WEBVTT

1 00:00:59.160 --> 00:01:00.210 <v Vin>Donna's looking over it.</v>

2 00:01:00.210 --> 00:01:01.083 I'll just start.

3 00:01:04.440 --> 00:01:06.573 So can we hear us okay online?

4 00:01:08.336 --> 00:01:10.012 <v Donna>Yeah, if you want you can go to the podium.</v>

5 00:01:10.012 --> 00:01:11.050 <v ->Yeah.</v> <v ->Okay.</v>

6 00:01:11.050 --> 00:01:12.860 'Cause this is last-minute,

7 00:01:12.860 --> 00:01:15.096 so I need to just get your bio (indistinct).

8 00:01:15.096 --> 00:01:17.944 (laughing)

9 00:01:17.944 --> 00:01:19.361 So, hi, everyone.

10 00:01:20.430 --> 00:01:23.760 It's my pleasure to welcome Dr. Ashley Buchanan

11 00:01:23.760 --> 00:01:26.700 today as our speaker in this seminar series.

12 00:01:26.700 --> 00:01:31.230 And Dr. Buchanan is associate professor of biostatistics

13 00:01:31.230 --> 00:01:34.020 in the Department of Pharmacy Practice

14 00:01:34.020 --> 00:01:35.490 in University Rhode Island

15 00:01:35.490 --> 00:01:39.690 and also as an adjunct in Brown University Biostatistics.

16 00:01:39.690 --> 00:01:40.770 Hi, Donna. <v ->Hi.</v>

17 00:01:40.770 --> 00:01:43.440 <v Vin>And she specializes in the area</v>

18 00:01:43.440 --> 00:01:45.270 of epidemiology and causal inference.

19 00:01:45.270 --> 00:01:46.860 And she has a lot of experiences

20 00:01:46.860 --> 00:01:49.410 collaborating on HIV/AIDS research,

21 00:01:49.410 --> 00:01:52.290 work closely with colleagues both domestically

22 00:01:52.290 --> 00:01:54.120 and internationally to develop

23 00:01:54.120 --> 00:01:57.330 and apply causal methods to improve treatment

24 00:01:57.330 --> 00:01:59.460 and prevention of HIV and AIDS.

25 00:01:59.460 --> 00:02:01.740 And without further ado,

26 00:02:01.740 --> 00:02:04.185 I'll give the floor to you, Ashley.

27 00:02:04.185 --> 00:02:06.420 (indistinct)

28 00:02:06.420 --> 00:02:08.550 <v ->Thanks, Vin, for that nice introduction.</v>

29 00:02:08.550 --> 00:02:10.380 And thanks for the invitation, Donna,

30 00:02:10.380 --> 00:02:12.120 to speak at (indistinct) today.

31 00:02:12.120 --> 00:02:14.550 It's nice to be here in person with folks

32 00:02:14.550 --> 00:02:16.650 that I normally just see on Zoom.

33 00:02:16.650 --> 00:02:17.880 So great to be here.

34 00:02:17.880 --> 00:02:20.763 And welcome to all the folks on Zoom, as well.

35 00:02:22.440 --> 00:02:23.703 Just get my slide.

36 00:02:26.400 --> 00:02:29.310 I see that the slides are already sharing.

37 00:02:29.310 --> 00:02:30.460 Let's do the slideshow.

38 00:02:32.040 --> 00:02:32.873 Oops.

39 00:02:34.170 --> 00:02:36.330 <v Vin>It's the lower right.</v>

40 00:02:36.330 --> 00:02:37.470 It's a little-

41 00:02:37.470 --> 00:02:40.677 <v ->Is this gonna work? (drowned out)</v>

42 00:02:40.677 --> 00:02:43.143 Can the folks on Zoom still see the slides?

43 00:02:44.010 --> 00:02:44.940 <v Vin>You did share, right?</v>

44 00:02:44.940 --> 00:02:47.091 <v ->Yeah, I think it's sharing.</v>

45 00:02:47.091 --> 00:02:48.115 <v Gabrielle>We see a full screen.</v>

46 00:02:48.115 --> 00:02:50.050 (indistinct)

47 00:02:50.050 --> 00:02:51.393 <v ->Okay, great.</v> <v ->Perfect.</v>

48 00:02:53.070 --> 00:02:55.830 <v ->Okay, so today, I'm gonna be presenting work</v>

49 00:02:55.830 --> 00:02:57.420 about study design, power,

50 00:02:57.420 --> 00:02:59.580 and sample size calculation for evaluating

51 00:02:59.580 --> 00:03:01.980 spillover in networks in the context

52 00:03:01.980 --> 00:03:04.500 of the interventions not randomized.

53 00:03:04.500 --> 00:03:07.680 This is definitely work in progress, ongoing work.

54 00:03:07.680 --> 00:03:11.670 So we have some initial simulation results

55 00:03:11.670 --> 00:03:13.350 and some promising findings

56 00:03:13.350 --> 00:03:15.360 and then a lot of open questions

57 00:03:15.360 --> 00:03:18.570 that I'd love to have some discussion about towards the end,

58 00:03:18.570 --> 00:03:21.090 sort of about where the practical world

59 00:03:21.090 --> 00:03:22.560 meets the statistical world,

60 00:03:22.560 --> 00:03:26.430 and how can we bring these ideas into practice

61 00:03:26.430 --> 00:03:28.563 for designing these network type studies.

62 00:03:30.376 --> 00:03:32.220 I'd like to start off with acknowledgements.

63 00:03:32.220 --> 00:03:36.134 So Ke Zhang is a graduate student at URI,

64 00:03:36.134 --> 00:03:38.010 and she's been primarily leading

65 00:03:38.010 --> 00:03:39.210 a lot of the simulation work.

66 00:03:39.210 --> 00:03:42.901 She's been a key individual in this work.

67 00:03:42.901 --> 00:03:46.050 We also have collaborators, Doctors Katenka, Wu, and Lee.

68 00:03:46.050 --> 00:03:48.810 And then I also wanna thank a larger list of collaborators

69 00:03:48.810 --> 00:03:52.140 that have been part of this ongoing work with Avenir,

70 00:03:52.140 --> 00:03:54.157 including Dr. Lee, Forastieri,

71 00:03:54.157 --> 00:03:56.743 Halleran, Friedman, and Nichopoulos.

72 00:03:57.600 --> 00:04:00.180 And then just to acknowledge our funding support

73 00:04:00.180 --> 00:04:04.473 and funding support that collected the motivating data set.

74 00:04:06.780 --> 00:04:08.103 So an outline for today,

75 00:04:09.730 --> 00:04:10.680 I'm gonna give a little bit of background

76 00:04:10.680 --> 00:04:12.990 and talk about the motivating study of TRIP,

77 00:04:12.990 --> 00:04:15.660 talk about the objectives of this particular work.

78 00:04:15.660 --> 00:04:18.810 And then we'll look at some of the simulation results

79 00:04:18.810 --> 00:04:22.173 and then discuss conclusions and future directions.

80 00:04:23.610 --> 00:04:28.140 So this work is focused on people who inject drugs,

81 00:04:28.140 --> 00:04:30.780 and these individuals are at risk for HIV

82 00:04:30.780 --> 00:04:33.750 due to drug use, sharing equipment,

83 00:04:33.750 --> 00:04:35.580 and sexual risk behaviors.

84 00:04:35.580 --> 00:04:38.970 In addition, these individuals are often part of networks.

85 00:04:38.970 --> 00:04:41.550 So when they receive an intervention,

86 00:04:41.550 --> 00:04:43.890 the intervention can benefit not only them

87 00:04:43.890 --> 00:04:47.430 but their partners and possibly even beyond that.

88 00:04:47.430 --> 00:04:49.590 So in these networks, interventions often have

89 00:04:49.590 --> 00:04:51.450 what's known as spillover effects,

90 00:04:51.450 --> 00:04:53.040 sometimes called the indirect effect

91 00:04:53.040 --> 00:04:55.170 in interference literature.

92 00:04:55.170 --> 00:04:57.540 So spillover,

93 00:04:57.540 --> 00:05:00.390 historically in the causal inference literature,

94 00:05:00.390 --> 00:05:02.130 it's been called interference.

95 00:05:02.130 --> 00:05:04.080 Here I'll be calling it spillover.

96 00:05:04.080 --> 00:05:06.360 So that's when one individual's exposure

97 00:05:06.360 --> 00:05:09.270 affects another's outcome.

98 00:05:09.270 --> 00:05:12.210 And recently, there's been several papers

99 00:05:12.210 --> 00:05:14.250 that have been looking at how do we assess

100 00:05:14.250 --> 00:05:17.283 these spillover effects in network studies.

101 00:05:20.640 --> 00:05:21.750 So our motivating study

102 00:05:21.750 --> 00:05:25.020 is the Transmission Reduction Intervention Project.

103 00:05:25.020 --> 00:05:27.480 This was a network-based study of injection drug users

104 00:05:27.480 --> 00:05:32.367 and their contacts in Athens, Greece, 2013 to 2015.

105 00:05:32.367 --> 00:05:34.950 And the individuals were connected

106 00:05:34.950 --> 00:05:37.470 through sexual and drug use partnerships.

107 00:05:37.470 --> 00:05:40.020 The original study was focused on using

108 00:05:40.020 --> 00:05:41.940 this new network tracing technique

109 00:05:41.940 --> 00:05:44.370 to find recently infected individuals

110 00:05:44.370 --> 00:05:45.780 and get them on treatment.

111 00:05:45.780 --> 00:05:49.320 So the idea is when individuals are acutely infected,

112 00:05:49.320 --> 00:05:50.700 they're more likely to transmit.

113 00:05:50.700 --> 00:05:51.780 So if we can find more

114 00:05:51.780 --> 00:05:53.940 of these recently infected individuals,

115 00:05:53.940 --> 00:05:55.140 get them on treatment,

116 00:05:55.140 --> 00:05:57.720 they'll be less likely to infect their partners.

117 00:05:57.720 --> 00:05:59.370 And the punchline from the main study

118 00:05:59.370 --> 00:06:01.440 was this was very successful in finding

119 00:06:01.440 --> 00:06:03.543 more recently infected individuals.

120 00:06:04.500 --> 00:06:06.043 <v Ke>Excuse me.</v>

121 00:06:06.043 --> 00:06:07.170 <v Ashley>What?</v>

122 00:06:07.170 --> 00:06:08.670 <v Ke>I'm so sorry for the bothering,</v>

123 00:06:08.670 --> 00:06:11.433 but from my end, the slides are not moving.

124 00:06:13.290 --> 00:06:15.090 <v ->Not at all, okay, let me try again.</v>

125 00:06:15.090 --> 00:06:16.753 One second. (indistinct)

126 00:06:16.753 --> 00:06:18.687 (Donna laughing)

127 00:06:18.687 --> 00:06:20.487 <v Donna>For one more day (laughs).</v>

128 00:06:21.960 --> 00:06:24.030 <v Vin>At least the (indistinct), so that's okay.</v>

129 00:06:24.030 --> 00:06:25.914 <v ->Yeah, yeah, we haven't made it too far.</v>

130 00:06:25.914 --> 00:06:27.507 (laughing)

131 00:06:27.507 --> 00:06:28.346 <v Donna>Thanks for telling us.</v>

132 00:06:28.346 --> 00:06:30.053 <v Vin>Thanks for letting us know.</v>

133 00:06:34.250 --> 00:06:35.373 How 'bout now?

134 00:06:37.170 --> 00:06:40.326 <v Gabrielle>Yep, we can see the motivating study slide.</v>

135 00:06:40.326 --> 00:06:41.159 <v ->[Donna And Ashley] Okay.</v>

136 00:06:41.159 --> 00:06:41.992 Is it the slide?

137 00:06:41.992 --> 00:06:44.850 Is it in presentation view or is it the slide?

138 00:06:44.850 --> 00:06:46.200 <v ->On the right-hand side,</v>

139 00:06:46.200 --> 00:06:49.557 we can see the next slide and then some notes.

140 00:06:49.557 --> 00:06:50.940 <v ->Oh, so it's in presentation.</v>

141 00:06:50.940 --> 00:06:52.080 I mean, that's not the worst thing,

142 00:06:52.080 --> 00:06:55.950 but sometimes, it's better if they can

143 00:06:55.950 --> 00:06:58.410 just see the whole slide (laughs).

144 00:06:58.410 --> 00:06:59.373 Sorry about that.

145 00:07:03.120 --> 00:07:04.123 <v ->Think you'll have to maybe go</v>

146 00:07:04.123 --> 00:07:06.570 out of the presentation mode.

147 00:07:06.570 --> 00:07:07.890 <v ->Exit presentation mode.</v>

148 00:07:07.890 --> 00:07:11.730 <v Vin>Yeah, so then it's the same in the computer</v>

149 00:07:11.730 --> 00:07:12.993 and the screen sharing.

150 00:07:26.926 --> 00:07:27.759 <v ->Sorry.</v>

151 00:07:30.030 --> 00:07:31.380 How do you do it, Vin?

152 00:07:31.380 --> 00:07:33.150 <v Vin>Just that little button, yeah.</v>

153 00:07:33.150 --> 00:07:34.620 You're actually on it right now.

154 00:07:34.620 --> 00:07:35.870 <v ->I think they're still see-</v> <v ->If you could just click</v>

155 00:07:35.870 --> 00:07:36.703 on that.

156 00:07:36.703 --> 00:07:38.966 <v Donna>Or you can go to the top bar, too, I think.</v>

157 00:07:38.966 --> 00:07:39.836 And there we go. <v ->No, I think they'll</v>

158 00:07:39.836 --> 00:07:40.669 still see that.

159 00:07:40.669 --> 00:07:42.480 <v ->And then I think over here,</v>

160 00:07:42.480 --> 00:07:46.574 maybe there's a way to even exit presentation mode.

161 00:07:46.574 --> 00:07:49.068 (indistinct)

162 00:07:49.068 --> 00:07:50.144 It's that.

163 00:07:50.144 --> 00:07:52.928 <v ->(indistinct) slidehow.</v>

164 00:07:52.928 --> 00:07:55.345 (indistinct)

165 00:08:00.238 --> 00:08:04.201 <v ->(indistinct) if there's any.</v>

166 00:08:04.201 --> 00:08:06.330 (indistinct) presenter view.

167 00:08:06.330 --> 00:08:07.503 <v ->There we go.</v>

168 00:08:07.503 --> 00:08:08.339 <v ->Okay, thanks, Vin.</v>

169 00:08:08.339 --> 00:08:10.413 <v ->Does that look okay for (drowned out)?</v>

170 00:08:10.413 --> 00:08:12.123 (laughing) (indistinct)

171 00:08:12.123 --> 00:08:14.919 <v Gabrielle>Yep, now, it's in presentation mode.</v>

172 00:08:14.919 --> 00:08:16.779 <v ->Okay, great.</v>

173 00:08:16.779 --> 00:08:20.310 Sorry about that, thanks for your patience.

