Supplement to "Evidence of recruitment bias in a network study of people who inject drugs"

[Blinded for peer review]

1 Detailed methods

1.1 Data structure

We assume that the target population social network exists and is a simple undirected graph $G = (V, E)$ [2, 3]. The vertex set V corresponds to members of the population, and edges in E represent social links across which recruitments might take place. A recruiter is a subject that can recruit other yet-unrecruited vertices because it has at least one coupon. A susceptible individual is not yet recruited, and has at least one recruiter neighbor in G. Each recruitment of a susceptible vertex costs the recruiter one coupon. No subject can be recruited more than once.

Let M be the set of "seeds" in an RDS study whose sample size is n. Since recruitments are assumed to happen over edges in the population network G, RDS recruitments reveal edges in G to researchers. Let $G_R = (V_R, E_R)$ be the recruitment subgraph, consisting of the set V_R of recruited subjects and the set E_R of recruitment links. It follows that the sizes of these sets are $|V_R| = n$ and $|E_R| = n - |M|$. We can interpret U as the "boundary" of the observed set V_R in G. Let E_U be the set of links connecting vertices in U to vertices in V_R. Let $G_{SU} = (V_{SU}, E_{SU})$ be the augmented recruitment-induced subgraph, consisting of $V_{SU} = V_R \cup U$ and $E_{SU} = E_R \cup E_U$. We can interpret G_{SU} as the subgraph consisting of the sampled vertices and the boundary of the sampled set in G. Let $\mathbf{X}_i$ be a vector representing the individual traits of subject i, which typically includes gender, ethnicity, HIV status and etc. Let $\mathbf{Z}_{ij}$ be a vector characterizing the attributes of the link connecting subjects i and j, which usually includes shared drug using activities, history of living together, and etc.

Researchers also observe characteristics of recruited subjects. Let $\mathbf{d} = (d_1, \ldots, d_n)$ be the vector of recruited subjects' reported degrees in G. For $i \in V_{SU}$, let $\mathbf{X}_i$ be a $p \times 1$ vector consisting of continuous or categorical trait values of i. Likewise for $i \in V_R$ and $j \in V_{SU}$ with $i \neq j$, let $\mathbf{Z}_{ij}$ be a $q \times 1$ vector consisting of continuous or categorical trait values corresponding to the edge $\{i, j\} \in E_{SU}$. Let $\mathbf{X}_R = \{\mathbf{X}_i : i \in V_R\}$ be the set of recruited subjects' trait values, and let $\mathbf{Z}_R = \{\mathbf{Z}_{ij} : i \in V_R, j \in V_R, i \neq j\}$ be the set of edge attributes. Likewise, let $\mathbf{X}_{SU} = \{\mathbf{X}_i : i \in V_{SU}\}$ be the set of trait values for subjects in V_{SU}, and let $\mathbf{Z}_{SU} = \{\mathbf{Z}_{ij} : i \in V_R, j \in V_{SU}, i \neq j\}$ be the set of edge attributes for vertices connected in G_{SU}. Finally, let $\mathbf{t} = (t_1, \ldots, t_n)$ be the vector of the times of recruitment of each subject into the study. Researchers typically only observe G_R, $\mathbf{d}$, $\mathbf{t}$, $\mathbf{X}_R$ and $\mathbf{Z}_R$ from the RDS recruitment

process, and do not observe any traits of unsampled subjects, non-recruitment links, or their attributes.

To construct a rigorous definition of uniform recruitment, we require knowledge of the augmented recruitment-induced subgraph G_{SU}. Within V_{SU} is every vertex that was connected to a recruiter at some point during the study. Therefore subsets of V_{SU} were eligible to be recruited at different times, and we will define uniform recruitment with respect to the traits of these subsets. We first define the set of vertices that can be recruited by a recruiter at each moment in time. Let $S_i(t)$ be the set of susceptible neighbors of i just before time t. If i recruits uniformly at random from the set $S_i(t)$ of susceptibles linked to it by an edge, then it exhibits no "preference" for recruiting one vertex over another.

Definition 1 (Uniform recruitment). *Recruitment is uniform if the probability that i recruits $k \in S_i(t)$ at time t is equal to $1/|S_i(t)|$.*

To obtain a statistical test, we need to develop a regression model whose coefficients control deviations from uniformity. The key idea is that estimating certain regression coefficients to be different from zero indicates non-uniform recruitment.

