WEBVTT

1 00:00:02.361 --> 00:00:03.840 <v ->[Maria Ciarleglio] My name is Maria Ciarleglio</v>

2 00:00:03.840 --> 00:00:07.740 and I'm a faculty member in the Department of Biostatistics

3 00:00:07.740 --> 00:00:10.023 at the Yale School of Public Health.

4 00:00:10.890 --> 00:00:12.210 In this video series,

5 00:00:12.210 --> 00:00:14.880 I will introduce the clinical research process

6 00:00:14.880 --> 00:00:18.873 to prepare you to collaborate with a statistician.

7 00:00:20.370 --> 00:00:21.900 In this fourth video,

8 00:00:21.900 --> 00:00:25.170 we'll discuss elements of sample size determination,

9 00:00:25.170 --> 00:00:28.623 which is an important part of study design.

10 00:00:30.000 --> 00:00:34.740 In statistics, we apply methods that allow us to use data

11 00:00:34.740 --> 00:00:38.550 from a sample to answer a variety of questions.

12 00:00:38.550 --> 00:00:42.900 Sample data are used to estimate population parameters

13 00:00:42.900 --> 00:00:46.830 such as population means or population proportions.

14 00:00:46.830 --> 00:00:51.060 Develop models relating one or more explanatory variable

15 00:00:51.060 --> 00:00:55.140 to a response variable, and test hypotheses.

16 00:00:55.140 --> 00:00:57.480 In all we do, we answer these questions

17 00:00:57.480 --> 00:00:59.670 using a representative sample

18 00:00:59.670 --> 00:01:01.740 from the population of interest.

19 00:01:01.740 --> 00:01:03.840 This leads to the natural question

20 00:01:03.840 --> 00:01:06.870 how large should the sample be?

21 00:01:06.870 --> 00:01:09.300 Sample size methods that we'll discuss today

22 00:01:09.300 --> 00:01:12.960 are presented with the idea of a parallel-design

23 00:01:12.960 --> 00:01:16.200 two-arm randomized clinical trial in mind,

24 00:01:16.200 --> 00:01:18.720 but they can also be applied to other designs,

25 00:01:18.720 --> 00:01:21.090 such as observational studies.

26 00:01:21.090 --> 00:01:24.330 Our goal is to determine the sample size needed

27 00:01:24.330 --> 00:01:27.840 to be able to detect a hypothesized difference

28 00:01:27.840 --> 00:01:31.260 of clinical interest between the two study arms.

29 00:01:31.260 --> 00:01:33.330 If a difference truly exists

30 00:01:33.330 --> 00:01:36.150 we want to be able to detect that difference

31 00:01:36.150 --> 00:01:37.620 with high probability,

32 00:01:37.620 --> 00:01:39.840 and that probability is a term

33 00:01:39.840 --> 00:01:43.890 we'll introduce shortly known as statistical power.

34 00:01:43.890 --> 00:01:46.110 The sample size calculation depends

35 00:01:46.110 --> 00:01:48.300 on the planned hypothesis test,

36 00:01:48.300 --> 00:01:50.430 but the planned test depends

37 00:01:50.430 --> 00:01:52.830 on the study's primary endpoint.

38 00:01:52.830 --> 00:01:56.550 In video three, we reviewed quantitative endpoints,

39 00:01:56.550 --> 00:02:00.450 dichotomous endpoints, and time to event endpoints.

40 00:02:00.450 --> 00:02:03.720 When the primary endpoint is a continuous measure,

41 00:02:03.720 --> 00:02:07.380 such as change in portal pressure within a subject,

42 00:02:07.380 --> 00:02:10.620 we summarize the response using the average

43 00:02:10.620 --> 00:02:14.970 or mean change in portal pressure in each treatment group.

44 00:02:14.970 --> 00:02:18.210 If there's no difference between the treatments on average

45 00:02:18.210 --> 00:02:22.500 then the difference in means in group two versus group one,

46 00:02:22.500 --> 00:02:25.800 mu two, minus mu one, equals zero.

47 00:02:25.800 --> 00:02:29.070 This is called our null hypothesis.

48 00:02:29.070 --> 00:02:32.370 Our goal is usually to demonstrate a difference,

49 00:02:32.370 --> 00:02:35.790 so we would like to reject the null hypothesis

50 00:02:35.790 --> 00:02:38.880 and conclude that a difference exists.

51 00:02:38.880 --> 00:02:41.220 The hypothesis of a difference

52 00:02:41.220 --> 00:02:44.850 or effect is called the alternative hypothesis,

53 00:02:44.850 --> 00:02:47.370 and we'll discuss the alternative hypothesis

54 00:02:47.370 --> 00:02:48.513 on the next slide.

