Understanding the Opioid Syndemic in North Carolina: A Novel Approach to Modeling and Identifying Factors

Eva Murphy*, David Kline**, Kathleen L. Egan **, Kathryn E. Lancaster **, William C. Miller +, Lance Waller ▲, Staci A. Hepler*

WAKE FOREST UNIVERSITY

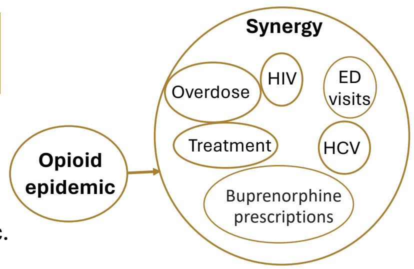
Motivation

- **Opioid epidemic** is a critical public health problem, e.g. in 2021 4041 NC residence died from opioid overdose

Challenge: no data to fully capture this complex problem. → **Need of an alternative approach**

We know that:

- Opioid epidemic is associated with HIV and HCV infections.
- Fatal and nonfatal overdose, ED admissions and treatment admissions are also linked to opioid misuse.
- These epidemics interact synergistically, forming a syndemic.



⇓

Identify underlying factors that explain dependencies in the observed data.

Methodology

- **Bayesian Hierarchical Model:**
  - Observed count for outcome k in year j and in county i :

$$Y_{ijk} \sim Poisson(E_{ijk}\lambda_{ijk})$$
$$\log(\lambda_{ijk}) = \Gamma_k' F_{ij} + \epsilon_{ijk}$$
$$F_{ij} \sim ICAR - AR(1)$$
$$\epsilon_{ijk} \sim N(0, \sigma_k^2)$$

Factor Analysis

- allows us to design factors that yield **meaningful interactions** among predefined subsets of the the observed outcome.

- **Describe the opioid syndemic through 4 factors:**
  - Factor 1 to characterize **variation** shared by all **six outcomes: overall burden.**
  - Factor 2 to characterize **variation** shared by **Death (D) and ED visits (E)**, after accounting for F1: **overdose aspect** of the epidemic.
  - Factor 3 to characterize **variation** shared by **Treatment (T) and Buprenorphine (B)**, after accounting for F1: **treatment aspect** of the epidemic.
  - Factor 4 to characterize the remaining shared **variation** between **HCV (C) and HIV (I), after** accounting for F1: **infectious disease** elements of the epidemic.

Identifiability of $log(\lambda_{ij}) = \Gamma \mathbf{F}_{ij} + \epsilon_{ij}$

- A well-known challenge in factor analysis modeling is the identifiability[1] of model parameters, i.e., The likelihood $f(\boldsymbol{Y}_{ij}|\boldsymbol{\Gamma}, \mathbf{F}_{ij}, \boldsymbol{\epsilon}_{ij})$ is equivalent to $f(\boldsymbol{Y}_{ij}|\boldsymbol{\Gamma}\mathbf{H}, \mathbf{H}'\mathbf{F}_{ij}, \boldsymbol{\epsilon}_{ij})$ for any $m \times m$ orthogonal $\boldsymbol{H}$.
- **Common solution**: specify the loadings matrix to be lower triangular[2] ⇒ factors lacking meaningful interpretation.
- **Proposed solution**: specify the loadings matrix $\boldsymbol{\Gamma}$ as needed to quantify meaningful relationships among subsets of outcomes and employ LQ decomposition[3] on $\boldsymbol{\Gamma}$ to decompose it into a lower triangular matrix with **positive diagonal elements**, $\boldsymbol{L}$, and an orthogonal matrix, $\boldsymbol{Q}$, i.e., $\boldsymbol{log(\lambda_{ij})} = \boldsymbol{\Gamma F_{ij}} + \boldsymbol{\epsilon_{ij}} = \boldsymbol{LQF_{ij}} + \boldsymbol{\epsilon_{ij}} = \boldsymbol{L\,\widetilde{F}_{ij}} + \boldsymbol{\epsilon_{ij}}$, ensuring that the new model adheres to the identifiability constraints[4] of a factor model.

Illustration:

$\log(\lambda_{ijk}) = \Gamma_k' F_{ij} + \epsilon_{ijk}, k = 1, \ldots, 4$, written in matrix form:

Loadings matrix

$$\log(\boldsymbol{\lambda}_{ij}) \equiv \begin{bmatrix} \log(\lambda_{ijD}) \\ \log(\lambda_{ijE}) \\ \log(\lambda_{ijT}) \\ \log(\lambda_{ijB}) \\ \log(\lambda_{ijC}) \\ \log(\lambda_{ijI}) \end{bmatrix} = \begin{bmatrix} 1 & \gamma_{12} & 0 & 0 \\ \gamma_{21} & 1 & 0 & 0 \\ \gamma_{31} & 0 & 1 & 0 \\ \gamma_{41} & 0 & \gamma_{43} & 0 \\ \gamma_{51} & 0 & 0 & 1 \\ \gamma_{61} & 0 & 0 & \gamma_{64} \end{bmatrix} \cdot \begin{bmatrix} F_1 \\ F_2 \\ F_3 \\ F_4 \end{bmatrix} + \begin{bmatrix} \epsilon_D \\ \epsilon_E \\ \epsilon_T \\ \epsilon_B \\ \epsilon_C \\ \epsilon_I \end{bmatrix}$$

↓

Must be identifiable ≡ *lower triangular with positive diagonal elements*

↙ ↘

L Q

factorization

Algorithm: Metropolis-Hastings Update of Loadings

0\. Let $\boldsymbol{\Gamma}$ denote the current value of the loadings with $\boldsymbol{\Gamma} = \boldsymbol{LQ}$.

