Some Devices and Desires of an FDA Advisory Committee Statistician

Frank E Harrell Jr

Division of Biostatistics and Epidemiology
Department of Health Evaluation Sciences
School of Medicine
University of Virginia

9th Annual DIA Meeting on Statistical Issues in the Pharmaceutical Industry
Hilton Head, SC

1 Apr 96

Abstract

This talk will first list some things I would rather not see when reviewing study reports or watching sponsors' presentations. Examples include defining "intention to treat" as "treatment actually received", presenting analyses of "responders", presenting long tabular summaries, inappropriately using standard errors, presenting multiple subgroup tests without formal tests of interaction, overemphasizing site effects (Senn, Stat in Med 14:2661; 1995), and attaching too much importance to very low P–values in very large studies.

I would like to see more use of quantiles and Prob($X < Y$) as descriptive statistics, and charts depicting how patients were dropped from the analysis. Examples will be given showing how dot charts and other graphical methods can replace large tables.

For repeated measurement data with dropouts, sponsors often present analyses of completers or use the "last value carried forward" method. There are better methods for handling informative dropouts (e.g., Lavori et al, Stat in Med 14:1913;1995).

Sponsors often present analyses which adjust for baseline variables with the intention of making up for possible randomization imbalances. Instead, I would like to see more analyses motivated by the desire to estimate the treatment effect in the presence of patient heterogeneity (see Ford et al, Stat in Med 14:735;1995), along

with pre-specified plans for how covariable adjustment is to be done in a blinded fashion (e.g., CPMP Working Party, Stat in Med 14:1659;1995 or Knaus et al, JAMA 270:1233;1993).

Finally, the talk will briefly discuss the need for adding Bayesian interpretations of efficacy and safety data to traditional P-values and confidence limits (Spiegelhalter et al, JRSS A 157:357;1994 and Hughes, Stat in Med 12:1651;1993). For example, I would like to see estimates of the probability that a new drug reduces mortality by at least 10equivalent (within $\pm$ 10%).

Dislikes

- Intention to treat = Treatment actually received
- Analyses of "responders"
- Long tabular summaries
- Using standard errors to hide inter–patient variability
- Subgroup analyses without interaction tests
- Over–emphasizing site effects (Senn, Stat in Med 1995)
- Attaching too much importance to P–values

Would Like to See More of:

- Quantiles for descriptive statistics
- Pr($X < Y$) from Wilcoxon–Mann–Whitney as descriptive stat.
- Charts showing exactly how pts. dropped from analysis
- Efficient estimation of Pr(ΔBP $>$ 5mm Hg) from a model for continuous BP
- Charts instead of tables

Repeated Measurement Data with Dropouts

- Completers or "last value carried forward" commonly used

- Very promising approach: Lavori et al., Stat in Med 1995

- Incorporated a model for the propensity to be missing

- Handled informative dropout, showed disadvantages of older methods

Adjustment for Baseline Variables

- **Not** for randomization imbalances
- To get the model right, and sometimes to gain power
- To estimate treatment effect in presence of pt. heterogeneity (Ford et al., Stat in Med 1995)
- Example: proportional hazards violated without adjustment, satisfied with adjustment
- Covariable adjustment method developed when treatment blinded (CPMP Working Party, Stat in Med 1995, Knaus et al. JAMA 1993)

Bayesian Interpretation

- Flat and skeptical priors
- Posterior probability density is an excellent display
- Compute Pr(mortality reduction $\geq 10\%$) Pr(drug A equivalent to drug B $\pm\, 10\%$)
- Prevent over–excitement from low P–values with large n by computing prob. of clinically useful effects
- Spiegelhalter et al. JRSS A 1994, Hughes Stat in Med 1993
- If base monitoring and final analysis on Bayesian methods, don't need complex adjustments for interim analyses