1.2 Survival interpretation of recruitment dynamics

The timing and order of recruitments gives information about recruiters' propensity to recruit some alters more readily than others. Consider a recruiter i connected to a susceptible individual $j \in S_i(t_i^+)$, where t_i^+ is a time instantaneously after t_i. Three possible outcomes can occur along the edge connecting i to j: 1) i successfully recruits j; 2) j is recruited by another recruiter before i can recruit j; or 3) j remains unrecruited by the time i depletes their coupons or the end of study, whichever happens first.

Let successful recruitment be the event of interest, and let T_{ij} be the random time to this event, starting from the time t_i that i enters the study. In case 1 above, successful recruitment of j by i allows researchers to observe $T_{ij} = t_j - t_i$ exactly. In cases 2 and 3, the recruitment is intervened, thus we only observe $T_{ij} > t^* - t_i$, where t^* is the time of the intervening event (recruitment by other subjects, depletion of coupons or end of study). In cases 2 and 3, T_{ij} is "censored" by an intervening event at time $\min\{t_j - t_i, t_i^* - t_i, t_f - t_i\}$, where t_i^* is the time at which i depletes its coupons and t_f is the end of the study. When T_{ij} is censored, researchers only observe that $T_{ij} > \min\{t_j, t_i^*, t_f\} - t_i$. Let $t_{ij} = \min\{t_j, t_i^*, t_f\} - t_i$. Thus researchers observe $T_{ij} = t_{ij}$ when i recruits j, and $T_{ij} > t_{ij}$ otherwise. This characterization of the recruitment process as a sequence of fully observed and censored waiting times suggests viewing the time dynamics of recruitment in a "survival analysis" framework. Figure 1 illustrates this argument for an example network, showing the recruitment graph G_R overlaid on G, and the waiting (survival) times T_{ij} along edges connecting recruiters i with susceptible vertices j. Suppose the random edge-wise recruitment waiting time T_{ij} has density function $f_{ij}(t)$ and cumulative distribution function $F_{ij}(t)$. We define the distribution of T_{ij} by specifying the recruitment hazard function, which is the probability of recruitment in an infinitesimal time interval: $\lambda_{ij}(t) = \lim_{h \to 0} P(t \le T_{ij} < t + h | T_{ij} \ge t)/h$. We stipulate that T_{ij}, and hence $\lambda_{ij}(t)$ are only defined when i and j are linked by an edge in G. Furthermore, $\lambda_{ij}(t)$ is nonzero only if i is a recruiter at time t.
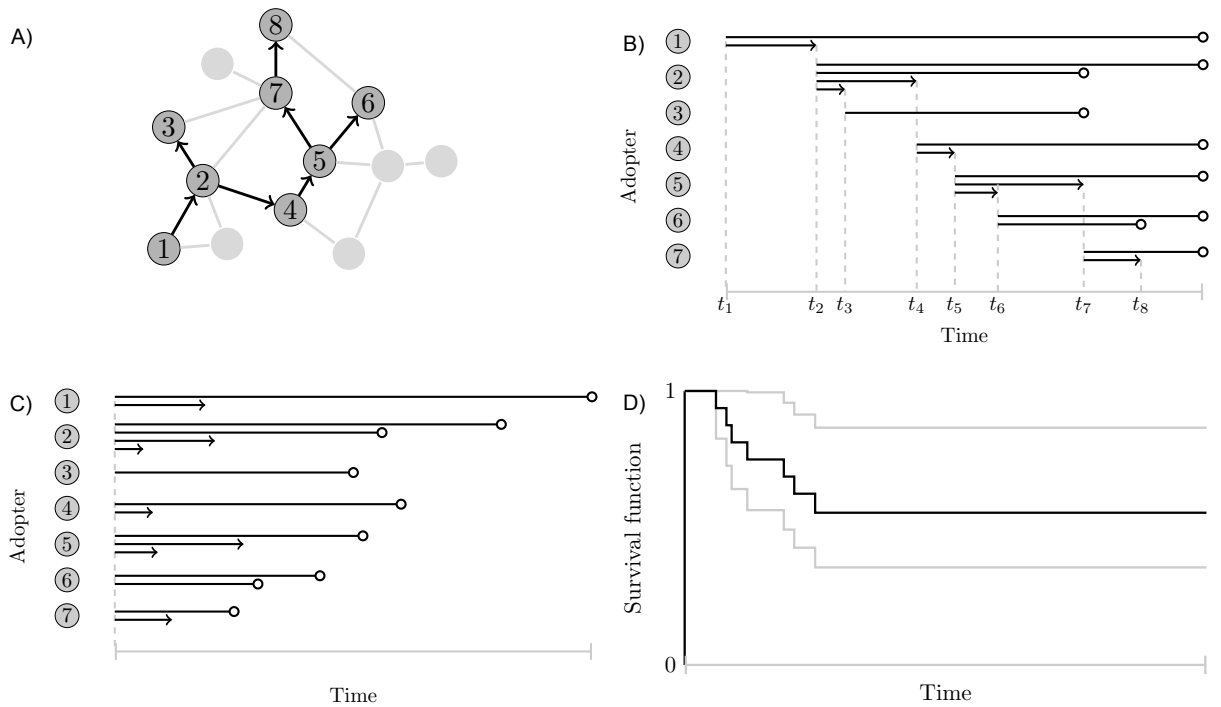
Figure 1. Survival interpretation of per-edge recruitment waiting times. A) Recruitment graph overlaid on the population graph, where $a \to b$ means a recruited b, and vertices are labeled in the order of their recruitment; B) per-edge waiting times to recruitment, where open circles indicate censored waiting times, and arrow tips indicate fully observed recruitment times; C) waiting times shifted to begin at time 0; D) edge-wise recruitment survival curve (black) with upper and lower confidence curves (gray).

