Overview of Recurrent Events in Clinical Trials

September 26, 2017

Brian Claggett, PhD
Brigham and Women's Hospital
Harvard Medical School

Clinical Trials

- A clinical trial is an experiment

- Years of effort and millions of dollars devoted to answering a single clinical question:
  - Crudely: “Does the treatment work?”
  - Specifically: “With respect to the **pre-specified primary clinical outcome**, is there any difference (that is unlikely to arise solely by chance) between two groups of patients who are similar in every respect except for the treatment that they receive?”

Choosing a Clinical Primary Outcome/Endpoint

- Three primary considerations:
  - **Meaningful**: That the chosen endpoint is an accurate measure of the burden of disease
  - **Modifiable**: That the therapies we are testing will differentially affect this endpoint
  - **Practical**: That this endpoint occurs often enough that sufficient data can be collected in a reasonable time frame

We Rarely Have the Luxury of Just Counting Bodies

the ONION

World Death Rate Holding Steady At 100 Percent

GENEVA, SWITZERLAND--World Health Organization officials expressed disappointment Monday at the group's finding that, despite the enormous efforts of doctors, rescue workers and other medical professionals worldwide, the global death rate remains constant at 100 percent.

Death, a metabolic affliction causing total shutdown of all life functions, has long been considered humanity's number one health concern. Responsible for 100 percent of all recorded fatalities worldwide, the condition has no cure.

"I was really hoping, what with all those new radiology treatments, rescue helicopters, aerobics TV shows and what have you, that we might at least make a dent in it this year," WHO Director General Dr. Gernst Bladt said. "Unfortunately, it would appear that the death rate remains constant and total, as it has inviolably since the dawn of time."

Many are suggesting that the high mortality rate represents a massive failure on the part of the planet's health care workers.

<u>All-cause Mortality</u>

Meaningful: YES

Modifiable: Somewhat?

Practical: Depends?

Heart Failure Examples: CIBIS-II

THE LANCET • Vol 353 • January 2, 1999

The Cardiac Insufficiency Bisoprolol Study II (CIBIS-II): a randomised trial

*CIBIS-II Investigators and Committees**

Summary

Background In patients with heart failure, β-blockade has improved morbidity and left-ventricular function, but the impact on survival is uncertain. We investigated the efficacy of bisoprolol, a β_1 selective adrenoceptor blocker in decreasing all-cause mortality in chronic heart failure.

Heart Failure Examples: BEST

N Engl J Med, Vol. 344, No. 22 · May 31, 2001 · www.nejm.org · 1659

A TRIAL OF THE BETA-BLOCKER BUCINDOLOL IN PATIENTS WITH ADVANCED CHRONIC HEART FAILURE

THE BETA-BLOCKER EVALUATION OF SURVIVAL TRIAL INVESTIGATORS*

ABSTRACT

Background Although beta-adrenergic–receptor antagonists reduce morbidity and mortality in patients with mild-to-moderate chronic heart failure, their effect on survival in patients with more advanced heart failure is unknown.

Methods A total of 2708 patients with heart failure designated as New York Heart Association (NYHA) functional class III (in 92 percent of the patients) or IV (in 8 percent) and a left ventricular ejection fraction of 35 percent or lower were randomly assigned to double-blind treatment with either bucindolol (1354 patients) or placebo (1354 patients) and followed for the primary end point of death from any cause.

The First Event is the Best Event?

