Survival Analysis in Clinical Trials: The Need to Implement Improved Methodology

Lucinda (Cindy) Billingham
Professor of Biostatistics
Director, MRC Midland Hub for Trials Methodology Research
Lead Biostatistician, Cancer Research UK Clinical Trials Unit
University of Birmingham
Agenda

• Context: phase III cancer clinical trials
• Describe and critique the ‘typical’ statistical approach to survival analysis in cancer clinical trials and debate the need for change
• Trials to assess treatments that improve survival time in rare diseases
• Discussion: implementing changes in practice
Maintenance pemetrexed plus best supportive care versus placebo plus best supportive care for non-small-cell lung cancer: a randomised, double-blind, phase 3 study

Tudor Ciuleanu, Thomas Brodowicz, Christoph Zielinski, Joo Hang Kim, Maciej Krzakowski, Eckart Laack, Yi-Long Wu, Isabel Bover, Stephen Begbie, Valentina Tzekova, Branka Cucevic, Jose Rodrigues Pereira, Sung Hyun Yang, Jayaprakash Madhavan, Katherine P Sugarman, Patrick Peterson, William J John, Kurt Krejcy, Chandra P Belani

Lancet October 2009, Vol 374, No 9699, p1432-1440
Survival Analysis Results

Median survival (95% confidence interval)
Pemetrexed: 13.4 (11.9-15.9)
Placebo: 10.6 (8.7-12.0)

Hazard Ratio = 0.79 (0.65-0.95)
p-value = 0.012
(Cox model with single treatment covariate)

Number at risk

<table>
<thead>
<tr>
<th></th>
<th>Pemetrexed</th>
<th>Placebo</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>441</td>
<td>222</td>
</tr>
<tr>
<td>Events</td>
<td>340</td>
<td>160</td>
</tr>
<tr>
<td>1 year</td>
<td>221</td>
<td>93</td>
</tr>
<tr>
<td>2 years</td>
<td>141</td>
<td>60</td>
</tr>
<tr>
<td>3 years</td>
<td>63</td>
<td>29</td>
</tr>
<tr>
<td>4 years</td>
<td>29</td>
<td>13</td>
</tr>
<tr>
<td>5 years</td>
<td>11</td>
<td>4</td>
</tr>
<tr>
<td>6 years</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Is the Methodology Appropriate?

- Kaplan-Meier estimates of survivor functions
  - Parametric
  - Hazard functions, differences in hazards and survival
- Summary measures: median survival and hazard ratio
  - Absolute difference in hazard
  - Restricted mean survival time (Royston 2011)
- Cox model with treatment as covariate i.e. Log-rank test
  - 83 centres from 20 countries
  - Stratification factors (disease stage, ECOG performance status, sex, best response to CT, non-platinum component of CT, history of brain mets)
  - Other prognostic factors
Another Example: The ISEL Trial

Gefitinib plus best supportive care in previously treated patients with refractory advanced non-small-cell lung cancer: results from a randomised, placebo-controlled, multicentre study (Iressa Survival Evaluation in Lung Cancer)

Nick Thatcher, Alex Chang, Purvish Parikh, José Rodrigues Pereira, Tudor Ciuleanu, Joachim von Pawel, Sumitra Thongprasert, Eng Huat Tan, Kristine Pemberton, Venice Archer, Kevin Carroll*

Lancet 2005; Vol 366, p1527-1537
ISEL: Survival in Overall Population

Supportive Cox regression analysis gives $p=0.03$

Stratified Log-rank test

HR=0.89 (0.77, 1.02)

Stratification Factors:
- Histology
- Smoking history
- Reason for previous CT failure
- Performance Status
- Sex

‘Supportive’ Cox regression analysis gives $p=0.03$
Figure 5: Comparison of hazard ratios from ISEL and BR21
Another Example: The IPASS Trial

The NEW ENGLAND JOURNAL of MEDICINE 2009

Gefitinib or Carboplatin–Paclitaxel in Pulmonary Adenocarcinoma


Untreated patients in East Asia with pulmonary adenocarcinoma, non-smokers or former light smokers
‘Gefitinib is superior to carboplatin-paclitaxel’
Debate Re: IPASS Publication

