

# Measuring ROI: This is Our Final Answer

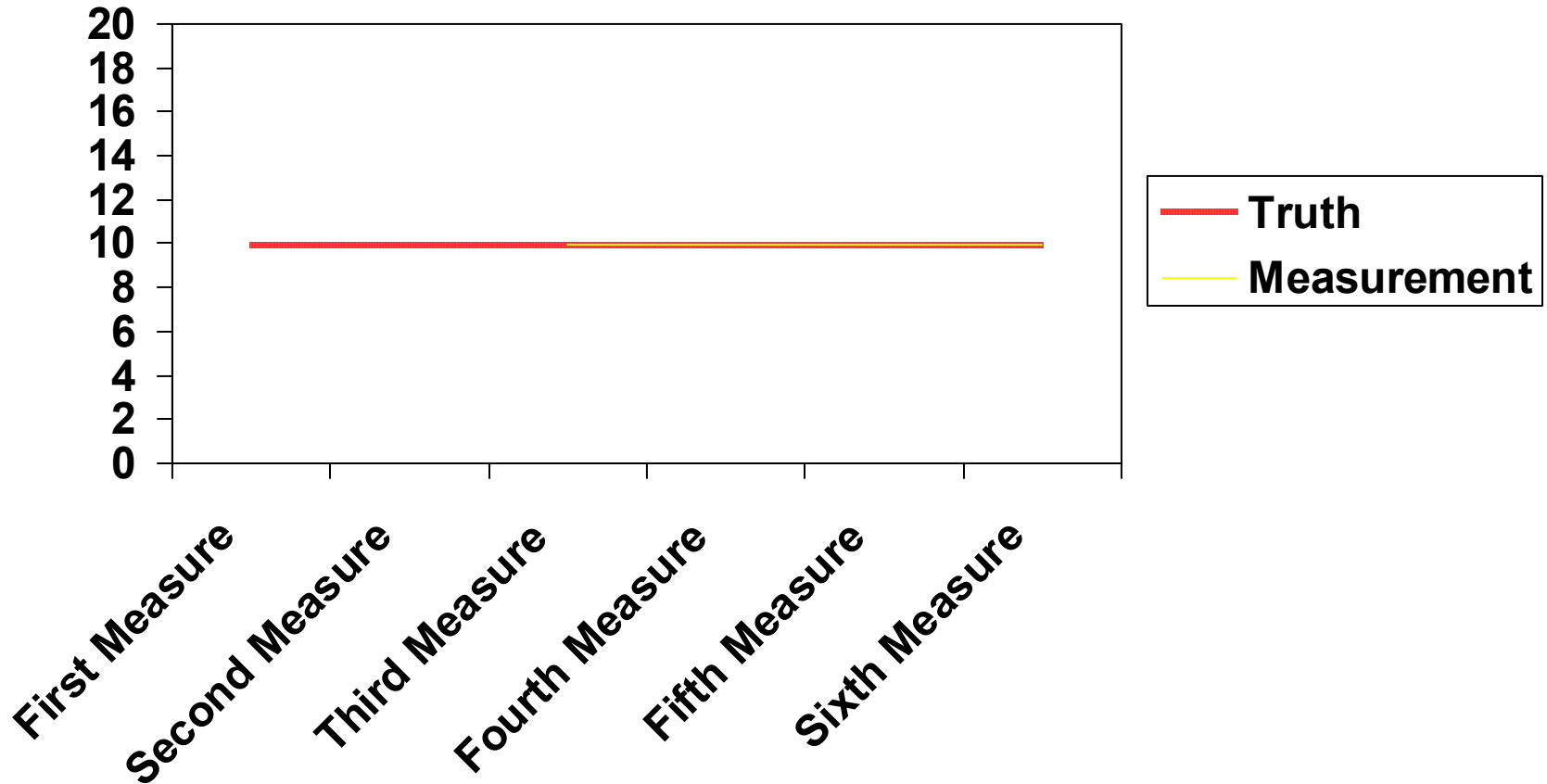
# Today's Agenda

1. Introduction of the first-ever valid pre-post study design in disease management--\$1000 reward if I am wrong (Lewis)
2. Validity  $\neq$  True Accuracy. Next presentation will show how to approach the latter **to improve the former** (Linden)
3. Wilson presentation on the inevitability of not being 100% accurate and needing to focus on probabilistic outcomes

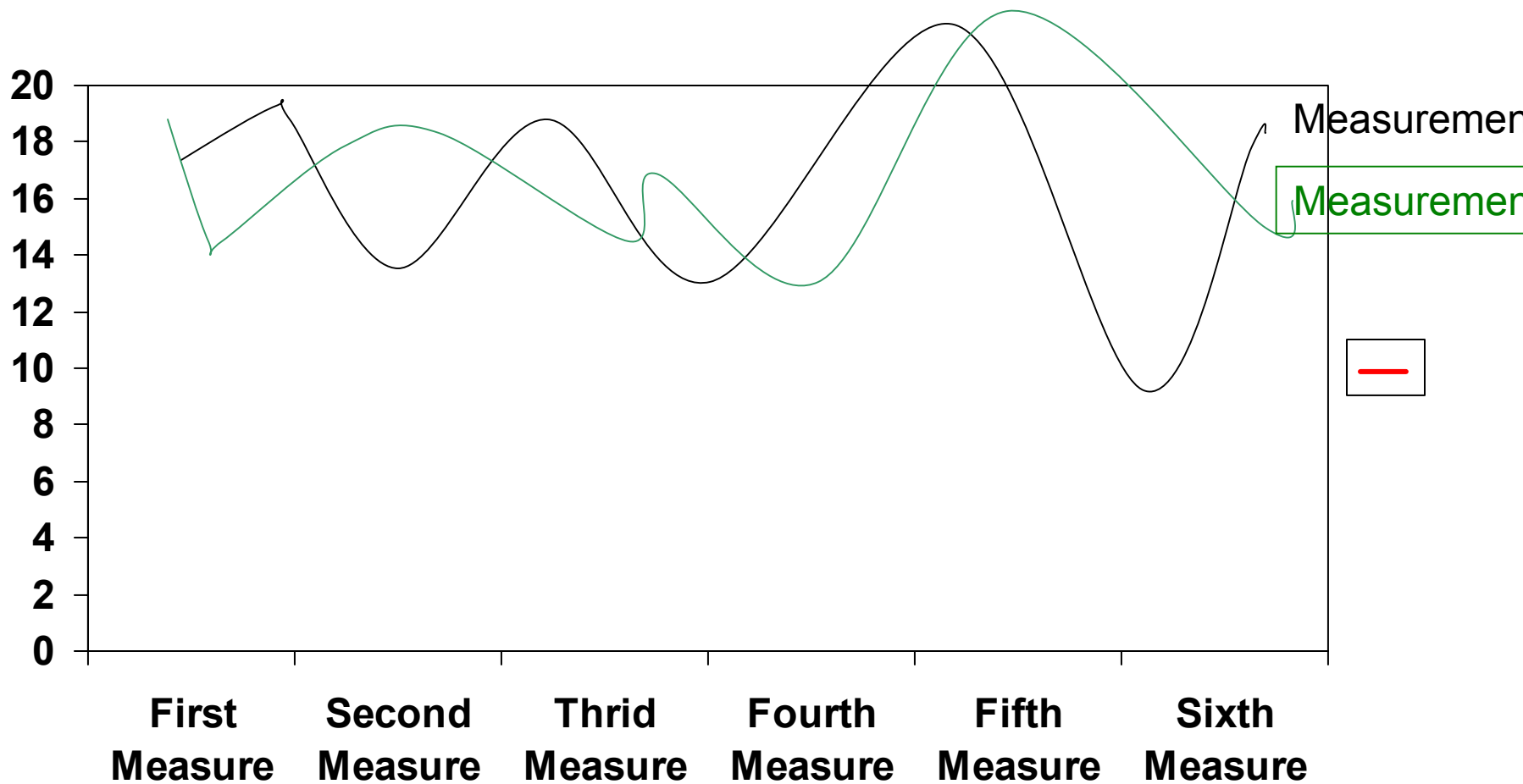
# Validity and Accuracy: Ideally you could measure the true impact

“from bias free of every kind”

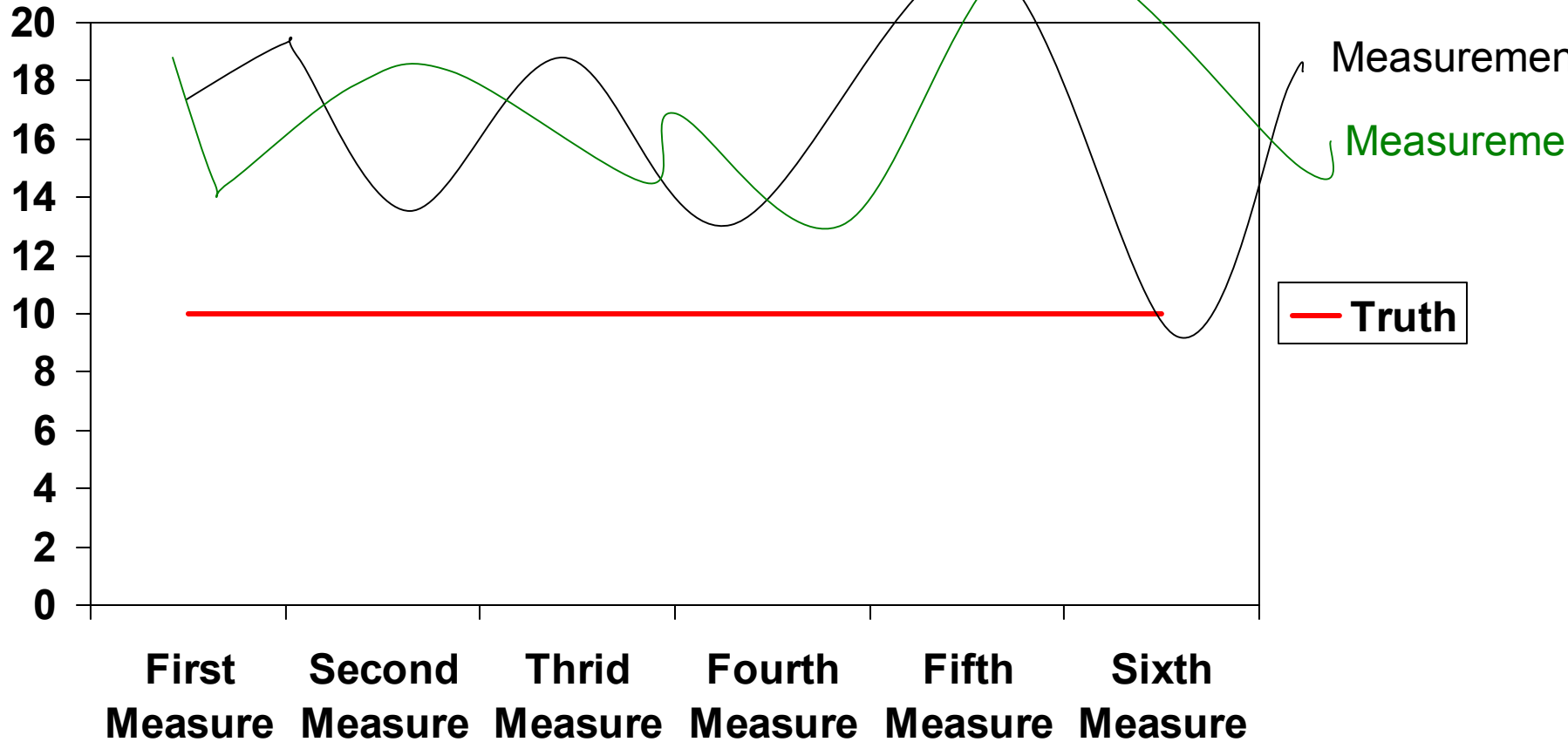
(but if that were the case none of us would be here)



# In reality measurements look more like this

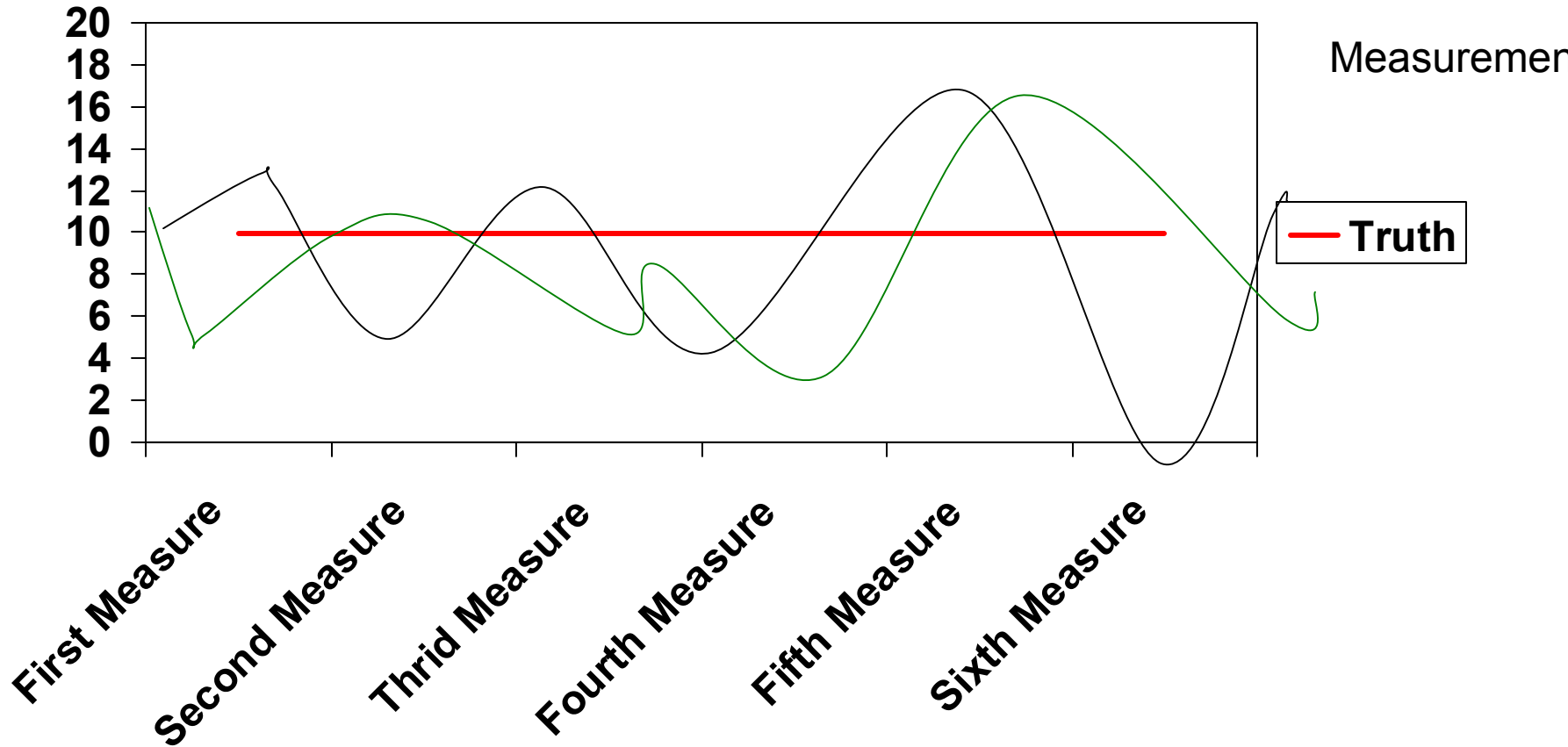


# Validity and Accuracy: Systematic Bias means that the measurements rarely intersect the truth

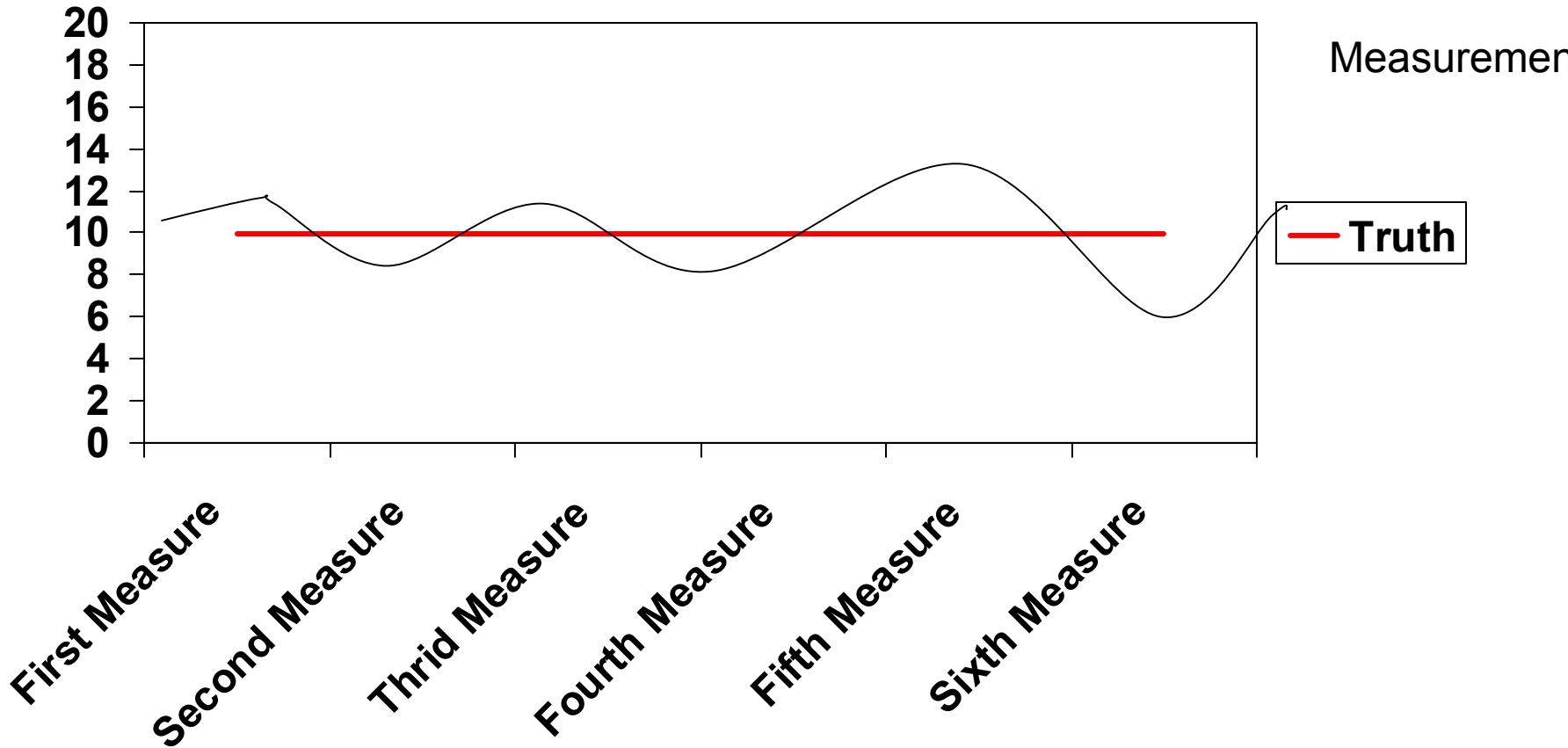


# Validity and Accuracy:

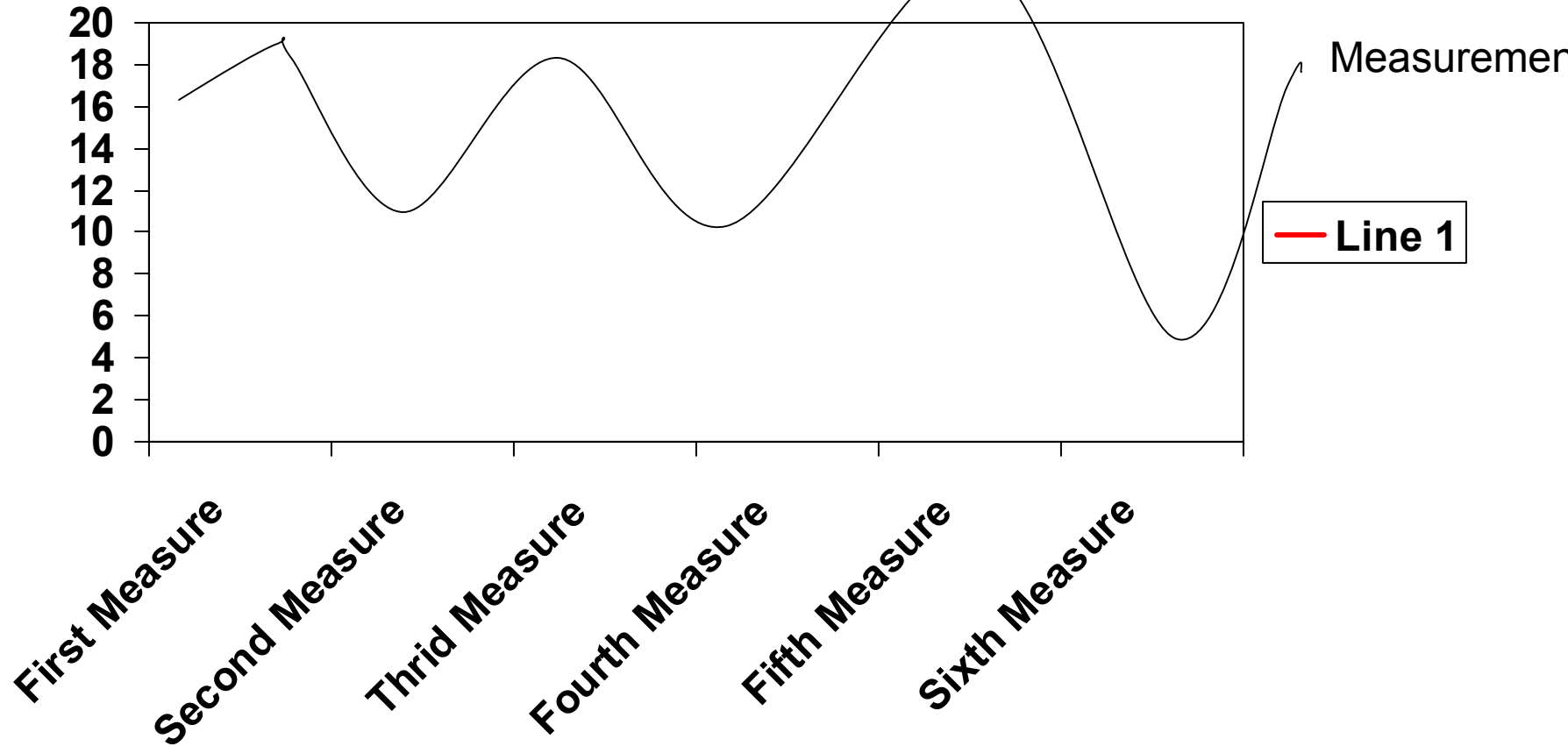
First Presentation shows how to move the random fluctuations so they are around the line of truth



Validity and Accuracy:  
Second Presentation (Linden) shows how to smooth out those fluctuations around that line



# Third presentation shows why these happen based on patterns of individuals and populations (Wilson)



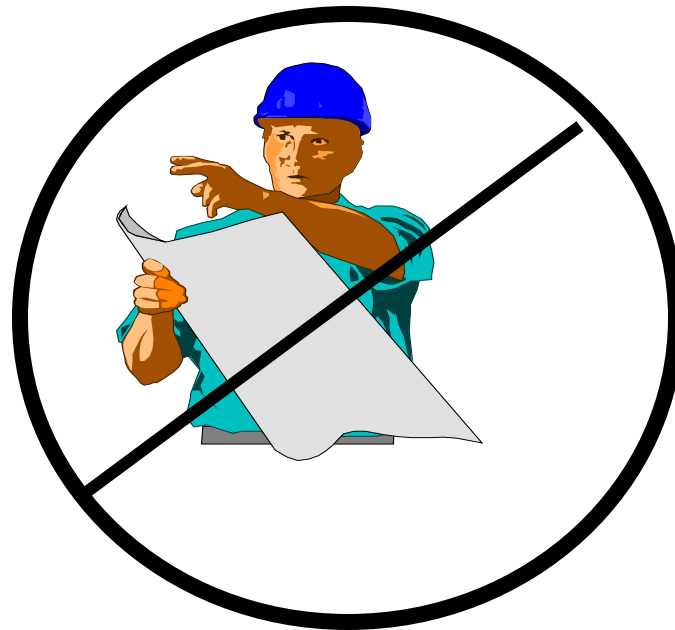


# Approaching Total Accuracy

- Validity (First Presentation)
  - Means if you do this 100 times it will be accurate *in toto*
  - Means all **known** SYSTEMATIC biases are removed (**or accounted for**)
  - Easier to achieve but not certain
- Accuracy (Second Presentation)
  - Means it is close **to** “right” each time
  - Means all **known** NON-SYSTEMATIC biases are addressed too
  - Harder to achieve, certain, requires more analysis **and/or more adjustments**