When the subgraph of all recruiter-susceptible pairs G_{SU} is known, the observed data in an RDS study can be used to learn about the hazard of recruitment for different pairs i and j. Recruitment bias corresponds to an increase or decrease in the edge-wise waiting time to recruitment T_{ij}. By parameterizing the hazard function $\lambda_{ij}(t)$ of T_{ij} in terms of characteristics of i, j, and the edge connecting them, standard regression models can reveal recruitment bias, and perform statistical tests of uniform recruitment.

1.3   Regression models and testing for uniform recruitment

We include three different types of effects, based on 1) characteristics $\mathbf{X}_i$ of the recruiter i and $\mathbf{X}_j$ of the susceptible vertex j; 2) similarity of trait values $|\mathbf{X}_i - \mathbf{X}_j|$, where $|\cdot|$ indicates element-wise absolute value with respect to its vector argument; and 3) attributes $\mathbf{Z}_{ij}$ of the edge connecting i and j. We specify the edge-wise recruitment hazard of i recruiting $j \in S_i(t)$ at time $t > t_i$ as

$$\lambda_{ij}(t - t_i) = \lambda_0(t - t_i) \exp\left[\boldsymbol{\beta}' |\mathbf{X}_i - \mathbf{X}_j| + \boldsymbol{\theta}_1' \mathbf{X}_i + \boldsymbol{\theta}_2' \mathbf{X}_j + \boldsymbol{\kappa}' \mathbf{Z}_{ij} \right], \tag{1}$$

where $\lambda_0(\cdot)$ is a non-negative propensity to recruit (baseline hazard) shared by all recruiters. The regression coefficients $\boldsymbol{\beta}$, $\boldsymbol{\theta}_1$, and $\boldsymbol{\theta}_2$ have dimension $p \times 1$, and the coefficient $\boldsymbol{\kappa}$ has dimension $q \times 1$. The following two results make explicit connection between equation 1 and Definition 1 Before developing regression models to estimate the coefficients in equation 1, we state a result that makes explicit the connection between equation 1 and Definition 1.

Proposition 1. *Given that i recruits a susceptible vertex at time $t > t_i$, the probability that $j \in S_i(t)$ is the recruited vertex is*

$$\frac{\lambda_{ij}(t - t_i)}{\sum_{k \in S_i(t)} \lambda_{ik}(t - t_i)}. \tag{2}$$

