252E Final Project

Relationship between Frequency of Care and Progression to Hepatocellular Carcinoma among Chronic Hepatitis C Patients

Stephanie Holm and Shelley Facente

12/11/2019

Description of our Dataset

We are using a dataset of patients with chronic hepatitis C virus (HCV), receiving care in the UCSF system since 2009.

- n = 1848
- Adults seen in the UCSF system by the end of 2015
- At least one visit in the primary care or hepatology (liver) clinic between 2015 and 2019
- Did not already have HCC at the start of the study period

Roadmap

1. Specify Causal Model

1. Specify Causal Model

Defining Our Variables

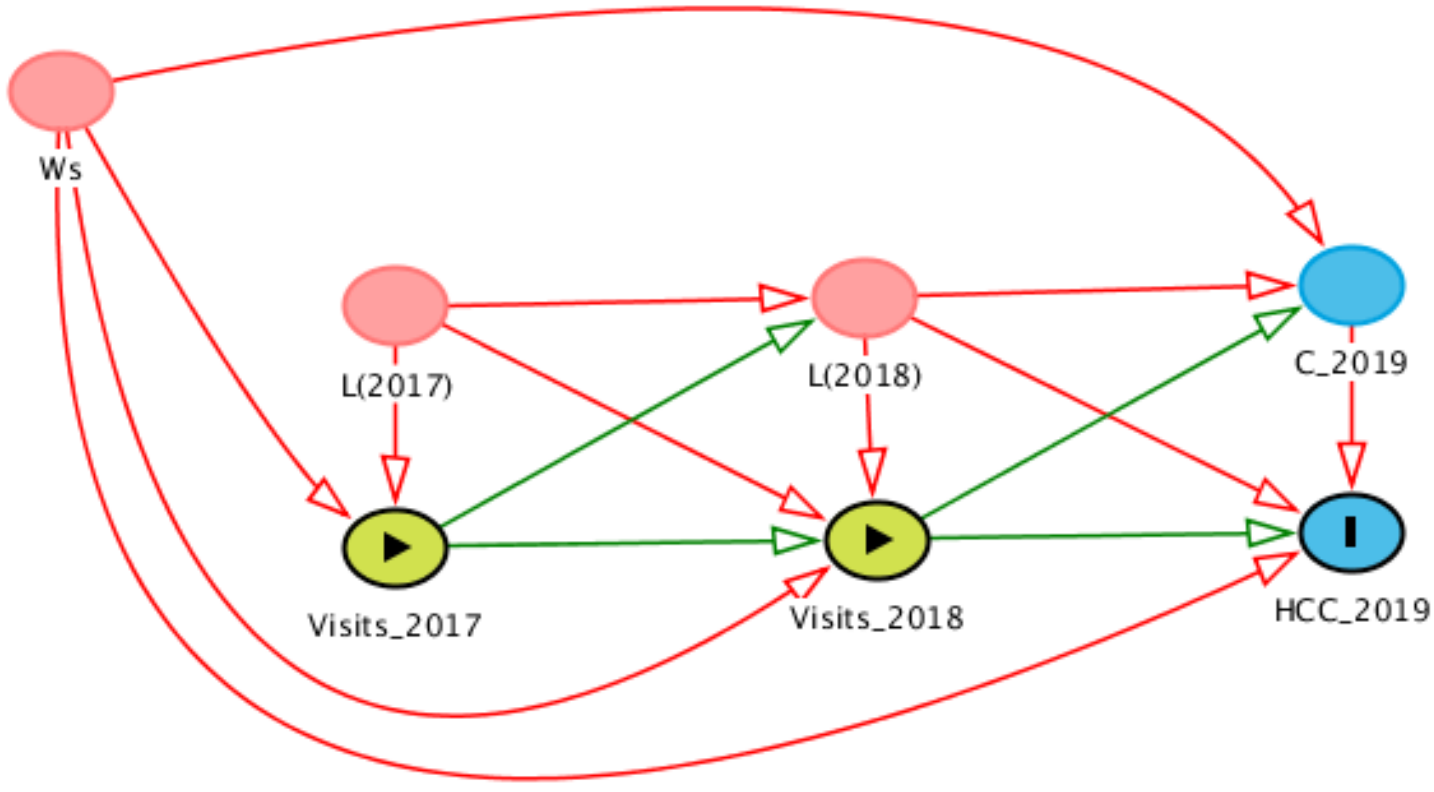
Exposure: $A(t),\ t = 1, 2, ..., 4$: The annual number of clinical visits that each chronic HCV patient had in the UCSF system annually from 2015 through 2018.