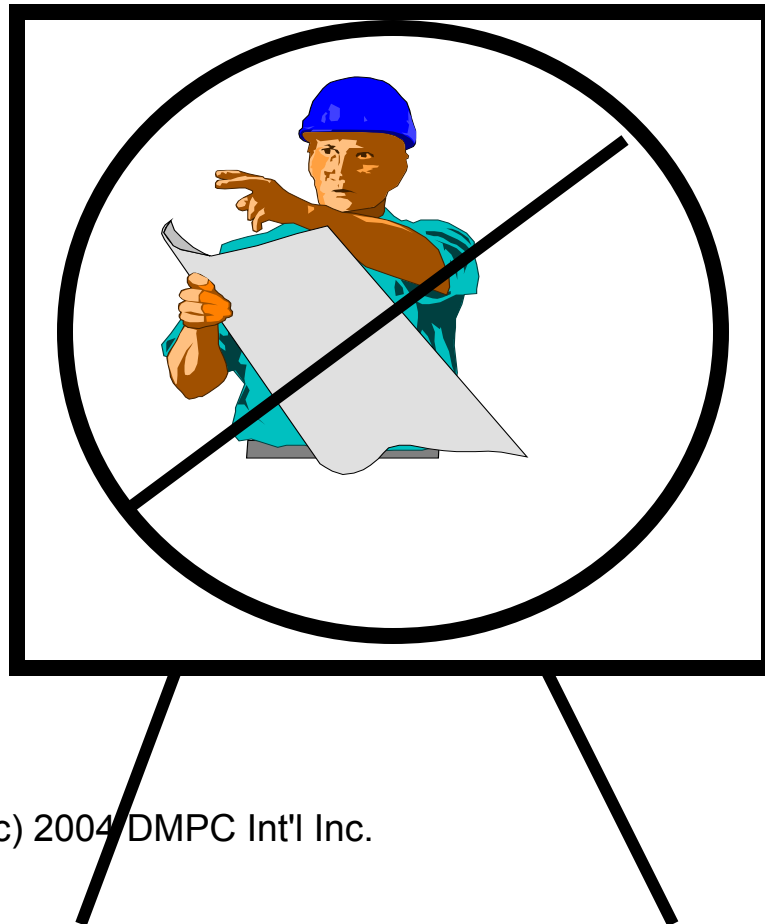
# Warning

- I am not a biostatistician



# Warning

- I don't even play one on TV



# So my goals are to

- Simplify
- Be understandable
- Give you something which is explainable to your CFO in English
- Note that we don't even get to the "data" until well into the workshop...using real data without context is confusing, not illuminating...while also
- ...**Increasing** the validity to highest levels in field

# So my goal is to

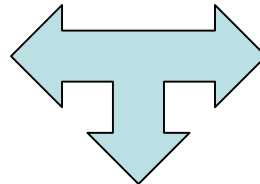
- **Increase** the validity to highest levels in field
- Simplify
- Be understandable
- Give you something which is explainable to your CFO in English
  - Note that we don't even get to the “data” until well into the workshop...using real data without context is confusing, not illuminating



Let's start with a review of the blatantly obvious (to a CFO)

# Your health plan's total medical spending

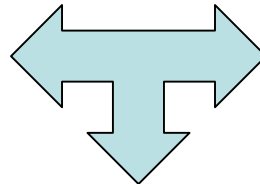
- \$1-billion on 500,000 members
  - 400,000 of which had claims



# Your health plan's medical spending

- \$1-billion on 500,000 members
  - 400,000 of which had claims

$$\begin{aligned} \$1\text{-billion}/500,000 &= \\ & \$2000 \end{aligned}$$



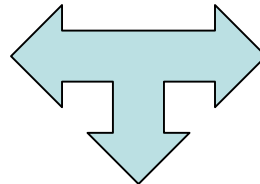
$$\begin{aligned} \$1\text{-billion}/400,000 &= \\ & \$2500 \end{aligned}$$

Which way do you calculate *per capita* spending?

# Your health plan's medical spending

- \$1-billion on 500,000 members
  - 400,000 of which had claims

$$\text{\$1-billion} / 500,000 = \text{\$2000}$$



$$\text{\$1-billion} / 400,000 = \text{\$2500}$$

Which way do you calculate *per capita* spending?

Raise your hand if you think this is blatantly obvious





We will come back to that later...

## Many different ways to measure ROI

- There are several acceptable **population-based** measurement approaches (I prefer Hopkins)
- All have advantages and disadvantages
- All have adherents and detractors

# I really don't have an opinion on how you measure ROI **within reason**

- There are several acceptable **population-based** measurement approaches
- All have advantages and disadvantages
- All have adherents and detractors

There are plenty of non-population-based methodologies which are wrong too

- measuring enrollees against those who declined to enroll
- measuring enrollees against a passive “matched” control group which matches for everything... *except motivation* (if you match for motivation this is an excellent methodology)
- measuring **ONLY** people who had high costs last year

# HOWEVER

- Even the acceptable methodologies end up being wrong because they all overlook the biases created by sentinel events (*even methodologies which purport to include them*)

NONE of them (except a pure passive control/passive study) control for the “Sentinel event”

- The “sentinel event” is the event which tells the health plan that someone has a disease
- It is often the most expensive claim from that member during the first 12 months with the disease
- It is (almost) invariably excluded or included incorrectly...even in methodologies which claim to address it

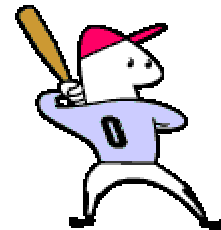
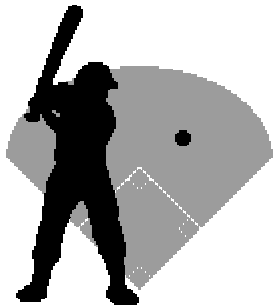
# The Sentinel Event Fallacy Infecting Everyone's Metrics

Presentation will show (using obviously simplifying assumptions):

- **THAT** it happens
- HOW it happens
- WHY it happens
- EXAMPLES from real life
- What to do about it
  - Using simple, understandable, adjustments

# Let's show THAT it happens with baseball

- Analogy that a loss a team has is like a claim for a disease. You are searching your database for people with a disease, called “lossitis”



# Standings after 20 games in '03

Team	Won	Lost	Team	Won	Lost
Yankees	15	5	Red Sox	12	8
Tampa	14	8	Blue Jays	11	9
Baltimore	13	7	White Sox	11	9
Royals	8	12	Cleveland	11	9
Seattle	8	12	Detroit	10	10
Anaheim	7	13	Texas	9	11
Minnesota	7	13	Oakland	7	13

# How to Identify the prevalence of lossitis

- Look for a “claim” for a loss (=\$1000)



# All 14 teams are in the findable lossitis prevalence

Team	Won	Lost	Team	Won	Lost
Yankees	15	5	Red Sox	12	8
Tampa	14	8	Blue Jays	11	9
Baltimore	13	7	White Sox	11	9
Royals	8	12	Cleveland	11	9
Seattle	8	12	Detroit	10	10
Anaheim	7	13	Texas	9	11
Minnesota	7	13	Oakland	7	13

# How to Identify the prevalence of lossitis

- Look for a “claim” for a loss (=\$1000)
  - 14 teams are in the prevalence

# How to identify the cost/person with the disease

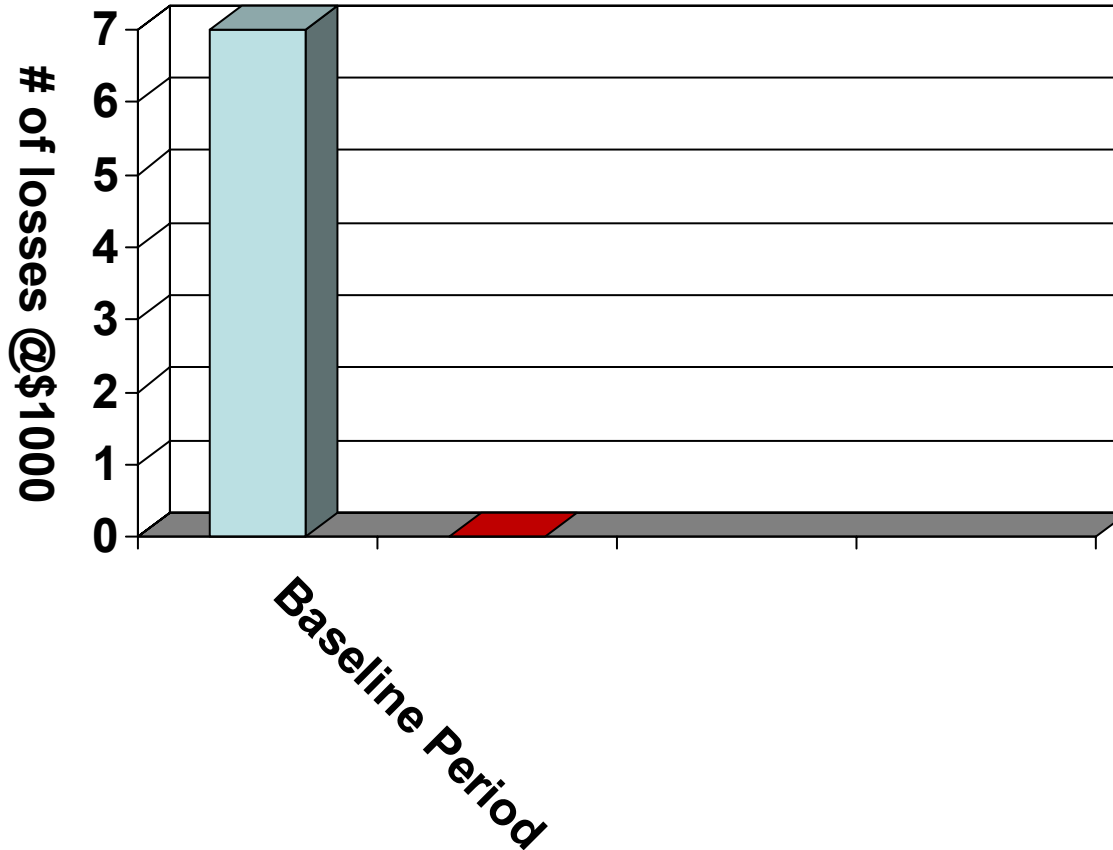
- Look at baseline year claims cost for people with the condition

# Standings after twenty games—identifying who won and lost 20<sup>th</sup> game, the 20<sup>th</sup> period being the “baseline”

Team	Won 20 <sup>th</sup> game		Team		Lost 20 <sup>th</sup> game (baseline claims for lossitis)
Yankees	15	5	Red Sox	12	8
Tampa	14	8	Blue Jays	11	9
Baltimore	13	7	White Sox	11	9
Royals	8	12	Cleveland	11	9
Seattle	8	12	Detroit	10	10
Anaheim	7	13	Texas	9	11
Minnesota	7	13	Oakland	7	13

In the baseline year there were 7 \$1000 claims for lossitis

So the baseline losses are 7 games (\$7000)  
or \$500/team with prevalence (14 teams  
with the prevalence)



# How to Identify the prevalence of lossitis

- Look for a “claim” for a loss (= \$1000)
- All 14 teams have losses so they are all in the prevalence
  - In the baseline period seven teams had \$0 claims and seven had \$1000
    - The “baseline” cost/team was  $\$7000/14$ , or \$500

# Now Apply Disease Management

- Look for a “claim” for a loss (=\$1000)
- All 14 teams have losses so they are all in the prevalence
  - In the baseline period there were seven \$1000 claims among the 14 teams
    - The “baseline” cost/team was \$7000/14, or \$500

- Intervention is rooting real hard
- You root for all the identified teams the next day

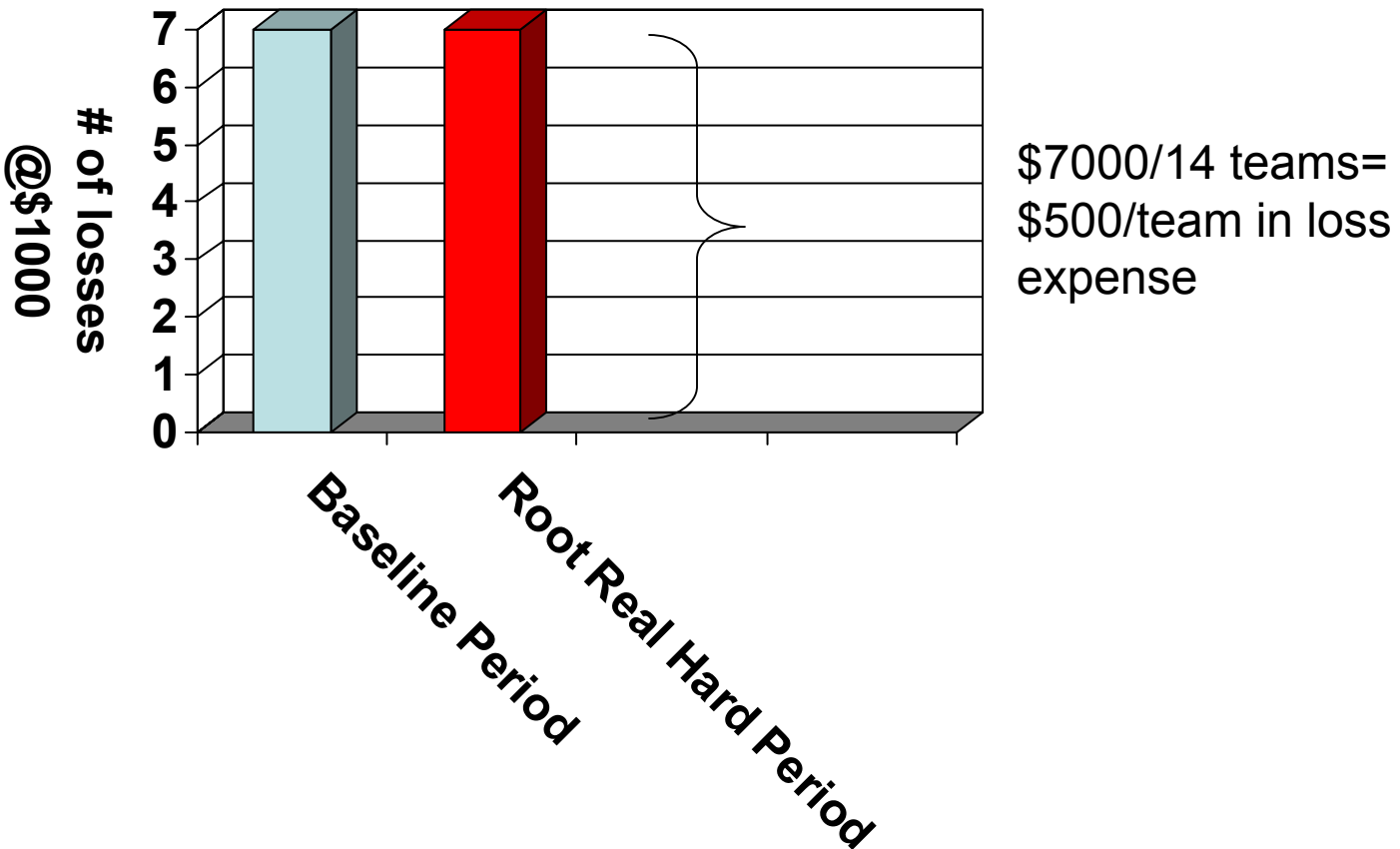
# Standings after 21 games

Team	Won 20 <sup>th</sup> game	Lost 21 <sup>st</sup> game	Team	Won	Lost 20 <sup>th</sup> and 21 <sup>st</sup> game
Yankees	16	5	Red Sox	12	9
Tampa	14	9	Blue Jays	12	9
Baltimore	13	8	White Sox	12	9
Royals	8	13	Cleveland	12	9
Seattle	8	13	Detroit	11	10
Anaheim	8	13	Texas	9	12
Minnesota	8	13	Oakland	7	14

7 Teams in Red lost 21<sup>st</sup> game



So you were unable to reduce the prevalence of lossitis among identified members the next day



# Biostatistics for \$200 please, Alex

- This is the percentage of all teams identified in this manner which will lose on any given day

# Biostatistics for \$200 please, Alex

- This is the percentage of all teams identified in this manner which will lose on any given day
  - “What is 50%?”

# Biostatistics for \$200 please, Alex

- This is the percentage of all teams identified in this manner which will lose on any given day
  - “What is 50%?”
  - Raise your hand if you think this is blatantly obvious



# Biostatistics for \$200 please, Alex

- This is the percentage of all teams identified in this manner which will lose on any given day
  - “What is 50%?”

**wrong**



# Suppose instead you did the same intervention after Opening Day

- We use losses to identify the prevalent population, same as before
  - Exact same methodology
  - Exact same “membership” -- the American League still has 14 teams

# Teams identified with findable lossitis after Opening Day

Team	Won	Lost	Team	Won	Lost
Yankees	1	0	Red Sox	0	1
Tampa	1	0	Blue Jays	0	1
Baltimore	1	0	White Sox	0	1
Royals	1	0	Cleveland	0	1
Seattle	1	0	Detroit	0	1
Anaheim	1	0	Texas	0	1
Minnesota	1	0	Oakland	0	1

# After Opening Day vs. 20 games in

	20 games in	After Opening Day
<b>Teams “findable” with lossitis in prevalence</b>	<b>14</b>	<b>7</b>
Total losses @\$1000 in baseline period	\$7000	\$7000



# After Opening Day

- Remember, you have no idea who those 7 unidentified teams are – they didn't file any claim related to the condition of lossitis

# Suppose instead you did the same intervention after the first game

- We use losses to identify the prevalent population, same as before
  - Exact same methodology
  - Exact same “membership” in the major leagues

- Exact same intervention is rooting real hard
- You root for all the identified teams the next day

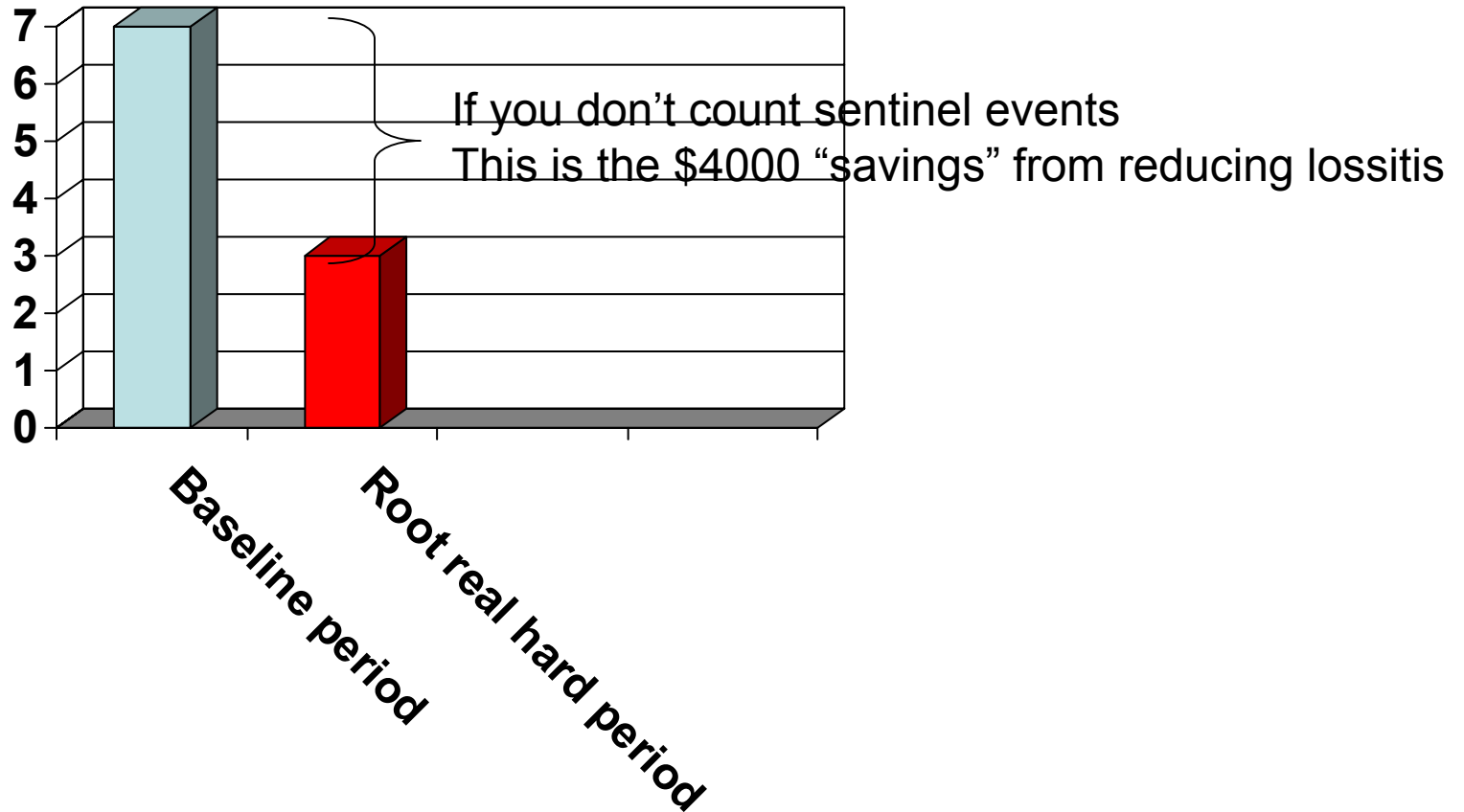
# Standings after second game

Team	Won	Lost	Team	Won	Lost
Yankees	2	0	Red Sox	1	1
Tampa	1	1	Blue Jays	1	1
Baltimore	1	1	White Sox	0	2
Royals	2	0	Cleveland	0	2
Seattle	1	1	Detroit	0	2
Anaheim	1	1	Texas	1	1
Minnesota	2	0	Oakland	1	1

# After the first game...

- After the first game you have identified 7 teams with “claims” (i.e., losses)
  - So you apply that intervention to the next day’s claims cycle
- Now you find that those teams only had 3 “claims” in this cycle so among identified people with lossitis, claims fell by \$4000

# Just counting previously 7 identified teams with lossitis (\$1000/identified team)



# What just happened?

- Example showed the impact on results when you CAN'T find the people in advance because they DON'T have any claims before getting sick...
- You get a completely invalid result *using the exact same methodology which was perfectly valid when used well into the season!*
  - Note: We will see later what happens when you add in sentinel events using conventional methodologies

# What are the implications for disease management ROI measurement?

- discussion:
  - Which diseases are more like the 20-game example (where you can identify everyone) and which diseases are more like the 1-game example (where some events will occur among people who are not identified)?

# Example from Asthma

First asthmatic has a claim in 2002

	2002	2003
Asthmatic #1	1000	0
Asthmatic #2		
Baseline		



# Second asthmatic has a claim in 2003

	2002	2003
Asthmatic #1	1000	0
Asthmatic #2	0	1000
Baseline		

# Baseline

	2002	2003
Asthmatic #1	1000	0
Asthmatic #2	0	1000
Baseline cost/asthmatic—usual methodology	1000	???

