Meta-analysis in Drug Safety Assessment with a Focus on ESA's and Diabetes

Jesse A. Berlin, ScD
Johnson & Johnson Pharmaceutical Research and Development
Seattle Clinical Trials Symposium
22 November 2010

The Disclaimer

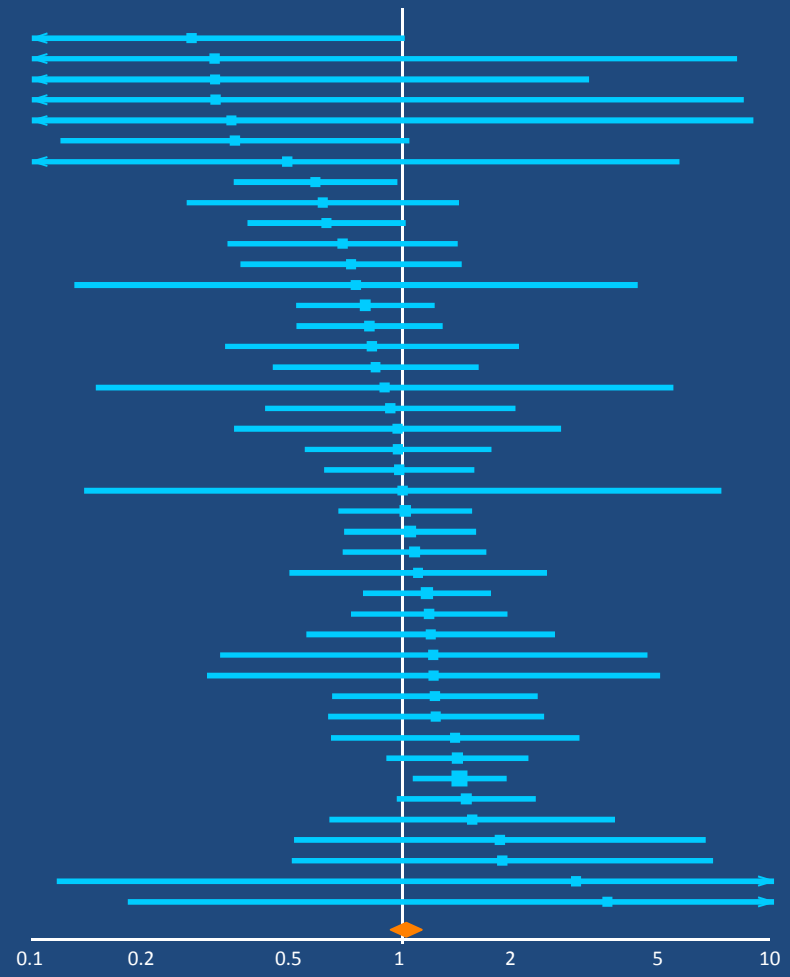
- These are the views of the speaker and do NOT (necessarily) reflect policy of Johnson & Johnson Pharmaceutical Research and Development.

Is it sampling variability?

- Problem: How do we distinguish sampling variability from “real” variability (possibly) associated with different effects of treatment in different subgroups of patients (or with different dosing algorithms or other specific aspects of treatment)?

Study-level Meta-analysis of OUTCOME X
46 Controlled DRUG A Studies *(n=12,034; 6505 treated; 5529 Control)*

Study name	Odds ratio	Lower limit	Upper limit
Dammacco (INT-2)	0.27	0.07	1.01
* Throuvalas 2000	0.31	0.01	7.95
Del Mastro	0.31	0.03	3.17
Dunphy 1999	0.31	0.01	8.28
* Vadhan-Raj	0.34	0.01	8.80
Oberhoff	0.35	0.12	1.04
Cazzola (Roche)	0.49	0.04	5.56
* EPO-GER-022	0.58	0.35	0.96
INT-3	0.61	0.26	1.41
Vansteenkiste (AMG 980297)	0.62	0.38	1.01
Henry_1995	0.69	0.34	1.40
* Blohmer (AGO-NOGGO)	0.73	0.37	1.44
Ten Bokkel (Roche)	0.75	0.13	4.28
Pirker(AMG 20010145)	0.79	0.52	1.21
Littlewood	0.81	0.52	1.28
Kotasek_2003 (AMG 980291)	0.83	0.33	2.05
Taylor 2005 (AMG 20030232)	0.84	0.45	1.59
Pangalis (P-174)	0.89	0.15	5.36
* EPO-CAN-15	0.93	0.43	2.00
Coiffer	0.97	0.35	2.66
Chang 2005 (EPO-CAN-17)	0.97	0.55	1.73
Witzig 2005	0.98	0.62	1.55
Razzouk 2006 updt	1.00	0.14	7.23
Aapro 2006 (BRAVE)	1.02	0.67	1.53
Mobus	1.05	0.70	1.57
Osterborg 2005	1.08	0.69	1.67
Osterborg 96 (Roche)	1.10	0.50	2.44
Milroy (INT-49)	1.16	0.79	1.72
Savonije 2005	1.18	0.73	1.90
* Thomas (GOG-0191)	1.19	0.55	2.56
Engert 2007 (HD 15 IA)	1.21	0.32	4.55
Thatcher combined	1.21	0.30	4.93
Case	1.22	0.65	2.30
Prozanto (INT-47)	1.23	0.63	2.39
Rose	1.39	0.64	2.98
PREPARE	1.40	0.90	2.15
Leyland-Jones (1 year ITT)	1.42	1.07	1.90
Hedenus 2003 (AMG 20000161)	1.48	0.97	2.27
Grote 2005 (N93-004)	1.54	0.64	3.72
Bamia	1.83	0.51	6.55
INT-I	1.86	0.50	6.85
O'Shaughnessy 2005	2.94	0.12	73.93
Wilkinson 2006	3.57	0.18	70.31
Random Effects Model	1.02	0.93	1.13



