R markdown for Supplementary PDF

Samuel Tingle

2025-06-16

ONCE THIS IS FINALISED –> MAKE A MARKDOWN WHICH WILL HAVE THE PDF IN IT WILL ALSO NEED TO HAVE V1.0.0 ON CRAN note will need to clean up this top bit later and just load rmsMD in with CRAN rather than through github NOTE WHEN LETTER FINALISED, CHANGE THINGS TO BE LIKE "FOR FIGURE 2" ETC FOR CLARITY ADD REFERENCE TO THE MANUSCRIPT ONCE PUBLISHED

Data loading and set-up

Load in simulated data from the simulate_data.R file. This data was simulated for this manuscript. The true relationship is that age has a linear relationship and BMI has a U-shapred relationship.

We will assume that data has been appropriately examined, plotted and cleaned

```
# load in the simulated data
source("simulate_data.R")

# Set up data distribution information for rms. These two lines are standard
# when using the rms package
dd <- datadist(data)
options(datadist = "dd")
```

Data summary and demographics table

This is using table1() from the `table1` package to give a summary of the simulated data.

```
table1(~age + bmi + sex + smoking + as.factor(majorcomplication), data = data)
```

	Overall
	(N=5000)
age	
Mean (SD)	49.9 (12.1)
Median [Min, Max]	50.1 [7.40, 90.2]
bmi	
Mean (SD)	25.0 (3.96)
Median [Min, Max]	25.0 [11.2, 39.2]
sex	
Female	2467 (49.3%)
Male	2533 (50.7%)
smoking	
Never	1678 (33.6%)
Former	1677 (33.5%)
Current	1645 (32.9%)
as.factor(majorcomplication)	
0	4169 (83.4%)
1	831 (16.6%)

Model building and basic rmsMD outputs

```
# Fit logistic regression model using restricted cubic splines

fit <- lrm(majorcomplication ~ rcs(age, 3) + rcs(bmi, 3) + sex + smoking,
               data = data,
               x = TRUE, y = TRUE)

# note, x = TRUE, y = TRUE is recommended for lrm and cph models to allow subsequent
# LR tests to be performed

# Check the model fit and diagnostics including number of observations and events
# Note the spline term coefficients (age, age', bmi, bmi') are difficult to interpret
fit
```

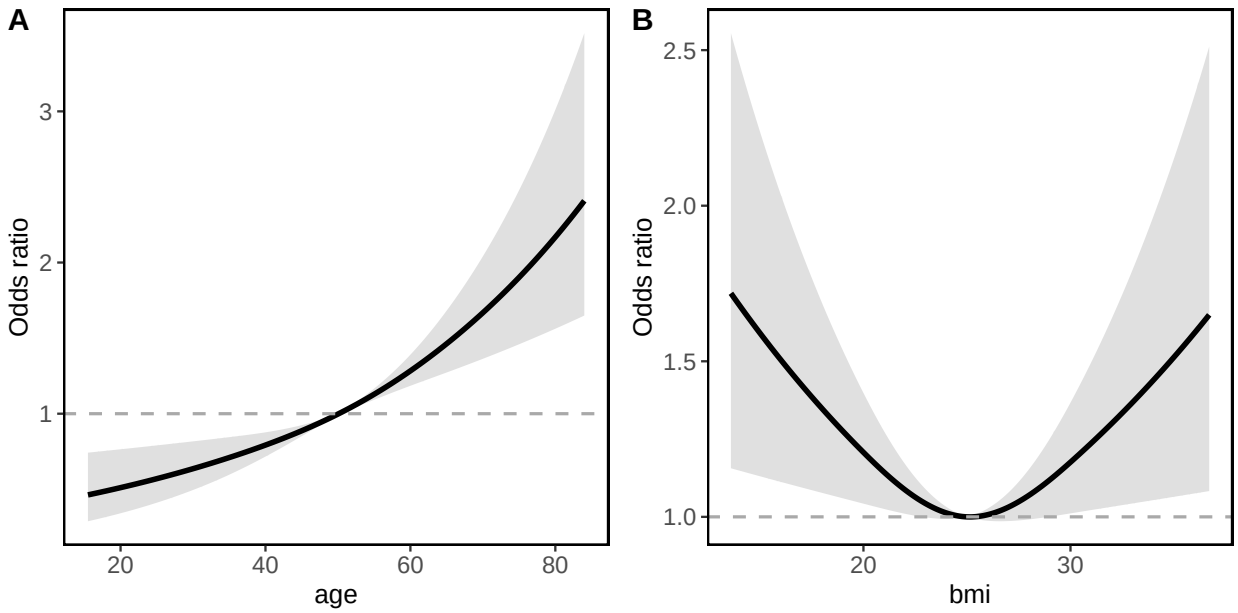
```
## Logistic Regression Model
##
## lrm(formula = majorcomplication ~ rcs(age, 3) + rcs(bmi, 3) +
##     sex + smoking, data = data, x = TRUE, y = TRUE)
##
##                       Model Likelihood       Discrimination    Rank Discrim.
##                             Ratio Test              Indexes          Indexes
## Obs          5000    LR chi2     155.74       R2       0.052    C       0.636
##  0           4169    d.f.             7      R2(7,5000)0.029    Dxy     0.271
##  1            831    Pr(> chi2) <0.0001    R2(7,2078.7)0.069    gamma   0.271
## max |deriv| 3e-07                            Brier    0.134    tau-a   0.075
##
##                Coef    S.E.   Wald Z Pr(>|Z|)
## Intercept      -1.8860 0.5600 -3.37  0.0008
## age             0.0219 0.0078  2.80  0.0052
```

```
## age'              0.0028 0.0084  0.34  0.7354
## bmi              -0.0552 0.0200 -2.76  0.0059
## bmi'              0.0711 0.0243  2.92  0.0035
## sex=Male          0.0755 0.0774  0.98  0.3293
## smoking=Former   -0.0142 0.1031 -0.14  0.8902
## smoking=Current   0.7317 0.0933  7.84  <0.0001
```

