Prototype simtest with a motivation effect

Jacqueline Buros

9/6/2019

The purpose of this document is to review the quality of model fits to simulated data. This document builds on an earlier doc by including a "motivation" effect on the outcome (completion rate) in addition to the effects of study duration & notification level.

Review simulation parameters

First we should review the parameters used to simulate our data. In this round of simulations, a fixed set of parameter values are used to simulate data for all of our draws. We have parameters for the completion rate according to duration consistent with 30%, 20%, and 10% response in the 5-day, 15-day and 27-day groups respectively where notification level is low.

Increasing the notification level to "moderate" confers a modest increase in completion rate uniformly across these three groups. This is reflected by a fixed OR of 1.1 (roughly a 10% increase).

In addition, we are assuming that the motivation level will be set into two groups (mostly for simplicity of the simulation) high vs low, and that roughly 20% of the participants will be in the "high" motivation group. Motivation increases completion rate by an OR of 1.2.

Given these parameter values, these are the "target" completion rates for each of our groups used to simulate our data.

duration_group

notify_moderate

motivation_high: 0

motivation_high: 1

5_days

0

30.0%

34.0%

5_days

1

32.0%

36.1%

15_days

0

20.0%

23.1%

15_days

1

21.6%

24.8%

27_days

0

10.0%

11.8%

27_days

1

10.9%

12.8%

Of course, each draw of simulated data has a different *observed* completion rate. There are two sources of variation contributing to the observed completion rate in each draw. One is the composition of the particular study population – both the notification rate and the assignment to duration groups is randomized and so one would expect *some* imbalance in these assignments in practice.

Looking at the group sizes over our draws can give a sense of how this would be expected to vary in practice for a sample of this size (640 participants).

