9-1-2015

Per-contact infectivity of HCV associated with injection exposures in a prospective cohort of young injection drug users in San Francisco, CA (UFO Study)

Yuridia Leyva

Yuridia Leyva
Candidate
Mathematics and Statistics
Department

This thesis is approved, and it is acceptable in quality and form for publication:

Approved by the Thesis Committee:

Erik Barry Erhardt , Chairperson

Kimberly Ann Page

Gabriel Huerta

Per-contact infectivity of HCV associated with injection exposures in a prospective cohort of young injection drug users in San Francisco, CA (UFO Study)

BY

YURIDIA LEYVA

B.S., Applied Mathematics, University of New Mexico, 2013
B.A., Chemistry, University of New Mexico, 2013

THESIS

Submitted in Partial Fulfillment of the
Requirements for the Degree of

Master of Science
Statistics

The University of New Mexico
Albuquerque, New Mexico

July 2015

Dedication

I dedicate this thesis to my father, Aaron D. Leyva, for believing in me, for being my greatest motivator, and for showing me, by example, the true meaning of hard work.

I also dedicate this thesis to my husband, Carlos M. Gutierrez, whose humor, support, and great attitude have been inestimable throughout this entire project. Thank you for being my "load-bearing wall", standing by me and allowing me to lean on you.

I would like to thank my entire family, who has been a constant source of encouragement and genuine support throughout the challenges of my graduate career.

Acknowledgments

I would like to express my deepest appreciation to my advisor and committee chair, Dr. Erik Barry Erhardt, for all of the positive energy, hard work, and time spent in helping me complete this thesis. Ever since I took my first class with Dr. Erhardt, I knew I wanted to work with him someday, and I feel complete because this goal has been fulfilled. Dr. Erhardt's attention to detail and passion for statistics have made me a stronger statistician. I am truly grateful for having worked with him.

I am also grateful to Dr. Kimberly Ann Page for allowing me the privilege of working with her using the UFO dataset. Her time, expertise, and all her years' work on Hepatitis C Virus have been invaluable throughout this research project.

I would like to take the opportunity to thank Dr. Gabriel Huerta for checking in on me periodically, for supporting this research project, and for guiding and advising me throughout my graduate career.

I will always treasure their time, assistance, and encouragement on my master's thesis project.

Per-contact infectivity of HCV associated with injection exposures in a prospective cohort of young injection drug users in San Francisco, CA (UFO Study)

by

Yuridia Leyva

B.S., Applied Mathematics, University of New Mexico, 2013

B.A., Chemistry, University of New Mexico, 2013

M.S., Statistics, University of New Mexico, 2015

Abstract

Sharing needles and ancillary injection drug equipment places injection drug users (IDU) at risk for Hepatitis C Virus (HCV), a highly infectious blood-borne virus. A limited number of studies have analyzed the per-contact infectivity of HCV associated with the use of previously-used needles, but per-contact infectivity of ancillary injecting equipment has not been previously investigated. Our goal is to estimate the per-contact infectivity of HCV associated with (1) injecting with another person's previously-used needle, classified as receptive needle sharing (RNS), and (2) using another person's previously-used ancillary injecting equipment, such as cookers to melt drugs and cottons to strain impurities from the melted drugs, termed receptive equipment sharing (RES). Estimates of per-contact probabilities were calculated based on self-reported exposures to RNS and RES. A probabilistic exposure model was used on the UFO (yoU Find Out) dataset composed of 784 IDU under the age of

30 who were surveyed quarterly between 2003-2008 and 2010-2014. For each participant, we selected the first survey with an HCV seronegative status up through their next seropositive survey, leaving us with 505 subjects on whom to conduct the analysis. A marginal maximum likelihood estimate (MLE) considering only RNS gives a per-contact infectivity of HCV as 0.39% (95% CI: 0.188% - 0.679%). A joint MLE gives RNS as 0.44% (95% CI: 0.0001% - 0.600%) and RES as 0% (95% CI: 0.00% - 0.69%), thus needles are a much bigger cause of concern than equipment. Though both probabilities are small, 13% (65/505) of the subjects studied seroconverted to an HCV-postitive status. Strategies for reducing RNS, and RES to a lesser extent, are important for reducing the spread of HCV and its related maladies.

Contents

Chapter 1

Introduction

Injection drug use and sharing ancillary injecting equipment contaminated with blood are common means of Hepatitis C Virus (HCV) transmission among young injection drug users (IDUs) (Klevens et al., 2012). In the United States, HCV is the most prevalent bloodborne virus whose severity ranges from an acute illness to a chronic illness attacking the liver. In the majority ($75\% - 85\%$) of people, acute infection advances to chronic infection, which is the leading cause for liver transplantation in the United States (Viral Hepatitis Action Coalition, 2015). As of 2015, an estimated 3.2 to 5 million people are infected with HCV in the United States (Centers for Disease Control and Prevention, 2014), with 130 to 170 million infected worldwide (Boelen et al., 2014).

Newer IDUs can rapidly obtain HCV infection after they inject for the first time, and they have the highest HCV incidence rates (Hagan et al., 2008). Injecting with another person's previously-used needle, classified as receptive needle sharing (RNS), is the easiest way to become infected with HCV (Hahn et al., 2010). Additionally, using another person's previously-used ancillary injecting equipment, such as cookers to melt drugs and cottons to strain impurities from the melted drugs, has been connected with HCV infectivity (Pouget et al., 2012). This receptive form of equipment

sharing is termed receptive equipment sharing (RES).

A limited number of studies have been conducted to assess the risk of HCV infection associated with RNS and RES (Hahn et al., 2010). The per-contact probability of HCV infection following RES is still unknown. Just as important as assessing the per-contact infectivities of RNS and RES are the per-contact probabilities of HCV infection associated with receptive backloading (RBL), which is the process of injecting drug doses from a previously-used syringe into the barrel of another syringe in order to measure and split drugs equally.

In this particular analysis, we examined the different types of exposures consisting of RNS and RES among young IDUs in San Francisco, California in order to determine the per-contact infectivity of HCV based strictly on receptive exposures associated with injection drug use.

Chapter 2

Methods

2.1 UFO Study

The UFO (yoU Find Out) study was used to obtain an estimate of the per-contact infectivity of HCV infection associated with young IDUs engaging in RNS and/or RES. The UFO study is a prospective study of viral Hepatitis C among non-infected young adults generally under the age of 30 who engaged in the use of injection drugs within the past month. Subjects were sampled from San Francisco, CA by outreach workers and by word-of-mouth between 2003-2008 and 2010-2014 (Hahn et al., 2002). A possible limitation of this sampling strategy is the uncertainty of how representative the sample is of the young injection drug user population in San Francisco or elsewhere. Participating young IDUs required an HCV-negative status at baseline screening; they were tested using a viremia test (HCV RNA) or tested for HCV antibodies (anti-HCV). Every three months, they received follow-up testing for HCV and were questioned by an interviewer regarding their exposures to HCV via injection drug use. Subjects received $10 USD for the screening visit and $20 upon return for their HCV test results. Descriptions of the study design and methods for the UFO cohort have been published in detail (Evans et al., 2009).

2.1.1 Data Description and Data Cleaning

Some of the participants in the UFO study were tested using both anti-HCV and HCV RNA tests, but had serological results that were non-coincidental. Individuals who were HCV RNA or anti-HCV positive were placed in the seroconverted category (status = 1). The remaining individuals were placed in the non-seroconverted category (status = 0), as they had results where either both types of HCV tests were negative, or one of the HCV tests was negative and the other was unknown.

The variables used for this analysis consisted strictly of receptive exposures to HCV. Receptive exposures consist of the young IDU having used a needle or injecting paraphernalia after it was used by someone else. The variables were placed into one of two exposure categories: needle or equipment exposures for subjects who engaged in RNS or RES, respectively. Table 2.1 lists the questions corresponding to each of the variables used, as well as its classification of exposure type.

Total number of injection exposures in the last 30 days (both receptive and non-receptive) were calculated as the product of the number of days injected in the last 30 days and the number of times injected per day.

Variables with Likert scale responses were expressed as probabilities: (1) "Always" 1.00, (2) "Usually" 0.75, (3) "Sometimes" 0.50, (4) "Rarely" 0.25, (5) "Never" 0.00.

UFO staff interviewed and surveyed subjects every three months, but some subjects had long time gaps between surveys. Within 180 days, a small proportion of UFO subjects have cleared and become reinfected with HCV (Page et al., 2009). Series of surveys were thus separated into multiple monitoring windows. For example, suppose a person has been interviewed a total of 8 times, but there is a gap of 180 days or more between the 4th and 5th survey. To account for this long time period between surveys, the first 4 surveys are placed in monitoring window 1, and the last 4 surveys are placed in monitoring window 2. For each subject, a decision was then made as

Table 2.1: Survey questions used in the analysis for receptive frequencies of needle and equipment sharing.