174 00:08:20.310 --> 00:08:21.390 So where were we; so we were talking

175 00:08:21.390 --> 00:08:24.120 about the Transmission Reduction Intervention Project.

176 00:08:24.120 --> 00:08:25.590 So this worked well to find

177 00:08:25.590 --> 00:08:27.750 these recently infected individuals

178 00:08:27.750 --> 00:08:29.220 and refer them to treatment.

179 00:08:29.220 --> 00:08:31.890 So it was this successful strategic network

180 00:08:31.890 --> 00:08:33.240 tracing approach.

181 00:08:33.240 --> 00:08:34.200 In addition in this study,

182 00:08:34.200 --> 00:08:36.210 they also delivered community alerts.

183 00:08:36.210 --> 00:08:37.530 So if there is an individual

184 00:08:37.530 --> 00:08:41.310 who was recently infected in the network...

185 00:08:41.310 --> 00:08:43.890 Get this outta the way so you guys can see the figure.

186 00:08:43.890 --> 00:08:45.750 There's an individual who was recently infected

187 00:08:45.750 --> 00:08:50.010 in the proximity of a particular individual in the network,

188 00:08:50.010 --> 00:08:51.870 these community alerts would be distributed,

189 00:08:51.870 --> 00:08:54.600 which were basically flyers, handouts,

190 00:08:54.600 --> 00:08:59.600 or flyers even posted on the wall of frequented venues.

191 00:09:00.480 --> 00:09:02.190 So then individuals in the network

192 00:09:02.190 --> 00:09:04.350 either received these community alerts

193 00:09:04.350 --> 00:09:06.180 from the investigators or they did not.

194 00:09:06.180 --> 00:09:09.150 So the little red dots are those individuals

195 00:09:09.150 --> 00:09:10.380 who received the alerts.

196 00:09:10.380 --> 00:09:13.080 And then the blue ones are those who were not alerted.

197 00:09:14.100 --> 00:09:17.190 And then we looked at this in our previous paper.

198 00:09:17.190 --> 00:09:19.459 We looked at the spillover effects of the community alerts

199 00:09:19.459 --> 00:09:23.010 on HIV injection risk behavior at six months

200 00:09:23.010 --> 00:09:25.620 to see if receiving this alert yourself

201 00:09:25.620 --> 00:09:27.180 reduced your injection risk behavior.

202 00:09:27.180 --> 00:09:30.540 Or if you had contacts who were alerted,

203 00:09:30.540 --> 00:09:33.300 then did that information spill over to you,

204 00:09:33.300 --> 00:09:36.603 and then you also reduced your injection risk behavior?

205 00:09:44.190 --> 00:09:46.830 <v Donna>So is that the actual network, that picture?</v>

206 00:09:46.830 --> 00:09:50.340 <v ->Yep, that's the visualization of the network among...</v>

207 00:09:50.340 --> 00:09:51.390 There's some missing data

208 00:09:51.390 --> 00:09:53.130 and this problem system among the individuals

209 00:09:53.130 --> 00:09:55.863 that had all the outcomes observed.

210 00:09:57.150 --> 00:09:58.750 Okay, good, the slides can move.

211 00:09:59.940 --> 00:10:01.170 So I'm just gonna,

212 00:10:01.170 --> 00:10:03.300 for those who are not familiar with networks,

213 00:10:03.300 --> 00:10:06.300 I'll define some terminology using this slide.

214 00:10:06.300 --> 00:10:08.910 So this is a visualization of the network here,

215 00:10:08.910 --> 00:10:10.440 the TRIP network.

216 00:10:10.440 --> 00:10:13.770 There's 216 individuals here.

217 00:10:13.770 --> 00:10:16.110 So the individuals are denoted by the blue dots.

218 00:10:16.110 --> 00:10:17.700 Those are people who inject drugs

219 00:10:17.700 --> 00:10:20.190 and their sexual and drug use partners.

220 00:10:20.190 --> 00:10:24.360 And then the edges represent when two individuals,

221 00:10:24.360 --> 00:10:26.040 or nodes, share a partnership.

222 00:10:26.040 --> 00:10:28.980 And we call those connections edges sometimes.

223 00:10:28.980 --> 00:10:33.870 And then the little pink one is an example of a component.

224 00:10:33.870 --> 00:10:35.820 So that's a connected subnetwork

225 00:10:35.820 --> 00:10:38.040 for individuals in that group are connected

226 00:10:38.040 --> 00:10:39.660 to each other through at least one path

227 00:10:39.660 --> 00:10:42.570 but not connected to others in the network.

228 00:10:42.570 --> 00:10:45.330 So right away, we see that TRIP primarily comprised

229 00:10:45.330 --> 00:10:47.970 this one, large, connected component

230 00:10:47.970 --> 00:10:50.340 and several other small components.

231 00:10:50.340 --> 00:10:54.030 We can sort of see them out on the edges of the network.

232 00:10:54.030 --> 00:10:57.000 And then when we zoom in on the component,

233 00:10:57.000 --> 00:11:01.020 the individual in red is the,

234 00:11:01.020 --> 00:11:02.790 we'll call that the index person.

235 00:11:02.790 --> 00:11:05.160 And then the individuals shaded

236 00:11:05.160 --> 00:11:07.590 in this lighter pink are their neighbors

237 00:11:07.590 --> 00:11:09.750 or their first-degree contacts.

238 00:11:09.750 --> 00:11:11.790 So as I go through presenting these methods,

239 00:11:11.790 --> 00:11:14.490 there are some times when I'll be talking about components.

240 00:11:14.490 --> 00:11:16.920 And then in terms of defining the spillover effects,

241 00:11:16.920 --> 00:11:18.030 in this particular paper,

242 00:11:18.030 --> 00:11:21.333 we defined it using the exposure of the nearest neighbors.

243 00:11:24.030 --> 00:11:24.870 <v Donna>By nearest neighbors,</v>

244 00:11:24.870 --> 00:11:26.100 you mean just first-degree (drowned out)?

245 00:11:26.100 --> 00:11:28.470 <v ->First-degree, yeah, it may be said</v>

246 00:11:28.470 --> 00:11:30.660 even more applied to their partners.

247 00:11:30.660 --> 00:11:32.100 <v ->Okay.</v> <v ->Right, so we're really</v>

248 00:11:32.100 --> 00:11:35.550 thinking about their immediate partners,

249 00:11:35.550 --> 00:11:36.570 and these would be individuals

250 00:11:36.570 --> 00:11:38.907 that they either used drugs with or had sex with,

251 00:11:38.907 --> 00:11:40.590 and they reported that in the study

252 00:11:40.590 --> 00:11:41.913 for that edge to be there.

253 00:11:42.930 --> 00:11:43.763 Yep.

254 00:11:47.190 --> 00:11:48.900 So a little bit of notation.

255 00:11:48.900 --> 00:11:52.663 So we have N is denoting the participants in the study.

256 00:11:52.663 --> 00:11:54.510 A is going to be the intervention

257 00:11:54.510 --> 00:11:57.270 based on the community alerts in our example.

258 00:11:57.270 --> 00:11:58.980 We have baseline covariates,

259 00:11:58.980 --> 00:12:02.310 and then we index the neighbor, the partners who were...

260 00:12:02.310 --> 00:12:04.110 I guess in the networks they call it the neighbors.

261 00:12:04.110 --> 00:12:06.480 But in this case, it's really just their partners,

262 00:12:06.480 --> 00:12:09.330 set of participants that share an edge

263 00:12:09.330 --> 00:12:11.610 or partnership with person I.

264 00:12:11.610 --> 00:12:12.960 We have the degree.

265 00:12:12.960 --> 00:12:13.980 And then we have a vector

266 00:12:13.980 --> 00:12:16.380 of the baseline covariates for the neighbors,

267 00:12:16.380 --> 00:12:19.440 vector of baseline covariates for...

268 00:12:19.440 --> 00:12:21.300 Sorry, the treatment for the neighbors,

269 00:12:21.300 --> 00:12:23.280 baseline covariates for the neighbors.

270 00:12:23.280 --> 00:12:27.993 And then we denote the non-overlapping subnetworks by G.

271 00:12:31.470 --> 00:12:35.190 So we're doing causal inference with an intervention

272 00:12:35.190 --> 00:12:36.930 that's not randomized in a network.

273 00:12:36.930 --> 00:12:39.150 So this requires numerous assumptions

274 00:12:39.150 --> 00:12:42.423 in order to be able to identify these causal effects.

275 00:12:43.380 --> 00:12:45.510 So first, as in the figure,

276 00:12:45.510 --> 00:12:48.120 what I alluded to is we're assuming

277 00:12:48.120 --> 00:12:49.950 the nearest neighbor interference set.

278 00:12:49.950 --> 00:12:54.210 So basically, it's only the person's exposure themselves

279 00:12:54.210 --> 00:12:57.840 or the exposure of their neighbors that can impact

280 00:12:57.840 --> 00:13:01.170 the potential outcome or affect the potential outcome.

281 00:13:01.170 --> 00:13:04.560 We have an exchange ability assumption that applies

282 00:13:04.560 --> 00:13:07.050 not only to the exposure for the person

283 00:13:07.050 --> 00:13:09.870 but, also, the vector of exposures for their neighbors.

284 00:13:09.870 --> 00:13:14.700 So we have comparability between individuals

285 00:13:14.700 --> 00:13:16.730 who are exposed and not exposed.

286 00:13:16.730 --> 00:13:20.280 This is, of course, conditional on baseline covariates.

287 00:13:20.280 --> 00:13:22.590 We require a positivity assumption

288 00:13:22.590 --> 00:13:25.140 so that there's a positive probability of exposure.

289 00:13:25.140 --> 00:13:26.760 Each level of the covariates, again,

290 00:13:26.760 --> 00:13:29.100 both for the individual and their neighbors.

291 00:13:29.100 --> 00:13:32.940 And we also assume if there are different versions

292 00:13:32.940 --> 00:13:35.310 of the community alerts, for example,

293 00:13:35.310 --> 00:13:37.380 they don't matter for the potential outcome.

294 00:13:37.380 --> 00:13:40.080 So it's really whether you just got the alert,

295 00:13:40.080 --> 00:13:42.540 whether you got it as a paper flyer handed to you,

296 00:13:42.540 --> 00:13:45.030 or you saw it as a poster,

297 00:13:45.030 --> 00:13:47.380 we're just assuming it's the same intervention.

298 00:13:48.750 --> 00:13:49.950 So with these assumptions,

299 00:13:49.950 --> 00:13:51.510 we can write the potential outcome index

300 00:13:51.510 --> 00:13:54.990 by the exposure for the individual and their neighbors.

301 00:13:54.990 --> 00:13:57.030 And then by consistency,

302 00:13:57.030 --> 00:13:59.460 the observed outcome is one of the potential outcomes

303 00:13:59.460 --> 00:14:02.220 corresponding to the intervention received.

304 00:14:02.220 --> 00:14:05.130 And there's a little bit of notation

305 00:14:05.130 --> 00:14:09.090 that goes into the background of defining these effects.

306 00:14:09.090 --> 00:14:10.410 But long story short,

307 00:14:10.410 --> 00:14:12.630 we define the average potential outcomes

308 00:14:12.630 --> 00:14:15.330 using a Bernoulli allocation strategy,

309 00:14:15.330 --> 00:14:18.780 which is why those, when we define the spillover effect,

310 00:14:18.780 --> 00:14:20.190 it's a wide bar.

311 00:14:20.190 --> 00:14:22.680 And then what this effect is,

312 00:14:22.680 --> 00:14:25.290 is it's comparing the average potential outcome

313 00:14:25.290 --> 00:14:27.270 of unexposed individuals

314 00:14:27.270 --> 00:14:30.270 under two different allocation strategies.

315 00:14:30.270 --> 00:14:32.430 So that's the spillover effect

316 00:14:32.430 --> 00:14:35.310 that is in the first paper that we worked on.

317 00:14:35.310 --> 00:14:36.780 And then now when we're doing the power

318 00:14:36.780 --> 00:14:38.040 and sample size stuff,

319 00:14:38.040 --> 00:14:41.523 this is, basically, the parameter of interest.

320 00:14:48.060 --> 00:14:50.670 In the first paper, there's two different estimators.

321 00:14:50.670 --> 00:14:52.881 To get started with this study design stuff,

322 00:14:52.881 --> 00:14:56.070 we're looking at the second IPW estimator,

323 00:14:56.070 --> 00:14:58.860 which uses a generalized propensity score

324 00:14:58.860 --> 00:15:01.740 extending work in Laura's paper from 2021

325 00:15:01.740 --> 00:15:05.040 from a stratified estimator

326 00:15:05.040 --> 00:15:07.350 to an inverse probability weighted estimator.

327 00:15:07.350 --> 00:15:08.460 And we actually made the decision

328 00:15:08.460 --> 00:15:09.690 to start with this one first,

329 00:15:09.690 --> 00:15:12.650 because in the simulations of the first paper,

330 00:15:12.650 --> 00:15:15.630 it actually had slightly better finite sample performance.

331 00:15:15.630 --> 00:15:17.640 And then in actual application,

332 00:15:17.640 --> 00:15:19.860 we were able to add more covariates

333 00:15:19.860 --> 00:15:22.260 to this model to control for confounding.

334 00:15:22.260 --> 00:15:23.490 So we decided to start here.

335 00:15:23.490 --> 00:15:27.210 We'll also look at IPW-1 as a different estimator

336 00:15:27.210 --> 00:15:28.440 for the study design stuff.

337 00:15:28.440 --> 00:15:32.478 But we decided to start with IPW-2.

338 00:15:32.478 --> 00:15:37.470 And IPW-2, what this does is it uses

339 00:15:37.470 --> 00:15:39.420 a stratified interference assumption.

340 00:15:39.420 --> 00:15:41.970 So it looks at,

341 00:15:41.970 --> 00:15:43.770 instead of looking at the vector

342 00:15:43.770 --> 00:15:45.540 of exposures of the neighbors,

343 00:15:45.540 --> 00:15:47.340 it looks at SI which is the number

344 00:15:47.340 --> 00:15:49.620 of your neighbors that were exposed.

345 00:15:49.620 --> 00:15:53.580 Then, there's also a reducible propensity score assumption,

346 00:15:53.580 --> 00:15:57.390 which allows us to factor that generalized propensity score

347 00:15:57.390 --> 00:16:00.700 into a propensity score for the individual

348 00:16:02.070 --> 00:16:03.150 and then a propensity score

349 00:16:03.150 --> 00:16:06.963 for the neighbor's conditional on the individual.

350 00:16:08.100 --> 00:16:09.120 I may have just mixed that up,

351 00:16:09.120 --> 00:16:10.770 but it's on the next slide.

352 00:16:10.770 --> 00:16:13.170 Yeah, this is the neighbor's conditional on the individual

353 00:16:13.170 --> 00:16:17.265 and then the individual conditional on their covariates.

354 00:16:17.265 --> 00:16:18.665 Okay, got it right (laughs).

