WEBVTT

1 00:00:00.470 --> 00:00:01.460 <v ->Lets get started</v>

2 00:00:01.460 --> 00:00:03.310 and thank you everyone for coming today.

3 00:00:03.310 --> 00:00:06.520 And this is will be your final seminar

4 00:00:06.520 --> 00:00:09.330 for this semester for the (indistinct) the house seminar.

5 00:00:09.330 --> 00:00:11.230 And we are very, very pleasant

6 00:00:11.230 --> 00:00:14.683 to have very our own affiliate faculty,

7 00:00:15.740 --> 00:00:18.930 Dr. Josh Warren joining us.

8 00:00:18.930 --> 00:00:21.330 Dr. Warren is a associate professor

9 00:00:21.330 --> 00:00:23.950 at the Biostatistics Department here,

10 00:00:23.950 --> 00:00:27.870 and his research focuses on statistical method

11 00:00:27.870 --> 00:00:30.260 in public health with the emphasis

12 00:00:30.260 --> 00:00:32.397 on environmental health programs,

13 00:00:32.397 --> 00:00:35.690 and much of his work involves introducing spatial

14 00:00:35.690 --> 00:00:38.780 and spatial temporal models in the basin setting

15 00:00:38.780 --> 00:00:40.640 to learn about the association

16 00:00:40.640 --> 00:00:42.490 between environmental exposures,

17 00:00:42.490 --> 00:00:45.640 such as air pollution and various health outcomes,

18 00:00:45.640 --> 00:00:49.550 including the stillbirth that we are here today.

19 00:00:49.550 --> 00:00:52.260 He's also interested in applying and developing

20 00:00:52.260 --> 00:00:56.230 some spatial temper models in collaborative settings,

21 00:00:56.230 --> 00:00:58.480 such as the infectious disease

22 00:00:58.480 --> 00:01:01.570 we been considered during the COVID pandemic.

23 00:01:01.570 --> 00:01:03.820 So without further ado, Josh,

24 00:01:03.820 --> 00:01:05.320 the floor is yours, thank you.

25 00:01:06.410 --> 00:01:08.470 <v ->Thank thank you Kai for the introduction.</v>

26 00:01:08.470 --> 00:01:10.350 Can everyone hear me?

27 00:01:10.350 --> 00:01:11.183 <v Kai>Yes.</v>

28 00:01:11.183 --> 00:01:12.070 <v ->All right, perfect.</v>

29 00:01:13.531 --> 00:01:15.350 And thanks to Kai for the invitation

30 00:01:15.350 --> 00:01:17.110 and Mulholland for setting all of this up

31 00:01:17.110 --> 00:01:19.430 and allowing me to do this virtually.

32 00:01:19.430 --> 00:01:22.280 It's nice to be here talking about something

33 00:01:22.280 --> 00:01:23.190 other than COVID.

34 00:01:23.190 --> 00:01:25.530 And I guess more recently in my past,

35 00:01:25.530 --> 00:01:27.860 I've been doing a lot of infectious disease work,

36 00:01:27.860 --> 00:01:30.160 so it's kind of nice to be back into something

37 00:01:30.160 --> 00:01:31.830 that I'm still passionate about

38 00:01:31.830 --> 00:01:33.580 and still working heavily on.

39 00:01:33.580 --> 00:01:35.620 And so hopefully some of this today

40 00:01:35.620 --> 00:01:37.670 will be a little bit of review of what we've done

41 00:01:37.670 --> 00:01:39.870 and really current project

42 00:01:39.870 --> 00:01:42.620 that we've just completed and published,

43 00:01:42.620 --> 00:01:45.180 but hopefully there are some elements in here

44 00:01:45.180 --> 00:01:48.410 that you can find overlap within your own work.

45 00:01:48.410 --> 00:01:49.620 And so if you have,

46 00:01:49.620 --> 00:01:51.730 if you see something that brings a bell,

47 00:01:51.730 --> 00:01:54.150 just please reach out and we can kind of talk.

48 00:01:54.150 --> 00:01:55.560 My goal and all of this work

49 00:01:55.560 --> 00:01:58.730 is to kind of develop user friendly methods

50 00:01:58.730 --> 00:02:01.080 that are useful for people outside

51 00:02:01.080 --> 00:02:02.440 of statistics and biostatistics.

52 00:02:02.440 --> 00:02:05.880 So the EPI community and at large usually.

53 00:02:05.880 --> 00:02:08.030 So, yeah, just feel free to reach out afterwards,

54 00:02:08.030 --> 00:02:10.200 and I can share more information,

55 00:02:10.200 --> 00:02:12.080 but today we're gonna be talking about

56 00:02:12.080 --> 00:02:14.980 critical window variable selection for mixtures

57 00:02:14.980 --> 00:02:17.470 and particularly air pollution and stillbirth.

58 00:02:17.470 --> 00:02:19.573 So we'll go ahead and jump into it.

59 00:02:21.000 --> 00:02:23.927 I think probably most people here will know air pollution,

60 00:02:23.927 --> 00:02:25.240 reproductive outcomes.

61 00:02:25.240 --> 00:02:28.660 There's a pretty substantial literature at this point

62 00:02:28.660 --> 00:02:32.020 that suggests exposure to ambient air pollution

63 00:02:32.020 --> 00:02:33.400 during pregnancies associated

64 00:02:33.400 --> 00:02:35.100 with a number of adverse birth outcomes,

65 00:02:35.100 --> 00:02:37.840 including preterm pregnancy, low birth weight,

66 00:02:37.840 --> 00:02:41.560 congenital heart defects, stillbirth, and others.

67 00:02:41.560 --> 00:02:42.710 These are some of the main ones.

68 00:02:42.710 --> 00:02:44.580 Stillbirth is a more recently

69 00:02:44.580 --> 00:02:47.800 kind of emerging outcome of study.

70 00:02:47.800 --> 00:02:49.090 Traditionally, it's been pre-term birth

71 00:02:49.090 --> 00:02:52.180 and low birth weight have gotten a lot of attention,

72 00:02:52.180 --> 00:02:54.690 but these associations are stable robust,

73 00:02:54.690 --> 00:02:56.860 and have been observed across a number of different study

74 00:02:56.860 --> 00:02:58.900 settings, designs, pollutants

75 00:02:58.900 --> 00:03:00.700 and there are a number of good review papers.

76 00:03:00.700 --> 00:03:03.070 If you're interested in a lot of the EPI literature

77 00:03:03.070 --> 00:03:03.903 on this topic,

78 00:03:06.100 --> 00:03:08.700 I would kind of summarize previous a number

79 00:03:08.700 --> 00:03:10.040 of the previous EPI studies,

80 00:03:10.040 --> 00:03:13.970 but as they like to use pollution exposures

81 00:03:13.970 --> 00:03:17.760 that are summarized kind of A priorities,

82 00:03:17.760 --> 00:03:19.570 so they wanna focus on a trimester,

83 00:03:19.570 --> 00:03:21.800 they wanna focus on the entire pregnancy,

84 00:03:21.800 --> 00:03:24.620 like, what is the exposure across the entire pregnancy?

85 00:03:24.620 --> 00:03:27.400 What impact does that have with respect to this outcome?

86 00:03:27.400 --> 00:03:30.490 So these are usually pre-specified averaging periods

87 00:03:30.490 --> 00:03:32.310 and they're explored separately

88 00:03:33.350 --> 00:03:35.310 in these different usually kind

89 00:03:35.310 --> 00:03:37.830 of traditional statistical models like logistic regression

90 00:03:37.830 --> 00:03:40.660 or (indistinct) if you're using some kind of count model.

91 00:03:40.660 --> 00:03:42.980 And so lots of different pollutants

92 00:03:42.980 --> 00:03:44.490 are floating around in these analyses,

93 00:03:44.490 --> 00:03:46.240 lots of different averaging periods

94 00:03:46.240 --> 00:03:50.030 in terms of the exposure, relevance exposure period.

95 00:03:50.030 --> 00:03:51.550 Luckily working with pregnancy,

96 00:03:51.550 --> 00:03:56.470 we have a relatively stable idea

97 00:03:56.470 --> 00:04:00.103 of when exposure potentially affects the fetus.

98 00:04:02.405 --> 00:04:05.070 So lots of models floating around lots of pollutants

99 00:04:05.070 --> 00:04:06.830 and exposure weeks,

100 00:04:06.830 --> 00:04:08.520 but this method is inefficient

101 00:04:08.520 --> 00:04:11.340 and doesn't allow for a joint identification

102 00:04:11.340 --> 00:04:13.110 of more kind of specific periods

103 00:04:13.110 --> 00:04:16.290 across the entire pregnancy in a continuous manner.

104 00:04:16.290 --> 00:04:19.160 So more recently there has been a focus on

105 00:04:19.160 --> 00:04:21.700 critical window estimation and identification.

106 00:04:21.700 --> 00:04:25.430 So this is where I have done quite a bit of work, I think,

107 00:04:25.430 --> 00:04:26.880 in this world.

108 00:04:26.880 --> 00:04:28.790 And then even more recently, I would say,

109 00:04:28.790 --> 00:04:31.120 and I know a number of people I work with even here

110 00:04:31.120 --> 00:04:35.010 at Yale pollution mixers are becoming a really big deal.

111 00:04:35.010 --> 00:04:36.330 So in this talk,

112 00:04:36.330 --> 00:04:37.890 we're trying to combine both of these things,

113 00:04:37.890 --> 00:04:39.000 things that we know really well

114 00:04:39.000 --> 00:04:40.240 or that my group knows really well,

115 00:04:40.240 --> 00:04:42.430 critical windows, estimation identification,

116 00:04:42.430 --> 00:04:43.480 and then pollution mixers,

117 00:04:43.480 --> 00:04:46.883 things that we're getting into more and more it seems.

118 00:04:47.900 --> 00:04:50.010 So starting with critical windows of exposure

119 00:04:50.010 --> 00:04:51.580 and exactly what am I talking about

120 00:04:51.580 --> 00:04:54.690 when I'm talking about critical windows?

121 00:04:54.690 --> 00:04:56.760 So there's an increasing interest in identifying

122 00:04:56.760 --> 00:04:59.560 more specific periods of increased vulnerability.

123 00:04:59.560 --> 00:05:01.210 Usually we're thinking about pregnancy,

124 00:05:01.210 --> 00:05:04.140 but this can go for any really health outcome

125 00:05:04.140 --> 00:05:05.520 that you're interested in,

126 00:05:05.520 --> 00:05:08.030 but more vulnerable periods of the pregnancy

127 00:05:08.030 --> 00:05:09.750 to environmental exposures

128 00:05:09.750 --> 00:05:12.370 and doing this within a single modeling framework.

129 00:05:12.370 --> 00:05:13.980 So estimation of these effects,

130 00:05:13.980 --> 00:05:15.180 we're calling critical windows

131 00:05:15.180 --> 00:05:17.260 or windows of susceptibility.

132 00:05:17.260 --> 00:05:20.710 The NIHS included this identification of critical windows

133 00:05:20.710 --> 00:05:24.250 as a part of its strategic goals back in 2012.

134 00:05:24.250 --> 00:05:27.140 And the focus has remained since then.

135 00:05:27.140 --> 00:05:30.880 So understanding like specific timing of exposure

136 00:05:30.880 --> 00:05:32.520 with respect to outcome development

137 00:05:32.520 --> 00:05:35.200 has a number of features but importantly,

138 00:05:35.200 --> 00:05:37.890 it could lead to improve mechanistic explanations

139 00:05:37.890 --> 00:05:39.500 of disease development,

140 00:05:39.500 --> 00:05:41.930 and ultimately focus guidelines for protection

141 00:05:41.930 --> 00:05:43.893 of the unborn child.

142 00:05:45.340 --> 00:05:46.330 So we have, like I mentioned,

143 00:05:46.330 --> 00:05:48.553 we've done a lot of methods work here,

144 00:05:49.560 --> 00:05:52.540 trying to understand variability in these windows

145 00:05:52.540 --> 00:05:55.580 essentially, and how to estimate them appropriately.

146 00:05:55.580 --> 00:05:58.190 So you'll start to see, I show some pictures,

147 00:05:58.190 --> 00:06:01.940 some figures here that the models become really

148 00:06:01.940 --> 00:06:03.410 lots of parameters in these models.

149 00:06:03.410 --> 00:06:06.080 So you, it really becomes an estimation challenge.

150 00:06:06.080 --> 00:06:07.290 Like how do you,

151 00:06:07.290 --> 00:06:09.010 the model makes sense, you can write it down,

152 00:06:09.010 --> 00:06:10.450 but can you actually fit these models?

153 00:06:10.450 --> 00:06:13.910 So we've done these or consider these models

154 00:06:13.910 --> 00:06:15.170 in a number of different settings,

155 00:06:15.170 --> 00:06:17.260 including the space temporal settings,

156 00:06:17.260 --> 00:06:20.190 survival statistics setting, semi parametric,

157 00:06:20.190 --> 00:06:23.010 non-parametric bays with multi-varied outcomes,

158 00:06:23.010 --> 00:06:25.540 and then more recently variable selection.

159 00:06:25.540 --> 00:06:27.600 And so inferences typically carried out

160 00:06:27.600 --> 00:06:30.760 in the Bayesian setting where I do most of my work

161 00:06:30.760 --> 00:06:33.430 due to increased computational flexibility

162 00:06:33.430 --> 00:06:35.460 and importantly incorporation

163 00:06:35.460 --> 00:06:37.620 of stabilizing prior structure.

164 00:06:37.620 --> 00:06:42.310 So not only have these been done on the method side

165 00:06:42.310 --> 00:06:43.870 where a lot of my time is spent,

166 00:06:43.870 --> 00:06:46.390 but I really like seeing them translated

167 00:06:46.390 --> 00:06:48.250 to actual practice too.

168 00:06:48.250 --> 00:06:51.460 So these methods and kind of variants

169 00:06:51.460 --> 00:06:52.700 of these methods have been,

170 00:06:52.700 --> 00:06:55.740 has successfully identified these critical windows

171 00:06:55.740 --> 00:06:58.100 in a number of outcomes and settings

172 00:06:58.100 --> 00:06:59.350 and different populations,

173 00:06:59.350 --> 00:07:01.630 but pre-term birth, low birth weight,

174 00:07:01.630 --> 00:07:04.240 CHDs so across a number of studies now.

