WEBVTT

1 00:00:00.000 --> 00:00:03.180 - We want to the last Biostatistics seminar

2 00:00:03.180 --> 00:00:04.740 for the fall series.

3 00:00:04.740 --> 00:00:07.480 It's my great pleasure to welcome our speaker,

4 00:00:07.480 --> 00:00:09.450 Dr. Liangyuan Hu.

5 00:00:09.450 --> 00:00:12.680 Dr. Hu is an Assistant Professor of Biostatistics

6 00:00:12.680 --> 00:00:16.160 in the Department of Population Health Sciences and Policy

7 00:00:16.160 --> 00:00:19.050 at Mount Sinai School of Medicine.

8 00:00:19.050 --> 00:00:22.890 She received her PhD in Biostatistics from Brown University.

9 00:00:22.890 --> 00:00:25.170 Her methods research focuses on causal inference

10 00:00:25.170 --> 00:00:28.280 with complex longitudinal and survival data

11 00:00:28.280 --> 00:00:30.150 and Bayesian machine learning.

12 00:00:30.150 --> 00:00:33.210 Her independent research has been funded by NIH

13 00:00:33.210 --> 00:00:36.200 and Patient Centered Outcomes Research Institute.

14 00:00:36.200 --> 00:00:39.194 And her paper in Biometrics has been selected to receive

15 00:00:39.194 --> 00:00:44.194 the 2019 Outstanding Statistical Application Award

16 00:00:44.880 --> 00:00:47.810 by the American Statistical Association.

17 00:00:47.810 --> 00:00:50.270 Today, she's going to share with us her recent work

18 00:00:50.270 --> 00:00:54.190 on developing a continuous time marginal structure of models

19 00:00:54.190 --> 00:00:56.279 for complex survival outcomes.

20 00:00:56.279 --> 00:00:58.210 Liangyuan, the floor is yours.

21 00:00:58.210 --> 00:00:59.930 - Well, thank you Li Fan.

22 00:00:59.930 --> 00:01:02.079 Thank you so much Fan for your introduction,

23 00:01:02.079 --> 00:01:05.300 for the invite also.

24 00:01:05.300 --> 00:01:08.000 Let me just share my slides full screen.

25 00:01:08.000 --> 00:01:10.220 I'm really excited to be here today

26 00:01:10.220 --> 00:01:14.700 to talk about some of the projects I've been working on

27 00:01:14.700 --> 00:01:16.453 in the causal inference field,

28 00:01:17.390 --> 00:01:21.420 namely, how do we use marginal structure models

29 00:01:21.420 --> 00:01:25.970 for more complex comparative effectiveness

30 00:01:25.970 --> 00:01:29.500 research questions involving continuous-time treatment

31 00:01:29.500 --> 00:01:31.920 and censored survival outcomes.

32 00:01:31.920 --> 00:01:34.860 So I'd like to first acknowledge my colleagues,

33 00:01:34.860 --> 00:01:37.330 especially Doctors Hogan and Daniels

34 00:01:37.330 --> 00:01:40.170 who had been instrumental to me

35 00:01:40.170 --> 00:01:42.683 during the time I was working on this project.

36 00:01:44.289 --> 00:01:47.453 And let me just shift to the bar a little if I can.

37 00:01:50.120 --> 00:01:51.116 Okay.

38 00:01:51.116 --> 00:01:56.116 So this is just for those who aren't very familiar

39 00:01:56.340 --> 00:01:57.540 with causal inference,

40 00:01:57.540 --> 00:02:00.570 and simple slide to introduce the concept.

41 00:02:00.570 --> 00:02:02.020 Some key concept.

42 00:02:02.020 --> 00:02:06.040 Suppose we are interested in estimating the causal effect

43 00:02:06.040 --> 00:02:10.920 of a binary treatment A on some outcome Y.

44 00:02:10.920 --> 00:02:13.340 Using the potential outcomes framework,

45 00:02:13.340 --> 00:02:16.229 we can define the average treatment effect

46 00:02:16.229 --> 00:02:19.440 as the difference between the mean

47 00:02:19.440 --> 00:02:21.680 of the two sets of potential outcomes.

48 00:02:21.680 --> 00:02:25.620 So, Y1 here is the potential outcome

49 00:02:25.620 --> 00:02:27.030 that would have been observed

50 00:02:27.030 --> 00:02:30.450 had everyone in the population received the treatment.

51 00:02:30.450 --> 00:02:32.850 Similarly, Y0 here is the potential outcome

52 00:02:32.850 --> 00:02:34.290 that would have been observed

53 00:02:34.290 --> 00:02:37.510 had no one in the population received the treatment.

54 00:02:37.510 --> 00:02:39.490 To estimate the causal effect,

55 00:02:39.490 --> 00:02:43.450 the gold standard is the randomized controlled file.

56 00:02:43.450 --> 00:02:47.820 So in an RCT, we would randomly allocate patients

57 00:02:47.820 --> 00:02:52.820 to receive either treatment or the control or placebo,

58 00:02:52.870 --> 00:02:56.800 the randomization would make the two groups of patients

59 00:02:56.800 --> 00:03:01.110 more or less very similar in terms of their characteristics.

60 00:03:01.110 --> 00:03:06.110 So in a sense that these two groups are exchangeable,

61 00:03:07.600 --> 00:03:11.170 so that an individual's potential outcome

62 00:03:11.170 --> 00:03:13.550 to either treatment or control

63 00:03:13.550 --> 00:03:15.930 would not depend on which treatment group

64 00:03:15.930 --> 00:03:17.810 this person was assigned to.

65 00:03:17.810 --> 00:03:20.800 But just depends on how the treatment works.

66 00:03:20.800 --> 00:03:23.820 And this way we can simply look at the difference

67 00:03:23.820 --> 00:03:28.820 and the mean of the observed outcome

68 00:03:29.580 --> 00:03:31.490 between the two treatment groups

69 00:03:31.490 --> 00:03:34.130 and just to estimate the causal effect.

70 00:03:34.130 --> 00:03:35.867 But in many, many situations,

71 00:03:35.867 --> 00:03:38.550 we cannot conduct an RCT

72 00:03:38.550 --> 00:03:41.620 and we have to rely on observational data

73 00:03:41.620 --> 00:03:44.680 to get the causal inference about treatment effects.

74 00:03:44.680 --> 00:03:46.530 So in these situations,

75 00:03:46.530 --> 00:03:49.140 the independence between the potential outcome

76 00:03:49.140 --> 00:03:51.823 and treatment assignment would no longer hold.

77 00:03:52.870 --> 00:03:55.660 Because there might be exists a confounder

78 00:03:55.660 --> 00:03:58.240 that is predictive of the outcome,

79 00:03:58.240 --> 00:04:01.210 such that the probability of receiving the treatment

80 00:04:01.210 --> 00:04:02.640 depends on the confounder.

81 00:04:02.640 --> 00:04:06.093 So for example, age might be such a confounder.

82 00:04:06.093 --> 00:04:09.900 For example, younger patients may be more likely

83 00:04:09.900 --> 00:04:12.620 to receive the treatment.

84 00:04:12.620 --> 00:04:13.490 So in this case,

85 00:04:13.490 --> 00:04:16.410 if you take the difference in the average

86 00:04:16.410 --> 00:04:19.600 of the observed outcome between the two groups,

87 00:04:19.600 --> 00:04:23.880 then this estimate would not bear a causal interpretation

88 00:04:23.880 --> 00:04:27.173 because the difference might be confounded by age.

89 00:04:28.630 --> 00:04:31.100 So we would have to use specialized

90 00:04:31.100 --> 00:04:34.500 causal inference techniques to remove the confounding.

91 00:04:34.500 --> 00:04:37.420 And there are just many, many techniques out there,

92 00:04:37.420 --> 00:04:41.282 but today I'm just gonna focus on marginal structure model,

93 00:04:41.282 --> 00:04:46.150 because it is simple to implement.

94 00:04:46.150 --> 00:04:48.960 It has good statistical properties,

95 00:04:48.960 --> 00:04:50.740 and it is versatile enough

96 00:04:50.740 --> 00:04:53.550 to accommodate many, many complications

97 00:04:53.550 --> 00:04:57.520 posed by observational data that I'll talk about later.

98 00:04:57.520 --> 00:05:00.170 So we can propose a marginal structure model

99 00:05:00.170 --> 00:05:04.410 relating the potential outcome to the treatment assignment.

100 00:05:04.410 --> 00:05:08.020 And here theta one would capture the causal effect.

101 00:05:08.020 --> 00:05:09.070 But in reality,

102 00:05:09.070 --> 00:05:12.040 we can only fit a model to the observer data.

103 00:05:12.040 --> 00:05:15.611 And as I talked earlier,

104 00:05:15.611 --> 00:05:19.110 the parameter estimator beta one here

105 00:05:19.110 --> 00:05:23.270 would not bear a causal interpretation,

106 00:05:23.270 --> 00:05:25.600 it just measures association.

107 00:05:25.600 --> 00:05:27.840 But we can get to causation

108 00:05:27.840 --> 00:05:32.280 if by solving the weighted estimating equation,

109 00:05:32.280 --> 00:05:37.280 using the weight as W inverse of conditional probability

110 00:05:39.620 --> 00:05:43.380 of treatment assignment given the measured covariance.

111 00:05:43.380 --> 00:05:45.560 And this works because the IP weighting

112 00:05:45.560 --> 00:05:47.430 or inverse probability weighting

113 00:05:47.430 --> 00:05:51.350 removes confounding by measured covariance X

114 00:05:51.350 --> 00:05:53.373 in the weighted pseudo-population.

115 00:05:54.220 --> 00:05:57.433 So that's just a simple example

116 00:05:57.433 --> 00:06:02.433 to illustrate the use of marginal structure model.

117 00:06:02.880 --> 00:06:05.880 And traditionally treatment assignment,

118 00:06:05.880 --> 00:06:09.760 treatment is assigned at baseline and it's time fixed.

119 00:06:09.760 --> 00:06:13.663 So it means that the treatment doesn't change over time,

120 00:06:13.663 --> 00:06:18.663 but with increased availability of healthcare data sets,

121 00:06:20.910 --> 00:06:23.490 there are increased demands for more refined

122 00:06:23.490 --> 00:06:27.590 causal inference methods to evaluate complex

123 00:06:27.590 --> 00:06:28.930 treatment regimens.

124 00:06:28.930 --> 00:06:32.817 So one example is that treatment initiation

125 00:06:34.010 --> 00:06:38.673 can actually depend on time, so it changes over time.

126 00:06:39.981 --> 00:06:40.814 In this case,

127 00:06:40.814 --> 00:06:43.930 it would just be impractical to conduct RCTs

128 00:06:43.930 --> 00:06:45.590 because there are just simply too many

129 00:06:45.590 --> 00:06:47.323 treatment initiation time points.

130 00:06:48.290 --> 00:06:51.963 So I'm going to use two motivating examples in this talk.

131 00:06:53.920 --> 00:06:58.920 The first example is about timing of treatment initiation

132 00:06:59.880 --> 00:07:04.880 for patients who present both HIV and TB, tuberculosis.

133 00:07:05.130 --> 00:07:06.570 For these patients,

134 00:07:06.570 --> 00:07:09.380 TB treatment will be initiated immediately

135 00:07:09.380 --> 00:07:13.120 after the diagnosis, but during the TB treatment,

136 00:07:13.120 --> 00:07:17.260 when is the optimal time to initiate the HIV treatment

137 00:07:17.260 --> 00:07:19.620 or ART, anti-retroviral therapy?

138 00:07:19.620 --> 00:07:22.430 That is a very important question to answer,

139 00:07:22.430 --> 00:07:25.010 because if you initiate the treatment too early,

140 00:07:25.010 --> 00:07:28.610 there might be drug interactions, drug toxicity,

141 00:07:28.610 --> 00:07:30.870 but if you delay the treatment too much,

142 00:07:30.870 --> 00:07:33.050 then there's also increased the mortality

143 00:07:33.050 --> 00:07:34.913 associated with AIDS.