Variable Name	Exposure Type	Question	Response options
a0lj00	RNS	How many times in the past month did you inject with a needle that had already been used by the FIRST person with whom you inject most (partner 1)?	Numeric
a0lk00	RNS	How many times in the past month did you inject with a needle that had already been used by the SECOND person with whom you inject most (partner 2)?	Numeric
a0ll00	RNS	How many times in the past month did you inject with a needle that had already been used by the THIRD person with whom you inject most (partner 3)?	Numeric
a0mr00	RES	In the last 3 months, how often did you share a cooker or other container for dissolving drugs or use one that had already been used by someone else?	Always (1) Usually (2) Sometimes (3) Rarely (4) Never (5)
inj30d	general	In the last 30 days, how many DAYS did you shoot up anything including medication?	Numeric
tinj30d	general	How many times per day did you usually inject, on the days that you injected?	Numeric

to which observations to keep in order to conduct the analysis. Subjects required a serological status of 0 (HCV-negative) at baseline in order for probabilities to be calculated. Only the first observation with serological status of 1 (HCV-positive) was useful to help quantify the number of exposures it took to seroconvert an HCV-negative person to HCV-positive. Therefore, leading 1s, as well as any values after the first HCV positive status of 1 following any zeros, were removed. An example of this is illustrated in Table 2.2.

Table 2.2: Survey selection example based on HCV Status.

ID	Interview Date	Number of Exposures to: Needles	Equipment	Backloading	HCV Status	Action
FAKEID	15 Dec 10	6	0	1	1	*discard*
FAKEID	23 Mar 11	0	0	0	0	*keep*
FAKEID	15 Jun 11	5	0	2	0	*keep*
FAKEID	07 Sep 11	7	5	2	0	*keep*
FAKEID	07 Dec 11	8	5	4	1	*keep*
FAKEID	06 Mar 12	4	3	0	1	*discard*
FAKEID	06 Jun 12	11	0	3	1	*discard*
FAKEID	29 Aug 12	2	0	1	0	*discard*
FAKEID	05 Dec 12	5	0	2	0	*discard*

It may be the case that IDUs who naturally clear HCV (have a positive serostatus before a negative in their survey timeline) have better capability to "deal with" the virus than other IDUs. If this is the case, then including these IDUs in the sample may underestimate the per-contact infectivity rate. Therefore, a second analysis with the additional selection criterion that excludes IDUs who were positive at their first survey was also conducted. (We call this L0s for "leading 0s, only" in the raw data.)

We adjusted the number of exposures relative to the actual intersurvey times. First, the number of days between surveys was divided by 90 days (the length of time about which the surveys asked) to obtain a "stretch" multiplier. Then the reported 90-day exposure values were multiplied by this stretch factor to account for intersurvey times more or less than 90 days. An example of this is in Figure 2.1.

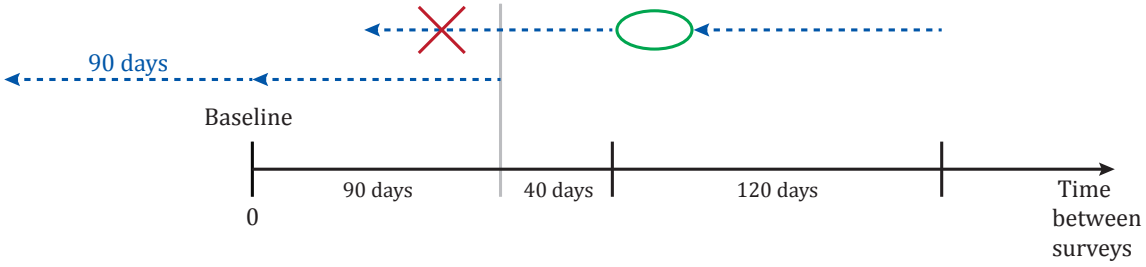


Figure 2.1: Adjustment of intersurvey times using a "stretch" factor. In the example above, there were 40 days between the second and third surveys, so the reported 90-day exposures (blue dashed lines) at the third survey are shortened by a multiplicative factor of 40/90 (red X). Similarly the fourth survey is increased by 120/90 (green O).

Figures 2.2 and 2.3 illustrate that the total number of exposures associated with RNS generally exceeds the total number of exposures associated with RES.

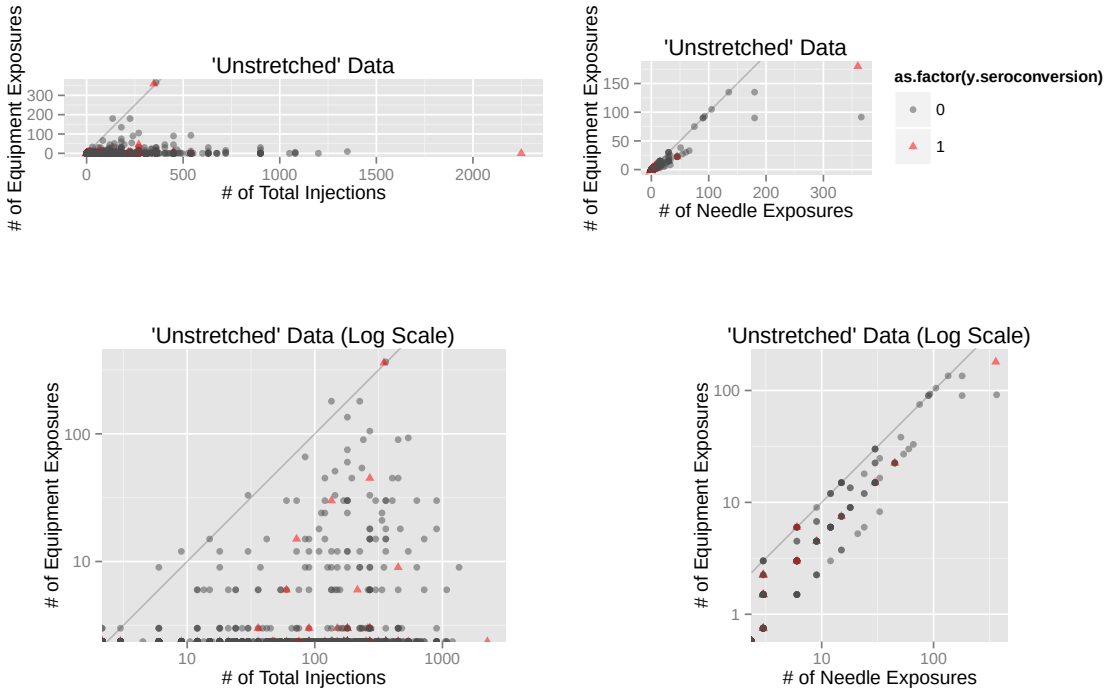


Figure 2.2: Left: Plot of receptive equipment shares against needle shares before stretching of intersurvey times, Right: Plot of receptive equipment shares against needle shares before stretching of intersurvey times.

Data cleaning revealed several anomalies. Thirty-six (36) subjects claimed to have zero exposures associated with receptive equipment and needle shares in a given survey, yet they seroconverted to an HCV-positive status. This means that they may have under-reported their exposures associated with RNS and RES, or they may have obtained HCV via another route. This will be discussed further in Chapter 4. Additionally, Figure 2.6 revealed that almost 50% of subjects never returned for a follow-up interview.

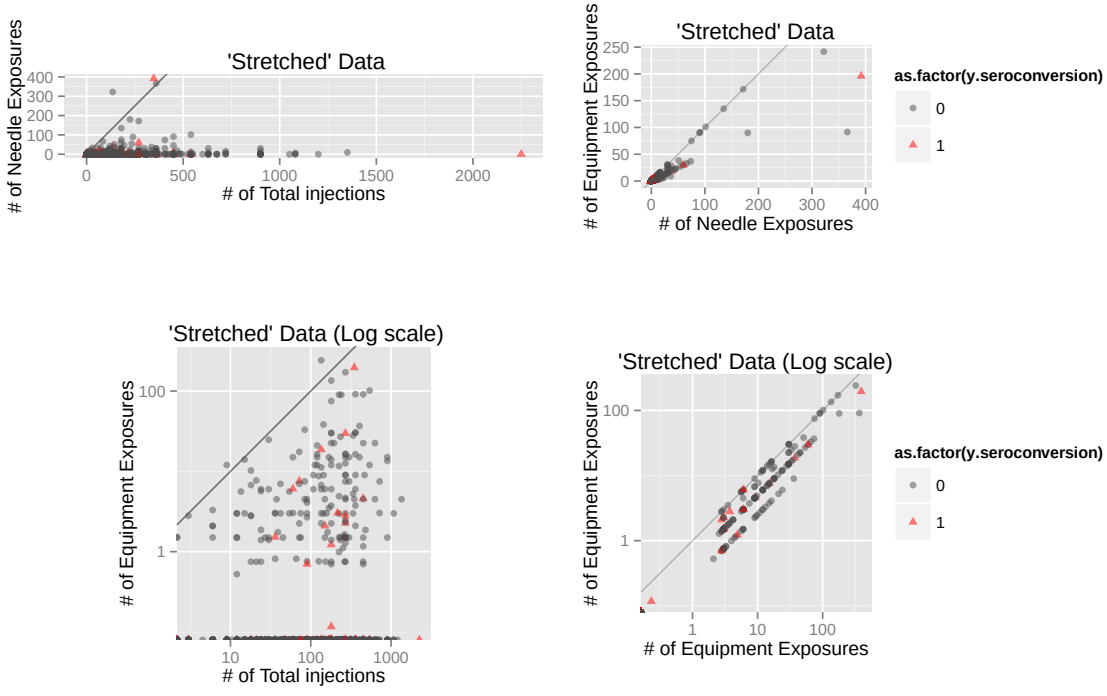


Figure 2.3: Left: Plot of receptive equipment shares against needle shares after stretching of intersurvey times, Right: Plot of receptive equipment shares against needle shares after stretching of intersurvey times.

