WEBVTT

1 00:00:02.010 --> 00:00:03.810 <v Maria>My name is Maria Ciarlegio</v>

2 00:00:03.810 --> 00:00:05.430 and I'm a faculty member

3 00:00:05.430 --> 00:00:07.710 in the Department of Biostatistics

4 00:00:07.710 --> 00:00:10.023 at the Yale School of Public Health.

5 00:00:10.890 --> 00:00:12.240 In this video series,

6 00:00:12.240 --> 00:00:14.880 I will introduce the clinical research process

7 00:00:14.880 --> 00:00:19.680 to prepare you to collaborate with a statistician.

8 00:00:19.680 --> 00:00:21.150 In the second video,

9 00:00:21.150 --> 00:00:24.750 we'll discuss some study designs that are commonly used

10 00:00:24.750 --> 00:00:27.180 in clinical research.

11 00:00:27.180 --> 00:00:30.570 The study design used to answer your research questions

12 00:00:30.570 --> 00:00:33.630 is determined by the goal of the research

13 00:00:33.630 --> 00:00:36.270 and the feasibility of different designs,

14 00:00:36.270 --> 00:00:39.240 including the availability of resources.

15 00:00:39.240 --> 00:00:42.150 For example, a randomized clinical trial

16 00:00:42.150 --> 00:00:45.990 can require a good amount of administrative support

17 00:00:45.990 --> 00:00:49.650 in addition to the required clinical involvement.

18 00:00:49.650 --> 00:00:53.010 The first major distinction between study types

19 00:00:53.010 --> 00:00:55.830 is whether the study investigator decides

20 00:00:55.830 --> 00:00:59.040 if a participant receives the exposure

21 00:00:59.040 --> 00:01:01.320 or the intervention.

22 00:01:01.320 --> 00:01:04.860 Studies where the investigator assigns the exposure

23 00:01:04.860 --> 00:01:06.990 are experimental studies,

24 00:01:06.990 --> 00:01:08.940 and studies where the investigator

25 00:01:08.940 --> 00:01:13.530 does not assign the exposure are observational studies.

26 00:01:13.530 --> 00:01:15.150 In observational studies,

27 00:01:15.150 --> 00:01:18.270 the investigator observes or documents

28 00:01:18.270 --> 00:01:20.910 the presence of exposures and outcomes

29 00:01:20.910 --> 00:01:24.903 as they naturally occur in a population or a sample.

30 00:01:25.740 --> 00:01:27.570 There are both descriptive

31 00:01:27.570 --> 00:01:30.660 and analytical observational studies.

32 00:01:30.660 --> 00:01:33.060 A descriptive observational study

33 00:01:33.060 --> 00:01:35.700 is used to describe characteristics

34 00:01:35.700 --> 00:01:38.250 of a sample or population.

35 00:01:38.250 --> 00:01:40.140 There's no comparator group

36 00:01:40.140 --> 00:01:42.960 because the goal is simply to describe,

37 00:01:42.960 --> 00:01:44.790 not formally compare.

38 00:01:44.790 --> 00:01:47.670 For example, you can use a descriptive study

39 00:01:47.670 --> 00:01:50.220 to estimate the prevalence of a disease

40 00:01:50.220 --> 00:01:54.570 or report the results of a patient survey of symptoms.

41 00:01:54.570 --> 00:01:56.700 They can also be used to inform the design

42 00:01:56.700 --> 00:02:00.420 of a future comparative or analytical study.

43 00:02:00.420 --> 00:02:03.210 For example, you may not have a good estimate

44 00:02:03.210 --> 00:02:05.190 of baseline prevalence of disease

45 00:02:05.190 --> 00:02:10.190 for a future study testing a new treatment on that disease.

46 00:02:10.230 --> 00:02:13.530 A descriptive study can be used to estimate this value

47 00:02:13.530 --> 00:02:16.260 to help you design that future study.

48 00:02:16.260 --> 00:02:20.070 The first type of descriptive study is the case report,

49 00:02:20.070 --> 00:02:24.570 which usually describes a patient presenting with an unusual

50 00:02:24.570 --> 00:02:26.610 or complicated disease.

51 00:02:26.610 --> 00:02:29.520 When more than one patient is described,

52 00:02:29.520 --> 00:02:31.650 it becomes a case series.