55 00:02:49.740 --> 00:02:52.770 When the primary endpoint is a dichotomous measure,

56 00:02:52.770 --> 00:02:55.500 such as treatment response, yes or no,

57 00:02:55.500 --> 00:02:58.500 we summarize the response using the proportion

58 00:02:58.500 --> 00:03:02.250 of patients who respond in each treatment group.

59 00:03:02.250 --> 00:03:04.620 Again, if there's no difference between the treatments

60 00:03:04.620 --> 00:03:06.240 the difference in proportions,

61 00:03:06.240 --> 00:03:11.240 P two, minus P one, equals zero under the null hypothesis.

62 00:03:11.970 --> 00:03:14.640 When the primary endpoint is time to event,

63 00:03:14.640 --> 00:03:17.400 such as time to death or time to relapse,

64 00:03:17.400 --> 00:03:19.740 then we often represent the effect

65 00:03:19.740 --> 00:03:22.830 in terms of the hazard rate, lambda.

66 00:03:22.830 --> 00:03:25.200 If there's no difference between the treatments

67 00:03:25.200 --> 00:03:27.330 then under the null hypothesis,

68 00:03:27.330 --> 00:03:30.630 the difference in the hazard rates equals zero,

69 00:03:30.630 --> 00:03:34.503 or equivalently, the hazard ratio equals one.

70 00:03:35.610 --> 00:03:38.280 Ideally, if there's a treatment effect,

71 00:03:38.280 --> 00:03:40.260 we reject the null hypothesis

72 00:03:40.260 --> 00:03:42.780 and conclude the alternative hypothesis,

73 00:03:42.780 --> 00:03:46.380 that states there is a difference in the two populations.

74 00:03:46.380 --> 00:03:49.500 For example, suppose our goal is to determine

75 00:03:49.500 --> 00:03:51.150 if there's a difference in the proportion

76 00:03:51.150 --> 00:03:56.010 of responders in those on Sorafenib compared to placebo.

77 00:03:56.010 --> 00:03:59.730 The alternative hypothesis tested is a two-sided alternative

78 00:03:59.730 --> 00:04:02.280 that the difference in the proportion of responders

79 00:04:02.280 --> 00:04:04.470 is not equal to zero.

80 00:04:04.470 --> 00:04:06.660 When performing two-sided tests,

81 00:04:06.660 --> 00:04:08.730 if our test statistic falls

82 00:04:08.730 --> 00:04:12.510 into either of the blue rejection regions, shown here,

83 00:04:12.510 --> 00:04:15.810 we would reject the null hypothesis of no difference,

84 00:04:15.810 --> 00:04:18.390 and conclude that there's a significant difference

85 00:04:18.390 --> 00:04:20.580 between treatments.

86 00:04:20.580 --> 00:04:23.820 Suppose we, instead, were only interested

87 00:04:23.820 --> 00:04:25.920 in a significant conclusion,

88 00:04:25.920 --> 00:04:28.830 if we showed that the proportion of responders

89 00:04:28.830 --> 00:04:31.320 is higher in the treatment group

90 00:04:31.320 --> 00:04:33.570 compared to the placebo group,

91 00:04:33.570 --> 00:04:36.180 this would give a difference in proportions

92 00:04:36.180 --> 00:04:37.470 greater than zero.

93 00:04:37.470 --> 00:04:39.690 In this case, we would only be interested

94 00:04:39.690 --> 00:04:43.050 in effects in the upper tail, shown in red.

95 00:04:43.050 --> 00:04:47.220 This one directional test is called a one-sided test

96 00:04:47.220 --> 00:04:51.270 or more specifically an upper tail test.

97 00:04:51.270 --> 00:04:54.810 Similarly, we may be interested in an effect

98 00:04:54.810 --> 00:04:56.220 in the negative direction,

99 00:04:56.220 --> 00:04:58.080 in which case we would only look

100 00:04:58.080 --> 00:05:01.800 for a significant conclusion in the lower tail.

101 00:05:01.800 --> 00:05:05.403 This one-sided test is a lower tail test.

102 00:05:06.510 --> 00:05:08.130 The direction of your test

103 00:05:08.130 --> 00:05:11.100 affects your sample size calculation.

104 00:05:11.100 --> 00:05:14.190 We will talk about this alpha symbol shortly,

105 00:05:14.190 --> 00:05:16.380 but it's called the significance level,

106 00:05:16.380 --> 00:05:18.990 or the type one error of your test.

107 00:05:18.990 --> 00:05:20.310 It's the probability

108 00:05:20.310 --> 00:05:24.540 of incorrectly rejecting the null hypothesis.

