Bayesian analysis of the NESTA study of interventions against verbal aggression online

Rafal Urbaniak

Contents

1 Exploration

Load the dataset and take a look first.

```
summaries <- read.csv(file = "datasets/Summaries.csv")
head(summaries) %>% kable( "latex", booktabs = T) %>%
  kable_styling(latex_options = c("striped", "scale_down") ,font_size = 9)
```

The basic variables we are dealing with are in the following table.

Further variables are defined in terms of those, in particular, we will be predicting AdiffS which is the standardized difference AA-AB, and AdiffS, which is the standardized difference CA-CB. Before we proceed, we will also standardize the predictors, and add a numerical index for the group:

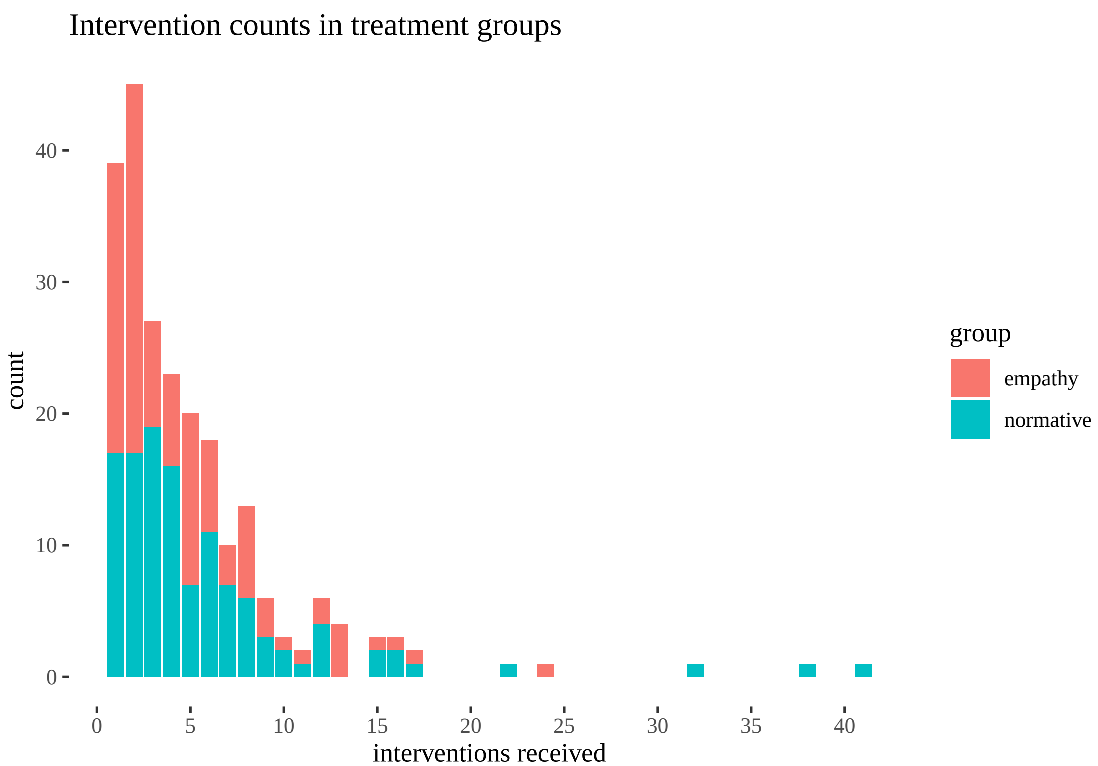
```
summaries$ABS <- standardize(summaries$AB)
summaries$CBS <- standardize(summaries$CB)
summaries$AAS <- standardize(summaries$AA)
summaries$CAS <- standardize(summaries$CA)
summaries$CDS <- standardize(summaries$CD)
summaries$ADS <- standardize(summaries$AD)
summaries$group <- as.factor(summaries$group)
summaries$groupID <-  as.integer( as.factor(summaries$group) )
```

First, let's take a look at the distribution of IC in the treatment groups:

```
ggplot(summaries[summaries$group != "control",], aes(x = IC, fill = group))+
  geom_bar()+theme_tufte()+
  xlab("interventions received")+
  labs(title = "Intervention counts in treatment groups")+
  scale_x_continuous(breaks = seq(0,40,5))
```

