Critical Outcomes Report Analysis (CORA) Training: Spotting the errors

May 2009

DMPC

Disease Management

Purchasing Consortium Advisory Council

Agenda

- Background on Disease Management Purchasing Consortium
- First Review of Materials emailed ahead of meeting
  - Powerpoint example
  - Word Example
  - Mercer example
- Why these mistakes exist – “planes on the ground” and “heads or tails” and “trend assumptions”
- How to measure validly at low cost
- Examples of mistakes – second review of earlier materials
- Test
- Time-out to take test (2 hours)
- Review of answers

Hey, Butch, Who Are These Guys?

- DMPC is largest "buy-side" advisor of DM/wellness procurement in US. (STRS of Ohio would be an example of public-employee client)
- DMPC is the only organization offering certification of savings in both CORA and Savings Measurement Validity (see website www.dismgmt.com for details and list)
- DMPC ranked #1 by *Managed Healthcare Executive* in all surveys this century for DM
- I "wrote the book" on DM (now in 3rd printing)

When DMPC gets called in to health plan or employer

- Plan needs valid measurement (as opposed to measurement "showing savings" – that's the benefits consultants' or vendor's job)
- Credibility required to convince higher-ups that results are "real"
- Budget cuts limit spending on evaluation to <$10,000
- Plan want to "make a change" and need justification that current program isn't working
- Plan needs to procure a program but doesn't have $50,000 to spend on consultants AND needs the program to be cost-effective

Agenda

- Background on Disease Management Purchasing Consortium
- First Review of Materials emailed ahead of meeting
  - Powerpoint example
  - Word Example
  - Mercer example
- Why these mistakes exist – “planes on the ground” and “heads or tails” and “trend assumptions”
- How to measure validly at low cost
- Examples of mistakes – second review
- Test
- Time-out to take test (2 hours)
- Review of answers

#1 Example-- A real submission disguised

Category	Base	Intervention
Total Comm. Membership	505,000	511,000
Prevalence of selected case mgmt conditions	23%	23%
Annual claims cost	$972	$935
Annual admission rate	99	79
Annual Admission cost	$261	$244
Annual MD visit rate per 1000 members	4535	4475
Annual MD visit costs/member	$132	$128
Annual ED visit rate/1000	452	339
Annual ED costs/member	$39	$33

#2 and #3 examples

- Word report
- Mercer report

Agenda

- Background on Disease Management Purchasing Consortium
- First Review of Materials emailed ahead of meeting
  - Powerpoint example
  - Word Example
  - Mercer example
- Why these mistakes exist – “planes on the ground” and “heads or tails”
- How to measure validly at low cost
- Examples of mistakes – second review
- Test
- Time-out to take test (2 hours)
- Review of answers

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)

What is the average altitude in this example?

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

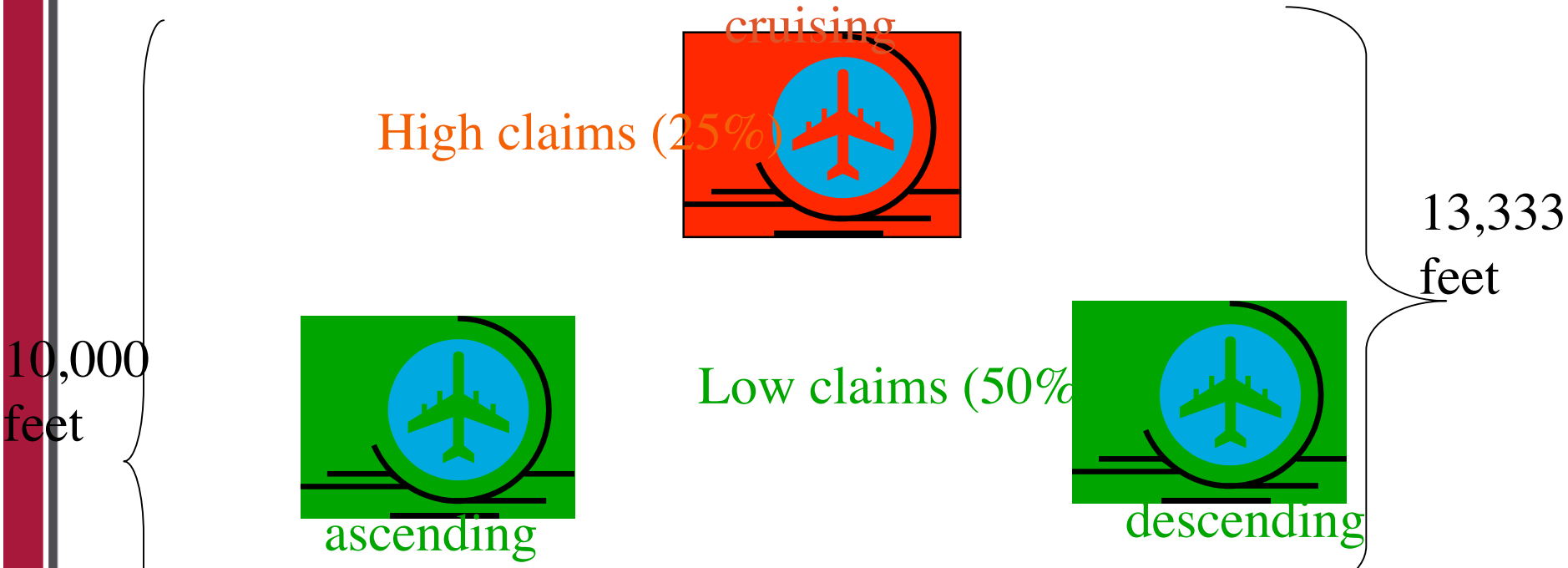
Uncovering the hidden flaw in the current methodology

- 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

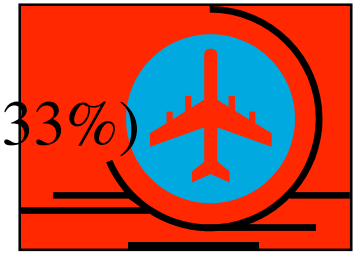


The current approach being used by vendors and benefits consultants

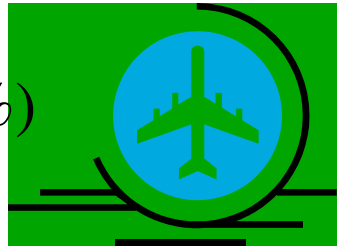
- Tracks ALL people *with claims for the disease, high or low*, in the baseline
- It's what *they* call a population-based approach
  - Equivalent to finding all *flights* including ascending and descending but not all *planes*
  - Average baseline altitude (2/3 at 10,000, 1/3 at 20,000) is: 13,333 feet

They measure the claims on ALL patients with claims

High claims (33%)



Low claims (67%)

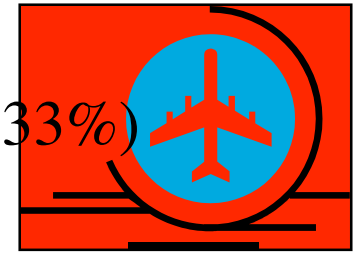


Above the line are datapoints which are found and measured

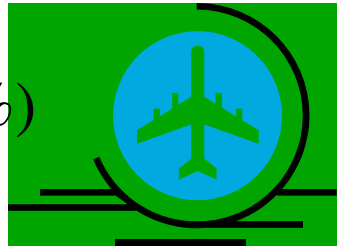
No claim

They measure the claims on ALL patients with claims

High claims (33%)

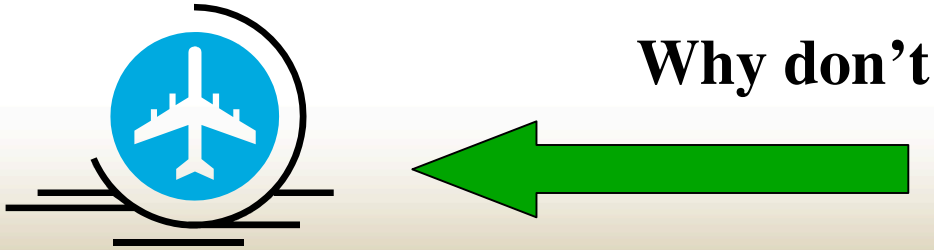


Low claims (67%)



Above the line are datapoints which are found and measured

Why don't they measure these guys?

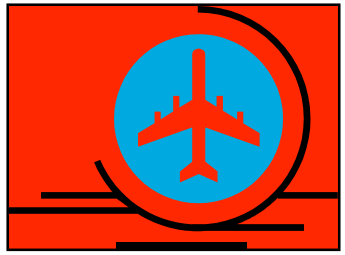


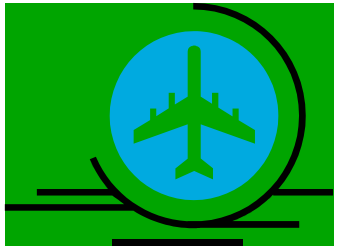
No claim

They measure the claims on ALL members with claims

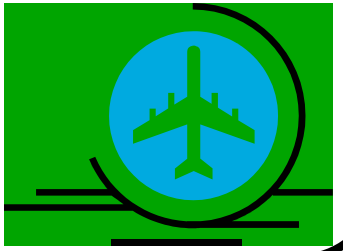
13,333
Feet
On average

High claims





Low claims



These get
Found in
The clain
pull

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

No claim

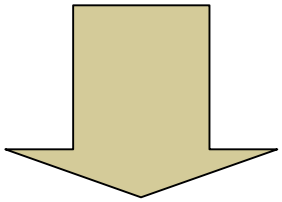
DMPC | Disease Management
Purchasing Consortium Advisory Council

Why might members not have "claims to be found" ?