- NEJM Dec 2009
- Seruga, Amir and Tannock
  - ‘Primary analysis is inappropriate and does not support the superiority of gefitinib over carboplatin-paclitaxel’
  - ‘If the curves cross, there is clear violation of the proportional hazards model and the hazard ratio should not be used as a measure of relative benefit’ (ref Concato et al 1993)
- Suggest re-analysis with the use of statistical methods that do not rely on assumption of proportional hazards, such as the modified Kolmogorov-Smirnov test (ref Le CT 2004)
ICH E9 Guidance: Statistical Principles for Clinical Trials: Statistical Analysis Plan

1.2 Scope and Direction
For each clinical trial contributing to a marketing application, all important details of its design and conduct and the principal features of its proposed statistical analysis should be clearly specified in a protocol written before the trial begins. The extent to which the procedures in the protocol are followed and the primary analysis is planned a priori will contribute to the degree of confidence in the final results and conclusions of the trial. The

5.1 Prespecification of the Analysis
When designing a clinical trial the principal features of the eventual statistical analysis of the data should be described in the statistical section of the protocol. This section should include all the principal features of the proposed confirmatory analysis of the primary variable(s) and the way in which anticipated analysis problems will be handled. In case of exploratory trials this section could describe more general principles and directions.

The statistical analysis plan (see Glossary) may be written as a separate document to be completed after finalising the protocol. In this document, a more technical and detailed elaboration of the principal features stated in the protocol may be included (see section 7.1). The plan may include detailed procedures for executing the statistical analysis of the primary and secondary variables and other data. The plan should be reviewed and possibly updated as a result of the blind review of the data (see 7.1 for definition) and should be finalised before breaking the blind. Formal records should be kept of when the statistical analysis plan was finalised as well as when the blind was subsequently broken.
5.7 Subgroups, Interactions and Covariates

The primary variable(s) is often systematically related to other influences apart from treatment. For example, there may be relationships to covariates such as age and sex, or there may be differences between specific subgroups of subjects such as those treated at the different centres of a multicentre trial. In some instances an adjustment for the influence of covariates or for subgroup effects is an integral part of the planned analysis and hence should be set out in the protocol. Pre-trial deliberations should identify those covariates and factors expected to have an important influence on the primary variable(s), and should consider how to account for these in the analysis in order to improve precision and to compensate for any lack of balance between treatment groups. If one or more factors are used to stratify the design, it is appropriate to account for those factors in the analysis. When the potential value of an adjustment is in doubt, it is often advisable to nominate the unadjusted analysis as the one for primary attention, the adjusted analysis being supportive. Special attention should be paid to centre effects and to the role of baseline measurements of the primary variable. It is not advisable to adjust the main analyses for covariates measured after randomisation because they may be affected by the treatments.
An unreasonable prejudice against modelling?

Stephen Senn*, †
Department of Statistics, University of Glasgow, Glasgow, UK

This coverage of covariates reflects, I believe, the fact that in the pharmaceutical industry we are more prepared to adjust for covariates than elsewhere, and I personally think this is good.

Outside the industry you will sometimes encounter a prejudice against sophistication in analysis. For example, odds ratios have been

So I will nail my colours to the mast. Where they disagree, I generally prefer the results of analysis of covariance to simpler models, of proportional hazards analyses to the log-rank test and of on the first period alone. Analyses that make intelligent use of prognostic variables can considerably increase the precision of our inferences.
Improper analysis of trials randomised using stratified blocks or minimisation

Brennan C. Kahan*† and Tim P. Morris

Many clinical trials restrict randomisation using stratified blocks or minimisation to balance prognostic factors across treatment groups. It is widely acknowledged in the statistical literature that the subsequent analysis should reflect the design of the study, and any stratification or minimisation variables should be adjusted for in the analysis. However, a review of recent general medical literature showed only 14 of 41 eligible studies reported adjusting their primary analysis for stratification or minimisation variables. We show that balancing treatment groups using stratification leads to correlation between the treatment groups. If this correlation is ignored and an unadjusted analysis is performed, standard errors for the treatment effect will be biased upwards, resulting in 95% confidence intervals that are too wide, type I error rates that are too low and a reduction in power. Conversely, an adjusted analysis will give valid inference. We explore the extent of this issue using simulation for continuous, binary and time-to-event outcomes where treatment is allocated using stratified block randomisation or minimisation. Copyright © 2011 John Wiley & Sons, Ltd.
Debate

• Is current practice to take simplistic approach to survival analysis?
• Why?
• Should we change practice to plan more complex approaches for primary analysis?
• What barriers do we need to overcome to do this?
Rare Diseases: The Dilemma

Randomised Phase III trials are the optimal method for establishing best patient care.

Patients with rare diseases have the same right to evidence based treatment as those with common diseases.

