DMPC

Disease Management

Purchasing Consortium Advisory Council

www.dismgmt.com

Disease Management Outcomes

Gimme Three Steps, Gimme Three Steps towards validity

DMPC
Disease Management
Purchasing Consortium Advisory Council

 www.dismgmt.com

Hey, Butch, Who Are These Guys

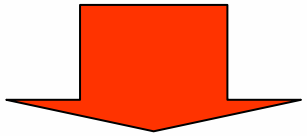
- DMPC is me (with a little help from my friends)
- Invented DM contracting (source: *Managed Healthcare Executive* March 2003)
- Founded DMAA
- Currently offers procurement/measurement consulting, and Certifications for Savings Validity and for Critical Outcomes Report Analysis

What You've Heard

- View(s) which supports the DMAA methodology
- View(s) which opposes the DMAA methodology

What you are about to hear

- View(s) which supports the DMAA methodology
- View(s) which opposes the DMAA methodology



Let the data speak for itself and then make up your own mind

The Three Steps: Observations which form the basis for this presentation

1. Your data could show Regression to the Mean.
   - It will vary according to several factors (annual vs. prospective, disease, Length of ID period, algorithm) and may be 0 or even more than offset by disease progressivity in some cases
2. Instead of denying it, acknowledge the possibility of it and...
3. ...TEST for it, and then check the plausibility of your result with a "confirming analysis"

Example of data showing regression to the mean

- Assume no inflation,
  no claims other than asthma
  - These assumptions just simplify.
    They don't distort

Example from Asthma
First asthmatic has a $1000 IP claim in 2004

	2004 (baseline)	2005 (contract)
Asthmatic #1	1000	
Asthmatic #2		
Cost/asthmatic		

Example from Asthma

Second asthmatic has an IP claim in 2005 while first asthmatic goes on drugs (common post-event)

	2004 (baseline)	2005 (contract)
Asthmatic #1	1000	100
Asthmatic #2	**0**	**1000**
Cost/asthmatic		

What is the Cost/asthmatic In the baseline?

Cost/asthmatic in baseline?

	2004 (baseline)	2005 (contract)
Asthmatic #1	1000	100
Asthmatic #2	0	1000
Cost/asthmatic	**$1000**	

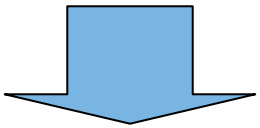
Vendors don't count #2 in 2004 bec. he can't be found

Cost/asthmatic in contract period?

	2004 (baseline)	2005 (contract)
Asthmatic #1	1000	**100**
Asthmatic #2	0	**1000**
Cost/asthmatic	**$1000**	**$550**

How can you find people like Asthmatic #2 in advance?

- HRAs
- Two years of identification for baseline



This will help but not eliminate RTM. (Ask me for My proof that a 2-year baseline doesn't eliminate RTM If I don't have time to show it.)

The Three Steps: Observations which form the basis for this presentation

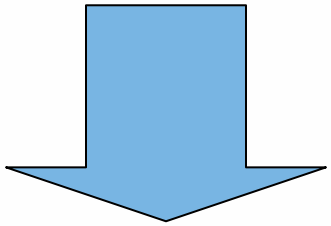
1. Your data could show Regression to the Mean.
   - It will vary according to several factors (annual vs. prospective, disease, Length of ID period, algorithm) and may be 0 or even more than offset by disease progressivity in some cases
2. Instead of denying it, acknowledge the possibility of it and...
3. ...TEST for it, and then check the plausibility of your result with a "confirming analysis"

Do current methodologies fix regression to the mean?

- "Annual Requalification"
  - A person only counts in years in which they "requalify" with claims
- "Prospective Requalification"
  - Once chronic, always chronic
  - Once you are identified, you are counted in all subsequent years

Do current methodologies address this problem?

- “Annual Requalification”
- “Prospective Requalification”



Let’s re-look at that exact example and see how requalification Policies affect it

In annual requalification, the first asthmatic requalifies in 2005 and the second qualifies for the first time in 2005

	2004 (baseline)	2005 (contract)
Asthmatic #1	1000	**100**
Asthmatic #2	0	**1000**
Cost/asthmatic	**$1000**	**$550**

In "prospective qualification," the same thing happens

	2004 (baseline)	2005 (contract)
Asthmatic #1	1000	**100**
Asthmatic #2		**1000**
Cost/asthmatic	$1000	**$550**

In this case both approaches give the same result

But now assume that #1 doesn't fill his prescriptions

	2004 (baseline)	2005 (contract)
Asthmatic #1	1000	**0**
Asthmatic #2	0	**1000**
Cost/asthmatic	**$1000**	

But now assume that #1 doesn't fill his prescriptions: Prospective shows this result

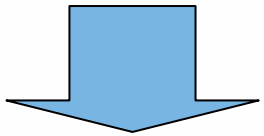
	2004 (baseline)	2005 (contract)
Asthmatic #1	1000	**0**
Asthmatic #2	0	**1000**
Cost/asthmatic	**$1000**	**$500**

But now assume that #1 doesn't fill his prescriptions: Annual shows this result

	2004 (baseline)	2005 (contract)
Asthmatic #1	1000	**0**
Asthmatic #2	0	**1000**
Cost/asthmatic	**$1000**	**$1000**

Annual vs. Prospective

- They can't both be right.* This is math. There is only one right way to do things
  - Prospective makes epidemiological sense
  - But annual is more likely to give the right math answer



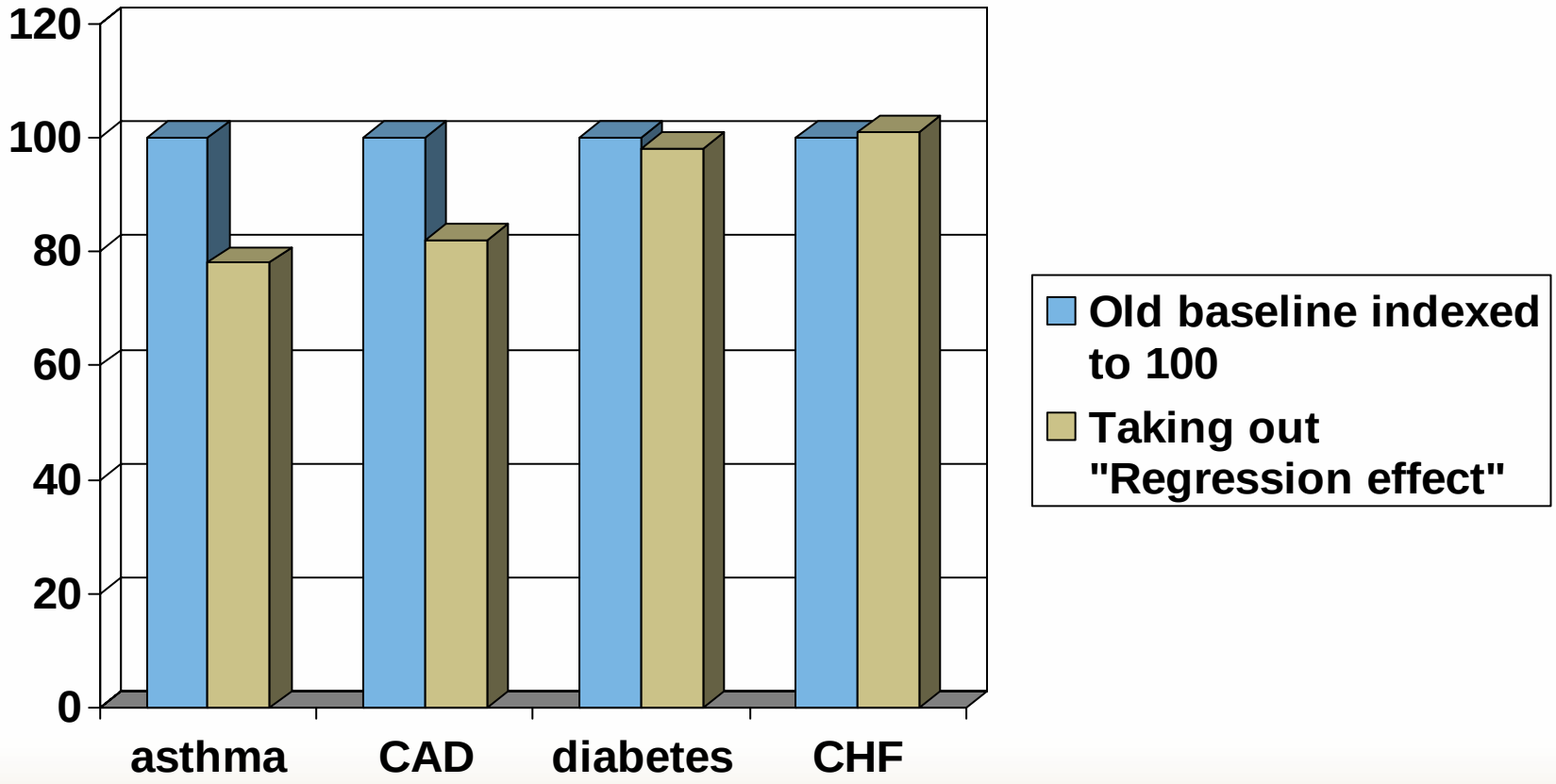
Verdict: Use Annual Requalification, of the two choices