- Ignoring their disease and not filling Rx
- Not long enough in the plan to have claims (or have claims adjudicated)
- Don't know they have the disease
- Mild enough to be treated with lifestyle only
- Not enough claims (example: One 250.xx for diabetes)

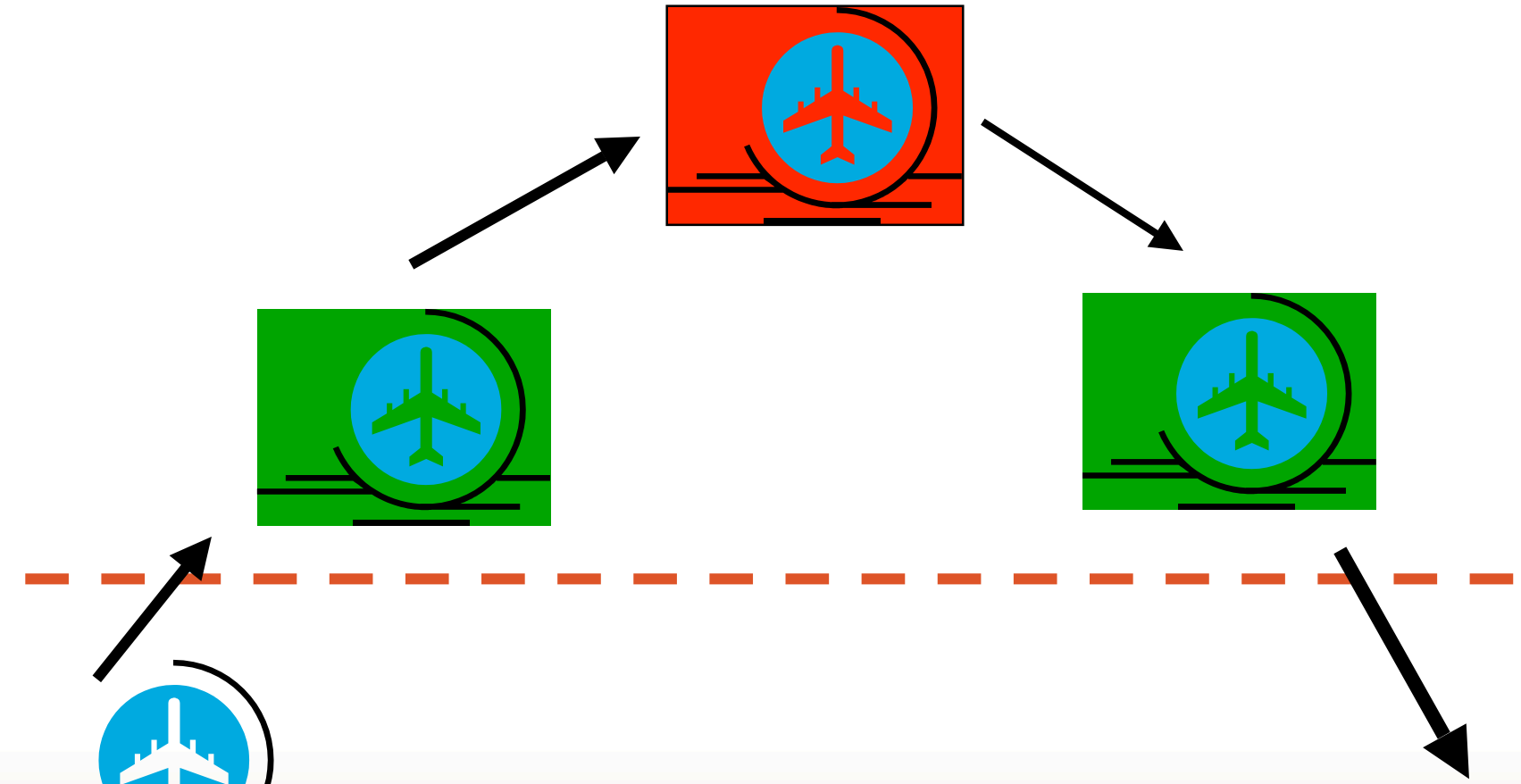
The conventional approach

- Tracks all members *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 members with baseline claims during the study period)

One hour later…(next claims cycle)

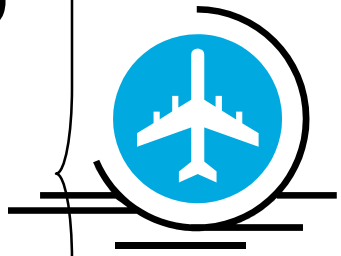


We can all agree that…

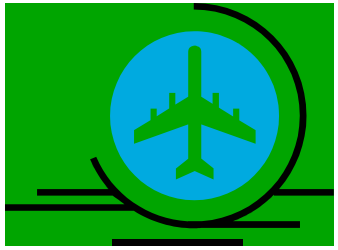
- The aviation system is in a steady state
- Still 25% at each point
- Average altitude has not changed

One hour later…(next claims cycle)

Average Plane is Still 10,000 feet

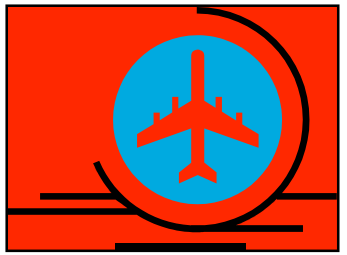


25%



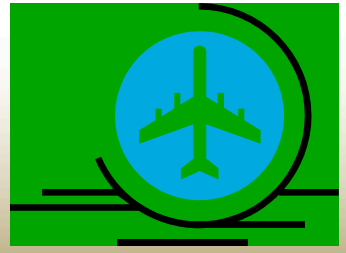
High Claims 25%

Low claims 25%



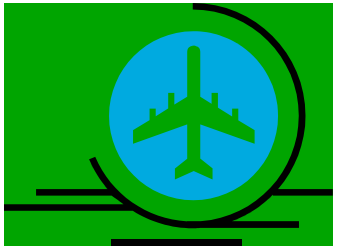
Average Flight is Still 13,333 feet

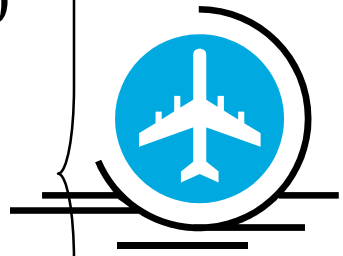
No claim 25%

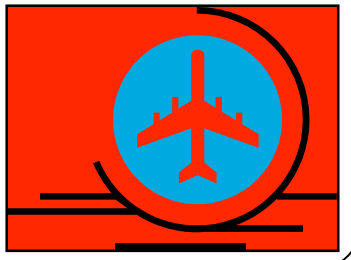


One hour later…(next claims cycle)

Average Plane is Still 10,000 feet

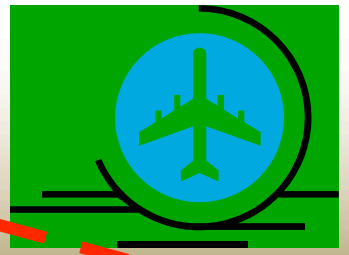






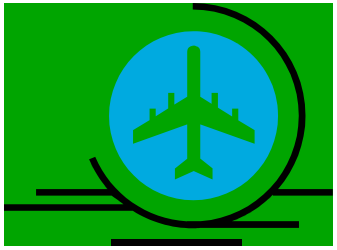
Average Flight is Still 13,333 feet

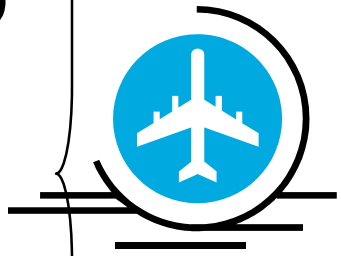
Except that now <u>all the flights</u> are being Tracked including the ones which have Landed!

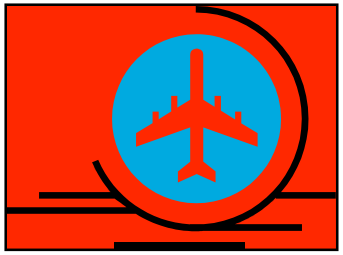


One hour later…(next claims cycle)

Average Plane is Still 10,000 feet



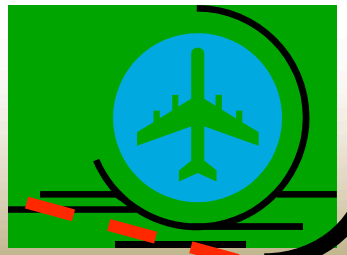




Average Flight is Still 13,333 feet

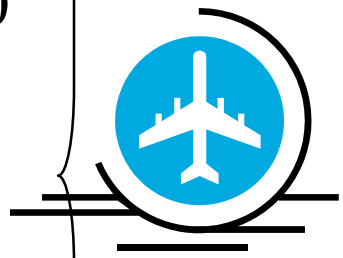
Meas Ment 10,00 feet

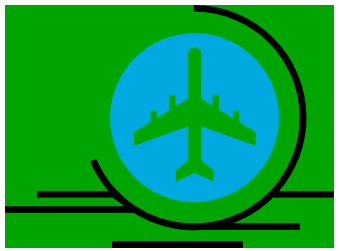
Except that now <u>all the flights</u> are being Tracked including the ones which have Landed!

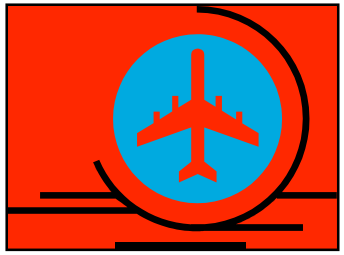


“But this shouldn’t happen if a member has to requalify every year”

Average Plane is Still 10,000 feet

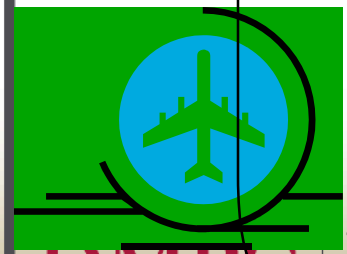




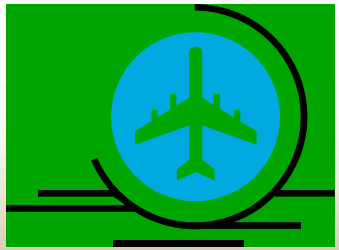


Average Flight is Still 13,333 feet

Not in measurement here



Not here either



Wrong

- What is the fallacy with that “adjustment” ?

