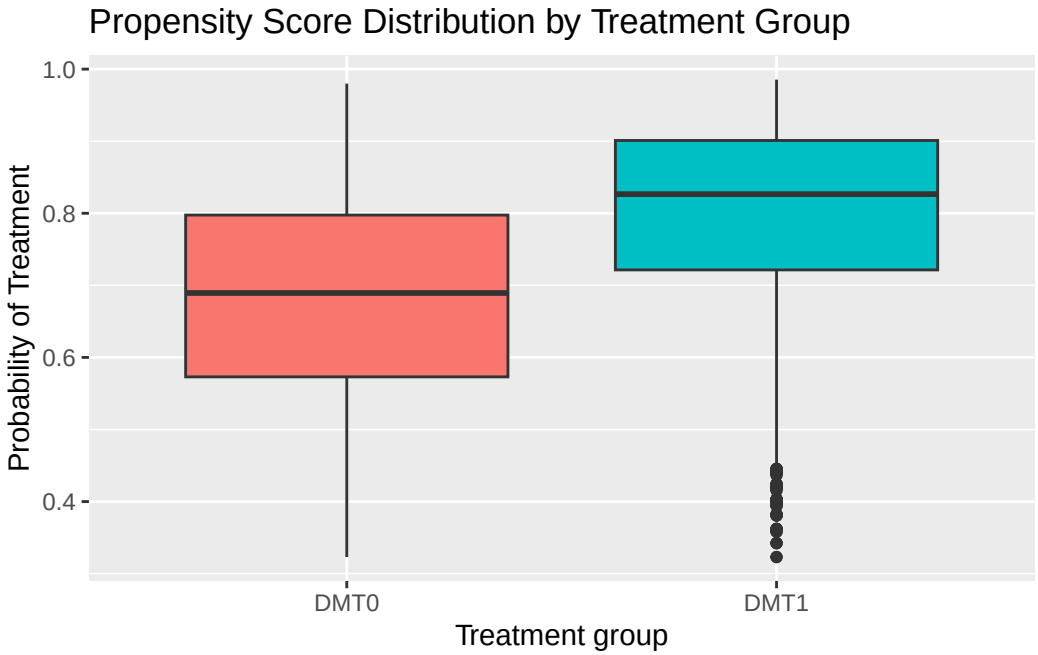
We examined the degree of overlap in the distribution of propensity scores across treatment groups using histograms and side-by-side box plots.

```r
# Histogram
ggplot(dat, aes(x = ps, fill = as.factor(treatment), color = as.factor(treatment))) +
  geom_histogram(alpha = 0.3, position='identity', bins = 15) +
  facet_grid(as.factor(treatment) ~ .) +
```

```
  xlab("Probability of Treatment") +
  ylab("Count") +
  ggtitle("Propensity Score Distribution by Treatment Group") +
  theme(legend.position = "bottom", legend.direction = "vertical")
```




```
# Side-by-side box plots
ggplot(dat, aes(x=as.factor(treatment), y=ps, fill=as.factor(treatment))) +
  geom_boxplot() +
  ggtitle("Propensity Score Distribution by Treatment Group") +
  ylab("Probability of Treatment") +
  xlab("Treatment group") +
  theme(legend.position = "none")
```




```r
# Distribution of propensity scores by treatment groups
summary(dat$ps[dat$treatment == "DMT1"])
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.3230  0.7214  0.8265  0.7970  0.9010  0.9854
```

```r
summary(dat$ps[dat$treatment == "DMT0"])
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.3230  0.5730  0.6894  0.6795  0.7975  0.9799
```

3.4 Propensity score matching

3.4.1 1:1 Optimal full matching without replacement

```
library(MatchIt)

# Use MatchIt package for PS matching
opt <- matchit(treatment ~ age + female + prevDMTefficacy + prerelapse_num,
               data = dat,
               method = "full",
               estimand = "ATT")

opt
```

```
A matchit object
 - method: Optimal full matching
 - distance: Propensity score
             - estimated with logistic regression
 - number of obs.: 10000 (original), 10000 (matched)
 - target estimand: ATT
 - covariates: age, female, prevDMTefficacy, prerelapse_num
```

3.4.2 Assess balance after matching

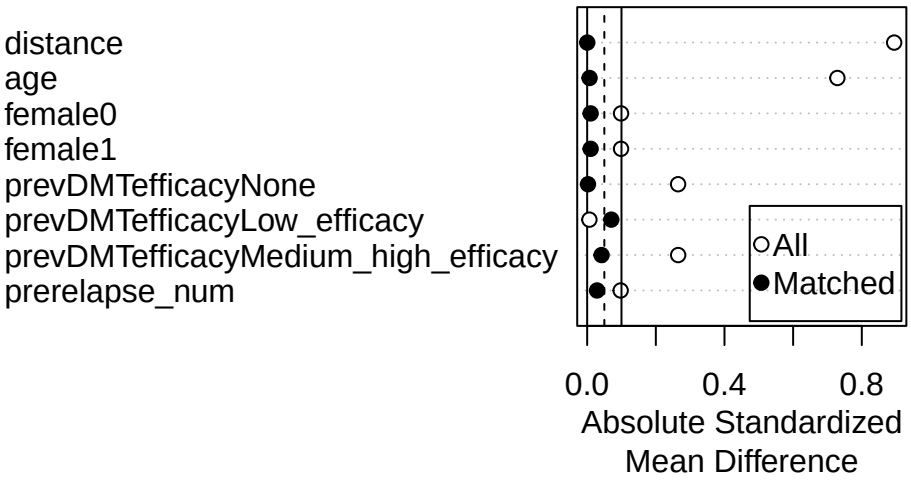
```
summary(opt)
```

```
Call:
matchit(formula = treatment ~ age + female + prevDMTefficacy +
    prerelapse_num, data = dat, method = "full", estimand = "ATT")

Summary of Balance for All Data:
                                  Means Treated Means Control Std. Mean Diff.
distance                             0.7970        0.6795          0.8943
age                                 44.2496       51.3883         -0.7289
female0                              0.2318        0.2735         -0.0987
female1                              0.7682        0.7265          0.0987
prevDMTefficacyNone                  0.4118        0.5422         -0.2649
prevDMTefficacyLow_efficacy          0.1114        0.1135         -0.0065
```

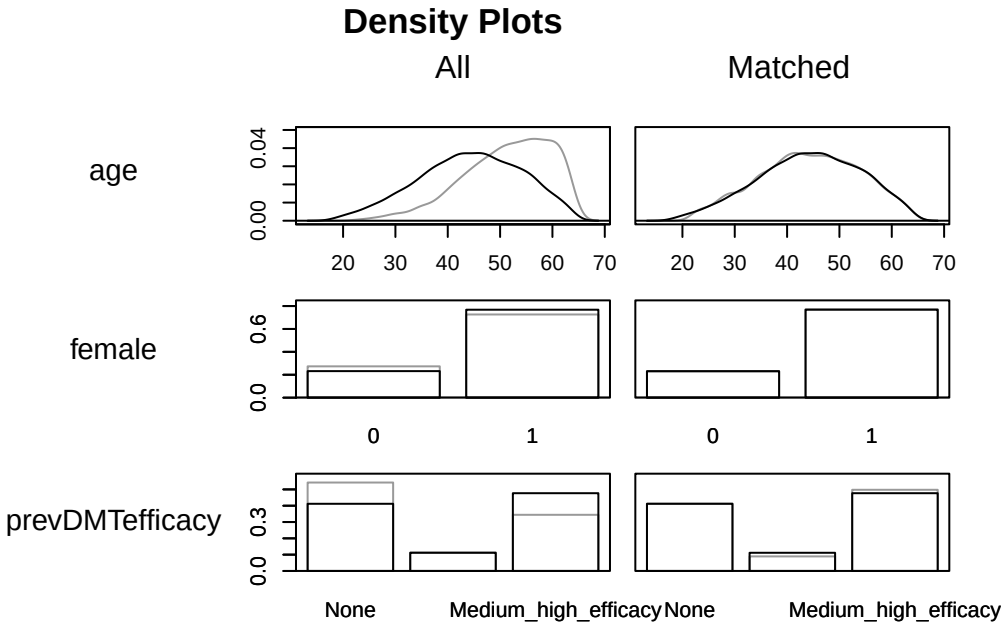
```
prevDMTefficacyMedium_high_efficacy        0.4768        0.3443          0.2651
prerelapse_num                             0.4595        0.3930          0.0976
                                  Var. Ratio eCDF Mean eCDF Max
distance                              0.7873    0.1917   0.3379
age                                   1.3868    0.1519   0.3085
female0                                    .    0.0417   0.0417
female1                                    .    0.0417   0.0417
prevDMTefficacyNone                        .    0.1304   0.1304
prevDMTefficacyLow_efficacy                .    0.0020   0.0020
prevDMTefficacyMedium_high_efficacy        .    0.1324   0.1324
prerelapse_num                        1.1990    0.0133   0.0383

Summary of Balance for Matched Data:
                                  Means Treated Means Control Std. Mean Diff.
distance                             0.7970        0.7970          0.0003
age                                 44.2496       44.3185         -0.0070
female0                              0.2318        0.2275          0.0101
female1                              0.7682        0.7725         -0.0101
prevDMTefficacyNone                  0.4118        0.4130         -0.0024
prevDMTefficacyLow_efficacy          0.1114        0.0893          0.0703
prevDMTefficacyMedium_high_efficacy  0.4768        0.4977         -0.0419
prerelapse_num                       0.4595        0.4399          0.0288
                                  Var. Ratio eCDF Mean eCDF Max
distance                              0.9976    0.0005   0.0075
age                                   1.0392    0.0038   0.0153
female0                                    .    0.0043   0.0043
female1                                    .    0.0043   0.0043
prevDMTefficacyNone                        .    0.0012   0.0012
prevDMTefficacyLow_efficacy                .    0.0221   0.0221
prevDMTefficacyMedium_high_efficacy        .    0.0209   0.0209
prerelapse_num                        1.1319    0.0060   0.0229
                                  Std. Pair Dist.
distance                                  0.0008
age                                       0.0667
female0                                   0.1775
female1                                   0.1775
prevDMTefficacyNone                       0.1100
prevDMTefficacyLow_efficacy               0.1846
prevDMTefficacyMedium_high_efficacy       0.1614
prerelapse_num                            0.2170

Sample Sizes:
              Control Treated
```

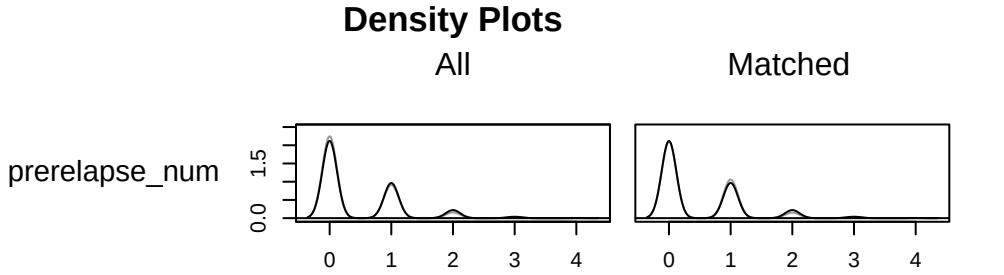
```
All           2300.      7700
Matched (ESS)  307.06    7700
Matched       2300.      7700
Unmatched        0.         0
Discarded        0.         0
```

```r
plot(summary(opt))
```




```r
# black line is treated group, grey line is control group
plot(opt, type = "density", which.xs = vars)
```


Density Plots
All
Matched
age
female
prevDMTefficacy
0.00
0.04
20
30
40
50
60
70
0.0
0.6
0
1
0.0
0.3
None
Medium_high_efficacy


Density Plots
All
Matched
prerelapse_num
0.0
1.5
0
1
2
3
4

3.4.3 Estimating the ATT

We can estimate the ATT in the matched sample using Poisson regression in which the number of post-treatment relapses is regressed on treatment status and follow-up time for each patient (captured by the variable `years`). More details are provided at https://cran.r-project.org/web/packages/MatchIt/vignettes/estimating-effects.html.

```
# Matched data
matched.data <- match.data(opt)

# Poisson regression model
opt.fit <- glm(y ~ treatment + offset(log(years)),
               family = poisson(link = "log"),
               data = matched.data,
               weights = weights)

# Treatment effect estimation
opt.comp <- comparisons(opt.fit,
                        variables = "treatment",
                        vcov = ~subclass,
                        newdata = subset(matched.data, treatment == "DMT1"),
                        wts = "weights",
                        transform_pre = "ratio")

opt.comp |> tidy()
```

```
# A tibble: 1 x 8
  term      contrast    estimate std.error statistic  p.value conf.low conf.high
  <chr>     <chr>          <dbl>     <dbl>     <dbl>    <dbl>    <dbl>     <dbl>
1 treatment mean(DMT1)~    0.804     0.102      7.88 3.25e-15    0.604      1.00
```

As indicated in the summary output above, the annualized relapse rate ratio for DMT1 vs DMT0 among patients treated with DMT0 (ATT) is given as 0.8 with a 95% confidence interval ranging from 0.6 to 1.

3.5 Propensity score stratification

3.5.1 Divide sample into quintiles of propensity scores

We will form five mutually exclusive groups of the estimated propensity score.

```r
# Create five strata
dat <- dat %>% mutate(ps.strata = cut(ps,
                                      breaks = c(quantile(ps, probs=seq(0,1,0.2))),
                                      labels = seq(1:5),
                                      include.lowest = TRUE))

# Number of patients in each stratum
table(dat$ps.strata)
```

```
   1    2    3    4    5
2002 2015 1991 1997 1995
```

3.5.2 Assess balance within each propensity score stratum

Within each propensity score stratum, treated and control patients should have similar values of the propensity score and the distribution of baseline covariates should be approximately balanced between treatment groups.

3.5.2.1 Propensity Score Stratum #1

```r
tab1.strata1 <- CreateTableOne(vars, data = dat %>% filter(ps.strata == 1),
                               factorVars = c("female", "prevDMTefficacy"),
                               strata = "treatment", test = FALSE)

tab1.strata1.print <- print(tab1.strata1, catDigits = 2, contDigits = 2,
                            smd = TRUE)
```

	DMT0	DMT1	SMD
n	901	1101	
age (mean (SD))	58.38 (3.67)	57.45 (3.73)	0.251
female = 1 (%)	605 (67.15)	775 (70.39)	0.070
prevDMTefficacy (%)			0.056
None	650 (72.14)	771 (70.03)	
Low_efficacy	106 (11.76)	130 (11.81)	
Medium_high_efficacy	145 (16.09)	200 (18.17)	
prerelapse_num (mean (SD))	0.29 (0.53)	0.33 (0.56)	0.074

3.5.2.2 Propensity Score Stratum #2

```
tab1.strata2 <- CreateTableOne(vars, data = dat %>% filter(ps.strata == 2),
                               factorVars = c("female", "prevDMTefficacy"),
                               strata = "treatment", test = FALSE)

tab1.strata2.print <- print(tab1.strata2, catDigits = 2, contDigits = 2,
                            smd = TRUE)
```

	DMT0	DMT1	SMD
n	617	1398	
age (mean (SD))	52.18 (4.35)	51.97 (4.22)	0.049
female = 1 (%)	458 (74.23)	1048 (74.96)	0.017
prevDMTefficacy (%)			0.054
None	292 (47.33)	624 (44.64)	
Low_efficacy	69 (11.18)	162 (11.59)	
Medium_high_efficacy	256 (41.49)	612 (43.78)	
prerelapse_num (mean (SD))	0.40 (0.64)	0.41 (0.66)	0.004

3.5.2.3 Propensity Score Stratum #3

```
tab1.strata3 <- CreateTableOne(vars, data = dat %>% filter(ps.strata == 3),
                               factorVars = c("female", "prevDMTefficacy"),
                               strata = "treatment", test = FALSE)

tab1.strata3.print <- print(tab1.strata3, catDigits = 2, contDigits = 2,
                            smd = TRUE)
```

	DMT0	DMT1	SMD
n	392	1599	
age (mean (SD))	46.73 (4.06)	46.36 (4.08)	0.092
female = 1 (%)	305 (77.81)	1193 (74.61)	0.075
prevDMTefficacy (%)			0.041
None	168 (42.86)	687 (42.96)	
Low_efficacy	52 (13.27)	191 (11.94)	
Medium_high_efficacy	172 (43.88)	721 (45.09)	
prerelapse_num (mean (SD))	0.49 (0.68)	0.47 (0.66)	0.031

3.5.2.4 Propensity Score Stratum #4

```
tab1.strata4 <- CreateTableOne(vars, data = dat %>% filter(ps.strata == 4),
                               factorVars = c("female", "prevDMTefficacy"),
                               strata = "treatment", test = FALSE)

tab1.strata4.print <- print(tab1.strata4, catDigits = 2, contDigits = 2,
                            smd = TRUE)
```

	DMT0	DMT1	SMD
n	269	1728	
age (mean (SD))	41.07 (4.11)	40.88 (4.29)	0.046
female = 1 (%)	203 (75.46)	1356 (78.47)	0.071
prevDMTefficacy (%)			0.084
None	105 (39.03)	634 (36.69)	
Low_efficacy	22 (8.18)	181 (10.47)	
Medium_high_efficacy	142 (52.79)	913 (52.84)	
prerelapse_num (mean (SD))	0.50 (0.69)	0.51 (0.71)	0.012

3.5.2.5 Propensity Score Stratum #5

```
tab1.strata5 <- CreateTableOne(vars, data = dat %>% filter(ps.strata == 5),
                               factorVars = c("female", "prevDMTefficacy"),
                               strata = "treatment", test = FALSE)

tab1.strata5.print <- print(tab1.strata5, catDigits = 2, contDigits = 2,
                            smd = TRUE)
```

	DMT0	DMT1	SMD
n	121	1874	
age (mean (SD))	33.26 (4.95)	32.04 (5.58)	0.233
female = 1 (%)	100 (82.64)	1543 (82.34)	0.008
prevDMTefficacy (%)			0.050
None	32 (26.45)	455 (24.28)	
Low_efficacy	12 (9.92)	194 (10.35)	
Medium_high_efficacy	77 (63.64)	1225 (65.37)	
prerelapse_num (mean (SD))	0.52 (0.66)	0.52 (0.73)	0.004

3.5.3 Estimating and pooling of stratum-specific treatment effects

The overall ATT across strata can be estimated by weighting stratum-specific estimates by the proportion of treated patients in each stratum over all treated patients in the sample.

We first define a function `att.strata.function()` to calculate stratum-specific estimates of the treatment effect:

```
att.strata.function <- function(data, stratum, confint = TRUE) {

  fit <- glm("y ~ treatment + offset(log(years))",
      family = poisson(link = "log"),
      data = data %>% filter(ps.strata == stratum))

  arr <- round(as.numeric(exp(coef(fit)["treatmentDMT1"])), digits = 3)
  ll <- ul <- NA

  if (confint) {
    ll <- round(exp(confint(fit))["treatmentDMT1",1], digits = 3)
    ul <- round(exp(confint(fit))["treatmentDMT1",2], digits = 3)
  }

  return(c("stratum" = stratum,
           "arr" = arr,
           "ci_lower"  = ll,
           "ci_upper"  = ul))
}

arr.strata <- as.data.frame(t(sapply(1:5, att.strata.function, data = dat)))
arr.strata
```

```
  stratum   arr ci_lower ci_upper
1       1 0.904    0.760    1.076
2       2 0.822    0.696    0.975
3       3 0.798    0.666    0.961
4       4 0.716    0.587    0.881
5       5 0.589    0.463    0.761
```

Subsequently, we define a function `weights.strata.function()` to calculate the weights for each stratum. The weight is the proportion of treated patients in each stratum over all treated patients in the sample:

```r
weights.strata.function <- function(data, stratum) {
  n_DMT1_stratum <- nrow(data %>% filter(ps.strata == stratum & treatment == "DMT1"))
  n_DMT1_all <- nrow(data %>% filter(treatment == "DMT1"))
  weight <- n_DMT1_stratum/n_DMT1_all
  return(c("stratum" = stratum, "weight" = weight))
}

weights.strata <- as.data.frame(t(sapply(1:5, weights.strata.function, data = dat)))
weights.strata
```

```
  stratum    weight
1       1 0.1429870
2       2 0.1815584
3       3 0.2076623
4       4 0.2244156
5       5 0.2433766
```

```r
# Create table with ARRs and weights for each PS stratum
arr.weights.merged <- merge(arr.strata, weights.strata, by = "stratum")

# Calculate the weighted ARR for each stratum
arr.weights.merged <- arr.weights.merged %>%
  mutate(weighted.arr = as.numeric(arr) * weight)

# Sum the weighted ARRs across strata to get the overall ATT
sum(arr.weights.merged$weighted.arr)
```

```
[1] 0.7482462
```

We now define a new function `ps.stratification.bootstrap()` that integrates estimation of the ATT and the PS weights for bootstrapping purposes:

```r
ps.stratification.bootstrap <- function(data, inds) {
  d <- data[inds,]

  d$ps.strata <- cut(d$ps,
                     breaks = c(quantile(dat$ps, probs = seq(0, 1, by = 0.2))),
                     labels = seq(5),
                     include.lowest = TRUE)
```

```
  arr.strata <- as.data.frame(t(sapply(1:5, att.strata.function,
                                       data = d, confint = FALSE)))

  weights.strata <- as.data.frame(t(sapply(1:5, weights.strata.function, data = d)))

  return(arr.strata$arr[1] * weights.strata$weight[1] +
           arr.strata$arr[2] * weights.strata$weight[2] +
           arr.strata$arr[3] * weights.strata$weight[3] +
           arr.strata$arr[4] * weights.strata$weight[4] +
           arr.strata$arr[5] * weights.strata$weight[5])
}
```

We can now estimate the treatment effect and its confidence interval using the bootstrap procedure:

```
library(boot)
```

```
Attaching package: 'boot'

The following object is masked from 'package:survival':

    aml
```

```
set.seed(1854)
arr.stratification.boot <- boot(data = dat,
                                statistic = ps.stratification.bootstrap,
                                R = 1000)

# Bootstrapped ARR
median(arr.stratification.boot$t)
```

```
[1] 0.7558609
```

```
# Bootstrapped ARR 95% CI
quantile(arr.stratification.boot$t[,1], c(0.025, 0.975))
```

```
     2.5%     97.5%
0.6835885 0.8362947
```

3.6 Propensity score weighting

3.6.1 Calculate propensity score weights for ATT

Propensity score weighting reweights the study sample to generate an artificial population (i.e., pseudo-population) in which the covariates are no longer associated with treatment, thereby removing confounding by measured covariates. For the ATT, the weight for all treated patients is set to one. Conversely, the weight for patients in the control group is set to the propensity score divided by one minus the propensity score, that is, (PS/(1 − PS)). We estimated stabilized weights to address extreme weights.

```
library(WeightIt)

w.out <- weightit(treatment ~ age + female + prevDMTefficacy + prerelapse_num,
                  data = dat,
                  method = "ps",
                  estimand = "ATT")
                  #stabilize = TRUE)

w.out
```

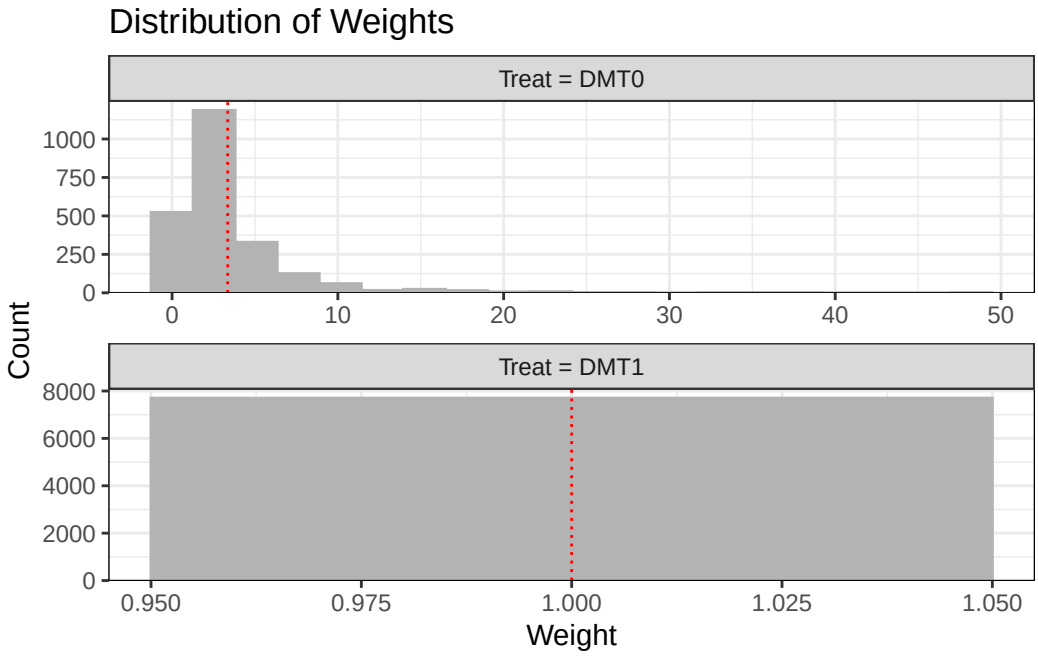
```
A weightit object
 - method: "glm" (propensity score weighting with GLM)
 - number of obs.: 10000
 - sampling weights: none
 - treatment: 2-category
 - estimand: ATT (focal: DMT1)
 - covariates: age, female, prevDMTefficacy, prerelapse_num
```

```
summary(w.out)
```

```
                Summary of weights

- Weight ranges:

        Min                                  Max
DMT0 0.4772 |--------------------------| 48.6856
DMT1 1.0000  ||                           1.0000

- Units with the 5 most extreme weights by group:
```

```
         9492    8836    6544    9610    4729
 DMT0 32.1027 32.1027 34.3126 38.1817 48.6856
            8       7       4       2       1
 DMT1       1       1       1       1       1

- Weight statistics:

     Coef of Var   MAD Entropy # Zeros
DMT0       1.098 0.673   0.383       0
DMT1       0.000 0.000  -0.000       0

- Effective Sample Sizes:

              DMT0 DMT1
Unweighted 2300.   7700
Weighted   1043.16 7700
```

```
plot(summary(w.out))
```



3.6.2 Assess balance in the weighted sample

```
bal.tab(w.out, stats = c("m", "v"), thresholds = c(m = .05))
```

```
Balance Measures
                                        Type Diff.Adj     M.Threshold
prop.score                          Distance  -0.0045 Balanced, <0.05
age                                  Contin.   0.0054 Balanced, <0.05
female                                Binary   0.0005 Balanced, <0.05
prevDMTefficacy_None                  Binary  -0.0003 Balanced, <0.05
prevDMTefficacy_Low_efficacy          Binary   0.0023 Balanced, <0.05
prevDMTefficacy_Medium_high_efficacy  Binary  -0.0020 Balanced, <0.05
prerelapse_num                       Contin.  -0.0034 Balanced, <0.05
                                    V.Ratio.Adj
prop.score                               0.9926
age                                      1.0102
female                                        .
prevDMTefficacy_None                          .
prevDMTefficacy_Low_efficacy                  .
prevDMTefficacy_Medium_high_efficacy          .
prerelapse_num                           1.0941

Balance tally for mean differences
                     count
Balanced, <0.05          7
Not Balanced, >0.05      0

Variable with the greatest mean difference
 Variable Diff.Adj     M.Threshold
      age   0.0054 Balanced, <0.05

Effective sample sizes
              DMT0 DMT1
Unadjusted 2300.   7700
Adjusted   1043.16 7700
```

3.6.3 Estimate the ATT

One way to estimate the ATT is to use the survey package. The function `svyglm()` generates model-robust (Horvitz-Thompson-type) standard errors by default, and thus does not require additional adjustments.

```r
library(survey)

weighted.data <- svydesign(ids = ~1, data = dat, weights = ~w.out$weights)

weighted.fit <- svyglm(y ~ treatment + offset(log(years)),
                       family = poisson(link = "log"),
                       design = weighted.data)

exp(coef(weighted.fit)["treatmentDMT1"])
```

```
treatmentDMT1
    0.7083381
```

```r
exp(confint(weighted.fit))["treatmentDMT1",]
```

```
    2.5 %    97.5 %
0.6245507 0.8033662
```

As indicated above, propensity score weighting yielded an ATT estimate of 0.71 (95% CI: 0.62; 0.8).

