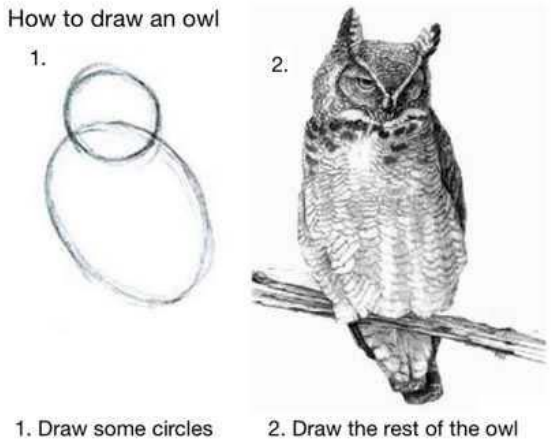


Modelling Review

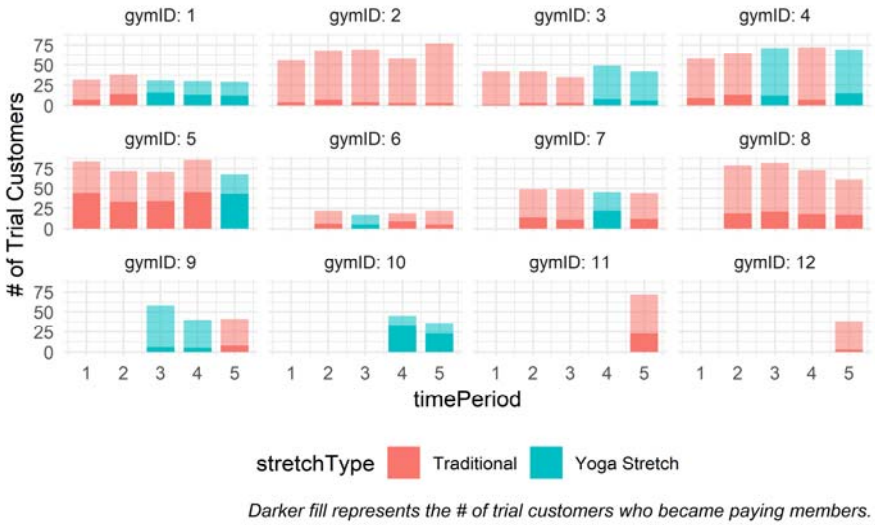
Preparing for Multi-level Models

BEACHBODY® YOGASTUDIO

Is it worthwhile to pay Beachbody to use its name?

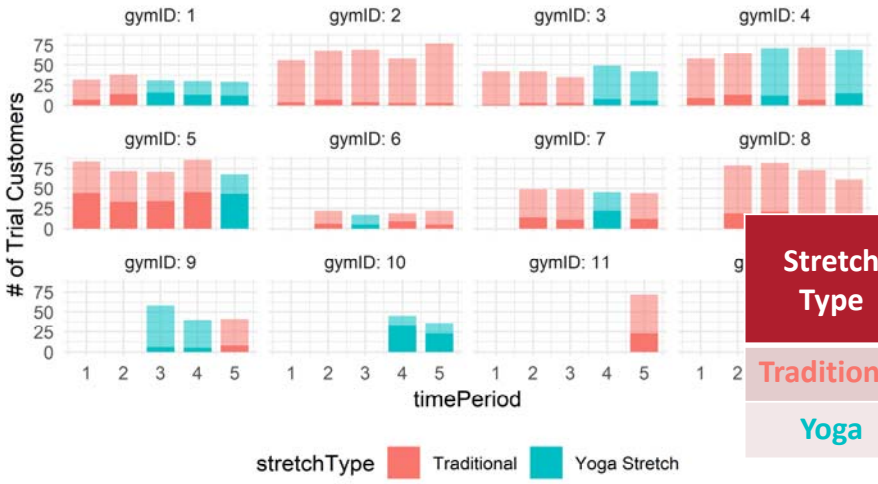
The Data

```
> gymDF
# A tibble: 44 x 5
   gymID timePeriod nTrialCustomers nSigned yogaStretch
   <int>      <dbl>           <int>   <int>       <dbl>
 1     1          1              32       7           0
 2     2          1              56       4           0
 3     3          1              42       1           0
 4     4          1              58       9           0
 5     5          1              84      44           0
 6     1          2              38      14           0
 7     2          2              68       7           0
 8     3          2              42       3           0
 9     4          2              64      13           0
10     5          2              72      33           0
# ... with 34 more rows
```