I^2 = 0.3%; Fixed Effects Model = Random Effects Model
If BEST LFU data are used, then I^2 =0%; Random Effects Model = Fixed Effects Model = 0.97 (0.88, 1.07)
[a]: 3 studies (Cascinu; Hedenus; Kurtz) reported no EVENTS in either arm
*: Radiochemotherapy study

Background on ESAs

- Darbepoetin alfa (Aranesp) and epoetin alfa (Procrit) are erythropoiesis-stimulating agents (ESAs)
  - Both are licensed for the treatment of anemia in non-myeloid cancer patients with chemotherapy-induced anemia (CIA)
  - Also approved (originally) in treating anemia in chronic renal failure (essentially as a kind of hormone replacement therapy)
- ESAs provide the only alternative to blood transfusion
  - risks including infection, transfusion reactions, and transfusion-associated immunosuppression

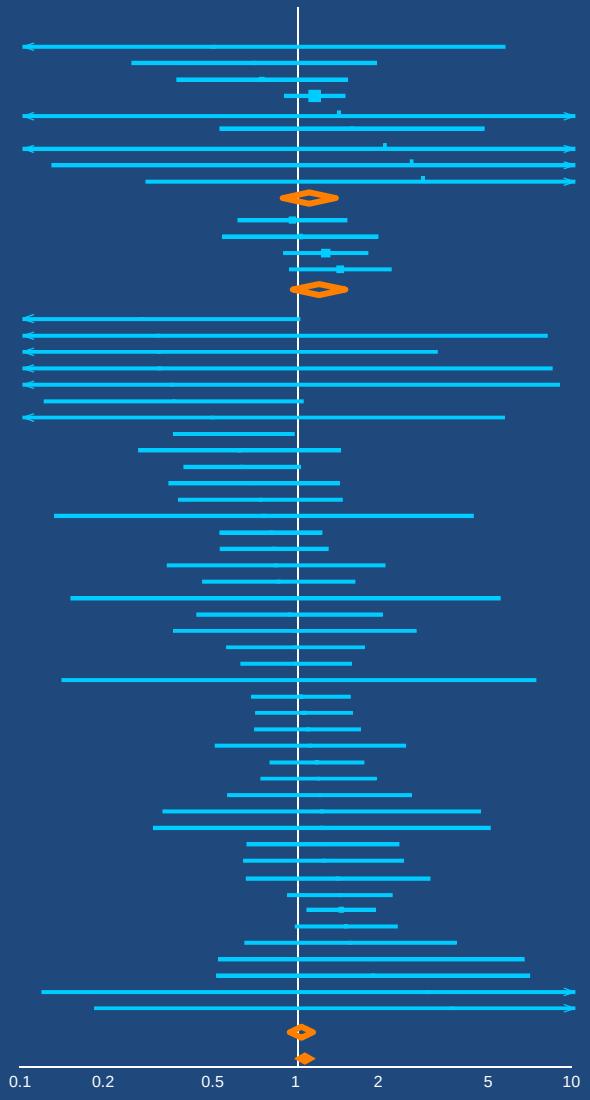
Known Safety Concerns with ESAs in CIA

- An increased risk of cardiovascular/thromboembolic (CV/TE) events has been consistently observed and appropriately represented in class labeling for all ESAs
- Growing FDA concern over time that ESAs may negatively impact overall survival and lead to more rapid tumor progression; warning in current label
- Current boxed warning specifically states:
  - ESAs shortened overall survival and/or time-to-tumor progression in clinical studies in patients with breast, NSC lung, head and neck, lymphoid, and cervical cancers when dosed to target a hemoglobin of ≥ 12 g/dL
- There are several published systematic reviews / meta-analyses, including an individual patient data analysis done by members of one of the Cochrane review groups