53 00:02:31.650 --> 00:02:36.090 Case reports and case series are useful for defining cases,

54 00:02:36.090 --> 00:02:39.510 generating hypotheses about the causes of disease,

55 00:02:39.510 --> 00:02:42.420 or use in clinical education.

56 00:02:42.420 --> 00:02:45.600 Finally, descriptive cross-sectional studies

57 00:02:45.600 --> 00:02:50.070 collect information on the presence or the level of one

58 00:02:50.070 --> 00:02:54.210 or more characteristic at one point in time.

59 00:02:54.210 --> 00:02:56.970 Those characteristics can include risk factors

60 00:02:56.970 --> 00:02:59.700 or different exposures and outcomes,

61 00:02:59.700 --> 00:03:02.580 such as the presence of disease.

62 00:03:02.580 --> 00:03:05.010 If the goal is to report the distribution

63 00:03:05.010 --> 00:03:07.320 of one or more of the characteristics,

64 00:03:07.320 --> 00:03:10.530 then the cross-sectional study is descriptive.

65 00:03:10.530 --> 00:03:13.440 However, if the goal is to assess the relationship

66 00:03:13.440 --> 00:03:16.110 between say presence of an exposure

67 00:03:16.110 --> 00:03:17.760 and presence of disease,

68 00:03:17.760 --> 00:03:20.940 then the cross-sectional study is analytical.

69 00:03:20.940 --> 00:03:23.610 And we'll talk about analytical studies next.

70 00:03:23.610 --> 00:03:25.740 But the takeaway on cross-sectional studies

71 00:03:25.740 --> 00:03:28.200 is that they provide a snapshot

72 00:03:28.200 --> 00:03:31.530 of the frequency of disease and patient characteristics

73 00:03:31.530 --> 00:03:33.363 at one point in time.

74 00:03:35.100 --> 00:03:38.640 Analytical observational studies include a comparator

75 00:03:38.640 --> 00:03:40.410 or control group.

76 00:03:40.410 --> 00:03:42.540 The goal is to formally establish

77 00:03:42.540 --> 00:03:46.800 or quantify an association between exposures and outcomes.

78 00:03:46.800 --> 00:03:49.230 Again, the exposures are naturally determined

79 00:03:49.230 --> 00:03:50.910 in observational studies,

80 00:03:50.910 --> 00:03:52.650 unlike experimental studies

81 00:03:52.650 --> 00:03:55.860 where the investigator assigns exposure usually

82 00:03:55.860 --> 00:03:58.320 in the form of different treatments

83 00:03:58.320 --> 00:04:00.870 the temporal direction of the study determines

84 00:04:00.870 --> 00:04:03.990 the type of analytical observational study.

85 00:04:03.990 --> 00:04:06.150 If the study assesses exposures

86 00:04:06.150 --> 00:04:08.370 and outcomes at the same point in time,

87 00:04:08.370 --> 00:04:10.170 the study is cross-sectional.

88 00:04:10.170 --> 00:04:12.360 Analytical cross-sectional studies

89 00:04:12.360 --> 00:04:15.570 as opposed to descriptive cross-sectional studies, again,

90 00:04:15.570 --> 00:04:17.700 are used to assess the relationship

91 00:04:17.700 --> 00:04:20.403 between exposure and disease.

92 00:04:21.408 --> 00:04:25.710 An analytical cross-sectional study was conducted in Japan

93 00:04:25.710 --> 00:04:27.150 to assess the relationship

94 00:04:27.150 --> 00:04:30.150 between non-alcoholic fatty liver disease

95 00:04:30.150 --> 00:04:31.980 and periodontal disease.

96 00:04:31.980 --> 00:04:34.110 Magnetic resonance elastography

97 00:04:34.110 --> 00:04:36.270 was used to measure liver stiffness

98 00:04:36.270 --> 00:04:38.460 in these patients with NAFLD.

99 00:04:38.460 --> 00:04:42.450 They found an association between P. gingivalis positivity

100 00:04:42.450 --> 00:04:46.350 and number of periodontal pockets and liver stiffness.

101 00:04:46.350 --> 00:04:48.900 The idea here is that the investigators

102 00:04:48.900 --> 00:04:52.290 are looking at these patients at a snapshot in time,

103 00:04:52.290 --> 00:04:55.323 not over time, in this cross-sectional study.

104 00:04:56.340 --> 00:04:59.910 A cohort study follows patients forward in time

105 00:04:59.910 --> 00:05:03.540 for development or occurrence of the outcome.