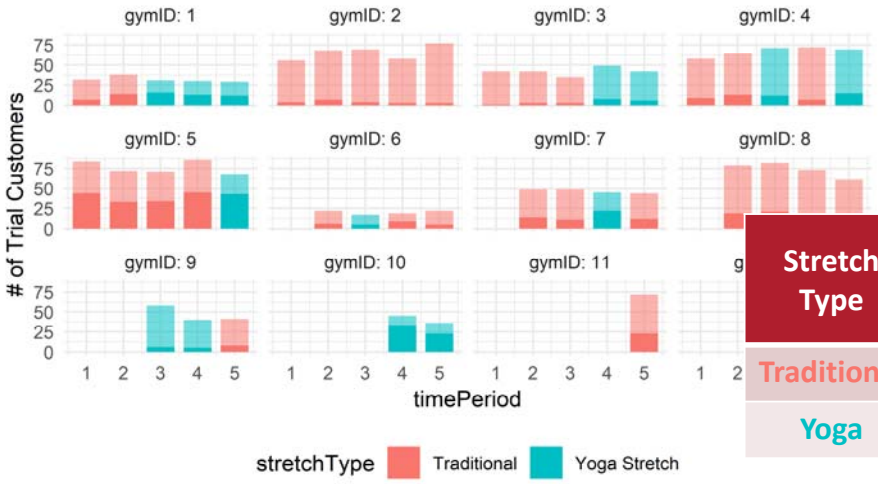
Darker fill represents the # of trial customers who became paying members.

QUESTION: Can you tell whether Yoga stretching helps?



Stretch Type	# of Free Trials	Percent Signed
Traditional	1675	**24%**
Yoga	630	**35%**

Darker fill represents the # of trial customers who became paying members.



QUESTION: Why might the below #'s be misleading?

Stretch Type	# of Free Trials	Percent Signed
Traditional	1675	**24%**
Yoga	630	**35%**

Darker fill represents the # of trial customers who became paying members.

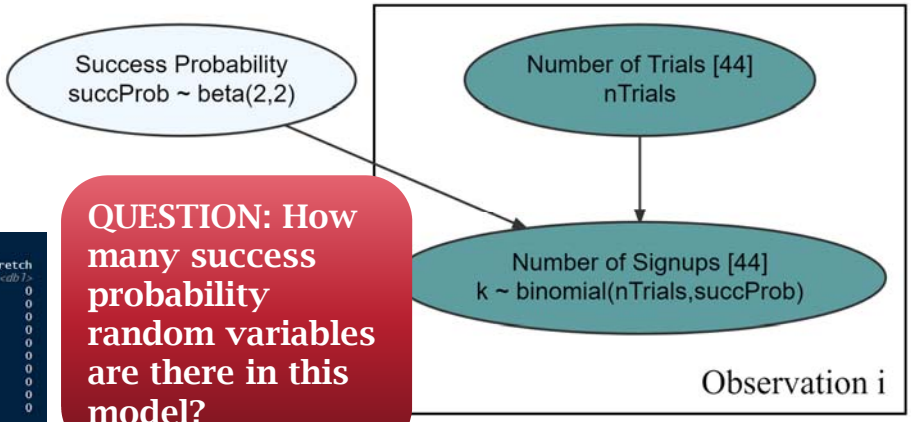
The Data

```
> gymDF
# A tibble: 44 x 5
   gymID timePeriod nTrialCustomers nSigned yogaStretch
   <int>      <dbl>           <int>   <int>       <dbl>
 1     1          1              32       7           0
 2     2          1              56       4           0
 3     3          1              42       1           0
 4     4          1              58       9           0
 5     5          1              84      44           0
 6     1          2              38      14           0
 7     2          2              68       7           0
 8     3          2              42       3           0
 9     4          2              64      13           0
10     5          2              72      33           0
# ... with 34 more rows
```


Really Complete Pooling Model

QUESTION: Why is there a 44 in brackets next to the observed node descriptions?

