Statistics•Collaborative

*design and analysis for biomedical research*

# Heterogeneity in Reporting of Adverse Events in Multiregional Clinical Trials

Janet Wittes

ASA Biopharm Workshop

Washington DC

September 14, 2018

# Outline

- What we already know about regional heterogeneity
- Is region just another subgroup?
- If not, why not?
- Some examples of US vs. other - efficacy
- Hypotheses related to safety

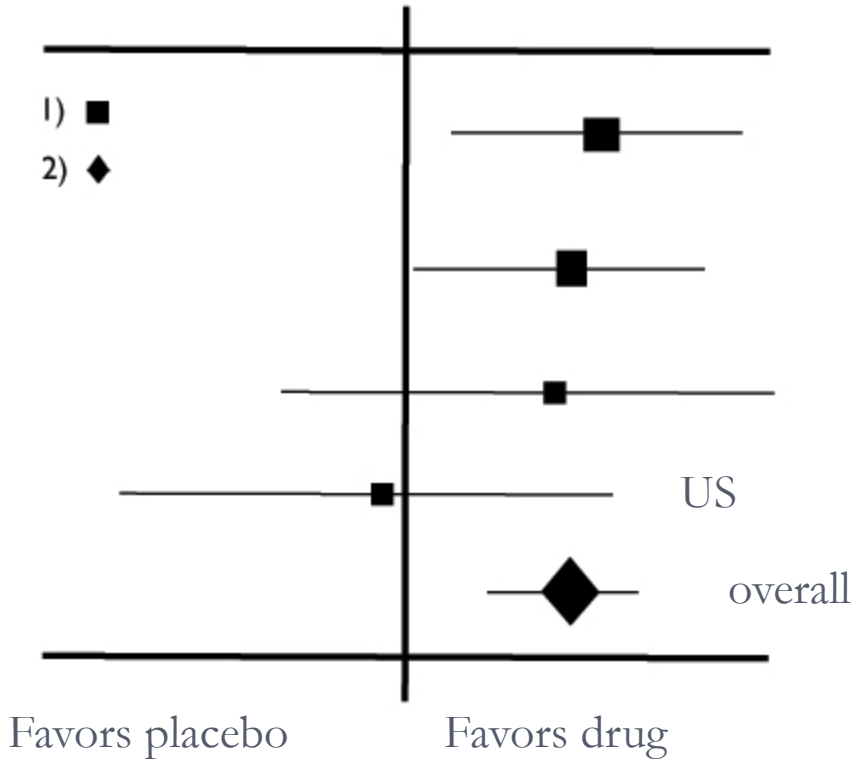
# What we already know: disease factors

- Genetic diseases: different genotypes
- Infectious disease: different organisms
- Chronic disease: different stage of disease
  - cervical cancer in India vs. US
  - invasive breast cancer in Russia vs US
  - heart failure US+W Europe vs Russia+E Eur

# Why are regional subgroups different from all other subgroups?

- US vs. ROW
  - Different diets
  - Different cultures
  - Different standard of care
- But US is not homogeneous (e pluribus unum)
- And study population is not representative

# Typical forest plot for US and others



# How do we split the world?

- US vs ROW
- US+Canada vs ROW
- US+W Eur+(Israel)+(Australia) vs ROW
- What is "Asia"?
  - Far East (China, Japan, Korea)
  - Subcontinent
  - What about Turkey?
- Africa – does it include the Mahgreb?
- Where does South America go? Mexico?
- Hint of the future of this talk: orthography

# Population factors

- Diet
- Risk factors
  - Smoking
  - Drinking
  - Comorbidities
- Racial (genetic) and ethnic (cultural) differences

# Treatment factors

- Standard of care
  - Time of diagnosis
  - Use of drugs
  - Surgical interventions
- Adherence to protocol

# Reminder of examples: MERIT-HF

- Randomized, double-blind, placebo-controlled trial
  - Symptomatic heart failure
  - Metoprolol (different doses depending on NYHA class)
  - ~2000 participants / group
  - 13 European countries + US
- Co-primary outcomes (time to) - either
  - total mortality OR
  - combined endpoint of total mortality or all-cause hospitalizations
- Randomization February 1997 - April 14, 1998.