Study-level Meta-analysis of Overall Death

59[a] Controlled ESA Studies *(n=15,249; 8309 ESA, 6940 Control)*

Treatment Type	Study name	Odds ratio	95% CI Lower limit	95% CI Upper limit
AOC	Mystakidou 2005	0.49	0.04	5.58
	Gordon 2006 (AMG 20030204)	0.69	0.25	1.91
	Abels	0.74	0.36	1.49
	Smith 2008 (AMG 20010103)	1.14	0.89	1.47
	EPO-CAN-203	1.40	0.05	36.45
	Charu 2004 (AMG 20000219)	1.56	0.52	4.69
	EPO-CAN-303	2.05	0.07	58.65
	Smith 2003 (AMG 990111)	2.56	0.13	51.56
	Wright 2007 (EPO-CAN-20)	2.82	0.28	28.56
Random Effects Model: AOC		1.09	0.87	1.36
RTx	EPO-GBR-7	0.95	0.60	1.49
	Machtay (RTOG 99-03)	1.01	0.53	1.93
	Overgaard (DAHANCA 10, IA)	1.25	0.88	1.78
	Henke 2003 (ENHANCE)	1.41	0.93	2.16
Random Effects Model: RTx		1.18	0.95	1.47
CIA	Dammacco (INT-2)	0.27	0.07	1.01
	Throuvalas 2000	0.31	0.01	7.95
	Del Mastro	0.31	0.03	3.17
	Dunphy 1999	0.31	0.01	8.28
	Vadhan-Raj	0.34	0.01	8.80
	Oberhoff	0.35	0.12	1.04
	Cazzola (Roche)	0.49	0.04	5.56
	EPO-GER-022	0.58	0.35	0.96
	INT-3	0.61	0.26	1.41
	Vansteenkiste (AMG 980297)	0.62	0.38	1.01
	Henry_1995	0.69	0.34	1.40
	Blohmer (AGO-NOGGO)	0.73	0.37	1.44
	Ten Bokkel (Roche)	0.75	0.13	4.28
	Pirker(AMG 20010145)	0.79	0.52	1.21
	Littlewood	0.81	0.52	1.28
	Kotasek_2003 (AMG 980291)	0.83	0.33	2.05
	Taylor 2005 (AMG 20030232)	0.84	0.45	1.59
	Pangalis (P-174)	0.89	0.15	5.36
	EPO-CAN-15	0.93	0.43	2.00
	Coiffer	0.97	0.35	2.66
	Chang 2005 (EPO-CAN-17)	0.97	0.55	1.73
	Witzig 2005	0.98	0.62	1.55
	Razzouk 2006 updt	1.00	0.14	7.23
	Aapro 2006 (BRAVE)	1.02	0.67	1.53
	Mobus	1.05	0.70	1.57
	Osterborg 2005	1.08	0.69	1.67
	Osterborg 96 (Roche)	1.10	0.50	2.44
	Milroy (INT-49)	1.16	0.79	1.72
	Savonije 2005	1.18	0.73	1.90
	Thomas (GOG-0191)	1.19	0.55	2.56
	Engert 2007 (HD 15 IA)	1.21	0.32	4.55
	Thatcher combined	1.21	0.30	4.93
	Case	1.22	0.65	2.30
	Prozanto (INT-47)	1.23	0.63	2.39
	Rose	1.39	0.64	2.98
	PREPARE	1.40	0.90	2.15
	Leyland-Jones (1 year ITT)	1.42	1.07	1.90
	Hedenus 2003 (AMG 20000161)	1.48	0.97	2.27
	Grote 2005 (N93-004)	1.54	0.64	3.72
	Bamia	1.83	0.51	6.55
	INT-I	1.86	0.50	6.85
	O'Shaughnessy 2005	2.94	0.12	73.93
	Wilkinson 2006	3.57	0.18	70.31
Random Effects Model: CIA		1.02	0.93	1.13
Random Effects Model: Overall		1.05	0.97	1.14



I^2 = 0% for all subgroups and Overall except CIA where I^2 = 0.3%; Overall Fixed Effects Model = Random Effects Model
Meta Analysis Using OR

[a] 3 studies (Cascinu; Hedenus; Kurtz) reported no deaths in either arm

If BEST LTFU data are used, then Overall I^2 = 0%; Random Effects Model = Fixed Effects Model = 1.02 (0.94, 1.10)

* = Radiochemotherapy study

Risks of ESAs (from 2008 FDA presentation)