355 00:16:21.180 --> 00:16:24.900 So then this estimator looks like this.

356 00:16:24.900 --> 00:16:28.740 And then just to kind of break apart what's going on here,

357 00:16:28.740 --> 00:16:31.260 so it's an inverse probability weighted estimator

358 00:16:31.260 --> 00:16:33.750 where we have this generalized propensity score,

359 00:16:33.750 --> 00:16:35.400 where we have the individual exposure

360 00:16:35.400 --> 00:16:37.650 following a Bernoulli distribution

361 00:16:37.650 --> 00:16:39.090 with a certain probability.

362 00:16:39.090 --> 00:16:41.400 And then the SI variable,

363 00:16:41.400 --> 00:16:42.810 the number of the neighbors exposed,

364 00:16:42.810 --> 00:16:45.060 following a binomial distribution.

365 00:16:45.060 --> 00:16:47.790 And then with that reducible propensity score assumption,

366 00:16:47.790 --> 00:16:49.710 we can factor,

367 00:16:49.710 --> 00:16:51.930 one approach is to factor it this way.

368 00:16:51.930 --> 00:16:54.420 And then we can use these forms

369 00:16:54.420 --> 00:16:56.193 to estimate the propensity score.

370 00:16:58.230 --> 00:16:59.940 And then we still have this pi term here,

371 00:16:59.940 --> 00:17:01.440 because we're standardizing

372 00:17:01.440 --> 00:17:03.420 to a certain allocation strategy.

373 00:17:03.420 --> 00:17:06.060 So we're thinking about specific policies here

374 00:17:06.060 --> 00:17:08.280 when defining the counterfactuals.

375 00:17:08.280 --> 00:17:09.450 <v Donna>Ashley, I have a question.</v>

376 00:17:09.450 --> 00:17:13.940 The very first equation where you have Y at IPW-2,

377 00:17:15.780 --> 00:17:18.840 open paren zero comma alpha one.

378 00:17:18.840 --> 00:17:20.400 What does the zero mean?

379 00:17:20.400 --> 00:17:21.870 <v ->That means that the individual...</v>

380 00:17:21.870 --> 00:17:26.370 So A refers to the exposure for the individual.

381 00:17:26.370 --> 00:17:28.923 So it means the individual is not exposed,

382 00:17:30.420 --> 00:17:31.620 possibly contrary to facts.

383 00:17:31.620 --> 00:17:33.165 So they're all counterfactuals,

384 00:17:33.165 --> 00:17:35.280 but the individual themselves is not exposed.

385 00:17:35.280 --> 00:17:37.440 <v Donna>They're not directly exposed.</v>

386 00:17:37.440 --> 00:17:39.923 <v ->I don't like the words, "Directly exposed."</v>

387 00:17:39.923 --> 00:17:42.630 So in my mind, it's like we're either exposed or we're not.

388 00:17:42.630 --> 00:17:44.670 I don't know, it cleans it up in my mind a little bit,

389 00:17:44.670 --> 00:17:45.870 but I know what you're saying.

390 00:17:45.870 --> 00:17:48.030 So the individual themselves did not receive the...

391 00:17:48.030 --> 00:17:49.740 Let's make it in the context of the problem.

392 00:17:49.740 --> 00:17:52.830 Individual themselves did not receive the community alert

393 00:17:52.830 --> 00:17:54.855 from the TRIP investigative staff.

394 00:17:54.855 --> 00:17:55.830 <v Donna>Okay.</v>

395 00:17:55.830 --> 00:17:57.480 <v ->They may have gotten it secondhand,</v>

396 00:17:57.480 --> 00:17:59.940 which is the whole thing we're trying to estimate.

397 00:17:59.940 --> 00:18:01.920 So they didn't get it from the investigators,

398 00:18:01.920 --> 00:18:03.903 but then their neighbors,

399 00:18:04.980 --> 00:18:09.480 so these orange folks, alpha output percent of them,

400 00:18:09.480 --> 00:18:11.820 a certain percentage of them received the alert.

401 00:18:11.820 --> 00:18:16.200 So maybe we're interested in if 75% of your neighbors

402 00:18:16.200 --> 00:18:18.803 were alerted versus just 20%.

403 00:18:19.740 --> 00:18:22.980 And then there's sort of some practical considerations

404 00:18:22.980 --> 00:18:24.810 that I try to follow in our work.

405 00:18:24.810 --> 00:18:26.850 So we actually look at the distribution

406 00:18:26.850 --> 00:18:28.937 of coverage of treatment for the neighbors,

407 00:18:28.937 --> 00:18:31.050 and we only wanna be estimating effects

408 00:18:31.050 --> 00:18:32.820 sort of within the range of what we're seeing.

409 00:18:32.820 --> 00:18:37.820 So say 20% to maybe 60% were alerted

410 00:18:37.860 --> 00:18:39.360 and we have a lot of data there,

411 00:18:39.360 --> 00:18:40.740 then we could do contrast

412 00:18:40.740 --> 00:18:42.663 for those alpha levels in the data.

413 00:18:44.190 --> 00:18:45.510 Maybe some people feel more comfortable

414 00:18:45.510 --> 00:18:46.650 going out of the range of data,

415 00:18:46.650 --> 00:18:49.620 but I like to know we have information there.

416 00:18:49.620 --> 00:18:50.880 'Cause I think a lot of the times,

417 00:18:50.880 --> 00:18:51.870 it'll give you an estimate,

418 00:18:51.870 --> 00:18:55.050 but it feels better knowing we have this many neighbors,

419 00:18:55.050 --> 00:18:59.313 neighborhoods that had this type of exposure.

420 00:19:00.390 --> 00:19:02.010 Does that make sense? <v ->Yeah.</v>

421 00:19:02.010 --> 00:19:04.397 It does, so I don't agree with last thing.

422 00:19:04.397 --> 00:19:05.359 (laughing)

423 00:19:05.359 --> 00:19:06.192 <v ->Okay.</v>

424 00:19:07.820 --> 00:19:10.370 We all have different preferences I guess (laughs).

425 00:19:11.670 --> 00:19:12.750 <v Donna>I mean, yeah, you take that</v>

426 00:19:12.750 --> 00:19:14.220 to its logical extreme,

427 00:19:14.220 --> 00:19:17.473 I would say that it (indistinct) having a simple regression.

428 00:19:17.473 --> 00:19:20.910 You would have to observe X at every single value.

429 00:19:20.910 --> 00:19:23.580 <v ->Not every single value, but just the range.</v>

430 00:19:23.580 --> 00:19:25.140 So say that it stops at six-

431 00:19:25.140 --> 00:19:26.142 <v ->You don't wanna-</v> <v ->Say it stops at-</v>

432 00:19:26.142 --> 00:19:26.975 (drowned out).

433 00:19:26.975 --> 00:19:29.760 <v ->Yeah, yeah, say it stops at 60%,</v>

434 00:19:29.760 --> 00:19:32.490 and then we're trying to estimate 95% coverage.

435 00:19:32.490 --> 00:19:33.950 It almost feels too far out.

436 00:19:33.950 --> 00:19:35.250 <v Donna>So you don't wanna extrapolate,</v>

437 00:19:35.250 --> 00:19:36.690 but you're willing to interpolate.

438 00:19:36.690 --> 00:19:38.370 <v ->Yeah, yep.</v> <v ->Okay, I thought you</v>

439 00:19:38.370 --> 00:19:39.750 were saying you weren't willing to interpolate.

440 00:19:39.750 --> 00:19:41.310 <v ->No, then the coverage levels,</v>

441 00:19:41.310 --> 00:19:42.170 if you look at the distribution,

442 00:19:42.170 --> 00:19:44.490 it kind of bumps around and there's some that are missing.

443 00:19:44.490 --> 00:19:47.686 But I'm okay going over that range of the data, but-

444 00:19:47.686 --> 00:19:48.519 <v Donna>Then I do.</v>

445 00:19:48.519 --> 00:19:49.993 <v ->Okay, that's good.</v>

446 00:19:49.993 --> 00:19:50.910 <v Colleague>I mean, you can still do it,</v>

447 00:19:50.910 --> 00:19:52.620 people do it like to extrapolate,

448 00:19:52.620 --> 00:19:54.928 but you know that the (indistinct) we'll get

449 00:19:54.928 --> 00:19:56.027 is gonna be higher, right?

450 00:19:56.027 --> 00:19:57.930 'Cause you don't have data there.

451 00:19:57.930 --> 00:19:58.763 <v ->Yep.</v>

452 00:20:01.350 --> 00:20:02.400 That's a little digression

453 00:20:02.400 --> 00:20:03.720 from where I wanted to go with the slides,

454 00:20:03.720 --> 00:20:05.610 but it's still interesting (laughs).

455 00:20:05.610 --> 00:20:07.727 <v Donna>Ashley, can ask you a question about the,</v>

456 00:20:07.727 --> 00:20:10.049 so (indistinct) design IPW-1,

457 00:20:10.049 --> 00:20:12.943 but you said that you weren't able

458 00:20:12.943 --> 00:20:15.990 to include more covariates (indistinct).

459 00:20:15.990 --> 00:20:17.070 <v ->In the TRIP data.</v>

460 00:20:17.070 --> 00:20:19.080 <v Donna>And what (indistinct)?</v>

461 00:20:19.080 --> 00:20:20.220 <v ->So I think it has to do with,</v>

462 00:20:20.220 --> 00:20:22.200 so just to say it's not really even on this slide,

463 00:20:22.200 --> 00:20:25.830 but IPW-1 uses a generalized logit model

464 00:20:25.830 --> 00:20:27.660 to estimate the propensity score.

465 00:20:27.660 --> 00:20:29.910 And basically, that thing's kind of a bugger.

466 00:20:30.767 --> 00:20:32.417 It's pretty sensitive it.

467 00:20:32.417 --> 00:20:33.420 It doesn't...

468 00:20:33.420 --> 00:20:35.550 Linear mixed models tend to do pretty well,

469 00:20:35.550 --> 00:20:38.250 but these ones with the logit link

470 00:20:38.250 --> 00:20:41.250 I find in practice they can be,

471 00:20:41.250 --> 00:20:43.443 they run into these convergence issues.

472 00:20:44.790 --> 00:20:48.000 And then this one that extended Laura's estimator,

473 00:20:48.000 --> 00:20:49.380 in practice at least,

474 00:20:49.380 --> 00:20:51.360 we haven't run it in hundreds of data sets or anything,

475 00:20:51.360 --> 00:20:52.920 but the few that we have,

476 00:20:52.920 --> 00:20:54.507 we tend to be able to add more covariates.

477 00:20:54.507 --> 00:20:56.940 And because the nonrandomized intervention,

478 00:20:56.940 --> 00:20:59.370 that just seems like the right thing to do,

479 00:20:59.370 --> 00:21:01.720 because we want better control for confounding.

480 00:21:02.880 --> 00:21:03.713 <v Donna>Thanks.</v>

481 00:21:03.713 --> 00:21:04.656 <v ->Yours is winning.</v>

482 00:21:04.656 --> 00:21:05.944 (laughing)

483 00:21:05.944 --> 00:21:08.430 (indistinct)

484 00:21:08.430 --> 00:21:10.653 At least with our team recently.

485 00:21:11.880 --> 00:21:13.620 And that's not to say IPW-1...

486 00:21:13.620 --> 00:21:15.900 It's a great estimator, as well.

487 00:21:15.900 --> 00:21:17.490 It has some nice properties,

488 00:21:17.490 --> 00:21:19.320 but there's just sort of this practical issue

489 00:21:19.320 --> 00:21:22.710 of the generalized logit model.

490 00:21:22.710 --> 00:21:24.150 <v Donna>Yeah, the benefit of that one, though,</v>

491 00:21:24.150 --> 00:21:25.674 is that you don't have to assume

492 00:21:25.674 --> 00:21:26.970 the stratified interference.

493 00:21:26.970 --> 00:21:29.070 <v ->Right, you don't have to assume stratified interference,</v>

494 00:21:29.070 --> 00:21:30.030 and then we don't have to make

495 00:21:30.030 --> 00:21:32.490 this reducible propensity score assumption.

496 00:21:32.490 --> 00:21:36.828 So pros and cons, right?

497 00:21:36.828 --> 00:21:37.661 Yeah, and then it's interesting

498 00:21:37.661 --> 00:21:39.690 to think about what are our practical recommendations

499 00:21:39.690 --> 00:21:43.482 when folks have a menu of estimators to choose from.

500 00:21:43.482 --> 00:21:48.030 What do we tell folks to do in their substantive papers?

501 00:21:48.030 --> 00:21:50.090 Do we ask them to check both?

502 00:21:50.090 --> 00:21:52.080 I think that's what I've been advising for now,

503 00:21:52.080 --> 00:21:53.880 just as it's one is your main analysis,

504 00:21:53.880 --> 00:21:55.530 one is for sensitivity analysis,

505 00:21:55.530 --> 00:21:59.013 but I think that's another open question.

506 00:22:01.020 --> 00:22:03.180 So I spared us all the notation on this slide,

507 00:22:03.180 --> 00:22:06.600 but just to say the variance estimation

508 00:22:06.600 --> 00:22:09.090 is used in the study design issue.

509 00:22:09.090 --> 00:22:11.280 So we use M estimation here.

510 00:22:11.280 --> 00:22:13.290 And then to do M estimation,

511 00:22:13.290 --> 00:22:17.210 we're using the union of the connected subnetworks

512 00:22:17.210 --> 00:22:19.293 to break up the graph.

513 00:22:21.900 --> 00:22:22.733 But at the same time,

514 00:22:22.733 --> 00:22:25.920 we also preserve the underlying connection.

515 00:22:25.920 --> 00:22:29.160 So we maintained that nearest neighbor structure

516 00:22:29.160 --> 00:22:31.020 when calculating the variance.

517 00:22:31.020 --> 00:22:32.970 And then in the simulation study,

518 00:22:32.970 --> 00:22:36.090 we found that accounting for that

519 00:22:36.090 --> 00:22:39.330 as compared to just doing complete partial interference

520 00:22:39.330 --> 00:22:40.920 was more efficient.

521 00:22:40.920 --> 00:22:43.140 So the complete partial interference

522 00:22:43.140 --> 00:22:44.700 would be you would assume

523 00:22:44.700 --> 00:22:47.760 the entire component is the interference set

524 00:22:47.760 --> 00:22:49.880 versus, here, we maintain that the neighbors

525 00:22:49.880 --> 00:22:51.060 of the interference set.

526 00:22:51.060 --> 00:22:52.800 But then we still leverage

527 00:22:52.800 --> 00:22:54.780 the components as independent units,

528 00:22:54.780 --> 00:22:57.363 because it's required for M estimation.

529 00:23:01.050 --> 00:23:02.490 Okay.

