2024 Summer Institute In Statistics for Clinical & Epidemiological Research

Module 3:

Design, Conduct, and Analysis of Randomized Clinical Trials with Time to Event Primary Endpoints

Lecture 25:

Protopathic Bias

Scott S. Emerson, M.D., Ph.D.
Professor Emeritus of Biostatistics
University of Washington

1

Abstract

The various phases of randomized clinical trials (RCTs) are the mainstays of the process of drug discovery. A major challenge of that process is how to best focus on the eventual use of the therapy in clinical practice, all the while ensuring the scientific and statistical rigor inherent in a randomized intervention. Scientific rigor demands analysis of the clinical trial results according to randomized treatment assignment, and missing data greatly detracts from the rigor needed to address the estimand of greatest interest in a particular RCT. However, during the course of the clinical trial, some participants discontinue their therapy. Clinical investigators often express the desire to restrict attention to the causal effects of a treatment in fully compliant patients, treating the observations on noncompliant patients as “missing”. There has been much statistical literature devoted to methods to obtain causal inference in such a setting. In this talk, I discuss the difficult issues that must be addressed in such analyses, using published clinical trials in type 2 diabetes and chronic obstructive pulmonary disease as examples.

2

2

Perspective

- Since 1985: Target of my statistical methodologic research

  Distribution-free adaptive causal inference (frequentist and Bayesian) in the presence of missing data

3

"Modern Statistics"

4

"Modern Statistics"

"You know, sex was not invented in the 1960's"

- Virginia Emerson (my mother)
(often said to her 7 children born between 1950 and 1959)

5

Disclaimers

- Since 1985: Target of my statistical methodologic research

  Distribution-free adaptive causal inference (frequentist and Bayesian) in the presence of missing data

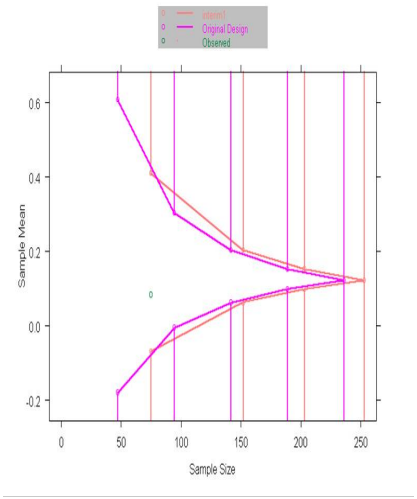
- What did these buzzwords mean for me in 1985 (and beyond)?
  - **Causal inference** from randomized clinical trials
    - Distribution-free based on central limit theorem
    - Frequentist inference that provides credible Bayesian inference
  - **Adaptive clinical tr**ials using
    - Modified specific aims across phases of clinical trials
    - Group sequential designs within individual trials
  - **Missing data** from right censored measurement of time to event
    - Especially Kaplan-Meier, proportional hazards regression

6

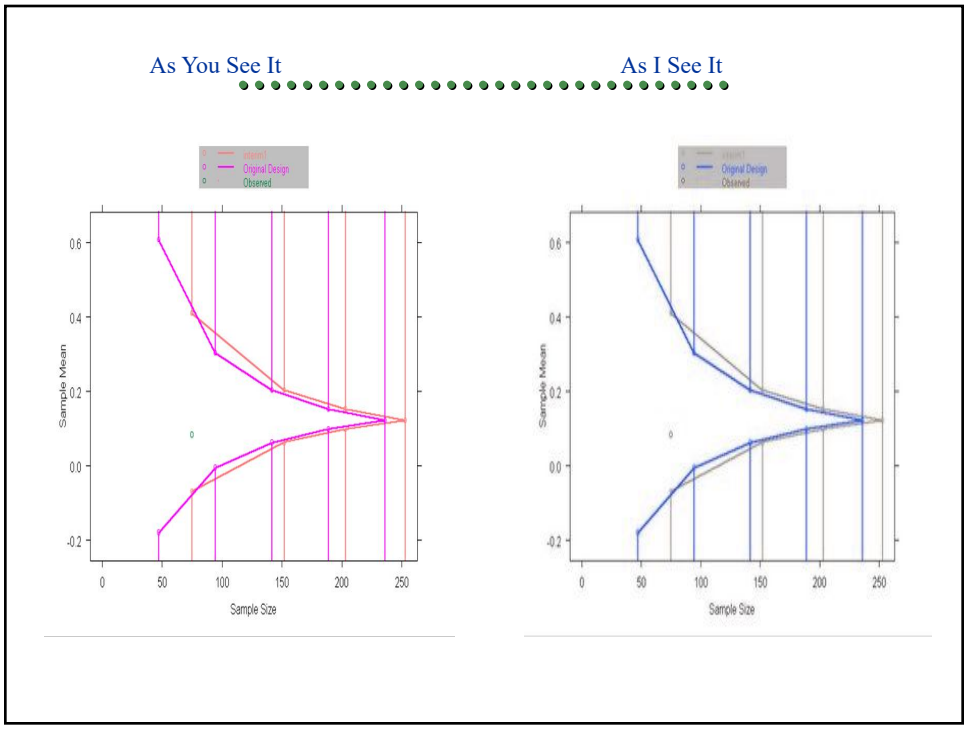
Fair Warning

7

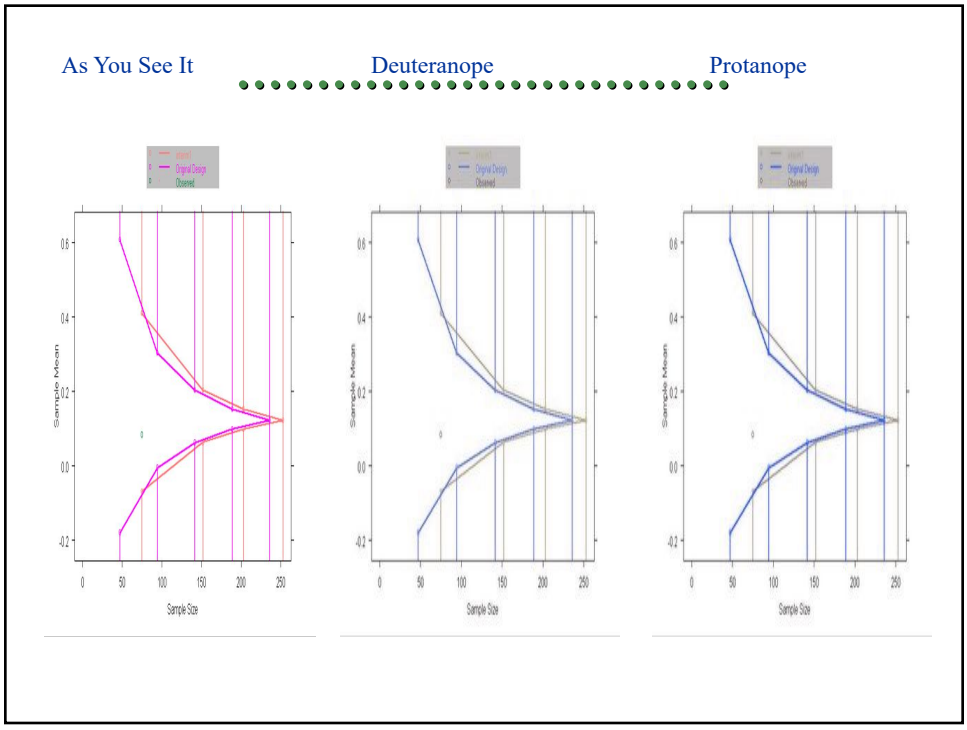
As You See It



8


As You See It
As I See It
Original Design
Observed
Sample Mean
Sample Size

9


As You See It
Deuteranope
Protanope
Original Design
Observed
Sample Mean
Sample Size

10

Scientific Efficacy / Safety Estimands

- What is impact in a population assigned to treatment protocol?
  - "Intent to treat" (ITT) estimand
  - Rigorous causal estimand
  - Full data on all randomized subjects
  - Analysis by randomized group
- What is impact among patients who follow protocol?
  - "Per protocol" (PP) estimand
  - No matter what: An interesting basic science question
  - Clinically may be used to explore mechanism of action
  - Clinically may be desired to describe prognosis
  - However, not scientifically rigorous
    - Conditions on a post-randomization variable

11

11

Per Protocol Estimands: Use of Data

- Patients who don't follow protocol may be irrelevant to goal
  - Patients who do not follow directions
  - Patients who have intolerable adverse reactions
    - Perhaps "intolerable" only because uncertain of efficacy, or
    - Perhaps leading to serious consequences with continued therapy
  - Patients with real or perceived lack of efficacy
    - Early clinical course is discouraging, or
    - Definitive progression to serious condition prior to primary endpoint
  - Development of contraindication to treatment (e.g., pregnancy)
  - Patients with early evidence of cure
  - Patients with competing risk that prevents measurement
- Ignoring vs imputing missing data?

