

# How to Measure Outcomes in Disease Management

May 2008

Al Lewis

Disease Management Purchasing Consortium

[www.dismgmt.com](http://www.dismgmt.com)

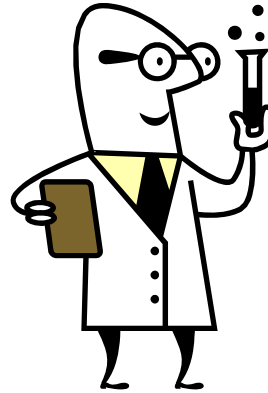
All slides © 2008 DMPC

# Agenda

- Who am I? Why am I here?
- Why your program results may not be valid
  - Back by popular demand, planes on the ground analogy
  - Examples of mistakes driven by pre-post
- How to Measure validly using utilization rates
  - Real world examples and interesting findings

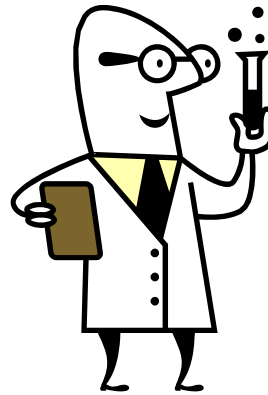
# Who am I? Why am I here?

- Invented disease management (according to Google) and founded DMAA



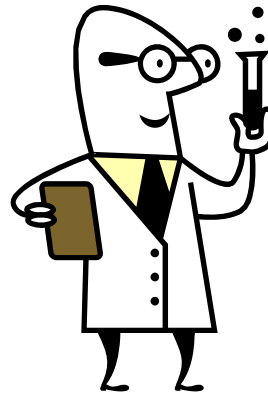
# Who am I? Why am I here?

- Invented disease management (according to Google) and founded DMAA
- Have done more procurements than anyone on earth



# Who am I? Why am I here?

- Invented disease management (according to Google) and founded DMAA
- Have done more procurements than anyone on earth
- Have Certified 30 payors for savings validity and 60 individuals as qualified to measure outcomes



# Agenda

- Who am I? Why am I here?
- Why your program results may not be valid
  - Back by popular demand, planes on the ground analogy
  - Examples of mistakes driven by pre-post
- How to Measure validly using utilization rates
  - Real world examples and interesting findings

# Current Situation

- Show of hands time –agree or disagree:
  - Is your reported ROI generally valid?

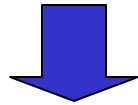
# Why pre-post is not the best way to look at DM outcomes

- Watch what happens in the following analogy and then...
- ... Watch what happens in the following hypothetical using the exact pre-post methodology which you are being shown



# Let's look at this by analogy to airplanes

- 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)



What is the average altitude in this example?

# 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

# 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

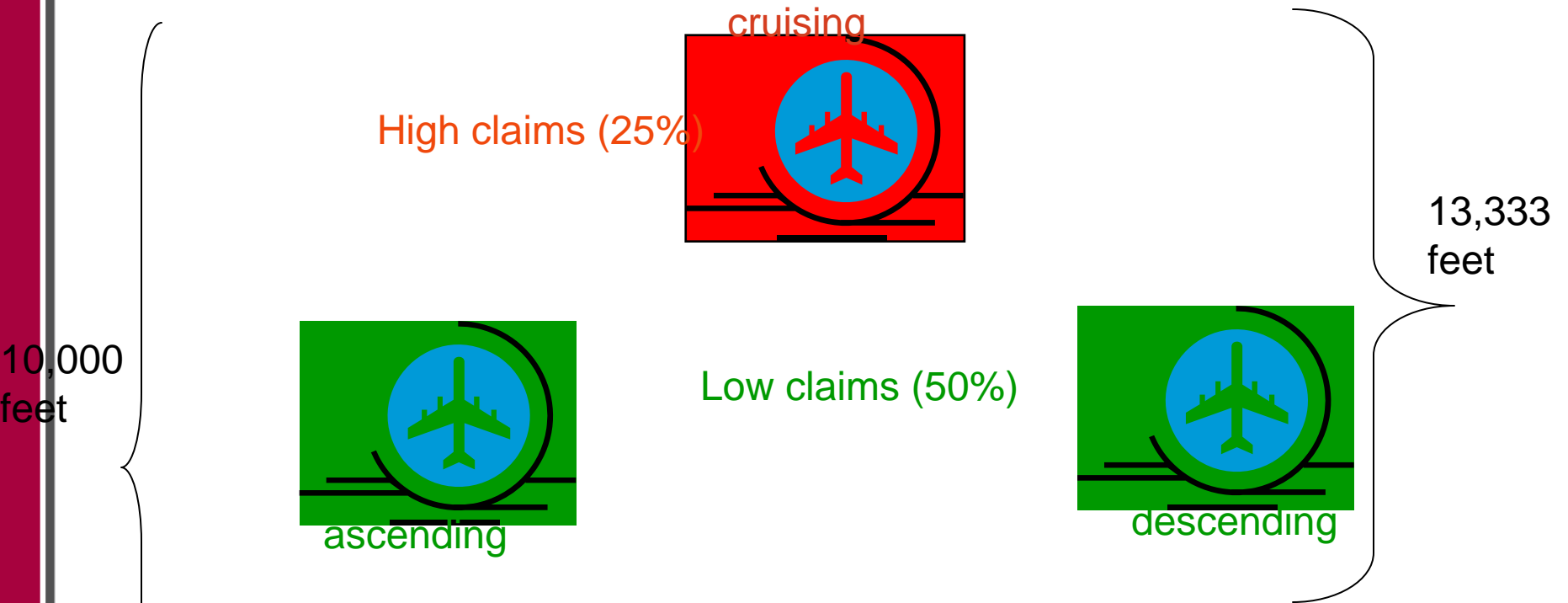
# Further assume...

- 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



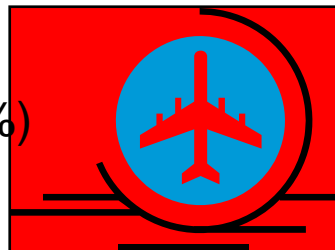
No claim (25%)

## The benefits consultant/DM vendor approach – moving from planes to claims

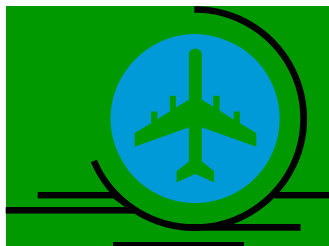
- Tracks ALL people with claims for the disease, high or low, in the baseline
- Emphasizes finding low utilizers for a population-based approach
  - Equivalent to finding all *flights* including ascending and descending
  - Average baseline altitude (2/3 at 10,000, 1/3 at 20,000) is: 13,333 feet

# You measure the claims on ALL patients with claims

High claims (33%)



Low claims (67%)



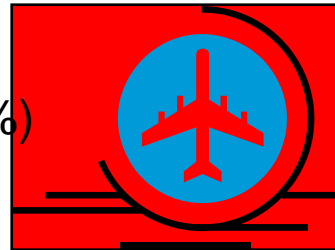
Above the line are datapoints which are found and measured



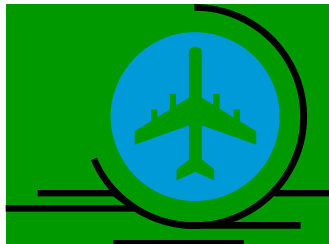


# You measure the claims on ALL patients with claims

High claims (33%)



Low claims (67%)



Above the line are datapoints which are found and measured

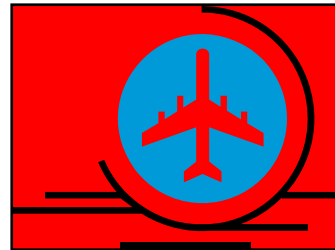
Why don't you measure these guys?



