STAT 426: Final Project

Alice Cao, Colleen Jung, Jingyang Yu

Introduction

We aim to define and model the presence of the bacteria H. influenzae in children with a middle ear infection in the Northern Territory of Australia.

- What is H. influenzae?

Haemophilus influenzae disease is a name for any illness caused by bacteria called H. influenzae. It was once the most common cause of bacterial infection in children.

- What is Otitis media?

Otitis media is inflammation or infection located in the middle ear. The predominant bacteria that cause otitis media are cause of otitis media are Streptococcus pneumoniae, Moraxella catarrhalis, and non-typeable Haemophilus influenzae(H. influenzae).

Research Questions:

1. Is the treatment effective?

2. How do they behave over time?

3. Any compliance effect?

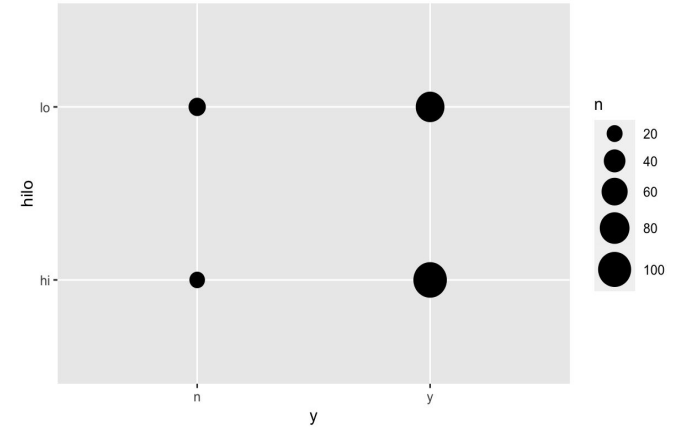
The Data

Check Baseline Proportions
There isn't appropriate approach of detecting outliers for categorical data (especially binary data).
Though we tried to apply criterions like leverage and Mahalanobis distance, neither of which turned out to be effective.

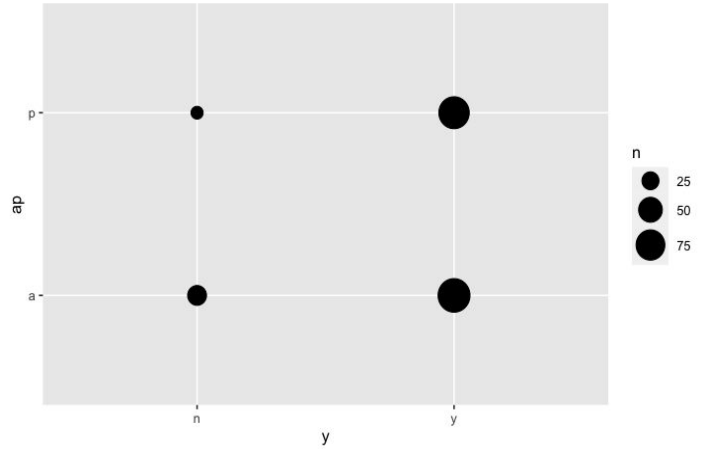
The proportion of Outcome n to y in placebo is: 0.09523 and 0.9047619
The proportion of Outcome n to y in drug and drug+ is: 0.1034483 and 0.89655

		y	n	y
week	trt			
0	placebo		2	19
	drug		2	12
	drug+		1	14
2	placebo		1	19
	drug		2	11
	drug+		1	10
4	placebo		4	14
	drug		5	7
	drug+		2	10
6	placebo		1	16
	drug		5	6
	drug+		5	7
11	placebo		4	16
	drug		4	8
	drug+		4	8