An alternative approach is to use `glm()` to estimate the treatment effect and calculate robust standard errors.

```r
# Alternative way to estimate treatment effect
weighted.fit2 <- glm(y ~ treatment + offset(log(years)),
              family = poisson(link = "log"),
              data = dat,
              weights = w.out$weights)

# Extract the estimated ARR
exp(coef(weighted.fit2))["treatmentDMT1"]
```

```
treatmentDMT1
    0.7083381
```

```r
# Calculate robust standard error and p-value of the log ARR
coeftest(weighted.fit2, vcov. = vcovHC)["treatmentDMT1",]
```

```
     Estimate    Std. Error       z value      Pr(>|z|)
-3.448337e-01  6.442745e-02 -5.352280e+00  8.685284e-08
```

```r
# Derive 95% confidence interval of the ARR
exp(lmtest::coefci(weighted.fit2,
       level = 0.95, # 95% confidence interval
       vcov. = vcovHC)["treatmentDMT1",])
```

```
    2.5 %    97.5 %
0.6243094 0.8036767
```

Using this approach, the ATT estimate was 0.71 (95% CI: 0.62; 0.8).

3.7 Regression adjustment for the propensity score for the ATE

In this approach, a regression model is fitted to describe the observed outcome as a function of the received treatment and the estimated propensity score:

```r
ps.reg.fit <- glm(y ~ treatment + ps + offset(log(years)),
                  family = poisson(link = "log"),
                  data = dat)

summary(ps.reg.fit)
```

```
Call:
glm(formula = y ~ treatment + ps + offset(log(years)), family = poisson(link = "log"),
    data = dat)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.0160  -0.7336  -0.4441  -0.1352   4.2634

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)   -1.99585    0.10359 -19.266  < 2e-16 ***
treatmentDMT1 -0.25598    0.04431  -5.777 7.60e-09 ***
ps             1.07521    0.13878   7.748 9.36e-15 ***
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 7514.7  on 9999  degrees of freedom
Residual deviance: 7443.0  on 9997  degrees of freedom
AIC: 12378

Number of Fisher Scoring iterations: 6
```

```r
# ATE
exp(coef(ps.reg.fit))["treatmentDMT1"]
```

```
treatmentDMT1
    0.7741606

Waiting for profiling to be done...
Waiting for profiling to be done...
```

Bootstrapped confidence intervals can be obtained as follows:

```r
# Function to bootstrap for 95% CIs
ps.reg.bootstrap <- function(data, inds) {
  d <- data[inds,]

  fit <- glm(y ~ treatment + ps + offset(log(years)),
             family = poisson(link = "log"),
             data = d)

  return(exp(coef(fit))["treatmentDMT1"])
}

set.seed(1854)

# Generate 1000 bootstrap replicates
arr.boot <- boot(dat, statistic = ps.reg.bootstrap, R = 1000)

# Extract the median annualized relapse rate across 1000 bootstrap replicates
median(arr.boot$t)
```

```
[1] 0.7750426
```

```r
# Take 2.5th and 97.5th percentiles to be 95% CI
quantile(arr.boot$t[,1], c(0.025, 0.975))
```

```
     2.5%     97.5%
0.7010540 0.8545169
```

3.8 Overview

Method	Estimand	Estimate	95% CI (lower)	95% CI (upper)
Optimal full matching	ATT	0.8039901	0.6040414	1.0039388
Propensity score stratification	ATT	0.7482462	NA	NA
Propensity score stratification (with bootstrapping)	ATT	0.7558609	0.6835885	0.8362947
Propensity score weighting	ATT	0.7083381	0.6245507	0.8033662
Propensity score weighting (robust SE)	ATT	0.7083381	0.6243094	0.8036767
PS regression adjustment	ATE	0.7741606	0.7101080	0.8448218
PS regression adjustment (bootstrapping)	ATE	0.7750426	0.7010540	0.8545169

Version info

This chapter was rendered using the following version of R and its packages:

```
R version 4.2.3 (2023-03-15 ucrt)
Platform: x86_64-w64-mingw32/x64 (64-bit)
Running under: Windows 10 x64 (build 19045)

Matrix products: default

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

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

```
other attached packages:
 [1] WeightIt_0.14.2        boot_1.3-28.1          MatchIt_4.5.4
 [4] sandwich_3.0-2         truncnorm_1.0-9        tableone_0.13.2
 [7] survey_4.2-1           survival_3.5-5         Matrix_1.5-4.1
[10] MASS_7.3-60            marginaleffects_0.13.0 lmtest_0.9-40
[13] zoo_1.8-12             knitr_1.43             ggplot2_3.4.2
[16] data.table_1.14.8      cobalt_4.5.1           dplyr_1.1.2

loaded via a namespace (and not attached):
 [1] tidyselect_1.2.0 xfun_0.39        mitools_2.4      splines_4.2.3
 [5] haven_2.5.2      lattice_0.21-8   labelled_2.11.0  colorspace_2.1-0
 [9] vctrs_0.6.3      generics_0.1.3   htmltools_0.5.5  yaml_2.3.7
[13] utf8_1.2.3       rlang_1.1.1      e1071_1.7-13     pillar_1.9.0
[17] glue_1.6.2       withr_2.5.0      DBI_1.1.3        lifecycle_1.0.3
[21] munsell_0.5.0    gtable_0.3.3     codetools_0.2-19 evaluate_0.21
[25] labeling_0.4.2   forcats_1.0.0    fastmap_1.1.1    class_7.3-22
[29] fansi_1.0.4      optmatch_0.10.6  Rcpp_1.0.10      checkmate_2.2.0
[33] backports_1.4.1  scales_1.2.1     jsonlite_1.8.5   farver_2.1.1
[37] chk_0.9.0        hms_1.1.3        digest_0.6.31    insight_0.19.2
[41] cli_3.6.1        tools_4.2.3      magrittr_2.0.3   proxy_0.4-27
[45] tibble_3.2.1     crayon_1.5.2     pkgconfig_2.0.3  rlemon_0.2.1
[49] rmarkdown_2.22   rstudioapi_0.14  R6_2.5.1         compiler_4.2.3
```

References

4 Effect Modification Analysis within the Propensity score Framework

Mohammad Ehsanul Karim (University of British Columbia)

Observational comparative effectiveness studies often adopt propensity score analysis to adjust for confounding. Although this approach is relatively straightforward to implement, careful thought is needed when treatment effect heterogeneity is present. This chapter illustrates the estimation of subgroup-specific treatment effects using (traditional) covariate adjustment methods, propensity score matching, propensity score weighting, propensity score stratification, and covariate adjustment using propensity scores.

4.1 Simulation

First, we need to install the R package `simcausal`, which can be obtained from GitHub:

```
devtools::install_github('osofr/simcausal', build_vignettes = FALSE)
```

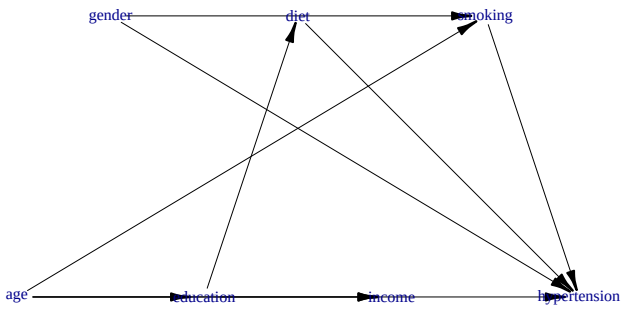
We will use the following data-generation model:

```
require(simcausal)
D <- DAG.empty()
D <- D +
  node("age", distr = "rnorm",
       mean = 2, sd = 4) +
  node("gender", distr = "rbern",
       prob = plogis(4)) +
  node("education", distr = "rbern",
       prob = plogis(3 + 5 * age)) +
  node("diet", distr = "rbern",
       prob = plogis(1 - 3 * education)) +
  node("income", distr = "rbern",
```

```
      prob = plogis(2 - 5 * education - 4 * age)) +
  node("smoking", distr = "rbern",
      prob = plogis(1 + 1.2 * gender + 2 * age)) +
  node("hypertension", distr = "rbern",
      prob = plogis(1 + log(3) * diet +
                    log(1.3) * age +
                    log(3.5) * smoking +
                    log(0.5) * gender))
Dset <- set.DAG(D)
```

Below is the diagram, with pink lines representing open backdoor path.

```
using the following vertex attributes:

NAdarkbluenone100.50

using the following edge attributes:

black0.210.60.5
```



We can now generate an example dataset:

```r
Obs.Data <- sim(DAG = Dset, n = 50000, rndseed = 123)
Obs.Data$smoking <- as.character(Obs.Data$smoking)
Obs.Data$income <- as.factor(Obs.Data$income)
Obs.Data$income <- relevel(Obs.Data$income, ref = "1")
```

Sample data from the hypothetical example of association between hypertension and smoking, where other variables such as income, age [centered], gender, education and diet also plays a role in the data generation process.

	age	gender	education	diet	income	smoking	hypertension
34901	12.29	1	1	1	0	1	1
149	10.40	1	1	0	0	1	1
10060	2.99	1	1	0	0	1	0
22220	-4.31	0	0	0	1	0	1
9979	-6.44	0	0	0	1	0	1

4.2 Covariate adjustment

4.2.1 Interaction approach

Below, we estimate a logistic regression model to assess whether the effect of smoking (the exposure) on hypertension is modified by income levels. This model considers the following variables:

- Outcome: `hypertension`
- Exposure variables: `smoking` and `income`
- Confounders: `age` and `gender`

```r
require(jtools)

fit.w.em <- glm(hypertension ~ smoking * income + age + gender,
            family = binomial(link = "logit"), data = Obs.Data)

results.model <- summ(fit.w.em, exp = TRUE)
```

Results indicate that the interaction between smoking status and income level is statistically significant ($p = 0.02$).

If we expand previous model to adjust for an additional confounder `education`, we have:

	exp(Est.)	2.5%	97.5%	z val.	p
(Intercept)	5.46	4.37	6.82	14.97	0.00
smoking1	2.93	2.60	3.30	17.69	0.00
income0	0.48	0.41	0.57	-8.28	0.00
age	1.29	1.27	1.31	36.77	0.00
gender	0.54	0.43	0.67	-5.55	0.00
smoking1:income0	1.27	1.04	1.56	2.33	0.02

```
fit.w.int <- glm(hypertension ~ smoking * income + age + gender + education,
                 family = binomial(link = "logit"),
                 data = Obs.Data)

results.int.model <- summ(fit.w.int, exp = TRUE)
```

	exp(Est.)	2.5%	97.5%	z val.	p
(Intercept)	5.69	4.56	7.11	15.31	0.00
smoking1	3.35	2.95	3.79	18.85	0.00
income0	1.09	0.85	1.40	0.68	0.49
age	1.30	1.28	1.32	37.32	0.00
gender	0.54	0.43	0.67	-5.58	0.00
education	0.42	0.35	0.51	-8.87	0.00
smoking1:income0	1.10	0.90	1.35	0.93	0.35

The interaction term between income and smoking is no longer statistically significant (p = 0.35).

We can generate a summary report from aforementioned effect modification analysis.

```
require(interactionR)

em.object <- interactionR(fit.w.em,
                          exposure_names = c("income0", "smoking1"),
                          ci.type = "mover", ci.level = 0.95,
                          em = TRUE, recode = FALSE)
```

The table below depicts the adjusted odds ratios for income levels (`high = 0`, and `low = 1`). The variables `CI.ll` and `CI.ul` depict the lower and upper limits of the 95 percent confidence intervals, `OR11` = $OR_{A=1,M=1}$, `OR10` = $OR_{A=1}$, `OR01` = $OR_{M=1}$ and `OR00` captures the reference.

Table 4.1: Summary report from an interaction analysis when investigating association between two exposure variables (smoking and income) and hypertension.

Measures	Estimates	CI.ll	CI.ul
OR00	1.00	NA	NA
OR01	2.93	2.60	3.30
OR10	0.48	0.41	0.57
OR11	1.80	1.63	1.98
OR(smoking1 on outcome [income0==0]	2.93	2.60	3.30
OR(smoking1 on outcome [income0==1]	3.72	3.14	4.41
Multiplicative scale	1.27	1.04	1.56
RERI	-0.61	-0.98	-0.29

Similarly, for the analysis adjusting for an additional confounder `education`, we have:

Table 4.2: Summary report from an interaction analysis when investigating association between two exposure variables (smoking and income) and hypertension.

Measures	Estimates	CI.ll	CI.ul
OR00	1.00	NA	NA
OR01	1.09	0.85	1.40
OR10	3.35	2.95	3.79
OR11	4.02	3.29	4.92
OR(income0 on outcome [smoking1==0]	1.09	0.85	1.40
OR(income0 on outcome [smoking1==1]	1.20	1.00	1.45
OR(smoking1 on outcome [income0==0]	3.35	2.95	3.79
OR(smoking1 on outcome [income0==1]	3.69	3.11	4.37
Multiplicative scale	1.10	0.90	1.35
RERI	0.59	0.03	1.27
AP	0.15	0.00	0.26
SI	1.24	1.01	1.53

```
# test run with additive model
Obs.Data$smoking <- as.numeric(as.character(Obs.Data$smoking))
Obs.Data$income <- as.numeric(as.character(Obs.Data$income))
fit.w.int.add <- glm(hypertension ~ smoking * income + age + gender + education,
                     family = gaussian(link = "identity"), data = Obs.Data)
sim_slopes(fit.w.int.add, pred = smoking, modx = income,
           exp = TRUE, robust = TRUE,
```

```
            confint = TRUE, data = Obs.Dat)
```

```
JOHNSON-NEYMAN INTERVAL

When income is INSIDE the interval [-3.27, 16.87], the slope of smoking is
p < .05.

Note: The range of observed values of income is [0.00, 1.00]

SIMPLE SLOPES ANALYSIS

Slope of smoking when income = 0.00 (0):

  Est.   S.E.   2.5%   97.5%   t val.      p
------ ------ ------ ------- -------- ------
  0.25   0.02   1.24    1.34    12.76   0.00

Slope of smoking when income = 1.00 (1):

  Est.   S.E.   2.5%   97.5%   t val.      p
------ ------ ------ ------- -------- ------
  0.28   0.01   1.30    1.34    34.53   0.00
```

4.2.2 Stratification

This approach involves estimating a regression model in different strata of the discrete effect modifier `income`:

```
# Estimate the prognostic effect of smoking in low income individuals
fit.income1 <- glm(hypertension ~ smoking + age + gender,
            family = binomial(link = "logit"),
            data = subset(Obs.Data, income == 1))

# Estimate the prognostic effect of smoking in high income individuals
fit.income0 <- glm(hypertension ~ smoking + age + gender,
            family = binomial(link = "logit"),
            data = subset(Obs.Data, income == 0))
```

The table below summarizes the adjusted odds ratios for smoking across the different income levels (`low = 1`, and `high = 0`) as obtained using the stratified approach.

Value of income	Estimate	2.5 %	97.5 %	z value	p value
1	3.07	2.71	3.47	17.65	0
0	3.59	3.02	4.26	14.57	0

Note that we can obtain the same results by estimating a regression model with an interaction term between the modifier and all covariates:

```
fit.all.int <- glm(hypertension ~ income * (smoking + age + gender),
                   family = binomial(link = "logit"), data = Obs.Data)

# Odds ratio for smoking in individuals with low income
exp(coef(fit.all.int)["smoking"])
```

```
smoking
3.59026
```

```
# Odds ratio for smoking in individuals with high income
exp(coef(fit.all.int)["smoking"] + coef(fit.all.int)["income:smoking"])
```

```
 smoking
3.066878
```

4.3 Propensity score matching

4.3.1 Stratification with exact matching within subgroups

We simulate another example dataset using aforementioned DAG, but restrict the sample size to 5000 individuals to reduce computational burden.

```
set.seed(123)
Obs.Data <- sim(DAG = Dset, n = 5000, rndseed = 123)
```

We first estimate the propensity of smoking in the high-income group (`income == 0`):

```
require(MatchIt)

match.income.0 <- matchit(smoking ~ age + gender,
                          data = subset(Obs.Data, income == 0),
```

```
                          method = "full", distance = "glm", link = "logit")
data.income.0 <- match.data(match.income.0)
```

Below, we draw a sample from the high-income group based on the hypothetical example of an association between hypertension and smoking. Here age [centered], gender, education, and diet are covariates.

```
            age gender education diet income smoking hypertension  distance
657   6.0810120      0         1    1      0       1            1 0.9999874
4932  1.6109860      1         1    0      0       1            0 0.9943155
252  -0.2475055      1         1    1      0       0            1 0.8525107
2693 -0.2511048      1         1    0      0       1            1 0.8516785
1646 -0.2836155      1         0    1      0       1            1 0.8439843
        weights subclass
657  1.00000000       36
4932 1.00000000       50
252  0.03296089       25
2693 1.00000000       25
1646 1.00000000        4
```

Now, we do the same for the low-income group (`income == 1`):

```
match.income.1 <- matchit(smoking ~ age + gender,
                          data = subset(Obs.Data, income == 1),
                          method = "full", distance = "glm", link = "logit")
data.income.1 <- match.data(match.income.1)
```

We estimated the exposure effect from a weighted outcome model for the matched data. While the `weights` are essential for estimating the point estimate from the outcome model, the `subclass` variable assists in calculating the robust variance of the exposure effect estimate.

```
# Treatment effect estimation
fit.income.0 <- glm(hypertension ~ smoking + age + gender,
                    data = data.income.0, weights = weights,
                    family = binomial("logit"))
fit.income.1 <- glm(hypertension ~ smoking + age + gender,
                    data = data.income.1, weights = weights,
                    family = binomial("logit"))
# Robust variance calculation
fit.nexp.adj.res1 <- summ(fit.income.1,
                          robust = TRUE,
                          cluster = "subclass",
```

```
                            confint = TRUE)
fit.nexp.adj.res0 <- summ(fit.income.0,
                            robust = TRUE,
                            cluster = "subclass",
                            confint = TRUE)
```

Table 4.3: Subgroup-specific treatment effect estimates (expressed in log-OR) from the hypothetical example using the stratified approach.

Value of income	Est.	2.5%	97.5%	z val.	p
0	3.74	-37.58	45.06	0.18	0.86
1	1.39	0.94	1.85	6.04	0.00

4.3.2 Joint approach without exact matching within subgroups

Here, entire cohort data is used to estimate the propensity scores, and the effect modifier `income` is considered as a covariate in the propensity score model:

```
ps.formula <- as.formula("smoking ~ age + gender + income")
match.obj.j <- matchit(ps.formula, data = Obs.Data,
                       method = "full",
                       distance = "glm",
                       link = "logit")
match.data.j <- match.data(match.obj.j)
```

```
fit.joint.no.exact <- glm(hypertension ~ smoking*income + age + gender,
                            data = match.data.j,
                            weights = weights,
                            family = binomial("logit"))
require(interactions)
nem.nexp.adj.res <- sim_slopes(fit.joint.no.exact,
                                 pred = smoking,
                                 modx = income,
                                 robust = "HC1",
                                 cluster = "subclass",
                                 johnson_neyman = TRUE,
                                 confint = TRUE,
                                 data = match.data.j)
```

Table 4.4: Subgroup-specific treatment effect estimates (expressed in log-OR) from the hypothetical example using the joint approach.

Value of income	Est.	S.E.	2.5%	97.5%	z val.	p
0	3.85	1.00	1.89	5.82	3.84	0
1	1.40	0.28	0.85	1.95	4.99	0

4.3.3 Joint approach with exact matching within subgroups

We specify the moderator variable's name in the `exact` argument of the `matchit` function.

```
ps.formula.no.mod <- as.formula("smoking ~ age + gender")
match.obj.js <- matchit(ps.formula.no.mod, data = Obs.Data,
                        method = "full", distance = "glm",link = "logit",
                        exact = "income")
match.data.js <- match.data(match.obj.js)
fit.joint.exact <- glm(hypertension ~ smoking*income + age + gender,
                       data = match.data.js, weights = weights,
                       family = binomial("logit"))
js.nexp.adj.res <- sim_slopes(fit.joint.exact,
                              pred = smoking, modx = income,
                              robust = "HC1", cluster = "subclass",
                              johnson_neyman = FALSE, confint = TRUE,
                              data = match.data.js)
```

Table 4.5: Subgroup-specific exposure effect estimates (expressed in log-OR) from the hypothetical example using the Joint model, separate matching approach.

Value of income	Est.	S.E.	2.5%	97.5%	z val.	p
0	3.89	1.01	1.92	5.87	3.87	0
1	1.38	0.28	0.84	1.93	4.95	0

4.3.4 Interaction approach without exact matching within subgroups

Analysts incorporate relevant moderator-covariate interactions into the propensity score model that align with biological plausibility. For instance, in the case study we considered an interaction between age (a covariate) and income (a moderator), but did not include other interactions terms.

```
ps.formula.with.int <- formula("smoking ~ age*income + gender")
match.obj.i <- matchit(ps.formula.with.int, data = Obs.Data,
                       method = "full", distance = "glm",link = "logit")
match.data.i <- match.data(match.obj.i)
fit.int.no.exact <- glm(hypertension ~ smoking*income + age + gender,
                        data = match.data.i, weights = weights,
                        family = binomial("logit"))
i.nexp.adj.res <- sim_slopes(fit.int.no.exact,
                             pred = smoking, modx = income,
                             robust = "HC1", cluster = "subclass",
                             johnson_neyman = FALSE, confint = TRUE,
                             data = match.data.i)
```

Table 4.6: Subgroup-specific exposure effect estimates (expressed in log-OR) from the hypothetical example using the interaction approach.

Value of income	Est.	S.E.	2.5%	97.5%	z val.	p
0	3.87	1.00	1.90	5.83	3.86	0
1	1.39	0.28	0.84	1.94	4.95	0

4.3.5 Interaction approach with exact matching within subgroups

This method bears resemblance to the interaction approach for propensity score estimation. However, when it comes to matching, researchers match within each moderator subgroup.

```
match.obj.is <- matchit(ps.formula.with.int, data = Obs.Data,
                        method = "full", distance = "glm",link = "logit",
                        exact = "income")
match.data.is <- match.data(match.obj.is)
fit.int.exact <- glm(hypertension ~ smoking*income + age + gender,
                     data = match.data.is, weights = weights,
                     family = binomial("logit"))
is.nexp.adj.res <- sim_slopes(fit.int.exact,
                              pred = smoking, modx = income,
                              robust = "HC1", cluster = "subclass",
                              johnson_neyman = FALSE, confint = TRUE,
                              data = match.data.is)
```

Table 4.7: Subgroup-specific exposure effect estimates (expressed in log-OR) from the hypothetical example using the interaction model, separate matching approach.

Value of income	Est.	S.E.	2.5%	97.5%	z val.	p
0	3.86	1.00	1.90	5.83	3.85	0
1	1.40	0.28	0.85	1.95	4.99	0

4.4 Propensity Score Weighting

4.4.1 Common model

This approach adds confounder-moderator interactions in the common weight model.

```
require(WeightIt)
W.out <- weightit(ps.formula.with.int,
                  data = Obs.Data,
                  method = "ps",
                  estimand = "ATT")
require(survey)
d.w <- svydesign(~1, weights = W.out$weights, data = Obs.Data)
fit2w <- svyglm(hypertension ~ smoking*income, design = d.w,
                family = binomial("logit"))
w.nexp.adj.res <- sim_slopes(fit2w, pred = smoking, modx = income,
                             confint = TRUE)
```

Table 4.8: Subgroup-specific exposure effect estimates (expressed in log-OR) from the hypothetical example using the weighting approach.

Value of income	Est.	S.E.	2.5%	97.5%	t val.	p
0	2.66	0.63	1.42	3.89	4.23	0
1	1.32	0.25	0.83	1.82	5.24	0

We can adjust previous analysis model to adopt stabilized weights for the propensity score (`stabilize = TRUE`):

```
W.out.st <- weightit(ps.formula.with.int, data = Obs.Data,
                     method = "ps",
                     estimand = "ATT",
                     stabilize = TRUE)
d.sw <- svydesign(~1, weights = W.out.st$weights, data = Obs.Data)
```

```
fit2sw <- svyglm(hypertension ~ smoking*income + age + gender,
                 design = d.sw,
                 family = binomial("logit"))
ws.nexp.adj.res <- sim_slopes(fit2sw,
                              pred = smoking, modx = income,
                              confint = TRUE)
```

Table 4.9: Subgroup-specific exposure effect estimates (expressed in log-OR) from the hypothetical example using stabilized propensity score weights.

Value of income	Est.	S.E.	2.5%	97.5%	t val.	p
0	2.27	0.73	0.84	3.69	3.12	0
1	1.32	0.25	0.83	1.82	5.23	0

4.4.2 Separate models

Propensity score weighting approach with weights estimated separately from each subgroup:

```
ps.formula.with.no.int <- formula("smoking ~ age + gender")
W.out1 <- weightit(ps.formula.with.no.int,
                   data = subset(Obs.Data, income == 1),
                   method = "ps",
                   estimand = "ATT")
trimmed.weight.1.percent1 <- trim(W.out1$weights,
                                  at = 1, lower = TRUE)
```

Table 4.10: Weight summaries before and after truncation.

Weight	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Raw weights	0	0.01	0.11	0.45	1	11.69
1% truncated weights	0	0.01	0.11	0.44	1	7.61

```
# Outcome model for income = 1
d.w1 <- svydesign(~1, weights = trimmed.weight.1.percent1,
                  data = subset(Obs.Data, income == 1))
fit2unadj1 <- svyglm(hypertension ~ smoking, design = d.w1,
                     family = binomial("logit"))

# weight model for income = 0
```

```r
W.out0 <- weightit(ps.formula, data = subset(Obs.Data, income == 0),
                   method = "ps", estimand = "ATT")
trimmed.weight.1.percent0 <- trim(W.out0$weights, at = 1, lower = TRUE)

# Outcome model for income = 0
d.w0 <- svydesign(~1, weights = trimmed.weight.1.percent0,
                  data = subset(Obs.Data, income == 0))
fit2unadj0 <- svyglm(hypertension ~ smoking, design = d.w0,
                     family = binomial("logit"))

fit.exp.adj.res1 <- summ(fit2unadj1, confint = TRUE)
fit.exp.adj.res0 <- summ(fit2unadj0, confint = TRUE)
```

Table 4.11: Subgroup-specific exposure effect estimates (expressed in log-OR) from the hypothetical example using the propensity score weighting approach (Separate weight models).

Value of income	Est.	2.5%	97.5%	t val.	p
0	2.21	1.27	3.15	4.60	0
1	1.34	0.85	1.83	5.36	0

4.4.3 Weights from the subgroup balancing propensity scores

Subgroup balancing propensity scores for propensity score weighting:

```r
w.out <- weightit(smoking ~ age + gender + income,
                  data = Obs.Data,
                  method = "ps", estimand = "ATT")
w.out.sb <- sbps(w.out, moderator = "income")
d.w.sb <- svydesign(~1, weights = w.out.sb$weights, data = Obs.Data)
fit2unadj.sb <- svyglm(hypertension ~ smoking*income, design = d.w.sb,
                       family = binomial("logit"))
sb.w.nexp.adj.res <- sim_slopes(fit2unadj.sb,
                                pred = smoking,
                                modx = income,
                                confint = TRUE,
                                johnson_neyman = FALSE,)
```

Table 4.12: Subgroup-specific exposure effect estimates (expressed in log-OR) from the hypothetical example using the subgroup balancing weighting approach.

Value of income	Est.	S.E.	2.5%	97.5%	t val.	p
0	2.68	0.64	1.44	3.92	4.22	0
1	1.32	0.25	0.82	1.82	5.22	0

4.5 Covariate adjustment for the propensity score

4.5.1 As continuous covariate

An implementation of propensity scores as a continuous covariate in the outcome model:

```
# Separate models for each subgroup

# For subgroup income = 1
Obs.Data$ps[Obs.Data$income == 1] <- glm(ps.formula,
                                        data = subset(Obs.Data, income == 1),
                                        family = "binomial")$fitted.values
fit2adj1 <- glm(hypertension ~ smoking + age + gender,
                family = binomial("logit"),
                data = subset(Obs.Data, income == 1))

# For subgroup income = 0
Obs.Data$ps[Obs.Data$income == 0] <- glm(ps.formula,
                                        data = subset(Obs.Data, income == 0),
                                        family = "binomial")$fitted.values
fit2adj0 <- glm(hypertension ~ smoking + age + gender,
                family = binomial("logit"),
                data = subset(Obs.Data, income == 0))

fit.nexp.adj.res1 <- summ(fit2adj1, robust = TRUE, confint = TRUE)
fit.nexp.adj.res0 <- summ(fit2adj0, robust = TRUE, confint = TRUE)
```

```
# Common model
Obs.Data$ps <- glm(ps.formula.with.int, data = Obs.Data,
                       family = "binomial")$fitted.values
```

```
Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

Table 4.13: Subgroup-specific exposure effect estimates (expressed in log-OR) from the hypothetical example using Propensity Score as a covariate adjustment approach (considering separate models for each subgroup).