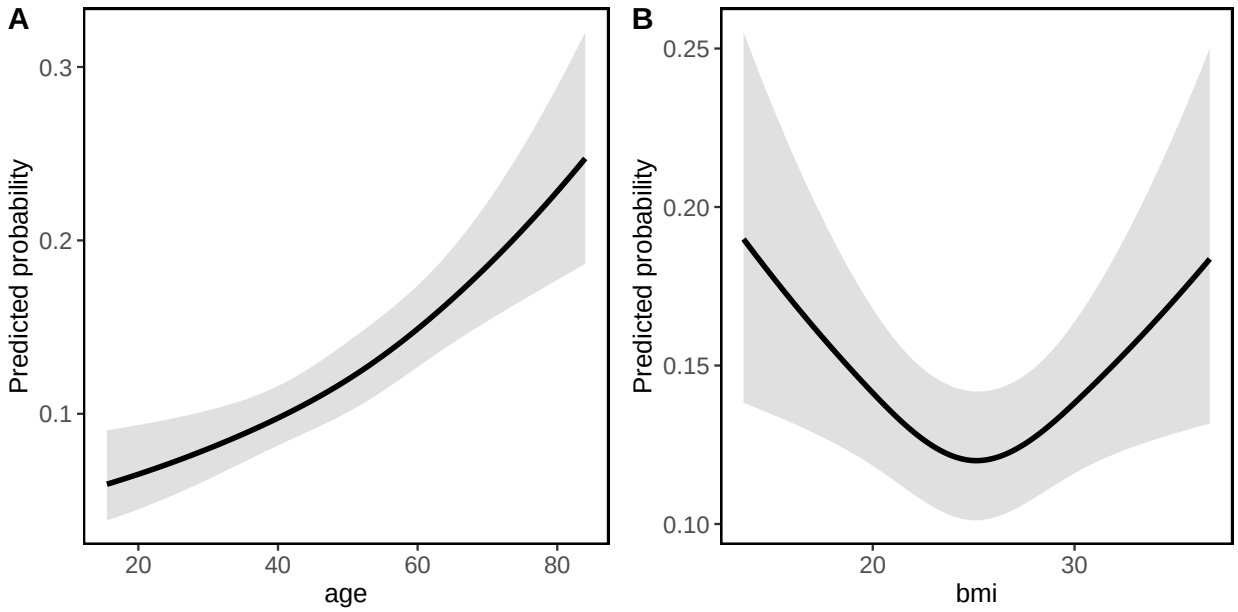
```r
# Use modelsummary_rms function from rmsMD package to get an output for medical journals
# for spline terms this gives an overall p-value for the association of that variable with outcome
modelsummary_rms(fit)
```

```
##           variable                OR_95CI Pvalue
## 1       sex=Female                    Ref      -
## 2         sex=Male 1.078 (0.927 to 1.255)  0.329
## 3    smoking=Never                    Ref      -
## 4   smoking=Former 0.986 (0.806 to 1.207)  0.890
## 5  smoking=Current 2.079 (1.731 to 2.496) <0.001
## 6 RCSoverallP: age                LR test <0.001
## 7 RCSoverallP: bmi                LR test  0.015
```

```r
# Use ggrmsMD from rmsMD package to plot the splines
# This determines which variables were analysed as RCS, and plots them appropriately
# As this is a logistic regression it plots OR and 95% confidence interval
# combined = TRUE means that a single combined plot with all spline terms in the model is outputted
ggrmsMD(fit, data, combined = TRUE)
```




```r
# for logistic regression models. the lrm_prob argument can be used to plot predicted probabilities
# rather than odds ratios
ggrmsMD(fit, data, combined = TRUE, lrm_prob = TRUE)
```




```
# To assess whether rcs variables are significantly non-linear use anova
# for each RCS term, a p-value for "Nonlinear" is given
anova(fit, test = "LR")
```

```
##                 Likelihood Ratio Statistics          Response: majorcomplication
##
##  Factor          Chi-Square d.f. P
##  age              57.40     2    <.0001
##   Nonlinear        0.11     1    0.7359
##  bmi               8.41     2    0.0149
##   Nonlinear        8.29     1    0.0040
##  sex               0.95     1    0.3291
##  smoking          88.35     2    <.0001
##  TOTAL NONLINEAR   8.38     2    0.0151
##  TOTAL           155.74     7    <.0001
```

```
# As expected, BMI, but not age, is significalty non-linear
```

