



HHS Public Access

Author manuscript

Ann Epidemiol. Author manuscript; available in PMC 2022 August 01.

Published in final edited form as:

Ann Epidemiol. 2021 August ; 60: 1–7. doi:10.1016/j.annepidem.2021.04.011.

Inverse probability weighting for selection bias in a Delaware community health center electronic medical record study of community deprivation and hepatitis C prevalence

Neal D. Goldstein^a, Deborah Kahal^{b,c}, Karla Testa^{c,d}, Igor Burstyn^{a,e}

^aDrexel University Dornsife School of Public Health, Department of Epidemiology and Biostatistics, Philadelphia, PA, USA

^bWilliam J. Holloway Community Program, ChristianaCare, Wilmington, DE, USA

^cSydney Kimmel College of Medicine, Thomas Jefferson University, Philadelphia, PA, USA

^dWestside Family Healthcare, Wilmington, DE, USA

^eDrexel University Dornsife School of Public Health, Department of Environmental and Occupational Health, Philadelphia, PA, USA

Abstract

Purpose: To demonstrate how selection into a healthcare facility can induce bias in an electronic medical record-based study of community deprivation and chronic hepatitis C virus infection, in order to more accurately identify local risk factors and prevalence.

Methods: We created a catchment model that attempted to define the probability of selection into a retrospective cohort. Then using the inverse of this probability, we compared naïve unweighted and weighted models to demonstrate the impact of selection bias.

Results: ZIP code-level ecological plots of the cohort demonstrated that there was a pattern of the community deprivation, hepatitis C outcome, and distance to the health center (an intuitive proxy for being within catchments). The naïve multilevel analysis found that living in an area with greater deprivation resulted in 1.25 times greater odds of HCV (95% CI: 1.06, 1.48), whereas the weighted analysis found less certainty of this effect due to a selection bias.

Conclusions: We observed that selection into the catchment area of the studied healthcare facility may bias the association of community deprivation and hepatitis C. This may be mitigated through inverse probability weighting.

Keywords

Catchment Area, Health; Residence Characteristics; Electronic Health Records; Hepatitis C, Chronic; Cohort Studies; Selection Bias

*Corresponding Author: Neal D. Goldstein, 3215 Market St., Philadelphia, PA, 19104, United States, 1-267-359-6207, ng338@drexel.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

INTRODUCTION

With the increasing utilization of electronic medical records (EMRs) in outpatient medicine,¹ and corresponding analyses that may inform decisions about resources and services offered, it is imperative that researchers understand the patient population that contributed data into the EMR, i.e. the healthcare facility's catchment area. In particular, there is the potential for selection bias in the design phase of a study if the catchment area is not representative of the underlying population.² Access to care and entrance into a healthcare setting is a complex process exacerbated by the multifaceted payer system in the U.S.,^{3,4} and can result in differential patient population selection that may affect both outcomes and exposures.⁵ Existing evidence suggests marked differences between population-based surveys and health care system-based estimates of disease prevalence^{6,7} and risk factors.^{8,9} The representativeness of EMR data must be thoroughly understood in an effort to minimize bias, especially given the rise in population health surveillance from these data.¹⁰

While this potential for selection bias is known, it has rarely been quantified. Ness et al. examined differential participation in clinic-based care for a cohort of childhood cancer survivors across the United States and Canada between 1994 and 2003.² They observed that cohort participants who visited a long-term follow-up clinic were on average 1.4 times more likely to report a chronic health condition compared to those who were not seen in the clinic. The authors adjusted the endpoints of the reported confidence intervals to account for the differential participation and corresponding uncertainty of the resulting estimates. This approach necessitated individual-level data from an underlying cohort on those individuals who did not participate in the follow-up clinic. In the present study, we took a different perspective supposing that individual-level data on non-participants were not available, as is frequently the case (i.e. those who have not visited a given healthcare facility and thus do not have data in the EMR). We sought to derive a model to predict the probability of visiting a healthcare facility and then use this information to adjust for potential selection bias in an analysis, based only on information from the EMR supplemented with publicly available aggregated data.

A motivating study

Chronic infection with hepatitis C virus (HCV) is a preventable disease affecting 2 – 4 million people in the U.S.¹¹ Despite decreasing HCV prevalence (from both curative treatment and mortality), acute infections are increasing, largely driven by people who inject drugs, which has fueled by the opioid epidemic.¹² Less well understood is how the community influences individual HCV risk. We focused on community socioeconomic context measured via the area deprivation index (ADI) and its relation to HCV diagnosis at an urban federally qualified health center (FQHC). Community healthcare facilities play a crucial role in delivering health care to the medically underserved, amplifying concern about differential selection as compared to the source population.¹³ If patients that have elected to seek care at this FQHC are different with respect to both ADI and HCV compared to patients who have not elected to seek care at this FQHC, there is potential for selection bias

that would effect inference about the effect of ADI on HCV prevalence (Figure 1). This scenario would arise if the ADI for patients *not* seeking care at this FQHC (unobserved) is different from that of the ADI assigned to their community. To ascertain this, we can examine the catchment area of the FQHC locations, and evaluate whether catchment is related to ADI and HCV. To put this more generally, can the catchment area of a community healthcare facility induce a selection bias in community-wide inference of an EMR-based cohort study?