duration_group

notify_moderate

motivation_high

1

2

3

4

5

5_days

0

0

49

51

50

50

52

5_days

0

1

14

17

10

13

15

5_days

1

0

54

53

52

40

62

5_days

1

1

16

12

12

10

14

15_days

0

0

160

164

149

163

143

15_days

0

1

44

34

39

41

39

15_days

1

0

167

132

173

149

155

15__days

1

1

35

39

29

34

42

27__days

0

0

41

60

59

68

48

27__days

0

1

11

14

14

6

10

27__days
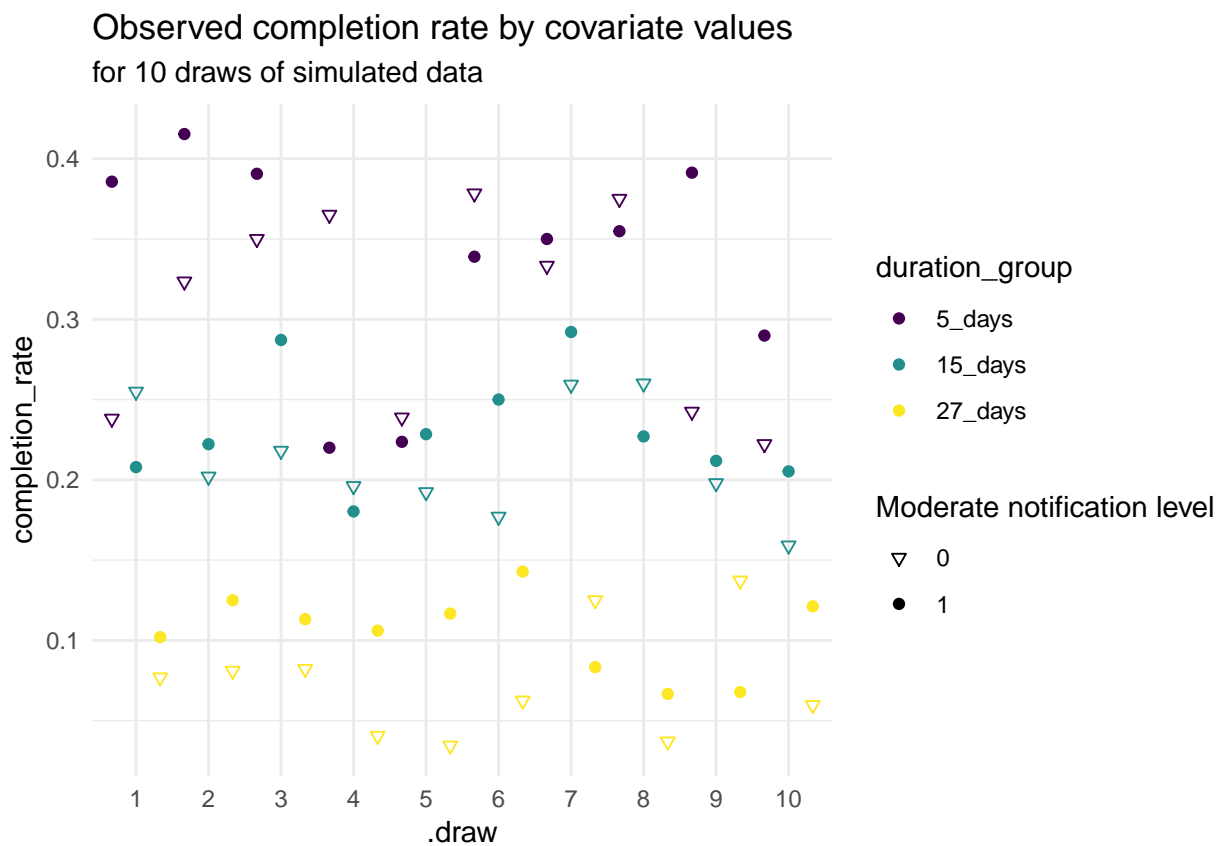
1

0

43

51

41

51

52

27__days

1

1

6

13

12

15

8

This is very likely a modest influence on the observed rates, but it is nonetheless there.

The second source comes from the simulation of the outcome itself.

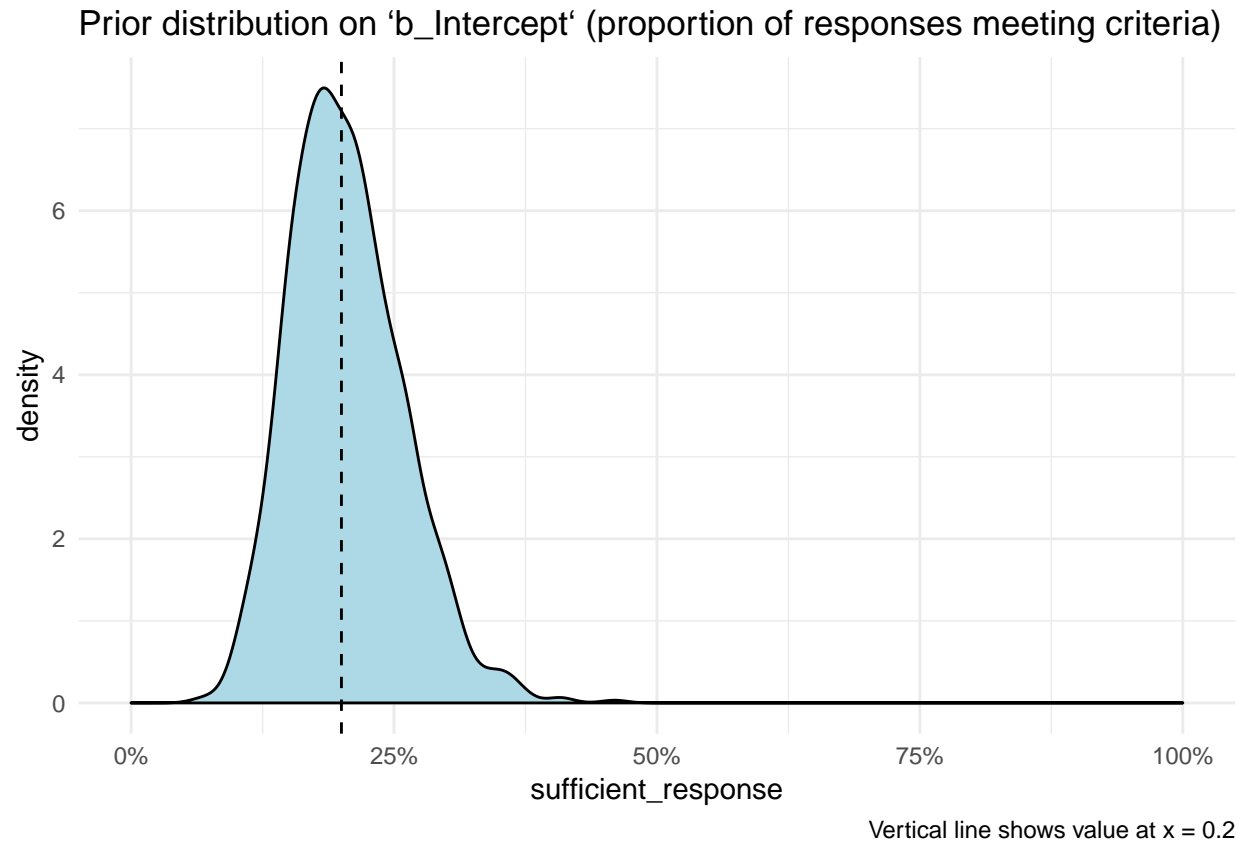Here is a sample of the observed completion rates for each of our 6 covariate combinations.

