THE SPREAD OF EVIDENCE-POOR MEDICINE VIA FLAWED SOCIAL-NETWORK ANALYSIS

Russell Lyons
Department of Mathematics,
Indiana University

Presented by:
Nafisa Afrin Chowdhury
Introduction

- “The misuse of statistics is usually inadvertent, not intentional”
- Critics on papers by Christakis and Fowler (2007)
  - Misuse of statistical methods
  - Misinterpretation of statistical results
Framingham Heart Study

- 12000 participants studied in 32 years
- Residents of Framingham, MA
- Three generation of cohorts
  - Original, Offspring, third generation
- Friends, family and co-workers
- Research papers by Christakis and Fowler (C&F)
  - Obesity, smoking cessation, happiness, and loneliness
  - All C&F papers use similar methods
    - Studying only the Obesity one here
Papers by C&F

Two major inferences:

• There is a process of infection or contagion within this social network that transmits various personal characteristics, including obesity, smoking cessation, happiness, and loneliness.

• Such transmission occurs up to three steps in the network, providing evidence of a universal ‘three degrees of influence’ rule of social network contagion.
Rewards of the obesity paper

- Reported on the front page of “New York times”
- Both authors were named in “Top 100 Global Thinkers” by Foreign Policy Magazine
- Their book (2009) translated in 20 different languages
- Awarded $11,000,000 grant from National Institute of Aging

Despite all these accolades...
The critics claim that
- Both of their claims are unfounded!
- Parts of their studies even proved untrue!
C&F Findings and Lyon’s Criticism

• “Chances of spreading obesity is 57% in ego perceived network and 13% in alter perceived network”
  • In fact, there is no statistical difference found in “57%” and “13%”

• “Addition of lagged obesity term in their model controls homophily”
  • In fact, rather than subtracting the effect of homophily it amplifies it.

• “Transmission persists in three degree of friendship separation, in not in geographic separation”
  • This suggests that homophily is there and, in fact, playing a major role.
Points of criticism

- Directionality
- Random Networks
- Modeling
- The Role of Review
Directionality

- C&F argued about perceived directional effects in friendship
  - Each participant were asked to name his closest friend
  - Focal participant (FP) or ego perceived friendship
  - Linked participant (LP) or alter perceived friendship
  - Some LPs are also FPs, so LP is relative to FP
  - Some ties were omitted
    - Ties between two LPs
    - $\frac{1}{4}$ participant named somebody outside the study
Flaws in Directionality

• C&F claims that
  • “Causality is the best explanation for the directional effect of friendship”
• Three flaws in this claim
  1. The difference is not statistically significant
  2. This difference still shows homophily, not only causality
  3. This difference is consistent with all three possibilities: homophily, confounding and causality
Directionality: **Not statistically significant**

- C&F showed 95% confidence level for
  - FP->LP obesity risk is 57% (CI 6% to 123%)
  - LP->FP obesity risk is 13% (CI -28% to 68%)
- However, 13% itself fall in between 6% to 123% and 57% also fall in between -28% to 68%

Fig. 1: Probability that an ego will become obese according to the types of relationship with alter.
Three possibilities of clustering obese people

- C&F considered three possibilities
- Homophily
  - People tend to associate with other like themselves. Also called selection.
- Confounding
  - People sharing similar environment. Also called contextual influence.
- Induction
  - The transmission from one to another. Also called influence or endogenous social effect.
Directionality: Homophily still exists

- C&F claims
  - “Our models account for homophily by including a time-lagged measurement of the alters' obesity.”
- FP’s current obesity = LP’s lagged obesity + LP’s current obesity + other Variables
- LP’s current obesity measures the “effect”
- LP’s lagged obesity “controls” the homophily
- LP’s lagged obesity + LP’s current obesity ≈ 0
- Coefficients have opposite signs!
- It is amplifying the Homophily!
Directionality: **Consistent with all three explanations**

- Three explanations
  - Induction, Homophily and environment
- Everyone named only one friend
- General phenomenon depends on person’s choice
  - If there was homophily
    - He named a person with similar interest
  - If there was shared environment
    - He named his closest neighbour
Directionality: Consistent with all three explanations

Fig. 2: 100 random locations on a disc, each pointing to its nearest neighbor.

- Distance = Degree of homophily Or arises from location
- Either way it can be mathematically proven that “distance between mutual friends is stochastically smaller than the distance between non-mutual friends.”
Random Network

- C&F preserved the actual network
  - But randomly redistribute obesity
- C&F comparing coefficients from different models: original model and random model
- They cannot estimate the difference between these coefficients.
- Data used in C&F is incomplete
  - Every person named only one friend
  - In reality one person has more than one
Random Network

- C&F reported that
  - The associations in their network are unrelated to geographic distance.

- In fact, this is the evidence that the associations of obesity are due to homophily, more than shared environment and unlikely due to induction.
Random Network

- C&F reported that
  - “The risk of obesity among alters who were directly connected to an obese ego was about 45% higher than the observed network than in the random network.”

- Calculating confidence interval on a random network is wrong!
  - CI definition has restrictions
    - Random sampling of true population
Modeling

- C&F aim to reveal causation by technical means.
- Critics also tried to mimic C&F’s modeling assumptions

\[
\log \frac{P[Y_{i,t} = 1 | C_{i,t}]}{P[Y_{i,t} = 0 | C_{i,t}]} = \alpha + \left( \beta_1 Y_{j,t} + \beta_2 Y_{j,t-1} \right) + \beta_3 Y_{i,t-1} \\
+ \sum_{n=3}^{7} \gamma_n W_n(t) + \delta_1 A_{i,t} + \delta_2 F_i + \delta_3 E_{i,t}.
\]

- \( Y_{i,t} \) = obesity status (0,1)
- \( C_{i,t} \) = Collection of indicator variables
- \( Y_{j,s} \) = status for another individual \((i,t) \neq (j,s)\)
- C&Fs’ main interest to estimate
  - \( \beta_1 \), the ‘effect’
  - \( \beta_2 \) the ‘control’
Modeling

- C&F found that
  - FP <-> LP $\beta_1 = 1.19$, SE = 0.33
  - Inferred increase in risk 171% (95% CI [59%, 326%])
- If i named only one j at time t, $\left(\beta_1 Y_{j,t} + \beta_2 Y_{j,t-1}\right)$, must be same and $\beta_1 = \beta_2 = 0$
- Means no individual effect on anyone on any time!
- Their model contradicts the data!
Conclusion

1. The data are not available to others.
2. The unavailable data are sparse to friendships.
3. The model used to analyze the sparse data contradict the data and the conclusions.
4. The statistical significance test from the questionable estimates do not show proposed differences.
5. The wrongly proposed differences do not distinguish homophily, confounding and induction.
6. Associations at a distant are better explained by homophily than by induction.

• How did these errors arise and pass the inspection?