*Source: Mark Knopfler

The Three Steps: Observations which form the basis for this presentation

1. Your data could show Regression to the Mean.
   - It will vary according to several factors (annual vs. prospective, **disease**, length of ID period, algorithm) and may be 0 or even positive in some cases
2. Instead of denying it, acknowledge the possibility of it and...
3. ...TEST for it, and then check the plausibility of your result with a "confirming analysis"

Examples: By Disease (using 1-year baseline and standard algorithms; ask me for my standard algorithms) – what is the difference which is caused automatically by just trending forward like we just did?



The Three Steps: Observations which form the basis for this presentation

1. Your data could show Regression to the Mean.
   - It will vary according to several factors (annual vs. prospective, disease, Length of ID period, algorithm) and may be 0 or even more than offset by disease progressivity in some cases
2. Instead of denying it, acknowledge the possibility of it and...
3. ...TEST for it, and then check the plausibility of your result with a "confirming analysis"

“Example” by number of years of baseline ID

- Use a simplified version of a health plan to see what happens when you move from 1 to 2 years of member identification

Example: The actual situation is that (taking trend out) nothing changed…but see how Much RTM there is in 1 year vs. 2 years (prospective qualification)

Claimant	Baseline Period(s)		study period	Claimed	Actual Savings
	2003	2004	2005	Savings	
#1	100	0	200		
#2	0	0	50		
#3	20	0	0		
#4	40	100	50		
#5	40	100	0		
#6	100	100	0		
total	300	300	300		
true cost/member	50	50	50		there are none!

Example with one year of baseline: running the numbers

Claimant	2004-baseline
#1	0
#2	0
#3	0
#4	100
#5	100
#6	100
total	300
Number of identified members to divide by	3
cost/identified member, 1 year of baseline	$100

These are the only Three people who Are findable

Now we are adding the members identified
In the study period (2005)

Claimant	2004	2005
#1	0	200
#2	0	50
#3	0	0
#4	100	50
#5	100	0
#6	100	0
total	300	300
Number of identified members to divide by	3	5
cost/identified member, 1 year of baseline	$100	$60

You now find two More people

To go with the Three you Already found

Now we are adding the members identified In the study period (2005)

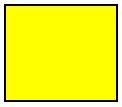
Claimant	2004		2005		Claimed Savings	Actual Savings
#1	0		200			
#2	0		50			
#3	0		0			
#4	100		50			
#5	100		0			
#6	100		0			
total	300		300			
Number of identified members to divide by	3		5			
cost/identified member, 1 year of baseline	$100	-	$60	=	$40	$0

So now we see that

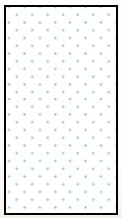
- One year of baseline doesn't work
- Let's see if two years solves it

Now try two years of ID-ing for baseline

Claimant	2003 Baseline years	2004 Baseline years
#1	100	0
#2	0	
#3	20	0
#4	40	100
#5	40	100
#6	100	100
total	300	300
Number of ID'd members, two years baseline		5
cost/identified member, 2 years of baseline		$60

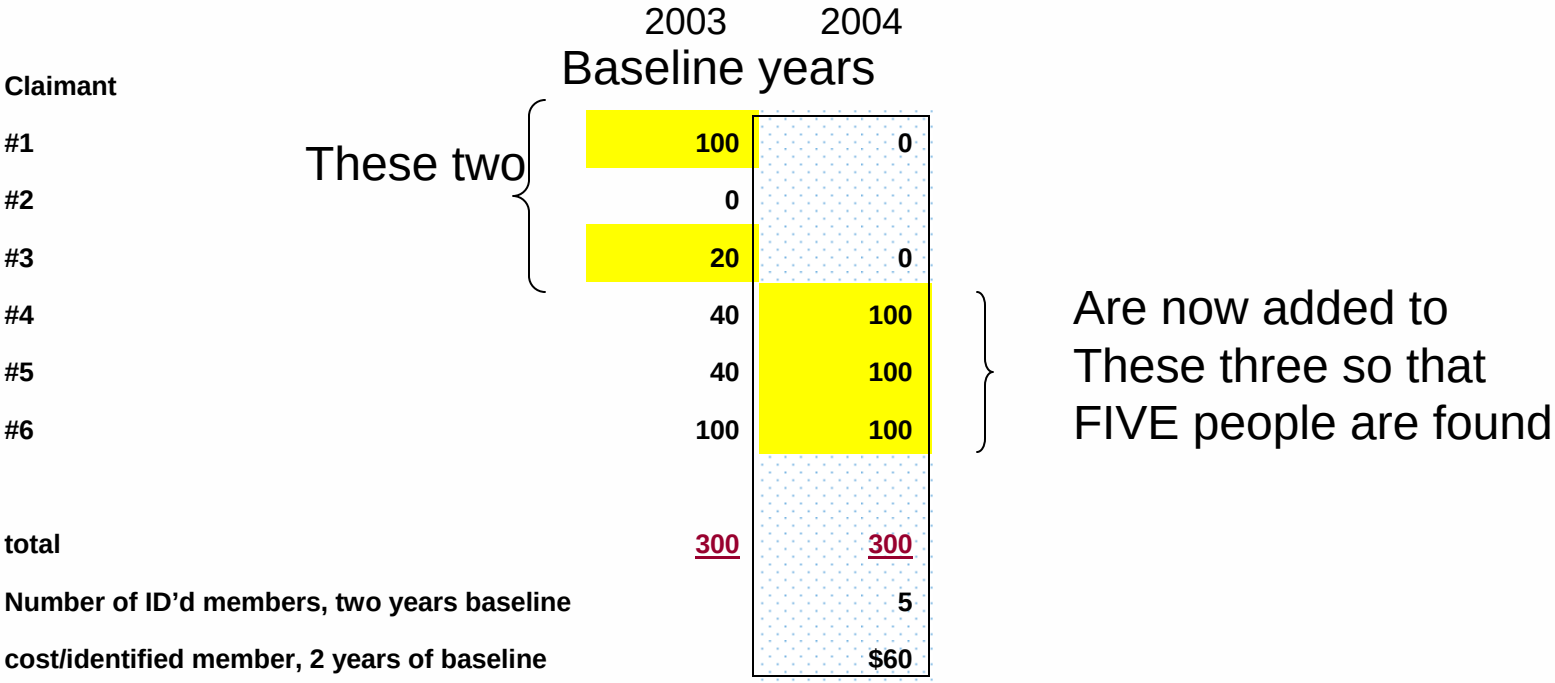


= members identified



= member claims and member-years in baseline

Now try two years of ID-ing for baseline, which is in some contracts



= members identified

= member claims and member-years in baseline

Two years of baseline does not solve the problem in prospective qualification

	Baseline years		Study		
Claimant	**2003**	**2004**	**2005**	**Claimed Savings**	**Actual Savings**
#1	**100**	**0**	**200**		
#2	**0**	**0**	**50**		
#3	**20**	**0**	**0**		
#4	**40**	**100**	**50**		
#5	**40**	**100**	**0**		
#6	**100**	**100**	**0**		
total	**300**	**300**	**300**		
Number of ID'd members, two years baseline		**5**	**6**		**$0**
cost/identified member, 2 years of baseline		**$60**	**$50**	**$10**	**$0**

