Structural Equation Modeling for HIV Stigma, Clinical Utilization, and Viral Suppression

Will Gertsch

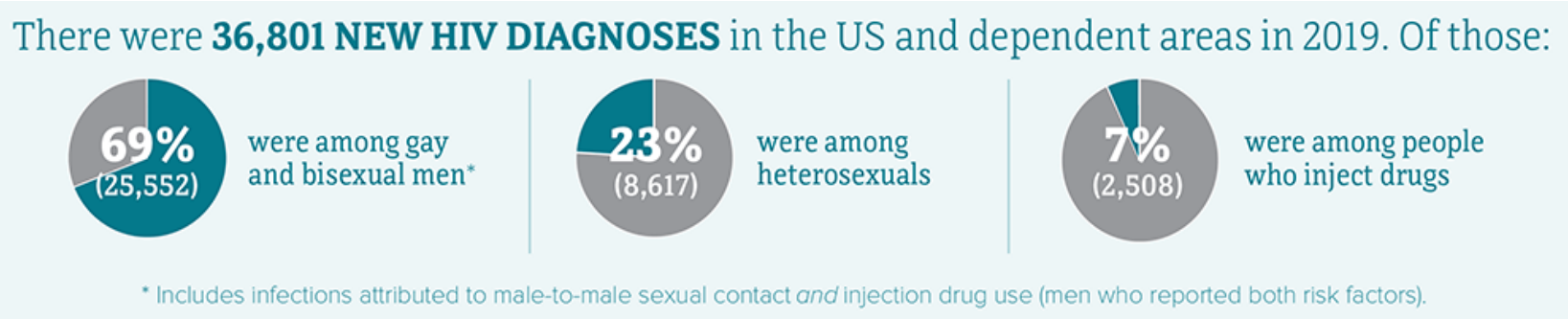
UCLA

11/22/2021

Facts about HIV

Human immunodeficiency virus (HIV) is a virus that can lead to acquired immunodeficiency syndrome (AIDS).

- The virus attacks certain types of white blood cell, leaving the body open to opportunistic infections.
- First case in US reported in 1981.
- 1.2 million living with HIV in the US (CDC, May 2021).



Source: www.cdc.gov/hiv

HIV care continuum

Source: www.hiv.gov

Viral suppression

It is important for young people to know their HIV status so they can take medicine to treat HIV if they have the virus. Taking HIV medicine every day can make the viral load undetectable. People who get and keep an undetectable viral load (or remain virally suppressed) can stay healthy for many years and have effectively no risk of transmitting HIV to their sex partners.

Although more than half of adults and adolescents with diagnosed HIV are virally suppressed, more work is needed to increase these rates. **For every 100 adults and adolescents with diagnosed HIV in 2019:**

76 received some HIV care

58 were retained in care †

66 were virally suppressed ‡

Source: www.cdc.gov/hiv

HIV Stigma

"**Stigma is a discrediting social label which changes the way the individual looks at him/her self and disqualifies them from full social acceptance**. HIV has a particular, insidious stigmatization since it is associated with factors which imbue judgment and criticism such as sexual behavior and substance use. ... **leads to fear of accessing services**, ... **psychosocial consequences**. In a study in an urban clinic of adult HIV patients, **HIV stigma** ... **was associated with depressive symptomatology and a lower quality of life**."
- Abstract from Wright, Kathryn et al. "*Stigma scale revised: reliability and validity of a brief measure of stigma for HIV+ youth.*"

Study background

"Use of Social Media to Improve Engagement, Retention, and Health Outcomes along the HIV Care Continuum"

- Part of a program run by the Health Resources & Services Administration to develop innovative methods of HIV care.
- Ran 2015-2019 across 10 demonstration centers nationwide.
- Each center had a unique intervention.
- UCLA is responsible for evaluation and dissemination of findings.
- Study population: HIV+, underserved, underinsured, hard-to-reach youth and young adults (aged 13-34).

Research questions

We are interested in the following:

- How does stigma (or a change in stigma) relate to 12 month clinical utilization and viral suppression?
- Mediation relationships?

Goal: capture the complex relationships between variables $\rightarrow$ SEM

Notes on the data

Study was longitudinal with visits for surveys and lab tests.

- Survey questions at baseline, 6 months, and 12 months measure stigma, clinical utilization, demographics, etc.
- Viral suppression is measured by a viral load test at a separate clinic.
- Self-report visits and viral load test visits may not be on the same day (or week).
- $\therefore$ we use the closest viral suppression measurement within 30 days.

