

Bias in Clinical Trials

Talat Ashraf MD, MS

Executive Director

Astex Pharmaceuticals

October 20th 2017

Disclosure

- **I work for Astex Pharmaceuticals which is currently developing drugs in hematological malignancies and oncology**
- **The opinion expressed here are of Talat Ashraf MD, MS and not of Astex Pharmaceuticals**

Talat Ashraf MD, MS

Speakers Bio

- **I am an Executive Director at Astex Pharmaceuticals and serve as head of drug safety and medical monitor for clinical trials for FDA and global regulatory approval since 2013**
- **Prior to joining Astex I was Head of Global Clinical and Regulatory Strategy for Vertex Pharmaceuticals for a year**
- **Previously I have worked as Global Medical Director in clinical development and medical affairs for Medtronic, Johnson and Johnson and Abbott for total of 15 years**
- **I have a (MS) Masters of Science from the University of Cincinnati, Ohio in Health Planning/Administration and Economics**
- **I have a MD from Quaid-e-Azam Medical, Pakistan and was on the faculty as Clinical Instructor (Medicine and Community Medicine) at the Department of Community Health Science, Aga Khan University & Hospital, Karachi, Pakistan for 4 years**

Learning Objectives

- **Identify clinical study bias**
- **Design studies with minimization of bias**
- **Provide remedies to fix bias**

What is a Bias?

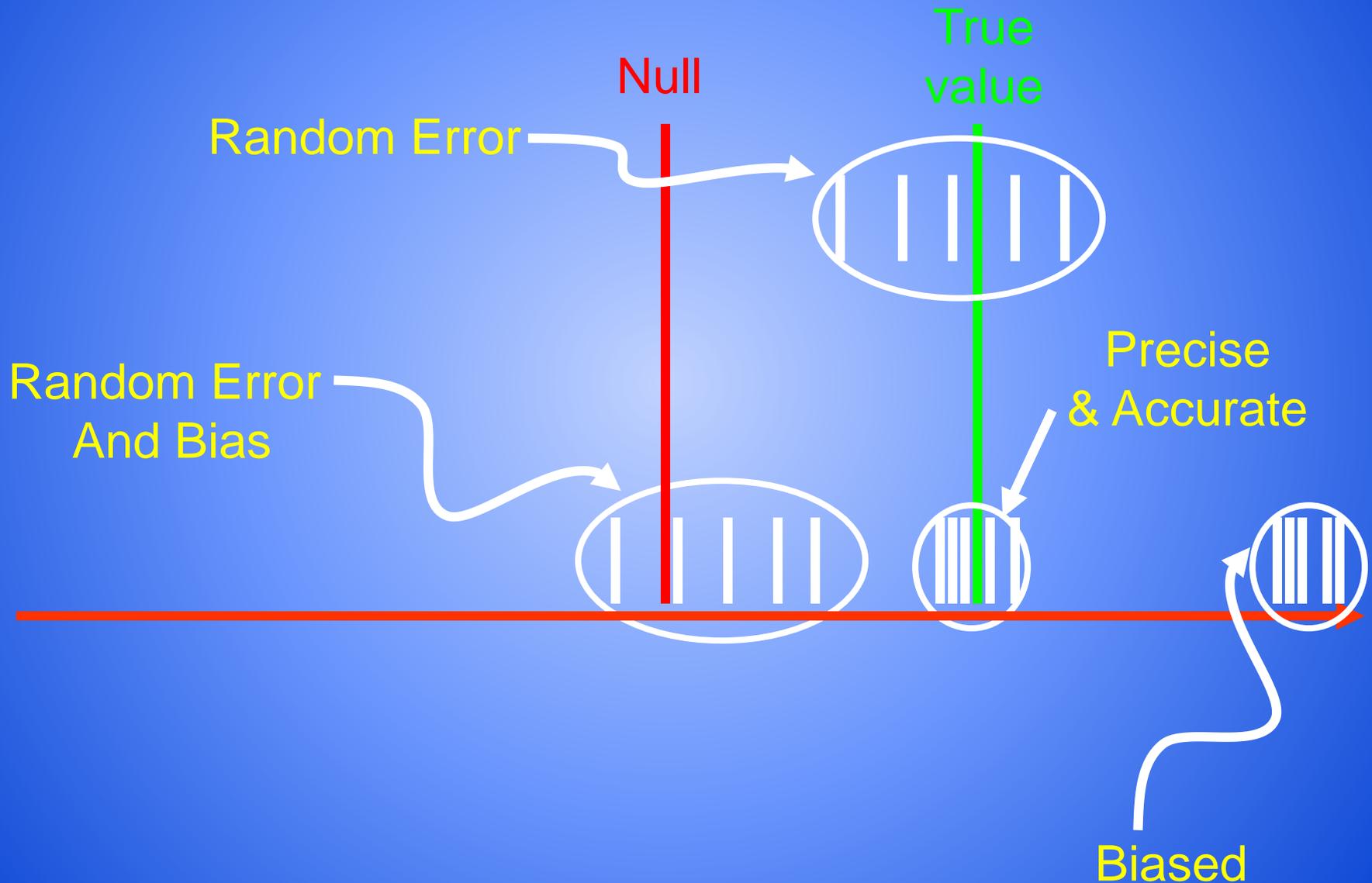
A systematic error (caused by the investigator or the subjects) that causes an incorrect (over- or under-) estimate of an association.