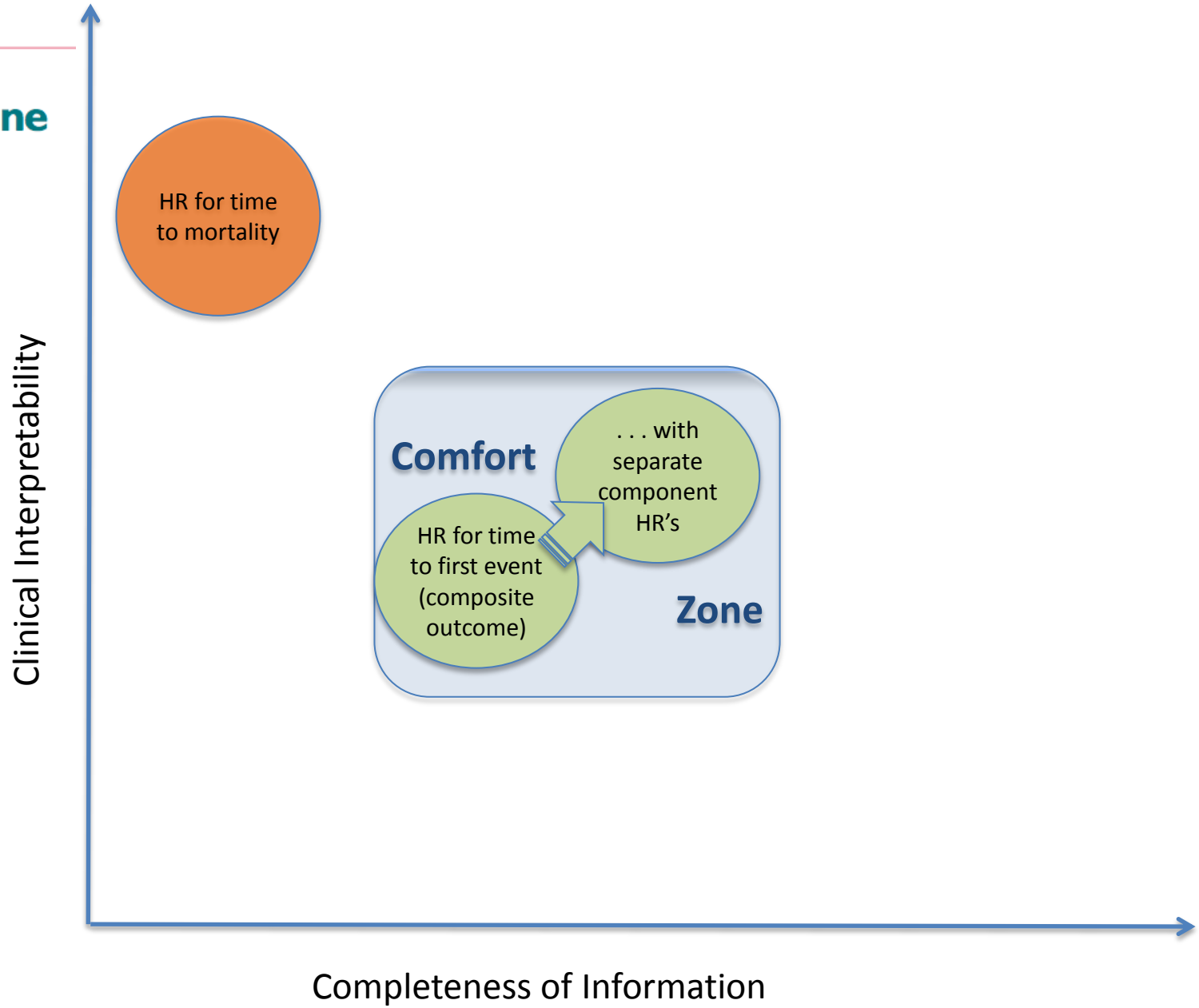
- Many investigators then began declaring the primary endpoint to be the **first** occurrence of “Death or . . . “
  - Myocardial Infarction
  - Heart Failure
  - Unstable Angina
  - Revascularization
  - Stroke
  - Myocardial Ischemia
  - Cardiovascular Hospitalization

- This approach has been successful in dozens of therapies that have changed the practice of cardiovascular medicine
  - Trials: CONSENSUS, SAVE, MERIT-HF, TIMI-1, ISIS, V-HEFT, HOPE, 4-S, CARE, PROVE-IT
  - Drug Classes: ACE Inhibitors, Beta-Blockers, ARBs, Thrombolytics, Aspirin, Statins