530 00:23:02.490 --> 00:23:05.804 So that was all the background to build up to (laughs)

531 00:23:05.804 --> 00:23:09.690 (indistinct) to do study design in these networks

532 00:23:09.690 --> 00:23:11.250 with these particular methods

533 00:23:11.250 --> 00:23:14.793 that have been developed over the recent years.

534 00:23:16.230 --> 00:23:17.243 So basically, I don't know.

535 00:23:17.243 --> 00:23:19.470 I don't think I need to sell it to this group,

536 00:23:19.470 --> 00:23:22.820 but to understand how features

537 00:23:22.820 --> 00:23:25.860 of the study design impact the power is important.

538 00:23:25.860 --> 00:23:27.210 As far as we can tell,

539 00:23:27.210 --> 00:23:31.620 this hasn't been a real emphasis in network-based studies,

540 00:23:31.620 --> 00:23:34.440 particularly in the area of substance use in HIV.

541 00:23:34.440 --> 00:23:36.600 Folks kind of get the sample that they can get.

542 00:23:36.600 --> 00:23:37.770 It's a ton of work,

543 00:23:37.770 --> 00:23:39.930 so they're not thinking about designing them

544 00:23:39.930 --> 00:23:42.510 like a cluster randomized trial.

545 00:23:42.510 --> 00:23:45.480 Or even in observational studies,

546 00:23:45.480 --> 00:23:47.700 there's some proposals where they'll wanna see

547 00:23:47.700 --> 00:23:49.680 at least power calculations to show

548 00:23:49.680 --> 00:23:52.323 that there's a large enough sample size.

549 00:23:53.160 --> 00:23:55.320 So there are approaches coming out

550 00:23:55.320 --> 00:23:57.120 in the statistics literature.

551 00:23:57.120 --> 00:24:00.360 Of course, there are some older ones about overall effects

552 00:24:00.360 --> 00:24:02.040 in cluster randomized trials.

553 00:24:02.040 --> 00:24:03.090 I just put one reference there,

554 00:24:03.090 --> 00:24:05.250 but that's a very large literature.

555 00:24:05.250 --> 00:24:08.070 But then getting into the causal spillover effects,

556 00:24:08.070 --> 00:24:10.890 there are some papers by Baird et al.

557 00:24:10.890 --> 00:24:13.080 looking at a two-stage randomized design.

558 00:24:13.080 --> 00:24:15.840 And I found another paper by Sinclair in 2012

559 00:24:15.840 --> 00:24:18.180 that was a multi-level randomized design,

560 00:24:18.180 --> 00:24:20.160 which kind of had the similar flavor

561 00:24:20.160 --> 00:24:21.600 to a cluster randomized design,

562 00:24:21.600 --> 00:24:22.920 but it was from the econ literature,

563 00:24:22.920 --> 00:24:26.010 so they had a slightly different name for it.

564 00:24:26.010 --> 00:24:29.670 However, when we're doing a sociometric network study,

565 00:24:29.670 --> 00:24:33.210 these larger network-based studies,

566 00:24:33.210 --> 00:24:35.010 it would be difficult to implement

567 00:24:35.010 --> 00:24:36.960 a two-stage randomized design

568 00:24:36.960 --> 00:24:39.600 just because of how folks are recruited.

569 00:24:39.600 --> 00:24:42.180 And then we're also interested in being able to evaluate

570 00:24:42.180 --> 00:24:44.080 interventions that are not randomized.

571 00:24:44.970 --> 00:24:48.330 So we wanna have adequately powered studies

572 00:24:48.330 --> 00:24:51.033 to evaluate these interventions.

573 00:24:54.720 --> 00:24:57.450 So this overall paper,

574 00:24:57.450 --> 00:25:00.390 we're gonna start off with simulation studies,

575 00:25:00.390 --> 00:25:02.790 thinking about the varying the number of components

576 00:25:02.790 --> 00:25:04.590 and the number of nodes,

577 00:25:04.590 --> 00:25:07.470 and then changing different parameters

578 00:25:07.470 --> 00:25:09.720 in the network including effect size,

579 00:25:09.720 --> 00:25:13.230 features of the network like degree,

580 00:25:13.230 --> 00:25:14.790 intracluster correlation,

581 00:25:14.790 --> 00:25:16.980 and see how these impact the power.

582 00:25:16.980 --> 00:25:20.490 And then lastly, trying to work on driving

583 00:25:20.490 --> 00:25:23.790 an expression for the minimal detectable effect

584 00:25:23.790 --> 00:25:26.553 as well as expressions for sample size.

585 00:25:29.790 --> 00:25:31.680 So the ongoing work I'll be presenting today

586 00:25:31.680 --> 00:25:34.890 are focusing on mostly on the first aim,

587 00:25:34.890 --> 00:25:37.410 so simulation study to detect spillover effects,

588 00:25:37.410 --> 00:25:38.967 varying the number of components

589 00:25:38.967 --> 00:25:41.310 for the number of nodes in the network.

590 00:25:41.310 --> 00:25:43.980 And then as the next step for this,

591 00:25:43.980 --> 00:25:46.980 we have some initial results for a wall test statistic

592 00:25:46.980 --> 00:25:48.540 and showing that that test statistic

593 00:25:48.540 --> 00:25:49.953 is normally distributed.

594 00:25:51.270 --> 00:25:54.390 So just an overview of how we've generated some of the data.

595 00:25:54.390 --> 00:25:55.470 We started off by generating

596 00:25:55.470 --> 00:25:57.150 a network with certain features.

597 00:25:57.150 --> 00:25:59.400 Then on that network, we simulate random variables

598 00:25:59.400 --> 00:26:03.030 and then generate the potential outcomes

599 00:26:03.030 --> 00:26:05.490 and then, subsequently, the observed outcomes.

600 00:26:05.490 --> 00:26:08.280 In each data set, we estimate the spillover effects using,

601 00:26:08.280 --> 00:26:11.850 in this case we used IPW-2 and confidence intervals.

602 00:26:11.850 --> 00:26:13.200 And then we calculate the power

603 00:26:13.200 --> 00:26:15.603 in the empirical coverage probability.

604 00:26:18.241 --> 00:26:20.220 (coughs)

605 00:26:20.220 --> 00:26:21.063 Sip of water.

606 00:26:25.170 --> 00:26:26.670 So in the first setting,

607 00:26:26.670 --> 00:26:30.150 we're looking to see if power varies by components,

608 00:26:30.150 --> 00:26:31.740 which I thought was a good place to start,

609 00:26:31.740 --> 00:26:33.810 because our M estimation,

610 00:26:33.810 --> 00:26:35.640 the effective sample size is M,

611 00:26:35.640 --> 00:26:37.173 or the number of components.

612 00:26:38.190 --> 00:26:40.230 So we had two different approaches.

613 00:26:40.230 --> 00:26:41.760 We keep the component size the same

614 00:26:41.760 --> 00:26:43.530 and increase the number of components,

615 00:26:43.530 --> 00:26:45.570 or we fix the number of nodes

616 00:26:45.570 --> 00:26:48.030 and then increase the number of components.

617 00:26:48.030 --> 00:26:50.910 So the first one is really how the statistics

618 00:26:50.910 --> 00:26:52.680 of the M estimation are working.

619 00:26:52.680 --> 00:26:55.290 And the second one I think is empirically interesting.

620 00:26:55.290 --> 00:26:58.480 I don't think it's as founded in the theory

621 00:26:58.480 --> 00:27:01.860 of the estimation, just to be clear,

622 00:27:01.860 --> 00:27:04.396 but nonetheless, I think interesting to look at.

623 00:27:04.396 --> 00:27:05.940 <v Donna>Could you go back a second?</v>

624 00:27:05.940 --> 00:27:06.773 <v ->Yeah.</v> <v ->So what did</v>

625 00:27:06.773 --> 00:27:09.360 the motivating study have in terms

626 00:27:09.360 --> 00:27:12.643 of the number of components and the number of nodes?

627 00:27:12.643 --> 00:27:17.610 <v ->The motivating study has 10 components, 216 nodes.</v>

628 00:27:17.610 --> 00:27:19.500 And then what we did in our first paper

629 00:27:19.500 --> 00:27:21.390 was to try to increase the number of components.

630 00:27:21.390 --> 00:27:25.680 We tried to break up that largest connected component using

631 00:27:25.680 --> 00:27:30.150 network science community detection methods, which is okay.

632 00:27:30.150 --> 00:27:32.910 I don't think it's the most satisfying answer.

633 00:27:32.910 --> 00:27:34.440 And then once we do the community detection,

634 00:27:34.440 --> 00:27:36.420 then we had 20 components.

635 00:27:36.420 --> 00:27:38.610 So the actual motivating data set

636 00:27:38.610 --> 00:27:43.610 is really 10 to 20 components, about 216 individuals.

637 00:27:43.920 --> 00:27:47.460 <v Donna>Okay, so nodes and individuals are the same thing?</v>

638 00:27:47.460 --> 00:27:49.641 <v ->Yep, sorry, I may have probably using those-</v>

639 00:27:49.641 --> 00:27:51.060 <v ->No, that's okay.</v> <v ->Individual, yeah.</v>

640 00:27:51.060 --> 00:27:52.533 216 nodes, yep.

641 00:27:53.520 --> 00:27:55.020 <v Donna>Ashley, can ask you another question?</v>

642 00:27:55.020 --> 00:27:55.853 <v ->Yeah.</v>

643 00:27:55.853 --> 00:27:56.686 <v Donna>So is that in general?</v>

644 00:27:56.686 --> 00:27:58.187 And you see that treatment, right?

645 00:27:59.613 --> 00:28:02.016 Like in the previous slide said (indistinct) treatment

646 00:28:02.016 --> 00:28:05.280 and potential outcomes I guess, right?

647 00:28:05.280 --> 00:28:07.650 (indistinct) treatment.

648 00:28:07.650 --> 00:28:10.440 So do you do that (indistinct) thing of observational study

649 00:28:10.440 --> 00:28:14.124 like simulating the treatment from propensity score?

650 00:28:14.124 --> 00:28:18.570 <v ->Yeah, so we fit the propensity score in the TRIP data,</v>

651 00:28:18.570 --> 00:28:20.460 and then you'll see in a couple slides

652 00:28:20.460 --> 00:28:23.280 I have the actual values of the parameters that we used.

653 00:28:23.280 --> 00:28:25.640 And then we, obviously, can't fit a model to,

654 00:28:25.640 --> 00:28:27.660 we just fit a model to the observed outcome

655 00:28:27.660 --> 00:28:30.090 to try to get the betas for the model,

656 00:28:30.090 --> 00:28:32.640 the potential outcome out of the TRIP.

657 00:28:32.640 --> 00:28:34.830 Again, the motivating data.

658 00:28:34.830 --> 00:28:36.480 Yep, good question.

659 00:28:36.480 --> 00:28:38.250 And this is like a roadmap.

660 00:28:38.250 --> 00:28:39.270 I'm gonna actually go through

661 00:28:39.270 --> 00:28:41.520 a lot of detail for each one now (laughs).

662 00:28:41.520 --> 00:28:42.780 <v Vin>Sorry, I also have a question.</v>

663 00:28:42.780 --> 00:28:45.330 So in the simulation for component,

664 00:28:45.330 --> 00:28:47.460 and there's nobody in that component received

665 00:28:47.460 --> 00:28:49.560 the treatment in the simulations,

666 00:28:49.560 --> 00:28:50.880 is that possible?

667 00:28:50.880 --> 00:28:52.380 <v ->Yep, that could happen.</v>

668 00:28:52.380 --> 00:28:54.450 <v Vin>And then like for that component,</v>

669 00:28:54.450 --> 00:28:56.280 is that excluded from this,

670 00:28:56.280 --> 00:28:57.780 because perhaps it violate

671 00:28:57.780 --> 00:29:00.090 the positivity assumption I guess?

672 00:29:00.090 --> 00:29:01.140 <v ->Well, it depends on...</v>

673 00:29:01.140 --> 00:29:02.610 They would come into play

674 00:29:02.610 --> 00:29:05.333 if you're interested in a coverage of 0%.

675 00:29:06.690 --> 00:29:08.760 Right, so it depends on what your...

676 00:29:08.760 --> 00:29:11.490 So that would be if you're interested in estimating

677 00:29:11.490 --> 00:29:15.863 Y of zero with alpha equals 0%.

678 00:29:16.890 --> 00:29:19.170 It's like a pure control group.

679 00:29:19.170 --> 00:29:21.183 So it would be that case.

680 00:29:23.070 --> 00:29:23.903 Yep.

681 00:29:24.750 --> 00:29:27.660 Yeah, so we didn't exclude anyone on that case,

682 00:29:27.660 --> 00:29:30.170 but in another paper, we did exclude...

683 00:29:30.170 --> 00:29:32.000 We were actually looking at HIV seroconversion

684 00:29:32.000 --> 00:29:33.360 in the other paper,

685 00:29:33.360 --> 00:29:35.130 and we did an analysis by components.

686 00:29:35.130 --> 00:29:39.960 So if the component had no HIV-infected individuals

687 00:29:39.960 --> 00:29:41.850 at baseline and the components

688 00:29:41.850 --> 00:29:43.980 in the study were not allowed to change,

689 00:29:43.980 --> 00:29:46.223 then that was like a,

690 00:29:46.223 --> 00:29:47.160 I forget what the epi term for it,

691 00:29:47.160 --> 00:29:49.680 there's no way anyone can get infected.

692 00:29:49.680 --> 00:29:51.870 So it was a perfectly protected component.

693 00:29:51.870 --> 00:29:54.240 So we excluded those.

694 00:29:54.240 --> 00:29:57.000 So we wanted components in that study that were at risk.

695 00:29:57.000 --> 00:30:00.510 So we had to have at least one individual

696 00:30:00.510 --> 00:30:02.580 in the component with HIV at baseline.]

697 00:30:02.580 --> 00:30:06.109 so there was some chance that it could spread.

698 00:30:06.109 --> 00:30:07.723 <v Colleague>But it seems that even if you don't exclude</v>

699 00:30:07.723 --> 00:30:09.553 these components where no one is treated,

700 00:30:09.553 --> 00:30:12.492 the (indistinct) weights will be very low, right?

701 00:30:12.492 --> 00:30:16.260 <v ->Yep, they'll just get downgraded for the treatment thing.</v>

702 00:30:16.260 --> 00:30:17.430 But then I guess it might made

703 00:30:17.430 --> 00:30:19.293 my mind go to thinking about,

704 00:30:20.370 --> 00:30:22.460 particularly for HIV seroconversion,

705 00:30:22.460 --> 00:30:24.960 if you have a case where there is a really small,

706 00:30:24.960 --> 00:30:27.360 maybe it's one of these little components,

707 00:30:27.360 --> 00:30:29.580 and it's just these two people,

708 00:30:29.580 --> 00:30:33.453 like the two, like a little dyad, neither have HIV.

709 00:30:34.770 --> 00:30:36.450 I guess then, if you're assuming

710 00:30:36.450 --> 00:30:38.760 that there's no other edges into there,

711 00:30:38.760 --> 00:30:40.860 then there can be no events.