X	author	AB	AD	AA	CB	CD	CA	Adiff	Cdiff	AdiffS	CdiffS	group	IC
1	_swf	19	1	0	720	25	28	-19	-692	-0.0245122	-0.3501491	normative	1
2	-Allergic	24	24	8	1614	1451	1237	-16	-377	0.0719197	0.1057675	normative	3
3	-funny-username-	23	6	12	847	497	721	-11	-126	0.2326395	0.4690535	control	0
4	-Johnny-	18	2	8	1465	408	684	-10	-781	0.2647835	-0.4789637	empathy	2
5	1secwhileiyeet3	15	3	4	1384	198	120	-11	-1264	0.2326395	-1.1780359	control	0
6	20CharsIsNotEnough	16	10	25	779	907	972	9	193	0.8755188	0.9307596	empathy	4

variable	explanation
AB	attacks before (pre-treatment)
AD	attacks during (the treatment period)
AA	attacks after (post-treatment)
CB	comments before
CD	comments during
CA	comments after
group	treatment group
IC	intervention count



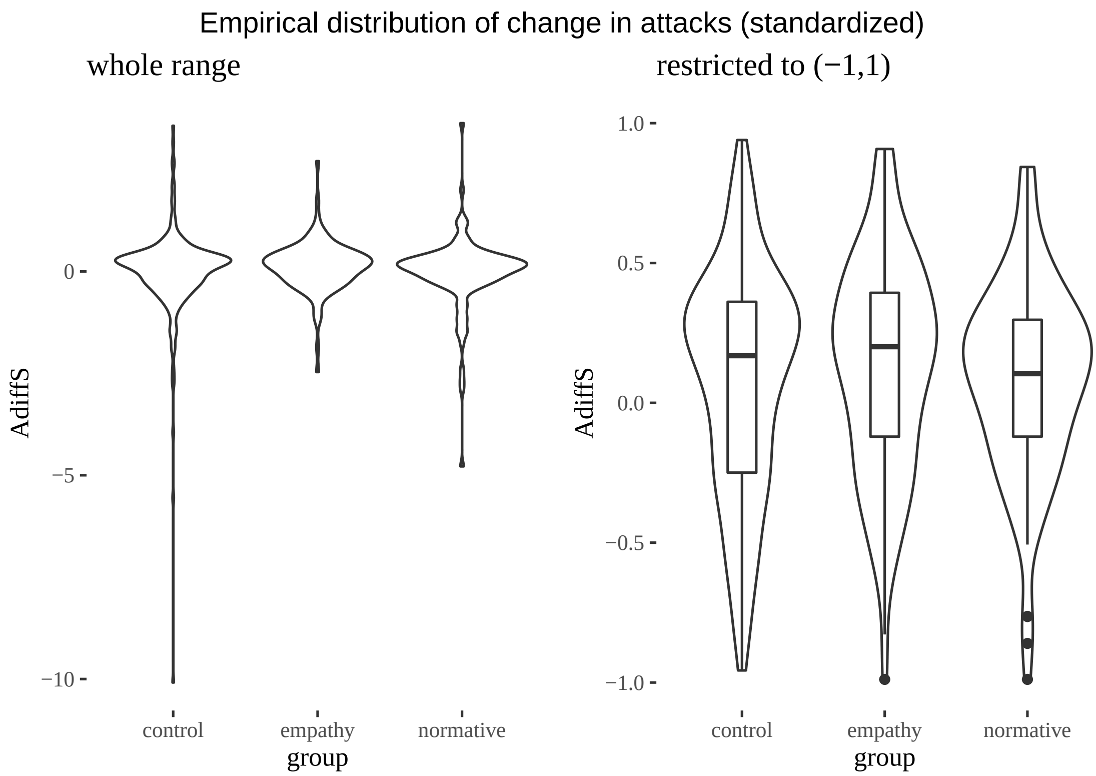
Second, when we look at the distribution of standardized difference in attacks, when restricted to (-1,1), the peaks of distributions are shifted a bit, with lowest median for the normative group, but not too much:

Note there were much more empathetic interventions, this needs an explanation

```
violAdiffS <- ggplot(summaries, aes(x=group, y = AdiffS))+
  geom_violin() +theme_tufte()
violJoint <- ggarrange(violAdiffS+ggtitle("whole range"),
              violAdiffS + ylim(c(-1,1))+geom_boxplot(width = .2)+
              ggtitle("restricted to (-1,1)"))
```

```
## Warning: Removed 58 rows containing non-finite values (stat_ydensity).
## Warning: Removed 58 rows containing non-finite values (stat_boxplot).
```

```
violJointTitled <- annotate_figure(violJoint,
  top = text_grob("Empirical distribution of change in attacks (standardized)",
                size = 12))
violJointTitled
```