Type 1 Error- False Positive

Type 2 Error - False Negative

Table of error types		Null hypothesis (H_0) is	
		True	False
Decision About Null Hypothesis (H_0)	Reject	Type I error (False Positive)	Correct inference (True Positive)
	Fail to reject	Correct inference (True Negative)	Type II error (False Negative)

Suppose a study was conducted multiple times in an identical way.



What does this error do to the study?

- **Bias affects the Internal Validity;**
 - the rigor with which the study was conducted; the study's design, the care taken to conduct measurements, and decisions concerning what was and wasn't measured
- **Validity refers to the degree to which a study accurately reflects or assesses the specific concept that the researcher is attempting to measure**
- **External validity refers to the extent to which the results of a study are generalizable or transferable**

Selection Bias

Occurs when selection, enrollment, or continued participation in a study is somehow dependent on the likelihood of having the exposure of interest or the outcome of interest.

Selection bias can cause an overestimate or underestimate of the association.

Randomization

- **Randomization or other similar methods abolishes selection bias**
 - **Simple randomization**
 - **Block randomization**
 - **Stratified randomization**
 - **Covariate adaptive randomization – Minimization**
 - **Pocock and Simon**

<https://www.graphpad.com/quickcalcs/index.cfm>

<http://www.randomization.com/>

After Randomization

- **Once we have randomized participants we eliminate selection bias but the validity of the experiment can be threatened by other forms of bias, which we must guard against**

Forms of Bias After Randomization

David Torgerson, Director York Trial Unit

- **Subversion Bias**
- **Technical Bias**
- **Attrition Bias**
- **Consent Bias**
- **Ascertainment Bias**
- **Dilution Bias**
- **Recruitment Bias**
- **Resentful demoralization**
- **Delay Bias**
- **Chance Bias**
- **Hawthorne effect**
- **Analytical Bias**

Subversion Bias

- **Subversion Bias occurs when a researcher or clinician manipulates participant recruitment such that groups formed at baseline are NOT equivalent**

Subversion - Qualitative Evidence

Schulz JAMA 1995;274:1456.

- **Schulz has described, anecdotally, a number of incidents of researchers subverting allocation by looking at sealed envelopes through x-ray lights.**
- **Researchers have confessed to breaking open filing cabinets to obtain the randomisation code**

Poor concealment

- **Schulz et al. Examined 250 RCTs and classified them into having adequate concealment (where subversion was difficult), unclear, or inadequate where subversion was able to take place.**
- **They found that badly concealed allocation led to increased effect sizes – showing CHEATING by researchers**

Comparison of concealment
Schulz et al. JAMA 1995;273:408.

Allocation Concealment	Effect Size OR	
Adequate	1.0	
Unclear	0.67	P < 0.01
Inadequate	0.59	

Case Study

David Torgerson, Director York Trial Unit

- **One study where it has been reported was for a large, multi-centred surgical trial.**
- **Participants were being randomised to 5+ centres using sealed envelopes**

Mean ages of groups

Clinician	Experimental	Control
All $p < 0.01$	59	63
1 $p = .84$	62	61
2 $p = 0.60$	43	52
3 $p < 0.01$	57	72
4 $p < 0.001$	33	69
5 $p = 0.03$	47	72
Others $p = 0.99$	64	59

Concealment

Hewitt et al. BMJ;2005: March 10th

- **Hewitt and colleagues examined the association between p values and adequate concealment in 4 major medical journals.**
- **Inadequate concealment largely used opaque envelopes.**
- **The average p value for inadequately concealed trials was 0.022 compared with 0.052 for adequate trials (test for difference $p = 0.045$)**
- **False Positive**

Subversion

Brown et al. Stats in Medicine, 2005,24:3715

- **In a survey of 25 researchers 4 admitted to keeping ‘a log’ of previous allocations to try and predict future allocations**

Testing for subversion

- **Comparison of baseline characteristics may help if subversion is suspected. Although this will only identify gross subversion**
- **If blocked allocation is used a statistical test – Bergner-Exner test, may help identify subversion**

Technical Bias

COMET Lancet 2001;358:19

- **This occurs when the allocation system breaks down often due a computer fault.**
- **A great example is the COMET I trial (COMET II was done because COMET 1 suffered bias)**

COMET 1

- A trial of two types of epidural anaesthetics for women in labour
- The trial was using **MIMINISATION** via a computer programme
- The groups were minimised on age of mother and her ethnicity
- Programme had a fault
- **COMET II was done with 1000 women**
- **LESSON – Always check the balance of your groups as you go along if computer allocation is being used**

Minimisation (clinical trials) **Minimisation** is a method of adaptive stratified sampling that is used in clinical trials, as described by Pocock and Simon. The aim of **minimisation** is to **minimise** the imbalance between the number of patients in each treatment group over a number of factors.

