WEBVTT

1 00:00:00.690 --> 00:00:03.090 <v Laura>All right, let's get started.</v>

2 00:00:03.090 --> 00:00:05.100 Thank you, everyone, for coming.

3 00:00:05.100 --> 00:00:07.473 So let me introduce our speaker today.

4 00:00:08.460 --> 00:00:11.430 Ariel Chao is a PhD student in the department

5 00:00:11.430 --> 00:00:16.320 of biostatistics, advised by me and Donna Spiegelman.

6 00:00:16.320 --> 00:00:19.653 So let me say few things about her, about Ariel.

7 00:00:20.670 --> 00:00:23.850 So I've been working with Ariel for three years now

8 00:00:23.850 --> 00:00:27.060 and I have to say it's been a real pleasure.

9 00:00:27.060 --> 00:00:30.300 Ariel is an extraordinary student, very patient,

10 00:00:30.300 --> 00:00:34.020 definitely her characteristic and independent.

11 00:00:34.020 --> 00:00:36.780 I've always been impressed by her creativity

12 00:00:36.780 --> 00:00:40.710 and the way she would always find solutions by herself.

13 00:00:40.710 --> 00:00:42.240 We have been having issues

14 00:00:42.240 --> 00:00:44.550 with getting data from our collaborators

15 00:00:44.550 --> 00:00:48.240 and she never gave up and found ways to keep working

16 00:00:48.240 --> 00:00:50.850 on what she had while waiting.

17 00:00:50.850 --> 00:00:53.400 So she deeply cares about the applications

18 00:00:53.400 --> 00:00:57.450 and she's working on and she has a great intuition.

19 00:00:57.450 --> 00:00:58.860 I also been impressed

20 00:00:58.860 --> 00:01:03.060 on how she can work in several things at the same time.

21 00:01:03.060 --> 00:01:05.970 And as you will see today, she's very talented

22 00:01:05.970 --> 00:01:09.565 and I wish her the best for her future career.

23 00:01:09.565 --> 00:01:12.660 Before that, today, she will present her work

24 00:01:12.660 --> 00:01:14.807 on addressing bias in causal effects,

25 00:01:14.807 --> 00:01:18.480 estimated underspecified interference sets

26 00:01:18.480 --> 00:01:22.470 with application to HIV prevention trials.

27 00:01:22.470 --> 00:01:25.710 So let's give Ariel a more welcome.

28 00:01:25.710 --> 00:01:26.860 Ariel, (crackling drowns out speaker).

29 00:01:30.390 --> 00:01:32.070 <v Speaker>Let me just add,</v>

30 00:01:32.070 --> 00:01:34.350 Ariel has a lot of material to present

31 00:01:34.350 --> 00:01:38.460 so we decided to not take questions while she's talking

32 00:01:38.460 --> 00:01:40.451 or she'll never get through the necessarily.

33 00:01:40.451 --> 00:01:41.760 And then we're gonna allow

34 00:01:41.760 --> 00:01:44.220 for around 10 minutes at the end for questions.

35 00:01:44.220 --> 00:01:45.360 So write down the questions

36 00:01:45.360 --> 00:01:48.180 and then we'll try to give as many people a chance

37 00:01:48.180 --> 00:01:49.780 to ask her questions at the end.

38 00:01:51.330 --> 00:01:53.100 <v Laura>I will keep track of it.</v>

39 00:01:53.100 --> 00:01:55.380 <v Speaker>I'm just monitoring the chat.</v>

40 00:01:55.380 --> 00:01:57.900 <v ->Oh yes, can someone, 'cause I don't think I can see you.</v>

41 00:01:57.900 --> 00:02:00.363 <v Speaker>I can see you.</v>

42 00:02:01.470 --> 00:02:02.303 <v ->All right.</v>

43 00:02:02.303 --> 00:02:04.830 So thank you, Laura, and it's been a real pleasure

44 00:02:04.830 --> 00:02:06.390 working with you as well.

45 00:02:06.390 --> 00:02:10.080 So today, I'll be presenting on my dissertation research,

46 00:02:10.080 --> 00:02:13.530 which is on addressing bias in causal effects

47 00:02:13.530 --> 00:02:16.380 estimated under misspecified interference sets.

48 00:02:16.380 --> 00:02:18.540 And we've applied our methods through the analysis

49 00:02:18.540 --> 00:02:20.373 of HIV prevention trials.

50 00:02:22.560 --> 00:02:25.740 So as an introduction, so interference

51 00:02:25.740 --> 00:02:28.860 or spillover is often present in either randomized

52 00:02:28.860 --> 00:02:30.510 or observational studies.

53 00:02:30.510 --> 00:02:32.280 Whereby interference, we mean

54 00:02:32.280 --> 00:02:34.830 that a participant's outcome can be determined

55 00:02:34.830 --> 00:02:36.570 by not only their own exposure

56 00:02:36.570 --> 00:02:38.370 but also the exposure of others.

57 00:02:38.370 --> 00:02:41.550 So a common example is with vaccines.

58 00:02:41.550 --> 00:02:44.070 So say, my disease status is not only affected

59 00:02:44.070 --> 00:02:45.870 by my own vaccination status,

60 00:02:45.870 --> 00:02:48.870 but also the vaccination status of others around me.

61 00:02:48.870 --> 00:02:51.630 And in the context of HIV prevention trials,

62 00:02:51.630 --> 00:02:54.180 it's been found in several network-based studies

63 00:02:54.180 --> 00:02:57.000 that when only some participants of a network

64 00:02:57.000 --> 00:03:00.000 are trained on say HIV knowledge

65 00:03:00.000 --> 00:03:02.520 or safe practices, that the members

66 00:03:02.520 --> 00:03:05.130 who are weren't trained in the network

67 00:03:05.130 --> 00:03:07.440 also demonstrated increased knowledge

68 00:03:07.440 --> 00:03:09.493 and reduced risk behaviors.

69 00:03:09.493 --> 00:03:12.000 And this is known as disability effect.

70 00:03:12.000 --> 00:03:16.770 So causal inference, that is conducted the presence

71 00:03:16.770 --> 00:03:19.230 of interference is often done under assumptions

72 00:03:19.230 --> 00:03:22.380 on the extent and mechanism of interference.

73 00:03:22.380 --> 00:03:25.200 And typically, this will require a specification

74 00:03:25.200 --> 00:03:28.110 of an interference set for each participant.

75 00:03:28.110 --> 00:03:29.370 Whereby interference sets,

76 00:03:29.370 --> 00:03:32.490 we mean that a group of individuals

77 00:03:32.490 --> 00:03:35.490 who can affect the outcome of that participant.

78 00:03:35.490 --> 00:03:37.050 And then to this interference set,

79 00:03:37.050 --> 00:03:40.620 we also typically apply an exposure mapping function

80 00:03:40.620 --> 00:03:42.990 that will take the exposure vector

81 00:03:42.990 --> 00:03:44.940 observed in this interference set

82 00:03:44.940 --> 00:03:46.950 and map it to some scaler quantity.

83 00:03:46.950 --> 00:03:49.203 And we'll see some examples of this later.

84 00:03:50.940 --> 00:03:53.310 So existing literature interference sets

85 00:03:53.310 --> 00:03:55.980 are typically assumed to be correctly specified

86 00:03:55.980 --> 00:03:57.180 so that the exposures

87 00:03:57.180 --> 00:03:59.130 that are mapped from these interference sets

88 00:03:59.130 --> 00:04:01.290 are also correctly measured.

89 00:04:01.290 --> 00:04:04.290 But often, this correctly specifying

90 00:04:04.290 --> 00:04:06.240 an interference set is challenging.

91 00:04:06.240 --> 00:04:09.540 For example, networks can be mismeasured

92 00:04:09.540 --> 00:04:12.480 and when interference sets are misspecified,

93 00:04:12.480 --> 00:04:15.210 we show under various settings that causal effects estimated

94 00:04:15.210 --> 00:04:17.880 by usual purchase are typically biased.

95 00:04:17.880 --> 00:04:20.010 And there have been several publications

96 00:04:20.010 --> 00:04:22.018 that have addressed this issue.

97 00:04:22.018 --> 00:04:24.540 And the majority of these publications aim

98 00:04:24.540 --> 00:04:26.910 to first estimate the true networks

99 00:04:26.910 --> 00:04:28.770 and then using these estimated networks

100 00:04:28.770 --> 00:04:30.750 to estimate the causal effects.

101 00:04:30.750 --> 00:04:32.370 And there have also been methods proposed

102 00:04:32.370 --> 00:04:34.263 for a sensitivity analysis as well.

103 00:04:36.090 --> 00:04:38.730 However, we pursue a different approach where we assume

104 00:04:38.730 --> 00:04:41.130 that we have a validation study

105 00:04:41.130 --> 00:04:44.040 in which the true interference sets are measured alongside

106 00:04:44.040 --> 00:04:46.380 the observed or surrogate ones for a subset

107 00:04:46.380 --> 00:04:47.760 of the study sample.

108 00:04:47.760 --> 00:04:48.810 And this will allow us

109 00:04:48.810 --> 00:04:51.510 to empirically estimate the measurement error process

110 00:04:51.510 --> 00:04:54.540 and use the estimated measurement error parameters

111 00:04:54.540 --> 00:04:57.690 to bias correct causal effects.

112 00:04:57.690 --> 00:05:00.690 So again, this dissertation is a collection of three papers

113 00:05:00.690 --> 00:05:03.480 where we first consider the setting

114 00:05:03.480 --> 00:05:05.940 of an egocentric network randomized trial

115 00:05:05.940 --> 00:05:08.190 where at most one person per network

116 00:05:08.190 --> 00:05:10.200 can receive the intervention.

117 00:05:10.200 --> 00:05:11.610 Then we extend our methods

118 00:05:11.610 --> 00:05:13.410 to consider cluster randomized trials

119 00:05:13.410 --> 00:05:15.240 where multiple participants per cluster

120 00:05:15.240 --> 00:05:17.070 can receive the intervention.

