Chapter 40
Confounding

1 General Purpose

In the Chaps. 23 and 24 confounding of continuous outcome data has been assessed. Briefly, with confounding the treatment efficacies are better in one subgroup than they are in the other. For binary data (= proportional data) this means that the proportions of responders in one subgroups are better than they are in the other subgroup. In the current chapter we will assess confounding of binary outcome data instead of continuous outcome data.

2 Schematic Overview of Type of Data

Treatment modality	outcome (binary)
(0 and 1)	(1 and 2)
.	.
.	.
.	.
.	.
.	.
.	.
.	.
.	.

T.J. Cleophas, A.H. Zwinderman, *Clinical Data Analysis on a Pocket Calculator*,
DOI 10.1007/978-3-319-27104-0_40

3 Primary Scientific Question

Do treatment efficacies perform better in one subgroup than in the other.

4 Data Example, Demonstrating Confounding

The numbers of responders to two different treatments is assessed in a parallel-group study of 384 patients.

	responders	non-responders	total	proportion	SE $(=\sqrt{[p(1-p)/n)]}$
males:					
treat 1	36	50	86	36/ 86 = 0.42	
					$\sqrt{(0.42 \times 0.58/86)}$ = 0.05
treat 2	14	50	64	14/ 64 = 0.22	= 0.05
total	50	100	150		
females:					
treat 1	24	10	34	0.71	=0.08
treat 2	120	80	200	0.60	=0.03
total	144	90	234		
together:					
treat 1	60	60	120		
treat 2	134	130	264		
total	194	190	384		

p = proportion, n = sample size, SE = standard error

For the males the treatments 1 and 2 perform significantly different, because

t = difference of proportions/its pooled variance =
$(0.42 - 0.22)/(SE_{treat\ 1}{}^2 + SE_{treat\ 2}{}^2) =$
$0.20/0.07 = 2.86$.

With proportional data t-values are more often called z-values.

df	One-Tail = .4	.25	.1	.05	.025	.01	.005	.0025	.001	.0005
	Two-Tail = .8	.5	.2	.1	.05	.02	.01	.005	.002	.001
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657	127.32	318.31	636.62
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925	14.089	22.327	31.598
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841	7.453	10.214	12.924
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.265	0.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.263	0.711	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.262	0.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.260	0.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.259	0.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.259	0.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.258	0.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.258	0.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.258	0.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.257	0.689	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	0.257	0.688	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	0.257	0.688	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.257	0.687	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	0.257	0.686	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	0.256	0.686	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	0.256	0.685	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767
24	0.256	0.685	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	0.256	0.684	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	0.256	0.684	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	0.256	0.684	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	0.256	0.683	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	0.256	0.683	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	0.256	0.683	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	0.255	0.681	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
60	0.254	0.679	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
120	0.254	0.677	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
∞	0.253	0.674	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291

The t-table has a left-end column giving degrees of freedom (≈ sample sizes), and two top rows with p-values (areas under the curve = p − values), one-tail meaning that only one end of the curve, two-tail meaning that both ends are assessed simultaneously. The t-table is, furthermore, full of t-values, that, with ∞ degrees of freedom, are equal to z-values. The z-values and t-values are to be understood as mean results of studies, but not expressed in mmol/l, kilograms, or proportions of responders, but in so-called SEM-units (Standard error of the mean units), that are obtained by dividing your mean result by its own standard error. For continuous outcome data, with many degrees of freedom (large samples) the curve will be a little bit narrower, and more in agreement with nature. For binary outcome data, nature has determined that the curves will always be as narrow as can be, according to the row at the bottom.

A z-value of 2.86 is larger than 2.807, and, therefore, statistically significant from zero with a two-tail p-value <0.05.

For the females the treatments 1 and 2 do not perform significantly differently, because

z = difference of proportions/its pooled variance =
$(0.71 - 0.60)/(SE_{treat\ 1}{}^2 + SE_{treat\ 2}{}^2) =$
$0.11/0.09 = 1.22.$

This difference is smaller than 1.960, and therefore, not statistically significantly better than a difference of 0.

For the combined data, the treatments 1 and 2 do not perform significantly differently from zero, because

z = difference of proportions/its pooled variance =
$(0.50 - 0.51)/(SE_{treat\ 1}{}^2 + SE_{treat\ 2}{}^2) =$
0.01/.... = very small.

