409 presentation results

Will Gertsch

11/20/2021

Summary stats

```
## Warning in chisq.test(table(y, g)): Chi-squared approximation may be incorrect

## Warning in chisq.test(table(y, g)): Chi-squared approximation may be incorrect
```

	Suppressed	Unsuppressed	Total	P-value
	(N=117)	(N=145)	(N=262)	
Stigma baseline				
Mean (SD)	25.2 (7.26)	25.6 (6.16)	25.4 (6.66)	0.622
Median [Min, Max]	25.0 [10.0, 40.0]	26.0 [10.0, 40.0]	25.0 [10.0, 40.0]	
Stigma 6 month				
Mean (SD)	23.3 (6.10)	24.1 (6.24)	23.7 (6.18)	0.282
Median [Min, Max]	23.0 [10.0, 40.0]	25.0 [10.0, 40.0]	24.0 [10.0, 40.0]	
Stigma 12 month				
Mean (SD)	23.7 (6.48)	23.3 (6.16)	23.5 (6.29)	0.647
Median [Min, Max]	24.0 [10.0, 40.0]	23.0 [10.0, 40.0]	23.0 [10.0, 40.0]	
HIV clinical utilization 12m				
Mean (SD)	1.79 (1.78)	2.94 (5.67)	2.42 (4.41)	0.022
Median [Min, Max]	1.00 [0, 12.0]	2.00 [0, 51.0]	2.00 [0, 51.0]	
Social support				
Mean (SD)	22.0 (11.5)	21.1 (7.21)	21.5 (9.37)	0.449
Median [Min, Max]	24.0 [1.00, 112]	21.0 [3.00, 30.0]	22.5 [1.00, 112]	
Doctors office care utilization 12m				
Mean (SD)	2.08 (3.31)	1.64 (2.30)	1.84 (2.80)	0.229
Median [Min, Max]	1.00 [0, 25.0]	1.00 [0, 18.0]	1.00 [0, 25.0]	
Emergency/urgent care utilization 12m				
Mean (SD)	0.521 (1.06)	0.669 (1.38)	0.603 (1.25)	0.33
Median [Min, Max]	0 [0, 6.00]	0 [0, 10.0]	0 [0, 10.0]	
Times admitted to hospital 12m				
Mean (SD)	0.188 (0.642)	0.462 (1.45)	0.340 (1.17)	0.0429
Median [Min, Max]	0 [0, 5.00]	0 [0, 10.0]	0 [0, 10.0]	
Total care utilization				
Mean (SD)	4.57 (4.91)	5.71 (7.91)	5.20 (6.75)	0.156
Median [Min, Max]	3.00 [0, 32.0]	4.00 [0, 66.0]	3.00 [0, 66.0]	
Gender				
cis	102 (87.2%)	129 (89.0%)	231 (88.2%)	0.902
other	3 (2.6%)	3 (2.1%)	6 (2.3%)	
trans	12 (10.3%)	13 (9.0%)	25 (9.5%)	
Sexual orientation				
bi	16 (13.7%)	23 (15.9%)	39 (14.9%)	0.245
gay	72 (61.5%)	84 (57.9%)	156 (59.5%)	
other	1 (0.9%)	4 (2.8%)	5 (1.9%)	
queer	9 (7.7%)	4 (2.8%)	13 (5.0%)	
straight	19 (16.2%)	30 (20.7%)	49 (18.7%)	
Race/ethnicity				
latino	25 (21.4%)	36 (24.8%)	61 (23.3%)	0.0272
non-latino black	54 (46.2%)	81 (55.9%)	135 (51.5%)	
other/mixed	12 (10.3%)	15 (10.3%)	27 (10.3%)	
white	26 (22.2%)	13 (9.0%)	39 (14.9%)	
Employment				
disabled	10 (8.5%)	11 (7.6%)	21 (8.0%)	0.469
full-time	43 (36.8%)	43 (29.7%)	86 (32.8%)	
part-time	18 (15.4%)	35 (24.1%)	53 (20.2%)	
student	6 (5.1%)	8 (5.5%)	14 (5.3%)	
unemployed	40 (34.2%)	48 (33.1%)	88 (33.6%)	