12

12

Analysis Using Hazards

- Hazard
  - Instantaneous risk of failure
  - "Given alive now, what is risk of failure in next time interval?"
- Estimable with censored data providing noninformative censoring
  - Only need a representative population of individuals at risk at each time point
  - Can then combine hazard estimates to obtain survivor distribution
- Popular analysis methods to compare distributions
  - Exponential regression – constant hazard is inverse of mean
    - Hazard estimate: Number of events divided by person- years
  - Weibull regression – linear log hazard in log time
  - Proportional hazards regression – includes Weibull, exponential

13

Protopathic Bias

"Why is the alphabet in the order it is?
Is it because of that song?"

-Steven Wright

14

EXSCEL: CVOT in T2DM (*NEJM*, 2017)

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

Effects of Once-Weekly Exenatide on Cardiovascular Outcomes in Type 2 Diabetes

Rury R. Holman, F.Med.Sci., M. Angelyn Bethel, M.D., Robert J. Mentz, M.D., Vivian P. Thompson, M.P.H., Yuliya Lokhnygina, Ph.D., John B. Buse, M.D., Ph.D., Juliana C. Chan, M.D., Jasmine Choi, M.S., Stephanie M. Gustavson, Ph.D., Nayyar Iqbal, M.D., Aldo P. Maggioni, M.D., Steven P. Marso, M.D., Peter Öhman, M.D., Ph.D., Neha J. Pagidipati, M.D., M.P.H., Neil Poulter, F.Med.Sci., Ambady Ramachandran, M.D., Bernard Zinman, M.D., and Adrian F. Hernandez, M.D., M.H.S., for the EXSCEL Study Group*

ABSTRACT

BACKGROUND
The cardiovascular effects of adding once-weekly treatment with exenatide to usual care in patients with type 2 diabetes are unknown.

15

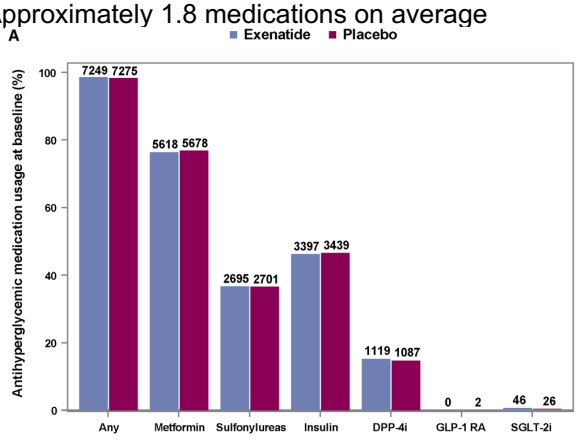
EXSCEL: A Pragmatic Trial

- N= 9,600 ➔14,000 patients with T2DM to observe 1,360 MACE
  - Stable diabetes regimen of glucose lowering medications (GLM)
    - 0-3 oral GLM **or** insulin plus 0-2 oral GLM
  - Approximately 70% with prior cardiovascular disease (CVD)
- Randomize 1:1 to weekly subcutaneous exenatide vs placebo
  - A glucagon-like peptide 1 receptor agonist (GLP-1 RA)
- Endpoint hierarchy
  - Noninferiority for major adverse cardiovascular events (MACE)
    - 95% CI excludes hazard ratio of 1.3
  - Superiority for MACE: 95% CI excludes HR of 1.0
  - Superiority for all cause mortality
  - Individual components of MACE, hospitalization for ACS,

16

EXSCEL: Baseline

- N= 14,752: 73% with prior CVD
- GLM use at baseline
  - Approximately 1.8 medications on average



17

EXSCEL: Disposition after ***1,744*** Events

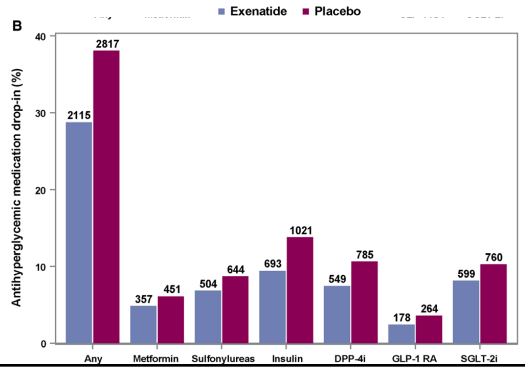
- 3.2 years *median* time of observation (I wish I had the mean)
- Availability of data of exenatide/placebo arms

– No study drug	18	/	18
– Lost to follow up	39	/	33
– Withdrew consent	223	/	270
– Percent of planned person-years	95.1%	/	94.5%
– No primary endpoint			
• LTFU	38	/	29
• WD	217	/	257
– Vital status unknown			
• LTFU	39	/	33
• WD	44	/	55

18

EXSCEL: Post Randomization GLM

- Premature d/c study drug 43.0% / 45.2%
- Years on study drug *(median!)* 2.4 / 2.3
- Open label GLM therapies 28.1% / 38.8%
- Mean GLMs exposure (approx.) 3.2 / 2.4



19

EXSCEL: Results

- Primary endpoint:
  - MACE exen : plc HR 0.91 (95% CI 0.83, 1.00; p= 0.06)
  - Noninferiority met
  - Superiority not met
- Secondary endpoint:
  - All cause mortality: HR 0.86 (95% CI 0.77, 0.97; p= NA)
- Supportive endpoints- Slight benefit of exenatide on some, but not all, cardiovascular risk factors
  - Weight loss -1.27 kg
  - Glycosylated hemoglobin -0.53 %
  - Systolic blood pressure -1.57 mmHg
  - Low density lipoprotein -1.55 mg/dL
  - Heart rate +2.51 bpm

20

EXSCEL: Author Conclusions

- Noninferiority for MACE established (but not superiority)
  - RCTs for other GLP-1 RA:
    - Liraglutide HR 0.87 (95% CI 0.78, 0.97)
    - Semaglutide HR 0.74 (95% CI 0.58, 0.95)
    - Lixisenatide no statistically significant superiority
- Superiority for ACM not established due to hierarchy of endpoints
- Effect on modifiable cardiovascular risk factors was modest
- "The disproportionate use in the placebo group of diabetes therapies known to reduce cardiovascular risk, such as SGLT-2 inhibitors and GLP-1 receptor agonists, may have preferentially resulted in lower event rates in the placebo group."

21

Exploratory Re-analyses (*Circ* 2020)

Circulation

ORIGINAL RESEARCH ARTICLE

Exploring the Possible Impact of Unbalanced Open-Label Drop-In of Glucose-Lowering Medications on EXSCEL Outcomes

BACKGROUND: EXSCEL (Exenatide Study of Cardiovascular Event Lowering) assessed the impact of once-weekly exenatide 2 mg versus placebo in patients with type 2 diabetes mellitus, while aiming for glycemic equipoise. Consequently, greater drop-in of open-label glucose-lowering medications occurred in the placebo group. Accordingly, we explored the potential effects of their unbalanced use on major adverse cardiovascular events (MACE), defined as cardiovascular death, nonfatal myocardial infarction or nonfatal stroke, and all-cause mortality (ACM), given that some of these agents are cardioprotective.

M. Angelyn Bethel, MD
Susanna R. Stevens, MS
John B. Buse, MD, PhD
Jasmine Choi, MS
Stephanie M. Gustavson, PhD
Nayyar Iqbal, MD
Yuliya Lokhnygina, PhD
Robert J. Mentz, MD

22

Exploratory: Motivation

- (Sponsor unhappy at failure to demonstrate superiority for MACE
- Pragmatic trial allowed *ad lib* modifications to GLM regimens
  - Placebo arm had more such modifications
  - Some GLP-1 RA and SGLT-2i known to provide MACE protection
- "With these cardioprotective glucose-lowering agents being used increasingly in routine clinical practice, there is a greater likelihood that their open-label drop-in in placebo groups could impact clinical trial CV event rates and potentially bias study outcomes."
- "used several statistical and modeling methods to evaluate whether their unbalanced use during the trial might have affected the primary outcome (MACE) or ACM time-to-event analyses."

23

Bias? What Bias?

- NRC monograph on the prevention and treatment of missing in data in RCT emphasized the need to identify the "estimand"
  - What are we trying to estimate?
- Choice of estimand reflects clinical, scientific, regulatory, and statistical issues
- What is the relevant estimand here?
  - EXSCEL can provide unbiased estimate of the effect of immediate addition of exenatide to a stable GLM regimen vs later changes to the regimen based on clinical judgement of progression of T2DM and CVD.
  - The authors did not specify any other estimand explicitly.