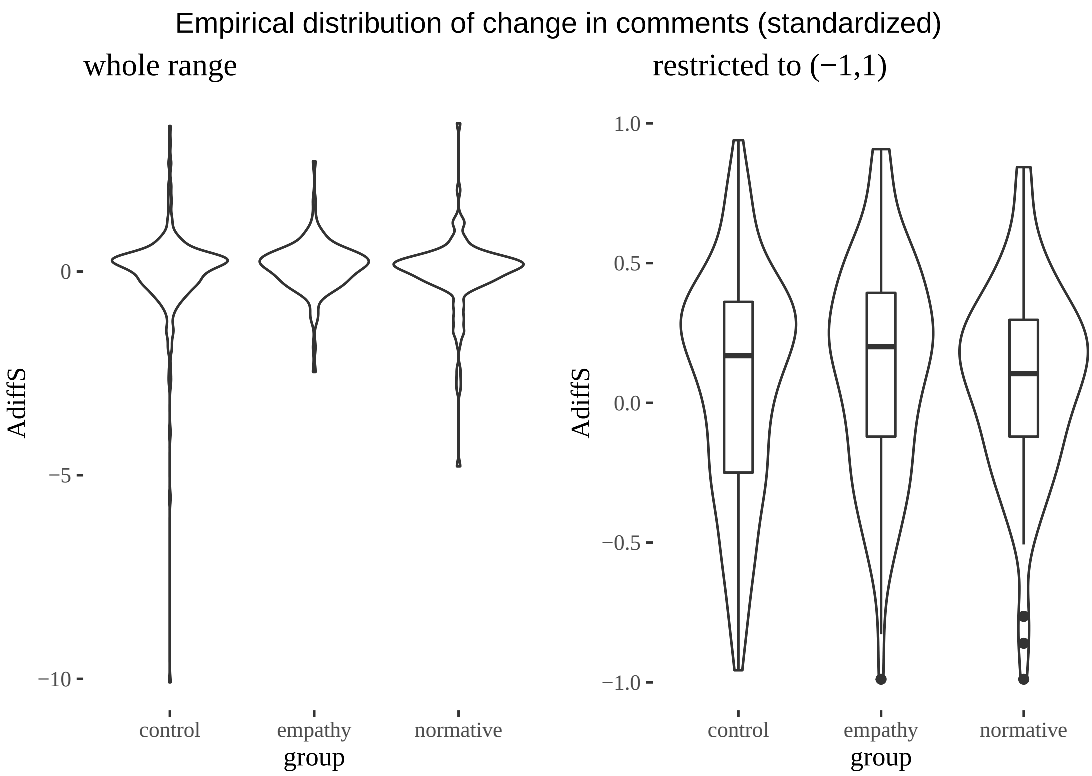


Analogous plot for comments does not reveal this slight downward shift for normative, but otherwise the visualisation migth suggest no strong impact of interventions on attacks, and no impact on comments.

```
violCdiffS <- ggplot(summaries, aes(x=group, y = CdiffS))+
  geom_violin() +theme_tufte()
violJointC <- ggarrange(violCdiffS+ggtitle("whole range"),
              violCdiffS + ylim(c(-1,1))+geom_boxplot(width = .2)+
              ggtitle("restricted to (-1,1)"))
```