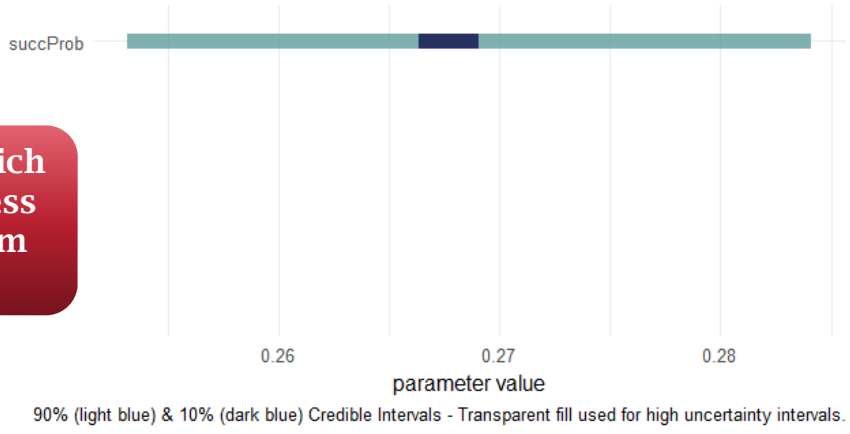
Really Complete Pooling Model



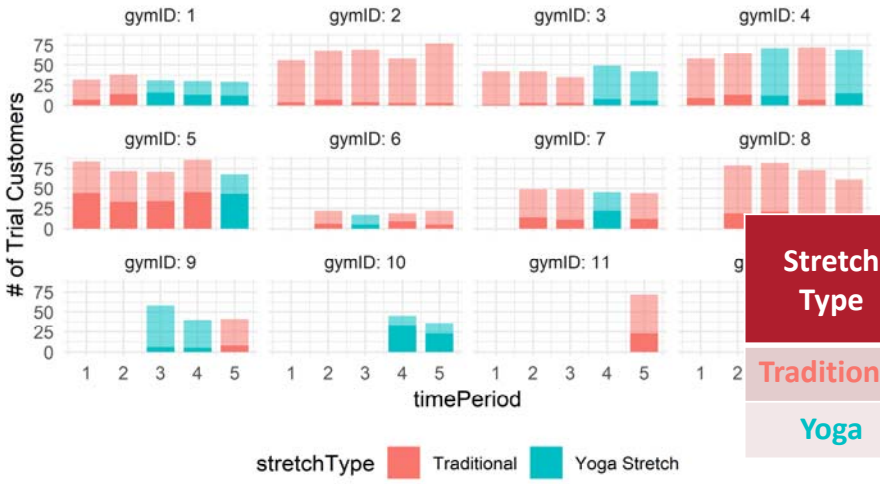

```
> gymDF
# A tibble: 44 x 5
   gymID timePeriod nTrialCustomers nSigned yogaStretch
   <int>      <dbl>           <int>   <int>        <dbl>
 1     1          1              32       7            0
 2     2          1              56       4            0
 3     3          1              42       1            0
 4     4          1              58       9            0
 5     5          1              84      44            0
 6     1          2              38      14            0
 7     2          2              68       7            0
 8     3          2              42       3            0
 9     4          2              64      13            0
10     5          2              72      33            0
# ... with 34 more rows
```

QUESTION: How many success probability random variables are there in this model?

The Posterior Distribution for the Really Complete Pooling Model



QUESTION: Which values of success probability seem plausible?



Darker fill represents the # of trial customers who became paying members.

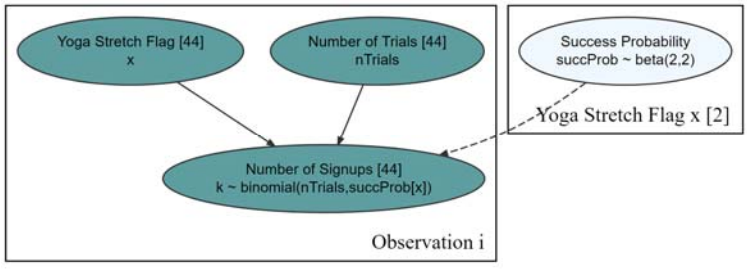
Really Complete Pooling ???

Stretch Type	# of Free Trials	Percent Signed
Traditional	1675	**24%**
Yoga	630	**35%**

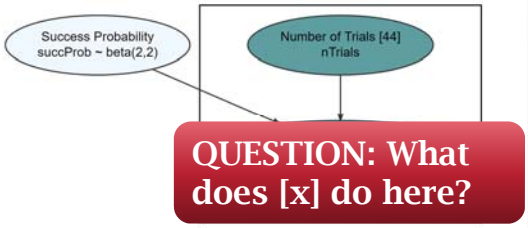
Really Complete Pooling Model

Almost Complete Pooling Model

QUESTION: How many success probability random variables are there in the almost complete model?

