SUPPLEMENTAL METHODS

Missing Values

Less than 5% of covariates were missing. Complete case analysis is unlikely to introduce bias at such low levels of missingness, so imputation of missing values was not conducted.[1–4] Moreover, we did not impute missing values for outcome variables, as this may add noise to resulting estimates.[5]

Instrumental Variables Analysis

The IV analysis followed a standard two-stage least-squares approach and is represented by the following two equations for each individual i in year t:

$$Inc_{it} = \alpha_0 + \alpha_1 EITC_{it} + \alpha_2 X_{it} + \theta_t + \lambda_i + \varepsilon_{it} \quad (1)$$

$$Health_{it} = \beta_0 + \beta_1 \widehat{Inc}_{it} + \beta_2 X_{it} + \theta_t + \lambda_i + \upsilon_{it} \quad (2)$$

Equation (1) represents the first stage of the IV analysis, where post-tax household income (the endogenous variable) was regressed on EITC refund size (the instrument), adjusting for individual-level characteristics X_{it}, individual-level fixed effects λ_i, and year fixed effects θ_t. A predicted value for post-tax income was produced for each person-year observation, which was then used in Equation (2), the second stage of the IV analysis. Here, $Health_{it}$ represents each of the outcomes of interest. Robust standard errors ε were again clustered at the individual level to account for correlated observations. The coefficient of interest was β_1, representing the effect of income on each outcome. In practice, this analysis is carried out with a single command in Stata, *ivreg2*.

In the first stage of the IV analysis, every $1,000 increase in EITC refund size was associated with an increase in post-tax household income of $521 (95%CI: 435.92, 605.57). The *F*-statistic

Shields-Zeeman L, *et al. J Epidemiol Community Health* 2021; 75:929–935. doi: 10.1136/jech-2020-214841

for the first stage was above 10 for all outcomes (range: 10-135), a rule of thumb indicating that EITC refund size is a strong instrument for post-tax income.

Use of Fixed Effect Models

A key component of our analysis, which contributes to a stronger causal model, is the inclusion of individual fixed effects (i.e., indicator variables). This in essence results in a within-person rather than across-person comparison. Logistic models with fixed effects failed to converge due to the large number of covariates. Sacrificing the fixed effects in the models in order to accommodate logistic regressions would come at a substantial loss to the strength of causal inference, especially when prior work has shown that the statistical properties of linear regression for binary outcomes are less problematic with large samples like ours.[6] In other disciplines in which fixed effects are similarly used to strengthen causal inference—e.g., economics and political science—the use of linear models in this way for large samples is accepted. Also, it is well established that linear probability models (i.e., OLS models with binary outcomes) and logistic models tend to produce very similar estimates of average marginal effects.[7]

There is also a more technical problem with the use of logistic models with fixed effects even if the models had converged.[8] It is not possible to run an "unconditional" maximum likelihood estimation with dummy variables, analogous to a linear model because of the "incidental parameters problem"[8,9] in which the number of dummies increases directly with the sample size, violating one of the conditions that underlie asymptotic theory of maximum likelihood estimation. The solution is conditional maximum likelihood;[10] however, this procedure drops all units (individuals) that have no variation in exposure over time, and this would automatically exclude a large number of observations.

Shields-Zeeman L, *et al. J Epidemiol Community Health* 2021; 75:929–935. doi: 10.1136/jech-2020-214841

Additional Analyses

The EITC policy is not a linear combination of age, marital status and other variables; rather, it is a complex non-linear formula used by the U.S. Internal Revenue Service (IRS), so these variables are not perfectly collinear. Nevertheless, we have carried out an additional analysis that does not adjust for variables that are inputs to the EITC eligibility formula (age, number of children, and marital status), as has been conducted in prior work.[11,12] Results for this analysis were similar to our primary models (Supplemental Table 2).