Value of income	Est.	2.5%	97.5%	z val.	p
0	1.16	0.56	1.75	3.83	0
1	1.37	0.96	1.77	6.61	0

```
fit2adjc <- glm(hypertension ~ smoking*income + age + gender + ps,
                family = binomial("logit"),
                data = Obs.Data)
c.nexp.adj.res <- sim_slopes(fit2adjc,
                             pred = smoking, modx = income,
                             confint = TRUE,
                             data = Obs.Data)
```

Table 4.14: Subgroup-specific exposure effect estimates (expressed in log-OR) from the hypothetical example using Propensity Score as a covariate adjustment approach (considering a common model).

Value of income	Est.	S.E.	2.5%	97.5%	z val.	p
0	1.17	0.29	0.61	1.74	4.07	0
1	1.43	0.23	0.98	1.87	6.30	0

4.5.2 As quantiles

The propensity scores as a categorical covariate, broken by quintiles, in the outcome model.

```
Obs.Data$ps <- glm(ps.formula.with.int,
                   data = Obs.Data,
                   family = "binomial")$fitted.values
quintiles <- quantile(Obs.Data$ps,
                      prob = seq(from = 0, to = 1, by = 0.2),
                      na.rm = T)
Obs.Data$psq <- cut(Obs.Data$ps, breaks = quintiles,
                    labels = seq(1,5), include.lowest = T)
Obs.Data$psq <- as.factor(Obs.Data$psq)

fit2adjq <- glm(hypertension ~ (smoking*psq)*income,
```

```
                family = binomial("logit"),
                data = Obs.Data)
cq.nexp.adj.res <- sim_slopes(fit2adjq,
                              pred = smoking,
                              modx = income,
                              confint = TRUE,
                              data = Obs.Data)
```

Table 4.15: Subgroup-specific exposure effect estimates (expressed in log-OR) from the hypothetical example using Propensity Score as a covariate adjustment approach (as quintiles).

Value of income	Est.	S.E.	2.5%	97.5%	z val.	p
0	3.08	0.63	1.85	4.32	4.91	0
1	2.60	0.47	1.68	3.51	5.56	0

4.6 Propensity Score Stratification

Here is an implementation of propensity score stratification approach by using the marginal mean weighting through stratification (MMWS):

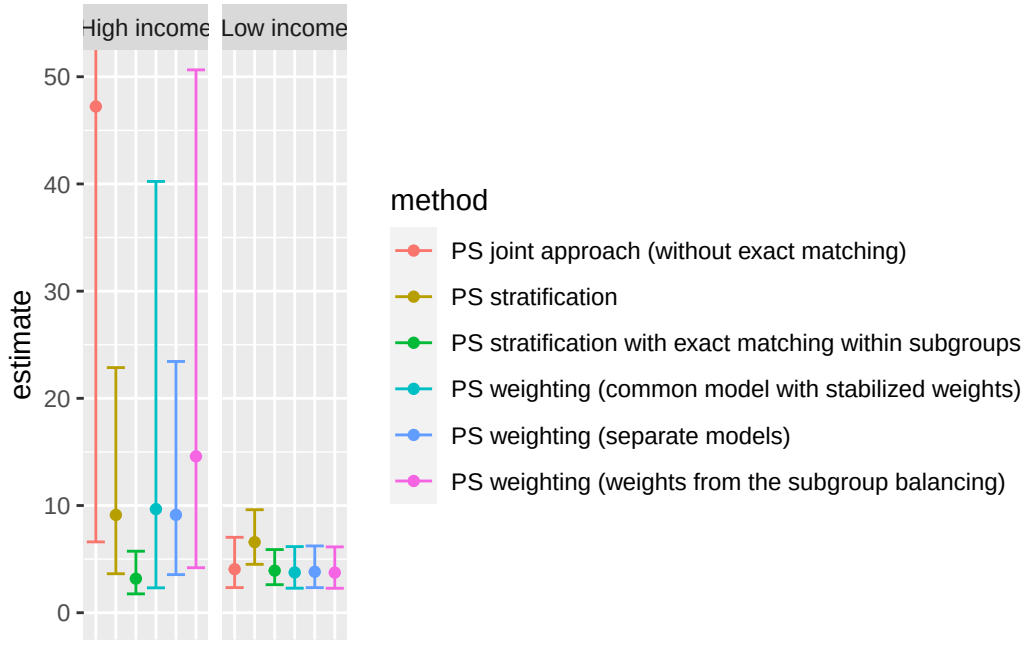
```
match.obj <- matchit(ps.formula, data = Obs.Data,
                     method = "subclass", subclass = 3,
                     estimand = "ATT", min.n = 10)
data.subclass <- match.data(match.obj)
subclass.fit <- glm(hypertension ~ smoking*income, family = binomial("logit"),
              data = data.subclass,
              weights = weights)
subclass.nexp.adj.res <- sim_slopes(subclass.fit,
                                    pred = smoking,
                                    modx = income,
                                    confint = TRUE,
                                    robust = "HC3",
                                    johnson_neyman = FALSE,
                                    data = data.subclass)
```

Table 4.16: Subgroup-specific exposure effect estimates (expressed in log-OR) from the hypothetical example using propensity score stratification approach.

Value of income	Est.	S.E.	2.5%	97.5%	z val.	p
0	2.21	0.47	1.29	3.13	4.71	0
1	1.89	0.19	1.51	2.26	9.78	0

4.7 Summary

The marginal odds ratios for `smoking` are summarized below



Version info

This chapter was rendered using the following version of R and its packages:

```
R version 4.2.3 (2023-03-15 ucrt)
Platform: x86_64-w64-mingw32/x64 (64-bit)
Running under: Windows 10 x64 (build 19045)

Matrix products: default
```

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

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

other attached packages:
 [1] interactionR_0.1.6 simcausal_0.5.6    scales_1.2.1       ggplot2_3.4.2
 [5] xtable_1.8-4       dplyr_1.1.2        kableExtra_1.3.4   knitr_1.43
 [9] cowplot_1.1.1      readstata13_0.10.1 survey_4.2-1       survival_3.5-5
[13] Matrix_1.5-4.1     broom_1.0.5        MatchIt_4.5.4      interactions_1.1.5
[17] jtools_2.2.1       sandwich_3.0-2     lmtest_0.9-40      zoo_1.8-12
[21] optmatch_0.10.6    WeightIt_0.14.2    cobalt_4.5.1       table1_1.4.3

loaded via a namespace (and not attached):
 [1] fontquiver_0.2.1        webshot_0.5.4           httr_1.4.6
 [4] tools_4.2.3             backports_1.4.1         utf8_1.2.3
 [7] R6_2.5.1                DBI_1.1.3               colorspace_2.1-0
[10] withr_2.5.0             tidyselect_1.2.0        curl_5.0.1
[13] compiler_4.2.3          textshaping_0.3.6       cli_3.6.1
[16] rvest_1.0.3             expm_0.999-7            flextable_0.9.2
[19] xml2_1.3.4              officer_0.6.2           fontBitstreamVera_0.1.1
[22] labeling_0.4.2          mvtnorm_1.2-2           askpass_1.1
[25] systemfonts_1.0.4       stringr_1.5.0           digest_0.6.31
[28] rmarkdown_2.22          svglite_2.1.1           gfonts_0.2.0
[31] pkgconfig_2.0.3         htmltools_0.5.5         fastmap_1.1.1
[34] rlang_1.1.1             rstudioapi_0.14         httpcode_0.3.0
[37] shiny_1.7.4             farver_2.1.1            generics_0.1.3
[40] jsonlite_1.8.5          car_3.1-2               zip_2.3.0
[43] magrittr_2.0.3          Formula_1.2-5           Rcpp_1.0.10
[46] munsell_0.5.0           fansi_1.0.4             abind_1.4-5
[49] gdtools_0.3.3           lifecycle_1.0.3         chk_0.9.0
[52] stringi_1.7.12          yaml_2.3.7              carData_3.0-5
[55] promises_1.2.0.1        crayon_1.5.2            lattice_0.21-8
[58] splines_4.2.3           pander_0.6.5            pillar_1.9.0
[61] uuid_1.1-0              igraph_1.5.0            codetools_0.2-19
[64] crul_1.4.0              glue_1.6.2              evaluate_0.21
[67] msm_1.7                 mitools_2.4             fontLiberation_0.1.0
[70] data.table_1.14.8       vctrs_0.6.3             httpuv_1.6.11
```

```
[73] openssl_2.0.6          gtable_0.3.3           purrr_1.0.1
[76] tidyr_1.3.0            assertthat_0.2.1       xfun_0.39
[79] mime_0.12              later_1.3.1            ragg_1.2.5
[82] viridisLite_0.4.2      tibble_3.2.1           ellipsis_0.3.2
[85] rlemon_0.2.1
```

5 Dealing with missing data

Johanna Munoz (Julius Center for Health Sciences and Primary Care)
Thomas Debray (Smart Data Analysis and Statistics B.V.)

5.1 Main Analysis

The main objective of this analysis is to assess whether the number of episodes (y) occurring within specific time periods (years) differs between the treatment groups (1: DMF and 0: TERI). To address potential confounding factors, the researchers consider variables such as patient age, the log of premedical cost (`logPremedicalcost`), previous DMT efficacy (`prevDMTefficacy`), and the number of episodes in previous relapses (prerelapseNum).

When estimating treatment effects from observational data, an assumption is made that the patient populations in both treatment groups are as similar as possible. Various methods for balancing data across treatment groups are proposed, including matching, inverse propensity weighting, stratification, and regression adjustment.

In this case, the focus is specifically on the matching method, which offers advantages over regression adjustment by potentially alleviating issues related to model mis-specification. This includes addressing non-linear relationships between certain confounders and the outcome variable and accounting for treatment effects that may depend on specific confounders (treatment-confounder interaction terms). Propensity scores are used to match subjects in the treatment groups.

Moreover, intentionally introducing incomplete covariate variables in this example adds complexity to the propensity score estimation. Depending on the propensity score estimation technique employed, it may be necessary to incorporate an imputation step. For instance, logistic regression estimation requires complete data for all observations, while XGBoost is robust to missing data [**?**].

To estimate marginal treatment effects, the g-computation method is employed [**?**]. This method involves specifying a model for the outcome dependent on the treatment and covariates. The potential outcomes, i.e., the predicted values of the outcome on treatment (y_i^1) and control

(y_i^0) for each sample unit i, are estimated. The marginal treatment effect is then calculated by contrasting the averaged estimated potential outcomes.

In this example, we consider the estimation of comparative treatment effects in the absence of treatment-effect heterogeneity.

5.2 Estimation workflow

The proposed workflow consists of the following steps:

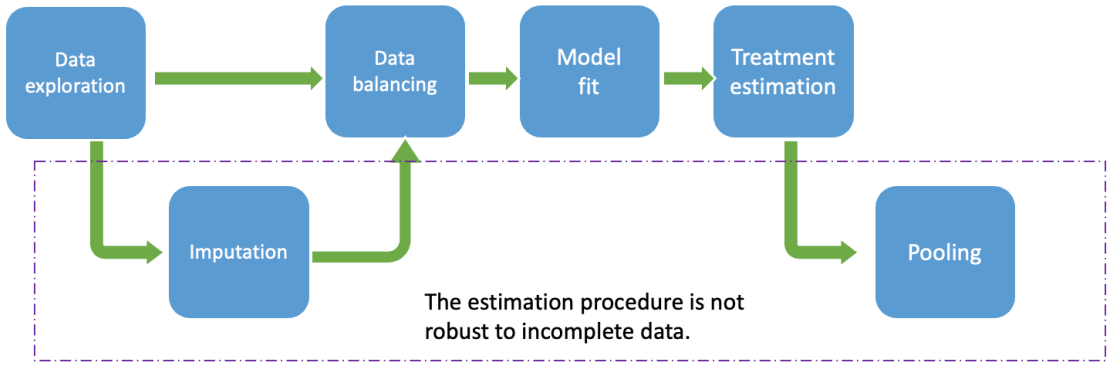


Figure 5.1: Estimation Workflow

1. **Data Exploration:** In this step, we examine the observed data to comprehend the variables within the dataset. Our primary focus lies on identifying missing patterns and relationships among observed variables, including missing indicator variables and others. This exploration aids in discerning the most plausible missing mechanisms and suitable imputation techniques. Additionally, field experts' insights may be incorporated to enhance understanding of the missing process, potentially considering MNAR assumptions.
2. **Imputation:** It is essential to evaluate whether the imputation procedure is necessary or if simpler methods, such as complete case analysis, are more suitable. In case imputation procedures are required, selecting plausible imputation methods that align with the main model analysis is crucial. This involves choosing individual imputation methods for each incomplete variable, determining the predictor variables on the imputation model. Pre_imputation (where imputation values can be deterministically derived from other variables) and Post-imputation (e.g.ensuring imputed values fall within a reasonable range) steps may also considered.
3. **Data Balancing:** Several methods, including PS matching or inverse weighting propensity score, can be utilized. It is required to evaluate the balance, which could be done

via visual inspection.(eg.cobalt package). In this example, we estimate propensity scores using logistic regression. For most balancing procedures in R, counterparts specifically designed for imputed datasets are available, such as those in the matchthem R package, which includes PS matching and IPW as done in the matchit R package.

4. **Model Fit:** : It is fit a model to predict the outcomes for each sample unit under each possible treatment value (DMF and TERI), as predictors include the treatment and optionally the baseline covariates and also the propensity score.
5. **Treatment Estimation & Pooling:** For simplicity in this tutorial, we will use the comparison functions from the R **matchingmethods** package [**?**], which can be used for completed data and also from outputs from the imputation process. In the last case, internally the functions calculate the treatment effects on each imputed dataset and pool the estimates using Rubin's Rules.

Let's start by preparing the R environment. All the functions used in this tutorial can be found in `functions.r` within the resources directory.

```
# Load the required packages and additional functions
source("resources/chapter 09/functions.r")
```

5.3 Homogeneous Treatment Effect

In this example, we focus on estimating comparative treatment effects in the absence of treatment-effect heterogeneity.

Version info

This chapter was rendered using the following version of R and its packages:

```
R version 4.2.3 (2023-03-15 ucrt)
Platform: x86_64-w64-mingw32/x64 (64-bit)
Running under: Windows 10 x64 (build 19045)

Matrix products: default

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

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

other attached packages:
 [1] marginaleffects_0.15.0 ggplot2_3.4.3          missForest_1.5
 [4] sandwich_3.0-2         PSweight_1.1.8         cobalt_4.5.1
 [7] WeightIt_0.14.2        MatchIt_4.5.4          optmatch_0.10.6
[10] truncnorm_1.0-9        MASS_7.3-60            survey_4.2-1
[13] survival_3.5-5         Matrix_1.5-4.1         data.table_1.14.8
[16] tidyr_1.3.0            MatchThem_1.1.0        ggmice_0.1.0
[19] dplyr_1.1.2            mice_3.16.0            table1_1.4.3
[22] kableExtra_1.3.4

loaded via a namespace (and not attached):
 [1] nlme_3.1-162          webshot_0.5.5         httr_1.4.7
 [4] numDeriv_2016.8-1.1   doRNG_1.8.6           tools_4.2.3
 [7] backports_1.4.1       utf8_1.2.3            R6_2.5.1
[10] rpart_4.1.19          DBI_1.1.3             colorspace_2.1-0
[13] jomo_2.7-6            nnet_7.3-19           withr_2.5.0
[16] gbm_2.1.8.1           tidyselect_1.2.0      compiler_4.2.3
[19] glmnet_4.1-7          cli_3.6.1             rvest_1.0.3
[22] xml2_1.3.4            scales_1.2.1          nnls_1.5
[25] randomForest_4.7-1.1  systemfonts_1.0.4     stringr_1.5.0
[28] digest_0.6.31         minqa_1.2.5           rmarkdown_2.24
[31] svglite_2.1.1         pkgconfig_2.0.3       htmltools_0.5.5
[34] lme4_1.1-33           itertools_0.1-3       fastmap_1.1.1
[37] rlang_1.1.1           rstudioapi_0.15.0     shape_1.4.6
[40] generics_0.1.3        zoo_1.8-12            jsonlite_1.8.5
[43] magrittr_2.0.3        Formula_1.2-5         Rcpp_1.0.10
[46] munsell_0.5.0         fansi_1.0.4           lifecycle_1.0.3
[49] stringi_1.7.12        yaml_2.3.7            parallel_4.2.3
[52] mitml_0.4-5           crayon_1.5.2          lattice_0.21-8
[55] splines_4.2.3         knitr_1.44            pillar_1.9.0
[58] boot_1.3-28.1         rngtools_1.5.2        codetools_0.2-19
[61] pan_1.6               glue_1.6.2            evaluate_0.21
[64] mitools_2.4           vctrs_0.6.3           nloptr_2.0.3
[67] foreach_1.5.2         gtable_0.3.4          purrr_1.0.1
[70] xfun_0.39             SuperLearner_2.0-28.1 broom_1.0.5
[73] viridisLite_0.4.2     tibble_3.2.1          iterators_1.0.14
[76] gam_1.22-2
```

6 Systematic review and meta-analysis of Real-World Evidence

Dimitris Mavridis (University of Ioannina)
Thomas Debray (Smart Data Analysis and Statistics B.V.)

6.1 Introduction

We first load the required packages

```
library(dplyr)
library(gemtc)
library(netmeta)
```

6.2 Pairwise meta-analysis of clinical trials

6.2.1 Toculizumab for coronavirus disease 2019

In this example, we consider the results from a systematic literature review of clinical trials investigating any pharmacological in hosptialized patients with coronavirus disease 2019 (Selvarajan et al. 2022). A total of 23 randomized controlled trials were included and studied seven different interventions: dexamethasone, remdesivir, tocilizumab, hydroxychloroquine, combination of lopinavir/ritonavir, favipiravir and interferon- . We here focus on the synthesis of 7 trials that comparted toculizumab (`TOCI`) to standard care (`STD`) and collected mortality data.

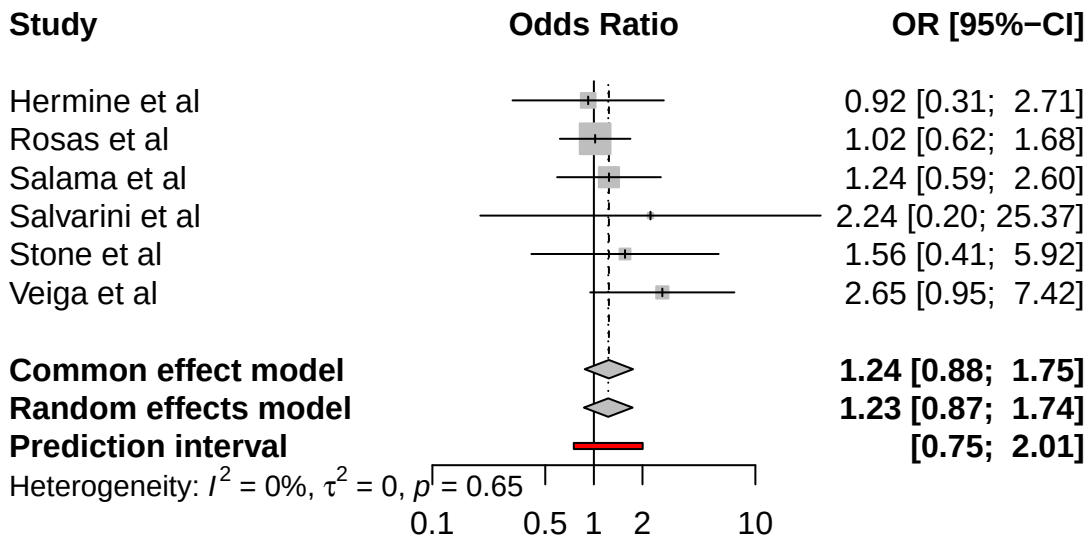
studlab	treat1	treat2	event1	n1	event2	n2
Hermine et al	TOCI	STD	7	63	8	67
Rosas et al	TOCI	STD	58	294	28	144
Salama et al	TOCI	STD	26	249	11	128

(continued)

studlab	treat1	treat2	event1	n1	event2	n2
Salvarini et al	TOCI	STD	2	60	1	66
Stone et al	TOCI	STD	9	161	3	82
Veiga et al	TOCI	STD	14	65	6	64

We now conduct a pairwise meta-analysis to assess the pooled effect of tocilizumab versus standard care. For each study, the log odds ratio and corresponding standard error is derived after which the corresponding estimates are pooled using the Mantel-Haenszel method.

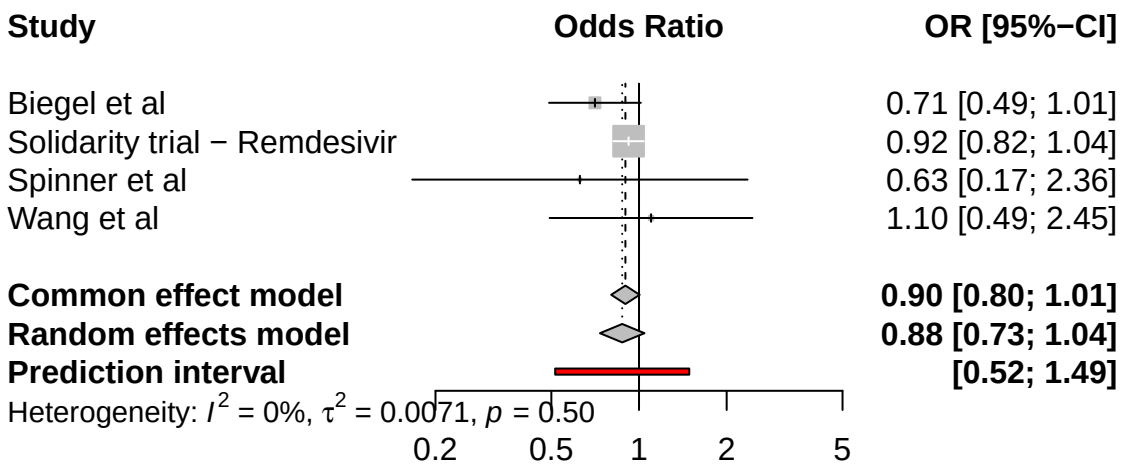
```
results.TOCI <- metabin(event1,n1,event2,n2,studlab,data=tocilizumab,
                        sm="OR",main="tocilizumab vs standard care",
                        prediction=TRUE)
forest(results.TOCI, leftcols = "studlab", rightcols = "effect.ci")
```



Altough a random effects meta-analysis was conducted, no heterogeneity was found (τ=0, with a 95% confidence interval ranging from 0 to 0.85).

6.2.2 Remdesivir for coronavirus disease 2019

In aforementioned example, a total of 4 trials compared remdesivir to standard care:



6.3 Network meta-analysis of clinical trials

We here use the R packages `netmeta` for conducting a frequentist network meta-analysis. A detailed tutorial on the use of `netmeta` is available from the book Doing Meta-Analysis with R: A Hands-On Guide.

6.3.1 Interventions for coronavirus disease 2019

We here consider data from a study which aimed to assess the comparative effectiveness of remdesivir and tocilizumab for reducing mortality in hospitalised COVID-19 patients. 80 trials were identified from two published network meta-analyses (Selvarajan et al. 2022), (Siemieniuk et al. 2020), a living COVID-19 trial database (COVID-NMA Initiative) [Covid-NMA.com], and a clinical trial database [clinicaltrials.gov]. Trials were included in this study if the patient population included hospitalized COVID-19 patients, active treatment was remdesivir or tocilizumab, comparator treatment was placebo or standard care, short-term mortality data was available, and the trial was published. 21 trials were included. For included trials, a risk of bias score was extracted from the COVID-NMA Initiative.

studlab	treat1	treat2	event1	n1	event2	n2
Ader	REM	STD	34	414	37	418

(continued)

studlab	treat1	treat2	event1	n1	event2	n2
Beigel (ACTT-1)	REM	STD	59	541	77	521
Broman	TOCI	STD	1	57	0	29
Criner	REM	STD	4	384	4	200
Declerq (COV-AID)	TOCI	STD	10	81	9	74
Gordon (REMAP-CAP)	TOCI	STD	83	353	116	358
Hermine (CORIMUNO)	TOCI	STD	7	63	8	67
Horby (RECOVERY)	TOCI	STD	621	2022	729	2094
Islam	REM	STD	0	30	0	30
Mahajan	REM	STD	5	34	3	36
Pan (WHO Solidarity)	REM	STD	602	4146	643	4129
Rosas (COVACTA)	TOCI	STD	58	294	28	144
Rutgers	TOCI	STD	21	174	34	180
Salama (EMPACTA)	TOCI	STD	26	249	11	128
Salvarani	TOCI	STD	2	60	1	63
Soin (COVINTOC)	TOCI	STD	11	92	15	88
Spinner	REM	STD	5	384	4	200
Stone (BACC-BAY)	TOCI	STD	9	161	4	82
Talaschian	TOCI	STD	5	17	4	19
Veiga (TOCIBRAS)	TOCI	STD	14	65	6	64
Wang	REM	STD	22	158	10	78

The corresponding network is displayed below:

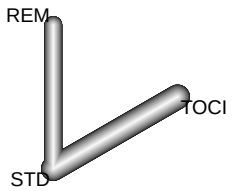


Figure 6.1: Evidence network of the 21 coronavirus-19 trials

We use the following command to calculate the log odds ratios and corresponding standard errors for each study:

```
covid <- pairwise(treat = treat, event = event, n = n, studlab = studlab, sm = "OR")
head(covid)
```

TE	seTE	studlab	treat1	treat2	event1	n1	event2	n2	incr	alls
-0.0819293	0.2483849	Ader	REM	STD	34	414	37	418	0.0	FAI
-0.3483875	0.1851030	Beigel (ACTT-1)	REM	STD	59	541	77	521	0.0	FAI
0.4487619	1.6487159	Broman	TOCI	STD	1	57	0	29	0.5	FAI
-0.6620566	0.7125543	Criner	REM	STD	4	384	4	200	0.0	FAI
0.0170679	0.4904898	Declerq (COV-AID)	TOCI	STD	10	81	9	74	0.0	FAI
-0.4442338	0.1688337	Gordon (REMAP-CAP)	TOCI	STD	83	353	116	358	0.0	FAI

Below, we conduct a random effects network meta-analysis where we consider standard care (`STD`) as the control treatment. Note that we have one study where zero cell counts occur, this study will not contribute to the NMA as the log odds ratio and its standard error cannot be determined.