European Heart Journal (2013) **34**, 869–871
doi:10.1093/eurheartj/ehs485

Moving beyond our comfort zone

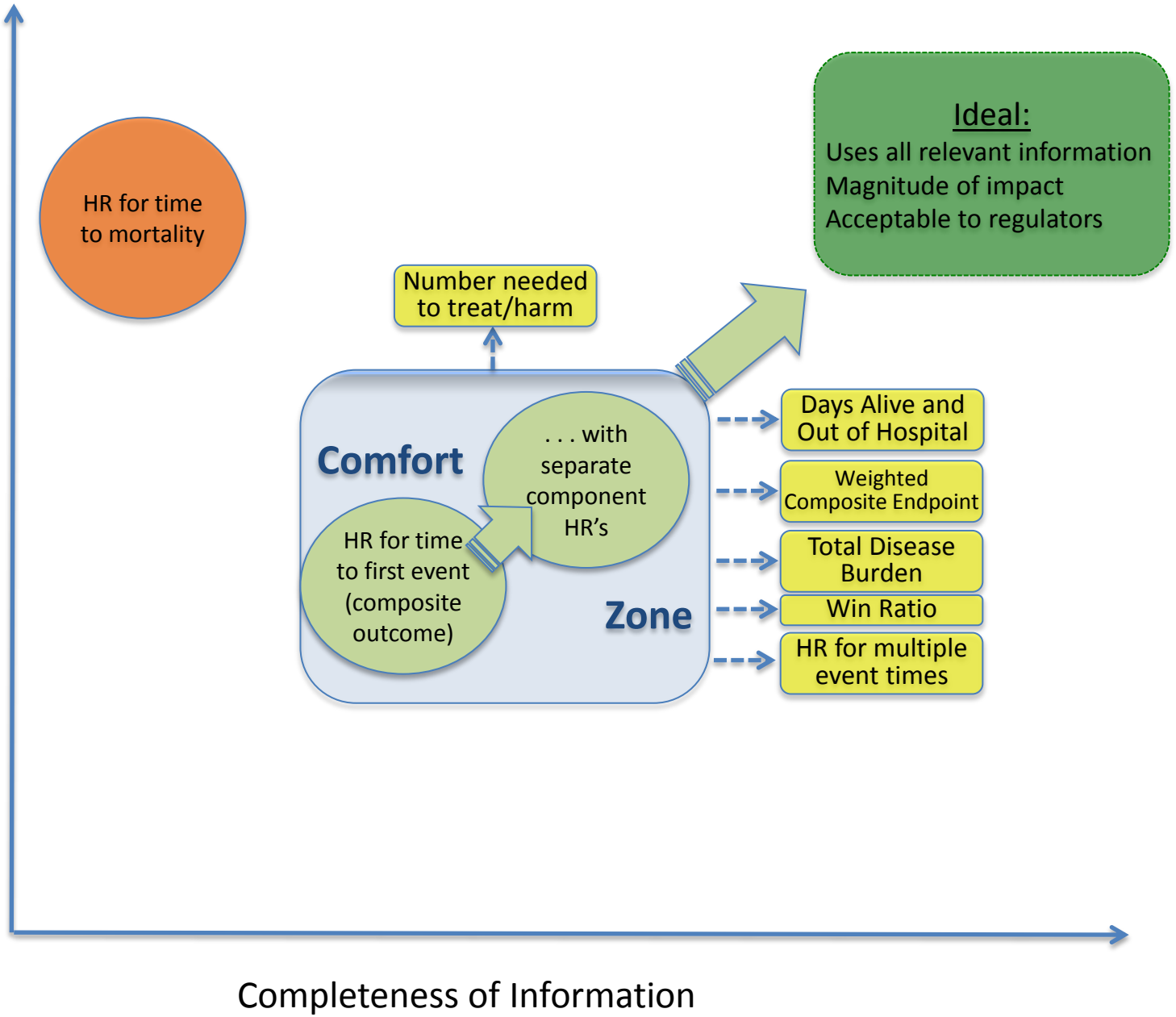
Brian Claggett[1], Lee-Jen Wei[1], and Marc A. Pfeffer[2]*



European Heart Journal (2013) **34**, 869–871
doi:10.1093/eurheartj/ehs485

Moving beyond our comfort zone

Brian Claggett[1], Lee-Jen Wei[1], and Marc A. Pfeffer[2]*



Choosing a Clinical Primary Outcome/Endpoint

- Three primary considerations:
  - **Meaningful**: That the chosen endpoint is an accurate measure of the burden of disease
  - **Modifiable**: That the therapies we are testing will differentially affect this endpoint
  - **Practical**: That this endpoint occurs often enough that sufficient data can be collected in a reasonable time frame

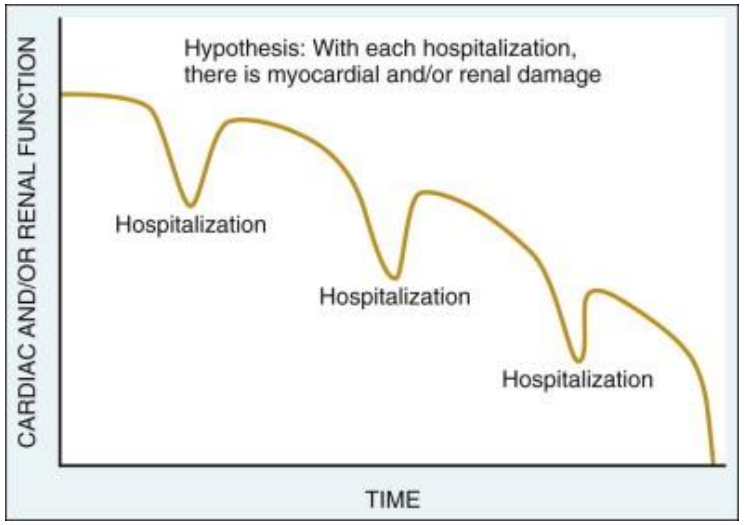
- These assumptions are relevant when deciding whether/how to use a "composite" endpoint
- They are also relevant when considering whether/how to use a recurrent event endpoint

