Causal Effects of Mental Health Conditions on Food Insecurity and the Role of SNAP

Helen H. Jensen
Iowa State University

Kimberly A. Greder
Iowa State University

Brent E. Kreider
Iowa State University

John V. Pepper
University of Virginia

Oleksandr Zhylyevskyy
Iowa State University

MVEA 2018
Memphis, TN
November 2, 2018

Motivation

Mental health conditions affect many individuals:

- 18% of U.S. adults have suffered from mental illness

Studies find associations between:

- Depressive symptoms and food insecurity
- Mother's mental health problems and food insecurity

No research on **causal** impact of mental health on food security that accounts for:

- Misreporting of mental health status
- Role of SNAP in effects of mental illness

Methodological Challenge

Identifying **causal** effect of mental health is difficult:

- **Endogeneity**: same unobservables affect food security and mental health
  - OLS produces **inconsistent** estimates
- **Mismeasurement**: mental disorders are often misdiagnosed, survey instruments have flaws, underreporting due to stigma
  - Treatment variables are binary → measurement error is **nonclassical**
  - IV methods produce **inconsistent** estimates

Assessing whether SNAP mitigates effect of mental illness on food security is challenging because SNAP participation is **endogenous** and **misreported**

We develop partial identification methodology to quantify **joint effect** of **two** potentially mismeasured, endogenous treatments—mental illness and SNAP participation— on food security

National Health Interview Survey (**NHIS**)

- CDC's main source of info on health of U.S. civilian population
- Cross-sectional, nationally representative, 80% response rate
- Annual sample of 35,000 households containing 87,500 individuals

Core components of NHIS questionnaire:

- **Household**: demographics, geocodes (restricted access)
- **Family**: demographics, **food security**, program participation, health, injuries, healthcare use, health insurance
- **Sample adult** (one randomly selected adult per family): **psychological distress**, **mental health problems**, other aspects of health, healthcare services, health behaviors
- **Sample child** (one randomly selected child per family): health, healthcare services, health behaviors

NHIS also provides income measures

Analytical Sample

We pool linked sample adult–family records, NHIS 2011–2014:

- Sample adult is **18–64** years old (working age)
- Every family member is U.S. citizen
- Income ≤ **130% of poverty** (gross income cutoff for SNAP)
- ***N* = 21,520**

Selected sample characteristics (weighted):

Variable	Mean	(Std.Dev.)
SNAP participation (indicator)	0.485	(0.500)
Income-to-poverty ratio	0.689	(0.372)
Child (age < 18) present	0.355	(0.479)
Sample adult's **age** (years)	37.05	(14.32)
Sample adult is **male**	0.436	(0.496)
Residence in **large metro** area	0.461	(0.498)

Food Security (FS) Indicators

NHIS includes **10-item** FS survey module:

- Referenced to **last 30 days**
- Family- and adult-specific questions; no child questions

We create two indicators of family's FS status:

1) **Food secure**: 1 if raw score ≤ 2 (high or marginal FS)
2) **Not very low food secure**: 1 if score ≤ 5 (absence of very low FS)

Descriptive statistics (weighted):

Indicator	Mean	(Std.Dev.)
Family is **food secure**	0.677	(0.468)
Family is **not very low food secure**	0.831	(0.375)

	SNAP subsample	Non-SNAP subsample	Difference
Food secure	0.574	0.775	-0.201***

Indicators of Psychological Distress

NHIS administers 6 questions underlying **Kessler (K-6) psycho-logical distress scale**:

- How frequently in past 30 days you felt sad, nervous, restless, hopeless, that everything was an effort, worthless (5-point Likert answer scale)
- K-6 is standardized and validated measure of **nonspecific psychological distress** (CDC, 2013)

We follow McMorrow et al. (2016) and create indicators for:

1) Sample adult in **severe distress**: 1 if K-6 scale ≥ 13 (max is 24)
2) Sample adult in **moderate or severe distress**: 1 if K-6 scale ≥ 8

Descriptive statistics (weighted):

Indicator	Mean	(Std.Dev.)
Adult is in **severe distress**	0.097	(0.296)
Adult is in **moderate/severe distress**	0.226	(0.418)