109 00:05:24.540 --> 00:05:26.340 In one-sided tests,

110 00:05:26.340 --> 00:05:28.620 since we're only looking in one direction

111 00:05:28.620 --> 00:05:31.110 for evidence against the null hypothesis,

112 00:05:31.110 --> 00:05:34.680 all of our type one error is in a single tail.

113 00:05:34.680 --> 00:05:37.920 However, with two-sided tests, because it's possible

114 00:05:37.920 --> 00:05:40.530 for us to reject the null hypothesis,

115 00:05:40.530 --> 00:05:43.800 if there is extreme evidence in either tail

116 00:05:43.800 --> 00:05:47.493 we split our type one error between the two tails.

117 00:05:48.390 --> 00:05:50.040 When looking at each tail

118 00:05:50.040 --> 00:05:52.770 we actually require stronger evidence

119 00:05:52.770 --> 00:05:57.510 against the null to reject in the case of a two-sided test.

120 00:05:57.510 --> 00:06:01.080 Since it's more difficult to reject the null hypothesis

121 00:06:01.080 --> 00:06:03.450 we will need a larger sample size

122 00:06:03.450 --> 00:06:06.390 when performing a two-sided test compared

123 00:06:06.390 --> 00:06:09.120 to a one-sided test.

124 00:06:09.120 --> 00:06:12.900 We recommend performing two-sided tests.

125 00:06:12.900 --> 00:06:15.720 Although you might expect a new treatment

126 00:06:15.720 --> 00:06:19.560 to demonstrate superiority over the control treatment,

127 00:06:19.560 --> 00:06:22.530 it's always good to have the option to formally

128 00:06:22.530 --> 00:06:24.330 reject the null hypothesis

129 00:06:24.330 --> 00:06:27.783 if an effect is seen in the opposite direction.

130 00:06:29.250 --> 00:06:30.900 There are several key factors

131 00:06:30.900 --> 00:06:33.240 that affect the required sample size.

132 00:06:33.240 --> 00:06:35.970 The hypothesized treatment difference, delta,

133 00:06:35.970 --> 00:06:40.260 the variability or noise in the endpoint measurement, sigma,

134 00:06:40.260 --> 00:06:43.290 the level of statistical significance, alpha,

135 00:06:43.290 --> 00:06:47.610 and the level of statistical power, one minus beta.

136 00:06:47.610 --> 00:06:49.500 We'll discuss each of these components,

137 00:06:49.500 --> 00:06:52.650 starting with the expected clinical difference

138 00:06:52.650 --> 00:06:55.113 between the two treatments being tested.

139 00:06:56.250 --> 00:07:00.480 In order to estimate sample size, you must first specify

140 00:07:00.480 --> 00:07:03.630 the magnitude of the difference you wish to detect.

141 00:07:03.630 --> 00:07:06.660 We denote this difference as delta.

142 00:07:06.660 --> 00:07:08.310 Sample size is calculated

143 00:07:08.310 --> 00:07:11.460 under a specific alternative hypothesis,

144 00:07:11.460 --> 00:07:13.620 that the difference in your parameters,

145 00:07:13.620 --> 00:07:17.520 the difference in means here, is equal to delta.

146 00:07:17.520 --> 00:07:20.550 The blue curve shows us the distribution

147 00:07:20.550 --> 00:07:24.300 of the difference in means under the null hypothesis.

148 00:07:24.300 --> 00:07:27.990 Under the null, the distribution is centered at zero,

149 00:07:27.990 --> 00:07:32.550 assuming no difference on average between the treatments.

150 00:07:32.550 --> 00:07:33.960 Under the alternative,

151 00:07:33.960 --> 00:07:36.630 for the purpose of sample size calculations,

152 00:07:36.630 --> 00:07:39.210 the distribution of the difference in means

153 00:07:39.210 --> 00:07:43.320 is centered at delta and is represented by the red curve.

154 00:07:43.320 --> 00:07:47.220 The more different the two distributions are assumed to be,

155 00:07:47.220 --> 00:07:48.690 the larger delta,

156 00:07:48.690 --> 00:07:53.130 and the less overlap we see between the two distributions.

157 00:07:53.130 --> 00:07:56.700 Smaller differences are more difficult to detect

158 00:07:56.700 --> 00:07:59.430 because the distributions are closer together,

159 00:07:59.430 --> 00:08:02.760 and, as a result, we require a larger sample size

160 00:08:02.760 --> 00:08:05.370 to be able to detect the small difference

161 00:08:05.370 --> 00:08:08.910 and distinguish that difference from random variation.