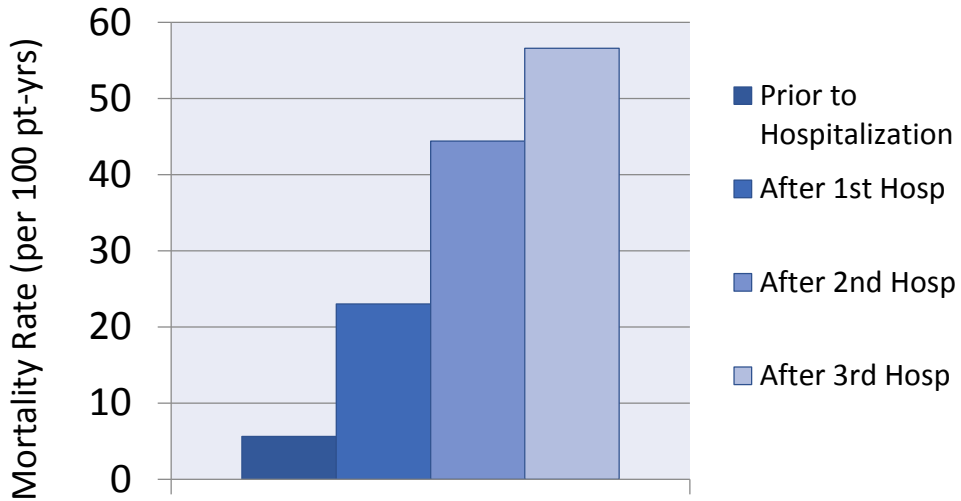
Rationale for Recurrent Event Analyses

- Chronic diseases are characterized by recurrent “encounters”
  - hospitalizations, ED visits, office visits etc.

- Analysing all such events (i.e. the “patient journey”) may be a more accurate measure of the true burden of disease

The Patient Journey





Solomon et al. Circulation 2005

- Each subsequent HF hospitalization heralds a substantial worsening of the long term prognosis
- This suggests that the patient journey is important and can portend outcomes. It is therefore important to clinicians and patients.

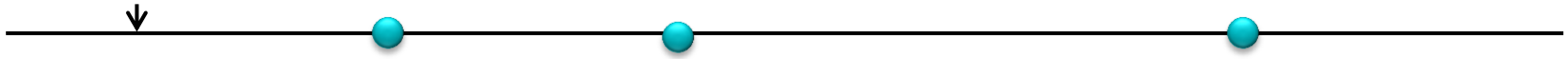
Why do we care about the “full burden” of disease?

- Becoming more difficult to consider efficacy of a therapy without considering the cost –
  - direct costs to the healthcare system
  - also immeasurable costs, monetary and otherwise, to patients, families, society

- Important to determine if a therapy has same effect on recurrent events as first event (desirable in chronic diseases) – in other words, does the therapy have a “persistent” effect

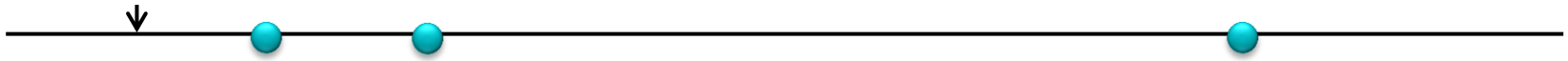
Differential Influence of Therapies on First and Recurrent Events: Three Possible Scenarios

Scenario 1



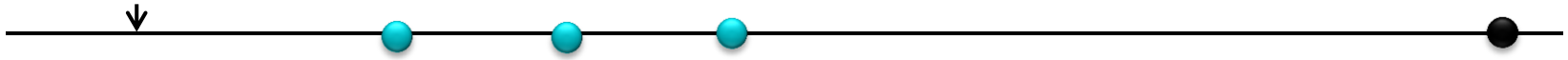
- **Treatment prevents/delays all events (1st or subsequent) equally**
  - Recurrent event analysis should tell SAME STORY as 1st event analysis, but with greater power and precision