Key variables

- **Clinical utilization**:
  Number of times received care for HIV at a clinic in past 6 months.
- **Stigma**: 10 questions on a scale from 1 to 4 representing strong disagreement to strong agreement.
- Ex: "I have been hurt by how people reacted to learning I have HIV."
- Stigma questions will be combined by summing the individual scores.
- **Viral suppression** is binary (1=Yes, 0=No).

Data and sample size

Lots of missing data to keep track of.

- Total sample size: $N = 964$.
- Drop missing baseline self-report: $n = 963$.
- Drop missing 6m self-report: $N = 730$.
- Drop missing 12m self-report: $N = 588$.
- Drop missing viral suppression measure: $N = 533$.
- Drop those without viral suppression measure within 30 days of 12m self-report: $N = 262$.

Only 27% of observations have complete data.
$\rightarrow$ This makes me worry about selection bias.

Summary statistics

	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]	

Introduction to Structural Equation Models

Structural equation models use multiple regression equations to model relationships between variables.

- Example: mediation
  - $X \rightarrow M \rightarrow Y$
  - $E[Y|M] = \beta_0 + \beta_1 M$
  - $E[M|X] = \alpha_0 + \alpha_1 X$
- Two frameworks for combining regression equations for inference.
  1. Covariance based (classical).
  2. Conditional independence based (Shipley 2000).
- The classical approach is more developed, but latter approach is easier to apply to non-normal data.

Basic ideas

To model the relationship between variables $Y_1, \ldots, Y_m$, we specify a joint distribution $P(Y) = P(Y_1, \ldots, Y_m)$.

- We can always factorize the joint using conditional probabilities.

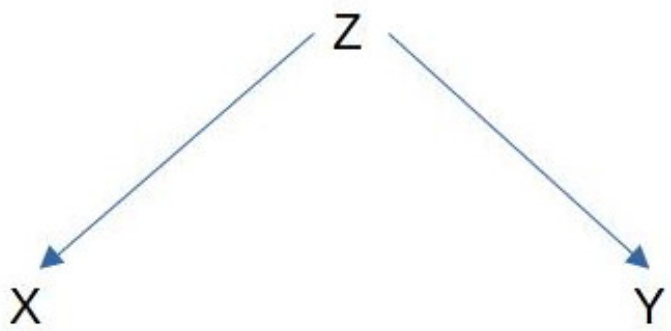
$$P(Y_1, \ldots, Y_m) = P(Y_2, \ldots Y_m | Y_1) P(Y_1) = \ldots$$

- If two variables are conditionally independent, we can factorize the joint distribution further.
- EX: If $Y \perp\!\!\!\perp X | Z$, then

$$P(X, Y, Z) = P(X|Z) P(Y|Z)$$

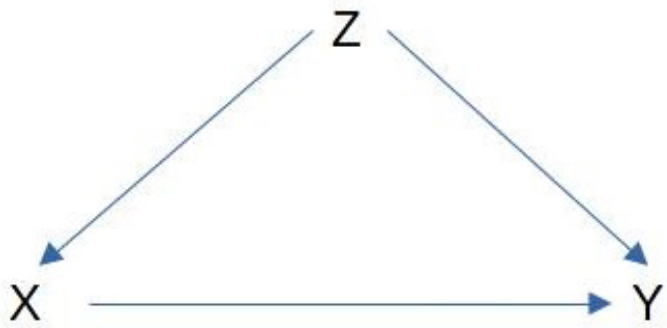
DAG notation

It is helpful to visualize the conditional independencies using a directed acyclic graph. The factorization $P(X, Y, Z) = P(X|Z)P(Y|Z)$ can be drawn as



X and Y will covary in the data because the confounder Z.

Regression test for independence



Suppose we want to test whether $X \perp\!\!\!\perp Y|Z$.

1. Assume a distribution for outcome Y.
2. Specify model $E[Y|X, Z] = \beta_0 + \beta_1 X + \beta_2 Z$
3. Test $H_0 : \beta_1 = 0$.
4. If reject H_0 : conclude X, Y are not conditionally independent.
5. Equivalently, compare models with and without $X \to Y$.