121 00:05:17.070 --> 00:05:19.260 And we also consider general settings

122 00:05:19.260 --> 00:05:21.600 where interference sets can be mismeasured

123 00:05:21.600 --> 00:05:24.993 and the exposure is not necessarily randomized.

124 00:05:27.030 --> 00:05:28.800 So I'll begin with the first paper

125 00:05:28.800 --> 00:05:31.710 on egocentric network randomized trials.

126 00:05:31.710 --> 00:05:34.380 So under this design, we have index participants

127 00:05:34.380 --> 00:05:36.480 who are recruited into this study

128 00:05:36.480 --> 00:05:40.080 and they're each asked to nominate a set of network members,

129 00:05:40.080 --> 00:05:43.710 which can be their drug injection partners or sex partners,

130 00:05:43.710 --> 00:05:46.110 and they form egocentric networks.

131 00:05:46.110 --> 00:05:49.410 And the index participants are the ones in the study

132 00:05:49.410 --> 00:05:52.110 who are randomized to receive their intervention.

133 00:05:52.110 --> 00:05:56.550 And examples of this are typically

134 00:05:56.550 --> 00:05:59.310 be peer education or behavioral-based.

135 00:05:59.310 --> 00:06:02.940 And the index participants are asked to encourage

136 00:06:02.940 --> 00:06:05.103 behavioral change to their network members.

137 00:06:06.930 --> 00:06:08.220 So for some notation,

138 00:06:08.220 --> 00:06:10.860 we have participant ik being the i participant

139 00:06:10.860 --> 00:06:12.810 in the k network.

140 00:06:12.810 --> 00:06:16.350 And we'll let i equal one denote the index participant

141 00:06:16.350 --> 00:06:17.970 in each network and I incur then one

142 00:06:17.970 --> 00:06:20.220 denote the network members.

143 00:06:20.220 --> 00:06:24.360 We'll also define a network neighborhood for participant ik,

144 00:06:24.360 --> 00:06:26.010 which comprises of participants

145 00:06:26.010 --> 00:06:28.293 who share a network link with ik.

146 00:06:30.060 --> 00:06:33.180 And then we also have a true membership matrix

147 00:06:33.180 --> 00:06:36.576 which essentially represents whether a participant

148 00:06:36.576 --> 00:06:39.360 is a network member of a certain index.

149 00:06:39.360 --> 00:06:43.443 And we also have an intervention assignment indicator,

150 00:06:44.520 --> 00:06:46.290 which again the intervention is randomized

151 00:06:46.290 --> 00:06:48.213 and only received by the index member.

152 00:06:51.240 --> 00:06:54.480 So we'll let, throughout this dissertation,

153 00:06:54.480 --> 00:06:57.450 represent an individual exposure to the intervention.

154 00:06:57.450 --> 00:07:01.470 So because in here in NNRT, only index participants

155 00:07:01.470 --> 00:07:05.070 who are randomized to treatment can receive the treatment

156 00:07:05.070 --> 00:07:08.880 and therefore, A is only equal to one for a treated index

157 00:07:08.880 --> 00:07:11.913 and A is equal to zero for everyone else.

158 00:07:12.810 --> 00:07:15.600 And so to define potential outcomes under interference,

159 00:07:15.600 --> 00:07:18.450 we need to make assumptions on the interference structure.

160 00:07:18.450 --> 00:07:22.260 So here, we assume neighborhood interference

161 00:07:22.260 --> 00:07:26.100 with an exposure mapping function, which essentially says

162 00:07:26.100 --> 00:07:28.380 that I case potential outcome is determined

163 00:07:28.380 --> 00:07:30.070 by their own individual exposure

164 00:07:31.127 --> 00:07:33.870 and the exposures of those in I case network neighborhood

165 00:07:33.870 --> 00:07:35.640 and not anyone outside of it,

166 00:07:35.640 --> 00:07:38.230 including participants from other networks

167 00:07:39.180 --> 00:07:42.030 and out of study individuals.

168 00:07:42.030 --> 00:07:45.690 So we further apply an exposure mapping function

169 00:07:45.690 --> 00:07:47.880 to this network neighborhood

170 00:07:47.880 --> 00:07:51.090 and in this paper, we consider an exposure mapping function

171 00:07:51.090 --> 00:07:54.090 defined by the number of treated neighbors.

172 00:07:54.090 --> 00:07:57.630 So under this assumption,

173 00:07:57.630 --> 00:08:01.170 ik's potential outcome is given by y indexed by A and G,

174 00:08:01.170 --> 00:08:02.457 which is their individual exposure

175 00:08:02.457 --> 00:08:06.000 and the spillover exposure given by the number

176 00:08:06.000 --> 00:08:08.703 of treated neighbors in their neighbor neighborhood.

177 00:08:11.010 --> 00:08:14.340 So this figure is a representation

178 00:08:14.340 --> 00:08:16.290 of two networks where in order

179 00:08:16.290 --> 00:08:19.050 to define the spillover exposure,

180 00:08:19.050 --> 00:08:21.060 we further make the assumption

181 00:08:21.060 --> 00:08:23.460 that the networks are not overlapping.

182 00:08:23.460 --> 00:08:25.470 And so this means that index participants

183 00:08:25.470 --> 00:08:27.720 cannot be connected amongst themselves

184 00:08:27.720 --> 00:08:29.610 and network members can only be connected

185 00:08:29.610 --> 00:08:31.503 to one index participant.

186 00:08:32.490 --> 00:08:35.550 And by making this assumption we can obtain G

187 00:08:35.550 --> 00:08:38.790 by multiplying M and R, which is the membership matrix

188 00:08:38.790 --> 00:08:40.560 and intervention assignment.

189 00:08:40.560 --> 00:08:42.000 And so the spillover exposure

190 00:08:42.000 --> 00:08:43.770 for each participant is only determined

191 00:08:43.770 --> 00:08:47.160 by whether they are connected to a treated index member,

192 00:08:47.160 --> 00:08:49.023 which is shown in this figure.

193 00:08:52.890 --> 00:08:55.590 So the causal estimate of interest in this paper,

194 00:08:55.590 --> 00:08:58.320 the average spillover effect which is the impact

195 00:08:58.320 --> 00:09:01.050 of the intervention on the network members.

196 00:09:01.050 --> 00:09:03.180 And we here, we define the spillover effect

197 00:09:03.180 --> 00:09:05.880 as a risk difference and as a risk ratio.

198 00:09:05.880 --> 00:09:08.190 And under assumptions of positivity

199 00:09:08.190 --> 00:09:12.630 and unconfoundedness which is guaranteed in the ENRG design

200 00:09:12.630 --> 00:09:14.733 under perfect treatment compliance,

201 00:09:20.190 --> 00:09:21.750 we can identify the spillover effects

202 00:09:21.750 --> 00:09:24.780 using observed outcomes and estimate them

203 00:09:24.780 --> 00:09:26.380 using observed outcomes as well.

204 00:09:29.400 --> 00:09:32.670 So under the unconfoundedness assumption,

205 00:09:32.670 --> 00:09:36.390 if we had data on the true exposures, we would estimate

206 00:09:36.390 --> 00:09:40.097 the spillover effect using sample average estimators

207 00:09:40.097 --> 00:09:42.333 using the two by two table at the top.

208 00:09:43.230 --> 00:09:45.690 The issue with this is that in ENRTs, the networks

209 00:09:45.690 --> 00:09:47.490 that are observed, which are the ones

210 00:09:47.490 --> 00:09:49.350 that are collected at study baseline,

211 00:09:49.350 --> 00:09:51.690 they may not represent the true connections that take place

212 00:09:51.690 --> 00:09:53.580 during the study period.

213 00:09:53.580 --> 00:09:56.340 So for example, a network member can fall out of touch

214 00:09:56.340 --> 00:09:58.710 with the index participant that they enrolled with

215 00:09:58.710 --> 00:10:03.540 and they can also befriend another index of another network.

216 00:10:03.540 --> 00:10:06.150 And so under these observed networks,

217 00:10:06.150 --> 00:10:10.020 there's spillover exposure may also be misclassified.

218 00:10:10.020 --> 00:10:13.560 And using these misclassified spillover exposures,

219 00:10:13.560 --> 00:10:15.930 we will instead estimate the spillover effect

220 00:10:15.930 --> 00:10:20.220 using the two by two table at the bottom of this slide.

221 00:10:20.220 --> 00:10:22.620 And we show that the estimated spillover effects

222 00:10:22.620 --> 00:10:24.873 under this table would be biased.

223 00:10:27.030 --> 00:10:29.430 And here's a representation of the types

224 00:10:29.430 --> 00:10:33.720 of network misclassification that can occur in ENRT.

225 00:10:33.720 --> 00:10:35.517 So here, the black links represent

226 00:10:35.517 --> 00:10:38.640 the correctly measured network ties.

227 00:10:38.640 --> 00:10:42.420 The blue links represent network links that are observed

228 00:10:42.420 --> 00:10:44.130 but are in fact not true.

229 00:10:44.130 --> 00:10:47.130 And the red links represent the ones that are not observed

230 00:10:47.130 --> 00:10:50.160 but in fact occur during the study period.

231 00:10:50.160 --> 00:10:54.600 And because of these types of misclassification

232 00:10:54.600 --> 00:10:56.490 person's truth, spillover exposure

233 00:10:56.490 --> 00:10:58.540 can be different from their observed one.

234 00:11:02.264 --> 00:11:04.410 In this paper, we further assume

235 00:11:04.410 --> 00:11:06.690 non-differential misclassification,

236 00:11:06.690 --> 00:11:08.640 which is that the misclassification process

237 00:11:08.640 --> 00:11:10.683 doesn't depend on potential outcomes.

238 00:11:11.550 --> 00:11:15.030 So under this assumption we derive an expression

239 00:11:15.030 --> 00:11:17.460 for the bias using four parameters,

240 00:11:17.460 --> 00:11:20.100 which are the baseline Malcolm rate,

241 00:11:20.100 --> 00:11:24.510 the true spillover risk ratio, PM, which is the probability

242 00:11:24.510 --> 00:11:27.870 of being classified into the correct network as well as PR,

243 00:11:27.870 --> 00:11:30.513 which is the intervention allocation probability.