162 00:08:08.910 --> 00:08:11.040 Larger hypothesized differences

163 00:08:11.040 --> 00:08:13.890 require smaller sample sizes.

164 00:08:13.890 --> 00:08:16.380 How do you choose a value for delta?

165 00:08:16.380 --> 00:08:18.120 Sometimes there's prior knowledge

166 00:08:18.120 --> 00:08:20.670 that allows an investigator to anticipate

167 00:08:20.670 --> 00:08:23.580 the treatment benefit that's likely to be observed,

168 00:08:23.580 --> 00:08:26.700 and the role of the study is to confirm that expectation.

169 00:08:26.700 --> 00:08:29.940 Other times, delta's taken to equal the smallest

170 00:08:29.940 --> 00:08:32.790 or minimum clinically relevant difference

171 00:08:32.790 --> 00:08:36.810 that would warrant adopting the new treatment.

172 00:08:36.810 --> 00:08:39.600 Investigators are often optimistic

173 00:08:39.600 --> 00:08:41.760 about the effect of a new treatment,

174 00:08:41.760 --> 00:08:43.080 and that's understandable,

175 00:08:43.080 --> 00:08:46.950 but I recommend you not be overly optimistic.

176 00:08:46.950 --> 00:08:50.760 If the treatment effect is not as large as expected,

177 00:08:50.760 --> 00:08:54.090 you could end up with a null or negative trial,

178 00:08:54.090 --> 00:08:56.040 which is a trial that does not show

179 00:08:56.040 --> 00:08:58.290 a significant difference.

180 00:08:58.290 --> 00:09:01.860 There may actually be a true and worthwhile

181 00:09:01.860 --> 00:09:04.050 treatment benefit that's been missed

182 00:09:04.050 --> 00:09:06.360 because the difference was mis-specified

183 00:09:06.360 --> 00:09:09.240 or hypothesized to be too large.

184 00:09:09.240 --> 00:09:11.310 This is why a lot of thought

185 00:09:11.310 --> 00:09:13.620 needs to go into the study design

186 00:09:13.620 --> 00:09:16.023 and what is considered meaningful.

187 00:09:17.640 --> 00:09:20.730 The next element involved in sample size determination

188 00:09:20.730 --> 00:09:24.420 is the standard deviation of the primary endpoint.

189 00:09:24.420 --> 00:09:27.240 The standard deviation is denoted by sigma,

190 00:09:27.240 --> 00:09:29.160 and needs to be specified

191 00:09:29.160 --> 00:09:33.240 when we're dealing with a continuous primary endpoint.

192 00:09:33.240 --> 00:09:34.230 In this figure,

193 00:09:34.230 --> 00:09:38.280 there are actually four normal distributions plotted.

194 00:09:38.280 --> 00:09:40.890 Let's begin with the solid blue curve

195 00:09:40.890 --> 00:09:42.750 and the solid red curve.

196 00:09:42.750 --> 00:09:46.050 These two curves have the same standard deviation.

197 00:09:46.050 --> 00:09:48.510 Their standard deviation is larger

198 00:09:48.510 --> 00:09:49.980 than the standard deviation

199 00:09:49.980 --> 00:09:53.670 of the dashed blue curve and the dashed red curve.

200 00:09:53.670 --> 00:09:55.410 As sigma decreases,

201 00:09:55.410 --> 00:09:59.430 there's less overlap between the two distributions.

202 00:09:59.430 --> 00:10:03.000 More noise or higher variability makes it more

203 00:10:03.000 --> 00:10:05.430 difficult to detect differences

204 00:10:05.430 --> 00:10:07.983 and requires a larger sample size.

205 00:10:08.880 --> 00:10:11.550 One thing to note, is that the treatment difference,

206 00:10:11.550 --> 00:10:14.040 delta, is sometimes standardized

207 00:10:14.040 --> 00:10:16.530 and presented as an effect size

208 00:10:16.530 --> 00:10:19.140 denoted here by capital delta.

209 00:10:19.140 --> 00:10:22.653 This is simply little delta divided by sigma.

210 00:10:23.730 --> 00:10:26.670 There are two errors that we can make

211 00:10:26.670 --> 00:10:28.650 when we perform a hypothesis test,

212 00:10:28.650 --> 00:10:31.470 and both of them influence sample size.

213 00:10:31.470 --> 00:10:32.970 We fix these errors

214 00:10:32.970 --> 00:10:35.190 at levels that we believe to be acceptable,

215 00:10:35.190 --> 00:10:39.030 and they're usually set to relatively small values.