- 6 studies→ statistically significant evidence of ↑ tumor promotion and/or ↓ survival
  - BEST (Breast)*
  - ENHANCE (Head/Neck)
  - DAHANCA (Head/Neck)
  - 161 (Lymphoid Ca)*
  - CAN-20 (NSCLC)
  - 103 (Anemia of Cancer) (Many tumor types)

- 2 studies→ trends of ↑ tumor promotion and/or ↓ survival
  - PREPARE (neoadjuvant breast)*
  - GOG 191 (cervical cancer)†

* = pts receiving chemo

†=pts receiving chemoRT

Study	N	1° Endpoint	ESA Adverse Outcome (label)
Chemo			
BEST (breast)	939	12 mo OS	↓ 12 mo OS
161 (Lymphoid)	344	Δ Hgb	↓ OS
PREPARE (breast)	733	RFS, OS	↓ RFS*, ↓ OS*
GOG 191 (cervical)	114	PFS	↓ OS*
RT			
ENHANCE (H/N) 351	LR PFS	↓ LR PFS, ↓ OS	
DAHANCA (H/N) 522	LRC	↓ LRC, ↓ OS*	
No Chemo or RT			
CAN-20 (NSCLC) 70	QOL	↓ OS	
103 (Heterogenous)	989	Transfusion	↓ OS

*=trend

What could be going on?

- Who is at risk?
- Why?
- Factors considered in the past:
  - Baseline hemoglobin
  - "Target" hemoglobin (≠ "achieved" hemoglobin)
  - Anemia due to cancer, not chemotherapy
  - Inability to respond to treatment
- All of this is exploratory (and at this point is *post hoc* and data-driven)

Mortality And ESA Response[a]: (Landmark At 4 Weeks On Treatment)
All Randomized Studies, ESA-treated Subjects

Study/Contrast	Hazards Ratio (b)	95% Confidence Interval (b)	P value (b)
All studies, ESA-treated subjects (N=3750, # of events=1293)			
ESA Response: Increase vs. Stable	**0.84**	**(0.72, 0.98)**	**0.02**
ESA Response: Decrease vs. Stable	**1.22**	**(1.03, 1.45)**	**0.02**
ESA Response: Missing vs. Stable	**1.13**	**(0.92, 1.38)**	**0.24**

a) Increase: Increase of >0.5 g/dL;
Stable: change of ≤ 0.5 g/dL;
Decrease: a drop of >0.5 g/dL;
In hemoglobin over the pre-treatment level, independent of transfusion

b) Analyses were based on Cox's proportional hazards model, adjusting for ECOG score, cancer type, advanced disease, pre-treatment hemoglobin and other baseline variables, stratified by study.

Mortality And ESA Response[a]: (Landmark At 4 Weeks On Treatment)
Beyond-anemia-correction studies, ESA-treated subjects

Study/Contrast	Hazards Ratio (b)	95% Confidence Interval (b)	P value (b)
Beyond-anemia-correction studies, ESA-treated subjects (N=2089, # of events=758)			
ESA Response: Increase vs. Stable	0.91	(0.75, 1.12)	0.38
ESA Response: Decrease vs. Stable	1.50	(1.21, 1.86)	0.00
ESA Response: Missing vs. Stable	1.19	(0.88, 1.60)	0.26

Mortality And ESA Response[a]: (Landmark At 4 Weeks On Treatment)
Anemia-correction studies, ESA-treated subjects

Study/Contrast	Hazards Ratio (b)	95% Confidence Interval (b)	P value (b)
Anemia-correction studies, ESA-treated subjects (N=1661, # of events=535)			
ESA Response: Increase vs. Stable	**0.75**	**(0.60, 0.95)**	**0.02**
ESA Response: Decrease vs. Stable	**0.92**	**(0.69, 1.22)**	**0.56**
ESA Response: Missing vs. Stable	**1.00**	**(0.75, 1.32)**	**0.99**

What about in chronic renal failure?

