

Supplemental Digital Content 1

Neuromuscular Blocking Agent Statistical Analyses

This document is divided into four. First I present a conventional analysis based on Fisher's 'exact' test (henceforth simply "Fisher's test"); I then justify my choice but also comment on the limitations of this conventional methodology and present alternate 'conventional' analyses that provide similar results to Fisher's test; I next provide a Monte Carlo simulation that permits a different analysis, but again comes to very similar conclusions; finally, I perform various "sensitivity analyses" to explore the robustness of the data. All analyses conducted by the author (Johan M van Schalkwyk) used R version 2.13.1 (R Foundation for Statistical Computing, Vienna, Austria). I provide worked examples.

Table 1. Anaphylaxis *versus* Exposure to Neuromuscular Blocking Agents

	succinylcholine	rocuronium	atracurium	other
Anaphylaxis	12	6	3	0
95% CI (Poisson)	6 – 21	2 – 13	0 – 9	0 – 4
Exposure	24,960	14,995	67,354	15,042
Rate	1:2,079	1:2,498	1:22,450	-
Range (from CI)	1: 1,190–4,030	1: 1150–6,810	1: 7680–109,000	1: 4,080–∞

A. A conventional analysis — Fisher's test

Study data for analysis are shown in table 1. The existence of over 60 tests for comparison of binomial proportions¹ underlines the incoherence of modern statistical comparison of such proportions. Two main classes of test exist, 'exact' tests and 'asymptotic' tests. Fisher's test falls into the former and has the advantage that it is considered to be a conservative test in that its use is unlikely to result in a Type I error (finding a difference that's not there).

The rates of anaphylaxis to succinylcholine, rocuronium, and atracurium are respectively about 1:2,080, 1:2,500 and 1:22,400, with no documented anaphylaxis to the other agents (vecuronium, pancuronium, and a small number of exposures to mivacurium). Simple inspection of these rates suggests that there is a large (tenfold) discrepancy between the rate for atracurium and the rates for succinylcholine and rocuronium, but statistical analysis is still desirable. Listing 1 shows entry of our data and an initial application of Fisher's test.

Use of this approach clearly indicates significance with a calculated p value of 6.5×10^{-6} confirming our initial impression. Fisher's test can readily be applied to tables that are larger than 2×2 , but simulation is required for the above data as otherwise the sizes of the denominators result in a workspace error. In the above, 10 million replicates are used. Unsurprisingly, individual 2×2 comparisons reveal that the differences reside in the rates of anaphylaxis to succinylcholine and rocuronium, compared with the other agents, *e.g.*, the p value for rocuronium *versus* atracurium is ~ 0.002 , and would be little affected by correction for multiple comparisons.

Listing 1

```
NMB <- matrix(c(12, 6, 3, 0, 24960-12, 14995-6, 67354-3, 15042), 4, 2,
dimnames = list(nmba=c("sux", "roc", "atra", "other"),
                 anaphylaxis=c("yes", "no")))
NMB
fisher.test(NMB, simulate.p.value=TRUE, B=1e7)
```

B. Limitations of conventional analyses

Lydersen *et al.*² thoroughly explore current knowledge of the problem of analysis of 2 x 2 tables. Perhaps most important is the realisation that many traditional tests are based on the assumption that the marginal values in the 2 x 2 table are fixed. For example, in Fisher's original 'tea-tasting' experiment, the subject knew that she would be presented with eight cups of tea, four of which would have the milk poured in first. Tests like Fisher's that depend on the marginal sums are termed 'conditional'. Lydersen *et al.* assert that Fisher's test should "practically never be used" as it is usually overly conservative.

In our circumstances, however, use of a conservative test seems appropriate. Agresti³ identifies two apparently favourable properties of exact tests: "Exact methods guarantee that the size of a hypothesis test is no greater than the nominal level and that the coverage probability for a confidence interval is at least the nominal confidence coefficient." Use of Barnard's test (another exact test, otherwise known as the Fisher-Boschloo test) is computationally too intensive with our denominators, and results in a workspace error in R.