175 00:07:04.240 --> 00:07:06.620 So they're getting good traction in other studies.

176 00:07:06.620 --> 00:07:08.070 Well, not just in the stat literature,

177 00:07:08.070 --> 00:07:09.263 which is nice to see.

178 00:07:10.850 --> 00:07:13.420 To give you a more kind of practical view

179 00:07:13.420 --> 00:07:14.480 of what I'm talking about,

180 00:07:14.480 --> 00:07:17.680 this is one of the first studies we published on

181 00:07:17.680 --> 00:07:18.800 way back in 2012.

182 00:07:18.800 --> 00:07:21.510 And this is for Harris County Texas,

183 00:07:21.510 --> 00:07:23.560 home of Houston, Texas.

184 00:07:23.560 --> 00:07:25.420 And on the left two panels,

185 00:07:25.420 --> 00:07:28.950 you'll see output from our newly developed method

186 00:07:28.950 --> 00:07:29.910 on the right two panels,

187 00:07:29.910 --> 00:07:32.240 you'll see output from more of a naive approach

188 00:07:32.240 --> 00:07:34.680 that was that we were considering at the time.

189 00:07:34.680 --> 00:07:36.930 So what we're talking about these critical windows

190 00:07:36.930 --> 00:07:38.940 are exactly what you're seeing.

191 00:07:38.940 --> 00:07:40.920 Maybe you can see my mouse here,

192 00:07:40.920 --> 00:07:45.250 but these periods where these risk ratios

193 00:07:45.250 --> 00:07:48.910 in this case kind of exclude zero

194 00:07:48.910 --> 00:07:50.380 or these risk parameters,

195 00:07:50.380 --> 00:07:51.750 they're not on any particular scale.

196 00:07:51.750 --> 00:07:54.370 That's easily to interpreted in this case, unfortunately,

197 00:07:54.370 --> 00:07:56.510 but this means that elevated exposure

198 00:07:56.510 --> 00:07:59.400 during pregnancy week 10 for example,

199 00:07:59.400 --> 00:08:01.080 leads to an increase in this case,

200 00:08:01.080 --> 00:08:03.730 was preterm birth, a preterm birth risk.

201 00:08:03.730 --> 00:08:06.610 So during your early kind of mid first

202 00:08:06.610 --> 00:08:09.110 and early second trimester pregnancy,

203 00:08:09.110 --> 00:08:13.910 we were noticing some interesting elevated risk to PM 2.5.

204 00:08:13.910 --> 00:08:16.460 And what we've seen across a number of studies now

205 00:08:16.460 --> 00:08:20.360 is that these windows vary by pollutant by outcome

206 00:08:20.360 --> 00:08:21.580 they're very different.

207 00:08:21.580 --> 00:08:24.230 There's lots of variability for ozone for example,

208 00:08:24.230 --> 00:08:26.773 it seemed to be early on in the first trimester.

209 00:08:27.870 --> 00:08:31.430 So this new methodology allows us to kind of hone in

210 00:08:31.430 --> 00:08:34.930 on the signal and reduce some of this noise.

211 00:08:34.930 --> 00:08:37.970 So if you try to basically imagine your data set,

212 00:08:37.970 --> 00:08:40.840 you have lots of pregnant women in your study,

213 00:08:40.840 --> 00:08:43.550 and you have linked with that pollution exposure

214 00:08:43.550 --> 00:08:45.990 for the first 36 weeks of pregnancy.

215 00:08:45.990 --> 00:08:47.320 A really naive thing to do would be,

216 00:08:47.320 --> 00:08:48.550 let's just throw all of those

217 00:08:48.550 --> 00:08:50.400 into a multiple regression model,

218 00:08:50.400 --> 00:08:51.630 some binary regression model,

219 00:08:51.630 --> 00:08:53.030 all at the same time.

220 00:08:53.030 --> 00:08:55.160 Clearly there's going to be correlation across time

221 00:08:55.160 --> 00:08:56.550 because exposure week one looks

222 00:08:56.550 --> 00:08:58.387 like exposure week two, et cetera.

223 00:08:58.387 --> 00:09:01.150 And if you do that, you can expect multicollinearity,

224 00:09:01.150 --> 00:09:04.020 which is jumping around of point estimates,

225 00:09:04.020 --> 00:09:05.400 increased variability,

226 00:09:05.400 --> 00:09:07.030 which is exactly what you see here.

227 00:09:07.030 --> 00:09:08.110 So our new methodology,

228 00:09:08.110 --> 00:09:10.670 which relied on like Gaussian processes

229 00:09:10.670 --> 00:09:12.930 and other smoothing techniques

230 00:09:12.930 --> 00:09:15.143 allowed us to in a data driven way,

231 00:09:16.340 --> 00:09:17.360 kind of tease out signal

232 00:09:17.360 --> 00:09:19.210 that you could almost make out by eye here.

233 00:09:19.210 --> 00:09:20.180 So if you look hard enough,

234 00:09:20.180 --> 00:09:23.900 you can see kind of a similar shape in both cases,

235 00:09:23.900 --> 00:09:25.860 but we were able to see a better shape here.

236 00:09:25.860 --> 00:09:27.510 So this is what we're generally in the past

237 00:09:27.510 --> 00:09:29.800 have been talking about with critical window estimation

238 00:09:29.800 --> 00:09:30.963 and identification.

239 00:09:32.931 --> 00:09:35.860 We mentioned that we worked on the survival outcome,

240 00:09:35.860 --> 00:09:37.910 we started to think about preterm birth

241 00:09:37.910 --> 00:09:40.690 instead of just a binary outcome yes or no.

242 00:09:40.690 --> 00:09:42.790 We wanted to consider it as a survival outcome.

243 00:09:42.790 --> 00:09:44.330 So what's the probability you make it

244 00:09:44.330 --> 00:09:46.090 to week 35 of your pregnancy,

245 00:09:46.090 --> 00:09:48.700 given that you've made it to 34 for example.

246 00:09:48.700 --> 00:09:50.050 So what this opened up was,

247 00:09:50.050 --> 00:09:52.830 well, maybe there are different exposure windows

248 00:09:52.830 --> 00:09:54.600 given different outcome weeks.

249 00:09:54.600 --> 00:09:57.870 So you can think of outcome week on the X axis

250 00:09:57.870 --> 00:10:00.610 on the Y axis here on an exposure week on the Y axis.

251 00:10:00.610 --> 00:10:04.100 So if you gave birth that week 27,

252 00:10:04.100 --> 00:10:07.100 you only had 27 weeks of exposure, for example.

253 00:10:07.100 --> 00:10:09.950 So people were leaving the set as pregnancy happened.

254 00:10:09.950 --> 00:10:11.780 And so we introduced methodology

255 00:10:11.780 --> 00:10:14.790 that not only kind of smoothed in the exposure direction,

256 00:10:14.790 --> 00:10:17.470 but also smooth across the outcome direction.

257 00:10:17.470 --> 00:10:20.790 And so these darker areas indicate weeks

258 00:10:20.790 --> 00:10:23.150 and outcome weeks, exposure weeks and outcome weeks

259 00:10:23.150 --> 00:10:26.510 where elevated exposure more adversely impacts

260 00:10:26.510 --> 00:10:29.330 like the risk of preterm birth in this case.

261 00:10:29.330 --> 00:10:32.380 So there was a distinct difference in this early preterm

262 00:10:32.380 --> 00:10:33.440 and then this late preterm,

263 00:10:33.440 --> 00:10:35.370 which kind of was impacted by exposures later

264 00:10:35.370 --> 00:10:36.363 in the pregnancy.

265 00:10:37.868 --> 00:10:41.700 And so underlying all of these kind of simplified plots

266 00:10:41.700 --> 00:10:43.010 I'm showing you were

267 00:10:43.010 --> 00:10:46.940 these individual outcome week specific critical window plots

268 00:10:46.940 --> 00:10:51.193 that we kind of are more accustomed to interpreting.

269 00:10:52.310 --> 00:10:56.700 So more recently we got into the spacial world noticing

270 00:10:56.700 --> 00:10:58.500 that, well, we started noticing that

271 00:10:58.500 --> 00:10:59.700 when we applied these methods

272 00:10:59.700 --> 00:11:02.740 to different data sets in different areas,

273 00:11:02.740 --> 00:11:05.330 we were seeing different shapes, different windows,

274 00:11:05.330 --> 00:11:07.660 different pollutants, emerging as important.

275 00:11:07.660 --> 00:11:08.550 And so we begin to think,

276 00:11:08.550 --> 00:11:13.180 well, is there spatial variability in even at a local scale?

277 00:11:13.180 --> 00:11:14.850 And so we develop new methodology

278 00:11:14.850 --> 00:11:16.330 that can kind of tease out

279 00:11:17.400 --> 00:11:20.660 not only temporal changes and exposure risk,

280 00:11:20.660 --> 00:11:22.380 but also spatial variability as well.

281 00:11:22.380 --> 00:11:26.060 So there's spatial correlation component here along

282 00:11:26.060 --> 00:11:28.520 with kind of these critical windows floating around as well.

283 00:11:28.520 --> 00:11:31.100 So this was 11 counties in North Carolina,

284 00:11:31.100 --> 00:11:34.750 including Wake County and the county to House Charlotte,

285 00:11:34.750 --> 00:11:36.610 and this was a low birth weight study.

286 00:11:36.610 --> 00:11:39.663 So there's methodology around that can do this.

287 00:11:40.600 --> 00:11:43.570 So we were working on these for a number of years

288 00:11:43.570 --> 00:11:47.430 and we got approached basically with a question,

289 00:11:47.430 --> 00:11:49.040 how are you actually defining

290 00:11:49.040 --> 00:11:50.880 a critical pregnancy window?

291 00:11:50.880 --> 00:11:52.370 And it seemed obvious at first,

292 00:11:52.370 --> 00:11:54.110 but then we started to really question

293 00:11:54.110 --> 00:11:56.450 the assumptions we had been making,

294 00:11:56.450 --> 00:11:57.960 but obviously what we had been doing,

295 00:11:57.960 --> 00:12:00.980 if I go back a few slides here is just looking,

296 00:12:00.980 --> 00:12:02.800 when did these individual week

297 00:12:02.800 --> 00:12:07.030 or time specific parameters exclude the critical value

298 00:12:07.030 --> 00:12:08.297 in zero in this case?

299 00:12:08.297 --> 00:12:10.720 And we were calling that a critical window

300 00:12:12.740 --> 00:12:16.080 but we started to worry that

301 00:12:16.080 --> 00:12:18.250 this might not be getting exactly what we're hoping

302 00:12:18.250 --> 00:12:20.070 is it capturing the true set?

303 00:12:20.070 --> 00:12:22.250 Is this doing a good job?

304 00:12:22.250 --> 00:12:24.670 In particular, we were worried about over smoothing

305 00:12:24.670 --> 00:12:26.920 with something like a Gaussian process

306 00:12:26.920 --> 00:12:29.220 and specifically with the endpoint.

307 00:12:29.220 --> 00:12:31.900 So if you can imagine, I'll go back one more time,

308 00:12:31.900 --> 00:12:33.710 sorry to scroll.

309 00:12:33.710 --> 00:12:36.950 Imagine the end points here and here,

310 00:12:36.950 --> 00:12:41.180 we begin to worry that the over smoothness

311 00:12:41.180 --> 00:12:44.940 could be pulling some of these actually null results

312 00:12:44.940 --> 00:12:47.680 into the critical set or vice versa,

313 00:12:47.680 --> 00:12:50.870 kind of pulling some important ones down to the null set.

314 00:12:50.870 --> 00:12:53.660 So we were very concerned about the endpoints here

315 00:12:53.660 --> 00:12:56.110 when we started working on this more recent work.

316 00:12:57.010 --> 00:12:58.990 So our solution to this

317 00:12:58.990 --> 00:13:00.740 was critical window variable selection.

318 00:13:00.740 --> 00:13:03.940 So we like the smoothness, we like the plots that emerge.

319 00:13:03.940 --> 00:13:05.560 We like how we can interpret these things,

320 00:13:05.560 --> 00:13:07.370 but a variable selection component

321 00:13:07.370 --> 00:13:10.250 would allow us to turn some of these effects off,

322 00:13:10.250 --> 00:13:13.570 even if they appear to be significant in the plots.

323 00:13:13.570 --> 00:13:15.350 And so what this meant is,

324 00:13:15.350 --> 00:13:18.790 we introduced like a bayesian variable selection technique

325 00:13:19.950 --> 00:13:21.510 called critical window variable selection,

326 00:13:21.510 --> 00:13:25.010 where basically you still have the critical window plots

327 00:13:25.010 --> 00:13:27.560 that you know and love, and you know how to interpret,

328 00:13:27.560 --> 00:13:30.010 but underlying each effect now,

329 00:13:30.010 --> 00:13:33.140 you actually have this binary exclusionary,

330 00:13:33.140 --> 00:13:34.390 or inclusion variable

331 00:13:34.390 --> 00:13:37.010 that tells you whether this thing should be included.

332 00:13:37.010 --> 00:13:39.000 This particular weekly effect should be included

333 00:13:39.000 --> 00:13:40.420 in the critical window set.

334 00:13:40.420 --> 00:13:43.890 And what we found is that there are a number of times,

335 00:13:43.890 --> 00:13:47.260 not in this particular real case study in North Carolina,

336 00:13:47.260 --> 00:13:48.140 but through simulation,

337 00:13:48.140 --> 00:13:49.890 we noticed that there were times

338 00:13:49.890 --> 00:13:52.640 when exactly what we had worried was happening

339 00:13:52.640 --> 00:13:54.170 had been happening so effects

340 00:13:54.170 --> 00:13:58.000 near the border here were being pulled into the set,

341 00:13:58.000 --> 00:13:59.880 but luckily they were not being included

342 00:13:59.880 --> 00:14:01.190 in the variable selection component.

343 00:14:01.190 --> 00:14:04.860 So to be in the variable selection set now,

344 00:14:04.860 --> 00:14:07.920 you had to have posterior inclusion probability bigger

345 00:14:07.920 --> 00:14:10.660 than point five, so bigger than this line

346 00:14:10.660 --> 00:14:12.840 and your individual weekly effects

347 00:14:12.840 --> 00:14:16.280 had to be exclude zero with a 95% credible vulnerable.