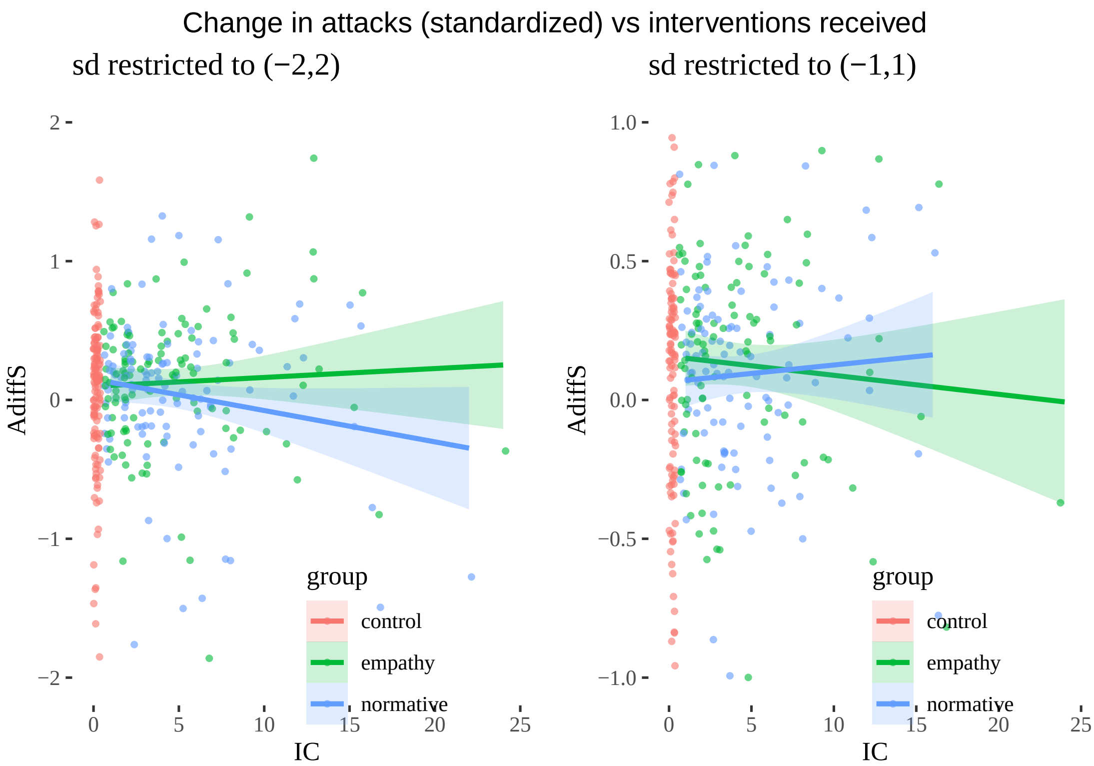
```
## Warning: Removed 90 rows containing non-finite values (stat_ydensity).
## Warning: Removed 90 rows containing non-finite values (stat_boxplot).
```

```
violJointCTitled <- annotate_figure(violJoint,
  top = text_grob("Empirical distribution of change in comments (standardized)",
                size = 12))
violJointCTitled
```