COMET I AND COMET II

COMET I

Characteristic	Traditional epidural (n=388)	Combined spinal epidural (n=335)	Low-dose infusion (n=331)
Age			
≤19	1 (0.3%)	69 (21%)	73 (22%)
20–24	12 (3%)	110 (33%)	100 (30%)
25–29	131 (34%)	111 (33%)	117 (35%)
30–34	208 (54%)	17 (5%)	16 (5%)
≥35	36 (9%)	28 (8%)	25 (8%)
Ethnic group			
White	322 (83%)	301 (90%)	298 (90%)
Asian	43 (11%)	24 (7%)	26 (8%)
Other	23 (6%)	10 (3%)	7 (2%)
Mode of delivery			
Spontaneous vaginal	118 (30%)	153 (46%)	159 (48%)
Instrumental vaginal	160 (41%)	107 (32%)	105 (32%)
Caesarean section	110 (28%)	75 (22%)	67 (20%)

< 25 13%

COMET II

	Traditional epidural (n=353)	Combined spinal epidural (n=351)	Low-dose infusion epidural (n=350)
Age (years)			
≤19	52 (15%)	49 (14%)	52 (15%)
20–24	78 (22%)	80 (23%)	78 (22%)
25–29	109 (31%)	107 (31%)	108 (31%)
30–34	82 (23%)	83 (24%)	79 (23%)
≥35	32 (9%)	32 (9%)	33 (9%)
Ethnic origin			
White	302 (86%)	302 (86%)	298 (85%)
Asian	36 (10%)	34 (10%)	38 (11%)
Other	15 (4%)	15 (4%)	14 (4%)
Maternal height (cm, mean [SD])	162.8 (6.7)	162.3 (6.7)	163.4 (7.2)
Maternal weight (kg, mean [SD])	66.5 (13.6)	65.3 (14.3)	67.6 (14.0)
Pregnancy-induced hypertension	45 (13%)	45 (13%)	61 (17%)
Induced labour	153 (43%)	140 (40%)	162 (46%)
Prostaglandins	88 (25%)	81 (23%)	88 (25%)
Oxytocin	120 (34%)	113 (32%)	120 (34%)
Cervical dilation			
≤2 cm	122 (35%)	100 (29%)	102 (29%)
3–5 cm	175 (50%)	192 (55%)	189 (54%)
Pre-block pain VAS score (median [range])	75 (0–100%)	78 (0–100%)	75 (0–100%)
Weeks' gestation*			
≤37	27 (8%)	24 (7%)	25 (7%)
≥41	142 (40.2%)	146 (41.6%)	145 (41.4%)
Birthweight (g, mean [SD])	3363 (542)	3365 (560)	3349 (512)

Values are numbers (%) of women unless otherwise indicated. VAS=visual analogue scale. *Completed weeks.

Table 2: **Characteristics of mothers and neonates (COMET2)**

Delivery	Traditional epidural (n=353)	Combined spinal epidural (n=351)	Low-dose infusion epidural (n=350)
Normal vaginal	124 (35%)	150 (43%)	150 (43%)
Instrumental vaginal	131 (37%)	102 (29%)	98 (28%)
Caesarean section	98 (28%)	99 (28%)	102 (29%)

*p=0.04, 1DF for normal vs other deliveries.

Table 3: **Mode of delivery**

Values are numbers (%). The age-standardised odds ratio for normal vaginal delivery for COMET1 was 0.91 (95% CI 0.62–1.37) for the combined spinal group and 1.05 (0.70–1.61) for the low-dose infusion group, relative to the traditional group; the respective values for the aggregated ratios for COMET1 and COMET2 were 1.20 (0.94–1.54) and 1.27 (0.98–1.61). The COMET1 and COMET2 age-standardised data did not differ (χ^2 test for interaction low-dose infusion p=0.10 and combined spinal epidural p=0.28).