348 00:14:16.280 --> 00:14:19.430 So with these two kind of definitions we were doing,

349 00:14:19.430 --> 00:14:22.590 we were getting a much better kind of recovering

350 00:14:22.590 --> 00:14:27.350 the true set of critical windows in simulation, at least.

351 00:14:27.350 --> 00:14:29.340 So this really outperformed

352 00:14:29.340 --> 00:14:30.430 what we had been doing previously.

353 00:14:30.430 --> 00:14:31.670 So we've been moving forward

354 00:14:31.670 --> 00:14:34.483 with this variable selection concept since then.

355 00:14:35.550 --> 00:14:37.840 All right, so we like critical window variable selection,

356 00:14:37.840 --> 00:14:39.150 we like a lot of these other methods.

357 00:14:39.150 --> 00:14:41.760 The problem is that as I know,

358 00:14:41.760 --> 00:14:43.280 a number of you are aware,

359 00:14:43.280 --> 00:14:45.510 the literature has really moved towards the science

360 00:14:45.510 --> 00:14:47.900 has moved towards pollution, mixtures

361 00:14:47.900 --> 00:14:49.800 and multiple exposures.

362 00:14:49.800 --> 00:14:52.230 And a lot of these methodologies were developed

363 00:14:52.230 --> 00:14:56.653 with one pollutant in mind at the most two to three,

364 00:14:57.830 --> 00:14:59.840 but they were not generally meant

365 00:14:59.840 --> 00:15:02.120 for pollution mixtures for example.

366 00:15:02.120 --> 00:15:03.630 So our goal in this work

367 00:15:03.630 --> 00:15:06.830 was to extend what we liked the CWVS,

368 00:15:06.830 --> 00:15:10.000 critical and variable selection to accommodate mixtures.

369 00:15:10.000 --> 00:15:12.020 And so when we started to thinking about mixtures,

370 00:15:12.020 --> 00:15:14.080 when you have time varying exposures

371 00:15:14.080 --> 00:15:15.900 and time varying effects,

372 00:15:15.900 --> 00:15:19.060 it became relatively conceptually complicated

373 00:15:19.060 --> 00:15:22.140 because you have lots of parameters floating around.

374 00:15:22.140 --> 00:15:23.410 So we wanted something that could do

375 00:15:23.410 --> 00:15:25.420 like a dimension reduction essentially.

376 00:15:25.420 --> 00:15:28.340 So what we thought is a nice solution,

377 00:15:28.340 --> 00:15:31.680 like in a single pollutant context, or I'm sorry,

378 00:15:31.680 --> 00:15:33.860 in a single exposure time period context

379 00:15:33.860 --> 00:15:36.200 is this weighted quantile sum regression,

380 00:15:36.200 --> 00:15:37.580 which I know a lot of you are familiar with,

381 00:15:37.580 --> 00:15:40.230 'cause I've helped write pieces of grants

382 00:15:40.230 --> 00:15:43.130 that have discussed weighted quantile sum regression here,

383 00:15:44.070 --> 00:15:46.090 but it offers a nice interpretable solution

384 00:15:46.090 --> 00:15:49.290 for estimating the impact of a mixture on an outcome.

385 00:15:49.290 --> 00:15:52.500 And it has this really nice sum to one constraint

386 00:15:52.500 --> 00:15:54.580 on the regression parameters.

387 00:15:54.580 --> 00:15:56.010 And so you get in the end,

388 00:15:56.010 --> 00:15:58.290 you have 20 pollutants for example,

389 00:15:58.290 --> 00:16:00.670 and you get to see the relative contribution

390 00:16:00.670 --> 00:16:04.190 of each of these pollutants in terms of the entire mixture.

391 00:16:04.190 --> 00:16:06.387 So you have these little sum to one between zero

392 00:16:06.387 --> 00:16:08.790 and one probabilities or proportions

393 00:16:08.790 --> 00:16:12.080 that describe the role of individual pollutants.

394 00:16:12.080 --> 00:16:14.640 And then you have this global regression parameter

395 00:16:14.640 --> 00:16:16.830 that describes the impact of that mixture

396 00:16:16.830 --> 00:16:20.720 as defined by those weights on the health outcome.

397 00:16:20.720 --> 00:16:23.800 So it does a little two stage process estimate weights

398 00:16:23.800 --> 00:16:26.240 and then global regression parameter,

399 00:16:26.240 --> 00:16:28.623 not important for this talk.

400 00:16:29.480 --> 00:16:32.410 More recently in 2020, this was extended

401 00:16:32.410 --> 00:16:35.420 to the lag weighted quantile sum regression.

402 00:16:35.420 --> 00:16:40.420 And yeah, it extended WQS to the multiple pollutants setting

403 00:16:41.950 --> 00:16:46.510 in a really, I think of it as a relatively ad hoc solution,

404 00:16:46.510 --> 00:16:50.830 but basically WQS has fit at each exposure week separately.

405 00:16:50.830 --> 00:16:52.580 The weights are estimated,

406 00:16:52.580 --> 00:16:55.490 the mixtures are combined based on those weights.

407 00:16:55.490 --> 00:16:58.290 And then those kind of package mixtures

408 00:16:58.290 --> 00:17:00.450 are thrown into like a distributed live model

409 00:17:00.450 --> 00:17:01.750 to estimate similar curves

410 00:17:01.750 --> 00:17:03.410 is what I've been showing you so far.

411 00:17:03.410 --> 00:17:05.400 So the estimation of the weights

412 00:17:05.400 --> 00:17:07.480 and their relative importance in the mixture

413 00:17:07.480 --> 00:17:11.340 are done separately outside of kind of the estimation

414 00:17:11.340 --> 00:17:12.550 of the regression parameters as well.

415 00:17:12.550 --> 00:17:15.970 So this more, again, more of a two stage approach.

416 00:17:15.970 --> 00:17:18.670 All right, so we like WQS

417 00:17:18.670 --> 00:17:21.830 because of its relative simplicity and its interpretability,

418 00:17:21.830 --> 00:17:23.970 we liked critical and variable selection.

419 00:17:23.970 --> 00:17:25.560 So the goals here were to combine

420 00:17:25.560 --> 00:17:29.320 that estimation identification ability of CWVS

421 00:17:29.320 --> 00:17:31.440 with the interpretability and shrinkage properties

422 00:17:31.440 --> 00:17:36.440 of WQS within a unified modeling framework and extending

423 00:17:36.870 --> 00:17:39.150 oh yeah, so WQS is nice.

424 00:17:39.150 --> 00:17:42.260 It has zero to some to one components

425 00:17:42.260 --> 00:17:44.170 that are between zero and one,

426 00:17:44.170 --> 00:17:46.880 but you don't actually get a sense of variable selection

427 00:17:46.880 --> 00:17:47.713 when doing this.

428 00:17:47.713 --> 00:17:49.600 So none of the weights can exactly equal zero.

429 00:17:49.600 --> 00:17:52.760 We wanted a more sparse solution

430 00:17:52.760 --> 00:17:54.470 and so we introduced also a way

431 00:17:54.470 --> 00:17:57.120 to make these weights exactly zero.

432 00:17:57.120 --> 00:17:58.300 So you can get a better sense of

433 00:17:58.300 --> 00:18:02.020 which pollutants are the main players in the mixture.

434 00:18:02.020 --> 00:18:06.030 And so what we're calling this is CWVS for mixtures

435 00:18:06.030 --> 00:18:07.423 or CWVS mix.

436 00:18:08.770 --> 00:18:11.300 And so some features before we get

437 00:18:11.300 --> 00:18:13.060 into a little bit of the details of the model,

438 00:18:13.060 --> 00:18:14.240 these are like the high,

439 00:18:14.240 --> 00:18:16.440 just if you take nothing else away from like

440 00:18:16.440 --> 00:18:17.840 what this model does, this is,

441 00:18:17.840 --> 00:18:19.840 I think the important slide here is that,

442 00:18:19.840 --> 00:18:22.690 we have main effects and first order interactions

443 00:18:22.690 --> 00:18:25.470 between the pollutants during each exposure period.

444 00:18:25.470 --> 00:18:27.340 So week one of pregnancy,

445 00:18:27.340 --> 00:18:29.620 week two of pregnancy, all of these interactions,

446 00:18:29.620 --> 00:18:31.220 all of these main effects are included.

447 00:18:31.220 --> 00:18:33.560 So there's lots of parameters you can already imagine

448 00:18:33.560 --> 00:18:34.980 are floating around here.

449 00:18:34.980 --> 00:18:38.380 We still hold onto this sum to one mixture weights

450 00:18:38.380 --> 00:18:40.580 at each exposure week separately.

451 00:18:40.580 --> 00:18:42.540 But we want to account for the fact that,

452 00:18:42.540 --> 00:18:44.470 what's happening in exposure week one

453 00:18:44.470 --> 00:18:47.670 may be similar to exposure week two to three to four,

454 00:18:47.670 --> 00:18:50.120 with this correlation dying out as you get further apart

455 00:18:50.120 --> 00:18:51.300 in exposure time.

456 00:18:51.300 --> 00:18:53.900 So we want these weights not to have to be estimated

457 00:18:54.740 --> 00:18:56.753 kind of independently at each exposure week.

458 00:18:56.753 --> 00:18:59.830 We want to enforce some smoothness,

459 00:18:59.830 --> 00:19:04.390 data driven smoothness preferably to estimate these weights.

460 00:19:04.390 --> 00:19:05.930 And as I mentioned, we want these weights

461 00:19:05.930 --> 00:19:07.580 to have a variable selection component.

462 00:19:07.580 --> 00:19:10.200 So we can actually identify individual elements

463 00:19:10.200 --> 00:19:13.840 of the mixture and we still have this global risk parameter,

464 00:19:13.840 --> 00:19:16.580 and this is going to follow the CWVS model

465 00:19:16.580 --> 00:19:18.240 so that we can estimate

466 00:19:18.240 --> 00:19:20.190 these critical windows more accurately.

467 00:19:21.620 --> 00:19:23.610 All right, so the goals of this study

468 00:19:23.610 --> 00:19:26.170 before you jump into some of the methodology here

469 00:19:26.170 --> 00:19:28.440 are to develop CWVS mix.

470 00:19:28.440 --> 00:19:29.273 As I mentioned,

471 00:19:30.380 --> 00:19:32.090 simulation is really important in this world.

472 00:19:32.090 --> 00:19:33.640 I wanna make sure that what we're doing

473 00:19:33.640 --> 00:19:35.960 is not just duplicating other efforts

474 00:19:35.960 --> 00:19:37.740 and that it's actually offering something new,

475 00:19:37.740 --> 00:19:39.530 something helpful to the literature

476 00:19:39.530 --> 00:19:41.150 that we can point to.

477 00:19:41.150 --> 00:19:44.670 I think I know the shortcomings of something like lag,

478 00:19:44.670 --> 00:19:45.820 weighted quantile sum regression,

479 00:19:45.820 --> 00:19:48.670 but until I see it actually happen in simulation

480 00:19:48.670 --> 00:19:50.500 it's just kind of hypothetical.

481 00:19:50.500 --> 00:19:53.370 So finally we wanna investigate the impact

482 00:19:53.370 --> 00:19:54.620 using this new methodology

483 00:19:54.620 --> 00:19:58.160 of multiple ambient air pollutants on stillbirth risk.

484 00:19:58.160 --> 00:19:58.993 And in this case,

485 00:19:58.993 --> 00:20:03.450 we're focusing on New Jersey from 2005 to 2014.

486 00:20:03.450 --> 00:20:05.980 And actually we have really nice output

487 00:20:05.980 --> 00:20:07.740 from a novel data fusion model.

488 00:20:07.740 --> 00:20:09.500 There are lots of data fusion models floating

489 00:20:09.500 --> 00:20:12.390 around right now, but this is a one from 2019,

490 00:20:12.390 --> 00:20:16.270 from our collaborator at Georgia tech and at Emory

491 00:20:16.270 --> 00:20:18.120 that provided 12 pollutants,

492 00:20:18.120 --> 00:20:22.330 12 kilometer grid cell size across the entire US

493 00:20:22.330 --> 00:20:25.680 daily no missing this things like that.

494 00:20:25.680 --> 00:20:27.543 So for these particular pollutants.

495 00:20:28.690 --> 00:20:31.330 All right, so let's talk a little bit about

496 00:20:31.330 --> 00:20:32.670 the model and what it does

497 00:20:32.670 --> 00:20:34.030 and some of the intuitive features

498 00:20:34.030 --> 00:20:37.380 that I think it has and why it might work well.

499 00:20:37.380 --> 00:20:42.090 So yeah, we're starting with some outcome,

500 00:20:42.090 --> 00:20:43.890 it could be some adverse health outcome

501 00:20:43.890 --> 00:20:45.760 like preterm pregnancy or not,

502 00:20:45.760 --> 00:20:48.750 or stillbirth or not some be newly outcome

503 00:20:48.750 --> 00:20:51.500 where this PI describes kind of the probability

504 00:20:51.500 --> 00:20:54.243 that person I experiences this outcome.

505 00:20:55.160 --> 00:20:57.780 We model this probability using logistic regression

506 00:20:57.780 --> 00:20:58.923 as we normally would,

507 00:21:00.050 --> 00:21:02.320 these green I'm kind of trying to different.

508 00:21:02.320 --> 00:21:03.940 I'm trying to keep people's attention

509 00:21:03.940 --> 00:21:05.010 to the parameters

510 00:21:05.010 --> 00:21:07.930 and how I'm mentally grouping them as well.

511 00:21:07.930 --> 00:21:11.700 So these green represent these typical like demographics.

512 00:21:11.700 --> 00:21:13.280 We know there are certain risk factors

513 00:21:13.280 --> 00:21:14.780 for different health outcomes,

514 00:21:15.620 --> 00:21:18.720 particularly pregnancy outcomes being over 35 for example,

515 00:21:18.720 --> 00:21:23.120 with preterm pregnancy, alcohol, smoking, et cetera.

516 00:21:23.120 --> 00:21:26.240 So this would go into this exi transpose data.