712 00:30:40.860 --> 00:30:42.570 So thinking about like, you know.

713 00:30:42.570 --> 00:30:44.310 I think it makes sense to exclude that,

714 00:30:44.310 --> 00:30:47.913 because they're not at risk as a group, as a dyad.

715 00:30:50.715 --> 00:30:52.815 And maybe that's another tangent (laughs).

716 00:30:55.320 --> 00:30:57.870 Okay, so approach one.

717 00:30:57.870 --> 00:31:01.290 We have this regular connected network with degree four,

718 00:31:01.290 --> 00:31:03.300 which is approximately the observed degree

719 00:31:03.300 --> 00:31:04.560 in the TRIP network.

720 00:31:04.560 --> 00:31:08.850 And then we sampled nodes from a place on 10 distribution.

721 00:31:08.850 --> 00:31:10.050 And then we repeat this

722 00:31:10.050 --> 00:31:14.223 and then combine the M subnetworks to form the full network.

723 00:31:15.330 --> 00:31:17.080 So this is the first case where we,

724 00:31:19.230 --> 00:31:20.790 yeah, we have the number,

725 00:31:20.790 --> 00:31:22.410 we keep the component size the same,

726 00:31:22.410 --> 00:31:24.963 and then we're increasing the number of components.

727 00:31:26.250 --> 00:31:28.320 Alternatively for approach two,

728 00:31:28.320 --> 00:31:30.810 we have the same four-degree network.

729 00:31:30.810 --> 00:31:35.463 We have M components but for a fixed set of number of nodes,

730 00:31:36.797 --> 00:31:39.030 and then we generate the connected network,

731 00:31:39.030 --> 00:31:41.953 and then, again, combine the subnetworks.

732 00:31:45.300 --> 00:31:47.400 So in either case, there's sort these two scenarios

733 00:31:47.400 --> 00:31:49.530 where we're generating the network,

734 00:31:49.530 --> 00:31:53.429 and then we generate the potential outcomes

735 00:31:53.429 --> 00:31:54.570 and the observed outcomes.

736 00:31:54.570 --> 00:31:57.060 We assign random effects to induce

737 00:31:57.060 --> 00:31:59.400 correlation within each component,

738 00:31:59.400 --> 00:32:01.020 and then simulate...

739 00:32:01.020 --> 00:32:03.409 We just have one binary covariate for now.

740 00:32:03.409 --> 00:32:04.290 Of course, we wanna extend this

741 00:32:04.290 --> 00:32:06.590 to multiple covariates, continuous covariates.

742 00:32:07.586 --> 00:32:09.030 And then we generate the potential outcome

743 00:32:09.030 --> 00:32:10.830 using this formula here

744 00:32:10.830 --> 00:32:13.020 where the values of the parameters

745 00:32:13.020 --> 00:32:15.690 are from an estimated model in the TRIP data.

746 00:32:15.690 --> 00:32:17.820 And then we generate the treatment

747 00:32:17.820 --> 00:32:21.240 or exposure using this per newly random variable.

748 00:32:21.240 --> 00:32:23.340 Again, with the parameter values

749 00:32:23.340 --> 00:32:26.493 from a model in the TRIP data.

750 00:32:27.360 --> 00:32:30.180 And then depending on what the value of A is,

751 00:32:30.180 --> 00:32:31.890 and A and I is,

752 00:32:31.890 --> 00:32:34.920 then we can pull off the observed outcome

753 00:32:34.920 --> 00:32:38.793 from the vector of potential outcomes for each individual.

754 00:32:40.260 --> 00:32:41.093 <v Donna>I have a question.</v>

755 00:32:41.093 --> 00:32:41.926 <v ->Yep.</v>

756 00:32:42.960 --> 00:32:44.970 <v Donna>So earlier, you said you were</v>

757 00:32:44.970 --> 00:32:49.470 only allowing spillover between first-degree,

758 00:32:49.470 --> 00:32:52.226 nodes that were connected by first-degree.

759 00:32:52.226 --> 00:32:54.270 <v ->Mn-hm.</v>

760 00:32:54.270 --> 00:32:58.260 <v Donna>But then if you're same kind of variable</v>

761 00:32:58.260 --> 00:32:59.730 to describe spillovers,

762 00:32:59.730 --> 00:33:03.180 the proportion of nodes,

763 00:33:03.180 --> 00:33:06.270 or the proportion of, I don't know what you call them,

764 00:33:06.270 --> 00:33:10.713 participants in a component that are exposed,

765 00:33:12.073 --> 00:33:14.490 then it's ignoring that.

766 00:33:14.490 --> 00:33:16.440 <v ->So yeah, maybe I was mixing papers.</v>

767 00:33:16.440 --> 00:33:19.620 In this paper, it's really the proportion

768 00:33:19.620 --> 00:33:21.360 of the neighbors that are treated.

769 00:33:21.360 --> 00:33:22.350 So you have each person.

770 00:33:22.350 --> 00:33:24.780 It's the proportion of their neighbors that are treated

771 00:33:24.780 --> 00:33:26.640 that's going to define their potential outcome.

772 00:33:26.640 --> 00:33:28.770 <v Donna>That has to be a first-degree neighbors-</v>

773 00:33:28.770 --> 00:33:30.750 <v ->In this-</v> <v ->Anybody (indistinct).</v>

774 00:33:30.750 --> 00:33:33.600 <v ->In this paper you, we could extend this to second,</v>

775 00:33:33.600 --> 00:33:36.510 third-degree, different interference structures.

776 00:33:36.510 --> 00:33:39.390 But in this particular paper, that's how it's defined.

777 00:33:39.390 --> 00:33:40.230 But I think what I was doing,

778 00:33:40.230 --> 00:33:42.210 I was actually giving an example from another paper

779 00:33:42.210 --> 00:33:44.820 where we assume partial interference by component.

780 00:33:44.820 --> 00:33:47.970 In this paper, it's the nearest neighbor interference.

781 00:33:47.970 --> 00:33:50.940 So the potential outcomes depend on the number

782 00:33:50.940 --> 00:33:53.280 of the neighbors that are treated

783 00:33:53.280 --> 00:33:55.740 out of the total, the proportion.

784 00:33:55.740 --> 00:33:57.128 <v Donna>One other question.</v>

785 00:33:57.128 --> 00:34:01.737 So at this point, five squared between subjects' variance,

786 00:34:03.420 --> 00:34:06.210 what kind have ICC does that give, do you know?

787 00:34:06.210 --> 00:34:07.137 <v ->I don't remember off the top of my head,</v>

788 00:34:07.137 --> 00:34:08.313 but we can check.

789 00:34:10.080 --> 00:34:11.340 And I'm trying to remember.

790 00:34:11.340 --> 00:34:13.530 I think we got that from looking at the TRIP data,

791 00:34:13.530 --> 00:34:17.133 but I'd have to go back and check how we landed on that.

792 00:34:19.104 --> 00:34:20.363 But yeah, it's a good idea to check.

793 00:34:27.401 --> 00:34:30.357 And then we estimate the spillover effect

794 00:34:30.357 --> 00:34:33.240 and the corresponding 95% confidence interval

795 00:34:33.240 --> 00:34:35.730 in each data set using the methods

796 00:34:35.730 --> 00:34:37.410 that were presented earlier.

797 00:34:37.410 --> 00:34:38.880 And then we calculate the power

798 00:34:38.880 --> 00:34:40.980 in the empirical coverage probability.

799 00:34:40.980 --> 00:34:42.960 We simulated across 500 data sets,

800 00:34:42.960 --> 00:34:45.960 and we're still working on driving

801 00:34:45.960 --> 00:34:47.490 and evaluating the test statistic.

802 00:34:47.490 --> 00:34:49.530 So for now, we just use the confidence interval

803 00:34:49.530 --> 00:34:51.960 to see if the null value is in the confidence interval

804 00:34:51.960 --> 00:34:55.260 or not as a way to assess the power.

805 00:34:55.260 --> 00:34:57.060 And then just as a sanity check,

806 00:34:57.060 --> 00:34:58.230 we checked it in the first paper,

807 00:34:58.230 --> 00:35:00.870 but we also look at the empirical coverage probability

808 00:35:00.870 --> 00:35:05.400 just to make sure the estimators are behaving as we expect.

809 00:35:05.400 --> 00:35:07.680 <v Donna>So is there a test statistic?</v>

810 00:35:07.680 --> 00:35:09.150 <v ->It's derived and we're looking</v>

811 00:35:09.150 --> 00:35:11.730 at the normality of it first, assessing it.

812 00:35:11.730 --> 00:35:13.830 And then the next step, which we ran outta time

813 00:35:13.830 --> 00:35:16.380 to do for today is we wanna redo these simulations.

814 00:35:16.380 --> 00:35:18.840 So that's step four.

815 00:35:18.840 --> 00:35:22.320 Sub two is based on the test statistic,

816 00:35:22.320 --> 00:35:24.090 not the confidence interval.

817 00:35:24.090 --> 00:35:25.440 I mean, they should largely agree,

818 00:35:25.440 --> 00:35:28.050 but what makes me nervous is it's a confidence interval

819 00:35:28.050 --> 00:35:31.980 for a estimation of two parameters.

820 00:35:31.980 --> 00:35:33.930 And sometimes in that case, the confidence interval

821 00:35:33.930 --> 00:35:35.820 may not always agree with the test statistics.

822 00:35:35.820 --> 00:35:40.110 So it should typically, but to be...

823 00:35:40.110 --> 00:35:41.670 I think it's correct.

824 00:35:41.670 --> 00:35:46.050 It's more appropriate to be using the test statistic.

825 00:35:46.050 --> 00:35:48.570 <v Vin>The confidence interval or the indirect effect?</v>

826 00:35:48.570 --> 00:35:49.830 <v ->Yeah.</v>

827 00:35:49.830 --> 00:35:52.230 <v Vin>So you will...</v>

828 00:35:52.230 --> 00:35:53.250 I mean, I think there are...

829 00:35:53.250 --> 00:35:54.840 They should agree, right?

830 00:35:54.840 --> 00:35:56.430 <v ->But I worry about-</v> <v ->(drowned out) the null</v>

831 00:35:56.430 --> 00:35:59.250 distribution for the test statistic.

832 00:35:59.250 --> 00:36:00.090 <v ->Yeah.</v>

833 00:36:00.090 --> 00:36:01.440 <v Donna>That's the main thing.</v>

834 00:36:01.440 --> 00:36:03.692 If it's a wall test statistic,

835 00:36:03.692 --> 00:36:06.590 then we use the null distribution,

836 00:36:06.590 --> 00:36:08.578 which you can't do (indistinct) have

837 00:36:08.578 --> 00:36:09.438 different statistical (drowned out).

838 00:36:09.438 --> 00:36:11.605 <v Vin>Yeah, I see, yeah.</v>

839 00:36:13.375 --> 00:36:14.220 <v ->So I think this is a good way</v>

840 00:36:14.220 --> 00:36:16.860 that we got started as we're working on...

841 00:36:16.860 --> 00:36:18.990 We first wanna evaluate we got the test statistic correct

842 00:36:18.990 --> 00:36:21.003 before we blow through all this.

843 00:36:21.870 --> 00:36:24.090 <v Donna>The other thing is that the robust standard errors</v>

844 00:36:24.090 --> 00:36:27.090 are problematic in smaller samples, too.

845 00:36:27.090 --> 00:36:29.220 And there are all these different fixes to it.

846 00:36:29.220 --> 00:36:30.810 So I don't know if the test statistic

847 00:36:30.810 --> 00:36:32.463 would also have that problem.

848 00:36:33.330 --> 00:36:34.800 <v ->Yeah, potentially.</v> <v ->We've mostly seen it</v>

849 00:36:34.800 --> 00:36:36.270 about confidence intervals.

850 00:36:36.270 --> 00:36:38.670 Have you seen it about test statistics?

851 00:36:38.670 --> 00:36:39.503 <v ->Yeah.</v> <v ->With the robust</v>

852 00:36:39.503 --> 00:36:41.117 standardized- <v ->The same thing (indistinct).</v>

853 00:36:42.390 --> 00:36:43.320 They would agree,

854 00:36:43.320 --> 00:36:45.450 because we're always talking about,

855 00:36:45.450 --> 00:36:48.330 assuming normality, the variance doesn't change

856 00:36:48.330 --> 00:36:51.093 across the hypothesis (indistinct) space.

857 00:36:52.650 --> 00:36:54.450 But then, CI here,

858 00:36:54.450 --> 00:36:57.431 you're refer to the CI of the impact (indistinct).

859 00:36:57.431 --> 00:36:58.264 <v ->Correct, yeah.</v>

860 00:36:58.264 --> 00:37:01.380 <v Vin>And that's already accounting for the covariance.</v>

861 00:37:01.380 --> 00:37:03.423 The two potential outcome estimates.

862 00:37:05.430 --> 00:37:07.430 So if normality holds, they would agree.

863 00:37:08.935 --> 00:37:13.140 If you can derive the normality of the estimator,

864 00:37:13.140 --> 00:37:15.000 then the CI I think (indistinct).

865 00:37:15.000 --> 00:37:17.370 <v ->Yeah, so we have the normality of the estimator already,</v>

866 00:37:17.370 --> 00:37:18.330 and then in a couple slides,

867 00:37:18.330 --> 00:37:20.340 I'll show what we have for the test statistic.

868 00:37:20.340 --> 00:37:22.080 And I have some preliminary results showing

869 00:37:22.080 --> 00:37:23.460 that it looks approximately normal,

870 00:37:23.460 --> 00:37:26.410 but I don't think it's quite ready for prime time (laughs).

871 00:37:28.664 --> 00:37:29.520 <v Donna>So then that error is reliant</v>

872 00:37:29.520 --> 00:37:30.720 on M estimation, right?

873 00:37:30.720 --> 00:37:31.770 <v ->Correct.</v>

874 00:37:31.770 --> 00:37:32.940 Yep. (drowned out)

875 00:37:32.940 --> 00:37:34.530 Yeah, and that's the AOS paper.

876 00:37:34.530 --> 00:37:37.830 All the M estimations worked out for this.

877 00:37:37.830 --> 00:37:39.360 The IPW-2, for example.

878 00:37:39.360 --> 00:37:40.193 <v ->Right.</v> <v ->Yep.</v>

879 00:37:44.310 --> 00:37:46.310 In our first results, we actually had a,

880 00:37:48.650 --> 00:37:49.620 this is a smaller, yep, smaller effect size.

881 00:37:49.620 --> 00:37:51.330 The effect size is -0.1,

882 00:37:51.330 --> 00:37:52.950 and this is on the different scale.

883 00:37:52.950 --> 00:37:55.380 So the smaller effect size,

884 00:37:55.380 --> 00:37:58.140 the power was actually surprisingly low.

885 00:37:58.140 --> 00:37:59.850 Even as we increased the number of components,

886 00:37:59.850 --> 00:38:02.010 it didn't even reach 40%.

