

Pilot study report

Background

The pilot study was performed in Zürich on 10 subjects.

TPR

First we load necessary packages.

```
library(dplyr)
```

Then, we load the preprocessed pilot study data:

```
datPilot <- readRDS("data/pilot_long.rds")
datPilot$period_ctr <- datPilot$period - mean(unique(datPilot$period))
head(datPilot)
```

```
# A tibble: 6 x 7
  center doctor scenario avatar correct period period_ctr
  <dbl> <fct>   <fct>   <fct> <fct>   <dbl>   <dbl>
1     1 1     A       0     0       1     -1.5
2     1 1     A       0     0       1     -1.5
3     1 1     A       0     0       1     -1.5
4     1 1     A       0     0       1     -1.5
5     1 1     A       0     1       1     -1.5
6     1 1     A       0     1       1     -1.5
```

The pilot study is represented as follows: Each row represents an event, where an event is a change in one of the vital signs. The event is characterized by the *scenario* it happened in, and whether the *avatar* was present or not. The *correct* column denotes whether the *doctor* made a correct prediction or not for this specific event.

Table and Analysis for TPR

We model this using a generalized linear mixed model as follows:

```
require(lmerTest)
m <- glmer(
  correct ~ scenario + avatar + period_ctr + (avatar | doctor),
  family = binomial, data = datPilot,
  control = glmerControl(optimizer = "bobyqa")
)
summary(m)
```

Generalized linear mixed model fit by maximum likelihood (Laplace
Approximation) [glmerMod]
Family: binomial (logit)
Formula: correct ~ scenario + avatar + period_ctr + (avatar | doctor)
Data: datPilot
Control: glmerControl(optimizer = "bobyqa")

AIC	BIC	logLik	-2*log(L)	df.resid
1154.3	1188.5	-570.1	1140.3	973

Scaled residuals:

Min	1Q	Median	3Q	Max
-0.8690	-0.6557	-0.5123	1.2010	2.0813

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
doctor	(Intercept)	0.000e+00	0.000e+00	
	avatar1	6.411e-15	8.007e-08	NaN

Number of obs: 980, groups: doctor, 10

Fixed effects:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-1.40186	0.12167	-11.522	< 2e-16 ***
scenarioB	0.43499	0.15311	2.841	0.0045 **
avatar1	0.62185	0.14707	4.228	2.35e-05 ***
period_ctr	0.04275	0.06443	0.664	0.5070

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

Correlation of Fixed Effects:

```

(Intr) scenrB avatr1
scenarioB -0.436
avatar1 -0.678 -0.001
period_ctr -0.020 -0.144 0.162
optimizer (bobyqa) convergence code: 0 (OK)
boundary (singular) fit: see help('isSingular')

```

```

confint(m, oldNames = FALSE, method = "Wald")

```

	2.5 %	97.5 %
sd_(Intercept) doctor	NA	NA
cor_avatar1.(Intercept) doctor	NA	NA
sd_avatar1 doctor	NA	NA
(Intercept)	-1.64031737	-1.1633967
scenarioB	0.13488948	0.7350842
avatar1	0.33360321	0.9101059
period_ctr	-0.08353036	0.1690318

We see that the point estimate of the effect of the avatar “treatment” is equal to an odds ratio of $e^{0.62} \approx 1.86$. We can interpret this as follows: **The odds of correctly predicting a vital sign change increase by a factor of 1.86 when presented with the avatar** (95%-CI: ([1.39, 2.48])).

We can look at the table of true positives with and without avatar:

```

ds <- datPilot |> filter(correct==1)
xtabs(~avatar + doctor, data = ds)

```

	doctor									
avatar	1	2	3	4	5	6	7	8	9	10
0	14	9	8	13	13	9	13	12	13	5
1	20	18	17	16	17	16	15	18	15	16

We see that the best performing doctor without the avatar (Doctor 1), is still worse (TP=14), than the worst performing doctor with the avatar (Doctors 7 and 9, TP=15).

Sample Size calculation for primary outcome

We use the effect estimates from the above model for the intercept, the period effect and the scenario effect. Other effects, such as the random effects per doctor or center, we had to guess.

For the avatar effect, we take a conservative approach and evaluate power for a smaller odds ratio equal to 1.25. This is quite a bit lower than what was observed in the pilot study, but it allows us to account for the fact that we might observe quite strong center-to-center variability in the full study.