517 00:21:26.240 --> 00:21:31.100 This specter here where a lot of our work came in

518 00:21:31.100 --> 00:21:32.810 are on these blue parameters,

519 00:21:32.810 --> 00:21:34.640 which are the weights that I've been talking about.

520 00:21:34.640 --> 00:21:37.300 So these weights, these blue parameters

521 00:21:37.300 --> 00:21:41.730 actually sum to one at each exposure week.

522 00:21:41.730 --> 00:21:44.380 So each exposure week T

523 00:21:44.380 --> 00:21:46.550 we basically have a vector of Lambdas

524 00:21:46.550 --> 00:21:48.630 that are weights between zero and one

525 00:21:48.630 --> 00:21:51.930 could be actually equal to zero exactly.

526 00:21:51.930 --> 00:21:55.190 And they sum to one at each exposure week separately,

527 00:21:55.190 --> 00:21:56.880 you notice their index Byte

528 00:21:56.880 --> 00:21:59.300 because we're allowing the possibility

529 00:21:59.300 --> 00:22:03.320 that the exposure profile changes across the pregnancy.

530 00:22:03.320 --> 00:22:05.120 So it early on in the pregnancy,

531 00:22:05.120 --> 00:22:10.120 maybe the risk is primarily driven by pollutant A

532 00:22:10.340 --> 00:22:11.700 but later on in the pregnancy,

533 00:22:11.700 --> 00:22:13.210 perhaps that shifts.

534 00:22:13.210 --> 00:22:15.810 And so the weights would shift well as well,

535 00:22:15.810 --> 00:22:18.790 but we expect this shift to be smoother

536 00:22:18.790 --> 00:22:21.690 rather than complete choppiness across the exposure weeks.

537 00:22:23.010 --> 00:22:25.360 And so what these weights do are

538 00:22:25.360 --> 00:22:28.520 they kind of multiply here with the main effects

539 00:22:28.520 --> 00:22:30.510 and these first order interactions.

540 00:22:30.510 --> 00:22:33.040 And if you think about taking this sum across

541 00:22:33.040 --> 00:22:34.390 main effects and interactions,

542 00:22:34.390 --> 00:22:38.010 you have this package of weighted exposure essentially.

543 00:22:38.010 --> 00:22:42.190 And the alpha here tells us whether at exposure period T

544 00:22:42.190 --> 00:22:44.490 this package has any impact on

545 00:22:44.490 --> 00:22:47.710 your ultimate probability of developing the outcome.

546 00:22:47.710 --> 00:22:49.890 So we have this nice sense of the weights,

547 00:22:49.890 --> 00:22:53.160 help us describe what's happening with the mixture profiles.

548 00:22:53.160 --> 00:22:54.700 And, but the alpha keeps us honest

549 00:22:54.700 --> 00:22:57.040 and keeps us able to say,

550 00:22:57.040 --> 00:22:59.720 well, you know, this mixture's interesting,

551 00:22:59.720 --> 00:23:03.143 but it has no impact on the health outcome of interest here.

552 00:23:05.610 --> 00:23:08.790 So how do we do these mixture weights?

553 00:23:08.790 --> 00:23:11.220 As I mentioned, two features that we're interested in

554 00:23:11.220 --> 00:23:13.650 the ability to actually equal zero

555 00:23:13.650 --> 00:23:15.090 and smoothness across time.

556 00:23:15.090 --> 00:23:18.800 And so first point is to,

557 00:23:18.800 --> 00:23:21.620 well, we introduce these latent weight parameters

558 00:23:21.620 --> 00:23:23.050 that I'm calling Lambda star,

559 00:23:23.050 --> 00:23:25.040 not to don't get too caught up in them.

560 00:23:25.040 --> 00:23:28.150 Basically they're continuously varying parameters

561 00:23:28.150 --> 00:23:31.400 that as soon as they cross the zero threshold,

562 00:23:31.400 --> 00:23:32.680 they turn on in our model.

563 00:23:32.680 --> 00:23:34.640 So that's what this maximum is doing.

564 00:23:34.640 --> 00:23:37.070 So they turn on and they give you some weight

565 00:23:37.070 --> 00:23:39.510 and then as soon as they cross into negative territory,

566 00:23:39.510 --> 00:23:40.343 they go to zero.

567 00:23:40.343 --> 00:23:42.960 So this is how we're getting actual zeros in these weights.

568 00:23:42.960 --> 00:23:44.960 So the Lambdas and the Lambda Tilda

569 00:23:44.960 --> 00:23:46.310 can actually equal zero

570 00:23:47.210 --> 00:23:51.830 based on these underlying latent weight parameters.

571 00:23:51.830 --> 00:23:54.680 All right, so we keep them summing to one

572 00:23:54.680 --> 00:23:57.500 by dividing by the sum of the numerator, essentially.

573 00:23:57.500 --> 00:24:00.440 So whatever weights are positive gets summed

574 00:24:00.440 --> 00:24:01.480 and we're dividing by,

575 00:24:01.480 --> 00:24:03.520 we're basically self kind of correcting here

576 00:24:03.520 --> 00:24:06.570 so that the weights always come to one,

577 00:24:06.570 --> 00:24:08.450 these weights combined.

578 00:24:08.450 --> 00:24:11.840 For the interactions, we don't want the case.

579 00:24:11.840 --> 00:24:13.650 We prefer sparse model,

580 00:24:13.650 --> 00:24:16.540 particularly as the number of pollutants get really large.

581 00:24:16.540 --> 00:24:18.570 So the number of interactions will grow.

582 00:24:18.570 --> 00:24:21.150 So what we want is our interactions

583 00:24:21.150 --> 00:24:23.810 that are only turned on essentially

584 00:24:23.810 --> 00:24:25.180 when the main effects are turned on.

585 00:24:25.180 --> 00:24:27.480 So you can see these two indicators I've added

586 00:24:27.480 --> 00:24:29.590 basically say if the main effect themselves

587 00:24:29.590 --> 00:24:31.080 aren't both turned on,

588 00:24:31.080 --> 00:24:34.120 this interaction effect gets zeroed out already.

589 00:24:34.120 --> 00:24:36.900 So the interaction has a kind of a higher bar clear

590 00:24:36.900 --> 00:24:40.230 this strict hierarchy basically

591 00:24:40.230 --> 00:24:42.860 where both main effects have to be on

592 00:24:42.860 --> 00:24:45.670 and the interaction latent variable has to be on.

593 00:24:45.670 --> 00:24:47.640 So there's the zero component now,

594 00:24:47.640 --> 00:24:49.990 how do we do smoothness across time?

595 00:24:49.990 --> 00:24:52.360 Well, it's all about this correlation structure.

596 00:24:52.360 --> 00:24:54.570 So these latent Lambda star parameters

597 00:24:54.570 --> 00:24:56.970 that control the weights are actually modeled

598 00:24:56.970 --> 00:24:59.260 as a multi Gaussian process.

599 00:24:59.260 --> 00:25:01.490 And I think the key thing to focus on here is that

600 00:25:01.490 --> 00:25:04.260 there's this underlying correlation structure

601 00:25:04.260 --> 00:25:08.670 that tells us as two exposure time points get further apart.

602 00:25:08.670 --> 00:25:10.570 This exponential of a negative number

603 00:25:10.570 --> 00:25:11.610 will get closer to zero.

604 00:25:11.610 --> 00:25:15.650 So correlation dies out as exposure time gets further apart

605 00:25:15.650 --> 00:25:17.050 now, as they get closer together,

606 00:25:17.050 --> 00:25:18.720 this correlation is gonna be higher.

607 00:25:18.720 --> 00:25:21.690 And the main parameter that controls this level

608 00:25:21.690 --> 00:25:23.230 of correlation is this fee parameter.

609 00:25:23.230 --> 00:25:26.770 And we actually put prior distributions on this

610 00:25:26.770 --> 00:25:28.630 to allow the data, to drive the inference,

611 00:25:28.630 --> 00:25:31.630 rather than like our view of what we expect

612 00:25:31.630 --> 00:25:33.490 this smoothness to look like.

613 00:25:33.490 --> 00:25:35.270 So yeah, this is data driven

614 00:25:35.270 --> 00:25:37.403 kind of smoothness across exposure time.

615 00:25:38.680 --> 00:25:41.700 All right, so now, so we've got the weights handled

616 00:25:41.700 --> 00:25:43.410 they have both properties that we care about.

617 00:25:43.410 --> 00:25:45.770 Now let's talk about the mixture impact itself.

618 00:25:45.770 --> 00:25:48.400 So this alpha recall tells us whether the mixture

619 00:25:48.400 --> 00:25:50.050 that we observe at time point T

620 00:25:50.050 --> 00:25:52.570 or that we estimated exposure time point T

621 00:25:52.570 --> 00:25:55.390 is actually relevant to the health outcome.

622 00:25:55.390 --> 00:25:56.700 So we want, again,

623 00:25:56.700 --> 00:25:58.510 we want this variable selection here

624 00:25:58.510 --> 00:26:00.550 because we've noticed the problem with the end points

625 00:26:00.550 --> 00:26:02.170 that I described earlier.

626 00:26:02.170 --> 00:26:05.960 So to do this, we decompose this effect into two pieces,

627 00:26:05.960 --> 00:26:07.560 a continuously varying piece.

628 00:26:07.560 --> 00:26:08.720 And then this binary piece

629 00:26:08.720 --> 00:26:11.120 that I mentioned earlier on in the talk,

630 00:26:11.120 --> 00:26:13.230 the binary piece are just independent

631 00:26:13.230 --> 00:26:15.090 but newly random variables.

632 00:26:15.090 --> 00:26:19.600 But we imagine that if you're in the critical window set

633 00:26:19.600 --> 00:26:22.010 at time one, then you may be in it at time two

634 00:26:22.010 --> 00:26:24.580 and may be more likely to be in at time three.

635 00:26:24.580 --> 00:26:26.750 So there may be some sense of correlation

636 00:26:26.750 --> 00:26:28.830 across exposure time here as well.

637 00:26:28.830 --> 00:26:30.880 So while we model these things as independent,

638 00:26:30.880 --> 00:26:34.240 the probabilities that underlie these zero

639 00:26:34.240 --> 00:26:36.940 and one variables are actually smoothly varying

640 00:26:36.940 --> 00:26:38.730 and correlated across time.

641 00:26:38.730 --> 00:26:39.563 So again,

642 00:26:39.563 --> 00:26:42.063 we use this kind of exponential correlation structure.

643 00:26:42.960 --> 00:26:46.130 We allow for cross correlation between the continuous

644 00:26:46.130 --> 00:26:48.110 and the binary piece.

645 00:26:48.110 --> 00:26:49.290 Not important to get into here,

646 00:26:49.290 --> 00:26:51.543 you can kind of read back over.

647 00:26:52.450 --> 00:26:53.940 I can share a paper with you if you want to,

648 00:26:53.940 --> 00:26:55.740 or talk more about it offline,

649 00:26:55.740 --> 00:26:57.830 but essentially there's some cross correlation

650 00:26:57.830 --> 00:26:59.060 there's correlation across time,

651 00:26:59.060 --> 00:27:02.247 but this allows for smoothness in the effects

652 00:27:02.247 --> 00:27:04.310 and the kind of the regression parameter effects

653 00:27:04.310 --> 00:27:05.300 that we've been looking at,

654 00:27:05.300 --> 00:27:08.030 but also in the variable selection as well.

655 00:27:08.030 --> 00:27:10.510 And these, both of these things come together to kind

656 00:27:10.510 --> 00:27:13.890 of define the critical window variable selection model.

657 00:27:13.890 --> 00:27:17.270 To finish the model recall everything's in the base setting

658 00:27:17.270 --> 00:27:20.900 so really weekly informative prior distributions

659 00:27:20.900 --> 00:27:24.180 kind of standard prior distributions when possible,

660 00:27:24.180 --> 00:27:25.520 nothing too interesting here.

661 00:27:25.520 --> 00:27:29.060 So the model you may be looking at on this previous slide

662 00:27:29.060 --> 00:27:31.290 and thinking there's a lot of parameters floating

663 00:27:31.290 --> 00:27:32.123 around here.

664 00:27:32.123 --> 00:27:35.280 There's a lot of output that you're going to be estimating.

665 00:27:35.280 --> 00:27:37.780 So how do you make sense of this as a practitioner,

666 00:27:37.780 --> 00:27:39.830 someone who actually wants to know if a mixture

667 00:27:39.830 --> 00:27:42.130 is having an impact on your health?

668 00:27:42.130 --> 00:27:45.640 Well, luckily we still have relatively nice

669 00:27:45.640 --> 00:27:49.760 and estimable kind of effects here,

670 00:27:49.760 --> 00:27:51.640 associations that we can talk about.

671 00:27:51.640 --> 00:27:52.710 So for example,

672 00:27:52.710 --> 00:27:55.350 for a change in the log odds for a one unit increase

673 00:27:55.350 --> 00:27:59.240 in each pollutant during a particular exposure period,

674 00:27:59.240 --> 00:28:00.570 this would be the quantity

675 00:28:00.570 --> 00:28:02.021 that you would make (indistinct).

676 00:28:02.021 --> 00:28:02.941 You would exponentiate this,

677 00:28:02.941 --> 00:28:05.060 and you would have like an odd ratio, for example,

678 00:28:05.060 --> 00:28:08.150 now recall for any model that includes interactions.

679 00:28:08.150 --> 00:28:11.980 The interpretation is always increasingly complicated

680 00:28:11.980 --> 00:28:14.290 because it matters where you start

681 00:28:14.290 --> 00:28:15.230 when you have interactions.

682 00:28:15.230 --> 00:28:17.500 So if you're already at a high level,

683 00:28:17.500 --> 00:28:21.143 so the values themselves of exposure have to come into play,

684 00:28:22.210 --> 00:28:25.040 but nonetheless, you can still get nice quantities

685 00:28:25.040 --> 00:28:26.850 to estimate in the end.

686 00:28:26.850 --> 00:28:28.820 And if you're only interested in what happens

687 00:28:28.820 --> 00:28:31.920 if pollutant A increased during exposure period T

688 00:28:31.920 --> 00:28:33.610 you can write down actually

689 00:28:33.610 --> 00:28:34.680 what that looks like as well.