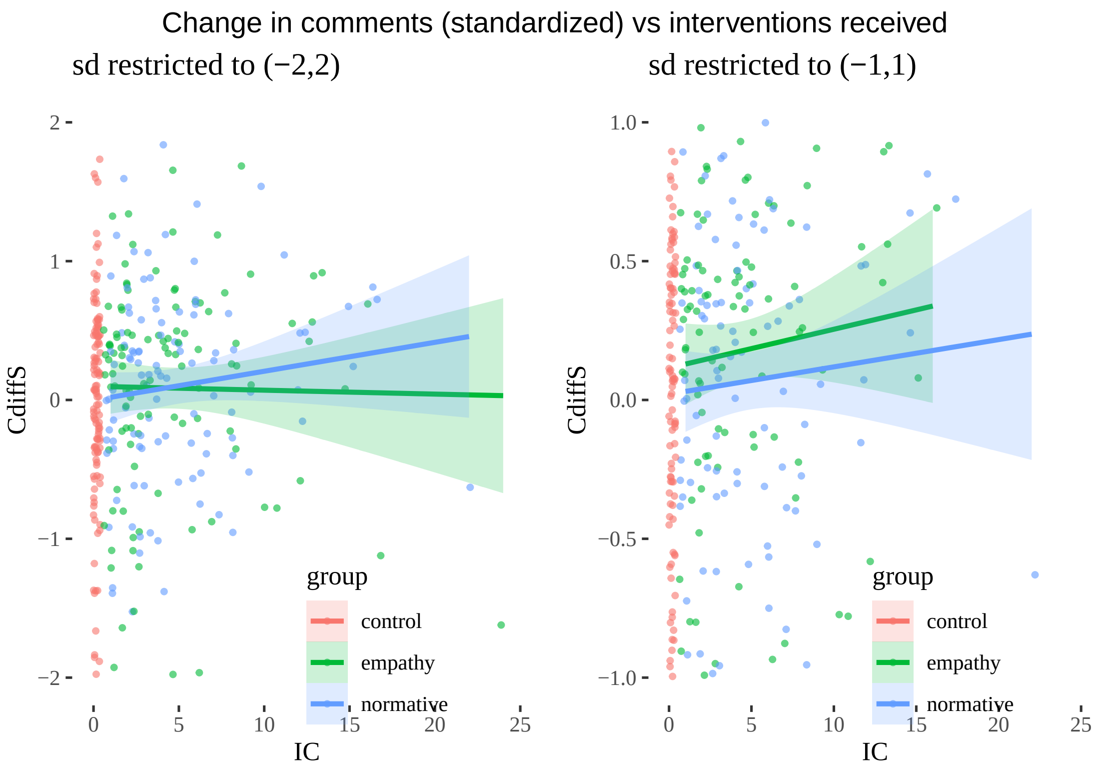
However, plotting changes against intervention counts reveals that restricting attention to various activity levels drastically changes the regression lines.

```
icplot1 <- ggplot(summaries, aes(x = IC, y = AdiffS, color = group, fill = group))+
  geom_jitter(alpha = 0.6, size =.8)+theme_tufte()+
  geom_smooth(alpha = 0.2, method = "lm")+
  xlim(c(0,25))+ylim(c(-2,2))+
  ggtitle("sd restricted to (-2,2)")+
  theme(legend.position = c(0.65, 0.1))

icplot2 <-  ggplot(summaries, aes(x = IC, y = AdiffS, color = group, fill = group))+
  geom_jitter(alpha = 0.6, size =.8)+theme_tufte()+
  geom_smooth(alpha = 0.2, method = "lm")+
  xlim(c(0,25))+ylim(c(-1,1))+ggtitle("sd restricted to (-1,1)")+
  theme(legend.position = c(0.65, 0.1))

icplotJoint <- ggarrange(icplot1, icplot2)
icplotTitled <- annotate_figure(icplotJoint,
  top = text_grob("Change in attacks (standardized) vs interventions received",  size = 12))
icplotTitled
```



Some interactions are also suggested by the differences in linear smoothing when attention is restricted when it comes to change in comments.

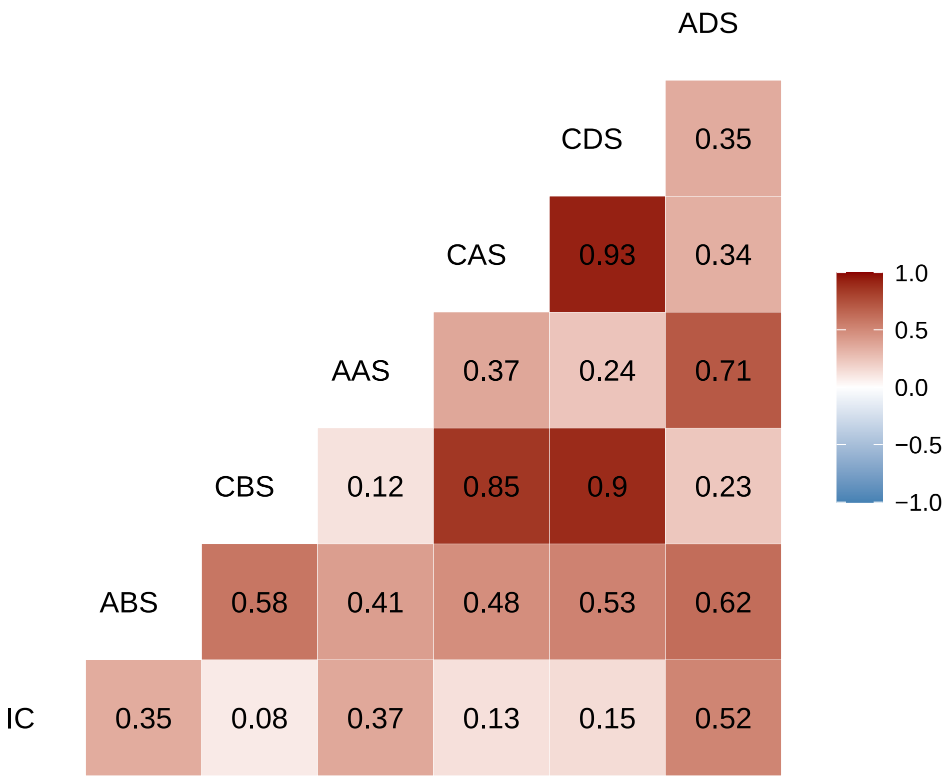
```
icCplot1 <- ggplot(summaries, aes(x = IC, y = CdiffS, color = group, fill = group))+
  geom_jitter(alpha = 0.6, size =.8)+theme_tufte()+
  geom_smooth(alpha = 0.2, method = "lm")+
  xlim(c(0,25))+ylim(c(-2,2))+
  ggtitle("sd restricted to (-2,2)")+
  theme(legend.position = c(0.65, 0.1))

icCplot2 <-  ggplot(summaries, aes(x = IC, y = CdiffS, color = group, fill = group))+
  geom_jitter(alpha = 0.6, size =.8)+theme_tufte()+
  geom_smooth(alpha = 0.2, method = "lm")+
  xlim(c(0,25))+ylim(c(-1,1))+ggtitle("sd restricted to (-1,1)")+
  theme(legend.position = c(0.65, 0.1))

icCplotJoint <- ggarrange(icCplot1, icCplot2)
icCplotTitled <- annotate_figure(icCplotJoint,
  top = text_grob("Change in comments (standardized) vs interventions received",
   size = 12))
icCplotTitled
```