Publication ready outputs using rmsMD

Publication ready tables

The packages `flextable` and `officer` can be used to output results from modelsummary_rms() to word documents. These are used in Table 2 of the accompanying manuscript.

```
# modelsummary_rms will output a dataframe. first make this into a flextable
results <- modelsummary_rms(fit)
results_flextable <- flextable(results)

# output to a word document
```

```r
doc <- read_docx()
doc <- body_add_flextable(doc, results_flextable)
print(doc, target = "Results_of_main_model.docx")

# creating a model which does not use RCS terms to use for comparison
# note this model assumes linear relationships, and incorrectly finds no association
# between bmi and outcome
fit_linear <- lrm(majorcomplication ~ age + bmi + sex + smoking,
           data = data,
           x = TRUE, y = TRUE)
fit_linear # check diagnostics etc
```

```
## Logistic Regression Model
##
## lrm(formula = majorcomplication ~ age + bmi + sex + smoking,
##     data = data, x = TRUE, y = TRUE)
##
##                       Model Likelihood       Discrimination    Rank Discrim.
##                             Ratio Test              Indexes          Indexes
## Obs          5000    LR chi2     147.35       R2       0.049    C       0.633
##  0           4169    d.f.             5      R2(5,5000)0.028    Dxy     0.267
##  1            831    Pr(> chi2) <0.0001    R2(5,2078.7)0.066    gamma   0.267
## max |deriv| 1e-08                            Brier    0.134    tau-a   0.074
##
##                 Coef    S.E.   Wald Z Pr(>|Z|)
## Intercept       -3.0831 0.3114 -9.90  <0.0001
## age              0.0244 0.0032  7.52  <0.0001
## bmi             -0.0035 0.0097 -0.36  0.7173
## sex=Male         0.0770 0.0773  1.00  0.3194
## smoking=Former  -0.0229 0.1029 -0.22  0.8241
## smoking=Current  0.7242 0.0932  7.77  <0.0001
```

```r
results_linear <- modelsummary_rms(fit_linear) # get results dataframe
results_linear_flextable <- flextable(results_linear)
doc <- read_docx() # set up word doc
doc <- body_add_flextable(doc, results_linear_flextable) # add results
print(doc, target = "Results_of_model_with_linear_assumption.docx")
```

Publication ready plots

Using ggrmsMD() from `rmsMD` to make Figure 2 in the accompanying manuscript.