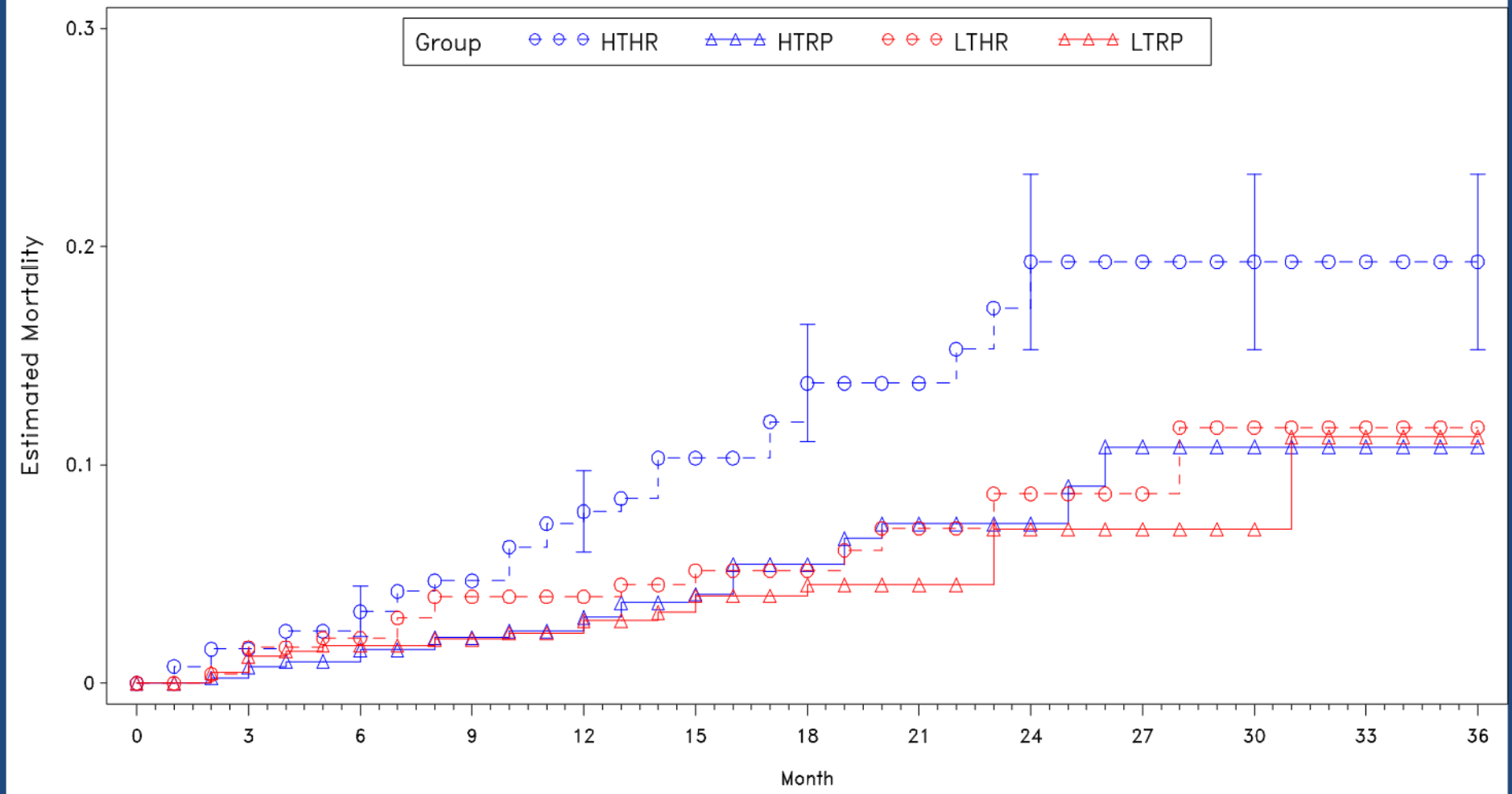
- Similar pattern with respect to “hyporesponse”?

CHOIR Study

- Chronic kidney disease, not on dialysis (but late stage disease)
- All participants treated with same initial dose for first 4 weeks of study
- This allowed an assessment of "responsiveness" at that time point

CHOIR Study: Plot of KM Mortality by Responsiveness Group

HTHR=High Target Hyporesponders; HTRP=High Target Responders
LTHR=Low Target Hyporesponders; LTRP=Low Target Responders



# at Risk												
HTHR	261	239	212	185	164	125	95	59	38	30	21	6	0
HTRP	424	396	367	335	300	223	170	107	61	41	28	12	0
LTHR	261	238	213	193	176	135	107	70	47	30	20	11	0
LTRP	427	421	369	347	321	232	164	93	53	36	23	9	0

Conclusions (1)

- Meta-analysis of safety endpoints can provide increased power and precision, especially useful for uncommon events
- Meta-analysis (e.g., ESAs), when studies are already completed, permits exploration of patient (or other) characteristics that may modify treatment differences (EXPLORATORY, even if pre-defined)
- Sometimes an overall pooled result is not as informative as we would like
  - Where is the increase in risk?
  - Where is there relative “safety”?

META-ANALYSIS IN DIABETES DRUG STUDIES

What is a Prospective Meta-analysis (PMA)?