144 00:07:35.980 --> 00:07:39.370 The second example is timing of HIV treatment

145 00:07:39.370 --> 00:07:40.543 for adolescents.

146 00:07:41.830 --> 00:07:44.750 The timing now is defined with respect

147 00:07:44.750 --> 00:07:48.577 to the evolving value of a biomarker CD4.

148 00:07:48.577 --> 00:07:51.430 And this is also an important question to answer

149 00:07:51.430 --> 00:07:56.430 because the WHO guideline is in the form of

150 00:07:57.130 --> 00:08:01.350 treat this person when the person's CD4 cell count

151 00:08:01.350 --> 00:08:04.150 drops below 350, for example,

152 00:08:04.150 --> 00:08:06.450 and for the population of adolescents

153 00:08:06.450 --> 00:08:10.440 currently there's no concrete evidence

154 00:08:10.440 --> 00:08:13.753 for supporting the optimal threshold.

155 00:08:15.370 --> 00:08:20.040 So to statistically formulate these two motivating examples,

156 00:08:20.040 --> 00:08:21.380 the first one,

157 00:08:21.380 --> 00:08:24.720 when is the best time to initiate a treatment?

158 00:08:24.720 --> 00:08:27.120 So this is actually a static treatment regimen

159 00:08:27.120 --> 00:08:28.703 with respect to time,

160 00:08:31.919 --> 00:08:32.752 and the initiation can occur on the continuous timescale.

161 00:08:34.670 --> 00:08:38.690 And second example is actually a dynamic treatment regimen.

162 00:08:38.690 --> 00:08:43.690 It's dynamic because it depends on the evolving history

163 00:08:44.890 --> 00:08:47.528 of treatment and a biomarker,

164 00:08:47.528 --> 00:08:52.528 but initiation can also occur on the continuous timescale.

165 00:08:53.570 --> 00:08:56.380 So marginal structure models are suitable

166 00:08:56.380 --> 00:08:59.300 for addressing a time dependent treatment,

167 00:08:59.300 --> 00:09:02.000 but in order to use the models,

168 00:09:02.000 --> 00:09:04.980 we have to overcome some statistical challenges.

169 00:09:04.980 --> 00:09:07.530 The first challenge is that we need

170 00:09:07.530 --> 00:09:11.070 to estimate the causal effect of the actual timing,

171 00:09:11.070 --> 00:09:15.360 not compare protocols defined by some specific intervals,

172 00:09:15.360 --> 00:09:18.653 which is a lot of existing studies did.

173 00:09:21.331 --> 00:09:25.140 And also a lot of RCT reported these kinds of results.

174 00:09:25.140 --> 00:09:26.690 Because as I said earlier,

175 00:09:26.690 --> 00:09:29.700 it's just impractical for RCTs

176 00:09:29.700 --> 00:09:32.713 to report continuous time causal effects.

177 00:09:33.620 --> 00:09:36.320 We would also need to address complications

178 00:09:36.320 --> 00:09:38.840 posed by observational data.

179 00:09:38.840 --> 00:09:41.240 This is something I'll talk about later.

180 00:09:41.240 --> 00:09:44.750 And also we are dealing with censored survival outcomes

181 00:09:44.750 --> 00:09:47.393 that adds another layer of complexity.

182 00:09:48.550 --> 00:09:53.300 So these are four sensory patterns observed in our data.

183 00:09:53.300 --> 00:09:57.200 So our goal is to estimate the causal effect of A,

184 00:09:57.200 --> 00:10:00.610 treatment initiation time and T, death time.

185 00:10:00.610 --> 00:10:04.590 And we have almost 5,000 patients

186 00:10:04.590 --> 00:10:07.160 and only a very small proportion of patients

187 00:10:07.160 --> 00:10:10.103 have both observed A and T.

188 00:10:10.103 --> 00:10:12.930 A lot of patients don't have observed T.

189 00:10:12.930 --> 00:10:15.510 So their death time is censored by C.

190 00:10:15.510 --> 00:10:18.030 And we have about 20% of our patients,

191 00:10:18.030 --> 00:10:20.256 they don't even have observed A.

192 00:10:20.256 --> 00:10:21.920 Their treatment initiation time

193 00:10:21.920 --> 00:10:26.550 can be censored by death time or censored by C, dropout,

194 00:10:26.550 --> 00:10:27.513 for example.

195 00:10:28.360 --> 00:10:32.160 So our goal is to estimate effect of A on T,

196 00:10:32.160 --> 00:10:34.530 but we only have about 300 patients

197 00:10:34.530 --> 00:10:36.250 have complete information.

198 00:10:36.250 --> 00:10:39.590 Most of the patients we have incomplete information

199 00:10:39.590 --> 00:10:42.500 on either A or T or both.

200 00:10:42.500 --> 00:10:46.090 How do we probably use these incomplete information

201 00:10:46.090 --> 00:10:49.360 to draw causal inference about A on T,

202 00:10:49.360 --> 00:10:50.860 the effect of A on T,

203 00:10:50.860 --> 00:10:54.693 that's a problem we solve in this project.

204 00:10:55.954 --> 00:10:59.230 So three challenges.

205 00:10:59.230 --> 00:11:02.160 First one, treatment initiation time,

206 00:11:02.160 --> 00:11:05.600 this is observational data, so it's not randomly allocated.

207 00:11:05.600 --> 00:11:09.940 We don't know the actual functional form of causal effect

208 00:11:09.940 --> 00:11:13.190 of initiation timing or mortality rate.

209 00:11:13.190 --> 00:11:16.890 And we see that, Oh, there's incomplete information

210 00:11:16.890 --> 00:11:20.073 on either exposure or outcome or both.

211 00:11:21.500 --> 00:11:23.680 The general solutions we proposed

212 00:11:26.010 --> 00:11:29.450 that we first formulate a flexible structural

213 00:11:29.450 --> 00:11:30.990 causal hazard model

214 00:11:30.990 --> 00:11:35.990 that can capture the effects of both timing and duration

215 00:11:36.130 --> 00:11:36.963 of the treatment.

216 00:11:36.963 --> 00:11:39.370 And then we can derive methods

217 00:11:39.370 --> 00:11:43.780 to consistently estimate the model parameters

218 00:11:43.780 --> 00:11:48.290 under non random allocation and complex censoring patterns.

219 00:11:48.290 --> 00:11:52.900 Using the model outputs we can estimate the functional form

220 00:11:52.900 --> 00:11:56.452 of the causal relationship between our initiation timing

221 00:11:56.452 --> 00:11:58.023 and mortality.

222 00:11:58.920 --> 00:12:02.463 So some notation before we introduce our approach,

223 00:12:02.463 --> 00:12:06.300 note that we have three time to events in our study,

224 00:12:06.300 --> 00:12:08.500 we have treatment initiation time, death time,

225 00:12:08.500 --> 00:12:09.690 censoring time.

226 00:12:09.690 --> 00:12:12.890 We'll use T sub cap A to denote death time

227 00:12:12.890 --> 00:12:17.220 associated with the actual treatment time.

228 00:12:17.220 --> 00:12:19.900 And potential outcomes T sub small A,

229 00:12:19.900 --> 00:12:21.180 this is the death time.

230 00:12:21.180 --> 00:12:24.360 If treatment initiated at time A,

231 00:12:24.360 --> 00:12:27.600 and we use T infinity to denote death time

232 00:12:27.600 --> 00:12:30.950 if treatment is initiated beyond sometime point

233 00:12:30.950 --> 00:12:32.193 of our interest.

234 00:12:33.269 --> 00:12:36.380 Because of all the censoring,

235 00:12:36.380 --> 00:12:39.950 all the three time to events can be censored by one another.

236 00:12:39.950 --> 00:12:44.100 We use T star to denote the minimum of T and C.

237 00:12:44.100 --> 00:12:47.060 Delta T is a corresponding event indicator.

238 00:12:47.060 --> 00:12:50.600 So A star is the minimum of the three time to events.

239 00:12:50.600 --> 00:12:53.838 Delta A is a corresponding event in the data.

240 00:12:53.838 --> 00:12:57.600 Adopting the convention in the causal inference literature,

241 00:12:57.600 --> 00:13:00.220 we use overbar to denote history.

242 00:13:00.220 --> 00:13:05.220 So overbar L of T here is a covariate history

243 00:13:05.880 --> 00:13:07.960 up to a time T.

244 00:13:07.960 --> 00:13:09.210 Putting everything together,

245 00:13:09.210 --> 00:13:12.040 we have a set of observed data.

246 00:13:12.040 --> 00:13:14.990 Now back to the censoring patterns.

247 00:13:14.990 --> 00:13:18.210 In case one, we observed both A and T.

248 00:13:18.210 --> 00:13:21.683 So we would observe A, we would observe T sub A.

249 00:13:22.630 --> 00:13:25.310 Case two T is censored by C,

250 00:13:25.310 --> 00:13:27.660 so we observe A, we just know TA

251 00:13:27.660 --> 00:13:29.390 is going to be greater than C.

252 00:13:29.390 --> 00:13:30.502 Case three,

253 00:13:30.502 --> 00:13:32.180 we will observe A,

254 00:13:32.180 --> 00:13:34.710 but we know A is greater than TA.

255 00:13:34.710 --> 00:13:38.170 And case four we don't observe A, we don't observe T

256 00:13:38.170 --> 00:13:41.723 but we know A is greater than C and TA is greater than C.

257 00:13:42.850 --> 00:13:43.683 Okay.

258 00:13:43.683 --> 00:13:46.020 So now we propose a structural causal

259 00:13:46.020 --> 00:13:47.840 proportional hazards model

260 00:13:48.970 --> 00:13:52.493 to capture the survival effect of treatment initiation time.

261 00:13:53.520 --> 00:13:55.690 Lambda AT here is a hazard function

262 00:13:55.690 --> 00:13:58.510 for the potential outcome T sub A,

263 00:13:58.510 --> 00:14:01.320 we start from lambda infinity T right here.

264 00:14:01.320 --> 00:14:04.514 This is a reference hazard for T infinity.

265 00:14:04.514 --> 00:14:06.610 So we start from here.

266 00:14:06.610 --> 00:14:10.040 Once the treatment is initiated at A,

267 00:14:10.040 --> 00:14:13.910 there is an instantaneous effect of treatment initiation

268 00:14:13.910 --> 00:14:16.870 captured by the G1 function here,

269 00:14:16.870 --> 00:14:19.990 and the effect of staying on the treatment

270 00:14:19.990 --> 00:14:22.360 at any given time point T,

271 00:14:22.360 --> 00:14:27.000 is captured by the G2 function of ART duration.

272 00:14:27.000 --> 00:14:30.260 And the G3 function here captures the interaction

273 00:14:30.260 --> 00:14:34.323 between treatment initiation and treatment duration.

274 00:14:35.470 --> 00:14:40.470 So we leave this structural model relatively flexible.

275 00:14:40.790 --> 00:14:44.360 First, the reference hazard is left unspecified

276 00:14:44.360 --> 00:14:46.570 and the 3G functions, we also left them

277 00:14:46.570 --> 00:14:49.140 as unspecified smooth function

278 00:14:49.140 --> 00:14:53.233 of treatment initiation time duration and their interaction.

279 00:14:54.220 --> 00:14:57.600 So now we can parametrize these three functions

280 00:14:57.600 --> 00:14:59.754 using natural cubic splines,

281 00:14:59.754 --> 00:15:04.754 and by rewriting the risk function of our structural model,

282 00:15:06.640 --> 00:15:09.460 we can use beta this parameter

283 00:15:09.460 --> 00:15:13.140 to include the causal effects of ART initiation time

284 00:15:13.140 --> 00:15:14.870 on mortality hazard.