24

Possible Implicit Estimands

- Focus on changes in GLM regimen suggests EXSCEL investigators wanted to estimate exenatide effect in a population where GLM regimen never changes
- Such a goal is compatible with type of statistical methods used
  - Imputation of what might have happened based on "similar" patients whose GLM regimen did not change
- This is not enough to go on, however
  - Who is "similar"?
- Need to consider why GLM regimen might have changed
  - Reasons a given patient might change GLM might differ across treatment arms

25

Reasons for Changing GLM Regimen

- Unwilling to follow protocol
  - Poorly compliant patient (in general or due to worsening health)
  - Aversion to therapy (e.g., injections, secular trends)
  - Mild adverse events attributed to treatment by patient
  - Perception of lack of efficacy or, alternatively, cure
- Medically inadvisable to continue on protocol
  - Severe adverse reactions
  - Development of contraindications to treatment
  - Need to progress to other therapies
- The above likely varies by treatment within individual patients
  - The patients changing GLM on experimental arm may not be similar to the patients changing on control arm

26

Exploratory: Statistical Methods

- "Right Censoring Analyses"
  - Censor subjects at time of additions to their GLM regimen
  - Proportional hazards analysis using artificially censored data
    - (a time dependent covariate analysis would have been better)
- "Inverse Probability of Treatment Weighting Analyses"
  - Proportional hazards analysis to estimate probability of additions to the GLM regimen
    - Censor observations at times of MACE events
    - Build predictive models for each treatment arm separately
  - Weighted proportional hazards analysis
    - Censor subjects at time of additions to their GLM regimen
    - Weight observations according to the inverse of the time dependent probability of receiving additional GLM

27

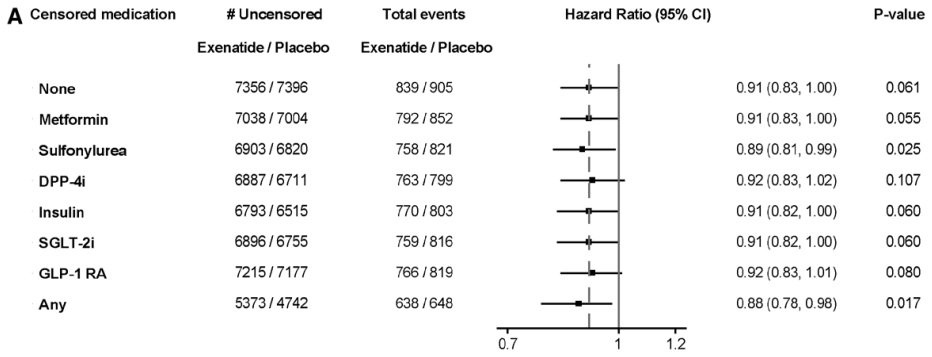
Exploratory: What is the Estimand?

- Imputations presume it is relevant to replace an observation after added GLM with experience of those who did not add GLMs
- Seems to suggest investigators are imagining a world in which we would force everyone to keep taking the drug, thus ignoring
  - Patient preference
  - Seeming lack of efficacy on surrogate measures
  - Severe AEs that lead to changes in GLM regimen
  - Progression of T2DM
- Further presumes outcomes would be similar to outcomes in patients who did not need/want to add a new GLM
- NOT imagining restricting to those who would not need to add to GLM regimen (otherwise would try to exclude from RCT analysis)

28

MACE: Censoring at Additional GLM

- Proportional hazards analysis using artificially censored data
  - ”HR was not altered meaningfully … for any of the drop-in”
  - “Small changes in event numbers did lead to nominally significant P values when censoring for sulfonylureas or any drop-in”



29

MACE: IPTW Analyses Weights

- PH prediction models for additions to GLM regimen computed for each arm separately
  - 23 baseline (pre-randomization) variables
  - Time varying HbA_{1c}, eGFR (renal function)
  - Weights truncated at 99th percentile

- Covariates contributing most to prediction models (P < .01)
  - Plc: HbA_{1c}, Region, Diabetes duration, Race, CABG
  - Ex: HbA_{1c}, Region, Chronic liver disease, Prior CV event, Diabetes duration

30

MACE: IPTW Analyses Results

- Weighted PH with artificial censoring for any added GLM
- Exenatide : placebo HR 0.85 (95% CI 0.76, 0.96; P= 0.008)
- Discussion in *Circulation* article:
  - "Effect sizes were increased … achieving statistical significance"
  - "analyses suggest that the CV effects of some agents might have had a discernible impact on the MACE findings in EXSCEL"
  - "For the relatively few EXSCEL participants experiencing drop-in of GLM known to be cardioprotective, the likely effects appear to be too small to have impacted on the trial findings"
  - (And quite a few comments about right censoring and IPTW methods that I do not believe are correct in this setting.)

31

Limitations of the Analyses

- Analyses not pre-specified in Statistical Analysis Plan
- GLM regimens misclassified
  - Did not consider discontinuation of treatments
  - Did not consider GLM classes jointly
- Conditioning on a post randomization variable
  - Observational data subject to confounding
    - Did not present any data to assess possible confounding
  - Changes in GLM regimen may be mediators or markers of treatment outcomes
- Possible violation of assumption of noninformative censoring
  - Authors did not present statistics for events following changes

32

Analyses of Censored Time to Event

- Kaplan-Meier and proportional hazards analyses presume noninformative censoring
  - Censored subject looks like a random selection from the risk set
- A missing at random (MAR) model in which the missing event time can be imputed from the uncensored subjects still at risk
- This assumption is violated by conditioning on protopathic signs or symptoms
  - "Protopathic" – vague signs or symptoms caused by an impending event that has not yet been recognized
  - Protopathic bias results from censoring subjects who were just about to have an event
  - (Similar to, but not the exact same as, indication bias)

33

MAR vs MNAR

- By definition, there is nothing in your data that can distinguish missing at random (MAR) mechanisms from missing not at random (MNAR) mechanisms
- We can however explore the data to assess whether there are patterns suggestive of MNAR mechanisms
- In this example, we can consider how the event rate changes after modifications to the GLM regimens

34

"Back of the Envelope" Analysis

- Can perform "person-years" analysis using descriptive statistics, with some caveats
  - Equivalent to exponential regression
- Person-years of exposure provided as median rather than mean
- Person-years of exposure did not distinguish between
  - Description of time until administrative censoring, or
  - Description of time until earlier of event or administrative censoring
- Person-years of exposure after GLM changes not provided separately for each treatment arm
- Only two significant figures of precision

35

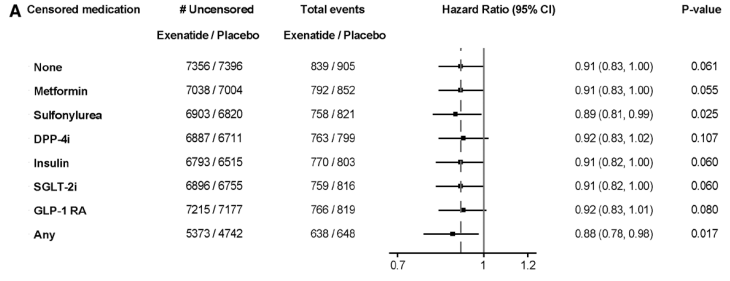
"Back of the Envelope": Source of Data

- From NEJM
  - Median(!) person years of exposure in total population combined
- From *Circulation*
  - Median(!) person years of exposure post drop-in combined arms

Bethel et al — Impact of Drop-in Medications in EXSCEL

Table. Baseline Patient Characteristics and Exposure Time According to Type of Diabetes Mellitus Medication Initiated Postrandomization

	Metformin (N=808)	Sulfonylurea (N=1148)	Insulin (N=1714)	DPP-4i (N=1334)	GLP-1 RA (N=442)	SGLT-2i (N=1359)
Exposure time, y	2.0 (0.9–3.2)	2.1 (1.0–3.5)	2.0 (0.9–3.3)	1.8 (0.8–3.0)	1.5 (0.6–3.0)	1.0 (0.4–1.6)
Age at randomization, y	62 (56–69)	60 (54–67)	61 (54–67)	62 (56–68)	60 (53–66)	60 (53–66)



36

"Back of the Envelope": Exponential Data Models

- For covariate vector $\vec{X}$, we model the hazard (mean) according to

$$log(\lambda) = \vec{X}\ \vec{\beta}$$

- Estimating equations for one sample

$$L(\vec{\Phi}, \vec{T}, \vec{D}) \propto \prod_{i=1}^{n} \lambda^{\{D_i\}} e^{-\lambda T_i}$$

$$\frac{\partial}{\partial \lambda} L(\vec{\Phi}, \vec{T}, \vec{D}) = \frac{\sum D_i}{\lambda} - \sum T_i$$

$$\hat{\lambda} = \frac{\sum D_i}{\sum T_i} \ \dot{\sim}\ N\left(\lambda, \frac{\sum D_i}{(\sum T_i)^2}\right)$$