Scenario 2



- **Treatment prevents/delays later events more than first events**
  - Drug may require time to become effective – build up – early events may be less likely to be affected by drug than later events

Scenario 3



- **Treatment prevents/delays first events more than subsequent events**
  - Patient may develop tolerance or resistance to therapy (e.g., chemotherapy for cancer, antibiotics for infections, ?ACE inhibitors)

Example: CHARM-Preserved

- Component arm of CHARM, EF ≥ 40%
- Compared candesartan vs. placebo in 3021 patients
- Primary endpoint composite of HF hospitalisation or CV death

HF Hospitalisations	Candesartan (N=1513)	Placebo (N=1508)
≥ 1 Admission	229	278
≥ 2 Admissions	94	114
All Admissions	390	547
Unused Admissions	161	269

What does Recurrent Event Analysis Buy Us?

- Recurrent event analysis can lead to a gain in statistical power → smaller sample size . . .
  - If (and only if) a therapy continues to affect subsequent events (not just first)
  - Meaningful **and** modifiable

- Practical:
  - Increased statistical power →
  - Demonstrate benefit of therapies that would be too expensive to test otherwise

Recurrent Event Methods

Courtesy of Hajime Uno, PhD

Method	Details	How to possibly include death*	Notes (relative attributions etc.)
Poisson regression	The most popular count model. This models incidence rates (total number of events/total follow-up time).	as one of recurrent events	Assume that all events are independent, that event time is identically distributed within group, that mean parameter is identical for all members within group, and is constant over time.
Negative binomial regression	An extension of Poisson regression model -- the mean parameter (or Poisson intensity parameter) is NOT identical for all subject within group, but itself is distributed with gamma distribution	as one of recurrent events	More flexible than Poisson in the sense that each individual can have their own mean parameter (while it is constant over time), which allows us to take into account heterogeneity in a population.
Andersen-Gill (1982)	Extension of Cox PH model. This models gap times (i.e., time from the previous event)	as one of recurrent events	More flexible than Poisson in another direction --- the intensity does NOT have to be constant over time (while the intensity function is identical for all members within group). Robust variance can handle heterogeneity
Wei-Lin-Weissfeld (WLW, 1989)	Each event has its own stratum and each patient appears in all strata. For example, when we focus on up to the 3rd event, we consider 3 strata. Analysis is done by strata and combined later. Subject is at risk for 2nd event regardless of the occurrence of the 1st event (marginal risk set model)	as one of recurrent events, or another type of event (i.e., consider one stratum only for death).	Intensity of the 1st, 2nd and 3rd events … can be different.
Prentice-Williams-Peterson (PWP, 1981)	Similar to WLW, each event has its own stratum, but PWP considers gap times (time from the previous event) instead of time from entry to each event. Also, unlike WLW, each patient does not necessary appears in all strata. For example, subject is at risk for 2nd event only after the occurrence of the 1st event (i.e., conditional risk set model).	as one of recurrent events	PWP can only deal with the recurrent event times. Unlike WLW, it cannot include death as another type of event.
Lin-Wei-Yang-Ying (LWYY, 2000)	Modeling the mean count function of recurrent events over time instead of intensity function.	as one of recurrent events	LWYY and AG give the same treatment effect estimate, while variance estimates are different.
Ghosh and Lin (2001)	Similar to LWYY, this models the mean frequency function, but death is included in competing risk.	Competing risk	In the methods listed above, patients who died are excluded from the risk set after death. But, this one does not exclude those patients from the risk set. The mean frequency function derived in this method is a recurrent-event analog of the cumulative incidence function.
Joint modeling	Consider models for non-fatal events and model for death, those of which share some of parameters (i.e., shared frailty models)	Via modeling	This allows us to adjust informative drop-out in the analysis of non-fatal recurrent events data. Both models have to be correctly specified for valid inference.