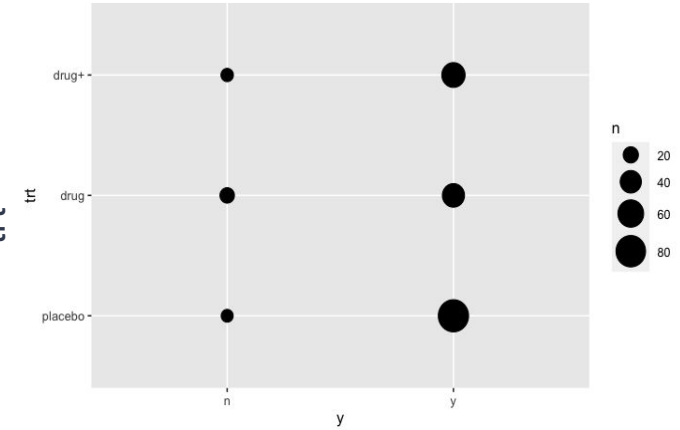
high\low



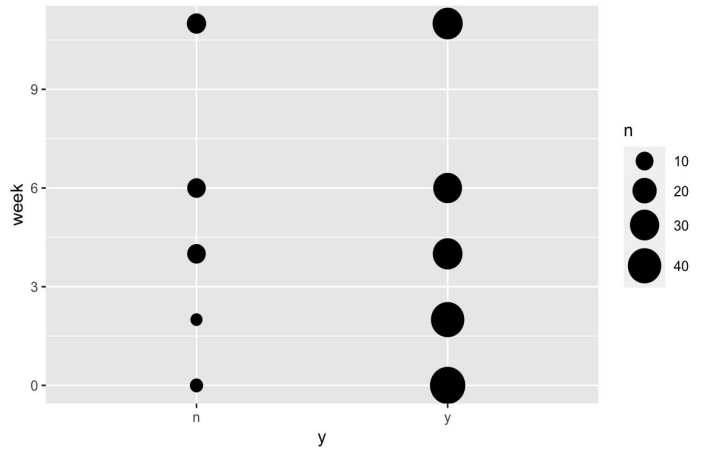
A/P



trt



week



Methods

Variable Selection: Odds Ratio and CI

Odds Ratio A/P

ap	y	n	y
a	31	93	
p	12	84	

OR: 2.33
OR CI: (1.125,4.835)
Significant

Odds Ratio HiLo

hilo	y	n	y
hi	19	103	
lo	24	74	

OR: 0.568
OR CI: (0.29,1.113)
Not Significant

Methods

Variable Selection: Chi-square
Independence Test for Predictors vs y

Indep Test for week

```
Pearson's Chi-squared test

data:  contingency.table
X-squared = 10.415, df = 4, p-value = 0.03399
```

P-value < 0.05, indicating that week is significant

Indep Test for treatment

```
Pearson's Chi-squared test

data:  contingency.table
X-squared = 6.6585, df = 2, p-value = 0.03582
```

P-value < 0.05, indicating that treatment is significant

GLMM: ap + hilo + factor(week)

Main Effects

mod <- glmer(y~ap+hilo+factor(week)+(1|ID), family = binomial, data=data)

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	2.6909	0.7011	3.838	0.000124	***
app	1.0167	0.5837	1.742	0.081547	.
hilolo	-0.5969	0.5493	-1.087	0.277168	
factor(week)2	0.1552	0.7613	0.204	0.838423	
factor(week)4	-1.4351	0.6664	-2.154	0.031263	*
factor(week)6	-1.5587	0.6754	-2.308	0.021008	*
factor(week)11	-1.5742	0.6631	-2.374	0.017592	*

Marked ones are significant. The OR of placebo against active > 1. The OR of low against high and week 4-11 against week 0 < 1 (descending when week increases). So, active, low compliance and time duration of treatment help reduce the probability of the presence of the bacteria.

Interactions

```
mod: y ~ ap + hilo + factor(week) + (1 | ID)
mod3: y ~ hilo + ap * factor(week) + (1 | ID)
     npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
mod     8 207.71 234.86 -95.854   191.71
mod3   12 210.78 251.51 -93.392   186.78 4.9254  4      0.295

mod: y ~ ap + hilo + factor(week) + (1 | ID)
mod2: y ~ ap + hilo * factor(week) + (1 | ID)
     npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
mod     8 207.71 234.86 -95.854   191.71
mod2   12 212.34 253.06 -94.170   188.34 3.3697  4      0.498

mod: y ~ ap + hilo + factor(week) + (1 | ID)
mod4: y ~ hilo * factor(week) + ap * factor(week) + (1 | ID)
     npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
mod     8 207.71 234.86 -95.854   191.71
mod4   16 214.99 269.29 -91.497   182.99 8.7151  8     0.3669
```

All p-values > 0.05, implying that the original model is true and that we don't need interaction terms

GLMM: ap + hilo + numeric week

Main Effects

mod5 <- glmer(y~ap+hilo+(week)+(1|ID), family = binomial, data=data)

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	2.36254	0.54722	4.317	1.58e-05	***
app	0.99936	0.55461	1.802	0.07156	.
hilolo	-0.56989	0.51992	-1.096	0.27303	
week	-0.14376	0.05096	-2.821	0.00479	**

Marked ones are significant. The OR of placebo against active > 1. The OR of low against high and week i against week i-1 < 1. So, active, low compliance and time duration of treatment help reduce the probability of the presence of the bacteria.