285 00:15:14.870 --> 00:15:17.180 The problem here now,

286 00:15:17.180 --> 00:15:19.860 our goal is to how do we obtain a consistent

287 00:15:19.860 --> 00:15:22.653 estimate of beta using observed a data?

288 00:15:23.690 --> 00:15:25.660 Once we have obtained that

289 00:15:25.660 --> 00:15:29.990 we can use beta hat to estimate the 3G functions,

290 00:15:29.990 --> 00:15:33.440 to understand the relative contribution of timing

291 00:15:33.440 --> 00:15:36.810 versus duration and interactions.

292 00:15:36.810 --> 00:15:40.520 And we could also estimate the causal does-response

293 00:15:40.520 --> 00:15:43.230 of initiation time versus mortality

294 00:15:43.230 --> 00:15:46.450 by relating the survival function to the hazard function.

295 00:15:46.450 --> 00:15:51.283 We can derive this from our structural model.

296 00:15:52.160 --> 00:15:54.640 And now we can also estimate the model-based

297 00:15:54.640 --> 00:15:56.650 optimal initiation time

298 00:15:56.650 --> 00:16:01.330 that will lead to the maximal survival probability

299 00:16:01.330 --> 00:16:05.423 at say 52 weeks after diagnosis.

300 00:16:06.500 --> 00:16:10.170 Okay, how to obtain a consistent estimate of beta.

301 00:16:10.170 --> 00:16:14.950 So first let's assume if A is randomly allocated

302 00:16:14.950 --> 00:16:17.410 and both A and T are observed,

303 00:16:17.410 --> 00:16:21.910 then we can write the partial likelihood score function

304 00:16:21.910 --> 00:16:24.120 of our structural model.

305 00:16:24.120 --> 00:16:28.135 And this is a sample average of score function

306 00:16:28.135 --> 00:16:31.550 is an unbiased estimator of the expectation

307 00:16:31.550 --> 00:16:32.730 of the score function.

308 00:16:32.730 --> 00:16:36.508 So E sub R here is the expectation

309 00:16:36.508 --> 00:16:39.950 under the randomized treatment assignment.

310 00:16:39.950 --> 00:16:44.950 So this would be an unbiased estimator function,

311 00:16:46.550 --> 00:16:50.070 and solving this unbiased estimating equation

312 00:16:50.070 --> 00:16:53.153 would give us a consistent estimator of beta.

313 00:16:54.760 --> 00:16:57.900 Now, if A is still randomly allocated,

314 00:16:57.900 --> 00:17:00.173 but T can occur before A,

315 00:17:01.933 --> 00:17:03.823 so A may be censored by T.

316 00:17:04.830 --> 00:17:05.670 In this case,

317 00:17:05.670 --> 00:17:08.050 we would need to break the mean

318 00:17:08.050 --> 00:17:11.280 of an individual score contribution into two parts.

319 00:17:11.280 --> 00:17:13.170 In one part A is observed.

320 00:17:13.170 --> 00:17:15.740 The second part is A is not observed.

321 00:17:15.740 --> 00:17:18.890 And then we can apply the law of total expectation

322 00:17:18.890 --> 00:17:21.100 to the second part.

323 00:17:21.100 --> 00:17:23.840 The inner expectation would be conditioning

324 00:17:23.840 --> 00:17:26.390 on the observed information.

325 00:17:26.390 --> 00:17:30.230 Then using this strategy and taking in account

326 00:17:30.230 --> 00:17:32.220 the survival hazard structure,

327 00:17:32.220 --> 00:17:37.220 we can revise the estimating equation.

328 00:17:37.350 --> 00:17:41.253 And by solving this to obtain a consistent estimate of beta.

329 00:17:42.470 --> 00:17:45.660 In the case of non random allocation of treatment,

330 00:17:45.660 --> 00:17:50.380 then if we want to estimate the causal effect of A on T,

331 00:17:50.380 --> 00:17:53.943 then we would have to make a key assumption,

332 00:17:55.600 --> 00:17:57.150 ignore ability assumption.

333 00:17:57.150 --> 00:17:58.620 Essentially the assumption says

334 00:17:58.620 --> 00:18:03.620 that the initiation of treatment at any given time T

335 00:18:04.190 --> 00:18:06.400 is sequentially randomized in the sense

336 00:18:06.400 --> 00:18:09.060 that as a potential outcome beyond this time

337 00:18:09.060 --> 00:18:11.870 is independent of treatment initiation.

338 00:18:11.870 --> 00:18:15.930 Conditioning on all covariate history up to T.

339 00:18:15.930 --> 00:18:17.363 So with this assumption,

340 00:18:18.610 --> 00:18:21.110 we will be able to use observed data

341 00:18:21.110 --> 00:18:23.180 to derive the causal effect.

342 00:18:23.180 --> 00:18:27.460 So say PR is the data distribution under randomized A,

343 00:18:27.460 --> 00:18:29.940 and PO is the data distribution.

344 00:18:29.940 --> 00:18:33.120 And they're not random allocation of A.

345 00:18:33.120 --> 00:18:35.440 Note that in both settings,

346 00:18:35.440 --> 00:18:38.710 there is a same set of observed data.

347 00:18:38.710 --> 00:18:42.650 And as long as the observed data under PR

348 00:18:42.650 --> 00:18:47.650 is absolutely continues with the observed data under PO.

349 00:18:48.170 --> 00:18:51.940 Now we can derive a random-nikodym derivative.

350 00:18:51.940 --> 00:18:54.500 And so Murphy's 2001 paper

351 00:18:54.500 --> 00:18:57.730 developed a version of R-N derivative

352 00:18:57.730 --> 00:19:01.360 that connects the distribution of the observed data

353 00:19:01.360 --> 00:19:04.710 under PR and under PO for discrete time

354 00:19:04.710 --> 00:19:06.840 and ordinary GEE score.

355 00:19:06.840 --> 00:19:11.840 Johnson's 2005 paper extended this version of R-N derivative

356 00:19:11.950 --> 00:19:15.470 to continuous time still for ordinary GEE score.

357 00:19:15.470 --> 00:19:20.287 In this paper we extended the R-N derivative

358 00:19:20.287 --> 00:19:23.451 for time to event setting.

359 00:19:23.451 --> 00:19:26.350 So this is a version of R-N derivative

360 00:19:26.350 --> 00:19:28.084 for survival data.

361 00:19:28.084 --> 00:19:31.940 The reason why we wanted to use R-N derivative

362 00:19:31.940 --> 00:19:34.080 is that we can then use it

363 00:19:34.080 --> 00:19:36.840 to derive an unbiased estimating equation

364 00:19:36.840 --> 00:19:40.930 using some weighted version of the observed data.

365 00:19:40.930 --> 00:19:43.490 So we can estimate the causal effect.

366 00:19:43.490 --> 00:19:48.490 So now we want to use this R-N derivative for survival data.

367 00:19:48.628 --> 00:19:51.380 We want to apply that to Cox score

368 00:19:51.380 --> 00:19:54.760 and to derive S rated estimating equation.

369 00:19:54.760 --> 00:19:59.550 That's a little bit more complex than the GEE score,

370 00:19:59.550 --> 00:20:01.770 but we can observe that the Cox score

371 00:20:01.770 --> 00:20:06.300 can essentially be represented in three averages.

372 00:20:06.300 --> 00:20:07.710 The one in blue,

373 00:20:07.710 --> 00:20:12.670 the one in orange and the whole average.

374 00:20:12.670 --> 00:20:17.240 And each average converges to its expectation.

375 00:20:17.240 --> 00:20:19.080 And as I showed earlier,

376 00:20:19.080 --> 00:20:22.550 we can always break the expectation into two parts.

377 00:20:22.550 --> 00:20:24.770 In one part A is observed,

378 00:20:24.770 --> 00:20:26.880 second part is not observed.

379 00:20:26.880 --> 00:20:28.070 For the second part,

380 00:20:28.070 --> 00:20:32.418 we can apply the total law of expectation,

381 00:20:32.418 --> 00:20:34.700 the law of total expectation,

382 00:20:34.700 --> 00:20:39.700 and recognizing the survival structure

383 00:20:40.310 --> 00:20:42.560 to derive the second part.

384 00:20:42.560 --> 00:20:46.340 And then we can apply the R-N derivative for survival data

385 00:20:46.340 --> 00:20:48.400 to each piece separately,

386 00:20:48.400 --> 00:20:52.143 to construct the unbiased score equation.

387 00:20:53.390 --> 00:20:58.390 So after some derivation, we would arrive at the weights

388 00:20:59.110 --> 00:21:03.480 and actually the weights come down in a very neat form.

389 00:21:03.480 --> 00:21:06.040 Essentially, it suggests that for patients

390 00:21:06.040 --> 00:21:09.630 who have initiated treatment by time T,

391 00:21:09.630 --> 00:21:12.960 we would weight them by the marginals density function

392 00:21:12.960 --> 00:21:17.500 of A divided by the conditional density of A

393 00:21:17.500 --> 00:21:22.500 given their covariate history after time T.

394 00:21:22.710 --> 00:21:25.080 And for those who are censored,

395 00:21:25.080 --> 00:21:27.630 so not initiated by the time T,

396 00:21:27.630 --> 00:21:30.280 we would weight them by some survival function

397 00:21:31.229 --> 00:21:34.673 of the treatment initiation process.

398 00:21:36.120 --> 00:21:38.910 And then by applying this weighting scheme,

399 00:21:38.910 --> 00:21:43.300 we will be able to derive a weighted estimating equation.

400 00:21:43.300 --> 00:21:45.880 And just a note that we have to apply

401 00:21:45.880 --> 00:21:49.370 the same weighting scheme to the people

402 00:21:49.370 --> 00:21:52.653 who are still in the risk set at any time T.

403 00:21:54.310 --> 00:21:57.750 And so now that said, previously we have assumed

404 00:21:57.750 --> 00:21:58.900 there's no censoring.

405 00:21:58.900 --> 00:22:00.368 Now with censoring,

406 00:22:00.368 --> 00:22:05.368 we need to assume another similar assumption,

407 00:22:05.895 --> 00:22:08.824 similar to the ignore ability assumption,

408 00:22:08.824 --> 00:22:12.140 and then using the similar strategy

409 00:22:12.140 --> 00:22:15.013 to derive another set of weight for censoring.

410 00:22:16.050 --> 00:22:18.450 For those who stay, remain in the study,

411 00:22:18.450 --> 00:22:22.520 we would weight them by the survival function

412 00:22:22.520 --> 00:22:23.940 for censoring.

413 00:22:23.940 --> 00:22:26.190 And this would lead to the final modification

414 00:22:27.136 --> 00:22:29.470 of the estimating equation for beta.

415 00:22:29.470 --> 00:22:33.330 So censoring contributes information about the parameter

416 00:22:33.330 --> 00:22:34.770 in two ways,

417 00:22:34.770 --> 00:22:39.510 FC is observed as the person is actually censored.

418 00:22:39.510 --> 00:22:42.250 It contributes to the risk set up to C.

419 00:22:42.250 --> 00:22:45.960 If C is not observed, so C could be censored by T.

420 00:22:45.960 --> 00:22:47.210 If death's occurred,

421 00:22:47.210 --> 00:22:49.990 then it contributes to the individual partial likelihood

422 00:22:49.990 --> 00:22:53.633 to weight for C but evaluated at death time.

423 00:22:54.810 --> 00:22:56.460 Okay, now we know how to weight.

424 00:22:56.460 --> 00:22:58.640 Back to the four censoring patterns.

425 00:22:58.640 --> 00:23:01.530 The first one, both A and T are observed.

426 00:23:01.530 --> 00:23:06.110 We would weight them by the first set of weight for A

427 00:23:06.110 --> 00:23:06.943 evaluated at A,

428 00:23:08.170 --> 00:23:13.170 T occurred, so the weight for C but evaluated at T.