Phase III trials in rare diseases will never be large enough to determine best practice with adequate certainty.

Trials in rare diseases are not a worthwhile investment due to high cost-utility.
Recognising the Need to Undertake Trials in Rare Cancers

- **International Rare Cancers Initiative**
  - Led by Professor Matt Seymour, Director of NCRN
  - CRUK, EORTC, NCI
  - 8 rare cancers selected for phase III trial plus more

- **European Network for Cancer Research in Children and Adolescents (ENCCCA)**
  - FP7 grant for 11 million Euros
  - Work Package on statistical design and analysis for rare paediatric cancers

- **European Clinical Trials in Rare Sarcomas within an Integrated Translational Trial Network (EuroSarc)**
  - FP7 grant for 5 million Euros
  - Work Package on statistical design for rare sarcomas

- **October 5th 2012 – RSS Medical Section Meeting on Trials in Rare Cancers**
Example: Phase III Trial in Merkel Cell Carcinoma

Merkel Cell Carcinoma Stages I-III; completed definitive loco-regional therapy (surgery+/-RT) with curative intent

RANDOMISED

STANDARD
No adjuvant systemic treatment

NEW
4 cycles of etoposide and carboplatin

Overall survival

Hypothesis test: to detect an increase in 5-year survival rates from 55% to 62% (HR=0.8), 90% power, 5% significance
Need 848 events, 2044 patients

230 new cases of stage I-III MCC per year in UK
Given 20% recruitment rate, take 44 years to recruit
With 100 events in 250 patients: power is 20%
Proposed Methodology

• Limited literature
• Latest review provided by: Gupta S et al, A framework for applying unfamiliar trial designs in studies of rare diseases; JClinEpi 2011; 64: 1085-1094
  – Cross-over designs
  – N-of-1 trials
  – Adaptive designs
    • Response-adaptive randomisation
    • Ranking and selection
    • Internal pilot
    • Sequential
  – Bayesian analysis
• Implementation in practice: hypothesis testing with relaxed type I and II errors
Example: The 111 Trial
Professor Michael Cullen (CI), University Hospital Birmingham
Dr Emma Hall (Statistician), Institute for Cancer Research

High risk, stage 1, non-seminoma germ cell tumours of the testis

Registration into trial

Experimental treatment: 1 cycle of adjuvant BEP chemotherapy

2-year recurrence-free survival = 95%

Historical Control: 2 cycles of adjuvant BEP chemotherapy

2-year recurrence-free survival = 98%

CTAAC funded, single arm, phase III trial

Would need 1110 patients to demonstrate equivalence
236 patients are sufficient to exclude a recurrence-free survival rate of less than 95% (A’Hern’s test)
Lilford’s Proposal
Lilford R, Thornton JG, Braunholtz D
Clinical trials and rare diseases: a way out of a conundrum BMJ 1995

- Ethics of small clinical trials
  - Small well designed study versus no study
  - Contribute to a pool of knowledge

- Proposes an alternative view to clinical trials:
  - Carry out a trial NOT to gain a definitive answer but to change the level of uncertainty

- Bayesian perspective is useful in these circumstances
  \[ p \left( \text{treatment effect lies in a particular range} \mid \text{data, prior} \right) \]
  \[ p\text{-value} = p \left( \text{data} \mid \text{no treatment effect} \right) \]

- Make use of all knowledge, results from non-randomised studies should not be discarded
Example: Bayesian Approach to Survival Analysis

Advanced NSCLC, Stage IIIb/IV, N=422

MIC
Mitomycin 6mg/m² IV d1
Ifosfamide 3g/m² IV d1
Cisplatin 50mg/m² IV d1
Every 3 weeks, max 4 cycles

GC
Gemcitabine 1,200mg/m² IV d1,d8
Carboplatin AUC5 IV d1
Every 3 weeks, max 4 cycles

Gemcitabine Plus Carboplatin Versus Mitomycin, Ifosfamide, and Cisplatin in Patients With Stage IIIB or IV Non–Small-Cell Lung Cancer: A Phase III Randomized Study of the London Lung Cancer Group

Simplest Approach: Conjugate Models

Conjugate models make the calculations easier

Conjugate models occur when the posterior distribution is of the same family as the prior distribution