We omit the code for brevity and computational reasons, and report only the results. Using the approach and estimates from above, with **5 study centers and 10 doctors per center**, our simulations (500 simulation runs) come to the conclusion, that for **detecting an avatar odds ratio of 1.25 the power is 0.834** (95%-CI: [0.798, 0.866]).

Lead-time

Data

Loading the lead time data:

```
lead_table <- readRDS("data/lead_table.rds")
lead_table$period_ctr <- lead_table$period - mean(unique(lead_table$period))
head(lead_table)
```

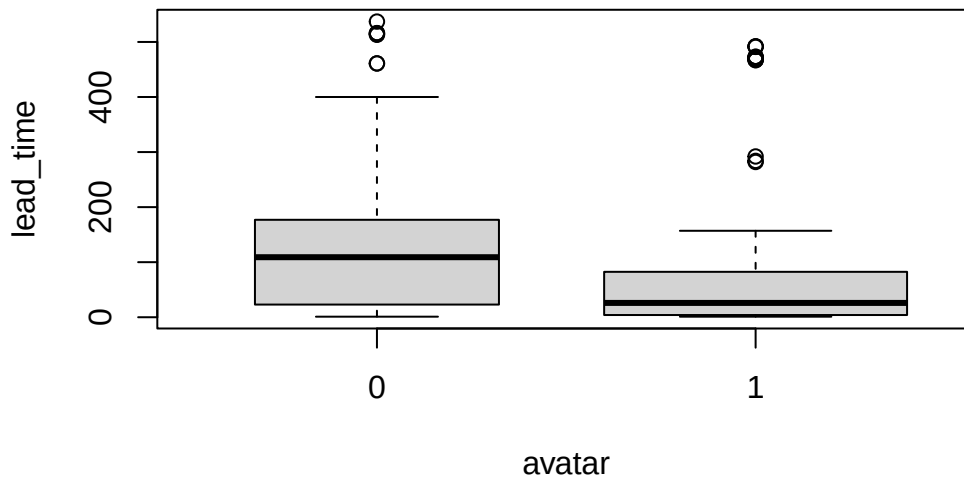
```
# A tibble: 6 x 7
  center doctor scenario avatar lead_time period period_ctr
  <int> <fct>   <fct>   <fct>   <dbl> <dbl>    <dbl>
1     1  1 1     A       0      124     1      -1.5
2     1  1 1     A       0       17     1      -1.5
3     1  1 1     A       0     109     1      -1.5
4     1  1 1     A       0     121     1      -1.5
5     1  1 1     A       0     265     1      -1.5
6     1  1 1     A       0     235     1      -1.5
```

The lead time table contains only the correctly predicted events with information on how early the event was predicted.

Plot and Analysis

We can look at the lead time visually:

```
boxplot(lead_time ~ avatar, data = lead_table, varwidth = TRUE)
```



From the boxplot we see that lead time tends to be larger without the avatar, meaning that, among the correctly predicted events, doctors tend to make earlier predictions without the avatar.

We can fit a linear mixed effects model:

```
m <- lmer(lead_time ~ scenario + avatar + period_ctr + (1 | doctor),
          data = lead_table)
summary(m)
```

Linear mixed model fit by REML. t-tests use Satterthwaite's method [lmerModLmerTest]

Formula: lead_time ~ scenario + avatar + period_ctr + (1 | doctor)

Data: lead_table

REML criterion at convergence: 3445.2

Scaled residuals:

Min	1Q	Median	3Q	Max
-1.1587	-0.6382	-0.2552	0.2201	3.4149

Random effects:

Groups	Name	Variance	Std.Dev.
doctor	(Intercept)	0	0.0
	Residual	16470	128.3

Number of obs: 277, groups: doctor, 10

Fixed effects:

Estimate	Std. Error	df	t value	Pr(> t)
----------	------------	----	---------	----------

```

(Intercept) 148.9050      13.6178 273.0000  10.935 < 2e-16 ***
scenarioB    -43.2131      16.1506 273.0000  -2.676 0.00791 **
avatar1      -51.6726      16.0347 273.0000  -3.223 0.00142 **
period_ctr    0.5277       6.8968 273.0000   0.077 0.93907
---

```

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

Correlation of Fixed Effects:

```

      (Intr) scenrB avatr1
scenarioB -0.418
avatar1   -0.692 -0.043
period_ctr -0.032 -0.166  0.176
optimizer (nloptwrap) convergence code: 0 (OK)
boundary (singular) fit: see help('isSingular')

```

```
confint(m, method = "Wald")
```

```

              2.5 %    97.5 %
.sig01         NA         NA
.sigma         NA         NA
(Intercept) 122.21469 175.59541
scenarioB    -74.86770 -11.55842
avatar1      -83.10011 -20.24512
period_ctr   -12.98977  14.04507

```

We see that when using avatar, average lead-time is smaller by around 52 seconds (point estimate: -51.67 seconds; 95%-CI: $[-83.80, -21.56]$).