This suggests we should keep an eye out for interactions in the analysis, and that the intial comparison of means or medians between groups might be misleading if the effects in different volume groups are different and cancel each other.

Now, let's inspect correlations between the variables involved in the model:

```
summariesCorr <- select(summaries, IC, ABS, CBS, AAS, CAS, CDS, ADS)
ggcorr(summariesCorr, method = c("pairwise"),
       digits = 4, low = "steelblue", mid = "white",
       high = "darkred", midpoint =0,
       geom = "tile", label = TRUE, label_size=4, label_round =2, layout.exp =1,
       label_alpha = FALSE,hjust = 0.75)
```

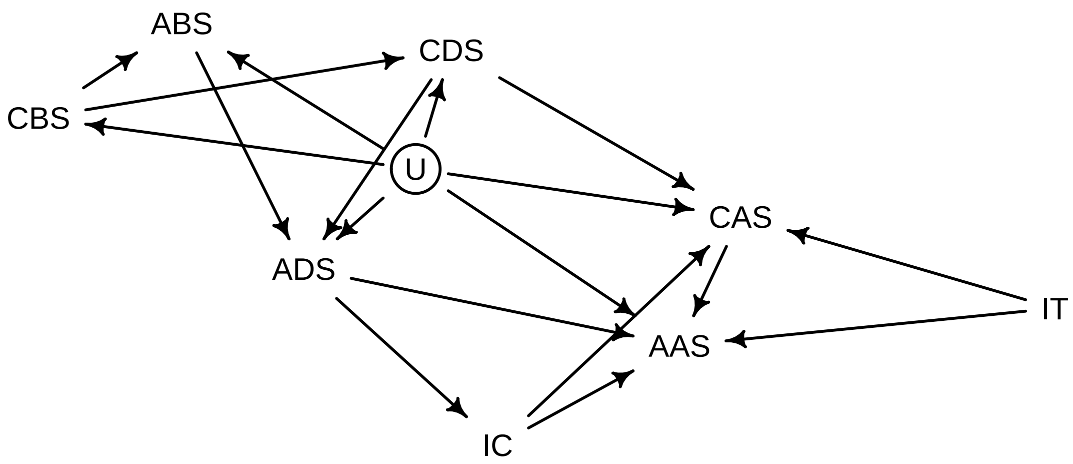


This tells us that almost no predictors are strongly correlated, except for pairs CBS-CDS, so we drop CDS from the analysis and avoid using them in the same model to avoid multicolinearity issues. These are just comments during the intervention period, which, unsurprisingly are also a good proxy for comments before and comments after.

2 Causal inference

To identify the right variables to condition (or not condition) on to identify the causal effect of the interventions, we first need to think about the causal structure of the problem. Here's a plausible causal structure that we will be working with:

```
dag <- dagitty("
  dag{
  CDS -> ADS -> IC  ons
               U [unobserved]
               U -> CBS -> ABS
               U -> ABS
               U -> CDS -> ADS
               U -> ADS
               U -> CAS -> AAS
               U -> AAS
               IC -> AAS
               IC -> CAS
               IT -> CAS
               IT -> AAS
               CBS -> CDS -> CAS
               ABS -> ADS -> AAS
               }")
set.seed(123)
drawdag(dag)
```



ons

Comments during impact attacks during, which trigger interventions. Unmeasured user features cause comments before, which impact attacks before, and also attacks before directly. Comments during (their impact on ADS is aready included) impact attacks during during directly and comments after, which impact attacks after and attacks after directly. Intervention count impacts attacks after and comments after. The same directions of impact are included for intervention type. Finally, comments through time are connected causally, and so are attacks.

