STATISTICAL PRACTICE AND CHALLENGES IN CHRONIC DISEASE CLINICAL TRIALS: Diabetes And Cardiovascular Diseases

John M. Lachin
Professor of Biostatistics, Epidemiology and Statistics
The Biostatistics Center
The George Washington University

www.bsc.gwu.edu
Diabetes and CVD

Long-term chronic diseases
  Multiple outcomes
  Multiple mechanisms
Require long-term studies
Statistical Practice and Challenges

Robustness versus efficiency
Missing Data
Composite and Multiple Outcomes
The GRADE study
Robustness versus Efficiency

Efficiency:
“Optimal” method in some sense under a specific model, e.g. estimation efficiency.

Robustness:
Good efficiency (power) over a range of settings or alternatives
$t$-test versus Wilcoxon
Survival Analysis

Logrank ($Z^0$) optimal under a PH model.

Peto-Peto-Prentice Wilcoxon ($Z^1$) optimal under a proportional odds model.

Robust Tests:

Supremum (Renyi) versions (F-H-O’S, ‘87)

Max($|Z^0|, |Z^1|$), linear combinations (Lee, ‘06)

Gastwirth MERT convex combination (’85)

All provide good power over a range of alternatives

None provide an estimate of a natural parameter of interest to describe the observed difference
Survival Analysis

Others:

AFT Model of covariate effect on $\log(t)$

Jin et al. (‘03) rank based semi-parametric model

$$100(e^{\beta} - 1) = \% \text{ change in } t \text{ per unit } x$$

Under a PH model,

$$1/HR = \text{ proportional time delay}$$

MDs prefer HR
Preferences

Investigators understand models (e.g. PH) that generally provide estimates and tests of simple parameters that allow clear exposition.

Investigators understand need for robust inferences, e.g. variance estimates.

PH model with Lin-Wei sandwich estimator one option.

But, robust tests alone not appealing.
Preferences

Inferentially, robust methods preferred. Lack of a meaningful index or parameter is an obstacle to use of robust methods for inference.

We need more experience with robust methods.

Perhaps extensive re-analysis of public repository data could provide more comfort, if not support for use.
Incomplete (Missing) Data
The Fundamental Issue - BIAS

Incomplete Data:
The observed results reflect a combination of:

The properties of the larger population
PLUS

Bias due to missing data.

If missing data is informative, then:

Parameter estimates, confidence limits and \( p \)-values may be biased.
Bias and $\alpha$

Assume a two group study under $H_0$. Test the difference in proportions.
Control group probability is 0.20.
Non-random missing data introduces a bias of 0.05.
Treatment group probability is 0.25.
If $p = 0.05$ then

N = 200, type I error probability $\alpha = 0.14$
N = 800, type I error probability $\alpha = 0.40$
Can’t Statistics Handle This?

*Not definitively.*

The magnitude of the bias cannot be estimated, no correction possible. Analyses can be conducted *under certain assumptions.* But there is no way to prove that the assumptions apply. Best way to deal with missing data is to prevent it.
What About Last Observation Carried Forward?

*LOCF* is a HOAX.

Imputed values do not contain nearly as much statistical information as real values.

Parameter estimates may be biased.

Variance estimates (standard errors) are too small.

*p*-values are too small (significance overstated).

*Yet it persists.*
CANA (SGLT-2) vs. SITA (DPP-4)

*Diabetes Care*, 2013

LOCF: 292 of 756 (39%) missing at 52 weeks

Baseline (%)

<table>
<thead>
<tr>
<th></th>
<th>SITA 100 mg</th>
<th>CANA 300 mg</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline (mmol/mol)</td>
<td>8.1</td>
<td>8.1</td>
</tr>
<tr>
<td>Baseline (%)</td>
<td>65</td>
<td>65</td>
</tr>
</tbody>
</table>

168 124 missing
FDA Guidance on “Diabetes Mellitus: Developing Drugs and Therapeutic Biologics for Treatment and Prevention”

Statistical analysis using last observation carried forward (LOCF) is easy to apply and transparent in the context of diabetes trials. Assuming an effective investigational therapy, it is often the case that more placebo patients will drop out early because of a lack of efficacy, and as such, LOCF will tend to underestimate the true effect of the drug relative to placebo providing a conservative estimate of the drug’s effect. Perhaps true for the mean difference (point estimate), not the inference.
How Can We Prevent Another LOCF And Encourage Appropriate Methods

LOCF not published in peer review (to my knowledge). Red flag.

FDA Statistical Advisory Board:
The original BEMAC
Professional Continuing Education Societies, Academia
Graduate Education

*We teach statisticians to teach statisticians, not to be statisticians*
The Intent-to-Treat Design

The only incontrovertibly unbiased study is one with no or minimal missing data.

In a clinical trial, this leads to the *intent-to-treat design*.

All patients randomized are evaluated as scheduled as objectively as possible and are included in all analyses, regardless.

Requires continued follow-up of all patients entered, regardless.
Withdrawal From Treatment Versus Withdrawal from Follow-up

Withdrawal from treatment may occur for "real life” clinical or ethical reasons including

Safety, poor compliance, lack of benefit, etc.

BUT, withdrawal from treatment should not require withdrawal from study follow-up.