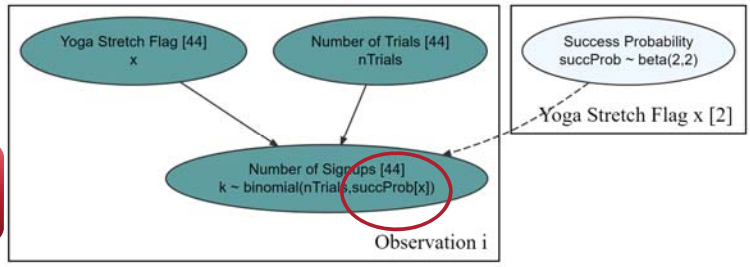


Really Complete Pooling Model

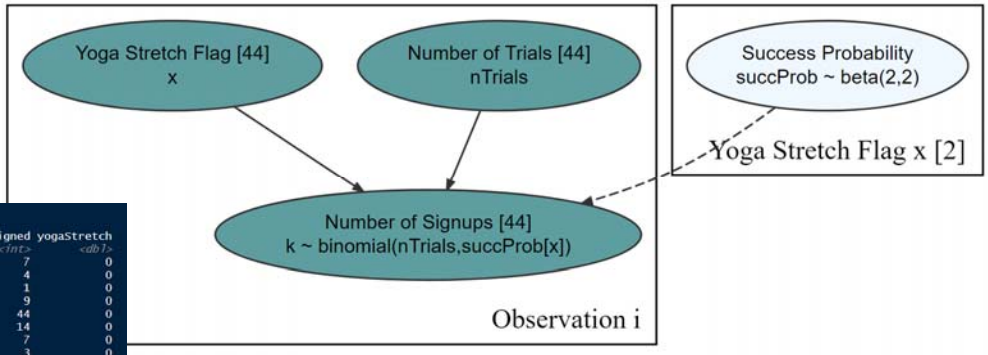


QUESTION: What does [x] do here?

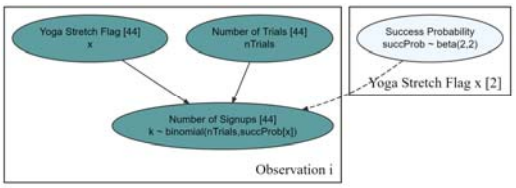
Almost Complete Pooling Model

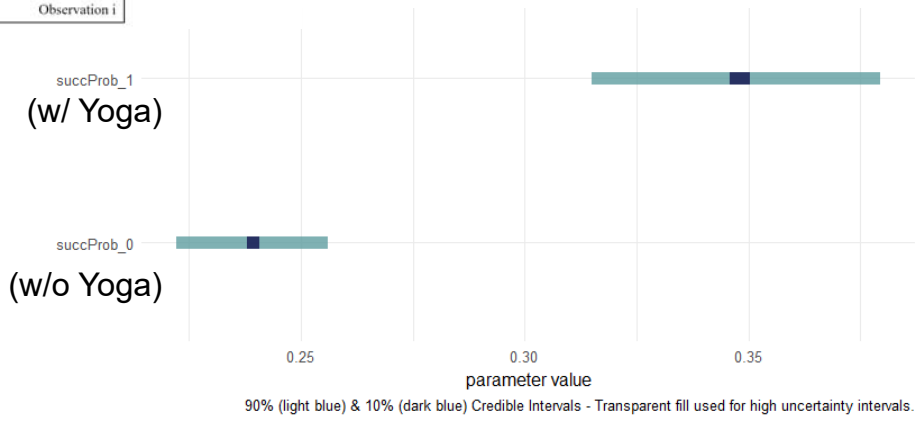


Almost Complete Pooling Model – Two Types of Rows

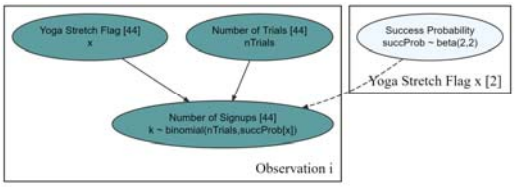



```
> gymDF
# A tibble: 44 x 5
   gymID timePeriod nTrialCustomers nSigned yogaStretch
   <int>      <dbl>           <int>   <int>       <dbl>
 1     1          1              32       7           0
 2     2          1              56       4           0
 3     3          1              42       1           0
 4     4          1              58       9           0
 5     5          1              84      44           0
 6     1          2              38      14           0
 7     2          2              68       7           0
 8     3          2              42       3           0
 9     4          2              64      13           0
10     5          2              72      33           0
# ... with 34 more rows
```