Regressions

```
# Viral suppression
mod = glm(ViralSupp ~  CAREHV06.12m + stigmasum_6m + stigmasum_baseline,
          family = binomial(), data = d)

summary(mod)
```

```
##
## Call:
## glm(formula = ViralSupp ~ CAREHV06.12m + stigmasum_6m + stigmasum_baseline,
##     family = binomial(), data = d)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -1.285  -1.110  -0.935   1.213   1.791
##
## Coefficients:
##                     Estimate Std. Error z value Pr(>|z|)
## (Intercept)         0.575762   0.579074   0.994   0.3201
## CAREHV06.12m       -0.120155   0.061514  -1.953   0.0508 .
## stigmasum_6m       -0.017362   0.025061  -0.693   0.4885
## stigmasum_baseline -0.004793   0.023462  -0.204   0.8381
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 360.21  on 261  degrees of freedom
## Residual deviance: 352.80  on 258  degrees of freedom
## AIC: 360.8
##
## Number of Fisher Scoring iterations: 5
```

```
# clinical utilization
mod = glm(CAREHV06.12m ~ stigmasum_6m + stigmasum_baseline,
          family = poisson(), data = d)
summary(mod)
```

```
##
## Call:
## glm(formula = CAREHV06.12m ~ stigmasum_6m + stigmasum_baseline,
##     family = poisson(), data = d)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.1616  -1.2074  -0.4917   0.1385  14.1171
##
## Coefficients:
##                     Estimate Std. Error z value Pr(>|z|)
## (Intercept)         0.857857   0.176040   4.873 1.10e-06 ***
## stigmasum_6m        0.047206   0.007874   5.995 2.04e-09 ***
```

```
## stigmasum_baseline -0.044513   0.007336  -6.068 1.29e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 878.17  on 261  degrees of freedom
## Residual deviance: 832.07  on 259  degrees of freedom
## AIC: 1416.5
##
## Number of Fisher Scoring iterations: 6
```

```
# stigma
mod = glm(stigmasum_6m ~ stigmasum_baseline,
          family = gaussian(), data = d)
summary(mod)
```

```
##
## Call:
## glm(formula = stigmasum_6m ~ stigmasum_baseline, family = gaussian(),
##     data = d)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -21.4043   -2.8784   -0.1628    3.1270   16.5546
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)        10.43823    1.24650   8.374 3.49e-15 ***
## stigmasum_baseline  0.52415    0.04749  11.038  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 26.11461)
##
##     Null deviance: 9971.4  on 261  degrees of freedom
## Residual deviance: 6789.8  on 260  degrees of freedom
## AIC: 1602.3
##
## Number of Fisher Scoring iterations: 2
```

SEM

The DAG we are interested in is

```
g1 = dagitty('dag {
             S0 [pos="0,0"]
             S1 [pos="1,0"]
             CU [pos="2,0"]
             VS [pos="3,0"]
             S0 -> S1 -> CU -> VS
```

```
}')
plot(g1)
```

Let's fit using piecewiseSEM and see how well this models the data.

```
sem_g1 = psem(
  glm(ViralSupp ~  CAREHV06.12m,
          family = binomial(), data = d),
  glm(CAREHV06.12m ~ stigmasum_6m,
          family = poisson(), data = d),
  glm(stigmasum_6m ~ stigmasum_baseline,
          family = gaussian(), data = d)
)
summary(sem_g1)
```