Table 1: **COMET1 data**

COMET 1 – Technical Bias

AGE	Traditional	Combined	Low dose
Total	388	335	331
<25 years	13 (3%)	179 (53%)	173 (52%)
• COMET II	37%	38%	37%

Attrition Bias

David Torgerson, Director York Trial Unit

- **Usually most trials lose participants after randomization. This can cause bias, particularly if attrition differs between groups**
- **If a treatment has side-effects this may make drop outs higher among the less well participants, which can make a treatment appear to be effective when it is not**

Selection bias after randomization

- **Selection bias is avoided if ALL participants who are randomized are completely followed up**
- **Often there is some attrition – after randomization some refuse to continue to take part**
- **Or some may refuse the intervention but can still be tracked – IMPORTANT to distinguish between these**

Ascertainment Bias

- **This occurs when the person reporting the outcome can be biased**
- **A particular problem when outcomes are not 'objective' and there is uncertainty as to whether an event has occurred**
- **Un-blinded pain scores are always different than blinded pain scores**

Resentful Demoralization

- **This can occur when participants are randomized to treatment they do not want**
- **This may lead to them reporting outcomes badly in ‘revenge’**
- **This can lead to bias**

- **One solution is to use a patient preference design where only participants who are ‘indifferent’ to the treatment they receive are allocated**
- **This should remove its effects**

Hawthorne Effect

- **This is an effect that occurs by being part of the study rather than the treatment. Interventions that require more TLC than controls could show an effect due to the TLC than the drug or surgical procedure**
- **Example is Best Supportive Care Design**
- **Placebos largely eliminate this or TLC should be given to controls as well**

Analytical Bias

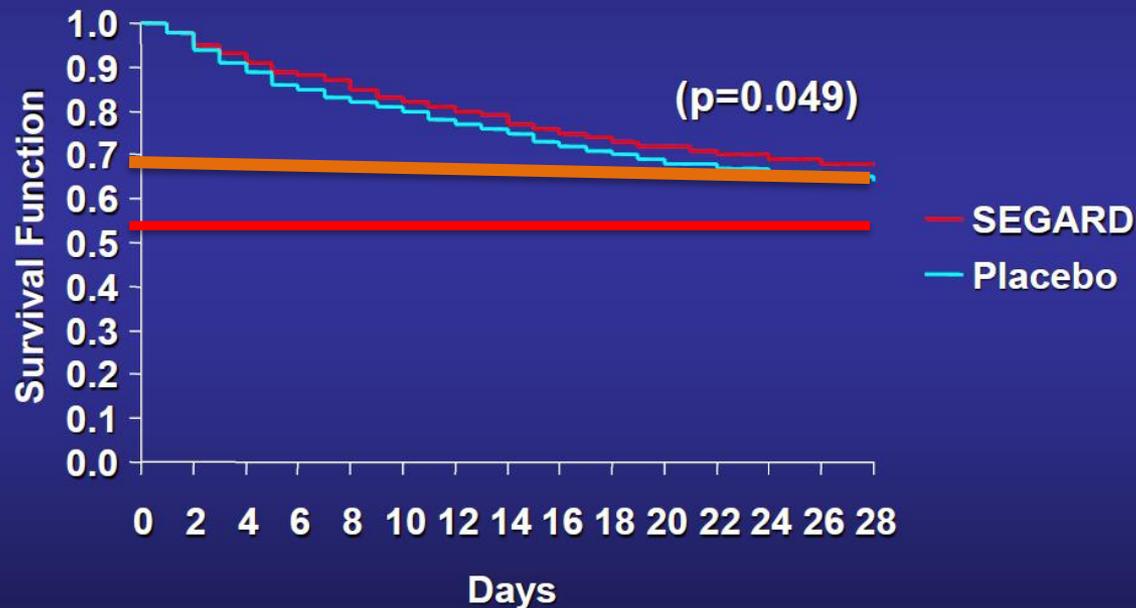
- **Once a trial has been completed and data gathered in it is still possible to arrive at the wrong conclusions by analyzing the data incorrectly**
- **Most IMPORTANT is ITT**
- **Also inappropriate sub-group analyses is a common practice**

- **Example of IL-6 (+) and (-) in Sepsis. Subgroup analysis suggested IL-6 (+) patients treated with Afelimomab has survival benefit**