Observed completion rate by covariate values
for 10 draws of simulated data

Review priors for the model fit

Prior on `b_Intercept`

The `Intercept` term defines the proportion of the population in the 15-day study length meeting the criteria for a "sufficient" response.

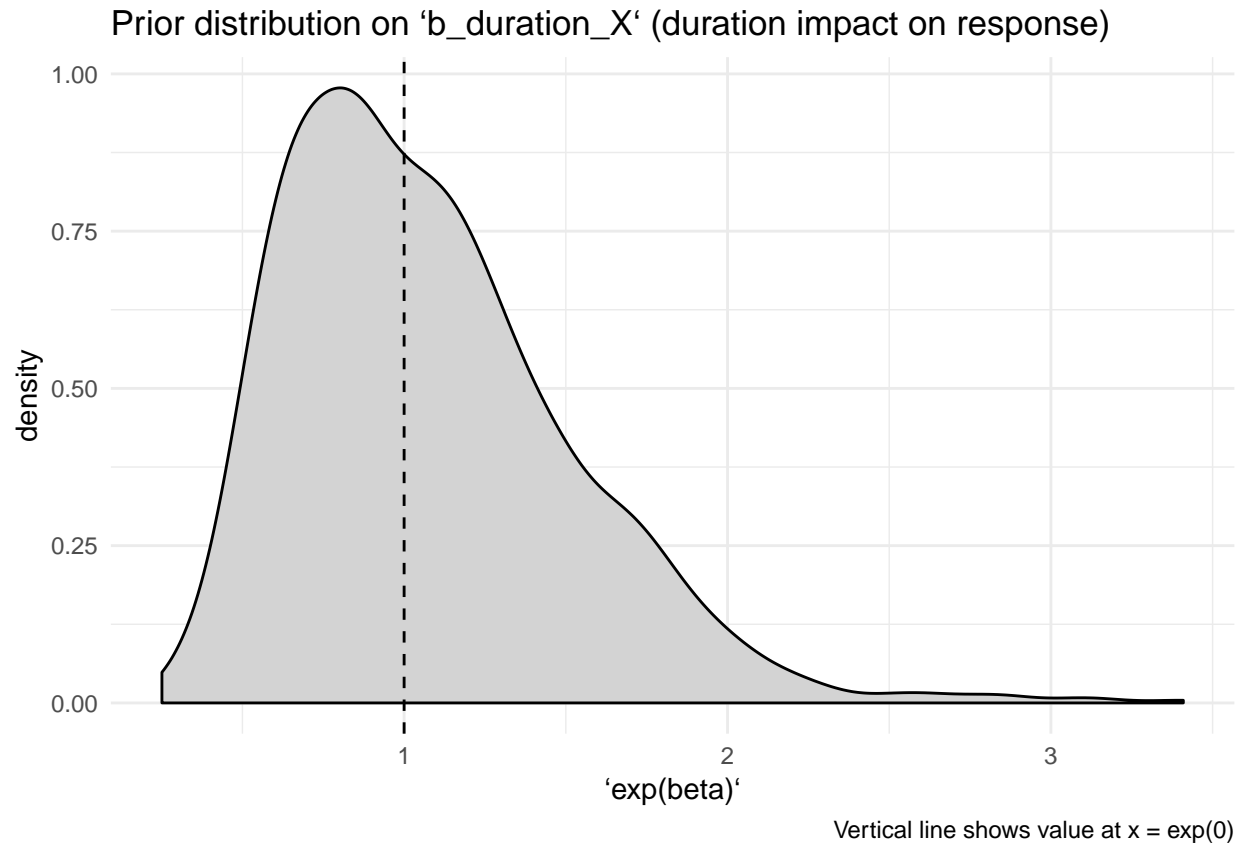I have put a fairly narrow prior centered at the 20% mark:

