An application of Bayesian multivariate analysis to an Experimental Medicine clinical trial

Ken Edwards *GlaxoSmithKline*

Introduction

Early in drug development studies need to provide insight into how a drug might alter pathogenic disease processes at the same time that they provide a robust go/no-go decision for future development. These demands can be conflicting because adding to the number of endpoints for decision-making increases the chance that one of them could show a biologically important difference by chance alone but overly restricting the number of endpoints means that a medicine may be discarded because a key endpoint wasn't considered.

This trade-off can be framed in a quantitative manner by adopting methods that deal with decision making across multiple endpoints so that stakeholders can agree robust biomarker strategies that have clear decision rules for progression in drug development while also learning about the drug's effect on disease processes. This poster illustrates one approach to doing this using a multivariate Bayesian model.

An Experimental Medicine Trial

Rare diseases often have no effective treatments and there can be little knowledge of the relationship between biomarkers and clinical endpoints.

A drug could be hypothesised to affect an important pathway in the disease but while there are several biomarkers that should be affected by an efficacious drug, there is often not enough evidence in the literature to narrow this down to a single marker where end of study decision making would be simple to quantify.

The approach outlined here involved working closely with the biology and clinical team members to define a subset of biomarkers for decision making and then to fit a model to simulated data of several, potentially correlated, endpoints.

The multivariate Bayesian model provides a flexible framework to understand:

- The probability of observing a meaningful difference across each endpoint and across several endpoints for a quantified scenario.
- The operating characteristics of chosen decision rules for each endpoint.
- Properties of different ways of combining the results of individual endpoints.
- Providing a clear understanding of what successful scenarios might look like, and whether they would be sufficient to support the asset as it competes in the portfolio.
- An estimate of the correlation of the drug effect between different endpoints.
- The ability to use data from external sources to boost the placebo arm through informative priors.

Working with the study team

Initially the biology and clinical members of the study team were focussed on biomarkers that should or could move in response to drug, but were not as focussed on the implications that having many endpoints would have for decision making. In particular, the implication of rules such as 'at least one biomarker of this set should move convincingly' (where this kind of rule is associated with a high false positive rate) were not well understood by all stakeholders. Once this was clear, careful work was performed by the biology team to narrow down biomarkers to an evidence-based primary set of four and the team agreed that the biological hypothesis would be supported if three of these four biomarkers showed a true treatment difference.

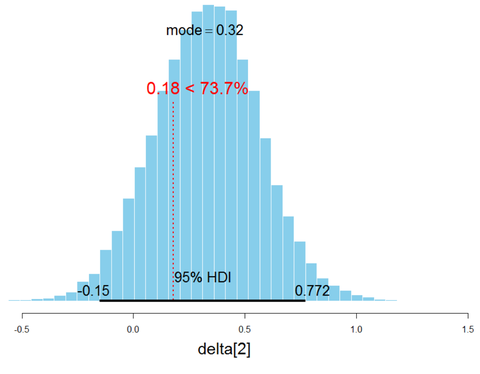
Rules and thresholds

For each biomarker, we can create a rule which passes when the probability is high enough that the true average difference between active and placebo arms is biologically meaningful:

$$P(\delta > \text{critical\_difference}) > \text{critical\_probability}$$

For example, for a particular biomarker, the rule would pass if the posterior probability is greater than 60% that the true difference (δ) between treatments is above 0.18 (where 0.18 is a biologically meaningful difference). Figure 1 shows the posterior probability of the difference for one simulated trial, where the probability that the difference is above 0.18 is ~74%.

Figure 1. Decision making using the posterior probability of the treatment difference for a single endpoint



Simulations

The additional benefit that a multivariate model provides is that it allows the posterior probability of each endpoint to be considered simultaneously, while correctly accounting for the correlation between them.