Interactions

```
mod5: y ~ ap + hilo + (week) + (1 | ID)
mod6: y ~ ap + hilo * (week) + (1 | ID)
     npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
mod5    5 207.23 224.20 -98.616   197.23
mod6    6 209.16 229.53 -98.582   197.16 0.0678  1     0.7946
mod5: y ~ ap + hilo + (week) + (1 | ID)
mod7: y ~ hilo + ap * (week) + (1 | ID)
     npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
mod5    5 207.23 224.20 -98.616   197.23
mod7    6 208.86 229.22 -98.430   196.86 0.3733  1     0.5412
mod5: y ~ ap + hilo + (week) + (1 | ID)
mod8: y ~ hilo * (week) + ap * (week) + (1 | ID)
     npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
mod5    5 207.23 224.20 -98.616   197.23
mod8    7 210.75 234.51 -98.376   196.75 0.4802  2     0.7866
```

All p-values > 0.05, implying that the original model is true and that we don't need interaction terms

GLMM: trt

Trt + factor(week)

modtrt5 <- glmer(y~trt+factor(week)+(1|ID), family = binomial, data=data)

```
Fixed effects:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)       3.4813     0.7717   4.511 6.45e-06 ***
trtdrug          -1.3703     0.6790  -2.018   0.0436 *
trtdrug+         -0.7825     0.6851  -1.142   0.2534
factor(week)2     0.1577     0.7623   0.207   0.8361
factor(week)4    -1.4431     0.6673  -2.163   0.0306 *
factor(week)6    -1.5680     0.6764  -2.318   0.0204 *
factor(week)11   -1.5729     0.6639  -2.369   0.0178 *
```

```
modtrt5: y ~ trt + factor(week) + (1 | ID)
modtrt6: y ~ trt * factor(week) + (1 | ID)
        npar    AIC    BIC  logLik deviance Chisq Df Pr(>Chisq)
modtrt5    8 208.15 235.30 -96.076   192.15
modtrt6   16 217.62 271.92 -92.811   185.62  6.53  8     0.5881
```

Trt + numeric week

modtrt3 <- glmer(y~trt+(week)+(1|ID), family = binomial, data=data)

```
Fixed effects:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  3.14392    0.62250   5.050 4.41e-07 ***
trtdrug     -1.32014    0.64240  -2.055  0.03988 *
trtdrug+    -0.79544    0.65198  -1.220  0.22245
week        -0.14369    0.05099  -2.818  0.00484 **
```

```
Models:
modtrt3: y ~ trt + (week) + (1 | ID)
modtrt4: y ~ trt * (week) + (1 | ID)
        npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
modtrt3    5 207.77 224.74 -98.885   197.77
modtrt4    7 211.03 234.78 -98.514   197.03 0.7422  2       0.69
```

Marked ones are significant. The week predictor acts identical in the previous models. OR of drugs against placebo < 1(OR of drug < OR of drug+). Low compliance active drug had better effect in eliminating bacteria than high compliance active drugs.