The Three Steps: Observations which form the basis for this presentation

1. Your data could show Regression to the Mean.
   - It will vary according to several factors (annual vs. prospective, disease, Length of ID period, algorithm) and may be 0 or even more than offset by disease progressivity in some cases
2. Instead of denying it, acknowledge the possibility of it and...
3. ...TEST for it, and then check the plausibility of your result with a "confirming analysis"

Two algorithms, two results (CAD): Algorithm #1

All eligible recipients who have a CAD Diagnosis (414.x) OR have the following events or procedures in their claims

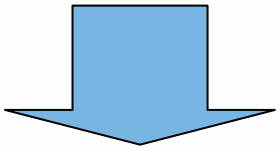
Event	CPT	Revenue Code	ICD 9 Procedure or ICD-9 Dx code	DRG
Catheterization	93501 93510 93511 93514 93524-29 93542 93543 93545-56	481	37.21-37.23 38.91	124 125
PTCA/Stent	92980-92984 92995,92996		36.01-36.09	112
CABG	33510-33530 33533-33536		36.10- 36.19	106 107 108
Acute MI	No procedure		410 (primary or secondary)	121, 122, 123
IHD admissions (angina, rule-out MI etc.)	No procedure		411-414 (primary or secondary)	132, 133, 140, 143

Algorithm #2: #1 plus…

- >50 yrs old with diagnosed diabetes, HTN or morbid obesity
- Drug codes: 3 or more concurrent fills for antihyperlipidemics plus antihypertensives

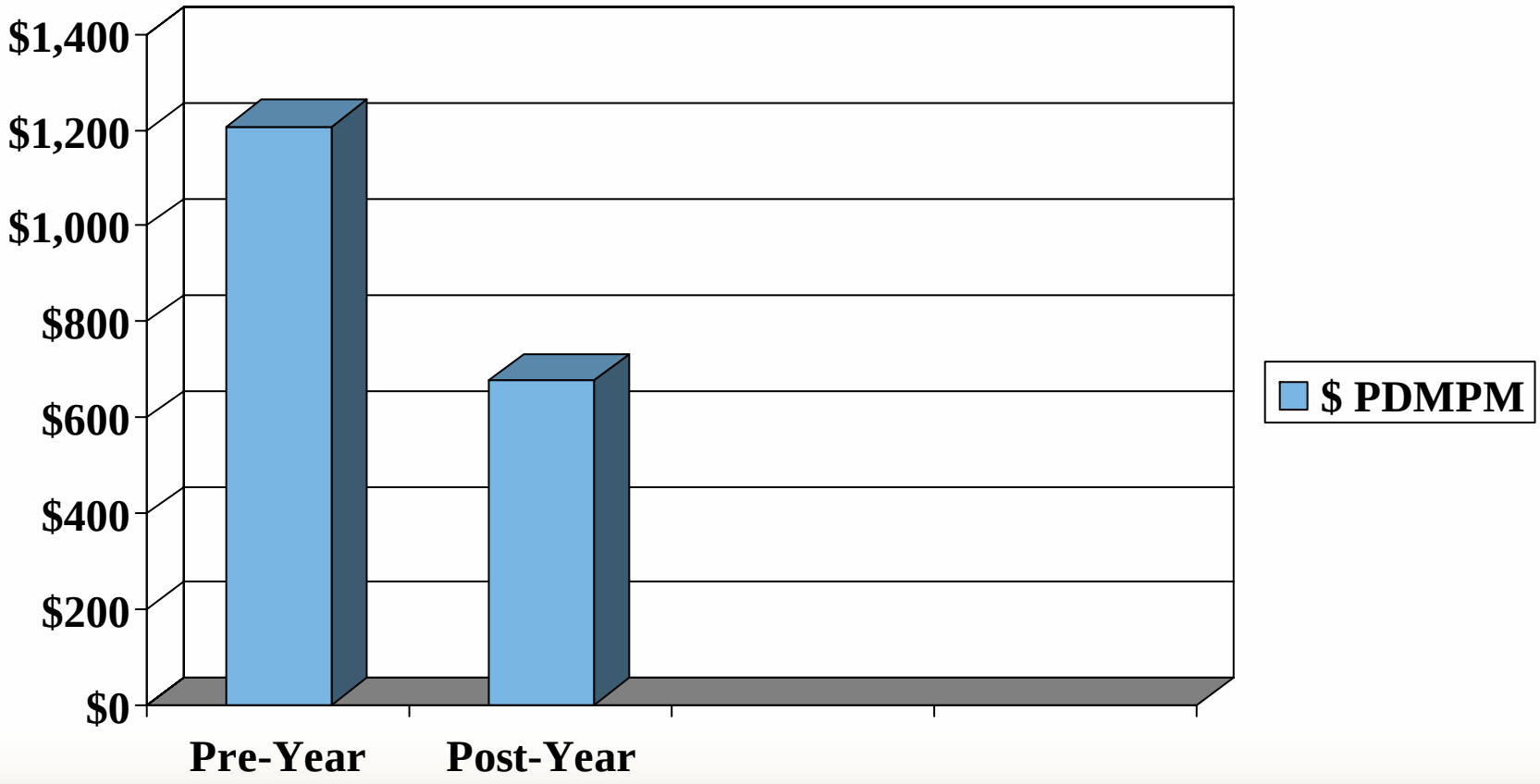
Using the second algorithm

- Increases the likelihood of finding people who will have events or lots of drugs in the study year – hence increases the study year cost (tradeoff is reduced specificity, of course)
- Reduces the baseline cost because not everyone in the baseline will have had an event or procedure