What's been done

- CHAMPION

Wireless pulmonary artery haemodynamic monitoring in chronic heart failure: a randomised controlled trial

*William T Abraham, Philip B Adamson, Robert C Bourge, Mark F Aaron, Maria Rosa Costanzo, Lynne W Stevenson, Warren Strickland, Suresh Neelagaru, Nirav Raval, Steven Krueger, Stanislav Weiner, David Shavelle, Bradley Jeffries, Jay S Yadav, for the CHAMPION Trial Study Group**

www.thelancet.com Vol 377 February 19, 2011

Negative Binomial Model

	Not enrolled (n=25)	Treatment group (n=270)	Control group (n=280)	All patients (n=575)	Risk (95% CI)	p value
Primary efficacy endpoints*						
Heart-failure-related hospitalisations up to 6 months (number; events per patient per 6 months)	NA	84 (0·32)	120 (0·44)	NA	0·72† (0·60–0·85)	0·0002
Secondary efficacy endpoints						
Change from baseline in pulmonary artery mean pressure at 6 months (mm Hg×days; mean area under the curve)	NA	−156	33	NA	NA	0·008
Patients admitted to hospital for heart failure at 6 months	NA	55 (20%)	80 (29%)	NA	0·71‖ (0·53–0·96)	0·03

Time to first HF hosp

Personal Experience

- ~~Poisson Model~~ → Negative Binomial model

TOPCAT

on 630 participants with primary outcomes would be needed to achieve 85% power. The target enrollment was 3515 patients.[18]

All randomly assigned participants were included in all analyses according to the intention-to-treat principle. For prespecified comparisons of multiple hospitalizations, the planned Poisson regression model was replaced with a negative binomial model to allow for correlated events. All other analyses were prespecified unless stated otherwise. Participants were followed for clinical and

The NEW ENGLAND JOURNAL of MEDICINE

ESTABLISHED IN 1812 APRIL 10, 2014 VOL. 370 NO. 15

Spironolactone for Heart Failure with Preserved Ejection Fraction

Bertram Pitt, M.D., Marc A. Pfeffer, M.D., Ph.D., Susan F. Assmann, Ph.D., Robin Boineau, M.D., Inder S. Anand, M.D., Brian Claggett, Ph.D., Nadine Clausell, M.D., Ph.D., Akshay S. Desai, M.D., M.P.H., Rafael Diaz, M.D., Jerome L. Fleg, M.D., Ivan Gordeev, M.D., Ph.D., Brian Harty, M.A., John F. Heitner, M.D., Christopher T. Kenwood, M.S., Eldrin F. Lewis, M.D., M.P.H., Eileen O'Meara, M.D., Jeffrey L. Probstfield, M.D., Tamaz Shaburishvili, M.D., Ph.D., Sanjiv J. Shah, M.D., Scott D. Solomon, M.D., Nancy K. Sweitzer, M.D., Ph.D., Song Yang, Ph.D., and Sonja M. McKinlay, Ph.D., for the TOPCAT Investigators*

European Journal of Heart Failure (2014) **16**, 33–40
doi:10.1002/ejhf.29

Analysing recurrent hospitalizations in heart failure: a review of statistical methodology, with application to CHARM-Preserved

Jennifer K. Rogers[1]*, Stuart J. Pocock[1], John J.V. McMurray[2], Christopher B. Granger[3], Eric L. Michelson[4], Jan Östergren[5], Marc A. Pfeffer[6], Scott D. Solomon[6], Karl Swedberg[7,8], and Salim Yusuf[9]

[1]London School of Hygiene and Tropical Medicine, London, UK; [2]University of Glasgow, Glasgow, UK; [3]Duke University Medical Centre, Durham, USA; [4]AstraZeneca, Wilmington, USA; [5]Karolinska University Hospital, Stockholm, Sweden; [6]Brigham and Women's Hospital, Harvard Medical School, Boston, USA; [7]Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden; [8]National Heart and Lung Institute, Imperial College, London; and [9]Hamilton Health Sciences and McMaster University, Hamilton, Canada

Received 26 March 2013; revised 14 June 2013; accepted 21 June 2013

Table 2 Comparison of treatment effects (candesartan versus placebo) for the composite endpoints and recurrent events, with 95% confidence interval and *P*-value

	Treatment effect	95% CI	*P*-value
Hazard ratios for composite of first heart failure hospitalization or cardiovascular death			
Adjudicated composite	0.89	(0.77–1.03)	0.118
Unadjudicated composite	0.86	(0.74–1.00)	0.050
Rate ratios for recurrent heart failure hospitalizations			
Poisson	0.71	(0.62–0.81)	<0.001
Negative binomial	0.68	(0.54–0.85)	<0.001
Andersen–Gill (robust SE)	0.71	(0.57–0.88)	0.002
Rate ratios for composite of recurrent heart failure hospitalizations and cardiovascular death			
Poisson	0.78	(0.69–0.87)	<0.001
Negative binomial	0.75	(0.62–0.91)	0.003
Andersen–Gill (robust SE)	0.78	(0.65–0.93)	0.006
Joint frailty model			
Rate ratio	0.69	(0.55–0.85)	<0.001
Hazard ratio	0.96	(0.73–1.26)	0.769
Win ratio			
Matched pairs approach	1.18	(1.00–1.40)	0.049
Unmatched approach	1.16	(1.00–1.35)	0.062

CI, confidence interval.