# Study stopped at 2nd interim analysis

- 50% information; $p<0.001$
- Mean follow-up: 1 year
- Deaths
  - Metoprolol: 145
  - Placebo: 217
  - Relative risk: 0.66
  - 95% CL: (0.53, 0.81)

# MERIT-HF

| Region | Relative Risk | 95% CI |
|---|---|---|
| **Overall** | **0.66** | **(0.53, 0.81)** |
| USA | 1.05 | (0.71, 1.56) |
| Ex-US | 0.55 | (0.43, 0.70) |
| Interaction p-value: 0.003 | | |

# *FDA statistical review – May 30, 2000*

If the mortality endpoint is the most important among all endpoints, the US sub-population should be the most important subgroup in a multinational trial because the goal of the NDA submission is to gain approval for marketing the drug in the US. The efficacy outcome in this population must be examined carefully as part of the evaluation of the totality of the evidence and possible extrapolation of the efficacy evidence from foreign population[s] to [the] US population.

# Belimumab for lupus

**Response Rate**

**(SELENA-SLEDAI improvement 4 or more points, no clinically significant worsening in BILAG or Physician's Global)**

| Placebo | Hi dose |
|---|---|
| (N=275) | (N=273) |
| 34% | 43% |

*FDA briefing document 19-Oct-10*

# Belimumab for lupus

| | | Placebo | Hi dose |
|---|---|---|---|
| | | (N=275) | (N=273) |
| **Overall** | | **34%** | **43%** |
| USA/Can | (300) | 32% | 35% |
| W Eur/Israel | (200) | 23% | 51% |
| E. Eur | ( 60) | 42% | 53% |
| LA/SA | ( 60) | 57% | 53% |

*FDA briefing document 19-Oct-10*

# The PROTECT Study

- Rolofylline+placebo both + loop diuretic
- Heart failure signs and symptoms

# Design (hospitalized heart failure)

- 600 patients (later 2000)
- 2:1 active to placebo
- 75 sites (US, Israel, E and W Europe, Russia)
- 3 day infusion of drug or placebo
- Outcome is day 2 and 3
  - Other measures at Day 7, 60, and 180
  - Study ends when last patient has 60 days f-up

# Primary outcome: 3 category variable

- Failure
  - Worsening symptoms
  - Death, hospital readmission, or other bad things
- No change – not a success or a failure
- Success
  - Dyspnea Day 2 & 3 moderately or markedly better
  - Not a treatment failure

# Safety concern

- Drug is an adenosine A1 receptor agonist
- Known to lower seizure threshold

# DMC's concern: so few from US

- DMC to Sponsor:
  - ~2/3 of participants from Europe, Israel, Russia
- Sponsor: healthcare in Israel like that in US
  - Therefore, ½ are from "US-like" countries
- DMC: what percentage should be US-like?
- Sponsor: No specific requirement; hope ~40%

# May 16, 2008: DSMB meeting

| | Placebo | Rolo |
|---|---|---|
| N | 343 | 694 |
| With data | 250 | 507 |
| Deaths | 32 | 60 |
| Seizures | 0 | 4 |
| | | |
| Success | | |
| Day 3 | 34% | 36% |
| Day 14 | 54% | 58% |

# Demographics

- Russia 29%
- USA 15%
- ROW 56%

# What the DSMB saw: Russia vs ROW

| | Russia | ROW | Comment |
|---|---|---|---|
| Mean age | 68 | 70 | ok |
| % male | 58% | 69% | ? |
| Mean weight (kg) | 84 | 81 | odd |
| Hypertension | 89% | 75% | oops |
| Diabetes | 20% | 45% | ? |