```
NMA.covid <- netmeta(TE = TE, seTE = seTE, treat1 = treat1, treat2 = treat2,
                     studlab = studlab, data = covid, sm = "OR", ref = "STD",
                     comb.random = TRUE, common = FALSE, warn = FALSE)
NMA.covid
```

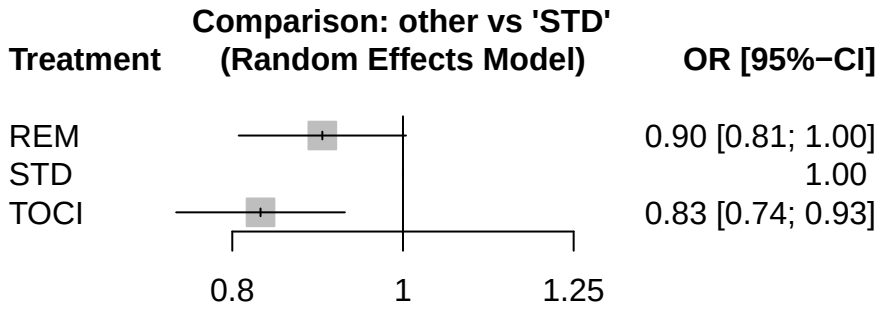
```
Number of studies: k = 20
Number of pairwise comparisons: m = 20
Number of treatments: n = 3
Number of designs: d = 2

Random effects model

Treatment estimate (sm = 'OR', comparison: other treatments vs 'STD'):
         OR           95%-CI     z p-value
REM  0.8999 [0.8067; 1.0039] -1.89  0.0588
STD       .                .     .       .
TOCI 0.8301 [0.7434; 0.9268] -3.31  0.0009

Quantifying heterogeneity / inconsistency:
tau^2 = 0; tau = 0; I^2 = 0% [0.0%; 48.9%]

Tests of heterogeneity (within designs) and inconsistency (between designs):
                    Q d.f. p-value
Total           16.38   18  0.5663
Within designs  16.38   18  0.5663
Between designs  0.00    0      --
```

A league table of the treatment effect estimates is given below:

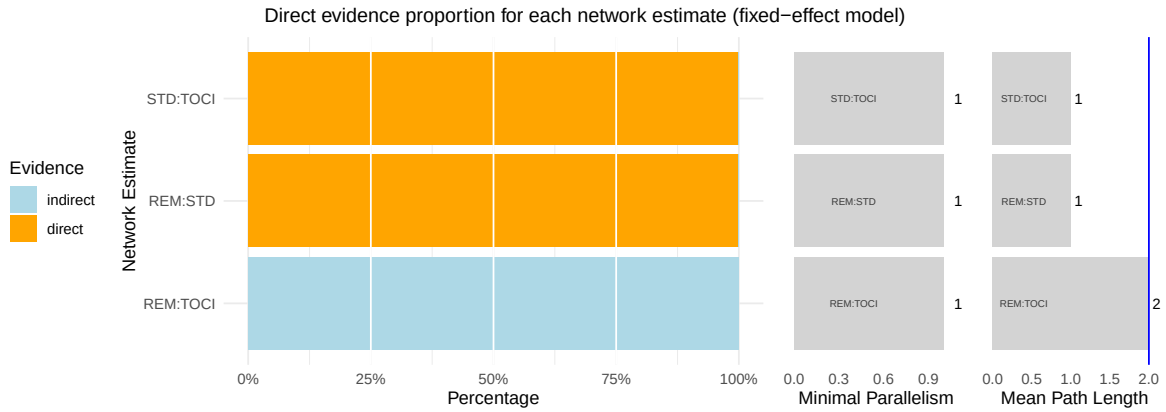
```
netleague(NMA.covid)
```

```
League table (random effects model):

                    REM 0.8999 [0.8067; 1.0039]                         .
 0.8999 [0.8067; 1.0039]                     STD 1.2047 [1.0789; 1.3451]
 1.0842 [0.9282; 1.2663] 1.2047 [1.0789; 1.3451]                    TOCI
```

We can also present the results in a forest plot:



The figure below shows the percentage of direct and indirect evidence used for each estimated comparison.



We now consider a Bayesian random effects network meta-analysis that analyzes the observed event counts using a binomial link function.

```
bdata <- data.frame(study = studlab,
                    treatment = treat,
                    responders = event,
                    sampleSize = n)

network <- mtc.network(data.ab  = bdata)

model <- mtc.model(network,
                   likelihood = "binom",
                   link = "log",
                   linearModel = "random",
                   n.chain = 3)
```

```
# Adaptation
mcmc1 <- mtc.run(model, n.adapt = 1000, n.iter = 1000, thin = 10)
```

```
Compiling model graph
   Resolving undeclared variables
   Allocating nodes
Graph information:
   Observed stochastic nodes: 42
   Unobserved stochastic nodes: 45
   Total graph size: 930

Initializing model
```

```r
# Sampling
mcmc2 <- mtc.run(model, n.adapt = 10000, n.iter = 100000, thin = 10)
```

```
Compiling model graph
   Resolving undeclared variables
   Allocating nodes
Graph information:
   Observed stochastic nodes: 42
   Unobserved stochastic nodes: 45
   Total graph size: 930

Initializing model
```

We can extract the pooled treatment effect estimates from the posterior distribution. When using `STD` as control group, we have:

```r
summary(relative.effect(mcmc2, t1 = "STD"))
```

```
Results on the Log Risk Ratio scale

Iterations = 10010:110000
Thinning interval = 10
Number of chains = 3
Sample size per chain = 10000

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

              Mean      SD  Naive SE Time-series SE
d.STD.REM  -0.1075 0.09755 0.0005632      0.0008145
d.STD.TOCI -0.1126 0.08254 0.0004765      0.0008494
sd.d        0.1129 0.08934 0.0005158      0.0016915

2. Quantiles for each variable:

                2.5%      25%      50%      75%   97.5%
d.STD.REM  -0.317752 -0.15982 -0.10226 -0.05105 0.08262
d.STD.TOCI -0.256122 -0.16379 -0.12024 -0.06997 0.07558
sd.d        0.005699  0.04386  0.09278  0.15994 0.33299
```

The corresponding odds ratios are as follows:

Comparison	95% CrI
REM vs. STD	0.9 (0.73; 1.09)
TOCI vs. STD	0.89 (0.77; 1.08)
REM vs. TOCI	1.02 (0.75; 1.27)

Finally, we expand the COVID-19 network with trials investigating the effectiveness of hydroxychloroquine (HCQ), lopinavir/ritonavir (LOPI), dexamethasone (DEXA) or interferon-β (INTB) (Selvarajan et al. 2022). The corresponding network is displayed below:

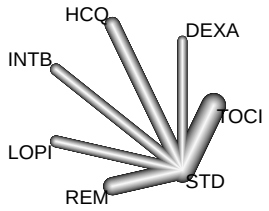


Figure 6.2: Evidence network of the 33 coronavirus-19 trials

We conducted a random effects network meta-analysis, results are depicted below:

```
Number of studies: k = 33
Number of pairwise comparisons: m = 33
Number of treatments: n = 7
Number of designs: d = 6

Random effects model

Treatment estimate (sm = 'OR', comparison: other treatments vs 'STD'):
         OR           95%-CI     z p-value            95%-PI
DEXA 0.8557 [0.7558; 0.9688] -2.46  0.0139  [0.7463; 0.9812]
HCQ  1.1809 [0.8934; 1.5610]  1.17  0.2428  [0.8786; 1.5872]
INTB 1.1606 [0.9732; 1.3841]  1.66  0.0973  [0.9604; 1.4026]
LOPI 1.0072 [0.8906; 1.1392]  0.11  0.9085  [0.8794; 1.1537]
REM  0.8983 [0.8014; 1.0070] -1.84  0.0658  [0.7913; 1.0199]
```

```
STD       .                 .     .       .                 .
TOCI 0.8304 [0.7410; 0.9306] -3.20  0.0014  [0.7316; 0.9426]

Quantifying heterogeneity / inconsistency:
tau^2 = 0.0004; tau = 0.0205; I^2 = 0.6% [0.0%; 42.3%]

Tests of heterogeneity (within designs) and inconsistency (between designs):
                    Q d.f. p-value
Total           27.18   27  0.4543
Within designs  27.18   27  0.4543
Between designs  0.00    0      --
```

We can calculate the P score for each treatment as follows:

```
netrank(NMA.covidf)
```

```
     P-score
TOCI  0.9070
DEXA  0.8357
REM   0.7143
STD   0.4027
LOPI  0.3899
HCQ   0.1336
INTB  0.1166
```

6.3.2 Pharmacologic treatments for chronic obstructive pulmonary disease

In this example, we consider the resuls from a systematic review of randomized controlled trials on pharmacologic treatments for chronic obstructive pulmonary disease (Baker, Baker, and Coleman 2009). The primary outcome, occurrence of one or more episodes of COPD exacerbation, is binary (yes / no). For this outcome, five drug treatments (fluticasone, budesonide, salmeterol, formoterol, tiotropium) and two combinations (fluticasone + salmeterol, budesonide + formoterol) were compared to placebo. The authors considered the two combinations as separate treatments instead of evaluating the individual components.

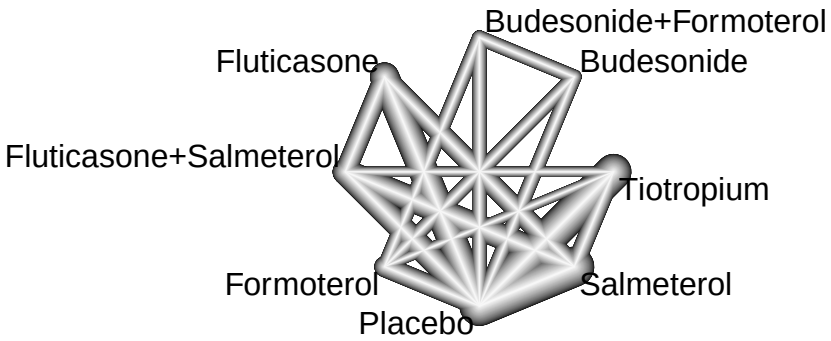
```
data(Baker2009)
```

study	year	id	treatment	exac	total
Llewellyn-Jones 1996	1996	1	Fluticasone	0	8
Llewellyn-Jones 1996	1996	1	Placebo	3	8
Boyd 1997	1997	2	Salmeterol	47	229
Boyd 1997	1997	2	Placebo	59	227
Paggiaro 1998	1998	3	Fluticasone	45	142
Paggiaro 1998	1998	3	Placebo	51	139

```
Baker <- pairwise(treat = treatment,
                  event = exac,
                  n = total,
                  studlab = id,
                  sm = "OR",
                  data = Baker2009)

NMA.COPD <- netmeta(TE = TE, seTE = seTE, treat1 = treat1, treat2 = treat2,
                    studlab = studlab, data = Baker, sm="OR", ref = "Placebo",
                    comb.random = TRUE)
```

```
Warning: Comparisons with missing TE / seTE or zero seTE not considered in
network meta-analysis.

Comparisons not considered in network meta-analysis:
 studlab                  treat1     treat2 TE seTE
      39 Fluticasone+Salmeterol    Placebo NA   NA
      39 Fluticasone+Salmeterol Salmeterol NA   NA
      39              Salmeterol    Placebo NA   NA
```

```
netgraph(NMA.COPD)
```



6.3.3 Advanced Therapies for Ulcerative Colitis

In this example, we consider a systematic literature review of Phase 3 randomized controlled trials investigating the following advanced therapies: infliximab, adalimumab, vedolizumab, golimumab, tofacitinib, ustekinumab, filgotinib, ozanimod, and upadacitinib (Panaccione et al. 2023). This review included 48 RCTs, from which 23 were found eligible for inclusion in a network meta-analysis. The included RCT populations were largely comparable in their baseline characteristics, though some heterogeneity was noted in weight, disease duration, extent of disease, and concomitant medications. A risk of bias assessment showed a low risk of bias for all included RCTs, which were all industry sponsored.

We here focus on the synthesis of 18 trials that contributed efficacy data for induction in bio-naive populations. The following FDA- and/or EMA-approved biologic or SMD doses were investigated:

- Adalimumab subcutaneous 160 mg at week 0, 80 mg at week 2, and 40 mg at week 4 (`ADA160/80`)
- Infliximab intravenous 5 mg/kg (`INF5`) at weeks 0, 2, and 6 then every 8 weeks
- Infliximab intravenous 10 mg/kg (`INF10`) at weeks 0, 2, and 6 then every 8 weeks
- Filgotinib oral 100 mg once daily (`FIL100`)
- Filgotinib oral 200 mg once daily (`FIL200`)
- Golimumab subcutaneous 200 mg at week 0 and 100 mg at week 2 (`GOL200/100`)

- Ozanimod oral 0.23 mg once daily for 4 days, 0.46 mg once daily for 3 days, then 0.92 mg once daily (`OZA0.92`)
- Tofacitinib oral 10 mg twice daily for 8 weeks (`TOF10`)
- Upadacitinib oral 45 mg once daily for 8 weeks (`UPA45`)
- Ustekinumab intravenous 6 mg/kg at week 0 (`UST6`)
- Vedolizumab intravenous 300 mg at weeks 0, 2, and 6 (`VED300`)

The reference treatment is placebo (`PBO`).

Table 6.1: Efficacy outcomes (i.e., clinical remission) data of induction bio-naïve populations

studlab	treat1	treat2	event1	n1	event2	n2
ACT-1	INF10	INF5	39	122	47	121
ACT-1	INF10	PBO	39	122	18	121
ACT-1	INF5	PBO	47	121	18	121
ACT-2	INF10	INF5	33	120	41	121
ACT-2	INF10	PBO	33	120	7	123
ACT-2	INF5	PBO	41	121	7	123
GEMINI 1	VED300	PBO	30	130	5	76
Japic CTI-060298	INF5	PBO	21	104	11	104
Jiang 2015	INF5	PBO	22	41	9	41
M10-447	ADA160/80	PBO	9	90	11	96
NCT01551290	INF5	PBO	11	50	5	49
NCT02039505	VED300	PBO	22	79	6	41
OCTAVE 1	TOF10	PBO	56	222	9	57
OCTAVE 2	TOF10	PBO	43	195	4	47
PURSUIT-SC	GOL200/100	PBO	45	253	16	251
SELECTION	FIL100	FIL200	47	277	60	245
SELECTION	FIL100	PBO	47	277	17	137
SELECTION	FIL200	PBO	60	245	17	137
TRUE NORTH	OZA0.92	PBO	66	299	10	151
U-ACCOMPLISH	UPA45	PBO	54	166	3	81
U-ACHIEVE Study 2	UPA45	PBO	41	145	4	72
ULTRA-1	ADA160/80	PBO	24	130	12	130
ULTRA-2	ADA160/80	PBO	32	150	16	145
UNIFI	UST6	PBO	27	147	15	151

The corresponding network is displayed below:

Below, we conduct a random effects network meta-analysis of the reported study effects (expressed as odds ratio) and consider placebo (`treat = "PBO"`) as the control treatment.

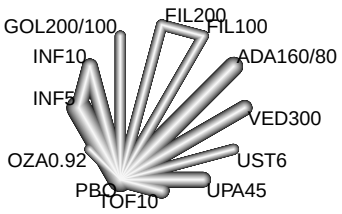


Figure 6.3: Evidence network of 18 trials that contributed efficacy data for induction in bio-naive populations

```
NMA.uc <- netmeta(TE = TE, seTE = seTE, treat1 = treat1, treat2 = treat2,
                  studlab = studlab, data = UlcerativeColitis, sm = "OR",
                  ref = "PBO", common = FALSE, comb.random = TRUE)
NMA.uc
```

All treatments except `FIL100` and `UST6` are significantly more efficacious than `PBO` at inducing clinical remission. We can now estimate the probabilities of each treatment being at each possible rank and the SUCRAs (Surface Under the Cumulative RAnking curve):

```
sucra.uc <- rankogram(NMA.uc, nsim = 100, random = TRUE, common = FALSE,
                      small.values = "undesirable")

# Exctract the SUCRA values
sucra.uc$ranking.random
```

```
 ADA160/80     FIL100     FIL200 GOL200/100      INF10       INF5    OZA0.92
0.26000000 0.17363636 0.44818182 0.65363636 0.59181818 0.76454545 0.74909091
       PBO      TOF10      UPA45       UST6     VED300
0.01363636 0.40545455 0.97272727 0.34909091 0.61818182
```

These results indicate that 97.3% of the evaluated treatments are worse than `UPA45`.

Version info

This chapter was rendered using the following version of R and its packages:

```
R version 4.2.3 (2023-03-15 ucrt)
Platform: x86_64-w64-mingw32/x64 (64-bit)
Running under: Windows 10 x64 (build 19045)

Matrix products: default

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

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

other attached packages:
[1] dmetar_0.0.9000  netmeta_2.8-2    meta_6.5-0       gemtc_1.0-1
[5] coda_0.19-4      dplyr_1.1.2      kableExtra_1.3.4

loaded via a namespace (and not attached):
 [1] httr_1.4.6          magic_1.6-1         jsonlite_1.8.5
 [4] viridisLite_0.4.2   splines_4.2.3       stats4_4.2.3
 [7] metafor_4.2-0       slam_0.1-50         yaml_2.3.7
[10] robustbase_0.99-0   ggrepel_0.9.3       numDeriv_2016.8-1.1
[13] pillar_1.9.0        lattice_0.21-8      glue_1.6.2
[16] digest_0.6.31       rvest_1.0.3         minqa_1.2.5
[19] colorspace_2.1-0    MuMIn_1.47.5        htmltools_0.5.5
[22] Matrix_1.5-4.1      plyr_1.8.8          pkgconfig_2.0.3
[25] mvtnorm_1.2-2       Rglpk_0.6-5         scales_1.2.1
[28] webshot_0.5.4       svglite_2.1.1       rjags_4-14
[31] metadat_1.2-0       lme4_1.1-33         tibble_3.2.1
[34] farver_2.1.1        generics_0.1.3      ggplot2_3.4.2
[37] withr_2.5.0         nnet_7.3-19         cli_3.6.1
[40] magrittr_2.0.3      mclust_6.0.0        evaluate_0.21
[43] fansi_1.0.4         nlme_3.1-162        MASS_7.3-60
[46] truncnorm_1.0-9     forcats_1.0.0       xml2_1.3.4
[49] class_7.3-22        tools_4.2.3         lifecycle_1.0.3
[52] stringr_1.5.0       kernlab_0.9-32      munsell_0.5.0
[55] cluster_2.1.4       fpc_2.2-10          compiler_4.2.3
[58] systemfonts_1.0.4   rlang_1.1.1         grid_4.2.3
[61] nloptr_2.0.3        rstudioapi_0.14     CompQuadForm_1.4.3
[64] igraph_1.5.0        labeling_0.4.2      rmarkdown_2.22
[67] boot_1.3-28.1       gtable_0.3.3        codetools_0.2-19
[70] abind_1.4-5         flexmix_2.3-19      R6_2.5.1
```

```
[73] gridExtra_2.3       knitr_1.43          prabclus_2.3-2
[76] fastmap_1.1.1       utf8_1.2.3          mathjaxr_1.6-0
[79] poibin_1.5          modeltools_0.2-23   stringi_1.7.12
[82] parallel_4.2.3      Rcpp_1.0.10         vctrs_0.6.3
[85] DEoptimR_1.0-14     tidyselect_1.2.0    xfun_0.39
[88] diptest_0.76-0
```

References

7 Dealing with irregular and informative visits

Janie Coulombe (Université de Montréal)
Thomas Debray (Smart Data Analysis and Statistics B.V.)

7.1 Introduction

We first load the required packages

```
library(dplyr)
library(broom)
library(ggplot2)
library(mice)
```

Subsequently, we load the relevant R scripts:

```
source("resources/chapter12_sim.r")
```

```
Loading required package: nlme
```

```

Attaching package: 'nlme'
```

```
The following object is masked from 'package:dplyr':

    collapse
```

```
Loading required package: MASS
```

```

Attaching package: 'MASS'
```

```
The following object is masked from 'package:dplyr':

    select

Loading required package: truncnorm
```

```r
source("resources/chapter12_fig_functions.r")
source("resources/chapter12_mlmi.r")
```

7.2 Example dataset

Below, we generate an example dataset that contains information on the treatment allocation `x` and three baseline covariates `age`, `sex` and `edss` (EDSS at treatment start). The discrete outcome `y` represents the Expanded Disability Status Scale (EDSS) score after `time` months of treatment exposure. Briefly, the EDSS is a semi-continuous measure that varies from 0 (no disability) to 10 (death).

```r
set.seed(9843626)

dataset  <- sim_data_EDSS(npatients = 500,
                          ncenters = 10,
                          follow_up = 12*5, # Total follow-up (number of months)
                          sd_a_t = 0.5,   # DGM - Within-visit variation in EDSS scores
                          baseline_EDSS = 1.3295,    # DGM - Mean baseline EDDS score
                          sd_alpha_ij = 1.46,    # DGM - Between-subject variation in base
                          sd_beta1_j = 0.20,    # DGM - Between-site variation in baseline
                          mean_age = 42.41,
                          sd_age = 10.53,
                          min_age = 18,
                          beta_age = 0.05, # DGM - prognostic effect of age
                          beta_t = 0.014,  # DGM - prognostic effect of time
                          beta_t2 = 0,    # DGM - prognostic effect of time squared
                          delta_xt = 0, # DGM - interaction treatment time
                          delta_xt2 = 0, # 0.0005    # DGM - interaction treatment time2
                          p_female = 0.75,
                          beta_female = -0.2 ,  ## DGM - prognostic effect of male sex
                          delta_xf = 0,      ## DGM - interaction sex treatment
                          rho = 0.8,             # DGM - autocorrelation of between alpha_
                          corFUN = corAR1,       # DGM - correlation structure of the late
                          tx_alloc_FUN = treatment_alloc_confounding_v2 ) ## or treatment_
```

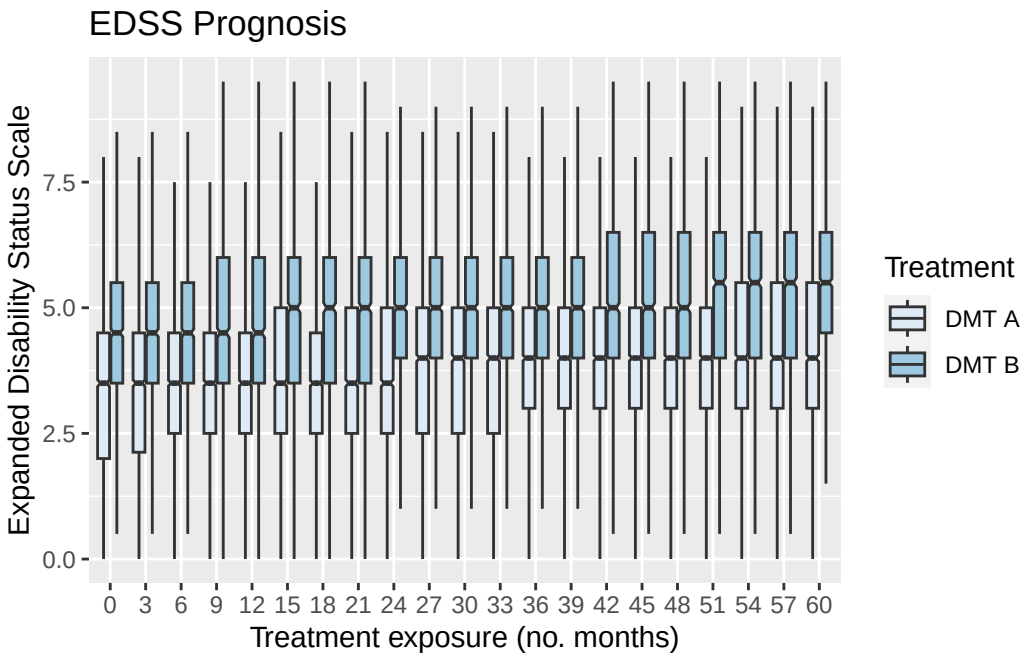


Figure 7.1: Distribution of the EDSS score at each time point

We remove the outcome `y` according to the informative visit process that depends on the received treatment, gender, and age.

```
dataset_visit <- censor_visits_a5(dataset, seed = 12345) %>%
  dplyr::select(-y) %>%
  mutate(time_x = time*x)
```

In the censored data, a total of 17 out of 5000 patients have a visit at `time=60`.

7.3 Estimation of treatment effect

We will estimate the marginal treatment effect at time `time=60`.

7.3.1 Original data

```
origdat60 <- dataset %>% filter(time == 60)

# Predict probability of treatment allocation
fitps <- glm(x ~ age + sex + edss, family = 'binomial',
             data = origdat60)

# Derive the propensity score
origdat60 <- origdat60 %>% mutate(ipt = ifelse(x == 1, 1/predict(fitps, type = 'response')
                                               1/(1-predict(fitps, type = 'response'))))

# Estimate
fit_ref_m <- tidy(lm(y ~ x, weight = ipt, data = origdat60), conf.int = TRUE)
```

7.3.2 Doubly-weighted marginal treatment effect

We here implement inverse probability of response weights into the estimating equations to adjust for nonrandom missingness Coulombe, Moodie, and Platt (2020).

```
obsdat60 <- dataset_visit %>% mutate(visit = ifelse(is.na(y_obs),0,1)) %>% filter(time ==

gamma <- glm(visit ~ x + sex + age + edss, family = 'binomial', data = obsdat60)$coef

obsdat60 <- obsdat60 %>% mutate(rho_i = 1/exp(gamma["(Intercept)"] +
                                                gamma["x"]*x +
                                                gamma["sex"]*sex +
                                                gamma["age"]*age))

# Predict probability of treatment allocation
fitps <- glm(x ~ age + sex + edss, family='binomial', data = obsdat60)

# Derive the propensity score
obsdat60 <- obsdat60 %>% mutate(ipt = ifelse(x==1, 1/predict(fitps, type='response'),
                                             1/(1-predict(fitps, type='response'))))


fit_w <- tidy(lm(y_obs ~ x, weights = ipt*rho_i, data = obsdat60), conf.int = TRUE)
```

7.3.3 Multilevel multiple imputation

We adopt the imputation approach proposed by Debray et al. (2023). Briefly, we impute the entire vector of `y_obs` for all 61 potential visits and generate 10 imputed datasets. Note: `mlmi` currently does not support imputation of treatment-covariate interaction terms.