Proof of Proposition 1. Let $J \in S_i(t)$ be the next recruited vertex. The joint density of the recruitment time t and the recruited vertex J is

$$\begin{aligned}
g_i(t, J = j) &= f_{ij}(t - t_i) \prod_{\substack{k \in S_i(t) \\ k \neq j}} (1 - F_{ik}(t - t_i)) \\
&= \lambda_{ij}(t - t_i) \exp\left[-\Lambda_{ij}(t - t_i) \right] \prod_{\substack{k \in S_i(t) \\ k \neq j}} \exp\left[-\Lambda_{ik}(t - t_i) \right] \\
&= \lambda_{ij}(t - t_i) \exp\left[-\sum_{k \in S_i(t)} \Lambda_{ik}(t - t_i) \right],
\end{aligned} \tag{3}$$

where $F_{ik}(\cdot)$ is the cumulative density function of the waiting time T_{ik} for k to be recruited by i. Since t is study time, thus we have subtracted it by t_i to make it edge-wise waiting time in the

above equations. Let T, r_j denote the random variables for recruitment time and recruiter of j respectively . Applying Bayes' Theorem, we find the conditional probability is

$$
\begin{aligned}
\Pr(J = j | T = t, r_j = i) &= \frac{g_i(t, J = j)}{\sum_{k \in S_i(t)} g_i(t, J = k)} \\
&= \frac{\lambda_{ij}(t - t_i) \exp\left[-\sum_{k \in S_i(t)} \Lambda_{ik}(t - t_i)\right]}{\sum_{l \in S_i(t)} \lambda_{il}(t - t_i) \exp\left[-\sum_{k \in S_i(t)} \Lambda_{ik}(t - t_i)\right]} \\
&= \frac{\lambda_{ij}(t - t_i)}{\sum_{l \in S_i(t)} \lambda_{il}(t - t_i)}
\end{aligned}
\tag{4}
$$

as claimed. □

We can now formalize a sufficient condition for uniform recruitment.

Corollary 1. *Suppose β, θ_2, κ in (1) are equal to zero. Then uniform recruitment holds under Definition 1.*

Proof of Corollary 1 follows by substituting zero for each coefficient in (1), and substituting the hazards into equation 2, so that conditional recruitment probability for every $j \in S_i(t)$ is $1/|S_i(t)|$. We call equation 2 the conditional probability of recruitment in what follows. Below we develop regression models for estimating these unknown coefficients, and statistical tests that β, θ_2, κ are equal to zero follow directly. Rejection of the test constitutes evidence against uniform recruitment.

1.4 Discrete-time regression model

One way to test uniform recruitment under Definition 1 is to condition on the times of recruitment events and the identity of the recruiter at each time, while treating the identity of the recruited subject as random. The likelihood $L_d(\beta, \theta_2, \kappa)$ of the discrete-time recruitment process is given by the product of the conditional recruitment probabilities, for each recruitment event. The likelihood is

$$
\begin{aligned}
L_d(\beta, \theta_2, \kappa) &= \prod_{j \notin M} \frac{\lambda_{r_j j}(t_j)}{\sum_{k \in S_{r_j}(t_j)} \lambda_{r_j k}(t_j)} \\
&= \prod_{j \notin M} \frac{\exp[\beta'|\mathbf{X}_{r_j} - \mathbf{X}_j| + \theta_2'\mathbf{X}_j + \kappa'\mathbf{Z}_{r_j j}]}{\sum_{k \in S_{r_j}(t_j)} \exp[\beta'|\mathbf{X}_{r_j} - \mathbf{X}_k| + \theta_2'\mathbf{X}_k + \kappa'\mathbf{Z}_{r_j k}]}
\end{aligned}
\tag{5}
$$

where r_j is the recruiter of j. The above expression follows because the terms $\lambda_0(t - t_i)$ and $\exp(\theta_1'X_i)$ cancel in each conditional recruitment probability. Fitting this model by maximum likelihood is straightforward using a Newton-Raphson algorithm. To determine whether uniform recruitment holds, a simultaneous test over all coefficients can be conducted using a likelihood ratio test.

The null hypothesis of uniform recruitment under Definition 1 is that every element of $\boldsymbol{\beta}$, $\boldsymbol{\theta}_2$, and $\boldsymbol{\kappa}$ is zero. Suppose we estimate $\hat{\boldsymbol{\gamma}} = (\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\kappa}})$ by maximum likelihood. The elements of these estimated coefficient vectors are asymptotically normally distributed with estimated variance-covariance matrix $\widehat{\mathbf{V}}$. To simultaneously test this hypothesis for every coefficient, we perform an "omnibus" F-test that all elements of $\boldsymbol{\gamma}$ are zero. The test statistic $F_0 = \hat{\boldsymbol{\gamma}}' \widehat{V}^{-1} \hat{\boldsymbol{\gamma}} / (2p + q)$ has F distribution with $2p + q$ and $n - (2p + q)$ degrees of freedom. Rejecting the null hypothesesis provides evidence of non-uniform recruitment.