# Baseline

	2002	2003
Asthmatic #1	1000	0
Asthmatic #2	0	1000
Study Period cost/asthmatic— usual methodology	1000	500

# Who thinks this is an example of the “Opening Day” effect?

- IRVING, Texas--(BUSINESS WIRE)--Nov. 18, 2003--A pediatric asthma disease management program offered by [vendor] saved the State of North Carolina nearly one-third of the amount the government health plan expected to spend on children diagnosed with the disease

# The Sentinel Event Fallacy Infecting Everyone's Metrics

Presentation will show:

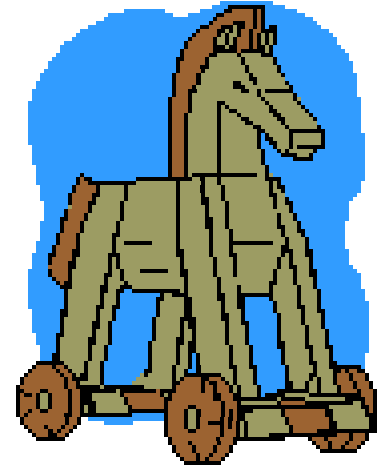
- THAT it happens
- **HOW it happens**
- WHY it happens
- **EXAMPLES** from real life
- What to do about it

# Let's Look at this another way

- We have shown THAT it happens.
- Now...*how* it happens

# Let's Look at this another way

- We have shown THAT it happens.
- Now...*how* it happens
  - A dynamic example
  - This is NOT beating a dead horse



# Uncovering the hidden flaw in the current measurement methodology: How this fallacy skews results

- Use an airplane analogy. Assume at any given time:
  - 25% of planes are cruising at 20,000 feet
  - 25% of planes are ascending at 10,000 feet
  - 25% of planes are descending at 10,000 feet
  - (25% of planes are on the ground)



# Uncovering the hidden flaw in the current methodology

- Use an airplane analogy. Assume at any given time:
  - 25% of planes are cruising at 20,000 feet
  - 25% of planes are ascending at 10,000 feet
  - 25% of planes are descending at 10,000 feet
- The average FLIGHT is at 13,333 feet

# Uncovering the hidden flaw in the current methodology

- Use an airplane analogy. Assume at any given time:
  - 25% of planes are cruising at 20,000 feet
  - 25% of planes are ascending at 10,000 feet
  - 25% of planes are descending at 10,000 feet
  - 25% of planes are on the ground
- The average FLIGHT is at 13,333 feet
- The average PLANE is at 10,000 feet

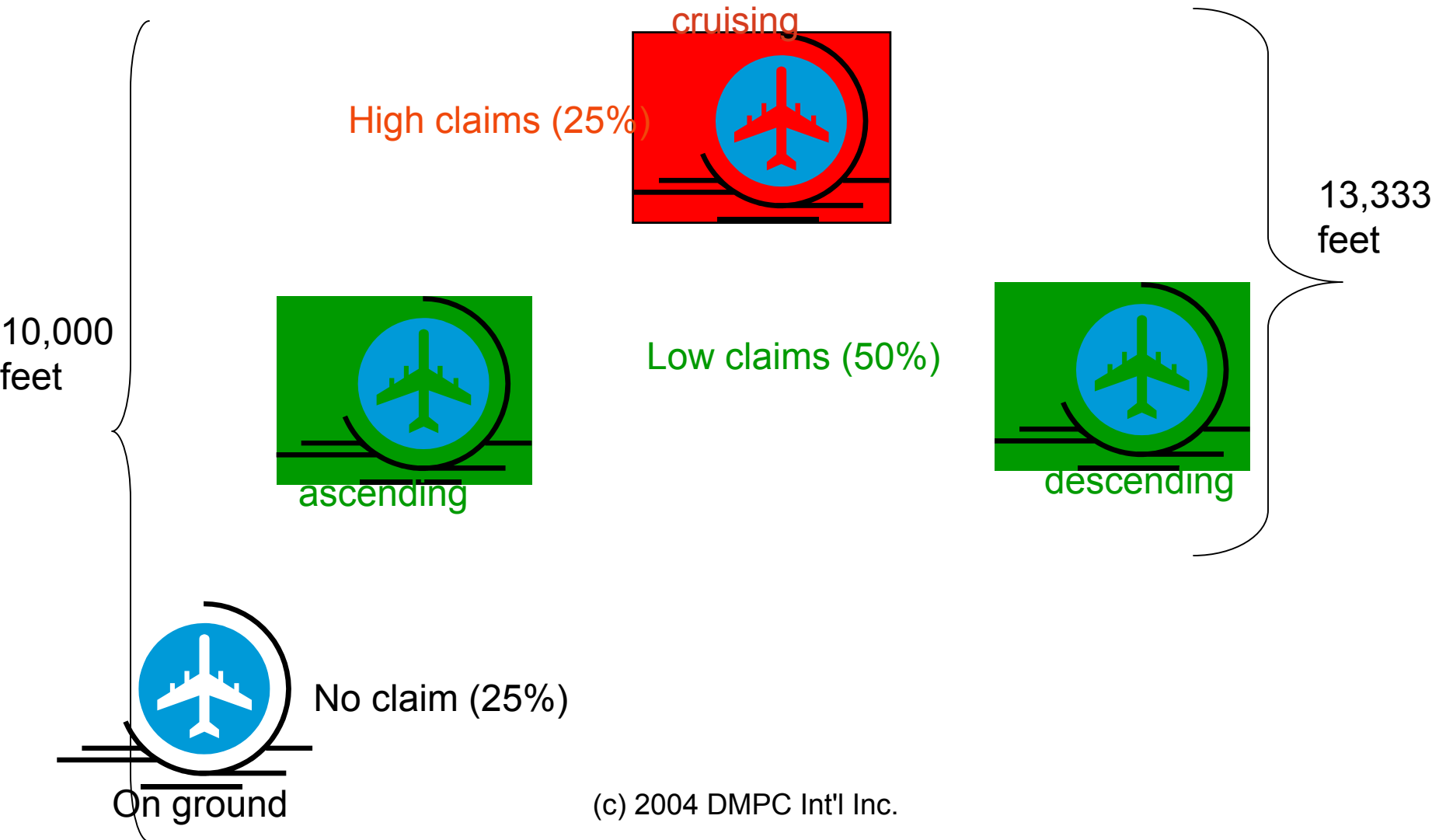
# Uncovering the hidden flaw in the current methodology

- Use an airplane analogy. Assume at any given time:
  - 25% of planes are cruising at 20,000 feet
  - 25% of planes are ascending at 10,000 feet
  - 25% of planes are descending at 10,000 feet
  - 25% of planes are on the ground
- The average FLIGHT is at 13,333 feet
- The average PLANE is at 10,000 feet
- Further assume that planes spend an hour (= one claims cycle) on the ground, ascending, descending, cruising

# The Analogy between flights and claims

- 25% of planes are cruising at 20,000 feet
  - These are High-claims members
- 25% of planes are ascending at 10,000 feet
  - These are Low-claims members
- 25% of planes are descending at 10,000 feet
  - These are Low-claims members
- 25% of planes are on the ground
  - These members have no claims for the disease in question

# Here's where current methodologies start—the baseline (first) tracking

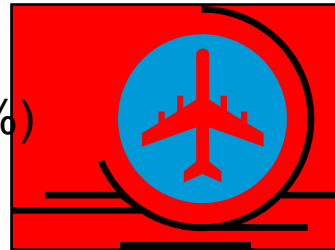


# The current best-practice approach

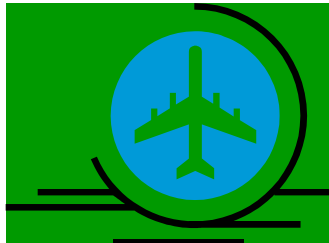
- Tracks ALL people with claims for the disease, high or low, in the baseline
- Properly emphasizes finding low utilizers for a population-based approach
  - Equivalent to finding all *flights* including ascending and descending
  - Average baseline altitude (2/3 at 10,000, 1/3 at 20,000) is: 13,333 feet

# You measure the claims on ALL patients with claims

High claims (33%)



Low claims (67%)



Above the line are datapoints which are found and measured

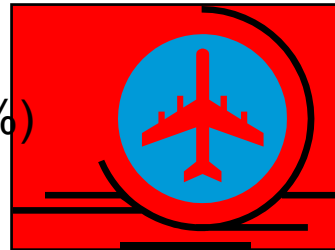
---



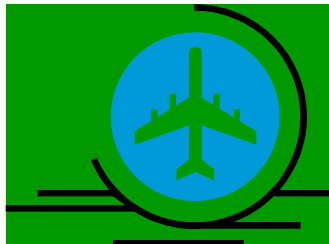
No claim

# You measure the claims on ALL patients with claims

High claims (33%)



Low claims (67%)

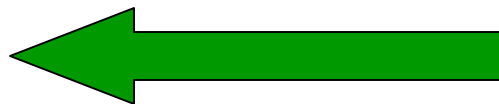


Above the line are datapoints which are found and measured



No claim

Why don't you measure these guys?



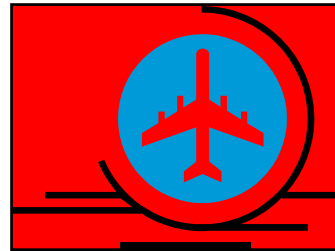
(c) 2004 DMPC Int'l Inc.



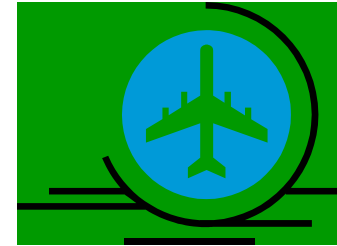
# You measure the claims on ALL patients with claims

13,333 Feet  
On average

High claims



Low claims



These get Found in The claims pull

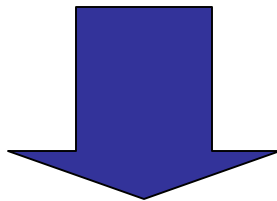
Above the line are datapoints which are measured  
Below the line is not included in measurement  
**Because they have no relevant claims to be found**



No claim

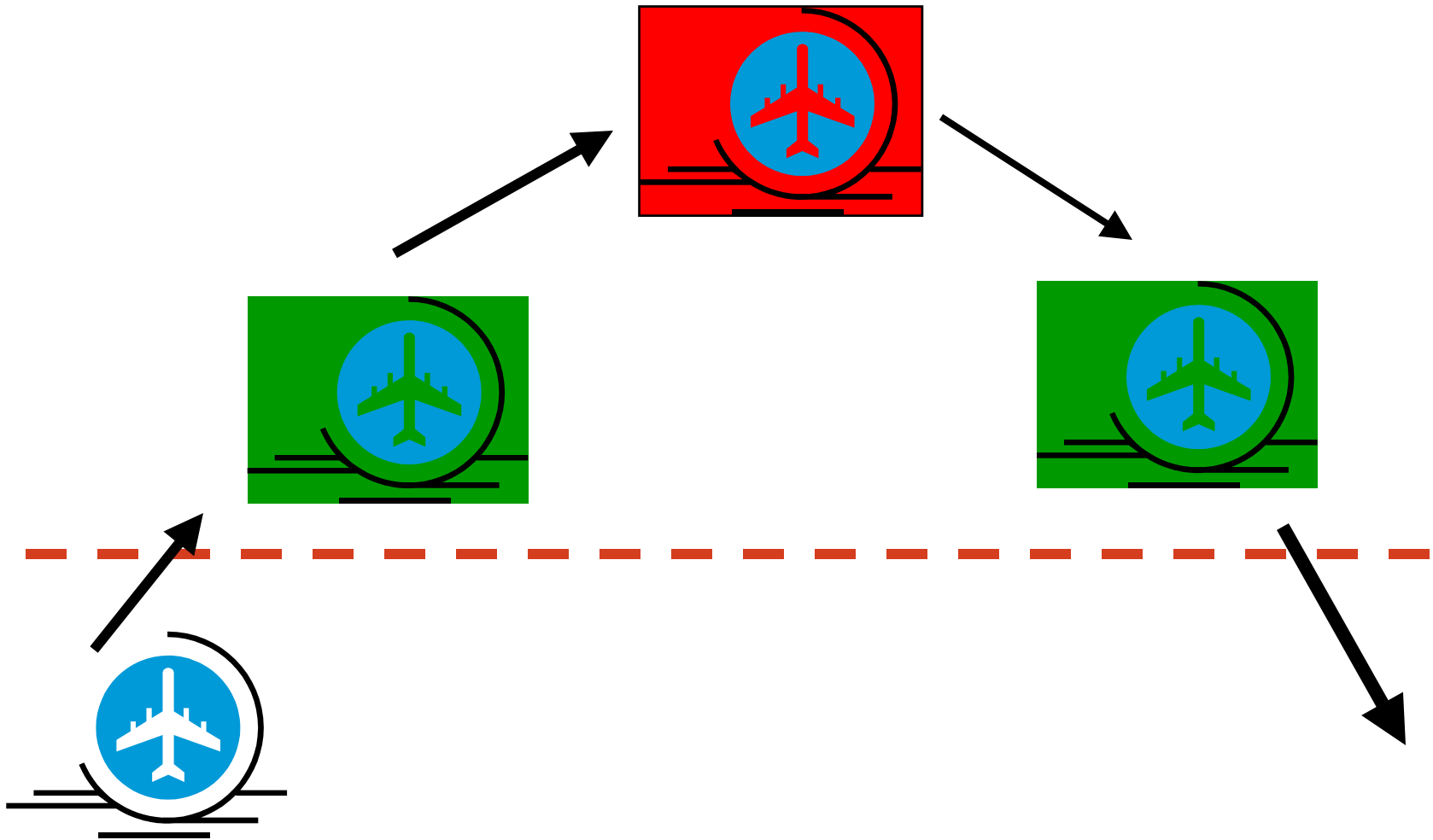
# The conventional approach

- Tracks ALL claims with claims for the disease, high or low, in the baseline
  - Equivalent to finding all *flights*
  - Average baseline altitude (2/3 at 10,000, 1/3 at 20,000) is: 13,333 feet



Now, track the baseline flights an hour later  
(analogous to tracking the claims during the study period)

# One hour later...(next claims cycle)

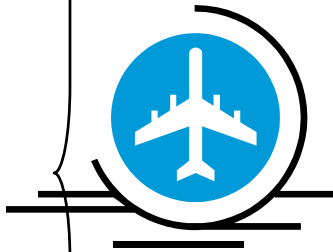


# We can all agree that...

- The aviation system is in a steady state
- Still 25% at each point
- Average altitude has not changed

# One hour later...(next claims cycle)

Average Plane is Still 10,000 feet

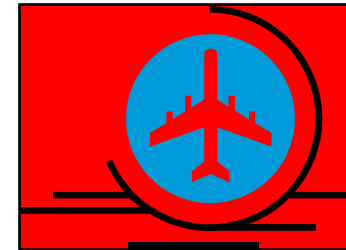


25%



High Claims  
25%

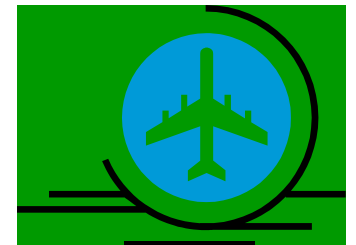
Low claims 25%



Average Flight is Still 13,333 feet

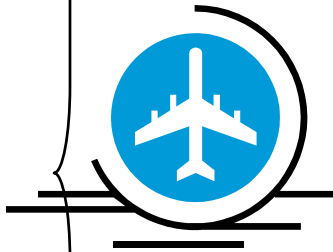
No claim 25%

(c) 2004 DMPC Int'l Inc.

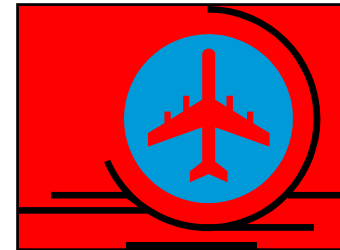


# One hour later...(next claims cycle)

Average  
Plane is  
Still  
10,000  
feet

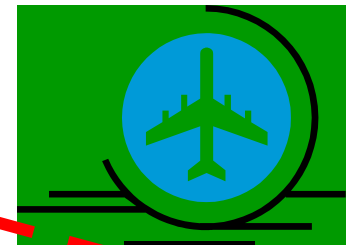


Average  
Flight is  
Still  
13,333  
feet



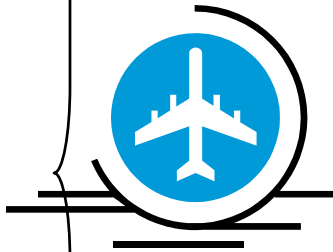
***Except that now all the flights are being Tracked including the ones which have Landed!***

(c) 2004 DMPC Int'l Inc.

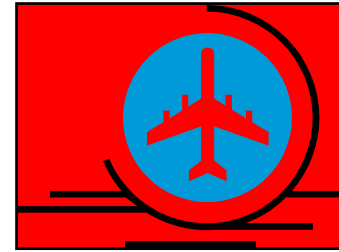


# One hour later...(next claims cycle)

Average Plane is Still 10,000 feet

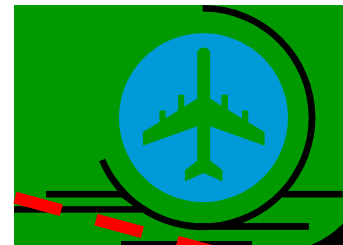


Average Flight is Still 13,333 feet



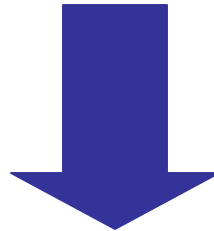
Measurement is 10,000 feet

***Except that now all the flights are being Tracked including the ones which have Landed!***



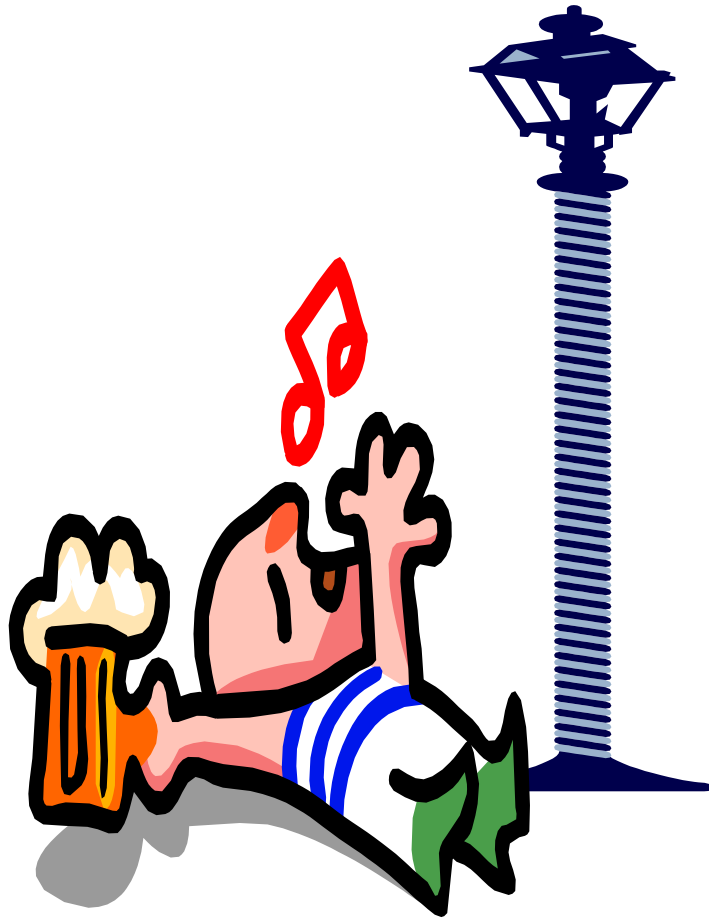
# Another way of looking at it

- Everyone with \$1 in claims identifying the disease is counted in a “whole population” methodology
  - But people with the disease with \$0 are not unless they are known about in advance



**What is the biostatistical rationale for this?**





(c) 2004 DMPC Int'l Inc.

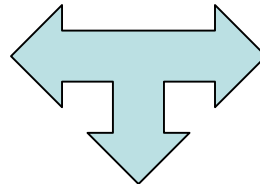


(c) 2004 DMPC Int'l Inc.

# A review of the allegedly blatantly obvious: Your health plan's medical spending

- \$1-billion on 500,000 members
  - 400,000 of which had claims

$$\begin{aligned} \$1\text{-billion}/500,000 &= \\ & \$2000 \end{aligned}$$



$$\begin{aligned} \$1\text{-billion}/400,000 &= \\ & \$2500 \end{aligned}$$

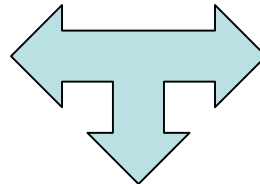
Which way do you calculate spending?



# Suppose it was Your health plan's *disease management* spending – Year 1

- \$1-billion on 500,000 *diseased* members
  - 400,000 of which had claims *identifying them as having the disease*

$$\begin{aligned} \$1\text{-billion}/500,000 &= \\ & \$2000 \end{aligned}$$



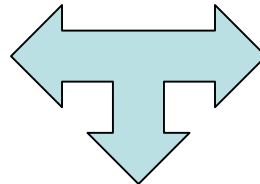
$$\begin{aligned} \$1\text{-billion}/400,000 &= \\ & \$2500 \end{aligned}$$

Which way is spending being calculated  
According to this approach?

# Suppose it was Your health plan's *disease management* spending – Year 1

- \$1-billion on 500,000 *diseased* members
  - 400,000 of which had claims *identifying them as having the disease*

$$\begin{aligned} \$1\text{-billion}/500,000 &= \\ & \$2000 \end{aligned}$$



$$\begin{aligned} \$1\text{-billion}/400,000 &= \\ & \$2500 \end{aligned}$$

Which way is spending being calculated  
According to this approach?

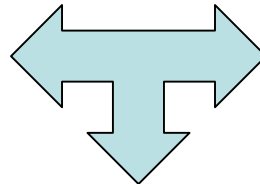
# Now look at year 2 for the health plan overall

- Assume no inflation, no turnover.
- Still \$1-billion in spending, still 500,000 members, 400,000 of which have claims (but it's a different 400,000)

# Suppose it was Your health plan's medical spending – Year 2

- \$1-billion on 500,000 members  
– 400,000 of which had claims

$$\begin{aligned} \$1\text{-billion}/500,000 &= \\ &\$2000 \end{aligned}$$



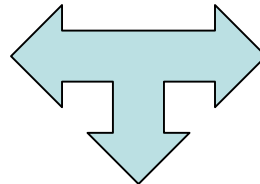
$$\begin{aligned} \$1\text{-billion}/400,000 &= \\ &\$2500 \end{aligned}$$

Still \$2000 in *per capita* medical spending, right?

# Suppose it was Your health plan's *disease management* spending – Year 2

- \$1-billion on 500,000 *diseased* members
  - 400,000 of which had claims *identifying them as having the disease in Year 2 but they are a different 400,000*

$$\begin{aligned} \$1\text{-billion}/500,000 &= \\ &\$2000 \end{aligned}$$



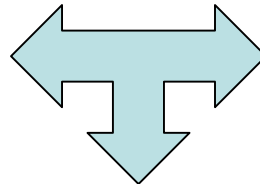
$$\begin{aligned} \$1\text{-billion}/400,000 &= \\ &\$2500 \end{aligned}$$



Suppose it was Your health plan's *disease management* medical spending – Year 2

- \$1-billion on 500,000 *diseased* members
  - 400,000 of which had claims *identifying them as having the disease but they are a different 400,000 (as in asthma, CAD)*

$$\begin{aligned} \$1\text{-billion}/500,000 &= \\ &\$2000 \end{aligned}$$



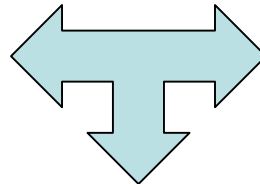
$$\begin{aligned} \$1\text{-billion}/400,000 &= \\ &\$2500 \end{aligned}$$

Which way is spending being calculated  
According to this approach?

Suppose it was Your health plan's *disease management* medical spending – Year 2

- \$1-billion on 500,000 *diseased* members
  - 400,000 of which had claims *identifying them as having the disease but they are a different 400,000 (as in asthma, CAD)*

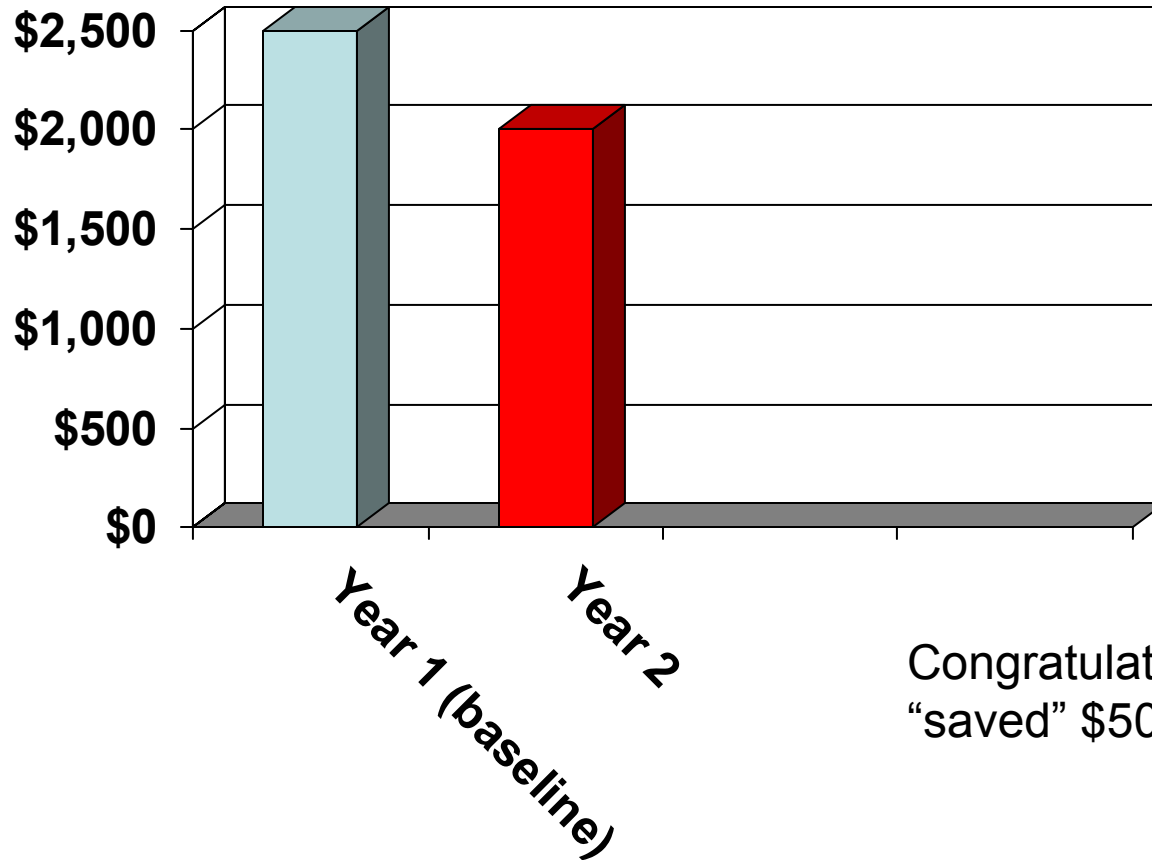
$$\begin{aligned} \$1\text{-billion}/500,000 &= \\ & \$2000 \end{aligned}$$



$$\begin{aligned} \$1\text{-billion}/400,000 &= \\ & \$2500 \end{aligned}$$

Which way is spending being calculated  
According to this approach?