429 00:23:13.360 --> 00:23:16.610 Second case, T is not observed,

430 00:23:16.610 --> 00:23:19.100 A is observed.

431 00:23:19.100 --> 00:23:23.340 So first set of weight for A evaluated at A

432 00:23:23.340 --> 00:23:26.163 and C just contributes information to the risks set.

433 00:23:28.260 --> 00:23:31.370 Third case, A is not observed,

434 00:23:31.370 --> 00:23:35.220 so second weight for A evaluated at T.

435 00:23:35.220 --> 00:23:40.220 And weight for C, censoring evaluated at T.

436 00:23:40.250 --> 00:23:43.820 The fourth case or final case, A is not observed,

437 00:23:43.820 --> 00:23:46.780 again, second set of weight for A,

438 00:23:46.780 --> 00:23:50.453 but evaluated at C, and C also contributes to the risks set.

439 00:23:51.420 --> 00:23:53.930 Okay, so now we know how to weight.

440 00:23:53.930 --> 00:23:57.843 We would have to estimate the weights.

441 00:24:00.320 --> 00:24:02.490 The approach we used in the paper

442 00:24:02.490 --> 00:24:05.540 is that we model the intensity processes

443 00:24:05.540 --> 00:24:09.410 associated with the two counting processes,

444 00:24:09.410 --> 00:24:11.500 one for A, one for C.

445 00:24:11.500 --> 00:24:14.680 And then when we fit Cox proportional hazards models

446 00:24:14.680 --> 00:24:17.220 for the two intensity processes,

447 00:24:17.220 --> 00:24:20.053 we use fitted hazard to estimate the weights.

448 00:24:21.110 --> 00:24:23.680 We use empirical cumulative hazards

449 00:24:23.680 --> 00:24:26.810 to estimate the conditional density and function.

450 00:24:26.810 --> 00:24:28.760 And for the marginal density function,

451 00:24:28.760 --> 00:24:32.110 we use some nonparametric Nelson-Aalen estimator,

452 00:24:32.110 --> 00:24:34.910 and use similar fashion to estimate rates for censoring.

453 00:24:36.220 --> 00:24:39.380 Then we apply our methods to the AMPATH data.

454 00:24:39.380 --> 00:24:44.210 AMPATH is a large HIV care program based in West Kenya,

455 00:24:44.210 --> 00:24:47.100 our data has almost 5,000 patients

456 00:24:47.100 --> 00:24:51.017 and for covariates, we have demographic information

457 00:24:51.017 --> 00:24:53.590 and some disease-specific information.

458 00:24:53.590 --> 00:24:56.440 Some of them are time varying like, weight, the CD4,

459 00:24:56.440 --> 00:24:58.890 these are time varying variables.

460 00:24:58.890 --> 00:25:03.777 We categorize the baseline CD4 subgroups into two groups,

461 00:25:05.980 --> 00:25:08.310 the less than, or below 50 group,

462 00:25:08.310 --> 00:25:10.900 this is the highest risk group.

463 00:25:10.900 --> 00:25:13.600 So CD4 the higher, the better.

464 00:25:13.600 --> 00:25:16.170 So below 50, this is a highest risk group.

465 00:25:16.170 --> 00:25:18.690 And between 200 and 350,

466 00:25:18.690 --> 00:25:20.890 there's relatively healthy patients.

467 00:25:20.890 --> 00:25:23.200 The reason we categorize them into three groups

468 00:25:23.200 --> 00:25:26.190 is because the program guidelines

469 00:25:26.190 --> 00:25:28.180 are based on these subgroups

470 00:25:28.180 --> 00:25:31.733 and RCT is reported results for below 50 group.

471 00:25:33.039 --> 00:25:37.030 We want to compare our results to our CT findings.

472 00:25:37.030 --> 00:25:41.950 So this plot shows the three estimated G functions.

473 00:25:41.950 --> 00:25:46.830 The G1 A here suggests that the instantaneous effect

474 00:25:46.830 --> 00:25:49.920 of a treatment initiation has a U shape,

475 00:25:49.920 --> 00:25:53.290 achieving maximum benefit, or the lowest mortality hazard

476 00:25:53.290 --> 00:25:55.620 at just about 10 weeks.

477 00:25:55.620 --> 00:25:59.630 And after that, the longer the treatment is delayed,

478 00:25:59.630 --> 00:26:03.170 the less the benefit of the treatment initiation.

479 00:26:03.170 --> 00:26:05.510 And this is the effect of duration,

480 00:26:05.510 --> 00:26:07.660 in general, it says that the longer

481 00:26:07.660 --> 00:26:10.700 you stay on the treatment, the more benefit you get.

482 00:26:10.700 --> 00:26:15.170 There's an upward trend for the interaction effect.

483 00:26:15.170 --> 00:26:18.830 Essentially suggesting that delayed treatment initiation

484 00:26:18.830 --> 00:26:21.800 would reduce the benefit associated

485 00:26:21.800 --> 00:26:25.023 with long ART duration.

486 00:26:26.990 --> 00:26:30.930 And so the net causal effect of treatment initiation

487 00:26:30.930 --> 00:26:33.310 is summarized in this plot.

488 00:26:33.310 --> 00:26:37.570 Top panel shows the mortality rate at one year

489 00:26:37.570 --> 00:26:40.210 versus treatment initiation time.

490 00:26:40.210 --> 00:26:44.320 Bottom panel compares immediate initiation

491 00:26:44.320 --> 00:26:47.990 versus delayed initiation at A.

492 00:26:47.990 --> 00:26:52.670 So we can see that the benefit of early initiation

493 00:26:52.670 --> 00:26:56.650 is most pronounced for the CD4 below 50 group,

494 00:26:56.650 --> 00:26:58.310 or the highest risk group.

495 00:26:58.310 --> 00:27:00.550 And the curves here are pretty flat,

496 00:27:00.550 --> 00:27:03.077 suggesting that there's not much benefit

497 00:27:03.077 --> 00:27:06.773 of early initiation for relatively healthy patients.

498 00:27:08.770 --> 00:27:12.063 Several advantages for this approach.

499 00:27:12.063 --> 00:27:16.960 It's easy to get optimal initiation time

500 00:27:16.960 --> 00:27:19.123 based on the model outputs.

501 00:27:20.325 --> 00:27:22.410 And we could also use the model outputs

502 00:27:22.410 --> 00:27:27.370 to emulate comparisons between regimens reported in RCTs.

503 00:27:27.370 --> 00:27:31.940 So we could mimic random allocation

504 00:27:31.940 --> 00:27:35.690 of treatment initiation time to specific intervals

505 00:27:35.690 --> 00:27:38.670 by assuming a distribution for A,

506 00:27:38.670 --> 00:27:41.170 for treatment initiation time A,

507 00:27:41.170 --> 00:27:44.020 that is independent of covariates and outcome

508 00:27:44.020 --> 00:27:48.820 and compare interval specific mortality rates

509 00:27:48.820 --> 00:27:53.180 and draw inferences about treatment initiation.

510 00:27:53.180 --> 00:27:56.210 But with the continuous time marginal structure model,

511 00:27:56.210 --> 00:28:00.070 we'll also be able to conduct a higher resolution analysis

512 00:28:00.070 --> 00:28:02.620 that can potentially generate new insights

513 00:28:02.620 --> 00:28:05.893 in relation to a randomized control trial.

514 00:28:09.160 --> 00:28:10.480 For the sake of timing,

515 00:28:10.480 --> 00:28:13.630 I just gonna briefly talk about the simulation.

516 00:28:13.630 --> 00:28:15.440 We conduct simulation to examine

517 00:28:15.440 --> 00:28:18.800 the finite-sample properties of weighted estimators,

518 00:28:23.890 --> 00:28:26.759 we evaluate sensitivity of our estimators

519 00:28:26.759 --> 00:28:29.580 to the violations of the ignore ability,

520 00:28:29.580 --> 00:28:31.870 or no unmeasured confounding assumption,

521 00:28:31.870 --> 00:28:34.780 but we only considered confounding at baseline.

522 00:28:34.780 --> 00:28:38.640 So the sensitivity analysis strategy

523 00:28:38.640 --> 00:28:41.120 for time-varying confounding,

524 00:28:41.120 --> 00:28:43.760 especially with the censored survival outcome

525 00:28:43.760 --> 00:28:48.220 is kind of very complex topic,

526 00:28:48.220 --> 00:28:51.340 and we were still working on this project right now,

527 00:28:51.340 --> 00:28:54.683 but in this paper we just consider confounding at baseline.

528 00:28:56.330 --> 00:28:58.690 Under random allocation of treatment,

529 00:28:58.690 --> 00:29:02.060 our estimator produced a new zero bias

530 00:29:02.060 --> 00:29:04.910 and nominal coverage probability,

531 00:29:04.910 --> 00:29:06.990 in the presence of measured confounding,

532 00:29:06.990 --> 00:29:09.260 it eliminated nearly all the biases

533 00:29:09.260 --> 00:29:12.949 and provided close to nominal coverage probability,

534 00:29:12.949 --> 00:29:16.130 but in the presence of unmeasured confounding,

535 00:29:16.130 --> 00:29:19.100 there was bias in our estimator.

536 00:29:19.100 --> 00:29:22.140 And the biases were in proportion

537 00:29:22.140 --> 00:29:24.333 to the degree of measured confounding.

538 00:29:26.490 --> 00:29:27.323 Okay,

539 00:29:27.323 --> 00:29:29.440 so moving to the second example,

540 00:29:29.440 --> 00:29:33.610 this is a continuous time dynamic treatment regimen

541 00:29:33.610 --> 00:29:34.443 of the form,

542 00:29:34.443 --> 00:29:38.373 initiate treatment when a biomarker crosses a threshold.

543 00:29:39.512 --> 00:29:41.930 It's dynamic treatment regimen

544 00:29:41.930 --> 00:29:44.970 because it depends on evolving history of treatment

545 00:29:44.970 --> 00:29:46.980 and a tailoring variable.

546 00:29:46.980 --> 00:29:49.600 So in our case, CD4 is a tailoring variable.

547 00:29:49.600 --> 00:29:52.790 That means we make our treatment decision

548 00:29:52.790 --> 00:29:54.073 based on this variable.

549 00:29:55.150 --> 00:30:00.100 A little bit different from our previous motivating example.

550 00:30:00.100 --> 00:30:03.580 The outcome interest is different.

551 00:30:03.580 --> 00:30:05.150 This is a pediatric data.

552 00:30:05.150 --> 00:30:08.980 So for the kids, the mortality rate is very low

553 00:30:08.980 --> 00:30:12.242 and our data I think it's around 3%.

554 00:30:12.242 --> 00:30:14.470 And for kids, we're also interested

555 00:30:14.470 --> 00:30:17.220 in their CD4 measurements,

556 00:30:17.220 --> 00:30:21.300 because CD4 is important marker of immune system function

557 00:30:21.300 --> 00:30:24.260 and both outcomes, both mortality rate and CD4

558 00:30:24.260 --> 00:30:26.333 are sparsely measured in our data,

559 00:30:27.200 --> 00:30:28.700 but we are interested in both.

560 00:30:29.620 --> 00:30:32.790 Other than that, we also have complications

561 00:30:32.790 --> 00:30:36.250 posed by observational data.

562 00:30:36.250 --> 00:30:41.250 So this is a picture of nine randomly selected individuals

563 00:30:41.430 --> 00:30:42.670 from our data,

564 00:30:42.670 --> 00:30:45.900 X axis here, follow-up time in days,

565 00:30:45.900 --> 00:30:49.440 Y axis here square root of CD4,

566 00:30:49.440 --> 00:30:53.220 purple line is end of follow-up,

567 00:30:53.220 --> 00:30:56.950 two gray lines here mark one year

568 00:30:56.950 --> 00:30:59.023 and two years post diagnosis.

569 00:31:00.270 --> 00:31:03.950 Empty circles here mean that the patient

570 00:31:03.950 --> 00:31:06.010 has not been treated.