37

37

"Back of the Envelope": Calculations

- "Back of the envelope" calculations using Excel

	NEJM Combined N	NEJM Baseline Exenatide N	NEJM Baseline Placebo N	Baseline Prim Prev N	Baseline Sec Prev N	Circulation New Exenatide N	Circulation New Placebo N	Circulation Exposure Avg PY
								3.249
ITT	14,752	7,356	7,396	3,969	10,781			
Metformin	11,296	5,618	5,678	3,375	7,914	357	451	2.000
Sulfonylureas	5,396	2,695	2,701	1,708	3,694	504	644	2.100
DPP-4i	2,206	1,119	1,087	692	1,510	549	785	1.800
Insulin	6,836	3,397	3,439	1,257	5,562	693	1,021	2.000
SGLT-2i	72	46	26	24	53	599	760	1.000
GLP-1 RA	2	0	2	0	0	178	264	1.500
Any	14,524	7,249	7,275					2.100

Exenatide									
Full Data Used					Artificial Cens				
N	Avg PY	Ev	% Ev	Rate	N	Avg PY	Ev	% Ev	Rate
7,356	3.116	839	0.114	0.037					
7,038	3.030	792	0.108	0.036	318	2.000	47	0.148	0.074
6,903	2.987	758	0.103	0.034	453	2.100	81	0.179	0.085
6,887	3.002	763	0.104	0.035	469	1.800	76	0.162	0.090
6,793	2.963	770	0.105	0.035	563	2.000	69	0.123	0.061
6,896	3.054	759	0.103	0.034	460	1.000	80	0.174	0.174
7,215	3.088	766	0.104	0.034	141	1.500	73	0.518	0.345
5,373	2.550	638	0.087	0.034	1,983	2.100	201	0.101	0.048

38

"Back of the Envelope" Results

- "Person-years" analysis (exponential regression model) was derived from the data provided in *NEJM* and *Circulation* articles

Ad hoc Censoring Criterion	Exenatide					Placebo					Exenatide : Placebo Rate Ratio	
	Prior* to ad hoc Censoring Times		After** ad hoc Censoring Times		After:Prior Rate Ratio	Prior* to ad hoc Censoring Times		After** ad hoc Censoring Times		After:Prior Rate Ratio		
	Events (N; Avg PY)	Event Rate	Events (N; Avg PY)	Event Rate		Events (N; Avg PY)	Event Rate	Events (N; Avg PY)	Event Rate		Prior to ad hoc Censoring Times	After ad hoc Censoring Times
None (ITT)	839 (7,356; 3.12)	0.037				905 (7,396; 3.04)	0.040				0.91	
Metformin	792 (7,356; 3.03)	0.036	47 (318; 2.00)	0.074	2.1	852 (7,396; 2.94)	0.039	53 (392; 2.00)	0.068	1.7	0.91	1.09
Sulfonylureas	758 (7,356; 2.99)	0.034	81 (453; 2.10)	0.085	2.5	821 (7,396; 2.88)	0.039	84 (576; 2.10)	0.069	1.8	0.90	1.23
DPP-4i	763 (7,356; 3.00)	0.035	76 (469; 1.80)	0.090	2.6	799 (7,396; 2.88)	0.038	106 (685; 1.80)	0.086	2.3	0.92	1.05
Insulin	770 (7,356; 2.96)	0.035	69 (563; 2.00)	0.061	1.7	803 (7,396; 2.81)	0.039	102 (881; 2.00)	0.058	1.5	0.91	1.06
SGLT-2i	759 (7,356; 3.05)	0.034	80 (460; 1.00)	0.174	5.1	816 (7,396; 2.96)	0.037	89 (641; 1.00)	0.139	3.7	0.91	1.25
GLP-1 RA	766 (7,356; 3.09)	0.034	73 (141; 1.50)	0.345	10.2	819 (7,396; 3.00)	0.037	86 (219; 1.50)	0.262	7.1	0.91	1.32
Any GLM	638 (7,356; 2.55)	0.034	201 (1,983; 2.10)	0.048	1.4	648 (7,396; 2.29)	0.038	257 (2,654; 2.10)	0.046	1.2	0.89	1.05

39

"Back of the Envelope" Results

- The person-years of observation for the ITT analysis were back calculated from published event rates.

Ad hoc Censoring Criterion	Exenatide					Placebo					Exenatide : Placebo Rate Ratio	
	Prior* to ad hoc Censoring Times		After** ad hoc Censoring Times		After:Prior Rate Ratio	Prior* to ad hoc Censoring Times		After** ad hoc Censoring Times		After:Prior Rate Ratio		
	Events (N; Avg PY)	Event Rate	Events (N; Avg PY)	Event Rate		Events (N; Avg PY)	Event Rate	Events (N; Avg PY)	Event Rate		Prior to ad hoc Censoring Times	After ad hoc Censoring Times
None (ITT)	839 (7,356; 3.12)	0.037				905 (7,396; 3.04)	0.040				0.91	
Metformin	792 (7,356; 3.03)	0.036	47 (318; 2.00)	0.074	2.1	852 (7,396; 2.94)	0.039	53 (392; 2.00)	0.068	1.7	0.91	1.09
Sulfonylureas	758 (7,356; 2.99)	0.034	81 (453; 2.10)	0.085	2.5	821 (7,396; 2.88)	0.039	84 (576; 2.10)	0.069	1.8	0.90	1.23
DPP-4i	763 (7,356; 3.00)	0.035	76 (469; 1.80)	0.090	2.6	799 (7,396; 2.88)	0.038	106 (685; 1.80)	0.086	2.3	0.92	1.05
Insulin	770 (7,356; 2.96)	0.035	69 (563; 2.00)	0.061	1.7	803 (7,396; 2.81)	0.039	102 (881; 2.00)	0.058	1.5	0.91	1.06
SGLT-2i	759 (7,356; 3.05)	0.034	80 (460; 1.00)	0.174	5.1	816 (7,396; 2.96)	0.037	89 (641; 1.00)	0.139	3.7	0.91	1.25
GLP-1 RA	766 (7,356; 3.09)	0.034	73 (141; 1.50)	0.345	10.2	819 (7,396; 3.00)	0.037	86 (219; 1.50)	0.262	7.1	0.91	1.32
Any GLM	638 (7,356; 2.55)	0.034	201 (1,983; 2.10)	0.048	1.4	648 (7,396; 2.29)	0.038	257 (2,654; 2.10)	0.046	1.2	0.89	1.05

40

"Back of the Envelope" Results

- Using sulfonylureas as an example:
  - 81 and 84 events after censoring for the Ex and Plc arms, resp
  - Person-years based on exposure in *Circulation* Table

Ad hoc Censoring Criterion	Exenatide					Placebo					Exenatide : Placebo Rate Ratio	
	Prior* to ad hoc Censoring Times		After** ad hoc Censoring Times		After:Prior Rate Ratio	Prior* to ad hoc Censoring Times		After** ad hoc Censoring Times		After:Prior Rate Ratio	Prior to ad hoc Censoring Times	After ad hoc Censoring Times
	Events (N; Avg PY)	Event Rate	Events (N; Avg PY)	Event Rate		Events (N; Avg PY)	Event Rate	Events (N; Avg PY)	Event Rate			
None (ITT)	839 (7,356; 3.12)	0.037				905 (7,396; 3.04)	0.040				0.91	
Metformin	792 (7,356; 3.03)	0.036	47 (318; 2.00)	0.074	2.1	852 (7,396; 2.94)	0.039	53 (392; 2.00)	0.068	1.7	0.91	1.09
Sulfonylureas	758 (7,356; 2.99)	0.034	81 (453; 2.10)	0.085	2.5	821 (7,396; 2.88)	0.039	84 (576; 2.10)	0.069	1.8	0.90	1.23
DPP-4i	763 (7,356; 3.00)	0.035	76 (469; 1.80)	0.090	2.6	799 (7,396; 2.88)	0.038	106 (685; 1.80)	0.086	2.3	0.92	1.05
Insulin	770 (7,356; 2.96)	0.035	69 (563; 2.00)	0.061	1.7	803 (7,396; 2.81)	0.039	102 (881; 2.00)	0.058	1.5	0.91	1.06
SGLT-2i	759 (7,356; 3.05)	0.034	80 (460; 1.00)	0.174	5.1	816 (7,396; 2.96)	0.037	89 (641; 1.00)	0.139	3.7	0.91	1.25
GLP-1 RA	766 (7,356; 3.09)	0.034	73 (141; 1.50)	0.345	10.2	819 (7,396; 3.00)	0.037	86 (219; 1.50)	0.262	7.1	0.91	1.32
Any GLM	638 (7,356; 2.55)	0.034	201 (1,983; 2.10)	0.048	1.4	648 (7,396; 2.29)	0.038	257 (2,654; 2.10)	0.046	1.2	0.89	1.05

41

"Back of the Envelope" Results

- Hazard ratios estimated from exponential rate model agree with Figure 2 HR within 0.01.
  - Suggests that constant hazard probability model is reasonable