This difference is again not statistically significant from a difference of zero, and, thus, the treatments do not perform significantly differently from one another.

5 Testing Confounding with a Z-Test

A weighted mean proportion is calculated, and tested (variance = SE^2). The underneath differences indicate, respectively, the differences in the males and the females.

$$\frac{\text{difference/variance} + \text{difference/variance}}{1/\text{variance} + 1/\text{variance}} = \frac{0.20/0.05^2 + 0.11/0.09^2}{1/0.05^2 + 1/0.09^2} = \frac{80 + 13.6}{4000 + 123.5} = \frac{93.6}{523.5} = 0.18$$

This weighted mean proportion of 0.18 is much closer to 0.20 than to 0.11, and this is due to the much larger sample size of females than that of the males. We will now test the weighted mean proportion against its SE (standard error).

SE $= \sqrt{[1/(1/SE_1{}^2 + 1/SE_2{}^2)]} = 0.044$
z = weighted mean/its SE = $0.18/0.044 = 4.9$

This z-value is larger than 3.291, and, therefore, produces a two-tail p-value of <0.01, and, so, after adjustment for confounding between males and females, the treatments 1 and 2 perform very significantly different from one another.

6 Testing Confounding with a Mantel-Haenszl Test

The Mantel-Haenszl chi-square test is equivalent to the z-test.

males

observed	expected	variance (n = 150)
36	(86 × 50)/150 = 28.7	$86 \times 64 \times 50 \times 100/n^2(n-1) = 8.21$

females

observed	expected	variance (n = 234)
24	(34 × 144)/234 = 20.9	$34 \times 200 \times 144 \times 90/n^2(n-1) = 6.91$.

Chi-square $= (28.7 - 20.9)^2/(\text{var } 1 + \text{var } 2) = 7.8^2/15.12 = 4.02$

With 1 degree of freedom the p-value should be < 0.01 (var = variance).

The underneath chi-square table has an upper row with areas under the curve, a left-end column with degrees of freedom, and a whole lot of chi-square values.

The chi-square value of 4.02 with one degree of freedom is larger than 3.841, and, thus the two-tail p-value is <0.05.

Chi-squared distribution

	Two-tailed *P*-value			
df	0.10	0.05	0.01	0.001
1	2.706	3.841	6.635	10.827
2	4.605	5.991	9.210	13.815
3	6.251	7.851	11.345	16.266
4	7.779	9.488	13.277	18.466
5	9.236	11.070	15.086	20.515
6	10.645	12.592	16.812	22.457
7	12.017	14.067	18.475	24.321
8	13.362	15.507	20.090	26.124
9	14.684	16.919	21.666	27.877
10	15.987	18.307	23.209	29.588
11	17.275	19.675	24.725	31.264
12	18.549	21.026	26.217	32.909
13	19.812	22.362	27.688	34.527
14	21.064	23.685	29.141	36.124
15	22.307	24.996	30.578	37.698
16	23.542	26.296	32.000	39.252
17	24.769	27.587	33.409	40.791
18	25.989	28.869	34.805	42.312
19	27.204	30.144	36.191	43.819
20	28.412	31.410	37.566	45.314
21	29.615	32.671	38.932	46.796
22	30.813	33.924	40.289	48.268
23	32.007	35.172	41.638	49.728

(continued)

	Two-tailed *P*-value			
df	0.10	0.05	0.01	0.001
24	33.196	36.415	42.980	51.179
25	34.382	37.652	44.314	52.619
26	35.536	38.885	45.642	54.051
27	36.741	40.113	46.963	55.475
28	37.916	41.337	48.278	56.892
29	39.087	42.557	49.588	58.301
30	40.256	43.773	50.892	59.702
40	51.805	55.758	63.691	73.403
50	63.167	67.505	76.154	86.660
60	74.397	79.082	88.379	99.608
70	85.527	90.531	100.43	112.32
80	96.578	101.88	112.33	124.84
90	107.57	113.15	124.12	137.21
100	118.50	124.34	135.81	149.45

7 Conclusion

Both z-test and Mantel-Haenszl chi-square test results show that after adjustment for confounding the non-significant difference between the effects of the treatments 1 and 2 turn into significant effects with p-values of <0.01 and <0.05.

8 Note

More background, theoretical and mathematical information of confounding assessments can be found in Statistics applied top clinical studies 5th edition, Chap. 28, Springer Heidelberg Germany, from the same authors.