887 00:38:02.010 --> 00:38:04.860 Although, the coverage of the estimator was approximately

888 00:38:04.860 --> 00:38:08.340 where we'd expect it to be performing.

889 00:38:08.340 --> 00:38:11.760 So the next thing we looked at was changing the effect size,

890 00:38:11.760 --> 00:38:13.280 making the effect size,

891 00:38:13.280 --> 00:38:14.940 in this case, actually making it larger

892 00:38:14.940 --> 00:38:16.983 and seeing how that impacts the power.

893 00:38:18.690 --> 00:38:21.900 So we basically picked...

894 00:38:21.900 --> 00:38:24.300 There's the supplemental slide if anyone has questions,

895 00:38:24.300 --> 00:38:26.700 but we have the original effect size,

896 00:38:26.700 --> 00:38:28.950 the largest effect size that we could obtain

897 00:38:28.950 --> 00:38:31.710 in this particular simulation setting,

898 00:38:31.710 --> 00:38:33.210 and then something in between.

899 00:38:34.200 --> 00:38:36.360 So we see as we increase the effect size

900 00:38:36.360 --> 00:38:40.020 that the largest effect size is -0.42.

901 00:38:40.020 --> 00:38:42.120 That actually achieves 80% power.

902 00:38:42.120 --> 00:38:43.080 Excuse me.

903 00:38:43.080 --> 00:38:46.413 A little bit, actually, it's right around 20 components.

904 00:38:47.250 --> 00:38:49.620 But then as we see, as the effect size gets smaller,

905 00:38:49.620 --> 00:38:54.003 it's harder for it to achieve that 80% power level.

906 00:38:55.830 --> 00:38:57.990 So I thought that was kinda interesting.

907 00:38:57.990 --> 00:39:00.003 And then approach two.

908 00:39:00.900 --> 00:39:04.980 We wanted to see changing the number of components

909 00:39:04.980 --> 00:39:07.260 for a fixed number of nodes.

910 00:39:07.260 --> 00:39:11.790 So here, we fixed a hundred, 300, 600, or a thousand nodes,

911 00:39:11.790 --> 00:39:13.740 and we see it doesn't really matter so much

912 00:39:13.740 --> 00:39:15.240 how many components are in the problem,

913 00:39:15.240 --> 00:39:17.430 which was a little bit surprising to me.

914 00:39:17.430 --> 00:39:19.050 So this is preliminary results.

915 00:39:19.050 --> 00:39:21.120 I'm not sure if this is gonna hold up as we keep

916 00:39:21.120 --> 00:39:25.440 pulling on the threads here, just as a disclaimer.

917 00:39:25.440 --> 00:39:30.150 But we see that with a hundred nodes,

918 00:39:30.150 --> 00:39:33.600 it doesn't achieve the appropriate power.

919 00:39:33.600 --> 00:39:35.110 Once we get up to 300 nodes

920 00:39:39.050 --> 00:39:40.650 and a thousand, sorry, 600 nodes,

921 00:39:40.650 --> 00:39:41.640 and then a thousand nodes,

922 00:39:41.640 --> 00:39:43.833 we see it's at 80% power or higher.

923 00:39:45.660 --> 00:39:49.890 <v Donna>So just to say cluster randomized designs,</v>

924 00:39:49.890 --> 00:39:54.060 in certain structures, you can find that no matter how much,

925 00:39:54.060 --> 00:39:56.733 like if you say the components are like the clusters,

926 00:39:57.606 --> 00:39:58.727 and then the nodes are like

927 00:39:58.727 --> 00:40:00.360 the number of people in that cluster,

928 00:40:00.360 --> 00:40:02.160 you can have a situation where,

929 00:40:02.160 --> 00:40:03.810 for a fixed number of components,

930 00:40:03.810 --> 00:40:07.493 no matter how many people you put into each component,

931 00:40:10.710 --> 00:40:12.090 you have an asymptote.

932 00:40:12.090 --> 00:40:14.160 Never get to the power you want.

933 00:40:14.160 --> 00:40:17.880 The only way to get to it is by increasing components.

934 00:40:17.880 --> 00:40:21.033 But you're finding an asymptote with components.

935 00:40:22.440 --> 00:40:25.830 <v ->Yeah, but here this is the number of people overall</v>

936 00:40:25.830 --> 00:40:28.473 in the whole study, not per component.

937 00:40:29.460 --> 00:40:31.200 So this was a little bit surprising

938 00:40:31.200 --> 00:40:34.170 that it seems to be a bigger driver

939 00:40:34.170 --> 00:40:36.480 is just the number of people enrolled in the network

940 00:40:36.480 --> 00:40:39.150 regardless of the number of components.

941 00:40:39.150 --> 00:40:41.370 <v Donna>So you fixed the total number of units,</v>

942 00:40:41.370 --> 00:40:44.700 and essentially you have them divided

943 00:40:44.700 --> 00:40:46.890 into different numbers of components.

944 00:40:46.890 --> 00:40:47.723 <v ->Yep.</v>

945 00:40:47.723 --> 00:40:50.301 <v Donna>And you're seeing that it doesn't change how many</v>

946 00:40:50.301 --> 00:40:51.150 components (indistinct). <v ->Yeah,</v>

947 00:40:51.150 --> 00:40:53.340 which I also acknowledge that's an artificial thing

948 00:40:53.340 --> 00:40:56.400 that probably would never happen in the real world, right?

949 00:40:56.400 --> 00:40:58.620 Because say we enroll 600 people,

950 00:40:58.620 --> 00:41:01.710 we can't force them into different sets

951 00:41:01.710 --> 00:41:04.051 of partners to get the statistics to work.

952 00:41:04.051 --> 00:41:06.693 So this is a very theoretical thought exercise.

953 00:41:08.340 --> 00:41:09.780 <v Vin>I also wonder if it's a function</v>

954 00:41:09.780 --> 00:41:12.090 of the residual correlation you were specifying

955 00:41:12.090 --> 00:41:13.260 in the simulation study.

956 00:41:13.260 --> 00:41:15.210 <v ->The random effect?</v>

957 00:41:15.210 --> 00:41:16.043 <v Donna>Yeah.</v>

958 00:41:17.304 --> 00:41:18.600 <v ->Interesting.</v> <v ->'Cause that'll definitely</v>

959 00:41:18.600 --> 00:41:20.460 affect the effect sample size, right?

960 00:41:20.460 --> 00:41:21.480 <v ->Mn-hm.</v> <v ->Yeah.</v>

961 00:41:21.480 --> 00:41:23.070 <v Vin>So maybe it's relatively small</v>

962 00:41:23.070 --> 00:41:24.867 and doesn't really matter in this simulation,

963 00:41:24.867 --> 00:41:25.830 and that could be-

964 00:41:25.830 --> 00:41:27.240 <v ->Oh, so if we-</v> <v ->a possibility.</v>

965 00:41:27.240 --> 00:41:30.240 <v ->If we increase the amount of correlation in the component,</v>

966 00:41:30.240 --> 00:41:32.040 this story could be very different.

967 00:41:32.040 --> 00:41:33.150 <v Donna>It might but might not.</v>

968 00:41:33.150 --> 00:41:34.740 So that’s something to check maybe.

969 00:41:34.740 --> 00:41:35.940 <v ->Yep.</v>

970 00:41:35.940 --> 00:41:37.791 That’s why, yeah, another disclaimer.

971 00:41:37.791 --> 00:41:39.120 This is very preliminary.

972 00:41:39.120 --> 00:41:40.680 And I think even at the end I remind us

973 00:41:40.680 --> 00:41:42.990 that needs more investigation.

974 00:41:42.990 --> 00:41:43.823 <v Vin>Right, but it’s cool,</v>

975 00:41:43.823 --> 00:41:46.050 because I guess the cost of randomized design

976 00:41:46.050 --> 00:41:48.930 is sort of a limiting design in some sense.

977 00:41:48.930 --> 00:41:50.130 They probably would not have

978 00:41:50.130 --> 00:41:52.653 the same outputting (indistinct) anyways.

979 00:41:53.724 --> 00:41:54.557 That’s good to-

980 00:41:54.557 --> 00:41:55.807 <v Colleague>What’s the minimum number</v>

981 00:41:55.807 --> 00:41:58.140 of components you could use?

982 00:42:01.182 --> 00:42:02.400 <v ->Looking at the dots, it looks like she went</v>

983 00:42:02.400 --> 00:42:05.220 all the way down to maybe about two,

984 00:42:05.220 --> 00:42:07.440 but it depends on, looks like there’s a...

985 00:42:07.440 --> 00:42:10.170 Depending on which number of nodes you have,

986 00:42:10.170 --> 00:42:12.360 she looks at different numbers of components,

987 00:42:12.360 --> 00:42:17.360 because when Ke generated it, it’s from here.

988 00:42:19.140 --> 00:42:20.670 Yeah, the cluster size is the number of nodes

989 00:42:20.670 --> 00:42:22.470 divided by the number of components.

990 00:42:24.120 --> 00:42:26.667 <v Colleague>So I’m wondering, with these few components</v>

991 00:42:26.667 --> 00:42:30.810 (indistinct) specified?

992 00:42:30.810 --> 00:42:32.133 <v ->Yeah, we should.</v>

993 00:42:33.240 --> 00:42:35.310 Based on other results, it should be.

994 00:42:35.310 --> 00:42:38.550 We start to see good coverage around 50 components.

995 00:42:38.550 --> 00:42:39.959 <v Colleague>That's what I see.</v>

996 00:42:39.959 --> 00:42:40.792 <v ->Yeah.</v>

997 00:42:40.792 --> 00:42:42.280 <v Donna>But I think it would depend</v>

998 00:42:42.280 --> 00:42:43.410 on if the cluster randomized designs

999 00:42:43.410 --> 00:42:46.387 or anything like this would also depend on the ICC.

1000 00:42:47.495 --> 00:42:49.907 Because if that ICC is zero,

1001 00:42:49.907 --> 00:42:52.243 then you could have one component (indistinct)

1002 00:42:52.243 --> 00:42:56.197 is equivalent to, again, a noncluster design.

1003 00:42:56.197 --> 00:42:57.280 <v ->Yeah.</v> <v ->Yep.</v>

1004 00:43:01.530 --> 00:43:03.120 Okay, so here's the preliminary results

1005 00:43:03.120 --> 00:43:04.920 for the wall test statistic.

1006 00:43:04.920 --> 00:43:07.530 So I changed the notation a little bit here

1007 00:43:07.530 --> 00:43:09.480 just to make this easier to read.

1008 00:43:09.480 --> 00:43:12.450 So uber expressed, the estimator is this theta hat.

1009 00:43:12.450 --> 00:43:15.180 Based on the AOS paper, we have that this will converge

1010 00:43:15.180 --> 00:43:17.340 in distribution to a multivariate normal.

1011 00:43:17.340 --> 00:43:20.250 And then we actually have an estimator

1012 00:43:20.250 --> 00:43:23.913 of the variance in that paper, as well.

1013 00:43:26.880 --> 00:43:29.130 Yeah, and then building a wall test statistic

1014 00:43:30.300 --> 00:43:33.780 from that parameter, we have a form that looks like this.

1015 00:43:33.780 --> 00:43:35.430 And then actually in the AOS paper,

1016 00:43:35.430 --> 00:43:37.830 just a minor note is the normalizing constant

1017 00:43:37.830 --> 00:43:41.520 of one over M is tucked into the sigma term.

1018 00:43:41.520 --> 00:43:44.370 I had to go back and double check that yesterday.

1019 00:43:44.370 --> 00:43:45.750 So then we have a wall test statistic

1020 00:43:45.750 --> 00:43:47.220 that's a form like this.

1021 00:43:47.220 --> 00:43:51.003 It should follow a normal distribution.

1022 00:43:54.300 --> 00:43:56.130 So then we started looking at this

1023 00:43:56.130 --> 00:43:58.500 empirically across the simulations.

1024 00:43:58.500 --> 00:44:01.860 And this looks, to my eye, to be approximately normal.

1025 00:44:01.860 --> 00:44:03.750 And what we're working on now,

1026 00:44:03.750 --> 00:44:05.190 the results aren't quite ready,

1027 00:44:05.190 --> 00:44:07.620 is actually doing a test for a normality

1028 00:44:07.620 --> 00:44:09.960 like a Kolmogorov-Smirnov test

1029 00:44:09.960 --> 00:44:13.473 to test for normality across these different scenarios.

1030 00:44:14.340 --> 00:44:16.230 So we're working on those results now,

1031 00:44:16.230 --> 00:44:17.790 and that's something we wanted to confirm

1032 00:44:17.790 --> 00:44:20.913 before we fold it into the rest of the simulations.

1033 00:44:22.867 --> 00:44:24.690 <v Donna>That test has very low power (indistinct).</v>

1034 00:44:24.690 --> 00:44:25.533 <v ->Low power?</v>

1035 00:44:27.030 --> 00:44:28.860 Yeah, and then there's other tests too,

1036 00:44:28.860 --> 00:44:29.820 but some of 'em are-

1037 00:44:29.820 --> 00:44:31.710 <v Donna>I think they all have low power.</v>

1038 00:44:31.710 --> 00:44:32.543 <v ->Yeah.</v>

1039 00:44:33.780 --> 00:44:35.550 So if anyone has any other thoughts about that,

1040 00:44:35.550 --> 00:44:37.020 about how to evaluate.

1041 00:44:37.020 --> 00:44:39.660 Like we derived this, but how do we-

1042 00:44:39.660 --> 00:44:42.090 <v Donna>In some sense, your simulations will tell you,</v>

1043 00:44:42.090 --> 00:44:45.330 because the property's relying

1044 00:44:45.330 --> 00:44:47.272 on that (indistinct) normality.

1045 00:44:47.272 --> 00:44:51.034 And so if you don't have 5% type one error, 1046 00:44:51.034 --> 00:44:52.067 and then you know (indistinct),

1047 00:44:53.416 --> 00:44:54.988 you now have...

1048 00:44:54.988 --> 00:44:55.821 I guess that would be the main thing

1049 00:44:55.821 --> 00:44:57.423 would 5% type one error.

1050 00:45:02.270 --> 00:45:04.020 <v Vin>I think maybe another way to visualize</v>

1051 00:45:04.020 --> 00:45:08.670 that is to try to increase the M,

1052 00:45:08.670 --> 00:45:11.610 and then actually gradually see if that looks more normal.

1053 00:45:11.610 --> 00:45:13.410 I guess that's just-

1054 00:45:13.410 --> 00:45:14.910 <v ->Yep.</v>

1055 00:45:14.910 --> 00:45:18.013 <v Vin>And I think people tend to do something like that.</v>

1056 00:45:18.013 --> 00:45:19.830 When they check convergence rate,

1057 00:45:19.830 --> 00:45:22.713 they would probably do something like plot

1058 00:45:22.713 --> 00:45:25.380 the results along with the sample size

1059 00:45:25.380 --> 00:45:27.576 and see how well they converge.