<table>
<thead>
<tr>
<th>Prior</th>
<th>Likelihood</th>
<th>Posterior</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
</tr>
<tr>
<td>Beta</td>
<td>Binomial</td>
<td>Beta</td>
</tr>
<tr>
<td>Dirichlet</td>
<td>Multinomial</td>
<td>Dirichlet</td>
</tr>
<tr>
<td>Gamma</td>
<td>Poisson</td>
<td>Gamma</td>
</tr>
</tbody>
</table>
Normal Conjugate Analysis For Survival Data

\[ \theta \] represents the true treatment effect in terms of \( \log(\text{HR}) \)

Likelihood : \[ y_m \sim N \left[ \theta, \frac{2^2}{m} \right] \] \( m = \text{total number of events} \)

(Tsiatis 1981)

Prior for \( \theta \) : \[ N[\mu_0, 2^2/m_0] \]

Posterior for \( \theta \) : \[ N \left[ \frac{m_0 \mu_0 + m y_m}{m_0 + m}, \frac{2^2}{m_0 + m} \right] \]

- Posterior is a weighted average of prior mean and data
- As \( m \to \infty \) so posterior \( \to \) data
**Frequentist Results**

Medians: 10.2 vs 6.9  
HR = 0.76  
95% CI: 0.61 to 0.93  
Log-rank test: p-value = 0.008

**Bayesian Results**

Non-informative Prior  
Mean HR = 0.76  
95% CI: 0.61 to 0.93  
P(HR < 0.9) = 0.93

‘Likelihood-based Bayesian Analysis’  
‘Standardised likelihood’
Possible Prior Distributions for LLCG Study 11

Pre-trial
HR=1 most likely

Data-based
HR=1.11 most likely

Enthusiastic
HR=0.90 most likely

Protocol: “95% CI on the HR at end of trial would include survival difference of 10% or greater in favour of MIC at 1 year”
Mechanics Behind the Bayesian Results

Pre-trial Prior

Protocol: “95% CI on the HR at end of trial would include survival difference of 10% or greater in favour of MIC at 1 year”
Bayesian Analysis of LLCG Study 11

Mean Hazard Ratios and 95% credible intervals

- Classical
- Non-informative
- Pre-trial
- Data-based
- Enthusiastic

Hazard Ratio: 0.50, 0.75, 1.00, 1.25

p(HR<0.9):
- NA
- 0.93
- 0.71
- 0.27
- 0.77

GC superior, MIC superior
Likelihood-Based Bayesian Analysis / Standardised Likelihood

- Estimating posterior distribution for a treatment effect parameter using a non-informative prior distribution
- Standardised likelihood
- Hughes MD, Reporting Bayesian analyses of clinical trials, Statistics in Medicine 1993, 12: 1651-1663
- Burton PR, Helping doctors to draw appropriate inferences from the analysis of medical studies, Statistics in Medicine 1994, 13: 1699-1713
Example: Phase III Trial in Merkel Cell Carcinoma

Merkel Cell Carcinoma Stages I-III; completed definitive loco-regional therapy (surgery+/-RT) with curative intent

RANDOMISED

STANDARD
No adjuvant systemic treatment

NEW
4 cycles of etoposide and carboplatin

Overall survival

Hypothesis test: to detect an increase in 5-year survival rates from 55% to 62% (HR=0.8), 90% power, 5% significance

Need 848 events, 2044 patients

230 new cases of stage I-III MCC per year in UK
Given 20% recruitment rate, take 44 years to recruit
With 100 events in 250 patients: power is 20%
Example: Bayesian Analysis of Trial Data with No Prior Information

Data: \( \text{HR}=0.8 \), \( d=50 \)

\[ P(\text{HR}<1)=0.78 \]

Posterior probability distribution for True HR

95%CI: (0.46, 1.39)

\( d=309 \) would give 95%CI that falls below 1

\[ P(\text{HR}<1)=0.78 \]
Changing Certainty with Changing Study Size

Data: HR=0.8, d=10 to 100
Changing Certainty with Changing Prior Information

Data: HR=0.8; Prior HR=0.8, d=0, 10, 20, 50, 100
Changing Certainty with Changing Prior Information

Data: HR=0.8; Prior HR=0.5, d=0, 10, 20, 50, 100
Summary / Issues for Debate

• Descriptive analysis on survival data in clinical trials should be extended to include more than Kaplan-Meier survival curves
• Pre-planned primary statistical analysis of survival outcome measures should be based on modelling
• Trial statisticians need to be provided with training and tools to ensure implementation (sample size)
• Statistical inference based on standardised likelihoods should be used more and may be preferable to hypothesis testing in trials of rare diseases
• Modelling rather conjugate analysis may be more appropriate