# “Improvement” from Year 1 baseline to Year 2

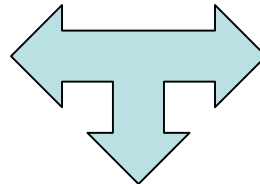


Congratulations—you just  
“saved” \$500!

# Your health plan's medical spending

- \$1-billion on 500,000 members
  - 400,000 of which had claims

$$\text{\$1-billion} / 500,000 = \text{\$2000}$$



$$\text{\$1-billion} / 400,000 = \text{\$2500}$$

Which way do you calculate *per capita* spending?

Raise your hand if you STILL think this was blatantly obvious



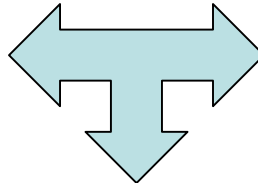
# But wait...Some people say...

- “We don’t track the people with no claims in the ‘post’ period in order to maintain equivalency with the ‘pre’ period”
- “The member has to re-trigger [with claims] each year to be counted”
  - So this bias shouldn’t happen because we don’t measure the zeros in EITHER period

“So, yes, we show \$2500 in the baseline”

- \$1-billion on 500,000 *diseased* members
  - 400,000 of which had claims *identifying them as having the disease*

$$\begin{aligned} \$1\text{-billion}/500,000 &= \\ & \$2000 \end{aligned}$$



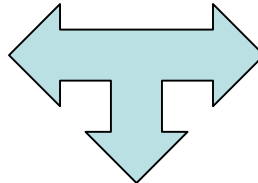
$$\begin{aligned} \$1\text{-billion}/400,000 &= \\ & \$2500 \end{aligned}$$

Which way is spending being calculated  
According to this approach?

“But we also show \$2500 in the study period”

- \$1-billion on 500,000 *diseased* members
  - 400,000 of which had claims *identifying them as having the disease in Year 2 but they are a different 400,000*

$$\begin{aligned} \$1\text{-billion}/500,000 &= \\ & \$2000 \end{aligned}$$



$$\begin{aligned} \$1\text{-billion}/400,000 &= \\ & \$2500 \end{aligned}$$

# Show of hands time...

- How many people think this is a valid “fix”?





# Show of hands time...

- How many people think this is a valid “fix”?

**Wrong Again**



# Biostatistics for \$400 please, Alex

Answer: This Phenomenon makes  
retriggering fix invalid

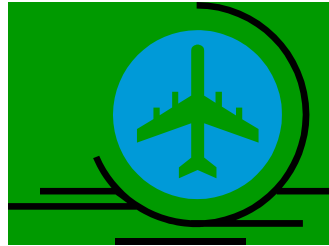
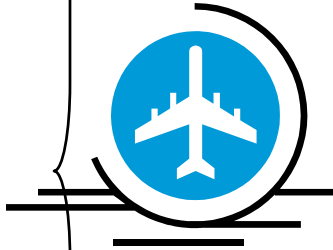
# Biostatistics for \$400 please, Alex

Answer: This phenomenon makes the fix invalid

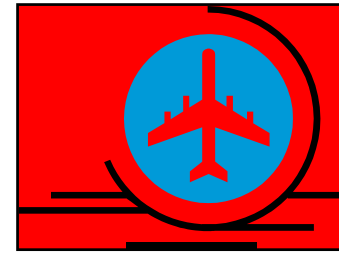
Question: The strong association between time since last event and compliance

“So this should happen because you don't measure the zeroes, right?”

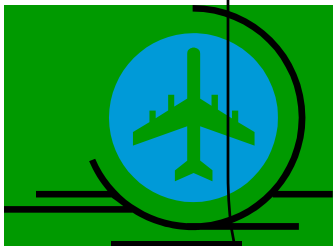
Average Plane is Still 10,000 feet



Average Flight is Still 13,333 feet



Not here



Not here



# Wrong

- What is the fallacy with that “adjustment” ?

# Explanation of why the bias is still there even if zeroes aren't measured

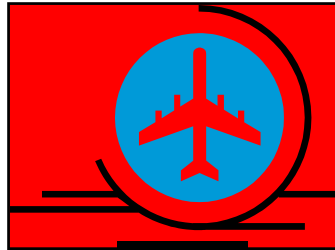
- Because AFTER a “zero” has an event and then recovers, that person is put on drugs (asthma, beta blockade, antihyperlipidemics etc.)

# This is called the “asymmetrical zeroes” fallacy

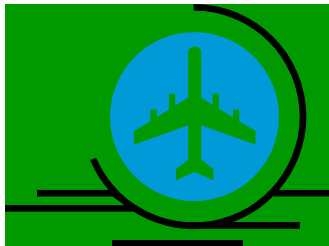
- If people were as likely to take drugs to prevent attacks *before* as *after*, then this adjustment would remove bias
- However, people are way more likely to take drugs (and hence have nonzero claims) after they land than before they take off

# Many more people have zero identifiable claims before an event than after it

High claims

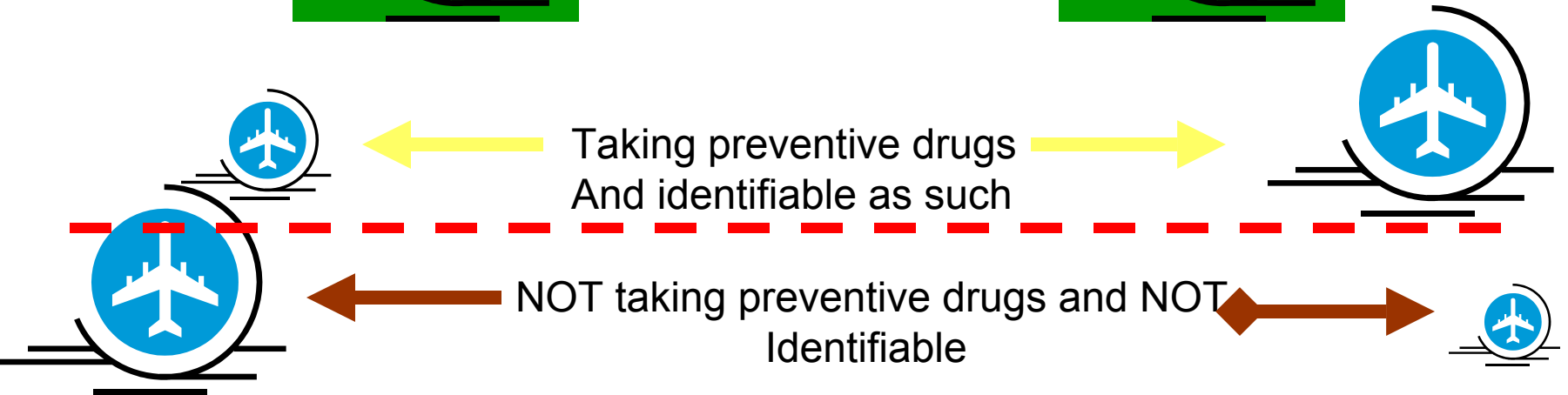


Middle claims



Taking preventive drugs  
And identifiable as such

NOT taking preventive drugs and NOT  
Identifiable





# Recall these 4 slides from earlier...

	2002	2003
Asthmatic #1	1000	0
Asthmatic #2		
Baseline		

# Second asthmatic has a claim in 2003

	2002	2003
Asthmatic #1	1000	0
Asthmatic #2	0	1000
Baseline		

# Baseline

	2002	2003
Asthmatic #1	1000	0
Asthmatic #2	0	1000
Baseline cost/asthmatic—usual methodology	1000	???

# Baseline

	2002	2003
Asthmatic #1	1000	0
Asthmatic #2	0	1000
Study Period cost/asthmatic— if you don't count the zeroes	1000	1000

You are removing Both zeroes

But here's what's more likely to happen  
Example from Asthma

First asthmatic has a claim in 2002 *and starts on meds in 2003*

	2002	2003
Asthmatic #1	1000	100
Asthmatic #2		
Baseline		

# Second asthmatic has a claim in 2003

	2002	2003
Asthmatic #1	1000	100
Asthmatic #2	0	1000
Baseline		

# Baseline

	2002	2003
Asthmatic #1	1000	100
Asthmatic #2	0	1000
Baseline— usual methodology	1000	???

# Baseline

	2002	2003
Asthmatic #1	1000	100
Asthmatic #2	0	1000
Study Period— usual methodology	1000	550



# The “zeroes” are asymmetrical

	2002	2003
Asthmatic #1	1000	100
Asthmatic #2	0	1000
Study Period— usual methodology	1000	550

(c) 2004 DMPC Int'l Inc.

Even if you don't  
Count zeroes you  
Get an invalid answer

# QED

- The “Zeroes” are not symmetrical due to people being put on drugs post-event
  - This IS the current methodology used by everyone--Including my own until 2003—except people who are making even more basic mistakes
  - It will distort results via the “Fallacy of the Asymmetrical Zeroes,” period...

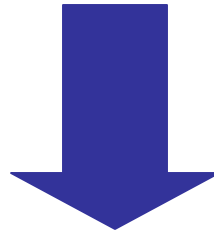
# The Sentinel Event Fallacy Infecting Everyone's Metrics

Presentation will show:

- THAT it happens
- HOW it happens
- **WHY it happens**
- **EXAMPLES** from real life
- What to do about it

# WHY this happens

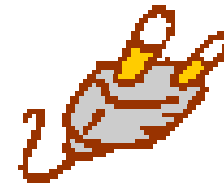
- Recall that Everyone with \$1 in claims identifying the disease is counted in a “whole population” methodology
  - But people with the disease with \$0 are not



**This is recognized by some vendors (and was recognized by me) and there was a “fix” put in place**

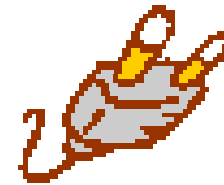
# Why the usual “cure” compounds the problem

- What is the usual “fix”
  - the plug-in number used for members who are identified “after the fact” to be added to the baseline?



# Why the usual “cure” compounds the program

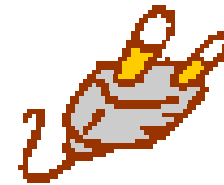
- What is the usual plug-in number used for members who are identified “after the fact” to be added to the baseline?



You add the person in THIS year even though they were not Added in LAST year

# Why the usual “cure” compounds the program

- What is the usual plug-in number used for members who are identified “after the fact” to be added to the baseline?



You add the person in as though they had the average Events last year

# Why the usual “cure” fails

- What is the usual plug-in number used for members who are identified “after the fact” to be added to the baseline?

– In the airplanes case?

Example from old DMPC RFP, pre-identification of fallacy

NEW AND NEWLY DIAGNOSED MEMBERS	Assumed to cost the Adjusted Baseline.
---------------------------------	----------------------------------------



# Why the usual “cure” fails

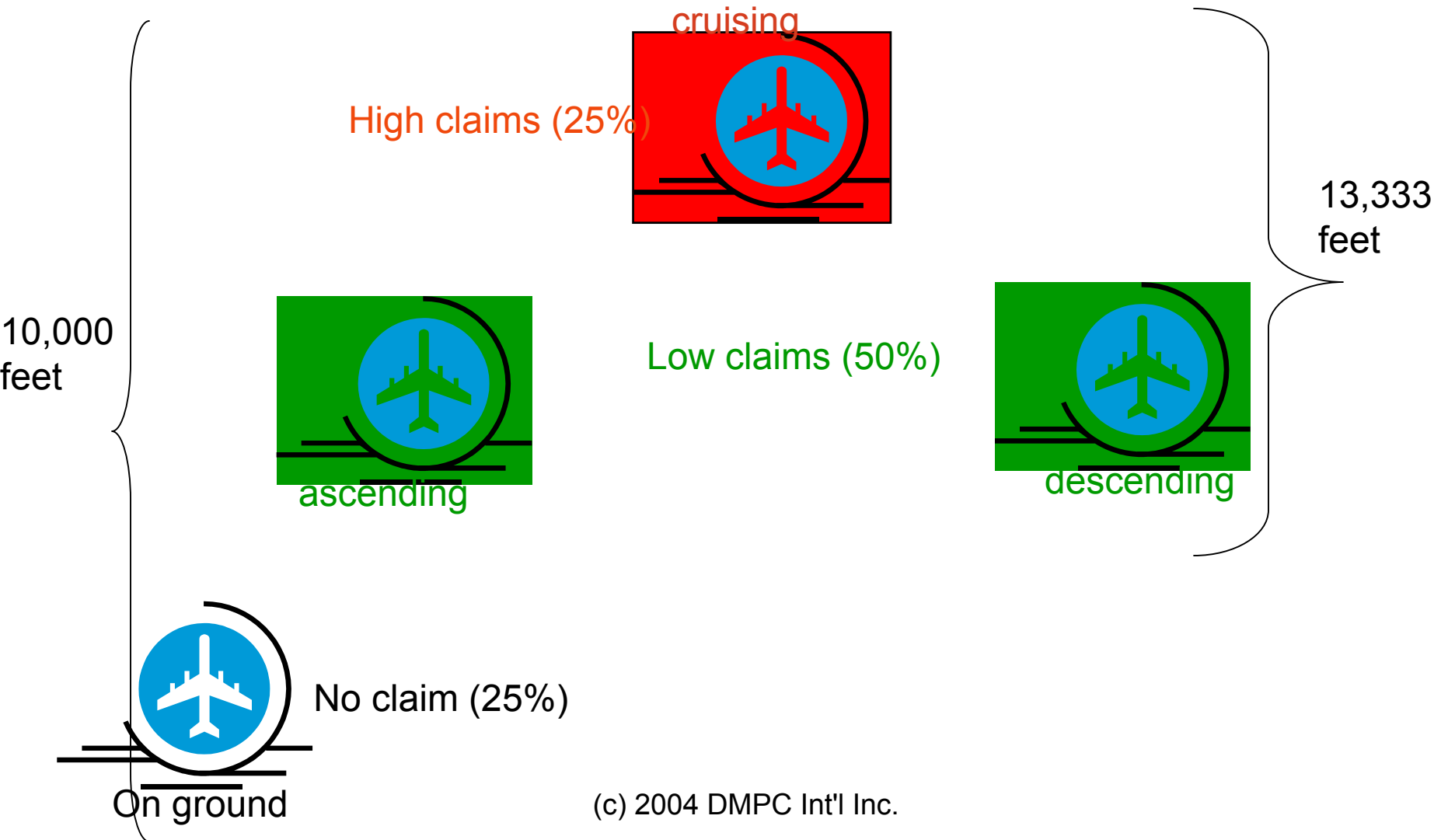
- What is the usual plug-in number used for members who are identified “after the fact” to be added to the baseline?

Example from old DMPC RFP, pre-identification of fallacy

What is this figure in the airplanes case?

NEW AND NEWLY DIAGNOSED MEMBERS	Assumed to cost the Adjusted Baseline.
---------------------------------	----------------------------------------

# The plug-in figure vs. what really happens



# Why the usual “cure” fails

- What is the usual plug-in number used for members who are identified “after the fact” to be added to the baseline?

Example from old DMPC RFP, pre-identification of fallacy

– In this case: \$13,333 because adding them does not change the baseline retro

NEW AND NEWLY DIAGNOSED MEMBERS	Assumed to cost the Adjusted Baseline.
---------------------------------	----------------------------------------

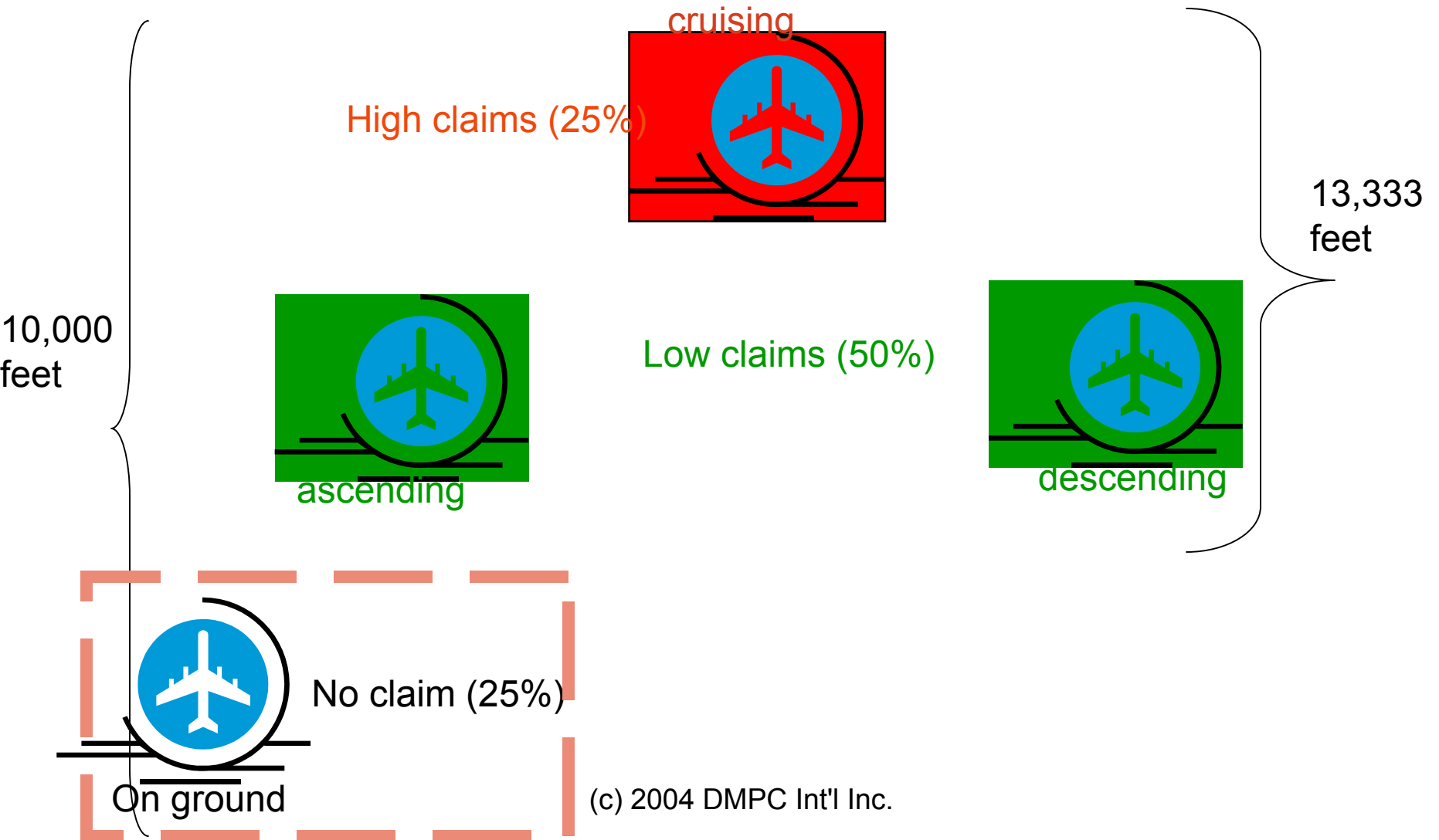
# Why the usual “cure” fails

- What is the usual plug-in number used for members who are identified “after the fact” to be added to the baseline?
  - In this case: \$13,333
- What should it be?

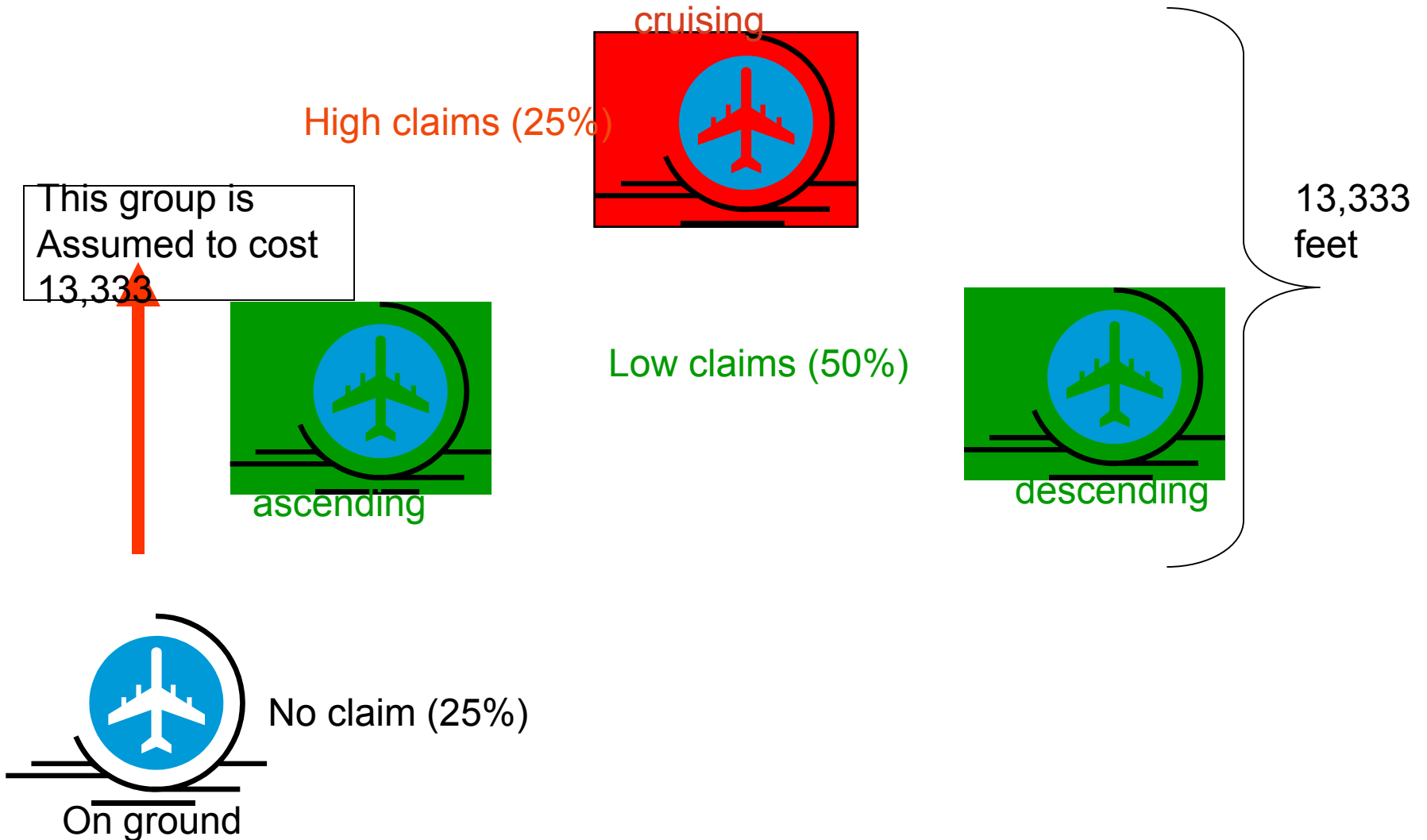
Example from old DMPC RFP, pre-identification of fallacy

NEW AND NEWLY DIAGNOSED MEMBERS	Assumed to cost the Adjusted Baseline.
---------------------------------	----------------------------------------