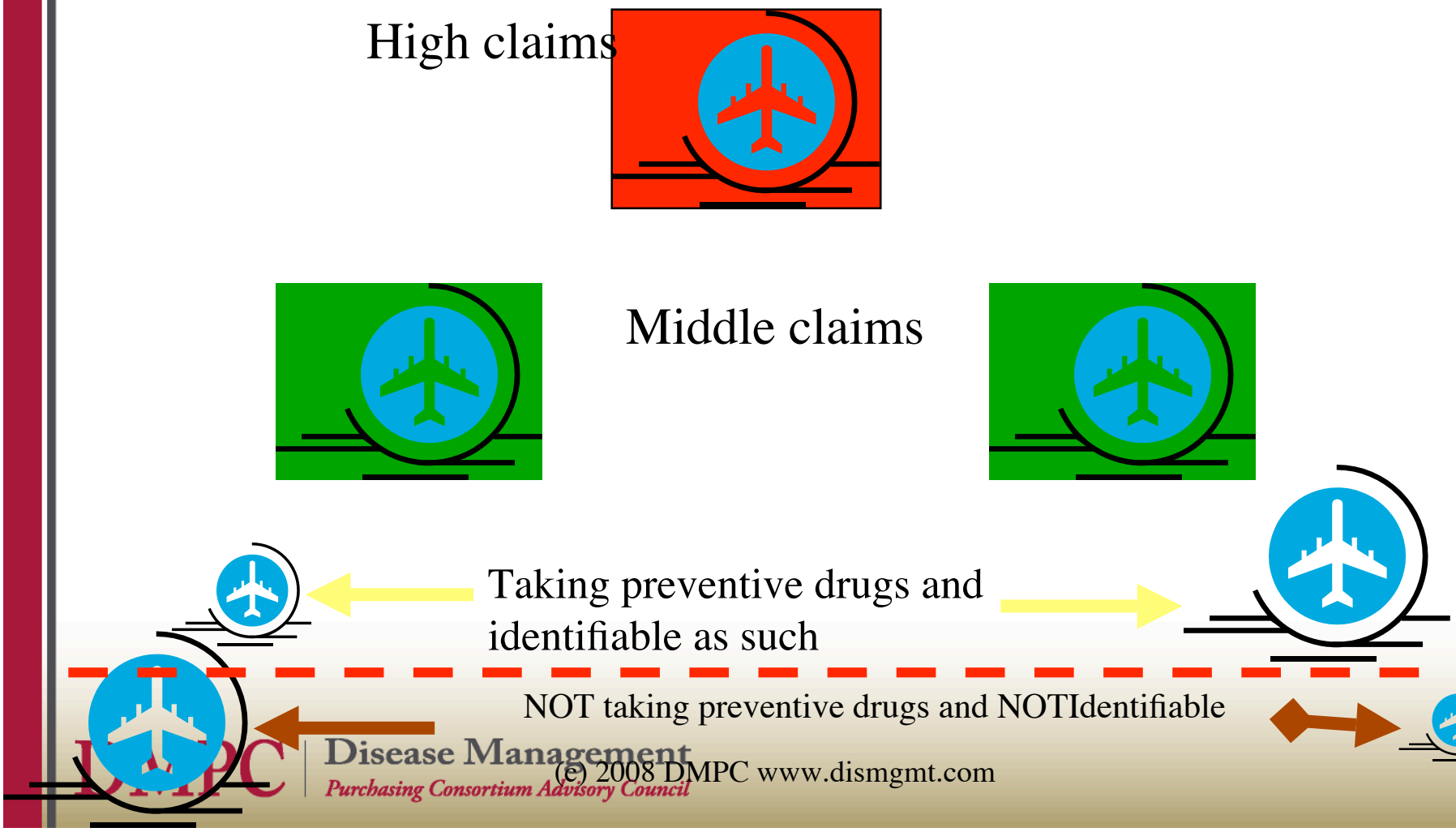
Explanation of why the bias is still there even if zeroes aren't measured

- Because AFTER someone with no claims has an event and then recovers, that person is put on drugs (asthma, beta blockade, antihyperlidemics etc.)
  - And for some period of time they comply and generate claims

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



Planes on the ground in healthcare

- Most contracts have a baseline period to which a contract period is compared (adjusted for trend)
  - Virtual Hand-raising time

In this example

- Assume that “trend” is already taken into account
  - Note that trend is usually also wrong because Mercer writes that one should “choose” a trend and that the “choice” of trend has a large impact on the savings
  - In reality the trend simply obscures the savings and has NO impact on the savings, which are either there or not
  - This was the focus of the 2/1 *Disease Management Journal*
- Look at the baseline and contract period comparison

Base Case: Example from Asthma
First asthmatic has a $1000 IP claim in 2007

	2007 (baseline)	2008 (contract)
Asthmatic #1	1000	
Asthmatic #2		
Cost/asthmatic		

Example from Asthma

Second asthmatic has an IP claim in 2008 while first asthmatic goes on drugs (common post-event)

	2007 (baseline)	2008 (contract)
Asthmatic #1	1000	100
Asthmatic #2	**0**	**1000**
Cost/asthmatic		

What is the Cost/asthmatic In the baseline?

Cost/asthmatic in baseline?

	2007 (baseline)	2008 (contract)
Asthmatic #1	1000	100
Asthmatic #2	0	1000
Cost/asthmatic	**$1000**	

Vendors don't count #2 in 2007 bec. he can't be found

Cost/asthmatic in contract period?

	2007 (baseline)	2008 (contract)
Asthmatic #1	1000	**100**
Asthmatic #2	0	**1000**
Cost/asthmatic	**$1000**	**$550**

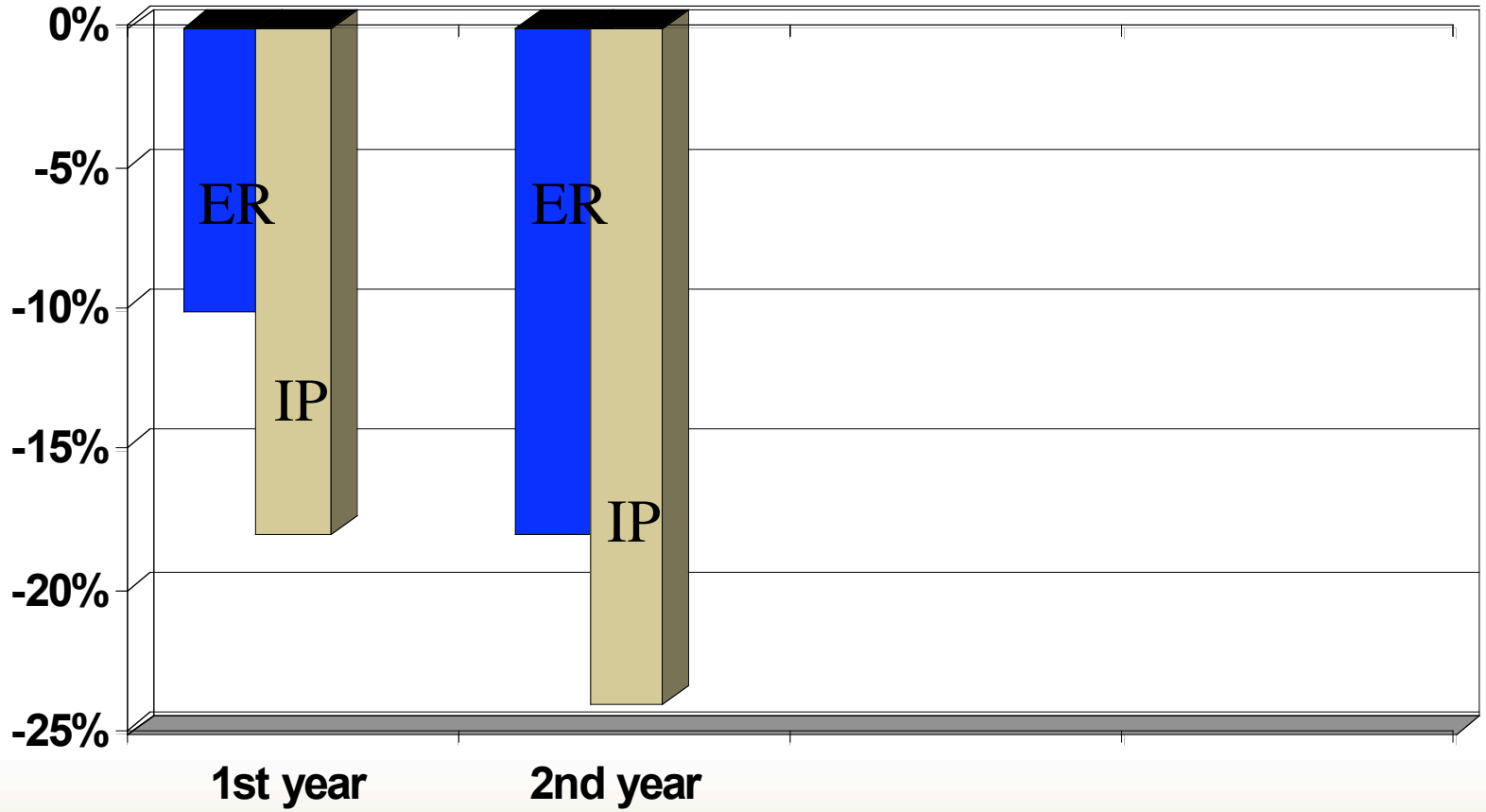
How to spot this issue:
Note the number of asthmatics

	2007 (baseline)	2008 (contract)
Asthmatic #1	1000	**100**
Asthmatic #2	0	**1000**
Number of asthmatics	**1**	**2**

Agenda

- Background on Disease Management Purchasing Consortium
- First Review of Materials emailed ahead of meeting
  - Powerpoint example
  - Word Example
  - Mercer example
- Why these mistakes exist – “planes on the ground” and “heads or tails”
  - “heads or tails” in theory and then in practice
- How to measure validly at low cost
- Examples of mistakes – second review
- Test
- Time-out to take test (2 hours)
- Review of answers

Example of just looking at people with previous claims for asthma ("heads"): Vendor Claims for Asthma Cost/patient Reductions