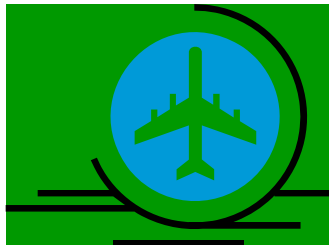
# You measure the claims on ALL patients with claims

13,333 Feet  
On average

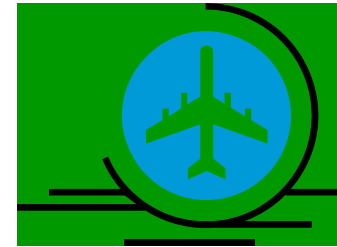
High claims



These get Found in The claims pull



Low claims



Above the line are datapoints which are measured

Below the line is not included in measurement

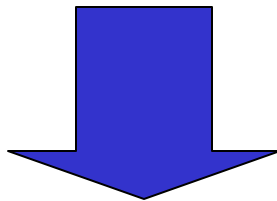
**Because they have no relevant claims to be found**



No claim

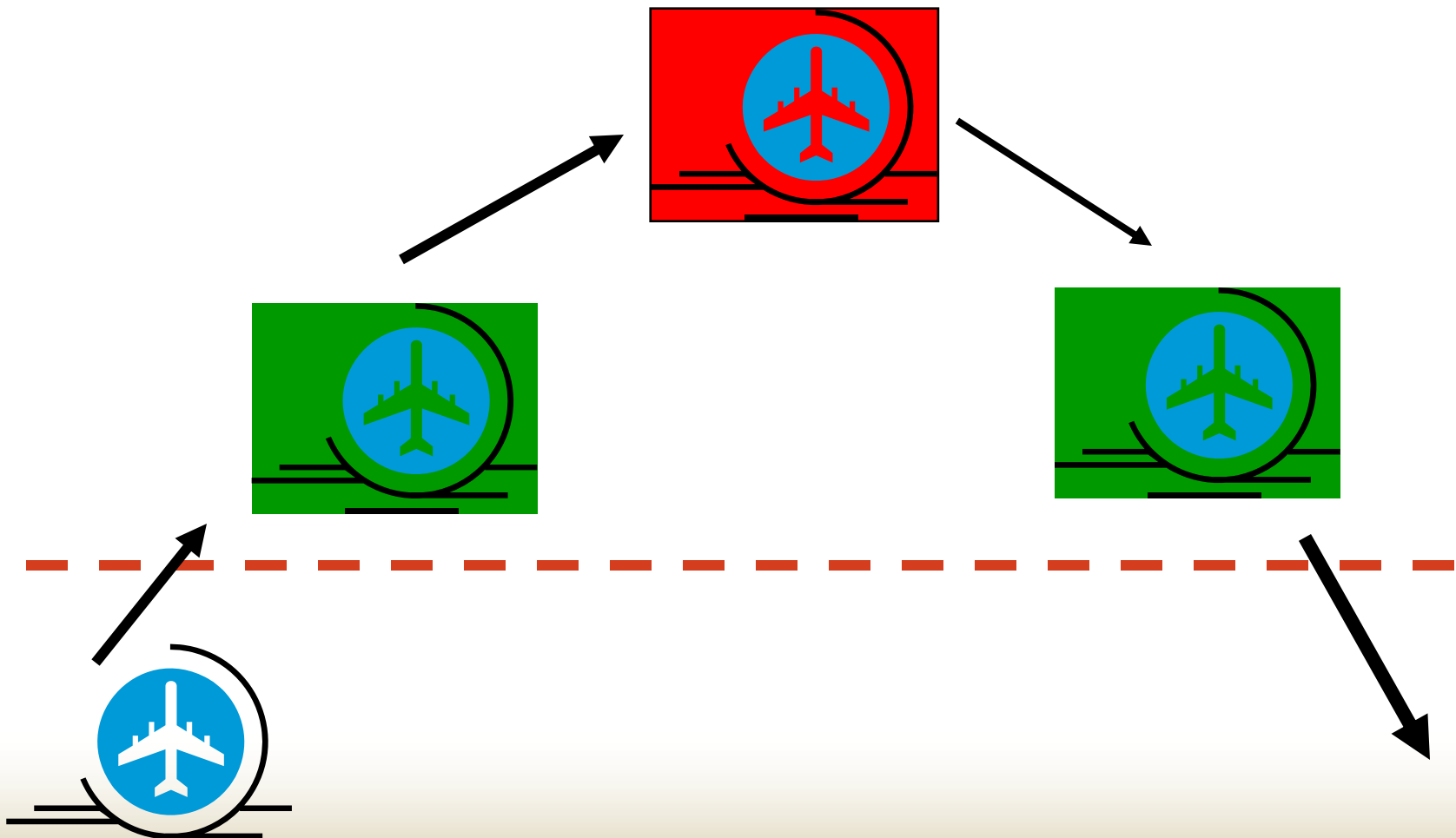
# The conventional approach

- 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)

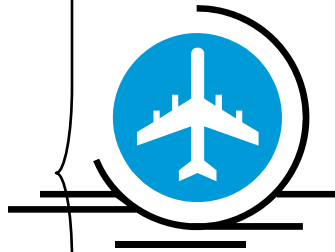


# We can all agree that...

- 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



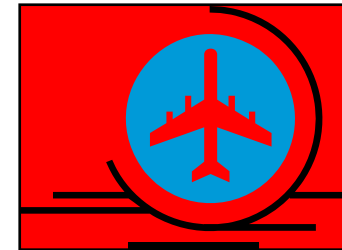
25%

Low claims

25%



High Claims  
25%



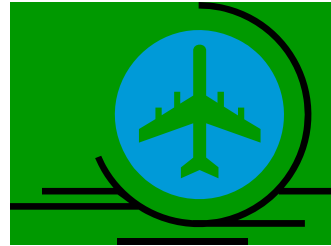
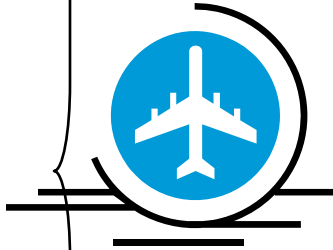
No claim 25%