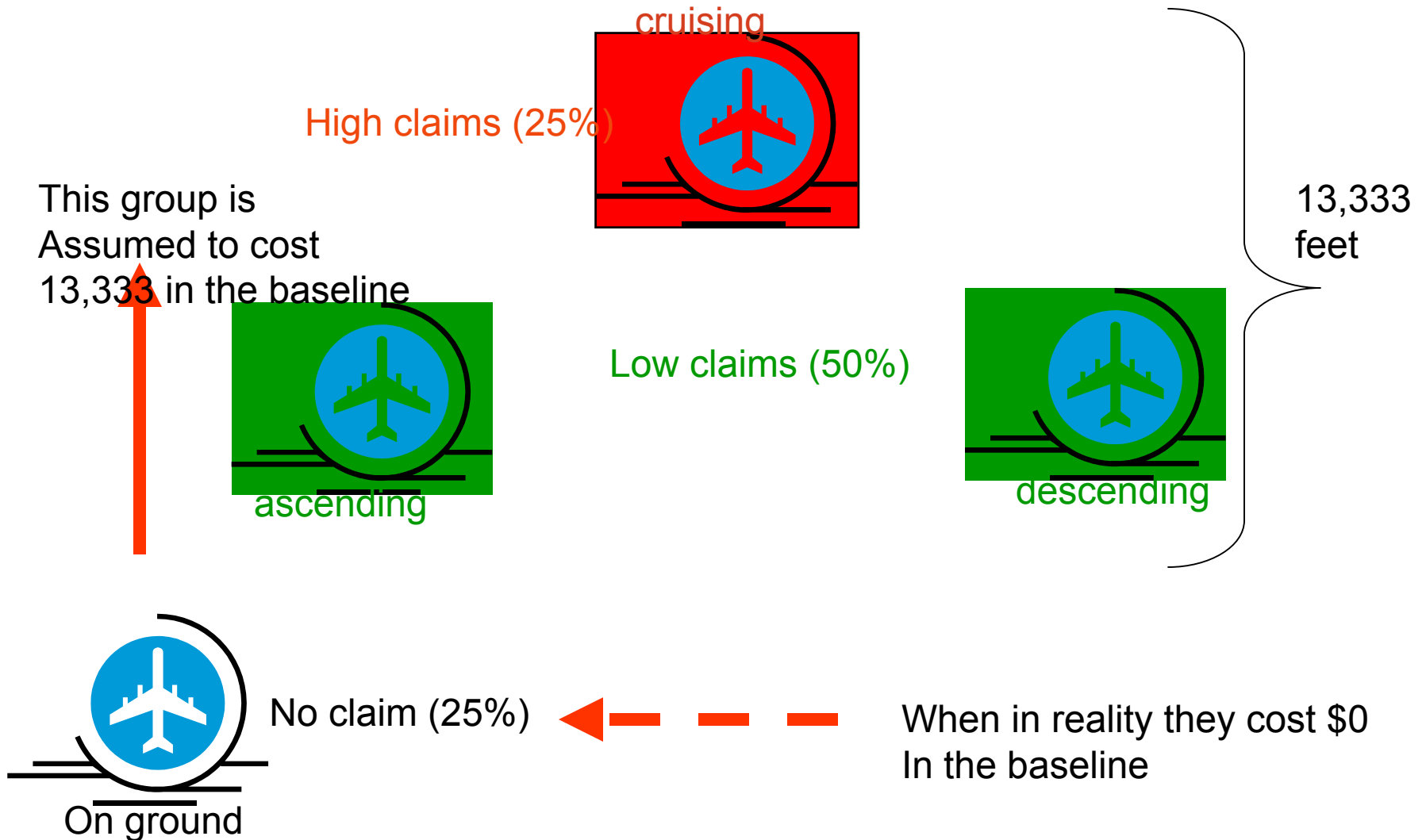
# The plug-in figure vs. what really happened in the baseline



The plug-in figure once you find them is the \$13,333 baseline...but what should it be?



# When they didn't cost \$13,333 in the baseline—they cost \$0



# Let's go back to the ball game

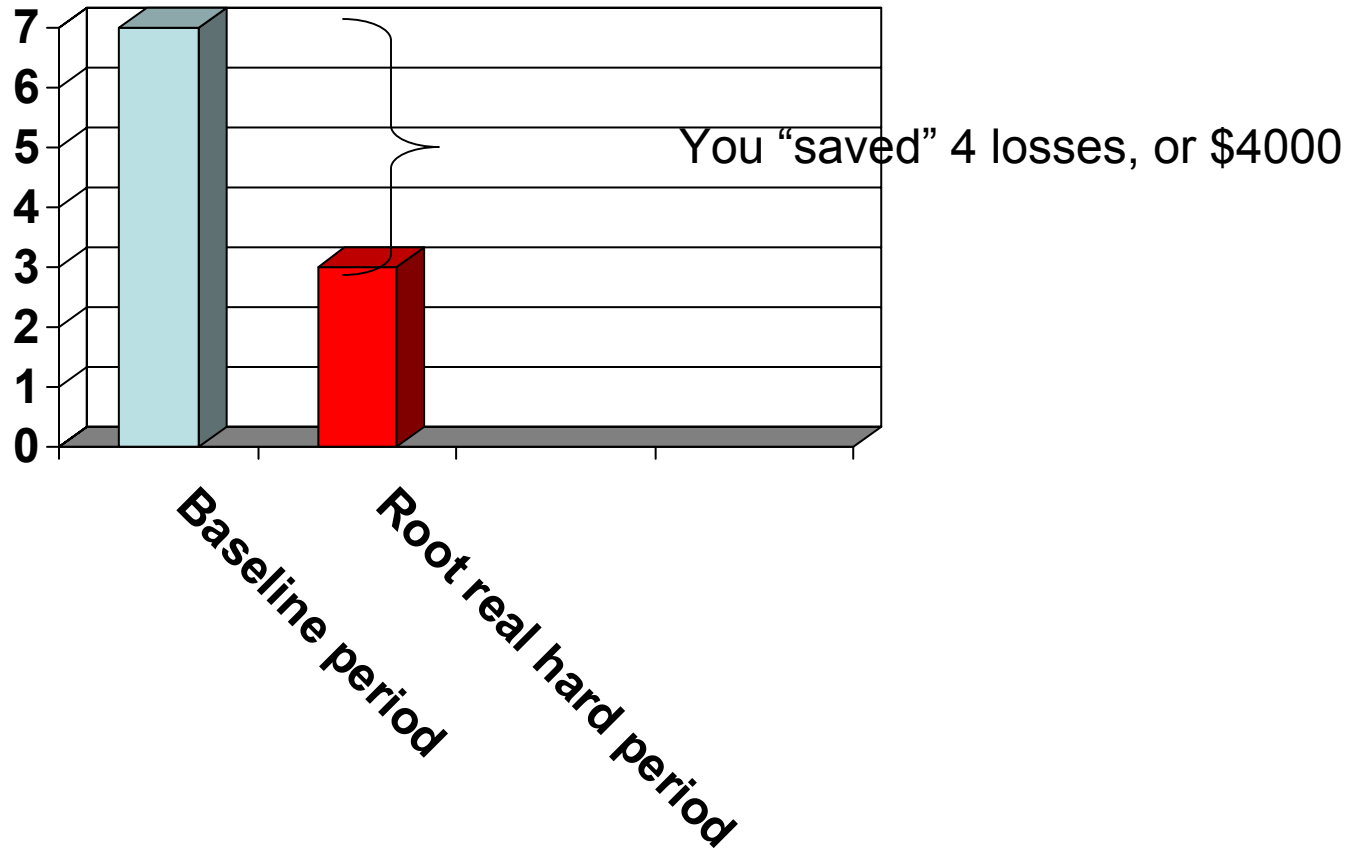
- See what happens if you apply that “fix” there



Recall the second game--That slide just looked at pre-identified members from the first game

Team	Won	Lost	Team	Won	Lost
Yankees	2	0	Red Sox	1	1
Tampa	1	1	Blue Jays	1	1
Baltimore	1	1	White Sox	0	2
Royals	2	0	Cleveland	0	2
Seattle	1	1	Detroit	0	2
Anaheim	1	1	Texas	1	1
Minnesota	2	0	Oakland	1	1

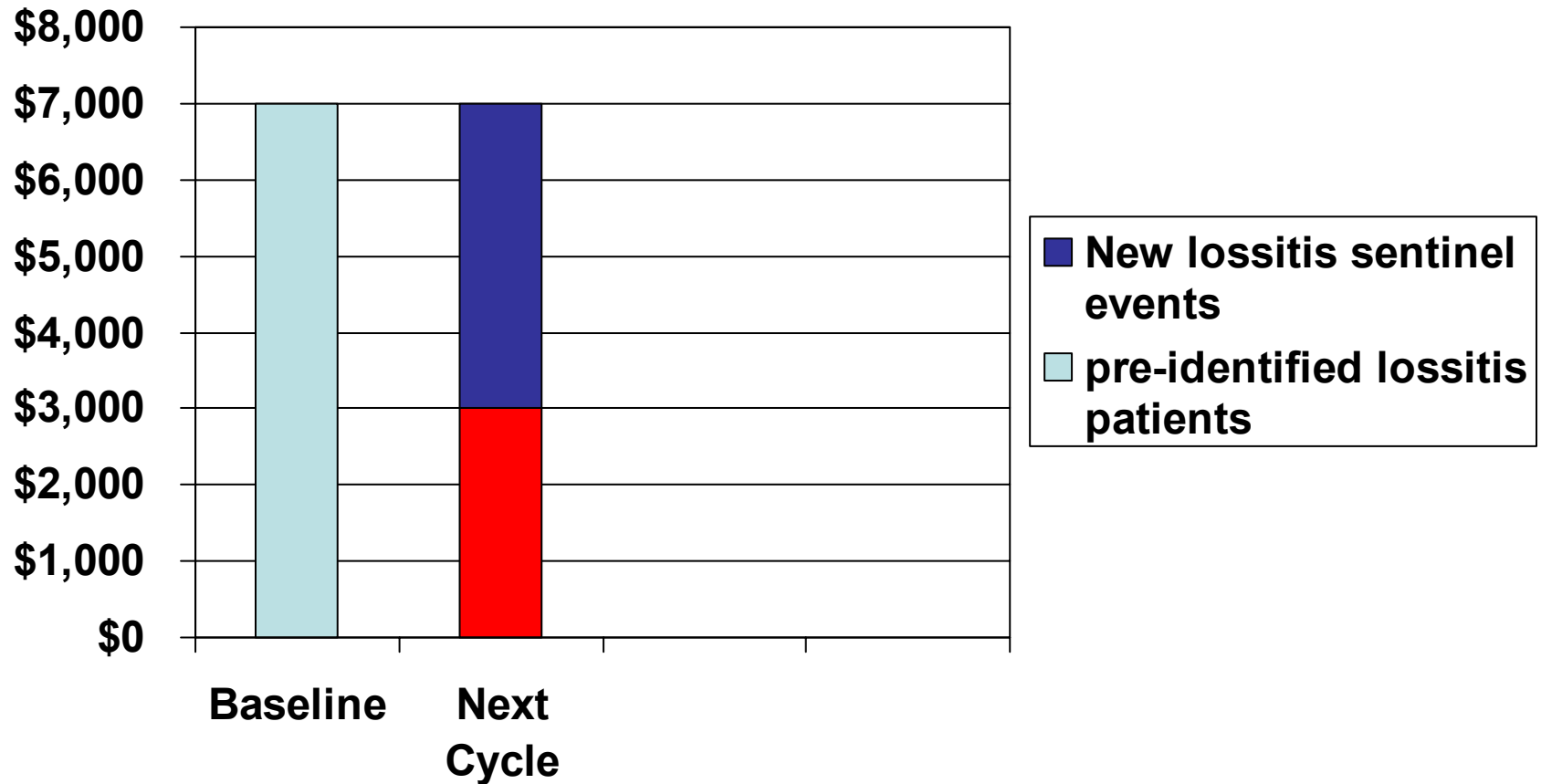
# Leading you to this conclusion...



# Standings after second game— including new “sentinel events”— patients with lossitis

Team	Won	Lost	Team	Won	Lost
Yankees	2	0	Red Sox	1	1
Tampa	1	1	Blue Jays	1	1
Baltimore	1	1	White Sox	0	2
Royals	2	0	Cleveland	0	2
Seattle	1	1	Detroit	0	2
Anaheim	1	1	Texas	1	1
Minnesota	2	0	Oakland	1	1

This is what really happens-- you add in new “sentinel event” claims —your overall lossitis rate (losses = \$1000) is still the same



# Apply the usual sentinel event “adjustment” to that slide...???

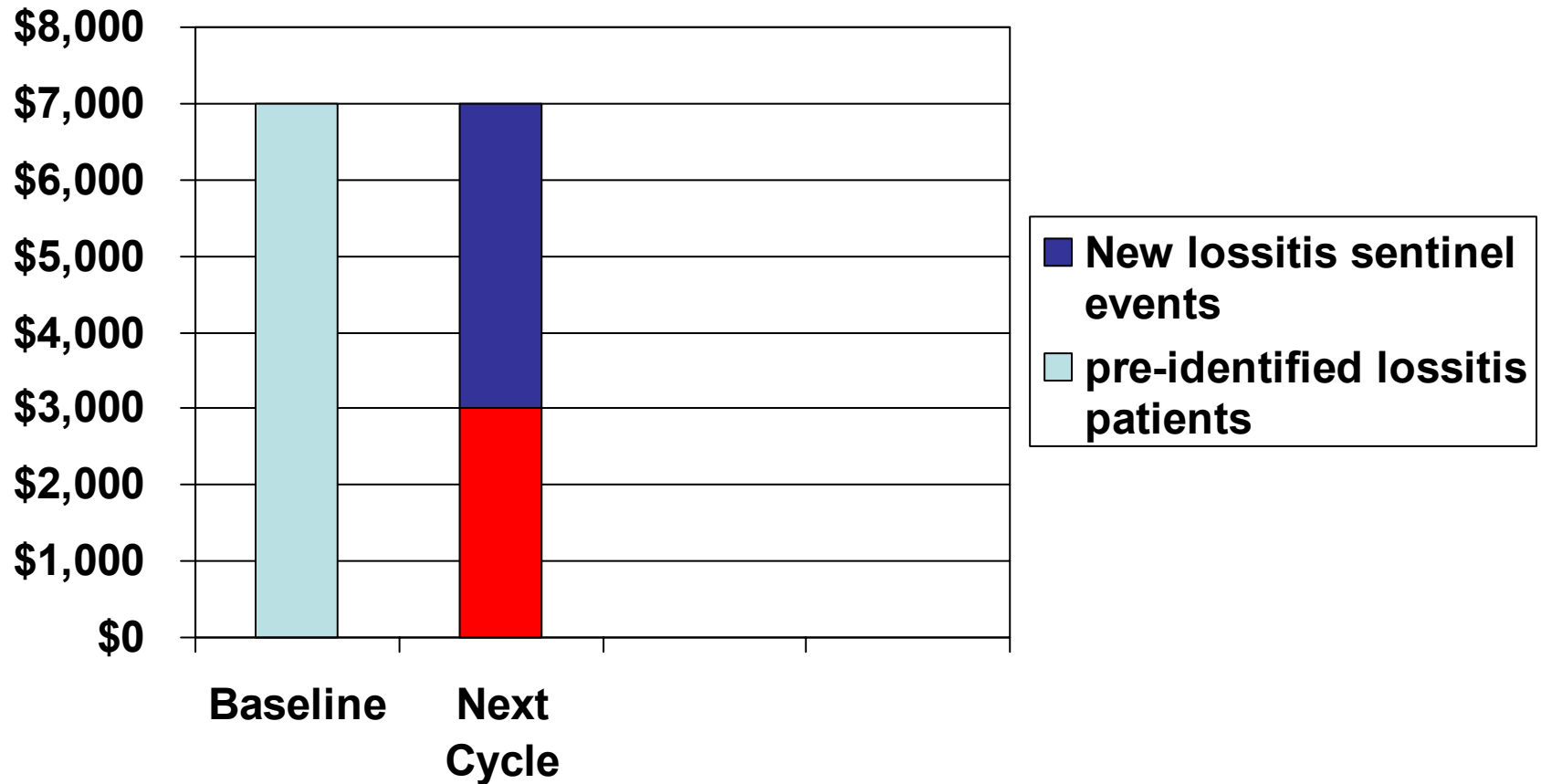
- What is the usual plug-in number used for members who are identified “after the fact” to be added to the baseline?

Example from old DMPC RFP, pre-identification of fallacy

– What do you get for the baseline?

NEW AND NEWLY DIAGNOSED MEMBERS	Assumed to cost the Adjusted Baseline.
------------------------------------------	-------------------------------------------

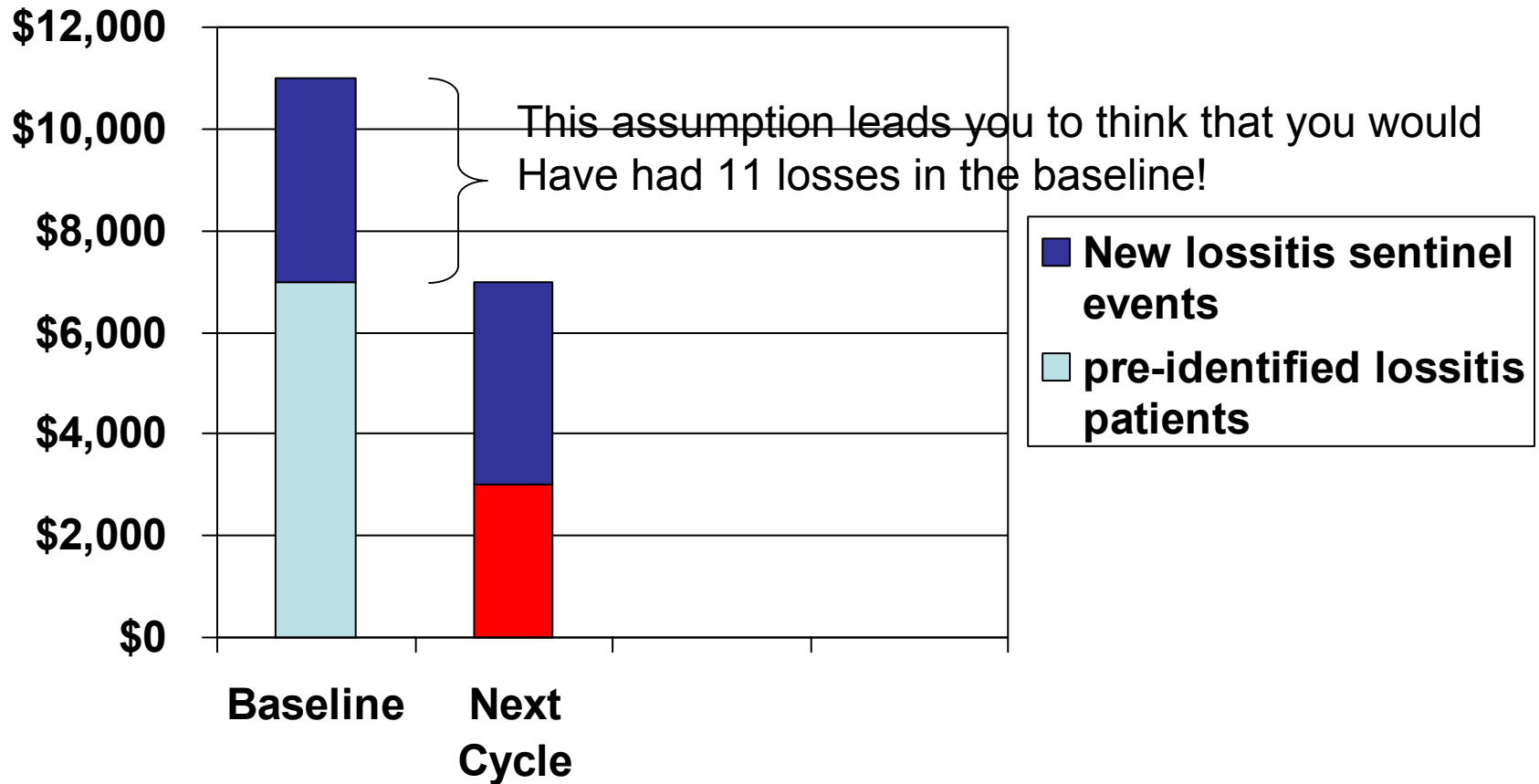
In this case the baseline is \$1000 so if you assume the teams in the second cycle WOULD HAVE HAD \$1000 in claims...



# Biostatistics for \$600 please, Alex

- Classic misunderstanding: “But the study period claims cost is accurate.”

This is what happens when you “assume” that previously unidentified means: “WOULD have had the average baseline cost (or their actual claims cost) the previous cycle...”





# Anyone still unconvinced?

- Who still thinks their metrics are as valid now as you thought they were an hour ago?

# The Sentinel Event Fallacy Infecting Everyone's Metrics

Presentation will show:

- THAT it happens
- HOW it happens
- WHY it happens
- **EXAMPLES from real life**
- **What to do about it**

# What to do about it-Part One

- Ways to lessen (but not eliminate) problem
  - Use 2+ years for baseline

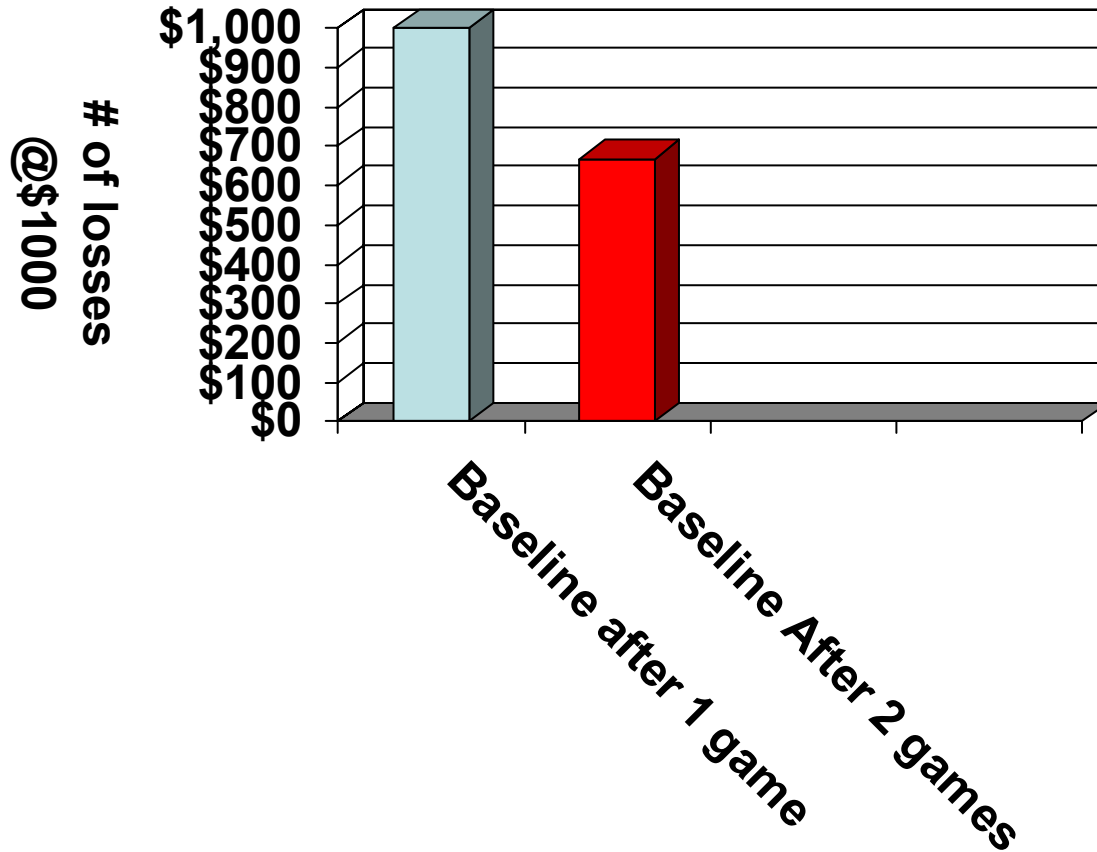
# Identifying people with lossitis using TWO years of data (first two games of season)

Team	Won	Lost	Team	Won	Lost
Yankees	2	0	Red Sox	1	1
Tampa	1	1	Blue Jays	1	1
Baltimore	1	1	White Sox	0	2
Royals	2	0	Cleveland	0	2
Seattle	1	1	Detroit	0	2
Anaheim	1	1	Texas	1	1
Minnesota	2	0	Oakland	1	1

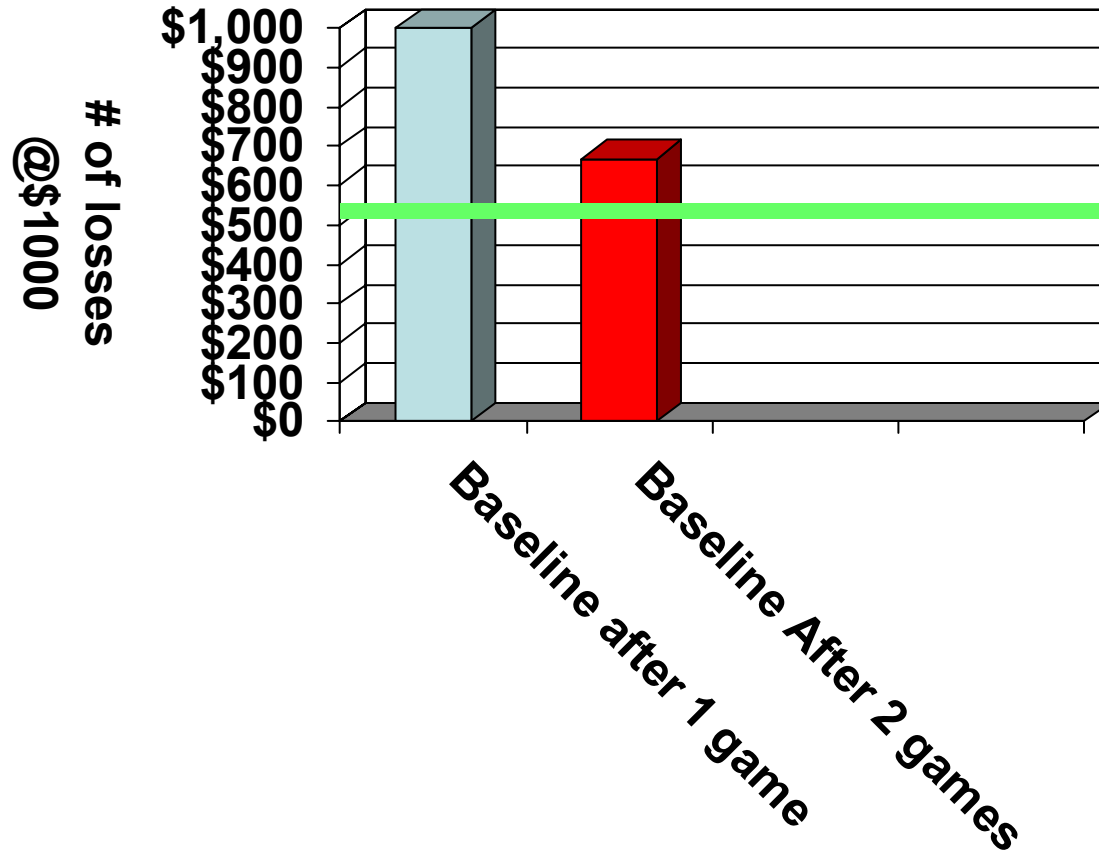
# Lossitis baseline with 11 identified teams

- Each loss in the baseline (2<sup>nd</sup> game) still \$1000
- Now you divide the 7 losses by the 11 identified teams instead of 7

# You've *lessened* the distortion



# You've *lessened* the distortion but it still remains



Obviously the "real" number is \$7000/14 teams, Or \$500 baseline

# What to do about it-Part One

- Ways to lessen (but not eliminate) problem
  - Use 2+ years for baseline
  - Use HRAs to find some “zeroes”
    - Would work if everyone did what three things?
      - 1.
      - 2.
      - 3.