Ad hoc Censoring Criterion	Exenatide					Placebo					Exenatide : Placebo Rate Ratio	
	Prior* to ad hoc Censoring Times		After** ad hoc Censoring Times		After:Prior Rate Ratio	Prior* to ad hoc Censoring Times		After** ad hoc Censoring Times		After:Prior Rate Ratio	Prior to ad hoc Censoring Times	After ad hoc Censoring Times
	Events (N; Avg PY)	Event Rate	Events (N; Avg PY)	Event Rate		Events (N; Avg PY)	Event Rate	Events (N; Avg PY)	Event Rate			
None (ITT)	839 (7,356; 3.12)	0.037				905 (7,396; 3.04)	0.040				0.91	
Metformin	792 (7,356; 3.03)	0.036	47 (318; 2.00)	0.074	2.1	852 (7,396; 2.94)	0.039	53 (392; 2.00)	0.068	1.7	0.91	1.09
Sulfonylureas	758 (7,356; 2.99)	0.034	81 (453; 2.10)	0.085	2.5	821 (7,396; 2.88)	0.039	84 (576; 2.10)	0.069	1.8	0.90	1.23
DPP-4i	763 (7,356; 3.00)	0.035	76 (469; 1.80)	0.090	2.6	799 (7,396; 2.88)	0.038	106 (685; 1.80)	0.086	2.3	0.92	1.05
Insulin	770 (7,356; 2.96)	0.035	69 (563; 2.00)	0.061	1.7	803 (7,396; 2.81)	0.039	102 (881; 2.00)	0.058	1.5	0.91	1.06
SGLT-2i	759 (7,356; 3.05)	0.034	80 (460; 1.00)	0.174	5.1	816 (7,396; 2.96)	0.037	89 (641; 1.00)	0.139	3.7	0.91	1.25
GLP-1 RA	766 (7,356; 3.09)	0.034	73 (141; 1.50)	0.345	10.2	819 (7,396; 3.00)	0.037	86 (219; 1.50)	0.262	7.1	0.91	1.32
Any GLM	638 (7,356; 2.55)	0.034	201 (1,983; 2.10)	0.048	1.4	648 (7,396; 2.29)	0.038	257 (2,654; 2.10)	0.046	1.2	0.89	1.05

42

"Back of the Envelope" Results

- Interestingly, in the placebo arm, the event rate after changes to the GLM regimen is higher for every newly prescribed GLM
  - Certainly not suggestive of a large cardioprotective effect

Ad hoc Censoring Criterion	Exenatide					Placebo					Exenatide : Placebo Rate Ratio	
	Prior* to ad hoc Censoring Times		After** ad hoc Censoring Times		After:Prior Rate Ratio	Prior* to ad hoc Censoring Times		After** ad hoc Censoring Times		After:Prior Rate Ratio		
	Events (N; Avg PY)	Event Rate	Events (N; Avg PY)	Event Rate		Events (N; Avg PY)	Event Rate	Events (N; Avg PY)	Event Rate		Prior to ad hoc Censoring Times	After ad hoc Censoring Times
None (ITT)	839 (7,356; 3.12)	0.037				905 (7,396; 3.04)	0.040				0.91	
Metformin	792 (7,356; 3.03)	0.036	47 (318; 2.00)	0.074	2.1	852 (7,396; 2.94)	0.039	53 (392; 2.00)	0.068	1.7	0.91	1.09
Sulfonylureas	758 (7,356; 2.99)	0.034	81 (453; 2.10)	0.085	2.5	821 (7,396; 2.88)	0.039	84 (576; 2.10)	0.069	1.8	0.90	1.23
DPP-4i	763 (7,356; 3.00)	0.035	76 (469; 1.80)	0.090	2.6	799 (7,396; 2.88)	0.038	106 (685; 1.80)	0.086	2.3	0.92	1.05
Insulin	770 (7,356; 2.96)	0.035	69 (563; 2.00)	0.061	1.7	803 (7,396; 2.81)	0.039	102 (881; 2.00)	0.058	1.5	0.91	1.06
SGLT-2i	759 (7,356; 3.05)	0.034	80 (460; 1.00)	0.174	5.1	816 (7,396; 2.96)	0.037	89 (641; 1.00)	0.139	3.7	0.91	1.25
GLP-1 RA	766 (7,356; 3.09)	0.034	73 (141; 1.50)	0.345	10.2	819 (7,396; 3.00)	0.037	86 (219; 1.50)	0.262	7.1	0.91	1.32
Any GLM	638 (7,356; 2.55)	0.034	201 (1,983; 2.10)	0.048	1.4	648 (7,396; 2.29)	0.038	257 (2,654; 2.10)	0.046	1.2	0.89	1.05

43

"Back of the Envelope" Results

- Similarly, event rates are higher for the exenatide arm after changes to GLM regimen
  - Furthermore, Ex : Plc rate ratio greater than 1 after changes

Ad hoc Censoring Criterion	Exenatide					Placebo					Exenatide : Placebo Rate Ratio	
	Prior* to ad hoc Censoring Times		After** ad hoc Censoring Times		After:Prior Rate Ratio	Prior* to ad hoc Censoring Times		After** ad hoc Censoring Times		After:Prior Rate Ratio		
	Events (N; Avg PY)	Event Rate	Events (N; Avg PY)	Event Rate		Events (N; Avg PY)	Event Rate	Events (N; Avg PY)	Event Rate		Prior to ad hoc Censoring Times	After ad hoc Censoring Times
None (ITT)	839 (7,356; 3.12)	0.037				905 (7,396; 3.04)	0.040				0.91	
Metformin	792 (7,356; 3.03)	0.036	47 (318; 2.00)	0.074	2.1	852 (7,396; 2.94)	0.039	53 (392; 2.00)	0.068	1.7	0.91	1.09
Sulfonylureas	758 (7,356; 2.99)	0.034	81 (453; 2.10)	0.085	2.5	821 (7,396; 2.88)	0.039	84 (576; 2.10)	0.069	1.8	0.90	1.23
DPP-4i	763 (7,356; 3.00)	0.035	76 (469; 1.80)	0.090	2.6	799 (7,396; 2.88)	0.038	106 (685; 1.80)	0.086	2.3	0.92	1.05
Insulin	770 (7,356; 2.96)	0.035	69 (563; 2.00)	0.061	1.7	803 (7,396; 2.81)	0.039	102 (881; 2.00)	0.058	1.5	0.91	1.06
SGLT-2i	759 (7,356; 3.05)	0.034	80 (460; 1.00)	0.174	5.1	816 (7,396; 2.96)	0.037	89 (641; 1.00)	0.139	3.7	0.91	1.25
GLP-1 RA	766 (7,356; 3.09)	0.034	73 (141; 1.50)	0.345	10.2	819 (7,396; 3.00)	0.037	86 (219; 1.50)	0.262	7.1	0.91	1.32
Any GLM	638 (7,356; 2.55)	0.034	201 (1,983; 2.10)	0.048	1.4	648 (7,396; 2.29)	0.038	257 (2,654; 2.10)	0.046	1.2	0.89	1.05

44

"Back of the Envelope" Results

- Most difficult to explain: Additions of open label GLP-1 RA
  - Patients stopped study drug to then take same class
  - What can possibly explain a 10-fold higher event rate?

Ad hoc Censoring Criterion	Exenatide					Placebo					Exenatide : Placebo Rate Ratio	
	Prior* to ad hoc Censoring Times		After** ad hoc Censoring Times		After:Prior Rate Ratio	Prior* to ad hoc Censoring Times		After** ad hoc Censoring Times		After:Prior Rate Ratio	Prior to ad hoc Censoring Times	After ad hoc Censoring Times
	Events (N; Avg PY)	Event Rate	Events (N; Avg PY)	Event Rate		Events (N; Avg PY)	Event Rate	Events (N; Avg PY)	Event Rate			
None (ITT)	839 (7,356; 3.12)	0.037				905 (7,396; 3.04)	0.040				0.91	
Metformin	792 (7,356; 3.03)	0.036	47 (318; 2.00)	0.074	2.1	852 (7,396; 2.94)	0.039	53 (392; 2.00)	0.068	1.7	0.91	1.09
Sulfonylureas	758 (7,356; 2.99)	0.034	81 (453; 2.10)	0.085	2.5	821 (7,396; 2.88)	0.039	84 (576; 2.10)	0.069	1.8	0.90	1.23
DPP-4i	763 (7,356; 3.00)	0.035	76 (469; 1.80)	0.090	2.6	799 (7,396; 2.88)	0.038	106 (685; 1.80)	0.086	2.3	0.92	1.05
Insulin	770 (7,356; 2.96)	0.035	69 (563; 2.00)	0.061	1.7	803 (7,396; 2.81)	0.039	102 (881; 2.00)	0.058	1.5	0.91	1.06
SGLT-2i	759 (7,356; 3.05)	0.034	80 (460; 1.00)	0.174	5.1	816 (7,396; 2.96)	0.037	89 (641; 1.00)	0.139	3.7	0.91	1.25
GLP-1 RA	766 (7,356; 3.09)	0.034	73 (141; 1.50)	0.345	10.2	819 (7,396; 3.00)	0.037	86 (219; 1.50)	0.262	7.1	0.91	1.32
Any GLM	638 (7,356; 2.55)	0.034	201 (1,983; 2.10)	0.048	1.4	648 (7,396; 2.29)	0.038	257 (2,654; 2.10)	0.046	1.2	0.89	1.05