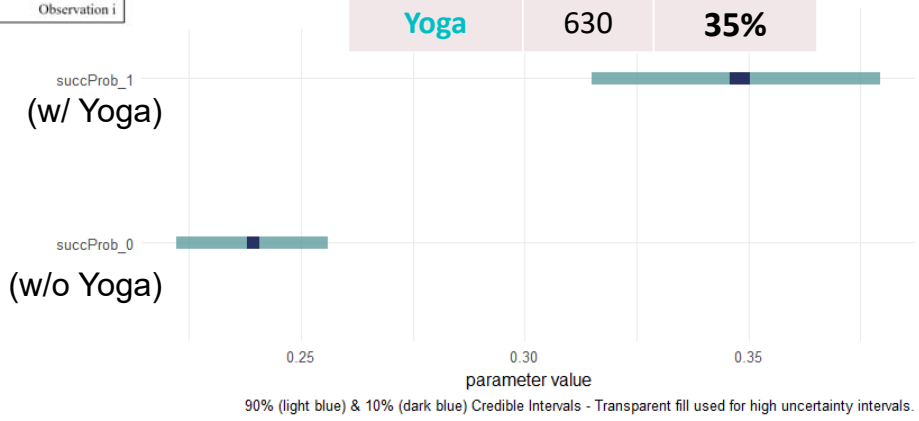




90% (light blue) & 10% (dark blue) Credible Intervals - Transparent fill used for high uncertainty intervals.

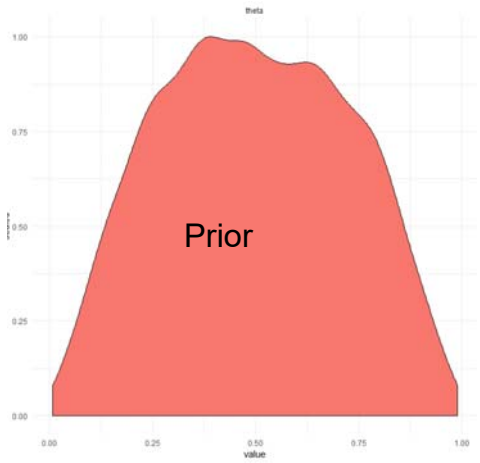


Stretch Type	# of Free Trials	Percent Signed
Traditional	1675	**24%**
Yoga	630	**35%**



90% (light blue) & 10% (dark blue) Credible Intervals - Transparent fill used for high uncertainty intervals.

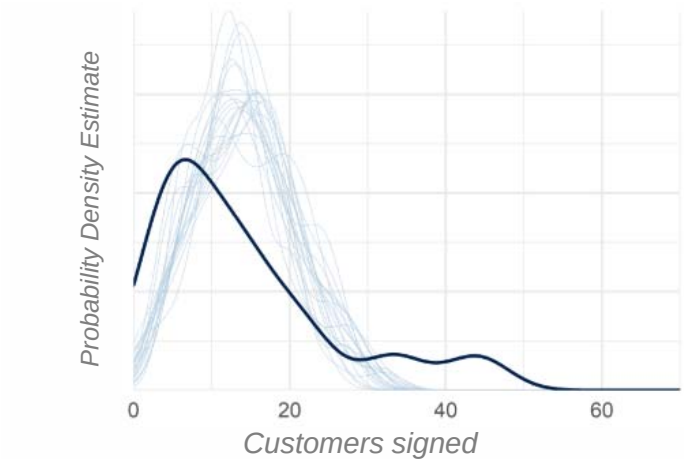
Prior versus Posterior for Success Probability