Average Flight is Still 13,333 feet

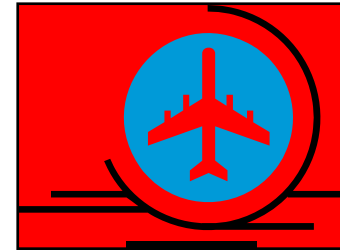


# One hour later...(next claims cycle)

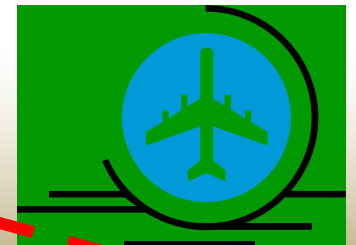
Average  
Plane is  
Still  
10,000  
feet



Average  
Flight is  
Still  
13,333  
feet

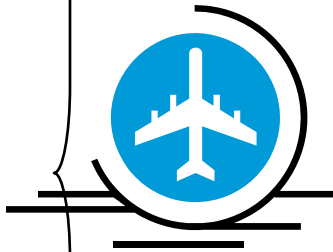


***Except that now all the flights are being  
Tracked including the ones which have  
Landed!***

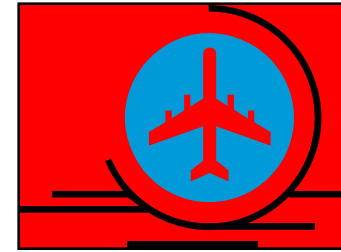


# One hour later...(next claims cycle)

Average  
Plane is  
Still  
10,000  
feet

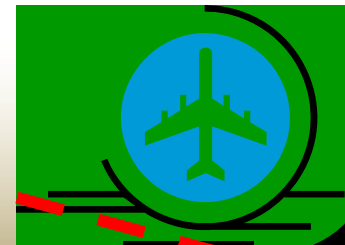


Average  
Flight is  
Still  
13,333  
feet



Measu  
ment is  
10,000  
feet

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



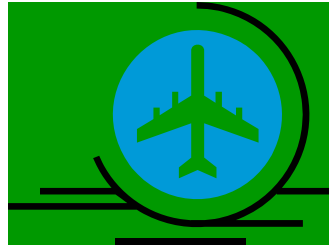
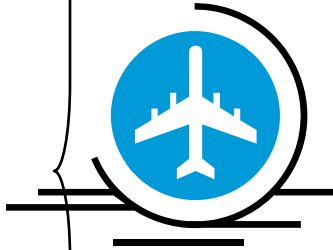


# But wait...Some people say...

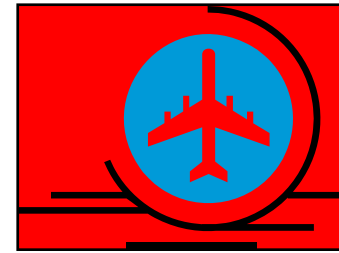
- “We don’t track the people with no claims in the ‘post’ period in order to maintain equivalency with the ‘pre’ period”
- “The member has to re-trigger [with claims] each year to be counted” (“annual requalification”)
  - So this bias shouldn’t happen because we don’t measure the zeros in EITHER period

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

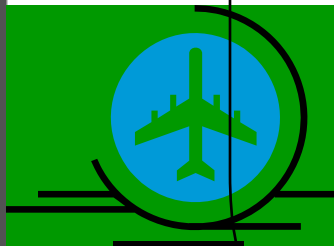
Average Plane is Still 10,000 feet



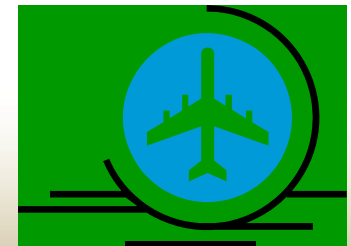
Average Flight is Still 13,333 feet



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 “plane on the ground” has an event and then recovers, that person is put on drugs (asthma, beta blockade, antihyperlipidemics etc.)
  - And for some period of time they comply

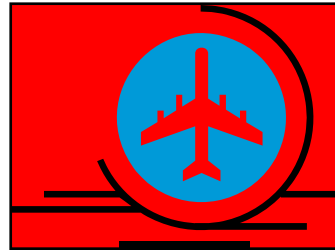


# The Fallacy

- 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

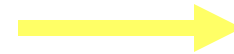
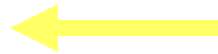
High claims



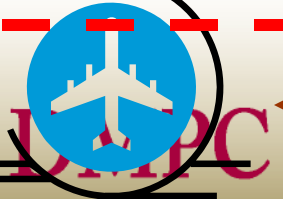
Middle claims



Taking preventive drugs  
And identifiable as such



NOT taking preventive drugs and NOT  
Identifiable



# Pre-post would be valid if only...

- People were as likely to take preventive meds before an event as after it
- People had an implanted chip which notified the health plan as soon as they had a chronic disease even if they didn't incur claims to treat it

# Why pre-post is not the best way to look at DM outcomes

- Watch what happens in the following analogy and then...
- ... Watch what happens in the following hypothetical using the exact pre-post methodology which you use



# In this example

- Assume that “trend” is already taken into account correctly
- Focus on the baseline and contract period comparison

**Base Case: Example from Asthma**  
First asthmatic has a \$1000 IP claim in 2005

	<b>2005 (baseline)</b>	<b>2006 (contract)</b>
Asthmatic #1	1000	
Asthmatic #2		
Cost/asthmatic		

## Example from Asthma

Second asthmatic has an IP claim in 2006 while first asthmatic goes on drugs (common post-event)

	<b>2005 (baseline)</b>	<b>2006 (contract)</b>
Asthmatic #1	1000	100
Asthmatic #2	0	1000
Cost/asthmatic		<b>What is the Cost/asthmatic In the baseline?</b>

# Cost/asthmatic in baseline?

	<b>2005 (baseline)</b>	<b>2006 (contract)</b>
Asthmatic #1	1000	100
Asthmatic #2	0	1000
<b>Cost/asthmatic</b>	<b>\$1000</b>	Vendors don't count #2 in 2005 bec. he can't be found

# Cost/asthmatic in contract period?

	<b>2005 (baseline)</b>	<b>2006 (contract)</b>
Asthmatic #1	1000	<b>100</b>
Asthmatic #2	0	<b>1000</b>
<b>Cost/asthmatic</b>	<b>\$1000</b>	<b>\$550</b>

# Why Pre-Post Overstates Savings

	<b>2005 (baseline)</b>	<b>2006 (contract)</b>
Asthmatic #1	1000	<b>100</b>
Asthmatic #2	0	<b>1000</b>
<b>Cost/asthmatic</b>	<b>\$1000</b>	<b>\$550</b>

In this case, a “dummy population” falls 45% on its own without DM due to #2 being a “plane on the ground”

# The Valid Way to Check Pre-Post Savings Claims

- You look at the event rates overall in the plan (or in your own organization if large enough) over time
  - As though you were measuring a birth rate. Very simple
  - As in this example, count total IP (and ER) events, divide by 1000

# Asthma events in the payor as a whole – the plausibility check

	<b>2005 (baseline)</b>	<b>2006 (contract)</b>
Asthmatic #1	1000	<b>100</b>
Asthmatic #2	0	<b>1000</b>
Inpatient events/year	1	1

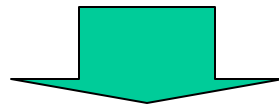


# This is called a “plausibility check”

- You use plausibility checks all the time in your everyday life and don't think twice about them
  - But for some reason in DM people rely on faith instead

# If you're still not convinced...

- Imagine applying a pre-post to a maternal program to measure outcomes
  - Find every woman who had a baby in the last two years (or underwent fertility or stopped filling BC bill prescriptions)
    - This is (for example) 10,000 babies last year and the year before



This cohort will account for about 5000 babies this year  
The “plausibility indicator” of counting babies yields 10,000

# That little insight is what's missing from your reports

- Pre-post is only valid if confirmed with event rates...
- The question is, did you reduce events in the conditions in which you were trying to reduce events?
  - Otherwise where did the savings come from, if not from event reduction in the conditions the program was managing?