Commonly used examples of asymptotic methods include Pearson's Chi-square test and the likelihood ratio (LR) statistic. An unfortunate feature of the Chi-square test is Cochran's rather empiric criterion that if any cell count is less than 5, the results produced by a Chi-square are unreliable, but this too can be accommodated by a simulation approach. Listing 2 shows the results of a Chi-square test with simulation, which provides a *p* value similar to that obtained with Fisher's test, although larger.

Listing 2

```
> chisq.test(NMB, simulate.p.value = TRUE, B = 1e7)
      Pearson's Chi-squared test with simulated p-value (based on
1e+07 replicates)
data:  NMB X-squared = 27.3838, df = NA, p-value = 3.77e-05
```

It is clear from the above analysis that, despite minute variation in *p* values, there is an unequivocal difference between the rates for succinylcholine and rocuronium, on the one hand, and other agents (notably atracurium) on the other.

For an optimal analysis, it would seem most sensible to use a Bayesian approach (*e.g.*, with OpenBUGS), but this is complex and I don't have confidence in my ability to conduct such an analysis reliably. In the next section I therefore provide an alternate analysis that is based on very simple assumptions.

Table 2. Expectation of Numbers of Anaphylaxis Cases at or above the Given Threshold

Agent	Cutoff(n)	Expectation (% of runs)	Expectation (n+3, %)	Expectation (n-3, %)
succinylcholine	≥ 12	1.3	-	5.7
rocuronium	≥ 6	13	-	56
atracurium	≤ 3	0.024	0.15	-
other	$=0$	3.3	45	-

C. A modeling approach

An approach that doesn't depend on "choice of test" is to assume the identity of all rates, and then use Monte Carlo simulation to derive the likelihood of the observed frequencies in the various groups:

1. Assume the same overall rate for every group, based on the group average (21 cases of anaphylaxis in 92,858 patients);
2. For the number in each group, repeatedly simulate the number of observed incidents, using the overall rate and random selection;
3. Determine in what percentage of groups the value achieved the observed rate in that group, or was more extreme than the observed rate.

An appropriate R function for this simulation is shown in Listing 3.

Listing 3

```
sim <- function(Ni)
{ for(i in 1:100000)
  { V = sample.int(n=92858, size=Ni, replace=TRUE)
    F[i] = length(V[V<=21])
  }
  return(F)
}
```

This approach is simplistic, because it does not account for the fact that some of the patients have been exposed to more than one agent, and all of the consequent ramifications. However, a compensatory increase in the value 92,858 would result in "more significant" values in table 2, so my simulation approach is conservative.

To simulate the observed number for 100,000 iterations, use the commands shown in Listing 4. The "Expectation" column in table 2 was constructed using this approach. It supports the previous analysis by showing that in just 1.3% of runs would one expect twelve or more cases of succinylcholine anaphylaxis, given the observed rate for the whole group. Similarly, the atracurium rate is vastly divergent from the expected rate. The assumption of a similar rate for the various agents is not supported by this analysis. The sensitivity analysis in the two columns on the right supports the robustness of this conclusion, notably for atracurium (This analysis is further discussed in Section D1 below).

Listing 4

```
roc <- sim(14995)
oth <- sim(15042)
sux <- sim(24960)
atra<- sim(67354)
```

C1. A worked example

Let's consider the calculation of “roc” derived above. We say:

```
troc <- table(roc)
```

This provides a table similar to the following (different runs will differ slightly):

0	1	2	3	4	5	6	7	8	9	10	11	12	13
3336	11373	19430	21698	18571	12682	7180	3499	1427	536	205	46	15	2

In the 100,000 iterations, we'd expect six or more cases in:

```
sum(troc[7,14])/1000
```

percent of cases, *i.e.*, 12.9%. Note that indexing in R starts at 1, not zero.

Confidence intervals can be calculated based on the Poisson distribution. For example, for $n = 6$, we say:

```
poisson.test(6, conf.level=0.95)
```

D. Sensitivity Analyses

In the following sections, I consider various scenarios in which the numerator data from our study are considered in some way “defective”, and explore the analytic consequences. It is important to note that these scenarios are speculative ones, designed to test the robustness of the data in the face of error, and do not reflect actual concerns about data quality. Three sensitivity analyses were conducted (A–C).