Motivation for Our Methodology

Parametric approach:

$$\underbrace{Y_i}_{\text{Outcome}} = \gamma \cdot \underbrace{S_i}_{\text{Treatment}} + \underbrace{x_i}_{\text{Covariates}} \cdot \beta + \underbrace{\varepsilon_i}_{\text{Error term}}$$

S_i is **binary**. For example, $S_i = 1$ if i is in distress, 0 if not

If same unobservables affect S_i and Y_i, then $cov(S_i, \varepsilon_i) \neq 0$ and OLS is inconsistent due to **endogeneity**

Measurement error in S_i is **nonclassical** → IV estimator is inconsistent too

Our **nonparametric bounding** methodology handles endogeneity and misreporting. We also develop methods to handle multiple treatments (not just one binary S_i): e.g., treatment = {in distress, on SNAP}

Our Approach: Setup

H^* = 1 if adult is truly in distress, 0 otherwise; H is self-reported measure of H^*

We assess **average treatment effect** (**ATE**) of distress on food security:

$$ATE(1,0 \mid X) = P[Y(H^* = 1) = 1 \mid X] - P[Y(H^* = 0) = 1 \mid X]$$

Y = 1: family is food secure, Y = 0: insecure

$Y(H^*$ = 1) indicates **potential** food security outcome if adult were to be in distress. $Y(H^*$ = 0) denotes potential outcome if adult were not in distress

X specifies subpopulation of interest. Say, families with income ≤ 130% of poverty, comprised of U.S. citizens, sample adult aged 18–64

Not a regression framework: X are not regressors, no regression error term here, no orthogonality conditions to satisfy

Decomposition Strategy

ATE cannot be point-identified without assumptions **even if** $H^* = H$

We decompose every formula into what is and isn't identified

Simplify notation: $ATE = P[Y(1)=1] - P[Y(0)=1]$

Consider decomposition:

$$P[Y(1)=1] = \underbrace{P[Y(1)=1 \mid H^*=1]}_{\text{identified}} \underbrace{P(H^*=1)}_{\text{identified}} + \underbrace{P[Y(1)=1 \mid H^*=0]}_{\text{not identified}} \underbrace{P(H^*=0)}_{\text{identified}}$$

Data cannot identify $P[Y(1)=1 \mid H^*=0]$ because it refers to unobserved **counterfactual**. We only know $P[Y(1)=1 \mid H^*=0] \in [0,1]$

Using methods of Manski (1995), we can still find worst-case bounds for $P[Y(1)=1]$, $P[Y(0)=1]$, and ATE

Addressing Misreporting

$$P[Y(1)=1] = P(Y=1, H^*=1) + P[Y(1)=1 \mid H^*=0]P(H^*=0)$$

$$= P(Y=1, H=1) + \theta_1^- - \theta_1^+ + P[Y(1)=1 \mid H^*=0]P(H^*=0)$$

$$\theta_1^- \equiv P(Y=1, H=0, H^*=1),\ \ \theta_1^+ \equiv P(Y=1, H=1, H^*=0)$$

Sharp **bounds** on ATE:

$$P(Y=1, H=1) - P(Y=1, H=0) - P^* + 2(\theta_1^- - \theta_1^+)$$

$$\leq ATE \leq$$

$$P(Y=1, H=1) - P(Y=1, H=0) + (1-P^*) + 2(\theta_1^- - \theta_1^+)$$

where $P^* \equiv P(H^*=1)$

Tightening Bounds

Without assumptions, ATE bounds are “too” **wide**

To tighten them, we can:

- Use logical constraints on probabilities and auxiliary (validation) data to restrict θ’s
- Apply “**no false positive**” assumption $\rightarrow$ $\theta_1^+ = \theta_0^+ = 0$
- Impose restrictions on selection process:
  - Monotone treatment selection (**MTS**)
  - Monotone instrumental variable (**MIV**)
  - Monotone treatment response (**MTR**)

By layering assumptions we show how they shape inference

Bounds under Endogenous Selection

		Self-reported prevalence rate: $P^* = P = 0.235$			10% Underreporting of true prevalence rate: $P^* = 1.1P = 0.258$		
Endogenous selection		LB	UB	width	LB	UB	width
(a) Arbitrary errors	p.e.	[-0.912,	0.558]	1.469	[-0.935,	0.581]	1.516
	CI	[-0.919	0.567]		[-0.942	0.590]	
(b) No false positives	p.e.	[-0.710,	0.290]	1.000	[-0.734,	0.313]	1.047
	CI	[-0.716	0.296]		[-0.739	0.319]	

Point estimates of the population bounds.