Figure 2.4: Visualization of seroconversion status at each interview, before data cleaning. Sample size: N=784 subjects; 39% seroconverted at least once.

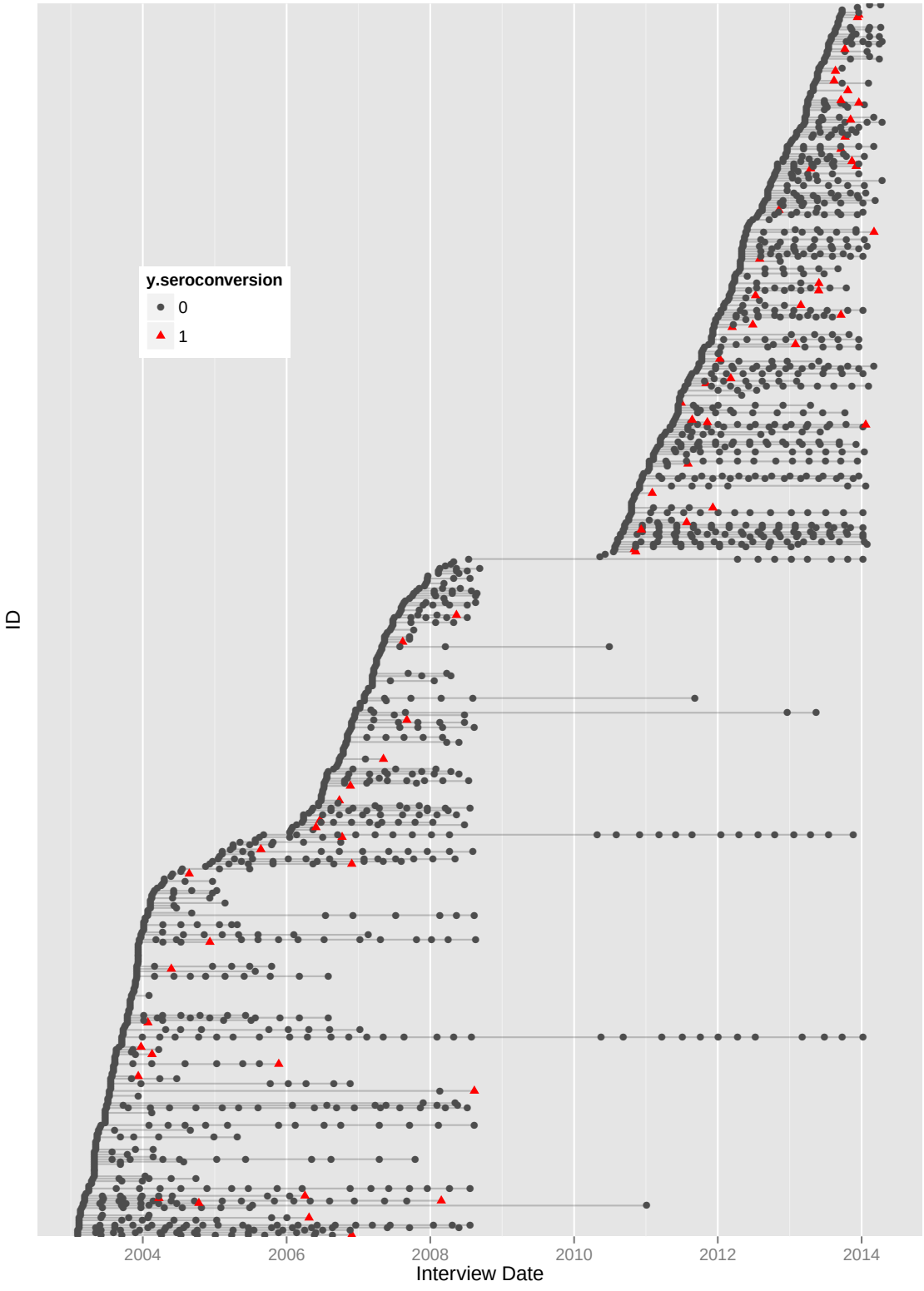


Figure 2.5: Visualization of seroconversion status at each interview, after data cleaning.Sample size: N=505 subjects; 13% seroconverted.

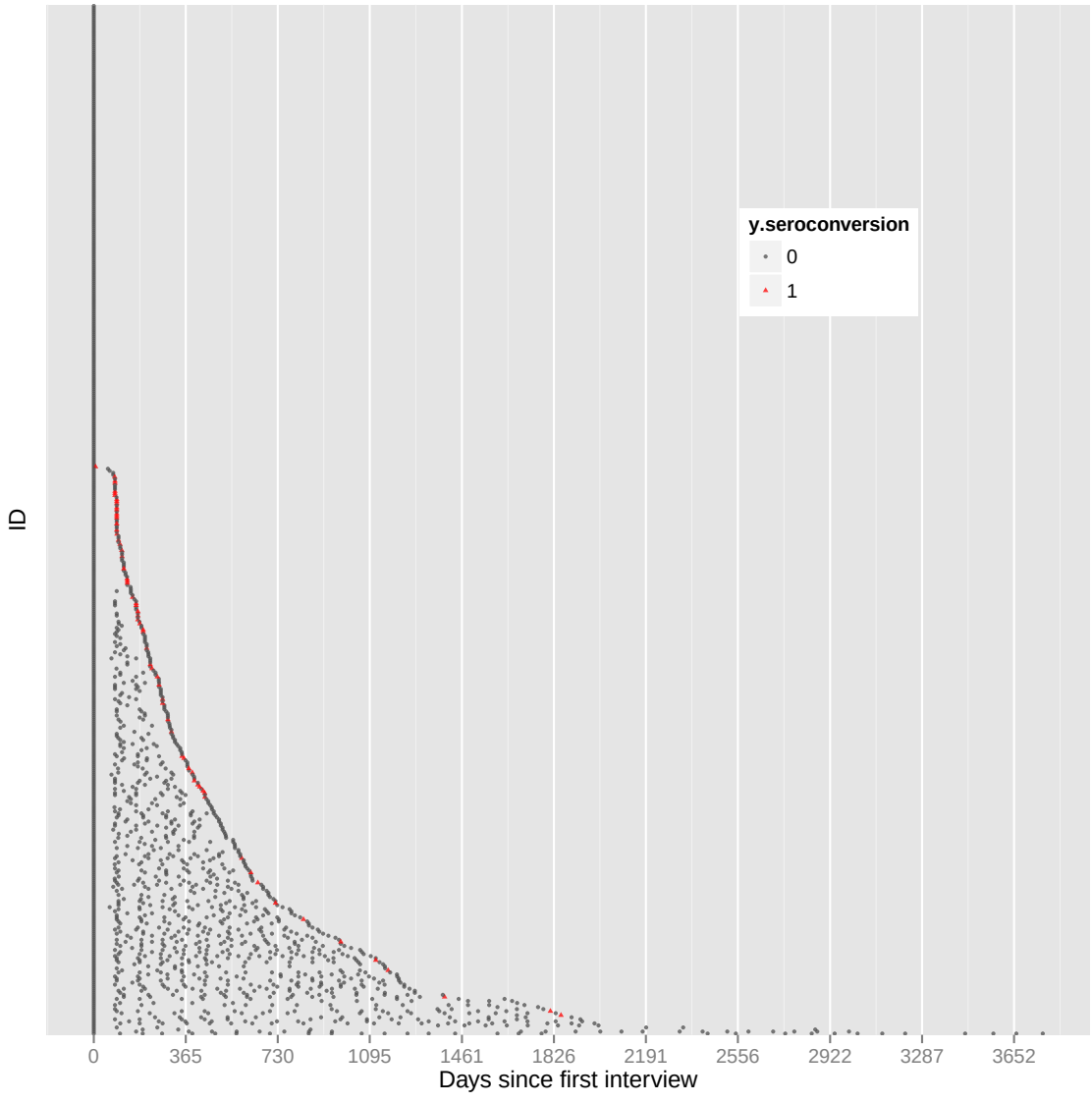


Figure 2.6: Visual representation of the length of time a subject is studied. About 45% of the sample used in the analysis had only one survey.