571 00:31:06.010 --> 00:31:09.310 Solid circles, mean that they're on the treatment.

572 00:31:09.310 --> 00:31:11.920 So we can see that there's a lot of variability

573 00:31:11.920 --> 00:31:15.940 in terms of the treatment initiation time.

574 00:31:15.940 --> 00:31:19.620 And some people are followed much longer

575 00:31:19.620 --> 00:31:22.290 than some other patients.

576 00:31:22.290 --> 00:31:27.290 And the follow-up time is pretty irregularly spaced

577 00:31:29.370 --> 00:31:33.880 and overall the CD4 measurements are quite sparse,

578 00:31:33.880 --> 00:31:36.440 and there's also incomplete information

579 00:31:36.440 --> 00:31:41.440 for example, these two they either died

580 00:31:41.490 --> 00:31:43.830 or were lost to follow up

581 00:31:43.830 --> 00:31:47.410 before they even got a chance to be treated.

582 00:31:47.410 --> 00:31:51.382 So there's also a lot of complication in the data.

583 00:31:51.382 --> 00:31:53.640 There's a continuous time measurement

584 00:31:53.640 --> 00:31:55.380 of the treatment initiation.

585 00:31:55.380 --> 00:31:57.725 It just happens all over the place.

586 00:31:57.725 --> 00:32:02.600 The longitudinal outcome of interest are sparsely measured,

587 00:32:02.600 --> 00:32:04.710 leading to incomplete data.

588 00:32:04.710 --> 00:32:08.570 There's also a censoring due to dropout or deaths.

589 00:32:08.570 --> 00:32:11.410 So our general solution is that we'll use weighting

590 00:32:11.410 --> 00:32:13.800 to handle time-varying confounding.

591 00:32:13.800 --> 00:32:16.830 And will show how to derive a continuous time versions

592 00:32:16.830 --> 00:32:18.820 of the weights.

593 00:32:18.820 --> 00:32:21.400 For the missing outcomes

594 00:32:21.400 --> 00:32:24.470 that is caused by sparse measurement and censoring

595 00:32:24.470 --> 00:32:27.920 we'll use imputations from a model of the joint distribution

596 00:32:27.920 --> 00:32:30.150 of CD4 and mortality.

597 00:32:30.150 --> 00:32:33.200 And because we're interested in both mortality status

598 00:32:33.200 --> 00:32:36.363 and CD4, we'll develop a composite outcome.

599 00:32:37.970 --> 00:32:42.460 So our general approach is to emulate a randomized trial

600 00:32:42.460 --> 00:32:45.000 in which we would randomize individuals

601 00:32:45.000 --> 00:32:47.900 to follow specific DTR Q.

602 00:32:47.900 --> 00:32:50.950 And Q equals zero means never treated,

603 00:32:50.950 --> 00:32:54.170 because CD4 can never drop below zero.

604 00:32:54.170 --> 00:32:57.730 Now, Q equals infinity means treat immediately.

605 00:32:57.730 --> 00:32:59.340 So after randomization,

606 00:32:59.340 --> 00:33:02.460 all the individuals will be followed

607 00:33:02.460 --> 00:33:04.890 for a fixed amount of time,

608 00:33:04.890 --> 00:33:07.160 at which point, say T star,

609 00:33:07.160 --> 00:33:09.830 both their mortality status.

610 00:33:09.830 --> 00:33:14.110 And among those who are alive at T star,

611 00:33:14.110 --> 00:33:18.230 their CD4 count will be assessed.

612 00:33:18.230 --> 00:33:21.240 So what define a composite outcome XQ,

613 00:33:21.240 --> 00:33:25.120 that is the product of the test indicator

614 00:33:25.120 --> 00:33:26.643 and the potential CD4.

615 00:33:27.770 --> 00:33:31.600 So the cumulative distribution of this composite outcome

616 00:33:31.600 --> 00:33:35.610 is a useful measure of treatment utility,

617 00:33:35.610 --> 00:33:38.580 because it has appointments at zero

618 00:33:38.580 --> 00:33:40.850 corresponding to mortality rate.

619 00:33:40.850 --> 00:33:45.310 Thereby capturing both mortality status

620 00:33:45.310 --> 00:33:50.310 and CD4 count among survivors at T star.

621 00:33:51.610 --> 00:33:55.720 So for example, the probability of a positive XQ,

622 00:33:55.720 --> 00:33:57.980 that's the survival fraction,

623 00:33:57.980 --> 00:34:01.510 and the probability of XQ greater than X,

624 00:34:01.510 --> 00:34:06.053 that's the fraction of survivors with CD4 above X.

625 00:34:07.560 --> 00:34:11.960 Okay, so similar to the first motivating example,

626 00:34:11.960 --> 00:34:14.653 we again have three timed events.

627 00:34:15.650 --> 00:34:19.170 Death time, censoring time, treatment initiation time.

628 00:34:19.170 --> 00:34:22.590 And now we have a tailoring variable, CD4 count.

629 00:34:22.590 --> 00:34:26.930 So the CD four process is defined for all continuous time,

630 00:34:26.930 --> 00:34:30.310 but it's just measured at discrete times.

631 00:34:30.310 --> 00:34:33.863 And we also have a P by one covariate process.

632 00:34:34.940 --> 00:34:37.760 Using a convention in the DTR literature,

633 00:34:37.760 --> 00:34:40.288 we assume that the treatment decision

634 00:34:40.288 --> 00:34:44.980 is always made after observing the covariate history

635 00:34:44.980 --> 00:34:48.419 and the CD4 count.

636 00:34:48.419 --> 00:34:50.140 Putting everything together,

637 00:34:50.140 --> 00:34:55.140 we have a history information indicator.

638 00:34:55.140 --> 00:34:59.510 For each individual, we'll have a observed a data process.

639 00:34:59.510 --> 00:35:01.700 And just note that each person

640 00:35:01.700 --> 00:35:03.770 can have a different lens of followup

641 00:35:03.770 --> 00:35:05.173 at different time points.

642 00:35:08.100 --> 00:35:11.800 Our goal is to evaluate the effect of DTRs,

643 00:35:11.800 --> 00:35:14.410 but we're dealing with observational data,

644 00:35:14.410 --> 00:35:17.630 so we'll have to map the observed treatment regimen

645 00:35:17.630 --> 00:35:21.927 to specific DTRs that we are interested in evaluating.

646 00:35:21.927 --> 00:35:26.927 Essentially we'll follow the deterministic function

647 00:35:28.090 --> 00:35:29.570 to create the mapping.

648 00:35:29.570 --> 00:35:31.686 Essentially there are three rules.

649 00:35:31.686 --> 00:35:34.416 First rule says not to treat the person

650 00:35:34.416 --> 00:35:37.750 if the person has not yet initiated treatment

651 00:35:37.750 --> 00:35:40.870 and their CD4 has not fallen below Q,

652 00:35:40.870 --> 00:35:42.403 or has not been observed.

653 00:35:43.710 --> 00:35:47.080 Second rule says, treat this person if their time T,

654 00:35:47.080 --> 00:35:51.170 CD4 has fallen below Q for the very first time.

655 00:35:51.170 --> 00:35:53.920 Once treated, always treat them.

656 00:35:53.920 --> 00:35:55.510 Following these three rules,

657 00:35:55.510 --> 00:35:59.880 we'll be able to create a regimen specific compliant process

658 00:35:59.880 --> 00:36:01.890 for each individual in the data.

659 00:36:01.890 --> 00:36:05.010 So essentially if the rule says treat,

660 00:36:05.010 --> 00:36:08.640 and if the person is actually treated by the time T,

661 00:36:08.640 --> 00:36:12.520 then this person is compliant at time T.

662 00:36:12.520 --> 00:36:14.260 If the rule says do not treat,

663 00:36:14.260 --> 00:36:16.610 and the person was not treated at the time T,

664 00:36:16.610 --> 00:36:18.963 so this person is still compliant to the rule.

665 00:36:20.090 --> 00:36:23.590 And so we'll be able to observe a compliant process

666 00:36:23.590 --> 00:36:25.101 for each person.

667 00:36:25.101 --> 00:36:29.430 Here a simple example to show you how to create the mapping.

668 00:36:29.430 --> 00:36:33.307 For example, we're interested in Q equals 350.

669 00:36:33.307 --> 00:36:35.440 This person came in at baseline,

670 00:36:35.440 --> 00:36:38.620 had a measurement 400 above the threshold.

671 00:36:38.620 --> 00:36:40.000 The rule says do not treat,

672 00:36:40.000 --> 00:36:41.450 the person was not treated.

673 00:36:41.450 --> 00:36:44.010 At this point, it's compliant with the rule.

674 00:36:44.010 --> 00:36:48.090 Next visit, no new CD4 observation.

675 00:36:48.090 --> 00:36:49.590 So the rule says do not treat,

676 00:36:49.590 --> 00:36:51.695 the person's still not treated,

677 00:36:51.695 --> 00:36:53.030 still compliant at this point.

678 00:36:53.030 --> 00:36:57.640 Third visit, the person's CD4 drops to 330,

679 00:36:57.640 --> 00:37:01.490 which is below the threshold for the very first time,

680 00:37:01.490 --> 00:37:04.540 the rules are start treating this person,

681 00:37:04.540 --> 00:37:06.370 the person was actually treated.

682 00:37:06.370 --> 00:37:09.610 So compliant at this point.

683 00:37:09.610 --> 00:37:14.370 Next visit the rule says once treated always treat them,

684 00:37:14.370 --> 00:37:16.390 the person kept being treated.

685 00:37:16.390 --> 00:37:19.820 So this person was compliant with the rule 350

686 00:37:19.820 --> 00:37:22.453 all throughout his or her followup.

687 00:37:23.350 --> 00:37:27.410 Next example, the first two rows are the same.

688 00:37:27.410 --> 00:37:32.410 The third visit, the person's CD4 jumps to 450,

689 00:37:32.900 --> 00:37:34.990 which is above the threshold.

690 00:37:34.990 --> 00:37:36.520 The rule says do not treat,

691 00:37:36.520 --> 00:37:40.450 but on the contrary, the person was actually treated

692 00:37:40.450 --> 00:37:42.480 and kept being treated.

693 00:37:42.480 --> 00:37:45.760 So from this time point onward,

694 00:37:45.760 --> 00:37:49.083 the person was not compliant with this rule.

695 00:37:50.970 --> 00:37:54.553 Okay, so that's just some simple example

696 00:37:54.553 --> 00:37:58.480 to show how to create the mapping.

697 00:37:58.480 --> 00:38:01.240 With missing outcomes for those alive

698 00:38:01.240 --> 00:38:04.660 at the target measurement time T star,

699 00:38:04.660 --> 00:38:09.660 the observed outcome XI is the CD4 measurement at T star.

700 00:38:10.800 --> 00:38:14.000 But because of CD4 is sparsely measured

701 00:38:14.000 --> 00:38:16.540 and irregularly spaced,

702 00:38:16.540 --> 00:38:18.856 Z of T star is directly observed

703 00:38:18.856 --> 00:38:23.856 only when the person's followup time is exactly at T star.

704 00:38:24.810 --> 00:38:27.250 So in this case, it is pretty common

705 00:38:27.250 --> 00:38:32.250 to predefine a interval and capture the CD4 that is measured

706 00:38:37.080 --> 00:38:40.920 at the time closest to the target measurement time.

707 00:38:40.920 --> 00:38:42.880 But even using this strategy,

708 00:38:42.880 --> 00:38:47.880 there's still a possibility that there is no measurement

709 00:38:48.374 --> 00:38:51.310 in predefined interval.

710 00:38:51.310 --> 00:38:53.990 Then we say this person has a missing outcome.

711 00:38:53.990 --> 00:38:56.690 And it's also possible that the person dropped out

712 00:38:56.690 --> 00:38:57.523 before TA.