1.5 Continuous-time regression model

The discrete-time regression model equation 5 captures the propensity of a recruiter i to recruit $j \in S_i(t)$, given that i recruits some susceptible vertex at time t. However, equation 5 does not capture the relationship of the time of recruitment to these propensities, and instead conditions on the time of recruitment and treats the identity of the recruiter as random. By taking a continuous-time view of recruitment, we develop a regression model that makes use of all the timing information contained in the data. The set $S_i(t)$ of vertices susceptible to recruitment by i may not be constant in time t. Consider a potential recruiter i and its susceptible neighbors $S_i(t)$ at some time $t > t_i$. Two types of events can change the value of $S_i(t)$. First, $j \in S_i(t)$ may be recruited by i at time $t_j > t$. If i has remaining coupons, then $|S_i(t_j^+)| = |S_i(t_j)| - 1$, otherwise coupons are depleted and $|S_i(t_j^+)| = 0$. Second, $j \in S_i(t_j)$ may be recruited by another recruiter at time $t_j > t$, and $|S_i(t_j^+)| = |S_i(t_j)| - 1$. The likelihood for the continuous-time process is

$$L_c(\boldsymbol{\beta}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\kappa}) = \prod_{i=1}^{n} \prod_{j \in S_i(t_i^+)} \lambda_{ij}(t_{ij})^{\mathbb{1}\{r_j = i\}} \exp\left[-\Lambda_{ij}(t_{ij})\right] \tag{6}$$

where again we have suppressed conditioning on the observed data, G_{SU}, $\mathbf{X}_{SU}$, and $\mathbf{Z}_{SU}$ for notational clarity. where r_j is the recruiter of j and $\Lambda_{ij}(\cdot)$ is the cumulative hazard function. The continuous-time likelihood (6) corresponds to a standard survival model for edge-wise recruitment time with censoring. Furthermore, in the continuous-time likelihood the baseline hazard and recruiter-specific coefficients $\boldsymbol{\theta}_1$ do not cancel. Allowing $\lambda_0(t)$ to be unspecified suggests use of the semi-parametric Cox proportional hazards model [1]. Alternatively, letting the baseline recruitment hazard take the parametric form $\lambda_0(t) = \rho t^{\rho - 1}$, the edge-wise waiting time T_{ij} to recruitment of j by i has Weibull distribution. As before, if coefficients $\boldsymbol{\beta}, \boldsymbol{\theta}_2, \boldsymbol{\kappa}$ in the regression model are zero, then recruitment is uniform. The likelihood ratio test introduced above also applies here, and we use standard software for maximization of the likelihood [4].

2 Likelihood ratio test for the discrete-time model

The null hypothesis of uniform recruitment under Definition 1 is that every element of $\boldsymbol{\beta}$, $\boldsymbol{\theta}_2$, and $\boldsymbol{\kappa}$ is zero. For simplicity, we denote $\boldsymbol{\gamma} = (\boldsymbol{\beta}, \boldsymbol{\theta}_2, \boldsymbol{\kappa})$ as a vector where $\boldsymbol{\beta}, \boldsymbol{\theta}_2$ and $\boldsymbol{\kappa}$ have length p, p and q respectively. The coefficients estimation under null hypothesis is zero vector $\boldsymbol{\gamma}_0 = \mathbf{0}$. And let $\hat{\boldsymbol{\gamma}} = (\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}_2, \hat{\boldsymbol{\kappa}})$ be the maximum likelihood estimation calculated from the

Newton Raphson algorithm mentioned above. Then the statistic

$$\chi^2 = 2 \log \frac{L_d(\hat{\boldsymbol{\gamma}})}{L_d(\boldsymbol{\gamma}_0)} \tag{7}$$

asymptotically follows a χ^2 distribution with degrees of freedom $2p + q$. Rejecting this null hypothesis provides evidence of non-uniform recruitment.