MATERIALS and METHODS

Study setting and population

This study was conducted at a FQHC in New Castle County, Delaware, USA, serving primarily an urban area (the city of Wilmington). Adult patients 18 years were eligible for inclusion if they had 1 visit at one of two FQHC locations – annotated as locations A and B – between November 1, 2016 and June 30, 2019. These dates were selected as identifying “active” patients in the FQHC’s EMR. Patients were geocoded into their respective ZIP codes for analysis: only patients residing in New Castle County were retained for analysis as they represent the target area of the FQHC. Data retrieved from the EMR included age in years, sex, race (white vs. non-white), Hispanic ethnicity, insurance status (non-private vs. private), number of visits in the past two years, days since last visit, and a diagnosis of HCV. Using the ZIP code, we then retrieved U.S. Census data from the 2017 American Community Survey 5-year estimates including population size; median household income in the past year; the number of women 15 – 50 years who had a birth in the past year (as a marker of fertility); percent white, black, or other race; percent employed; and percent with private or government funded health insured.¹⁴ As a proxy for FQHC accessibility, we also obtained the density of bus stops per square mile for each ZIP code from the Delaware open data portal.¹⁵ The FQHC estimates that over 50% of patients regularly use public transportation to access their services.

The ZIP code level ADI was our exposure of interest. This variable was operationalized as a Z-score composite of census indicators of education, employment, income and poverty, and household composition, where a higher score indicates greater deprivation.¹⁶ The primary outcome was a diagnosis of HCV in the EMR (yes vs. no) identified by the following ICD-10-CM codes from the EMR: B18.2 (Chronic viral hepatitis C), B19.20 (Unspecified viral hepatitis C without hepatic coma), B19.21 (Unspecified viral hepatitis C with hepatic coma), and B19.2 (Unspecified viral hepatitis C).

This study was deemed IRB exempt by ChristianaCare (Newark, DE).

Defining the catchment area

The catchment area represents the geography and population from which a healthcare facility draws patients. There are a variety of methods for operationalizing it.¹⁷ Based on availability of data, we used an approach that determined the three distance radii around each of the two FQHC locations that contained 75%, 80%, and 90% of patients (Figure

2), corresponding to the respective distances 7 miles, 8 miles, and 12 miles.¹⁸ Each individual in the cohort either falls within or outside of these three catchment definitions.

Statistical analysis

First, we examined individual as well as ecological ZIP code-level characteristics of the cohort using descriptive statistics. Multiple linear regression models and scatterplots summarized associations among the exposure (ADI), prevalence of the outcome (HCV), and patient distance to the FQHC (an intuitive proxy for being within catchment areas). These models provided insight into the possibility of selection bias via the association of HCV and ADI with distance to FQHC. Student's t-test, Wilcoxon rank-sum test, and Chi-squared test were used to identify individual- and ZIP code-level cohort characteristics associated with the exposure, outcome, and distance measures for possible confounder identification.

Second, we used generalized linear mixed effects regression models to estimate the log-odds of an individual being in the FQHC's catchment area, accounting for the clustering at the ZIP code-level. Three separate catchment models were fit, corresponding to the three distance radii around each of the two FQHC locations that contained 75%, 80%, and 90% of hospital patients (Figure 2). Individual-level characteristics hypothesized to be associated with visiting the FQHC included age at last visit, race, ethnicity, insurance status, and the number of visits in the past two years. Density of bus stops was also included as a ZIP-code level predictor. The predicted probability obtained from these models is equivalent to the propensity score of being in the catchment area.¹⁹ We then took the inverse of the propensity score to serve as the inverse probability weight (IPW) to attempt to account for potential selection bias.²⁰ In our cohort, a higher IPW corresponds to underrepresented individuals visiting the FQHC who will receive a greater weight in analysis compared to lower IPW values indicating overrepresented individuals. IPWs were stabilized by replacing the numerator in the calculation with the proportion of individuals falling within a given catchment definition.²¹ In a sensitivity analysis, stabilized IPWs were trimmed between the 2nd and 98th percentiles.

Third, we used generalized linear mixed effects regression models to estimate the log-odds of an individual having a diagnosis of HCV given their ADI, accounting for the clustering at the ZIP code-level. Potential confounders included age at last visit, sex, race, and insurance status. Four models were fit: a naïve model that did not consider the potential selection bias and served as the basis for comparisons (model 1), and three IPW models that allowed insight into possible selection bias via the three catchment areas previously defined (models 2 to 4).