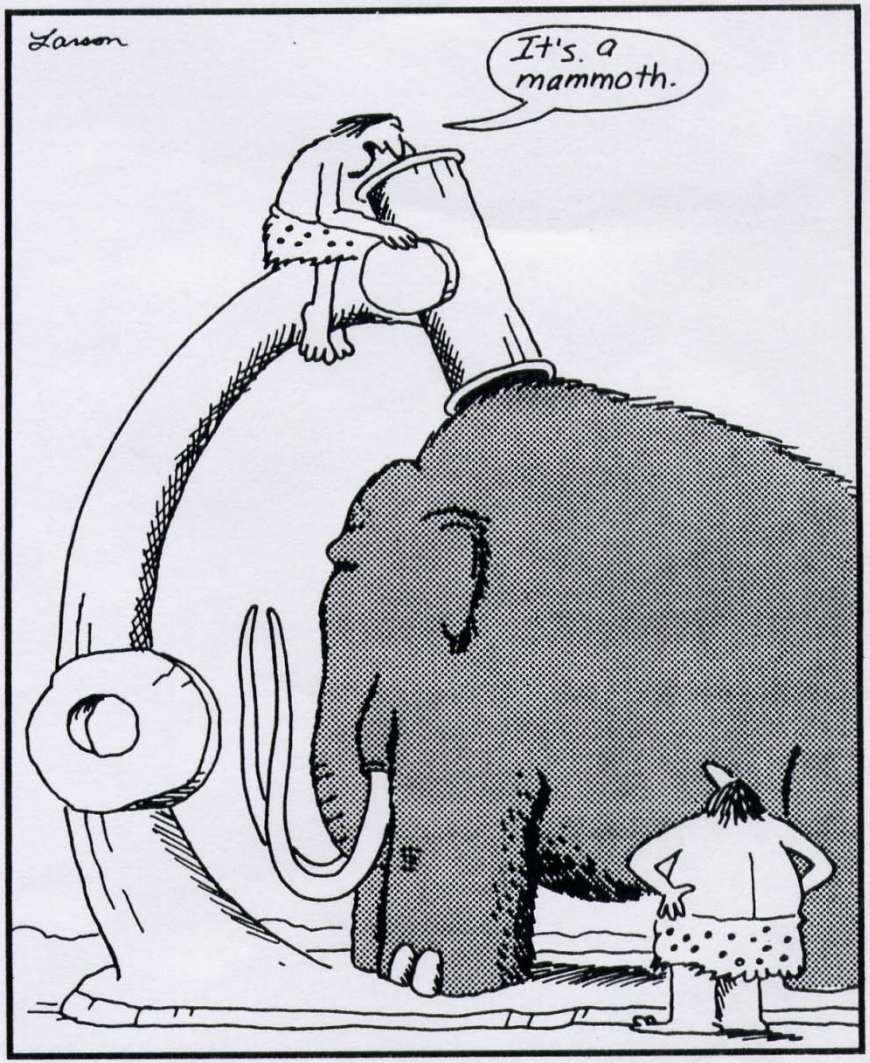
A PMA is a meta-analysis of RCTs identified, evaluated, and determined to be eligible for the meta-analysis before the results of any of those trials become known.

What can PMA do?

- PMA can help to overcome some of the recognized problems of retrospective meta-analyses by enabling PROSPECTIVE:
  - Specification of hypotheses (endpoints, etc.), ***ignorant of the results of individual trials***
  - Application of selection criteria
  - Specification of analysis plans (including subgroup definitions)

Another advantage

- PMA allows for standardization of data collection across studies.
  - For example, the same definitions of the endpoint (or the adverse event) can be specified in all trials!!!!!!!
  - And implemented by the same adjudication committee across all studies



Early microscope

Studies with dual purposes

- In diabetes, efficacy captured by
  - HbA1c
  - Microvascular events
  - Others
- But studies will also capture “macro” CV endpoints (and deaths)
  - Same definitions of the endpoint (or the adverse event) can be specified in all trials!!!!!!!
  - And implemented by the same adjudication committee across all studies

Plan the Meta-analysis to Analyze CV Events

- Typically will (could? should?) include:
  - Phase 3 (randomized)
  - Large randomized outcomes study (possibly "simple")
  - Phase 4 (randomized)
  - Phase 2 (randomized)? (recent debate in *Contemporary Clinical Trials*)

Designing “meta-experiments”

- Don’t (accidentally) confound important features with “study”
  - E.g., higher dose aspirin in men (1º prevention of MI), lower dose aspirin in women
  - Better statistically to stratify within study
    - But this is not intuitive to many people
    - Logistically more challenging

Features of interest in diabetes populations (not an exhaustive list)

- New onset vs. existing diabetes
  - Duration of diabetes
- Prior therapy (or therapies)
- Current “background” therapy
- Baseline HbA1c
- ENRICHMENT
  - Can we enroll patients at high CV risk?
  - How does that result apply to a lower risk population?
    - relative risk vs. absolute risk reduction
    - Interaction (on what scale?)