Posterior

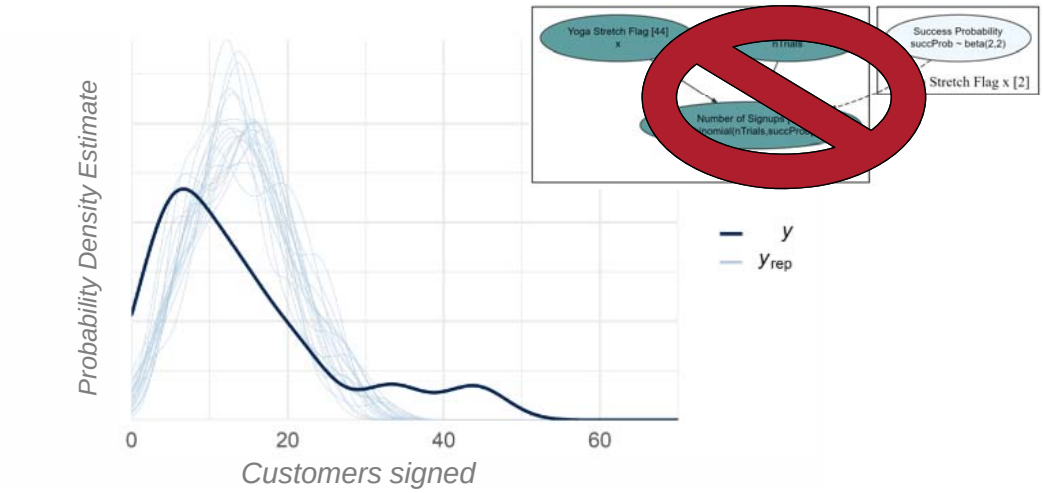
```
> drawsDF
# A tibble: 4,000 x 2
   succProb_0 succProb_1
        <dbl>      <dbl>
 1      0.236      0.369
 2      0.236      0.369
 3      0.236      0.369
 4      0.236      0.369
 5      0.236      0.369
 6      0.248      0.344
 7      0.236      0.349
 8      0.233      0.335
 9      0.254      0.363
10      0.255      0.319
# ... with 3,990 more rows
```


Posterior Predictive Check



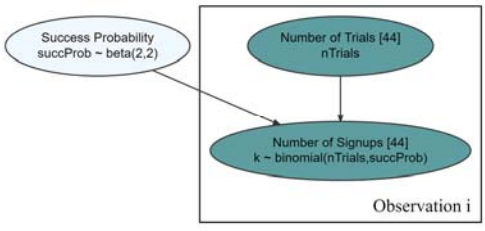
QUESTION: Does this model under-estimate or overestimate the number of observations that sign-up 20 customers in a month?