216 00:10:39.030 --> 00:10:42.240 The first error we'll discuss is type one error

217 00:10:42.240 --> 00:10:44.790 or the alpha level of the test.

218 00:10:44.790 --> 00:10:46.350 The blue curve is, again,

219 00:10:46.350 --> 00:10:48.780 the distribution of the difference in means

220 00:10:48.780 --> 00:10:51.480 under the null hypothesis.

221 00:10:51.480 --> 00:10:53.640 The red curve is the distribution

222 00:10:53.640 --> 00:10:56.640 under the specific alternative hypothesis

223 00:10:56.640 --> 00:11:00.990 that assumes the treatment effect is equal to delta.

224 00:11:00.990 --> 00:11:03.270 Hypothesis testing is performed

225 00:11:03.270 --> 00:11:05.880 assuming the null hypothesis is true.

226 00:11:05.880 --> 00:11:09.420 That is assuming the blue curve is true.

227 00:11:09.420 --> 00:11:12.900 The green shaded area in the tails of the blue curve

228 00:11:12.900 --> 00:11:16.470 are extreme values that aren't likely to be observed

229 00:11:16.470 --> 00:11:19.500 if the difference in means is equal to zero,

230 00:11:19.500 --> 00:11:22.410 that is if the null hypothesis is true.

231 00:11:22.410 --> 00:11:26.010 If we observe a result in the green shaded area

232 00:11:26.010 --> 00:11:28.800 then we'll reject the null hypothesis

233 00:11:28.800 --> 00:11:31.650 and conclude the alternative hypothesis.

234 00:11:31.650 --> 00:11:34.350 This is equivalent to observing a P value

235 00:11:34.350 --> 00:11:37.800 of the test less than or equal to alpha.

236 00:11:37.800 --> 00:11:40.950 However, if the null hypothesis is true,

237 00:11:40.950 --> 00:11:42.540 then we're committing an error

238 00:11:42.540 --> 00:11:45.270 by concluding there is an effect,

239 00:11:45.270 --> 00:11:48.510 there is a difference, when in fact there isn't one.

240 00:11:48.510 --> 00:11:51.480 This is called a type one error.

241 00:11:51.480 --> 00:11:53.970 The smaller you make the green shaded area,

242 00:11:53.970 --> 00:11:57.480 the less likely you will incorrectly reject

243 00:11:57.480 --> 00:11:58.740 the null hypothesis,

244 00:11:58.740 --> 00:12:00.120 because you're going to require

245 00:12:00.120 --> 00:12:02.910 greater and greater evidence to do so.

246 00:12:02.910 --> 00:12:05.760 We typically set alpha equal to 0.05

247 00:12:05.760 --> 00:12:09.120 because it's felt that a 5% chance

248 00:12:09.120 --> 00:12:13.560 of falsely rejecting the null hypothesis is acceptable.

249 00:12:13.560 --> 00:12:17.550 Choosing a smaller alpha will increase your protection

250 00:12:17.550 --> 00:12:20.040 against committing a type one error,

251 00:12:20.040 --> 00:12:21.450 but there's a trade off

252 00:12:21.450 --> 00:12:24.150 in that it will be more difficult for you to conclude

253 00:12:24.150 --> 00:12:27.330 there's a difference, even when there is one.

254 00:12:27.330 --> 00:12:31.563 Decreasing alpha will increase the required sample size.

255 00:12:32.730 --> 00:12:36.210 The second error is called type two error,

256 00:12:36.210 --> 00:12:38.550 and it's denoted beta,

257 00:12:38.550 --> 00:12:42.600 the gray shaded region in this figure.

258 00:12:42.600 --> 00:12:44.880 We do not reject the null hypothesis

259 00:12:44.880 --> 00:12:49.080 if the difference we observe falls in the gray region.

260 00:12:49.080 --> 00:12:52.050 We only reject the null if the difference observed

261 00:12:52.050 --> 00:12:54.780 falls in either of the green regions,

262 00:12:54.780 --> 00:12:56.880 the rejection region of the test.

263 00:12:56.880 --> 00:13:00.030 However, because the two distributions overlap

264 00:13:00.030 --> 00:13:02.220 there is this gray shaded region

265 00:13:02.220 --> 00:13:04.740 where the alternative hypothesis is true,

266 00:13:04.740 --> 00:13:07.680 the red curve is true, but we fail to reject the null

267 00:13:07.680 --> 00:13:11.850 because we don't observe an effect that's extreme enough.

268 00:13:11.850 --> 00:13:15.840 When this occurs, we are committing a type two error.