Statistical Issues (1)

- Sample size considerations
  - Rule out RR of 1.8 or 1.3
    - What if you meet non-inferiority but RR is statistically significantly above 1.0?
    - Should we also specify an upper limit to the point estimate?

Statistical Issues (2)

- Interim analyses and “rules” for stopping?
  - What about alpha penalties?
  - Even for meta-analyses? Especially for meta-analyses?
- Adaptation (at an interim analysis)?
  - Is it OK to change sample size and hypotheses to aim for superiority?
  - We can always test for superiority if we meet non-inferiority

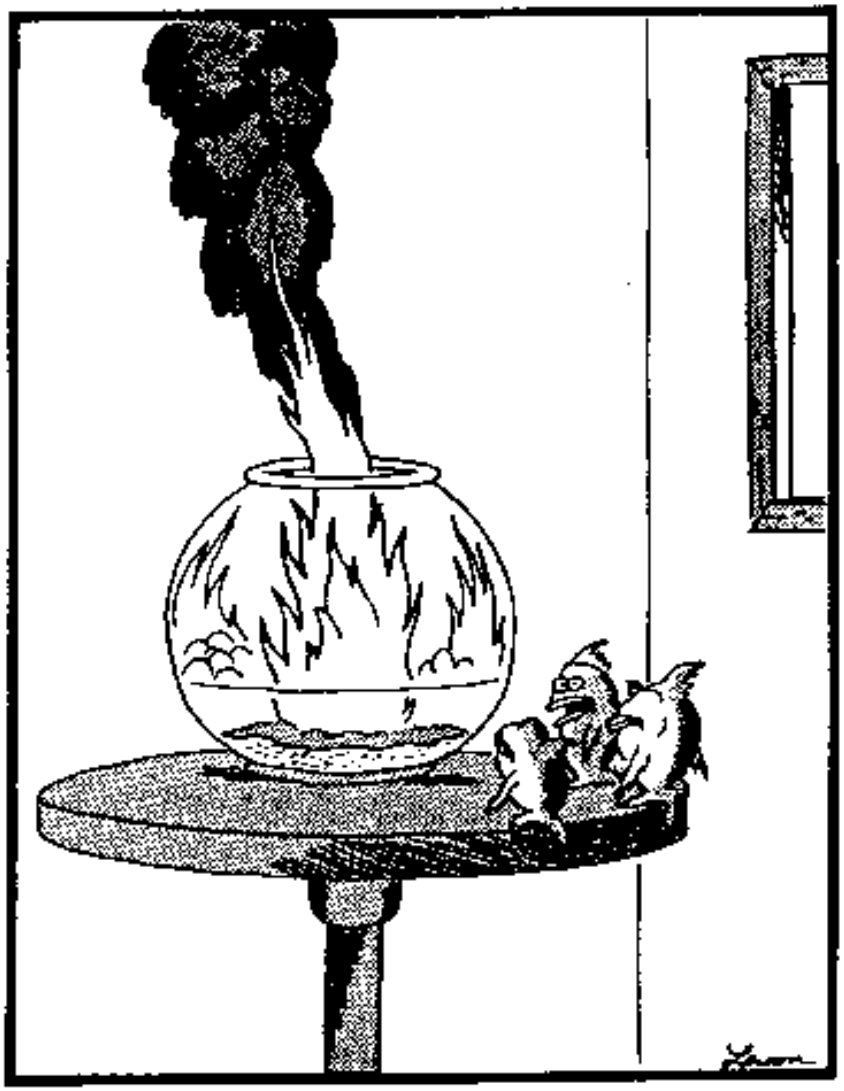
Statistical Issues (3)

- Rare events WILL still be an issue, despite overall statistical power
  - Zero counts within smaller, individual studies (especially phase 2 glycemic effect studies)

➢ Possible methodologic approaches

- ➢ Use continuity correction factors (with caution- see Sweeting 2004 and 2006)
- ➢ Grouping studies with similar randomization ratios (per recent Nissen paper – avoids confounding)
- ➢ Bayesian approaches (choice of difference or ratio scale?)