106 00:05:03.540 --> 00:05:06.240 Investigators identify a group of patients

107 00:05:06.240 --> 00:05:09.030 without the outcome or disease of interest.

108 00:05:09.030 --> 00:05:10.350 And of these patients,

109 00:05:10.350 --> 00:05:13.380 some are exposed to an exposure of interest

110 00:05:13.380 --> 00:05:15.990 and some are unexposed.

111 00:05:15.990 --> 00:05:20.280 We follow the exposed and unexposed groups forward in time

112 00:05:20.280 --> 00:05:22.470 for development of the outcome.

113 00:05:22.470 --> 00:05:25.620 If we observe a higher incidence of the outcome

114 00:05:25.620 --> 00:05:26.790 in the exposed,

115 00:05:26.790 --> 00:05:29.010 then the exposure is associated

116 00:05:29.010 --> 00:05:32.370 with an increased risk of the outcome.

117 00:05:32.370 --> 00:05:34.080 A strength of the cohort study

118 00:05:34.080 --> 00:05:37.410 is that we know the exposure preceded the outcome.

119 00:05:37.410 --> 00:05:40.260 However, it's necessary to wait for the development

120 00:05:40.260 --> 00:05:43.080 of the outcome in prospective cohort studies.

121 00:05:43.080 --> 00:05:45.840 So they can be slow to conduct,

122 00:05:45.840 --> 00:05:48.783 especially in the case of rare outcomes.

123 00:05:50.310 --> 00:05:54.000 We often perform retrospective cohort studies.

124 00:05:54.000 --> 00:05:56.490 Here, the start of the study can occur

125 00:05:56.490 --> 00:05:58.710 after some patients have already developed

126 00:05:58.710 --> 00:06:00.300 the outcome of interest.

127 00:06:00.300 --> 00:06:02.580 Exposure status in the past

128 00:06:02.580 --> 00:06:07.560 is established using existing data, such as medical records.

129 00:06:07.560 --> 00:06:10.830 The goal is to establish a cohort of individuals

130 00:06:10.830 --> 00:06:14.370 without the outcome at a fixed point in the past

131 00:06:14.370 --> 00:06:18.420 and determine their exposure status at that point in time.

132 00:06:18.420 --> 00:06:20.340 Then determine if the patient

133 00:06:20.340 --> 00:06:23.220 subsequently develops the outcome.

134 00:06:23.220 --> 00:06:26.460 A limitation of retrospective cohort studies

135 00:06:26.460 --> 00:06:29.430 is that exposure status is not assessed

136 00:06:29.430 --> 00:06:31.140 by the study investigators

137 00:06:31.140 --> 00:06:33.570 because we often rely on medical records.

138 00:06:33.570 --> 00:06:37.083 So the desired exposure data may not be available.

139 00:06:38.070 --> 00:06:41.580 A case control study is an alternative design

140 00:06:41.580 --> 00:06:44.850 that begins by identifying a group of individuals

141 00:06:44.850 --> 00:06:47.130 with the outcome of interest.

142 00:06:47.130 --> 00:06:50.160 A similar control group without the outcome

143 00:06:50.160 --> 00:06:52.290 is also identified.

144 00:06:52.290 --> 00:06:54.540 Through chart reviews or interviews,

145 00:06:54.540 --> 00:06:58.620 the investigator then determines past exposure status.

146 00:06:58.620 --> 00:07:01.260 Unlike a retrospective cohort study

147 00:07:01.260 --> 00:07:04.530 that begins by identifying study participants

148 00:07:04.530 --> 00:07:07.980 based on the exposure and then assesses outcomes,

149 00:07:07.980 --> 00:07:11.730 the case control study begins by assessing the outcome

150 00:07:11.730 --> 00:07:14.550 and then determines past exposures.

151 00:07:14.550 --> 00:07:18.510 Case control studies are useful when studying rare outcomes

152 00:07:18.510 --> 00:07:21.330 or diseases with long latency.

153 00:07:21.330 --> 00:07:24.480 However, limitations include recall bias

154 00:07:24.480 --> 00:07:26.310 in recalling exposure,

155 00:07:26.310 --> 00:07:29.430 problems with selecting comparable controls,

156 00:07:29.430 --> 00:07:31.713 as well as some analytical issues.