269 00:13:15.840 --> 00:13:18.930 Of course, we want the type two error to be low,

270 00:13:18.930 --> 00:13:23.250 but rather than set beta, we usually set one minus beta,

271 00:13:23.250 --> 00:13:26.760 which is called the statistical power of the test.

272 00:13:26.760 --> 00:13:30.303 This is represented by the purple shaded area.

273 00:13:31.200 --> 00:13:35.100 Power is the probability of rejecting the null hypothesis

274 00:13:35.100 --> 00:13:36.150 when we should.

275 00:13:36.150 --> 00:13:39.780 That is rejecting a false null hypothesis

276 00:13:39.780 --> 00:13:43.020 and we want this probability to be high.

277 00:13:43.020 --> 00:13:46.950 We typically set power to be at least 80%.

278 00:13:46.950 --> 00:13:50.520 Larger power will require a larger sample size

279 00:13:50.520 --> 00:13:54.453 to increase our chance of detecting a true difference.

280 00:13:56.190 --> 00:14:00.120 If you work with all of these ideas in their equation form

281 00:14:00.120 --> 00:14:03.510 you can derive a fundamental sample size equation

282 00:14:03.510 --> 00:14:06.210 that relates all four of these parameters

283 00:14:06.210 --> 00:14:09.750 to the sample size required in each treatment group.

284 00:14:09.750 --> 00:14:12.570 This equation shown here assumes a continuous

285 00:14:12.570 --> 00:14:14.190 primary outcome variable,

286 00:14:14.190 --> 00:14:17.220 but the relationships are the same for any outcome,

287 00:14:17.220 --> 00:14:20.460 including binary and time to event.

288 00:14:20.460 --> 00:14:22.890 We see sigma and delta.

289 00:14:22.890 --> 00:14:24.900 Delta is the treatment effect

290 00:14:24.900 --> 00:14:27.420 or the difference in group means.

291 00:14:27.420 --> 00:14:28.830 As I mentioned before,

292 00:14:28.830 --> 00:14:31.830 you can divide the difference in means, delta,

293 00:14:31.830 --> 00:14:34.590 by the common standard deviation, sigma,

294 00:14:34.590 --> 00:14:36.360 to write the equation as a function

295 00:14:36.360 --> 00:14:38.820 of the standardized effect size.

296 00:14:38.820 --> 00:14:40.500 As sigma increases,

297 00:14:40.500 --> 00:14:43.230 it's clear that the sample size will increase.

298 00:14:43.230 --> 00:14:45.780 This is because the data are more noisy,

299 00:14:45.780 --> 00:14:48.630 more heterogeneous, and it's more difficult

300 00:14:48.630 --> 00:14:51.660 to detect a signal when this is the case

301 00:14:51.660 --> 00:14:54.180 and we need a larger sample size.

302 00:14:54.180 --> 00:14:57.870 Delta is in the denominator, so as delta decreases,

303 00:14:57.870 --> 00:15:00.030 the sample size will increase.

304 00:15:00.030 --> 00:15:03.510 When delta is small, there will be more overlap

305 00:15:03.510 --> 00:15:05.340 between the two distributions

306 00:15:05.340 --> 00:15:08.520 and it will be more difficult to detect a difference,

307 00:15:08.520 --> 00:15:10.650 so we need a larger sample size

308 00:15:10.650 --> 00:15:13.500 to detect smaller differences.

309 00:15:13.500 --> 00:15:15.480 All of these relationships make sense

310 00:15:15.480 --> 00:15:16.980 if you talk them through

311 00:15:16.980 --> 00:15:19.530 and they're supported by the equations

312 00:15:19.530 --> 00:15:23.160 used to perform the sample size calculations.

313 00:15:23.160 --> 00:15:25.440 In terms of alpha and beta,

314 00:15:25.440 --> 00:15:27.960 our type one and type two errors,

315 00:15:27.960 --> 00:15:29.850 they are here in the numerator

316 00:15:29.850 --> 00:15:33.990 but they're represented by their corresponding Z values.

317 00:15:33.990 --> 00:15:36.300 Smaller alpha and beta errors

318 00:15:36.300 --> 00:15:40.353 correspond to larger Z values and larger sample sizes.

319 00:15:42.450 --> 00:15:44.400 We'll wrap up this video by going

320 00:15:44.400 --> 00:15:47.130 through the three common endpoint types

321 00:15:47.130 --> 00:15:50.130 and discussing the elements of sample size determination

322 00:15:50.130 --> 00:15:53.520 that you need to define for the sample size calculation

323 00:15:53.520 --> 00:15:55.320 in each case.

324 00:15:55.320 --> 00:15:59.040 You'll need to specify the type one error level, alpha,

325 00:15:59.040 --> 00:16:01.740 and the direction of the alternative hypothesis.