- Include sensitivity analyses



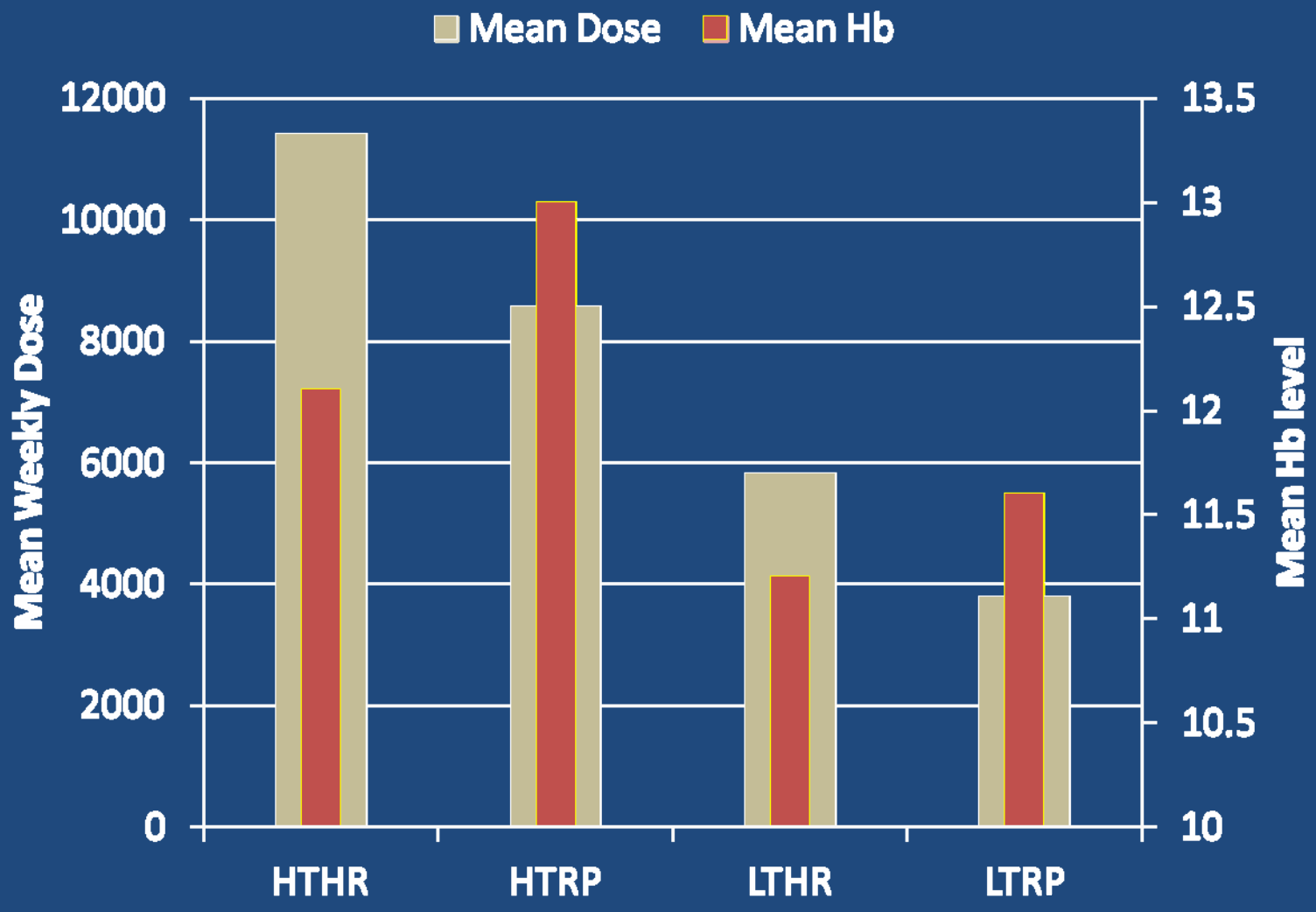
"Well, thank God we all made it out in time.
... 'Course, now we're equally screwed."

Conclusions (2)

- PMA in diabetes drug development provides a way of ensuring freedom from some biases (provided the component studies are well-designed) by pre-specifying a protocol for the MA
- Provides statistical power to examine uncommon (CV) events and subgroups
- Should rigorously collect data that would otherwise not have been collected (e.g., adjudicated CV events)
- Beware of confounding of study design factors

Backups

Mean Dose and Achieved Hb Level



Hierarchical Outcome Classification – An Example with CV Events

Endpoint	# Events Celecoxib	# Events Placebo	Hazard Ratio	95% CI	*p*-value
CV death	6	1	6.1	(0.7, 50.3)	0.10
CV death or MI	15	4	3.8	(1.3, 11.5)	0.015
CV death, MI, or stroke	20	6	3.4	(1.4, 8.5)	0.007
CV death, MI, stroke, or CHF	23	7	3.4	(1.4, 7.8)	0.006
CV death, MI, stroke, CHF, or angina	25	11	2.3	(1.1, 4.7)	0.027
CV death, MI, stroke, CHF, angina, or CV procedure	31	17	1.9	(1.0, 3.3)	0.05

- Proschan MA, Lan KKG, Wittes JT. Statistical Monitoring of Clinical Trials, 2006
- Solomon SD, McMurray JJV, Pfeffer, MA, et al. NEJM, 2005