244 00:11:31.500 --> 00:11:33.930 And using these expressions we can show

245 00:11:33.930 --> 00:11:36.240 that there's no bias when PM is one,

246 00:11:36.240 --> 00:11:38.460 which is when everyone is correctly classified

247 00:11:38.460 --> 00:11:40.173 due to the correct network.

248 00:11:41.220 --> 00:11:44.220 We can also show that the bias is always towards the null

249 00:11:44.220 --> 00:11:47.010 under the non differential misclassification assumption.

250 00:11:47.010 --> 00:11:49.680 So the ASP would always be underestimated

251 00:11:49.680 --> 00:11:50.760 under this assumption

252 00:11:50.760 --> 00:11:53.193 if spillover exposures were misclassified.

253 00:11:56.610 --> 00:11:58.590 So in order to correct for this bias,

254 00:11:58.590 --> 00:12:01.380 we use a validation study.

255 00:12:01.380 --> 00:12:03.210 So again, this is where the true network

256 00:12:03.210 --> 00:12:05.100 or spillover exposure is measured

257 00:12:05.100 --> 00:12:06.690 alongside the mismeasured ones

258 00:12:06.690 --> 00:12:09.030 for a subsample of the main study.

259 00:12:09.030 --> 00:12:12.180 And then in this paper, we estimate the sensitivity

260 00:12:12.180 --> 00:12:14.880 and specificity of spillover exposure classification

261 00:12:14.880 --> 00:12:18.090 among network members and we assume that the parameters

262 00:12:18.090 --> 00:12:20.040 that are estimated in the validation study

263 00:12:20.040 --> 00:12:22.353 is generalizable to the main study.

264 00:12:24.180 --> 00:12:25.920 We can show that the sensitivity

265 00:12:25.920 --> 00:12:30.800 and specificity can be expressed as functions of PM and PR

266 00:12:31.830 --> 00:12:35.370 where the intuition is that if a participant

267 00:12:35.370 --> 00:12:38.040 or if a network member is observed to be connected

268 00:12:38.040 --> 00:12:42.600 to a treated index, given that there really are connected

269 00:12:42.600 --> 00:12:43.947 to a treated index,

270 00:12:43.947 --> 00:12:47.190 this can be because they are correctly classified

271 00:12:47.190 --> 00:12:49.110 or it could be because they were misclassified

272 00:12:49.110 --> 00:12:50.130 but still connected

273 00:12:50.130 --> 00:12:52.473 to a treated index just from another network.

274 00:12:53.610 --> 00:12:55.290 We can also estimate the sensitivity

275 00:12:55.290 --> 00:12:58.260 and specificity using the two by two table.

276 00:12:58.260 --> 00:13:01.680 Here, when we assume that the misclassification process

277 00:13:01.680 --> 00:13:03.657 doesn't depend on covariate.

278 00:13:07.200 --> 00:13:10.470 So we propose three estimators in this paper.

279 00:13:10.470 --> 00:13:13.200 The first is called the matrix method estimator.

280 00:13:13.200 --> 00:13:17.460 And here, this estimator takes the estimated sensitivity

281 00:13:17.460 --> 00:13:19.890 and specificity from the validation study

282 00:13:19.890 --> 00:13:22.380 to bias correct accounts observed

283 00:13:22.380 --> 00:13:24.720 from the two by two table in the main study.

284 00:13:24.720 --> 00:13:27.480 And this would be the form of the bias corrected estimators

285 00:13:27.480 --> 00:13:29.310 for this spillover effect.

286 00:13:29.310 --> 00:13:30.990 And we can obtain its variance

287 00:13:30.990 --> 00:13:33.330 by the multivariate delta method.

288 00:13:33.330 --> 00:13:36.120 And if we believe that there is clustering in the study,

289 00:13:36.120 --> 00:13:40.290 we can also adjust for this by a design effect inflation

290 00:13:40.290 --> 00:13:42.630 or we can perform network bootstrapping

291 00:13:42.630 --> 00:13:44.480 where we re-sample networks as whole.

292 00:13:46.110 --> 00:13:48.960 And we know that to use this method

293 00:13:48.960 --> 00:13:51.750 and for this method to perform well,

294 00:13:51.750 --> 00:13:54.510 there needs to be constraints on the value of sensitivity

295 00:13:54.510 --> 00:13:57.510 and specificity for the estimator to be stable

296 00:13:57.510 --> 00:14:00.603 and to avoid estimating negative cell counts.

297 00:14:02.130 --> 00:14:05.220 So when these constraints are not met,

298 00:14:05.220 --> 00:14:08.650 we can instead consider an inverse matrix method estimator

299 00:14:09.720 --> 00:14:13.320 which corrects the cell counts in the two at two table

300 00:14:13.320 --> 00:14:14.940 in the main study using the positive

301 00:14:14.940 --> 00:14:16.320 and negative predictive values

302 00:14:16.320 --> 00:14:19.020 instead of the sensitivity and specificity.

303 00:14:19.020 --> 00:14:22.380 And this method uses the PPV and MPV

304 00:14:22.380 --> 00:14:25.860 estimated separately for those with and without the outcome.

305 00:14:25.860 --> 00:14:28.320 And therefore the matrix method estimated

306 00:14:28.320 --> 00:14:31.230 may be more efficient relative to this estimator

307 00:14:31.230 --> 00:14:32.190 if the outcome is rare

308 00:14:32.190 --> 00:14:33.850 and the validation study is small

309 00:14:36.570 --> 00:14:38.940 and the last estimator we considered

310 00:14:38.940 --> 00:14:41.880 was a likelihood-based estimator

311 00:14:41.880 --> 00:14:43.290 because while the matrix

312 00:14:43.290 --> 00:14:46.680 and inverse matrix estimators are easily implemented,

313 00:14:46.680 --> 00:14:49.470 there's no clear way of directly incorporating the effect

314 00:14:49.470 --> 00:14:51.570 of clustering into the entrance.

315 00:14:51.570 --> 00:14:55.140 And therefore, we can specify an outcome model

316 00:14:55.140 --> 00:14:57.100 including a network random effect

317 00:14:59.734 --> 00:15:02.100 to account for clustering by networks

318 00:15:02.100 --> 00:15:06.000 and using the likelihood specified here,

319 00:15:06.000 --> 00:15:08.593 we can obtain the MLE of the ASP

320 00:15:09.930 --> 00:15:12.930 and its variance by the inverse

321 00:15:12.930 --> 00:15:14.830 of the observation information matrix.

322 00:15:18.870 --> 00:15:22.936 Right, so I'll go over our application

323 00:15:22.936 --> 00:15:25.770 of our methods using the HPTN 037 study,

324 00:15:25.770 --> 00:15:29.640 which was an ENRT that was conducted in Philadelphia

325 00:15:29.640 --> 00:15:31.380 and Chiang Mai Thailand

326 00:15:31.380 --> 00:15:34.170 where in the study indexes were randomized

327 00:15:34.170 --> 00:15:36.240 to receive an intervention that consisted

328 00:15:36.240 --> 00:15:39.310 of a peer education training where they were encouraged

329 00:15:41.670 --> 00:15:43.950 to disseminate HIV knowledge

330 00:15:43.950 --> 00:15:46.740 and injection sexual risk reduction behaviors

331 00:15:46.740 --> 00:15:48.183 with their network members.

332 00:15:49.160 --> 00:15:53.130 In the study, we were interested in looking at the effect

333 00:15:53.130 --> 00:15:57.020 of the intervention on any self-reported HIV risk behaviors

334 00:15:57.020 --> 00:15:59.643 at one year after study enrollment.

335 00:16:00.990 --> 00:16:02.820 Here, we define G star,

336 00:16:02.820 --> 00:16:05.010 which is they observed spillover exposure

337 00:16:05.010 --> 00:16:07.110 based on the intervention assigned to the networks

338 00:16:07.110 --> 00:16:09.480 and receive by their index members.

339 00:16:09.480 --> 00:16:12.870 And we define G the true spillover exposure

340 00:16:16.500 --> 00:16:18.810 based on an exposure contamination survey

341 00:16:18.810 --> 00:16:21.990 that was taken at six months post baseline.

342 00:16:21.990 --> 00:16:24.120 So in the survey, participants were asked

343 00:16:24.120 --> 00:16:28.230 to recall five specific terminologies associated

344 00:16:28.230 --> 00:16:30.780 with the intervention training.

345 00:16:30.780 --> 00:16:33.150 So we suppose that if a network member were able

346 00:16:33.150 --> 00:16:36.090 to recall any of these five terms, that they were exposed

347 00:16:36.090 --> 00:16:38.430 to the intervention through a treated index

348 00:16:38.430 --> 00:16:39.690 and if they weren't able to recall

349 00:16:39.690 --> 00:16:42.690 any of the terms, then they weren't exposed.

350 00:16:42.690 --> 00:16:44.940 And because there was a possibility

351 00:16:44.940 --> 00:16:48.300 that network members may be exposed to the training

352 00:16:48.300 --> 00:16:50.580 but just didn't remember any of the terms

353 00:16:50.580 --> 00:16:53.340 or that there may be network members

354 00:16:53.340 --> 00:16:55.590 who were not exposed to the training

355 00:16:55.590 --> 00:16:57.360 but just said that they remember terms

356 00:16:57.360 --> 00:16:59.460 because of social desirability,

357 00:16:59.460 --> 00:17:01.170 we only included network members

358 00:17:01.170 --> 00:17:04.140 who recalled the positive control term which was exposed

359 00:17:04.140 --> 00:17:07.500 to everybody regardless of their randomization arm

360 00:17:07.500 --> 00:17:09.480 and none of the negative control terms,

361 00:17:09.480 --> 00:17:13.830 which none of the participants were supposed to know

362 00:17:13.830 --> 00:17:14.880 that only these participants

363 00:17:14.880 --> 00:17:16.320 were included in the validation study

364 00:17:16.320 --> 00:17:18.540 so we can more accurately estimate

365 00:17:18.540 --> 00:17:20.313 the sensitivity and specificity.