Posterior Predictive Check

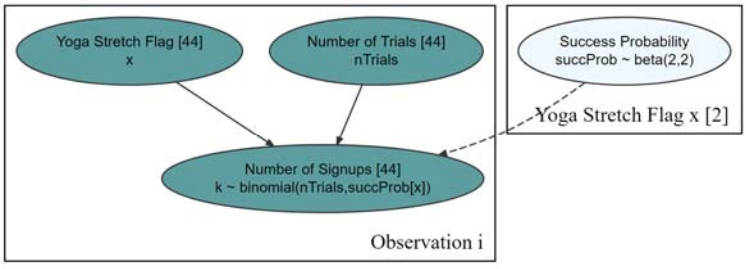


Recap So Far

Really Complete Pooling Model



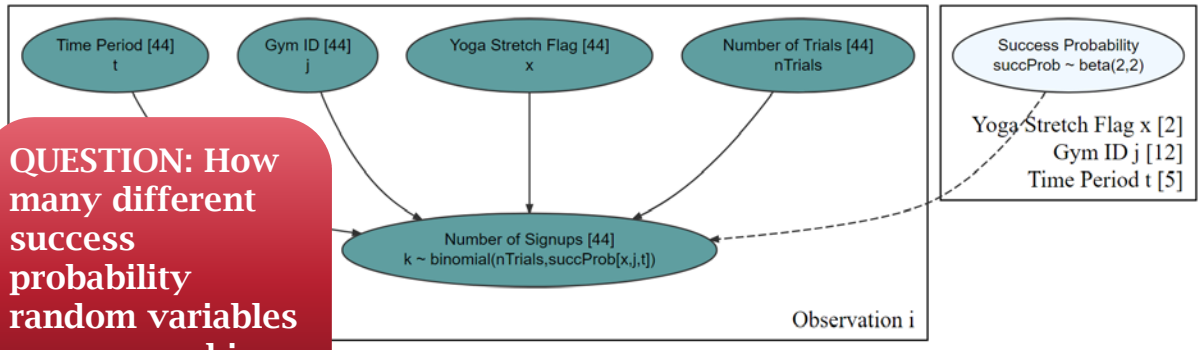
Almost Complete Pooling Model



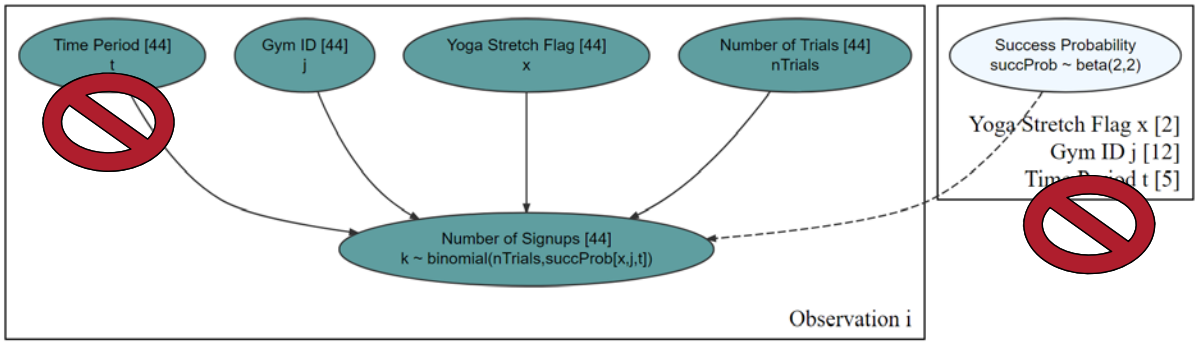
Zero Pooling (we will not run this model)

```
> gymDF
# A tibble: 44 x 5
   gymID timePeriod nTrialCustomers nSigned yogaStretch
   <int>      <dbl>           <int>   <int>       <dbl>
 1     1          1              32       7           0
 2     2          1              56       4           0
 3     3          1              42       1           0
 4     4          1              58       9           0
 5     5          1              84      44           0
 6     1          2              38      14           0
 7     2          2              68       7           0
 8     3          2              42       3           0
 9     4          2              64      13           0
10     5          2              72      33           0
# ... with 34 more rows
```

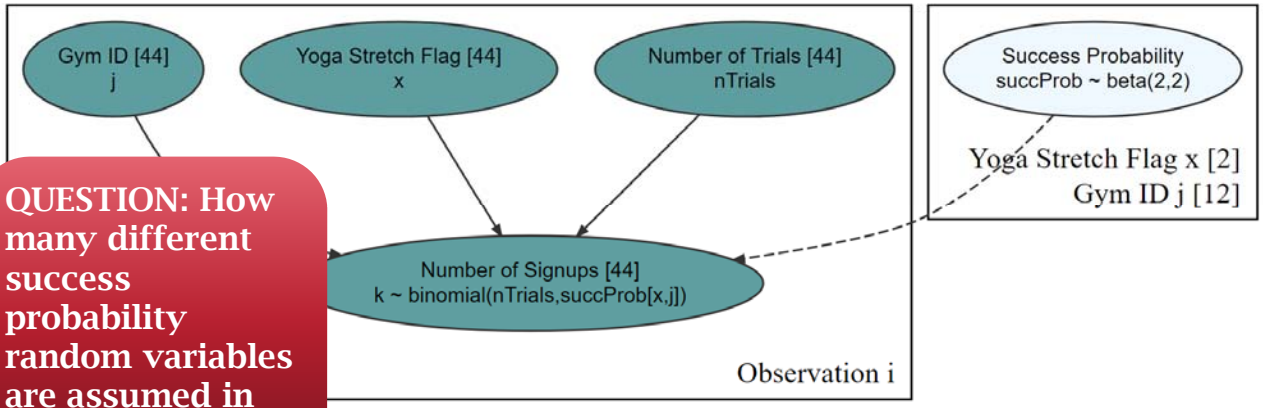

Zero Pooling (we will not run this model)



QUESTION: How many different success probability random variables are assumed in this model?