690 00:28:34.680 --> 00:28:37.720 So you can estimate both of these things relatively easily

691 00:28:37.720 --> 00:28:39.320 from our output, from our model.

692 00:28:40.700 --> 00:28:42.300 Alright, so we have a model

693 00:28:42.300 --> 00:28:44.160 that kind of checked all the boxes,

694 00:28:44.160 --> 00:28:46.460 at least in my head when I was writing it down

695 00:28:46.460 --> 00:28:49.200 and we can, I tested it, we can fit it,

696 00:28:49.200 --> 00:28:52.530 it seems to work and that it's converging

697 00:28:52.530 --> 00:28:54.903 and it's producing things that look reasonable,

698 00:28:55.890 --> 00:28:58.720 but the simulation study really allows us to dig deeper

699 00:28:58.720 --> 00:29:02.260 and say, is there anything, this it's obviously new,

700 00:29:02.260 --> 00:29:04.130 but is there anything beneficial to what we're doing?

701 00:29:04.130 --> 00:29:06.160 Or should we just be doing something simpler

702 00:29:06.160 --> 00:29:07.960 that already exists?

703 00:29:07.960 --> 00:29:09.950 So we wanted particularly to ask,

704 00:29:09.950 --> 00:29:12.810 how does CWVS mix compared

705 00:29:12.810 --> 00:29:14.930 to some of these existing approaches

706 00:29:14.930 --> 00:29:16.840 for three different factors that we're interested in?

707 00:29:16.840 --> 00:29:19.920 So first identifying the true critical window set,

708 00:29:19.920 --> 00:29:22.070 obviously probably the most important part

709 00:29:22.070 --> 00:29:24.960 of critical window research here is like,

710 00:29:24.960 --> 00:29:27.790 let's get the critical window set right

711 00:29:27.790 --> 00:29:31.210 when we're estimating and identifying these parameters.

712 00:29:31.210 --> 00:29:32.830 But obviously when you're talking about mixtures,

713 00:29:32.830 --> 00:29:34.520 we also care about these weights.

714 00:29:34.520 --> 00:29:37.720 We want to know that the mixture profile we're looking at

715 00:29:37.720 --> 00:29:41.053 on a certain exposure period actually is,

716 00:29:43.120 --> 00:29:45.230 reflective of the true mixture profile

717 00:29:45.230 --> 00:29:46.340 that makes sense here.

718 00:29:46.340 --> 00:29:49.320 So how well do we do at estimating these Lambdas

719 00:29:49.320 --> 00:29:51.710 and Lambdas Tilda parameters

720 00:29:51.710 --> 00:29:55.170 that describe the effects of main effects and interactions,

721 00:29:55.170 --> 00:29:56.003 and then finally,

722 00:29:56.003 --> 00:30:00.380 how well do we do it at estimating the magnitude of risk,

723 00:30:00.380 --> 00:30:01.580 these alpha T parameters.

724 00:30:01.580 --> 00:30:03.870 We wanna make sure we're getting these right as well.

725 00:30:03.870 --> 00:30:05.650 And as a side issue, I guess,

726 00:30:05.650 --> 00:30:07.460 just more of our curiosity,

727 00:30:07.460 --> 00:30:10.320 how well does this variable selection process work

728 00:30:10.320 --> 00:30:12.570 for the weights that we've introduced?

729 00:30:12.570 --> 00:30:15.130 So now we need to think about

730 00:30:15.130 --> 00:30:16.710 what are competing methods in this space.

731 00:30:16.710 --> 00:30:18.560 There aren't a lot of methods out there

732 00:30:18.560 --> 00:30:23.560 that aim to estimate critical windows with.

733 00:30:23.580 --> 00:30:27.260 So time bearing exposures and multiple pollutants

734 00:30:27.260 --> 00:30:29.500 and the ones that are out there

735 00:30:29.500 --> 00:30:31.027 give different enough output

736 00:30:31.027 --> 00:30:34.200 that's hard to compare one model to the next,

737 00:30:34.200 --> 00:30:38.060 but here are three approaches that we kind of came up with.

738 00:30:38.060 --> 00:30:40.280 One is the most naive kind of

739 00:30:40.280 --> 00:30:41.680 where I would always start

740 00:30:41.680 --> 00:30:43.870 as a practitioner with a new data set,

741 00:30:43.870 --> 00:30:45.440 this equal weights approach.

742 00:30:45.440 --> 00:30:49.390 So maybe just averaging all of the exposures for a person

743 00:30:49.390 --> 00:30:53.307 on a given exposure week and including that average

744 00:30:55.100 --> 00:30:59.400 and the interactions with the other exposure periods

745 00:30:59.400 --> 00:31:02.463 in a framework, a distributed lag framework.

746 00:31:03.860 --> 00:31:06.323 So yeah, this is called equal weights or EW.

747 00:31:07.866 --> 00:31:09.560 A PCA approach also makes sense.

748 00:31:09.560 --> 00:31:11.470 So let's allow the data to determine

749 00:31:11.470 --> 00:31:14.530 the correct weights of these Lambdas,

750 00:31:14.530 --> 00:31:17.620 but let's focus it only on the exposure period,

751 00:31:17.620 --> 00:31:18.820 only the exposure data.

752 00:31:18.820 --> 00:31:20.090 So at each exposure period,

753 00:31:20.090 --> 00:31:24.090 fit a PCA to the person specific exposures

754 00:31:24.090 --> 00:31:26.530 and generate these weights.

755 00:31:26.530 --> 00:31:29.370 That kind of describe the relative contribution

756 00:31:29.370 --> 00:31:33.770 of the different interactions and main effects in a mixture,

757 00:31:33.770 --> 00:31:35.720 and then weight the mixtures in that way

758 00:31:35.720 --> 00:31:37.790 and throw that weighted value

759 00:31:37.790 --> 00:31:40.500 into the distributed regression model.

760 00:31:40.500 --> 00:31:42.040 So for all of these methods,

761 00:31:42.040 --> 00:31:44.380 we're using the original CWVS,

762 00:31:44.380 --> 00:31:47.200 so that we're comparable so that the method

763 00:31:47.200 --> 00:31:49.290 so that the results are actually comparable across.

764 00:31:49.290 --> 00:31:52.180 And that the only thing that is changing essentially

765 00:31:52.180 --> 00:31:53.900 is how we define the weights.

766 00:31:53.900 --> 00:31:56.010 And then finally, the most sophisticated approach

767 00:31:56.010 --> 00:31:57.240 at that time was this lag,

768 00:31:57.240 --> 00:31:58.440 weighted quantal sum regression

769 00:31:58.440 --> 00:32:00.090 that we talked a little bit about

770 00:32:01.160 --> 00:32:03.350 where we applied weighted quantal sum regression

771 00:32:03.350 --> 00:32:05.660 separately to each exposure period,

772 00:32:05.660 --> 00:32:07.200 let that estimate the weights,

773 00:32:07.200 --> 00:32:08.970 create the little package of exposure,

774 00:32:08.970 --> 00:32:10.200 and then throw those packages

775 00:32:10.200 --> 00:32:13.400 into the regression model using CWVS.

776 00:32:13.400 --> 00:32:14.750 So once you have the weights,

777 00:32:14.750 --> 00:32:16.620 like once you condition on the weights

778 00:32:16.620 --> 00:32:17.970 and you know the weights,

779 00:32:17.970 --> 00:32:19.410 you basically have one exposure

780 00:32:19.410 --> 00:32:21.920 and that exposure is the package,

781 00:32:21.920 --> 00:32:24.080 the mixture package that you've made.

782 00:32:24.080 --> 00:32:25.300 So the model,

783 00:32:25.300 --> 00:32:27.650 the modeling becomes much simpler in that case.

784 00:32:29.250 --> 00:32:34.250 So how did we go about to test these different methods?

785 00:32:34.660 --> 00:32:36.100 Well, we started very simply.

786 00:32:36.100 --> 00:32:41.020 So these represent the weights cross exposure period.

787 00:32:41.020 --> 00:32:43.420 In this case, I'm pretending like there's only five weeks

788 00:32:43.420 --> 00:32:44.810 in the exposure set.

789 00:32:44.810 --> 00:32:47.320 In reality, I let that vary for each data set

790 00:32:47.320 --> 00:32:51.260 the length and the start time of the exposure window changed

791 00:32:51.260 --> 00:32:52.093 but for this case,

792 00:32:52.093 --> 00:32:54.550 we assumed it started at pregnancy week one

793 00:32:54.550 --> 00:32:56.060 and went to week five.

794 00:32:56.060 --> 00:32:57.380 And so in the simplest case,

795 00:32:57.380 --> 00:32:59.500 we had just assumed there was one pollutant at play

796 00:32:59.500 --> 00:33:01.640 and it stayed constant across the exposure period.

797 00:33:01.640 --> 00:33:03.180 This is really simple.

798 00:33:03.180 --> 00:33:07.670 One pollutant is driving the entire risk that we're seeing.

799 00:33:07.670 --> 00:33:11.480 In another setting, we assumed that there were two,

800 00:33:11.480 --> 00:33:13.370 but there was no changes over time.

801 00:33:13.370 --> 00:33:15.390 They were always static across time

802 00:33:15.390 --> 00:33:17.520 and three, there were three that were coming into play

803 00:33:17.520 --> 00:33:20.253 at four, four, and then five, five of them,

804 00:33:21.190 --> 00:33:23.430 obviously as more come online

805 00:33:23.430 --> 00:33:25.800 and become important players in the mixture.

806 00:33:25.800 --> 00:33:27.440 The weights generally go down

807 00:33:27.440 --> 00:33:29.893 because all of lots of these have to be non zero.

808 00:33:30.850 --> 00:33:32.090 In setting B,

809 00:33:32.090 --> 00:33:34.040 we wanted to allow for some variability

810 00:33:34.040 --> 00:33:35.130 among the important pollutants.

811 00:33:35.130 --> 00:33:38.330 So we still allow for the same important pollutants

812 00:33:38.330 --> 00:33:40.683 to be important at each exposure period,

813 00:33:41.650 --> 00:33:43.310 but we allowed their relative contribution

814 00:33:43.310 --> 00:33:44.290 to change across time.

815 00:33:44.290 --> 00:33:47.300 So early on in pregnancy, this one was important,

816 00:33:47.300 --> 00:33:48.730 but then it's contribution went down

817 00:33:48.730 --> 00:33:52.660 and it was kind of surpassed by number two here

818 00:33:52.660 --> 00:33:53.800 at pollutant two,

819 00:33:53.800 --> 00:33:55.630 and then they can keep swapping in and out

820 00:33:55.630 --> 00:33:57.200 across the exposure.

821 00:33:57.200 --> 00:34:01.960 And in setting C it was complete chaos essentially

822 00:34:01.960 --> 00:34:03.620 different pollutants could come online

823 00:34:03.620 --> 00:34:05.180 and then leave and become important

824 00:34:05.180 --> 00:34:07.140 or not important go to zero.

825 00:34:07.140 --> 00:34:09.580 We don't anticipate this would ever,

826 00:34:09.580 --> 00:34:11.210 or this would be the case,

827 00:34:11.210 --> 00:34:13.290 but it would be nice to know if our model

828 00:34:13.290 --> 00:34:17.100 can somehow collapse and kind of accommodate this reckless,

829 00:34:17.100 --> 00:34:18.713 this wild behavior, I guess.

830 00:34:20.370 --> 00:34:21.203 So, yeah,

831 00:34:21.203 --> 00:34:23.450 this is something that kind of testing the extreme

832 00:34:23.450 --> 00:34:26.630 of all these methods is what we were trying to do here.

833 00:34:26.630 --> 00:34:28.130 So we'll jump right into the results.

834 00:34:28.130 --> 00:34:30.750 Just to give you a sense of what happened

835 00:34:30.750 --> 00:34:32.040 when we tested these models

836 00:34:32.040 --> 00:34:34.390 with lots of simulated data sets,

837 00:34:34.390 --> 00:34:38.600 CWVS mix continuously and kind of consistently

838 00:34:39.521 --> 00:34:42.870 was able to get the critical windows set

839 00:34:42.870 --> 00:34:45.420 more accurately than the other methods,

840 00:34:45.420 --> 00:34:48.370 which struggled kind of in varying degrees

841 00:34:48.370 --> 00:34:50.400 across these different settings,

842 00:34:50.400 --> 00:34:53.680 in terms of estimating the weight parameters.

843 00:34:53.680 --> 00:34:58.680 There's a generally CWVS mix has a lower means scored error

844 00:34:58.770 --> 00:35:01.600 so it's doing a better job of estimating these parameters,

845 00:35:01.600 --> 00:35:03.630 as you would expect, like with equal weights,

846 00:35:03.630 --> 00:35:05.220 if you assume each weight,

847 00:35:05.220 --> 00:35:08.580 each pollutant and interaction is playing

848 00:35:08.580 --> 00:35:09.930 an equal part in the story,

849 00:35:09.930 --> 00:35:12.540 you can be very bad off a lot of times,

850 00:35:12.540 --> 00:35:15.740 which is given, which is why these weights

851 00:35:15.740 --> 00:35:18.453 these values are so high for some of these methods.

852 00:35:19.500 --> 00:35:20.560 And finally,

853 00:35:20.560 --> 00:35:23.220 with the estimation of the regression parameters

854 00:35:23.220 --> 00:35:25.073 that describe the magnitude of risk.

855 00:35:26.270 --> 00:35:30.740 Generally, we're seeing improved performance with CWVS mix,

856 00:35:30.740 --> 00:35:31.710 but interestingly,

857 00:35:31.710 --> 00:35:34.300 at least at the time when we first saw this

858 00:35:34.300 --> 00:35:37.870 is that the equal weights method does a pretty good job

859 00:35:37.870 --> 00:35:42.800 of estimating these risk magnitude parameters

860 00:35:42.800 --> 00:35:46.060 as the number of important pollutants increases.

861 00:35:46.060 --> 00:35:48.100 So if you tell me that every one of your pollutants

862 00:35:48.100 --> 00:35:49.220 are important,

863 00:35:49.220 --> 00:35:52.610 then it's going to be hard to beat that something

864 00:35:52.610 --> 00:35:55.030 that gives all of the pollutants equal weight.

865 00:35:55.030 --> 00:35:56.650 So that's kind of the intuition behind it.