366 00:17:22.860 --> 00:17:27.000 So here are the effects of the intervention

367 00:17:27.000 --> 00:17:28.950 or the spillover effects of the intervention

368 00:17:28.950 --> 00:17:33.950 on risk behaviors were from the validation study we see

369 00:17:34.050 --> 00:17:36.420 that there was indeed some degree

370 00:17:36.420 --> 00:17:40.620 of network misclassification where the sensitivity was 60%

371 00:17:40.620 --> 00:17:43.800 and specificity was 79%.

372 00:17:43.800 --> 00:17:47.100 So then intent-to-treat estimator uses G star,

373 00:17:47.100 --> 00:17:49.950 which is the intervention assigned to the networks.

374 00:17:49.950 --> 00:17:51.270 And we do see already

375 00:17:51.270 --> 00:17:53.940 that there was significant spillover effect

376 00:17:53.940 --> 00:17:57.030 of the intervention on reducing risk behaviors.

377 00:17:57.030 --> 00:17:59.010 And this effect was amplified

378 00:17:59.010 --> 00:18:01.653 after applying our bias correction method.

379 00:18:03.480 --> 00:18:07.170 So we applied the matrix method estimator as well as the MLE

380 00:18:07.170 --> 00:18:09.760 and the inverse matrix method

381 00:18:10.939 --> 00:18:13.590 was not an ideal choice in this study

382 00:18:13.590 --> 00:18:15.420 because of our small validation study

383 00:18:15.420 --> 00:18:17.730 and the number of participants

384 00:18:17.730 --> 00:18:19.580 who had the outcome within the study.

385 00:18:20.730 --> 00:18:25.590 Here, we also compared several standard errors were first,

386 00:18:25.590 --> 00:18:27.030 we consider standard standards obtained

387 00:18:27.030 --> 00:18:30.090 from the delta method and those inflated

388 00:18:30.090 --> 00:18:31.770 by the design effect.

389 00:18:31.770 --> 00:18:33.990 And we see that the confidence intervals here

390 00:18:33.990 --> 00:18:36.480 were pretty wide, which is due

391 00:18:36.480 --> 00:18:39.510 to the small validation study sample size.

392 00:18:39.510 --> 00:18:41.910 However, when we consider network bootstrapping

393 00:18:41.910 --> 00:18:43.770 or the likelihood base method,

394 00:18:43.770 --> 00:18:46.770 we see that the confidence interval significantly narrowed

395 00:18:46.770 --> 00:18:49.983 and we were able to see a significant spillover effect.

396 00:18:53.820 --> 00:18:55.800 So as a summary of this first paper,

397 00:18:55.800 --> 00:18:58.470 we proposed several bias correction estimators

398 00:18:58.470 --> 00:19:00.480 for this spillover effect

399 00:19:00.480 --> 00:19:04.290 to address network misclassification and NRTs.

400 00:19:04.290 --> 00:19:07.050 So our methods here assume that both the exposure

401 00:19:07.050 --> 00:19:08.730 and outcome are binary measures

402 00:19:08.730 --> 00:19:12.480 and we did not consider covariate adjustment

403 00:19:12.480 --> 00:19:15.060 because the intervention is randomized.

404 00:19:15.060 --> 00:19:17.610 And so as a segue to the second paper,

405 00:19:17.610 --> 00:19:19.620 we will be developing methods

406 00:19:19.620 --> 00:19:22.710 for non-binary exposures outcomes

407 00:19:22.710 --> 00:19:25.620 as well as allowing for covariate adjustment.

408 00:19:25.620 --> 00:19:27.390 And we develop these methods in the setting

409 00:19:27.390 --> 00:19:29.920 of cluster randomized trials.

410 00:19:34.766 --> 00:19:37.910 So causal inference in cluster randomized trials

411 00:19:37.910 --> 00:19:41.100 in CRTs often rely on the assumption

412 00:19:41.100 --> 00:19:42.510 of partial interference,

413 00:19:42.510 --> 00:19:45.220 which is that participants are separated

414 00:19:46.115 --> 00:19:47.460 into non-overlapping clusters

415 00:19:47.460 --> 00:19:49.890 and interference is assumed

416 00:19:49.890 --> 00:19:52.050 to be only contained within these clusters

417 00:19:52.050 --> 00:19:53.910 and not across clusters.

418 00:19:53.910 --> 00:19:56.100 And this assumption is typically made

419 00:19:56.100 --> 00:20:00.543 because there is an absence of social network data and CRTs.

420 00:20:02.220 --> 00:20:04.510 So interference sets define other departure

421 00:20:05.613 --> 00:20:06.870 interference assumptions are usually given

422 00:20:06.870 --> 00:20:09.780 by the randomization clusters in the trial.

423 00:20:09.780 --> 00:20:12.250 So this can be villages or communities

424 00:20:15.566 --> 00:20:17.130 and the interference says they're given

425 00:20:17.130 --> 00:20:20.280 by the randomization clusters can be measured with there

426 00:20:20.280 --> 00:20:23.970 because they might be a lot larger than the true networks

427 00:20:23.970 --> 00:20:26.370 if they were considered to be whole communities.

428 00:20:26.370 --> 00:20:28.020 And also interactions can exist

429 00:20:28.020 --> 00:20:29.883 across these communities as well.

430 00:20:30.720 --> 00:20:34.590 So this figure was taken from a file genetic analysis

431 00:20:34.590 --> 00:20:39.590 from BCPP where BCPP was in HIV prevention CRT

432 00:20:39.900 --> 00:20:43.620 that was conducted in 30 communities in Botswana.

433 00:20:43.620 --> 00:20:47.640 And so the randomization clusters were communities

434 00:20:47.640 --> 00:20:51.430 but from the phylogenetic analysis where they sequenced

435 00:20:52.380 --> 00:20:56.590 HIV genes viral sequences

436 00:20:57.840 --> 00:21:00.330 that they saw that the viral transmission chains obtained

437 00:21:00.330 --> 00:21:03.720 from the sequences, the majority of them

438 00:21:03.720 --> 00:21:06.300 actually crossed two or more communities,

439 00:21:06.300 --> 00:21:07.320 which was an indication

440 00:21:07.320 --> 00:21:09.840 of high-end cluster mixing in this study.

441 00:21:09.840 --> 00:21:11.490 And interference says that are defined

442 00:21:11.490 --> 00:21:14.883 just by communities would be misspecified in this case.

443 00:21:18.180 --> 00:21:20.387 So here we again have participant ik

444 00:21:20.387 --> 00:21:24.300 as the i participant indicate cluster.

445 00:21:24.300 --> 00:21:28.260 We have script one and to denote the study sample.

446 00:21:28.260 --> 00:21:33.260 In this study, we first consider a two-stage CRT

447 00:21:33.330 --> 00:21:35.640 where clusters are first randomized

448 00:21:35.640 --> 00:21:38.430 to an intervention allocation strategy.

449 00:21:38.430 --> 00:21:41.580 Here, we consider strategies alpha one and alpha two

450 00:21:41.580 --> 00:21:44.040 and alpha one and alpha two are probabilities.

451 00:21:44.040 --> 00:21:46.020 And under a balanced design,

452 00:21:46.020 --> 00:21:48.450 half of the clusters would be assigned to alpha one

453 00:21:48.450 --> 00:21:51.270 and half would be assigned to alpha alpha two.

454 00:21:51.270 --> 00:21:55.140 Then after the first stage randomization,

455 00:21:55.140 --> 00:21:56.670 participants within these clusters

456 00:21:56.670 --> 00:21:59.280 would be randomized to receive the intervention

457 00:21:59.280 --> 00:22:01.050 with the probability equal to the one

458 00:22:01.050 --> 00:22:02.800 that was assigned to their cluster.

459 00:22:03.900 --> 00:22:07.230 And we'll extend our methods to consider general CRTs,

460 00:22:07.230 --> 00:22:09.840 which can be considered as a special case

461 00:22:09.840 --> 00:22:12.190 of a two-stage design where alpha one

462 00:22:12.190 --> 00:22:14.880 and alpha two are one and zero, which means

463 00:22:14.880 --> 00:22:18.060 that clusters are randomized to intervention or control

464 00:22:18.060 --> 00:22:20.400 and there isn't a second stage randomization

465 00:22:20.400 --> 00:22:21.813 at the participant level.

466 00:22:25.110 --> 00:22:29.400 So we again have to denote the individual exposure.

467 00:22:29.400 --> 00:22:31.950 And to define potential outcomes in the setting,

468 00:22:31.950 --> 00:22:35.340 we first define a subset of the study sample

469 00:22:35.340 --> 00:22:39.153 for participant ik and we denote this by script I.

470 00:22:40.110 --> 00:22:42.850 So here, we make the partial interference assumption

471 00:22:45.497 --> 00:22:48.030 where ik's potential outcome is influenced

472 00:22:48.030 --> 00:22:51.690 by their own exposure as well as the exposures

473 00:22:51.690 --> 00:22:54.600 of the participants within this subset

474 00:22:54.600 --> 00:22:57.060 and not anyone outside of this subset.

475 00:22:57.060 --> 00:23:00.041 So because only the exposures of the participants

476 00:23:00.041 --> 00:23:03.910 will affect the outcome of ik, we call this subset

477 00:23:04.864 --> 00:23:06.164 for ik's interference set.

478 00:23:07.514 --> 00:23:10.710 And we can further apply an exposure mapping function

479 00:23:10.710 --> 00:23:13.000 to this interference set

480 00:23:13.920 --> 00:23:16.980 to obtain a scaler quantity of a spillover exposure.

481 00:23:16.980 --> 00:23:20.463 Here, we consider stratify interference,

482 00:23:21.600 --> 00:23:24.540 which essentially assumes that spillover occurs

483 00:23:24.540 --> 00:23:26.880 through the proportion of treated participants

484 00:23:26.880 --> 00:23:30.390 in the interference set regardless of who they are.

485 00:23:30.390 --> 00:23:33.540 So the spillover exposure would be given by disproportion

486 00:23:33.540 --> 00:23:34.650 and as in the first paper,

487 00:23:34.650 --> 00:23:37.893 we can index potential outcomes by A and G.