# What the DSMB saw: Russia vs ROW

| | Russia | ROW | Comment |
|---|---|---|---|
| Mean age | 68 | 70 | ok |
| % male | 58% | 69% | ? |
| Mean weight (kg) | 84 | 81 | odd |
| Hypertension | 89% | 75% | oops |
| Diabetes | 20% | 45% | ? |
| Class III/IV HF | 100% | 73% | ! |
| Class IV HF | 85% | 21% | !!!! |

# What the DSMB saw: Russia vs ROW

| | Russia | ROW | Comment |
|---|---|---|---|
| Mean age | 68 | 70 | ok |
| % male | 58% | 69% | ? |
| Mean weight (kg) | 84 | 81 | odd |
| Hypertension | 89% | 75% | oops |
| Diabetes | 20% | 45% | ? |
| Class III/IV HF | 100% | 73% | ! |
| Class IV HF | 85% | 21% | !!!! |

So was the trial studying Russia vs. ROW or Class IV HF vs other?

# Post-script: Outcomes – Treatment success

| | Russia | ROW |
|---|---|---|
| • At May DSMB meeting | | |
| • Placebo | 34% | 37% |
| • Rolo | 53% | 38% |
| • End of study | | |
| • Placebo | 31% | 37% |
| • Rolo | 53% | 39% |

# DMC should routinely look at region

- But: look should be careful
  - Are the patients different in meaningful ways?
- If PROTECT had continued recruiting in Russia
  - And data showed showed benefit overall
  - What could be concluded?
- PROTECT differs from MERIT-HF and Benlysta
  - There, the effect looked like chance
  - Here it looks like confounding

# But what about safety?

- Why don't we look at safety by region?
- The same considerations relative to efficacy are relevant to safety
  - Different underlying disease states
  - Different standard of care
  - Different cultures

# My alphabet hypothesis

The more different from a Roman alphabet,
the less accurate the safety data

# English

| Original | Translation |
| --- | --- |
| Lorem ipsum dolor sit amet, consectetur adipiscing elit. Praesent pulvinar sed quam ultricies lacinia. Aenean ullamcorper purus ac purus laoreet vestibulum. | Lorem ipsum dolor sit amet, consectetur adipiscing elit. Praesent pulvinar sed quam ultricies lacinia. Aenean ullamcorper purus ac purus laoreet vestibulum. |

# English and other Roman alphabet

| Original | Translation |
| --- | --- |
| Lorem ipsum dolor sit amet, consectetur adipiscing elit. Praesent pulvinar sed quam ultricies lacinia. Aenean ullamcorper purus ac purus laoreet vestibulum. | Lorem ipsum dolor sit amet, consectetur adipiscing elit. Praesent pulvinar sed quam ultricies lacinia. Aenean ullamcorper purus ac purus laoreet vestibulum. |
| és un text de farciment usat per la indústria de la tipografia i la impremta. Lorem Ipsum ha estat el text estàndard de la indústria des de l'any 1500, quan un impressor | Lorem ipsum dolor sit amet, consectetur adipiscing elit. Praesent pulvinar sed quam ultricies lacinia. Aenean ullamcorper purus |

# Non-Roman alphabet

| Original | Translation |
| --- | --- |
| Lorem ipsum dolor sit amet, consectetur adipiscing elit. Praesent pulvinar sed quam ultricies lacinia. Aenean ullamcorper purus ac purus laoreet vestibulum. | Lorem ipsum dolor sit amet, consectetur adipiscing elit. Praesent pulvinar sed quam ultricies lacinia. Aenean ullamcorper purus ac purus laoreet vestibulum. |
| és un text de farciment usat per la indústria de la tipografia i la impremta. Lorem Ipsum ha estat el text estàndard de la indústria des de l'any 1500, quan un impressor | Lorem ipsum dolor sit amet, consectetur adipiscing elit. Praesent pulvinar sed quam ultricies lacinia. Aenean ullamcorper purus |
| είναι απλά ένα κείμενο χωρίς νόημα για τους επαγγελματίες της τυπογραφίας και στοιχειοθεσίας. είναι επαγγελματικό πρότυπο όσον αφορά το κείμενο χωρίς νόημα, από τον 150 αιώνα, όταν ένας | Lorem ipsum dolor sit amet, consectetur adipiscing elit. Praesent pulvinar sed quam ultricies |