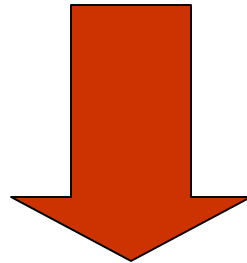


# What to do about it-Part One

- Ways to lessen (but not eliminate) problem
  - Use 2+ years for baseline
  - Use HRAs to find some “zeroes”
    - Would work if everyone
      1. Filled them out;
      2. told the truth;
      3. knew they were about to have their first attack

# What to do about it-Part One

- Ways to lessen (but not eliminate) problem
  - Use 2+ years for baseline
  - Use HRAs to find some “zeroes”



Helps reduce the distortion by finding some baseline people  
Before they have claims...but does not address the root cause which  
Is that many “zeroes” simply can’t be found

# Diagnosing It, Part One

- Plausibility indicators: Total unit claims paid which are relevant to a disease
  - This captures the zeroes by looking at OVERALL RATES PER 1000 so every claim is captured in every period
  - Based on total age/sex-adjusted population
  - Total population cannot regress to the mean because it *is* the mean

# How does looking at unit claims/1000 avoid this

- Unit claims can't hide

# Where are the claims from *previously undiagnosed* asthmatics?

- IRVING, Texas--(BUSINESS WIRE)--Nov. 18, 2003--A pediatric asthma disease management program offered by [Vendor with very good business judgment] saved the State of North Carolina nearly one-third of the amount the government health plan expected to spend on children **diagnosed** with the disease

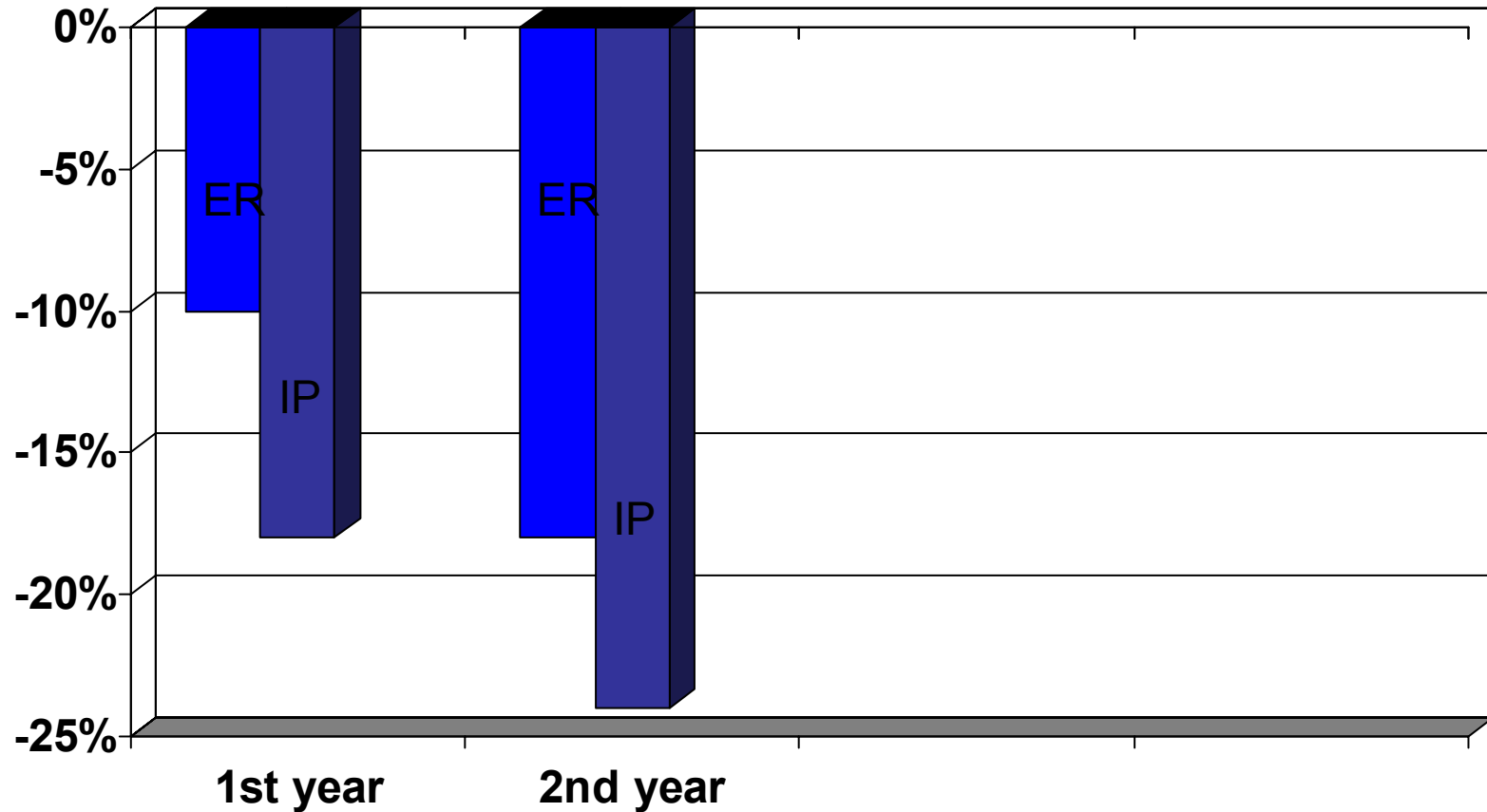
# Where are the claims from *previously undiagnosed* asthmatics?

- IRVING, Texas--(BUSINESS WIRE)--Nov. 18, 2003--A pediatric asthma disease management program offered by [Vendor with very good business judgment] saved the State of North Carolina nearly one-third of the amount the government health plan expected to spend on children **diagnosed** with the disease



Let's see what happens when you measure only people who were diagnosed

# Example of just looking at Diagnosed people: Vendor Claims for Asthma Cost/patient Reductions



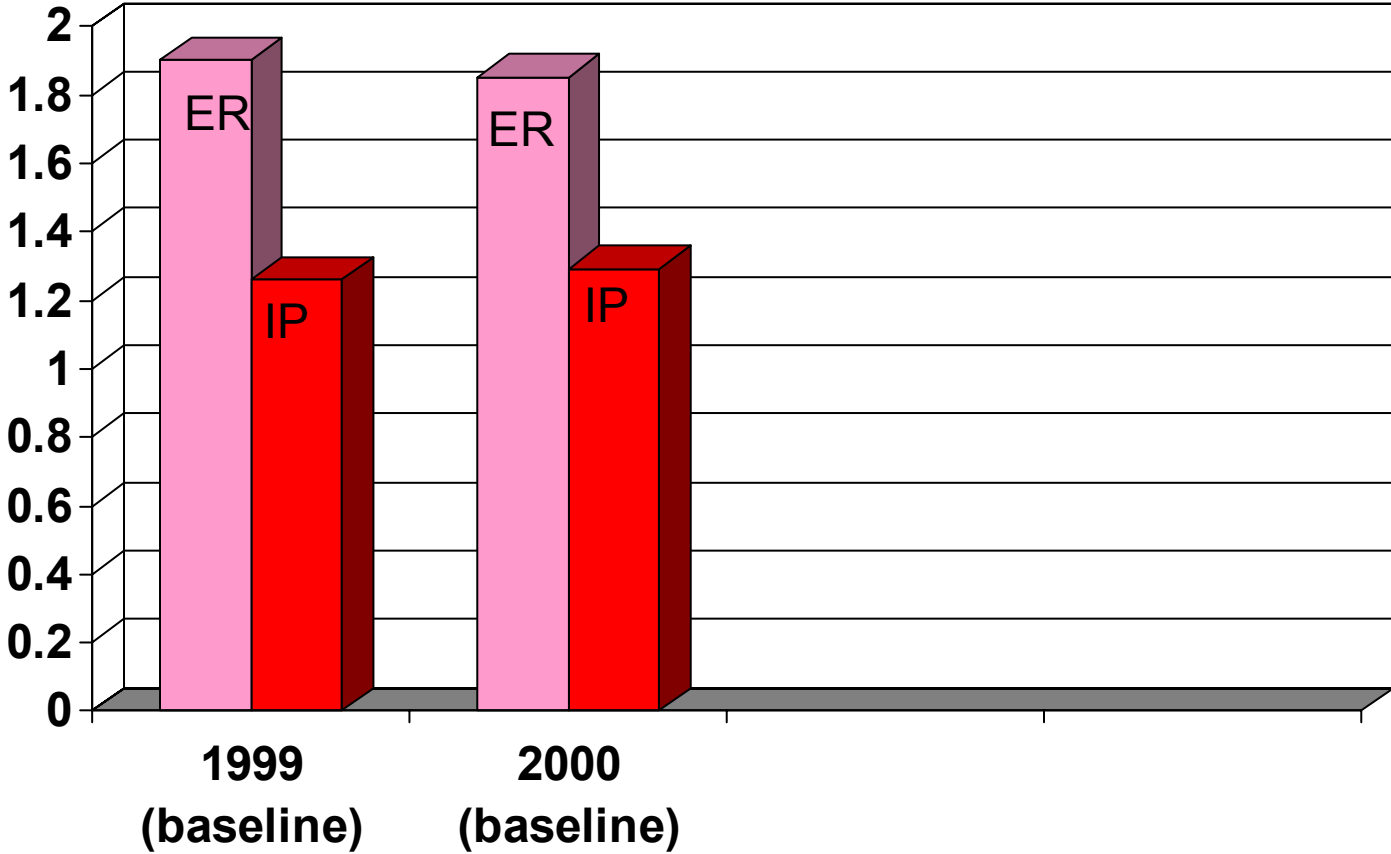
# What we did...

- We looked at the actual codes across the plan
- This includes everyone
- Two years of codes pre-program to establish trend
- Then two program years



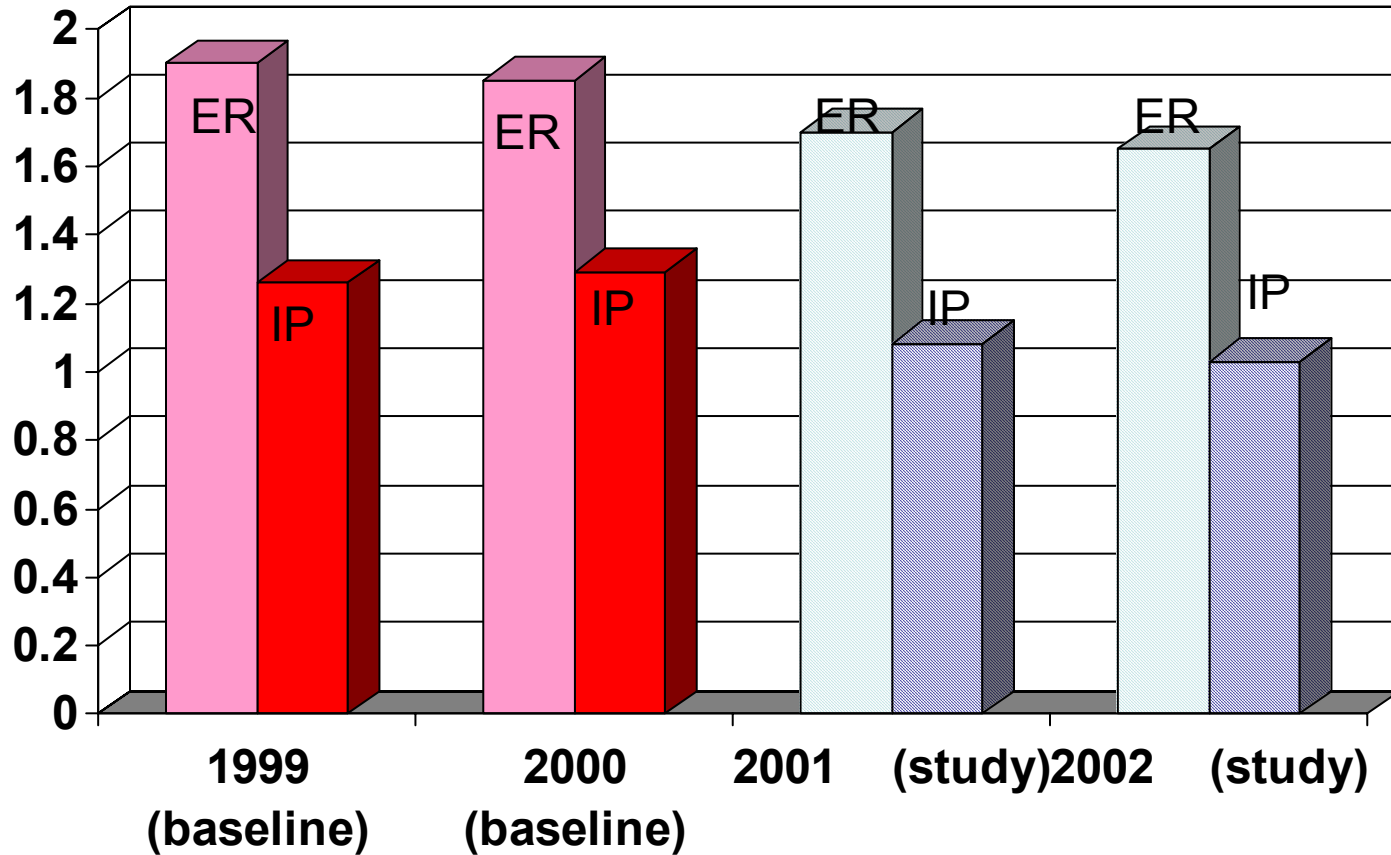
# Baseline trend

493.xx ER visits and IP stays/1000 planwide

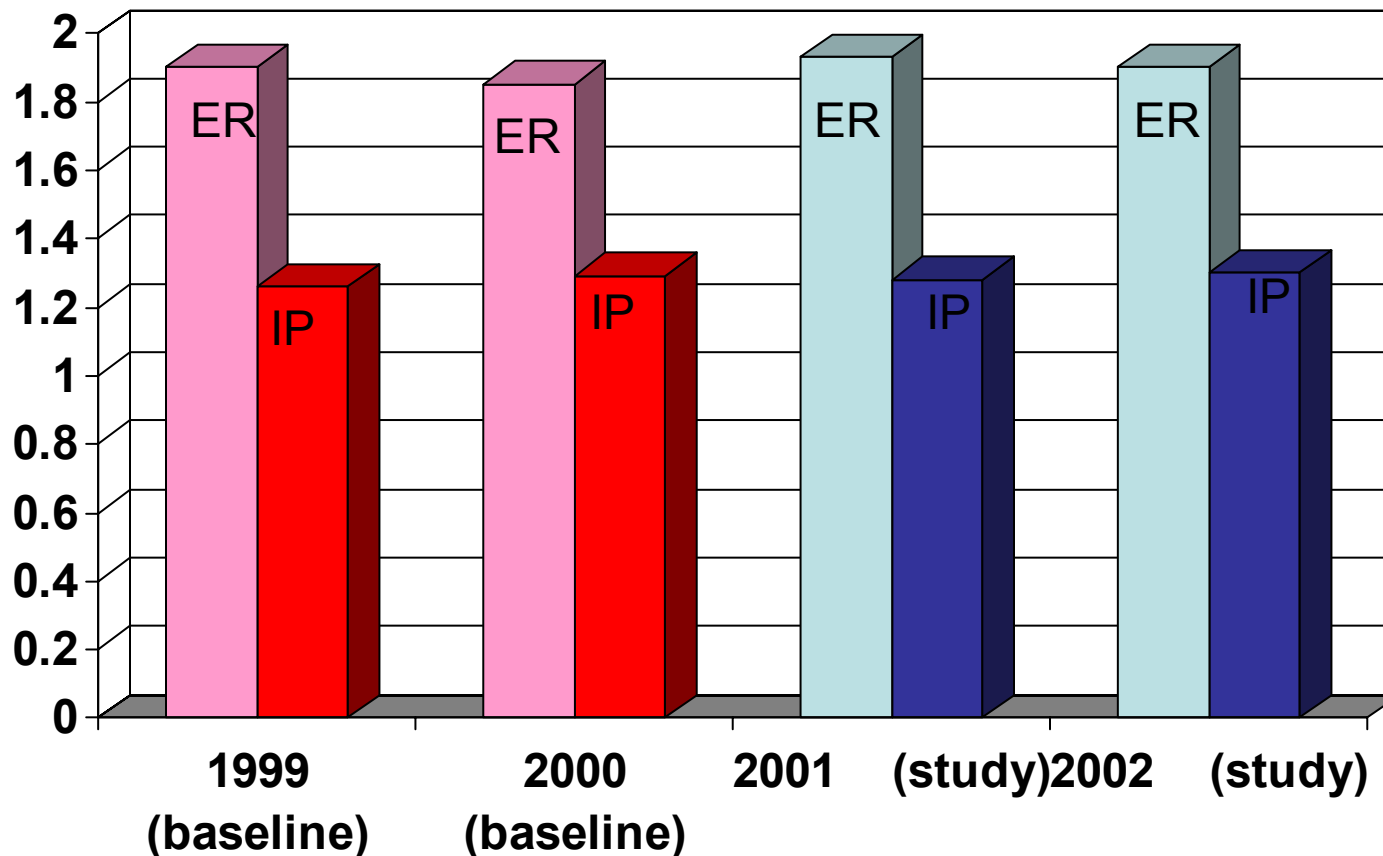


# Expectation is something like...

493.xx ER visits and IP stays/1000 planwide



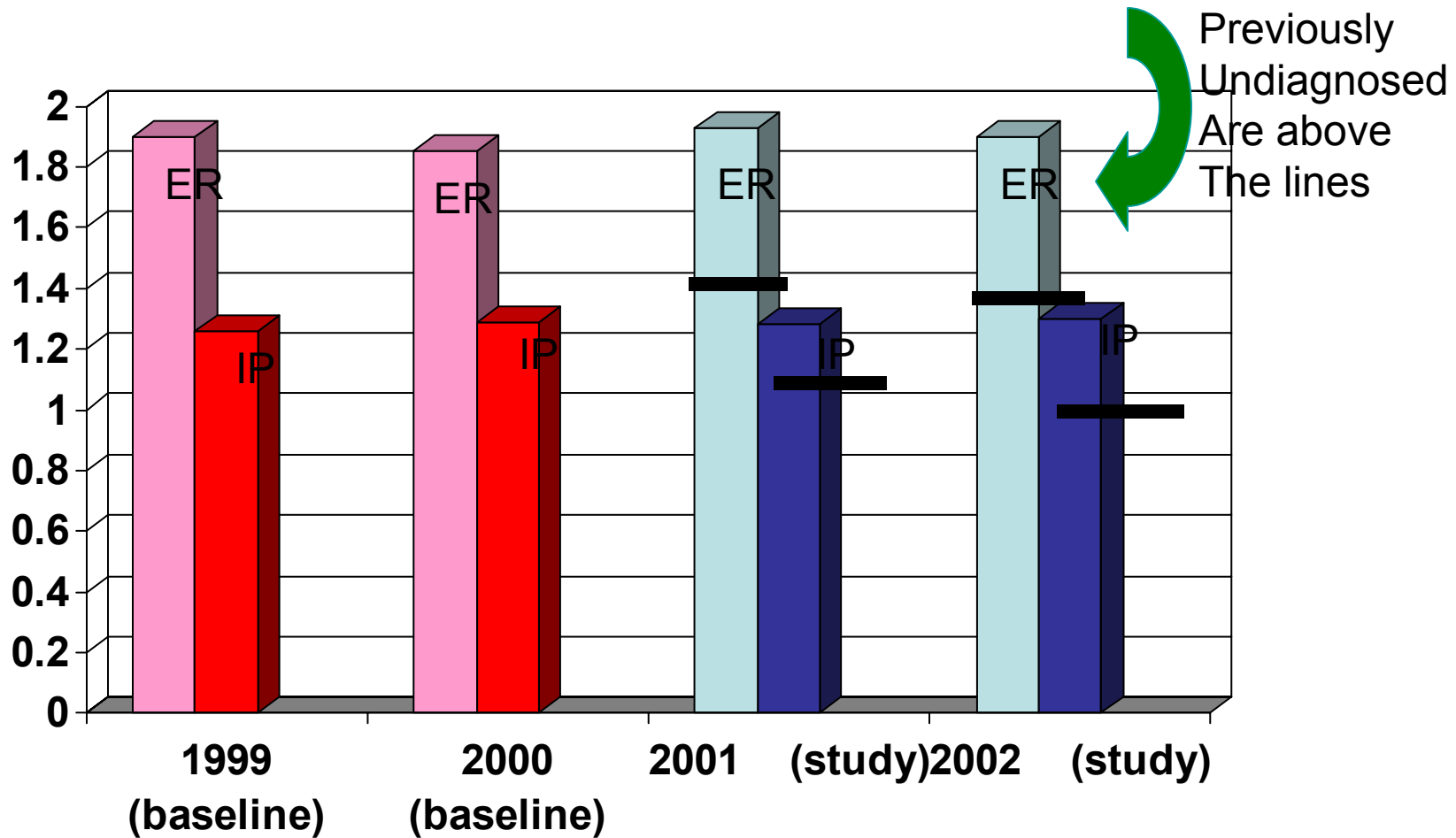
# Plausibility indicator Actual: Validation for Asthma savings from same plan including ALL CLAIMS for asthma 493.xx ER visits and IP stays/1000 planwide



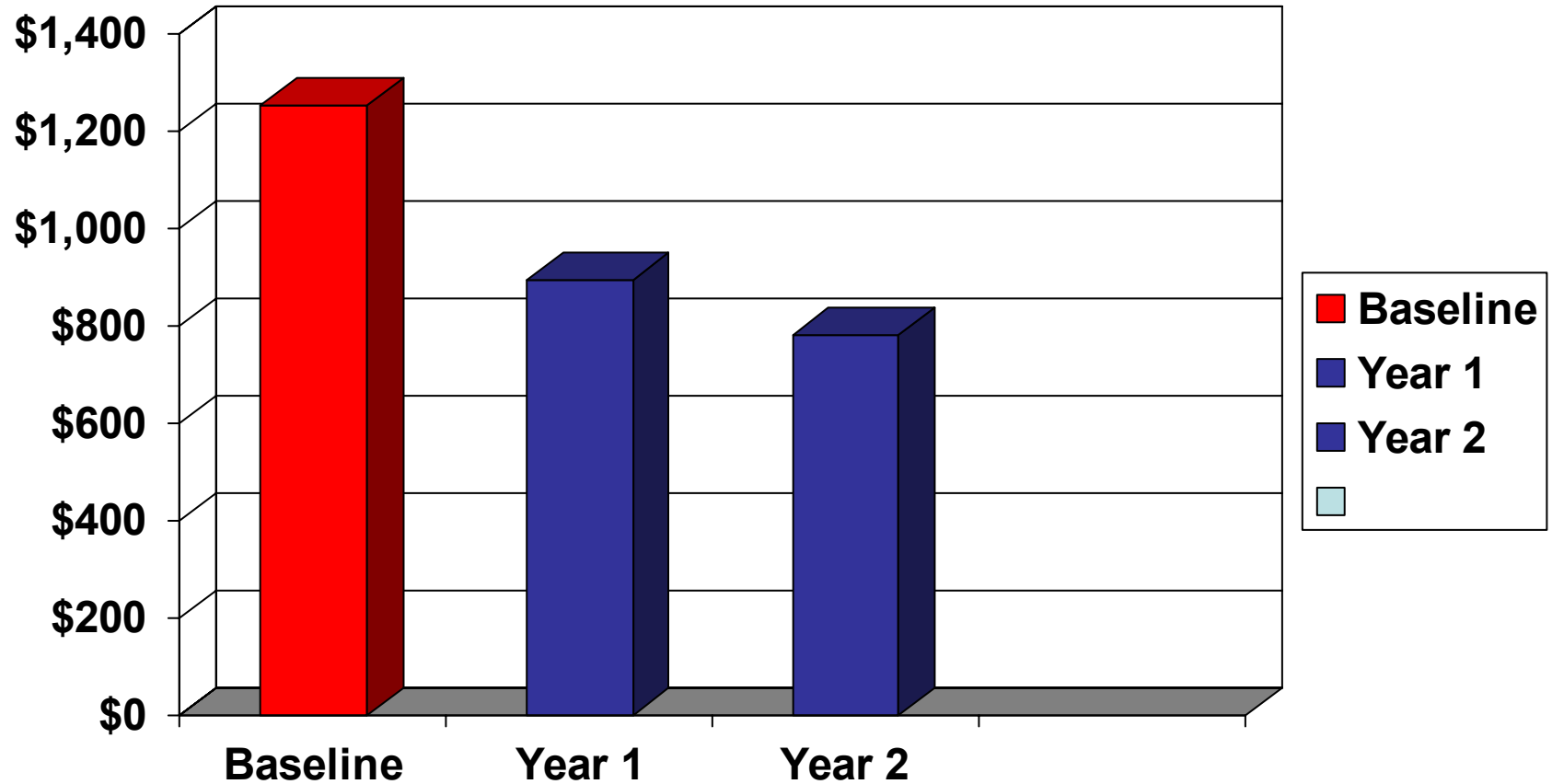
# We then went back and looked...