488 00:23:41.490 --> 00:23:44.700 Here, we consider four causal effects

489 00:23:44.700 --> 00:23:47.730 which the individual effect, spillover effect,

490 00:23:47.730 --> 00:23:49.443 total effect and overall effect.

491 00:23:50.280 --> 00:23:53.227 The individual effect is the effect

492 00:23:53.227 --> 00:23:57.240 of the individual exposure under a fixed spillover exposure.

493 00:23:57.240 --> 00:24:00.900 And on the other hand, the spillover effect is the effect

494 00:24:00.900 --> 00:24:05.100 of the spillover exposure under a fixed individual exposure.

495 00:24:05.100 --> 00:24:07.320 And then the total effect is the effect

496 00:24:07.320 --> 00:24:10.170 of having both an individual exposure to the intervention

497 00:24:10.170 --> 00:24:12.390 and some degree of spillover exposure

498 00:24:12.390 --> 00:24:15.060 versus neither type of exposure.

499 00:24:15.060 --> 00:24:17.280 And then the overall effect compares the effect

500 00:24:17.280 --> 00:24:19.390 of being assigned to a cluster randomized

501 00:24:20.777 --> 00:24:23.377 to treatment allocation strategy alpha versus alpha.

502 00:24:27.150 --> 00:24:31.650 So these causal effects can again be identified

503 00:24:31.650 --> 00:24:34.663 under the assumption the identifying assumptions

504 00:24:34.663 --> 00:24:36.900 we made in the paper earlier

505 00:24:36.900 --> 00:24:38.130 or as in the first paper,

506 00:24:38.130 --> 00:24:41.520 which were the unconfounded assumption

507 00:24:41.520 --> 00:24:44.430 which would hold under a two-stage design

508 00:24:44.430 --> 00:24:47.190 given perfect treatment compliance.

509 00:24:47.190 --> 00:24:48.600 And we estimate these effects

510 00:24:48.600 --> 00:24:52.350 using a regression-based estimation approach,

511 00:24:52.350 --> 00:24:53.183 which is consistent

512 00:24:53.183 --> 00:24:55.860 and efficient under a correctly specified model

513 00:24:55.860 --> 00:24:57.460 for the potential outcome.

514 00:24:58.560 --> 00:25:03.560 So here, we consider an outcome model in this form

515 00:25:05.962 --> 00:25:08.747 where we include a cluster random effect to account

516 00:25:10.050 --> 00:25:12.147 for the effect of clustering and the inference

517 00:25:12.147 --> 00:25:14.940 and we also have an interaction between A and G

518 00:25:14.940 --> 00:25:18.150 so that we can allow the individual effect to vary

519 00:25:18.150 --> 00:25:21.753 with G and the spillover effect to vary with A.

520 00:25:23.070 --> 00:25:27.270 So once we have the estimated coefficients from this model,

521 00:25:27.270 --> 00:25:29.760 we can estimate causal effects

522 00:25:29.760 --> 00:25:31.953 using these estimated coefficients.

523 00:25:35.820 --> 00:25:37.890 So again, in CRTs,

524 00:25:37.890 --> 00:25:40.800 because we don't have data on social connections

525 00:25:40.800 --> 00:25:43.200 and when we consider interference to be given

526 00:25:43.200 --> 00:25:46.950 by randomization clusters, they can be measured with error.

527 00:25:46.950 --> 00:25:48.750 And as a consequence, this spillover exposure

528 00:25:48.750 --> 00:25:51.450 can also be measured with error.

529 00:25:51.450 --> 00:25:53.490 So we have shown in this paper

530 00:25:53.490 --> 00:25:57.330 that when the outcome model is fit with G star instead of G,

531 00:25:57.330 --> 00:25:59.610 the estimated model coefficient will be biased

532 00:25:59.610 --> 00:26:01.440 and the causal effects estimated

533 00:26:01.440 --> 00:26:04.503 with these bias coefficients were therefore also be biased.

534 00:26:07.707 --> 00:26:09.810 And to correct for the bias

535 00:26:09.810 --> 00:26:11.786 in these regression coefficients,

536 00:26:11.786 --> 00:26:15.210 we apply a regression calibration approach

537 00:26:15.210 --> 00:26:17.310 which is developed under the assumption

538 00:26:17.310 --> 00:26:19.920 that the measurement error is additive

539 00:26:19.920 --> 00:26:22.020 and also the non differential measurement error

540 00:26:22.020 --> 00:26:23.433 as in the previous paper.

541 00:26:24.810 --> 00:26:27.180 So to apply this method,

542 00:26:27.180 --> 00:26:29.520 we will first regress the outcome

543 00:26:29.520 --> 00:26:31.200 on the mismeasured exposure

544 00:26:31.200 --> 00:26:33.660 in the main study as we would

545 00:26:33.660 --> 00:26:36.240 under the intent-to-treat analysis.

546 00:26:36.240 --> 00:26:37.425 In the validation study

547 00:26:37.425 --> 00:26:40.830 because we assumed that the measurement error is additive,

548 00:26:40.830 --> 00:26:45.300 we fit a linear measurement error model of the true exposure

549 00:26:45.300 --> 00:26:49.020 given the mismeasured spillover exposure.

550 00:26:49.020 --> 00:26:51.180 And then we can obtain bias corrected regression

551 00:26:51.180 --> 00:26:53.479 coefficients using the coefficients

552 00:26:53.479 --> 00:26:56.070 obtained from these two models.

553 00:26:56.070 --> 00:26:58.800 And we can obtain the variance

554 00:26:58.800 --> 00:27:01.713 of these corrected coefficients using the delta.

555 00:27:05.730 --> 00:27:09.030 We can also extend this approach to account

556 00:27:09.030 --> 00:27:10.680 for covariate adjustment

557 00:27:10.680 --> 00:27:12.630 and there may be several reasons why we need

558 00:27:12.630 --> 00:27:14.550 to adjust for covariates.

559 00:27:14.550 --> 00:27:18.390 First, if we step out of the two stage CRT setting

560 00:27:18.390 --> 00:27:21.990 and we consider a general CRT where intervention is work,

561 00:27:21.990 --> 00:27:24.780 clusters are assigned to either intervention or control.

562 00:27:24.780 --> 00:27:28.590 And a lot of public health studies that the interventions

563 00:27:28.590 --> 00:27:31.170 that are given to these clusters may be prone

564 00:27:31.170 --> 00:27:33.090 to non-compliance.

565 00:27:33.090 --> 00:27:36.360 And intervention uptake will depend

566 00:27:36.360 --> 00:27:38.220 on individual characteristics

567 00:27:38.220 --> 00:27:41.460 that may need to be accounted for.

568 00:27:41.460 --> 00:27:46.460 So in the when there are confounders between the outcome

569 00:27:46.470 --> 00:27:50.430 and the individual exposure, A or G,

570 00:27:50.430 --> 00:27:53.380 we would need to assume conditional unconfoundedness

571 00:27:54.870 --> 00:27:57.000 of the individual exposure exposures.

572 00:27:57.000 --> 00:28:01.650 So when there are covariates or confounders between Y and A,

573 00:28:01.650 --> 00:28:05.250 we would adjust them in the outcome model.

574 00:28:05.250 --> 00:28:09.360 And if they were confounders between Y and G,

575 00:28:09.360 --> 00:28:11.130 we would adjust for them in the outcome model

576 00:28:11.130 --> 00:28:13.833 as well as in the measurement error model.

577 00:28:15.268 --> 00:28:20.268 We might need to also make the nondifferential

578 00:28:20.340 --> 00:28:23.553 measurement error assumption conditional on covariates.

579 00:28:24.870 --> 00:28:27.420 And in this case, because these covariates are related

580 00:28:27.420 --> 00:28:29.820 to the measurement error as well as the outcome.

581 00:28:29.820 --> 00:28:33.960 They would need to be adjusted in both models as well.

582 00:28:33.960 --> 00:28:36.390 And lastly, we may only be able

583 00:28:36.390 --> 00:28:39.647 to generalize the measurement error parameters estimated

584 00:28:39.647 --> 00:28:42.510 in the validation study to the main study

585 00:28:42.510 --> 00:28:43.890 conditional on covariates.

586 00:28:43.890 --> 00:28:46.140 And in this case, we would adjust

587 00:28:46.140 --> 00:28:49.023 for these covariates as well.

588 00:28:49.860 --> 00:28:52.710 But regardless of the types of covariates that are adjusted

589 00:28:52.710 --> 00:28:56.310 for the regression calibration estimators

590 00:28:56.310 --> 00:28:57.807 and variance estimators

591 00:29:01.328 --> 00:29:04.020 for the coefficients that are of interest that are used

592 00:29:04.020 --> 00:29:07.650 to estimate the causal effects, they would not be changed

593 00:29:07.650 --> 00:29:10.323 as in the case without the variates.

594 00:29:13.650 --> 00:29:17.613 We've applied our methods to the BCPP study,

595 00:29:18.570 --> 00:29:22.080 which is, which was a HIV prevention CRT

596 00:29:22.080 --> 00:29:25.560 and 30 Botswana communities that was conducted

597 00:29:25.560 --> 00:29:27.810 between 2013 and 2018.

598 00:29:27.810 --> 00:29:32.810 And this trial was to assess whether an intervention package

599 00:29:33.090 --> 00:29:35.220 will reduce HIV incidents.

600 00:29:35.220 --> 00:29:38.670 So in this trial, 15 communities were randomized

601 00:29:38.670 --> 00:29:41.100 to receive intervention package

602 00:29:41.100 --> 00:29:45.210 that included HIV testing, linkage to care,

603 00:29:45.210 --> 00:29:48.840 and early ART initiation for those who are HIV positive

604 00:29:48.840 --> 00:29:50.430 as well as increased access

605 00:29:50.430 --> 00:29:53.130 to voluntary medical male circumcision.

606 00:29:53.130 --> 00:29:55.350 And the other 15 communities

607 00:29:55.350 --> 00:29:57.250 were randomized to a standard of care.