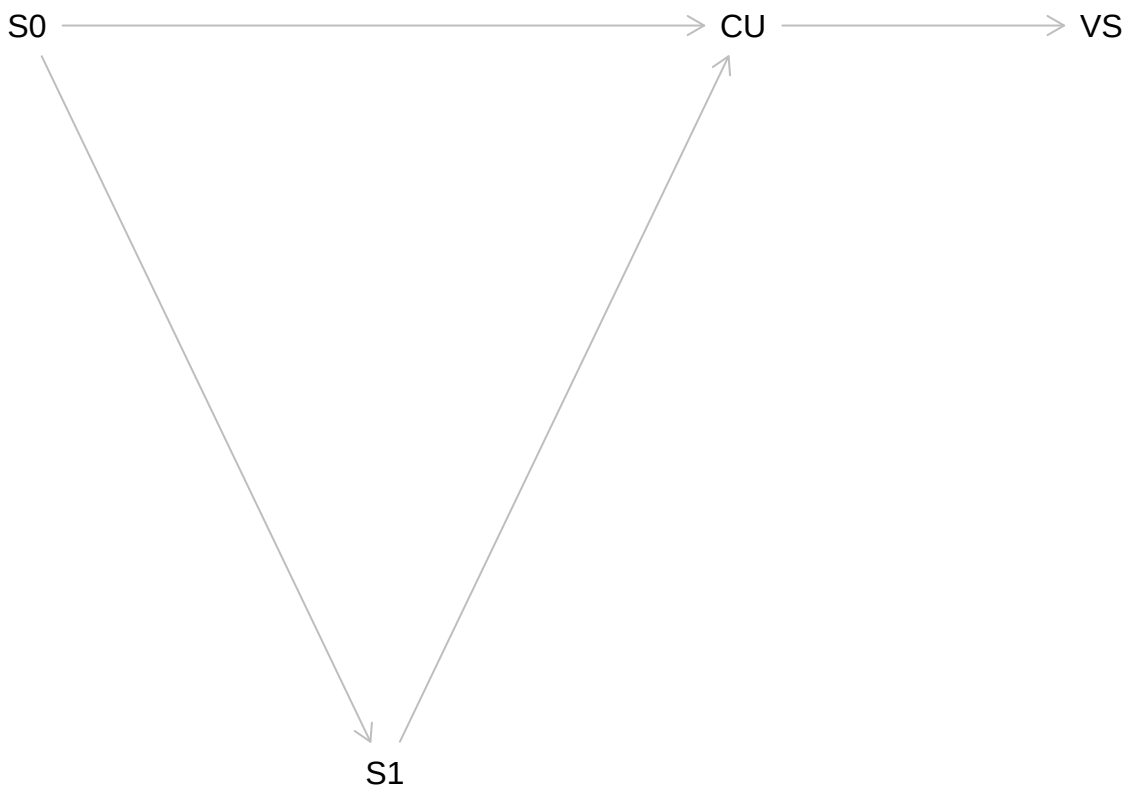
```
##   |                                                                      |
```

```
##
## Structural Equation Model of sem_g1
##
## Call:
##   ViralSupp ~ CAREHV06.12m
##   CAREHV06.12m ~ stigmasum_6m
##   stigmasum_6m ~ stigmasum_baseline
##
```

```
##     AIC      BIC
##  57.926   82.904
##
## ---
## Tests of directed separation:
##
##                            Independ.Claim Test.Type  DF Crit.Value P.Value
##   CAREHV06.12m ~ stigmasum_baseline + ...      coef 259    -6.0681  0.0000 ***
##      ViralSupp ~ stigmasum_baseline + ...      coef 259    -0.7423  0.4579
##            ViralSupp ~ stigmasum_6m + ...      coef 258    -0.6928  0.4885
##
## Global goodness-of-fit:
##
##   Fisher's C = 43.926 with P-value = 0 and on 6 degrees of freedom
##
## ---
## Coefficients:
##
##       Response          Predictor Estimate Std.Error  DF Crit.Value P.Value
##      ViralSupp       CAREHV06.12m  -0.1185    0.0607 260    -1.9534  0.0508
##   CAREHV06.12m       stigmasum_6m   0.0195    0.0064 260     3.0483  0.0023
##   stigmasum_6m stigmasum_baseline   0.5242    0.0475 260    11.0377  0.0000
##   Std.Estimate
##        -0.2769
##              -  **
##         0.5649 ***
##
##   Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05
##
## ---
## Individual R-squared:
##
##       Response     method R.squared
##      ViralSupp nagelkerke      0.03
##   CAREHV06.12m nagelkerke      0.03
##   stigmasum_6m       none      0.32
```

Goodness of fit p-value is 0, which suggests DAG g1 is not a good fit for the data. The independence claim that seems to be causing the issue is baseline stigma and care utilization. Let's try fitting a model that includes that path, i.e. following DAG g2.

```
g2 = dagitty('dag {
            S0 [pos="0,0"]
            S1 [pos="1,1"]
            CU [pos="2,0"]
            VS [pos="3,0"]
            S0 -> S1 -> CU -> VS
            S0 -> CU
}')
plot(g2)
```




```
sem_g2 = psem(
  glm(ViralSupp ~  CAREHV06.12m,
          family = binomial(), data = d),
  glm(CAREHV06.12m ~ stigmasum_6m + stigmasum_baseline,
          family = poisson(), data = d),
  glm(stigmasum_6m ~ stigmasum_baseline,
          family = gaussian(), data = d)
)
summary(sem_g2)
```

```
##   |                                                                      |
```