866 00:35:56.650 --> 00:35:58.210 As more pollutants become important,

867 00:35:58.210 --> 00:36:01.410 giving everything equal weight is not such a bad ideas,

868 00:36:01.410 --> 00:36:04.340 almost it's just averaging away some of that error,

869 00:36:04.340 --> 00:36:06.590 but generally, we're still doing well.

870 00:36:06.590 --> 00:36:08.600 And specifically in comparison

871 00:36:08.600 --> 00:36:10.360 to the lag weight quantile sum regression,

872 00:36:10.360 --> 00:36:11.910 that's really importantly,

873 00:36:11.910 --> 00:36:13.240 'cause at the time this was the kind

874 00:36:13.240 --> 00:36:14.810 of the main method out there

875 00:36:14.810 --> 00:36:17.440 that aimed to do the same thing we were doing.

876 00:36:17.440 --> 00:36:21.190 So in summary here with a simulation study,

877 00:36:21.190 --> 00:36:26.190 we did really well in critical in terms of accuracy, sorry,

878 00:36:26.250 --> 00:36:27.610 weight parameter estimation,

879 00:36:27.610 --> 00:36:32.420 and even in the risk magnitude parameter estimation.

880 00:36:32.420 --> 00:36:33.803 So models that don't have,

881 00:36:33.803 --> 00:36:36.170 that they don't actually estimate weights are more efficient

882 00:36:36.170 --> 00:36:37.250 when the complexity

883 00:36:37.250 --> 00:36:39.990 or the number of important pollutants grow

884 00:36:39.990 --> 00:36:42.260 and a little bit about the variable selection

885 00:36:42.260 --> 00:36:44.650 that we introduced with these latent variables.

886 00:36:44.650 --> 00:36:46.890 It appeared to do really well again,

887 00:36:46.890 --> 00:36:50.780 as the number of important pollutants was relatively small.

888 00:36:50.780 --> 00:36:53.820 So if you have lots of pollutants that are important

889 00:36:53.820 --> 00:36:55.660 and their interactions are important,

890 00:36:55.660 --> 00:36:58.510 it was hard for the variable section process

891 00:36:58.510 --> 00:36:59.460 to kind of tease out

892 00:36:59.460 --> 00:37:01.140 when something's included or excluded.

893 00:37:01.140 --> 00:37:03.820 It tended to just say everything was included.

894 00:37:03.820 --> 00:37:06.270 So something to keep in mind,

895 00:37:06.270 --> 00:37:09.650 I guess, as a limitation per perhaps of this approach.

896 00:37:09.650 --> 00:37:13.050 All right, so now onto the real data application

897 00:37:13.050 --> 00:37:14.110 that we had,

898 00:37:14.110 --> 00:37:18.380 and this is part of a larger kind of climate change,

899 00:37:18.380 --> 00:37:21.370 heat preterm birth study,

900 00:37:21.370 --> 00:37:25.980 we collected lots of state specific data birth records

901 00:37:25.980 --> 00:37:30.980 for all the way back to 1990 for maybe 12, 14 states.

902 00:37:31.240 --> 00:37:34.000 And so this one was set in New Jersey,

903 00:37:34.000 --> 00:37:35.890 but we focused on stillbirth given

904 00:37:35.890 --> 00:37:38.410 their really strong stillbirth data collection

905 00:37:39.680 --> 00:37:43.320 kind of methodology that New Jersey was using.

906 00:37:43.320 --> 00:37:46.120 So stillbirth the death or loss of a baby,

907 00:37:46.120 --> 00:37:48.370 at least 20 weeks of pregnancy affects about

908 00:37:48.370 --> 00:37:50.720 one in 160 births in the US.

909 00:37:50.720 --> 00:37:53.443 There are some known maternal risk factors,

910 00:37:54.450 --> 00:37:59.450 black mother, 35 years age or more of age,

911 00:38:00.420 --> 00:38:03.110 low SES, smoking, et cetera.

912 00:38:03.110 --> 00:38:06.170 And recent literature review meta analysis suggest that,

913 00:38:06.170 --> 00:38:10.860 PM 2.5 CO2 and O3 are associated with increased risk.

914 00:38:10.860 --> 00:38:12.560 This was really recent,

915 00:38:12.560 --> 00:38:14.440 but that more studies are definitely needed.

916 00:38:14.440 --> 00:38:16.690 There's not a lot as in comparison

917 00:38:16.690 --> 00:38:18.250 to some of the other adverse birth outcomes,

918 00:38:18.250 --> 00:38:21.290 there's not as much done with stillbirth, at least.

919 00:38:21.290 --> 00:38:23.240 However a majority of these previous studies

920 00:38:23.240 --> 00:38:26.900 have focused on again, single pollutant approaches,

921 00:38:26.900 --> 00:38:31.380 wide exposure periods like the entire relevant pregnancy

922 00:38:31.380 --> 00:38:33.513 before the delivery.

923 00:38:34.600 --> 00:38:35.640 So there is a need

924 00:38:35.640 --> 00:38:37.440 for kind of multiple pollutant critical window

925 00:38:37.440 --> 00:38:38.273 methods in this setting.

926 00:38:38.273 --> 00:38:42.540 So this is what kind of made us think about

927 00:38:42.540 --> 00:38:43.650 developing this methodology,

928 00:38:43.650 --> 00:38:46.483 but also applying it in this case study.

929 00:38:47.530 --> 00:38:50.490 So a little bit about the data we had access to.

930 00:38:50.490 --> 00:38:52.360 We had live birth and fetal death records

931 00:38:52.360 --> 00:38:55.290 from New Jersey from 2005 to 14.

932 00:38:55.290 --> 00:38:57.410 We included singletons with gestational age

933 00:38:57.410 --> 00:38:58.420 of at least 20 weeks,

934 00:38:58.420 --> 00:39:02.310 no birth defects, conception date in 25 to 2005 to 2013,

935 00:39:04.820 --> 00:39:06.530 we ran a case control analysis here

936 00:39:06.530 --> 00:39:09.590 where we five link live births were linked

937 00:39:09.590 --> 00:39:12.710 with each stillbirth matching only on race ethnicity.

938 00:39:12.710 --> 00:39:15.500 And we actually ended up running these analysis separately

939 00:39:15.500 --> 00:39:17.040 for each group non-Hispanic black,

940 00:39:17.040 --> 00:39:19.240 non-Hispanic white and Hispanic.

941 00:39:19.240 --> 00:39:21.660 And in terms of what our exposures,

942 00:39:21.660 --> 00:39:23.560 we included weekly pollution exposures

943 00:39:23.560 --> 00:39:28.560 through gestational week 20 were included in this analysis.

944 00:39:29.970 --> 00:39:32.300 All right, a little bit about the pollutants

945 00:39:32.300 --> 00:39:35.120 I mentioned we relied on a data fusion model

946 00:39:35.120 --> 00:39:39.880 that gave us kind of fine scale spatially

947 00:39:39.880 --> 00:39:44.880 and temporally estimates of 12 pollutants across New Jersey

948 00:39:45.590 --> 00:39:48.580 across the US actually, but focusing here on New Jersey.

949 00:39:48.580 --> 00:39:50.480 So you can see the pollutants listed here

950 00:39:50.480 --> 00:39:53.500 and we linked each woman's residence at delivery

951 00:39:53.500 --> 00:39:56.670 with the closest grid be where data were available

952 00:39:56.670 --> 00:39:58.870 or the estimates and predictions were available

953 00:39:58.870 --> 00:40:00.360 and assigned weekly exposures across

954 00:40:00.360 --> 00:40:02.330 the first 20 weeks of gestation.

955 00:40:02.330 --> 00:40:03.700 I know there's always a lot of pushback

956 00:40:03.700 --> 00:40:04.640 in these birth records

957 00:40:04.640 --> 00:40:08.120 because we don't have residential mobility,

958 00:40:08.120 --> 00:40:10.270 we don't have sense of like how often people move.

959 00:40:10.270 --> 00:40:12.840 And we know moving is differentialable

960 00:40:12.840 --> 00:40:15.140 by socioeconomic status for example,

961 00:40:15.140 --> 00:40:15.973 there are a lot of factors

962 00:40:15.973 --> 00:40:18.470 that influence moving during pregnancy,

963 00:40:18.470 --> 00:40:22.070 but if maybe this will make you feel somewhat better,

964 00:40:22.070 --> 00:40:23.683 but we did a study in 2019,

965 00:40:24.961 --> 00:40:26.510 the kind of assess the robustness

966 00:40:26.510 --> 00:40:29.160 of these critical window methods more generally

967 00:40:29.160 --> 00:40:31.860 to lots of different sources of error,

968 00:40:31.860 --> 00:40:34.042 including residential mobility

969 00:40:34.042 --> 00:40:35.550 and the results were actually very promising.

970 00:40:35.550 --> 00:40:39.190 I thought so the findings are robust generally

971 00:40:39.190 --> 00:40:43.140 to kind of this exposure misclassification

972 00:40:43.140 --> 00:40:46.023 or exposure error that's introduced through mobility.

973 00:40:47.440 --> 00:40:49.410 All right, so in summary,

974 00:40:49.410 --> 00:40:53.330 I guess for the data we had around 1300 non-Hispanic black,

975 00:40:53.330 --> 00:40:56.050 stillbirths in this time 928 Hispanic,

976 00:40:56.050 --> 00:40:58.990 and 1100 non-Hispanic white.

977 00:40:58.990 --> 00:41:02.120 our covariates that we included were a year of conception,

978 00:41:02.120 --> 00:41:04.940 season of conception to control for this kind of seasonality

979 00:41:04.940 --> 00:41:07.853 and long term time trends and pollution exposure,

980 00:41:08.700 --> 00:41:12.560 tobacco use indicator, age category, education.

981 00:41:12.560 --> 00:41:14.420 We had this sex of the fetus

982 00:41:14.420 --> 00:41:17.500 and to control for spatial kind of residual correlation.

983 00:41:17.500 --> 00:41:20.530 We actually included latitude, longitude

984 00:41:20.530 --> 00:41:23.999 of the residents had delivery and their interaction term

985 00:41:23.999 --> 00:41:25.730 as a pre-screening

986 00:41:25.730 --> 00:41:27.520 because we had 12 pollutants to work with.

987 00:41:27.520 --> 00:41:30.450 We didn't wanna introduce a lot of noise if possible,

988 00:41:30.450 --> 00:41:31.340 into the new framework.

989 00:41:31.340 --> 00:41:35.360 So we did a pre-run of the original critical window variable

990 00:41:35.360 --> 00:41:37.320 selection on each pollutant individually,

991 00:41:37.320 --> 00:41:40.380 as most analysis would do anyway,

992 00:41:40.380 --> 00:41:42.540 and identified a subset across all

993 00:41:42.540 --> 00:41:44.920 of the different data sets and by different data sets.

994 00:41:44.920 --> 00:41:46.870 I mean the non-Hispanic black,

995 00:41:46.870 --> 00:41:49.180 non-Hispanic white, and Hispanic.

996 00:41:49.180 --> 00:41:54.140 So all of the relevant and kind of significant exposures

997 00:41:54.140 --> 00:41:56.750 that came up and during any exposure period

998 00:41:56.750 --> 00:41:58.997 were included as a subset into this bigger framework.

999 00:41:58.997 --> 00:42:03.350 And so in total, we had PM 2.5 sulfate, nitrogen oxide,

1000 00:42:03.350 --> 00:42:05.410 ammonium, and nitrate that kind

1001 00:42:05.410 --> 00:42:08.623 of made this pre-screening period into the final subset.

1002 00:42:10.186 --> 00:42:12.000 So here is some of the output

1003 00:42:12.000 --> 00:42:13.930 that we thought was interesting.

1004 00:42:13.930 --> 00:42:14.970 There's a lot of output

1005 00:42:14.970 --> 00:42:16.940 that can be shown as you already know.

1006 00:42:16.940 --> 00:42:19.960 I guess now there's weight at each exposure period,

1007 00:42:19.960 --> 00:42:21.733 there's regression parameters,

1008 00:42:21.733 --> 00:42:23.597 there's just a lot that can happen here

1009 00:42:23.597 --> 00:42:25.730 and there's interactions, there's main effects,

1010 00:42:25.730 --> 00:42:29.080 but first let's focus on the first column here,

1011 00:42:29.080 --> 00:42:30.950 and this is at least something we can hold onto

1012 00:42:30.950 --> 00:42:35.180 that we understand from previous work in this space.

1013 00:42:35.180 --> 00:42:39.760 So what we can see for the non-Hispanic black population

1014 00:42:39.760 --> 00:42:42.020 that we were working with in New Jersey during this time,

1015 00:42:42.020 --> 00:42:43.790 that elevated exposure,

1016 00:42:43.790 --> 00:42:46.240 I'm not gonna say to what yet but elevated exposure

1017 00:42:46.240 --> 00:42:49.890 to some combination of these five pollutants

1018 00:42:49.890 --> 00:42:51.470 during pregnancy week two,

1019 00:42:51.470 --> 00:42:54.330 and then later on in the pregnancy, 16, 17,

1020 00:42:54.330 --> 00:42:59.150 and 20 actually led to increased odds.

1021 00:42:59.150 --> 00:43:03.130 So these are odds or ratios being presented of excuse me,

1022 00:43:03.130 --> 00:43:04.390 of stillbirth.

1023 00:43:04.390 --> 00:43:07.430 And so we can kind of take these in and say, 1024 00:43:07.430 --> 00:43:09.020 we get a sense of the critical windows

1025 00:43:09.020 --> 00:43:10.180 that are identified.

1026 00:43:10.180 --> 00:43:12.810 We also get a sense of the variable selection component

1027 00:43:12.810 --> 00:43:13.670 that I mentioned

1028 00:43:13.670 --> 00:43:16.500 and in this case, they line up pretty perfectly.

1029 00:43:16.500 --> 00:43:19.190 These are consistently in the model actually included

1030 00:43:19.190 --> 00:43:21.360 in the Bayesian variable selection model,

1031 00:43:21.360 --> 00:43:24.930 but also they're when they are in the model they're positive

1032 00:43:24.930 --> 00:43:26.860 So there this risk is in the right direction.