608 00:29:58.290 --> 00:30:00.450 So in the primary analysis they found

609 00:30:00.450 --> 00:30:02.760 that there were decreased incident rates

610 00:30:02.760 --> 00:30:04.790 and increased file suppression rates

611 00:30:04.790 --> 00:30:06.090 in the intervention communities

612 00:30:06.090 --> 00:30:08.340 compared to control communities.

613 00:30:08.340 --> 00:30:12.030 And in our application, we are accounting for non-compliance

614 00:30:12.030 --> 00:30:15.150 to the components where we analyzed the individual's

615 00:30:15.150 --> 00:30:16.860 spillover, total, and over effects

616 00:30:16.860 --> 00:30:18.210 of the package intervention

617 00:30:18.210 --> 00:30:21.993 that was received on behavioral and clinical outcomes.

618 00:30:23.340 --> 00:30:26.070 And here, we consider the communities

619 00:30:26.070 --> 00:30:28.680 to be the misspecified interference sets

620 00:30:28.680 --> 00:30:31.000 and we determined the true exposures

621 00:30:32.817 --> 00:30:35.610 using phylogenetics data, which we consider

622 00:30:35.610 --> 00:30:37.623 as our validation data.

623 00:30:40.680 --> 00:30:44.640 So the phylogenetic data was obtained from the study shown

624 00:30:44.640 --> 00:30:46.230 in the beginning of this section

625 00:30:46.230 --> 00:30:49.470 where they found viral transmission chains

626 00:30:49.470 --> 00:30:52.620 that crossed multiple clusters.

627 00:30:52.620 --> 00:30:56.340 So here, they approached HIV positive individuals

628 00:30:56.340 --> 00:30:59.340 in the study and obtained blood samples from them

629 00:30:59.340 --> 00:31:02.940 and they were able to sequence their viral genomes

630 00:31:02.940 --> 00:31:05.940 and construct HIV clusters

631 00:31:05.940 --> 00:31:08.910 where a participants group in the same viral cluster

632 00:31:08.910 --> 00:31:12.693 were implied to be from the same viral transmission chain.

633 00:31:13.860 --> 00:31:17.820 So here, each viral cluster they found to be composed

634 00:31:17.820 --> 00:31:19.498 of two to 27 participants who were

635 00:31:19.498 --> 00:31:21.873 from one to 16 communities.

636 00:31:23.190 --> 00:31:26.040 And there were several considerations that we had to make

637 00:31:27.069 --> 00:31:31.800 by using the phylogenetic data as our validation data

638 00:31:31.800 --> 00:31:35.100 because the viral clusters only captured participants

639 00:31:35.100 --> 00:31:37.680 were infected by the same HIV strain.

640 00:31:37.680 --> 00:31:42.090 And would not necessarily represent a participant's

641 00:31:42.090 --> 00:31:43.653 entire true interference set.

642 00:31:45.360 --> 00:31:47.160 So we had to make some assumptions

643 00:31:47.160 --> 00:31:50.070 to obtain the true spillover exposure

644 00:31:50.070 --> 00:31:53.430 using this phylogenetic data where the first,

645 00:31:53.430 --> 00:31:56.010 we consider the connections observed

646 00:31:56.010 --> 00:31:59.127 within the viral cluster were representative

647 00:32:00.564 --> 00:32:02.567 of the participants

648 00:32:02.567 --> 00:32:06.693 who were HIV positive in ik's true interference set.

649 00:32:08.190 --> 00:32:11.600 And then we also assume transportability

650 00:32:11.600 --> 00:32:15.180 of the measurement error process where we assume

651 00:32:15.180 --> 00:32:16.620 that the inter-cluster interactions

652 00:32:16.620 --> 00:32:18.930 that we observed from the viral clusters

653 00:32:18.930 --> 00:32:22.590 would've been the same among those who were HIV negative.

654 00:32:22.590 --> 00:32:24.600 And lastly, we considered

655 00:32:24.600 --> 00:32:26.460 that because those who are HIV positive

656 00:32:26.460 --> 00:32:28.200 might not have the same characteristics

657 00:32:28.200 --> 00:32:31.800 for intervention uptake as those who are HIV negative.

658 00:32:31.800 --> 00:32:35.400 We derived the true spillover exposure

659 00:32:35.400 --> 00:32:39.123 based on a weighted average of cluster intervention uptake.

660 00:32:40.170 --> 00:32:44.490 So for example, if there were five participants observed

661 00:32:44.490 --> 00:32:45.870 in the viral cluster

662 00:32:45.870 --> 00:32:48.870 from two different randomization communities,

663 00:32:48.870 --> 00:32:51.540 then we take the intervention uptake

664 00:32:51.540 --> 00:32:53.220 of these two communities

665 00:32:53.220 --> 00:32:55.800 and waited by the portion of participants

666 00:32:55.800 --> 00:32:59.583 from each community that were observed in the viral cluster.

667 00:33:03.510 --> 00:33:08.510 And so here are some details on the intervention components

668 00:33:09.300 --> 00:33:10.353 for the study.

669 00:33:11.280 --> 00:33:12.690 I don't know, do I have enough to?

670 00:33:12.690 --> 00:33:16.260 Okay, so basically,

671 00:33:16.260 --> 00:33:18.780 there are four components to this intervention package

672 00:33:18.780 --> 00:33:21.930 and these four components were eligible

673 00:33:21.930 --> 00:33:24.480 to different study populations.

674 00:33:24.480 --> 00:33:27.720 So here are the eligibility criteria that we had considered

675 00:33:27.720 --> 00:33:32.610 for our application where for testing, we considered

676 00:33:32.610 --> 00:33:34.950 that they were eligible for testing if they did not have

677 00:33:34.950 --> 00:33:38.490 documented HIV positive status prior to baseline

678 00:33:38.490 --> 00:33:41.460 and participants were eligible for HIV care

679 00:33:41.460 --> 00:33:45.633 and ART initiation if they were HIV positive at baseline.

680 00:33:46.560 --> 00:33:51.060 And for circumcision, we considered someone

681 00:33:51.060 --> 00:33:53.640 to be eligible for this treatment

682 00:33:53.640 --> 00:33:56.190 if they were an HIV negative male at baseline

683 00:33:56.190 --> 00:33:57.813 who had not been circumcised.

684 00:33:59.430 --> 00:34:01.680 And we also considered several definitions

685 00:34:01.680 --> 00:34:04.920 of the individual exposure which were receiving

686 00:34:04.920 --> 00:34:07.980 at least one of these intervention components

687 00:34:07.980 --> 00:34:10.080 or receiving all eligible components

688 00:34:10.080 --> 00:34:11.733 versus some or none of them.

689 00:34:13.345 --> 00:34:16.173 In this paper, we also considered three outcomes.

690 00:34:17.040 --> 00:34:19.890 First was a behavioral outcome that we defined

691 00:34:19.890 --> 00:34:22.440 as a sexual risk behavior score

692 00:34:22.440 --> 00:34:25.530 and these is defined as the number

693 00:34:25.530 --> 00:34:30.377 of self-reported behaviors that they had reported

694 00:34:30.377 --> 00:34:33.453 at their survey interview at one year post baseline.

695 00:34:34.560 --> 00:34:36.990 And then we also looked at two clinical outcomes,

696 00:34:36.990 --> 00:34:39.690 which were viral load at one year post baseline

697 00:34:39.690 --> 00:34:41.913 and HIV incidents by the end of the study.

698 00:34:45.840 --> 00:34:47.640 Before we looked at the effect

699 00:34:47.640 --> 00:34:50.370 of receiving the individual components,

700 00:34:50.370 --> 00:34:54.300 we first assessed the overall effect of being assigned

701 00:34:54.300 --> 00:34:56.820 to an intervention cluster versus control

702 00:34:56.820 --> 00:35:01.233 on these three outcomes where the ITT estimates

703 00:35:01.233 --> 00:35:05.670 were conducted assuming that the interference sets

704 00:35:05.670 --> 00:35:06.783 were communities.

705 00:35:07.890 --> 00:35:12.890 And we see that there was a minimal overall effect

706 00:35:15.060 --> 00:35:20.060 of cluster assignment on decreasing sexual risk behaviors.

707 00:35:20.070 --> 00:35:23.490 But there was significant effect on viral load

708 00:35:23.490 --> 00:35:27.040 and incidents where our findings echoed the ones

709 00:35:27.960 --> 00:35:30.450 from the primary analysis where they found

710 00:35:30.450 --> 00:35:31.620 increased viral suppression

711 00:35:31.620 --> 00:35:34.080 and decreased incidents for clusters assigned

712 00:35:34.080 --> 00:35:37.320 to intervention and versus control.

713 00:35:37.320 --> 00:35:39.000 And after bias correction we see

714 00:35:39.000 --> 00:35:41.403 that these effects are again amplified,

715 00:35:42.750 --> 00:35:45.360 which was expected due to the high levels

716 00:35:45.360 --> 00:35:48.150 of inter-cluster mixing where say,

717 00:35:48.150 --> 00:35:51.540 some preventative measures from intervention communities

718 00:35:51.540 --> 00:35:54.240 may have gone into the control communities

719 00:35:54.240 --> 00:35:56.670 and some incidents observed in intervention communities

720 00:35:56.670 --> 00:35:59.463 may have been attributable to control communities.

721 00:36:03.027 --> 00:36:04.650 And we also looked at the effect

722 00:36:04.650 --> 00:36:06.880 of receiving at least one component

723 00:36:07.808 --> 00:36:12.120 on essential risk behavior score where we see

724 00:36:12.120 --> 00:36:15.030 that after applying our bias correction method,

725 00:36:15.030 --> 00:36:16.620 that there was a significant total

726 00:36:16.620 --> 00:36:21.390 and overall effect of receiving at least one component

727 00:36:21.390 --> 00:36:23.763 on decreased sexual risk behaviors.

728 00:36:26.880 --> 00:36:30.030 And there was also a significant individual effect

729 00:36:30.030 --> 00:36:32.520 of receiving both HIV care

730 00:36:32.520 --> 00:36:37.083 and ART on decreased viral load which was expected.