# If you're still not convinced...

- Imagine applying a pre-post to a maternal program to measure outcomes
  - Find every woman who had a baby in the last two years (or underwent fertility or stopped filling BC bill prescriptions)
    - This is (for example) 10,000 babies last year and the year before

# To sum up

- This is not “He said. She said.” It’s a math proof
  - “Everyone is entitled to their own opinions, but not to their own facts.”
- The fundamental flaw in the “population-based pre-post” approach is that you cannot identify the whole population in advance, as in this example due to “planes on the ground.”
  - It is not “population-based” at all but rather a special case of regression to the mean

# Plausibility Analysis

- ALL ER and IP events primary-coded to a condition (see next slide) divided by all people in your employment
  - If you are <50,000 employees, make your health plan use Book of Business
- NOT people, just claims
- Write to me for the 16 watch-outs in compiling this simple metric (*e.g.*, what to do with transfers, admits at year-end etc.)

# All based on this one table

<b>Disease Program Category</b>	<b>ICD9s (all .xx unless otherwise indicated)</b>
<b>Asthma</b>	493 (including 493.2x <sup>[1]</sup> )
<b>Chronic Obstructive Pulmonary Disease</b>	491.1, 491.2, 491.8, 491.9, 492, 494, 496, 506.4
<b>Coronary Artery Disease (and related heart-health issues)</b>	410, 411, 413, 414
<b>Diabetes (CAD codes above will also indicate the success of the diabetes program)</b>	250
<b>Heart Failure</b>	428, 404.01, 404.03, 404.11, 404.13, 404.91, 404.93, 425.0, 425.4

<sup>[1]</sup> 493.2x is asthma with COPD. It could fit under either category but for simplicity we are keeping it with asthma

# Agenda

- Who am I? Why am I here?
- Why your program results may not be valid
  - Back by popular demand, planes on the ground analogy
  - **Examples of mistakes driven by pre-post**
- How to Measure validly using utilization rates
  - Real world examples and interesting findings



# Actual Report example – how to mislead with pre-post

Service category	Expected Cost (adjusted for trend)	Actual cost	Savings
Inpatient	\$137	\$131	\$6
ER	\$8.00	\$7.50	\$0.50
Outpatient	\$59	\$56	\$3
Labs	\$9.00	\$8.80	\$0.20
Office Visit	\$69	\$61	\$8
Drugs	\$131	\$127	\$4

# How you know this is wrong

- Savings in areas which DM doesn't affect (labs)
- Savings in areas which should increase with DM (drugs, MD visits)
  - Any reduction is due to generic substitution or other effects not part of DM
- No disease-specific event rate “plausibility check” on ER and IP
  - We did an event rate check and there was no change

Example: Cigna Presentation on DM Outcomes  
How do you know this is wrong?

Disease Category	All-cause Admission Reduction per disease member	All-cause Claims Cost Reduction per disease member
Asthma	2%	12%
cardiology	5%	15%

# How you know this is wrong

- Probably the admission reduction is wrong too but assume that it's right
- Hospital cost is only about half of total cost for a condition
  - Therefore total costs can fall only about half as fast as hospital costs. Certainly they cannot fall 3-5x faster
- Savings are shown in asthma, which is impossible with that level of admission reduction
  - See spreadsheet

If you are being shown savings in asthma your entire outcomes report (not just asthma) is invalid

- Assume:
  - \$0.25 PMPM for asthma alone (\$0.60 per contract holder)
  - \$2000/day inpatient and \$400/ER visit
  - Standard event rates and admission rates from ER
  - 2-day ALOS
  - Small reduction in comorbidities

It turns out that – and look at the spreadsheet – that...

- Spending on asthma ER and IP events is only a little higher than the cost of the program itself
  - IP and ER events would have to decline by 60%+ just to break even, assuming no increase in drug spending
  - You can see this for yourself on the spreadsheet

# Agenda

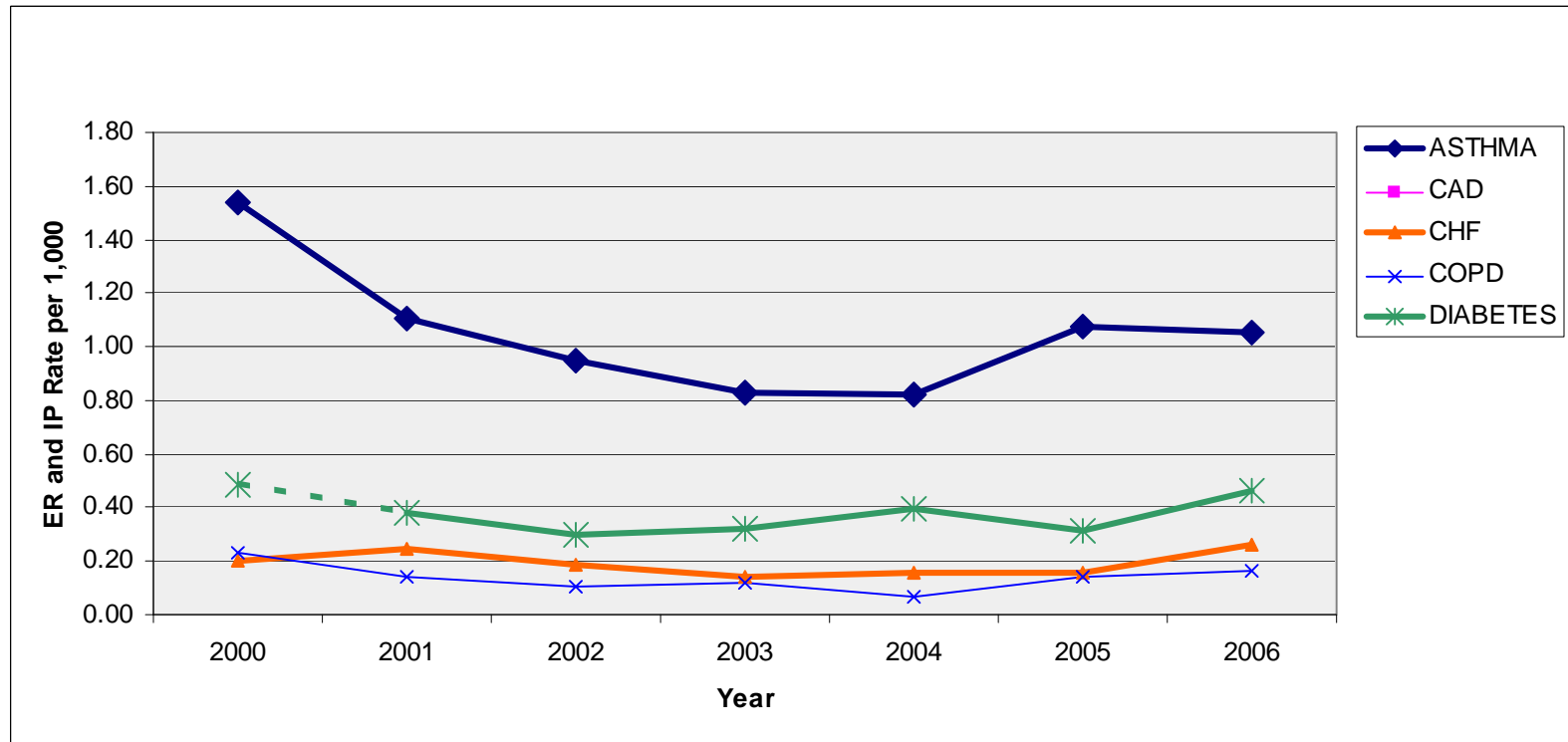
- Who am I? Why am I here?
- Why your program results may not be valid
  - Back by popular demand, planes on the ground analogy
  - Examples of mistakes driven by pre-post
- How to Measure validly using utilization rates
  - Real world examples and interesting findings

