Hiding safety signals: 5 easy lessons

Janet Wittes
Statistics Collaborative

Enhance benefit; hide harms

- Benefit: We try to enhance signal
  - Large sample size
  - We choose people likely to benefit
  - Composite outcomes to boost event rate

Enhance benefit; hide harms

- Benefit: We try to enhance signal
- Harms: We hide signals
  - Purposively
    - We exclude people likely to be harmed
  - Not purposively
    - Sample size small
    - We avoid composites –in fact, we dice

Examples of harms found late

- Short term studies for long-term use
  - Rofecoxib, celecoxib: *thrombotic events*
  - Diabetes drugs: *cardiovascular mortality*
  - Antipsychotics: *development of diabetes*
- Ignored early signals: ketek, troglitazone - *liver*
- Very rare event
  - Tysabri: *PML* (***progressive multifocal leukoencephalopathy)***
- Never studied: hormones: *heart attack, strokes*

Sample size problem inevitable

True rate	N needed so that E(events) =1
1/10	10
1/100	100
1/1,000	1,000
1/10,000	10,000

Sample size problem inevitable

True rate	N needed so that	
	E(events) =1	Prob(≥1)>0.90
1/10	10	22
1/100	100	230
1/1,000	1,000	2302
1/10,000	10,000	23,025

Suppose you WANT to hide harms

- We will deal with fair tricks
- Examples of unfair tricks

Lousy reports to DMC members

- Reporting statistician
  - Not present
  - Doesn't understand the study
- Reams (literally) of paper
- Most tables have
  - Lots of zeros
  - False precision

And if this isn't enough…

- Change from baseline where missing =0
  - (change in HR=64????)
- Values out of temporal order
- Lots and lots of decimal places
- P-values to 3 and 4 and even 5 significant digits
- Etc., etc., etc.

Another unfair trick: Who is the boss?

- If DMC asks for analysis often the answer is:
  - The program won't let me do that
  - My budget won't allow that

Five Tricks To Hide Harms

1. Animals are not people
2. Respect the investigators
3. Be precise
    1. Show all data
    2. Show only relevant data
4. Consider mechanism
5. Don’t dredge your data

1. Animals are not people

- Animals are very different from us
  - Appeal to comparative anatomy etc.

1. Animals are not people

- Animals are very different from us
  - Appeal to comparative anatomy etc.

1. Animals are not people

- Animals are very different from us
  - Appeal to comparative anatomy etc.

1. Animals are not people

- Animals are very different from us
  - Appeal to comparative anatomy etc.
- If drug ‘passes’ animal test report:
  - “We tested it in Species 1 and 2”

2. Respect the investigators

- They were at the bedside
- So don't change their characterization
  - if they say "acute MI", it is "acute MI"
  - If they say "MI", it is "MI"
  - KEEP THEM SEPARATE
- This will play into Rule #3: be precise
- Don't count deaths that are not on death form

3. Be precise

- Language
  - AMI≠MI
  - Hypertension ≠elevated blood pressure
  - Streptococcal infection ≠bacterial infection ≠infection

Classify precisely

- Misclassification attenuates effect
- Consequence for reporting
  - Put events in correct body system
  - Give precise definitions

e.g. Neuropathy

Event
Anosmia
…..
Autonomic neuropathy
…
Cranial neuropathy
…
…

e.g. Neuropathy

Event
Neuropathic pain
Neuropathy
Neuropathy NOS
Neuropathy peripheral
…
…
…

e.g. Neuropathy

Event
…
Parathesia
Parathesia NOS
Parathesia other
…
Peripheral motor neuropathy
Peripheral sensory neuropathy

And show ALL data

Safety report sample: Patients with abnormal hemoglobin

Time Point	A [N= 150]	B [N= 148]	Total [N= 298]
SCREENING	0	0	0
RANDOM	0	0	0
WEEK 2	0	0	0
WEEK 3	0	0	0
WEEK 4	0	0	0
WEEK 5	0	0	0
WEEK 6	0	0	0
WEEK 7	0	0	0
WEEK 8	0	0	0

But wait! You also get:

Time Point	A [N= 150]	B [N= 148]	Total [N= 298]
SCREENING	0 (0 %)	0 (0 %)	0 (0 %)
RANDOM	0 (0 %)	0 (0 %)	0 (0 %)
WEEK 2	0 (0 %)	0 (0 %)	0 (0 %)
WEEK 3	0 (0 %)	0 (0 %)	0 (0 %)
WEEK 4	0 (0 %)	0 (0 %)	0 (0 %)
WEEK 5	0 (0 %)	0 (0 %)	0 (0 %)
WEEK 6	0 (0 %)	0 (0 %)	0 (0 %)
WEEK 7	0 (0 %)	0 (0 %)	0 (0 %)
WEEK 8	0 (0 %)	0 (0 %)	0 (0 %)

And if you want more precision…

Time Point	A [N= 150]	B [N= 148]	Total [N= 298]
SCREENING	0 (0.00 %)	0 (0.00 %)	0 (0.00 %)
RANDOM	0 (0.00 %)	0 (0.00 %)	0 (0.00 %)
WEEK 2	0 (0.00 %)	0 (0.00 %)	0 (0.00 %)
WEEK 3	0 (0.00 %)	0 (0.00 %)	0 (0.00 %)
WEEK 4	0 (0.00 %)	0 (0.00 %)	0 (0.00 %)
WEEK 5	0 (0.00 %)	0 (0.00 %)	0 (0.00 %)
WEEK 6	0 (0.00 %)	0 (0.00 %)	0 (0.00 %)
WEEK 7	0 (0.00 %)	0 (0.00 %)	0 (0.00 %)
WEEK 8	0 (0.00 %)	0 (0.00 %)	0 (0.00 %)