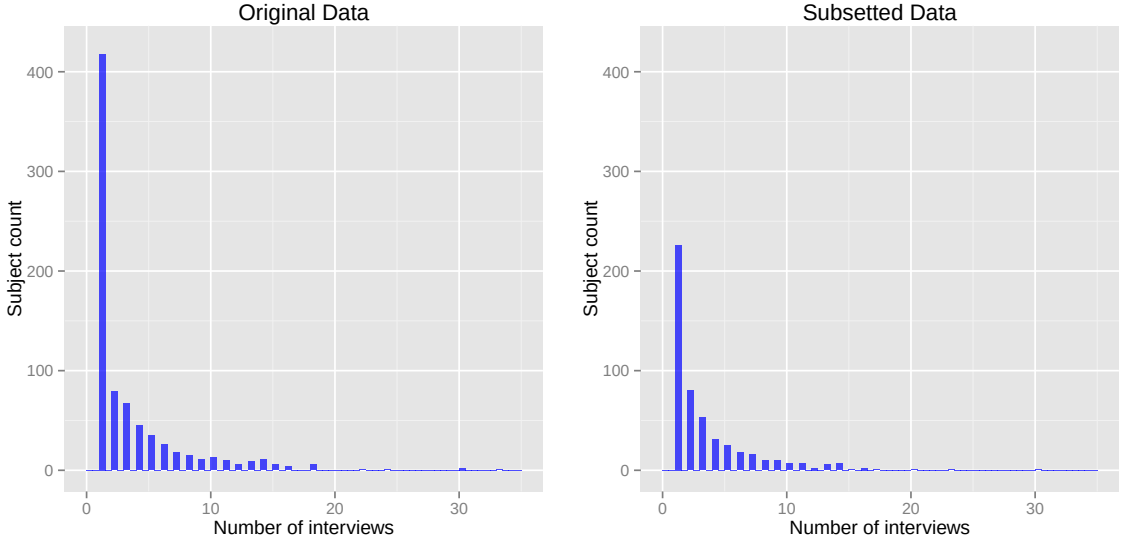


Figure 2.7: Distributions of interview frequencies per subject before and after data cleaning.

2.2 Statistical Model and Maximum Likelihood Estimate for Per-contact Infectivity Rates

Our goal is to estimate the per-contact infectivity rates, β_{N} and β_{E}, of HCV transmission for receptive needle sharing and receptive ancillary equipment sharing, respectively. We estimate β_{N} and β_{E} via a maximum likelihood estimate (MLE), which is the value of the parameter that makes the data most likely under the model. We use the following likelihood function, L, for a sample size of N participants and S_i surveys):

$$L(\beta_{\mathrm{N}}, \beta_{\mathrm{E}} \,|\, n_{\mathrm{N}ij}, n_{\mathrm{E}ij}, y_{ij}) = \prod_{i=1}^{N} \prod_{j=1}^{S_i} f_{ij}^{y_{ij}} (1 - f_{ij})^{(1-y_{ij})}.$$

To model the UFO data, the probabilities of the data must be considered a function of the β parameters in the model. The following probability mass function (pmf) was used for each subject $i = 1, ..., N$ and all subjects' surveys $j = 1, ..., S_i$:

$$f_{ij} = f_{ij}(\beta_{\mathrm{N}}, \beta_{\mathrm{E}} \,|\, n_{\mathrm{N}ij}, n_{\mathrm{E}ij}) = 1 - (1 - \beta_{\mathrm{N}})^{n_{\mathrm{N}ij}} (1 - \beta_{\mathrm{E}})^{n_{\mathrm{E}ij}},$$

where $n_{\mathrm{N}ij}$ is the number of receptive exposures to needles and $n_{\mathrm{E}ij}$ is the number of exposures associated with receptive sharing of ancillary injecting equipment (such as cookers or cottons). y_{ij} is the status of seroconversion; $y_{ij} = 0$ for HCV-negative participants, and $y_{ij} = 1$ for HCV-positive participants. This probability model assumes exposure probabilities are independent.

To obtain the MLEs, we use the log of L rather than the likelihood function itself in order to avoid numerical overflow issues. To maximize the log-likelihood, we minimize the negative log-likelihood using the optim() procedure in R. A uni-parameter analysis of per-incident seroconversion associated with needles alone assumes $\beta_{\mathrm{E}} \equiv 0$. There

is no general closed-form solution for β_{N}, even in the uni-parameter case, thus a numerical optimization procedure is necessary.

The per-contact infectivities of Hepatitis C Virus associated with RNS and RES were estimated using bootstrap intervals at the 95% confidence level. To construct a bootstrap estimate of the MLE sampling distribution, we obtained 1000 bootstrap re-samples by sampling with replacement from IDUs including all their "clean" surveys. Then we performed uni-parameter and bi-parameter maximum likelihood estimates on each resample.

2.3 Simulation study

To understand the behavior of the (negative log) likelihood function and the MLE, a series of small cases are followed by a larger data-realistic case.

2.3.1 Simulations summary

Estimation of per-contact infectivty is often accurate. For any number of types of exposures, when there is a unique exposure of one type at seroconversion, the estimation is unbiased. When there are multiple exposures of one type at seroconversion, it is not possible to know how many of the exposures would have resulted in seroconversion – thus the per-contact probability is estimated between the minimum and maximum probabilities, closer to the minimum.

With two exposure types when there are multiple exposures at seroconversion the MLE may locate on the boundary of the parameter space. In this case, the log-likelihood function favors the larger β and sends the other to zero. Also, when the number of exposures becomes very large, and the computer code for the (negative log) likelihood function is written in a direct way, the function becomes jagged due to numerical underflow (a probability raised to a high power goes to zero). Therefore, the R package `Rmpfr` for arbitrarily precise numbers was implemented to increase bit precision from 53 ("double") to 200, and this relieved the large-n underflow issue.

For the notation in the following tables, let there be two exposure types, and let n.1 and n.2 be the number of exposures of each type for four survey observations (4 rows of data). Let y.1 and y.2 indicate the unobserved true exposure set that caused a seroconversion for each type. Let y.seroconversion be the observed HCV status (0=negative, 1=positive).

The data-realistic simulation estimates the parameter on the boundary, which is con-

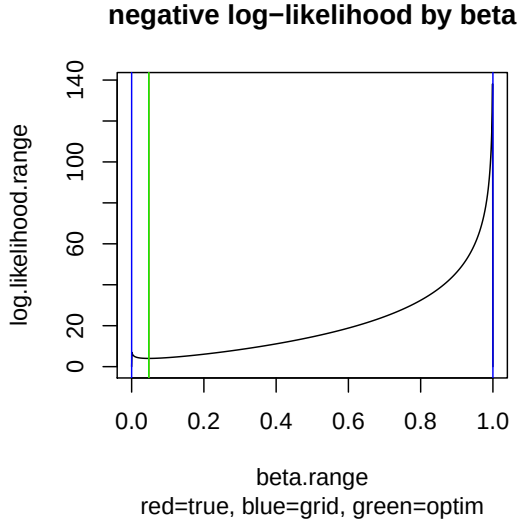
sistent with the simpler simulation cases. It favors the larger β and sends the other to zero.

2.3.2 Simulations: one types with nonoverlapping/separate exposures

Simulation: Univariate, one seroconverted obs: $\beta_1 = 1/21$

Univariate optimization is exact when there's one observation at seroconversion.

y.seroconversion	n.1	y.1
0	20	0
1	1	1



Model results:

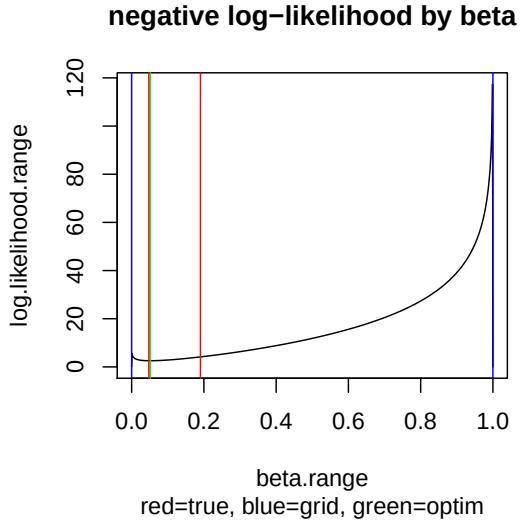
```
CONVERGENCE: REL_REDUCTION_OF_F <= FACTR*EPSMCH
```

Type	$-\log(L)$	β_1
True	4.02033	0.047619
Est	4.02033	0.0476348

Simulation: Univariate, four seroconverted obs: $\beta_1 = 4/21$

When there's multiple observations at seroconversion, convergence is inside the range of plausible correct values. In this case, the estimated β_1 is between 1/21 and 4/21.

y.seroconversion	n.1	y.1
0	17	0
1	4	1



Model results:

```
CONVERGENCE: REL_REDUCTION_OF_F <= FACTR*EPSMCH
```

Type	$-\log(L)$	β_1
True	4.15342	0.190476
Est	2.55629	0.051463
True 2	2.55936	0.047619

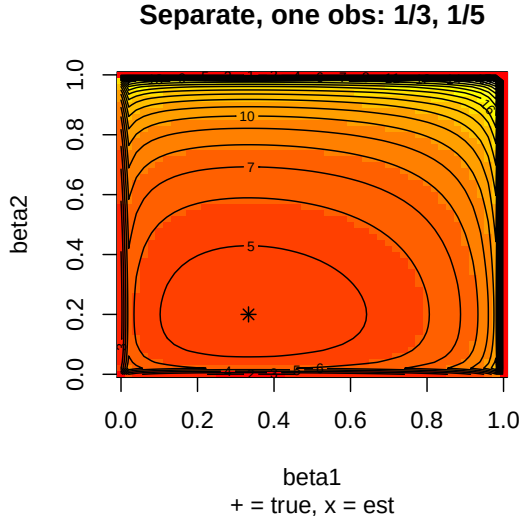
2.3.3 Simulations: two types with nonoverlapping exposures

In the following examples, when there is a single exposure at seroconversion the β MLE estimates the true parameters without bias. When there are multiple exposures at seroconversion for a single exposure type, then there is censoring, that is, the specific exposure(s) responsible for seroconversion is unknown. In this case, the MLE is between the lower and upper bounds of the minimum and maximum number of exposures responsible for seroconversion.