# Non-alphabetic language

| Original | Translation |
| --- | --- |
| Lorem ipsum dolor sit amet, consectetur adipiscing elit. Praesent pulvinar sed quam ultricies lacinia. Aenean ullamcorper purus ac purus laoreet vestibulum. | Lorem ipsum dolor sit amet, consectetur adipiscing elit. Praesent pulvinar sed quam ultricies lacinia. Aenean ullamcorper purus ac purus laoreet vestibulum. |
| és un text de farciment usat per la indústria de la tipografia i la impremta. Lorem Ipsum ha estat el text estàndard de la indústria des de l'any 1500, quan un impressor | Lorem ipsum dolor sit amet, consectetur adipiscing elit. Praesent pulvinar sed quam ultricies lacinia. Aenean ullamcorper purus |
| είναι απλά ένα κείμενο χωρίς νόημα για τους επαγγελματίες της τυπογραφίας και στοιχειοθεσίας. είναι επαγγελματικό πρότυπο όσον αφορά το κείμενο χωρίς νόημα, από τον 150 αιώνα, όταν ένας | Lorem ipsum dolor sit amet, consectetur adipiscing elit. Praesent pulvinar sed quam ultricies |
| 也称乱数假文或者哑元文本， 是印刷及排版领域所常用的虚拟文字。由于曾经一台匿名的打印机刻意打乱了也称乱数假文或者哑元文本， 是印刷及排版领域所常用的虚拟文字。由于曾经 | Lorem ipsum dolor |

# My unscientific conclusion+new hypothesis

- Conclusion: Complexity of events understated in many countries
  - Complexity depends at least partially on orthography
- New hypothesis: If translations are bad, there are other problems
  - Look at SAE reporting rates by country

# Reporting rates: tale of three typical studies

Serious adverse event rates/100 patient-years

| Results | Secondary prevention | Primary prevention | Very sick population |
|---|---|---|---|
| **Overall** | 20 | 15 | 30 |

# Reporting rate /100 person years

| Region | Sample sizes: 1000-4000 | 2000-5000 | 100-200 |
|---|---|---|---|
| **Overall** | 20 | 15 | 30 |
| Asia | 25 | 16 | - |
| Aust/NZ/SA | 25 | - | - |
| Western Europe | 25 | 20 | 35 |

# Reporting rate /100 person years

| Region | Sample sizes: 1000-4000 | 2000-5000 | 100-200 |
|---|---|---|---|
| **Overall** | 20 | 15 | 30 |
| Asia | 25 | 16 | - |
| Aust/NZ/SA | 25 | - | - |
| Western Europe | 25 | 20 | 35 |
| North America | 30 | 30 | 40 |

So why are overall rates so low?

# Reporting rate /100 person years

| Region | Sample sizes: 1000-4000 | 2000-5000 | 100-200 |
|---|---|---|---|
| **Overall** | 20 | 15 | 30 |
| Asia | 25 | 16 | - |
| Aust/NZ/SA | 25 | - | - |
| **Eastern Europe** | **20** | **10** | **-** |
| Western Europe | 25 | 20 | 35 |
| North America | 30 | 30 | 40 |
| **South America** | **-** | **10** | **10** |
| **Russia/Ukraine** | **10** | **-** | **-** |

# What about China

- Structure
  - In many trials, China enters the trial late
  - Therefore, follow-up is shorter
  - Higher percentage of prevalent cases
  - ….
- Efficacy – in time-to-event event-driven study
  - Less apparent efficacy
- Safety - ?

# Consequences to label

- Reports overall safety results
- My hypothesis: in multinational trials, this understates US rates

# What to do?

- During trial
  - Sponsors should look at SAE rates per study
  - Query clinical sites if the variability is too high
- At end of trial – report SAE rates
  - by country
  - (and orthography if you don't think I'm nuts)
- Label: ?