Imbens-Manski 95% confidence intervals around the true ATE.

Monotonicity Assumptions

Monotone treatment selection (**MTS**):

$$P[Y(j)=1 \mid H^{*}=1] \leq P[Y(j)=1 \mid H^{*}=0],\ j=0,1$$

Monotone instrumental variable (**MIV**):

Let *v* be income-to-poverty ratio. Higher *v* wouldn't harm food security:

$$u_1 \leq u \leq u_2 \Rightarrow$$

$$P[Y(j)=1 \mid v=u_1] \leq P[Y(j)=1 \mid v=u] \leq P[Y(j)=1 \mid v=u_2]$$

Monotone treatment response (**MTR**):

Poor mental health would not improve food security on average:

$$P[Y(1)=1 \mid H^{*}=h] \leq P[Y(0)=1 \mid H^{*}=h],\ h=0,1$$

Bounds under MTS+MIV+MTR

		Self-reported prevalence rate: $P^* = P = 0.235$			10% Underreporting of true prevalence rate: $P^* = 1.1P = 0.258$		
MTS + Food Density MIV + MTR:		LB	UB	width	LB	UB	width
(a) Arbitrary errors	p.e.	**[-0.852,**	**-0.142]**	0.710	**[-0.855,**	**-0.142]**	0.713
	CI	[-0.894	-0.054]		[-0.922	-0.054]	
(b) No false positives	p.e.	**[-0.224,**	**-0.142]**	0.083	**[-0.292,**	**-0.142]**	0.150
	CI	[-0.340	-0.052]		[-0.401	-0.054]	

Strictly negative ATEs are in **bold**
Estimates of population bounds are corrected for finite sample bias
CI: Imbens-Manski 95% confidence intervals around true ATE

Thank you!

Appendix

Indicators of Mental Health Problems

NHIS asks sample adults about degree of **difficulty** with 12 daily activities (e.g., walking) and what health problem caused this

NHIS also asks whether adults are **limited** in performing 7 activities (e.g., personal care) and what health problem caused this

We create indicators for existence of:

1) **Mental health problem causing difficulty** with activities
2) **Mental health problem causing limitation** in activities

'Problem' includes depression, anxiety, ADD, bipolar, schizophrenia, etc.

Selected descriptive statistics (weighted):

Indicator	Mean	(Std.Dev.)
Adult has mental health problem causing **difficulty**	0.069	(0.253)
Adult has mental health problem causing **limitation**	0.083	(0.275)