1033 00:43:26.860 --> 00:43:31.840 So more pollution during these pregnancy windows,

1034 00:43:31.840 --> 00:43:34.250 more risk of stillbirth in this population.

1035 00:43:34.250 --> 00:43:35.330 Now the question becomes,

1036 00:43:35.330 --> 00:43:36.180 well, what are you talking about

1037 00:43:36.180 --> 00:43:38.060 when you talk about the exposure?

1038 00:43:38.060 --> 00:43:40.430 Like, what is the mixture that you're talking about

1039 00:43:40.430 --> 00:43:42.040 in week two, for example?

1040 00:43:42.040 --> 00:43:43.240 Because we have five pollutants

1041 00:43:43.240 --> 00:43:45.200 and their interactions floating around.

1042 00:43:45.200 --> 00:43:48.910 So focusing first, so now let's move to the second column.

1043 00:43:48.910 --> 00:43:50.500 This represents the interactions,

1044 00:43:50.500 --> 00:43:53.240 this top part and the bottom part represents, I'm sorry,

1045 00:43:53.240 --> 00:43:54.077 this is main effects.

1046 00:43:54.077 --> 00:43:57.040 And the bottom part represents interactions.

1047 00:43:57.040 --> 00:44:00.760 So you can see ammonium is playing a big role throughout

1048 00:44:00.760 --> 00:44:02.300 until week 16, 1049 00:44:02.300 --> 00:44:05.960 which is dominated sharply by nitrogen oxides.

1050 00:44:05.960 --> 00:44:08.110 And then ammonium comes back into play here

1051 00:44:10.530 --> 00:44:12.480 in terms of the interactions that are important,

1052 00:44:12.480 --> 00:44:15.980 it looks like PM 2.5 and ammonium early on.

1053 00:44:15.980 --> 00:44:19.760 And then later on it's nitrogen oxides and ammonium

1054 00:44:19.760 --> 00:44:20.820 kind of come into play.

1055 00:44:20.820 --> 00:44:22.870 So a lot of this is noise.

1056 00:44:22.870 --> 00:44:25.310 I did not show you the variable section component,

1057 00:44:25.310 --> 00:44:27.520 but it probably would be nice

1058 00:44:27.520 --> 00:44:29.360 to kind of gray these out

1059 00:44:29.360 --> 00:44:32.170 if they're not selected in the model.

1060 00:44:32.170 --> 00:44:34.150 But a lot of these actually are selected in the model

1061 00:44:34.150 --> 00:44:35.030 with our variable selection.

1062 00:44:35.030 --> 00:44:37.000 So while these look to be non zero weights,

1063 00:44:37.000 --> 00:44:40.560 some of them are actually exactly zero essentially

1064 00:44:40.560 --> 00:44:42.760 because of the variable selection component.

1065 00:44:43.750 --> 00:44:44.583 But there's so much output,

1066 00:44:44.583 --> 00:44:46.800 it's hard to figure out what exactly

1067 00:44:46.800 --> 00:44:48.360 to show in a digestible way.

1068 00:44:48.360 --> 00:44:49.630 So this is where we landed.

1069 00:44:49.630 --> 00:44:52.010 So, interesting results you get to see how

1070 00:44:52.010 --> 00:44:54.990 the exposure kind of the mixture transitions

1071 00:44:54.990 --> 00:44:56.810 across exposure time,

1072 00:44:56.810 --> 00:44:59.010 you get to see what impact that has

1073 00:44:59.010 --> 00:45:02.820 on the actual risk of the outcome that you're talking about.

1074 00:45:02.820 --> 00:45:06.490 So a nice, I think coherent story can come,

1075 00:45:06.490 --> 00:45:09.560 can be told, if you're picturing your own analysis here,

1076 00:45:09.560 --> 00:45:11.720 you get to talk about the risk overall

1077 00:45:11.720 --> 00:45:14.170 to the mixture kind of combination or profile,

1078 00:45:14.170 --> 00:45:16.420 but also then dig deeper into individual weeks

1079 00:45:16.420 --> 00:45:17.850 and talk about which ones are important,

1080 00:45:17.850 --> 00:45:20.100 which interactions are reporting for example.

1081 00:45:20.970 --> 00:45:22.400 For the non-Hispanic white,

1082 00:45:22.400 --> 00:45:24.490 there was very little indication

1083 00:45:25.390 --> 00:45:27.730 that these pollutants were planning a role,

1084 00:45:27.730 --> 00:45:30.150 I guess, in the kind of development of stillbirth

1085 00:45:30.150 --> 00:45:32.850 or the risk of stillbirth in this population

1086 00:45:32.850 --> 00:45:35.720 and for the Hispanic population,

1087 00:45:35.720 --> 00:45:38.350 it looked like there potentially was some uptick here

1088 00:45:38.350 --> 00:45:41.780 at the end, but nothing significantly jumped out either.

1089 00:45:41.780 --> 00:45:44.830 And so at this point, it's almost...

1090 00:45:44.830 --> 00:45:47.060 You don't start to investigate

1091 00:45:47.060 --> 00:45:50.360 and over interpret these white parameters,

1092 00:45:50.360 --> 00:45:51.780 given that you're not seeing anything here.

1093 00:45:51.780 --> 00:45:55.890 So I kind of consider this to be noise essentially

1094 00:45:57.120 --> 00:46:00.120 for the Hispanic and non-Hispanic white results for example.

1095 00:46:01.380 --> 00:46:04.490 So a little brief kind of wrapping up here,

1096 00:46:04.490 --> 00:46:05.950 summary of our findings is that,

1097 00:46:05.950 --> 00:46:08.420 for the non-Hispanic black data set

1098 00:46:08.420 --> 00:46:10.300 and variable selection results

1099 00:46:10.300 --> 00:46:12.870 PM 2.5 and its chemical constituents

1100 00:46:12.870 --> 00:46:14.830 are primary drivers of risk.

1101 00:46:14.830 --> 00:46:17.240 And this was actually changing across exposure week.

1102 00:46:17.240 --> 00:46:20.730 So driven in week two by a lot of interactions

1103 00:46:20.730 --> 00:46:22.650 and kind of individual pieces.

1104 00:46:22.650 --> 00:46:27.080 Week 16, mainly heavily driven by nitrogen oxides

1105 00:46:27.080 --> 00:46:28.363 and then week 17,

1106 00:46:29.450 --> 00:46:31.400 one or two pollutants and their interactions.

1107 00:46:31.400 --> 00:46:33.110 So all the other interactions

1108 00:46:33.110 --> 00:46:36.410 that are not listed here among the five variables

1109 00:46:36.410 --> 00:46:39.280 were actually not significantly important here.

1110 00:46:39.280 --> 00:46:44.280 So no nothing kind of nothing seen

1111 00:46:45.620 --> 00:46:48.433 for the non-Hispanic white and Hispanic populations.

1112 00:46:49.570 --> 00:46:51.430 And I guess in conclusion,

1113 00:46:51.430 --> 00:46:55.850 we introduce CWVS mix with which combines smooth variable

1114 00:46:55.850 --> 00:46:58.050 Bayesian variable selection in the weights

1115 00:46:58.050 --> 00:46:59.310 and the regression parameters

1116 00:46:59.310 --> 00:47:01.680 with interpretable weighted quantile sum regression

1117 00:47:01.680 --> 00:47:04.210 shrinkage to identify critical windows,

1118 00:47:04.210 --> 00:47:05.900 but also kind of understand

1119 00:47:05.900 --> 00:47:10.310 and kind of dig deeper into the mixture itself.

1120 00:47:10.310 --> 00:47:13.350 And importantly, at least from our perspective

1121 00:47:13.350 --> 00:47:16.000 is that CWVS mix seemed to offer something

1122 00:47:17.260 --> 00:47:19.000 that the existing methods didn't,

1123 00:47:19.000 --> 00:47:21.760 which so consistently outperforming these other methods

1124 00:47:21.760 --> 00:47:24.730 for identifying the true critical window set,

1125 00:47:24.730 --> 00:47:25.970 estimating weight parameters, 1126 00:47:25.970 --> 00:47:29.420 which is really important for interpreting the mixtures

1127 00:47:29.420 --> 00:47:32.250 and then estimating the risk magnitude parameters as well.

1128 00:47:32.250 --> 00:47:34.020 And our stillbirth results from New Jersey

1129 00:47:34.020 --> 00:47:37.210 were in qualitative agreement with those in the literature,

1130 00:47:37.210 --> 00:47:42.210 in that PM 2.5 consistent signal across many studies

1131 00:47:42.470 --> 00:47:44.840 while developing kind of gaining new insights

1132 00:47:44.840 --> 00:47:48.850 regarding the exposure timing in this particular study,

1133 00:47:48.850 --> 00:47:50.090 obviously more work is needed.

1134 00:47:50.090 --> 00:47:52.660 And so I guess before jumping to this,

1135 00:47:52.660 --> 00:47:57.120 we were working on extending this framework.

1136 00:47:57.120 --> 00:47:59.690 So I'm working with the group at Emory here

1137 00:47:59.690 --> 00:48:03.190 on extending this framework to allow the windows themselves

1138 00:48:03.190 --> 00:48:06.750 to vary by something like socioeconomic status

1139 00:48:06.750 --> 00:48:10.380 or race ethnicity, or other individual level factors.

1140 00:48:10.380 --> 00:48:13.390 So there's this effect modification floating around now

1141 00:48:13.390 --> 00:48:14.770 plus the mixtures.

1142 00:48:14.770 --> 00:48:19.280 So it's becoming a really big task to kind of do all of this

1143 00:48:19.280 --> 00:48:20.113 in a single framework,

1144 00:48:20.113 --> 00:48:23.020 but we're trying to take baby steps, essentially.

1145 00:48:23.020 --> 00:48:25.060 We like where we're at now, we think it works well,

1146 00:48:25.060 --> 00:48:27.590 it's robust, it fits well

1147 00:48:27.590 --> 00:48:29.870 and can we extend it next to the questions

1148 00:48:29.870 --> 00:48:30.890 that are being asked?

1149 00:48:30.890 --> 00:48:34.330 So again, if you're someone who is asking similar questions,

1150 00:48:34.330 --> 00:48:35.410 please, we can talk.

1151 00:48:35.410 --> 00:48:38.250 And I really like enjoy sitting down with collaborators

1152 00:48:38.250 --> 00:48:39.570 and trying to figure out,

1153 00:48:39.570 --> 00:48:41.480 develop new methods that can answer the questions

1154 00:48:41.480 --> 00:48:42.950 that they have.

1155 00:48:42.950 --> 00:48:45.130 But if you find that,

1156 00:48:45.130 --> 00:48:46.840 your setting can already be answered

1157 00:48:46.840 --> 00:48:49.800 by some of these methods that I've discussed today

1158 00:48:49.800 --> 00:48:52.440 on my website and on my GitHub site,

1159 00:48:52.440 --> 00:48:55.220 I keep a lot of these packages that I've created

1160 00:48:55.220 --> 00:48:56.250 with help documentation

1161 00:48:56.250 --> 00:48:59.490 and then you are always free to reach out to me as well.

1162 00:48:59.490 --> 00:49:02.670 But if you're looking to do this original Gaussian process,

1163 00:49:02.670 --> 00:49:03.780 critical window estimation,

1164 00:49:03.780 --> 00:49:05.740 we have a package for that.

1165 00:49:05.740 --> 00:49:08.330 Howard Chang at Emory, go through my website again,

1166 00:49:08.330 --> 00:49:10.850 you'll find this his survival version

1167 00:49:10.850 --> 00:49:13.020 of the model up there as well.

1168 00:49:13.020 --> 00:49:16.420 CWVS in this original form is there for download

1169 00:49:16.420 --> 00:49:17.430 the spatial version,

1170 00:49:17.430 --> 00:49:20.550 which hopefully we're thinking about extending in

1171 00:49:22.040 --> 00:49:24.980 soon to account for something like oxidative potential

1172 00:49:24.980 --> 00:49:27.980 of these pollutants that's also there.

1173 00:49:27.980 --> 00:49:29.500 And then the newly developed methodology

1174 00:49:29.500 --> 00:49:32.540 is also there for download and for use as well.

1175 00:49:32.540 --> 00:49:34.730 And this obviously could not have happened

1176 00:49:34.730 --> 00:49:38.440 without collaborators, including Howard at Chang at Emory,

1177 00:49:38.440 --> 00:49:41.850 Lauren at RTI did a lot of data management,

1178 00:49:41.850 --> 00:49:44.750 Matthew Strickland, and Lindsey

1179 00:49:44.750 --> 00:49:47.010 at University of Nevada Reno,

1180 00:49:47.010 --> 00:49:49.940 and then James for providing the,

1181 00:49:49.940 --> 00:49:53.290 or helping with the data fusion output as well.

1182 00:49:53.290 --> 00:49:55.830 And here, this grant support here

1183 00:49:55.830 --> 00:49:58.010 that I mentioned in extreme heat duration,

1184 00:49:58.010 --> 00:49:59.370 and then data integration methods

1185 00:49:59.370 --> 00:50:01.250 for environmental exposures.

1186 00:50:01.250 --> 00:50:05.020 So yeah, please feel free to reach out

1187 00:50:05.020 --> 00:50:05.950 if you have any questions.

1188 00:50:05.950 --> 00:50:08.530 This work that I went over today

1189 00:50:08.530 --> 00:50:11.530 is in press at Annals of Applied Statistics,

1190 00:50:11.530 --> 00:50:12.430 not on their website yet,

1191 00:50:12.430 --> 00:50:14.640 but should be really soon.

1192 00:50:14.640 --> 00:50:16.440 But I think there's a version on archive

1193 00:50:16.440 --> 00:50:17.273 if you're interested

1194 00:50:17.273 --> 00:50:18.580 or if you want the most up to date version.

1195 00:50:18.580 --> 00:50:19.770 I actually think I sent it tomorrow

1196 00:50:19.770 --> 00:50:21.910 who may have passed it out to the class,

1197 00:50:21.910 --> 00:50:23.500 but yeah, definitely feel free to reach out

1198 00:50:23.500 --> 00:50:26.855 if there are any questions or anything I can help with.

1199 00:50:26.855 --> 00:50:28.083 Yeah, that's it.