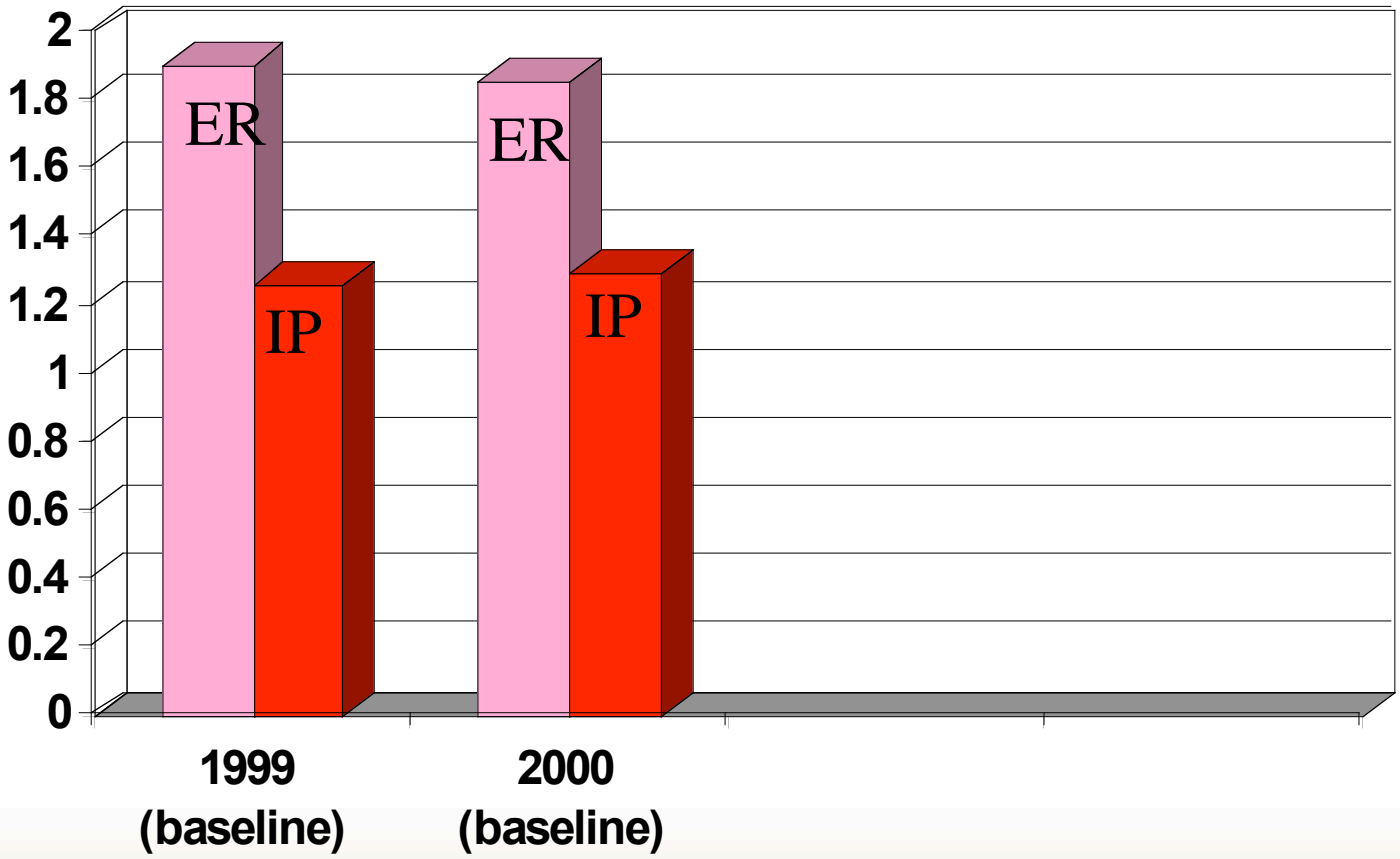


What we did to check

- We looked at the actual codes across the plan
- This includes even people who were “tails” who did not show up in the baseline as a “heads” (with a prior claim)
- Two years of codes pre-program to establish trend
- Then two program years

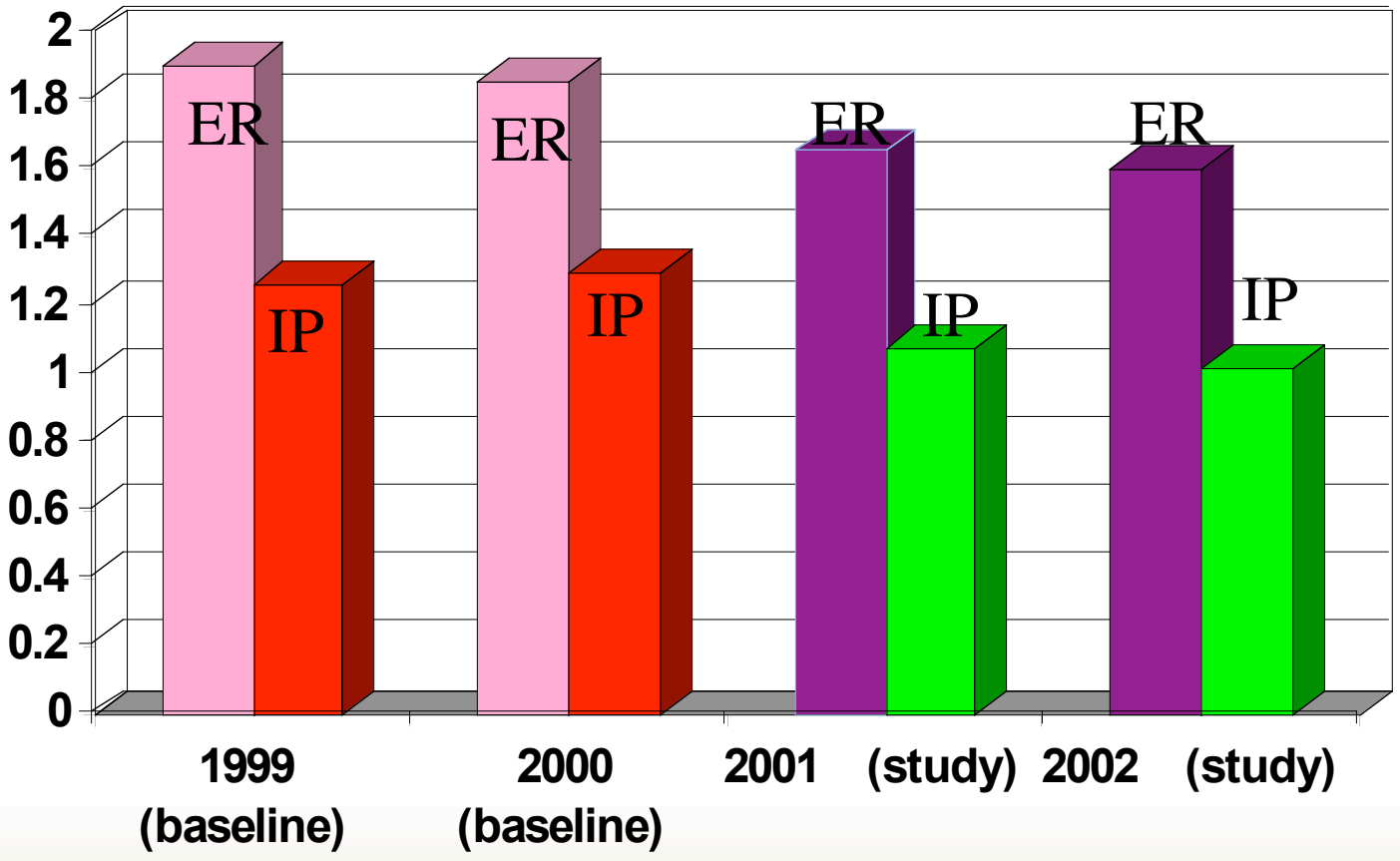
Baseline trend for asthma ER and IP Utilization

493.xx ER visits and IP stays/1000 planwide



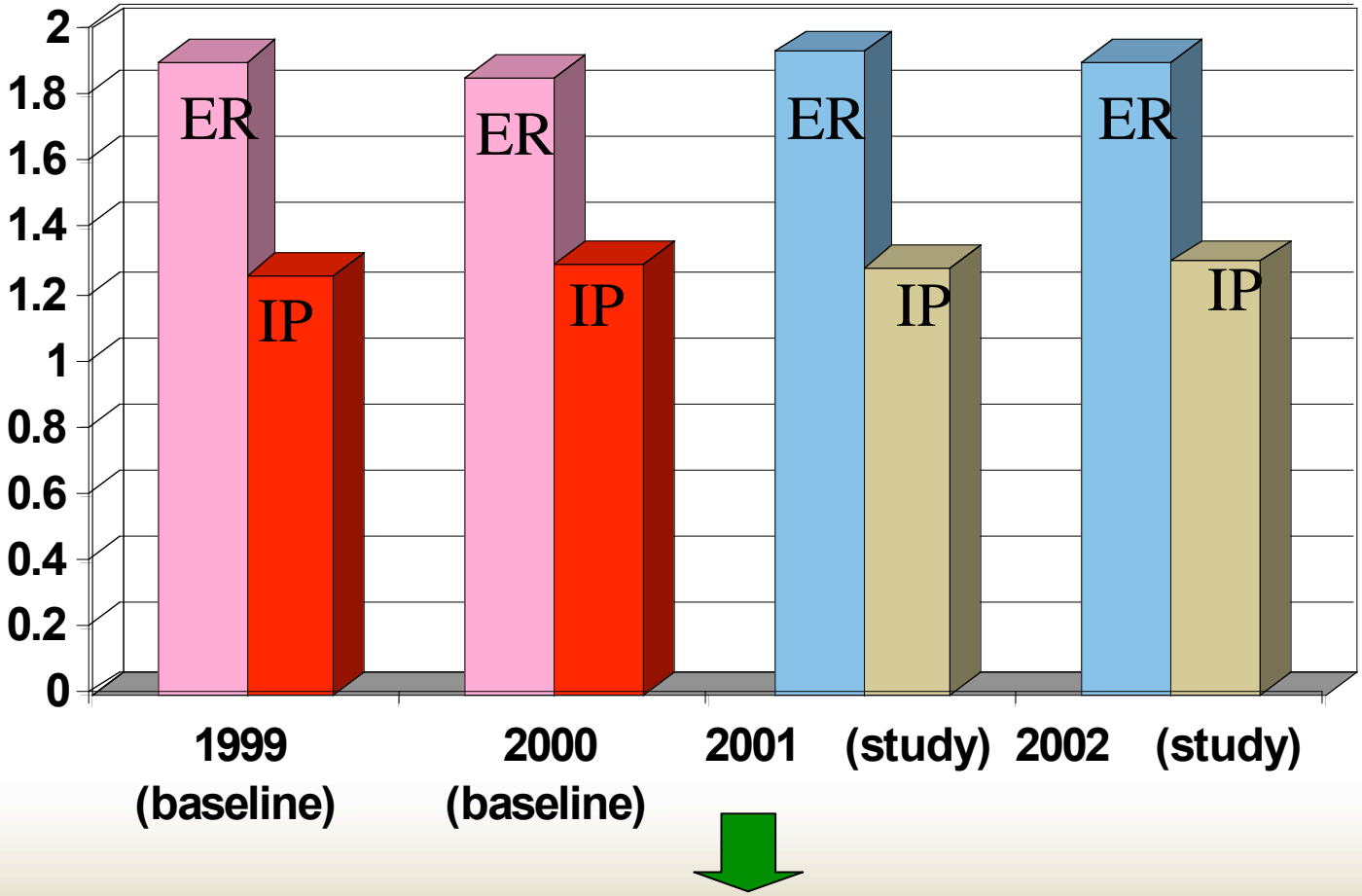
Expectation is something like…

493.xx ER visits and IP stays/1000 planwide



Plausibility indicator Actual:
Validation for Asthma savings from same plan including ALL CLAIMS for asthma, not just claims from people already known about – this includes both heads and tails from the baseline