All analyses were conducted in R version 3.6.1 (R Foundation for Statistical Computing, Vienna, Austria). Measures of central tendency are provided as means or medians, as appropriate, with corresponding standard deviations (SD) or interquartile ranges (IQR) respectively. Point estimates from regression models represent marginal effects presented as odds ratios (OR) with their corresponding confidence intervals (CI). CIs for IPW models were obtained via bootstrapping 1,000 samples.²² Analytic code may be downloaded from <https://doi.org/10.5281/zenodo.4630272>.

RESULTS

Cohort characteristics

Between November 1, 2016 and June 30, 2019, 16,070 patients visited one of two locations at the FQHC: 76% visited location A and 24% visited location B. Of those, 15,574 (97%) lived in New Castle County and were retained for analysis. The median age was 38 years old (IQR: 21 years) and most patients were female (64%). There was roughly an even split between those who identified as Hispanic and those who did not (51% vs. 49%), and a small majority were white compared to non-white (55% vs. 45%). As is expected at an FQHC, the preponderance of patients did not have private insurance (67%). Compared with the ZIP-code level census estimates, individuals with no private insurance were overrepresented in our cohort, with a median 17 percentage points higher across the study area (IQR: 13 percentage points). The median distance patients lived from their visited FQHC location was 3 miles (IQR: 5 miles) and the average ADI was 0.4 (SD: 0.9). An HCV diagnostic code was recorded for 286 patients (2%) of the cohort. Additional characteristics are reported in Table 1.

As distance to the FQHC increased, prevalence of HCV (Figure 3a) as well as ADI score (corresponding to less deprivation, Figure 3b) decreased; see also Table 1. This was corroborated by individual-level modeling. For each additional mile further from an FQHC, the averaged odds of HCV declined by 6% (OR=0.94, 95% CI: 0.91, 0.97) in logistic regression, and the averaged ADI score declined by 0.12 (95% CI: 0.11, 0.12) in linear regression. Figure 3c suggested an ecological association between ADI and prevalence of HCV. Again, individual-level regression corroborated this finding whereby each point increase in ADI (greater deprivation) was associated with an average increased odds of 51% (OR=1.51, 95% CI: 1.33, 1.71) for HCV. Stratum specific individual-level regression estimates by distance to clinic less than or greater than the 75th percentile indicated heterogeneity of this association (OR=2.69, 95% CI: 1.24, 6.93; 8 miles OR=1.44, 95% CI: 1.26, 1.67); 80th and 90th percentiles were similar.

Calculated prevalence of HCV in the cohort trended with distance, regardless of how catchment area was defined, in that prevalence was consistently higher among those within the catchment areas. For those within the distance thresholds, the prevalence proportions of HCV were 2.1%, 2.0%, and 1.9% for the 75th percentile, 80th percentile, and 90th percentile. For those outside of these distance thresholds, the prevalence proportions of HCV were 1.2%, 1.2%, and 0.9%. Taken as a whole, the data suggested a pattern of the exposure, the outcome, and distance to the FQHC.

Catchment area and probability of selection into cohort

Among the three catchment models, only non-white race and the density of bus stops were associated with the three distance-based threshold definitions (Table 2). Averaged across the 75th, 80th, and 90th percentile models, non-white race was associated with 21% decreased odds of being in the catchment radius compared to white race, controlling for age, ethnicity, insurance status, number of visits in the past two years, and density of bus stops. As the radius increased in size, there was a dose-response effect from the density of bus stops.

For example, the 90th percentile model demonstrated an average 5-fold increase in odds (OR=5.40, 95% CI: 1.68, 17.39) of being in the catchment for each unit increase in the density of bus stops, compared with an average 2-fold increase for the 80th percentile model, controlling for all other covariates. In other words, living in a ZIP code with greater bus accessibility translated to greater likelihood of being in the FQHC catchment. Model fit measured via the Akaike information criterion, improved as the catchment area widened.

Naïve analysis of the relation between individual HCV and community ADI

On average, for each ZIP code-level increase in ADI (moving towards an area with greater deprivation) odds of HCV increased by 25% (OR=1.25, 95% CI: 1.06, 1.48) controlling for age, sex, race, and insurance status, and taking into account correlated errors within ZIP codes. All four covariates were also independent predictors of HCV (Table 3).

IPW analysis of the relation between individual HCV and community ADI

Results from the three IPW models are shown in Table 3. While findings for the individual-level characteristics remained consistent with the naïve model, estimates of the primary exposure of ADI changed as a result of the weighting. The point estimate increased, while the CIs became wider in all three catchment models. Results from IPW trimming were consistent with untrimmed IPWs (not shown). These findings suggested that the naïve effect was overly precise due to ignoring at least one source of systematic error. The model fit measured via the Akaike information criterion improved with IPW, supporting model selection with the smallest catchment area definition (75th percentile). In short, although evidence supports that ADI and HCV may be related in the observed cohort, the evidence is weaker for existence of the association in the source population.