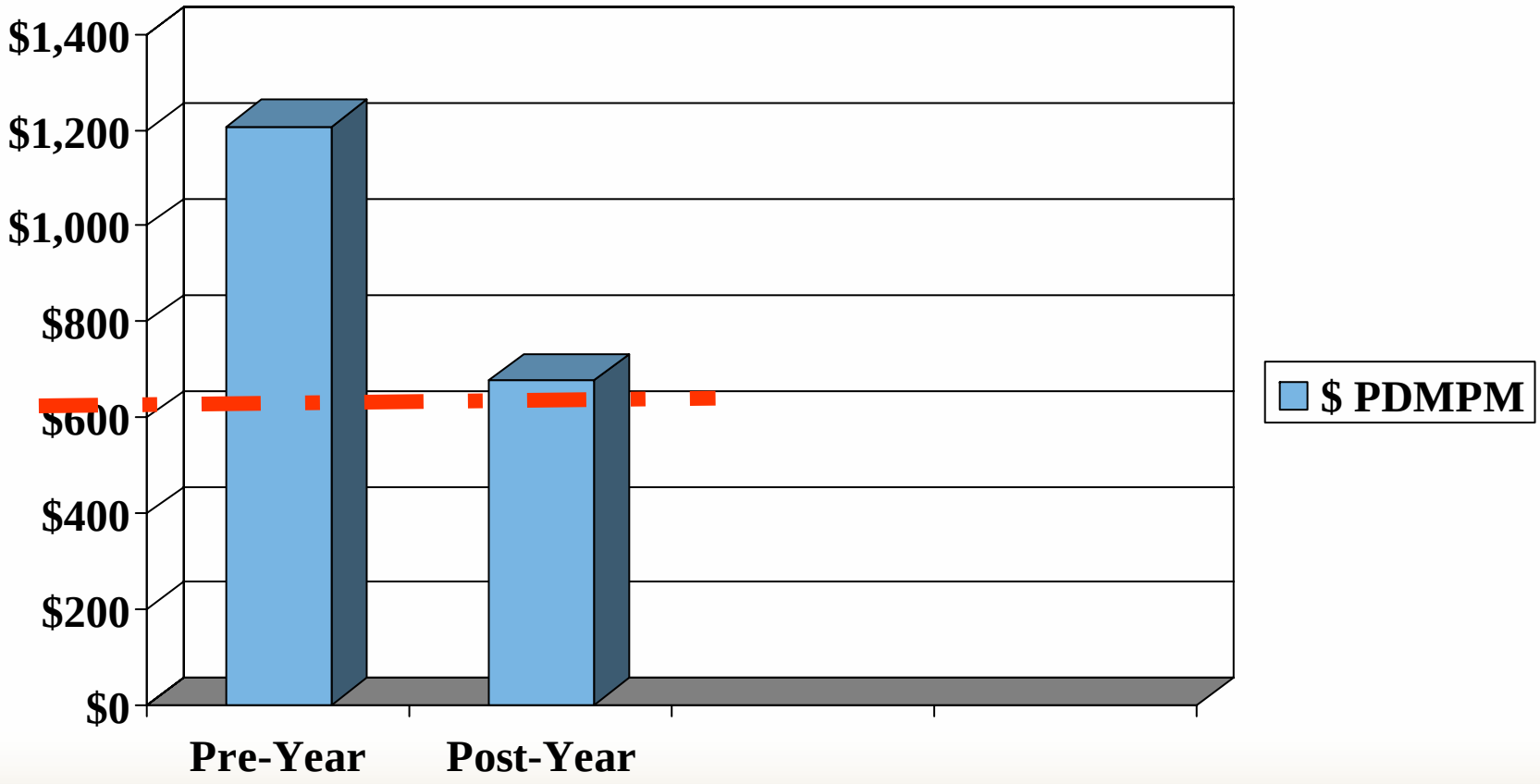


Both effects will serve to reduce the regression-generated savings as can be seen

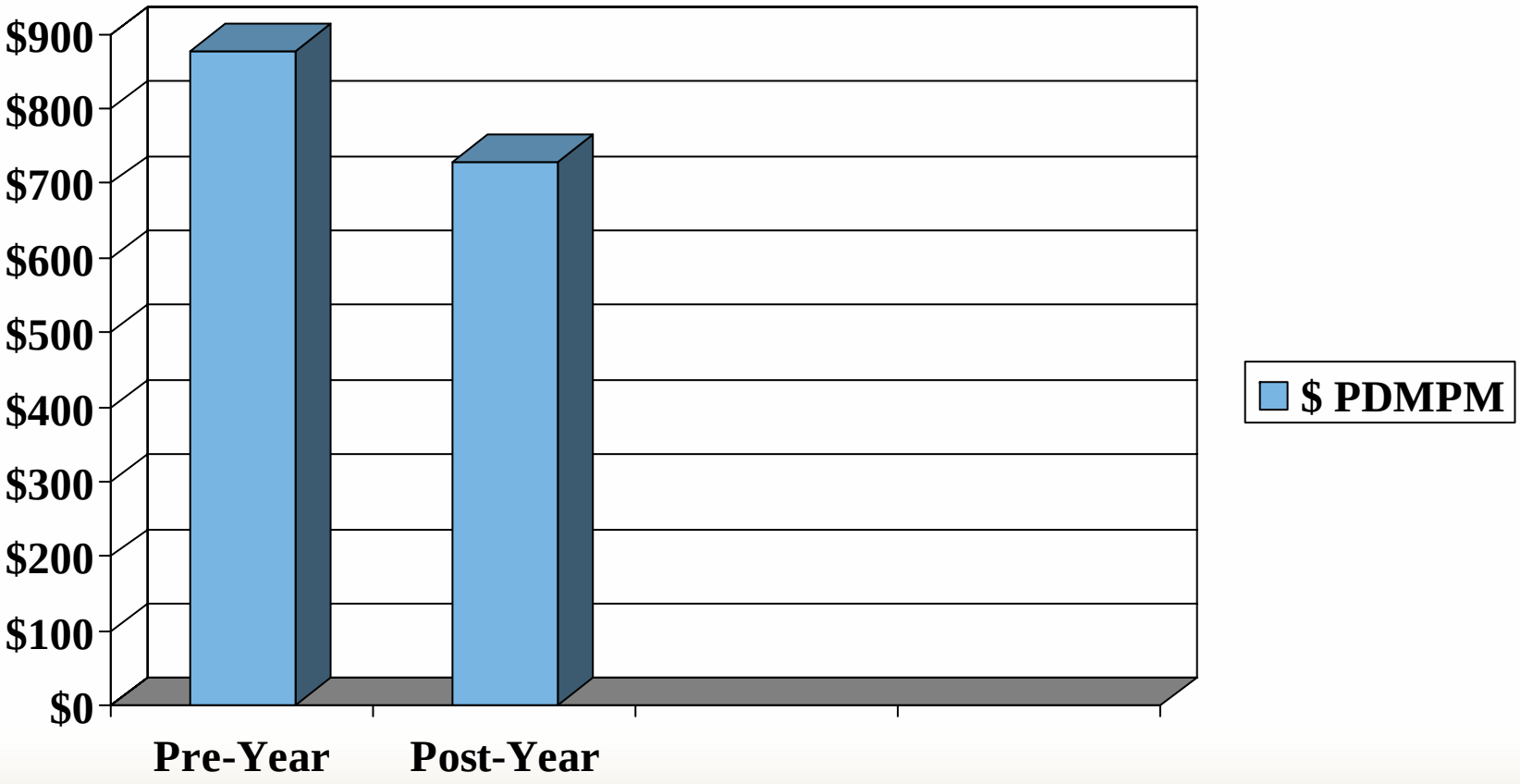
Using first algorithm



Sidebar: You know this one is bogus because hospitalization is only half of costs to begin with

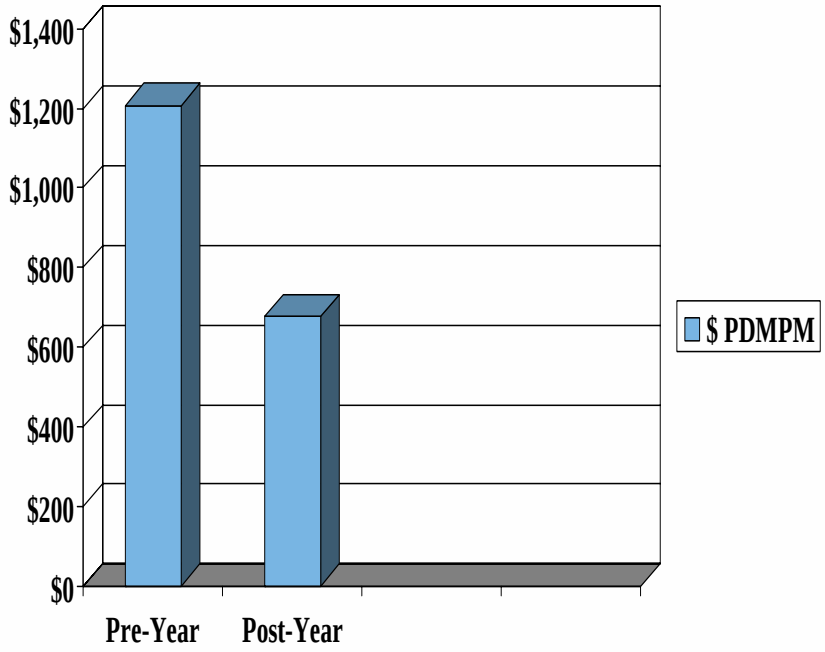


Using second algorithm

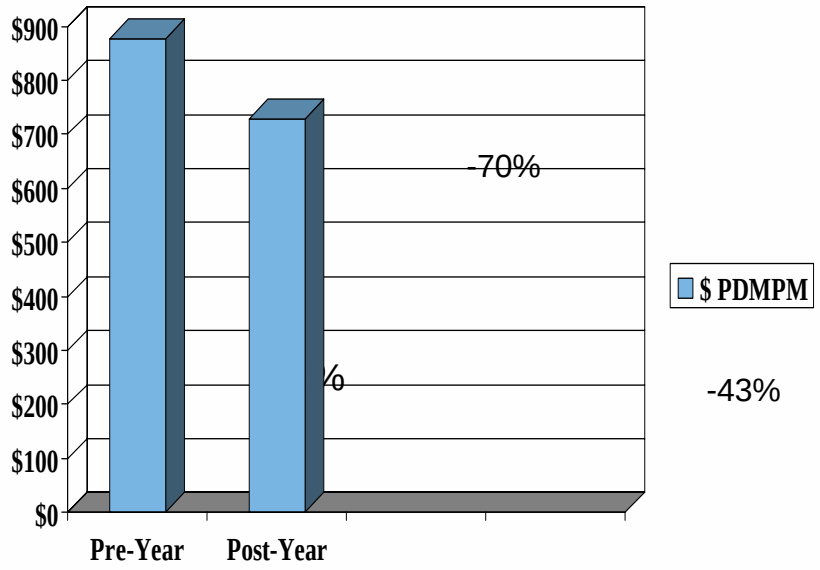


Impact of Algorithm on CAD baseline and study period cost

FIRST



SECOND



Fuzzy Math?