493.xx ER visits and IP stays/1000 planwide

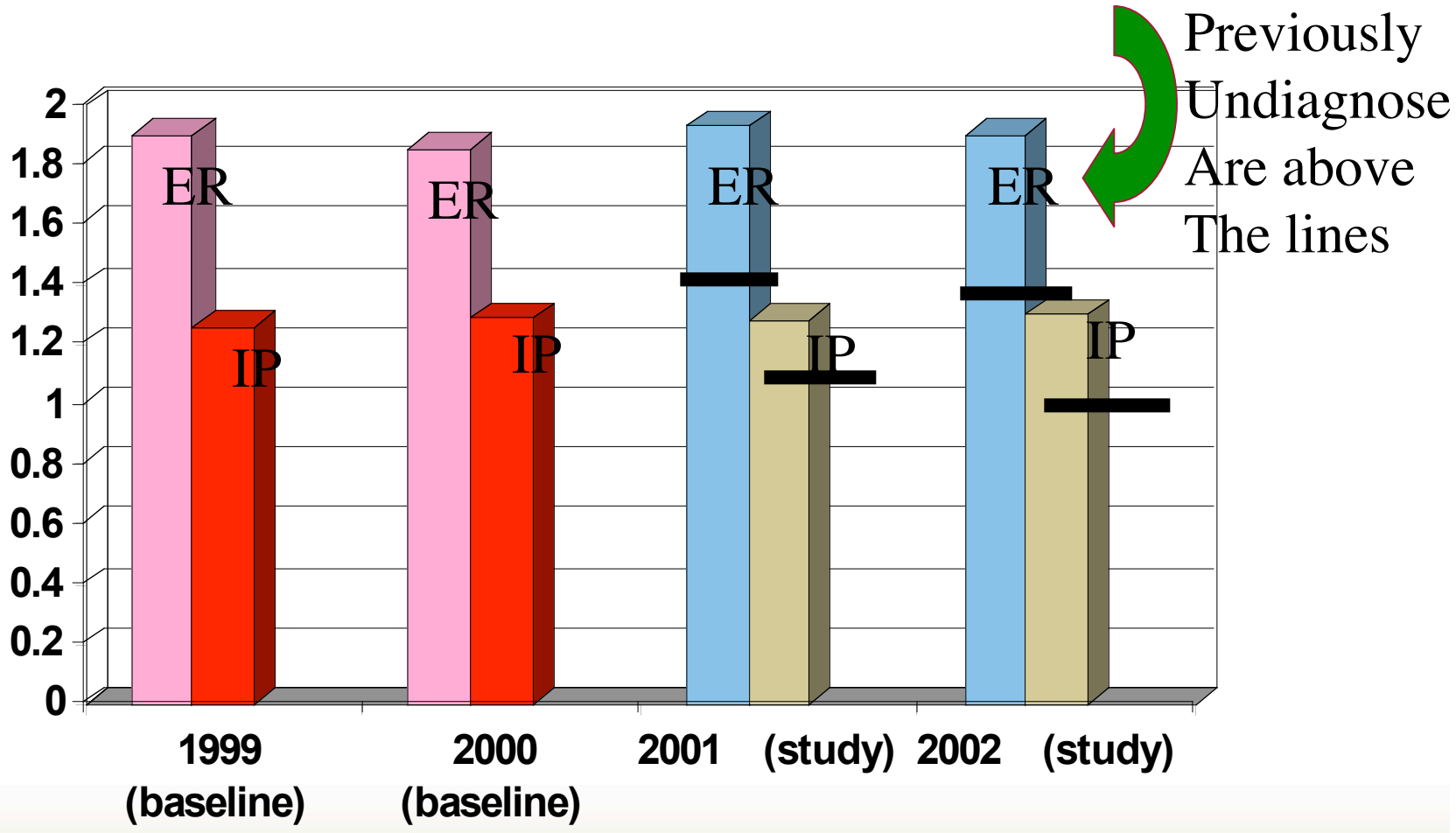


How could the vendor's methodology have been so far off

We then went back and looked…

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

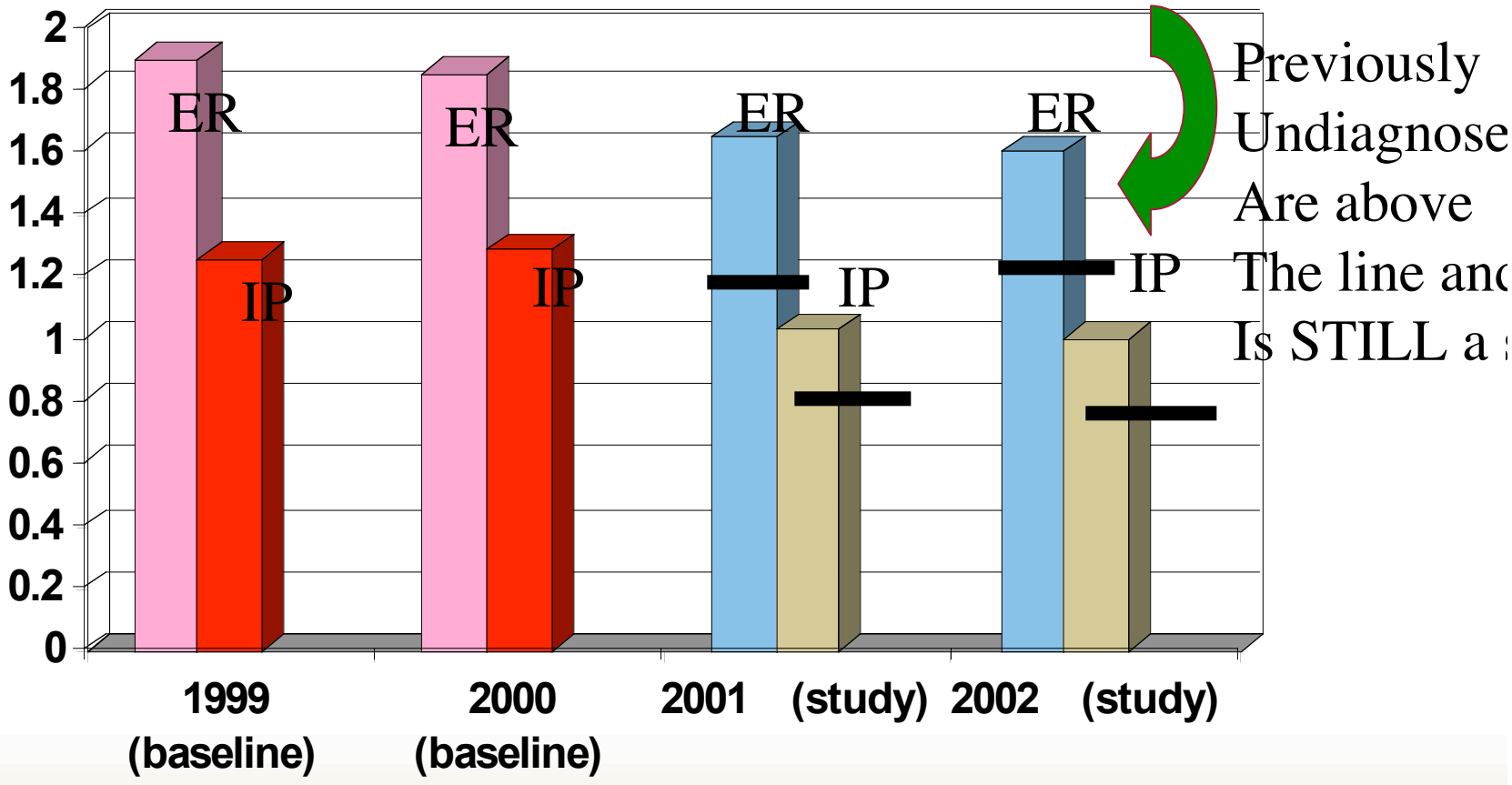
We were shocked, *shocked* to learn that the uncounted claims on previously undiagnosed people ("tails" in baseline who are "heads" now) accounted for virtually all the "savings"



Is it fair…

- To count the people the vendor didn't know about who were "tails" in the baseline?
- Of course – people with claims in the baseline ("heads") can randomly migrate to no claims ("tails") so the reverse should be counted too

You should be able to reduce visits in the *known* group ("heads") by enough so that adding back the new group (previously "tails" and now "heads") yields the reduction you claimed – otherwise you didn't do anything other than flip coins



Agenda

- Background on Disease Management Purchasing Consortium
- First Review of Materials emailed ahead of meeting
  - Powerpoint example
  - Word Example
  - Mercer example
- Why these mistakes exist – “planes on the ground” and “heads or tails” and “trend assumptions”
- How to measure validly at low cost
- Examples of mistakes – second review
- Test
- Time-out to take test (2 hours)
- Review of answers

Measuring Validly at Low Cost

- Look at ALL “heads” (events) in ALL years
- Total Event Rates
  - Not to be confused with what the consultants or vendors who say “we measure the whole population”
    - They don’t – they measure all the previous year’s heads
      - Meaning all the “flights in the air”

Event rates tracked by disease: ALL Primary-coded ER and IP events

Disease Program Category	ICD9s (all .xx unless otherwise indicated)
Asthma	493.xx (including 493.2x[1])
Chronic Obstructive Pulmonary Disease	491.1, 491.2, 491.8, 491.9,. 492, 494, 496, 506.4
Coronary Artery Disease (and related heart-health issues)	410, 411, 413, 414
Diabetes	250
Heart Failure	428, 404.01, 404.03, 404.11, 404.13, 404.91, 404.93, 425.0, 425.4

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

What is event rate measurement?