1\. Propose a new value γ_{km}^* using a normal random walk, centered at the current value γ_{km}. Let $\boldsymbol{\Gamma}^*$ be the loadings matrix containing the proposed γ_{km}^*.

2\. Perform the LQ-decomposition of $\boldsymbol{\Gamma}^*$:

Input($\boldsymbol{\Gamma}^*$)

- Let $\boldsymbol{\Gamma}^{*\prime} = [\Gamma_2^{*\prime} \,|\, \Gamma_2^{*\prime}]$.
- Apply the QR Gram-Schmidt process to get $\boldsymbol{Q}^*$ and $\boldsymbol{R}_1^*$ such that $\boldsymbol{\Gamma}_1^{*\prime} = \boldsymbol{Q}^* \boldsymbol{R}_1^*$.
  - Compute $\boldsymbol{R}_2^* = \left[\boldsymbol{Q}^{*\prime} \cdot \Gamma_2^{*\prime}\right]$.
  - Let $\mathbf{R}^* = [\boldsymbol{R}_1^* | \boldsymbol{R}_2^*]$.

Output ($\boldsymbol{L}^* = \boldsymbol{R}^{*\prime}, \boldsymbol{Q}^* = \boldsymbol{Q}^*$)

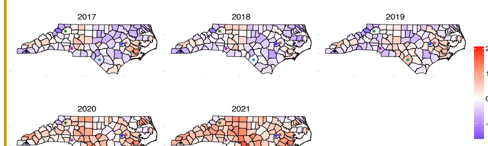
3\. Compute the Metropolis-Hastings acceptance probability:

$$\tau = \min\left(1, \frac{f(\boldsymbol{y}|\boldsymbol{L}(\boldsymbol{\Gamma}^*), \boldsymbol{Q}(\boldsymbol{\Gamma}^*), \boldsymbol{F}, \boldsymbol{\epsilon})\pi(\Gamma^*)}{f(\boldsymbol{y}|\boldsymbol{L}(\boldsymbol{\Gamma}), \boldsymbol{Q}(\boldsymbol{\Gamma}), \boldsymbol{F}, \boldsymbol{\epsilon})\pi(\Gamma)}\right).$$

4\. Set $\boldsymbol{\Gamma} = \boldsymbol{\Gamma}^*$ with probability τ .

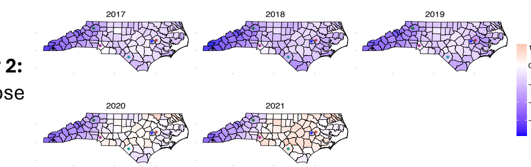
Results

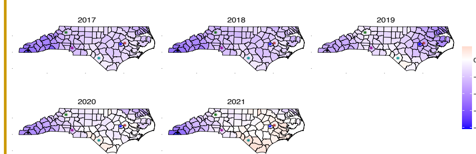
Posterior mean estimates of the four latent factors from 2017 to 2021. **Brown** star represents Clay county, **blue** star represents Green county, **dark green** star represents Mecklenburg county, **dark orange** star represents Pitt county, **dark magenta** star represents Robeson county **dark cyan** star represents Wilkes county.



Factor 1: Burden

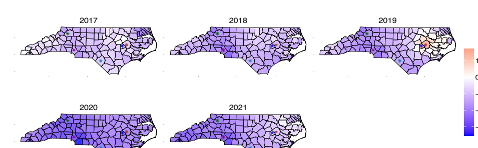
Factor 2: Overdose





Factor 3: Treatment

Factor 4: Infectious diseases



Next steps

- Combine county-level data with zip code-level data to understand the opioid epidemic at a more granular level.
- Use loadings that vary geographically.

References

1. T.W. Anderson. An introduction to multivariate statistical analysis. Wiley, Chichester, 3rd Edition, 2003.
2. D. Kline, L. A. Waller, E. McKnight, A. Bonny, W. C. Miller, and S. A. Hepler. A dynamic spatial factor model to describe the opioid syndemic in Ohio. Biometrics, 2021.
3. H. R. P. Ferguson, D. H. Bailey, and S. Arno. Analysis of PSLQ, and integer relation finding algorithm. Mathematics of Computation, 1999.
4. J. Bai and P. Wang. Identification and Bayesian estimation of dynamic factor models. Journal of Econometrics, 2016.

Funding and Acknowledgements:

- This work was supported by the National Institute On Drug Abuse of the National Institutes of Health under Award Number R01DA052214 and National.
- Computations were performed using the Wake Forest University (WFU) High Performance Computing Facility.

* Wake Forest University **Wake Forest University School of Medicine + UNC's Gillings School of Global Health ▲ Emory University