Outcome: $Y(K)$: Diagnosis of HCC by the final year (2019), which occured in 156 patients.

Covariates: FIB-4 score, years since FIB-4 score last measured, sex (Male, Female), race (Black, Latinx, Other), insurance type (Medi-Cal, Medicare, private) as a surrogate of SES.

1. Specify Causal Model

Our DAG



1. Specify Causal Model

Our SCM

$$\mathbf{O} = (\mathbf{W}, \mathbf{C(K)}, \mathbf{L(t)}, \mathbf{Y(t)}, \mathbf{A(t)})$$

This is survival data with missingness and censoring, where:

- W = baseline covariates (race, sex and SES)
- L(t) = the set of covariates (most recent FIB-4 score, years since last FIB-4) at time t
- A(t) = the exposure (number of visits) at time t
- C(K) = indicator of being censored at the final timepoint
- Y(t) = outcome (an indicator of HCC diagnosis)

$$U = (U_{L(t)}, U_{A(t)}, U_{C(t)}, U_{Y(t)}), \ t = 1, 2, 3, 4, 5 \sim P_U$$

1. Specify Causal Model

Structural Equations, $\mathcal{M}^{\mathcal{F}}$, for t from 1 to 5:

$$
\begin{aligned}
W &= f_W(U_W) \\
L(1) &= f_{L(1)}(U_{L(1)}) \\
A(1) &= f_{A(1)}(W, L(1), U_{A(1)}) \\
L(t) &= f_{L(t)}(\bar{L}(t-1), \bar{A}(t-1), U_{L(t)}) \\
A(t) &= f_{A(t)}(W, \bar{L}(t), \bar{A}(t-1), U_{A(t)}) \\
C(t) &= f_{C(t)}(W, \bar{A}(t-1), \bar{Y}(t-1)) \\
Y(t) &= f_{Y(t)}(W, \bar{L}(t-1), \bar{A}(t-1), Y(t-1), U_{Y(t)}) \\
Y(K) &= f_{Y(K)}(W, C(K), \bar{L}(K-1), \bar{A}(K-1), Y(K-1), U_{Y(K)})
\end{aligned}
$$

1. Specify Causal Model

We are also operating with one key exclusion restriction:

Once someone develops the outcome $[Y(t) = 1]$, we set their $A(t+1...K)$ and $L(t+1, ..., K)$ to *NA* for the remainder of the analysis, and set $C(K) = 1$ for that subject.

2. Specify Causal Question

2. Specify Causal Question

How does frequency of primary care and hepatology visits at UCSF affect the likelihood of developing hepatocellular carcinoma (HCC) by the end of the follow up period among patients diagnosed with chronic hepatitis C (HCV), who were HCC-free at the beginning of the follow up interval?

2. Specify Causal Question

We are interested in contrasting the counterfactual outcome from various numbers of primary care and/or hepatology visits with the counterfactual outcome from fewer visits, to see if there is some sort of exposure-response relationship.

To do this, we will use a MSM that helps us understand the relationship between frequency of primary care or hepatology visits and risk of HCC diagnosis, conditional on FIB-4 score (as a proxy for liver cirrhosis) and a variety of other demographics.