- Measure total event rates for diseases being managed, like you'd measure a birth rate. Couldn't be easier
  - Ask me for the specific directions. They're free from DMPC (and can be purchased from DMAA). See previous page
- Example from previous asthma hypothetical

Recall this Cost/asthmatic in contract period?

	2007 (baseline)	2008 (contract)
Asthmatic #1	1000	**100**
Asthmatic #2	0	**1000**
Cost/asthmatic	**$1000**	**$550**

Asthma events in the payor as a whole – the "event rate" or "plausibility" test

	2007 (baseline)	2008 (contract)
Asthmatic #1	1000	**100**
Asthmatic #2	0	**1000**
Inpatient 493.xx events/year	**1**	**1**

Plausible?

- How can you reduce asthma costs 45% without reducing planwide asthma event rate?
- Answer: You can't. Not plausible. Event-rate Plausibility test flunked
- Note that event rate measurement provides exactly the right answer (if drug classes are measured too) while pre-post doesn't come close
  - There is no counter-example

Event-rate Analysis example explanation: Heart Disease

- You have spent millions managing heart disease for several years, right?
- In order to reduce heart attacks (and related events), right?
- But…

Plausibility Analysis example explanation: Heart Disease

- You have spent millions managing heart disease for several years, right?
- In order to reduce heart attacks (and related events), right?
- But…
  - Do you even know your heart attack rate?

Plausibility Analysis example explanation: Heart Disease

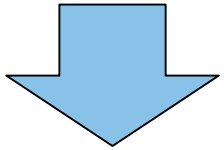
- You have spent millions managing heart disease for several years, right?
- In order to reduce heart attacks (and related events), right?
- But…
  - Do you even know your heart attack rate?
  - If you don't (and you don't), how do you know whether the rate has declined since you started the program?

Plausibility Analysis example explanation: Heart Disease

- You have spent millions managing heart disease for several years, right?
- In order to reduce heart attacks (and related events), right?
- But…
  - Do you even know your heart attack rate?
  - If you don't (and you don't), how do you know whether it has declined since you started the program?
  - How do you know how it compares to others?
- How can you do a program without knowing these three pieces of data?

This is what you learn with an event-rate plausibility test

- WHAT are my rates of adverse events (like heart attacks)
- HAVE they declined since I started a program
  - WHAT would they have likely been without a program
- HOW do they compare to others?

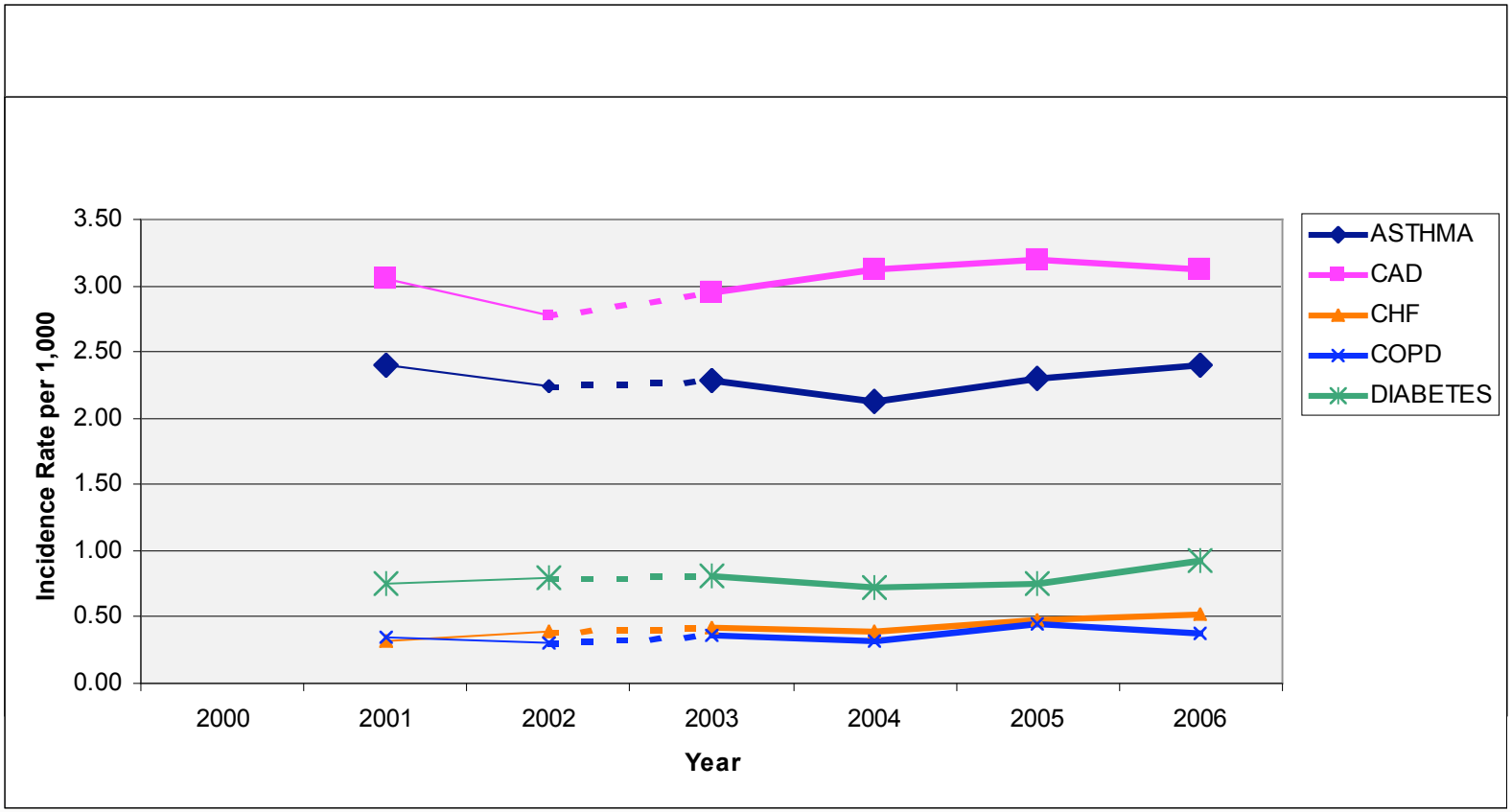


Lemme show you…

Key to Reading event rate slides

- Thin lines are pre-program
- Dotted lines are periods in which program was partially in place
- Thick lines are program fully implemented

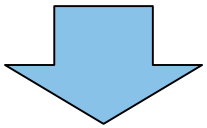
A HEALTH PLAN
Historical trend in event avoidance in DM-able conditions
Before and after DM program implementation
Rate of ER and IP events/1000 members ("event incidence")



DMPC | Disease Management Purchasing Consortium Advisory Council

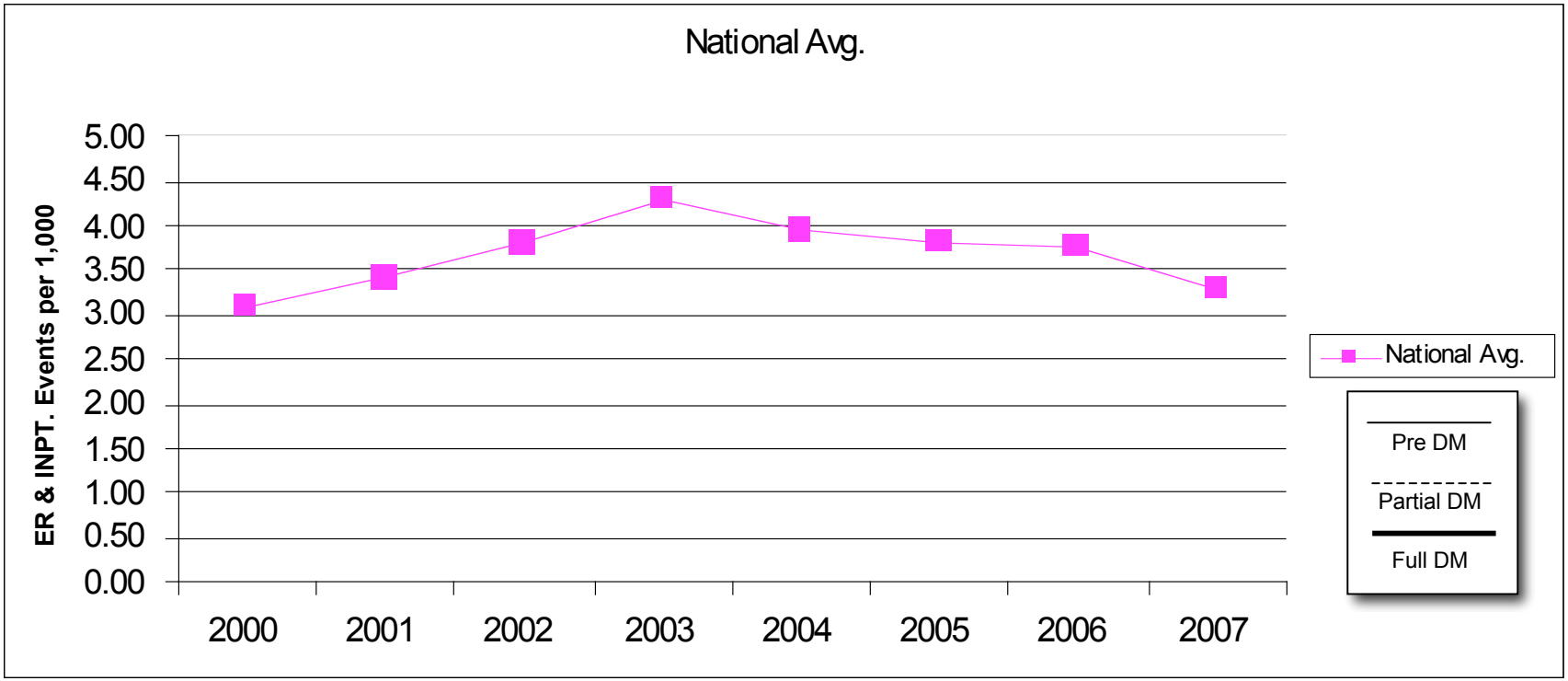
Three explanations

- Events would have gone up without the DM program
- Events were already low
- Program didn't save money (despite massive ROIs found by actuaries)