European Journal of Heart Failure (2014) **16**, 33–40
doi:10.1002/ejhf.29

Analysing recurrent hospitalizations in heart failure: a review of statistical methodology, with application to CHARM-Preserved

Jennifer K. Rogers[1]*, Stuart J. Pocock[1], John J.V. McMurray[2], Christopher B. Granger[3], Eric L. Michelson[4], Jan Östergren[5], Marc A. Pfeffer[6], Scott D. Solomon[6], Karl Swedberg[7,8], and Salim Yusuf[9]

[1]London School of Hygiene and Tropical Medicine, London, UK; [2]University of Glasgow, Glasgow, UK; [3]Duke University Medical Centre, Durham, USA; [4]AstraZeneca, Wilmington, USA; [5]Karolinska University Hospital, Stockholm, Sweden; [6]Brigham and Women's Hospital, Harvard Medical School, Boston, USA; [7]Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden; [8]National Heart and Lung Institute, Imperial College, London; and [9]Hamilton Health Sciences and McMaster University, Hamilton, Canada

Received 26 March 2013; revised 14 June 2013; accepted 21 June 2013

The gains in terms of statistical power of analysing all heart failure hospitalizations are clearly demonstrated. We have observed that analysing all the heart failure hospitalizations in CHARM-Preserved gives both more events to analyse and an apparently greater treatment effect than when analysing time to first event, so an increase in power is unsurprising. This increase in treatment effect

Table 2 Comparison of treatment effects (candesartan versus placebo) for the composite endpoints and recurrent events, with 95% confidence interval and *P*-value

	Treatment effect	95% CI	*P*-value
Hazard ratios for composite of first heart failure hospitalization or cardiovascular death			
Adjudicated composite	0.89	(0.77–1.03)	0.118
Unadjudicated composite	0.86	(0.74–1.00)	0.050
Rate ratios for recurrent heart failure hospitalizations			
Poisson	0.71	(0.62–0.81)	<0.001
Negative binomial	0.68	(0.54–0.85)	<0.001
Andersen–Gill (robust SE)	0.71	(0.57–0.88)	0.002
Rate ratios for composite of recurrent heart failure hospitalizations and cardiovascular death			
Poisson	0.78	(0.69–0.87)	<0.001
Negative binomial	0.75	(0.62–0.91)	0.003
Andersen–Gill (robust SE)	0.78	(0.65–0.93)	0.006
Joint frailty model			
Rate ratio	0.69	(0.55–0.85)	<0.001
Hazard ratio	0.96	(0.73–1.26)	0.769
Win ratio			
Matched pairs approach	1.18	(1.00–1.40)	0.049
Unmatched approach	1.16	(1.00–1.35)	0.062

CI, confidence interval.

European Journal of Heart Failure (2014) **16**, 33–40
doi:10.1002/ejhf.29

Analysing recurrent hospitalizations in heart failure: a review of statistical methodology, with application to CHARM-Preserved

Jennifer K. Rogers[1]*, Stuart J. Pocock[1], John J.V. McMurray[2], Christopher B. Granger[3], Eric L. Michelson[4], Jan Östergren[5], Marc A. Pfeffer[6], Scott D. Solomon[6], Karl Swedberg[7,8], and Salim Yusuf[9]

[1]London School of Hygiene and Tropical Medicine, London, UK; [2]University of Glasgow, Glasgow, UK; [3]Duke University Medical Centre, Durham, USA; [4]AstraZeneca, Wilmington, USA; [5]Karolinska University Hospital, Stockholm, Sweden; [6]Brigham and Women's Hospital, Harvard Medical School, Boston, USA; [7]Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden; [8]National Heart and Lung Institute, Imperial College, London; and [9]Hamilton Health Sciences and McMaster University, Hamilton, Canada

Received 26 March 2013; revised 14 June 2013; accepted 21 June 2013

The gains in terms of statistical power of analysing all heart failure hospitalizations are clearly demonstrated. We have observed that analysing all the heart failure hospitalizations in CHARM-Preserved gives both more events to analyse and an apparently greater treatment effect than when analysing time to first event, so an increase in power is unsurprising. This increase in treatment effect

