Using short term endpoint data in interrupted clinical trials

Nigel Stallard

Statistics & Epidemiology, Division of Health Sciences
Warwick Medical School, University of Warwick, Coventry, UK
n.stallard@warwick.ac.uk

Interrupted trials

Over 2,500 COVID clinical trials started in 2020

650 in April 2020 alone

Around 5,000 ongoing clinical trials in other disease areas

Many of these are disrupted due to

- cancellation of non-essential medical procedures
- restrictions on face-to-face assessments
- non-attendance due to restrictions or illness

Continue if possible, but early stopping may be necessary

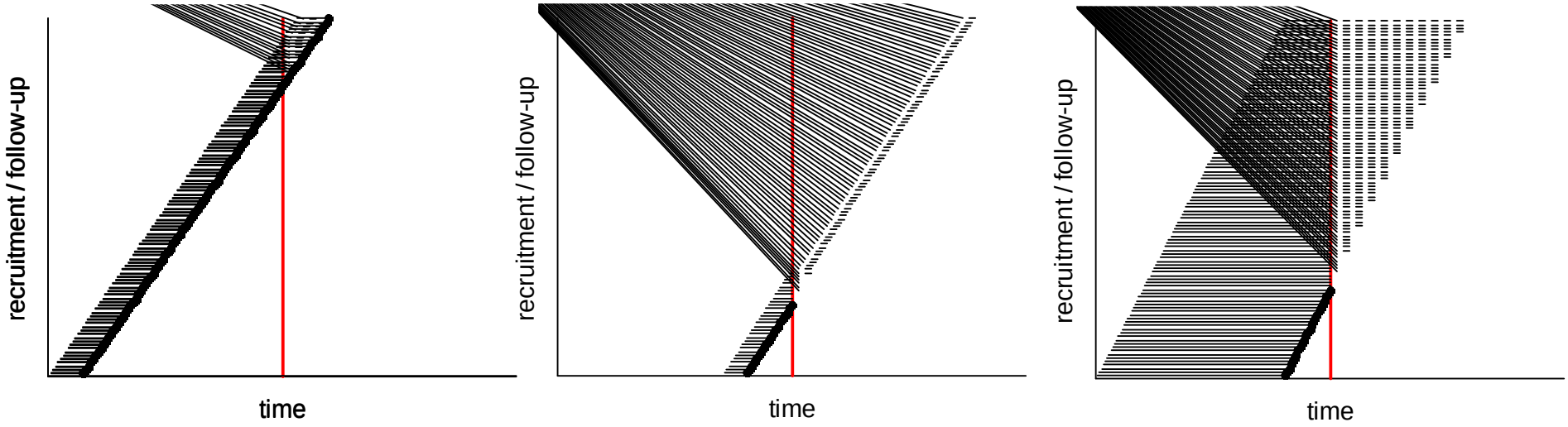
www.covid-trials.org, www.clinicaltrials.gov

EMA Points to consider on implications of COVID-19 on methodological aspects of ongoing clinical trials

FDA Guidance: Statistical considerations for clinical trials during the COVID-19 public health emergency

Interrupted trials

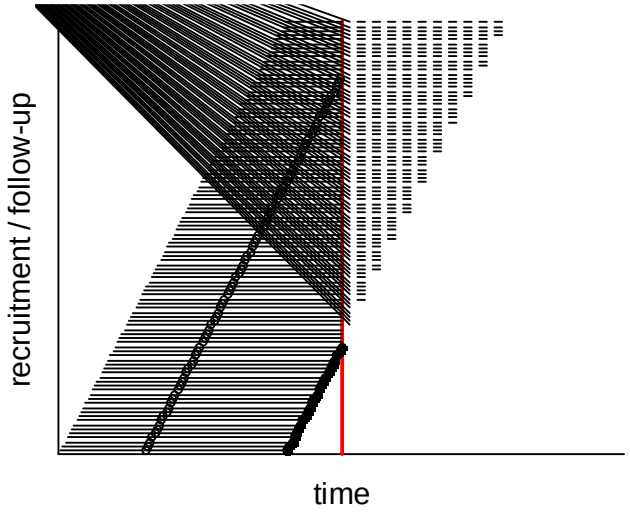
Trials may be impacted differently



Kunz et al (2020) Statistics in Biopharmaceutical Research, 12, 461-77.