# Real-world examples

- These are event rates by health plans over time
  - Note that a thin line is pre-DM, dotted is implementation, and thick line is post-DM
- Following DM, events would have to decline noticeably for the savings shown in a typical pre-post 2:1 ROI to be accurate



# Historic trend in ER and IP rates, by condition, Northeast health plan



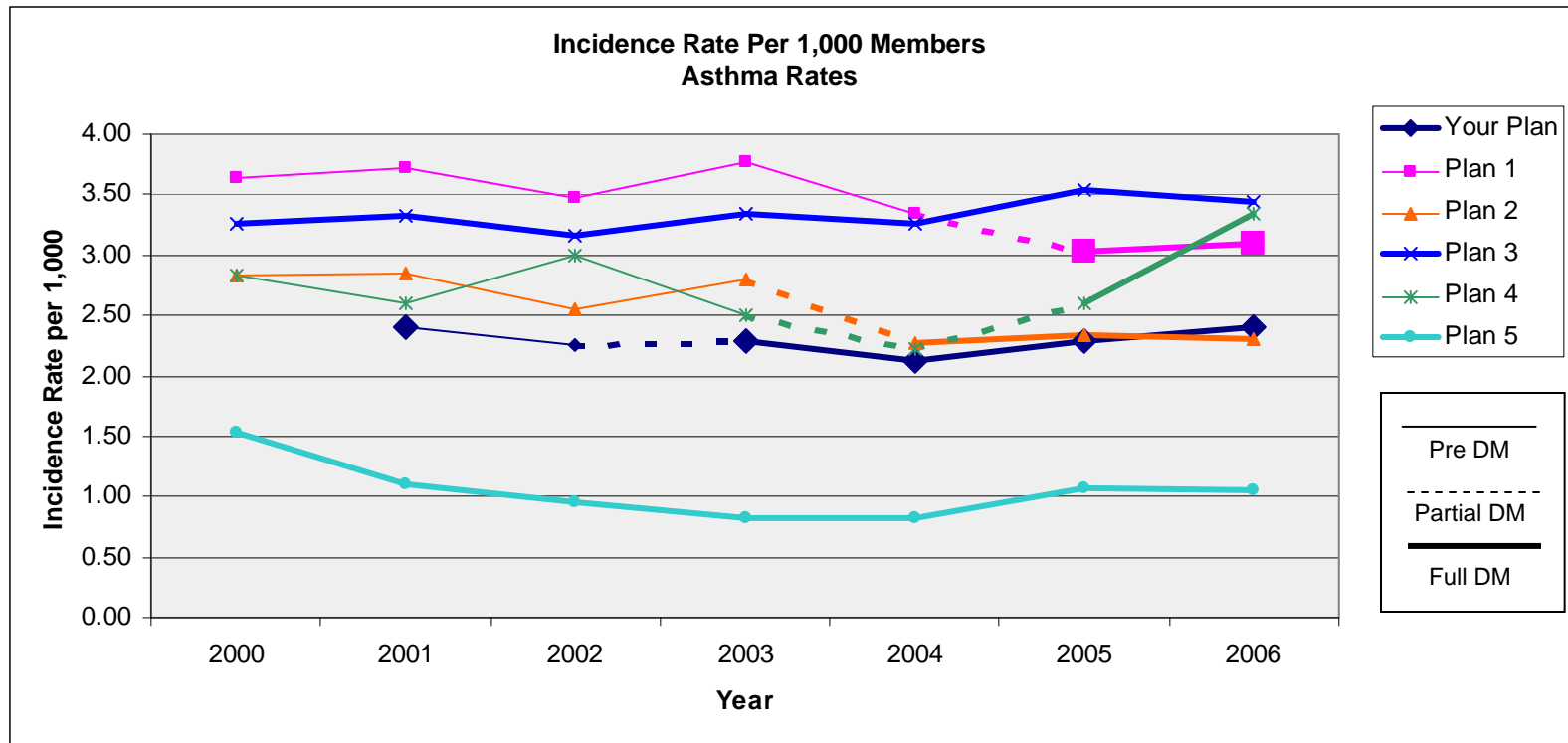
# Implications

- Huge progress in asthma, which has been a major focus
- Note that blips up in recent years due to adding a municipality (they were supposed to not count them in the data but they did)
- Solid improvements in other conditions, taking the above comment into account

# Some real-world examples

- These are event rates by health plans over time
  - Note that a thin line is pre-DM, dotted is implementation, and thick line is post-DM
- Following DM, events would have to decline noticeably for the savings shown in a typical pre-post 2:1 ROI to be accurate.
- It may be that the reason the improvements are modest is that the program is longstanding and has been successful for years even before DM, so that a plunge is unlikely