Analytical Bias

Subgroup Analysis Problem – IL-6

Effects of Neutralization of TNF with MAK 195F on 28 Day All-cause Mortality

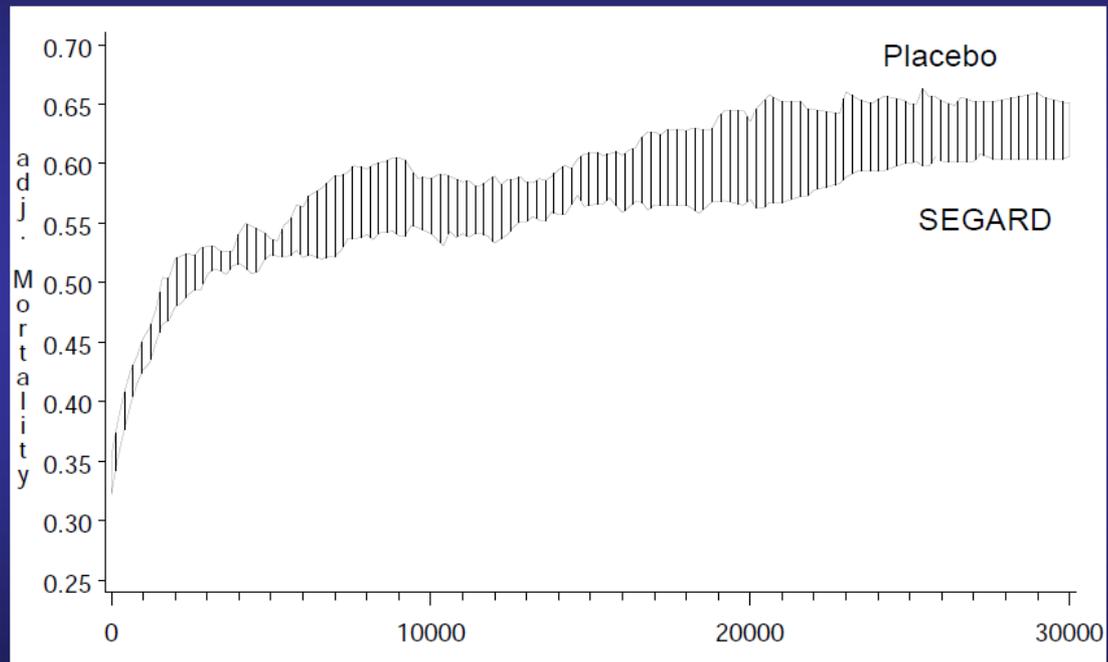


- Panacek, *Crit Care Med* 32:2173, 2004

Analytical Bias

Subgroup Analysis Problem – IL-6

Probability of Mortality by Baseline IL-6 Level



IL-6 Level (pg/ml)

Randomized, placebo-controlled trial of the anti-tumor necrosis factor antibody fragment afelimomab in hyper-inflammatory response during severe sepsis: The RAMSES Study

CRITICAL CARE MEDICINE. 29, 4, p. 765-769 5 p.

- The study was terminated prematurely after an interim analysis estimated that the primary efficacy end points would not be met**
- The 28-day mortality rate in the nonrandomized patients (39.6%, 197 of 498) was significantly lower ($p < .001$) than that found in the randomized patients (55.8%, 249 of 446)**
- The mortality rates in the IL-6 test kit positive patients randomized to afelimomab and placebo were similar, 54.0% (121 of 224) vs, 57.7% (128 of 222), respectively**

Dangers of Sub-Group Analysis

Lancet 1988;ii:349-60.

- In a large RCT of aspirin for myocardial infarction a sub-group analysis showed that people with the star signs Gemini and Libra aspirin was **INEFFECTIVE**
- This is complete **NONSENSE!**
- This shows dangers of subgroup analyses
- Remedy
 - **To avoid spurious findings analysis should be pre-specified and based on a reasonable hypothesis.**

Avoid Data Dredging

- **If you torture the data enough it will confess**

Study Amendment Bias

The tale of Drotrecogin Alfa (Activated)/DAA in Adults with Septic Shock

- **These regulatory decisions was based on the results of the PROWESS trial which demonstrated that treatment with DAA led to a 6.1% absolute risk reduction in 28-day mortality in patients with severe sepsis or septic shock as compared to placebo**
- **PROWESS trial was terminated early as it met a priori stopping criteria for efficacy**
- **A study amendment was made midway through the trial. The amendment included changes in the inclusion and exclusion criteria for study enrollment, a change in the placebo used (saline vs albumin), and changes in the formulation of the study drug**
- **While there was no benefit observed for DAA prior to the study amendment, after the study amendment the results favored the use of DAA – DRUG APPROVED**

Study Amendment Bias

The tale of Drotrecogin Alfa (Activated)/DAA in Adults with Septic Shock

- **In an October 25, 2011 the PROWESS-SHOCK study, which showed the study did not meet the primary endpoint of a statistically significant reduction in 28-day all-cause mortality in patients with septic shock.**
- **The study was published in 2012**