713 00:38:58.970 --> 00:39:02.483 And so in this case, the outcome is also missing.

714 00:39:03.815 --> 00:39:07.940 For these missing outcomes, our general strategy

715 00:39:07.940 --> 00:39:10.120 is to use multiple imputation.

716 00:39:10.120 --> 00:39:12.280 So we would specify and fit model

717 00:39:12.280 --> 00:39:15.630 for the joint distribution of the CD4 process

718 00:39:15.630 --> 00:39:17.720 and the mortality process.

719 00:39:17.720 --> 00:39:20.300 For those known to be alive,

720 00:39:20.300 --> 00:39:22.590 but without a CD4 measurement,

721 00:39:22.590 --> 00:39:27.590 we would impute the CD4 count from the fitted CD4 sub-model.

722 00:39:28.460 --> 00:39:30.840 And for those missing the CD4,

723 00:39:30.840 --> 00:39:32.820 because of right censoring,

724 00:39:32.820 --> 00:39:37.110 we would calculate the mortality probability

725 00:39:37.110 --> 00:39:38.890 from the fitted survival sub-model,

726 00:39:38.890 --> 00:39:41.310 and then impute the death indicator

727 00:39:42.350 --> 00:39:43.750 from the Bernoulli distribution

728 00:39:43.750 --> 00:39:46.020 with this calculated probability.

729 00:39:46.020 --> 00:39:48.830 If the death indicator was imputed to be zero,

730 00:39:48.830 --> 00:39:52.200 then we further impute a CD4 count for this person.

731 00:39:52.200 --> 00:39:54.853 Otherwise we'll set X to be zero.

732 00:39:56.220 --> 00:39:58.950 And again, we would have to assume

733 00:39:58.950 --> 00:40:01.850 some standard causal inference assumptions

734 00:40:01.850 --> 00:40:06.050 in order to draw causal effects about the DTRQ

735 00:40:07.570 --> 00:40:09.410 using observational data.

736 00:40:09.410 --> 00:40:13.763 And we can estimate and compare DTRs along a continuum.

737 00:40:14.900 --> 00:40:17.390 We can formulate a causal model

738 00:40:17.390 --> 00:40:22.050 for the smooth effect of Q on the task quantile of XQ.

739 00:40:22.050 --> 00:40:25.380 This is our composite outcome with separate parameters

740 00:40:25.380 --> 00:40:29.030 capturing the effect of treat immediately,

741 00:40:29.030 --> 00:40:31.320 and the effect of never treat.

742 00:40:31.320 --> 00:40:35.360 And then we can parametrize the model using splines of Q

743 00:40:35.360 --> 00:40:39.853 for the third term here, to gain statistical efficiency.

744 00:40:41.360 --> 00:40:46.360 And we can obtain a consistent estimator of effect of Q

745 00:40:46.940 --> 00:40:50.010 by solving the weighted quantile regression

746 00:40:50.010 --> 00:40:51.383 estimating equation.

747 00:40:52.820 --> 00:40:55.193 So what should be the weights?

748 00:40:57.020 --> 00:40:59.410 First, we assume there's no dropout or death

749 00:40:59.410 --> 00:41:01.739 prior to the target measurement time.

750 00:41:01.739 --> 00:41:06.739 In the discrete time setting with common time point,

751 00:41:06.930 --> 00:41:09.900 the form of the weights have already been done.

752 00:41:09.900 --> 00:41:13.570 It has been derived in several papers.

753 00:41:13.570 --> 00:41:16.680 Essentially, the denominator of the weight

754 00:41:16.680 --> 00:41:18.980 is this conditional probability.

755 00:41:18.980 --> 00:41:22.022 It's a conditional probability of the person being compliant

756 00:41:22.022 --> 00:41:27.022 all throughout the follow up, given the covariate history.

757 00:41:29.010 --> 00:41:34.010 So if we have a common set of discrete time points,

758 00:41:34.160 --> 00:41:37.640 it's a cumulative a product of the conditional probability

759 00:41:37.640 --> 00:41:42.150 of this person being compliant at every time point.

760 00:41:42.150 --> 00:41:45.740 And essentially if the rule says treat,

761 00:41:45.740 --> 00:41:48.350 it's a condition of probability of the person

762 00:41:48.350 --> 00:41:51.240 actually being treated at this time point,

763 00:41:51.240 --> 00:41:53.450 if the rule says not treat,

764 00:41:53.450 --> 00:41:56.850 as a conditional probability of this person not treated

765 00:41:56.850 --> 00:41:58.320 by this time point.

766 00:41:58.320 --> 00:42:02.140 So in order to estimate this probability,

767 00:42:02.140 --> 00:42:04.170 we just need to model the observed

768 00:42:04.170 --> 00:42:07.613 treatment initiation process among those regimen compliers,

769 00:42:09.360 --> 00:42:11.550 but this is for discrete time setting.

770 00:42:11.550 --> 00:42:14.380 What would be the continuous time weights?

771 00:42:14.380 --> 00:42:17.900 We note that the occurrence of treatment initiation

772 00:42:17.900 --> 00:42:21.170 in a small time interval T and T plus TD

773 00:42:21.170 --> 00:42:26.170 is actually a Bernoulli trial with outcome DNA of T.

774 00:42:27.690 --> 00:42:31.150 So then we can rewrite this probability,

775 00:42:31.150 --> 00:42:32.950 this probability here,

776 00:42:32.950 --> 00:42:36.850 in the form of individual partial likelihood

777 00:42:36.850 --> 00:42:38.840 for the counting process of A.

778 00:42:40.070 --> 00:42:44.680 And now we note that when DT becomes smaller and smaller,

779 00:42:44.680 --> 00:42:49.640 this finite product approaches a product integral.

780 00:42:49.640 --> 00:42:54.390 So then this finite product can be rewritten

781 00:42:54.390 --> 00:42:58.490 as a final product over jump times of the counting process

782 00:42:58.490 --> 00:43:01.900 for A times the survival function.

783 00:43:01.900 --> 00:43:04.970 And then by recognizing that each individual

784 00:43:04.970 --> 00:43:09.730 had at most one jump at exactly AI.

785 00:43:09.730 --> 00:43:14.563 Now we can further reduce this probability to this form.

786 00:43:15.480 --> 00:43:18.100 Which suggests weighting scheme.

787 00:43:18.100 --> 00:43:22.800 Essentially it says for those who have been treated

788 00:43:22.800 --> 00:43:25.170 by a T star, we would weight them

789 00:43:25.170 --> 00:43:28.370 by the conditional density function of A.

790 00:43:28.370 --> 00:43:32.040 For those who haven't been treated by the time T star,

791 00:43:32.040 --> 00:43:36.140 we would weight them by the survival or function of A.

792 00:43:36.140 --> 00:43:38.760 So if you recall the weighting scheme

793 00:43:38.760 --> 00:43:42.580 for the first motivating example, this is exactly the same,

794 00:43:42.580 --> 00:43:43.893 the same rating scheme,

795 00:43:44.930 --> 00:43:46.790 but we took different approaches.

796 00:43:46.790 --> 00:43:48.270 The first example,

797 00:43:48.270 --> 00:43:50.610 we use a random Aalen derivatives

798 00:43:50.610 --> 00:43:52.350 to derive the weighting scheme.

799 00:43:52.350 --> 00:43:56.187 The second project we derive the limit

800 00:43:57.930 --> 00:44:00.120 of the finite product,

801 00:44:00.120 --> 00:44:02.330 but using different approaches,

802 00:44:02.330 --> 00:44:04.343 we arrive at the same weighting scheme.

803 00:44:05.750 --> 00:44:10.130 And so similarly we modeled the intensity process

804 00:44:10.130 --> 00:44:11.710 of treatment initiation.

805 00:44:11.710 --> 00:44:13.193 We estimate the weights.

806 00:44:15.480 --> 00:44:17.740 So if there was a censoring or death

807 00:44:17.740 --> 00:44:19.810 prior to target measurement time,

808 00:44:19.810 --> 00:44:22.810 we would have to assume once lost to follow up

809 00:44:22.810 --> 00:44:25.000 at a time prior to T star,

810 00:44:25.000 --> 00:44:28.090 the treatment and regimen status remain constant.

811 00:44:28.090 --> 00:44:30.620 And this way we will just estimate the weights

812 00:44:30.620 --> 00:44:35.620 up to a time point CI, and if the person died before T star,

813 00:44:36.530 --> 00:44:39.920 then we would only evaluate compliance

814 00:44:39.920 --> 00:44:43.283 and treatment initiation processes up to time TI.

815 00:44:45.390 --> 00:44:48.590 Okay, so for missing outcomes,

816 00:44:48.590 --> 00:44:51.660 we propose a joint modeling approach.

817 00:44:51.660 --> 00:44:56.150 We specify a two-level model for the observed CD4 process.

818 00:44:56.150 --> 00:44:57.020 The first level,

819 00:44:57.020 --> 00:45:01.690 the observed CD4 process is a true CD4 trajectory

820 00:45:01.690 --> 00:45:03.570 plus some arrow process.

821 00:45:03.570 --> 00:45:04.720 The second level,

822 00:45:04.720 --> 00:45:07.020 we relate the true CD4 trajectory

823 00:45:07.020 --> 00:45:12.020 to baseline characteristics and treatment initiation time,

824 00:45:13.170 --> 00:45:16.440 and some subject specific random effects,

825 00:45:16.440 --> 00:45:18.900 capturing subject-specific deviations

826 00:45:18.900 --> 00:45:22.390 from the mean trajectories.

827 00:45:22.390 --> 00:45:25.610 And now we propose a hazard model for deaths

828 00:45:25.610 --> 00:45:30.270 uses the true CD4 trajectory as a covariate

829 00:45:30.270 --> 00:45:32.970 linking the two processes,

830 00:45:32.970 --> 00:45:37.580 Linking the death process and linking with a CD4 process.

831 00:45:37.580 --> 00:45:39.326 Now we use the joint model

832 00:45:39.326 --> 00:45:42.730 to impute the missing outcomes

833 00:45:42.730 --> 00:45:45.580 and estimate the variance of the target estimator

834 00:45:45.580 --> 00:45:47.580 using Rubin's combination wall.

835 00:45:49.530 --> 00:45:53.650 So we applied this method to the IeDEA dataset.

836 00:45:53.650 --> 00:45:58.650 IeDEA is another HIV consortium based in West Kenya.

837 00:46:00.230 --> 00:46:02.920 So we have almost 2000 data.

838 00:46:02.920 --> 00:46:06.910 We see that the CD4 is pretty sparsely measured

839 00:46:06.910 --> 00:46:11.910 and death rate is low around three and 4%.

840 00:46:11.960 --> 00:46:15.410 Most of patients have been treated by one year

841 00:46:16.480 --> 00:46:20.660 and we have a set of covariates.

842 00:46:20.660 --> 00:46:23.993 Some of them are time varying, some of them are time fixed.

843 00:46:25.710 --> 00:46:29.550 We proposed three target estimators,

844 00:46:29.550 --> 00:46:33.890 so first we're interested in mortality proportion.

845 00:46:33.890 --> 00:46:37.870 We're also interested in the median of the distribution

846 00:46:37.870 --> 00:46:40.631 of the composite outcome XQ.

847 00:46:40.631 --> 00:46:44.600 We also looked at CD4 among survivors,

848 00:46:44.600 --> 00:46:49.160 but this estimator does not have a causal interpretation

849 00:46:49.160 --> 00:46:53.320 because it conditions on having survived two T star.

850 00:46:53.320 --> 00:46:55.350 So it only measures association,

851 00:46:55.350 --> 00:47:00.350 but the first two estimators have causal interpretations.

852 00:47:02.830 --> 00:47:05.060 So we first look at the effectiveness

853 00:47:05.060 --> 00:47:09.960 of five specific regimens for both one year and two years

854 00:47:09.960 --> 00:47:11.698 after diagnosis.