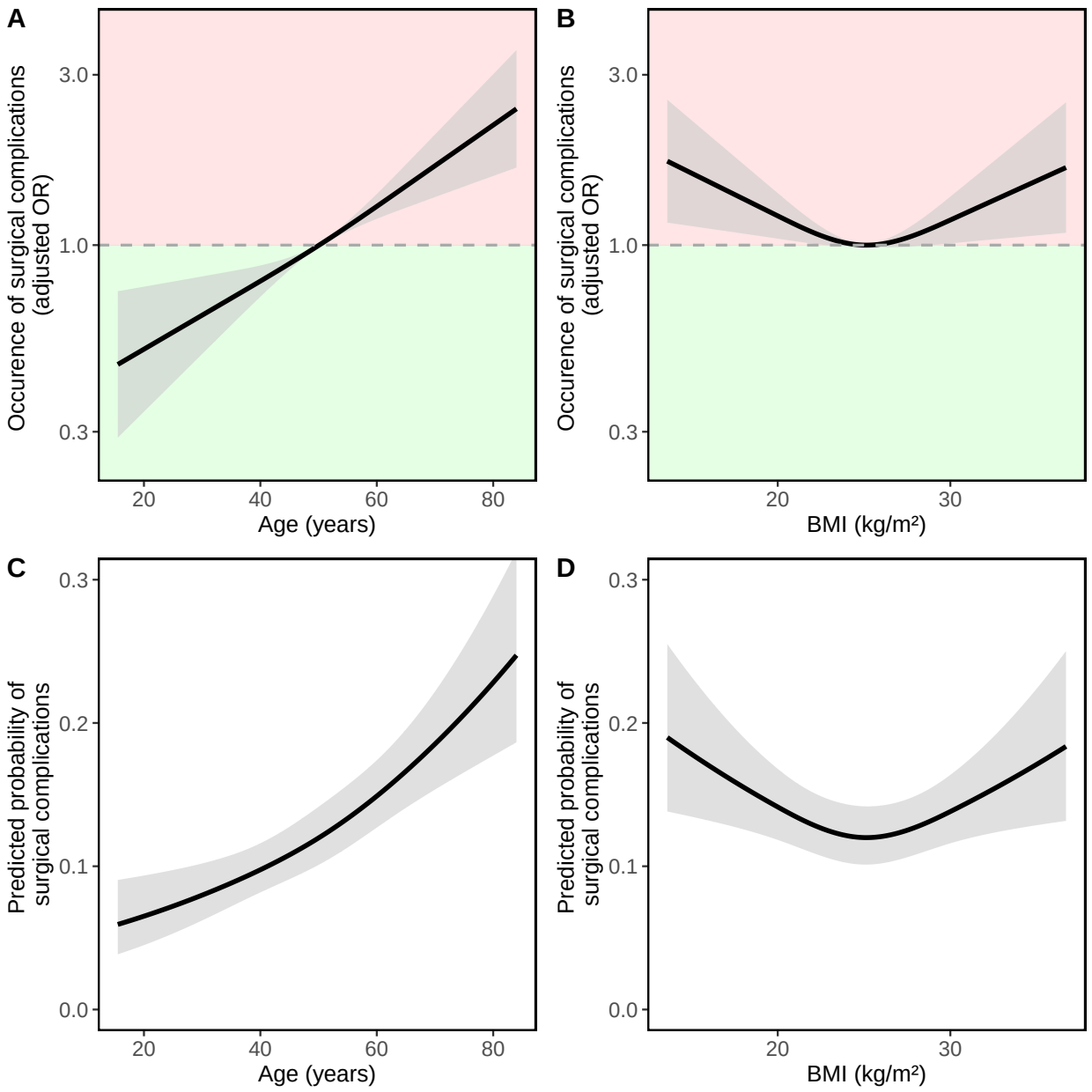
```r
# plot adjusted odds ratio. note the y axis is plotted here on a log-scale rather than
# a linear scale. The shade_inferior argument is used to shade either side of the no
# effect line (OR = 1) to give a visual cue for which side represents inferior outcome.
plots_OR <- ggrmsMD(fit, data, combined = FALSE,
                    shade_inferior = "higher",
                    ylab = "Occurence of surgical complications \n(adjusted OR)",
                    xlabs = list(age = "Age (years)", bmi = "BMI (kg/m²)"),
                    log_y = TRUE,
                    ylim = c(0.25, 4)
)
```

```r
# plot the predicted probability
plots_prob <- ggrmsMD(fit, data, combined = FALSE,
                ylab = "Predicted probability of \nsurgical complications",
                xlabs = list(age = "Age (years)", bmi = "BMI (kg/m²)"),
                lrm_prob = TRUE,
                ylim = c(0,0.3)
)

# combine the plot lists and making a single multipanel figure
plotlist <- c(plots_OR, plots_prob)

# using plot_grid from `cowplot` package
plots <- plot_grid(plotlist = plotlist, labels = "AUTO", align = "vh")

# to view the plot
plots
```




```
# to save as pdf
ggsave(file = "Case_study_RCS_plot.pdf", plot = plots, width = 8, height = 8)

# to save as jpg
ggsave(file = "Case_study_RCS_plot.jpg", plot = plots, width = 8, height = 8)
```