```
imp <- impute_y_mice_3l(dataset_visit, seed = 12345)
```

We can now estimate the treatment effect in each imputed dataset

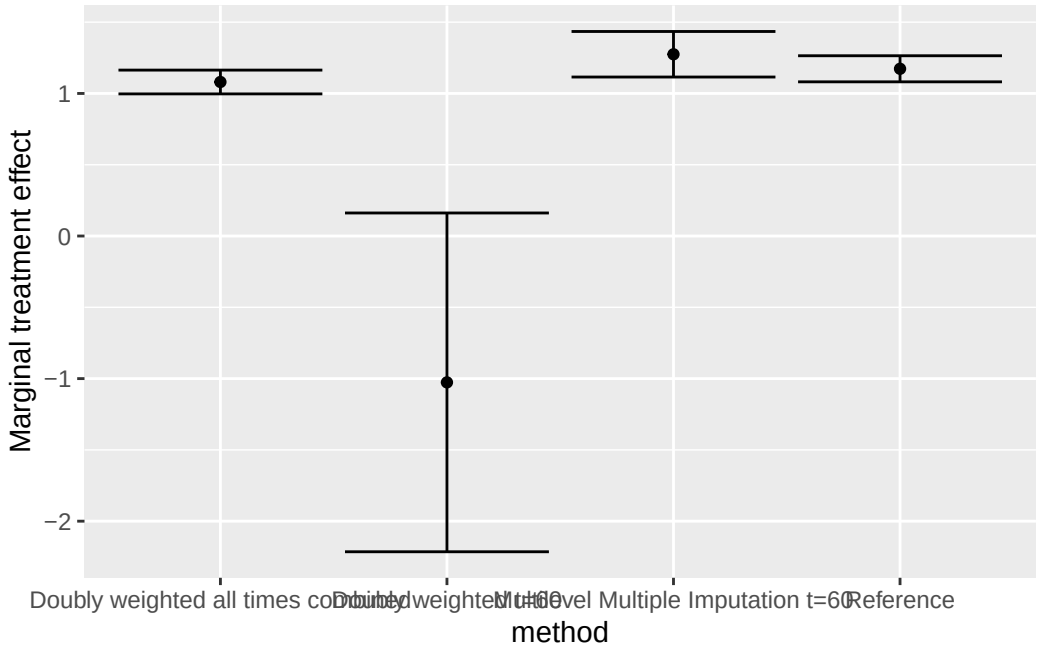
```
# Predict probability of treatment allocation
fitps <- glm(x ~ age + sex + edss, family='binomial', data = dataset_visit)

# Derive the propensity score
dataset_visit <- dataset_visit %>% mutate(ipt = ifelse(x==1, 1/predict(fitps, type='respon
                                                       1/(1-predict(fitps, type='response'

Q <- U <- rep(NA, 10) # Error variances

for (i in seq(10)) {
  dati <- cbind(dataset_visit[,c("x","ipt","time")], y_imp = imp[,i]) %>% filter(time == 6

  # Estimate
  fit <- tidy(lm(y_imp ~ x, weight = ipt, data = dati), conf.int = TRUE)

  Q[i] <- fit %>% filter(term == "x") %>% pull(estimate)
  U[i] <- (fit %>% filter(term == "x") %>% pull(std.error))**2
}

fit_mlmi <- pool.scalar(Q = Q, U = U)
```

7.4 Reproduce the results using all data to compute the marginal effect with IIV-weighted

7.4.1 Doubly -weighted marginal treatment effect total

```
obsdatall <- dataset_visit %>% mutate(visit = ifelse(is.na(y_obs),0,1))
gamma <- glm(visit ~ x + sex + age + edss, family = 'binomial', data = obsdatall)$coef
obsdatall <- obsdatall %>% mutate(rho_i = 1/exp(gamma["(Intercept)"] +
                                                gamma["x"]*x +
                                                gamma["sex"]*sex +
                                                gamma["age"]*age))
# Predict probability of treatment allocation
fitps <- glm(x ~ age + sex + edss, family='binomial', data = obsdatall)
# Derive the propensity score
obsdatall <- obsdatall %>% mutate(ipt = ifelse(x==1, 1/predict(fitps, type='response'),
                                               1/(1-predict(fitps, type='response'))))
fit_w <- tidy(lm(y_obs ~ x, weights = ipt*rho_i, data = obsdatall), conf.int = TRUE)
```

7.5 Results



Version info

This chapter was rendered using the following version of R and its packages:

```
R version 4.2.3 (2023-03-15 ucrt)
Platform: x86_64-w64-mingw32/x64 (64-bit)
Running under: Windows 10 x64 (build 19045)

Matrix products: default

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

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

other attached packages:
[1] truncnorm_1.0-9 MASS_7.3-60     nlme_3.1-162    mice_3.16.0
[5] ggplot2_3.4.2   broom_1.0.5     dplyr_1.1.2

loaded via a namespace (and not attached):
 [1] shape_1.4.6        tidyselect_1.2.0   xfun_0.39          purrr_1.0.1
 [5] splines_4.2.3      lattice_0.21-8     colorspace_2.1-0   vctrs_0.6.3
 [9] generics_0.1.3     htmltools_0.5.5    yaml_2.3.7         pan_1.6
[13] utf8_1.2.3         survival_3.5-5     rlang_1.1.1        jomo_2.7-6
[17] pillar_1.9.0       nloptr_2.0.3       glue_1.6.2         withr_2.5.0
[21] RColorBrewer_1.1-3 foreach_1.5.2      lifecycle_1.0.3    munsell_0.5.0
[25] gtable_0.3.3       codetools_0.2-19   evaluate_0.21      labeling_0.4.2
[29] knitr_1.43         fastmap_1.1.1      fansi_1.0.4        Rcpp_1.0.10
[33] scales_1.2.1       backports_1.4.1    jsonlite_1.8.5     farver_2.1.1
[37] lme4_1.1-33        digest_0.6.31      grid_4.2.3         cli_3.6.1
[41] tools_4.2.3        magrittr_2.0.3     glmnet_4.1-7       tibble_3.2.1
[45] tidyr_1.3.0        pkgconfig_2.0.3    ellipsis_0.3.2     Matrix_1.5-4.1
[49] minqa_1.2.5        rmarkdown_2.22     rstudioapi_0.14    iterators_1.0.14
[53] rpart_4.1.19       mitml_0.4-5        R6_2.5.1           boot_1.3-28.1
[57] nnet_7.3-19        compiler_4.2.3
```

References

8 Prediction of individual treatment effect using data from multiple studies

Orestis Efthimiou (Institute of Social and Preventive Medicine (ISPM))

In this chapter, we discuss statistical methods for developing models to predict patient-level treatment effects using data from multiple randomized and non-randomized studies. We will first present prediction models that assume a constant treatment effect and discuss how to address heterogeneity in baseline risk. Subsequently, we will discuss approaches that allow for treatment effect modification by adopting two different approaches in an IPD-MA context, namely the risk modelling and the effect modelling approach. For both approaches, we will first discuss how to combine IPD from RCTs comparing the same two treatments. We will then discuss how these methods can be extended to include randomized data from multiple treatments, real-world data, and published aggregate data. We will discuss statistical software to implement these approaches and provide example code as supporting information. Real examples will be used throughout to illustrate the main methods.

8.1 Estimating heterogeneous treatment effects in pairwise meta-analysis

We hereby provide code for estimating patient-level treatment effects for the case when we have patient-level data from multiple randomized trials.

8.1.1 Example of a continuous outcome

8.1.1.1 Setup

We start by simulating an artificial dataset using the R package **bipd**:

```
library(bipd)
ds <- generate_ipdma_example(type = "continuous")
```

Table 8.1: The simulated dataset with a continuous outcome

	0	1	Overall
	(N=321)	(N=279)	(N=600)
z1			
Mean (SD)	0.0681 (0.967)	0.0166 (0.983)	0.0441 (0.974)
Median [Min, Max]	0.0464 [-3.13, 2.87]	-0.0316 [-2.69, 2.50]	0.0244 [-3.13, 2.87]
z2			
Mean (SD)	-0.0961 (0.990)	0.0117 (1.01)	-0.0459 (1.00)
Median [Min, Max]	-0.117 [-2.55, 2.52]	0.0838 [-3.54, 3.38]	-0.0371 [-3.54, 3.38]
studyid			
1	57 (17.8%)	43 (15.4%)	100 (16.7%)
2	61 (19.0%)	39 (14.0%)	100 (16.7%)
3	50 (15.6%)	50 (17.9%)	100 (16.7%)
4	51 (15.9%)	49 (17.6%)	100 (16.7%)
5	57 (17.8%)	43 (15.4%)	100 (16.7%)
6	45 (14.0%)	55 (19.7%)	100 (16.7%)

Let us have a look at the dataset:

```
head(ds)
```

```
  studyid treat         z1         z2  y
1       1     0  0.3266390 -1.7236365 11
2       1     0  0.4951180 -2.2374921 11
3       1     0 -0.7056270 -1.3584076 11
4       1     1 -0.9785132 -0.3029899  8
5       1     0  0.2651324 -0.1735689 11
6       1     1  1.3853282 -0.2320760  6
```

The simulated dataset contains information on the following variables:

- the trial indicator `studyid`
- the treatment indicator `treat`, which takes the values 0 for control and 1 for active treatment
- two prognostic variables `z1` and `z2`
- the continuous outcome `y`

8.1.1.2 Model fitting

We synthesize the evidence using a Bayesian random effects meta-analysis model. The model is given in Equation 16.7 of the book. First we need set up the data and create the model:

```
ipd <- with(ds, ipdma.model.onestage(y = y, study = studyid, treat = treat,
                                     X = cbind(z1, z2),
                                     response = "normal",
                                     shrinkage = "none"),
                                     type="random")
```

The JAGS model can be accessed as follows:

```
ipd$model.JAGS
```

```
function ()
{
    for (i in 1:Np) {
        y[i] ~ dnorm(mu[i], sigma)
        mu[i] <- alpha[studyid[i]] + inprod(beta[], X[i, ]) +
            (1 - equals(treat[i], 1)) * inprod(gamma[], X[i,
                ]) + d[studyid[i], treat[i]]
    }
    sigma ~ dgamma(0.001, 0.001)
    for (j in 1:Nstudies) {
        d[j, 1] <- 0
        d[j, 2] ~ dnorm(delta[2], tau)
    }
    sd ~ dnorm(0, 1)
    T(0, )
    tau <- pow(sd, -2)
    delta[1] <- 0
    delta[2] ~ dnorm(0, 0.001)
    for (j in 1:Nstudies) {
        alpha[j] ~ dnorm(0, 0.001)
    }
    for (k in 1:Ncovariate) {
        beta[k] ~ dnorm(0, 0.001)
    }
    for (k in 1:Ncovariate) {
        gamma[k] ~ dnorm(0, 0.001)
    }
```

```
}
<environment: 0x000002aa85657fc0>
```

We can fit the treatment effect model as follows:

```
samples <- ipd.run(ipd, n.chains = 2, n.iter = 20,
                   pars.save = c("alpha", "beta", "delta", "sd", "gamma"))
```

```
Compiling model graph
   Resolving undeclared variables
   Allocating nodes
Graph information:
   Observed stochastic nodes: 600
   Unobserved stochastic nodes: 19
   Total graph size: 6034

Initializing model
```

Here are the estimated model parameters:

```
summary(samples)
```

```
Iterations = 2001:2020
Thinning interval = 1
Number of chains = 2
Sample size per chain = 20

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

            Mean      SD Naive SE Time-series SE
alpha[1] 11.0321 0.04281 0.006768       0.007664
alpha[2]  7.9608 0.05458 0.008630       0.014489
alpha[3] 10.4386 0.04525 0.007154       0.012180
alpha[4]  9.6425 0.04047 0.006399       0.006279
alpha[5] 12.8051 0.05709 0.009027       0.009144
alpha[6] 15.7531 0.05324 0.008418       0.013281
beta[1]   0.1937 0.02572 0.004067       0.004113
beta[2]   0.3048 0.01767 0.002793       0.004289
```

```
delta[1]  0.0000 0.00000 0.000000       0.000000
delta[2] -3.1631 0.39061 0.061760       0.083400
gamma[1] -0.5218 0.03882 0.006138       0.007766
gamma[2]  0.5545 0.02791 0.004413       0.006817
sd        0.9859 0.31067 0.049121       0.046491

2. Quantiles for each variable:

            2.5%     25%     50%     75%   97.5%
alpha[1] 10.9608 11.0077 11.0287 11.0546 11.1227
alpha[2]  7.8596  7.9252  7.9515  8.0081  8.0494
alpha[3] 10.3403 10.4122 10.4347 10.4741 10.5208
alpha[4]  9.5711  9.6176  9.6434  9.6655  9.7222
alpha[5] 12.7144 12.7776 12.8041 12.8494 12.8866
alpha[6] 15.6683 15.7274 15.7593 15.7833 15.8413
beta[1]   0.1560  0.1732  0.1945  0.2120  0.2302
beta[2]   0.2774  0.2932  0.3059  0.3176  0.3322
delta[1]  0.0000  0.0000  0.0000  0.0000  0.0000
delta[2] -4.0330 -3.3678 -3.1120 -2.9324 -2.3805
gamma[1] -0.5864 -0.5553 -0.5313 -0.4896 -0.4614
gamma[2]  0.5134  0.5341  0.5510  0.5742  0.6078
sd        0.6095  0.7709  0.9309  1.0851  1.7235
```

8.1.1.3 Prection

We can now predict the individualized treatment effect for a new patient with covariate values `z1=1` and `z2=0.5`.

```r
round(treatment.effect(ipd, samples, newpatient = c(z1 = 1, z2 = 0.5)), 2)
```

```
0.025   0.5 0.975
-4.28 -3.32 -2.60
```

We can also predict treatment benefit for all patients in the sample, and look at the distribution of predicted benefit.

```r
library(dplyr)
library(ggplot2)

ds <- ds %>% mutate(benefit = NA)
```

```
for (i in seq(nrow(ds))) {
  newpat <- as.matrix(ds[i, c("z1", "z2")])
  ds$benefit[i] <- treatment.effect(ipd, samples, newpatient = newpat)["0.5"]
}

ggplot(ds, aes(x = benefit)) + geom_histogram() + facet_wrap(~studyid) +
  xlab("Predicted treatment benefit")
```

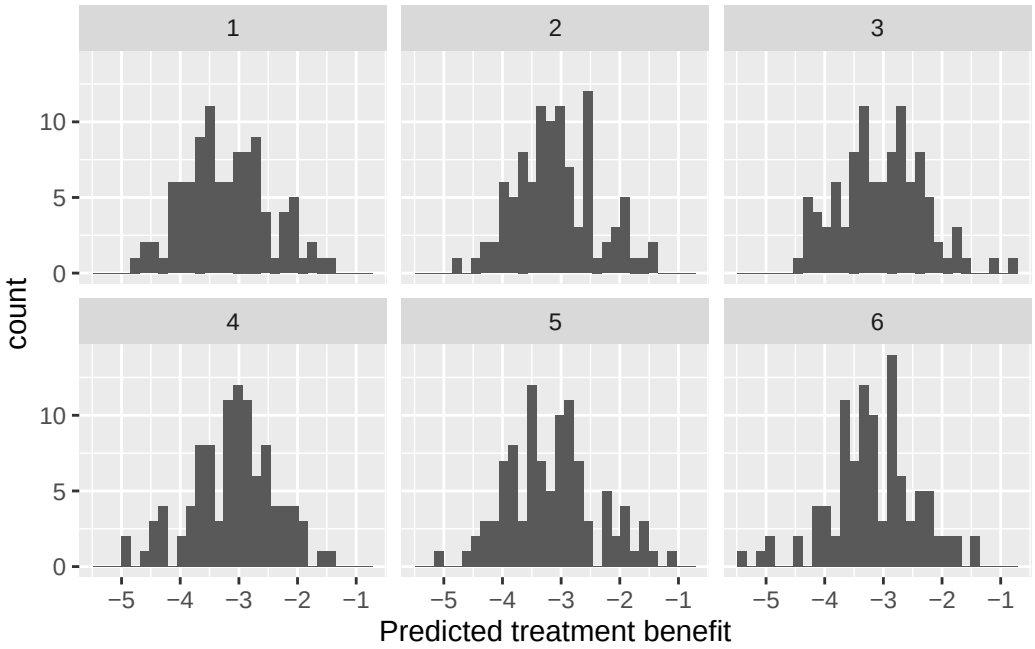


Figure 8.1: Distribution of predicted treatment benefit in each trial

8.1.1.4 Penalization

Let us repeat the analysis, but this time while penalizing the treatment-covariate coefficients using a Bayesian LASSO prior.

```
ipd <- with(ds, ipdma.model.onestage(y = y, study = studyid,
                                     treat = treat,
                                     X = cbind(z1, z2),
                                     response = "normal",
                                     shrinkage = "laplace"),
            type = "random")
```

```r
samples <- ipd.run(ipd, n.chains = 2, n.iter = 20,
                   pars.save = c("alpha", "beta", "delta", "sd", "gamma"))
```

```
Compiling model graph
   Resolving undeclared variables
   Allocating nodes
Graph information:
   Observed stochastic nodes: 600
   Unobserved stochastic nodes: 20
   Total graph size: 6039

Initializing model
```

```r
round(treatment.effect(ipd, samples, newpatient = c(1,0.5)), 2)
```

```
0.025   0.5 0.975
-5.07 -3.38 -2.45
```

8.1.2 Example of a binary outcome

8.1.2.1 Setup

We now present the case of a binary outcome. We first generate a dataset as before, using the **bipd** package.

```r
ds2 <- generate_ipdma_example(type = "binary")
head(ds2)
```

```
  studyid treat         w1           w2 y
1       1     0 -1.2303395  1.386443041 1
2       1     1  0.9995966  0.326715503 0
3       1     0 -1.1008584  1.060149816 0
4       1     1  2.8577405 -0.015462936 0
5       1     1 -1.4637407  0.398844344 0
6       1     0 -0.3153284  0.006682804 0
```

The simulated dataset contains information on the following variables:

Table 8.2: The simulated dataset with a binary outcome

	0	1	Overall
	(N=291)	(N=309)	(N=600)
w1			
Mean (SD)	0.00796 (0.958)	0.0314 (0.897)	0.0200 (0.926)
Median [Min, Max]	-0.00405 [-2.72, 2.78]	0.0593 [-2.00, 2.86]	0.0423 [-2.72, 2.86]
w2			
Mean (SD)	0.0459 (1.02)	-0.0446 (1.05)	-0.000697 (1.04)
Median [Min, Max]	0.107 [-2.70, 2.68]	-0.0511 [-3.32, 2.67]	0.000535 [-3.32, 2.68]
studyid			
1	52 (17.9%)	48 (15.5%)	100 (16.7%)
2	45 (15.5%)	55 (17.8%)	100 (16.7%)
3	51 (17.5%)	49 (15.9%)	100 (16.7%)
4	47 (16.2%)	53 (17.2%)	100 (16.7%)
5	45 (15.5%)	55 (17.8%)	100 (16.7%)
6	51 (17.5%)	49 (15.9%)	100 (16.7%)

- the trial indicator `studyid`
- the treatment indicator `treat`, which takes the values 0 for control and 1 for active treatment
- two prognostic variables `w1` and `w2`
- the binary outcome `y`

8.1.2.2 Model fitting

We use a Bayesian random effects model with binomial likelihood. This is similar to the model 16.7 of the book, but with a Binomial likelihood, i.e.

$$y_{ij} \sim Binomial(\pi_{ij})$$

$$logit(\pi_{ij}) == a_j + \delta_j t_{ij} + \sum_{l=1}^{L} \beta_l x_{ij} + \sum_{l=1}^{L} \gamma_l x_{ij} t_{ij}$$

The remaining of the model is as in the book. We can penalize the estimated parameters for effect modification (γ's), using a Bayesian LASSO. We can do this using again the *bipd* package:

```r
ipd2 <- with(ds2, ipdma.model.onestage(y = y, study = studyid, treat = treat,
                                       X = cbind(w1, w2),
                                       response = "binomial",
                                       shrinkage = "laplace"),
             type="random", hy.prior = list("dunif", 0, 1))

ipd2$model.JAGS
```

```
function ()
{
    for (i in 1:Np) {
        y[i] ~ dbern(p[i])
        logit(p[i]) <- alpha[studyid[i]] + inprod(beta[], X[i,
            ]) + (1 - equals(treat[i], 1)) * inprod(gamma[],
            X[i, ]) + d[studyid[i], treat[i]]
    }
    for (j in 1:Nstudies) {
        d[j, 1] <- 0
        d[j, 2] ~ dnorm(delta[2], tau)
    }
    sd ~ dnorm(0, 1)
    T(0, )
    tau <- pow(sd, -2)
    delta[1] <- 0
    delta[2] ~ dnorm(0, 0.001)
    for (j in 1:Nstudies) {
        alpha[j] ~ dnorm(0, 0.001)
    }
    for (k in 1:Ncovariate) {
        beta[k] ~ dnorm(0, 0.001)
    }
    tt <- lambda
    lambda <- pow(lambda.inv, -1)
    lambda.inv ~ dunif(0, 5)
    for (k in 1:Ncovariate) {
        gamma[k] ~ ddexp(0, tt)
    }
}
<environment: 0x000002aa847ff3e8>
```

```r
samples <- ipd.run(ipd2, n.chains = 2, n.iter = 20,
                   pars.save = c("alpha", "beta", "delta", "sd", "gamma"))
```

```
Compiling model graph
   Resolving undeclared variables
   Allocating nodes
Graph information:
   Observed stochastic nodes: 600
   Unobserved stochastic nodes: 19
   Total graph size: 6637

Initializing model
```

```r
summary(samples)
```

```
Iterations = 2001:2020
Thinning interval = 1
Number of chains = 2
Sample size per chain = 20

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

             Mean     SD Naive SE Time-series SE
alpha[1] -0.34078 0.2776  0.04389        0.06585
alpha[2] -0.80783 0.2718  0.04298        0.07790
alpha[3] -1.19985 0.3202  0.05063        0.08576
alpha[4] -0.85804 0.1990  0.03146        0.04716
alpha[5] -1.06959 0.2987  0.04724        0.10169
alpha[6] -0.94425 0.3131  0.04951        0.07678
beta[1]  -0.03590 0.1000  0.01582        0.01967
beta[2]  -0.03756 0.1218  0.01926        0.02130
delta[1]  0.00000 0.0000  0.00000        0.00000
delta[2] -0.01096 0.4379  0.06924        0.08327
gamma[1]  0.09028 0.1149  0.01816        0.02171
gamma[2]  0.21330 0.1498  0.02368        0.02245
sd        0.78142 0.2498  0.03950        0.07800

2. Quantiles for each variable:
```

```
               2.5%      25%      50%      75%   97.5%
alpha[1] -0.7605646 -0.50432 -0.37277 -0.15051  0.2115
alpha[2] -1.1484144 -1.00197 -0.86278 -0.60084 -0.2906
alpha[3] -1.8728435 -1.36581 -1.08526 -0.94346 -0.8677
alpha[4] -1.2660197 -0.99241 -0.81621 -0.74412 -0.5379
alpha[5] -1.6683012 -1.25329 -1.04303 -0.81286 -0.6004
alpha[6] -1.4772417 -1.16204 -0.92320 -0.84119 -0.3009
beta[1]  -0.2363999 -0.07803 -0.01780  0.03226  0.1232
beta[2]  -0.2262688 -0.13719 -0.04790  0.06893  0.1935
delta[1]  0.0000000  0.00000  0.00000  0.00000  0.0000
delta[2] -0.9416291 -0.24785  0.02087  0.23443  0.7910
gamma[1] -0.1021983  0.01017  0.09435  0.19198  0.2604
gamma[2]  0.0006925  0.10911  0.22227  0.29650  0.4860
sd        0.4516838  0.58578  0.74924  0.94277  1.2447
```

```r
round(treatment.effect(ipd2, samples, newpatient = c(w1= 1.6, w2 = 1.3)), 2)
```

```
0.025   0.5 0.975
 0.59  1.65  3.32
```

8.2 Estimating heterogeous treatment effects in network meta-analysis

8.2.1 Example of a continuous outcome

8.2.1.1 Setup

We use again the bipd package to simulate a dataset:

```r
ds3 <- generate_ipdnma_example(type = "continuous")
head(ds3)
```

```
  studyid treat         z1         z2  y
1       1     1  0.1829614 -0.3583425 11
2       1     1 -1.3498681 -0.4742485 11
3       1     1 -0.6371446  0.8219166 11
4       1     2 -2.2352659 -1.6490255  8
5       1     2  0.1722712  1.3452323  9
6       1     2  1.9461896  1.1121450  8
```

Table 8.3: The simulated dataset with a continuous outcome

	1	2	3	Overall
	(N=349)	(N=355)	(N=296)	(N=1000)
z1				
Mean (SD)	-0.0323 (1.00)	0.0113 (0.996)	-0.0787 (0.958)	-0.0306 (0.987)
Median [Min, Max]	-0.0262 [-3.17, 2.69]	-0.0127 [-2.67, 3.78]	-0.0860 [-2.41, 2.47]	-0.0318 [-3.17, 3.78]
z2				
Mean (SD)	-0.0524 (0.981)	-0.0161 (1.02)	-0.0959 (1.05)	-0.0524 (1.02)
Median [Min, Max]	-0.0323 [-2.69, 2.55]	-0.0175 [-3.32, 2.95]	-0.139 [-3.14, 2.34]	-0.0511 [-3.32, 2.95]
studyid				
1	46 (13.2%)	54 (15.2%)	0 (0%)	100 (10.0%)
2	46 (13.2%)	54 (15.2%)	0 (0%)	100 (10.0%)
3	50 (14.3%)	50 (14.1%)	0 (0%)	100 (10.0%)
4	54 (15.5%)	0 (0%)	46 (15.5%)	100 (10.0%)
5	47 (13.5%)	0 (0%)	53 (17.9%)	100 (10.0%)
6	0 (0%)	42 (11.8%)	58 (19.6%)	100 (10.0%)
7	0 (0%)	53 (14.9%)	47 (15.9%)	100 (10.0%)
8	40 (11.5%)	38 (10.7%)	22 (7.4%)	100 (10.0%)
9	37 (10.6%)	27 (7.6%)	36 (12.2%)	100 (10.0%)
10	29 (8.3%)	37 (10.4%)	34 (11.5%)	100 (10.0%)

Let us look into the data a bit in more detail:

8.2.1.2 Model fitting

We will use the model shown in Equation 16.8 in the book. In addition, we will use Bayesian LASSO to penalize the treatment-covariate interactions.