We advocate the use of the joint frailty model, as this method allows estimation of a treatment effect for recurrent events, whilst accounting for death as informative censoring. Where death rates are low, the negative binomial distribution could be used, as this is a simpler method of analysis. Our perspective is that the choice of any one specific primary analysis (whether that be negative binomial, joint frailty model, or whatever) should bear in mind a balance between clarity for non-statisticians and statistical robustness. In this latter regard, it is important that trialists undertake various sensitivity analyses to demonstrate that any apparent treatment benefit is not dependent on the chosen statistical technique.

Table 2 Comparison of treatment effects (candesartan versus placebo) for the composite endpoints and recurrent events, with 95% confidence interval and *P*-value

	Treatment effect	95% CI	*P*-value
Hazard ratios for composite of first heart failure hospitalization or cardiovascular death			
Adjudicated composite	0.89	(0.77–1.03)	0.118
Unadjudicated composite	0.86	(0.74–1.00)	0.050
Rate ratios for recurrent heart failure hospitalizations			
Poisson	0.71	(0.62–0.81)	<0.001
Negative binomial	0.68	(0.54–0.85)	<0.001
Andersen–Gill (robust SE)	0.71	(0.57–0.88)	0.002
Rate ratios for composite of recurrent heart failure hospitalizations and cardiovascular death			
Poisson	0.78	(0.69–0.87)	<0.001
Negative binomial	0.75	(0.62–0.91)	0.003
Andersen–Gill (robust SE)	0.78	(0.65–0.93)	0.006
Joint frailty model			
Rate ratio	0.69	(0.55–0.85)	<0.001
Hazard ratio	0.96	(0.73–1.26)	0.769
Win ratio			
Matched pairs approach	1.18	(1.00–1.40)	0.049
Unmatched approach	1.16	(1.00–1.35)	0.062

CI, confidence interval.

Method:
Lin-Wei-Yang-Ying

Enrollment:	4822
Actual Study Start Date:	July 18, 2014
Estimated Study Completion Date:	March 15, 2019
Estimated Primary Completion Date:	March 15, 2019 (Final data collection date for primary outcome measure)

Design	▪Double-blind period: Randomized to LCZ696 200 mg bid vs. valsartan 160 mg bid ▪ 2 years 9 months enrollment; estimated 2 years follow-up
Primary Endpoint	• Composite endpoint of CV death and total (first and recurrent) HF hospitalization
Secondary Endpoints	• Composite endpoint of CV death, total HF hospitalization, total stroke, and total MI • NYHA classification at 8 months • Time to new onset AF in patients with no history of AF and with sinus rhythm on ECG at V1 • All-cause mortality
Current major inclusion criteria	▪ ≥55 years of age, male or female, and LVEF ≥ 45% ▪ Current symptomatic HF (NYHA Class II-IV) ▪ Symptoms of HF ≥30 days prior to Visit 1 ▪ Treatment with diuretic(s) within 30 days prior to V1 ▪ Structural heart disease (LAE or LVH) ▪ HF hospitalization within 9 months OR Visit 1 elevated NT-proBNP (>300 pg/mL for patients in sinus rhythm or >900 pg/mL for patients with AF at Visit 1)
Sample size	• 4300 subjects
Leadership	•Chairs: S.Solomon, J. McMurray •Executive Cmt: I.Anand, A. Maggioni, F. Zannad •Steering cmt: M.Packer, M.Zile, B. Pieske, M.Redfield, J.Rouleau, M.Pfeffer, D. Van Veldhuisen, F. Martinez

Recurrent Events: Controversies

- Recurrent events are not statistically independent.
  - Events tend to ”cluster” within patients
  - What modelling assumptions are required to appropriately handle this?

- How much can/should 1 patient influence the results of a trial?
  - Should 17 events in 1 patient “count” as much as 1 event in 17 patients?
  - Can/should we stop counting events at some pointi?

- How do we handle death?
  - 2 Hosps + 1 Death = 3 Hosps ?
  - Estimate death separately?
  - Use weighting or ranking to make death more important than non-fatal events?

Conventional Approaches: Controversies

- “Time to death” analyses ignore morbidity (non-fatal events)
- “Time to first event” composite analyses
  - Mortality is no worse than any other first event
  - Mortality *after* a non-fatal event is completely ignored
- Simplistically: “Health” > Morbidity > Mortality