Precision

Data

Loading the precision data:

```

precision_table <- readRDS("data/precision_table.rds")
precision_table$period_ctr <- precision_table$period -
  mean(unique(precision_table$period))
precision_table$precision <- precision_table$true_pred /
  (precision_table$true_pred + precision_table$false_pred)
head(precision_table)

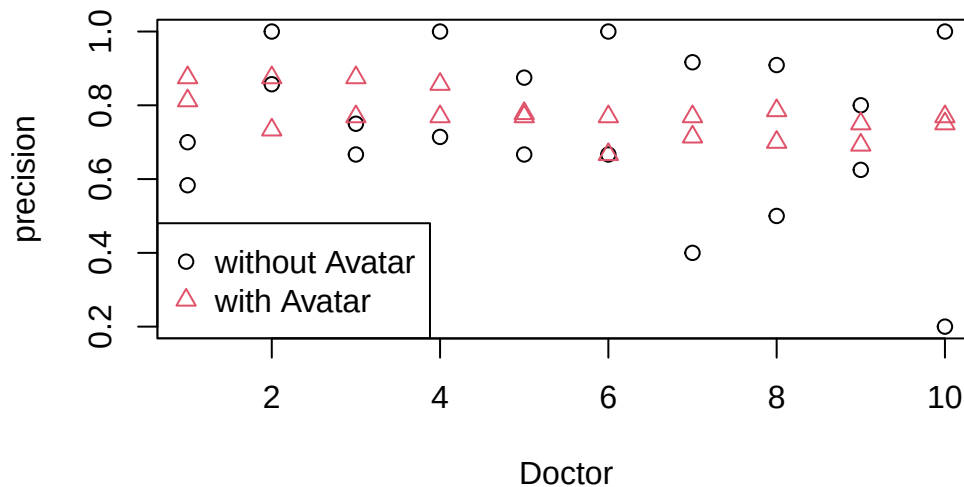
```

```
# A tibble: 6 x 9
  doctor scenario avatar period false_pred true_pred center period_ctr precision
  <fct>   <fct>   <fct>   <dbl>     <dbl>     <dbl>   <dbl>     <dbl>     <dbl>
1 1      A      0      1      3      7      1     -1.5      0.7
2 1      A      1      2      3     13      1     -0.5     0.812
3 1      B      0      3      5      7      1      0.5     0.583
4 1      B      1      4      1      7      1      1.5     0.875
5 2      A      1      1      4     11      1     -1.5     0.733
6 2      A      0      2      0      3      1     -0.5      1
```

The column “precision” contains the fraction of true active blocks divided by all active blocks.

Plot and Analysis

```
plot(precision ~ as.numeric(doctor), col = as.numeric(avatar),
     pch = as.numeric(avatar), data = precision_table, xlab = "Doctor")
legend("bottomleft", legend = c("without Avatar", "with Avatar"), col = 1:2,
     pch = 1:2)
```



For every doctor (x-axis) we see the precision on the y-axis. Every doctor has four observations (2 scenarios and with or without avatar). At first glance we do not see any striking difference between the values with and without avatar.

This is confirmed by a paired t-test. There is no significant difference in expected precision between the observations with and without avatar.

```

avg_precision_table <- precision_table |>
  group_by(doctor, avatar) |>
  summarise(avg_precision = mean(precision), .groups = 'drop')

w_avatar <- avg_precision_table |> filter(avatar == 1) |> select(avg_precision)
wo_avatar <- avg_precision_table |> filter(avatar == 0) |> select(avg_precision)

t.test(w_avatar$avg_precision, wo_avatar$avg_precision, paired = TRUE)

```

Paired t-test

```

data:  w_avatar$avg_precision and wo_avatar$avg_precision
t = 0.93785, df = 9, p-value = 0.3728
alternative hypothesis: true mean difference is not equal to 0
95 percent confidence interval:
 -0.04586329  0.11082264
sample estimates:
mean difference
 0.03247967

```