1060 00:45:27.576 --> 00:45:29.300 And then the limiting end would correspond

1061 00:45:29.300 --> 00:45:31.117 to the perfect results,

1062 00:45:31.117 --> 00:45:33.120 and then you'll see more of a bell curve shape.

1063 00:45:33.120 --> 00:45:35.520 But I think right now, looking at these 10 iterations,

1064 00:45:35.520 --> 00:45:37.080 it's a little spiky sometimes.

1065 00:45:37.080 --> 00:45:38.130 <v ->Yeah, and it doesn't seem...</v>

1066 00:45:38.130 --> 00:45:39.870 Like this one down in the far corner

1067 00:45:39.870 --> 00:45:40.797 is already a hundred components,

1068 00:45:40.797 --> 00:45:45.120 and it doesn't really seem like it's getting too much...

1069 00:45:45.120 --> 00:45:47.970 I mean, these are at least, yeah.

1070 00:45:47.970 --> 00:45:49.446 There's not a trend of constant-

1071 00:45:49.446 --> 00:45:50.940 <v Vin>(drowned out) specified model, right?</v>

1072 00:45:50.940 --> 00:45:53.520 It's definitely correctly specified

1073 00:45:53.520 --> 00:45:55.440 propensity score models and everything-

1074 00:45:55.440 --> 00:45:57.300 <v ->Should be, but we can double check.</v>

1075 00:45:57.300 --> 00:45:58.830 <v Vin>So the simulation models</v>

1076 00:45:58.830 --> 00:46:01.470 are basically identical to the models (drowned out).

1077 00:46:01.470 --> 00:46:02.303 <v ->Yep.</v>

1078 00:46:03.746 --> 00:46:04.579 <v Donna>But the spiking,</v>

1079 00:46:04.579 --> 00:46:07.980 this also just depends arbitrarily on the event size?

1080 00:46:07.980 --> 00:46:09.150 <v Vin>Yeah, that's right.</v>

1081 00:46:09.150 --> 00:46:11.190 <v Donna>So you could make it look very spiky</v>

1082 00:46:11.190 --> 00:46:12.762 if you have bigger events.

1083 00:46:12.762 --> 00:46:14.208 <v Vin>Right, and (indistinct)</v>

1084 00:46:14.208 --> 00:46:15.270 you could even Q-Q plot events sometimes.

1085 00:46:15.270 --> 00:46:16.858 <v ->Yeah.</v> <v ->Yep.</v>

1086 00:46:16.858 --> 00:46:18.356 (drowned out)

1087 00:46:18.356 --> 00:46:19.856 Vin says Q-Q plot.

1088 00:46:20.805 --> 00:46:21.638 (Donna laughs)

1089 00:46:21.638 --> 00:46:22.533 (indistinct)

1090 00:46:22.533 --> 00:46:23.366 Okay.

1091 00:46:25.010 --> 00:46:28.230 So that's the direction where we're heading in with this.

1092 00:46:28.230 --> 00:46:30.720 From simulation two, we also have some preliminary results.

1093 00:46:30.720 --> 00:46:32.550 So this is fixing the number of components

1094 00:46:32.550 --> 00:46:34.413 and varying the number of nodes.

1095 00:46:35.850 --> 00:46:40.680 In here, we see power increases with the number of nodes,

1096 00:46:40.680 --> 00:46:42.820 but we don't see any variation

1097 00:46:44.190 --> 00:46:45.510 between the number of components.

1098 00:46:45.510 --> 00:46:47.967 So the power is plotted against the number of nodes

1099 00:46:47.967 --> 00:46:51.660 and each line represents a different number of components,

1100 00:46:51.660 --> 00:46:54.060 which I think kind of echoes the other results

1101 00:46:54.060 --> 00:46:58.233 that we were seeing earlier in the talk.

1102 00:47:01.230 --> 00:47:04.470 <v Donna>That's the opposite of cluster randomized trials,</v>

1103 00:47:04.470 --> 00:47:07.950 'cause you're getting a lot of power by increasing nodes,

1104 00:47:07.950 --> 00:47:10.380 and you're barely seeing any impact of components.

1105 00:47:10.380 --> 00:47:12.210 Whereas with cluster randomized trials,

1106 00:47:12.210 --> 00:47:14.340 it's all in the clusters,

1107 00:47:14.340 --> 00:47:16.980 and it doesn't matter that much after a relatively

1108 00:47:16.980 --> 00:47:19.320 small number of people within cluster.

1109 00:47:19.320 --> 00:47:20.250 <v Vin>Right.</v>

1110 00:47:20.250 --> 00:47:22.140 <v ->Which this is still very surprising to me,</v>

1111 00:47:22.140 --> 00:47:23.280 because the M estimation,

1112 00:47:23.280 --> 00:47:25.740 the effective sample size is the number of components.

1113 00:47:25.740 --> 00:47:29.020 So yeah, this is pretty surprising.

1114 00:47:29.020 --> 00:47:29.967 <v Vin>(indistinct) interested to really check</v>

1115 00:47:29.967 --> 00:47:32.397 how that changes or not changes with the-

1116 00:47:32.397 --> 00:47:34.061 <v ->The IC?</v> <v ->Yeah.</v>

1117 00:47:34.061 --> 00:47:35.517 <v ->Change the. (drowned out)</v>

1118 00:47:35.517 --> 00:47:36.434 Yes. <v ->Yeah.</v>

1119 00:47:39.959 --> 00:47:40.792 <v Donna>What is the outcome?</v>

1120 00:47:40.792 --> 00:47:43.260 Like sort of this idea in this simulation,

1121 00:47:43.260 --> 00:47:45.309 what were you thinking of?

1122 00:47:45.309 --> 00:47:46.577 Is it a binary or a continuous?

1123 00:47:46.577 --> 00:47:48.573 <v ->Binary HIV risk behavior.</v>

1124 00:47:50.739 --> 00:47:52.170 So yeah, whether the person reports,

1125 00:47:52.170 --> 00:47:54.663 specifically injection risk behavior.

1126 00:47:56.550 --> 00:47:58.085 And then the intervention,

1127 00:47:58.085 --> 00:48:00.270 all the effects that we're looking at are negative,

1128 00:48:00.270 --> 00:48:05.270 because the intervention should be reducing the behavior.

1129 00:48:05.550 --> 00:48:09.810 <v Donna>Yeah, so with an ICC of 0.5 times 1 minus 0.5,</v>

1130 00:48:09.810 --> 00:48:12.750 that's the maximum amount of binomial variants.

1131 00:48:12.750 --> 00:48:13.800 So this should be...

1132 00:48:13.800 --> 00:48:17.913 The simulation is done under a very high ICC.

1133 00:48:19.530 --> 00:48:21.840 Like it might be the highest possible with binary-

1134 00:48:21.840 --> 00:48:23.613 <v ->For that binary data, yep.</v>

1135 00:48:27.450 --> 00:48:29.250 Okay, so zooming out a little bit,

1136 00:48:29.250 --> 00:48:32.760 thinking about network study design in practice,

1137 00:48:32.760 --> 00:48:35.100 some of the things that might come out of this work.

1138 00:48:35.100 --> 00:48:37.320 So there are definitely features that can be planned

1139 00:48:37.320 --> 00:48:39.060 when designing the study, right?

1140 00:48:39.060 --> 00:48:41.430 So we could increase the number of components

1141 00:48:41.430 --> 00:48:43.980 by having multiple sites or multiple cities

1142 00:48:43.980 --> 00:48:46.593 contributing to one particular study.

1143 00:48:48.540 --> 00:48:50.370 Although, that's, you know, can be very costly,

1144 00:48:50.370 --> 00:48:52.140 very time consuming.

1145 00:48:52.140 --> 00:48:54.810 We can, of course, increase more individuals recruited, 1146 00:48:54.810 --> 00:48:57.000 but that depends on who,

1147 00:48:57.000 --> 00:48:59.100 ’cause it’s a network study, who are their contacts,

1148 00:48:59.100 --> 00:49:00.240 if they don’t have contacts

1149 00:49:00.240 --> 00:49:03.480 to kind of come to an end in the network.

1150 00:49:03.480 --> 00:49:06.300 We can try to ensure distance between components some way.

1151 00:49:06.300 --> 00:49:07.740 And I put distance in quotes,

1152 00:49:07.740 --> 00:49:10.440 ’cause that could mean all sorts of things,

1153 00:49:10.440 --> 00:49:12.390 not just geographical distance.

1154 00:49:12.390 --> 00:49:13.320 And then we have some control

1155 00:49:13.320 --> 00:49:14.700 over the intervention treatment.

1156 00:49:14.700 --> 00:49:17.823 What proportion do we want to expose to the intervention?

1157 00:49:18.990 --> 00:49:19.920 And then I was thinking about features

1158 00:49:19.920 --> 00:49:22.440 that likely cannot be planned,

1159 00:49:22.440 --> 00:49:23.880 ’cause maybe someone’s really creative.

1160 00:49:23.880 --> 00:49:25.200 And we could think about ways

1161 00:49:25.200 --> 00:49:28.200 that these could be manipulated.

1162 00:49:28.200 --> 00:49:31.200 So once we have a given set of individuals,

1163 00:49:31.200 --> 00:49:34.230 pretty sure we can’t force them into different components,

1164 00:49:34.230 --> 00:49:36.510 unless we’re doing, actually now that’s coming to my mind,

1165 00:49:36.510 --> 00:49:38.550 unless we’re doing a network intervention

1166 00:49:38.550 --> 00:49:40.770 that’s meant to change the edges.

1167 00:49:40.770 --> 00:49:42.240 Then, we would have some control

1168 00:49:42.240 --> 00:49:44.760 over who’s interacting with whom,

1169 00:49:44.760 --> 00:49:46.050 but that’s a little bit complicated,

1170 00:49:46.050 --> 00:49:48.210 because then your structure is intertwined

1171 00:49:48.210 --> 00:49:49.503 with your intervention.

1172 00:49:51.060 --> 00:49:52.740 The features of the network like degree, 1173 00:49:52.740 --> 00:49:55.940 centrality, intracluster correlation,

1174 00:49:55.940 --> 00:49:58.110 we don’t have control over those.

1175 00:49:58.110 --> 00:49:59.190 Who’s connected to whom:

1176 00:49:59.190 --> 00:50:01.530 these are individual sexual and drug partnerships.

1177 00:50:01.530 --> 00:50:03.847 We don’t have control over that.

1178 00:50:03.847 --> 00:50:04.680 What the effect sizes are

1179 00:50:04.680 --> 00:50:08.310 or what the outcome prevalence is in the particular study.

1180 00:50:08.310 --> 00:50:11.640 <v Donna>Well, you can’t choose your study population,</v>

1181 00:50:11.640 --> 00:50:15.150 though, to have certain of these characteristics.

1182 00:50:15.150 --> 00:50:17.040 You can’t change them.

1183 00:50:17.040 --> 00:50:18.390 Let’s say you could do a study

1184 00:50:18.390 --> 00:50:21.180 of 10 different kind of places, communities,

1185 00:50:21.180 --> 00:50:22.650 and some might be more-

1186 00:50:22.650 --> 00:50:24.735 <v ->Different outcome prevalences or-</v>

1187 00:50:24.735 --> 00:50:26.550 <v Donna>Yeah, or different degrees of centrality,</v>

1188 00:50:26.550 --> 00:50:29.643 and they could have different ICCs and all of that.

1189 00:50:30.630 --> 00:50:33.090 So if people know what’s important,

1190 00:50:33.090 --> 00:50:37.020 they could look for study populations that have the features

1191 00:50:37.020 --> 00:50:39.180 that will maximize power of the study.

1192 00:50:39.180 --> 00:50:42.180 <v ->Yep, that’s a good point.</v>

1193 00:50:42.180 --> 00:50:43.110 That’s why I said likely,

1194 00:50:43.110 --> 00:50:45.158 ’cause I knew Donna would think of something.

1195 00:50:45.158 --> 00:50:46.669 (laughs)

1196 00:50:46.669 --> 00:50:49.086 (indistinct)

1197 00:50:50.970 --> 00:50:52.320 <v Colleague>What about the propensity score?</v>

1198 00:50:52.320 --> 00:50:53.890 You also don't have control.

1199 00:50:56.880 --> 00:50:58.620 <v ->Yeah, I mean that's the...</v>

1200 00:50:58.620 --> 00:51:00.870 It was non-randomized intervention.

1201 00:51:00.870 --> 00:51:04.560 So it's what the folks are are choosing or being exposed to

1202 00:51:04.560 --> 00:51:06.560 and then just their observed covariates.

1203 00:51:08.340 --> 00:51:11.790 <v Donna>Oh, there's one way, just randomize them.</v>

1204 00:51:11.790 --> 00:51:12.960 (drowned out) (laughing)

1205 00:51:12.960 --> 00:51:15.180 In epidemiology, we always talk about this

1206 00:51:15.180 --> 00:51:17.765 as one of the ways to control confounding,

1207 00:51:17.765 --> 00:51:20.040 which is to choose a homogeneous population

1208 00:51:20.040 --> 00:51:23.520 so you have no variation in the risk factors,

1209 00:51:23.520 --> 00:51:26.343 and that lowers the amount of confounding.

1210 00:51:27.660 --> 00:51:30.780 You might lose the ability to externally channelize,

1211 00:51:30.780 --> 00:51:32.463 but you'll reduce confounding.

1212 00:51:37.950 --> 00:51:41.190 <v ->Yeah, so I think there's a lot of thinking</v>

1213 00:51:41.190 --> 00:51:44.850 and papers that need to be written for design in networks.

1214 00:51:44.850 --> 00:51:46.920 I mean, I think in designing trials

1215 00:51:46.920 --> 00:51:48.480 and designing cluster randomized trials,

1216 00:51:48.480 --> 00:51:51.090 even thinking about observational studies,

1217 00:51:51.090 --> 00:51:52.860 I think it's clear to me how you have

1218 00:51:52.860 --> 00:51:55.230 more control over certain things.

1219 00:51:55.230 --> 00:51:58.440 But then here, I think there's a lot of work

1220 00:51:58.440 --> 00:52:01.380 to think about how do we take...

1221 00:52:01.380 --> 00:52:02.213 It's just in the beginning

1222 00:52:02.213 --> 00:52:03.720 with some of these statistical results,

1223 00:52:03.720 --> 00:52:05.850 but how do we take these statistical findings

1224 00:52:05.850 --> 00:52:07.770 and translate them into something that folks

1225 00:52:07.770 --> 00:52:10.050 can actually use in study designs,

1226 00:52:10.050 --> 00:52:13.800 grant proposals for network-based studies in public health.

1227 00:52:13.800 --> 00:52:15.480 So I think that's a call to action

1228 00:52:15.480 --> 00:52:18.483 to some of the folks in the room and on Zoom.

1229 00:52:20.820 --> 00:52:23.310 So just some highlights from what we found so far.