- ...at *which* claims the vendor included in the analysis...

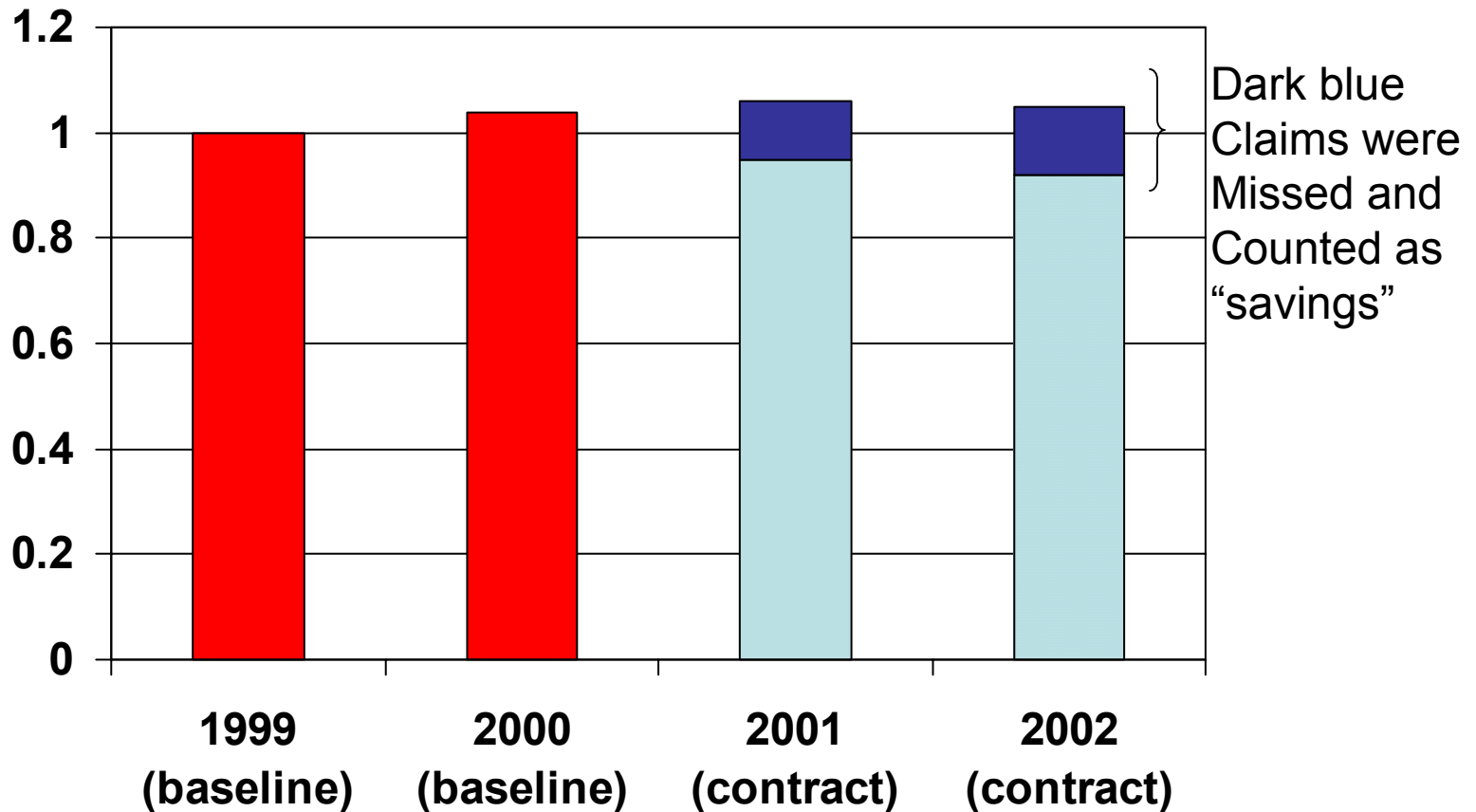
We were shocked, *shocked* to learn that the uncounted claims on previously undiagnosed people accounted for virtually all the “savings”



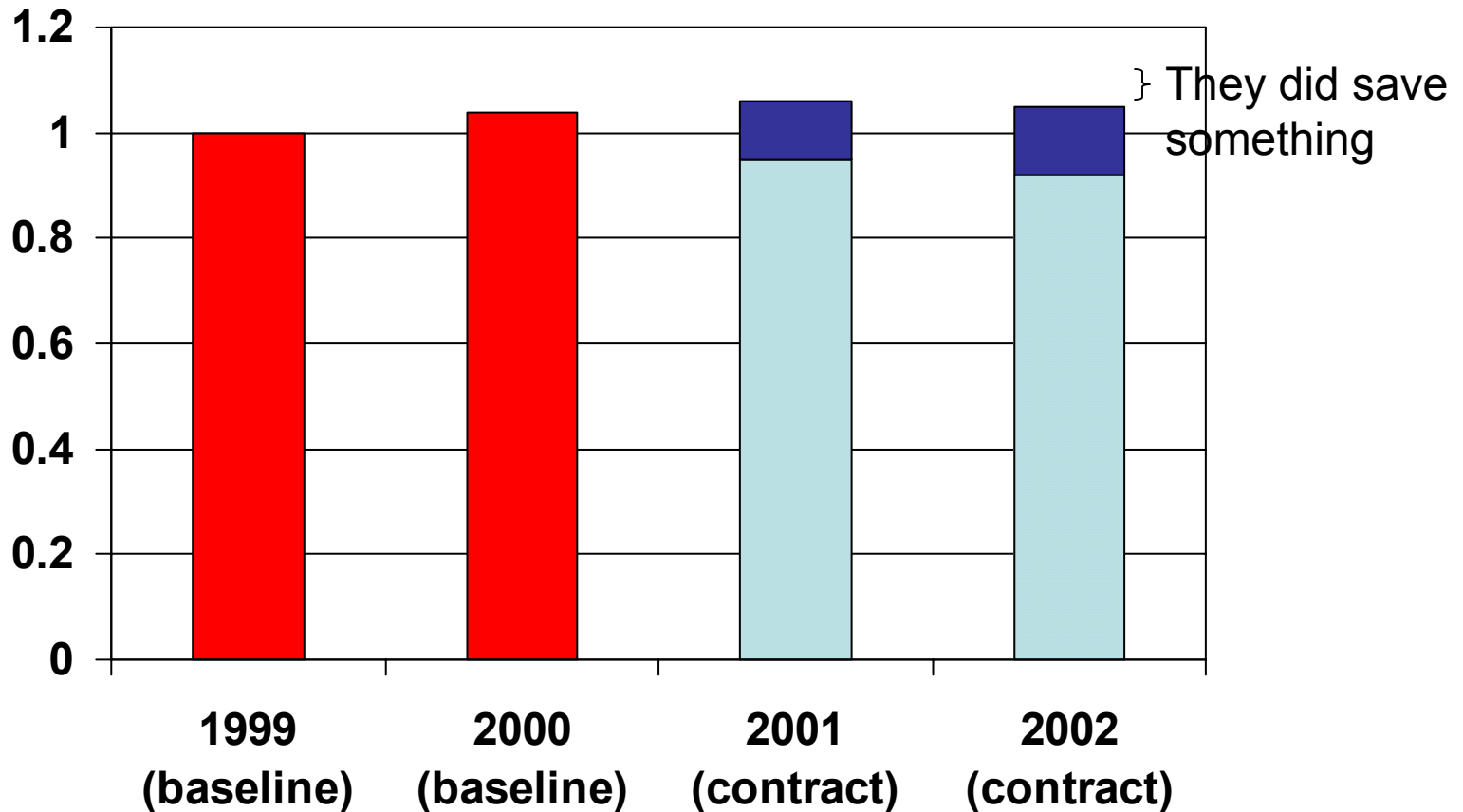
# Example 2: CAD Cost/Member/Month claimed by vendor



# 410 (MI) and 413 (angina) rates/1000 planwide indexed to 1999=1



# 410 (MI) and 413 (angina) rates/1000 planwide indexed to 1999=1





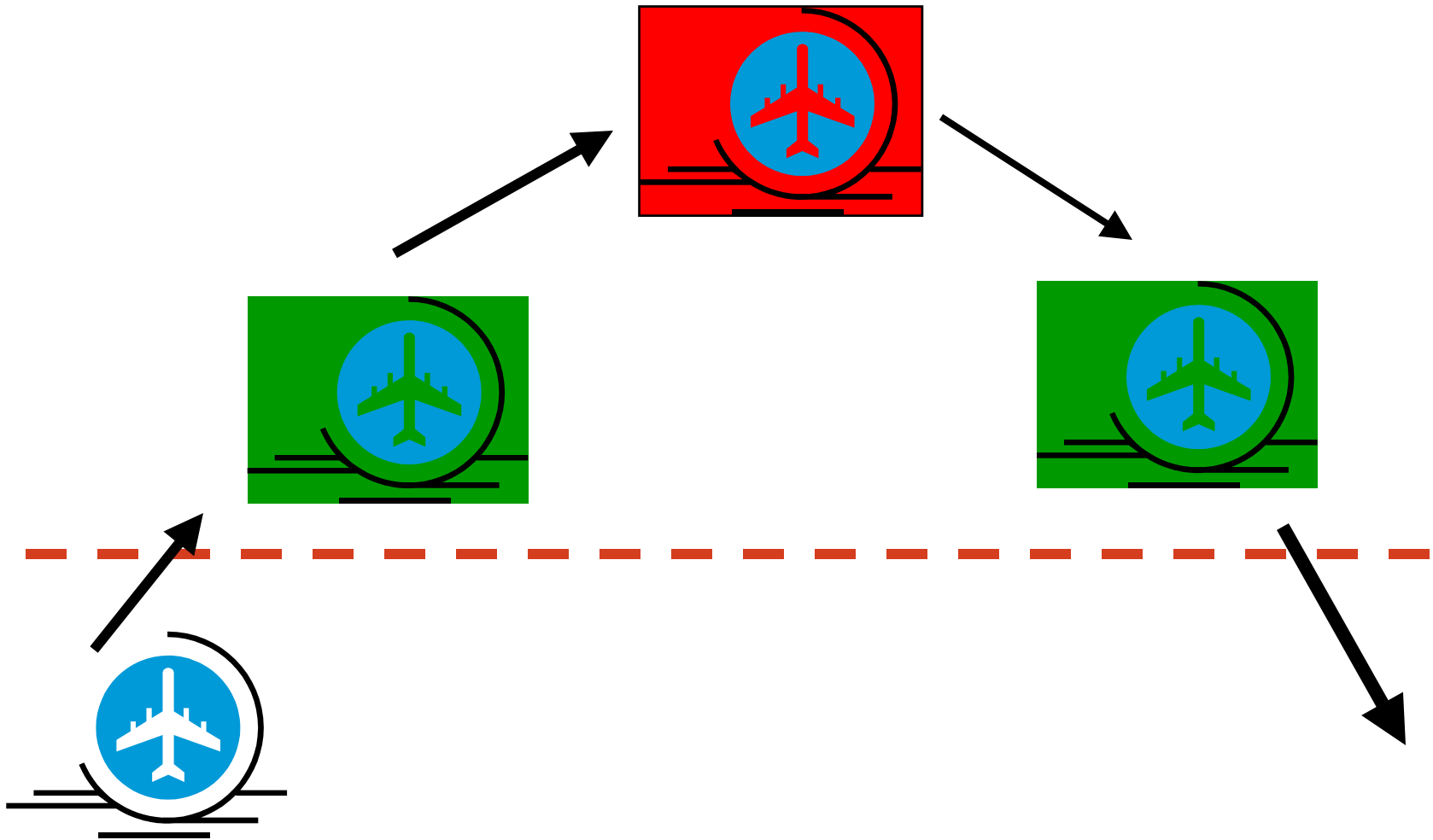
# Diagnosing It, Part Two

- Plausibility indicators: Total unit claims paid which are most relevant to a disease
  - Based on total age/sex-adjusted population
  - Total population cannot regress to the mean because it *is* the mean
  - Easy, intuitive, logical, valid...but this doesn't capture comorbidities...so it's just a diagnostic
- Try tracking your prevalence

# Tracking your prevalence

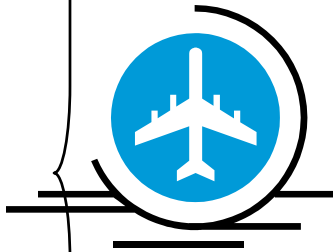
- Is it rising more than 1-2% a year for asthma and CAD?
  - Watch what's happening...

# Recall these slides...



# One hour later...(next claims cycle)

Average  
Plane is  
Still  
10,000  
feet

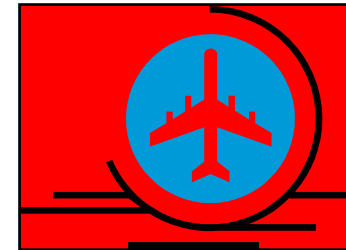


25%



High Claims  
25%

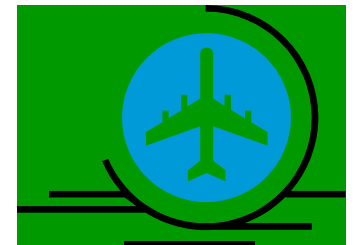
Low claims 25%



Average  
Flight is  
Still  
13,333  
feet

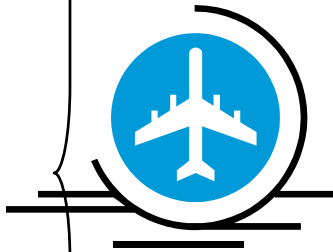
No claim 25%

(c) 2004 DMPC Int'l Inc.

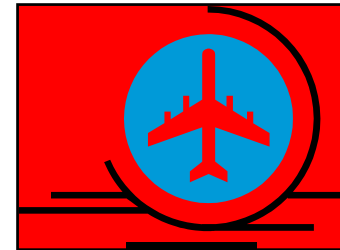


# One hour later...(next claims cycle)

Average  
Plane is  
Still  
10,000  
feet

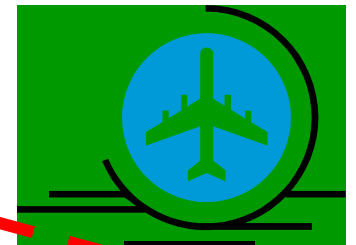


Average  
Flight is  
Still  
13,333  
feet



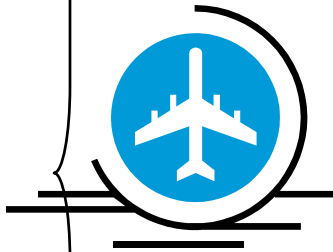
***Except that now all the flights are being Tracked including the ones which have Landed!***

(c) 2004 DMPC Int'l Inc.

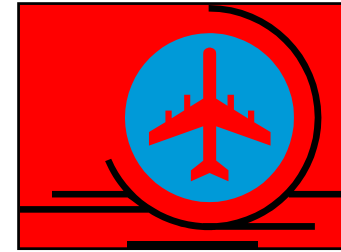


# One hour later...(next claims cycle)

Average  
Plane is  
Still  
10,000  
feet



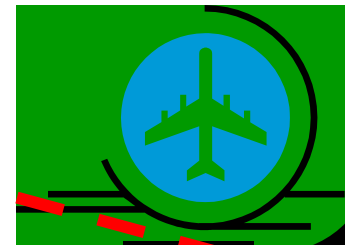
Average  
Flight is  
Still  
13,333  
feet



Measu  
ment is  
10,000  
feet

***Except that now all the flights are being  
Tracked including the ones which have  
Landed!***

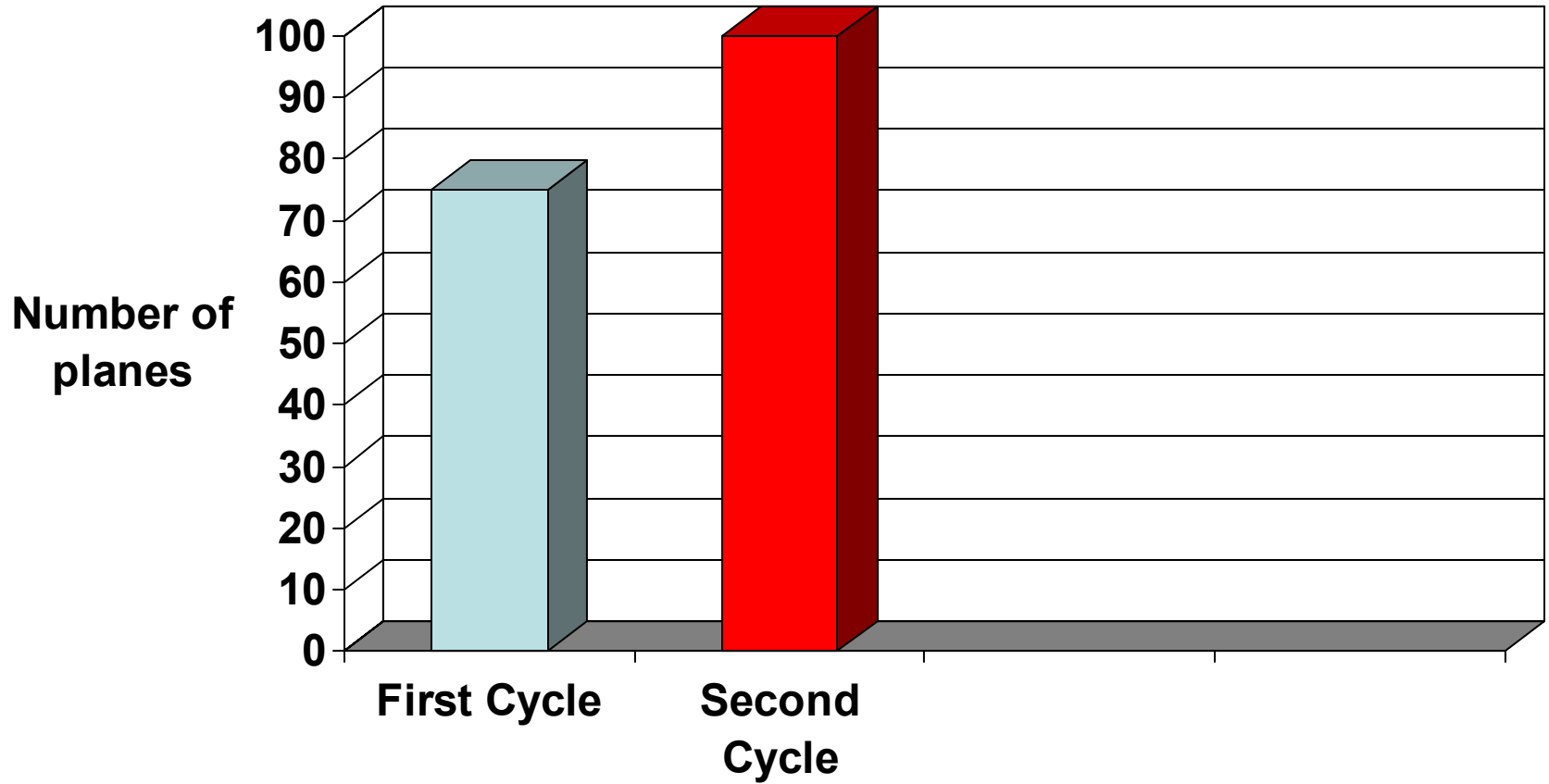
(c) 2004 DMPC Int'l Inc.



