11.1 A. Study 239 Paroxetine in Adolescents

EBD-2 Chapter 8 mentions Study 239, a Glaxo–Smith–Kline-funded randomized, double-blind study that compared the antidepressants paroxetine and imipramine with placebo in adolescents with major depression.

The **Results** section of that paper states:

> *Serious adverse effects occurred in 11 [of 93] patients in the paroxetine group, 5 [of 95] in the imipramine group, and 2 [of 87] in the placebo group . . . The serious adverse effects in the paroxetine group consisted of headache during discontinuation taper (1 patient) and various psychiatric events (10 patients) . . . Of the 11 patients, only headache (1 patient) was considered by the treating investigator to be related to paroxetine.*

Although no P-values for adverse events are presented in the paper, if we compare the proportion with serious adverse events with paroxetine (11 of 93) to that with placebo (2 of 87) using Stata, we get the following output:

```
. csi 11 2 82 85, ex

                 |   Exposed   Unexposed  |      Total
-----------------+------------------------+------------
           Cases |        11           2  |         13
        Noncases |        82          85  |        167
-----------------+------------------------+------------
           Total |        93          87  |        180
                 |                        |
            Risk |  .1182796    .0229885  |   .0722222
                 |                        |
                 |      Point estimate    |    [95% Conf. Interval]
                 |------------------------+------------------------
 Risk difference |         .0952911       |    .0224934    .1680887
      Risk ratio |         5.145161       |    1.173574    22.55732
 Attr. frac. ex. |         .8056426       |    .1479022    .9556685
 Attr. frac. pop |         .6816976       |
                 +-------------------------------------------------
                                  1-sided Fisher's exact P = 0.0124
                                  2-sided Fisher's exact P = 0.0188
```

a) The calculation above entirely ignores the fact that there was an imipramine group. If that group were included, the investigators would want to make three comparisons: paroxetine vs. imipramine, paroxetine vs. placebo, and imipramine vs. placebo. Using the Bonferroni correction for testing these three hypotheses at $\alpha = 0.05$, a 2-tailed P-value of $0.05/3 = 0.0167$ would be required to reject the null hypothesis, and results above would not be statistically significant. Do you think the Bonferroni correction is appropriate in this case? Why or why not?

No. The philosophy behind the Bonferroni correction is that investigators are overeager to conclude a treatment is effective when it isn't. In other words, the Bonferroni correction

protects against false positives. The equivalent of the "presumption of innocence" in criminal court is a presumption of ineffectiveness. This question is whether the drug causes an increase in adverse events. Investigators may be overeager to conclude that it doesn't and the interest of regulators and patients is to protect against false negatives, not false positives. In this case it makes sense to presume harm and require evidence against it. This would mean establishing an unacceptable increase in adverse events and showing that the 95% confidence interval excludes it. While this approach may not be practical, it certainly doesn't make sense to increase the threshold for establishing harm.

The **Discussion** states:

> *"Because these serious adverse events were judged by the investigators to be related to treatment in only 4 patients (Paroxetine, 1; imipramine, 2; placebo, 1), causality cannot be determined conclusively."*

b) Do you agree? How should the judgments of the investigators regarding whether adverse events were treatment-related be factored into judgments about causality of adverse events, assuming blinding was maintained?

No. I don't agree. Even if the investigators are blinded to treatment group, they shouldn't decide which adverse events are treatment related. This violates the principle of once randomized always analyzed. They could just exclude a large number of adverse events from the group or groups that had the most. Even if the exclusions were non-differential, they could decrease the total number of adverse events and bias the result towards the null, which favors the drug when you are looking at adverse events. In fact, they appear to have excluded adverse events that were plausibly treatment related. If investigators could determine whether adverse events were caused by treatment, can they also determine which benefits are caused by treatment? Let's skip randomization and blinding and just ask the investigators who the treatment helped and who it hurt.

Obviously, they can't tell which adverse effects were due to treatment any more than they can tell which improvements are due to treatment. Once randomized always analyzed.

c) Does your answer to part b change if you believe blinding of outcome ascertainment was compromised in this *industry-sponsored* trial?

Compromized blinding just makes the problem worse.