Prior distribution on 'b_Intercept' (proportion of responses meeting criteria)

Vertical line shows value at x = 0.2

Prior on `b_duration_X`

We originally started with what is a typical weakly informative prior (`normal(0, 1)`) on the `beta` effects describing the offset from the response at 15 days for the other two study durations of 5 & 27 days. However this led to much larger effect sizes than I would have expected (ie sufficient-response rates close to 80% in one group).

In this set of responses the prior on these betas is more narrow: normal(0, 0.4).

Let's see how that translates into a more familiar metric such as an OR.

Prior distribution on 'b_duration_X' (duration impact on response)

Vertical line shows value at x = exp(0)

This suggests even this distribution is too large for our betas – I say this because there is significant density at a two-fold difference in response depending on the study duration (however – maybe this is right?).

Prior on `b_notify_high`

Finally we will review the priors on our notification level.

Here we start with a more narrow prior than for the duration since I expect this will be a weaker effect.

Our prior is: .

Let's see how that translates into a more familiar metric such as an OR.

Prior distribution on 'b_notify_high' (notification impact on response)

Vertical line shows value at x = exp(0)
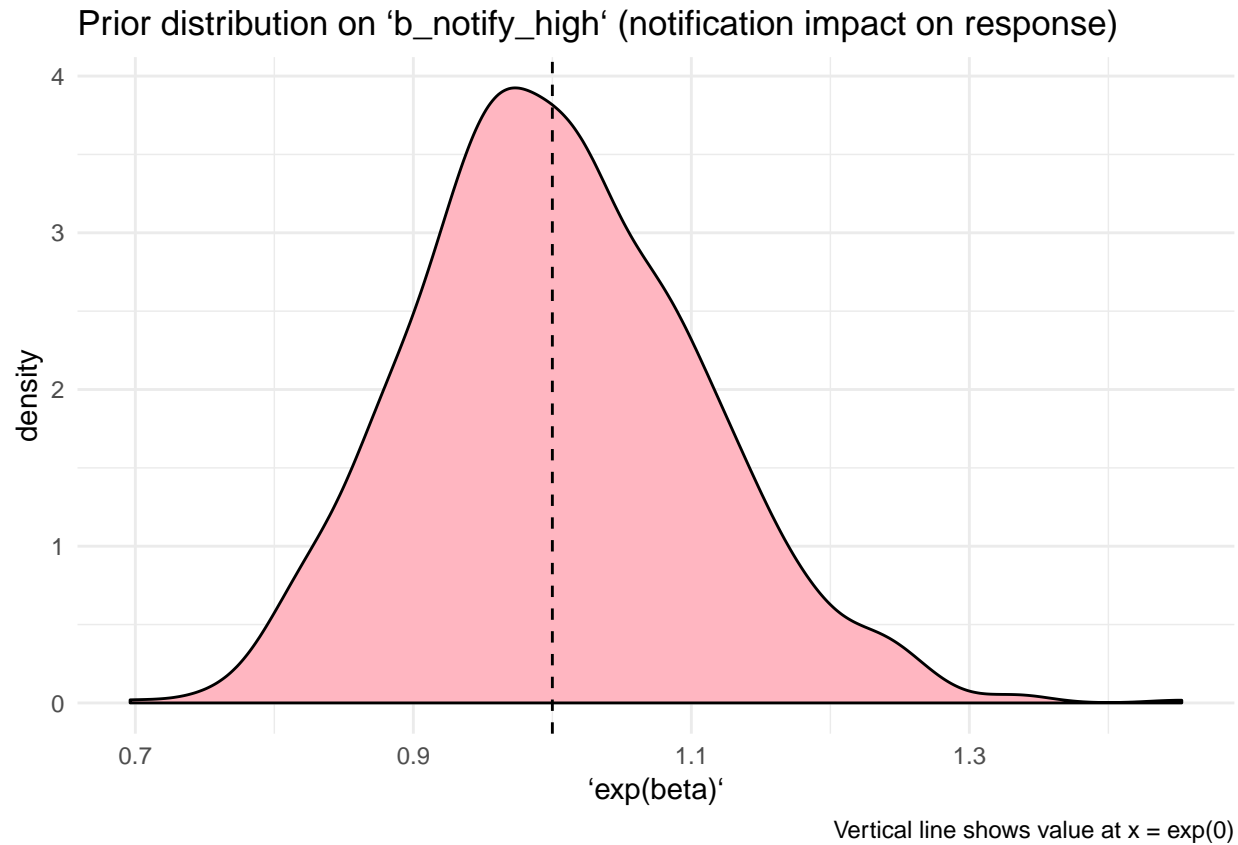
This is still centered at 0 (OR = 1) but with a more narrow distribution that we had used for the duration effect. Keep in mind that these are both starting points; we can improve upon them in future versions.

Posterior fits

We can now review the posterior fits for each of these scenarios. First, let's do a "typical" or generic summary of the posterior vs the parameter values.