Shipley's goodness of fit test

Suppose we want to test if the proposed joint fits the data. Shipley (2000) proposed the following:

1. Translate DAG into k conditional independencies.
2. For $i = 1, \ldots, k$, test for independence using regression.
3. Obtain p-value p_i from test i.
4. An overall test statistic is

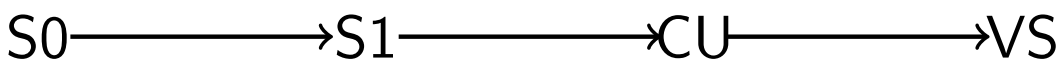
$$C = -2 \sum_{i=1}^{k} \log(p_i) \sim \chi^2_{2k}$$

5. If reject H_0, conclude model is incorrectly specified.

Shipley also developed AIC/BIC based on C.

Preliminary Results

DAG



- S0: baseline stigma.
- S1: 6m stigma.
- CU: number of HIV-related clinic visits for past 6m at 12m.
- VS: viral suppression at 12m.

The DAG represents the hypothesis that effect of stigma on viral suppression is fully mediated by clinical utilization.

Results

```
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)
```

```
## 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
```

Conclusion: not a good model for the data.

Results

Add the path $S0 \rightarrow CU$.

```
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)
)
```

```
##                         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
```

Model fits the data much better.

Comparison to saturated model

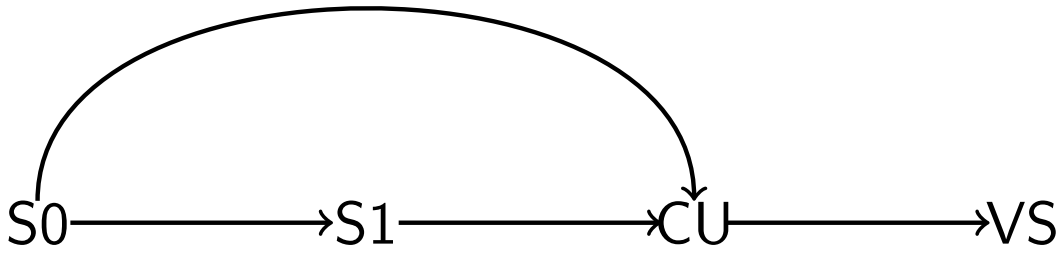
We can use AIC/BIC to compare to the saturated model which adds the paths $S0 \rightarrow VS$ and $S1 \rightarrow VS$.

```
##          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
```

Model 2 has the lowest AIC/BIC and was a reasonable fit to the data.

Conclusions

The DAG suggested by this analysis is



This result implies the effect of stigma on viral suppression is mediated by clinical utilization but 6 month stigma does not fully mediate baseline stigma.

Future work

- Further clarify research question.
- Total effect of stigma on viral suppression?
- Add other variables such as social support, demographics, etc. to address confounding of main effects of interest.
- Is there selection bias induced by requiring complete data?

Questions?

Coefficients from final model

```
## 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
```

- VS has a negative relationship with CU.
- S0 is associated with a decrease CU while S1 is associated with an increase in CU.

References (study)

- https://hab.hrsa.gov/about-ryan-white-hivaids-program/spns-social-media
- https://targethiv.org/library/spns-social-media-initiative-demonstration-site-resources
- Brooks RA, Swendeman D, Hannay J. Special Projects of National Significance Social Media Initiative: Improving Health Outcomes Along the HIV Care Continuum in Youth and Young Adults Living With HIV [published online ahead of print, 2020 Aug 6]. Health Promot Pract. 2020. doi.org/10.1177/1524839920944806.
- Wright, Kathryn et al. "Stigma scale revised: reliability and validity of a brief measure of stigma for HIV+ youth." The Journal of adolescent health : official publication of the Society for Adolescent Medicine vol. 40,1 (2007): 96-8. doi:10.1016/j.jadohealth.2006.08.001

References (SEM)

- Tarka, Piotr. "An overview of structural equation modeling: its beginnings, historical development, usefulness and controversies in the social sciences." Quality & quantity vol. 52,1 (2018): 313-354. doi:10.1007/s11135-017-0469-8
- Shipley, B. (2000a) A new inferential test for path models based on directed acyclic graphs. Structural Equation Modeling, 7, 206– 218.
- Shipley, B. (2013) The AIC model selection method applied to path analytic models compared using a d-separation tests. Ecology, 94, 560– 564.
- Lefcheck, J.S. (2016), piecewiseSEM: Piecewise structural equation modelling in r for ecology, evolution, and systematics. Methods Ecol Evol, 7: 573-579.