Separate, one seroconverted obs: $\beta_1 = 1/3$, $\beta_2 = 1/5$

Convergence to the correct value.

y.seroconversion	n.1	y.1	n.2	y.2
0	0	0	4	0
1	0	0	1	1
0	2	0	0	0
1	1	1	0	0



Model results:

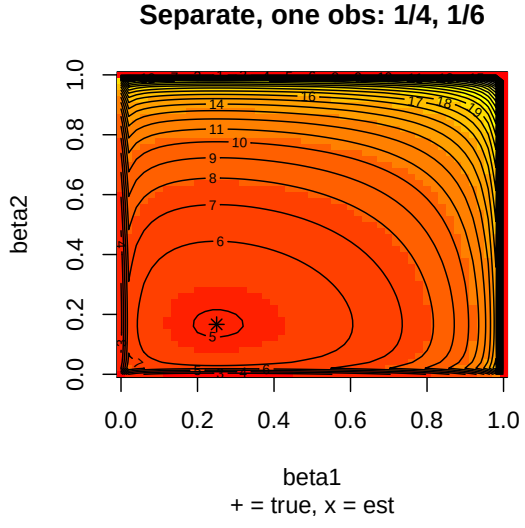
```
CONVERGENCE: REL_REDUCTION_OF_F <= FACTR*EPSMCH
```

Type	$-\log(L)$	β_1	β_2	plot symbol
True	4.41155	0.333333	0.2	+
Est	4.41155	0.333333	0.200001	x

Separate, one seroconverted obs: $\beta_1 = 1/4$, $\beta_2 = 1/6$

Convergence to the correct value.

y.seroconversion	n.1	y.1	n.2	y.2
0	0	0	5	0
1	0	0	1	1
0	3	0	0	0
1	1	1	0	0



Model results:

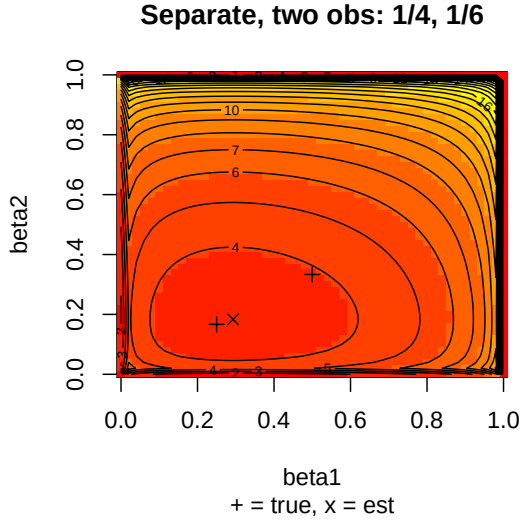
```
CONVERGENCE: REL_REDUCTION_OF_F <= FACTR*EPSMCH
```

Type	$-\log(L)$	β_1	β_2	plot symbol
True	4.95271	0.25	0.166667	+
Est	4.95271	0.250001	0.166668	x

Separate, two seroconverted obs: $\beta_1 = 2/4$, $\beta_2 = 2/6$

Convergence to inside the range of plausible correct values.

y.seroconversion	n.1	y.1	n.2	y.2
0	0	0	4	0
1	0	0	2	1
0	2	0	0	0
1	2	1	0	0



Model results:

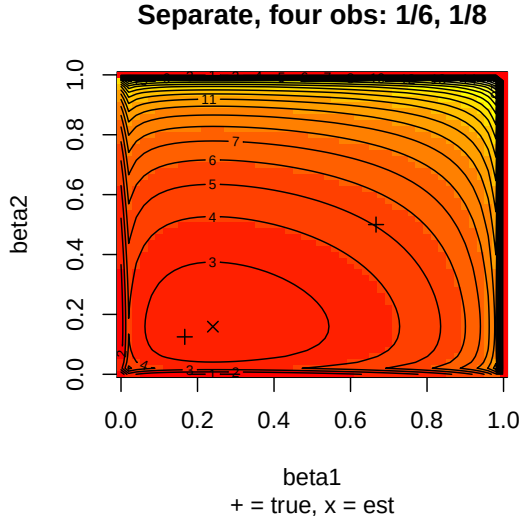
```
CONVERGENCE: REL_REDUCTION_OF_F <= FACTR*EPSMCH
```

Type	$-\log(L)$	β_1	β_2	plot symbol
True	3.31695	0.25	0.166667	+
Est	3.29584	0.292894	0.183505	x
True 2	3.88362	0.5	0.333333	+

Separate, four seroconverted obs: $\beta_1 = 4/6$, $\beta_2 = 4/8$

Convergence to inside the range of plausible correct values.

y.seroconversion	n.1	y.1	n.2	y.2
0	0	0	4	0
1	0	0	4	1
0	2	0	0	0
1	4	1	0	0



Model results:

```
CONVERGENCE: NORM OF PROJECTED GRADIENT <= PGTOL
```

Type	$-\log(L)$	β_1	β_2	plot symbol
True	2.43937	0.166667	0.125	+
Est	2.34107	0.240165	0.159105	x
True 2	5.04677	0.666667	0.5	+

2.3.4 Simulations: two types with overlapping exposures

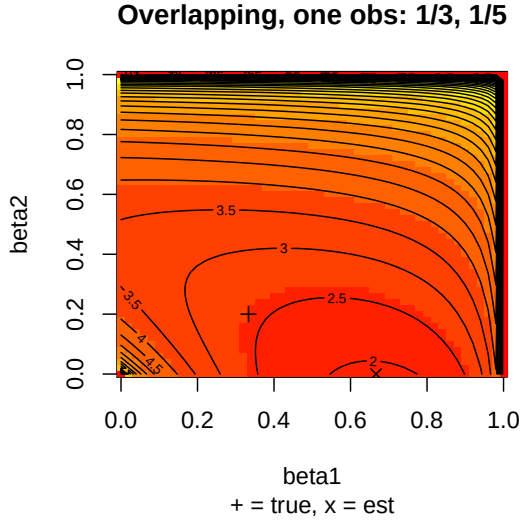
In the following examples, when there is are overlapping exposures at seroconversion the β MLE estimates the true parameters *with* bias. There are conditions (not shown) where the sample sizes are very large and the estimates are brought back off the boundary.

A future area of work is to use a Bayesian method to include prior information that may help this boundary issue.

Overlapping, one seroconverted obs: $\beta_1 = 1/3$, $\beta_2 = 1/5$

If both exposures have only 1 exposure on seroconversion, then the negative log likelihood function has a minimum on the boundary, sending the exposure with more observations to probability equal to 0.

y.seroconversion	n.1	y.1	n.2	y.2
0	0	0	3	0
1	1	0	1	1
0	1	0	0	0
1	1	1	1	0



Model results:

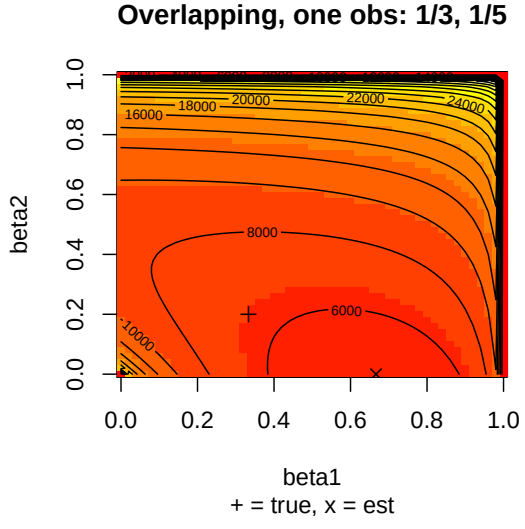
```
CONVERGENCE: NORM OF PROJECTED GRADIENT <= PGTOL
```

Type	$-\log(L)$	β_1	β_2	plot symbol
True	2.59918	0.333333	0.2	+
Est	1.90954	0.666666	10^{-10}	x

Overlapping, one seroconverted obs, 10000 samples: $\beta_1 = 1/3$, $\beta_2 = 1/5$

To test model consistence as the number of surveys goes to ∞, we repeated the earlier simulation replicating the rows 2500 times for 10000 total rows (with the same result for 10 times more data). The results are the same as before (hit the boundary), thus this model is not consistent. The MLE is not a consistent estimator for the true parameters.