We already know not to condition on CDS if we condition on CAS or CBS. What else? IT has no bacwkard paths, but IC does. Let's identify all paths from IC to AAS:

```r
paths(dag, from = c("IC"), to = "AAS")
```

```
## $paths
##  [1] "IC -> AAS"
##  [2] "IC -> CAS -> AAS"
##  [3] "IC -> CAS <- CDS -> ADS -> AAS"
##  [4] "IC -> CAS <- CDS -> ADS <- ABS <- CBS <- U -> AAS"
##  [5] "IC -> CAS <- CDS -> ADS <- ABS <- U -> AAS"
##  [6] "IC -> CAS <- CDS -> ADS <- U -> AAS"
##  [7] "IC -> CAS <- CDS <- CBS -> ABS -> ADS -> AAS"
##  [8] "IC -> CAS <- CDS <- CBS -> ABS -> ADS <- U -> AAS"
##  [9] "IC -> CAS <- CDS <- CBS -> ABS <- U -> AAS"
## [10] "IC -> CAS <- CDS <- CBS -> ABS <- U -> ADS -> AAS"
## [11] "IC -> CAS <- CDS <- CBS <- U -> AAS"
## [12] "IC -> CAS <- CDS <- CBS <- U -> ABS -> ADS -> AAS"
## [13] "IC -> CAS <- CDS <- CBS <- U -> ADS -> AAS"
## [14] "IC -> CAS <- CDS <- U -> AAS"
## [15] "IC -> CAS <- CDS <- U -> ABS -> ADS -> AAS"
## [16] "IC -> CAS <- CDS <- U -> ADS -> AAS"
## [17] "IC -> CAS <- CDS <- U -> CBS -> ABS -> ADS -> AAS"
## [18] "IC -> CAS <- IT -> AAS"
## [19] "IC -> CAS <- U -> AAS"
## [20] "IC -> CAS <- U -> ABS -> ADS -> AAS"
## [21] "IC -> CAS <- U -> ABS <- CBS -> CDS -> ADS -> AAS"
## [22] "IC -> CAS <- U -> ADS -> AAS"
## [23] "IC -> CAS <- U -> CBS -> ABS -> ADS -> AAS"
## [24] "IC -> CAS <- U -> CBS -> CDS -> ADS -> AAS"
## [25] "IC -> CAS <- U -> CDS -> ADS -> AAS"
## [26] "IC -> CAS <- U -> CDS <- CBS -> ABS -> ADS -> AAS"
## [27] "IC <- ADS -> AAS"
## [28] "IC <- ADS <- ABS <- CBS -> CDS -> CAS -> AAS"
## [29] "IC <- ADS <- ABS <- CBS -> CDS -> CAS <- IT -> AAS"
```

```
## [30] "IC <- ADS <- ABS <- CBS -> CDS -> CAS <- U -> AAS"
## [31] "IC <- ADS <- ABS <- CBS -> CDS <- U -> AAS"
## [32] "IC <- ADS <- ABS <- CBS -> CDS <- U -> CAS -> AAS"
## [33] "IC <- ADS <- ABS <- CBS -> CDS <- U -> CAS <- IT -> AAS"
## [34] "IC <- ADS <- ABS <- CBS <- U -> AAS"
## [35] "IC <- ADS <- ABS <- CBS <- U -> CAS -> AAS"
## [36] "IC <- ADS <- ABS <- CBS <- U -> CAS <- IT -> AAS"
## [37] "IC <- ADS <- ABS <- CBS <- U -> CDS -> CAS -> AAS"
## [38] "IC <- ADS <- ABS <- CBS <- U -> CDS -> CAS <- IT -> AAS"
## [39] "IC <- ADS <- ABS <- U -> AAS"
## [40] "IC <- ADS <- ABS <- U -> CAS -> AAS"
## [41] "IC <- ADS <- ABS <- U -> CAS <- IT -> AAS"
## [42] "IC <- ADS <- ABS <- U -> CBS -> CDS -> CAS -> AAS"
## [43] "IC <- ADS <- ABS <- U -> CBS -> CDS -> CAS <- IT -> AAS"
## [44] "IC <- ADS <- ABS <- U -> CDS -> CAS -> AAS"
## [45] "IC <- ADS <- ABS <- U -> CDS -> CAS <- IT -> AAS"
## [46] "IC <- ADS <- CDS -> CAS -> AAS"
## [47] "IC <- ADS <- CDS -> CAS <- IT -> AAS"
## [48] "IC <- ADS <- CDS -> CAS <- U -> AAS"
## [49] "IC <- ADS <- CDS <- CBS -> ABS <- U -> AAS"
## [50] "IC <- ADS <- CDS <- CBS -> ABS <- U -> CAS -> AAS"
## [51] "IC <- ADS <- CDS <- CBS -> ABS <- U -> CAS <- IT -> AAS"
## [52] "IC <- ADS <- CDS <- CBS <- U -> AAS"
## [53] "IC <- ADS <- CDS <- CBS <- U -> CAS -> AAS"
## [54] "IC <- ADS <- CDS <- CBS <- U -> CAS <- IT -> AAS"
## [55] "IC <- ADS <- CDS <- U -> AAS"
## [56] "IC <- ADS <- CDS <- U -> CAS -> AAS"
## [57] "IC <- ADS <- CDS <- U -> CAS <- IT -> AAS"
## [58] "IC <- ADS <- U -> AAS"
## [59] "IC <- ADS <- U -> ABS <- CBS -> CDS -> CAS -> AAS"
## [60] "IC <- ADS <- U -> ABS <- CBS -> CDS -> CAS <- IT -> AAS"
## [61] "IC <- ADS <- U -> CAS -> AAS"
## [62] "IC <- ADS <- U -> CAS <- IT -> AAS"
## [63] "IC <- ADS <- U -> CBS -> CDS -> CAS -> AAS"
## [64] "IC <- ADS <- U -> CBS -> CDS -> CAS <- IT -> AAS"
## [65] "IC <- ADS <- U -> CDS -> CAS -> AAS"
## [66] "IC <- ADS <- U -> CDS -> CAS <- IT -> AAS"
##
## $open
##  [1]  TRUE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [13] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [25] FALSE FALSE  TRUE  TRUE FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE FALSE
## [37]  TRUE FALSE  TRUE  TRUE FALSE  TRUE FALSE  TRUE FALSE  TRUE FALSE FALSE
## [49] FALSE FALSE FALSE  TRUE  TRUE FALSE  TRUE  TRUE FALSE  TRUE FALSE FALSE
## [61]  TRUE FALSE  TRUE FALSE  TRUE FALSE
```