731 00:36:41.880 --> 00:36:46.789 So here, we proposed methods to bias correct

732 00:36:46.789 --> 00:36:50.130 causal effects estimated underspecified interference

733 00:36:50.130 --> 00:36:54.060 sets in a CRT, although our methods are not restricted

734 00:36:54.060 --> 00:36:57.063 to the setting can be applied to broader settings as well.

735 00:36:59.010 --> 00:37:01.230 And to use our regression calibration method,

736 00:37:01.230 --> 00:37:03.690 we had to assume that both the measurement error

737 00:37:03.690 --> 00:37:06.003 and outcome models were correctly specified.

738 00:37:08.070 --> 00:37:09.540 And we also made some assumptions

739 00:37:09.540 --> 00:37:11.370 on the measurement error structure.

740 00:37:11.370 --> 00:37:16.370 So we proposed for a third paper and IPW-based method

741 00:37:19.290 --> 00:37:21.900 where parametric assumptions on the outcome model

742 00:37:21.900 --> 00:37:23.940 were not required and also we didn't need

743 00:37:23.940 --> 00:37:26.340 to make assumptions on the additive

744 00:37:26.340 --> 00:37:29.373 or non-differential nature of the measurement error process.

745 00:37:33.958 --> 00:37:38.958 Okay, so propensity score based methods are widely used

746 00:37:38.970 --> 00:37:42.590 to estimate intervention effects when characteristics

747 00:37:42.590 --> 00:37:46.560 of the exposed and unexposed participants may be unbalanced,

748 00:37:46.560 --> 00:37:49.860 which may be an observational setting

749 00:37:49.860 --> 00:37:53.493 where the exposure is not randomized.

750 00:37:55.050 --> 00:37:58.440 And in particular, we're focused on an IPW estimator

751 00:37:58.440 --> 00:38:00.060 that has been previously extended

752 00:38:00.060 --> 00:38:03.630 to estimate causal effects in the setting of interference.

753 00:38:03.630 --> 00:38:07.440 And this is typically done assuming the interference sets

754 00:38:07.440 --> 00:38:09.210 are known and true.

755 00:38:09.210 --> 00:38:13.600 And in this paper, we show that when interference sets

756 00:38:14.580 --> 00:38:17.701 are mismeasured and spillover exposures are mismeasured

757 00:38:17.701 --> 00:38:19.250 as a consequence, there is an error

758 00:38:19.250 --> 00:38:21.090 in not only the spillover exposure

759 00:38:21.090 --> 00:38:23.433 but also in the propensity score estimates.

760 00:38:27.660 --> 00:38:29.703 So for notations, here, we have,

761 00:38:30.960 --> 00:38:33.732 we're outside of the network and cluster setting

762 00:38:33.732 --> 00:38:37.260 so we have just i from one to end participants.

763 00:38:37.260 --> 00:38:39.840 Here, the individual exposure status may depend

764 00:38:39.840 --> 00:38:42.510 on observed individual covariates.

765 00:38:42.510 --> 00:38:45.120 And also, here, we assume

766 00:38:45.120 --> 00:38:47.970 the pressure interference assumption as in our second paper.

767 00:38:47.970 --> 00:38:50.880 Although this method doesn't require

768 00:38:50.880 --> 00:38:52.590 the pressure interference assumption.

769 00:38:52.590 --> 00:38:55.350 We can also make the neighborhood interference consumption

770 00:38:55.350 --> 00:38:58.743 if we were working in a setting of social networks.

771 00:39:00.780 --> 00:39:04.530 In this paper, we define a binary spillover exposure,

772 00:39:04.530 --> 00:39:06.630 although our methods can be generalized

773 00:39:06.630 --> 00:39:10.200 to categorical measures of the spillover exposures as well.

774 00:39:10.200 --> 00:39:12.690 And here we consider an extension

775 00:39:12.690 --> 00:39:14.100 of the stratify interference

776 00:39:14.100 --> 00:39:17.970 that we made in the previous paper where G,

777 00:39:17.970 --> 00:39:20.760 we define by one if the proportion

778 00:39:20.760 --> 00:39:23.400 of treated participants in interference set exceeds

779 00:39:23.400 --> 00:39:25.563 a certain pre-specified threshold.

780 00:39:27.150 --> 00:39:30.243 And again, credential outcomes are indexed by A and G.

781 00:39:31.303 --> 00:39:34.680 In this paper, we're interested in the individual spillover

782 00:39:34.680 --> 00:39:36.033 and total effects.

783 00:39:39.300 --> 00:39:41.940 So this is the IPW estimator

784 00:39:41.940 --> 00:39:43.900 for the average potential outcome

785 00:39:45.300 --> 00:39:47.460 where in the denominator, we have

786 00:39:47.460 --> 00:39:51.030 an estimated joint propensity score for the individual

787 00:39:51.030 --> 00:39:52.890 and spillover exposures.

788 00:39:52.890 --> 00:39:55.080 And this can be expressed as the product

789 00:39:55.080 --> 00:39:57.810 of the individual exposure propensity score

790 00:39:57.810 --> 00:40:00.120 and the spillover exposure propensity score.

791 00:40:00.120 --> 00:40:01.350 And these can be estimated

792 00:40:01.350 --> 00:40:04.260 using (indistinct) regression models.

793 00:40:04.260 --> 00:40:07.260 And we can obtain the variance of this estimator

794 00:40:07.260 --> 00:40:10.978 by bootstrap resampling where we can resample

795 00:40:10.978 --> 00:40:14.520 at the individual level or at the cluster level

796 00:40:14.520 --> 00:40:16.710 if we were working in a setting with clusters

797 00:40:16.710 --> 00:40:18.093 as in our second paper.

798 00:40:19.860 --> 00:40:23.130 And this estimator is consistent, if the models

799 00:40:23.130 --> 00:40:25.833 for the propensity scores are correctly specified.

800 00:40:29.100 --> 00:40:34.100 So as in the previous cases when interference specified,

801 00:40:34.980 --> 00:40:38.070 we would observe G star instead of G.

802 00:40:38.070 --> 00:40:42.630 And if we were to use G star in the IPW estimator,

803 00:40:42.630 --> 00:40:44.970 we would get a biased estimate

804 00:40:44.970 --> 00:40:48.840 because the expected value of this estimator is given

805 00:40:48.840 --> 00:40:52.680 by the form shown in the bottom here where we see

806 00:40:52.680 --> 00:40:56.610 that this estimator is only unbiased if the probability

807 00:40:56.610 --> 00:40:59.190 observing the true exposure equal to G,

808 00:40:59.190 --> 00:41:01.830 given that the spillover exposure

809 00:41:01.830 --> 00:41:03.810 is also equal to G is equal to one,

810 00:41:03.810 --> 00:41:06.083 which means that there's no measurement error.

811 00:41:07.386 --> 00:41:12.300 And also from the form of this expectation, we can also see

812 00:41:12.300 --> 00:41:17.040 that the bias can be eliminated if we divide both terms

813 00:41:17.040 --> 00:41:21.000 on the right-hand side by this measurement error probability

814 00:41:21.000 --> 00:41:23.973 and then subtracting away the second term.

815 00:41:25.415 --> 00:41:28.980 Which is the approach that we took.

816 00:41:28.980 --> 00:41:32.730 And this was an approach that was first proposed by brown

817 00:41:32.730 --> 00:41:35.970 and colleagues in the setting without interference.

818 00:41:35.970 --> 00:41:38.220 And here, we extended this estimator

819 00:41:38.220 --> 00:41:39.933 to the setting of interference.

820 00:41:41.640 --> 00:41:46.230 So from this bias corrected IPW estimator, we see

821 00:41:46.230 --> 00:41:49.050 that on the right-hand side in the first term,

822 00:41:49.050 --> 00:41:53.100 we have the IPW estimator that is estimated

823 00:41:53.100 --> 00:41:54.727 in the main study.

824 00:41:54.727 --> 00:41:56.880 We also have an IPW estimator

825 00:41:56.880 --> 00:42:00.180 that is estimated in the validation study alone.

826 00:42:00.180 --> 00:42:03.630 And the measurement error probabilities

827 00:42:03.630 --> 00:42:08.630 are also estimated in the validation study.

828 00:42:08.730 --> 00:42:12.480 And because here, we are estimating potential outcomes

829 00:42:12.480 --> 00:42:16.740 in the validation study, we need to assume generalizability

830 00:42:16.740 --> 00:42:17.850 of the potential outcome

831 00:42:17.850 --> 00:42:20.640 and measurement error process in this study

832 00:42:20.640 --> 00:42:22.530 so that the effects that are estimated

833 00:42:22.530 --> 00:42:24.630 in the validation study alone

834 00:42:24.630 --> 00:42:27.720 would be unbiased for the average effect

835 00:42:27.720 --> 00:42:29.770 that would be observed in the main study.

836 00:42:33.060 --> 00:42:36.810 So using these bias corrected IPW estimators,

837 00:42:36.810 --> 00:42:39.420 we can obtain a bias corrected estimator

838 00:42:39.420 --> 00:42:42.270 for the causal effect which is given as contrast

839 00:42:42.270 --> 00:42:44.880 between potential outcomes estimated

840 00:42:44.880 --> 00:42:47.790 using the bias corrected IPW estimators.

841 00:42:47.790 --> 00:42:50.963 And here, we can write this estimator using

842 00:42:54.510 --> 00:42:57.210 with weights, W here.

843 00:42:57.210 --> 00:42:59.760 Where the weights are meant to minimize the variance

844 00:42:59.760 --> 00:43:03.150 of the bias corrected causal effects

845 00:43:03.150 --> 00:43:06.960 and the weights are given at the bottom here

846 00:43:06.960 --> 00:43:10.620 where the variance of variance terms can also be estimated

847 00:43:10.620 --> 00:43:12.650 using bootstrap resampling.

848 00:43:16.770 --> 00:43:20.670 So while this estimator directly eliminates the bias,

849 00:43:20.670 --> 00:43:21.990 it does require the outcome

850 00:43:21.990 --> 00:43:25.440 to be available in the validation study.