Simulations were run for different levels of key design factors:
- Subject numbers
- Critical differences (i.e. thresholds for each endpoint)
- Critical probabilities

Under different levels of the factors describing the true environment:
- True differences and SD (for each endpoint)
- True correlations between endpoints

And the properties of different rules were investigated:
- The probability of individual endpoint success
- The joint probability of endpoint success (the probability that all endpoints are above their thresholds, not shown)
- n out of m rules, for example three successes out of four

The results presented here are from a multivariate Normal model which was implemented in R with JAGS used for MCMC. As the number of variables increase the 'design space' (i.e. combination of true values and study design thresholds to be investigated) expands combinatorically and so processing time can become an issue. Using the multivariate normal with a conjugate Wishart distribution for the prior of the covariance matrix is much faster than using a multivariate t distribution (where each variable of the covariance matrix is a separate node). For efficiency, I covered the wider space with the multivariate normal and am investigating chosen scenarios and final rules more thoroughly using the multivariate t-distribution.

The JAGS statements defining the multivariate normal likelihood with a treatment covariate is shown below:

```
for ( i in 1:Ntotal ) {
   zMu[i,1:Nvar] <- zBeta[1:Nvar] + zDelta[1:Nvar]*trt[i]
   zy[i,1:Nvar] ~ dmnorm(zMu[i,1:Nvar] , zInvCovMat[1:Nvar,1:Nvar])
 }
```

Where
- Nvar is the number of variables
- trt is the treatment covariate for each participant
- Vector zBeta is the standardised node representing the placebo effect
- Vector zDelta is the standardised node representing the treatment difference
- Matrix zy is the data
- Matrix zInvCovMat is the inverse covariance matrix (which is how JAGS parameterises the multivariate normal). The prior for this node is the Wishart distribution in the case of the multivariate normal.

Findings

One key advantage of making decisions on several biomarkers is that rules requiring all endpoints to be positive have very low false positive rates when the false positive rate would be high for individual endpoints if the same univariate decision rules were used. The unfortunate corollary of this is that, for the range of true values and plausible minimum differences investigated, the 'power' of this type of rule can be low. For a restricted sample size there is a trade-off between rules with good operating characteristics but where the thresholds appear to be a low hurdle and rules where the minimum difference would be very convincing to all stakeholders, but that have low 'power'. As more endpoints are added this effect becomes more profound. The rule that the study team is adopting is an 'n out of m' type rule and these appear to both have very low false positive rates and good power with more reasonable thresholds. Rules for 'any out of m' can have very high false positive rates and are not recommended.

With fixed sample size and univariate biomarker decision rules, table 1 illustrates the operating characteristics (probability of a trial being declared 'successful') of decision rules across several biomarkers under different 'true' scenarios.

Table 1. Operating characteristics

Number of endpoints with true treatment difference	At least one out of four	At least two out of four	At least three out of four	Four out of four
No true differences	0.28	0.04	0.01	0
One endpoint	0.87	0.12	0	0
Two endpoints	0.96	0.68	0.07	0
Three endpoints	0.99	0.91	0.57	0.02
Four endpoints	1	0.94	0.79	0.38

Further work is planned to summarise the range of observed trial outcomes that would pass the final rule in an effort to gain buy-in from senior management for potential future decisions and to support or challenge the study team's final rules in that context.

Presenting the results to the study team can be challenging. With several endpoints and design factors listed above there are no standard graphical ways of presenting the operating characteristics. One option is to use the posterior density of success to set the probability thresholds of the rules but this presents results on an abstract scale with no reference to the underlying endpoints. Another is to present multiple three dimensional surfaces of the properties of the rules with each axis being the true value of an endpoint but this is also complicated to discuss. So far using an interactive pivot table to summarise outputs for a subset of key scenarios is reasonably effective but work here is ongoing.

Acknowledgements

Graeme Archer, Fabio Rigat and Juan Abellan,
Advanced Biostatistics and Data Analytics, GSK

John K Kruschke
http://doingbayesiandataanalysis.blogspot.co.uk/