GLM vs GLMM: Random Effect

y~ap+hilo+week

```
Random effects:
 Groups Name        Variance Std.Dev.
 ID     (Intercept) 1.255    1.12
Number of obs: 220, groups:  ID, 50

glm(formula = y ~ ap + hilo + week, family = "binomial", data = data)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.3763   0.3813   0.5212   0.6576   1.1194

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   1.9278     0.3762   5.124 2.99e-07 ***
app           0.8343     0.3816   2.186  0.02879 *
hilolo       -0.5066     0.3546  -1.428  0.15317
week         -0.1167     0.0443  -2.633  0.00845 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

y~trt+week

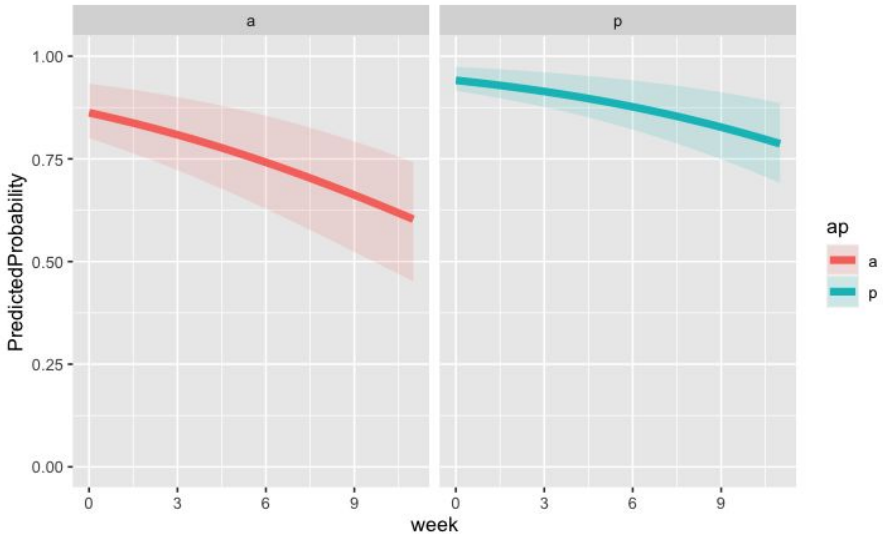
```
Random effects:
 Groups Name        Variance Std.Dev.
 ID     (Intercept) 1.557    1.248
Number of obs: 220, groups:  ID, 50

glm(formula = y ~ trt + week, family = "binomial", data = data)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.2899   0.3885   0.5400   0.7027   1.1077

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.54629    0.40555   6.279 3.42e-10 ***
trtdrug     -1.10667    0.42519  -2.603  0.00925 **
trtdrug+    -0.65166    0.44615  -1.461  0.14412
week        -0.11577    0.04414  -2.623  0.00872 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

The variance of the two models are significantly greater than 0, so we should consider the random effects to be significant. The OR interpretations of GLM’s are similar to which in the GLMM’s.

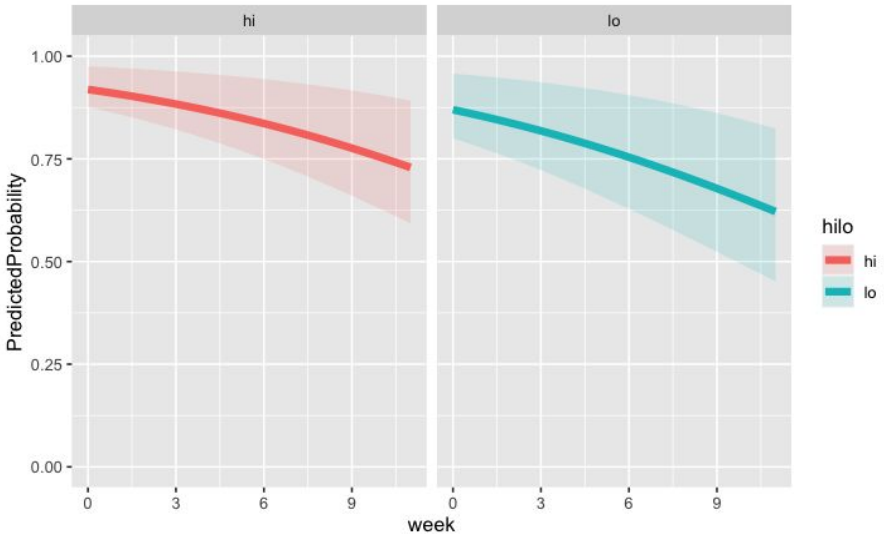
Results: y~ap+hilo

mod5 <- glmer(y~ap+hilo+(week)+(1|ID), family = binomial, data=data)

Predicted Probability vs Predictor (ap)