- What if all my examples are wrong and there is no RTM?
  - What is proposed: A test, not an alternative methodology
    - Keep the DMAA methodology
  - Watch how it would work
  - Even if my math is wrong a test is still valuable because it proves *no* RTM

The Three Steps: Observations which form the basis for this presentation

1. Your data could show Regression to the Mean.
   - It will vary according to several factors (annual vs. prospective, disease, algorithm, length of ID period) and may be 0 or even more than offset by disease progressivity in some cases
2. Instead of denying it, acknowledge the possibility of it and...
3. ...TEST for it, and then check the plausibility of your result with a "confirming analysis"

Test: Try applying the methodology – disease, length of baseline ID, algorithm -- in the *absence* of disease management

- Use data from a year or population in which there was/is no DM
- Create a "dummy" baseline and trend it forward to see what happens naturally absent DM

In this hypothetical, the effect is 45%

	2004 (baseline)	2005 (contract)
Asthmatic #1	1000	**100**
Asthmatic #2		**1000**
Cost/asthmatic	$1000	**$550**

45% reduction would happen Anyway even without DM

Let's look at this non-hypothetical example

- IRVING, Texas--(BUSINESS WIRE)--Nov. 18, 2003--A pediatric asthma disease management program offered by Advance PCS 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

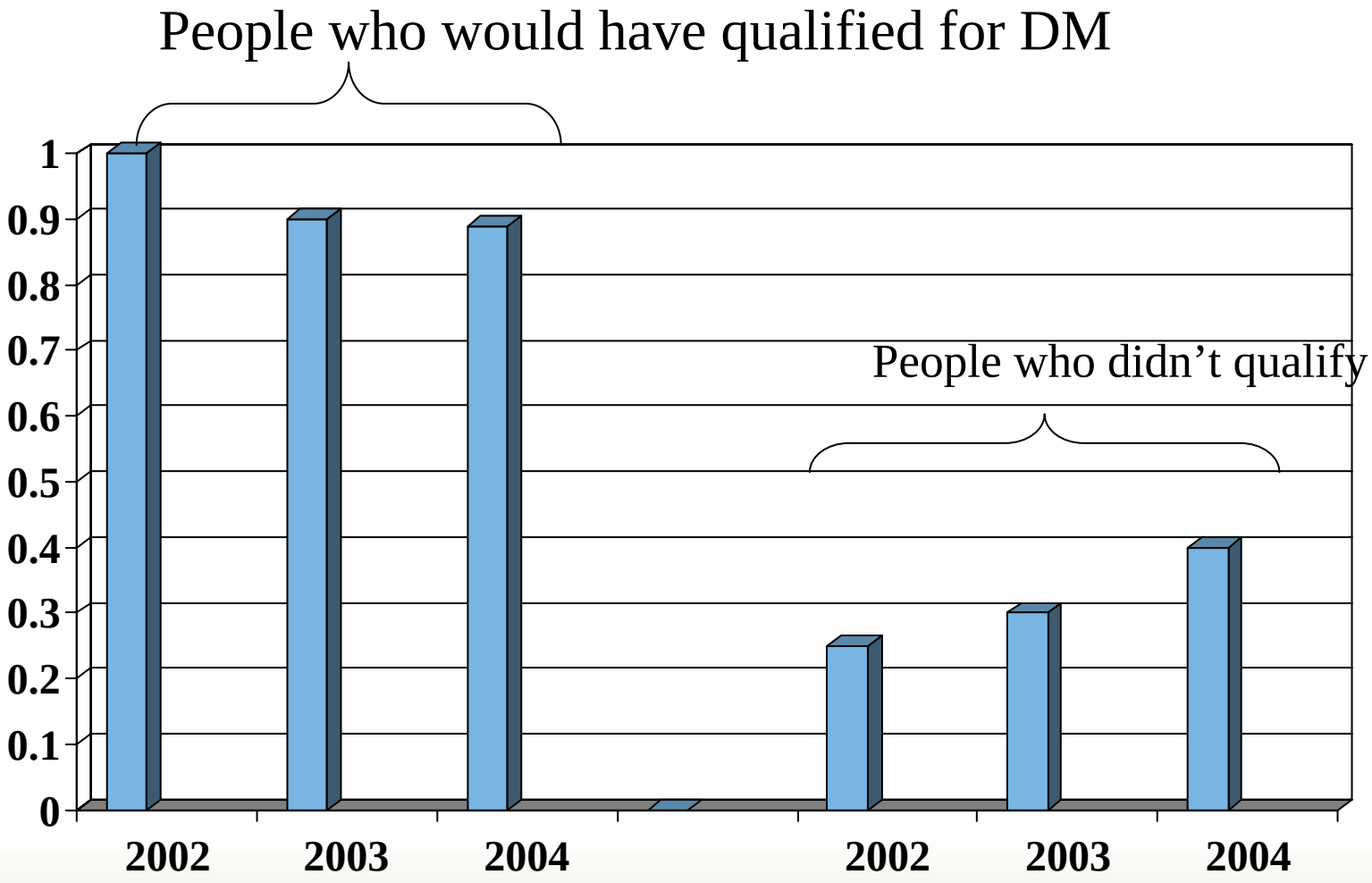
Example from North Carolina: What they think they accomplished

	2001 (baseline)	2002 (Advance PCS)
Diagnosed Asthmatics	100	67
Asthmatics not previously diagosed		
Savings:		33%