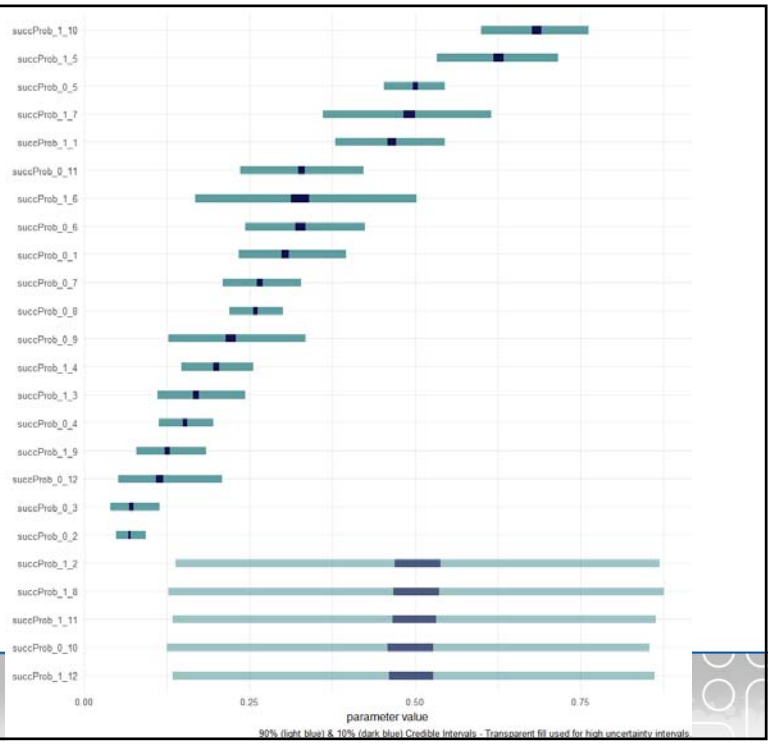
Almost Zero Pooling

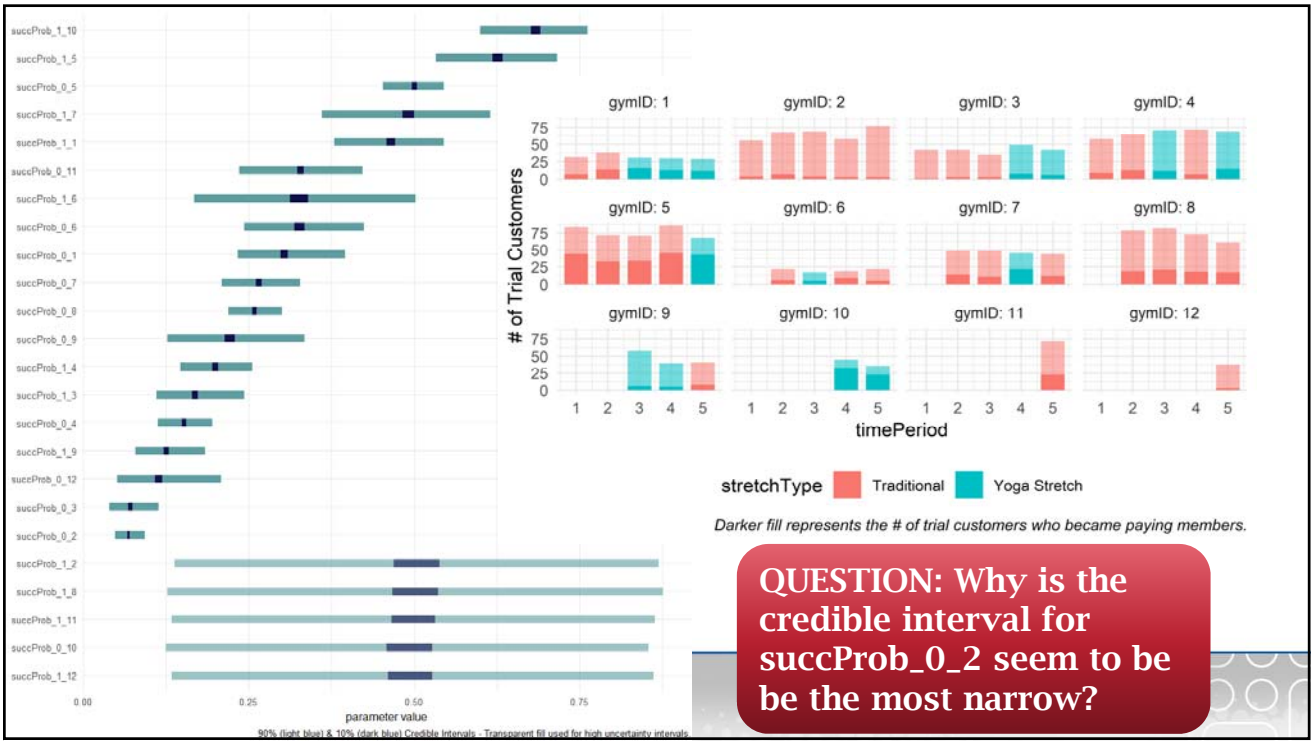


QUESTION: How many different success probability random variables are assumed in this model?

Posterior Distribution

QUESTION: What do the numbers mean on the y-axis (e.g. succProb_1_10)?





QUESTION: Why is the credible interval for succProb_0_2 seem to be be the most narrow?