y.seroconversion	n.1	y.1	n.2	y.2
0	0	0	3	0
1	1	0	1	1
0	1	0	0	0
1	1	1	1	0
0	0	0	3	0
1	1	0	1	1
0	1	0	0	0
1	1	1	1	0



Model results:

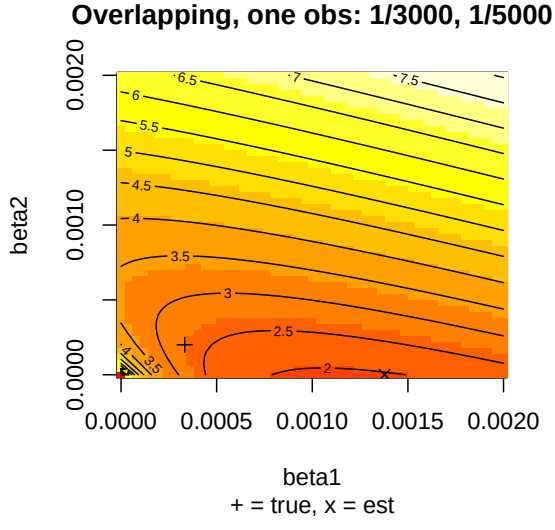
```
CONVERGENCE: REL_REDUCTION_OF_F <= FACTR*EPSMCH
```

Type	$-\log(L)$	β_1	β_2	plot symbol
True	6497.94	0.333333	0.2	+
Est	4773.86	0.666666	10^{-10}	x

Overlapping, one seroconverted obs, 10000 samples: $\beta_1 = 1/3000$, $\beta_2 = 1/5000$

The result also happens when the number of exposures goes to ∞, we repeated the earlier by multiplying the number of exposures by 1000. The results are the same as before (hit the boundary), thus this model is not consistent. The MLE is not a consistent estimator for the true parameters.

y.seroconversion	n.1	y.1	n.2	y.2
0	0	0	3000	0
1	1000	0	1000	1
0	1000	0	0	0
1	1000	1	1000	0



Model results:

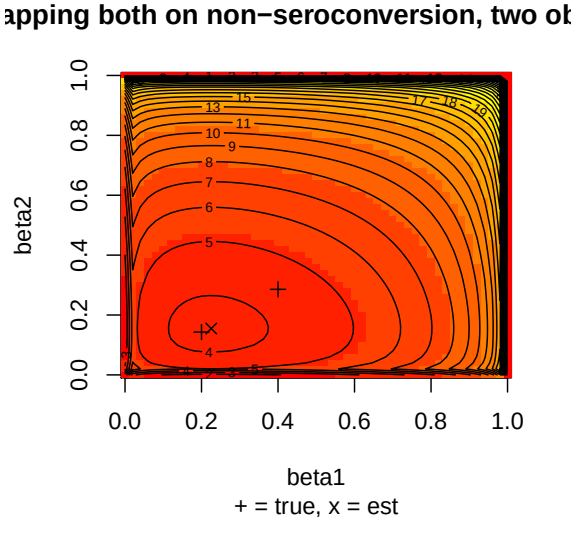
```
ERROR: ABNORMAL_TERMINATION_IN_LNSRCH
```

Type	$-\log(L)$	β_1	β_2	plot symbol
True	2.70014	3.33333×10^{-4}	2×10^{-4}	+
Est	1.95945	0.0013786	10^{-10}	x

Overlapping on nonserconversion, one seroconverted obs: $\beta_1 = 1/5$, $\beta_2 = 1/7$

If both exposures have only 1 exposure on seroconversion, but overlapping events on nonseroconversion, then there is no additional issue. The estimates are similar to as when there was no overlap.

y.seroconversion	n.1	y.1	n.2	y.2
0	1	0	4	0
1	0	0	2	1
0	2	0	1	0
1	2	1	0	0



Model results:

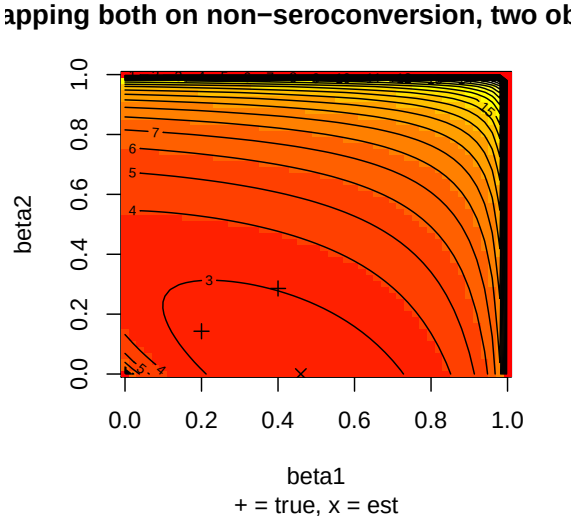
```
CONVERGENCE: NORM OF PROJECTED GRADIENT <= PGTOL
```

Type	$-\log(L)$	β_1	β_2	plot symbol
True	3.78871	0.2	0.142857	+
Est	3.77647	0.225404	0.154848	x
True 2	4.37489	0.4	0.285714	+

Overlapping nonseroconversion with both seroconvertion obs: $\beta_1 = 1/5$, $\beta_2 = 1/7$

Similar boundary issue when multiple exposure types overlap with one or more exposures.

y.seroconversion	n.1	y.1	n.2	y.2
0	0	0	4	0
1	1	0	2	1
0	2	0	0	0
1	2	1	1	0



Model results:

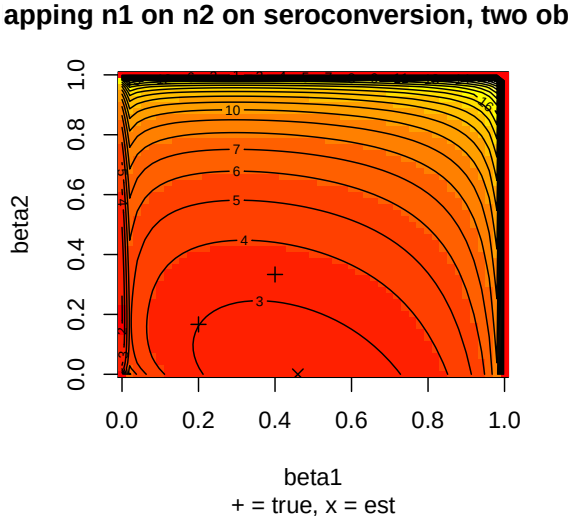
```
CONVERGENCE: NORM OF PROJECTED GRADIENT <= PGTOL
```

Type	$-\log(L)$	β_1	β_2	plot symbol
True	2.74437	0.2	0.142857	+
Est	2.35365	0.459688	10^{-10}	x
True 2	3.03025	0.4	0.285714	+

Overlapping n.1 on n.2 on seroconvertion obs: $\beta_1 = 1/5$, $\beta_2 = 1/6$

n.1 predicted when n.2 caused, but the model is blind to this.

y.seroconversion	n.1	y.1	n.2	y.2
0	0	0	4	0
1	1	0	2	1
0	2	0	0	0
1	2	1	0	0



Model results:

```
CONVERGENCE: NORM OF PROJECTED GRADIENT <= PGTOL
```

Type	$-\log(L)$	β_1	β_2	plot symbol
True	3.00815	0.2	0.166667	+
Est	2.35365	0.459688	10^{-10}	x
True 2	3.39995	0.4	0.333333	+

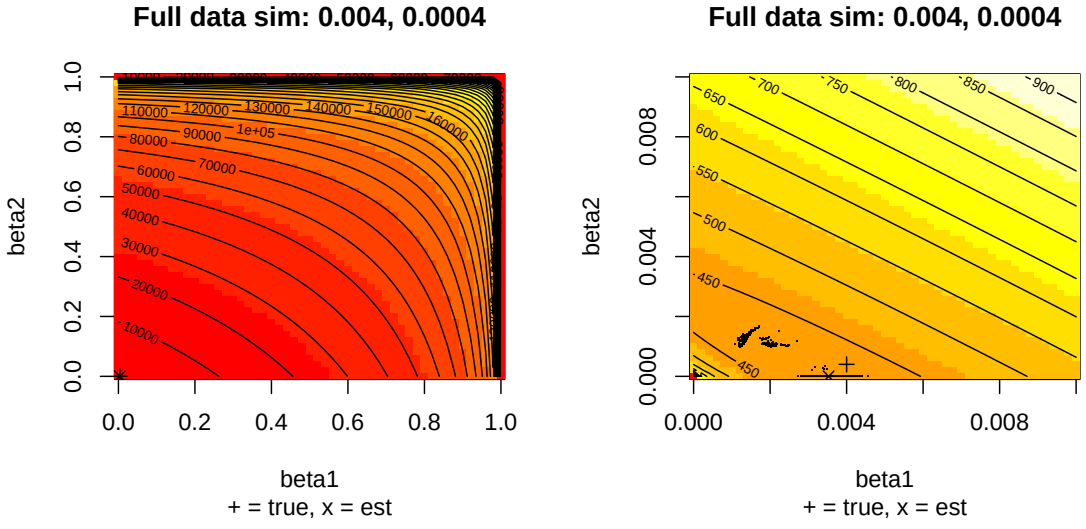
2.3.5 Simulation: data-realistic case

The data-realistic simulation strategy is to choose N subjects to study and survey each subject, $i = 1, \ldots, N$, for S_i surveys, where each subject has a common per-contact rate of seroconversion. Assume that each subject is surveyed at perfect 90-day intervals, so "stretching" is not necessary for the simulations, and that subjects report perfectly on their number of exposures. Let the maximum number of surveys for each subject be 15 and choose a rate of observations per subject λ (e.g., 3). Simulate the number of surveys iid for each subject as $S_i \sim \text{Poisson}(\lambda)$, and let $s = 1, \ldots, S_i$. For all subjects, choose a number of exposure types, E, and let $j = 1, \ldots, E$, to study (e.g., $E = 1$ for only needles, $E = 2$ for needles and equipment). For each exposure, choose a 30-day exposure rate r (e.g., 20 and 10). Simulate the number of exposures per survey for each subject as $n_{isj} \sim \text{Poisson}(r_j)$. For each exposure, choose a per-contact seroconversion probability β (e.g., 0.004 and 0.0004). Simulate the seroconversion due to contact for each subject as $y_{isj} \sim \text{Binomial}(n_{isj}, \beta_j)$. Finally, "clean" the data by excluding leading positive seroconversion surveys and any surveys after the first seroconversion.