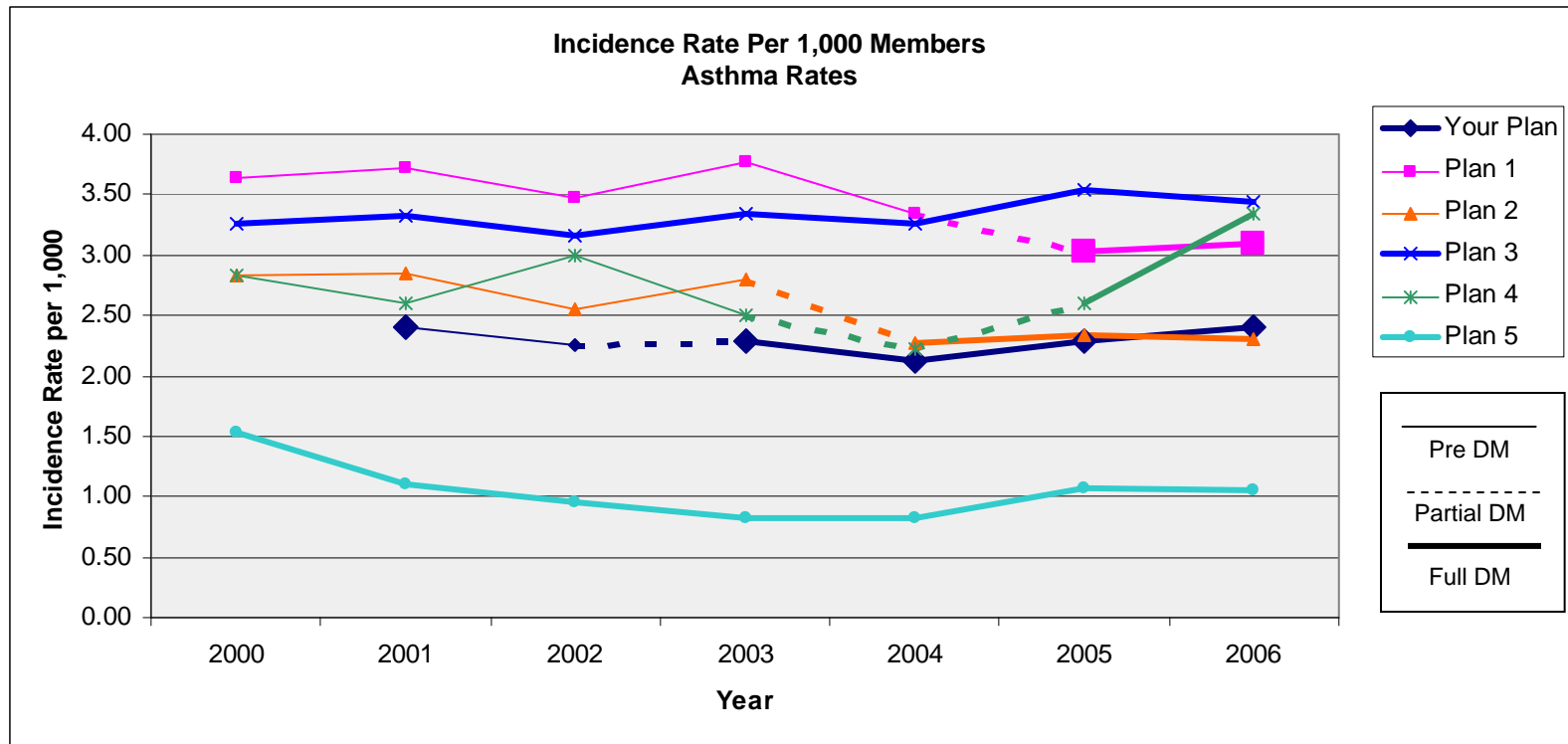
# Example: total asthma events across plans



# What did this slide show?

- Plan 5 has a very successful asthma program (assume that age profiles of these plans are similar)
- The other plans instituted asthma programs and failed to show actual savings despite giddy pre-post reports claiming ROIs.
- Even so, a “nexus” is needed between Plan 5 results and program

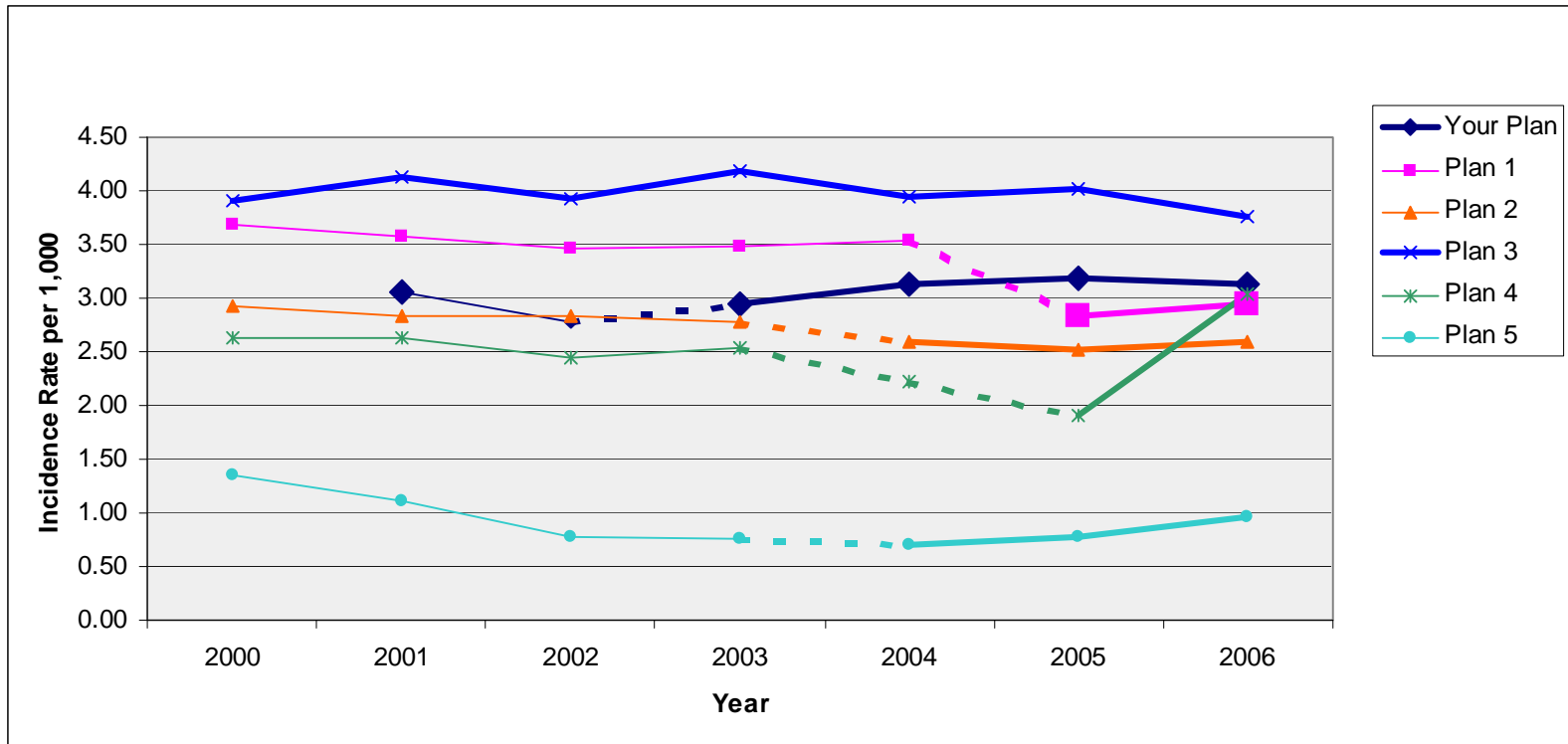
# Example: total asthma events across plans – Plan 5 is best: WHY?