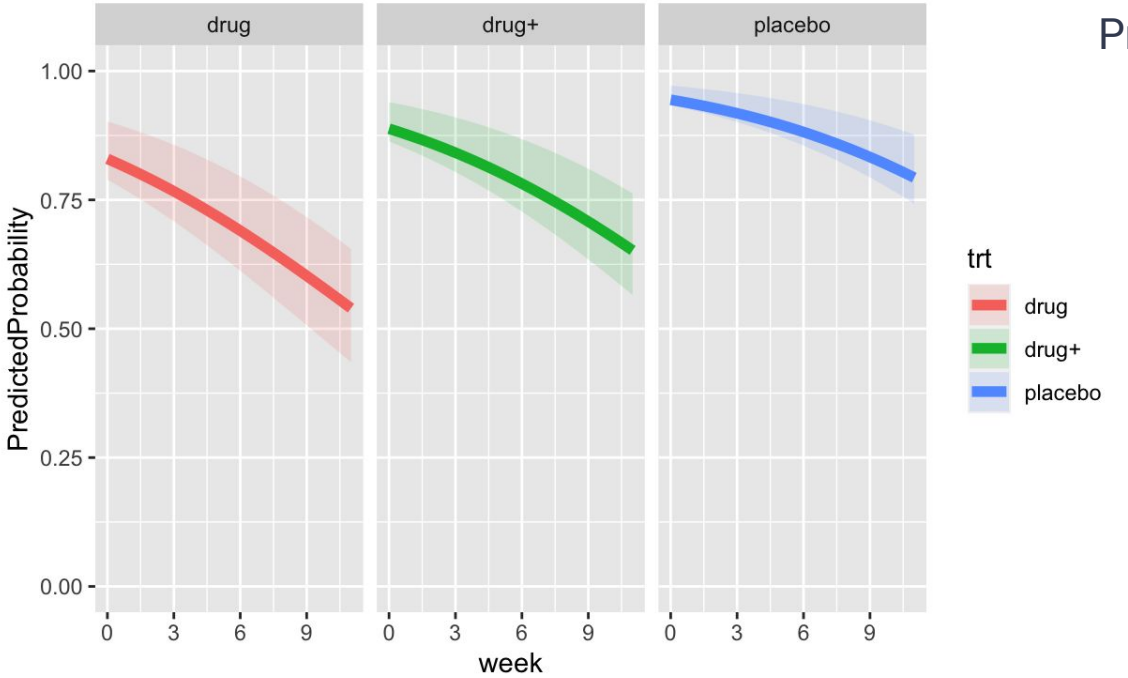


Predicted Probability vs Predictor (hilo)



Results: y~Trt

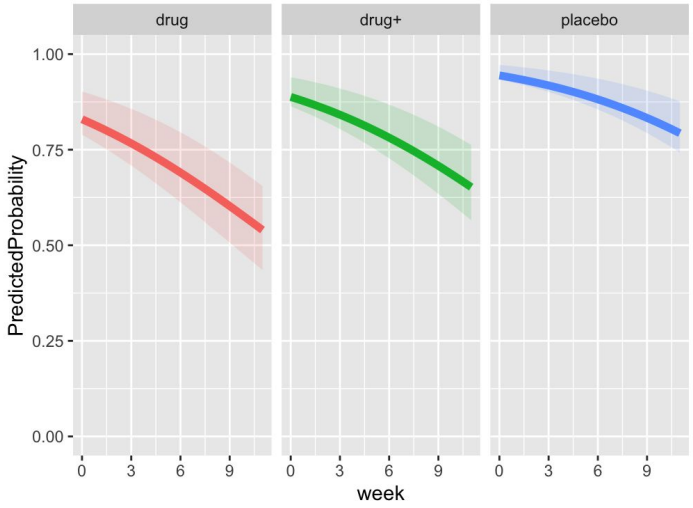
modtrt3 <- glmer(y~trt+(week)+(1|ID), family = binomial, data=data)



Predicted Probability vs Predictor (treatment)

This model shows similar result as y~ap+hilo had implied.

Summary



1. Is the treatment effective? We conclude that the treatment is effective for bacteria H. influenzae. Besides, we found out that the treatment is more effective for the group who take the drug than the placebo group. The likelihood of a positive test of the disease for placebo takers relative to non-placebo takers is 2.7165 times greater.

2. How do they behave over time? We conclude that the week is significant. The model shows that the likelihood of positive test of the disease is 0.8660956 times greater than the previous week. Generally, as the weeks increase, the likelihood to test positive decreases.

3. Any compliance effect? The compliance effect is not significant; however the model shows that the likelihood of a positive test of disease for low compliance takers relative to high-compliance is 0.565587 times greater, which means they are nearly half as likely to test positive.

Suggestions for further research: To improve the model, we should include possible confounding variables including other diseases that may affect results. For example, the predominant bacteria that cause otitis media are Streptococcus pneumonia and Moraxella catarrhalis could be integrated into the data.

Appendix

See attached pdf with r-code