Let's compare to national or regional average of All payors in DMPC database

Example of National Average Event Rates
Heart Attacks, Angina Attacks, other Ischemic Events
(CAD)



Implications (CAD example)

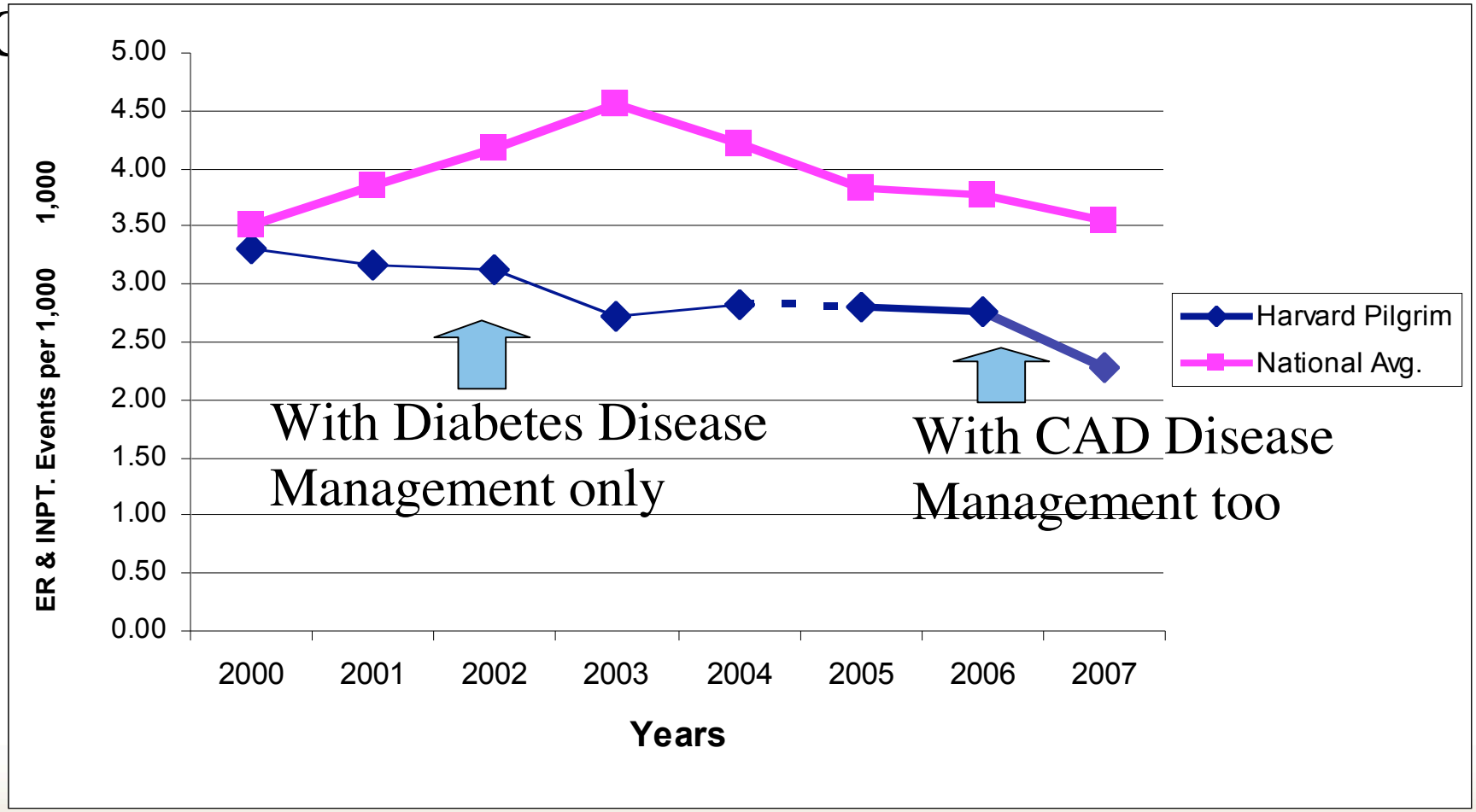
- Improvements in usual care, adherence to protocols and disease management have turned national trend around
  - It appears to diverges from trend towards more obesity, diabetes prevalence

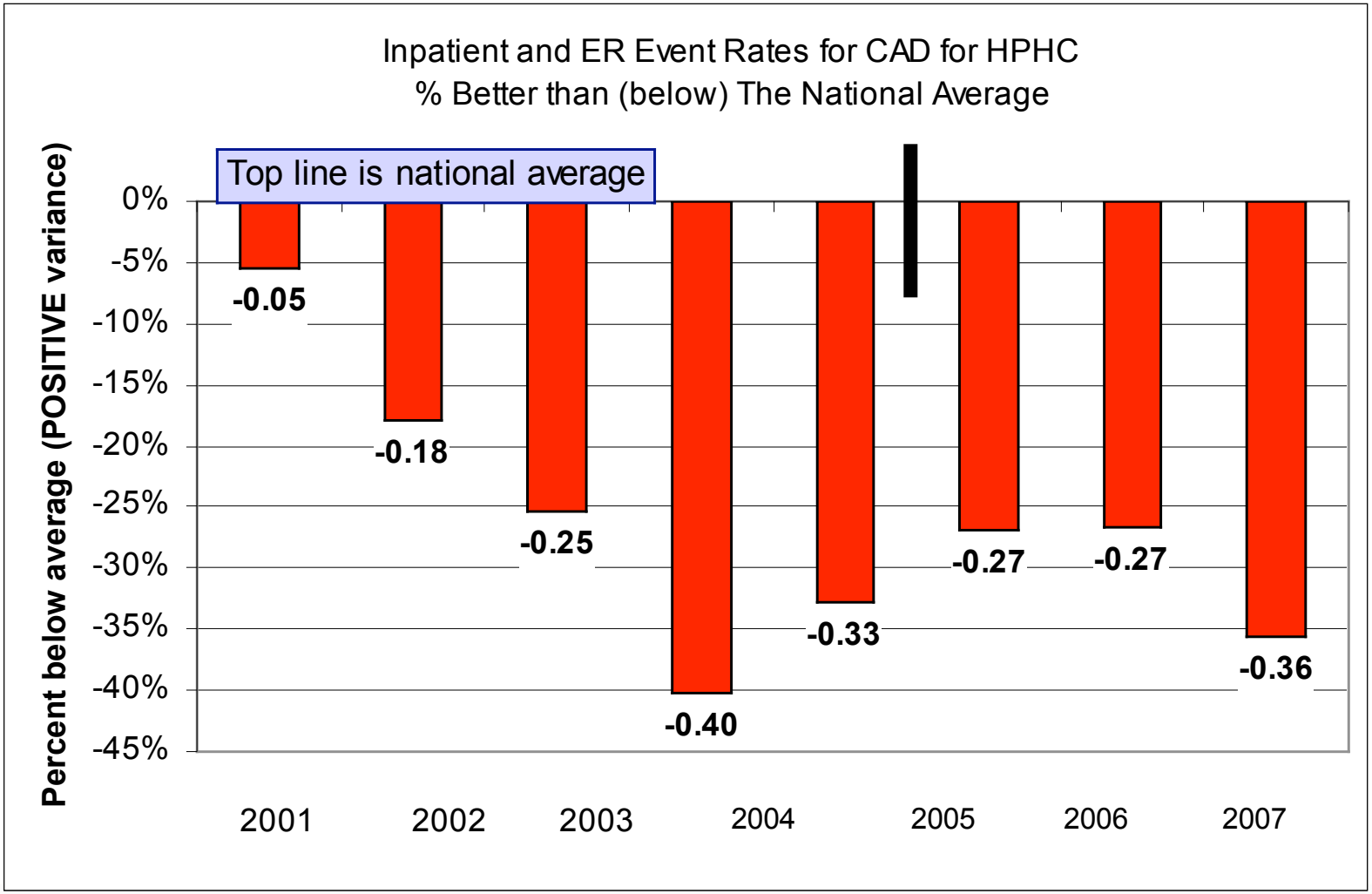
Examples of a payor (Harvard Pilgrim) vs. National Averages

- HPHC likes to be the example
- Ranks in tie for best in country (Cardiometabolic) or most improved in country (respiratory)