Study Amendment Bias

The tale of Drotrecogin Alfa (Activated)/DAA in Adults with Septic Shock

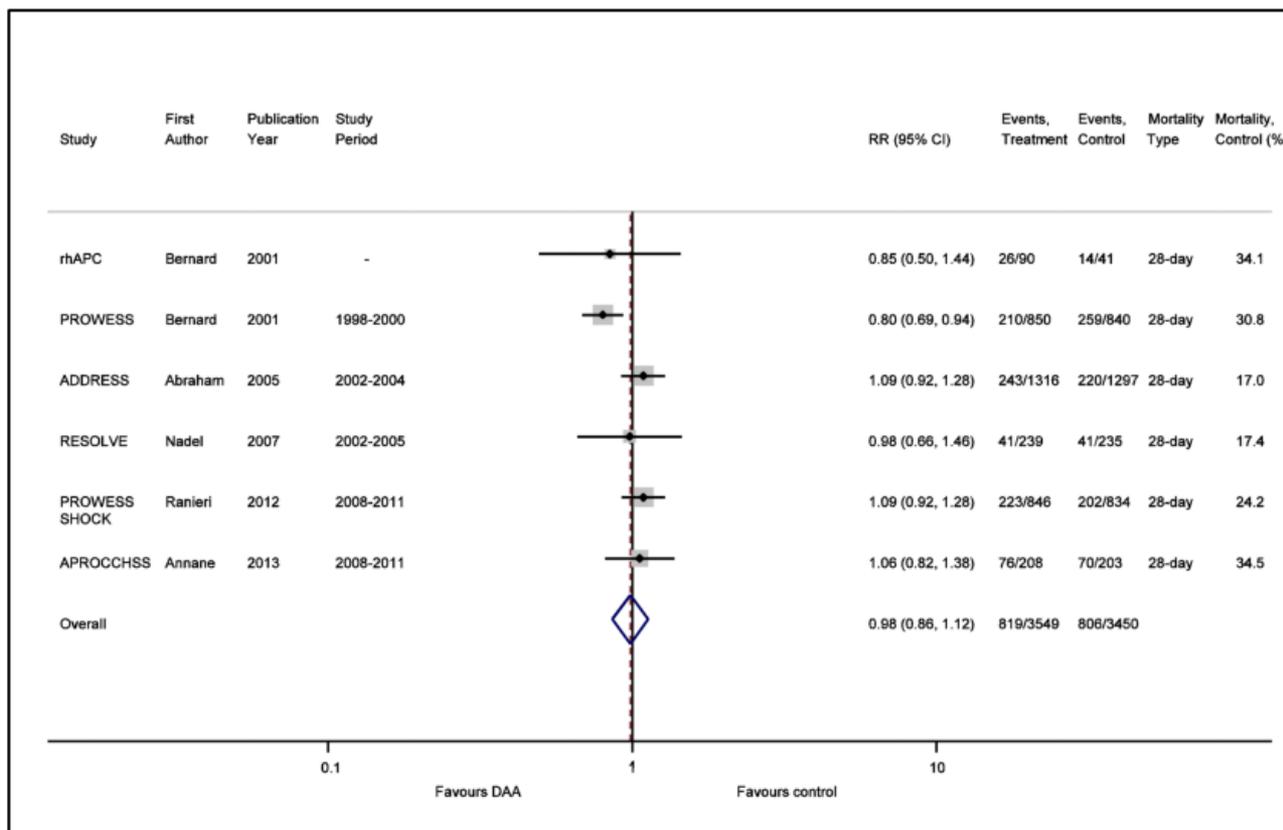
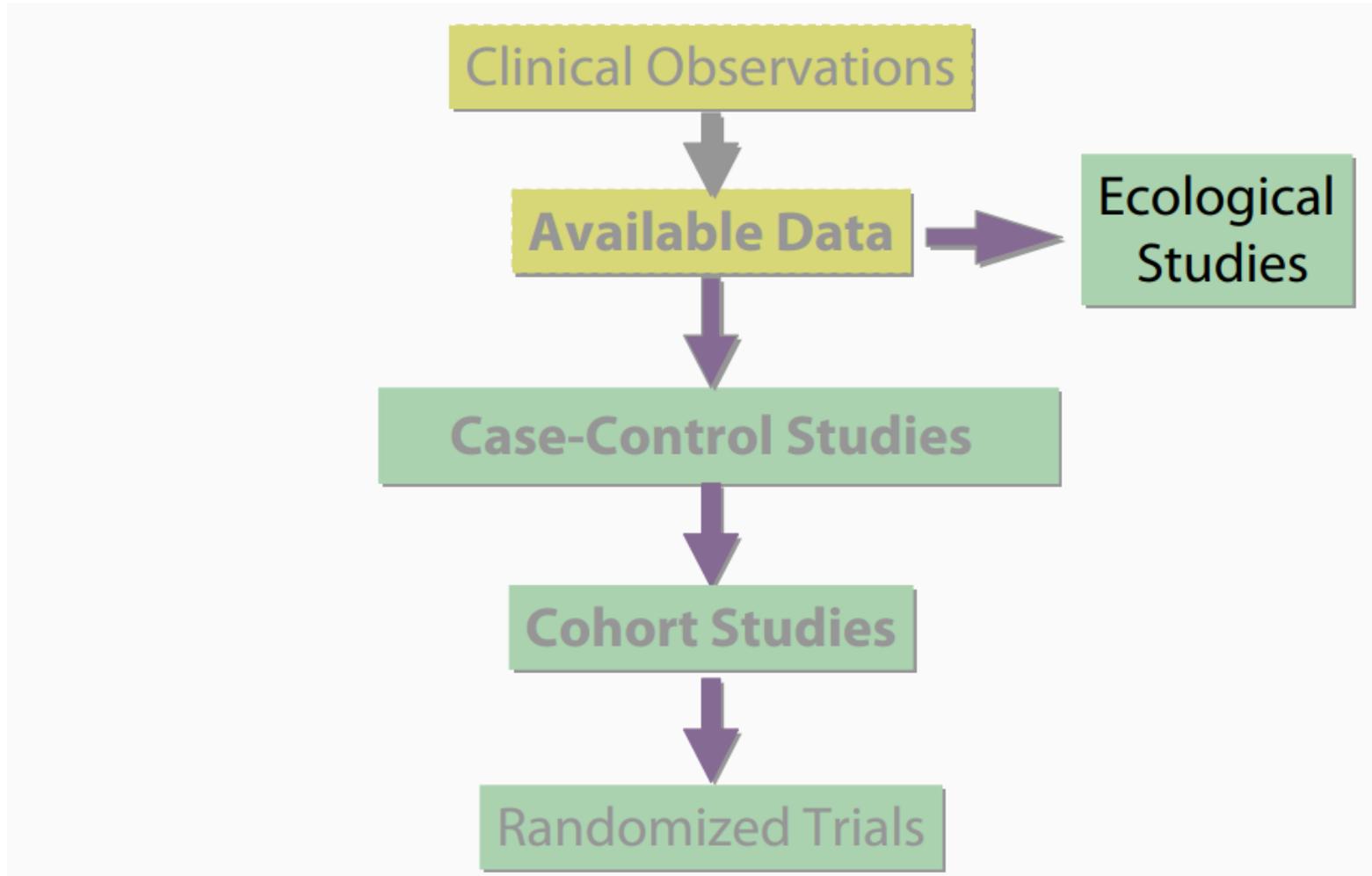