3   Frailty regression model

In survival analysis, frailty models are a common way of allowing for variation between units or clusters, and in RDS studies, recruiters are a natural unit of clustering for edge-wise waiting times. We again estimate parameters under the Weibull regression framework as in Section 3.2, but with a random recruiter-specific intercept ($\sigma\phi_i$) added to each log-hazard, essentially allowing recruiters to have different individual propensities to recruit. The hazard function for frailty model takes the following form:

$$\lambda_{ij}(t) = \exp(\alpha + \sigma\phi_i + \boldsymbol{\beta}'|\mathbf{X}_i - \mathbf{X}_j| + \theta_1'\mathbf{X}_i + \theta_2'\mathbf{X}_j + \kappa'\mathbf{Z}_{ij})\rho t^{\rho-1},$$

where ϕ_i's are assumed to be i.i.d samples from $N(0, 1)$.

4   Newton Raphson algorithm for the discrete model

For notation simplicity, we denote $\boldsymbol{\gamma} = (\beta, \boldsymbol{\theta}_2, \boldsymbol{\kappa})$ and $\eta_{r_j i} = (|X_{r_j} - X_i|, X_i, \mathbf{Z}_{r_j i})$, where r_j is the recruiter of subject j. Then maximizing equation (5) is the same as maximizing:

$$
\begin{aligned}
l(\boldsymbol{\gamma}) &= \log L_d(\boldsymbol{\gamma}) \\
&= \sum_{j \notin M} \{\eta_{r_j j}' \boldsymbol{\gamma} - \log \big[\sum_{k \in S_{r_j}(t_j)} \exp(\eta_{r_j k}' \boldsymbol{\gamma}) \big] \}
\end{aligned}
$$

The Newton Raphson algorithm requires calculation of first and second order derivatives of target function. The first order derivative (gradient) is

$$
l'(\boldsymbol{\gamma}) = \sum_{j \notin M} \Big[\eta_{r_j j} - \frac{\sum_{i \in S_{r_j}(t_j)} \eta_{r_j i} \exp(\eta_{r_j i}' \boldsymbol{\gamma})}{\sum_{k \in S_{r_j}(t_j)} \exp(\eta_{r_j k}' \boldsymbol{\gamma})} \Big]
$$

and the second order derivative (Hessian matrix) is

$$
\begin{aligned}
l''(\boldsymbol{\gamma}) = &\sum_{j \notin M} \frac{[\sum_{i \in S_{r_j}(t_j)} \eta_{r_j i} \exp(\eta_{r_j i}' \boldsymbol{\gamma})][\sum_{k \in S_{r_j}(t_j)} \eta_{r_j k}' \exp(\eta_{r_j k}' \boldsymbol{\gamma})]}{[\sum_{k \in S_{r_j}(t_j)} \exp(\eta_{r_j k}' \boldsymbol{\gamma})]^2} \\
&- \sum_{j \notin M} \frac{[\sum_{i \in S_{r_j}(t_j)} \eta_{r_j i} \eta_{r_j i}' \exp(\eta_{r_j i}' \boldsymbol{\gamma})][\sum_{k \in S_{r_j}(t_j)} \exp(\eta_{r_j k}' \boldsymbol{\gamma})]}{[\sum_{k \in S_{r_j}(t_j)} \exp(\eta_{r_j k}' \boldsymbol{\gamma})]^2}.
\end{aligned}
$$

Then we can implement a Newton-Raphson algorithm following the steps below:

Step 1 Assign γ_{old} as an initial value of γ

Step 2 Update $\gamma_{new} = \gamma_{old} - \alpha \, l''(\gamma_{old})^{-1} \cdot l'(\gamma_{old})$, where $l''(\gamma_{old})^{-1} \cdot l'(\gamma_{old})$ is the direction that causes the most increase in target function and α is a user-specified step-size parameter. In our practice, we set $\alpha = 0.5$.

Step 3 If $\max|\gamma_{old} - \gamma_{new}| > 10^{-5}$, let $\gamma_{old} = \gamma_{new}$ and return to Step 2. Otherwise, output $\hat{\gamma} = \gamma_{new}$ as solution.

After $\hat{\gamma}$ is calculated, the estimated variance-covariance matrix $\widehat{\Sigma}$ is calculated as

$$\widehat{\Sigma} = [-l''(\hat{\gamma})]^{-1}.$$

Diagonal elements of $\widehat{\Sigma}$ are the asymptotic variances of $\hat{\gamma}$.