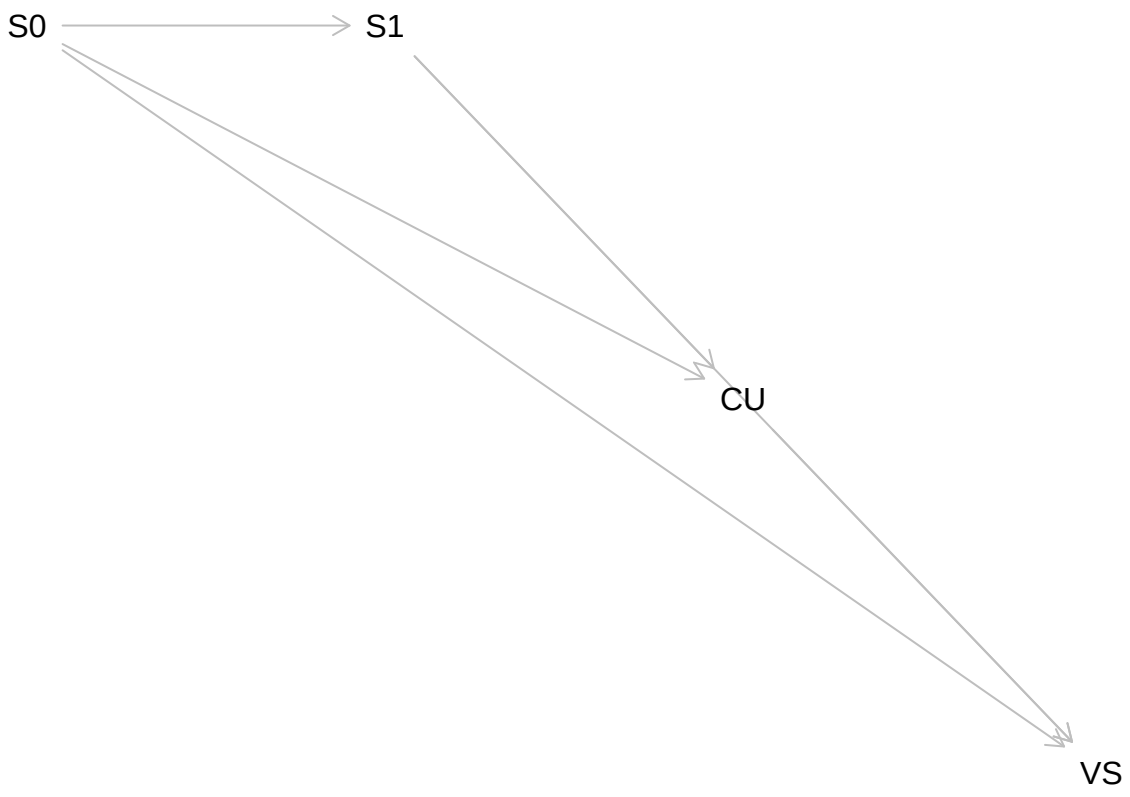
```
##
## Structural Equation Model of sem_g2
##
## Call:
##   ViralSupp ~ CAREHV06.12m
##   CAREHV06.12m ~ stigmasum_6m + stigmasum_baseline
##   stigmasum_6m ~ stigmasum_baseline
##
##     AIC      BIC
##  18.995   47.542
##
## ---
## Tests of directed separation:
```

```
##
##                         Independ.Claim Test.Type  DF Crit.Value P.Value
##   ViralSupp ~ stigmasum_baseline + ...      coef 259    -0.7423  0.4579
##         ViralSupp ~ stigmasum_6m + ...      coef 258    -0.6928  0.4885
##
## Global goodness-of-fit:
##
##   Fisher's C = 2.995 with P-value = 0.559 and on 4 degrees of freedom
##
## ---
## Coefficients:
##
##       Response          Predictor Estimate Std.Error  DF Crit.Value P.Value
##      ViralSupp       CAREHV06.12m  -0.1185    0.0607 260    -1.9534  0.0508
##   CAREHV06.12m       stigmasum_6m   0.0472    0.0079 259     5.9950  0.0000
##   CAREHV06.12m stigmasum_baseline  -0.0445    0.0073 259    -6.0681  0.0000
##   stigmasum_6m stigmasum_baseline   0.5242    0.0475 260    11.0377  0.0000
##   Std.Estimate
##        -0.2769
##              - ***
##              - ***
##         0.5649 ***
##
##   Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05
##
## ---
## Individual R-squared:
##
##       Response     method R.squared
##      ViralSupp nagelkerke      0.03
##   CAREHV06.12m nagelkerke      0.16
##   stigmasum_6m       none      0.32
```

The model now fits the data according to the goodness of fit statistic.