157 00:07:33.150 --> 00:07:35.670 The other major arm of study designs

158 00:07:35.670 --> 00:07:37.770 are the experimental studies.

159 00:07:37.770 --> 00:07:41.250 The gold standard of evidence is the clinical trial,

160 00:07:41.250 --> 00:07:44.340 in particular, the randomized clinical trial.

161 00:07:44.340 --> 00:07:47.520 Here, the investigator randomly assigns participants

162 00:07:47.520 --> 00:07:50.100 to different exposures, for example,

163 00:07:50.100 --> 00:07:54.540 sorafenib or placebo in patients with advanced liver cancer.

164 00:07:54.540 --> 00:07:57.510 Usually each participant has an equal chance

165 00:07:57.510 --> 00:07:59.700 of being assigned to the two groups,

166 00:07:59.700 --> 00:08:01.050 although the allocation ratio

167 00:08:01.050 --> 00:08:02.913 does not have to be one-to-one.

168 00:08:03.870 --> 00:08:07.380 Here we have an example of a randomized control trial.

169 00:08:07.380 --> 00:08:09.480 This study randomly assigned patients

170 00:08:09.480 --> 00:08:12.873 with liver cancer to sorafenib or placebo.

171 00:08:13.830 --> 00:08:17.430 A traditional parallel group randomized clinical trial

172 00:08:17.430 --> 00:08:19.740 resembles a prospective cohort study,

173 00:08:19.740 --> 00:08:23.700 except for the important difference of randomization.

174 00:08:23.700 --> 00:08:26.460 Randomization is an important element

175 00:08:26.460 --> 00:08:27.510 of clinical trials

176 00:08:27.510 --> 00:08:30.240 because it protects against selection bias

177 00:08:30.240 --> 00:08:32.460 and should balance both known

178 00:08:32.460 --> 00:08:36.573 and unknown confounding factors between the exposure groups.

179 00:08:37.650 --> 00:08:40.710 Additional commonly used randomized trial designs

180 00:08:40.710 --> 00:08:44.610 include crossover designs, factorial designs,

181 00:08:44.610 --> 00:08:47.343 and cluster randomized clinical trials.

182 00:08:48.960 --> 00:08:51.420 In non-randomized control trials,

183 00:08:51.420 --> 00:08:54.420 participants are assigned to different interventions

184 00:08:54.420 --> 00:08:57.060 without following a random procedure.

185 00:08:57.060 --> 00:08:59.850 For example, assignment may be made according

186 00:08:59.850 --> 00:09:01.740 to investigator preference.

187 00:09:01.740 --> 00:09:04.140 This design is susceptible to bias

188 00:09:04.140 --> 00:09:05.910 due to the potential differences

189 00:09:05.910 --> 00:09:08.793 in patient characteristics between the two groups.

190 00:09:09.780 --> 00:09:13.080 Another type of study is the systematic review.

191 00:09:13.080 --> 00:09:15.120 The goal here is to review

192 00:09:15.120 --> 00:09:17.820 and synthesize all available evidence

193 00:09:17.820 --> 00:09:20.430 on a specific research question.

194 00:09:20.430 --> 00:09:23.250 Systematic reviews are carried out according to

195 00:09:23.250 --> 00:09:27.060 a pre-specified protocol that defines the question,

196 00:09:27.060 --> 00:09:29.220 describes the scope of the review,

197 00:09:29.220 --> 00:09:32.520 and the criteria and methodology that will be used.

198 00:09:32.520 --> 00:09:35.670 Begin by specifying the review question,

199 00:09:35.670 --> 00:09:37.770 the population you're studying,

200 00:09:37.770 --> 00:09:41.040 the interventions or exposures of interest,

201 00:09:41.040 --> 00:09:42.420 outcomes of interest,

202 00:09:42.420 --> 00:09:46.170 and study designs that should be included in the review.

203 00:09:46.170 --> 00:09:48.840 Then you'll list the eligibility criteria

204 00:09:48.840 --> 00:09:50.850 for the studies to be included.

205 00:09:50.850 --> 00:09:54.150 For example, you may only want to include evidence

206 00:09:54.150 --> 00:09:56.160 from randomized control trials

207 00:09:56.160 --> 00:09:58.830 or large scale observational studies.

208 00:09:58.830 --> 00:10:02.610 You'll then search PubMed and other databases

209 00:10:02.610 --> 00:10:07.080 for articles or studies that meet these criteria.