Showing 50% and 90% intervals
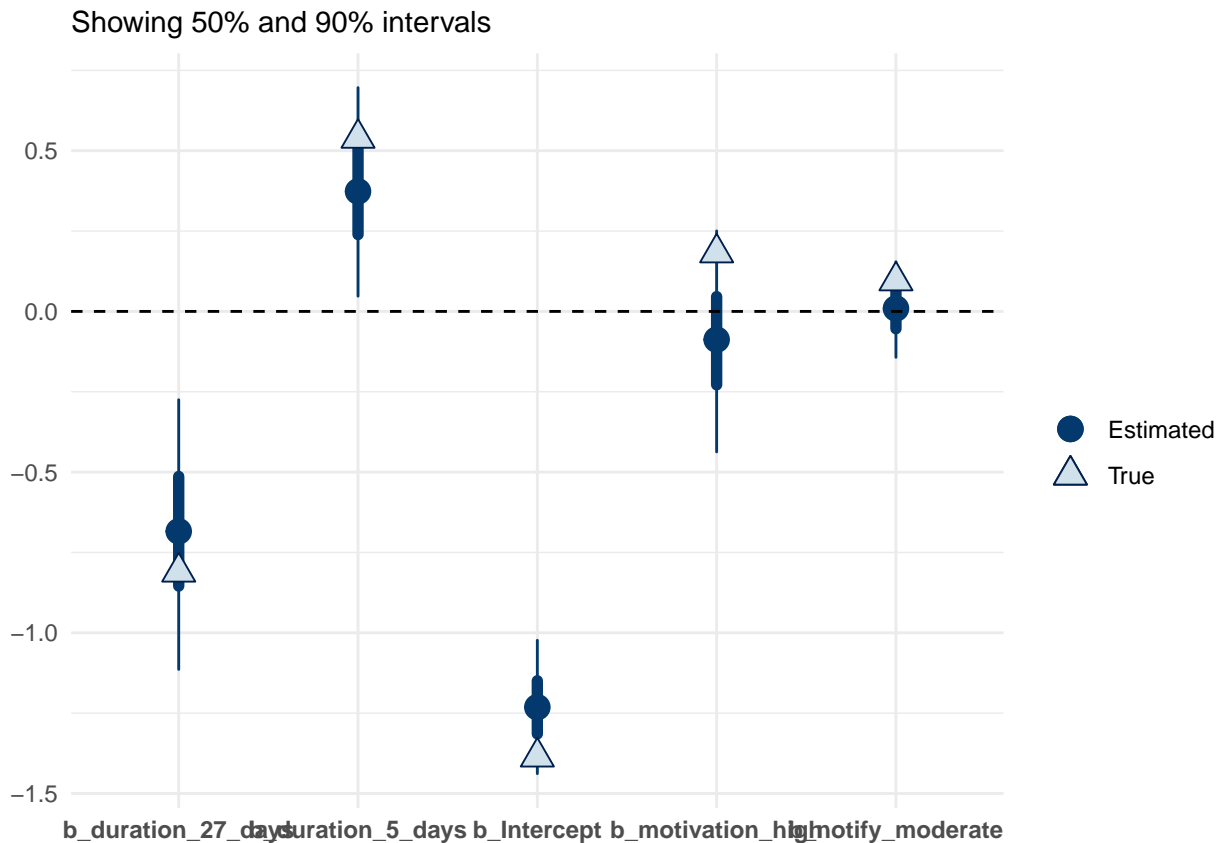
Next, we will write a function to summarise the metrics we "really" (supposedly) care about in our particular use case.

I would say these are (for each parameter):

1. Width of posterior 90% CI (credible interval)
2. Does the true value fall within the 90% posterior credible interval (check calibration)
3. For regression parameters, estimate the Type S (sign) error rate [if CI excludes 0]

   - IE does this interval contain 0?
     - if so, consider this analysis as "inconclusive"
     - if *not*, call the beta as being $<$ or $>$ 0
   - compare this determination to the direction of the "true effect"

4. For regression parameters, estimate the Type M (magnitude) error rate [if CI excludes 0]

   - IE does this interval contain 0?
     - if so, consider this analysis as "inconclusive"
     - if *not*, use the median value of the parameter as the "estimated effect"
   - compare this determination to the value of the "true effect". Is it $>$2 times as large?

Summary of model fits

Summarizing these metrics for our 100 fits at this sample size (n = 640):

parname

width

ci_width

ci_contains_true_value

ci_contains_0

ci_type_s_error

ci_type_m_error

b_duration_27_days

0.95

0.9189272

90.0%

17.0%

0.0%

0.0%

b_duration_5_days

0.95

0.7849221

94.0%

37.0%

0.0%

0.0%

b_Intercept

0.95

0.5035278

94.0%

0.0%

0.0%

0.0%

b_motivation_high

0.95

0.8125166

95.0%

96.0%

25.0%

75.0%

b_notify_moderate

0.95

0.3482306

100.0%

100.0%

NaN%

NaN%

The width of the credible intervals shown above, however, is on the log(OR) scale. It's not very easy to tell what that might look like in practice.

Let's review what the posterior density for the OR looks like in order to better appreciate this.

Posterior Densities of Regression Parameters for n = 640