```
ipd3 <- with(ds3, ipdnma.model.onestage(y = y, study = studyid, treat = treat,
                                        X = cbind(z1, z2),
                                        response = "normal",
                                        shrinkage = "laplace",
                                        type = "random"))
ipd3$model.JAGS
```

```
function ()
{
    for (i in 1:Np) {
```

```
    y[i] ~ dnorm(mu[i], sigma)
    mu[i] <- alpha[studyid[i]] + inprod(beta[], X[i, ]) +
        inprod(gamma[treat[i], ], X[i, ]) + d[studyid[i],
        treatment.arm[i]]
}
sigma ~ dgamma(0.001, 0.001)
for (i in 1:Nstudies) {
    w[i, 1] <- 0
    d[i, 1] <- 0
    for (k in 2:na[i]) {
        d[i, k] ~ dnorm(mdelta[i, k], taudelta[i, k])
        mdelta[i, k] <- delta[t[i, k]] - delta[t[i, 1]] +
            sw[i, k]
        taudelta[i, k] <- tau * 2 * (k - 1)/k
        w[i, k] <- d[i, k] - delta[t[i, k]] + delta[t[i,
            1]]
        sw[i, k] <- sum(w[i, 1:(k - 1)])/(k - 1)
    }
}
sd ~ dnorm(0, 1)
T(0, )
tau <- pow(sd, -2)
delta[1] <- 0
for (k in 2:Ntreat) {
    delta[k] ~ dnorm(0, 0.001)
}
for (j in 1:Nstudies) {
    alpha[j] ~ dnorm(0, 0.001)
}
for (k in 1:Ncovariate) {
    beta[k] ~ dnorm(0, 0.001)
}
lambda[1] <- 0
lambda.inv[1] <- 0
for (m in 2:Ntreat) {
    tt[m] <- lambda[m] * sigma
    lambda[m] <- pow(lambda.inv[m], -1)
    lambda.inv[m] ~ dunif(0, 5)
}
for (k in 1:Ncovariate) {
    gamma[1, k] <- 0
    for (m in 2:Ntreat) {
        gamma[m, k] ~ ddexp(0, tt[m])
```

```
        }
    }
}
<environment: 0x000002aa855142f8>
```

```
samples <- ipd.run(ipd3, n.chains = 2, n.iter = 20,
                   pars.save = c("alpha", "beta", "delta", "sd", "gamma"))
```

```
Compiling model graph
   Resolving undeclared variables
   Allocating nodes
Graph information:
   Observed stochastic nodes: 1000
   Unobserved stochastic nodes: 35
   Total graph size: 10141

Initializing model
```

```
summary(samples)
```

```
Iterations = 2001:2020
Thinning interval = 1
Number of chains = 2
Sample size per chain = 20

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

              Mean      SD Naive SE Time-series SE
alpha[1]   11.0143 0.05351 0.008461       0.012893
alpha[2]    8.0104 0.04313 0.006819       0.011318
alpha[3]   10.5057 0.04685 0.007408       0.007344
alpha[4]    9.5642 0.04372 0.006912       0.006977
alpha[5]   12.9822 0.03072 0.004858       0.004887
alpha[6]   13.2670 0.04836 0.007646       0.012434
alpha[7]    7.3961 0.05204 0.008228       0.014720
alpha[8]   11.1075 0.05268 0.008330       0.013409
alpha[9]   10.1329 0.05067 0.008012       0.014126
alpha[10]   9.1592 0.05562 0.008795       0.017214
```

```
beta[1]     0.2287 0.01429 0.002259       0.004478
beta[2]     0.3171 0.01742 0.002755       0.005921
delta[1]    0.0000 0.00000 0.000000       0.000000
delta[2]   -2.9582 0.06936 0.010967       0.009455
delta[3]   -1.1239 0.07083 0.011200       0.009991
gamma[1,1]  0.0000 0.00000 0.000000       0.000000
gamma[2,1] -0.6453 0.02223 0.003516       0.006829
gamma[3,1] -0.3277 0.01879 0.002971       0.003337
gamma[1,2]  0.0000 0.00000 0.000000       0.000000
gamma[2,2]  0.6075 0.02498 0.003950       0.007017
gamma[3,2]  0.4135 0.01868 0.002954       0.006306
sd          0.2230 0.05817 0.009198       0.016080

2. Quantiles for each variable:

              2.5%     25%     50%     75%   97.5%
alpha[1]   10.9302 10.9823 11.0220 11.0479 11.0988
alpha[2]    7.9224  7.9816  8.0152  8.0369  8.0830
alpha[3]   10.4128 10.4722 10.5106 10.5439 10.5787
alpha[4]    9.4920  9.5290  9.5637  9.5982  9.6411
alpha[5]   12.9236 12.9621 12.9842 13.0013 13.0324
alpha[6]   13.1556 13.2396 13.2700 13.2984 13.3439
alpha[7]    7.3262  7.3548  7.3923  7.4279  7.4879
alpha[8]   11.0004 11.0811 11.1067 11.1433 11.1924
alpha[9]   10.0257 10.1025 10.1393 10.1667 10.2160
alpha[10]   9.0730  9.1219  9.1572  9.1931  9.2463
beta[1]     0.2016  0.2203  0.2293  0.2358  0.2587
beta[2]     0.2853  0.3059  0.3196  0.3301  0.3481
delta[1]    0.0000  0.0000  0.0000  0.0000  0.0000
delta[2]   -3.1027 -3.0003 -2.9579 -2.9084 -2.8760
delta[3]   -1.2442 -1.1743 -1.1340 -1.0820 -0.9850
gamma[1,1]  0.0000  0.0000  0.0000  0.0000  0.0000
gamma[2,1] -0.6915 -0.6579 -0.6452 -0.6310 -0.6087
gamma[3,1] -0.3700 -0.3387 -0.3263 -0.3166 -0.2969
gamma[1,2]  0.0000  0.0000  0.0000  0.0000  0.0000
gamma[2,2]  0.5617  0.5949  0.6061  0.6285  0.6471
gamma[3,2]  0.3834  0.4000  0.4137  0.4247  0.4442
sd          0.1428  0.1828  0.2184  0.2439  0.3889
```

As before, we can use the `treatment.effect()` function of *bipd* to estimate relative effects for new patients.

```
treatment.effect(ipd3, samples, newpatient= c(1,2))
```

```
$`treatment 2`
    0.025       0.5     0.975
-2.525064 -2.418394 -2.266426

$`treatment 3`
     0.025        0.5      0.975
-0.7388297 -0.6376228 -0.5007683
```

This gives us the relative effects for all treatments versus the reference. To obtain relative effects between active treatments we need some more coding:

```
samples.all=data.frame(rbind(samples[[1]], samples[[2]]))
newpatient= c(1,2)
newpatient <- (newpatient - ipd3$scale_mean)/ipd3$scale_sd

median(
  samples.all$delta.2.+samples.all$gamma.2.1.*
    newpatient[1]+samples.all$gamma.2.2.*newpatient[2]
-
  (samples.all$delta.3.+samples.all$gamma.3.1.*newpatient[1]+
     samples.all$gamma.3.2.*newpatient[2])
)
```

```
[1] -1.765342
```

```
quantile(samples.all$delta.2.+samples.all$gamma.2.1.*
           newpatient[1]+samples.all$gamma.2.2.*newpatient[2]
         -(samples.all$delta.3.+samples.all$gamma.3.1.*newpatient[1]+
             samples.all$gamma.3.2.*newpatient[2])
         , probs = 0.025)
```

```
    2.5%
-1.940397
```

```
quantile(samples.all$delta.2.+samples.all$gamma.2.1.*
           newpatient[1]+samples.all$gamma.2.2.*newpatient[2]
```

```
          -(samples.all$delta.3.+samples.all$gamma.3.1.*newpatient[1]+
              samples.all$gamma.3.2.*newpatient[2])
          , probs = 0.975)
```

```
   97.5%
-1.62695
```

8.2.2 Modeling patient-level relative effects using randomized and observational evidence for a network of treatments

We will now follow Chapter 16.3.5 from the book. In this analysis we will not use penalization, and we will assume fixed effects. For an example with penalization and random effects, see part 2 of this vignettte.

8.2.2.1 Setup

We generate a very simple dataset of three studies comparing three treatments. We will assume 2 RCTs and 1 non-randomized trial:

```
ds4 <- generate_ipdnma_example(type = "continuous")
ds4 <- ds4 %>% filter(studyid %in% c(1,4,10)) %>%
  mutate(studyid = factor(studyid) %>%
           recode_factor(
             "1" = "1",
             "4" = "2",
             "10" = "3"),
         design = ifelse(studyid == "3", "nrs", "rct"))
```

The sample size is as follows:

```
          s1 s2 s3
treat A: 41 51 39
treat B: 59  0 25
treat C:  0 49 36
```

8.2.2.2 Model fitting

We will use the design-adjusted model, equation 16.9 in the book. We will fit a two-stage fixed effects meta-analysis and we will use a variance inflation factor. The code below is used to specify the analysis of each individual study. Briefly, in each study we adjust the treatment effect for the prognostic factors `z1` and `z2`, as well as their interaction with `treat`.

```
library(rjags)
```

```
Loading required package: coda
```

```
Linked to JAGS 4.3.1
```

```
Loaded modules: basemod,bugs
```

```
first.stage <- "
model{

for (i in 1:N){
    y[i] ~ dnorm(mu[i], tau)
    mu[i] <- a + inprod(b[], X[i,]) + inprod(c[,treat[i]], X[i,]) + d[treat[i]]
}
sigma ~ dunif(0, 5)
tau <- pow(sigma, -2)

a ~ dnorm(0, 0.001)

for(k in 1:Ncovariate){
    b[k] ~ dnorm(0,0.001)
}

for(k in 1:Ncovariate){
    c[k,1] <- 0
}

tauGamma <- pow(sdGamma,-1)
sdGamma ~ dunif(0, 5)

for(k in 1:Ncovariate){
    for(t in 2:Ntreat){
```

```
        c[k,t] ~ ddexp(0, tauGamma)
    }
}

d[1] <- 0
for(t in 2:Ntreat){
    d[t] ~ dnorm(0, 0.001)
}
}"
```

Subsequently, we estimate the relative treatment effects in the first (randomized) study comparing treatments A and B:

```
model1.spec <- textConnection(first.stage)
data1 <- with(ds4 %>% filter(studyid == 1),
              list(y = y,
                   N = length(y),
                   X = cbind(z1,z2),
                   treat = treat,
                   Ncovariate = 2,
                   Ntreat = 2))
jags.m <- jags.model(model1.spec, data = data1, n.chains = 2, n.adapt = 500,
                     quiet =  TRUE)
params <- c("d", "c")
samps4.1 <- coda.samples(jags.m, params, n.iter = 50)
samps.all.s1 <- data.frame(as.matrix(samps4.1))

samps.all.s1 <- samps.all.s1[, c("c.1.2.", "c.2.2.", "d.2.")]
delta.1 <- colMeans(samps.all.s1)
cov.1 <- var(samps.all.s1)
```

We repeat the analysis for the second (randomized) study comparing treatments A and C:

```
model1.spec <- textConnection(first.stage)
data2 <- with(ds4 %>% filter(studyid == 2),
              list(y = y,
                   N = length(y),
                   X = cbind(z1,z2),
                   treat = ifelse(treat == 3, 2, treat),
                   Ncovariate = 2,
                   Ntreat = 2))
```

```
jags.m <- jags.model(model1.spec, data = data2, n.chains = 2, n.adapt = 100,
                     quiet =  TRUE)
params <- c("d", "c")
samps4.2 <- coda.samples(jags.m, params, n.iter = 50)
samps.all.s2 <- data.frame(as.matrix(samps4.2))
samps.all.s2 <- samps.all.s2[, c("c.1.2.", "c.2.2.", "d.2.")]
delta.2 <- colMeans(samps.all.s2)
cov.2 <- var(samps.all.s2)
```

Finally, we analyze the third (non-randomized) study comparing treatments A, B, and C:

```
model1.spec <- textConnection(first.stage)
data3 <- with(ds4 %>% filter(studyid == 3),
              list(y = y,
                   N = length(y),
                   X = cbind(z1,z2),
                   treat = treat,
                   Ncovariate = 2,
                   Ntreat = 3))
jags.m <- jags.model(model1.spec, data = data3, n.chains = 2, n.adapt = 100,
                     quiet = TRUE)
params <- c("d", "c")
samps4.3 <- coda.samples(jags.m, params, n.iter = 50)
samps.all.s3 <- data.frame(as.matrix(samps4.3))

samps.all.s3 <- samps.all.s3[, c("c.1.2.", "c.2.2.", "d.2.", "c.1.3.",
                                 "c.2.3.", "d.3.")]
delta.3 <- colMeans(samps.all.s3)
cov.3 <- var(samps.all.s3)
```

The corresponding treatment effect estimates are depicted below:

Table 8.4: Treatment effect estimates.

study	B versus A	C versus A
study 1	-3.012 (SE = 0.062)	
study 2		-1.228 (SE = 0.051)
study 3	-3.013 (SE = 0.082)	-1.052 (SE = 0.073)

We can now fit the second stage of the network meta-analysis. The corresponding JAGS model is specified below:

```
second.stage <-
"model{

  #likelihood
  y1 ~ dmnorm(Mu1, Omega1)
  y2 ~ dmnorm(Mu2, Omega2)
  y3 ~ dmnorm(Mu3, Omega3*W)


  Omega1 <- inverse(cov.1)
  Omega2 <- inverse(cov.2)
  Omega3 <- inverse(cov.3)

  Mu1 <- c(gamma[,1], delta[2])
  Mu2 <- c(gamma[,2], delta[3])
  Mu3 <- c(gamma[,1], delta[2],gamma[,2], delta[3])

  #parameters
  for(i in 1:2){
    gamma[i,1] ~ dnorm(0, 0.001)
    gamma[i,2] ~ dnorm(0, 0.001)
  }

  delta[1] <- 0
  delta[2] ~ dnorm(0, 0.001)
  delta[3] ~ dnorm(0, 0.001)

}
"
```

We can fit as follows:

```
model1.spec <- textConnection(second.stage)
data3 <- list(y1 = delta.1, y2 = delta.2, y3 = delta.3,
              cov.1 = cov.1, cov.2 = cov.2, cov.3 = cov.3, W = 0.5)

jags.m <- jags.model(model1.spec, data = data3, n.chains = 2, n.adapt = 50,
                     quiet = TRUE)
params <- c("delta", "gamma")
samps4.3 <- coda.samples(jags.m, params, n.iter = 50)
```

```r
summary(samps4.3)
```

```
Iterations = 1:50
Thinning interval = 1
Number of chains = 2
Sample size per chain = 50

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

              Mean      SD Naive SE Time-series SE
delta[1]    0.0000 0.00000 0.000000       0.000000
delta[2]   -3.0479 0.06413 0.006413       0.007445
delta[3]   -1.1894 0.04620 0.004620       0.004642
gamma[1,1] -0.8124 0.05418 0.005418       0.005446
gamma[2,1]  0.8390 0.07016 0.007016       0.007048
gamma[1,2] -0.4248 0.06481 0.006481       0.006486
gamma[2,2]  0.3830 0.05754 0.005754       0.005753

2. Quantiles for each variable:

              2.5%     25%     50%     75%   97.5%
delta[1]    0.0000  0.0000  0.0000  0.0000  0.0000
delta[2]   -3.1476 -3.0843 -3.0490 -3.0206 -2.9350
delta[3]   -1.2775 -1.2212 -1.1912 -1.1595 -1.1086
gamma[1,1] -0.9119 -0.8504 -0.8097 -0.7759 -0.6993
gamma[2,1]  0.6673  0.8060  0.8435  0.8867  0.9439
gamma[1,2] -0.5186 -0.4632 -0.4402 -0.3948 -0.2968
gamma[2,2]  0.3057  0.3550  0.3932  0.4151  0.4543
```

```r
# calculate  treatment effects
samples.all=data.frame(rbind(samps4.3[[1]], samps4.3[[2]]))
newpatient= c(1,2)

median(
  samples.all$delta.2.+samples.all$gamma.1.1.*newpatient[1]+
    samples.all$gamma.2.1.*newpatient[2]
)
```

```
[1] -2.159397
```

```
  quantile(samples.all$delta.2.+samples.all$gamma.1.1.*newpatient[1]+
             samples.all$gamma.2.1.*newpatient[2]
           , probs = 0.025)
```

```
     2.5%
-2.498814
```

```
  quantile(samples.all$delta.2.+samples.all$gamma.1.1.*newpatient[1]+
             samples.all$gamma.2.1.*newpatient[2]
           , probs = 0.975)
```

```
    97.5%
-1.966954
```

Version info

This chapter was rendered using the following version of R and its packages:

```
R version 4.2.3 (2023-03-15 ucrt)
Platform: x86_64-w64-mingw32/x64 (64-bit)
Running under: Windows 10 x64 (build 19045)

Matrix products: default

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

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

other attached packages:
[1] rjags_4-14       coda_0.19-4      ggplot2_3.4.2    bipd_0.3
[5] kableExtra_1.3.4 dplyr_1.1.2      table1_1.4.3

loaded via a namespace (and not attached):
 [1] pillar_1.9.0      compiler_4.2.3    tools_4.2.3       digest_0.6.31
```

```
 [5] gtable_0.3.3      lattice_0.21-8    jsonlite_1.8.5    evaluate_0.21
 [9] lifecycle_1.0.3   tibble_3.2.1      viridisLite_0.4.2 pkgconfig_2.0.3
[13] rlang_1.1.1       cli_3.6.1         rstudioapi_0.14   yaml_2.3.7
[17] mvtnorm_1.2-2     xfun_0.39         fastmap_1.1.1     withr_2.5.0
[21] httr_1.4.6        stringr_1.5.0     knitr_1.43        xml2_1.3.4
[25] generics_0.1.3    vctrs_0.6.3       systemfonts_1.0.4 grid_4.2.3
[29] webshot_0.5.4     tidyselect_1.2.0  svglite_2.1.1     glue_1.6.2
[33] R6_2.5.1          fansi_1.0.4       rmarkdown_2.22    Formula_1.2-5
[37] farver_2.1.1      magrittr_2.0.3    scales_1.2.1      codetools_0.2-19
[41] htmltools_0.5.5   rvest_1.0.3       colorspace_2.1-0  labeling_0.4.2
[45] utf8_1.2.3        stringi_1.7.12    munsell_0.5.0
```

References

9 Visualization and interpretation of individualized treatment rule results

Xiaotong Jiang (Biogen)

In this tutorial, we will walk you through the code that implemented the precision medicine methods and generated the visualization results discussed in Chapter 18 of the book. This tutorial focuses more on helping you understand the code. We will not provide detailed interpretation of the results as they have been covered in the chapter already.

9.1 Introduction

We first load all relevant functions for this chapter.

```
source("resources/chapter 18/functions.r")
```

Subsequently, we use the function `simcountdata()` to generate an example dataset with a sample size of N=2000. In this example, we have two disease modifying therapies (DMT1 and DMT0) and the outcome is the number of post-treatment multiple sclerosis relapses during follow-up.

```
# Randomization seed
base.seed <- 999

set.seed(base.seed)
df.ori <- simcountdata(n = 2000,
                       seed = 63,
                       beta = c(log(0.4), log(0.5), log(1), log(1.1), log(1.2)),
                       beta.x = c(-1.54, -0.01, 0.06, 0.25, 0.5, 0.13, 0.0000003)
)$data
```

The dataset looks as follows:

```r
head(df.ori)
```

```
  trt ageatindex_centered female prerelapse_num prevDMTefficacy premedicalcost
1   0                   2      0              2    Low efficacy        4606.04
2   1                  10      1              1    Low efficacy       17065.19
3   1                  12      1              2            None        6308.39
4   1                 -12      0              0    Low efficacy       16633.97
5   1                  13      1              0    Low efficacy         642.96
6   1                  14      1              0    Low efficacy        2989.89
  numSymptoms postrelapse_num finalpostdayscount     group     score Iscore
1           0               1                305 Simulated 0.7129792      1
2           1               0                367 Simulated 0.7404238      2
3           0               0                325 Simulated 0.7564233      3
4           0               0                321 Simulated 0.7215764      1
5           0               0                 24 Simulated 0.7457823      2
6           0               0                 59 Simulated 0.7441632      2
```

Below is a summary table of the baseline characteristics by treatment group.

We now define key constants for the case study.

```r
# Baseline characteristics
covars <- c("age.z", "female", "prevtrtB", "prevtrtC", "prevnumsymp1",
            "prevnumsymp2p", "previous_cost.z", "previous_number_relapses")

# Precision medicine methods to be used
pm.methods <- c("all1", "all0", "poisson", "dWOLS", "listDTR2",
                "contrastReg")

# Precision medicine method labels
method.vec <- c("All 0", "All 1", "Poisson", "dWOLS",
                "Contrast\n Regression", "List DTR\n (2 branches)")

# Number of folds in each CV iteration
n.fold <- 5

# Number of CV iterations
n.cv <- 10

# Sample size of the large independent test set to get true value
big.n <- 100000
```

Table 9.1: Baseline characteristics of the case study data

	0	1	Overall
	(N=506)	(N=1494)	(N=2000)
Age (years)			
Mean (SD)	45.2 (9.82)	45.8 (9.73)	45.7 (9.75)
Median [Min, Max]	46.0 [20.0, 64.0]	46.0 [19.0, 64.0]	46.0 [19.0, 64.0]
Gender			
female	375 (74.1%)	1123 (75.2%)	1498 (74.9%)
male	131 (25.9%)	371 (24.8%)	502 (25.1%)
Previous number of relapses			
0	319 (63.0%)	973 (65.1%)	1292 (64.6%)
1	150 (29.6%)	427 (28.6%)	577 (28.9%)
2	31 (6.1%)	76 (5.1%)	107 (5.4%)
3	5 (1.0%)	17 (1.1%)	22 (1.1%)
4	1 (0.2%)	1 (0.1%)	2 (0.1%)
Efficacy of previous disease modifying therapy			
Low efficacy	216 (42.7%)	609 (40.8%)	825 (41.3%)
Medium and high efficacy	53 (10.5%)	179 (12.0%)	232 (11.6%)
None	237 (46.8%)	706 (47.3%)	943 (47.2%)
Previous medical cost (\$)			
Mean (SD)	13700 (20400)	14400 (24500)	14300 (23600)
Median [Min, Max]	7320 [343, 264000]	7560 [110, 556000]	7470 [110, 556000]
Previous number of symptoms			
0	348 (68.8%)	995 (66.6%)	1343 (67.2%)
1	119 (23.5%)	388 (26.0%)	507 (25.4%)
>=2	39 (7.7%)	111 (7.4%)	150 (7.5%)

```r
# Define formula for the CATE model
cate.formula <- as.formula(paste0("y ~", paste0(covars, collapse = "+"),
                                  "+ offset(log(years))"))

# Define formula for the propensity score model
ps.formula <- trt ~ age.z + prevtrtB + prevtrtC

# Color
myblue <- rgb(37, 15, 186, maxColorValue = 255)
mygreen <- rgb(109, 173, 70, maxColorValue = 255)
mygrey <- rgb(124, 135, 142, maxColorValue = 255)
```

The data need to be preprocessed to be more analyzable. We recategorized treatment, previous treatment, and number of symptoms; scaled medical cost and age; and standardized the data.

```r
df <- df.ori %>%
  rename(previous_treatment = prevDMTefficacy,
         age = ageatindex_centered,
         y = postrelapse_num,
         previous_number_relapses = prerelapse_num,
         previous_number_symptoms = numSymptoms,
         previous_cost = premedicalcost) %>%
  mutate(previous_treatment = factor(previous_treatment,
                                     levels = c("None", "Low efficacy", "Medium and high e
                                     labels = c("drugA", "drugB", "drugC")),
         previous_number_symptoms = factor(previous_number_symptoms,
                                           levels = c("0", "1", ">=2"),
                                           labels = c("0", "1", ">=2")),
         trt = factor(trt, levels = c(0, 1), labels = c("drug0", "drug1")),
         previous_cost.z = scale(log(previous_cost), scale = TRUE), # log-transformed due
         age.z = age + 48,
         age.z = scale(age.z, scale = TRUE),
         years = finalpostdayscount / 365.25,
         mlogarr0001 = -log(y / years + 0.001),
         drug1 = as.numeric(trt == "drug1"),
         prevtrtB = as.numeric(previous_treatment == "drugB"),
         prevtrtC = as.numeric(previous_treatment == "drugC"),
         prevnumsymp1 = as.numeric(previous_number_symptoms == "1"),
         prevnumsymp2p = as.numeric(previous_number_symptoms == ">=2")) %>%
  dplyr::select(age.z, female, contains("prevtrt"), previous_cost.z, contains("prevnumsymp
                previous_number_relapses, trt, drug1, y, mlogarr0001, years, Iscore)
```

```
# Standardize data
df.s <- df
df.s[, setdiff(covars, c("age.z", "previous_cost.z"))] <- df[, setdiff(covars, c("age.z",
```

9.2 Estmition of individualized treatment rules

The following code provides details of how to implement the precision medicine methods in the example data. Please feel free to jump to the next section if you want to focus on the results. The model results are available online for you to load and save time.

We used the function `listdtr()` in the **listdtr** package to estimate individualized treatment rules (ITRs) based on the listDTR method. We used the function `catefit()` in the **precmed** package to estimate ITRs based on the Poisson and contrast regression method. These were the methods used in Section 3 of the book where we talked about directly visualizing the ITR before bringing in the outcomes. The methods are discussed in further detail by Zhao et al. (2013) and Yadlowsky et al. (2020).

```
library(listdtr)

# Estimated ITR based on the listDTR method with 2 branches
modlist2 <- listdtr(y = df$mlogarr, # larger is more favorable
                    a = df$drug1,
                    x = df[, c("age.z", "female", "prevtrtB", "prevtrtC", "previous_cost.z"
                               "prevnumsymp1", "prevnumsymp2p", "previous_number_relapses")
                    stage.x = rep(1, 8), maxlen = 2L) # somewhat slow

# Estimated ITR based on the listDTR method with 3 branches
modlist3 <- listdtr(y = df$mlogarr,
                     a = df$drug1,
                     x = df[, c("age.z", "female", "prevtrtB", "prevtrtC", "previous_cost.z
                                "prevnumsymp1", "prevnumsymp2p", "previous_number_relapses"
                     stage.x = rep(1, 8), maxlen = 3L) # somewhat slow

# Estimated CATE score based on the Poisson and contrast regression
modpm <- catefit(response = "count",
             cate.model = cate.formula,
             ps.model = ps.formula,
             data = df,
             higher.y = FALSE,
             score.method = c("poisson", "contrastReg"),
```

```
            initial.predictor.method = "poisson",
            seed = 999)

# Estimated CATE score based on the Poisson and contrast regression
# (based on the scaled data so the coefficients are easier to compare)
modpm.s <- catefit(response = "count",
              cate.model = cate.formula,
              ps.model = ps.formula,
              data = df.s,
              higher.y = FALSE,
              score.method = c("poisson", "contrastReg"),
              initial.predictor.method = "poisson",
              seed = 999)
```

For results in Sections 4 and 5, we applied cross validation to mitigate over-fitting. For this chapter, we created our own customized function `cvvalue()` to estimate the ITR and calculate the estimated value function via cross validation for all methods, including the fixed method. The results were all saved under the prefix `cvmod`. The **precmed** package has a built-in cross validation procedure for CATE estimation so we used the function `catefit()`.