5 Supplemetary Tables

5.1 Results for Cox proportional hazards regression model

Covariate	γ	$\exp(\gamma)$	95% CI of $\exp(\gamma)$
drug using	-0.64	0.53	0.325 , 0.860
drug injection	-0.11	0.90	0.550 , 1.463
needle using	-0.03	0.97	0.553 , 1.699
sharing activity	0.52	1.68	1.074 , 2.629
sex	0.69	2.00	1.224 , 3.259
hiv positive	-0.07	0.93	0.583 , 1.498
recruiter's gender	0.39	1.47	1.045 , 2.075
recruiter's crack using	-0.09	0.92	0.698 , 1.205
recruiter's homelessness	0.09	1.09	0.821 , 1.451
recruiter's age	-0.01	0.99	0.973 , 1.005
alter's gender	0.03	1.03	0.734 , 1.437
alter's crack using	-0.28	0.76	0.582 , 0.988
alter's homelessness	0.23	1.26	0.961 , 1.647
alter's age	0.03	1.03	1.013 , 1.040
recruiter black	-0.66	0.51	0.212 , 1.248
recruiter white	-0.15	0.86	0.354 , 2.112
alter black	0.30	1.35	0.608 , 2.992
alter white	0.36	1.43	0.651 , 3.140
gender difference	0.10	1.11	0.792 , 1.545
crack using difference	0.19	1.20	0.929 , 1.560
homelessness difference	-0.47	0.62	0.477 , 0.814
age difference	-0.0021	1.00	0.981 , 1.015
black difference	0.67	1.96	0.800 , 4.769
white difference	-0.15	0.86	0.353 , 2.117
hispanic difference	-0.22	0.80	0.348 , 1.844

Table 1. Estimated regression coefficients for the Cox Proportional Hazards Model of edgewise recruitment times. The global test of the proportional hazard assumption gives a Chi-square statistic $\chi^2 = 32.98$ and p-value 0.132.

* CI: confidence interval

5.2 Results for Weibull frailty regression model

Covariate	γ	$\exp(\gamma)$	95% CI of $\exp(\gamma)$
Intercept	-6.66	0.001	0.000 , 0.006
drug using	-0.78	0.46	0.249 , 0.841
drug injection	-0.33	0.72	0.389 , 1.338
needle using	-0.20	0.82	0.405 , 1.649
sharing activity	0.99	2.70	1.514 , 4.821
sex	0.92	2.50	1.347 , 4.639
hiv positive	0.23	1.26	0.715 , 2.218
recruiter's gender	0.74	2.09	1.206 , 3.634
recruiter's crack using	-0.05	0.95	0.603 , 1.502
recruiter's homelessness	0.13	1.14	0.712 , 1.835
recruiter's age	-0.02	0.98	0.957 , 1.008
alter's gender	0.03	1.03	0.697 , 1.528
alter's crack using	-0.37	0.69	0.505 , 0.952
alter's homelessness	0.33	1.39	1.012 , 1.899
alter's age	0.04	1.04	1.022 , 1.055
recruiter black	-0.71	0.49	0.163 , 1.492
recruiter white	-0.13	0.88	0.284 , 2.711
alter black	0.40	1.49	0.595 , 3.744
alter white	0.48	1.61	0.647 , 4.013
gender difference	0.14	1.15	0.771 , 1.716
crack using difference	0.27	1.31	0.959 , 1.785
homelessness difference	-0.60	0.55	0.400 , 0.750
age difference	-0.01	0.99	0.973 , 1.015
black difference	0.93	2.53	0.906 , 7.075
white difference	0.01	1.01	0.352 , 2.891
hispanic difference	-0.21	0.81	0.307 , 2.131

Table 2. Estimated coefficients for the Weibull Frailty Model of edge-wise recruitment times.

* CI: confidence interval

References

[1] David R Cox. Regression models and life-tables. In *Breakthroughs in Statistics*, pages 527–541. Springer, 1992.

[2] Forrest W. Crawford. The graphical structure of respondent-driven sampling. *Sociological Methodology*, 46:187–211, 2016.

[3] Forrest W Crawford, Peter M Aronow, Li Zeng, and Jianghong Li. Identification of homophily and preferential recruitment in respondent-driven sampling. *American Journal of Epidemiology*, 187(1):153–160, 2017.

[4] Terry M Therneau and Thomas Lumley. Package 'survival', 2016.