1230 00:52:23.310 --> 00:52:25.140 So the power for estimating spillover effects

1231 00:52:25.140 --> 00:52:28.083 increases with more nodes or larger effect sizes.

1232 00:52:30.360 --> 00:52:32.490 It requires, of course, more investigation

1233 00:52:32.490 --> 00:52:33.600 like we've been discussing today.

1234 00:52:33.600 --> 00:52:35.370 There's some things we need to look into,

1235 00:52:35.370 --> 00:52:38.700 but the number of components may have less impact on power,

1236 00:52:38.700 --> 00:52:42.150 but that requires looking at some additional features.

1237 00:52:42.150 --> 00:52:43.560 When the effect size is large enough,

1238 00:52:43.560 --> 00:52:45.930 the spillover effect has reasonable power.

1239 00:52:45.930 --> 00:52:46.800 And then in the initial setting,

1240 00:52:46.800 --> 00:52:49.830 that was even with only 20 components.

1241 00:52:49.830 --> 00:52:51.900 And then just as a sanity check,

1242 00:52:51.900 --> 00:52:54.660 we saw the empirical coverage probability

1243 00:52:54.660 --> 00:52:56.700 was around the nominal level

1244 00:52:56.700 --> 00:52:59.403 as we would expect from our earlier paper.

1245 00:53:01.110 --> 00:53:02.430 So future directions.

1246 00:53:02.430 --> 00:53:03.780 We wanna keep looking at the impact

1247 00:53:03.780 --> 00:53:06.483 of other design parameters on the power,

1248 00:53:07.410 --> 00:53:09.780 continue working with this test statistic

1249 00:53:09.780 --> 00:53:12.660 and making sure it's performing as we expect,

1250 00:53:12.660 --> 00:53:14.727 and then using it in the simulation study

1251 00:53:14.727 --> 00:53:17.790 and working on getting a minimal detectable effect size,

1252 00:53:17.790 --> 00:53:21.306 as well as number of individuals

1253 00:53:21.306 --> 00:53:23.820 and/or components required for adequate power.

1254 00:53:23.820 --> 00:53:25.590 And if we have confined closed forms,

1255 00:53:25.590 --> 00:53:26.730 we'll have those expressions.

1256 00:53:26.730 --> 00:53:29.070 If not, we'll have some simulation-based programs

1257 00:53:29.070 --> 00:53:30.480 to look at this.

1258 00:53:30.480 --> 00:53:31.830 And then we want to...

1259 00:53:31.830 --> 00:53:33.780 We've done some kind of back-of-the-envelope things

1260 00:53:33.780 --> 00:53:35.700 in thinking about the power that we might have had

1261 00:53:35.700 --> 00:53:37.500 in TRIP to detect these effects

1262 00:53:37.500 --> 00:53:40.443 but doing that more carefully and formally.

1263 00:53:41.850 --> 00:53:43.260 And then last was sort of the issue

1264 00:53:43.260 --> 00:53:44.310 I was talking about at the end

1265 00:53:44.310 --> 00:53:47.610 is all of these statistical results are really interesting

1266 00:53:47.610 --> 00:53:49.770 and exciting for folks like us,

1267 00:53:49.770 --> 00:53:51.630 but then how do we make it practical

1268 00:53:51.630 --> 00:53:55.500 and useful and something that individuals

1269 00:53:55.500 --> 00:53:56.760 can use in their grant writing

1270 00:53:56.760 --> 00:53:59.193 when getting their network based studies funded.

1271 00:54:01.170 --> 00:54:03.210 Okay, and then this is my shameless plug.

1272 00:54:03.210 --> 00:54:06.060 If you thought this talk was interesting,

1273 00:54:06.060 --> 00:54:10.018 we're going to have an online workshop hosted by my group

1274 00:54:10.018 --> 00:54:12.600 at URI on Friday, March 10th from 2:00 to 5:00.

1275 00:54:12.600 --> 00:54:17.280 It's free and we have a star-studded lineup

1276 00:54:17.280 --> 00:54:20.340 of speakers that'll be joining for the workshop.

1277 00:54:20.340 --> 00:54:21.900 And I have some flyers,

1278 00:54:21.900 --> 00:54:24.360 and I can email the flyer around, as well.

1279 00:54:24.360 --> 00:54:26.232 <v Donna>We can circulate everything.</v>

1280 00:54:26.232 --> 00:54:27.840 (indistinct) also.

1281 00:54:27.840 --> 00:54:29.550 <v ->Yeah, that'd be great.</v>

1282 00:54:29.550 --> 00:54:31.290 Yeah, so welcome everyone on the call,

1283 00:54:31.290 --> 00:54:33.573 everyone in the room to join,

1284 00:54:34.470 --> 00:54:36.390 and I think it'll be a really informative

1285 00:54:36.390 --> 00:54:38.130 and interesting afternoon.

1286 00:54:38.130 --> 00:54:40.290 And if you're interested in this methods area,

1287 00:54:40.290 --> 00:54:42.210 it'd be a nice way to get caught up

1288 00:54:42.210 --> 00:54:43.647 on some of the literature

1289 00:54:43.647 --> 00:54:45.720 and start thinking about how you can use this

1290 00:54:45.720 --> 00:54:47.043 in some of your work.

1291 00:54:48.930 --> 00:54:51.573 So just a couple of references, as well.

1292 00:54:53.202 --> 00:54:54.690 And I know I've been taking questions as we go along,

1293 00:54:54.690 --> 00:54:57.960 but if there's any other questions from the audience,

1294 00:54:57.960 --> 00:54:59.733 happy to discuss.

1295 00:55:01.863 --> 00:55:03.990 <v Vin>So it's interesting to see that the component size</v>

1296 00:55:03.990 --> 00:55:07.110 doesn't have a very strong effect on the power,

1297 00:55:07.110 --> 00:55:11.550 but do you think in reality we need also consider

1298 00:55:11.550 --> 00:55:14.199 variability in that component size?

1299 00:55:14.199 --> 00:55:16.326 'Cause we always see a huge component.

1300 00:55:16.326 --> 00:55:18.042 <v ->Yep, that's a really good point.</v>

1301 00:55:18.042 --> 00:55:20.069 <v Vin>But there are a lot of very small components.</v>

1302 00:55:20.069 --> 00:55:21.750 <v ->Yep. (drowned out)</v>

1303 00:55:21.750 --> 00:55:22.893 Yep, great point.

1304 00:55:24.420 --> 00:55:28.350 And particularly in these HIV risk networks,

1305 00:55:28.350 --> 00:55:29.483 I mean, it's not like there's hundreds of them,

1306 00:55:29.483 --> 00:55:32.910 but the handful that we have and we've been able to look at,

1307 00:55:32.910 --> 00:55:34.320 there is a lot of variability.

1308 00:55:34.320 --> 00:55:36.720 We have, usually, there's one giant connected component

1309 00:55:36.720 --> 00:55:39.300 and then these smaller components.

1310 00:55:39.300 --> 00:55:41.640 And of course, whether or not that's the real network,

1311 00:55:41.640 --> 00:55:44.700 that's some of Laura's work, right?

1312 00:55:44.700 --> 00:55:46.860 These smaller components may actually even be connected

1313 00:55:46.860 --> 00:55:47.910 to the larger component,

1314 00:55:47.910 --> 00:55:49.950 or they might be connected to each other, as well.

1315 00:55:49.950 --> 00:55:54.330 But in this work, we assume that the network we observe

1316 00:55:54.330 --> 00:55:57.690 is the truer known network for now

1317 00:55:57.690 --> 00:55:59.520 just so we can look at some of these other issues.

1318 00:55:59.520 --> 00:56:01.050 But of course, there's always the caveat

1319 00:56:01.050 --> 00:56:03.870 that the network itself is mismeasured.

1320 00:56:03.870 --> 00:56:05.430 <v ->'Cause they-</v> <v ->Ashley, there's a bunch</v>

1321 00:56:05.430 --> 00:56:07.320 of things up in the chat maybe.

1322 00:56:07.320 --> 00:56:08.850 Just to give other people a chance.

1323 00:56:08.850 --> 00:56:11.173 <v ->Sure.</v> <v ->Some of it might have to do</v>

1324 00:56:11.173 --> 00:56:13.170 with the beginning when we were having technical problems,

1325 00:56:13.170 --> 00:56:15.713 but it might have some questions.

1326 00:56:15.713 --> 00:56:16.980 <v ->"See you have some technical problems.</v>

1327 00:56:16.980 --> 00:56:18.390 Slide's not moving."

1328 00:56:18.390 --> 00:56:20.280 Oh, and then thanks, Gabby.

1329 00:56:20.280 --> 00:56:21.510 Gabby's part of the URI team.

1330 00:56:21.510 --> 00:56:23.973 She put in a link to register for the workshop.

1331 00:56:24.960 --> 00:56:25.793 We actually,

1332 00:56:26.670 --> 00:56:28.770 we just have a couple survey questions as you register,

1333 00:56:28.770 --> 00:56:30.450 because what we wanna do is try to tailor

1334 00:56:30.450 --> 00:56:33.690 the content to the folks that are showing up.

1335 00:56:33.690 --> 00:56:35.490 So there's just a couple of quick questions,

1336 00:56:35.490 --> 00:56:37.380 and then that's all you have to do.

1337 00:56:37.380 --> 00:56:38.580 It's free (laughs).

1338 00:56:38.580 --> 00:56:40.440 Just answer a little survey.

1339 00:56:40.440 --> 00:56:44.310 And then Gabby put a link for some more details

1340 00:56:44.310 --> 00:56:45.710 about the workshop, as well.

1341 00:56:47.555 --> 00:56:48.388 (indistinct)

1342 00:56:48.388 --> 00:56:49.705 <v Donna>This has gotta be our last question,</v>

1343 00:56:49.705 --> 00:56:50.538 'cause we're down to 12.

1344 00:56:50.538 --> 00:56:51.987 <v Vin>Yeah, just a short comment.</v>

1345 00:56:51.987 --> 00:56:55.095 I think there's a potential to make this work more impactful

1346 00:56:55.095 --> 00:56:59.529 is that it doesn't have to be attached to IPW-2 I think,

1347 00:56:59.529 --> 00:57:01.080 because you're providing a simulation framework.

1348 00:57:01.080 --> 00:57:04.815 And theoretically, one can fit other IPW estimators, 1349 00:57:04.815 --> 00:57:06.687 certified estimators, regression based estimators,

1350 00:57:06.687 --> 00:57:09.000 and even double robust estimators.

1351 00:57:09.000 --> 00:57:12.330 And I would also imagine that they could have

1352 00:57:12.330 --> 00:57:14.100 different operating characteristics,

1353 00:57:14.100 --> 00:57:17.373 and so the impact of M and N could also,

1354 00:57:18.630 --> 00:57:20.580 that could also be specific

1355 00:57:20.580 --> 00:57:25.580 to not only the simulation parameters we choose,

1356 00:57:25.710 --> 00:57:27.750 but also to the estimators we choose.

1357 00:57:27.750 --> 00:57:31.020 I think it's an underappreciated point,

1358 00:57:31.020 --> 00:57:35.040 but it's very important to emphasize is that the power

1359 00:57:35.040 --> 00:57:38.167 we calculate is always gonna be based on approach.

1360 00:57:38.167 --> 00:57:39.655 <v Donna>It's true that it's underappreciated.</v>

1361 00:57:39.655 --> 00:57:40.963 Surprisingly, right?

1362 00:57:40.963 --> 00:57:42.780 <v Vin>Yeah, like you could say I use the approach</v>

1363 00:57:42.780 --> 00:57:45.570 to consider IPW-2 based power,

1364 00:57:45.570 --> 00:57:47.551 but I think a regression based approach

1365 00:57:47.551 --> 00:57:49.530 in terms of power would be different.

1366 00:57:49.530 --> 00:57:50.850 It's actually very specific, too.

1367 00:57:50.850 --> 00:57:54.720 And also, it curves to show could have some difference.

1368 00:57:54.720 --> 00:57:55.950 <v ->Yeah, that's interesting.</v>

1369 00:57:55.950 --> 00:57:59.400 So we can start, 'cause we have IPW-1, IPW-2 ready to go.

1370 00:57:59.400 --> 00:58:02.399 So we could start, for this work, we could look at that.

1371 00:58:02.399 --> 00:58:04.860 But I think maybe an idea would be to write the code.

1372 00:58:04.860 --> 00:58:07.320 Like if we have our programs that we're gonna share for this

1373 00:58:07.320 --> 00:58:09.868 to write it flexible enough so that the user-

1374 00:58:09.868 --> 00:58:10.920 <v Vin>That's something people should be able to choose.</v>

1375 00:58:10.920 --> 00:58:14.070 Or even if you have a estimate or specific program,

1376 00:58:14.070 --> 00:58:16.770 that should be sort of emphasized and clarified.

1377 00:58:16.770 --> 00:58:20.070 'Cause as a very simple example, if you are,

1378 00:58:20.070 --> 00:58:22.140 like in the cluster (indistinct) literature,

1379 00:58:22.140 --> 00:58:24.030 if you're assuming working independence

1380 00:58:24.030 --> 00:58:25.140 and working exchangeable,

1381 00:58:25.140 --> 00:58:26.370 the results can be very different

1382 00:58:26.370 --> 00:58:28.080 in terms of the efficiency.

1383 00:58:28.080 --> 00:58:32.610 And the extent to which the cluster size variation

1384 00:58:32.610 --> 00:58:36.900 impact the study power is also specific to whether you adopt

1385 00:58:36.900 --> 00:58:40.530 a independence working correlation or an exchangeable.

1386 00:58:40.530 --> 00:58:44.760 So sometimes, we have a unified conclusion,

1387 00:58:44.760 --> 00:58:49.760 but that's almost always coming from a specific estimate

1388 00:58:50.820 --> 00:58:52.500 and cannot really be overly generalized.

1389 00:58:52.500 --> 00:58:54.060 <v ->Yep, yeah, that's a great point.</v>

1390 00:58:54.060 --> 00:58:54.893 Thanks, Vin.

1391 00:58:55.950 --> 00:58:58.560 <v ->Well, this was a really interesting seminar, Ashley.</v>

1392 00:58:58.560 --> 00:58:59.967 <v ->Thank you.</v> <v ->You presented it</v>

1393 00:58:59.967 --> 00:59:01.230 very clearly,

1394 00:59:01.230 --> 00:59:02.580 so we really appreciate it.

1395 00:59:02.580 --> 00:59:06.493 Thank you so much and thanks to everybody else (indistinct).

1396 00:59:06.493 --> 00:59:07.883 <v ->Thank you, thanks, everyone.</v>

1397 00:59:12.637 --> 00:59:14.121 <v Donna>So go ahead and close the Zoom.</v>

1398 00:59:14.121 --> 00:59:15.510 <v ->Sure, yeah, thanks, everyone, for joining.</v>

1399 00:59:15.510 --> 00:59:17.610 We hope to see you at the online workshop.