DISCUSSION

In this analysis of a retrospective cohort of FQHC patients, we demonstrated how the catchment area may induce a selection bias in an outpatient EMR-based study. Understanding the potential impact of the selection process for this cohort was important because individuals with HCV, regardless of FQHC contact, benefit from HCV diagnosis and treatment.^{11,12} Best estimates of HCV prevalence in the source population suggest approximately 1% seropositivity (95% CI: 0.7%, 1.2%) based on a recent modeling study.²³ We observed that people in the catchment area (low IPW) have an expected higher prevalence of HCV than the community at large. Conversely, people not in the catchment area (high IPW) who are seen at the FQHC have an HCV prevalence similar to that of the general population. Individuals seen at the FQHC (low IPW) seem to be a select group of individuals who most likely will benefit from the mission of the FQHC to serve historically marginalized populations.

For individuals in the catchment area as modeled in our naïve analysis, which ignored possible selection bias, there is evidence of an association between ADI and HCV. This association should be viewed with caution given the weaker association for those outside of the catchment area (higher IPW). Selection bias affected the strength of association and induced apparent heterogeneity of effect in that there appeared to be

some effect modification by factors related to access to care at this FQHC. The IPW analysis that accounted for this phenomenon mitigated this heterogeneity of effect and permitted inference about what happened in the counterfactual population of patients in the community with equitable access to this FQHC.²⁰ This can be contrasted with the naïve analysis that ignored selection bias and the associated apparent heterogeneity of effect of ADI on HCV diagnosis.

Those individuals with HCV who reside outside of the FQHC catchment area yet chose to travel to the clinics may reflect an underlying different risk-group than the greater population in their respective ZIP codes: as we observed, individuals were more likely to have non-private insurance compared to the rest of the community. Our data are unable to offer any insight into the decision to seek care at an FQHC that was further away than other healthcare facilities closer to home. Indeed, there are other community healthcare facilities in our study area where individuals could have sought medical care. One may posit the decision to visit our location was due to comfort and familiarity with the providers or practice, cost, accessibility, or other unidentified reasons.^{24–26} Contrary to the fact, we expected persons who visited the FQHC to have similar prevalence of HCV, elevated relative to community at large, regardless of how far they travel. Therefore, it remains unknown to us whether these individuals represent the “tip of an iceberg” for sub-populations in those areas who also have HCV but did not visit this FQHC, i.e., persons in worse overall health relative to persons of similarly lower socioeconomic status who travel farther to visit FQHC (and incidentally exhibiting higher rates of HCV diagnosis). In other words, patients from outside of the catchment area who attended the FQHC may not be a random sampling of individuals (with similar socioeconomic characteristics) from outside of the catchment area.

This study was conducted during a dynamic period in the history of HCV management. The opioid epidemic has contributed substantially to the burden of HCV and may have influenced who was tested.^{12,27} At the time of our study, risk-factor based screening was used at the FQHC and there were likely individuals in the cohort with undiagnosed HCV. As certain HCV risk factors are likely related to community deprivation,²⁸ the impact of such differential misclassification is difficult to predict. Since study completion, the FQHC implemented universal one-time screening for adult in line with updated U.S. Preventive Services Task Force recommendations.²⁹ After sufficient time, these data would have been potentially useful to validate our diagnosis. With more widespread access to direct acting antiviral therapy, the ability to treat and cure HCV is markedly improved over interferon-based regimens.³⁰ Unlike other locales, in our study setting non-specialists are able to prescribe treatment and there is no urine drug screen requirement.^{31,32} Taken as a whole, this local context may have influenced who was seen at the FQHC compared to the community at large.

Our study is not without additional limitations. First, our models relied upon a simple definition of catchment area: the percentile of patients within a fixed radius of the FQHC. Future catchment analyses would be well served to employ more sophisticated approaches that can incorporate prior evidence of patient referral patterns through a Bayesian framework.¹⁷ Second, we modeled data that were available to us from both the EMR and public data sources to meet our aims. It is possible that residual confounding is

present in both our catchment and HCV models. Study strengths include use of healthcare worker entered-EMR data, a geospatial methodological approach to selection bias, and a large cohort size.

CONCLUSION

In summary, we observed that the FQHC catchment processes may induce selection bias in a study of ADI and HCV. Fortunately, the potential impact of this EMR-based study design issue is remediable through appropriate analytic methodology that allows for more complete presentation of uncertainty in the translation of results to the source population. We encourage other researchers working with EMR-based outpatient data to consider the patient selection process and how that may relate to their exposure and outcome under study.