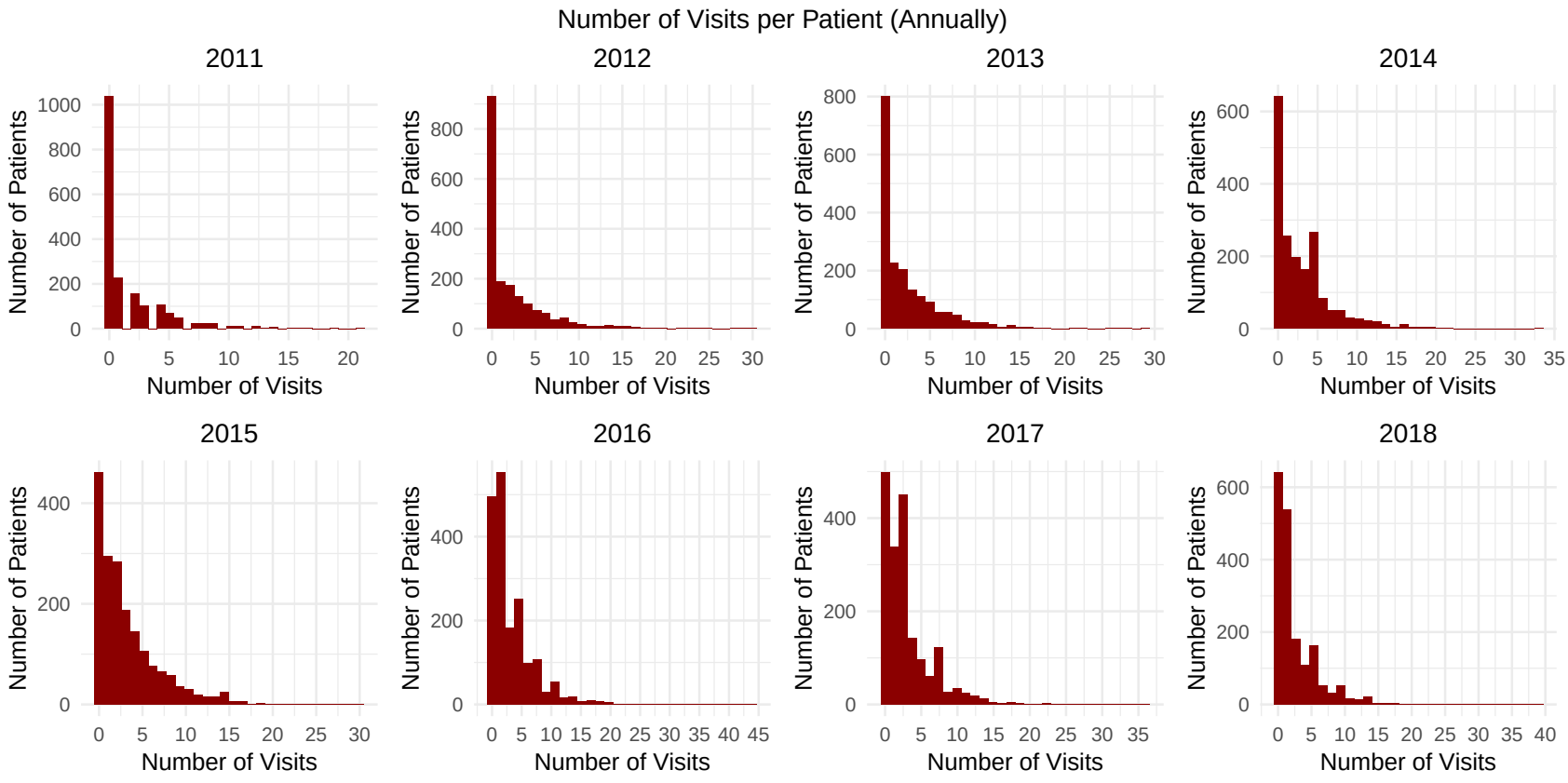
We will intervene to deterministically set $\bar{C}(K) = 0$ and $\bar{A}(t) = \bar{a}(t)$.

3. Specify Observed Data

3. Specify Observed Data

Number of Visits (annually)