Interrupted trials

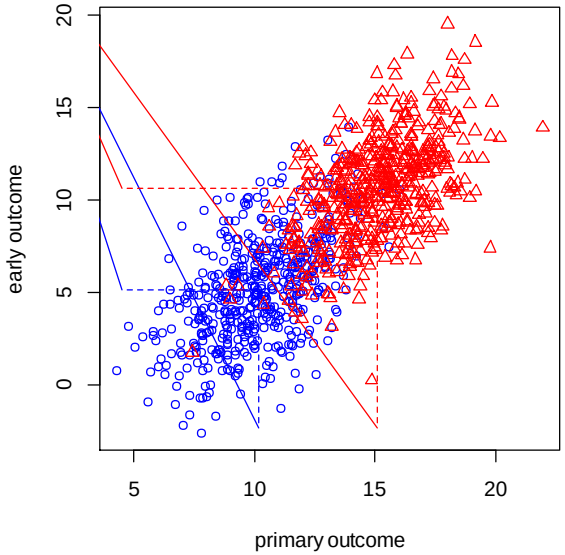
Early endpoint data might be available

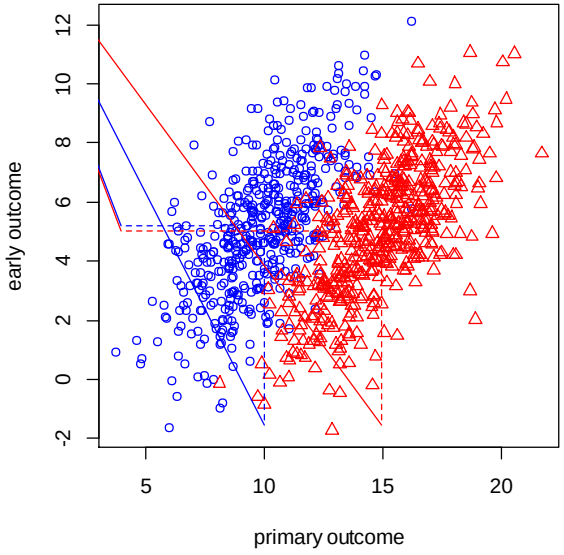


Using early endpoint data – change of endpoint

Criteria similar to those for surrogate endpoint

- Clinical plausibility
- Correlations within and between groups

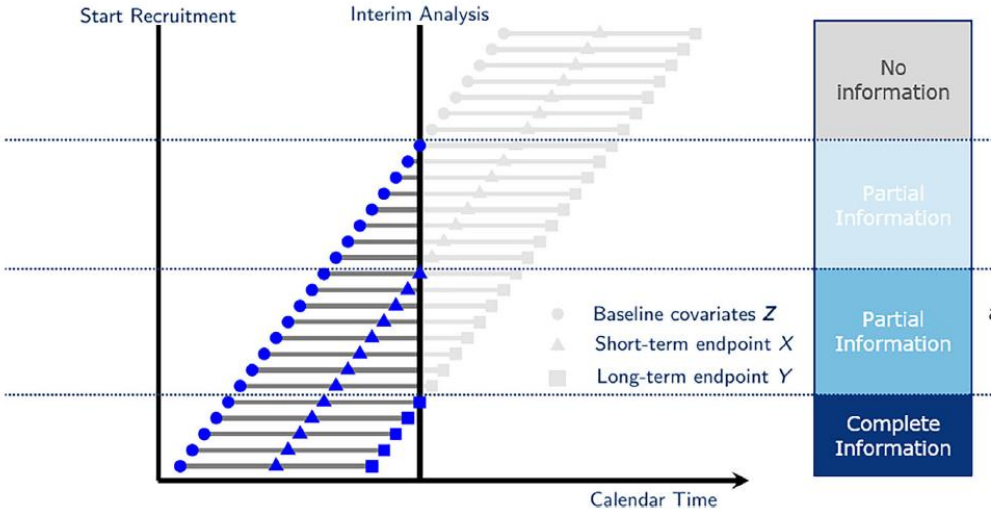




- Power will depend on effect size on early endpoint

Using early endpoint data – gaining precision

Similar to using early endpoint data in interim analysis



Idea: Specify model for early and final endpoints

Obtain likelihood for early and late data

Obtain MLEs for final endpoint parameters

Van Lancker et al (2020) Pharmaceutical Statistics, 12, 461-77.

Using early endpoint data – binary data

Estimate p = pr(death by time t) in single group of n patients

	Death	No death	Total
Final:	d	$n-d$	n

Likelihood = $p^d(1-p)^{n-d}$

$\hat{p} = d/n$ $\quad$ $\mathrm{E}(\hat{p})$ $\quad$ $\mathrm{Var}(\hat{p}) = p(1-p)/n$

Marschner, Becker (2001) Statistics in Medicine, 20, 177-92.

Using early endpoint data – binary data