# Plan 5: Do their practices support their results...or is it luck?

- Tight provider network
- Careful monitoring of use of evidence-based medicine
- Excellent disease management
  - Short “time to contact” between diagnosis and disease management
  - Disease mgmt at provider level too
- Urgent care center contracts prevent ER visits

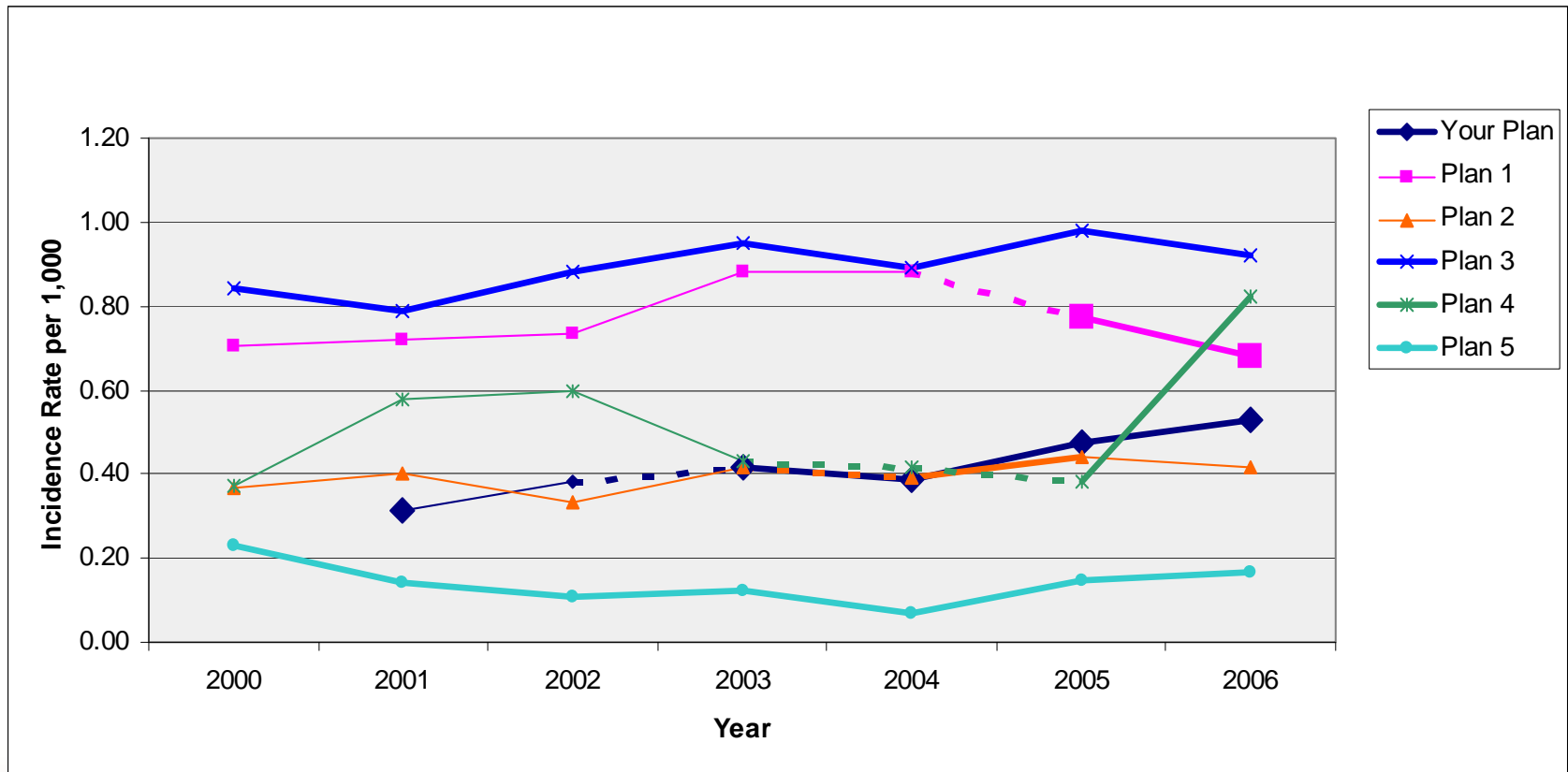
## ER/IP Usage --Incident Rate Per 1,000 Members CAD Rates





# Heart Failure Fluid Overload

ER/IP Incident Rate Per 1,000 Members  
CHF Rates



# How to find plans like Plan 5

- The regional Blues and provider-owned locals tend to be best (UHC may be best among nationals)
- Beware of plans with large networks – they can't monitor/incentivize best practices
- If they aren't monitoring their event rates, they can't control them
- The best DM

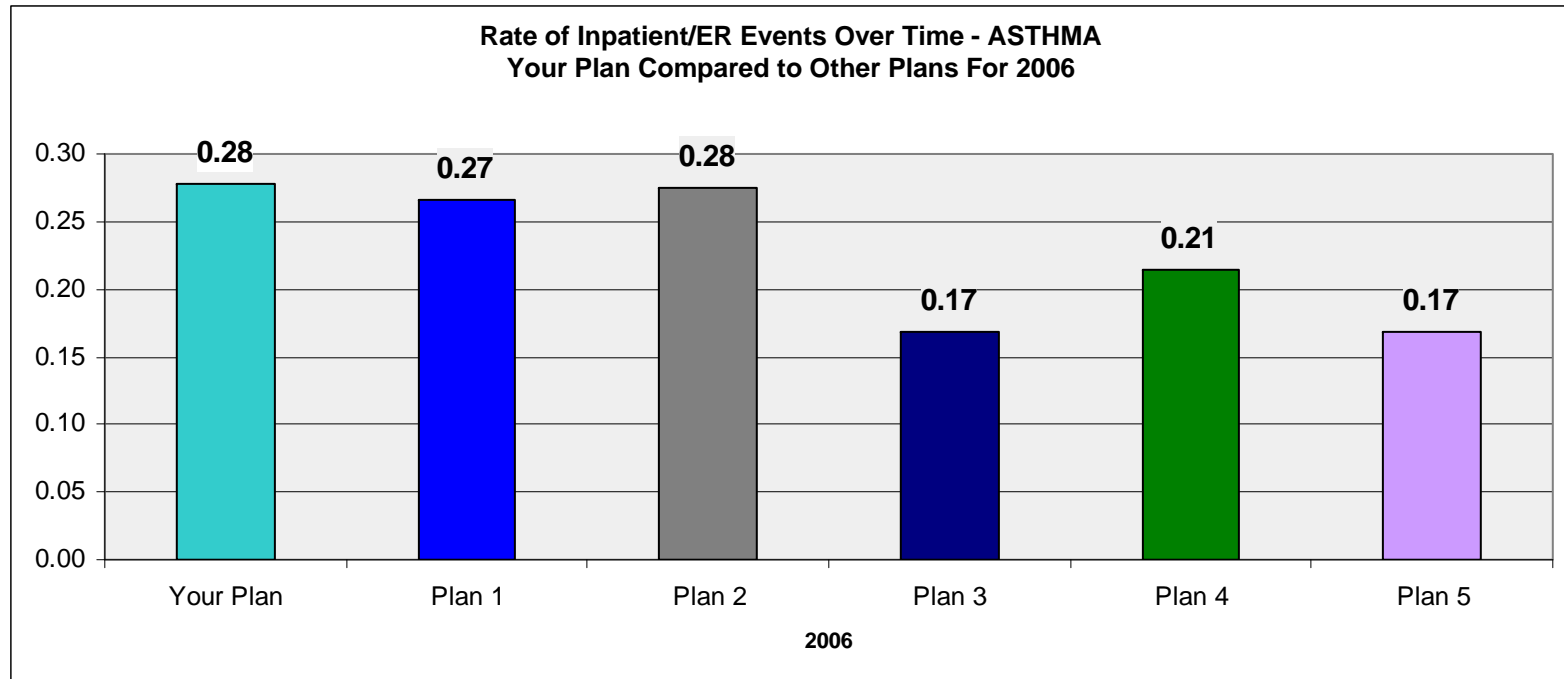
# Agenda

- Who am I? Why am I here?
- Why your program results may not be valid
  - Back by popular demand, planes on the ground analogy
  - Examples of mistakes driven by pre-post
- How to Measure validly using utilization rates
  - Real world examples and **interesting findings**