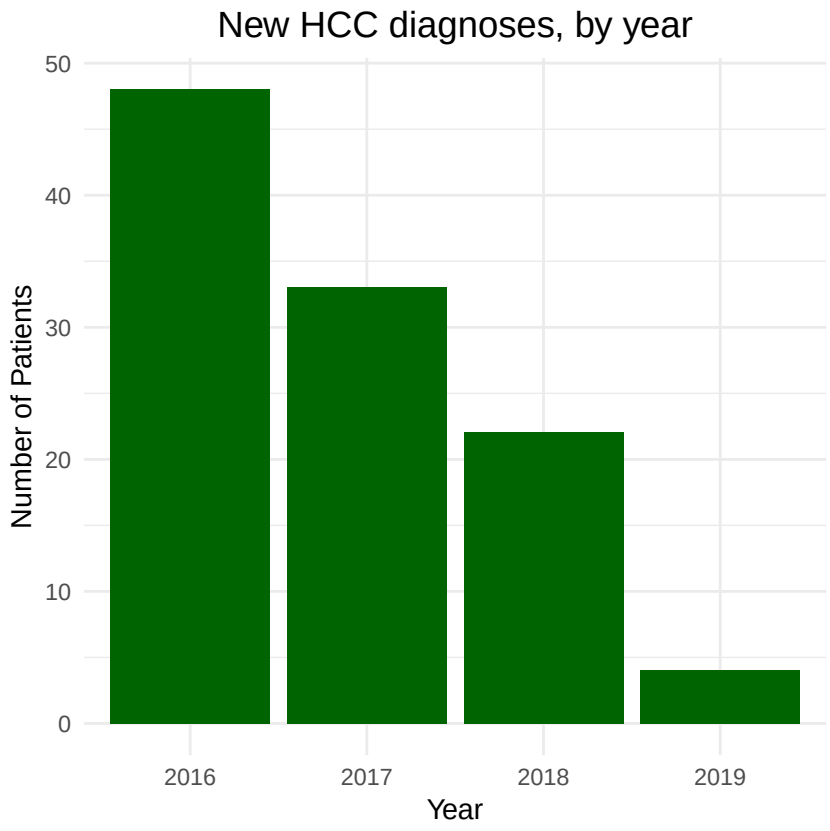
The number of visits per patient ranges from 0 to 44 in any given year, with a median of 1 visit per year overall.

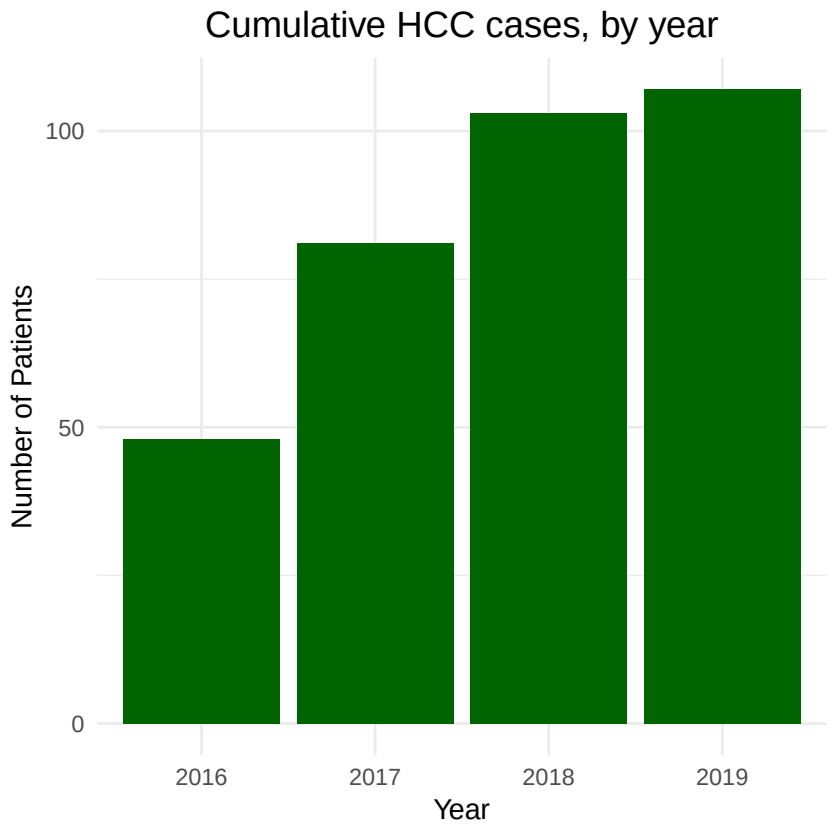


3. Specify Observed Data

Diagnosed with HCC (annually)

156 people (8.4%) developed the outcome by $Y(K)$.





*__note__ 2019 is an incomplete year!

3. Specify Observed Data

FIB-4 Score (annually)

FIB-4 scores are calculated using age, platelet count, AST, and ALT.