ACKNOWLEDGEMENTS

Research reported in this publication was supported by the National Institute Of Allergy And Infectious Diseases of the National Institutes of Health under Award Number K01AI143356 (to NDG). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Research was also supported by an award from the ChristianaCare Harrington Value Institute Community Partnership (to DK, NDG). A portion of this work was presented at the 2020 Society for Epidemiologic Research Annual Meeting, virtual conference.

ABBREVIATIONS and ACRONYMS

ADI	Area deprivation index
CI	Confidence interval
EMR	Electronic medical record
FQHC	Federally qualified health center
HCV	Chronic hepatitis C virus infection
IPW	Inverse probability weighting
IQR	Interquartile range
OR	Odds ratio
SD	Standard deviation

References

- [1]. Hsiao CJ, Hing E. Use and characteristics of electronic health record systems among office-based physician practices: United States, 2001–2013. *NCHS Data Brief*. 2014;(143):1–8.
- [2]. Ness KK, Leisenring W, Goodman P, et al. Assessment of selection bias in clinic-based populations of childhood cancer survivors: a report from the childhood cancer survivor study. *Pediatr Blood Cancer*. 2009;52(3):379–86. [PubMed: 18989878]
- [3]. Centers for Disease Control and Prevention (CDC). Vital signs: health insurance coverage and health care utilization --- United States, 2006--2009 and January-March 2010. *MMWR Morb Mortal Wkly Rep*. 2010;59(44):1448–54. [PubMed: 21063276]
- [4]. Miller S, Wherry LR. Health and Access to Care during the First 2 Years of the ACA Medicaid Expansions. *N Engl J Med*. 2017;376(10):947–956. [PubMed: 28273021]

- [5]. Ellenberg JH. Selection bias in observational and experimental studies. *Stat Med*. 1994;13(5-7):557-67. Review. [PubMed: 8023035]
- [6]. Eustache F, Auger J, Cabrol D, et al. Are volunteers delivering semen samples in fertility studies a biased population? *Hum Reprod* 2004;19:2831-2837. [PubMed: 15459173]
- [7]. Uter W, Ludwig A, Balda BR, et al. The prevalence of contact allergy differed between population-based and clinic-based data. *J Clin Epidemiol* 2004;57:627-632. [PubMed: 15246132]
- [8]. Wilfley DE, Pike KM, Dohm FA, et al. Bias in binge eating disorder: How representative are recruited clinic samples? *J Consult Clin Psychol* 2001;69:383-388. [PubMed: 11495168]
- [9]. Bak M, Drukker M, van Os J, et al. Hospital co-morbidity bias and the concept of schizophrenia. *Soc Psychiatry Psychiatr Epidemiol* 2005;40:817-821. [PubMed: 16172812]
- [10]. Kruse CS, Stein A, Thomas H, et al. The use of Electronic Health Records to Support Population Health: A Systematic Review of the Literature. *J Med Syst*. 2018;42(11):214. [PubMed: 30269237]
- [11]. Ly KN, Hughes EM, Jiles RB, et al. Rising mortality associated with hepatitis C virus in the United States, 2003-2013. *Clin Infect Dis*. 2016; 62(10):1287-1288. [PubMed: 26936668]
- [12]. Holtzman D, Asher AK, Schillie S. The Changing Epidemiology of Hepatitis C Virus Infection in the United States During the Years 2010 to 2018. *Am J Public Health*. 2021;111(1):e1-e7.
- [13]. Kaiser Family Foundation. Community Health Centers: Recent Growth and the Role of the ACA. Available at: <https://www.kff.org/medicaid/issue-brief/community-health-centers-recent-growth-and-the-role-of-the-aca/>. Accessed December 17, 2019.
- [14]. U.S. Census Bureau. American Community Survey (ACS). Available at: <https://www.census.gov/programs-surveys/acs/>. Accessed December 17, 2019.
- [15]. State of Delaware. Delaware Bus Stops. Available at: <https://firstmapdelaware.opendata.arcgis.com/datasets/delaware-bus-stops/data>. Accessed December 17, 2019.
- [16]. Messer LC, Laraia BA, Kaufman JS, et al. The development of a standardized neighborhood deprivation index. *J Urban Health*. 2006;83(6):1041-62. [PubMed: 17031568]
- [17]. Wang A, Wheeler DC. Catchment area analysis using bayesian regression modeling. *Cancer Inform*. 2015;14(Suppl 2):71-9.
- [18]. Phibbs CS, Robinson JC. A variable-radius measure of local hospital market structure. *Health Serv Res*. 1993;28(3):313-24. [PubMed: 8344822]
- [19]. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983;70(1):41-55.
- [20]. Cole SR, Hernán MA. Constructing inverse probability weights for marginal structural models. *Am J Epidemiol*. 2008;168(6):656-64. [PubMed: 18682488]
- [21]. Thoemmes F, Ong AD. A Primer on Inverse Probability of Treatment Weighting and Marginal Structural Models. *Emerging Adulthood*. 2016;4(1):40-59.
- [22]. Austin PC. Variance estimation when using inverse probability of treatment weighting (IPTW) with survival analysis. *Stat Med*. 2016;35(30):5642-5655. [PubMed: 27549016]
- [23]. Blach S, Zeuzum S, Manns M, et al. Global prevalence and genotype distribution of hepatitis C virus infection in 2015: a modeling study. *Lancet Gastroenterol Hepatol* 2017; 2:161-76 [PubMed: 28404132]
- [24]. Saultz JW, Albedaiwi W. Interpersonal continuity of care and patient satisfaction: a critical review. *Ann Fam Med*. 2004;2(5):445-51. [PubMed: 15506579]
- [25]. Anthony DL, Herndon MB, Gallagher PM, et al. How much do patients' preferences contribute to resource use? *Health Aff (Millwood)*. 2009;28(3):864-73. [PubMed: 19414899]
- [26]. Reed C, Rabito FA, Werthmann D, et al. Factors associated with using alternative sources of primary care: a cross-sectional study. *BMC Health Serv Res*. 2019;19(1):933. [PubMed: 31801526]
- [27]. Wu T, Konyon PG, Cattaneo AW, Saab S. New Face of Hepatitis C. *Dig Dis Sci*. 2019;64(7):1782-1788. doi: 10.1007/s10620-019-05511-y. Epub 2019 Feb 13. [PubMed: 30756208]