Example from North Carolina: What they actually accomplished

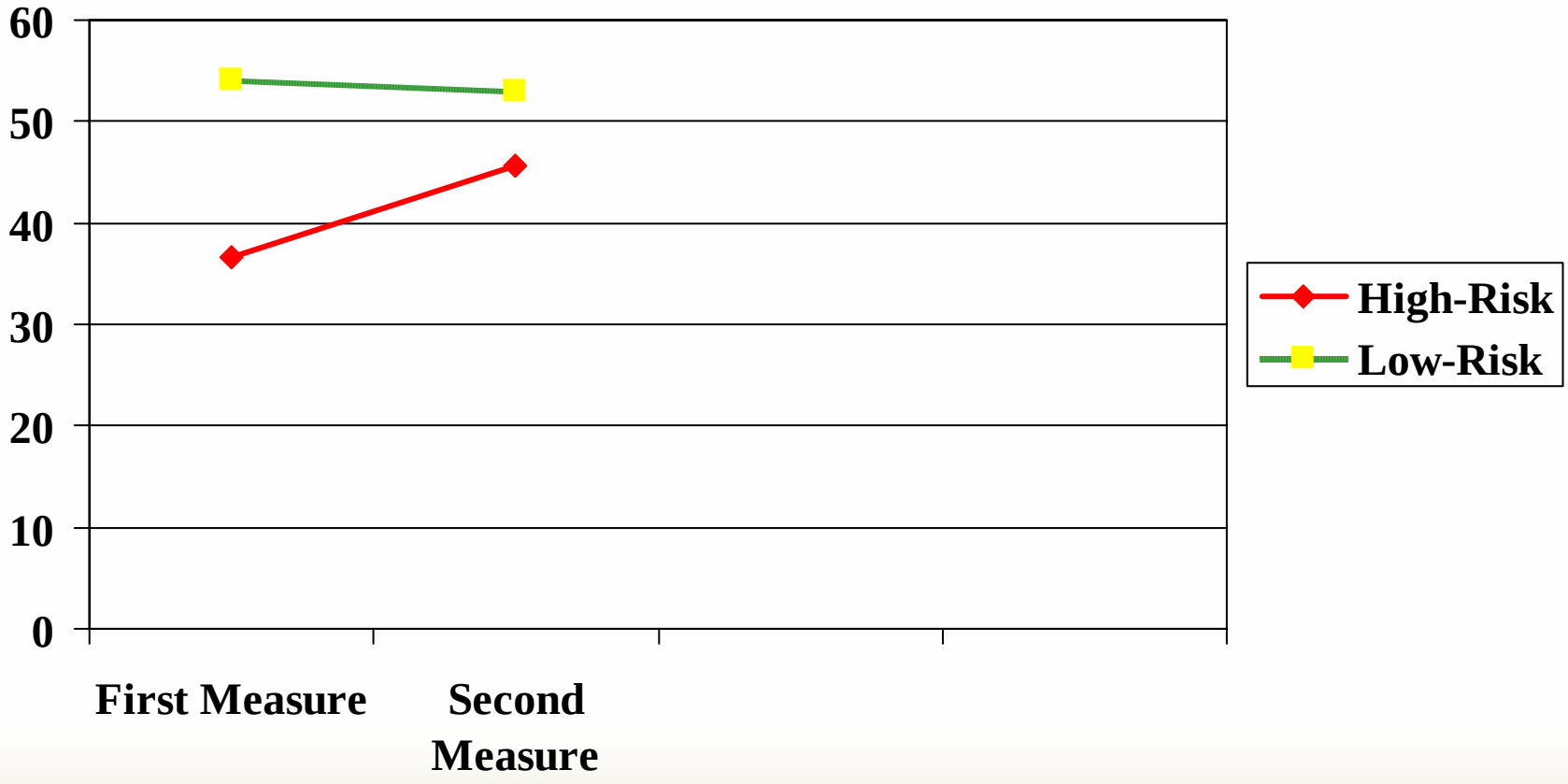
	2001 (baseline)	2002 (Advance PCS)
Diagnosed Asthmatics	100	67
Asthmatics not previously diagnosed		33
Savings:		0

Another Example: Vermont Medicaid (note: Numbers are approximations indexed to 1.0*)



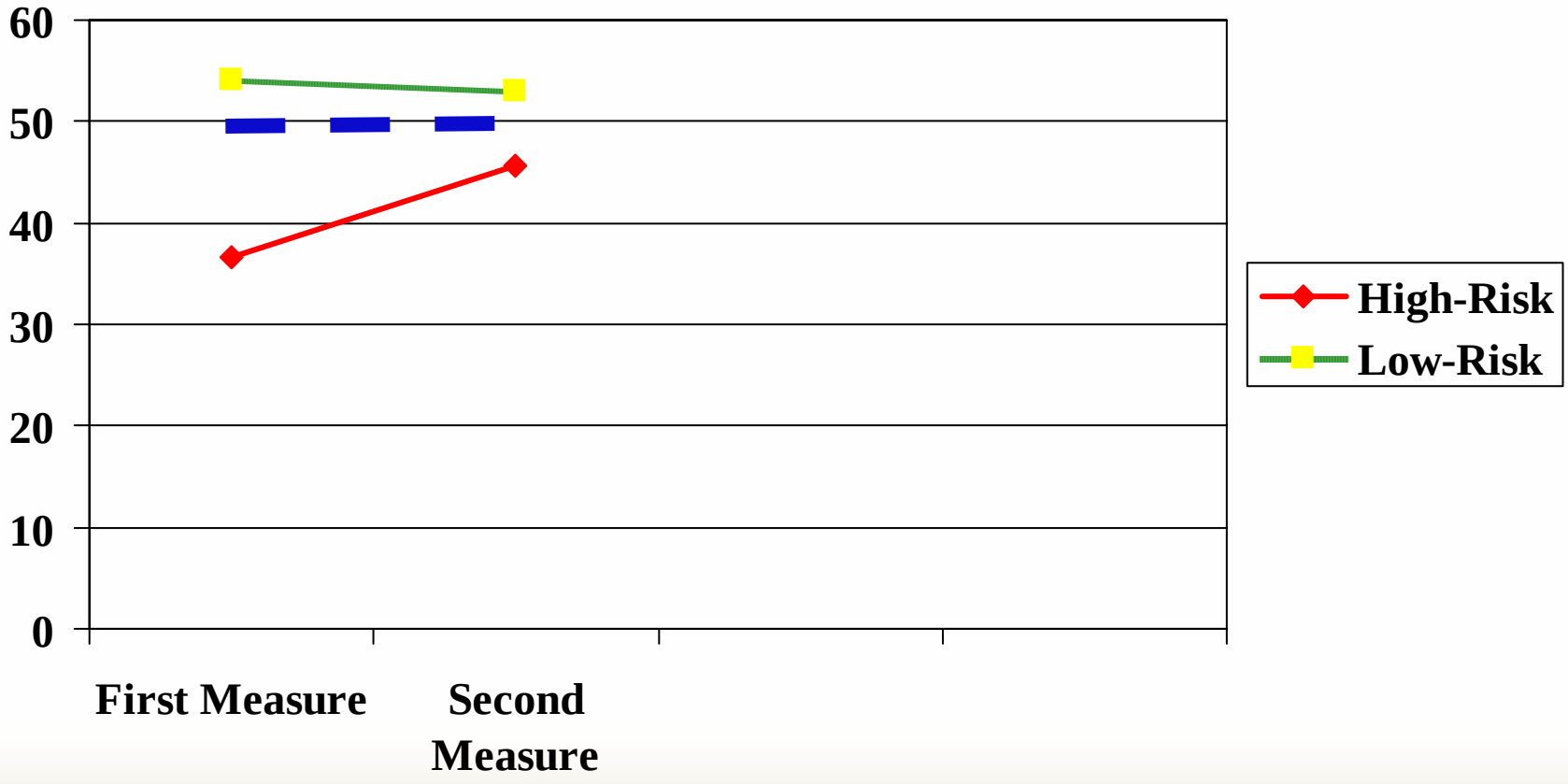
*Because I can't remember exactly and you have to sit through an entire presentation to see this slide: www.NESCSO.org

This is a pretty good wellness outcome as measured by an SF 12– the high-risk group (25% of total) dramatically improved their scores between periods