# Interesting findings from followup

- Your asthmatics get admitted to the hospital when they could be sent home
- Your HCC Scores/RAF may not correlate with how sick your seniors really are, in Days/1000
  - Leaving millions on the table
- Your angioplasties probably don't reduce your MIs and angina

# At what rate do your asthmatics get admitted when They show up in the ER?



# Insights from this data

- 40% variation
  - Probably correlates with other decisions to admit, way beyond just asthma
- Both health plans with lowest asthma ER/IP rate also had lowest admit rate from the ER
  - One is being studied as a “best practice” in many areas

# RAFs and Days/1000

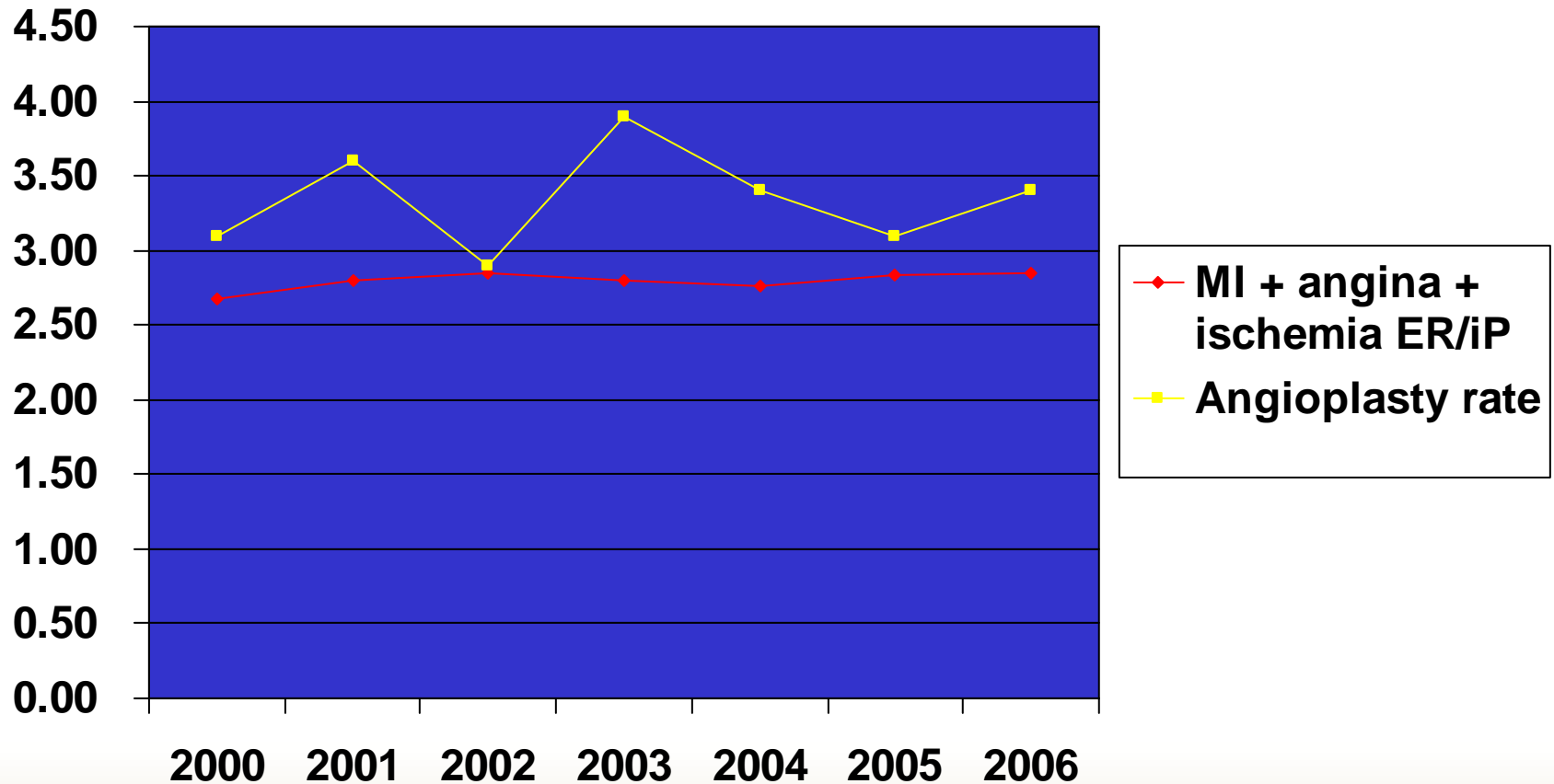
	RAF	Days/1000 (Acute Care)	<b>Ratio (low is good)</b>
Plan 1	0.91	1850	<b>2032</b>
Plan 2	1.05	1675	<b>1595</b>
Plan 3	1.12	1579	<b>1409</b>
Plan 4	0.93	1884	<b>2025</b>

# Insights from RAF Analysis

- Two plans are “best practice”
  - Do specific things with specific vendors
  - Per 10,000 Medicare lives, do about \$5-million better!



# Angioplasty Rate vs. MIs and Angina (one health plan – typical of others)



# Possible insights if analysis holds true

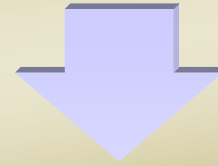
- In addition to “unwarranted variation” which Wennberg has found, there is also no incremental impact of angioplasties on MIs or angina as cardiologists do more of them

# Next Steps

- MEASURE event rates—this is what you are trying to impact
- DO NOT track events in the “managed population.” Some events occur outside that population. They all must be tracked to determine if a reduction is real or Regression to the Mean
- STOP wasting money on actuaries – this is not actuarial science. It’s biostatistics and common sense

# Impact of Measuring Correctly

**Size of ROI from DM: Lower**

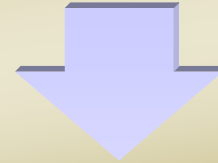


**Measurability of ROI from DM: Higher**



# Impact

**Size of ROI from DM: Lower**



**Measurability of ROI from DM: Higher**



**Credibility of ROI from DM: Priceless**



# 15-second shameless plug

- You can do everything here on your own but it's way easier with a DMPC Membership and the DMPC database for comparison
- “Join the Consortium or the terrorists win.”

George W. Bush