Figure 1. Forest plot comparing the effect of DAA vs. placebo on risk ratio (RR) for 28-day all-cause mortality in all placebo controlled randomized clinical trials of DAA for severe sepsis and septic shock.

Adverse Event Treatment to Influence Outcomes Bias

- **If the end-point is mortality for your Randomized Clinical Trial with no blinding – say in a cancer study**
- **You look at adverse event and find out that rate of febrile events are more on your drug compared to other drugs in the control arm**
- **Further there are more febrile event which are serious on your drug**
- **The protocol clearly states that standard practice of treatment will be used to treat adverse event**
- **By knowing this information you send out a letter or in personal communication with the sites, advise them to prophylactically use antibiotics**

Usual Sequence of Studies in Human Subjects



Why perform observational studies?

Richard Platt, MD, MS – Harvard Medical School

- **Understand experiences of actual users under conditions of actual use – nearly always different from clinical trials**
- **Provide timely information by assessing accumulated experience**
- **Assess very large populations**

Observational vs randomized studies: Differences

- **Randomized:**

- Treated/untreated groups more likely to be comparable;
- Treatment regimen and outcome assessment more certain;
- Risk factor, adherence info often better;
- Can establish causality.

- **Observational:**

- Subjects often more representative;
- Usage conditions usually more typical;
- Larger size/ longer duration possibilities permit observation of rare / delayed outcomes;
- Can establish association.

Types of observational studies

- **Spontaneous reports**
- **Case series**
- **Case-control studies – undefined source populations**
- **Nested case-control studies – well defined source populations**
- **Cohort studies – retrospective**
- **Cohort studies – prospective**

Case-control studies: design

- **Identify cases (outcome has occurred) and non-cases (hasn't occurred)**
- **Assess prior exposures**
- **Nested case-control studies**
 - Cases and controls come from a well-defined population.

Case-control studies: strengths/weaknesses

- **Some strengths relative to cohort:**
 - Efficient – study only cases and a moderate number of controls.
 - Individuals' exposure status can be classified.
- **Some weaknesses:**
 - Cases/controls may not be representative.
 - Knowing the outcome may bias the exposure ascertainment.