D1. Sensitivity analysis A

In the first scenario, I assume that for each drug in turn there was either under or over-diagnosis of three cases. This would (for example) account for cases missed due to deaths not being reported in one group, or the false presumption of a diagnosis where one didn't exist. Note that in such circumstances, the *sim* () function has to be adjusted accordingly, so we define two new functions, as shown in Listing 5.

No attempts were made to simulate the vast array of other possibilities, *e.g.*, overdiagnosis of anaphylaxis to one drug, combined with underdiagnosis with another agent. Because of the existing divergences, we are most interested in the “-3” cases for succinylcholine and rocuronium, and +3 for atracurium and other.

As previously, we can now sum the number of runs that meet the relevant criteria, *e.g.*

```
tsuxlo <- table(suxlo)
```

The results of this analysis are reflected in the two rightmost columns of table 2.

Listing 5

```

simhi <- function(Ni)
{ for(i in 1:100000)
  { V = sample.int(n=92858, size=Ni, replace=TRUE)
    F[i] = length(V[V<=24])
  }
  return(F)
}

simlo <- function(Ni)
{ for(i in 1:100000)
  { V = sample.int(n=92858, size=Ni, replace=TRUE)
    F[i] = length(V[V<=18])
  }
  return(F)
}

roclo <- simlo(14995)
suxlo <- simlo(24960)
atralo <- simhi(67354)
othlo <- simhi(15042)

```

D2. Sensitivity analysis B

A more severe sensitivity analysis considers the case where the cases referred to the clinic in whom no definite diagnosis of anaphylaxis to any agent was made are considered to have had anaphylaxis to one or more of the neuromuscular blocking drugs to which they were exposed. Results are shown in table 3.

Table 3. "Worst-case" Analysis

	succinylcholine	rocuronium	atracurium	other
All possibles	12+2+1+12	8+6	11+2+1+3	3
Exposure	24,960	14,995	67,354	15,042
Rate	1:920	1:1,070	1:4,000	1:5,000

The corresponding application of Fisher's exact test is shown in Listing 6.

This analysis nevertheless provides a p value of $6e-7$, suggesting that even were allergy testing completely uninformative, we would still be required to reject our null hypothesis.

Listing 6

```

WNMB <- matrix(c(27,14,17,3,24960-27,14995-14,67354-17,15042-3), 4, 2,
dimnames = list(nmba=c("sux", "roc", "atra", "other"),
                 anaphylaxis=c("yes", "no")))
WNMB
fisher.test(WNMB, simulate.p.value=TRUE, B=1e7)

```

Table 4. “Restrictive” Analysis

	Succinylcholine	Rocuronium	Atracurium	Other
Restrictive	10	2	0	0
Exposure	24,960	14,995	67,354	15,042

D3. Sensitivity analysis C

Here we take the converse, restrictive approach of strictly conforming to criteria for anaphylaxis, abandoning clinical judgement. Application of Fisher's test is shown in Listing 7. The corresponding numbers are shown in table 4.

Even this excessively restrictive approach still reveals a substantial difference between the agents due to the large contribution from succinylcholine.

Listing 7

```
REST <- matrix(c(10,2,0,0,24960-10,14995-2,67354,15042), 4, 2,
dimnames = list(nmba=c("sux", "roc", "atra", "other"),
                 anaphylaxis=c("yes", "no")))

REST

fisher.test(REST, simulate.p.value=TRUE, B=1e7)
```

References

1. Andrés M, Mato S: Choosing the optimal unconditioned test for comparing 2 independent proportions. *Comput Stat Data Anal* 1994; 17: 555–75
2. Lydersen S, Fagerland MW, Laake P: Recommended tests for association in 2×2 tables. *Stat Med* 2009; 28: 1159–75
3. Agresti A: Exact inference for categorical data: Recent advances and continuing controversies, *Stat Med* 2001; 20: 2709–22