Finally, we should compare to the fully saturated model given by DAG g3.

```
g3 = dagitty('dag {
            S0 [pos="0,0"]
            S1 [pos="1,0"]
            CU [pos="2,1"]
            VS [pos="3,2"]
            S0 -> S1 -> CU -> VS
            S0 -> CU
            S1 -> VS
            S0 -> VS
}')
plot(g3)
```




```
sem_g3 = psem(
  glm(ViralSupp ~  CAREHV06.12m + stigmasum_6m + stigmasum_baseline,
          family = binomial(), data = d),
  glm(CAREHV06.12m ~ stigmasum_6m + stigmasum_baseline,
          family = poisson(), data = d),
  glm(stigmasum_6m ~ stigmasum_baseline,
          family = gaussian(), data = d)
)
summary(sem_g3)
```

```
##
## Structural Equation Model of sem_g3
##
## Call:
##   ViralSupp ~ CAREHV06.12m + stigmasum_6m + stigmasum_baseline
##   CAREHV06.12m ~ stigmasum_6m + stigmasum_baseline
##   stigmasum_6m ~ stigmasum_baseline
##
##     AIC      BIC
##  20.000   55.683
##
## ---
## Tests of directed separation:
##
##  No independence claims present. Tests of directed separation not possible.
```

```
##
## Global goodness-of-fit:
##
##   Fisher's C = 0 with P-value = 1 and on 0 degrees of freedom
##
## ---
## Coefficients:
##
##       Response          Predictor Estimate Std.Error  DF Crit.Value P.Value
##      ViralSupp       CAREHV06.12m  -0.1202    0.0615 258    -1.9533  0.0508
##      ViralSupp       stigmasum_6m  -0.0174    0.0251 258    -0.6928  0.4885
##      ViralSupp stigmasum_baseline  -0.0048    0.0235 258    -0.2043  0.8381
##   CAREHV06.12m       stigmasum_6m   0.0472    0.0079 259     5.9950  0.0000
##   CAREHV06.12m stigmasum_baseline  -0.0445    0.0073 259    -6.0681  0.0000
##   stigmasum_6m stigmasum_baseline   0.5242    0.0475 260    11.0377  0.0000
##   Std.Estimate
##        -0.2796
##        -0.0566
##        -0.0168
##              - ***
##              - ***
##         0.5649 ***
##
##   Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05
##
## ---
## Individual R-squared:
##
##       Response     method R.squared
##      ViralSupp nagelkerke      0.04
##   CAREHV06.12m nagelkerke      0.16
##   stigmasum_6m       none      0.32
```

This model is fully specified so no independence claims are tested. Looking at the regression p-values, it does suggest we should get rid of the arrows from stigma to viral suppression. This is exactly what we did in DAG g2.

Let's compare all of our model

```
anova(sem_g1, sem_g2, sem_g3)
```

```
## Chi-square Difference Test
##
##         AIC    BIC Fisher.C Fisher.C.Diff DF.diff P.value
##  1    57.926 82.904   43.926
##  vs 2 18.995 47.542    2.995        40.931       2       0 ***
##  vs 3 20.000 55.683    0.000        43.926       6       0 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05
```

The fully specified model will always fit better by goodness of fit, but look at the AIC/BIC. Model g2 has lower AIC/BIC which suggests it is the better model.

Conclusions

For these basic models, it appears that the effect of stigma on viral suppression is primarily mediated through clinical utilization. The effect of baseline stigma on clinical utilization is not fully mediated by 6 month stigma and should be included as a predictor.

Baseline stigma seems to decrease clinical utilization, but 6 month stigma seems to increase it. This might suggest there is some selection bias here. Maybe people who show up to 6 month visits are more likely to go get care? Viral suppression also has a negative relationship with clinical utilization. This is not what I was expecting.

Future analysis should introduce other variables such as demographics in order to address possible confounders of the effects of interest. Another thing to do is to address the possible selection bias induced by only including those with complete data.