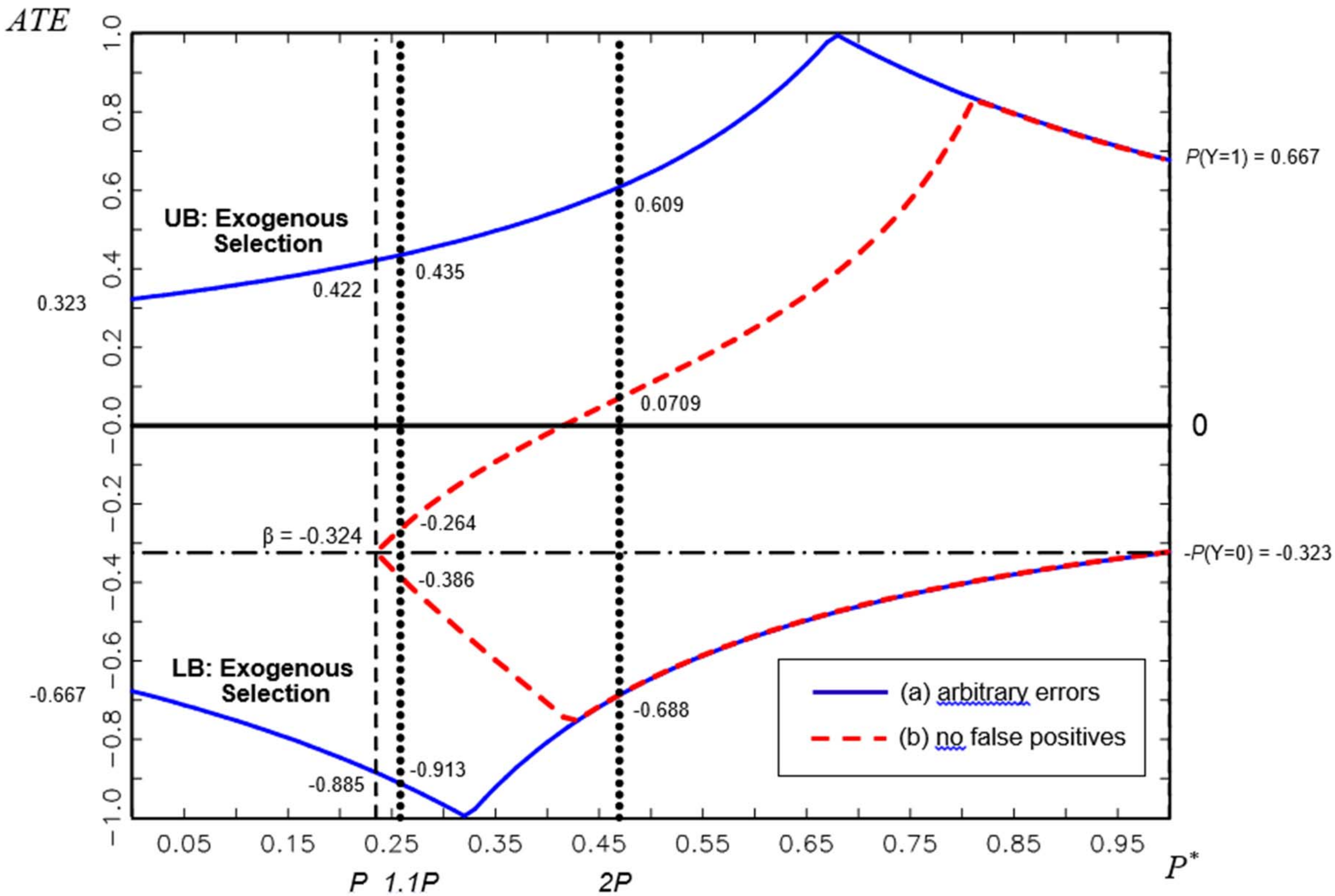
Food Security on Subsamples

Prevalence of food security (%, weighted) in subsamples by moderate/severe distress and SNAP participation:

		SNAP participation		(SNAP=Yes,·) – (SNAP=No,·)
		No	Yes	
Moderate or severe distress	No	83.03	65.46	-17.56
	Yes	49.43	39.11	-10.32
(Distress=Yes,·) – (Distress=No,·)		-33.60	-26.35	

Also, distress, mental health problem indicators are **positively associated** with SNAP participation

Bounds under Exogenous Selection (I)



Bounds under Exogenous Selection (II)