326 00:16:01.740 --> 00:16:04.170 That is, are you performing a one-sided

327 00:16:04.170 --> 00:16:06.540 or a two-sided hypothesis test?

328 00:16:06.540 --> 00:16:09.060 You'll also need to specify the level

329 00:16:09.060 --> 00:16:12.240 of statistical power of the test.

330 00:16:12.240 --> 00:16:15.840 When your primary endpoint is a continuous variable,

331 00:16:15.840 --> 00:16:18.120 such as change in portal pressure,

332 00:16:18.120 --> 00:16:20.430 you'll need to specify delta,

333 00:16:20.430 --> 00:16:23.040 the magnitude of the hypothesized difference

334 00:16:23.040 --> 00:16:26.850 in mean portal pressure change in the two treatment groups,

335 00:16:26.850 --> 00:16:28.950 and sigma, the standard deviation

336 00:16:28.950 --> 00:16:31.410 of the change in portal pressure.

337 00:16:31.410 --> 00:16:33.660 We often assume that the variability

338 00:16:33.660 --> 00:16:36.630 of the response is the same in both arms,

339 00:16:36.630 --> 00:16:38.370 but the sample size calculations,

340 00:16:38.370 --> 00:16:41.310 they can accommodate unequal standard deviations

341 00:16:41.310 --> 00:16:42.633 in each population.

342 00:16:43.620 --> 00:16:45.300 Again, we can specify the difference

343 00:16:45.300 --> 00:16:47.880 as a standardized effect size.

344 00:16:47.880 --> 00:16:51.780 Cohen suggested values of the effect size that correspond

345 00:16:51.780 --> 00:16:55.470 to small, moderate, and large effects.

346 00:16:55.470 --> 00:16:59.040 A small effect is estimated at 0.2,

347 00:16:59.040 --> 00:17:03.693 a moderate effect is 0.5, and a large effect is 0.8.

348 00:17:04.860 --> 00:17:07.050 Again, as delta decreases,

349 00:17:07.050 --> 00:17:11.493 the sample size necessary to detect that effect increases.

350 00:17:13.050 --> 00:17:16.500 When your primary endpoint is a binary variable,

351 00:17:16.500 --> 00:17:19.350 such as development of surgical site infection,

352 00:17:19.350 --> 00:17:22.890 we summarize the response using the proportion of responders

353 00:17:22.890 --> 00:17:24.870 in each treatment group.

354 00:17:24.870 --> 00:17:28.170 The anticipated effects between groups can be expressed

355 00:17:28.170 --> 00:17:30.990 as the difference in the two proportions,

356 00:17:30.990 --> 00:17:33.000 P two, minus P one,

357 00:17:33.000 --> 00:17:34.800 so you would need to specify

358 00:17:34.800 --> 00:17:37.560 the hypothesized proportion of responders

359 00:17:37.560 --> 00:17:40.443 in each group for the sample size calculation.

360 00:17:42.180 --> 00:17:45.720 Finally, when your primary endpoint is a survival,

361 00:17:45.720 --> 00:17:48.870 or time to event endpoint, such as time to death,

362 00:17:48.870 --> 00:17:50.220 or time to progression,

363 00:17:50.220 --> 00:17:53.160 the anticipated effect size between groups

364 00:17:53.160 --> 00:17:57.300 is usually in the form of a difference in hazard rates

365 00:17:57.300 --> 00:18:01.500 Lambda two, minus lambda one, or a hazard ratio.

366 00:18:01.500 --> 00:18:05.160 You would need to specify the hypothesized hazards

367 00:18:05.160 --> 00:18:09.060 in each group, or the hypothesized hazard ratio.

368 00:18:09.060 --> 00:18:12.690 For example, if the intervention reduces the mortality rate

369 00:18:12.690 --> 00:18:17.403 by 20%, the hazard ratio would equal 0.8.

370 00:18:18.570 --> 00:18:19.950 You may have prior data

371 00:18:19.950 --> 00:18:23.490 on a quantity called median survival time.

372 00:18:23.490 --> 00:18:26.580 This is often reported in the literature.

373 00:18:26.580 --> 00:18:29.490 The median survival time is the time point

374 00:18:29.490 --> 00:18:34.490 when we expect the survival probability to equal 50%.

375 00:18:35.070 --> 00:18:36.510 In the sample data,

376 00:18:36.510 --> 00:18:38.700 the estimated survival probability

377 00:18:38.700 --> 00:18:41.160 or probability of surviving

378 00:18:41.160 --> 00:18:43.560 beyond a certain number of weeks

379 00:18:43.560 --> 00:18:46.530 is plotted on the vertical axis.