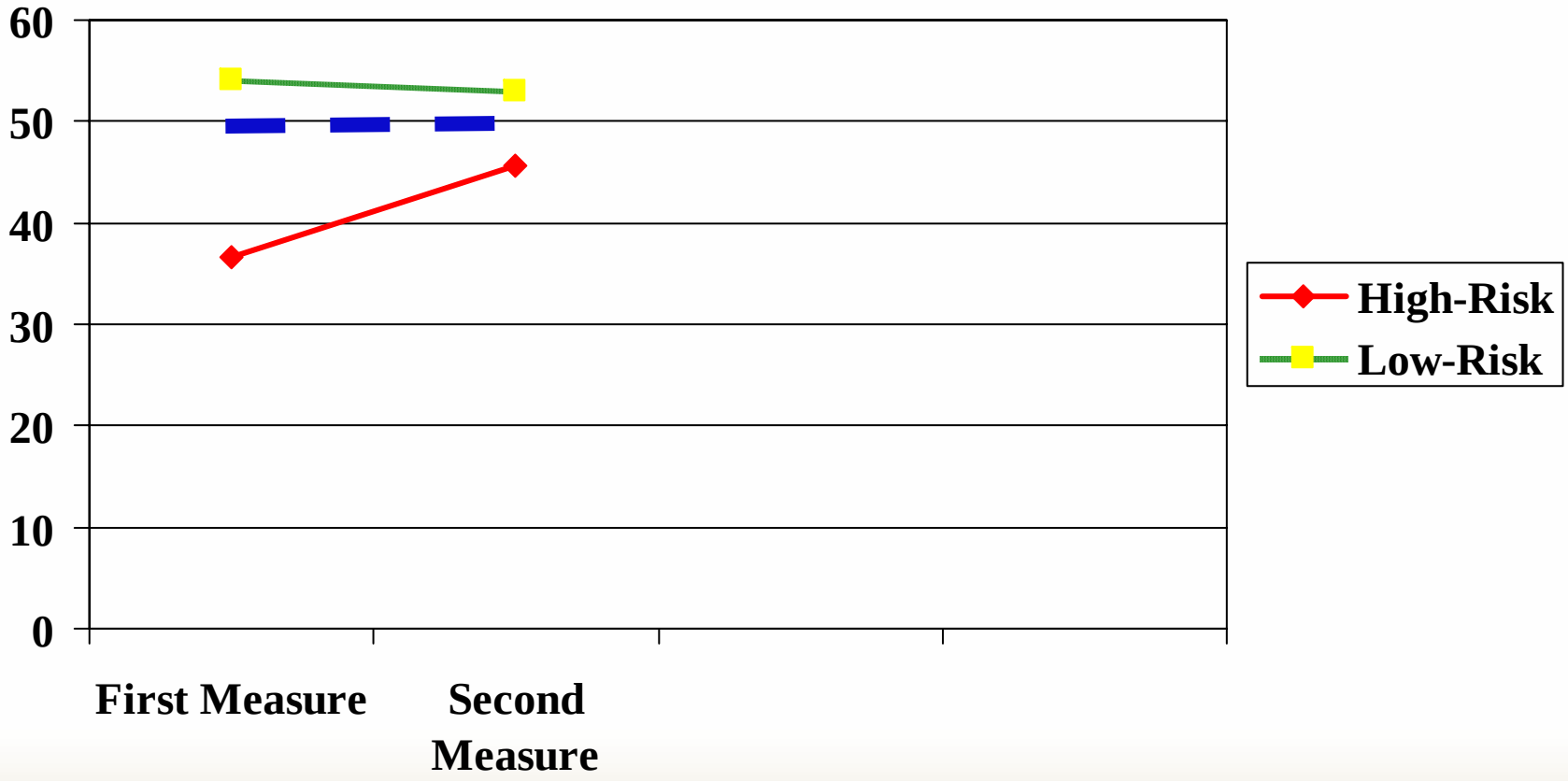
Source: Ariel Linden – citations Following presentation

But for some reason the average stayed the same (weighted because high-risk was only 25%)



Source: Ariel Linden – citation on request

Why did the average stay the same? *Because there was no program in this case* – just two samplings



Source: Ariel Linden

This SF-12 test example implies:

- Any wellness program where they coach people with high risk factors will overstate risk reduction dramatically on that high-risk group
- By analogy, any absence management program which starts with people who had high absences will show significant reduction automatically

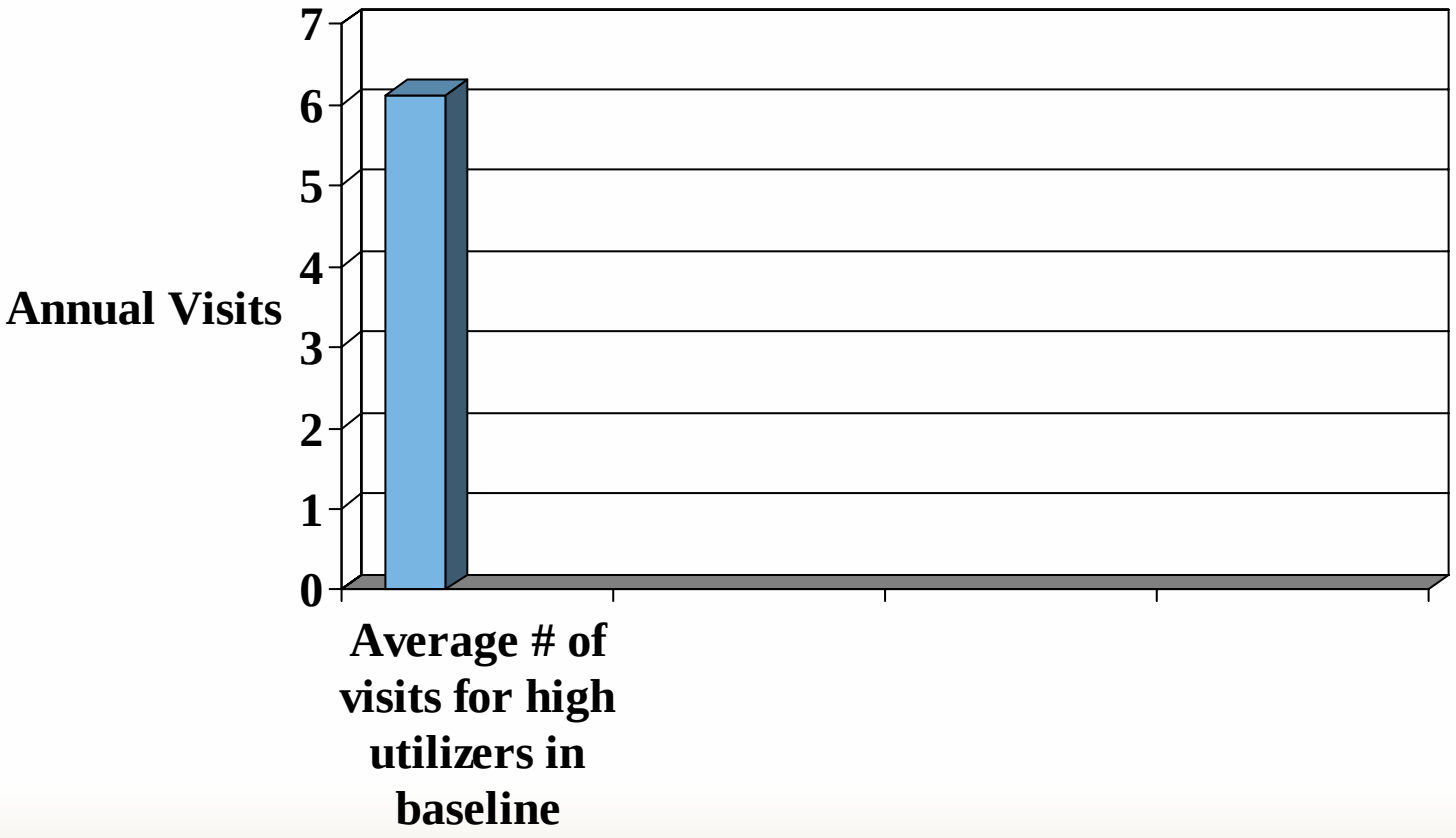
It's best to test for *multiple* "dummy" years and then take the effect out of the baseline as in this example from Illinois' Medicaid ER reduction program

- Goal is to reduce the visits of high utlizers
- But 5 years of "dummy data" suggests that people with 5+ visits in Year 0 fall 40% in subsequent year
  - So vendor value-added starts at 41% reduction