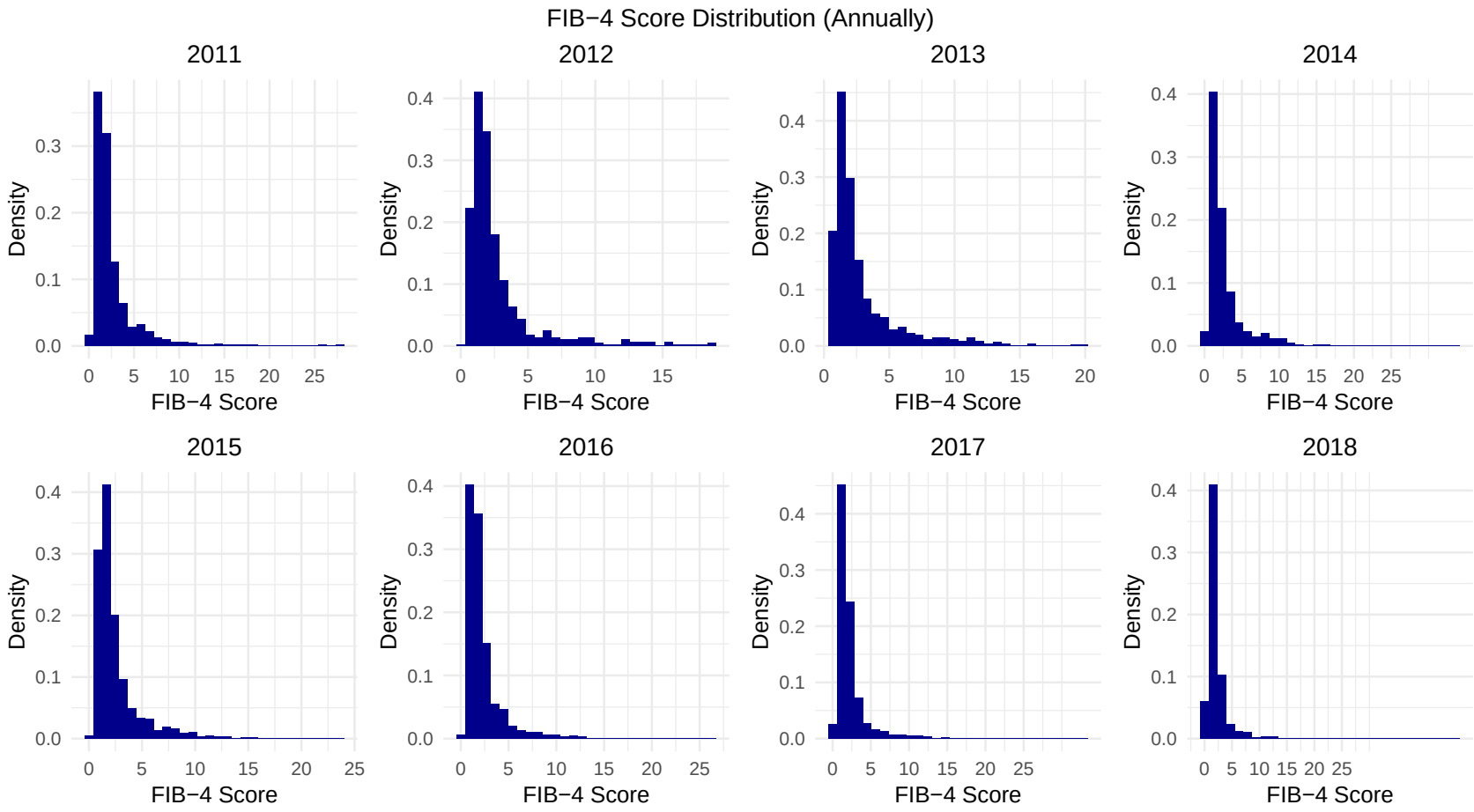
- Scores of $<$1.45 are considered to be strongly suggestive of no liver fibrosis
- Scores $>$3.25 are indicative of advanced fibrosis and/or cirrhosis.

FIB-4 scores in this dataset range of 0.181 to 45.956, with an IQR of 1.198-2.872.

- 1354 FIB-4 scores are $>$3.25 throughout all years of follow-up on all patients
- There are 256 instances where a patient's FIB-4 score was calculated to be $>$ 9.

3. Specify Observed Data

FIB-4 Score (annually)



3. Specify Observed Data

Other demographics

Demographic	Category	n	%
Sex	Male	1019	55.1%
	Female	829	44.9%
Race/ethnicity	Black/African American	288	15.6%
	Latinx	251	13.6%
	White	915	49.5%
	Other	372	20.1%
	Unknown	22	1.2%
SES (Payor type)	Medi-Cal	278	15%
	Medicare	1051	56.9%
	Commercial/Private	472	25.5%
	Unknown	47	2.5%
TOTAL		1848	100%

3. Specify Observed Data

Missingness

FIB-4 scores were available from 2009 onward.