And the unscheduled

Time Point	A [N= 150]	B [N= 148]	Total [N= 298]
SCREENING	0 (0.00 %)	0 (0.00 %)	0 (0.00 %)
RANDOM	0 (0.00 %)	0 (0.00 %)	0 (0.00 %)
WEEK 2	0 (0.00 %)	0 (0.00 %)	0 (0.00 %)
WEEK 3	0 (0.00 %)	0 (0.00 %)	0 (0.00 %)
WEEK 4	0 (0.00 %)	0 (0.00 %)	0 (0.00 %)
WEEK 5	0 (0.00 %)	0 (0.00 %)	0 (0.00 %)
WEEK 6	0 (0.00 %)	0 (0.00 %)	0 (0.00 %)
WEEK 7	0 (0.00 %)	0 (0.00 %)	0 (0.00 %)
WEEK 8	0 (0.00 %)	0 (0.00 %)	0 (0.00 %)
Unscheduled	3 (1.90 %)	5 (3.38%)	8 (2.68 %)

We get 150 pages of this! (Neat for FDA too)

Time Point	A [N= 150]	B [N= 148]	Total [N= 298]
SCREENING	0 (0.00 %)	0 (0.00 %)	0 (0.00 %)
RANDOM	0 (0.00 %)	0 (0.00 %)	0 (0.00 %)
WEEK 2	0 (0.00 %)	0 (0.00 %)	0 (0.00 %)
WEEK 3	0 (0.00 %)	0 (0.00 %)	0 (0.00 %)
WEEK 4	0 (0.00 %)	0 (0.00 %)	0 (0.00 %)
WEEK 5	0 (0.00 %)	0 (0.00 %)	0 (0.00 %)
WEEK 6	0 (0.00 %)	0 (0.00 %)	0 (0.00 %)
WEEK 7	0 (0.00 %)	0 (0.00 %)	0 (0.00 %)
WEEK 8	0 (0.00 %)	0 (0.00 %)	0 (0.00 %)
Unscheduled	3 (1.90 %)	5 (3.38%)	8 (2.68 %)

Another neat trick…

- This is from the “only relevant data” department
- Report only events that occur in >x%
  - 1%
  - 5%
  - Whatever

Another neat trick…

- This is from the “only relevant data” department
- Report only events that occur in >x%
  - 1%
  - 5%
  - Whatever
- That + dicing effectively hides harms

Examples where this is great

- Liver tox: would hide events in the SOC
  - ie, suppose there are lots of minor liver events
  - 2 bad ones – these don't show up
- GI: Nausea ≠vomiting ≠nausea+ vomiting
- MI ≠acute MI ≠Q-wave MI
- Works really well for lymphomas and ID
  - Great for neurology: neurologists are splitters

4. Consider mechanism

- If you don't get the drug, you can't react to it
  - Modified Daley's Rule: censor early and often
  - But there can be delayed effect
- Appeal to statistical conservatism

5. Don’t dredge your data

- We know that too many tests inflate α level
- So
  - only look at prespecified hypotheses, or
  - Do Bonferroni adjustments (That will absolve all!)

Choose your denominator & live with it

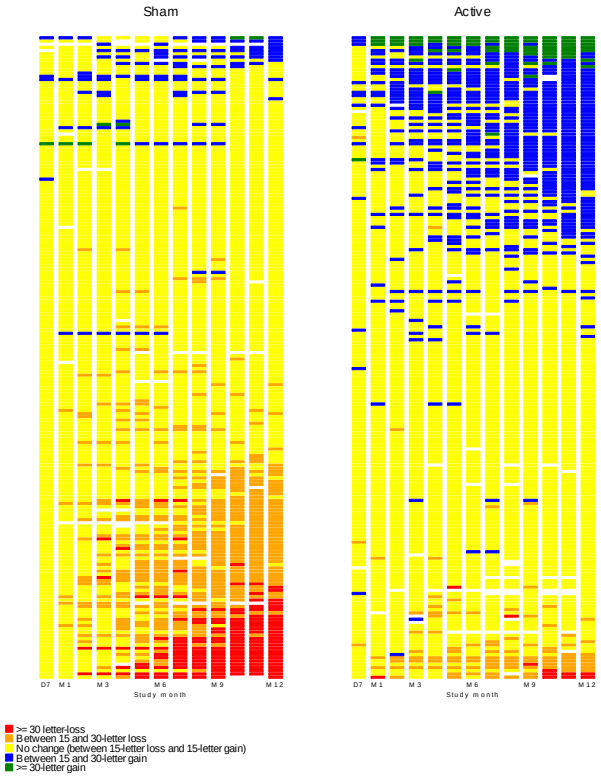
- Corollary of "you must prespecify"
  - Persons at risk?
  - Person-time?
  - Doses?
    - Answer: It depends!!!
    - They are asking different questions

What should be done

- MDs, statisticians, & programmers: no batons

1. Animals ARE like people
2. The investigator is not a researcher
3. Lumping (imprecision) often beats splitting
4. Be agnostic about mechanism
5. Dredge to your heart's content..
    - But don't be afraid of not being certain

Use graphs – e.g., effect over time



What should be done?

- Effort by PHarma should help a lot
- DMCs, regulators, journals should be more critical
  - Data should be complete, but not too complete
- Sponsors (industry/academe) should be more flexible
  - Think biology, not programming
  - Think of telling a story, not just presenting data
- Sensible fixes in data presentation will go a long way!