380 00:18:46.530 --> 00:18:51.150 The survival curve hits 50% at 23 weeks,

381 00:18:51.150 --> 00:18:52.950 so the median survival time

382 00:18:52.950 --> 00:18:56.910 in this group is estimated to be 23 weeks.

383 00:18:56.910 --> 00:18:59.670 Under the model that we typically use,

384 00:18:59.670 --> 00:19:01.290 the hazard ratio is equal

385 00:19:01.290 --> 00:19:05.790 to the ratio of the median survival times in the two groups.

386 00:19:05.790 --> 00:19:08.490 For example, if the median survival time

387 00:19:08.490 --> 00:19:12.420 in the drug group is twice that seen in the placebo group,

388 00:19:12.420 --> 00:19:16.593 the hypothesized hazard ratio would equal one half.

389 00:19:18.090 --> 00:19:20.700 Other important quantities to specify

390 00:19:20.700 --> 00:19:23.430 in a survival sample size calculation,

391 00:19:23.430 --> 00:19:25.980 is the duration of the accrual period

392 00:19:25.980 --> 00:19:28.680 and the duration of follow up.

393 00:19:28.680 --> 00:19:32.010 These will affect the number of events,

394 00:19:32.010 --> 00:19:35.100 since longer studies have a greater opportunity

395 00:19:35.100 --> 00:19:37.503 to observe study events.

396 00:19:39.270 --> 00:19:42.780 Lastly, I want to discuss an important issue

397 00:19:42.780 --> 00:19:44.850 that affects the required sample size,

398 00:19:44.850 --> 00:19:48.600 and that is the anticipated proportion of subjects

399 00:19:48.600 --> 00:19:50.400 who are lost to follow up.

400 00:19:50.400 --> 00:19:52.380 Since these subjects are lost,

401 00:19:52.380 --> 00:19:54.540 we'll never observe their endpoint,

402 00:19:54.540 --> 00:19:57.480 so we need to compensate for their loss.

403 00:19:57.480 --> 00:20:01.200 If the anticipated loss or withdrawal proportion

404 00:20:01.200 --> 00:20:05.670 is W, where W is a proportion between zero and one,

405 00:20:05.670 --> 00:20:08.520 then the required number of patients per group

406 00:20:08.520 --> 00:20:11.940 should be inflated to n adjusted,

407 00:20:11.940 --> 00:20:15.930 which equals the originally planned per group sample size,

408 00:20:15.930 --> 00:20:18.930 n, divided by one, minus W.

409 00:20:18.930 --> 00:20:22.920 The estimated size of W can often be obtained

410 00:20:22.920 --> 00:20:24.750 from prior studies.

411 00:20:24.750 --> 00:20:26.160 If there's no prior data,

412 00:20:26.160 --> 00:20:30.420 then you may want to set W equal to 0.1 or 10%.

413 00:20:32.280 --> 00:20:34.290 One thing to note is that we're assuming

414 00:20:34.290 --> 00:20:37.950 that the loss to follow up is occurring at random

415 00:20:37.950 --> 00:20:41.790 and it's not related to the health status of the subject.

416 00:20:41.790 --> 00:20:44.370 If it's true that, for example, sicker patients

417 00:20:44.370 --> 00:20:45.990 are dropping out of the study,

418 00:20:45.990 --> 00:20:48.210 then this may bias the results,

419 00:20:48.210 --> 00:20:50.490 especially if more of the sicker patients

420 00:20:50.490 --> 00:20:53.580 are dropping out of one group than the other.

421 00:20:53.580 --> 00:20:56.280 Inflating the sample size for dropouts

422 00:20:56.280 --> 00:20:58.380 will not fix a biased study,

423 00:20:58.380 --> 00:21:00.870 so it's important to try to minimize dropouts

424 00:21:00.870 --> 00:21:02.223 as much as possible.

425 00:21:04.380 --> 00:21:06.090 The sample size calculations

426 00:21:06.090 --> 00:21:10.140 are an important part of the study design process.

427 00:21:10.140 --> 00:21:12.180 The calculations can't be performed

428 00:21:12.180 --> 00:21:14.280 by the statistician alone.

429 00:21:14.280 --> 00:21:17.820 Input from the investigators and the study team is important

430 00:21:17.820 --> 00:21:20.910 when it comes to setting these sample size parameters.

431 00:21:20.910 --> 00:21:24.060 So it's my hope that you've come away with an understanding

432 00:21:24.060 --> 00:21:26.640 of the different factors that you need to consider

433 00:21:26.640 --> 00:21:29.883 and think about during the study planning process.