45

Differential Diagnosis for Increased Rates

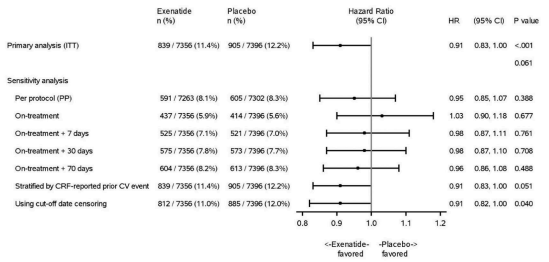
- *(Pure advantage of cardioprotective drugs can be ruled out)*
- Increasing hazard (e.g., perhaps Weibull not exponential)
  - Could potentially investigate among those who do not add GLMs
- Toxicity of added GLMs
  - Approved drugs, widespread use, many patients use at baseline
- Rebound after discontinuation of exenatide
  - Did not model discontinuation, except in GLP-1 RA
  - This would be a safety signal, and would not explain placebo arm
- Confounding
- Protopathic and/or indication bias
  - Placebo group may be diluted by reaction to surrogate measures

46

Additional Evidence of Protopathic Bias

- *NEJM* supplement provided noninferiority sensitivity analyses
  - (But again, did not include all relevant subgroup analyses)

Figure S9: Prespecified Sensitivity Analyses for the Primary Efficacy Endpoint



P values for non-inferiority (upper) and superiority (lower) are provided for the primary analysis.

In time-to-event analyses in the Intention-to-Treat population, patients are censored at the earliest of: 1) date of last contact where all elements of the endpoint could be assessed, and 2) the right censoring date. Patients without any assessment of the endpoint were censored at randomization.

Data analyzed in the Intention-to-Treat population using primary censoring scheme and using a Cox Proportional Hazards model that includes treatment as an explanatory factor and prior CV risk group at randomization based on CRF data as stratification variable. Patients without any assessment of the endpoint were censored at randomization.

47

Additional Evidence of Protopathic Bias

- "Back of the envelope" calculations of event rates after discontinuation of study drug

Ad hoc Censoring Criterion	Exenatide					Placebo					Exenatide : Placebo Rate Ratio	
	Cumulative		Interval		Post Tx : On Tx Rate Ratio	Cumulative		Interval		Post Tx : On Tx Rate Ratio		
	Events (N; Avg PY*)	Event Rate	Events (N; Avg PY)	Event Rate		Events (N; Avg PY)	Event Rate	Events (N; Avg PY)	Event Rate		Cumulative	Interval
On Treatment	437 (7,356; 2.33)	0.026	437 (7,356; 2.329)	0.026		414 (7,396; 2.24)	0.025	414 (7,396; 2.236)	0.025		1.02	1.02
1-7 Days Post Tx	525 (7,356; 2.35)	0.030	88 (6,919; 0.019)	0.668	26.2	521 (7,396; 2.25)	0.031	107 (6,982; 0.019)	0.806	32.2	0.97	0.83
8-30 Days Post Tx	575 (7,356; 2.40)	0.033	50 (6,831; 0.063)	0.117	4.6	573 (7,396; 2.31)	0.034	52 (6,875; 0.063)	0.121	4.8	0.97	0.97
31-70 Days Post Tx	604 (7,356; 2.51)	0.033	29 (6,781; 0.109)	0.039	1.5	613 (7,396; 2.41)	0.034	40 (6,823; 0.109)	0.054	2.1	0.95	0.73
> 70 Days Post Tx	839 (7,356; 3.12)	0.037	235 (6,752; 0.665)	0.052	2.1	905 (7,396; 3.04)	0.040	292 (6,783; 0.688)	0.063	2.5	0.91	0.84
ITT	839 (7,356; 3.12)	0.037	839 (7,356; 3.116)	0.037	1.4	905 (7,396; 3.04)	0.040	905 (7,396; 3.044)	0.040	1.6	0.91	0.91

48

Additional Evidence of Protopathic Bias

- EXTREMELY difficult to explain 32.2-fold higher event rates in the days following discontinuation of an inactive placebo except as protopathic or indication bias
  - Similarly higher rates on exenatide arm

	Exenatide					Placebo					Exenatide : Placebo Rate Ratio	
	Cumulative		Interval			Cumulative		Interval				
Ad hoc Censoring Criterion	Events (N; Avg PY*)	Event Rate	Events (N; Avg PY)	Event Rate	On Tx Rate Ratio	Events (N; Avg PY)	Event Rate	Events (N; Avg PY)	Event Rate	On Tx Rate Ratio	Cumulative	Interval
On Treatment	437 (7,356; 2.33)	0.026	437 (7,356; 2.329)	0.026		414 (7,396; 2.24)	0.025	414 (7,396; 2.236)	0.025		1.02	1.02
1-7 Days Post Tx	525 (7,356; 2.35)	0.030	88 (6,919; 0.019)	0.668	26.2	521 (7,396; 2.25)	0.031	107 (6,982; 0.019)	0.806	32.2	0.97	0.83
8-30 Days Post Tx	575 (7,356; 2.40)	0.033	50 (6,831; 0.063)	0.117	4.6	573 (7,396; 2.31)	0.034	52 (6,875; 0.063)	0.121	4.8	0.97	0.97
31-70 Days Post Tx	604 (7,356; 2.51)	0.033	29 (6,781; 0.109)	0.039	1.5	613 (7,396; 2.41)	0.034	40 (6,823; 0.109)	0.054	2.1	0.95	0.73
> 70 Days Post Tx	839 (7,356; 3.12)	0.037	235 (6,752; 0.665)	0.052	2.1	905 (7,396; 3.04)	0.040	292 (6,783; 0.688)	0.063	2.5	0.91	0.84
ITT	839 (7,356; 3.12)	0.037	839 (7,356; 3.116)	0.037	1.4	905 (7,396; 3.04)	0.040	905 (7,396; 3.044)	0.040	1.6	0.91	0.91

49

Implications for IPTW Analysis

- I am convinced protopathic / indication bias is a major component of increased rate after patient's GLM regimen being modified
- Models predicting such modifications are thus prognostic for patients having protopathic signs
- When using proportional hazards regression, adding strong prognostic variables deattenuates HR estimates
  - The investigators found IPTW had HR go from 0.91 to 0.85
  - (statistical significance depends on HR and number of events)
- To the extent that artificially censoring those who change GLM introduces protopathic bias, IPTW will only produce more extreme biased estimates

50

Further Example: COPD (*NEJM,* 2018)

The NEW ENGLAND JOURNAL of MEDICINE

ESTABLISHED IN 1812 MAY 3, 2018 VOL. 378 NO. 18

Once-Daily Single-Inhaler Triple versus Dual Therapy in Patients with COPD

David A. Lipson, M.D., Frank Barnhart, D.V.M., Noushin Brealey, M.D., Jean Brooks, M.Sc., Gerard J. Criner, M.D., Nicola C. Day, Ph.D., Mark T. Dransfield, M.D., David M.G. Halpin, M.D., MeiLan K. Han, M.D., C. Elaine Jones, Ph.D., Sally Kilbride, M.Sc., Peter Lange, M.D., David A. Lomas, M.D., Ph.D., Fernando J. Martinez, M.D., Dave Singh, M.D., Maggie Tabberer, M.Sc., Robert A. Wise, M.D., and Steven J. Pascoe, M.B., B.S., for the IMPACT Investigators

ABSTRACT

BACKGROUND

The benefits of triple therapy for chronic obstructive pulmonary disease (COPD) with an inhaled glucocorticoid, a long-acting muscarinic antagonist (LAMA), and a long-acting β_2-agonist (LABA), as compared with dual therapy (either inhaled glucocorticoid–LABA or LAMA–LABA), are uncertain.