210 00:10:07.080 --> 00:10:10.800 An important step is assessing the quality of the studies

211 00:10:10.800 --> 00:10:13.740 and critically assessing how well the studies were done.

212 00:10:13.740 --> 00:10:17.700 Any potential issues with the design may introduce bias,

213 00:10:17.700 --> 00:10:21.210 so it's important to critique the strengths and weaknesses

214 00:10:21.210 --> 00:10:23.370 of the evidence you've accumulated.

215 00:10:23.370 --> 00:10:24.203 At this stage,

216 00:10:24.203 --> 00:10:26.850 you can also comment on gaps in the evidence,

217 00:10:26.850 --> 00:10:30.720 such as patient populations not represented in the research.

218 00:10:30.720 --> 00:10:33.660 Then you extract the data or the study results

219 00:10:33.660 --> 00:10:35.520 that are summarized in the manuscripts

220 00:10:35.520 --> 00:10:37.230 and the supplemental materials.

221 00:10:37.230 --> 00:10:40.470 And a final step involves synthesizing the results,

222 00:10:40.470 --> 00:10:43.230 or combining and analyzing the results

223 00:10:43.230 --> 00:10:44.910 from multiple studies.

224 00:10:44.910 --> 00:10:47.580 This can be a qualitative synthesis

225 00:10:47.580 --> 00:10:48.960 in which you summarize

226 00:10:48.960 --> 00:10:52.350 how the research you found fit together,

227 00:10:52.350 --> 00:10:54.420 describe the strengths and the weaknesses

228 00:10:54.420 --> 00:10:55.740 of the body of evidence,

229 00:10:55.740 --> 00:10:59.100 identify gaps in areas of future research,

230 00:10:59.100 --> 00:11:02.343 or it can be a quantitative synthesis.

231 00:11:03.270 --> 00:11:06.240 Quantitative synthesis is usually in the form

232 00:11:06.240 --> 00:11:08.610 of a meta-analysis.

233 00:11:08.610 --> 00:11:13.610 A meta-analysis is an analytical way of formally combining

234 00:11:13.770 --> 00:11:17.160 or pooling results from different sources.

235 00:11:17.160 --> 00:11:19.710 In this study, the goal was to estimate the effect

236 00:11:19.710 --> 00:11:22.620 of antiviral therapy on liver stiffness

237 00:11:22.620 --> 00:11:25.380 in patients with hepatitis B at six months,

238 00:11:25.380 --> 00:11:27.690 one year, two years, three years,

239 00:11:27.690 --> 00:11:30.690 and five years after beginning treatment.

240 00:11:30.690 --> 00:11:34.380 This flow chart shows the process of study collection.

241 00:11:34.380 --> 00:11:38.043 In the end, 24 studies were included in the analysis.

242 00:11:38.910 --> 00:11:40.830 Typically, a forest plot,

243 00:11:40.830 --> 00:11:43.020 which is the graph shown here on the right,

244 00:11:43.020 --> 00:11:44.670 displays the point estimates

245 00:11:44.670 --> 00:11:48.150 and 95% confidence intervals from each study,

246 00:11:48.150 --> 00:11:50.550 along with the pooled estimate.

247 00:11:50.550 --> 00:11:52.740 At six months from the start of therapy,

248 00:11:52.740 --> 00:11:54.990 data pooled from eight studies

249 00:11:54.990 --> 00:11:59.040 with 968 patients showed a significant decline

250 00:11:59.040 --> 00:12:02.850 in liver stiffness by 2.21 kilopascals

251 00:12:02.850 --> 00:12:05.493 as compared to pre-treatment LSM.

252 00:12:07.110 --> 00:12:10.500 In this video, we discussed some commonly used designs

253 00:12:10.500 --> 00:12:13.800 in clinical research, including observational studies,

254 00:12:13.800 --> 00:12:17.400 experimental studies, and systematic reviews.

255 00:12:17.400 --> 00:12:20.790 The next video, which is the third video in this series,

256 00:12:20.790 --> 00:12:25.260 will discuss the data collection process and variable types.

257 00:12:25.260 --> 00:12:28.110 Understanding variable types will prepare us

258 00:12:28.110 --> 00:12:30.420 for the fourth video in this series

259 00:12:30.420 --> 00:12:32.493 on sample size determination.