Observe death by t_1 ($< t$) or t for n patients

	Death	No death	Total
Early:	e	$n - e$	n
Final:	$e + (d - e)$	$n - d$	n

Likelihood = $p_1{}^{e}(p - p_1)^{d-e}\ (1 - p)^{n-d}$

$\hat{p} = d/n$ $\quad$ $\mathrm{E}(\hat{p})$ $\quad$ $\mathrm{Var}(\hat{p}) = p(1 - p)/n$

p = pr(death at t_1) + pr(death at t | alive at t_1)pr(alive at t_1)

$\hat{p} = e/n + ((d - e)/(n - e)) \times ((n - e)/n)$

Marschner, Becker (2001) Statistics in Medicine, 20, 177-92.

Using early endpoint data – binary data

Observe death by t_1 for additional m patients

	Death	No death	Total
Early:	e	$n-e$	n
Final:	$e + (d-e)$	$n-d$	n
Early:	f	$m-f$	m

Likelihood: $p_1^{e}(1-p_1)^{n-e}\, p_1^{d-e}(1-p_1)^{n-d}\, p_1^{f}(1-p_1)^{m-f}$

pr(death at t_1) pr(alive at t_1) pr(death at t_1) pr(alive at t_1)

$$\hat{p} = (e+f)/(n+m) + ((d-e)/(n-e)) \times (n-e+m-f)/(n+m)$$

$$\mathrm{E}(\hat{p}) = p \qquad \mathrm{Var}(\hat{p}) = \left(1 - \frac{\rho^2 m}{n+m}\right)\frac{p(1-p)}{n}$$

Marschner, Becker (2001) Statistics in Medicine, 20, 177-92.

Using early endpoint data – normal data

Single group example:

Patients 1 to n: primary endpoint Y_i

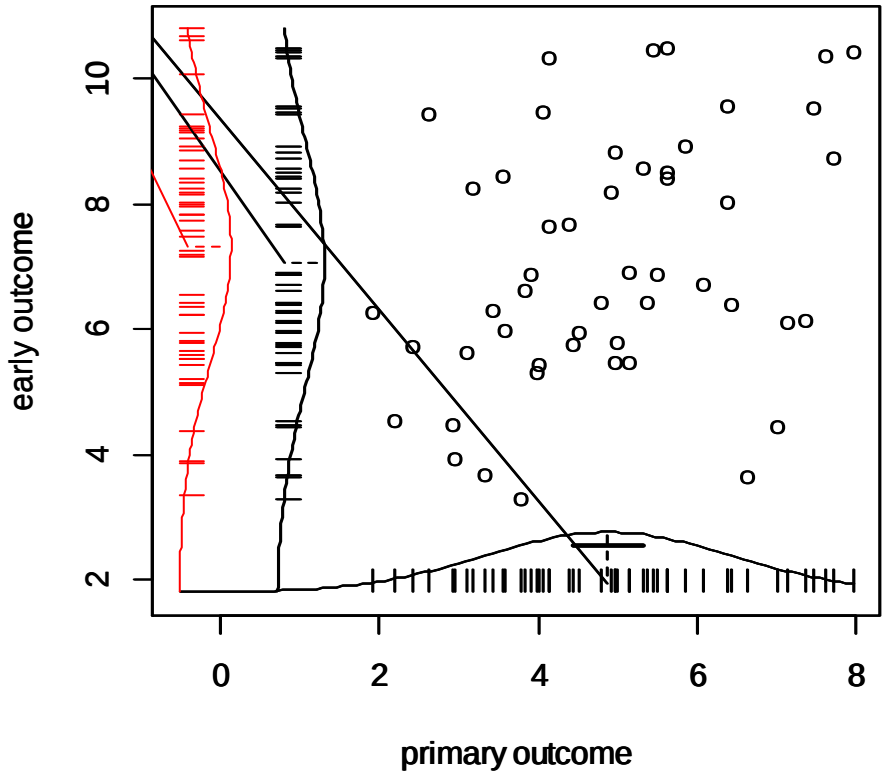
early endpoint X_i

$$\begin{pmatrix} X_i \\ Y_i \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_X \\ \mu \end{pmatrix}, \begin{pmatrix} \sigma_X^2 & \rho\sigma\sigma_X \\ \rho\sigma\sigma_X & \sigma^2 \end{pmatrix} \right)$$

Patients $n+1$ to m: early endpoint X_i only

$$X_i \sim N(\mu_X, \sigma_X^2)$$

Using early endpoint data – normal data



Using early endpoint data – normal data

'Double regression'

Fit regression model $E(Y) = \alpha + \beta X$ using n patients

Estimate μ by $\tilde{\mu} = \hat{\alpha} + \hat{\beta}\bar{x}$

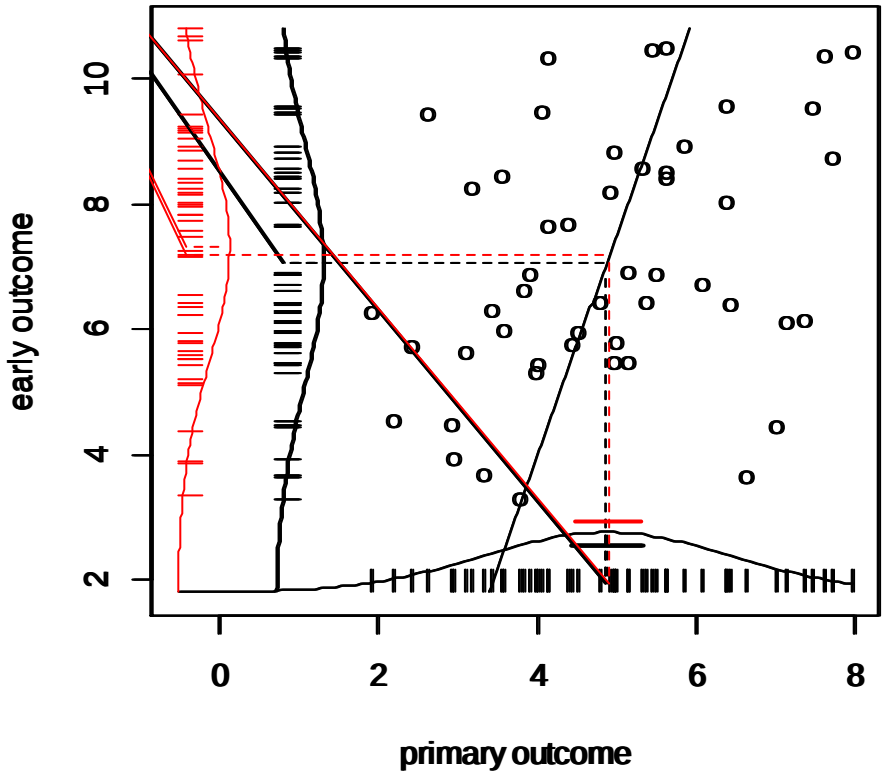
where $\bar{x}$ is mean from all $n + m$ patients