```
# Run cross validation for each method (used for Sections 4 & 5)

## Estimated CATE scores based on the Poisson and contrast regression with cross-validatio
modcv <- catecv(response = "count",
                cate.model = cate.formula,
                ps.model = ps.formula,
                data = df,
                higher.y = FALSE,
                score.method = c("poisson", "contrastReg"),
                initial.predictor.method = "poisson",
                cv.n = n.cv,
                plot.gbmperf = FALSE,
                seed = 999) # somewhat slow

## Estimated value function for each method
cvmodall0 <- cvvalue(data = df, xvar = covars,
                     method = "all0", n.fold = n.fold, n.cv = n.cv,
                     seed = base.seed)

cvmodall1 <- cvvalue(data = df, xvar = covars,
                     method = "all1", n.fold = n.fold, n.cv = n.cv,
                     seed = base.seed)
```

```
cvmoddwols <- cvvalue(data = df, xvar = covars,
                      method = "dWOLS", n.fold = n.fold, n.cv = n.cv,
                      seed = base.seed)

cvmodpois <- cvvalue(data = df, xvar = covars,
                     method = "poisson", n.fold = n.fold, n.cv = n.cv,
                     seed = base.seed)

cvmodlist2 <- cvvalue(data = df, xvar = covars,
                      method = "listDTR2", n.fold = n.fold, n.cv = n.cv,
                      seed = base.seed) # very slow

cvmodcontrastreg <- cvvalue(data = df, xvar = covars,
                            method = "contrastReg", n.fold = n.fold,
                            n.cv = n.cv,
                            seed = base.seed) # very slow
```

As a next step, we need to combine all estimated ITRs and value functions:

```
# Combine CV results
# Read in each CV result in a loop
vhats.dhat <- dhats <- NULL
mod_names <- c("cvmodall1", "cvmodall0", "cvmoddwols", "cvmodpois", "cvmodcontrastreg", "c
for (mod in mod_names){
  thismod <- get(mod)
  for (name in names(thismod)) {
    # Get estimated values, vhat.dhat
    vhats.dhat <- rbind(vhats.dhat,
                        thismod[[name]] %>%
                          map_df(~bind_rows(names(.x) %>% str_detect("vhat.dhat") %>% keep
                          mutate(method = mod, cv.i = name))
    # Get estimated rule from CV test fold, dhat
    dhats  <- rbind(dhats,
                    thismod[[name]] %>%
                      map_df(~bind_rows(names(.x) %>% str_detect("^dhat$") %>% keep(.x, .)
                      mutate(method = mod, cv.i = name))

  }
}

# One time run to get true optimal and worst value
```

```r
# Simulated data only
trueV <- getTrueOptimalValue(n = big.n, seed = base.seed)
trueWorstV <- getTrueWorstValue(n = big.n, seed = base.seed)

# Preprocess
vhats.dhat %<>%
  mutate(V = U/W,
         VR = (U/W - trueWorstV) / (trueV - trueWorstV)) %>%
  group_by(method) %>%
  summarize(n.batches = n(),
            n.nonnaU = sum(!is.na(U)),
            n.nonnaW = sum(!is.na(W)),
            meanV = mean(V, na.rm = T),
            sdV = sd(V, na.rm = T),
            meanVR = mean(VR, na.rm = T),
            sdVR = sd(VR, na.rm = T),
            .groups = "keep") %>%
  ungroup %>%
  arrange(desc(meanV)) %>%
  mutate(method = case_when(
    method == "cvmodcontrastreg" ~ "Contrast\n Regression",
    method == "cvmodall0" ~ "All 0",
    method == "cvmodall1" ~ "All 1",
    method == "cvmodlist2" ~ "List DTR\n (2 branches)",
    method == "cvmoddwols" ~ "dWOLS",
    method == "cvmodpois" ~ "Poisson"),
    method = factor(method,
                    levels = method.vec,
                    labels = method.vec)
  )

dhats %<>%
  mutate(method = case_when(
    method == "cvmodcontrastreg" ~ "Contrast\n Regression",
    method == "cvmodall0" ~ "All 0",
    method == "cvmodall1" ~ "All 1",
    method == "cvmodlist2" ~ "List DTR\n (2 branches)",
    method == "cvmoddwols" ~ "dWOLS",
    method == "cvmodpois" ~ "Poisson"),
    method = factor(method,
                    levels = method.vec,
```

```
                    labels = method.vec)
  )
```

9.3 Visualization of individualized treatment rules

9.3.1 Direct visualization

9.3.1.1 listDTR

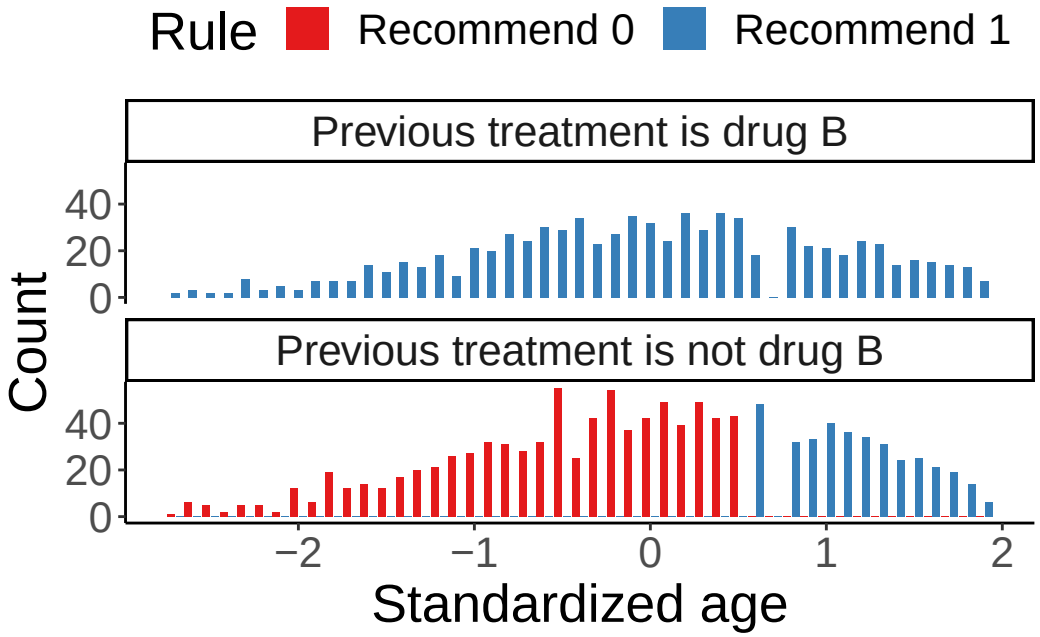
If the PM method already has built-in visualization (especially for tree-based methods), we can visualize the ITR directly. For example, we can simply use the function `plot()` to visualize the estimated ITR with the listDTR method.

```
#modlist3 %>% plot()
```

We can also create our own visualization like Figure 1A in the chapter.

```
df.list3 <- df %>%
  mutate(d.list = ifelse(age.z > 0.599 | prevtrtB > 0.5, "Recommend 1", "Recommend 0"), #
         Rule = factor(as.character(d.list), levels = c("Recommend 0", "Recommend 1")),
         prevtrtB = ifelse(prevtrtB == 1, "Previous treatment is drug B", "Previous treatm
  )

## Figure 1A
df.list3 %>%
  ggplot(aes(x = age.z, fill = Rule))+
  geom_histogram(position = position_dodge2(preserve = 'single'), binwidth = 0.1)+
  facet_wrap(~ prevtrtB, nrow = 2) +
  scale_fill_brewer(palette = "Set1") +
  labs(x = "Standardized age", y = "Count") +
  theme_classic() +
  theme(legend.position = 'top', text = element_text(size = 20))
```

Observed treatment	listDTR ITR	ARR	n	prop%
drug0	Recommend 0	0.32	197	10
drug0	Recommend 1	0.31	309	15
drug1	Recommend 0	0.39	615	31
drug1	Recommend 1	0.16	879	44



The subgroup-level annualized relapse rate (ARR) can be calculated based on the listDTR ITR:

```
df.list3 %>%
  group_by(trt, d.list) %>%
  summarise(ARR = round(sum(y) / sum(years), 2),
            n = n(),
            `prop%` = round(n / nrow(df), 2)*100, .groups = "drop") %>%
  rename("listDTR ITR" = d.list,
         "Observed treatment" = trt) %>%
  kable() %>%
  kable_styling(full_width = F)
```

Patients who received drug 0 and were recommended drug 0 by listDTR had a similar ARR on average than those who received drug 0 but were recommended drug 1 (0.32 vs 0.31). Patients who received drug 1 and were recommended drug 1 by listDTR had a much lower ARR on

average than those who received drug 1 but were recommended drug 0 (0.16 vs 0.39).

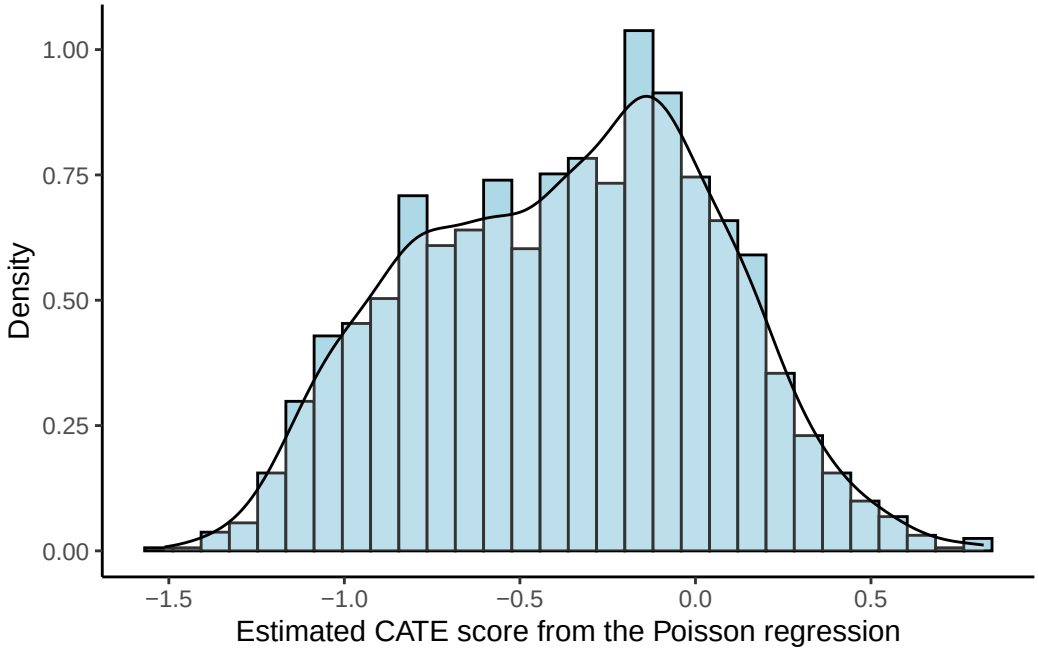
9.3.1.2 Score-based method

Although some PM methods do not have built-in visualization or not as "white-box" as some more interpretable methods, there still might be ways to visualize the ITR. For example, score-based methods (such as Poisson and contrast regression) produce an estimate of the CATE score for each patient, and a classification tree can be fitted on these scores and visualized. Below is a histogram-density plot of the CATE scores estimated from the Poisson regression and the fitted classification tree using the estimated CATE scores. We pruned the tree so it only had three nodes for simplicity. The `rpart.plot` package has a built-in visualization function of the `rpart` model, `rpart.plot()`, which is how Figure 1B in the chapter was generated.

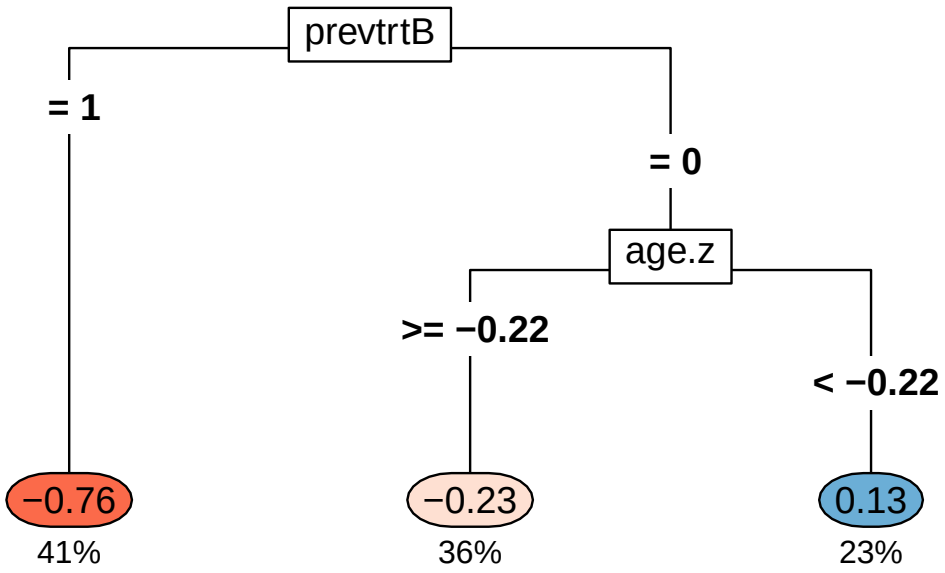
```
df["score.poisson"] <- modpm$score.poisson

ggplot(df, aes(x = score.poisson)) +
  geom_histogram(aes(y = ..density..), colour = "black", fill = "lightblue") +
  geom_density(alpha = .2, fill = "white") +
  labs(x = "Estimated CATE score from the Poisson regression", y = "Density") +
  theme_classic()
```

```
Warning: The dot-dot notation (`..density..`) was deprecated in ggplot2 3.4.0.
i Please use `after_stat(density)` instead.
```

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




```
modtree <- rpart(as.formula(paste0("score.poisson ~", paste0(covars, collapse = "+"))),
                 method = "anova", data = df, control = rpart.control(minsplit = 100, cp =

modtree.pr <- prune(modtree, cp = 0.09) # I ended up choosing a higher cp value to have on

# print(rpart.rules(modtree.pr, cover = TRUE))

## Figure 1B
rpart.plot(modtree.pr, box.palette = "RdBu", type = 5, under = TRUE, tweak = 1.2, compress
```



The CATE scores are now simplified as a tree classifier. Previous treatment of drug B and age seemed to be important in determining the CATE score values, which also showed up in the estimated from the listDTR method. Patients with previous treatment of drug B had the lowest CATE score on average (-0.76) and took up 41% of the samples (dark orange). Patients whose previous treatment was not drug B and age was >= -0.22 standard deviation of the mean also had a negative CATE score on average (-0.23) and took up 36% of the samples (light orange), but not as low as the dark orange group. Negative CATE scores mean that the number of relapses was expected to be lower for those recommended drug 1 than those recommended drug 0, so drug 1 was favored for them. For the blue group, the average CATE score was 0.13, taking up 23% of the samples, and they were expected to benefit from drug 0 based on the Poisson CATE scores.

9.3.2 ITR accuracy

The accuracy of ITR is the proportion of patients whose estimated ITR is the same as the true optimal ITR. The estimated ITRs have been obtained from the PM methods but we need to calculate the true optimal ITR. This is only possible for simulated data where the decision boundary is known. Based on the data generating mechanism in `simcountdata()`, `Iscore` is a score generated from a linear combination of baseline covariates where lower scores represented that drug 1 was better and higher scores represented that drug 0 was better. We then classified patients in 5 equal-size subgroups based on the `Iscore`, where groups 1 and 2 have drug 1 as their true optimal ITR and groups 3 and 4 have drug 0 as their true optimal

ITR. Group 3 is considered the neutral group, where patients are indifferent to either drug so we assign the true optimal ITR to be their observed treatment. Thus, we identify the true optimal ITR for every patient based on this subgrouping, which was derived from their true score `Iscore`. Since we used cross validation in estimating the ITR, we need to apply the exact same cross validation to the true optimal ITR. This is achieved by specifying the same randomization seed in the cross validation loop (see `seed`).

```
## Create new columns
dhats$d <- rep(NA, nrow(dhats)) # true d

# Identify the true optimal treatment
# See simcountdata() in the function script to learn more about Iscore
sim <- df %>%
  mutate(trueT = ifelse(as.numeric(Iscore) < 3, 1, 0),
         trueT = ifelse(Iscore == 3, drug1, trueT)) # neutral group

# Format data
input <- data.frame(y = sim$y, trt = sim$drug1, time = log(sim$years), sim[covars])

# Cross validation loop
for(i in unique(dhats$cv.i)) {
  seed <- base.seed*100 + as.numeric(str_extract(i, "[0-9]+"))
  set.seed(seed)

  # Create CV folds
  folds <- createFolds(input$trt, k = n.fold, list = TRUE) # Stratified CV, follow the sam

  for (fold.i in 1:n.fold){
    testdata <- sim[folds[[fold.i]],]
    # number of methods which succeeded for the given fold/batch. The "is.na(dhat) == FALS
    nr <- nrow(dhats %>% filter(fold == paste0("fold", fold.i), cv.i == i, is.na(dhat) ==
    dhats$d[which(dhats$fold == paste0("fold", fold.i) & dhats$cv.i == i & is.na(dhats$dha
    stopifnot(nr %% nrow(testdata) == 0)
  }
} # end of all cv iterations
```

Once we identified the true optimal ITR (d^{opt}), we can calculate the accuracy in each validation fold for each PM method ($\hat{d}_{pm}$). Mathematically, accuracy can be expressed as

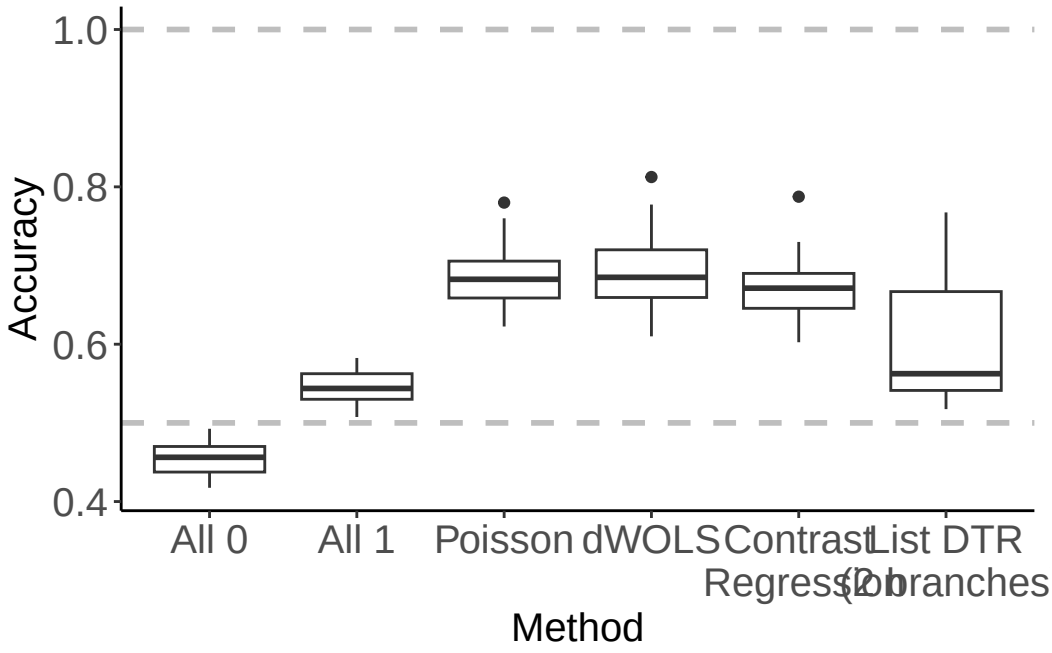
$$Accuracy_{pm}(x^{val}) = \frac{1}{n^{val}} \sum_{i=1}^{n^{val}} I(\hat{d}_{pm}(x_i^{val}) == d^{opt}(x_i^{val})),$$

where n^{val} is the sample size in the validation fold, x_i^{val} is the baseline characteristics of the

ith patient in the validation fold, and pm stands for one PM method.

Below is how Figure 2 in the chapter was generated. It summarized the accuracy across all validation folds as a box plot so we can also learn the variability of accuracy across folds.

```
##### Accuracy #####
## Calculate % accuracy for each iteration & summary statistics
dhats.accuracy <- dhats %>%
  group_by(method, cv.i, fold) %>%
  summarise(accuracy = sum(dhat == d)/n(), .groups = "drop") %>%
  ungroup

## Make the accuracy plot, Figure 2
dhats.accuracy %>%
  ggplot(aes(x = method, y = accuracy)) +
  geom_boxplot() +
  geom_hline(yintercept = 1, linetype = 2, linewidth = 1, color = "gray") +
  geom_hline(yintercept = 0.5, linetype = 2, linewidth = 1, color = "gray") +
  theme_classic() +
  labs(x = "Method", y = "Accuracy") +
  theme(axis.text = element_text(size = 15),
        axis.title.y = element_text(size = 15),
        axis.title.x = element_text(size = 15),
        axis.text.x = element_text(angle = 0, size = 15),
        strip.text.x = element_text(size = 15))
```



9.3.3 ITR agreement

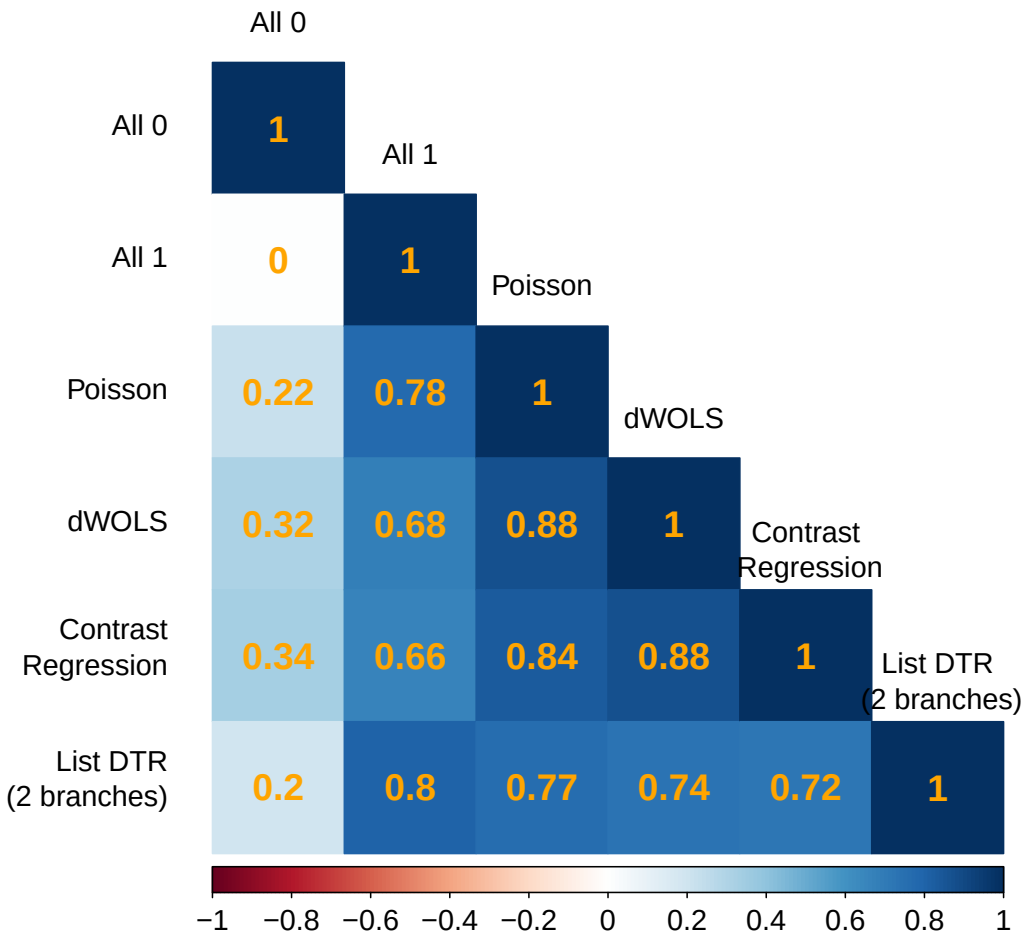
When we do not know the true data generating mechanism, e.g., real-world data, we cannot compare the estimated ITR with the true optimal ITR. However, we can compare the estimated ITR with another estimated ITR, and this is called agreement. Agreement is the proportion of patients whose estimated ITR of a method is the same as the estimated ITR of another method. Thus, agreement is between two methods. Mathematically,

$$Agreement_{1,2}(x^{val}) = \frac{1}{n^{val}} \sum_{i=1}^{n^{val}} I(\hat{d}_1(x^{val}) == \hat{d}_2(x^{val})),$$

where n^{val} is the sample size in the validation fold, x_i^{val} is the baseline characteristics of the ith patient in the validation fold, and $1, 2$ stands for method 1 and method 2.

```
##### Agreement #####
dhats.concat <- dhats %>%
  arrange(cv.i, fold, method) %>%
  mutate(iteration.fold = (as.numeric(str_extract(cv.i, "[0-9]+")) - 1) * 10 + as.numeric(
  dplyr::select(method, iteration.fold, dhat) %>%
  group_by(method, iteration.fold) %>%
  mutate(i = 1:n()) %>%
  ungroup
```

```r
m <- length(method.vec)
dhats.agreement <- matrix(nrow = m, ncol = m)
colnames(dhats.agreement) <- method.vec
rownames(dhats.agreement) <- method.vec

for(k in seq_len(m)){
  for(j in seq(k, m)){
    data.k <- dhats.concat %>% filter(method == method.vec[k])
    data.j <- dhats.concat %>% filter(method == method.vec[j])
    data.jk <- data.k %>% full_join(data.j, by = c("iteration.fold", "i"))
    dhats.agreement[k, j] <- dhats.agreement[j, k] <- sum(data.jk$dhat.x == data.jk$dhat.y
  }
}

# Make the agreement plot, Figure 3
corrplot(dhats.agreement, method = "color",  type = "lower",
         addCoef.col = "orange", number.cex = 1.5,
         tl.cex = 1.2, cl.cex = 1.2, tl.col = "black", tl.srt = 0, tl.offset = 1.5)
```



We used the `corrplot` package to generate Figure 3 in the chapter but agreement can be visualized in other creative ways that you prefer.

9.4 Patient well-being

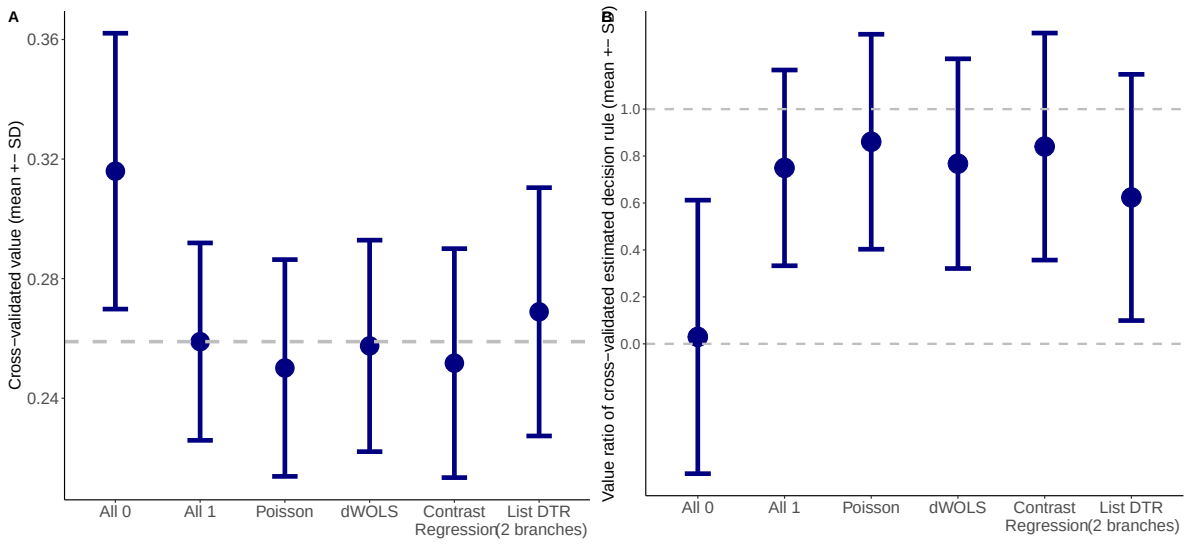
Patient well-being is evaluated via the value function, which is defined as the expected outcome had they followed the specified ITR. Like a fortune teller's crystal ball, this metric tells us

how well the patients would do on average under each ITR. We can then compare across different ITRs and identify an optimal ITR. Cross validation is necessary here to mitigate over-fitting, and we visualized the value function results as error bar plots. The mean and standard deviation of the value functions have been preprocessed previously. We use `ggplot()` to generate the error bar plots. Figure 4A is the original value function estimates, and Figure 4B is the standardized value ratio estimates, which convert value functions to a ratio where 1 is always more desirable.