Simulation: Real-data: $\beta_1 = 0.004$**,** $\beta_2 = 0.0004$

Data were simulated with the following parameters.

Name	Parameter	value
Users	N	500
Exposure types	E	2
Seroconversion probabilities	β	0.004, 4×10^{-4}
Number of surveys for each subject	λ	3
30-day Poisson exposure rate	r	20, 10
Max number of surveys		10
Resulting data:		
Number of Users		500
Number of surveys		2000



Model results:

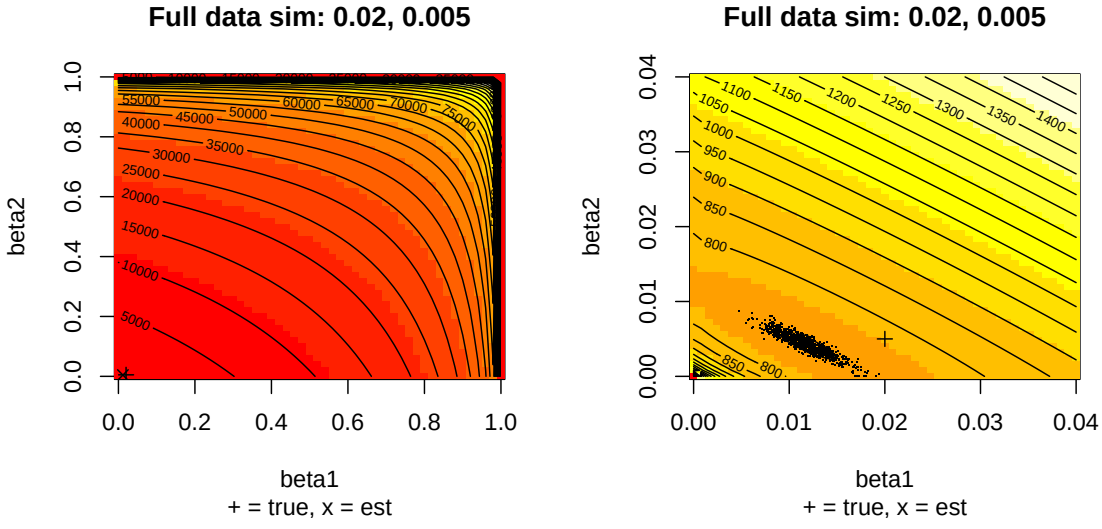
```
ERROR: ABNORMAL_TERMINATION_IN_LNSRCH
```

Type	$-\log(L)$	β_1	β_2	plot symbol
True	434.875	0.004	4×10^{-4}	+
Est	429.11	0.0035298	10^{-10}	x
95% CI	lower	0.001268	10^{-10}	.
95% CI	upper	0.0041371	0.0014191	.

Simulation: Real-data: $\beta_1 = 0.02$**,** $\beta_2 = 0.005$

Data were simulated with the following parameters.

Name	Parameter	value
Users	N	500
Exposure types	E	2
Seroconversion probabilities	β	0.02, 0.005
Number of surveys for each subject	λ	4
30-day Poisson exposure rate	r	10, 5
Max number of surveys		20
Resulting data:		
Number of Users		500
Number of surveys		2000



Model results:

```
ERROR: ABNORMAL_TERMINATION_IN_LNSRCH
```

Type	$-\log(L)$	β_1	β_2	plot symbol
True	777.228	0.02	0.005	+
Est	748.951	0.0116975	0.0043301	x
95% CI	lower	0.0077023	0.0013602	.
95% CI	upper	0.0165695	0.0070045	.

Chapter 3

Results

The prospective UFO dataset included a total of 784 subjects, of whom 39% seroconverted at least once during the course of their involvement in UFO study, as shown in Figure 2.4. After subsetting the data by excluding surveys with leading positive seroconversion results and any data after the first seroconversion, data on 505 subjects, of whom 65 seroconverted, was actually used in the analysis, as illustrated in Figure 2.5.

During the course of the study, 36 subjects (from the subsetted data) became infected with HCV yet reported no receptive IDU prior to the incidence time point, accounting for 7% of subjects. These 36 individuals may have under-reported their engagement in RNS or RES, or they may have obtained HCV via another route. (More in Chapter 4.)

If we keep in mind that data cleaning involved removing observations such as those described in Table 2.2, it makes sense that the distribution in Figure 2.7 of interviews for the subsetted data would only change frequencies for participants who had a small number of interviews. In order to estimate a per-contact infectivity, we must look at the number of exposures it took to change a person's HCV status from negative to positive. This requires more than one interview as well as a seroconversion from

HCV-negative to HCV-positive. At around 0 interviews, the distribution in Figure 2.7 has a lower frequency of interviews after data cleaning. This is because participants with only one interview and participants who did not meet the requirement of being non-infected at baseline, were omitted from the analysis.

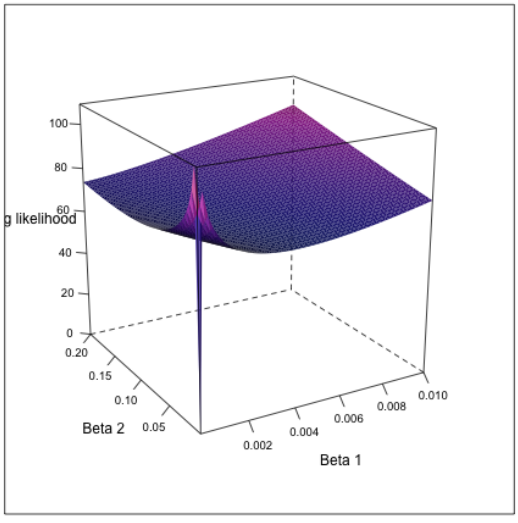


Figure 3.1: Three-dimensional plot of negative log-likelihood function.

Figure 3.2 gives us a downward view into the β_1 - β_2 plane and connects all the values with the same likelihoods, producing contour lines. Since we are looking for a maximum likelihood, we identify it by locating the minimum value in this plot of the negative likelihood, and the value is 70. This figure illustrates that the HCV infectivity rate for RNS is near 0.004, while the infectivity rate for RES is near 0, which is consistent with Figure 3.1.

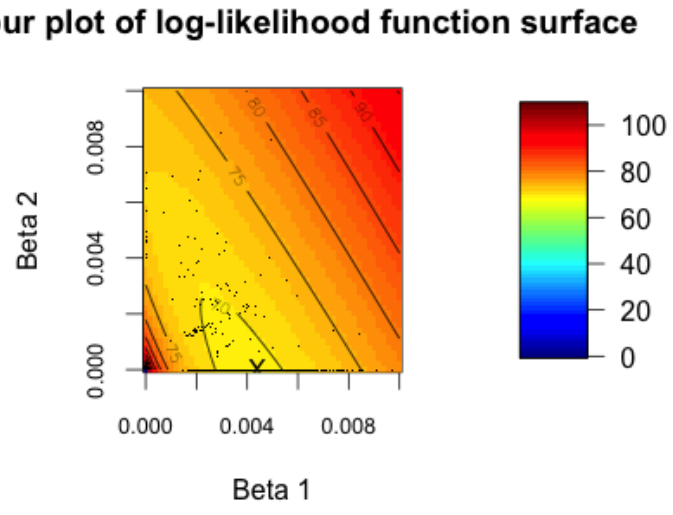


Figure 3.2: Two-dimensional plot of negative log-likelihood function.

For our bi-parameter model described in Chapter 2, the maximum likelihood estimate of per-contact infectivity of HCV associated with RNS was 0.44% (95% CI: 0.0001% - 0.600%), while the estimated per-contact infectivity of HCV associated with RES was 0% (95% CI: 0.00% - 0.69%).

Table 3.1: Results of estimates using the uni-parameter and bi-parameter models, compared to a uni-parameter estimate in literature.

