Statistical Analysis of the Age Data

Introduction

Age is not normally associated with survival in late-stage prostate cancer. In other large recent trials (TAX327, Abiraterone PhIII, Alpharadin PhIII) the groups receiving the same docetaxel regimen which is now the standard of care, showed no difference in survival between older and younger patients.

Below is a summary of the analysis conducted by a statistician and the conclusions which he would draw with no *a priori* expectations. The analysis is, by necessity, imperfect because it calculates backwards from the medians and confidence intervals which are the only available source of this information. However, the potential consequences of these imperfections are too small to alter the conclusions.

Statistician's Conclusion

"Out of context the most obvious conclusion from [the table of treatment effect by age] is that **the major effect found by the study is age dependence in the placebo arm**. Since the placebo arm age difference is expected to be insignificant, **the cause of this primary, novel "age effect" cannot be separated from the alleged treatment effect**."

Methodology for Statistical Analysis

The unpublished, unexpected age data was found in the table below from the FDA's 2010 Statistical Review

Table 10 Subgroup Analysis of Overall Survival by Age and Race

	Sipuleucel-T N	Sipuleucel-T Median Survival (95% CI)[1]	Placebo N	Placebo Median Survival (95% CI)[1]	Sipuleucel-T vs. placebo Hazard Ratio (95% CI)[2]
Age					
< 65	106	29.0 (22.8, 34.2)	66	28.2 (23.4, 32.5)	0.919 (0.618, 1.366)
≥ 65	382	23.4 (22.0, 27.1)	183	17.3 (13.5, 21.4)	0.661 (0.538, 0.813)

1. Assuming normal distributions for median survival times (survival time tends to be exponentially distributed rather than normally distributed so the imputed p-values are imperfect, but the conclusion will be the same), the 95% confidence intervals translate into +/- 1.96 standard deviation intervals. i.e. the difference of upper and lower limits is 3.92 standard deviations. We can therefore back out estimates of the standard deviations in question as:

Age	Treated (T)	Placebo (P)
< 65 (Y)	29 +/- 2.91	28.2 +/- 2.32
≥ 65 (O)	23.4 +/- 1.30	17.3 +/- 2.02

The exact CI of an exponential would use percentiles of a chi-squared distribution with 2*n* degrees of freedom. Here *n* is large enough that those values chi-squares would be very close to the normal ones, even in the tails. All p-values are for two-sided tests, but the rankings and relative values for one-sided tests would also be almost identical.

2. As usual for normals, the standard deviation of a difference is the Pythagorean diagonal of the two component standard deviations. Below are the differences sorted by statistical significance:

Comparison	rawDiff	Sigmas	P-value (2-sided)
PY-PO	10.9	3.54	0.00040
TY-PO	11.7	3.30	0.00097
TO-PO	6.1	2.54	0.011
PY-TO	4.8	1.80	0.072
TY-TO	5.6	1.76	0.078
TY-PY	0.8	0.21	0.83

- a small Python program was used to compute the differences, and standard software was used to compute p-values for a normal distribution.

3. All three prominent differences in the table involve the older group in the placebo arm, the single largest being the placebo arm "young vs old" difference. Indeed, out of context the most obvious conclusion from this table is that the major effect found by the study is age dependence in the placebo arm. Since the placebo arm age difference is expected to be insignificant, the cause of this primary, novel "age effect" cannot be separated from the alleged treatment effect. The placebo arm age difference is **twice as large** as the treatment effect in the older population and drastically less likely.

Afterword

There are many reasons why the unexpected association between age and treatment demands attention. The irrefutable statistical significance of the association between placebo treatment and age is merely one of arguments to consider the "Alternative Explanation", despite the age analysis being retrospective. Others are given below (reproduced from the Rebuttals & FAQ page)

Rebuttals & FAQ

1. "This is a retrospective analysis, and therefore invalid"

It might be suggested that the observations regarding the perplexing age data are invalid because the analysis was not pre-specified. This argument is sometimes appropriate, but it is woefully insufficient here.

1. The most important reason to investigate these anomalies is because advancing age is known to be the primary cause of deterioration in the immune system. In the trial of an invasive protocol removing and manipulating immune cells, age is perhaps the single most important subset to analyse.

2. In a teleconference with the FDA in 2006, years before the IMPACT data were analysed, Dendreon proposed submitting this subgroup analysis from the two smaller, earlier trials as part of their first application for approval. Then when submitting the IMPACT data for their second approval application, they included these subgroup analyses. In my personal opinion, specifying a subgroup several years before a trial's data are unblinded is the next best thing to being a pre-specified analysis. In their analysis of 9901 (the first trial to be unblinded and analysed) they identified the statistical significance of the relationship between age and survival in their trials, and noted that this association had not been found by other analyses of similar populations. When trial 9902 was split into parts A and B, it would have been an opportune time to **pre-specify** this analysis for 9902B.

3. Appropriate use of the "retrospective analysis" argument is generally limited to the association of a single prognostic factor with an outcome. The fact that there are four, separate, interwoven and overlapping,

unanticipated observations in the age data make their significance infinitely greater, and almost impossible to attribute to chance alone.

i. The drug does not appear to work in patients <65 – unexpected

ii. Patients >65 in the placebo group live 11 months (median) shorter than patients <65 – unexpected

iii. Patients >65 in the Provenge group live 5.6 months (median) shorter than patients <65 – unexpected (given the observation that the drug does not appear to be working in the younger patients, although not statistically significant due to high variability in the data)

iv. The age group in whom the drug "works" (patients >65) live shorter than the age group in whom the drug doesn't work – unexpected (and, to me, baffling)

It was unexpected observations, and scientists who, instead of dismissing them, dedicated themselves to discovering their causes, which lead to some of our biggest medical breakthroughs. e.g. antibiotics, vaccines and Viagra. Ignoring them simply keeps us from learning what they could teach us.

4. Provenge approval was contingent upon retrospective analyses.

- In IMPACT, the survival of patients <65 greatly favored placebo patients. Only by retrospectively pooling the patients from all 3 trials [a highly unorthodox analysis, since trials had significant differences and the FDA typically forbids any such pooling] was it possible to neutralize this observation into a more benign one in which the drug merely showed no efficacy in patients <65.
- When first applying for FDA approval in 2007, Dendreon had two small trials as proof of efficacy. D9902A (n=98) was not statistically significant, and the significance of the survival benefit in D9901 (n=127) was the result of retrospective analysis.

5. The observations regarding age speak only to the question of which of the two hypotheses to explain the trial results is more probable. The balance of evidence showing harm from cell losses in other trials would not be affected even if the problematic age data were entirely ignored.