1200 00:50:29.974 --> 00:50:30.807 <v ->Thank you so much.</v>

1201 00:50:30.807 --> 00:50:33.057 (applause)

1202 00:50:35.310 --> 00:50:38.780 Our students were impressed with this

1203 00:50:38.780 --> 00:50:41.613 heavy quantitative focused lecture.

1204 00:50:43.925 --> 00:50:45.240 We already collected some questions

1205 00:50:45.240 --> 00:50:47.030 from our students already,

1206 00:50:47.030 --> 00:50:49.500 but for folks who are joining online,

1207 00:50:49.500 --> 00:50:51.240 if you do have questions,

1208 00:50:51.240 --> 00:50:53.850 please feel free to put in the chat box.

1209 00:50:53.850 --> 00:50:55.530 So the first question,

1210 00:50:55.530 --> 00:50:58.070 one of the students is observing that

1211 00:50:58.070 --> 00:51:01.440 in your study, you found the elevator risk

1212 00:51:01.440 --> 00:51:04.640 was found in week two of the pregnancy,

1213 00:51:04.640 --> 00:51:05.630 which is very early.

1214 00:51:05.630 --> 00:51:10.630 So perhaps many pregnant women are not aware

1215 00:51:10.820 --> 00:51:12.700 of the pregnancy at that time.

1216 00:51:12.700 --> 00:51:15.740 So in terms of the intervention

1217 00:51:15.740 --> 00:51:18.810 at this early stage of pregnancy,

1218 00:51:18.810 --> 00:51:22.440 what's the kind of policy implications that we'll find?

1219 00:51:22.440 --> 00:51:24.130 <v ->Now that's a really great point.</v>

1220 00:51:24.130 --> 00:51:27.110 And this is something we've tried to,

1221 00:51:27.110 --> 00:51:29.390 we haven't figured out how to deal with either,

1222 00:51:29.390 --> 00:51:32.877 but has we've run into a number of interesting results

1223 00:51:34.930 --> 00:51:37.510 that we've seen early in the pregnancy.

1224 00:51:37.510 --> 00:51:38.490 We've particularly,

1225 00:51:38.490 --> 00:51:41.700 we've seen protective effects at some points

1226 00:51:41.700 --> 00:51:45.410 for like PM 2.5 exposure and pre-term pregnancy

1227 00:51:45.410 --> 00:51:46.700 very early on in the pregnancy.

1228 00:51:46.700 --> 00:51:50.570 And we believe it could be due to the exactly

1229 00:51:50.570 --> 00:51:51.403 what we're talking about.

1230 00:51:51.403 --> 00:51:52.970 People who don't actually know they're pregnant

1231 00:51:52.970 --> 00:51:54.320 at that point.

1232 00:51:54.320 --> 00:51:56.960 And so miscarriage is an issue

1233 00:51:56.960 --> 00:52:00.110 that isn't well kind of documented by a lot of these states.

1234 00:52:00.110 --> 00:52:02.920 There could be just fetal loss in general,

1235 00:52:02.920 --> 00:52:05.420 that we're not capturing in the birth records.

1236 00:52:05.420 --> 00:52:08.440 And so there's this population

1237 00:52:08.440 --> 00:52:11.840 that we're not even including in a lot of our analysis

1238 00:52:11.840 --> 00:52:13.050 that are lurking around

1239 00:52:13.050 --> 00:52:14.350 and kind of could be biasing

1240 00:52:14.350 --> 00:52:16.420 some of these early week results.

1241 00:52:16.420 --> 00:52:18.480 In terms of policy implications

1242 00:52:20.010 --> 00:52:20.940 it's a really good question.

1243 00:52:20.940 --> 00:52:25.300 I don't know other than if I guess it really,

1244 00:52:25.300 --> 00:52:26.900 if you're trying to get pregnant,

1245 00:52:26.900 --> 00:52:29.520 if you know you're on that, in that stage,

1246 00:52:29.520 --> 00:52:31.080 I mean, maybe it's helpful for you,

1247 00:52:31.080 --> 00:52:34.940 but if you're someone who doesn't know unanticipated

1248 00:52:34.940 --> 00:52:39.610 there's only so much that can go into outside

1249 00:52:39.610 --> 00:52:41.630 of just cleaner air altogether.

1250 00:52:41.630 --> 00:52:44.870 Which is something everyone can kind of agree on.

1251 00:52:44.870 --> 00:52:48.160 But I think it may only affect a subset of people

1252 00:52:48.160 --> 00:52:50.090 who are either attempting to get pregnant

1253 00:52:50.090 --> 00:52:52.250 or kind of really regimented and like,

1254 00:52:52.250 --> 00:52:54.870 know their schedule for example.

1255 00:52:54.870 --> 00:52:56.670 But there's this whole other issue about people

1256 00:52:56.670 --> 00:52:57.800 who aren't in our data set.

1257 00:52:57.800 --> 00:52:59.050 That's a really great point

1258 00:52:59.050 --> 00:53:02.380 and we have not figured out how to solve that yet.

1259 00:53:02.380 --> 00:53:03.800 <v Kai>Yet, tough question.</v>

1260 00:53:03.800 --> 00:53:04.830 Thanks, Josh.

1261 00:53:04.830 --> 00:53:06.000 We do have another question

1262 00:53:06.000 --> 00:53:08.520 from actually two students read this.

1263 00:53:08.520 --> 00:53:12.690 They really appreciate your talk about this new metrics.

1264 00:53:12.690 --> 00:53:15.500 And we realize this is the package.

1265 00:53:15.500 --> 00:53:19.210 Our package is available from your GitHub website.

1266 00:53:19.210 --> 00:53:21.207 So anyone who's interested in applying that

1267 00:53:21.207 --> 00:53:23.453 you can download the app package and run,

1268 00:53:24.600 --> 00:53:26.060 but the students are wondering like

1269 00:53:26.060 --> 00:53:29.980 beyond this time wearing air pollution mixtures

1270 00:53:29.980 --> 00:53:33.350 a lot other mixtures in terms of (indistinct)

1271 00:53:33.350 --> 00:53:35.560 like temperature, green space, other things.

1272 00:53:35.560 --> 00:53:39.050 So how does your approach this

1273 00:53:39.050 --> 00:53:43.190 the CWVS mix apply to a broader setting

1274 00:53:43.190 --> 00:53:45.026 of environment exposures?

1275 00:53:45.026 --> 00:53:48.030 <v ->I think, my push, and if you read the paper,</v>

1276 00:53:48.030 --> 00:53:49.710 you'll notice that I really push for people

1277 00:53:49.710 --> 00:53:53.930 to think about that in their own setting.

1278 00:53:53.930 --> 00:53:56.570 Cause I think it's generally applicable to any, 1279 00:53:56.570 --> 00:53:58.750 it doesn't have to be a pregnancy outcome.

1280 00:53:58.750 --> 00:54:01.300 It doesn't have to be air pollution.

1281 00:54:01.300 --> 00:54:04.150 What it does have to be is consistently measured

1282 00:54:04.150 --> 00:54:05.510 across some exposure period.

1283 00:54:05.510 --> 00:54:07.720 So I'll often get questions that,

1284 00:54:07.720 --> 00:54:10.480 I have two time periods measured,

1285 00:54:10.480 --> 00:54:13.800 in the first trimester and then in the third trimester,

1286 00:54:13.800 --> 00:54:15.710 can I fit your methodology?

1287 00:54:15.710 --> 00:54:18.950 Well, we need more fine grained exposure information.

1288 00:54:18.950 --> 00:54:21.130 That's consistent across the individuals

1289 00:54:21.130 --> 00:54:22.840 in order to estimate these critical windows.

1290 00:54:22.840 --> 00:54:25.120 So I think the only barrier for entry

1291 00:54:25.120 --> 00:54:27.390 is that you have consistently estimated

1292 00:54:27.390 --> 00:54:30.060 kind of exposures for the population of interest.

1293 00:54:30.060 --> 00:54:32.140 It doesn't matter so much what the exposure is now.

1294 00:54:32.140 --> 00:54:35.920 I say that, but if you're bringing binary exposures

1295 00:54:35.920 --> 00:54:38.220 and you have limit of detection issues,

1296 00:54:38.220 --> 00:54:39.520 there are obviously some issues

1297 00:54:39.520 --> 00:54:41.110 that will need to be sorted out,

1298 00:54:41.110 --> 00:54:43.930 but the framework itself should work really well.

1299 00:54:43.930 --> 00:54:46.470 The other covariate is, you'll notice that

1300 00:54:46.470 --> 00:54:49.490 a lot of my work has been focused on pregnancy outcomes

1301 00:54:49.490 --> 00:54:52.460 and that's because the exposure period is so well defined

1302 00:54:52.460 --> 00:54:55.200 if you're working with something like cancer for example, 1303 00:54:55.200 --> 00:54:58.083 well, how far do you extend back in time,

1304 00:54:59.520 --> 00:55:02.970 the exposures like how you could go years and years back.

1305 00:55:02.970 --> 00:55:05.793 So there's this cumulative idea as well.

1306 00:55:06.690 --> 00:55:08.100 That's really hard to understand

1307 00:55:08.100 --> 00:55:10.130 and these distributed lag models are great.

1308 00:55:10.130 --> 00:55:12.790 As long as you can a priority tell me

1309 00:55:12.790 --> 00:55:14.240 what the relevant exposure period is.

1310 00:55:14.240 --> 00:55:17.250 I can tell you if any of the interior parts

1311 00:55:17.250 --> 00:55:19.320 of that exposure period are important,

1312 00:55:19.320 --> 00:55:20.560 but if you're telling me you don't know

1313 00:55:20.560 --> 00:55:23.040 when the exposure period potentially started

1314 00:55:23.040 --> 00:55:24.640 or it's a completely different conversation.

1315 00:55:24.640 --> 00:55:26.750 So your outcome has to have,

1316 00:55:26.750 --> 00:55:28.530 or preferably would have some type

1317 00:55:28.530 --> 00:55:31.530 of relevant exposure period.

1318 00:55:31.530 --> 00:55:32.460 It's actually even better

1319 00:55:32.460 --> 00:55:34.850 for something like cardiac heart defects,

1320 00:55:34.850 --> 00:55:37.320 which we know the heart forms between like weeks three

1321 00:55:37.320 --> 00:55:38.600 and eight of pregnancy.

1322 00:55:38.600 --> 00:55:41.150 So you can really focus in on something like daily

1323 00:55:41.150 --> 00:55:42.350 or even sub daily

1324 00:55:42.350 --> 00:55:44.660 if you had that type of exposure information.

1325 00:55:44.660 --> 00:55:45.740 So yeah, those are the two,

1326 00:55:45.740 --> 00:55:46.840 generally it should work,

1327 00:55:46.840 --> 00:55:48.340 but just make sure you have a good sense

1328 00:55:48.340 --> 00:55:50.093 of the exposure period.

1329 00:55:51.940 --> 00:55:52.947 <v Kai>Very good point, thanks Josh.</v>

1330 00:55:52.947 --> 00:55:56.770 And we do have one comment from our on artist.

1331 00:55:56.770 --> 00:55:58.500 So I read Dr. Warren

1332 00:55:58.500 --> 00:56:01.410 could you please share your thought on applying

1333 00:56:01.410 --> 00:56:04.357 the critical window analysis?

1334 00:56:04.357 --> 00:56:07.305 (mutters)

1335 00:56:07.305 --> 00:56:08.255 <v ->Sorry, with what?</v>

1336 00:56:09.105 --> 00:56:10.100 (overlapping conversation)

1337 00:56:10.100 --> 00:56:11.290 That's a really great point.

1338 00:56:11.290 --> 00:56:14.150 So over, so I'm actually on sabbatical right now,

1339 00:56:14.150 --> 00:56:16.930 which is why I couldn't be there in person with you guys,

1340 00:56:16.930 --> 00:56:19.170 but over the sabbatical

1341 00:56:19.170 --> 00:56:22.320 I've developed the framework and the code

1342 00:56:22.320 --> 00:56:24.990 to account for binary outcomes,

1343 00:56:24.990 --> 00:56:28.440 continuous outcomes and count outcomes as well.

1344 00:56:28.440 --> 00:56:30.980 Luckily if you've taken my (indistinct) course

1345 00:56:30.980 --> 00:56:32.040 or you're gonna take it next fall,

1346 00:56:32.040 --> 00:56:34.120 you'll see how all of these connect

1347 00:56:34.120 --> 00:56:36.210 and lend themselves really nicely

1348 00:56:36.210 --> 00:56:39.680 to kind of full conditional distribution updates

1349 00:56:39.680 --> 00:56:41.050 that make the model fitting process

1350 00:56:41.050 --> 00:56:44.160 really kind of slick and nice.

1351 00:56:44.160 --> 00:56:46.700 So you can we have a negative binomial regression,

1352 00:56:46.700 --> 00:56:48.660 for example, that can do the same thing.

1353 00:56:48.660 --> 00:56:51.340 You just have count out outcome data,

1354 00:56:51.340 --> 00:56:53.340 if you have a continuous measure, for example,

1355 00:56:53.340 --> 00:56:54.990 so I'm really aiming this.

1356 00:56:54.990 --> 00:56:56.940 I hope this method doesn't just pop up

1357 00:56:56.940 --> 00:56:58.800 and then disappear, I want people to use it,

1358 00:56:58.800 --> 00:57:00.220 I want it to be useful.

1359 00:57:00.220 --> 00:57:01.680 And so that's why I'm trying to extend it

1360 00:57:01.680 --> 00:57:04.300 and trying to get people to use it in different contexts.

1361 00:57:04.300 --> 00:57:07.680 So, yeah, definitely I love those types of questions.

1362 00:57:07.680 --> 00:57:08.513 <v Kai>Thanks Josh.</v>

1363 00:57:08.513 --> 00:57:12.550 Because we actually have another (speech distorted)

1364 00:57:12.550 --> 00:57:14.490 So we have to end early

1365 00:57:14.490 --> 00:57:16.700 and we do have a lot of students questions

1366 00:57:16.700 --> 00:57:21.110 and I'm sure contact you for just once.

1367 00:57:21.110 --> 00:57:23.560 So thanks again, Josh for wonderful talk.

1368 00:57:23.560 --> 00:57:25.210 <v ->No, yeah thanks for being here.</v>