Criteria For Withdrawal From Study:

Death, complete withdrawal of consent, or end of study
THE DCCT (NEJM, 1993)
1441 subjects with type 1 diabetes.
Mean 6.5 years follow-up (4-9 years)
8 of 1430 survivors did not complete the end-of-study assessment
32 of 1441 patients were inactive for some period, the majority returned to follow-up.
Patients remained on assigned therapy for 97% of the time in study.

Results are incontrovertible.
MULTIPLE OUTCOMES

Comparative Effectiveness:

Compare treatments on multiple domains

Type 2 Diabetes:

Level of HbA1c (a measure of glucose control, lower better)

Weight gain (a frequent side effect, lower better)

Hypoglycemia (low glucose that leads to coma/siezure, less better).
MULTIPLE OUTCOMES - Issues

Separate tests:

Concordance of confidence limits with test for each component

Requires multiplicity adjustment.

Omnibus ($T^2$–like) test ($K df$):

Directed to any difference in any direction

Neither approach specifically directed to the restricted alternative of a beneficial effect on one or more outcomes, say $H_{1+}$.

\[ H_0: \delta_a = 0 \text{ and } \delta_b = 0 \]

\[ H_{1+}: \delta_a \geq 0 \text{ and } \delta_b \geq 0 \text{ and } \text{sum}(\delta_a, \delta_b) > 0 \]
MULTIPLE OUTCOMES - Issues

Multivariate One-sided test:

Test of $H_0$ versus $H_{1+}$.

Wei-Lachin test of stochastic ordering

$Z_j = Z$-test for the $j$-th component

Test statistic $= \Sigma_j Z_j$

Maxmin efficient

For test of two mean differences with correlation 0.5, the W-L test is

61% more efficient than pairwise tests

32% more efficient than 2 df $T^2$ test
COMPOSITE OUTCOMES

Major Adverse Cardiovascular Event (MACE):
   Fatal or non-fatal MI
   Fatal or non-fatal stroke

MACE+
   Unstable Angina
   Revascularization

Similar constructs in other diseases
COMPOSITE OUTCOMES

Survival analysis of time to the first component event.

Justification:

All components should show some benefit if the therapy has a beneficial effect on the underlying cardiovascular disease.

More subjects with an event means greater power, smaller study.
COMPOSITE OUTCOMES - Issues
Are there more efficient and robust tests?

Wei-Lachin multivariate one-sided test

\[ T_j = \text{Aalen-Gill test for time to first of the } j\text{-th type of event} \]

Test statistic = \( \sum_j T_j \)

But, assumes no component is “absorbing”. Different approach needed for MACE.
COMPOSITE OUTCOMES - Issues

Recurrent events analysis

Not considered helpful.

Once a patient experiences first event then at very high risk of subsequent events that may not be affected by treatment.
Glycemia Reduction Approaches in Diabetes: A Comparative Effectiveness Study (GRADE Study)
Diabetes

Hyperglycemia leads to microvascular and macrovascular complications
Near normal glycemia mitigates this risk.
10 new classes of drugs approved for treatment since 1995.
No long-term effectiveness studies
Minimal FDA requirement is to lower HbA1c (blood glucose) over 6 months.
Compare the effectiveness among 4 medications, in combination with metformin, to maintain Hba1c < 7% over an average of 5 years.

All drugs are commonly used in combination with metformin

The 4 most frequently used:

- Sulfonylurea (e.g. glyburide)
- DPP-4 inhibitor (e.g. sitagliptin)
- GLP-1 agonist (e.g. liraglutide)
- Basal insulin (e.g. lantus)
OTHER GRADE OBJECTIVES

Compare incidence of successive secondary and tertiary metabolic failures.

Compare adverse effects, tolerability and cost effectiveness.

Compare incidence of onset of diabetic nephropathy and other microvascular complications.

Compare incidence of CVD (though under-powered)
GRADE Design

5000 early T2D subjects, prior treatment with metformin for < 3 years.

37 clinical centers

3 years recruitment, 7 years total duration

Minimum 4 years potential follow-up

Intent-to-treat design – continued follow-up until study end, regardless.
Other Outcomes

Changes in weight and other risk factors

Incidence of hypoglycemia

Composite outcomes (e.g. free of metabolic outcomes with no weight gain and no hypoglycemia).

Patient oriented outcomes, quality of life

Economic cost-effectiveness evaluation
GRADE ANALYSES

Discrete time PH model with robust variance estimate for time to primary metabolic failure.

Models for time to secondary and tertiary metabolic failure.

No interim monitoring plans for termination for effectiveness or futility.

Multivariate one-sided test for multiple outcomes (metabolic failure, weight gain, hypoglycemia).
GRADE DESIGN

98% power to detect a difference among the 4 drugs using an omnibus test where one drug has a 25% relative risk difference versus the others.

90% power to detect a 25% relative risk difference between any pair of drugs, adjusting for 6 tests.

Assuming lagged recruitment with 4% losses per year yielding an average follow-up of 3.8 years.
CONCLUSION

As an intent-to-treat trial, GRADE should provide a definitive evaluation of the 4 most commonly used therapies for type 2 diabetes.

Should inform physician and patient decisions on therapeutic choices.

Should identify the most cost-effective care possible in early type 2 diabetes.