- **82.8%** of participants had FIB-4 scores at the start of the study.
- For the rest of the participants, we used multiple imputation to impute a 2015 FIB-4 score.
- Updated $L(t)$ every year the requisite labs were measured.
- **95.5%** of participants had a measured FIB-4 score rather than an imputed value at the end of the study.

No missingness expected in exposure, since EMR designed for clinical billing

No missingness assumed for outcome, unless patient is censored at end of study

- However, if no visits at a particular t, that $Y(t)$ unknown

4. Identification

4. Identification

Our target causal parameter is identified under the following conditions:

- Assumed independence between each of the exogenous variables (no shared unknowns)
- No practical positivity violations
- Sequential randomization assumption for our data generating process
- Controlling for baseline covariates (Ws)

5. Commit to an Estimand and Statistical Model

5. Commit to an Estimand and Statistical Model

Our statistical estimand is:

$$\Psi(P_0) = m(\bar{a}|\beta) = E[Y_{\bar{a}}] = \beta_0 + \beta_1 \sum_{t=1}^{8} a(t)$$

Our statistical model is an MSM:

$$\Psi^F(P_{U,X}) = E(Y_{\bar{a},C(K)=0}) = m(\bar{a}|\beta, V)$$
$$= \beta_0 \ + \ \beta_1 \sum_{t=1}^{K} a(t) + \beta_2 V$$

6. Estimation

6. Estimation

We estimated our target causal parameter using MSMs implemented with G-computation, IPTW, and LTMLE estimators, via the `ltmleMSM` function in the *ltmle* R package.

a. Simulation

Exogenous variables:

- $U_W, U_{C(t)},$ and $U_{Y(t)}$ variables have a uniform distribution with a min of 0 and max of 1.
- $U_{L(t)}$ variables have a gamma distribution with a shape parameter of 0.6 and scale parameter of 2.6.
- $U_{A(t)}$ variables have a negative binomial distribution with a dispersion parameter of 4 and a probability of 0.6.

6. Estimation

a. Simulation

Endogenous variables:

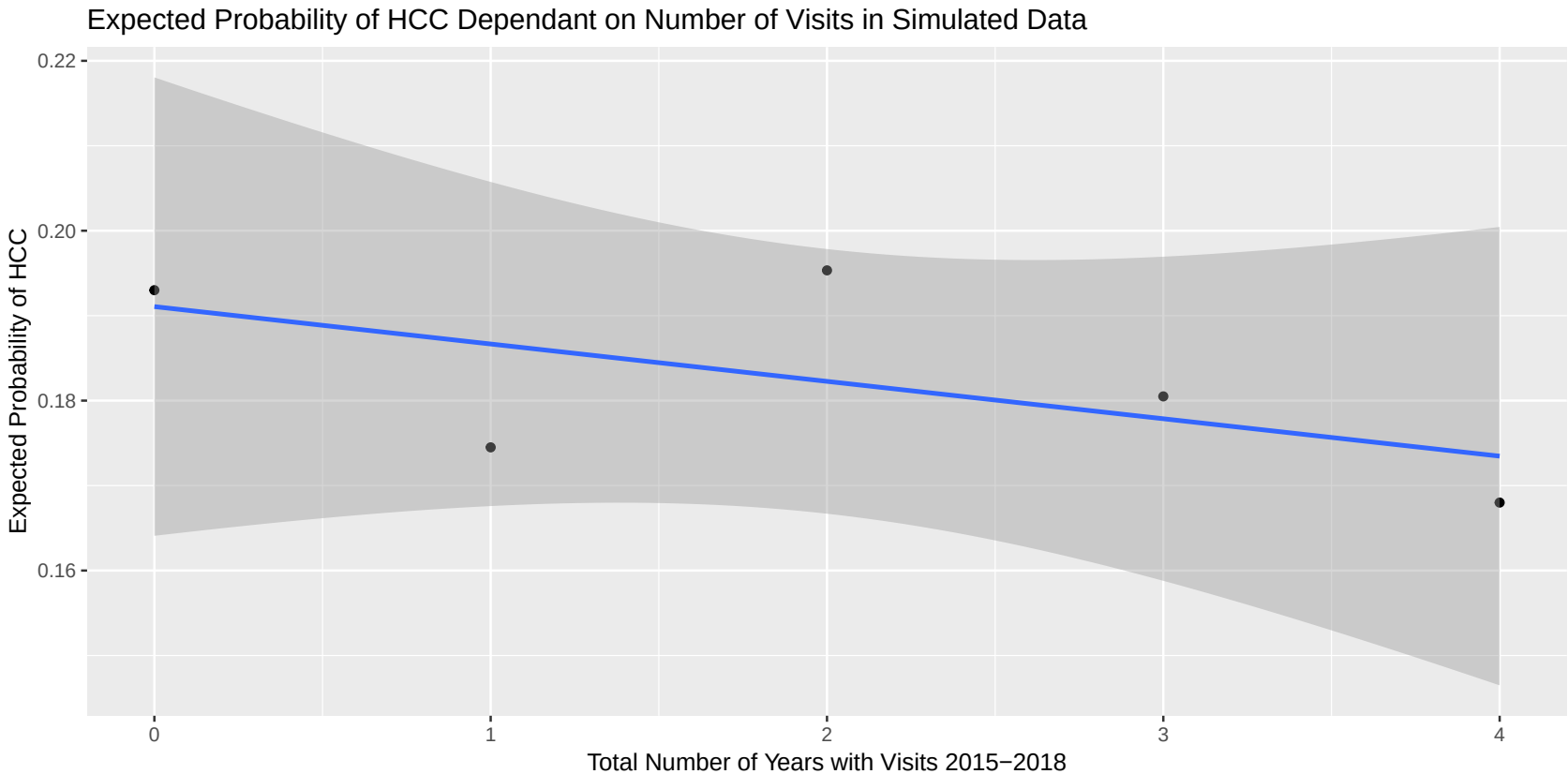
- W = a nominal categorical variable for the 16 possible different combinations of sex, gender, and SES.
- $L(t)$ = FIB-4 score from the prior year $(L(t-1)) + 10\%$ of the value generated by underlying gamma distribution for $U_{L(t)}$
- $A(t)$ = underlying negative binomial distribution for $U_{A(t)}$ plus FIB-4 score (more visits for higher FIB-4). **Dichotomitized into $A(t) = 0$ if visits are 0-1 or $A(t) = 1$ if visits are 2+.**
- $C(K) = 1$ if a subject had no primary care or hepatology visits in the final year, otherwise $C(K) = 0$.
- $Y(t) = 10\%$ chance of HCC if no visits over the prior interval, with a decreasing chance of HCC with every visit the subject had in the prior year and an increasing chance of HCC as the FIB-4 score rises, indicating worsening cirrhosis: $(Y(t) = I(U_{Y(t)} + 0.01(A(t-1)) - 0.05(L(t-1))) < 0.03)$.

6. Estimation

a. Simulation

We generated data with $n = 1000$. When we regressed to get the value of β_1 in our simulated MSM, it gave us ψ^F = -0.001625.

The plot below assesses the linearity of our simulated data.



6. Estimation

a. Simulation

We then assessed the performance of our estimators relative to our target causal parameter:

	G-Comp	IPTW	TMLE
Bias	-0.14	-0.687	-0.18
Variance	0.074	0.048	0.074
MSE	0.093	0.52	0.106

6. Estimation

a. Simulation

We then assessed the performance of our estimators relative to our target causal parameter:

	G-Comp	IPTW	TMLE
Bias	-0.14	-0.687	-0.18
Variance	0.074	0.048	0.074
MSE	0.093	0.52	0.106