# What else is happening besides that missed regression to the mean?

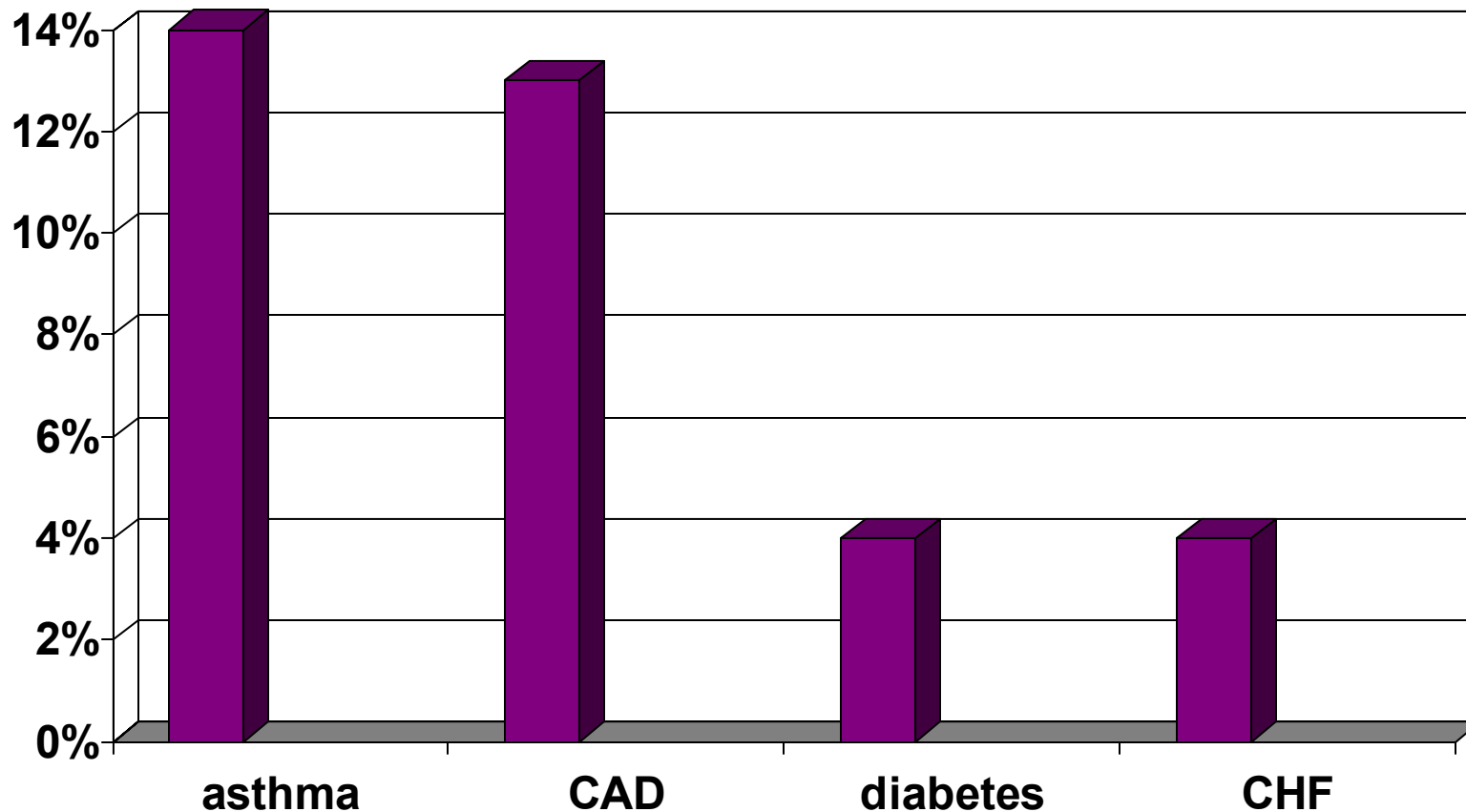
- Assume there are 100 planes in the system

# Number of planes increases in each claims cycle





# Actual data—year-over-year prevalence increase at one health plan



# Summary: Identifying the Problem using the two diagnostics

- Diagnostic #1: Unit claims across entire population...unit claims in targeted diseases should fall by *more than* gross savings claimed (in %)
  - Otherwise some people got missed
- Diagnostic #2: Prevalence increase year over year should be roughly 1-2% in asthma and CAD, maybe 3-4% in diabetes (assuming no change in demographics)

# What to do about it

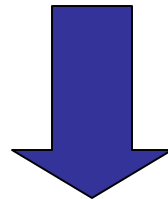
- Choice #1--Plausibility indicators: Total unit claims paid which are most relevant to a disease
  - You can just count these but you miss comorbidities
- Choice #2--Freezing the Population: DO NOT COUNT anybody who pops onto the radar screen following the first of the year (in baseline and in study period) *together* with the previous population
  - You should count “newly incident” members separately

# What to do about it

- Choice #1--Plausibility indicators: Total unit claims paid which are most relevant to a disease
  - You can just count these but you miss comorbidities
- Choice #2--Freezing the Population: DO NOT COUNT anybody who pops onto the radar screen following the first of the year (in baseline and in study period) *together* with the previous population
  - You should count “newly incident” members separately

# Freezing the Population

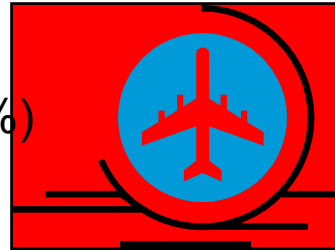
- FOUR steps
  1. Identified (“prevalent”) population (2002)
  2. Measure their claims in 2003 (“baseline”)
  3. Identify the population the same way in 2003 as you did in 2002
  4. Measure their claims in 2004 (“study period”)



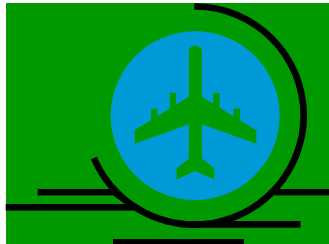
Watch what happens with the planes if we do this...

# 2002: Identify group to measure for baseline claims in 2003

High claims (33%)



Low claims (67%)



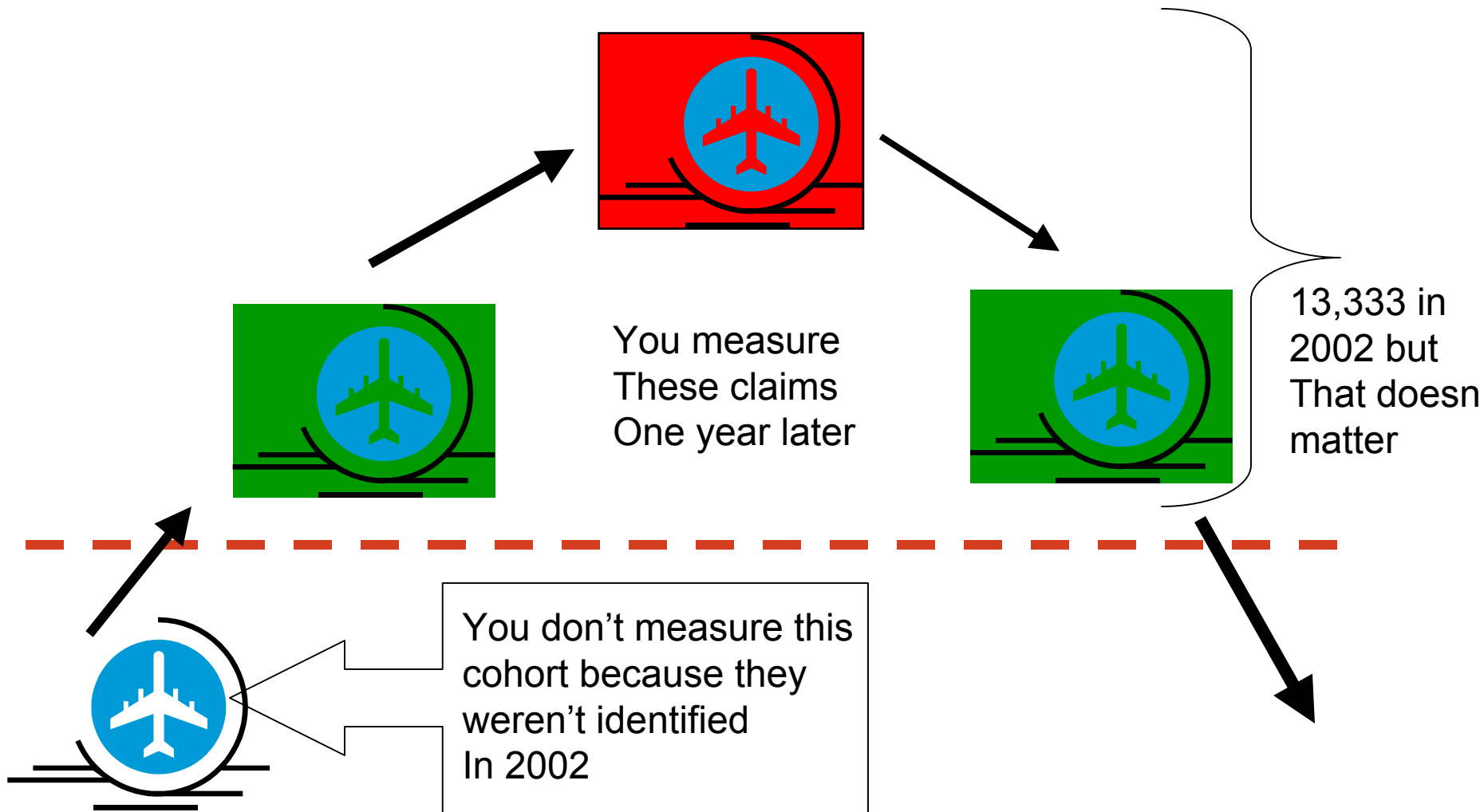
Above the line are datapoints which are found and measured

---

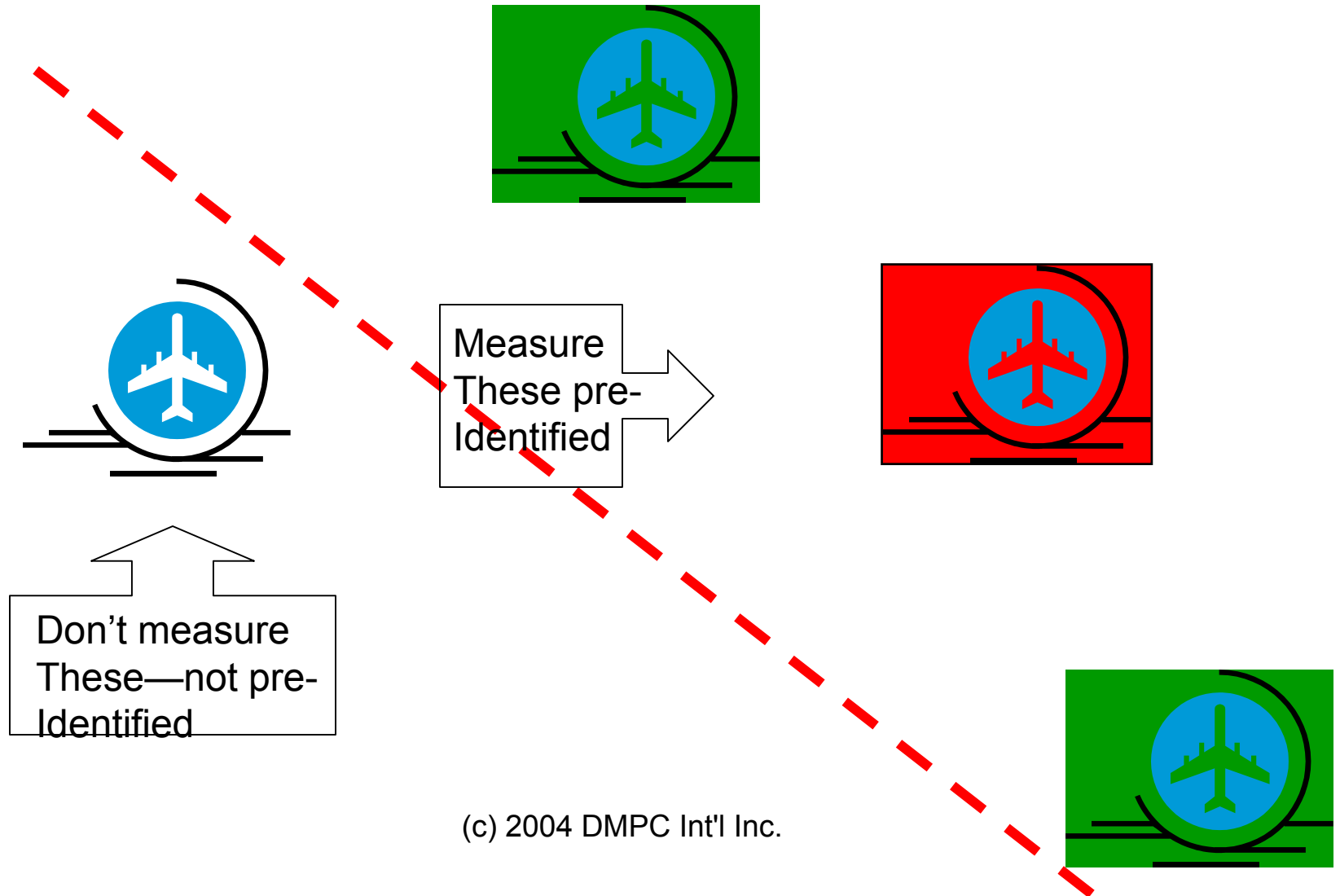


No claim

# Fast forward to 2003, where you measure the claims

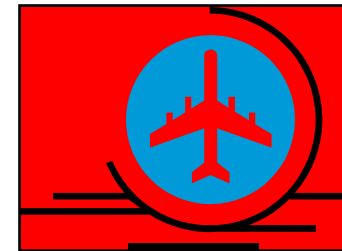
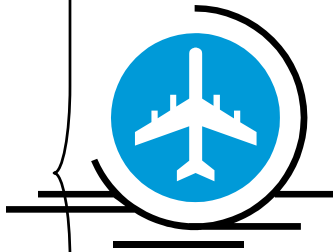


# Your baseline is the 2003 claims of the 2002 identified cohort, or 10,000 feet



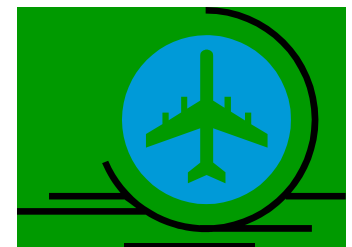
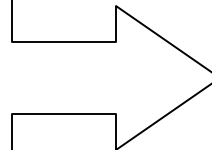


In 2003 you *identify* the prevalent population exactly the same way as you did in 2002

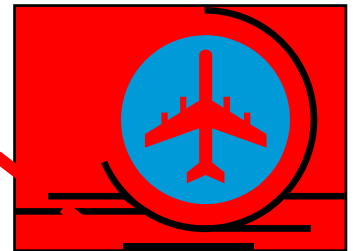
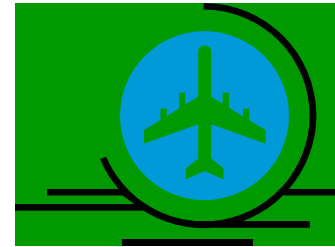


Average Flight is Still 13,333 Feet but That still Doesn't Matter— You are Just IDing

Why don't you Measure these Guys?



And in 2004 you measure the claims of the people you identified in 2003



You get the exact same  
10,000 feet that you got in the  
Baseline measurement of the  
Pre-identified population!

# Note that...

- Even though the dotted red line is crooked, it is equally crooked in BOTH periods because you are measuring the SAME way

# Recall this Baseline slide

	2002	2003
Asthmatic #1	1000	100
Asthmatic #2	0	1000
Study Period— usual methodology	1000	550

# Recall this Baseline slide

	2002	2003
Asthmatic #1	1000	100
Asthmatic #2	0	1000
What happens if you shake the RTM out?		

# Recall this Baseline slide

	2002	2003
Asthmatic #1	1000	100
Asthmatic #2	0	1000
What happens if you shake the RTM out?	No baseline— ID only	100

(c) 2004 DMPC Int'l Inc.

# What just happened?

- Instead of making incorrect assumptions about what claims *the newly incident population would have incurred* if they had been identified before they were incident, you DON'T ASSUME ANYTHING.
- You simply don't count them
  - You can also compare newly incident populations in 2003 and 2004 to each other...*but don't mix them with the prevalent population*

# How does this differ from the methodology of comparing trended pre to post?

- In the pre-post comparison, the identified and baseline period of the “pre” are the same, so the incident population is mixed in and you get RTM in the post period
- In this methodology, you take the “pre” population’s RTM *OUT* of the equation by doing the baseline measurement in the year after you identify them
  - So there is no incident population pollution



# Which is more purely parallel?

Baseline Group	Compared to inflation-adjusted...		
2002 prevalent group's 2003 claims	2003 prevalent group's 2004 claims		
2003 Newly incident members actual claims, 2003	2004 Newly incident members actual claims, 2004		

# Which is more purely parallel?

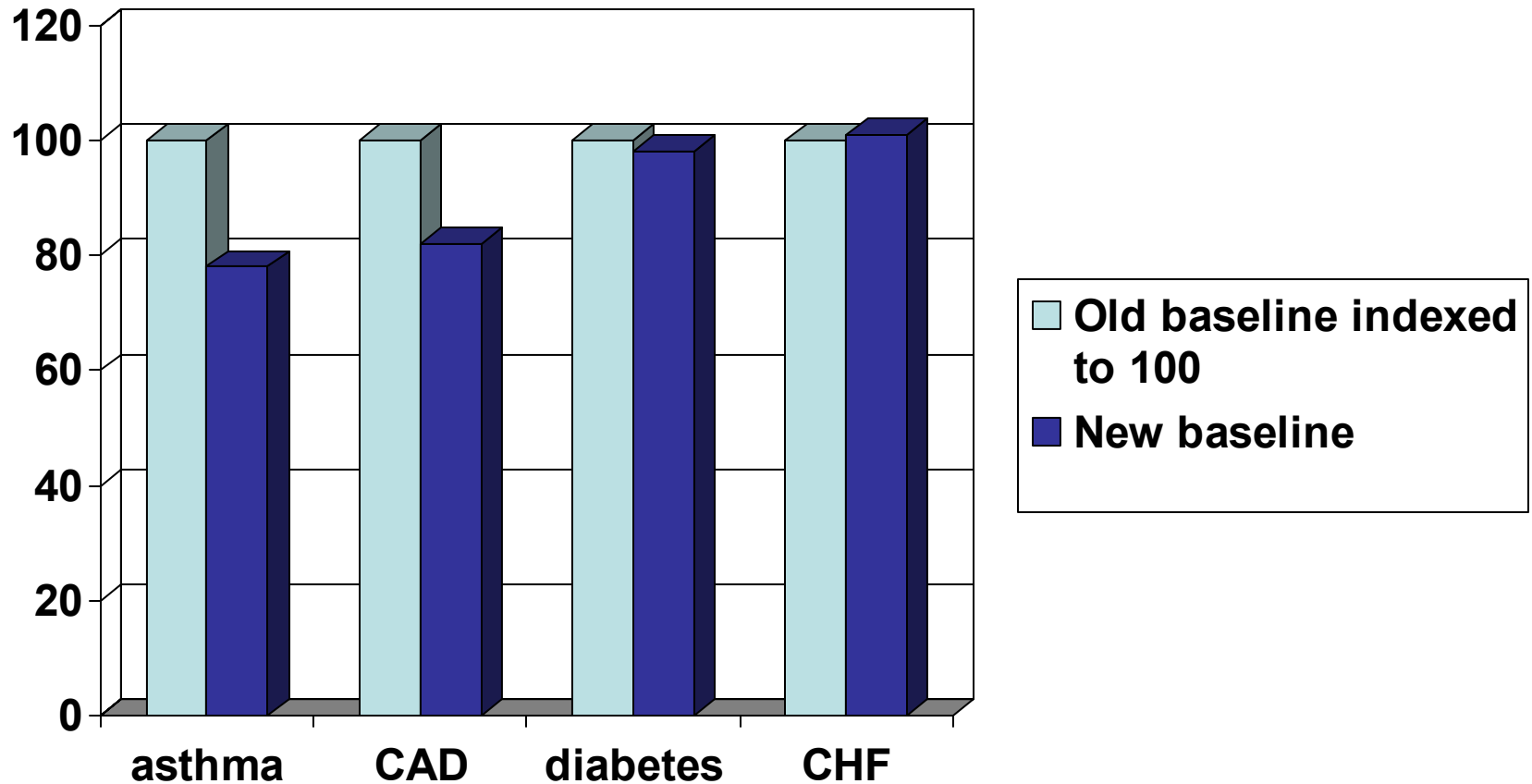
Baseline Group	Compared to inflation-adjusted...
2002 prevalent group's 2003 claims	2003 prevalent group's 2004 claims
2003 Newly incident members actual claims, 2003	2004 Newly incident members actual claims, 2004

Baseline Group	Compared to inflation-adjusted...
2003 prevalent group's 2003 claims	2003 prevalent group's 2004 claims plus 2004 incident group assumed to have cost 2003 prevalent group's claims in 2003

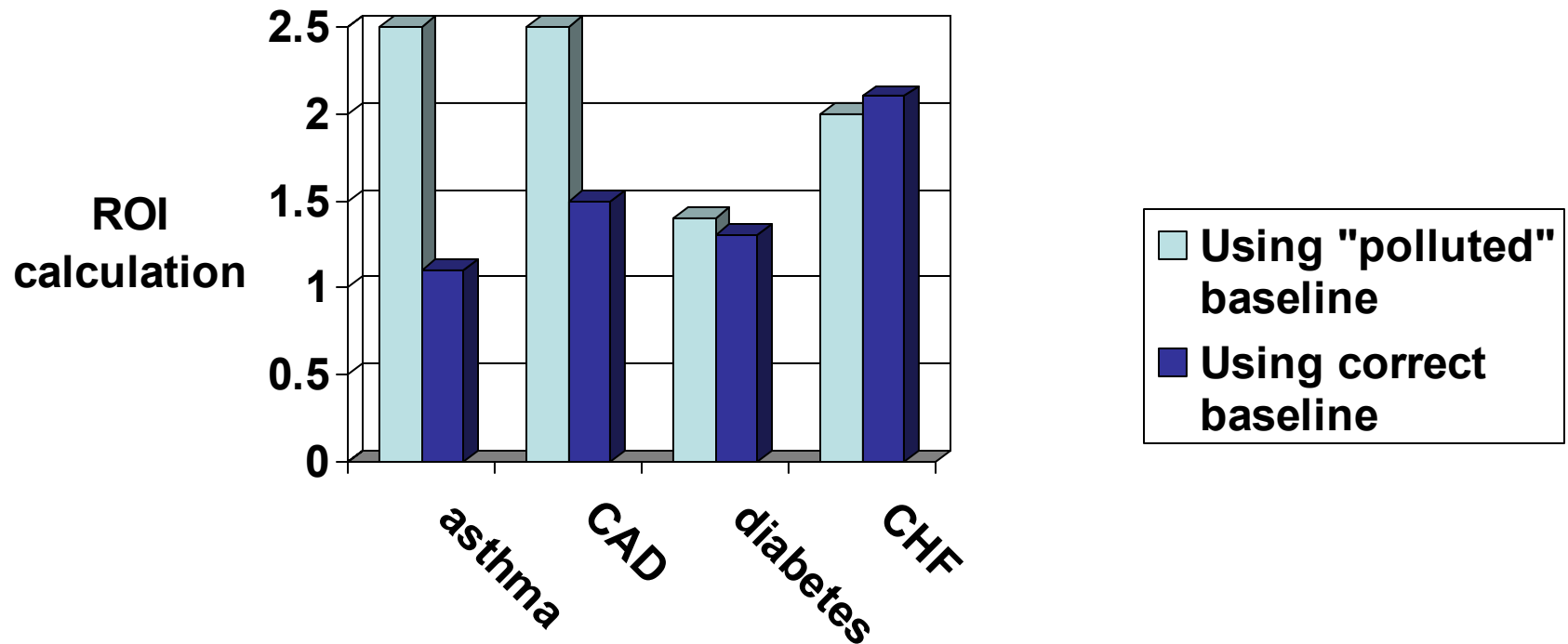
# What happens when you re-do baseline with new methodology?

- A health plan recalculated its baseline for four diseases to see what the impact would be
  - In each case “100” on the next slide represents the baseline with 2001 data
  - The number next to it represents how the baseline changed by using 2001 to identify people and 2002 to measure those people vs. 2001 to identify and measure

What happens in one health plan when you change the way you do this (n=1 plan c. 500,000 members) where you previously had 12 months of baseline data



# Impact on ROI from disease management



# What to do about it

- Choice #1--Plausibility indicators: Total unit claims paid which are most relevant to a disease
  - You can just count these but you miss comorbidities
- Choice #2--Freezing the Population: DO NOT COUNT anybody who pops onto the radar screen following the first of the year (in baseline and in study period) *together* with the previous population
  - You should count “newly incident” members separately
- Choice #3—Create a dummy baseline using the RTM effect between two non-DM years

# Create a dummy baseline using the RTM effect between two non-DM years

- Same as previous one except you simply calculate the difference

# Baseline—the old way

	2002	2003
Asthmatic #1	1000	100
Asthmatic #2	0	1000
Study Period— usual methodology	1000	550



# Baseline—Adding back in the Baseline year claims for new Dx

	2002	2003
Asthmatic #1	1000	100
Asthmatic #2	0	1000
Study Period— usual methodology	500	550

# What happens if you adopt one of these three fixes

- Choice #1--Plausibility indicators: Total unit claims paid which are most relevant to a disease
  - You can just count these but you miss comorbidities
- Choice #2--Freezing the Population: DO NOT COUNT anybody who pops onto the radar screen following the first of the year (in baseline and in study period) *together* with the previous population
  - You should count “newly incident” members separately
- Choice #3—Recalculate the baseline as new members are found

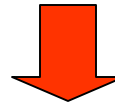
# Impact if you adopt one of these approaches

- Size of ROI from DM: **lower** 
- Measurability of ROI from DM: **Higher** 

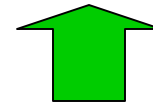
# Impact

- Size of ROI from DM:

**lower**



- Measurability of ROI from DM : **Higher**



- Credibility of ROI from DM: **Priceless**