$$\mathrm{E}(\tilde{\mu}) = \mu$$

$$\mathrm{Var}(\tilde{\mu}) = \left(1 - \frac{\rho^2 m}{n+m}\right)\frac{\sigma^2}{n}$$

Galbraith, Marschner (2003) Statistics in Medicine, 22, 1787-1805.
Engel, Walstra (1991) Biometrics, 47, 13-20.

Using early endpoint data – normal data



Using early endpoint data – gaining precision

Notes:

Gain in precision comes from correlation within group

Does not depend on treatment effect on early endpoint

'Effective sample size' is

$$n \leq \frac{n(n+m)}{n+(1-\rho^2)m} \leq n+m$$

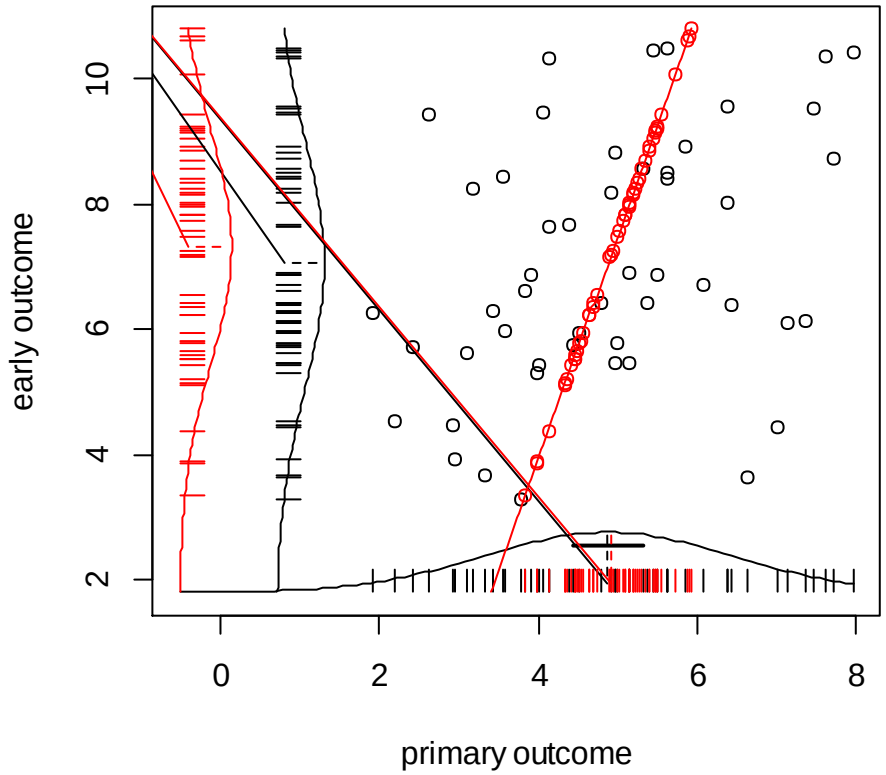
$(\rho^2 = 0)$ $(\rho^2 = 1)$

A missing data problem

	Death	No death	Total
Early:	e	$n-e$	n
Final:	$e+(d-e)$	$n-d$	n
Early:	f	$m-f$	m
Final:	$f+(m-f)\frac{(d-e)}{(n-e)}$	$(m-f)\frac{(n-d)}{(n-e)}$	

$$\hat{p}=\left(d+f+(m-f)\frac{(d-e)}{(n-e)}\right)/(n+m)$$

A missing data problem



Open questions

Generalisation to other models and data types
including time to event data

Link to methods for missing data

Regulatory issues

…