855 00:47:11.698 --> 00:47:16.660 We can see that the immediate treatment initiation

856 00:47:18.270 --> 00:47:21.200 lead to significant lower mortality rate

857 00:47:21.200 --> 00:47:24.050 and significantly higher median values

858 00:47:24.050 --> 00:47:29.050 of the composite alcohol compared to delayed treatment.

859 00:47:29.570 --> 00:47:32.460 And the never treat initiation

860 00:47:32.460 --> 00:47:36.800 will lead to a significantly higher mortality probability.

861 00:47:36.800 --> 00:47:41.440 And for those who do survive to T star,

862 00:47:41.440 --> 00:47:44.340 their CD4 count is higher.

863 00:47:44.340 --> 00:47:48.290 So resulting higher to theta Q2 and higher theta Q3

864 00:47:48.290 --> 00:47:53.290 compared to other delayed treatment regimen.

865 00:47:53.680 --> 00:47:57.430 So this may suggest that those who do survive

866 00:47:57.430 --> 00:47:59.670 to T-star without any treatment,

867 00:47:59.670 --> 00:48:01.900 maybe they are relatively healthier

868 00:48:01.900 --> 00:48:03.593 at the beginning of the followup.

869 00:48:05.290 --> 00:48:10.000 Okay, and then we also plot the dose response curve

870 00:48:10.000 --> 00:48:14.350 of the median value of the composite outcome

871 00:48:14.350 --> 00:48:17.250 versus DTR Q,

872 00:48:17.250 --> 00:48:22.250 also suggests that the immediate treatment

873 00:48:22.320 --> 00:48:26.920 would lead to significantly higher median values of XQ,

874 00:48:26.920 --> 00:48:28.530 and also as illustration

875 00:48:28.530 --> 00:48:30.480 of the gained statistical efficiency

876 00:48:30.480 --> 00:48:35.480 by modeling the smooth effect Q on the quantile of the XQ.

877 00:48:36.950 --> 00:48:39.880 The variance in the one year outcome

878 00:48:39.880 --> 00:48:44.880 associated with Q equals 350, achieved about 15% reduction

879 00:48:45.860 --> 00:48:49.453 compared to that from the regimen specific estimates.

880 00:48:50.970 --> 00:48:55.164 So we gain a bit of our statistical efficiency

881 00:48:55.164 --> 00:48:58.313 by modeling the smooth effect.

882 00:49:00.000 --> 00:49:02.390 So there are several strands of continuous time

883 00:49:02.390 --> 00:49:03.940 marginal structure model.

884 00:49:03.940 --> 00:49:07.850 We see that we can derive, using different approaches,

885 00:49:07.850 --> 00:49:11.598 closed form of weights for continuous-time treatment.

886 00:49:11.598 --> 00:49:16.052 It can handle complex dataset on its own terms

887 00:49:16.052 --> 00:49:20.100 without having to artificially align measurement times,

888 00:49:20.100 --> 00:49:22.943 which could possibly lead to loss of information.

889 00:49:23.840 --> 00:49:26.560 It is amenable to many different outcomes.

890 00:49:26.560 --> 00:49:28.260 We've used the survival outcomes,

891 00:49:28.260 --> 00:49:30.163 we've used composite outcomes.

892 00:49:31.280 --> 00:49:34.340 You can also handle many data complications

893 00:49:34.340 --> 00:49:37.150 introduced by various censoring patterns

894 00:49:37.150 --> 00:49:39.503 within the same marginal structure model.

895 00:49:40.650 --> 00:49:42.748 So these are the strengths,

896 00:49:42.748 --> 00:49:46.150 but there are also limitations with this approach of course.

897 00:49:46.150 --> 00:49:49.970 One notable limitation is extreme ways,

898 00:49:49.970 --> 00:49:53.023 which could possibly lead to unstable estimates.

899 00:49:54.290 --> 00:49:56.840 So how to address this issue,

900 00:49:56.840 --> 00:50:00.450 especially for time varying confounding

901 00:50:00.450 --> 00:50:04.330 with censored outcome, this would be a challenging task,

902 00:50:04.330 --> 00:50:06.150 but if we can solve this issue,

903 00:50:06.150 --> 00:50:09.770 it might be a very important contribution to the field.

904 00:50:09.770 --> 00:50:13.840 So this is something my colleagues and I

905 00:50:13.840 --> 00:50:18.396 have been thinking about and working on for some time.

906 00:50:18.396 --> 00:50:21.980 Another limitation is that we know

907 00:50:21.980 --> 00:50:25.060 that weighting-based estimator is less efficient

908 00:50:25.060 --> 00:50:27.620 than the so-called G methods.

909 00:50:27.620 --> 00:50:30.020 The G computation, G estimation,

910 00:50:30.020 --> 00:50:32.470 and both G methods require integrating

911 00:50:32.470 --> 00:50:35.250 over the space of longitudinal confounders.

912 00:50:35.250 --> 00:50:38.270 So the G methods are computationally

913 00:50:38.270 --> 00:50:40.160 much, much more expensive

914 00:50:40.160 --> 00:50:43.960 than the marginal structure model-based methods.

915 00:50:43.960 --> 00:50:46.603 And as far as I know,

916 00:50:47.849 --> 00:50:50.130 currently there's no continuous time version

917 00:50:50.130 --> 00:50:52.560 of the G computation methods.

918 00:50:52.560 --> 00:50:56.300 Judith Lok has a paper, back in 2008.

919 00:50:56.300 --> 00:50:59.980 She developed theory for continuous time G-estimation,

920 00:50:59.980 --> 00:51:03.470 but I have yet to see a practical implementation

921 00:51:03.470 --> 00:51:04.910 of this method.

922 00:51:04.910 --> 00:51:09.910 So this could be another avenue for future research,

923 00:51:11.000 --> 00:51:14.940 how to increase efficiency of the continuous time

924 00:51:14.940 --> 00:51:16.433 weighting-based methods.

925 00:51:17.650 --> 00:51:20.713 And here's some key references.

926 00:51:21.744 --> 00:51:23.786 Thank you.

927 00:51:23.786 --> 00:51:25.450 - Thank you Liangyuan for this very interesting

928 00:51:25.450 --> 00:51:29.250 and comprehensive presentation.

929 00:51:29.250 --> 00:51:32.460 Let's see if we have any questions from the audience.

930 00:51:32.460 --> 00:51:33.500 If there's any questions,

931 00:51:33.500 --> 00:51:36.450 please feel free to unmute yourself and speak

932 00:51:36.450 --> 00:51:38.393 or type in the chat.

933 00:51:43.010 --> 00:51:45.040 - [Donna] Thanks, it was a very interesting talk.

934 00:51:45.040 --> 00:51:47.300 This is Donna Spiegelman.

935 00:51:47.300 --> 00:51:48.474 - Hi, Donna.

936 00:51:48.474 --> 00:51:49.307 - Yeah, hi.

937 00:51:49.307 --> 00:51:50.910 I was wondering I might've missed it,

938 00:51:50.910 --> 00:51:55.160 but did you say much about estimating the variance?

939 00:51:55.160 --> 00:51:58.260 I see you have (indistinct) around the curve,

940 00:51:58.260 --> 00:52:01.479 so you must derive the variance.

941 00:52:01.479 --> 00:52:03.420 So I'm wondering if you could say a little bit about that

942 00:52:03.420 --> 00:52:04.640 or a little more about that

943 00:52:04.640 --> 00:52:07.350 if I missed what you did say.

944 00:52:07.350 --> 00:52:08.680 - Sure, sure, sure.

945 00:52:08.680 --> 00:52:11.520 So for this one, this is the second example,

946 00:52:11.520 --> 00:52:14.600 for this one we have multiple amputation

947 00:52:14.600 --> 00:52:16.133 and we also have weighting.

948 00:52:20.045 --> 00:52:21.195 So with weighting part,

949 00:52:22.080 --> 00:52:26.160 the variance was estimated using bootstrap

950 00:52:26.160 --> 00:52:29.310 for multiple amputation, and then we combined,

951 00:52:29.310 --> 00:52:32.713 so it's a bootstrap nested within multiple imputation.

952 00:52:32.713 --> 00:52:35.070 So then we use the Rubin's combination role

953 00:52:35.070 --> 00:52:37.093 to estimate the total variance.

954 00:52:38.150 --> 00:52:43.150 For the first example, we actually used a bootstrap,

955 00:52:45.210 --> 00:52:49.760 and the coverage probability was actually okay.

956 00:52:49.760 --> 00:52:51.473 It's good for the estimator.

957 00:52:52.340 --> 00:52:55.233 - Did you think about asymptotic variants derivations?

958 00:52:56.190 --> 00:52:57.023 - I did.

959 00:52:58.170 --> 00:53:00.023 It was a very difficult task,

960 00:53:02.556 --> 00:53:05.200 there's a story about our first paper

961 00:53:05.200 --> 00:53:06.423 found that about it.

962 00:53:09.940 --> 00:53:12.120 It was first submitted to Jaza

963 00:53:12.120 --> 00:53:16.370 and then they asked about the asymptotic variants

964 00:53:16.370 --> 00:53:17.520 about the estimator.

965 00:53:17.520 --> 00:53:21.830 And it's quite complex because they involve the splice

966 00:53:21.830 --> 00:53:25.340 and involves the survival data.

967 00:53:25.340 --> 00:53:28.773 And we have already approved as a consistency,

968 00:53:33.706 --> 00:53:36.539 and it also involves optimization.

969 00:53:38.340 --> 00:53:40.260 So it's just comes to-

970 00:53:40.260 --> 00:53:42.770 - What's the optimization piece.

971 00:53:42.770 --> 00:53:47.060 - Oh, it's the model based optimal treatment initiation time

972 00:53:47.060 --> 00:53:48.780 that will lead to the maximum survival

973 00:53:48.780 --> 00:53:53.780 at predefined time points.

974 00:53:53.930 --> 00:53:57.630 Right, so they are interested in the optimization.

975 00:53:57.630 --> 00:54:00.230 So the inference about the optimized

976 00:54:00.230 --> 00:54:02.410 treatment initiation time.

977 00:54:02.410 --> 00:54:04.320 We did some empirical evidence

978 00:54:04.320 --> 00:54:08.217 for like the largest sample convergence rate,

979 00:54:08.217 --> 00:54:13.217 but we weren't successful at deriving asymptotic variants.

980 00:54:14.450 --> 00:54:18.170 So that's another piece, I think maybe,

981 00:54:18.170 --> 00:54:19.003 I don't know.

982 00:54:19.003 --> 00:54:21.280 We had this discussion among colleagues

983 00:54:21.280 --> 00:54:23.720 and also my advisor at the time,

984 00:54:23.720 --> 00:54:27.840 we just not sure about whether it's worth the effort

985 00:54:27.840 --> 00:54:29.573 to go and do that route.

986 00:54:30.420 --> 00:54:32.220 - It's probably way more complex

987 00:54:32.220 --> 00:54:34.130 than just the usual derivation.

988 00:54:34.130 --> 00:54:36.470 'Cause you do have like two weighting models,

989 00:54:36.470 --> 00:54:39.590 which are also survival models,

990 00:54:39.590 --> 00:54:41.900 and also the derivation that these variances

991 00:54:41.900 --> 00:54:44.440 sometimes can be specific to the choice

992 00:54:44.440 --> 00:54:45.800 of these (indistinct) models.

993 00:54:45.800 --> 00:54:48.590 And so if you have a variance and the cup's model,

994 00:54:48.590 --> 00:54:51.696 it does not apply to other forms of models,

995 00:54:51.696 --> 00:54:54.290 I guess it's really a trade-off right?

996 00:54:54.290 --> 00:54:57.343 - Yeah, it is a trade off.