- [28]. Linton SL, Haley DF, Hunter-Jones J, et al. Social causation and neighborhood selection underlie associations of neighborhood factors with illicit drug-using social networks and illicit drug use among adults relocated from public housing. *Soc Sci Med.* 2017;185:81–90. [PubMed: 28554162]
- [29]. US Preventive Services Task Force (USPSTF), Owens DK, Davidson KW, Krist AH, Barry MJ, Cabana M, Caughey AB, Donahue K, Doubeni CA, Epling JW Jr, Kubik M, Ogedegbe G, Pbert L, Silverstein M, Simon MA, Tseng CW, Wong JB. Screening for Hepatitis C Virus Infection in Adolescents and Adults: US Preventive Services Task Force Recommendation Statement. *JAMA.* 2020;323:202032.
- [30]. Manns MP, Wedemeyer H, Cornberg M. Treating viral hepatitis C: efficacy, side effects, and complications. *Gut.* 2006;55(9):1350–9. [PubMed: 16905701]
- [31]. Breskin A, Westreich D, Hurt CB, Cole SR, Hudgens MG, Seaberg EC, Thio CL, Tien PC, Adimora AA. The Effects of Hepatitis C Treatment Eligibility Criteria on All-cause Mortality Among People With Human Immunodeficiency Virus. *Clin Infect Dis.* 2019;69(9):1613–1620. [PubMed: 30615096]
- [32]. Marshall AD, Pawlotsky JM, Lazarus JV, Aghemo A, Dore GJ, Grebely J. The removal of DAA restrictions in Europe - One step closer to eliminating HCV as a major public health threat. *J Hepatol.* 2018;69(5):1188–1196. [PubMed: 29959953]

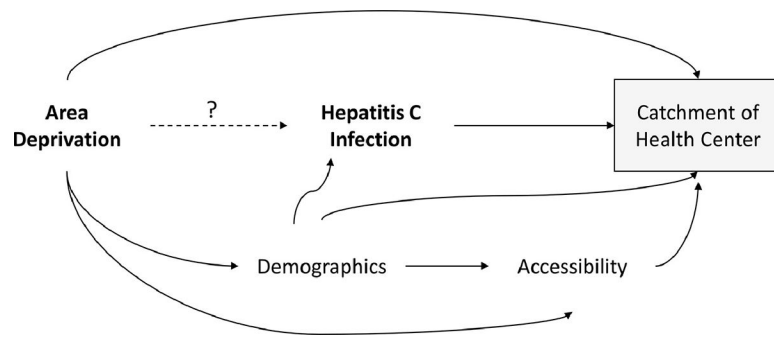
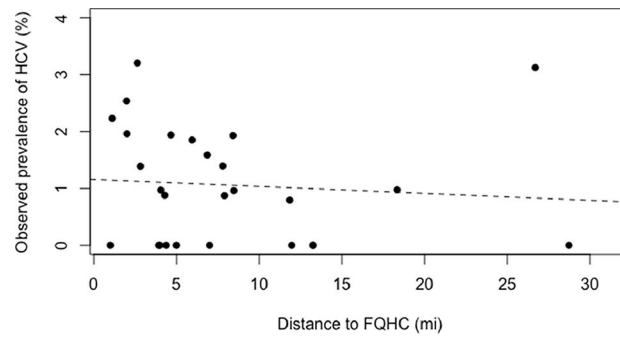


Figure 1.