6. Estimation

b. Observed data

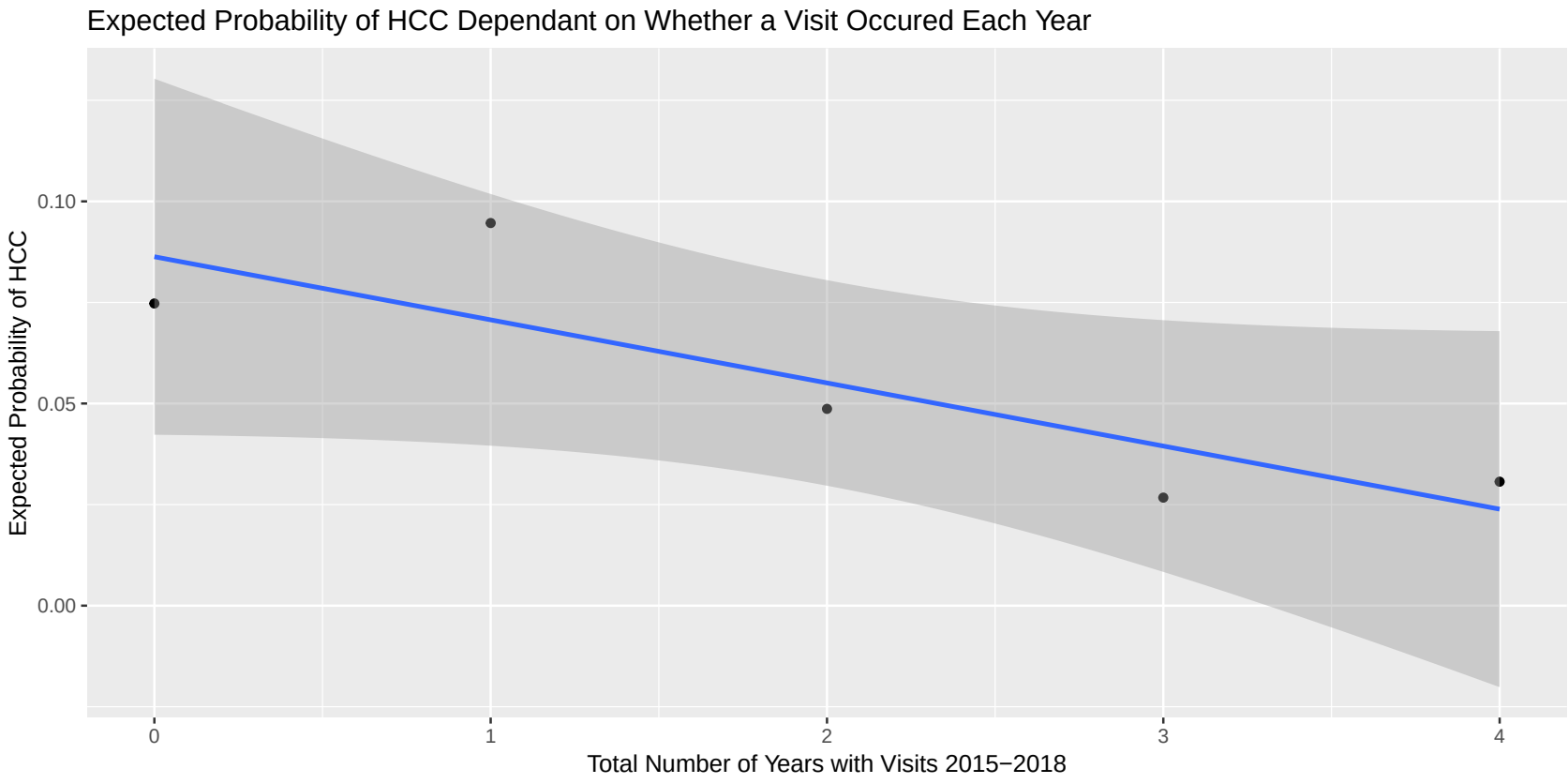
Implementing *ltmle* on our observed data produced the following estimates for β_1 on $\bar{A}(t) = \bar{a}(t)$ where $a \in \{0-1, 2+\}$ hepatology or primary care visits in each timepoint (i.e. per year):

G-Comp	IPTW	TMLE
-0.15	-0.169	-0.177

6. Estimation

b. Observed data

The linearity of our finding was encouraging:



7. Interpret Results

7. Interpret Results

Using the G-Comp estimator, which had the lowest MSE:

Over the five years under study, patients are 15% less likely to develop HCC each year that they had at least two primary care or hepatology visits, compared to patients who had one or fewer primary care of hepatology visits that year, adjusting for liver cirrhosis.

Current AASLD guidelines recommend liver monitoring every 6 months for **patients with cirrhosis** but these findings indicate such screening may be worthwhile for **all** patients with chronic HCV.