Type	β_1 (95% CI)	β_2 (95% CI)
Bi-parameter	0.44% $(0.0001\% - 0.600\%)$	0% ($0.0001\% - 0.69\%$)
Uni-parameter	0.39% $(0.188\% - 0.679\%)$	–
Boelen et al. uni-parameter	0.57% $(0.32 - 1.05\%)$	–

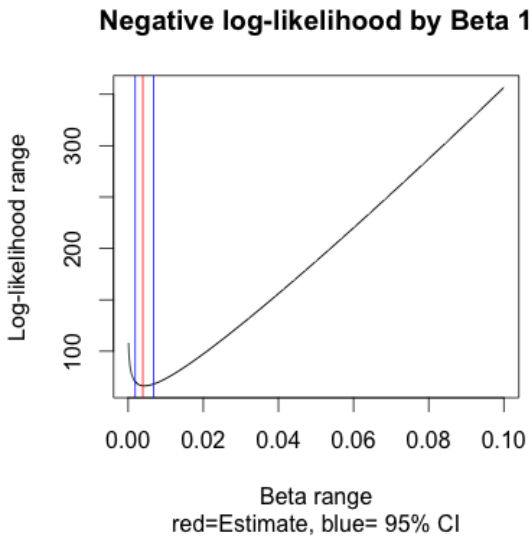


Figure 3.3: Plot of the uni-parameter analysis results for RNS. Red line illustrates the estimated per-contact infectivity of HCV associated with RNS. Blue lines illustrate lower and upper bounds of the 95% confidence interval.

A separate uni-parameter analysis was conducted on RNS alone, and the resulting infectivity rate was 0.39% (95% CI: 0.188% - 0.679%). These results are organized in Table 3.1 and illustrated in Table 3. Confidence intervals were obtained via bootstrapping.

Chapter 4

Discussion

For our study, the maximum likelihood estimate of per-contact infectivity of HCV associated with RNS was 0.44% (95% CI: 0.0001% - 0.600%) based on the bi-parameter model. This confidence interval and the confidence interval for the uni-parameter model, 0.39% (95% CI: 0.188% - 0.679%), are consistent with that of the ongoing HITS-p cohort established in 2005 within correctional centers in Australia, where the estimated probability of infection associated with needle-sharing was 0.57% (CI: 0.32 - 1.05%) (Boelen et al., 2014). The estimated per-contact infectivity of HCV associated with RES was 0% (95% CI: 0.00% - 0.69%). To our best knowledge, this analysis provides the first quantitative estimate of per-contact infectivity associated with using previously-used ancillary injecting equipment.

During the course of the study, 36 subjects (from the subsetted data) became infected with HCV yet reported no receptive IDU prior to the incidence time point, accounting for 7.13 % of subjects. These 36 individuals may have under-reported their engagement in RNS or RES, or they may have obtained HCV via another route. Though there are some exceptions, studies have shown that a very small number of new or old HCV infections are attributed to sexual transmission (Klevens et al., 2012).

Still, the results imply that RNS is generally a bigger problem than RES, given that the per-contact probability of infection due to equipment shares is extremely low. These low probabilities are a bit surprising, as HCV is extremely virulent; it has been shown to survive in a syringe for up to 63 days (Paintsil et al., 2010) and up to 5 days on inatimate surfaces (Doerrbecker et al., 2011). However, a 0% probability of infection is just an estimate. Our 95% confidence interval for RES illustrates a possibility that plausible values for the per-contact infectivity of HCV associated with RES are as high as .69%. Actions for reducing the number of exposures associated with RNS and RES must be actualized.

The UFO dataset has many good qualities that have allowed for extensive studies in a variety of areas. In our study, the UFO dataset shows that young IDU are learning to engage less in RNS, or at least they are reporting less RNS activitiy. Perhaps young IDU are becoming aware that utilizing previously-used needles may result in becoming infected with a virus that is not curable. However, young IDU need to be aware that they can also become infected when utilizing previously-used ancillary injection equipment.

References

Gregory L Armstrong, Annemarie Wasley, Edgar P Simard, Geraldine M McQuillan, Wendi L Kuhnert, and Miriam J Alter. The prevalence of Hepatitis C virus infection in the United States, 1999 through 2002. *Annals of internal medicine*, 144(10):705–714, 2006.

Lies Boelen, Suzy Teutsch, David P Wilson, Kate Dolan, Greg J Dore, Andrew R Lloyd, Fabio Luciani, HITS investigators, et al. Per-event probability of Hepatitis C infection during sharing of injecting equipment. *PloS one*, 9(7):e100749, 2014.

Centers for Disease Control and Prevention. Hepatitis C information for the public. `http://www.cdc.gov/hepatitis/C/cFAQ.htm`, 2014. Accessed: 2014-12-30.

Juliane Doerrbecker, Martina Friesland, Sandra Ciesek, Thomas J Erichsen, Pedro Mateu-Gelabert, Jörg Steinmann, Jochen Steinmann, Thomas Pietschmann, and Eike Steinmann. Inactivation and survival of Hepatitis C virus on inanimate surfaces. *Journal of Infectious Diseases*, 204(12):1830–1838, 2011.

Jennifer L Evans, Judith A Hahn, Paula J Lum, Ellen S Stein, and Kimberly Page. Predictors of injection drug use cessation and relapse in a prospective cohort of young injection drug users in San Francisco, CA (UFO Study). *Drug and alcohol dependence*, 101(3):152–157, 2009.

Jennifer L Evans, Meghan D Morris, Michelle Yu, Kimberly Page, and Judith A Hahn. Concordance of risk behavior reporting within HCV serodiscordant injecting

partnerships of young injection drug users in San Francisco, CA. *Drug and alcohol dependence*, 142:239–244, 2014.

Holly Hagan, Enrique R Pouget, Don C Des Jarlais, and Corina Lelutiu-Weinberger. Meta-regression of hepatitis c virus infection in relation to time since onset of illicit drug injection: the influence of time and place. *American journal of epidemiology*, 168(10):1099–1109, 2008.

Judith A Hahn, Kimberly Page-Shafer, Paula J Lum, Kristen Ochoa, and Andrew R Moss. Hepatitis C virus infection and needle exchange use among young injection drug users in San Francisco. *Hepatology*, 34(1):180–187, 2001.

Judith A Hahn, Kimberly Page-Shafer, Paula J Lum, Philippe Bourgois, Ellen Stein, Jennifer L Evans, Michael P Busch, Leslie H Tobler, Bruce Phelps, and Andrew R Moss. Hepatitis C virus seroconversion among young injection drug users: relationships and risks. *Journal of Infectious Diseases*, 186(11):1558–1564, 2002.

Judith A Hahn, Jennifer L Evans, Peter J Davidson, Paula J Lum, and Kimberly Page. Hepatitis C virus risk behaviors within the partnerships of young injecting drug users. *Addiction*, 105(7):1254–1264, 2010.

R Monina Klevens, Dale J Hu, Ruth Jiles, and Scott D Holmberg. Evolving epidemiology of Hepatitis C virus in the United States. *Clinical infectious diseases*, 55(suppl 1):S3–S9, 2012.

Kimberly Page, Judith A Hahn, Jennifer Evans, Stephen Shiboski, Paula Lum, Eric Delwart, Leslie Tobler, William Andrews, Lia Avanesyan, Stewart Cooper, et al. Acute Hepatitis C virus infection in young adult injection drug users: a prospective study of incident infection, resolution, and reinfection. *Journal of Infectious Diseases*, 200(8):1216–1226, 2009.

Kimberly Page, William Osburn, Jennifer Evans, Judith A Hahn, Paula Lum, Alice Asher, Eric Delwart, Leslie Tobler, Andrea L Cox, and Michael P Busch. Frequent

longitudinal sampling of HCV infection in IDU reveals intermittently detectable viremia and reinfection. *Clinical Infectious Diseases*, page cis921, 2012.

Kimberly Page, Meghan D Morris, Judith A Hahn, Lisa Maher, and Maria Prins. Injection drug use and Hepatitis C virus infection in young adult injectors: using evidence to inform comprehensive prevention. *Clinical infectious diseases*, 57(suppl 2):S32–S38, 2013.

Elijah Paintsil, Huijie He, Christopher Peters, Brett D Lindenbach, and Robert Heimer. Survival of Hepatitis C virus in syringes: implication for transmission among injection drug users. *Journal of Infectious Diseases*, 202(7):984–990, 2010.

Enrique R Pouget, Holly Hagan, and Don C Des Jarlais. Meta-analysis of hepatitis c seroconversion in relation to shared syringes and drug preparation equipment. *Addiction*, 107(6):1057–1065, 2012.

Daniel Tracy, Judith A Hahn, Crystal Fuller Lewis, Jennifer Evans, Alya Briceño, Meghan D Morris, Paula J Lum, and Kimberly Page. Higher risk of incident Hepatitis C virus among young women who inject drugs compared with young men in association with sexual relationships: a prospective analysis from the UFO Study cohort. *BMJ open*, 4(5):e004988, 2014.

Viral Hepatitis Action Coalition. Hepatitis C. `http://www.viralhepatitisaction.org/hepatitis-c`, 2015. Accessed: 2015-01-01.