METHODS

In this randomized trial involving 10,355 patients with COPD, we compared 52 weeks of a once-daily combination of fluticasone furoate (an inhaled glucocorticoid) at a dose of 100 μg, umeclidinium (a LAMA) at a dose of 62.5 μg, and vilanterol (a LABA) at a dose of 25 μg (triple therapy) with fluticasone furoate–

From GlaxoSmithKline, Collegeville (D.A. Lipson, J.B., S.J.P.), and the Perelman School of Medicine, University of Pennsylvania (D.A. Lipson), and Lewis Katz School of Medicine at Temple University (G.J.C.), Philadelphia — all in Pennsylvania; GlaxoSmithKline, Research Triangle Park, NC (F.B., C.E.J.); GlaxoSmithKline, Stockley Park West, Uxbridge (N.B., N.C.D., S.K., M.T.), the Department of Respiratory Medicine, Royal Devon and Exeter Hospital, Exeter (D.M.G.H.), UCL Respiratory, University College London,

51

COPD Example: All Cause Mortality

- Primary endpoint: on-treatment moderate / severe exacerbations
  - Abrupt discontinuation of inhaled corticosteroids associated with rebound exacerbations in asthma, less clear in COPD
- *NEJM* article stressed on treatment mortality differences
  - Data in supplement allows “back of the envelope” calculations
  - Death rate elevated after d/c of non-ICS treatment

Table 1: "Back of the envelope" calculations of events per 1,000 person years according to treatment arm and adherence status. Approximate inference based on exponential models.

	Rate / 1,000 PY (95% CI) (Events / PY; N)			Rate Ratio (95% CI) (2 Sided P value)		
	Arm A: ICS / LAMA / LABA	Arm B: ICS / LABA	Arm C: LAMA / LABA	Arm A: Arm B	Arm A: Arm C	Arm B: Arm C
ITT	**21.8 (17.7,26.8)** (89 / 4,088; N=4,151)	**24.1 (19.7,29.4)** (97 / 4,030; N=4,134)	**30.0 (23.3,38.7)** (60 / 1,999; N=2,070)	**0.90 (0.68,1.21)** (2P=0.494)	**0.73 (0.52,1.01)** (2P=0.055)	**0.80 (0.58,1.11)** (2P=0.179)
On Treatment	**13.2 (10.0,17.4)** (50 / 3,781; N=4,151)	**13.9 (10.5,18.4)** (49 / 3,523; N=4,134)	**22.5 (16.5,30.8)** (39 / 1,731; N=2,070)	**0.95 (0.64,1.41)** (2P=0.802)	**0.59 (0.39,0.89)** (2P=0.013)	**0.62 (0.41,0.94)** (2P=0.025)
Off Treatment	**126.7 (92.6,173.5)** (39 / 308; N=758)	**94.7 (71.4,125.7)** (48 / 507; N=1,040)	**78.2 (51.0,120.0)** (21 / 268; N=566)	**1.34 (0.88,2.04)** (2P=0.177)	**1.62 (0.95,2.75)** (2P=0.075)	**1.21 (0.72,2.02)** (2P=0.465)
Rate Ratio Off:On	**9.58 (6.30,14.57)**	**6.81 (4.57,10.14)**	**3.47 (2.04,5.90)**			

52

Take Home Message: RCTs

- The estimands that are most relevant clinically are best addressed by per randomization analyses
  - Avoids confounding, protopathic bias, indication bias
- NRC monograph on missing data in RCT stressed the need to collect data irrespective of adherence to compliance to study drug
  - The complete data allowed me to explore evidence of protopathic or indication bias
  - In the two cases presented here, I came to the conclusion that the observations were best explained by at least some degree of protopathic or indication bias that differed between arms
  - Absence of such strong evidence would not necessarily have swayed me:
    - Nothing in your data can prove that MNAR does not exist

53

Take Home Message: RCT CTRs

- The Clinical Trial Reports should make data available to perform analyses similar to those I have done here
- Specifically
  - Present detailed description of patient disposition
  - **Means are way more important than medians**
    - "Conditional confounding" is a function of difference in means
    - Means allow estimation of population totals
    - Most estimands are based on means of (transformed) variables
  - **Never report one subgroup without reporting opposite subgroup**
  - In time to event settings, provide description of administrative censoring times: Time from randomization to data analysis lock
  - In time to event settings, provide description of person-years of observation prior to event
    - To the extent the hazard is well approximated by a constant in subgroups, this allow estimating adjusted rates

54

Broader Implications

- These same problems exist in all observational data analyses, especially in the presence of missing data
  - Clinical trials have the advantage of a well-defined (and often clinically important) time 0
- A great many researchers do not seem to understand that censored data methods, IPW, mixed effects analyses, etc. are MAR methods that ultimately rely on untestable assumptions
- Sensitivity analyses for impact of missing data must include a spectrum of MNAR mechanisms
  - Merely providing several (nearly equivalent) MAR models does not suffice

55

Bottom Line

"You better think (think)
about what you're
trying to do…"

-Aretha Franklin, "Think"

56

Backup Slides

Frequentist vs Bayesian

57

57

Frequentist Inference

- Control type 1 error: False positive rate
  - Based on specificity of our methods
- Maximize statistical power: True positive rate
  - Sensitivity to detect specified effect
- At end of analysis provide
  - Unbiased (or consistent) estimates of effect
  - Standard errors: Estimate reproducibility of experiments
  - Confidence intervals: hypotheses that might generate similar data
  - P values
- Criticism: Compute probability of data already observed
  - "A precise answer to the wrong question"

58

Bayesian Inference

- Hypothesize prior prevalence of "good" ideas
  - Subjective probability
  - (Can consider a range of prior distributions)
- Using prior prevalence and frequentist sampling distribution
  - Condition on observed data
  - Compute probability that some hypothesis is true
    - "Posterior probability"
  - Estimates based on summaries of posterior distribution
- Criticism: Which presumed prior distribution is relevant?
  - "A vague answer to the right question"

59

Frequentist vs Bayesian

- Frequentist and Bayesian inference truly complementary
- Frequentist:
  - Design so the same data not likely from null / alt
- Bayesian:
  - Updated belief about probability of beneficial drug
  - Consider the range of prior distributions that would correspond to credible inference

60

Frequentist vs Bayesian

- Bayes rule: PPV from type I error, power, subjective prevalence
  - Maximize new information by maximizing Bayes factor
  - With simple (binary) hypotheses:

$$PPV = \frac{power \times prevalence}{power \times prevalence \quad + \quad type\ I\ err \times (1 - prevalence)}$$

$$\frac{PPV}{1 - PPV} = \frac{\textbf{power}}{\textbf{type I err}} \times \frac{prevalence}{1 - prevalence}$$

$$posterior\ \ odds = \textbf{Bayes Factor} \times prior\ \ odds$$

61

Distribution-free Bayesian Models

- Regard estimate of summary measure as the data
  - Use asymptotic distributions under population model

Some joint distribution for $(\theta, \hat{\theta})$

Frequentist usually considers $\hat{\theta} \mid \theta \sim N(\theta, V(\theta)/n)$

Bayesian can consider $p\left(\vec{\theta} \mid \hat{\vec{\theta}}\right) = \dfrac{p\left(\hat{\vec{\theta}} \mid \vec{\theta}\right) \lambda(\vec{\theta})}{\int p\left(\hat{\vec{\theta}} \mid \vec{\theta}\right) \lambda(\vec{\theta}) d\vec{\theta}}$

where $\lambda(\vec{\theta})$ is a prior distribution for $\vec{\theta}$

62

Backup Slides

Censored Survival Data

63

63

Motivating Example

- Hypothetical study of subject survival
- Subjects accrued to study and followed until time of analysis
- Study done at three centers, which started the studies in three successive years
- Censoring time thus differs across centers
- How do we use this data to estimate 3 year survival?

64

Data by Date (Real Time)

Staggered study entry by site

Year		Accrual Group A	B	C
1990	On study	100	--	--
	Died	43		
	Surviving	57		
1991	On study	57	100	--
	Died	27	53	
	Surviving	30	47	
1992	On study	30	47	100
	Died	13	22	55
	Surviving	17	25	45

65

Data by Study Time

Realign data according to time on study

Year		Accrual Group A	B	C
1	On study	100	100	100
	Died	43	53	55
	Surviving	57	47	45
2	On study	57	47	--
	Died	27	22	
	Surviving	30	25	
3	On study	30	--	--
	Died	13		
	Surviving	17		

66

Combined Data

Year		Accrual Group A	B	C	Combined
1	On study	100	100	100	300
	Died	43	53	55	151
	Surviving	57	47	45	149
2	On study	57	47	--	104
	Died	27	22		49
	Surviving	30	25		55
3	On study	30	--	--	30
	Died	13			13
	Surviving	17			17

67

Problem Posed by Missing Data

- Sampling scheme causes (informative) missing data
- Potentially, we might want to estimate three year survival probabilities
- Different centers contribute information for varying amounts of time
  - One year survival can be estimated at A, B, C
  - Two year survival can be estimated at A, B
  - Three year survival can be estimated at A

68

Possible Remedies

- WRONG: Ignore missing
  - E.g., 17 of 300 subjects alive at three years
- RIGHT BUT WRONG QUESTION: Use data only up to earliest censoring time
  - E.g., 149 of 300 subjects alive at one year
- RIGHT BUT INEFFICIENT: Use only center A
  - E.g., 17 of 100 subjects alive at three years