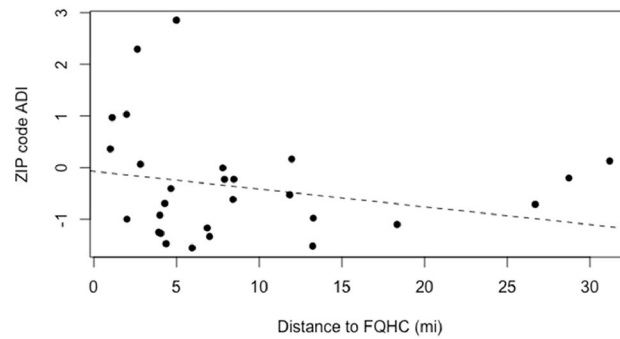
Directed acyclic graph of the primary research question: “Is area deprivation index (ADI) associated with a diagnosis of chronic hepatitis C (HCV) for a federally qualified health center (FQHC) serving New Castle County, Delaware, USA?” The shaded box indicates that we may be inadvertently conditioning the relationship between ADI and HCV on the catchment of the FQHC, inducing selection bias. Demographic factors include age, race (white, non-white), and Hispanic ethnicity. Accessibility factors include insurance (non-private, private), number of previous visits in past 2 years, and access to public transportation (density of bus stops).



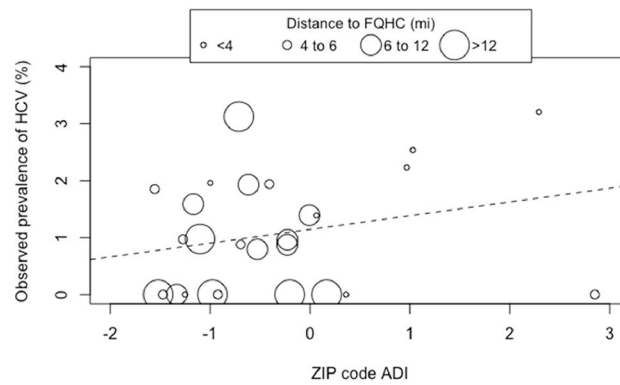
Figure 2. Distance-based threshold catchment areas (75th, 80th, 90th percentiles corresponding to the respective distances 7 miles, 8 miles, and 12 miles) for a federally qualified health center with locations shown in **A**) and **B**) serving New Castle County, Delaware, USA. Gradient represents a choropleth depiction of the density of patients for a given ZIP code (lighter shades of gray=low, darker shades of gray=high).



(a)



(b)



(c)

Figure 3. Ecological plots depicting ZIP code-level associations^a between **A)** median distance to federally qualified health center (FQHC) by observed prevalence of chronic Hepatitis C virus (HCV) infection, **B)** median distance to FQHC by area deprivation index (ADI), and **C)** ADI by observed prevalence of HCV and median distance to FQHC, for a cohort of patients in New Castle County, Delaware, USA. Dashed lines depict line of best fit.

Table 1.

Characteristics of a cohort of patients seen at a federally qualified health center serving New Castle County, Delaware, USA between November 1, 2016 – June 30, 2019, overall and within catchment areas defined by three distance-based thresholds (75%, 80%, and 90% of patients).

<i>Characteristic</i>	Overall (n=15,574)	75th percentile (n=11,683)	80th percentile (n=12,459)	90th percentile (n=14,017)
<i>Individual-level</i>				
Age in yrs, median (IQR)	38 (21)	38 (22)	38 (22)	38 (21)
Sex, n (%)				
Female	9,961 (64%)	7,443 (64%)	7,947 (64%)	8,945 (64%)
Male	5,613 (36%)	4,240 (36%)	4,512 (34%)	5,072 (36%)
Race, n (%)				
White	8,184 (55%)	6,022 (53%)	6,517 (54%)	7,418 (55%)
Non-white	6,819 (45%)	5,307 (47%)	5,557 (46%)	6,131 (45%)
Ethnicity, n (%)				
Non-Hispanic	7,850 (51%)	6,013 (52%)	6,340 (51%)	7,015 (50%)
Hispanic	7,603 (49%)	5,572 (48%)	6,017 (49%)	6,890 (50%)
Insurance, n (%)				
Non-private	6,460 (67%)	5,202 (70%)	5,453 (69%)	5,970 (68%)
Private	3,258 (33%)	2,282 (30%)	2,430 (31%)	2,798 (32%)
Primary health center, n (%)				
Location A	11,897 (76%)	8,923 (76%)	9,518 (76%)	10,708 (76%)
Location B	3,677 (24%)	2,760 (24%)	2,941 (24%)	3,309 (24%)
Number of visits in past 2 years, median (IQR)	3 (6)	3 (6)	3 (6)	3 (6)
Days since last visit, median (IQR)	233 (418)	214 (413)	218 (416)	225 (416)
Diagnosis of chronic Hepatitis C virus infection, n (%)	286 (2%)	241 (2%)	249 (2%)	272 (2%)
Distance to federally qualified health center in mi., median (IQR)	3 (5)	2 (3)	2 (3)	3 (4)
<i>ZIP code-level</i>				
Density of bus stops per sq. mi., median (IQR)	23 (24)	27 (20)	27 (20)	23 (21)
Area deprivation index ^a , mean (SD)	0.4 (0.9)	0.7 (0.9)	0.6 (0.9)	0.5 (0.9)