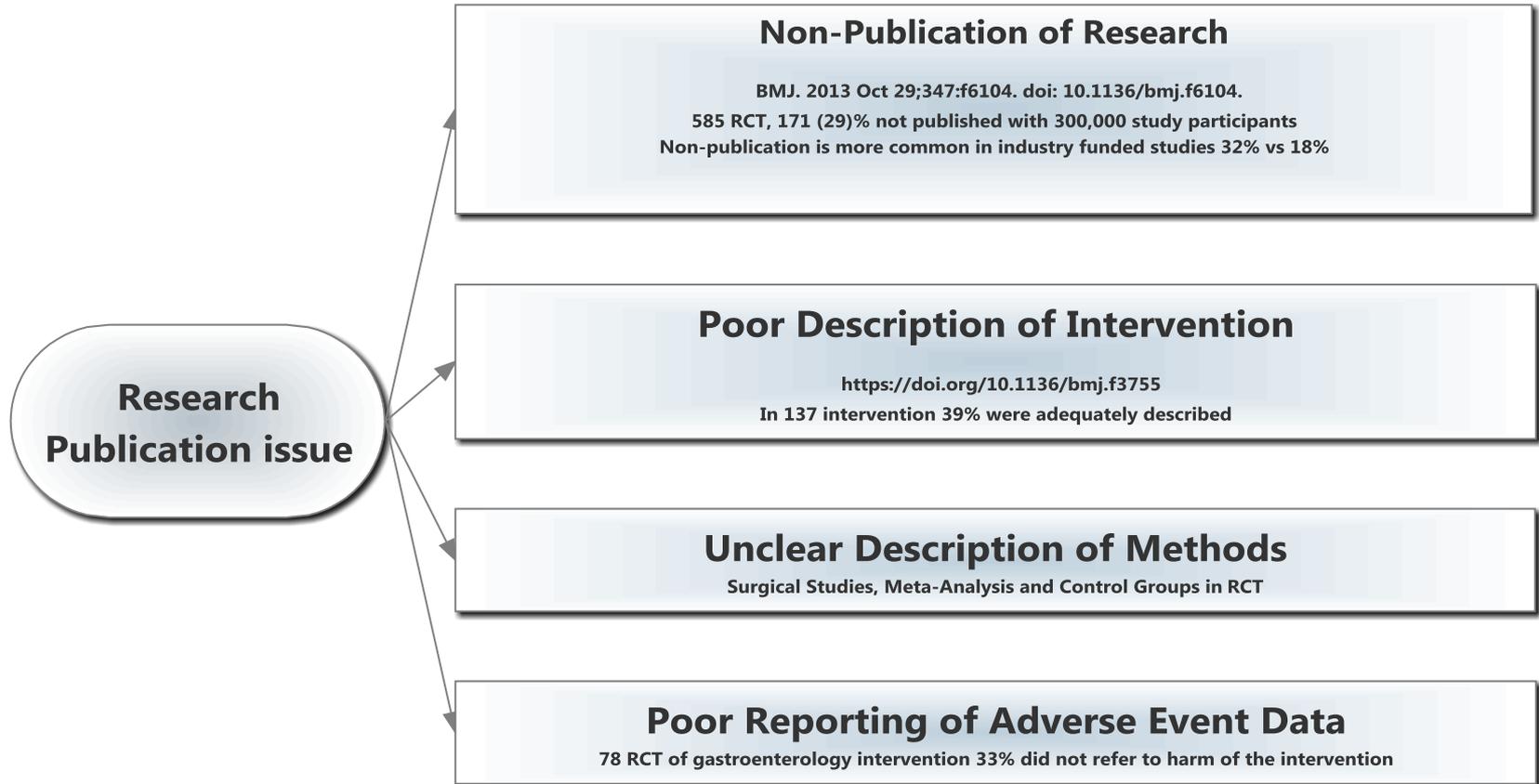
What should you do?

Avoiding Bias in Research

Pearls

- **Blind – Patient, Provider, data clean up people (1) (2) (3)**
- **RDBPC - Randomized double blind placebo control – Gold Standard**
- **Randomize**
- **Before randomization seek patient preference to treatment outlined in the protocol to minimize drop out or revenge bias**
- **Stratify**
- **No changes to inclusion and exclusion criteria during study**
- **Objective vs Non-Objective outcomes – Recall & Measurement Bias**
- **Blind the assessor (Bone Marrow, ECG readers etc.)**
- **Have a plan for lost to follow-ups – Keep it less than 2%**
- **Do not look at patient safety to advise sites to influence outcomes**
- **Don't design studies after data dredging the previous study**
- **Write data analysis plan prior to the initiation of the study – Analytical and Data Interpretation Bias**
- **Have Data Safety Monitoring Committee in place**
- **Limit access to data including safety data**
- **Publish even if the study is negative**

The problem with research literature are?



Summary

- **Despite the RCT being the BEST research method unless expertly used it can lead to biased results**
- **Care must be taken to avoid as many biases as possible**