851 00:43:25.440 --> 00:43:28.230 So when this is not available,

852 00:43:28.230 --> 00:43:30.000 we propose an alternative estimator

853 00:43:30.000 --> 00:43:32.880 that does not impose this requirement

854 00:43:32.880 --> 00:43:35.550 where we've extended methods proposed by rule

855 00:43:35.550 --> 00:43:39.540 and colleagues to the setting of interference.

856 00:43:39.540 --> 00:43:42.600 And so this is a regression calibration-based approach

857 00:43:42.600 --> 00:43:46.920 where first, we assume that we have a continuous measure

858 00:43:46.920 --> 00:43:48.453 of the spillover exposure.

859 00:43:49.440 --> 00:43:53.310 And we will predict the true continuous spillover exposures

860 00:43:53.310 --> 00:43:55.200 given the observed ones.

861 00:43:55.200 --> 00:43:57.330 And then under the exposure mapping

862 00:43:57.330 --> 00:44:00.210 that we had specified previously with the threshold,

863 00:44:00.210 --> 00:44:04.983 we would dichotomize this proportion.

864 00:44:06.870 --> 00:44:10.440 And the regression calibration based IPW estimator

865 00:44:10.440 --> 00:44:14.850 would use the predicted binary true exposures

866 00:44:14.850 --> 00:44:16.920 as well as the propensity scores estimated

867 00:44:16.920 --> 00:44:18.930 under these predictive values.

868 00:44:18.930 --> 00:44:22.980 And we've shown that as in the previous paper,

869 00:44:22.980 --> 00:44:25.440 that this estimator is only consistent

870 00:44:25.440 --> 00:44:28.120 if a linear measurement error model fits the data

871 00:44:31.860 --> 00:44:34.740 In this paper, we further consider the case

872 00:44:34.740 --> 00:44:37.110 where we might observe multiple surrogate

873 00:44:37.110 --> 00:44:39.180 interference sets in a study.

874 00:44:39.180 --> 00:44:44.180 And this was motivated by our illustrative example of BCPP

875 00:44:45.240 --> 00:44:48.180 where we may consider a surrogate interference set

876 00:44:48.180 --> 00:44:50.460 defined by a randomization cluster.

877 00:44:50.460 --> 00:44:54.420 And we can also consider a second surrogate interference set

878 00:44:54.420 --> 00:44:56.880 that is defined by household GPS data,

879 00:44:56.880 --> 00:44:59.790 which is available in the study.

880 00:44:59.790 --> 00:45:03.930 So when we have multiple surrogate interference sets,

881 00:45:03.930 --> 00:45:08.100 we propose to first apply our bias corrected estimators,

882 00:45:08.100 --> 00:45:12.780 either the first or the regression calibration-based one

883 00:45:12.780 --> 00:45:15.600 to each surrogate interference set individually.

884 00:45:15.600 --> 00:45:18.090 And then we will combine these individual estimates

885 00:45:18.090 --> 00:45:21.270 using a weighted average estimator to reduce the variance

886 00:45:21.270 --> 00:45:22.713 of the final estimate.

887 00:45:24.690 --> 00:45:29.310 So the weights are given by C in the bottom here

888 00:45:29.310 --> 00:45:32.970 where we would estimate the variance variance matrix

889 00:45:32.970 --> 00:45:37.203 of the individual bias corrected causal effects.

890 00:45:41.640 --> 00:45:45.060 Here, similar to the second paper, we've applied our methods

891 00:45:45.060 --> 00:45:48.510 to BCPP where we analyzed the individual's

892 00:45:48.510 --> 00:45:49.860 spillover total effects

893 00:45:49.860 --> 00:45:52.680 of receiving at least one intervention component

894 00:45:52.680 --> 00:45:56.703 on sexual risk behaviors one year after study enrollment.

895 00:45:57.630 --> 00:46:01.260 So as a reminder, the components here are HIV testing,

896 00:46:01.260 --> 00:46:05.730 HIV care, ART and circumcision.

897 00:46:05.730 --> 00:46:09.360 And here, we consider a binary outcome, which we define

898 00:46:09.360 --> 00:46:13.050 by one if a participant had reported having engaged

899 00:46:13.050 --> 00:46:16.383 in at least 30% of the surveyed risk behaviors.

900 00:46:18.150 --> 00:46:19.440 Here, for application,

901 00:46:19.440 --> 00:46:22.050 we consider the randomization clusters

902 00:46:22.050 --> 00:46:25.680 or communities as our first surrogate interference set.

903 00:46:25.680 --> 00:46:29.580 And we also consider a second surrogate interference set

904 00:46:29.580 --> 00:46:32.910 that is defined by smaller geographical plots.

905 00:46:32.910 --> 00:46:37.910 And in these geographical plots, they comprised

906 00:46:37.950 --> 00:46:41.700 of participant two to 18 participants on average,

907 00:46:41.700 --> 00:46:44.100 which were much smaller than randomization clusters,

908 00:46:44.100 --> 00:46:46.593 which were about 400 participants each.

909 00:46:48.927 --> 00:46:52.440 And in both of these interference sets,

910 00:46:52.440 --> 00:46:56.660 we define the spillover exposure to be one if at least 25%

911 00:46:56.660 --> 00:46:59.370 of participants in the inference set received

912 00:46:59.370 --> 00:47:02.550 at least one intervention component.

913 00:47:02.550 --> 00:47:04.230 And as in the second paper,

914 00:47:04.230 --> 00:47:06.520 we determine the true spillover exposures

915 00:47:07.967 --> 00:47:09.467 from the phylogenetic dataset.

916 00:47:12.660 --> 00:47:15.600 So here are the risk differences

917 00:47:15.600 --> 00:47:18.750 of receiving at least one intervention component

918 00:47:18.750 --> 00:47:20.980 on self-reported sexual risk behaviors

919 00:47:21.930 --> 00:47:24.750 where we compare the estimates obtained when we consider

920 00:47:24.750 --> 00:47:27.270 communities at the randomization clusters

921 00:47:27.270 --> 00:47:30.963 or the geographical plot as to the interference sets.

922 00:47:32.070 --> 00:47:35.190 And we compare these to the bias corrected estimates

923 00:47:35.190 --> 00:47:37.740 where here, I'm presenting the estimates

924 00:47:37.740 --> 00:47:39.990 of coming from the weighted average

925 00:47:39.990 --> 00:47:43.080 of the bias corrected estimates applied individually

926 00:47:43.080 --> 00:47:46.293 to the community and to the geographical plots.

927 00:47:48.120 --> 00:47:52.181 Where here, under the circuit interference,

928 00:47:52.181 --> 00:47:56.160 as we see that most effects were null.

929 00:47:56.160 --> 00:47:59.610 However, after bias correction, we see

930 00:47:59.610 --> 00:48:04.610 that there is a beneficial AIE when G is equal to one

931 00:48:04.770 --> 00:48:08.010 and beneficial ASP when A is equal to one.

932 00:48:08.010 --> 00:48:10.740 Which means that for participants

933 00:48:10.740 --> 00:48:13.920 who received at least one component of the intervention,

934 00:48:13.920 --> 00:48:16.950 if there were in the presence of at least 25%

935 00:48:16.950 --> 00:48:19.590 of participants who also received the intervention,

936 00:48:19.590 --> 00:48:22.050 that they had decreased risk behaviors.

937 00:48:22.050 --> 00:48:24.400 And likewise for ASP one,

938 00:48:27.660 --> 00:48:30.160 for participants who did receive the intervention,

939 00:48:32.430 --> 00:48:37.247 if they were exposed to at least 25% of participants

940 00:48:37.247 --> 00:48:41.280 in the interference set who also received the intervention,

941 00:48:41.280 --> 00:48:44.253 then there risk behavior fears were also reduced.

942 00:48:45.090 --> 00:48:46.203 But on the other hand,

943 00:48:47.647 --> 00:48:51.480 if a participant did not receive at least one component

944 00:48:51.480 --> 00:48:56.010 and greater than 75% of those interference

945 00:48:56.010 --> 00:48:58.740 that also did not receive the intervention,

946 00:48:58.740 --> 00:49:02.730 then this had an adverse effect on the risk behaviors.

947 00:49:02.730 --> 00:49:07.500 So overall, we see that a participant's risk behaviors

948 00:49:07.500 --> 00:49:10.110 are influenced by their own treatment

949 00:49:10.110 --> 00:49:14.280 and also in synergy with the treatment received

950 00:49:14.280 --> 00:49:16.773 by those in their interference set.

951 00:49:21.270 --> 00:49:23.820 So to wrap up,

952 00:49:23.820 --> 00:49:27.000 so we proposed several bias corrected estimators,

953 00:49:27.000 --> 00:49:29.220 which serve to decrease the bias in assessment

954 00:49:29.220 --> 00:49:32.100 of causal effects so that future intervention strategies

955 00:49:32.100 --> 00:49:35.133 can be more efficiently designed and interpreted.

956 00:49:36.300 --> 00:49:38.400 And our methods assume

957 00:49:38.400 --> 00:49:41.550 that we have suitable validation study that provides us

958 00:49:41.550 --> 00:49:44.130 with true measures of the interference set.

959 00:49:44.130 --> 00:49:46.630 However, as we see from our application

960 00:49:47.849 --> 00:49:49.560 that an exposure contamination dataset

961 00:49:49.560 --> 00:49:52.980 or a phylogenetic dataset are still imperfect measures

962 00:49:52.980 --> 00:49:55.290 of true social connections,

963 00:49:55.290 --> 00:49:56.310 although we do assume

964 00:49:56.310 --> 00:49:59.430 that these are more accurate than interference sets defined

965 00:49:59.430 --> 00:50:02.010 by general say spatial boundaries

966 00:50:02.010 --> 00:50:03.543 or administrative boundaries.

967 00:50:06.180 --> 00:50:08.430 And we propose for future extensions

968 00:50:08.430 --> 00:50:10.780 that we can perform sensitivity analysis

969 00:50:12.507 --> 00:50:13.860 on departures from the assumptions

970 00:50:13.860 --> 00:50:15.610 that are made in this dissertation.