69

Best Approach

- RIGHT AND EFFICIENT
  - Use all available data to estimate that portion of survival for which it is informative
  - Use Centers A, B, and C to estimate one year survival
  - Use Centers A and B to estimate proportion of one-year survivors who survive to two years
  - Use Center A to estimate proportion of two-year survivors who survive to three years

70

Theoretical Basis for Approach

- Properties of probabilities
  - Probability of event A and B occurring is product of
    - Probability that A occurs when B has occurred
    - Probability that B has occurred

$$\Pr(A \cap B) = \Pr(A \mid B) \times \Pr(B)$$

71

Estimate Conditional Survival

- Condition on surviving up until the start of the time interval
  - Denominator is number of subjects at start of interval
  - Numerator is deaths during the interval

- Requirement for validity
  - Subjects available at the start of each time interval are a random sample of the population surviving to that time
    - "Missing at Random" (MAR)
    - "Noninformative censoring"

72

Application to Example

- Within interval conditional probabilities
  - Use A, B, C to estimate $Pr\ (T^0 \geq 1)$
  - Use A, B to estimate $Pr\ (T^0 \geq 2 \mid T^0 \geq 1)$
  - Use A to estimate $Pr\ (T^0 \geq 3 \mid T^0 \geq 2)$

- Multiply to obtain unconditional cumulative survival
  - $Pr\ (T^0 \geq 1)$
  - $Pr\ (T^0 \geq 2) = Pr\ (T^0 \geq 2 \mid T^0 \geq 1)\ Pr\ (T^0 \geq 1)$
  - $Pr\ (T^0 \geq 3) = Pr\ (T^0 \geq 3 \mid T^0 \geq 2)\ Pr(T^0 \geq 2)$

73

Motivating Example Results

Survival Probabilities

Yr	Combined	Each Year	Cumulative
1	On study 300		
	Died 151		
	Surviving 149	149/300 = 49.67%	49.67%
2	On study 104		
	Died 49		
	Surviving 55	55/104 = 52.88%	.4967*.5288 = 26.27%
3	On study 30		
	Died 13		
	Surviving 17	17/ 30 = 56.67%	.2627*.5667 = 14.88%

74

Redistribute to the Right

- The Kaplan-Meier estimate is exactly equivalent to a MAR imputation scheme
  - Impute missing data for censored subjects by sampling from subjects who were still on study at the time of censoring
- Basic idea
  - Recall the empirical cdf assigns probability 1/n to each observation
  - A censored observation should be equally likely to have event time like any of the remaining uncensored observations
    - Recursively redistribute the mass of each censored observation among the subjects remaining at risk

75

Backup Slides

Regulatory

76

76

Overall Goal: "Drug Discovery"

- More generally
  - for some disease
  - in some population of patients
  - a therapy / preventive strategy or diagnostic / prognostic procedure
  - that provides a desired outcome

- A ***sequential***, ***adaptive*** series of experiments to establish
  - Safety of investigations / dose (phase 1)
  - Safety of therapy (phase 2)
  - Measures of efficacy (phase 2)
    - Treatment, population, and outcomes
  - Confirmation of efficacy (phase 3)
  - Confirmation of effectiveness (phase 3, post-marketing)

77

Regulation of Drugs in US

- Wiley Act (1906)
  - Labeling
- Food, Drug, and Cosmetics Act of 1938
  - Safety
- Kefauver – Harris Amendment (1962)
  - Efficacy / effectiveness
    - " [If] there is a lack of substantial evidence that the drug will have the effect ... shall issue an order refusing to approve the application. "
    - "...The term 'substantial evidence' means evidence consisting of **adequate and well-controlled investigations, including clinical investigations**, by experts qualified by scientific training"
- FDA Amendments Act (2007)
  - Registration of RCTs, Pediatrics, Risk Evaluation and Mitigation Strategies (REMS)

78

Medical Devices

- Medical Devices Regulation Act of 1976
  - Class I: General controls for lowest risk
  - Class II: Special controls for medium risk - 510(k)
  - Class III: Pre marketing approval (PMA) for highest risk
    - **"…valid scientific evidence** for the purpose of determining the safety or effectiveness of a particular device … adequate to support a determination that there is reasonable assurance that the device is safe and effective for its conditions of use…"
    - "Valid scientific evidence is evidence from well-controlled investigations, partially controlled studies, studies and objective trials without matched controls, well-documented case histories conducted by qualified experts, and reports of significant human experience with a marketed device, **from which it can fairly and responsibly be concluded by qualified experts that there is reasonable assurance of the safety and effectiveness…**"
- Safe Medical Devices Act of 1990
  - Tightened requirements for Class 3 devices

79

Missing Data in RCT

- National Academy of Science Oversight Committee
  - Experts in missing data methodology and clinical trial methodology

Roderick Little, Chair	
Ralph D'Agostino	Susan Murphy
Kay Dickersin	James Neaton
Scott Emerson	Andrea Rotnizky
John Farrar	Daniel Scharfstein
Constantine Frangakis	Weichung (Joe) Shih
Joseph Hogan	Jay Siegel
Geert Molenberghs	Hal Stern

80

80

Common Problems (Report)

- Missing data due to discontinuation of treatment
  - Adverse events vs lack of efficacy vs efficacy
  - Specified by protocol vs perception of subjects or investigators
    - Relevance of data *vis a vis* health status, rescue therapies
- Outcomes undefined or unmeasurable for some patients
  - Counterfactual estimands (e.g., QoL after death)
  - Competing risks (e.g., renal function after transplant)
- Missing data because of attrition in the course of the study
  - Missed visits, loss to follow-up, withdrawal of consent
- Missing data in composite outcomes
- Missing data due to death

81

81

Statistical Classification of Missing Data

- Missing completely at random (MCAR)
  - The indicator of missingness does not depend upon any measured data
    - Sometimes confused with ignorability
- Missing at random (MAR)
  - Within groups defined by some observed data, the data is missing completely at random
  - Information about missing data can be borrowed from data that is available
- Missing not at random (MNAR)
  - Even after conditioning on all observed data, the subjects missing data would have outcomes distributed differently than those for subjects with observed data

82

Statistical Impact of Missing Data

- Ignorable
  - Weak: Analyzing complete cases in the planned analyses provides unbiased estimates of the desired estimand
    - MCAR
    - MAR if we were going to adjust anyway
  - Strong: Just as precisely?
- Nonignorable
  - Failure to account for missingness results in biased estimation of the desired estimand

83

Estimands

- Clinical issues:
  - The indication
    - Disease, Population, Treatment, Outcome
  - Clinical importance of distributional summary measure
  - Clinical importance of stratification
- Statistical issues:
  - Summarizing the outcome distribution
    - Mean, geometric mean, median of continuous data
    - Proportion or odds above threshold
    - (Time averaged) hazard ratio of censored data
  - Covariate adjustment (precision)
  - Per randomization analyses

84

Scientific Efficacy / Safety Estimands

- What is impact in a population assigned to treatment protocol?
  - "Intent to treat" (ITT) estimand
  - Rigorous causal estimand
  - Full data on all randomized subjects
  - Analysis by randomized group
- What is impact among patients who follow protocol?
  - "Per protocol" (PP) estimand
  - No matter what: An interesting basic science question
  - Clinically may be used to explore mechanism of action
  - Clinically may be desired to describe prognosis

85

Per Protocol Estimands: Use of Data

- Patients who don't follow protocol may be irrelevant
  - Patients who do not follow directions
  - Patients who have intolerable adverse reactions
    - Perhaps "intolerable" only because uncertain of efficacy, or
    - Perhaps leading to serious consequences with continued therapy
  - Patients with real or perceived lack of efficacy
    - Early clinical course is discouraging, or
    - Definitive progression to serious condition prior to primary endpoint
  - Development of contraindication to treatment (e.g., pregnancy)
  - Patients with early evidence of cure
  - Patients with competing risk that prevents measurement
- Ignoring vs imputing missing data?

86

Impact of Estimand on RCT

- The patients who are "relevant" differ according to the estimand of interest
- The primary goal should be to devise an experiment that only randomizes patients who are relevant to the estimand
  - This is often difficult
  - It may mean using more than one RCT, answering different aspects of the safety/effectiveness profile in different studies
- Sometimes, however, a counterfactual estimand is of greatest scientific interest
  - In these cases, all results are subjective

87

ICH E9 (R1) Strategies

- Treatment Policy (per randomization, ITT)
  - Intercurrent events are generally irrelevant
- Hypothetical strategies
  - Imagine intercurrent event would not occur
- Composite endpoint strategies
  - Incorporate intercurrent event as part of outcome
- While on treatment strategies
  - Only incorporate experience prior to intercurrent event
- Principal stratum strategies

88

88