997 00:54:58.420 --> 00:55:03.330 It's still an open question and nobody had done it yet,

998 00:55:03.330 --> 00:55:06.240 but just, whether you're thinking it's was the effort

999 00:55:06.240 --> 00:55:10.020 just to devote a couple of years to work on that.

1000 00:55:10.020 --> 00:55:14.780 - So was bootstrap time consuming for these datasets,

1001 00:55:14.780 --> 00:55:18.211 for this data analysis, or they're pretty manageable.

1002 00:55:18.211 --> 00:55:19.780 - They're pretty manageable.

1003 00:55:19.780 --> 00:55:23.070 And it looks complicated because we have to weight everybody

1004 00:55:23.070 --> 00:55:24.110 that had event.

1005 00:55:24.110 --> 00:55:27.260 We also have to weight everywhere in the risk set

1006 00:55:27.260 --> 00:55:28.350 at any time point.

1007 00:55:28.350 --> 00:55:31.710 So it looks pretty complex, but still manageable.

1008 00:55:34.840 --> 00:55:38.000 Another reason is because we use parametric models.

1009 00:55:38.000 --> 00:55:41.220 If we wanted to,

1010 00:55:41.220 --> 00:55:45.810 I'm not aware of any machine learning algorithm

1011 00:55:45.810 --> 00:55:48.240 that can handle survival data,

1012 00:55:48.240 --> 00:55:50.543 but also with time varying covariates,

1013 00:55:51.930 --> 00:55:54.260 that's something I'm also thinking about.

1014 00:55:54.260 --> 00:55:56.210 Like, if we use those algorithm

1015 00:55:56.210 --> 00:55:58.550 might be more time consuming,

1016 00:55:58.550 --> 00:56:02.640 but with just a parametric models, it's pretty manageable.

1017 00:56:02.640 --> 00:56:03.540 - And when you're bootstrapped,

1018 00:56:03.540 --> 00:56:05.810 you go back to the weight models

1019 00:56:05.810 --> 00:56:08.340 and refit the weight models every time?

1020 00:56:08.340 --> 00:56:09.615 - Yeah.

1021 00:56:09.615 --> 00:56:12.430 - But the variable is pre-determined.

1022 00:56:12.430 --> 00:56:14.940 So that's what you mentioned, machine learning.

1023 00:56:14.940 --> 00:56:17.120 So the variables are predetermined

1024 00:56:17.120 --> 00:56:19.080 and they're functional forms in the model,

1025 00:56:19.080 --> 00:56:21.960 but the coefficients that correspond to them

1026 00:56:21.960 --> 00:56:24.400 are re estimated for each bootstrap.

1027 00:56:24.400 --> 00:56:25.720 - Very estimated.

1028 00:56:25.720 --> 00:56:27.046 Right, right, right.

1029 00:56:27.046 --> 00:56:27.879 Exactly.

1030 00:56:27.879 --> 00:56:28.712 Yeah.

1031 00:56:28.712 --> 00:56:30.090 - Great question.

1032 00:56:30.090 --> 00:56:31.070 - Yeah.

1033 00:56:31.070 --> 00:56:33.063 So a lot of open questions still.

1034 00:56:34.900 --> 00:56:38.083 - So any other questions from the audience?

1035 00:56:41.290 --> 00:56:42.623 - I have another comment.

1036 00:56:43.900 --> 00:56:46.810 So by getting back to this,

1037 00:56:46.810 --> 00:56:48.870 that you re estimated the coefficients

1038 00:56:48.870 --> 00:56:50.440 for the weight models.

1039 00:56:50.440 --> 00:56:54.053 So in sort of the standard marginal structural model,

1040 00:56:54.053 --> 00:56:58.800 the variability due to those weight models is ignored.

1041 00:56:58.800 --> 00:57:00.780 And the robust variance is used

1042 00:57:00.780 --> 00:57:02.430 and said to be an overestimate,

1043 00:57:02.430 --> 00:57:06.350 implying that if you took that variation into account,

1044 00:57:06.350 --> 00:57:08.450 you'd get a smaller variance

1045 00:57:08.450 --> 00:57:11.290 and you might see the same thing here with your bootstraps.

1046 00:57:11.290 --> 00:57:14.508 If you took the weight models as fixed,

1047 00:57:14.508 --> 00:57:17.990 you might find that you have a less efficient estimator,

1048 00:57:17.990 --> 00:57:19.760 which is kind of interesting

1049 00:57:19.760 --> 00:57:23.150 just in terms of say a methods paper to show,

1050 00:57:23.150 --> 00:57:26.060 because there's different ways to do bootstraps,

1051 00:57:26.060 --> 00:57:30.120 but here you're automatically taking the estimation

1052 00:57:30.120 --> 00:57:31.680 of the weight models into account,

1053 00:57:31.680 --> 00:57:35.300 which is not saying that say the classic paper

1054 00:57:35.300 --> 00:57:38.114 by Hernan in epidemiology,

1055 00:57:38.114 --> 00:57:42.310 that's ignored and the robust variance is recommended.

1056 00:57:42.310 --> 00:57:43.460 - Hmm.

1057 00:57:43.460 --> 00:57:46.090 It's a very great comment.

1058 00:57:46.090 --> 00:57:47.480 Something I have to think about.

1059 00:57:47.480 --> 00:57:51.860 So you're saying that in each bootstrap,

1060 00:57:51.860 --> 00:57:54.810 when we estimate the weight model, we fix the weight model.

1061 00:57:56.903 --> 00:57:59.763 So the coefficients from the weight model stay fixed-

1062 00:58:00.850 --> 00:58:02.810 - Yeah, so you don't even do a bootstrap for that.

1063 00:58:02.810 --> 00:58:06.465 You basically hold the weight model as a constant,

1064 00:58:06.465 --> 00:58:07.298 and then you'd-

1065 00:58:07.298 --> 00:58:08.570 - Robust variance.

1066 00:58:08.570 --> 00:58:10.950 - Yeah, you use the robust variance,

1067 00:58:10.950 --> 00:58:12.690 which I guess it's a little tricky

1068 00:58:12.690 --> 00:58:14.520 because now you don't have the robust variance

1069 00:58:14.520 --> 00:58:15.776 because you're not using it,

1070 00:58:15.776 --> 00:58:20.776 but it seems the bootstrap analog of the approach taken

1071 00:58:20.870 --> 00:58:24.230 would be to just fit the weight model once,

1072 00:58:24.230 --> 00:58:26.870 treat that fixed unknown,

1073 00:58:26.870 --> 00:58:31.190 and then only bootstrap on the outcome model.

1074 00:58:31.190 --> 00:58:32.260 - Right, right.

1075 00:58:32.260 --> 00:58:33.093 Yeah.

1076 00:58:33.093 --> 00:58:34.618 - [Fan Li] Totally. Yeah.

1077 00:58:34.618 --> 00:58:36.170 - Interesting.

1078 00:58:36.170 --> 00:58:37.753 Take that in as a note.

1079 00:58:39.300 --> 00:58:41.590 - So I do have a question as well.

1080 00:58:41.590 --> 00:58:44.200 I think Liangyuan you had presented two applications

1081 00:58:44.200 --> 00:58:46.970 at the HIV observational studies,

1082 00:58:46.970 --> 00:58:51.150 do you see the application that these new methods

1083 00:58:51.150 --> 00:58:53.580 to other areas as well

1084 00:58:54.450 --> 00:58:57.060 to solve the other questions? - Yeah.

1085 00:58:57.060 --> 00:59:01.770 Yeah, actually this is not pertaining to HIV area.

1086 00:59:01.770 --> 00:59:06.140 It's actually in the public health areas.

1087 00:59:06.140 --> 00:59:09.623 A lot of questions are involving

1088 00:59:13.430 --> 00:59:15.580 this statistical formulation.

1089 00:59:15.580 --> 00:59:16.693 So for example,

1090 00:59:17.600 --> 00:59:22.600 I've been collaborating with an epidemiologist at Columbia.

1091 00:59:22.623 --> 00:59:26.540 They are doing cardiovascular research.

1092 00:59:26.540 --> 00:59:29.223 So one research question is that,

1093 00:59:30.930 --> 00:59:35.930 I think it's blood pressure lowering intervention.

1094 00:59:36.690 --> 00:59:40.470 So blood lowering innovation is very useful

1095 00:59:40.470 --> 00:59:42.623 for preventing cardiovascular diseases,

1096 00:59:45.460 --> 00:59:46.850 but they don't know.

1097 00:59:46.850 --> 00:59:49.630 And there also a lack of randomized control trials.

1098 00:59:49.630 --> 00:59:52.920 What is the optimal threshold

1099 00:59:52.920 --> 00:59:57.200 to start giving the blood lowering treatment?

1100 00:59:57.200 --> 00:59:59.100 So this is exactly the same form

1101 00:59:59.100 --> 01:00:01.410 as our second motivating example.

1102 01:00:01.410 --> 01:00:04.350 Like what is the optimal CD4 threshold

1103 01:00:04.350 --> 01:00:06.250 to start the HIV treatment?

1104 01:00:06.250 --> 01:00:08.860 And their question is what is the optimal threshold

1105 01:00:08.860 --> 01:00:12.650 to start the blood lowering treatment?

1106 01:00:12.650 --> 01:00:17.070 So I think there's a lot of possibility

1107 01:00:19.780 --> 01:00:21.890 as to apply these kinds of methods

1108 01:00:21.890 --> 01:00:24.240 in other health research area.

1109 01:00:24.240 --> 01:00:26.320 - Yeah, it's a huge controversy

1110 01:00:26.320 --> 01:00:28.310 in terms of the treatment of hypertension,

1111 01:00:28.310 --> 01:00:30.520 what's the optimal blood pressure

1112 01:00:30.520 --> 01:00:33.147 to start antihypertensives.

1113 01:00:33.147 --> 01:00:35.310 And I think there was a very large trial

1114 01:00:35.310 --> 01:00:37.740 that showed that it was better to start it

1115 01:00:37.740 --> 01:00:42.120 at a much earlier threshold than what current practices.

1116 01:00:42.120 --> 01:00:46.710 And it's very troublesome for people around the world

1117 01:00:46.710 --> 01:00:49.170 because these medicines are expensive.

1118 01:00:49.170 --> 01:00:50.780 And if you see now,

1119 01:00:50.780 --> 01:00:54.060 like another like 40% of the population

1120 01:00:54.060 --> 01:00:58.180 should now be initiated a antihypertensive medication,

1121 01:00:58.180 --> 01:01:01.090 well, most countries can't even afford that.

1122 01:01:01.090 --> 01:01:04.730 So the implications of these different thresholds

1123 01:01:04.730 --> 01:01:08.630 is a very big topic of sort of substantive research

1124 01:01:08.630 --> 01:01:10.240 and debate right now.

1125 01:01:10.240 --> 01:01:11.720 - Well, that's great to know,

1126 01:01:11.720 --> 01:01:14.365 there's urgent need for that.

1127 01:01:14.365 --> 01:01:16.090 (indistinct)

1128 01:01:16.090 --> 01:01:17.040 - Totally.

1129 01:01:17.040 --> 01:01:19.420 All right, I think we are at the hour,

1130 01:01:19.420 --> 01:01:24.420 so thanks Liangyuan again for your great presentation

1131 01:01:25.020 --> 01:01:27.600 and if the audience has any questions,

1132 01:01:27.600 --> 01:01:30.610 I'm sure Liangyuan is happy to take any questions offline

1133 01:01:30.610 --> 01:01:31.623 by emails.

1134 01:01:32.570 --> 01:01:37.570 And I think this is the final seminar of our fall series,

1135 01:01:37.710 --> 01:01:40.430 and I hope to see everyone next spring,

1136 01:01:40.430 --> 01:01:42.040 have a good holiday.

1137 01:01:42.040 --> 01:01:43.000 Thank you.

1138 01:01:43.000 --> 01:01:44.140 - Thank you.

1139 01:01:44.140 --> 01:01:45.050 - Bye. - Bye.