SD, standard deviation; IQR, interquartile range

^aOperationalized as a Z-score composite of ZIP code-level 2017 American Community Survey 5-year estimates based on indicators of education, employment, income and poverty, and household composition.

Table 2.

Mixed effects regression models of belonging to a catchment area defined by three distance-based thresholds (75%, 80%, and 90% of patients) for a federally qualified health center serving New Castle County, Delaware, USA.

<i>Characteristic</i>	Marginal adjusted odds ratio (95% confidence interval)		
	<i>75th percentile</i>	<i>80th percentile</i>	<i>90th percentile</i>
<i>Individual-level</i>			
Age in years ^a	0.94 (0.87, 1.02)	0.93 (0.86, 1.01)	0.91 (0.83, 1.00)
Race			
White	Ref	Ref	Ref
Non-white	0.83 (0.67, 1.03)	0.76 (0.62, 0.94) ^b	0.77 (0.60, 0.98) ^b
Ethnicity			
Non-Hispanic	Ref	Ref	Ref
Hispanic	0.97 (0.78, 1.20)	0.92 (0.75, 1.15)	0.95 (0.74, 1.24)
Insurance			
Non-private	Ref	Ref	Ref
Private	1.00 (0.84, 1.20)	0.96 (0.81, 1.14)	0.93 (0.76, 1.13)
Number of visits in past 2 years ^a	1.05 (0.97, 1.14)	1.02 (0.93, 1.11)	1.06 (0.93, 1.20)
<i>ZIP code-level</i>			
Density of bus stops per sq. mi.	1.88 (1.41, 2.50) ^b	2.07 (1.31, 3.27) ^b	5.40 (1.68, 17.39) ^b
<i>Measure of model fit</i>			
Akaike information criterion	3,346	3,295	2,457

^aVariable standardized and centered for modeling

^bp<0.05

Table 3.

Naïve and inverse probability weighted mixed effects regression models for the association of area deprivation index and individual chronic Hepatitis C virus infection with catchment area defined by three distance-based thresholds (75%, 80%, and 90% of patients) for a federally qualified health center serving New Castle County, Delaware, USA.

<i>Characteristic</i>	Marginal adjusted odds ratio (95% confidence interval)			
	<i>Naïve model</i>	<i>Inverse probability weighted models</i>		
		<i>75th percentile</i>	<i>80th percentile</i>	<i>90th percentile</i>
<i>Individual-level</i>				
Age ^a	1.69 (1.51, 1.89) ^b	1.68 (1.51, 1.83) ^b	1.73 (1.56, 1.89) ^b	1.68 (1.54, 1.82) ^b
Sex				
Female	Ref	Ref	Ref	Ref
Male	1.81 (1.41, 2.33) ^b	1.83 (1.11, 2.30) ^b	1.65 (1.07, 2.07) ^b	1.70 (1.22, 2.09) ^b
Race				
White	Ref	Ref	Ref	Ref
Non-white	0.74 (0.57, 0.97) ^b	0.63 (0.37, 0.78) ^b	0.70 (0.42, 0.89) ^b	0.70 (0.44, 0.88) ^b
Insurance				
Non-private	Ref	Ref	Ref	Ref
Private	0.43 (0.31, 0.59) ^b	0.35 (0.18, 0.48) ^b	0.38 (0.21, 0.51) ^b	0.42 (0.26, 0.57) ^b
<i>ZIP code-level</i>				
Area deprivation index	1.25 (1.06, 1.48) ^b	1.39 (1.00, 1.56)	1.39 (1.04, 1.59) ^b	1.27 (1.01, 1.44) ^b
<i>Measure of model fit</i>				
Akaike information criterion	2,206	1,556	1,660	1,963

^aVariable standardized and centered for modeling

^bp<0.05

^cOperationalized as a Z-score composite of ZIP code-level 2017 American Community Survey 5-year estimates based on indicators of education, employment, income and poverty, and household composition.