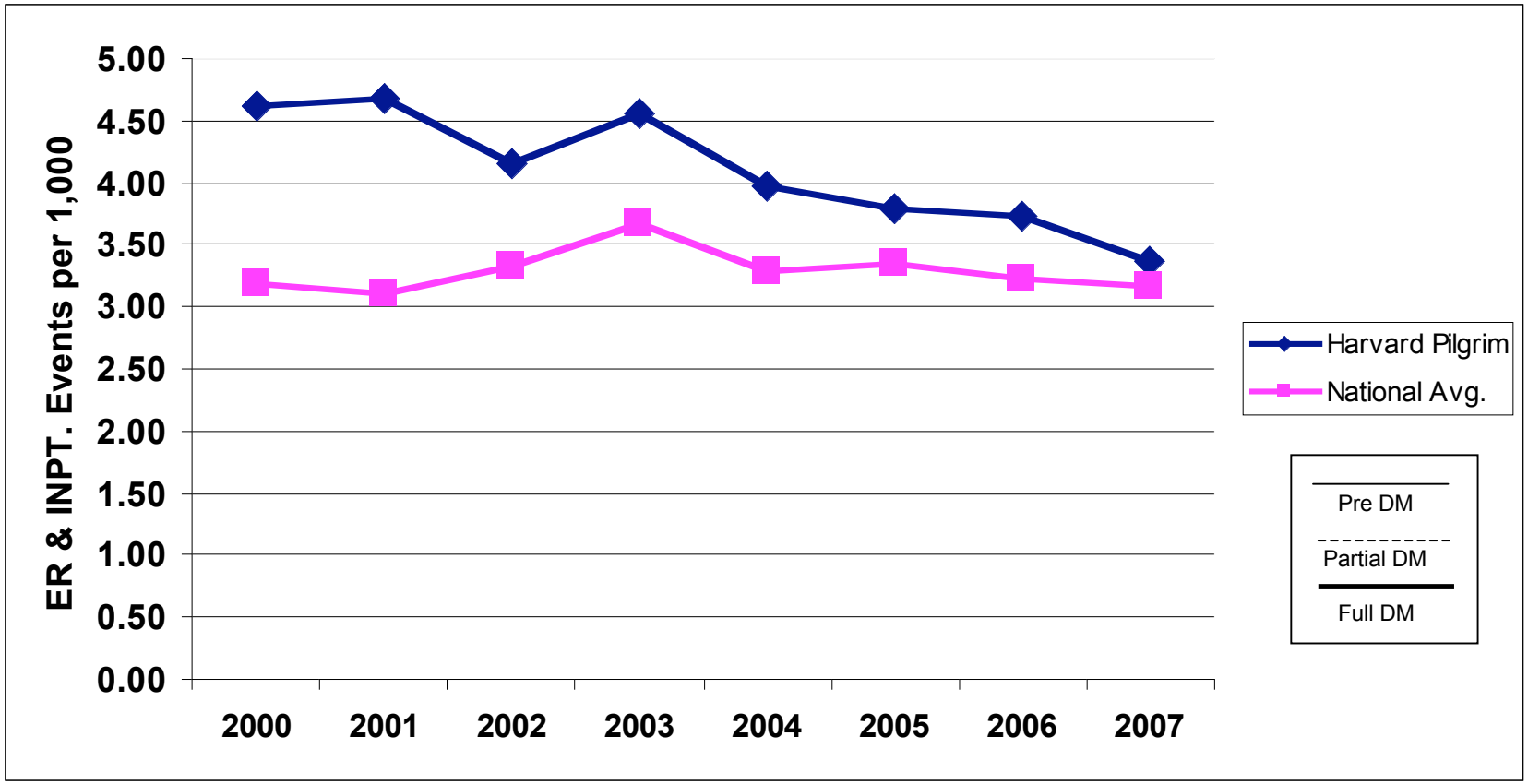
ER and Inpatient Event Rates (Commercial) Harvard Pilgrim vs. National Average of 29 health plans

- C

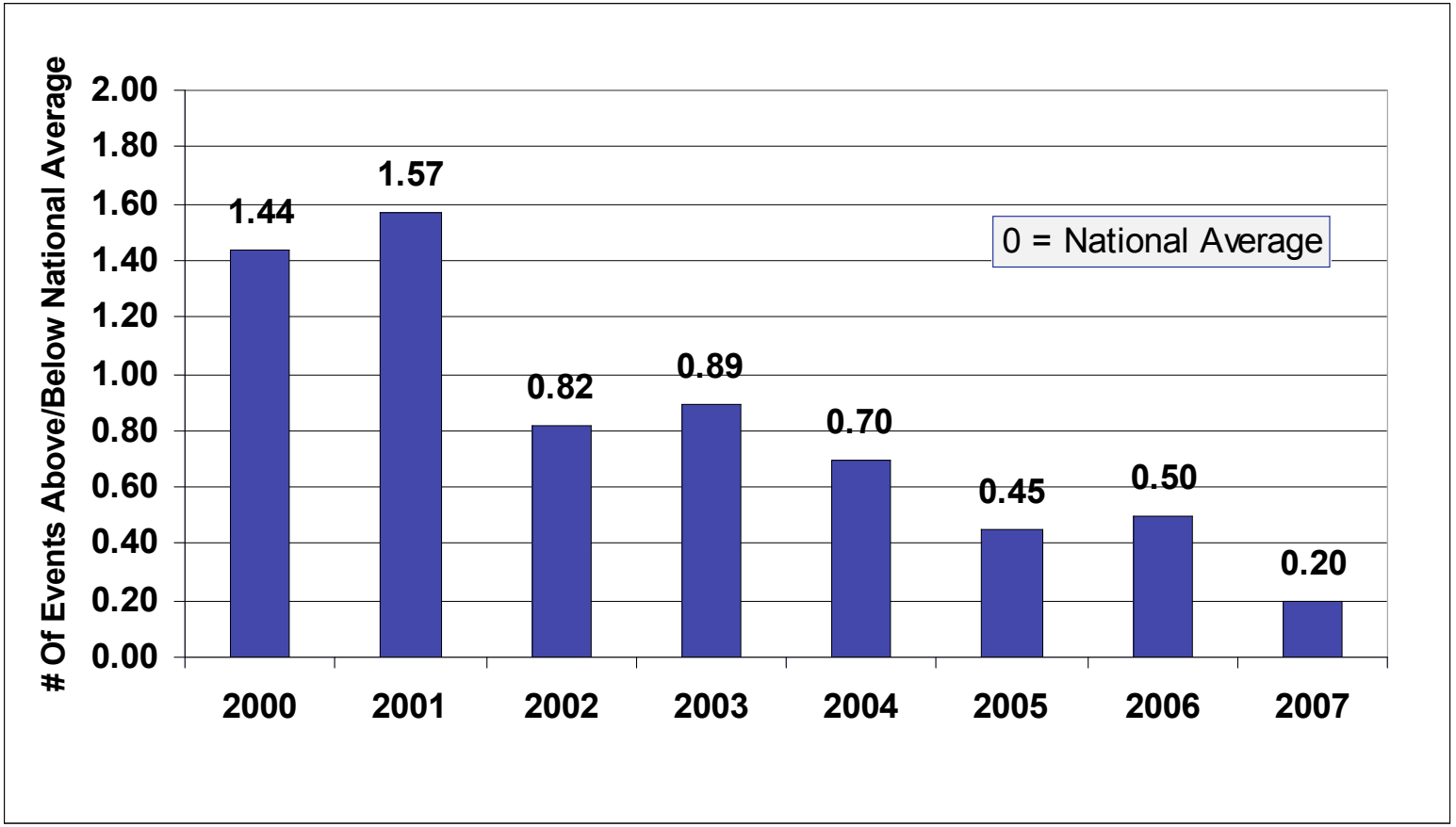



Inpatient and ER Event Rates for CAD for HPHC
% Better than (below) The National Average
Top line is national average
Percent below average (POSITIVE variance)
0%
-5%
-10%
-15%
-20%
-25%
-30%
-35%
-40%
-45%
-0.05
-0.18
-0.25
-0.40
-0.33
-0.27
-0.27
-0.36
2001
2002
2003
2004
2005
2006
2007

ER and Inpatient Events Per 1,000 Commercial Members
Harvard Pilgrim Compared To National Average
Asthma



Comparison of Inpatient and ER Event Rates for Asthma
# of Events Above/Below The National Average
Your Plan vs.National Average - Commercial



Calculating ROI validly from event rates

- Size of ROI from DM: lower
- Emphasis on ROI from DM: higher

Impact

- Size of ROI
- Emphasis on ROI: higher

- Credibility of ROI:
  Priceless

Conclusion: Event rate measurement

- Provides the only valid measurement (can be translated into ROI with a tool which I can send on request)
- This is NOT "he said-she said." The "pre-post" methodology (which I invented, by the way – google "invented disease management") is simply wrong
- Some vendors perpetuate pre-post because it shows high savings and many consultants perpetuate it because it generates a lot of consulting fees
- "Then why doesn't everyone use event rates?"
  - Many payors DO use it, including Harvard-Pilgrim. No payor has ever switched back, and 30 switched to using it last year
    - Because no one makes money on it, no one advocates it, like "counterdetailing" for generics vs. detailing for name brands

Agenda

- Background on Disease Management Purchasing Consortium
- First Review of Materials emailed ahead of meeting
  - Powerpoint example
  - Word Example
  - Mercer example
- Why these mistakes exist – “planes on the ground” and “heads or tails” and “trend assumptions”
- How to measure validly at low cost
- Examples of mistakes – second review of earlier materials and some watch-outs
- Test
- Time-out to take test (2 hours)
- Review of answers

Baseline knowledge: What you need to know to review reports (all <65)

- Annual spending per person (average and by the top 5 conditions)
- Cost per day in hospital
- Cost per ER visit
- Heart attack rates (short term memory)
- Asthma attack rates
- MD visit rates
- Admit rates per 1000

Seven other things to consider in reviewing

- More important to know when a number is wrong than what the right number could be
- Look across pages – often comparisons of pages reveal obvious errors not evident on a single page
- Things which should go down, go up (like drugs)
- “Heads or tails” mistakes (start with high numbers)
- Too-high ROIs
- Massive changes attributed to program
- Changes in areas where changes not expected
- Math errors

#1 Example-- A real submission disguised

Category	Base	Intervention
Total Comm. Membership	505,000	511,000
Prevalence of selected case mgmt conditions	23%	23%
Annual claims cost	$972	$935
Annual admission rate	99	79
Annual Admission cost	$261	$244
Annual MD visit rate per 1000 members	4535	4475
Annual MD visit costs/member	$132	$128
Annual ED visit rate/1000	452	339
Annual ED costs/member	$39	$33

#2 Actual report example in Word

- Switch over to Word
- Real example. Names and any other identifiers removed and some numbers changed (though not the point of the numbers) for copyright reasons
  - No copyright on cluelessness

#3 Now (or after the test) look at the Mercer report again

- See how many more red flags you can find

What else to look for in the test and in general (keep this page handy for the test and in general)

- Magnitudes which don't make sense
  - Axis way off
- Changes of magnitudes which don't make sense
- Changes in categories which don't make sense (like drugs declining)
- Changes in prevalences which don't make sense
- Inconsistencies between pages or between data on the same page (like drug cost goes down but compliance increases)
- Math problems

More Watchouts not covered in the test or course

- "Our results have been validated by…"
- Results dependent on trend
- ROIs which "bounce" a lot between conditions
- ROIs calculated at different severity levels (unbelievably dumb idea)
- ROIs not confirmed by an "event-rate plausibility test"
- Large savings with low number of calls completed

Now it's test time for those who are taking it

- We will reconvene in 2 hours to do answers
- You need to have the tests sent to alewis@dismgmt.com by 4 PM EDT
- If you want to do something similar for wellness outcomes measurement, we will do that separately *gratis* at your convenience – I advertised wellness but realized I couldn't' fit it in, so I will go over it one-on-one with anyone who wants to