Crucially, all backdoor paths go through ADS, which then becomes either a fork or a pipe, so all backdoor paths can be closed by conditioning on ADS. Moreover there is only one directed indirect path, it goes through CAS, so we should not condition on it if we are to identify causal effect on attacks mediated by impact on comments (unless we care about the direct effect of IC and IT on AAS, but that's a separate question). This is in line with the adjustment set identified algorithmically, and the same move makes sense when we want to predict CAS.

```r
adjustmentSets(dag, exposure = c("IC", "IT"), outcome = "AAS")
```

```
## { ADS }
```

```r
adjustmentSets(dag, exposure = c("IC", "IT"), outcome = "CAS")
```

```
## { ADS }
```

It's open season for other variables, and our decision to include them in the model will be guided by information-theoretic criteria of predictive power.

3 Bayesian models and their priors

We will focus on a class of additive models where the outcome variable is normally distributed around the predicted mean, which is a linear function of predictors (possibly with some interactions). To spoil the story, we will end up using a model, whose specification is as follows:

$$\mathsf{AdiffS} \sim Norm(\mu, \sigma)$$
$$mu_i = \alpha + \beta_{\mathsf{ADS}}[\mathsf{ADS}] + \beta_{\mathsf{IT}}[\mathsf{group}] + \beta_{IC}[\mathsf{group}] \times \mathsf{IC} + b_{\mathsf{ADSIC}} \times \mathsf{ADS} \times \mathsf{IC} + \beta_{\mathsf{CBS}}[\mathsf{group}] \times \mathsf{CBS}$$

References