Shields-Zeeman L, *et al. J Epidemiol Community Health* 2021; 75:929–935. doi: 10.1136/jech-2020-214841

SUPPLEMENTAL REFERENCES

1 Bennett DA. How can I deal with missing data in my study? *Aust N Z J Public Health* 2001;**25**:464–9.

2 Dong Y, Peng CYJ. Principled missing data methods for researchers. *Springerplus* 2013;**2**:1–17.

3 Langkamp DL, Lehman A, Lemeshow S. Techniques for handling missing data in secondary analyses of large surveys. *Acad Pediatr* 2010;**10**:2015–2010.

4 Millsap R, Maydeu-Olivares A. *The SAGE Handbook of Quantitative Methods in Psychology*. SAGE Publications Ltd 2012.

5 Von Hippel PT. Regression with missing Ys: An improved strategy for analyzing multiply imputed data. *Sociol Methodol* 2007;**87**:83–117.

6 Hellevik O. Linear versus logistic regression when the dependent variable is a dichotomy. *Qual Quant* 2009;**43**:59–74.

7 Angrist J., Pischke J-S. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press 2008.

8 Allison P. *Fixed Effects Regression Models*. SAGE Publications 2009.

9 Kalbfleisch JD, Sprott DA. Application of Likelihood Methods to Models Involving Large Numbers of Parameters. *J R Stat Soc Ser B* 1970;**32**:175–94.

10 Chamberlain G. Analysis of Covariance with Qualitative Data. *NBER Work Pap Ser* 1979.

11 Hamad R, Collin DF, Rehkopf DH. Estimating the Short-Term Effects of the Earned Income Tax Credit on Child Health. *Am J Epidemiol* 2018;**187**:2633–41.

12 Rehkopf DH, Strully KW, Dow WH. The short-term impacts of earned income tax credit disbursement on health. *Int J Epidemiol* 2014;**43**:1884–94.

Shields-Zeeman L, *et al. J Epidemiol Community Health* 2021; 75:929–935. doi: 10.1136/jech-2020-214841

Supplemental Table 1. Available Outcomes by Survey Year, Panel Study of Income Dynamics

Outcomes	Survey Years Available
Psychological distress	2001, 2003, 2007-2015 (biennial)
Very good/excellent overall health	1985-1997 (annual), 1999-2015 (biennial)
Currently drink alcohol	1999-2015 (biennial)
3+ drinks per day	1999-2015 (biennial)
Currently smoke	1999-2015 (biennial)
Cigarettes per day	1999-2015 (biennial)

Supplemental Table 2. Association of income and the earned income tax credit with mental health and health behaviors, without adjustment for EITC inputs (age, number of children, marital status)

Outcomes	Model 1 Income	Model 2 EITC	Model 3 IV
Psychological distress	-0.024**	-0.077**	-0.097**
	(-0.031, -0.017)	(-0.12, -0.039)	(-0.15, -0.049)
Very good/excellent health	0.00044*	-0.00090	-0.0030
	(0.000021, 0.00085)	(-0.0035, 0.0017)	(-0.0070, 0.0010)
Currently drink alcohol	0.00092**	0.0021	0.0044
	(0.00044, 0.0014)	(-0.00074, 0.0050)	(-0.0015, 0.010)
3+ drinks per day	0.00039	-0.0018	-0.0037
	(-0.000026, 0.00080)	(-0.0041, 0.00053)	(-0.0084, 0.0011)
Currently smoke	-0.000014	0.00060	0.0012
	(-0.00033, 0.00030)	(-0.0014, 0.0026)	(-0.0029, 0.0054)
Cigarettes per day	0.014**	0.014	0.030
	(0.0083, 0.021)	(-0.021, 0.049)	(-0.043, 0.10)

* $p < 0.05$; ** $p < 0.01$
Study sample was drawn from the Panel Study of Income Dynamics for survey years 1985-2015.
Model 1 represents the OLS model, regressing each outcome on pre-tax inflation-adjusted household income.
Model 2 represents the reduced form of the IV analysis, regressing each outcome on EITC refund size.
Model 3 represents the IV analysis, with EITC refund size as the instrument for post-tax inflation-adjusted income.
All models adjusted for education, household inflation-adjusted pre-tax earned income and income-squared, year fixed effects, and individual fixed effects. Robust standard errors were clustered at the individual level.
EITC: earned income tax credit; IV: instrumental variable; OLS: ordinary least squares.