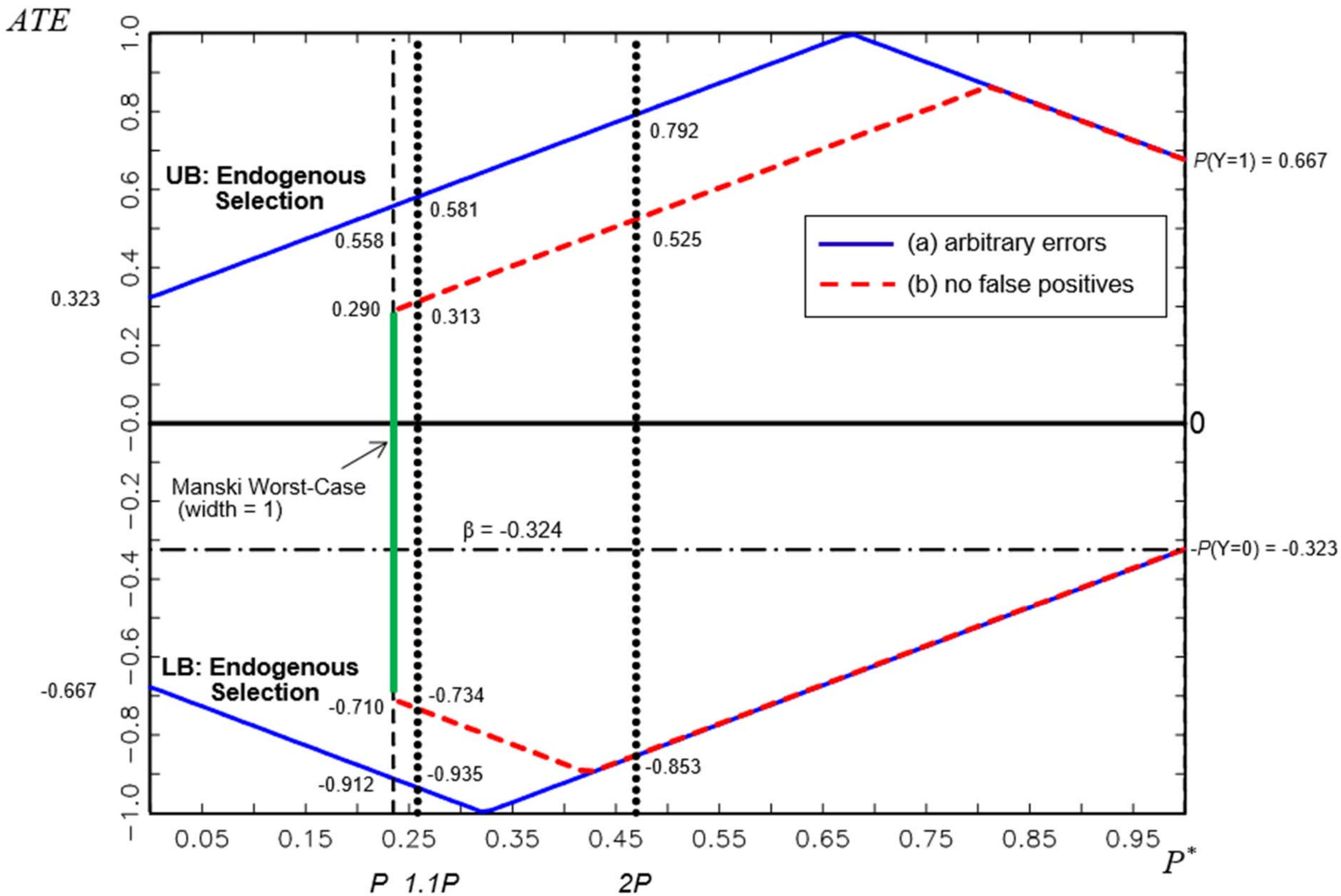
		Self-reported prevalence rate: $P^* = P = 0.235$			10% Underreporting of true prevalence rate: $P^* = 1.1P = 0.258$		
Exogenous selection		LB	UB	width	LB	UB	width
(a) Arbitrary errors	p.e.†	[-0.885,	0.422]	1.307	[-0.913,	0.435]	1.348
	CI‡	[-0.894	0.431]		[-0.922	0.445]	
(b) No false positives	p.e.	**[-0.324,**	**-0.324]**	0.000	[-0.386,	-0.264]	0.123
	CI	[-0.341	-0.308]		[-0.400	-0.250]	

Strictly negative average treatment effects in **bold**.

† Point estimates of the population bounds.

‡ Imbens-Manski 95% confidence intervals around the true ATE.

Bounds under Endog. Selection: Graph



Income MIV + Other Assumptions

		Self-reported prevalence rate: $P^* = P = 0.235$			10% Underreporting of true prevalence rate: $P^* = 1.1P = 0.258$		
MTS + Income MIV:		LB	UB	width	LB	UB	width
(a) Arbitrary errors	p.e.	[-0.851,	0.476]	1.328	[-0.878,	0.500]	1.379
	CI	[-0.878	0.507]		[-0.901	0.530]	
(b) No false positives	p.e.	[-0.210,	0.189]	0.400	[-0.290,	0.213]	0.503
	CI	[-0.266	0.229]		[-0.331	0.252]	
MTS + Income MIV + MTR:		LB	UB	width	LB	UB	width
(a) Arbitrary errors	p.e.	**[-0.851,**	**-0.0956]**	0.756	**[-0.878,**	**-0.0956]**	0.783
	CI	[-0.878	-0.0649]		[-0.901	-0.0649]	
(b) No false positives	p.e.	**[-0.210,**	**-0.0956]**	0.115	**[-0.290,**	**-0.0956]**	0.194
	CI	[-0.266	-0.0649]		[-0.331	-0.0649]	