```
##### Errorbar plot #####
# Figure 4A
p4a <- vhats.dhat %>%
  ggplot(aes(x = method, y = meanV)) +
  geom_point(size = 8, shape = 16, color = "navy") +
  geom_errorbar(aes(ymin = meanV - sdV, ymax = meanV + sdV), width = 0.3, size = 2, positi
  theme_classic() + xlab("") + ylab("Cross-validated value (mean +- SD)") +
  theme(axis.text = element_text(size = 15), axis.title.y = element_text(size = 15)) +
  geom_hline(yintercept = vhats.dhat$meanV[which(vhats.dhat$method == "All 1")], linetype
```

```
Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
i Please use `linewidth` instead.
```

```
##### Value ratio #####
# Figure 4B
p4b <- vhats.dhat %>%
  dplyr::select(method, contains("VR"), n.nonnaU) %>%
  ggplot(aes(x = method, y = meanVR)) +
  geom_point(size = 8, color = "navy") +
  geom_errorbar(aes(ymin = meanVR - sdVR, ymax = meanVR + sdVR), width = 0.3, size = 2, po
  geom_hline(yintercept = 1, color = "gray", linetype = 2, size = 1) +
  geom_hline(yintercept = 0, color = "gray", linetype = 2, size = 1) +
  scale_y_continuous(breaks = seq(0, 1, length = 6)) +
  theme_classic() +
  labs(x = "", y = "Value ratio of cross-validated estimated decision rule (mean +- SD)")
  theme(axis.text = element_text(size = 13),
        axis.title.y = element_text(size = 15),
        axis.title.x = element_text(size = 15),
        axis.text.x = element_text(size = 15),
        strip.text.x = element_text(size = 12))

# Figure 4
ggarrange(p4a, p4b, ncol = 2, nrow = 1, labels = c("A", "B"))
```

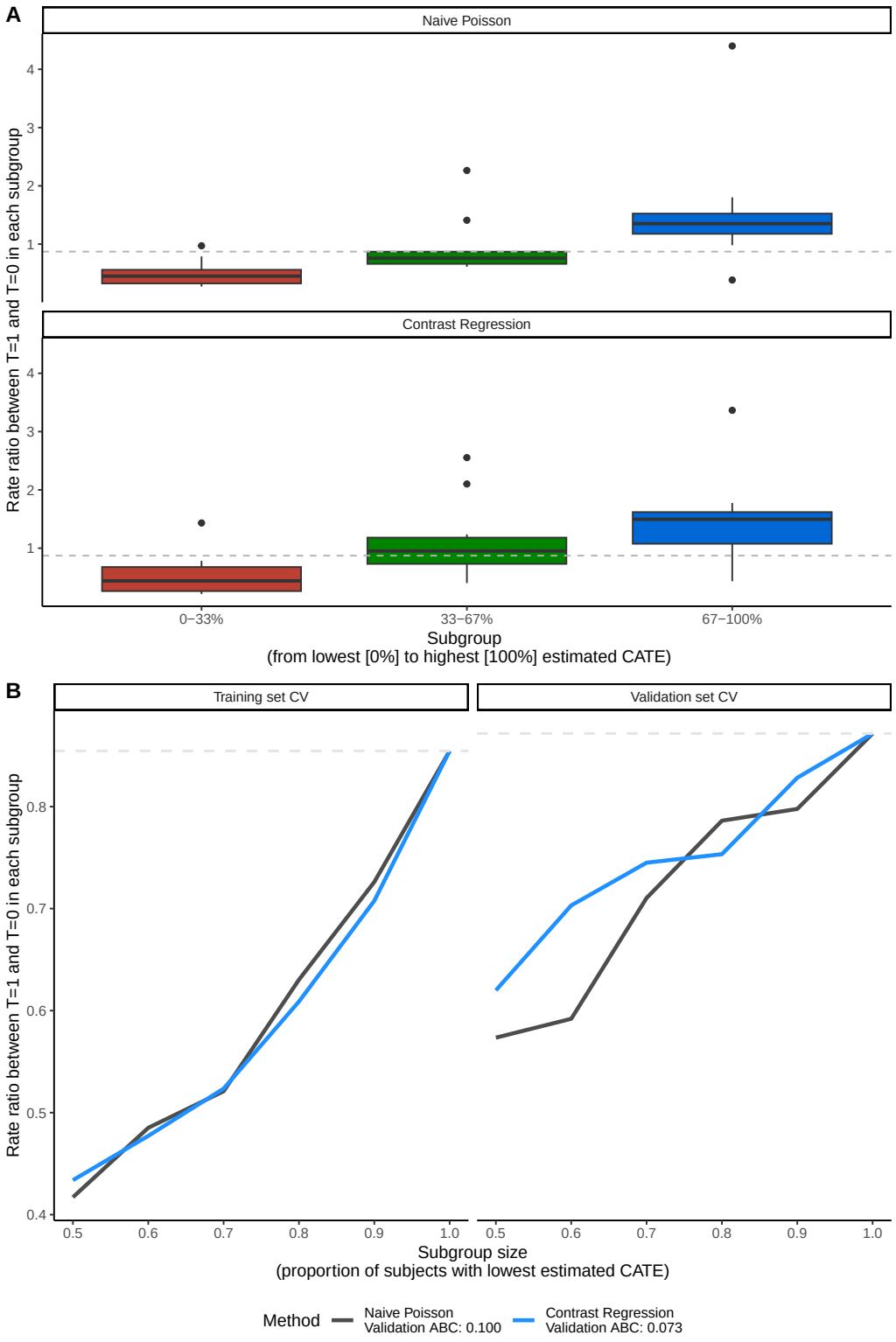


9.5 Responder diagnostics

9.5.1 Validation

The package we used for the two score-based methods (Poisson and contrast regression), `PrecMed`, has built-in visualization tools to diagnose the results: validation box plots `boxplot()`, validation curves `plot()`, and area between curves (ABC) statistics `abc()`.

```
##### Validation of ITR scores #####
# Figure 5A
p5a <- boxplot(modcv, ylab = "Rate ratio between T=1 and T=0 in each subgroup")
# Figure 5B
p5b <- plot(modcv, ylab = "Rate ratio between T=1 and T=0 in each subgroup")

# Figure 5
ggarrange(p5a, p5b, ncol = 1, nrow = 2, labels = c("A", "B"))
```


A
Naive Poisson
Contrast Regression
Rate ratio between T=1 and T=0 in each subgroup
0–33%
33–67%
67–100%
Subgroup
(from lowest [0%] to highest [100%] estimated CATE)
B
Training set CV
Validation set CV
Rate ratio between T=1 and T=0 in each subgroup
Subgroup size
(proportion of subjects with lowest estimated CATE)
Method
Naive Poisson
Validation ABC: 0.100
Contrast Regression
Validation ABC: 0.073

The `PrecMed` package has more PM methods implemented other than Poisson and contrast regression that you can try, such as negative binomial and two regressions. See its documentation for more details.

9.5.2 Univariate comparision of patient characteristics

The 60/40 cutoff was used in the chapter to split patients into "high responders" and "standard responders". The function `CreateTableOne()` in the `tableone` package was used to generate a table comparing side-by-side the baseline characteristics between the two responder groups.

```
##### Side-by-side baseline characteristic comparison between responder subgroups #####
cutoff <- quantile(modpm$score.poisson, 0.6) # 60/40 high vs standard responder split
df["responder to T=1"] <- ifelse(modpm$score.poisson < cutoff, "High", "Standard")
df["age.z"] <- as.numeric(df$age.z)
df["previous_cost.z"] <- as.numeric(df$previous_cost.z)

labs <- list("age.z" = "Standardized baseline age", "female" = "Female",
             "prevtrtB" = "Previous treatment drug B", "prevtrtC" = "Previous treatment dr
             "prevnumsymp1" = "Previous number of symptoms == 1",
             "prevnumsymp2p" = "Previuos number of symptoms >= 2",
             "previous_cost.z" = "Standardized previous medical cost\n(excluding medicatio
             "previos_number_relapses" = "Previous number of relapses")

tab <- CreateTableOne(vars = covars, strata = "responder to T=1", data = df, test = F) %>%
```

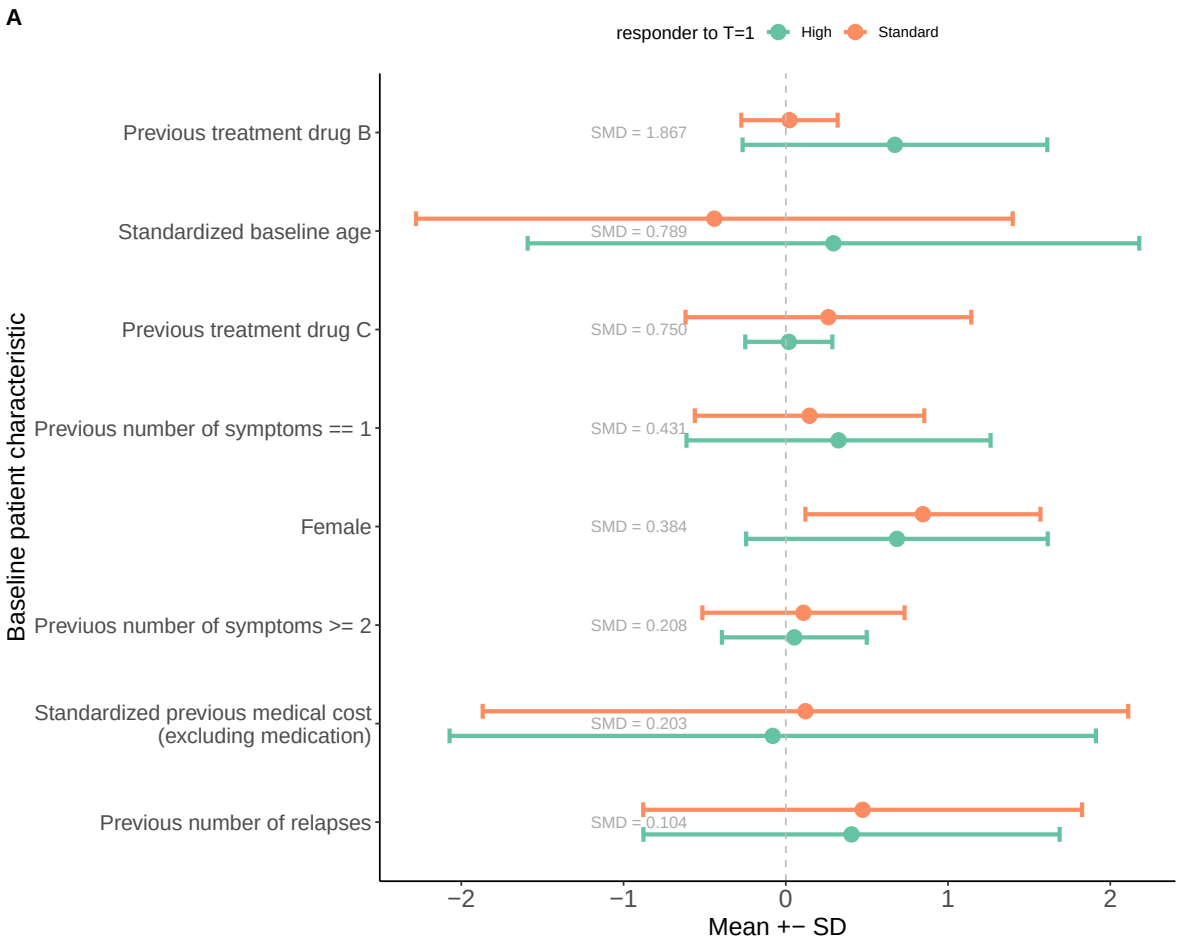
```
                                  Stratified by responder to T=1
                                   High          Standard      SMD
n                                    1200            800
age.z (mean (SD))                    0.29 (0.94) -0.44 (0.92)  0.789
female (mean (SD))                   0.69 (0.46)  0.84 (0.36)  0.384
prevtrtB (mean (SD))                 0.67 (0.47)  0.02 (0.15)  1.867
prevtrtC (mean (SD))                 0.02 (0.13)  0.26 (0.44)  0.750
prevnumsymp1 (mean (SD))             0.32 (0.47)  0.15 (0.35)  0.431
prevnumsymp2p (mean (SD))            0.05 (0.22)  0.11 (0.31)  0.208
previous_cost.z (mean (SD))         -0.08 (1.00)  0.12 (0.99)  0.203
previous_number_relapses (mean (SD)) 0.41 (0.64)  0.47 (0.68)  0.104
```

We can directly present the table or visualize the comparison with errorbar plots, which is what the chapter presented (Figure 6A). Here we show both. Tables are helpful if specific numbers are important but readers would have to perform mental comparison to understand

which value is higher, whereas plots are helpful if you want people to quickly identify the larger differences and not focus on the specific values of certain results.

```
smd <- as_tibble(tab, rownames = "var") %>%
  rowwise() %>%
  mutate(variable = as.factor(str_extract(var, ".*(?= \\(mean \\(SD\\)\\))"))) %>%
  filter(!is.na(variable)) %>%
  arrange(desc(SMD)) %>%
  mutate(smd = paste0("SMD =", SMD),
         variable = labs[[variable]]) %>%
  dplyr::select(variable, smd) %>%
  mutate(ID = 1)

levels <- unique(smd$variable)

p6a <- df %>%
  mutate(ID = 1:n()) %>%
  dplyr::select(all_of(covars), ID, contains("responder")) %>%
  melt(id = c("ID", "responder to T=1")) %>%
  rowwise() %>%
  mutate(variable = labs[[variable]]) %>%
  left_join(smd, by = c("variable", "ID")) %>%
  mutate(variable2 = factor(variable, levels = levels)) %>%
  ggplot(aes(x = reorder(variable2, desc(variable2)), color = `responder to T=1`, y = valu
  stat_summary(fun = mean, geom = "point", size = 4, position = position_dodge(width = 0.5
  stat_summary(fun.data = mean_sdl, geom = "errorbar", position = position_dodge(width = 0
  geom_hline(yintercept = 0, color = "gray", linetype = "dashed") +
  geom_text(aes(label = smd), hjust = -0.5, y = -1.5, color = "darkgray", size = 3.5) +
  # facet_wrap(~ variable2, nrow = 4) +
  labs(x = "Baseline patient characteristic",
       y = "Mean +- SD") +
  coord_flip() +
  scale_color_brewer(palette = "Set2")  +
  theme_classic() +
  theme(legend.position = "top",
        axis.text = element_text(size = 13),
        axis.title.y = element_text(size = 15),
        axis.title.x = element_text(size = 15),
        axis.text.x = element_text(size = 15),
        strip.text.x = element_text(size = 12))

# Figure 6A
```

```
ggarrange(p6a, nrow = 1, labels = c("A"))
```



We can also show the density of ITR scores obtained from the score-based methods. The results can be found in `modpm` and we used histogram to visualize (Figure 6B).

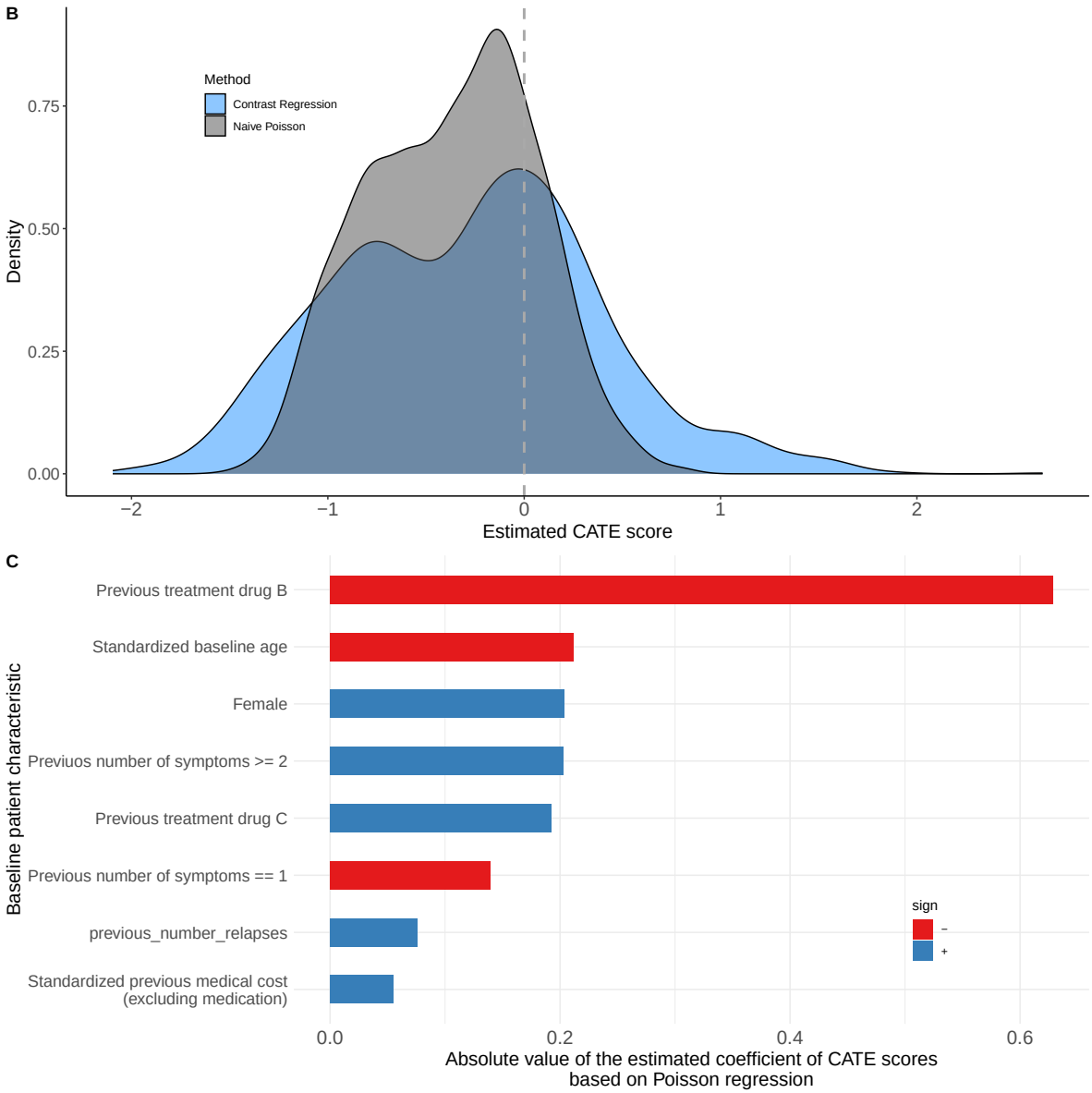
```
##### Density of ITR score #####
dataplot <- data.frame(score = factor(rep(c("Naive Poisson", "Contrast Regression"),
                                          each = length(modpm$score.poisson))),
                       value = c(modpm$score.poisson, modpm$score.contrastReg))

p6b <- dataplot %>%
  ggplot(aes(x = value, fill = score)) +
  geom_density(alpha = 0.5) +
```

```
  scale_fill_manual(values = c("dodgerblue", "gray30")) +
  geom_vline(xintercept = 0, color = "darkgray", linetype = "dashed", size = 1) +
  labs(x = "Estimated CATE score", y = "Density", fill = "Method") +
  theme_classic() +
  theme(legend.position = c(0.2, 0.8),
        axis.text = element_text(size = 13),
        axis.title.y = element_text(size = 15),
        axis.title.x = element_text(size = 15),
        axis.text.x = element_text(size = 15),
        strip.text.x = element_text(size = 12))
```

The ITR scores are essentially a linear combination of the baseline characteristics, thus it might be also of interest for one to know the corresponding coefficients (or weights) which shows how much each baseline variable contributed to the ITR score. To make it comparable across different scales of the baseline variables, we used the scaled data and the model result `modpm.s` was used to extract the coefficients and visualize as a bar plot. The coefficients can be presented in a table as well.

```
# Coefficients
coef <- modpm.s$coefficients

p6c <- coef %>%
  as_tibble(rownames = "varname") %>%
  melt(id.vars = "varname") %>%
  filter(variable == "poisson", varname != "(Intercept)") %>%
  mutate(absval = abs(value),
         sign = ifelse(value > 0, "+", "-")) %>%
  arrange(absval) %>%
  mutate(varname = factor(varname, levels = unique(varname))) %>%
  ggplot(aes(x = varname, y = absval, fill = sign)) +
  geom_bar(stat = "identity", width = 0.5) +
  scale_fill_brewer(palette = "Set1") +
  scale_x_discrete(labels = labs) +
  coord_flip() +
  labs(y = "Absolute value of the estimated coefficient of CATE scores\nbased on Poisson r
  theme_minimal() +
  theme(legend.position = c(0.8, 0.2),
        axis.text = element_text(size = 13),
        axis.title.y = element_text(size = 15),
        axis.title.x = element_text(size = 15),
        axis.text.x = element_text(size = 15),
```

```
        strip.text.x = element_text(size = 12))

# Figure 6B, 6C
ggarrange(p6b, p6c, nrow = 2, labels = c("B", "C"))
```



	poisson	contrastReg	SE_contrastReg
(Intercept)	-0.30	-0.25	0.46
age.z	-0.21	-0.23	0.17
female	0.20	0.29	0.43
prevtrtB	-0.63	-0.86	0.36
prevtrtC	0.19	0.21	0.54
prevnumsymp1	-0.14	-0.42	0.39
prevnumsymp2p	0.20	1.05	0.58
previous_cost.z	0.06	0.13	0.18
previous_number_relapses	0.08	0.26	0.23

```
# Coefficients presented as a table
coef %>% round(2) %>% kable() %>% kable_styling(full_width = F)
```

Version info

This chapter was rendered using the following version of R and its packages:

```
R version 4.2.3 (2023-03-15 ucrt)
Platform: x86_64-w64-mingw32/x64 (64-bit)
Running under: Windows 10 x64 (build 19045)

Matrix products: default

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

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

other attached packages:
 [1] fastDummies_1.6.3 reshape2_1.4.4    truncnorm_1.0-9   table1_1.4.3
 [5] kableExtra_1.3.4  knitr_1.43        ggpubr_0.6.0      MASS_7.3-60
 [9] corrplot_0.92     caret_6.0-94      lattice_0.21-8    gbm_2.1.8.1
[13] tableone_0.13.2   rpart.plot_3.1.1  rpart_4.1.19      precmed_1.0.0
[17] DTRreg_1.7        magrittr_2.0.3    lubridate_1.9.2   forcats_1.0.0
[21] stringr_1.5.0     dplyr_1.1.2       purrr_1.0.1       readr_2.1.4
```

```
[25] tidyr_1.3.0       tibble_3.2.1      ggplot2_3.4.2     tidyverse_2.0.0

loaded via a namespace (and not attached):
  [1] colorspace_2.1-0      ggsignif_0.6.4        class_7.3-22
  [4] ggridges_0.5.4        htmlTable_2.4.1       ggstance_0.3.6
  [7] base64enc_0.1-3       proxy_0.4-27          rstudioapi_0.14
 [10] listenv_0.9.0         farver_2.1.1          prodlim_2023.03.31
 [13] fansi_1.0.4           xml2_1.3.4            codetools_0.2-19
 [16] splines_4.2.3         polyclip_1.10-4       Formula_1.2-5
 [19] jsonlite_1.8.5        pROC_1.18.2           broom_1.0.5
 [22] cluster_2.1.4         geepack_1.3.9         ggforce_0.4.1
 [25] data.tree_1.0.0       httr_1.4.6            DiagrammeR_1.0.10
 [28] compiler_4.2.3        randomForestSRC_3.2.2 backports_1.4.1
 [31] Matrix_1.5-4.1        fastmap_1.1.1         survey_4.2-1
 [34] cli_3.6.1             tweenr_2.0.2          visNetwork_2.1.2
 [37] htmltools_0.5.5       tools_4.2.3           gtable_0.3.3
 [40] glue_1.6.2            MESS_0.5.9            Rcpp_1.0.10
 [43] carData_3.0-5         vctrs_0.6.3           svglite_2.1.1
 [46] nlme_3.1-162          iterators_1.0.14      timeDate_4022.108
 [49] gower_1.0.1           xfun_0.39             globals_0.16.2
 [52] rvest_1.0.3           timechange_0.2.0      lifecycle_1.0.3
 [55] geeM_0.10.1           mosaicCore_0.9.2.1    rstatix_0.7.2
 [58] future_1.32.0         scales_1.2.1          ipred_0.9-14
 [61] hms_1.1.3             parallel_4.2.3        RColorBrewer_1.1-3
 [64] yaml_2.3.7            gridExtra_2.3         labelled_2.11.0
 [67] gam_1.22-2            stringi_1.7.12        foreach_1.5.2
 [70] checkmate_2.2.0       e1071_1.7-13          hardhat_1.3.0
 [73] lava_1.7.2.1          shape_1.4.6           systemfonts_1.0.4
 [76] rlang_1.1.1           pkgconfig_2.0.3       evaluate_0.21
 [79] labeling_0.4.2        recipes_1.0.6         htmlwidgets_1.6.2
 [82] cowplot_1.1.1         tidyselect_1.2.0      parallelly_1.36.0
 [85] ggformula_0.10.4      plyr_1.8.8            R6_2.5.1
 [88] Hmisc_5.1-0           generics_0.1.3        DBI_1.1.3
 [91] foreign_0.8-84        pillar_1.9.0          haven_2.5.2
 [94] withr_2.5.0           abind_1.4-5           survival_3.5-5
 [97] nnet_7.3-19           future.apply_1.11.0   car_3.1-2
[100] utf8_1.2.3            tzdb_0.4.0            rmarkdown_2.22
[103] grid_4.2.3            data.table_1.14.8     ModelMetrics_1.2.2.2
[106] webshot_0.5.4         digest_0.6.31         stats4_4.2.3
[109] munsell_0.5.0         glmnet_4.1-7          viridisLite_0.4.2
[112] mitools_2.4
```

References

Baker, William L, Erica L Baker, and Craig I Coleman. 2009. "Pharmacologic Treatments for Chronic Obstructive Pulmonary Disease: A Mixed-Treatment Comparison Meta-Analysis." *Pharmacotherapy* 29 (8): 891–905. https://doi.org/10.1592/phco.29.8.891.

Coulombe, Janie, Erica E. M. Moodie, and Robert W. Platt. 2020. "Weighted Regression Analysis to Correct for Informative Monitoring Times and Confounders in Longitudinal Studies." *Biometrics* 77 (1): 162–74. https://doi.org/10.1111/biom.13285.

Coulombe, Janie, Erica E. M. Moodie, Robert W. Platt, and Christel Renoux. 2022. "Estimation of the Marginal Effect of Antidepressants on Body Mass Index Under Confounding and Endogenous Covariate-Driven Monitoring Times." *The Annals of Applied Statistics* 16 (3). https://doi.org/10.1214/21-aoas1570.

Debray, Thomas PA, Gabrielle Simoneau, Massimiliano Copetti, Robert W Platt, Changyu Shen, Fabio Pellegrini, and Carl de Moor. 2023. "Methods for Comparative Effectiveness Based on Time to Confirmed Disability Progression with Irregular Observations in Multiple Sclerosis." *Statistical Methods in Medical Research*, June, 096228022311720. https://doi.org/10.1177/09622802231172032.

Panaccione, Remo, Eric B Collins, Gil Y Melmed, Severine Vermeire, Silvio Danese, Peter D R Higgins, Christina S Kwon, et al. 2023. "Efficacy and Safety of Advanced Therapies for Moderately to Severely Active Ulcerative Colitis at Induction and Maintenance: An Indirect Treatment Comparison Using Bayesian Network Meta-Analysis." *Crohn's & Colitis 360* 5 (2). https://doi.org/10.1093/crocol/otad009.

Selvarajan, Sandhiya, Annuja Anandaradje, Santhosh Shivabasappa, Deepthy Melepurakkal Sadanandan, N. Sreekumaran Nair, and Melvin George. 2022. "Efficacy of Pharmacological Interventions in COVID-19: A Network Meta-Analysis." *British Journal of Clinical Pharmacology* 88 (9): 4080–91. https://doi.org/10.1111/bcp.15338.

Siemieniuk, Reed AC, Jessica J Bartoszko, Dena Zeraatkar, Elena Kum, Anila Qasim, Juan Pablo Díaz Martinez, Ariel Izcovich, et al. 2020. "Drug Treatments for Covid-19: Living Systematic Review and Network Meta-Analysis." *BMJ*, July, m2980. https://doi.org/10.1136/bmj.m2980.

Yadlowsky, Steve, Fabio Pellegrini, Federica Lionetto, Stefan Braune, and Lu Tian. 2020. "Estimation and Validation of Ratio-Based Conditional Average Treatment Effects Using Observational Data." *Journal of the American Statistical Association* 116 (533): 335–52. https://doi.org/10.1080/01621459.2020.1772080.

Zhao, Lihui, Lu Tian, Tianxi Cai, Brian Claggett, and L. J. Wei. 2013. "Effectively Selecting a Target Population for a Future Comparative Study." *Journal of the American Statistical Association* 108 (502): 527–39. https://doi.org/10.1080/01621459.2013.770705.