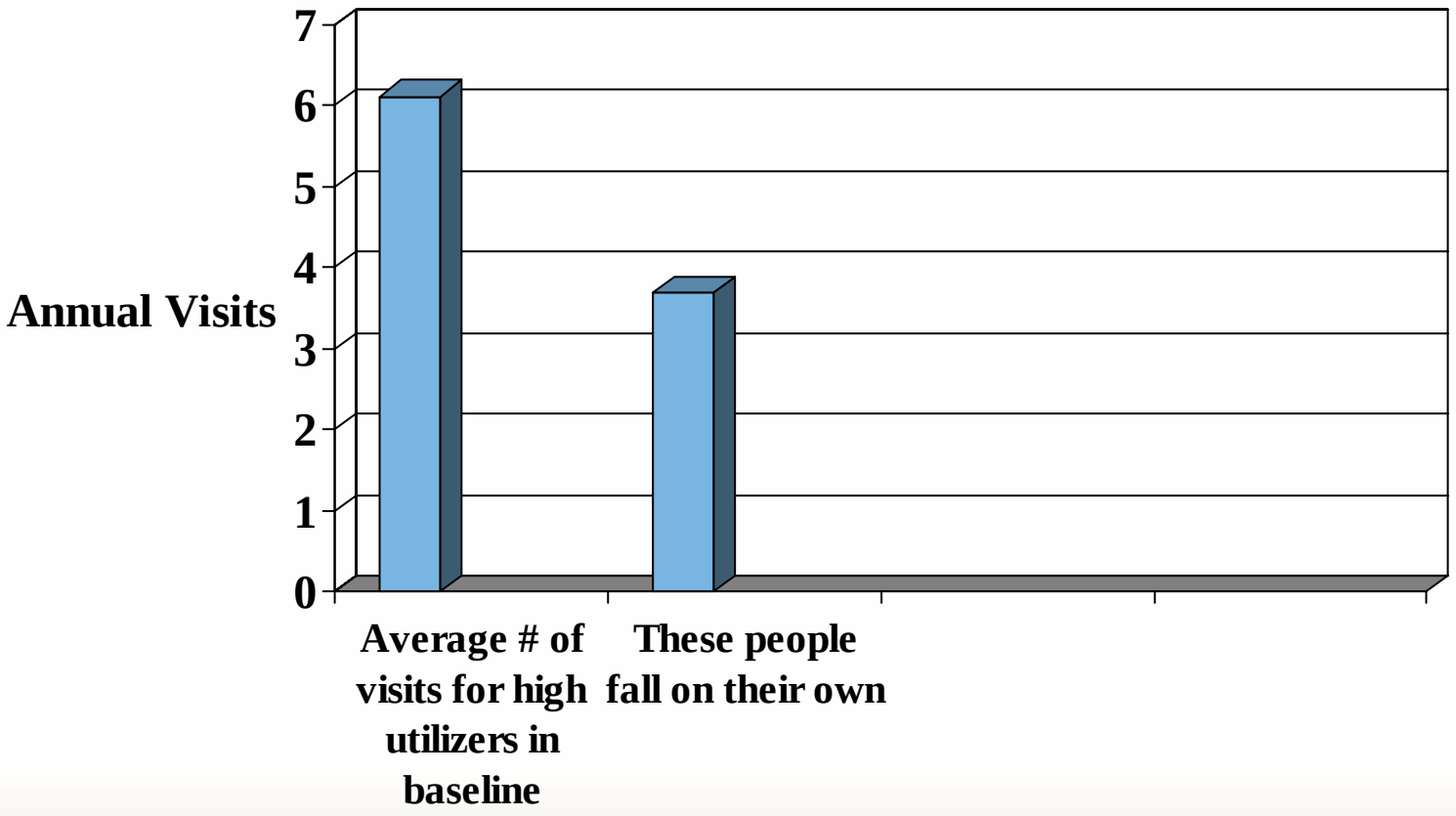
Here is the actual calculation

Average Number of Visits for Identified "Frequent ER" Population	***6.11***
Less Adjustment for "Natural Regression"	
.3960 x 6.67	***2.42***
(Average Regression from 5 cohorts as defined in contract, is equal to .3960)	
= Trended Number of Visits per Frequent ER Utilizer for FY2007	***3.69***

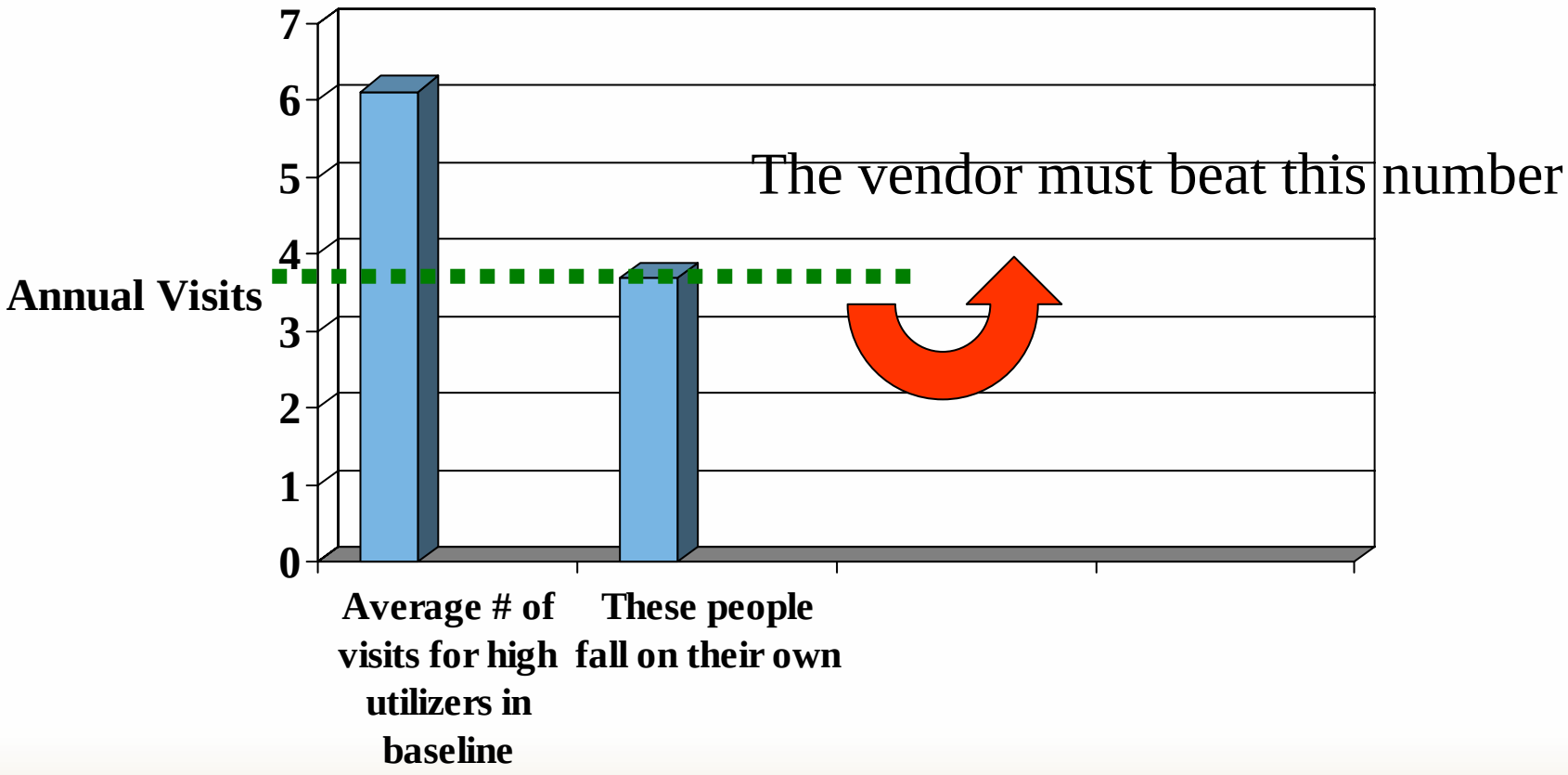
Or if you like it visually…



Or if you like it visually…



Or if you like it visually…



Advantages of testing (especially if done multiple times and results confirm one another)

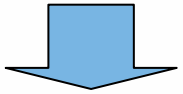
- Compensates for inevitable RTM
- Methodology-independent, algorithm-independent, years-of-baseline independent
- Vendors can’t argue the point

The Three Steps: Observations which form the basis for this presentation

1. Your data could show Regression to the Mean.
   - It will vary according to several factors (annual vs. prospective, disease, algorithm, length of ID period) and may be 0 or even more than offset by disease progressivity in some cases
2. Instead of denying it, acknowledge the possibility of it and...
3. ...TEST for it, and then check the plausibility of your result with a "confirming analysis"

Plausibility testing: Could the results have been achieved? (This is called "validation" by Ian and "preponderance of the evidence" by Ariel.)

- Plausibility check: Does this result make sense?
  - Do the quality changes support the cost changes?
  - Could it have been achieved or be achieved?
  - What do you look for in a report to identify mistakes and invalidities?



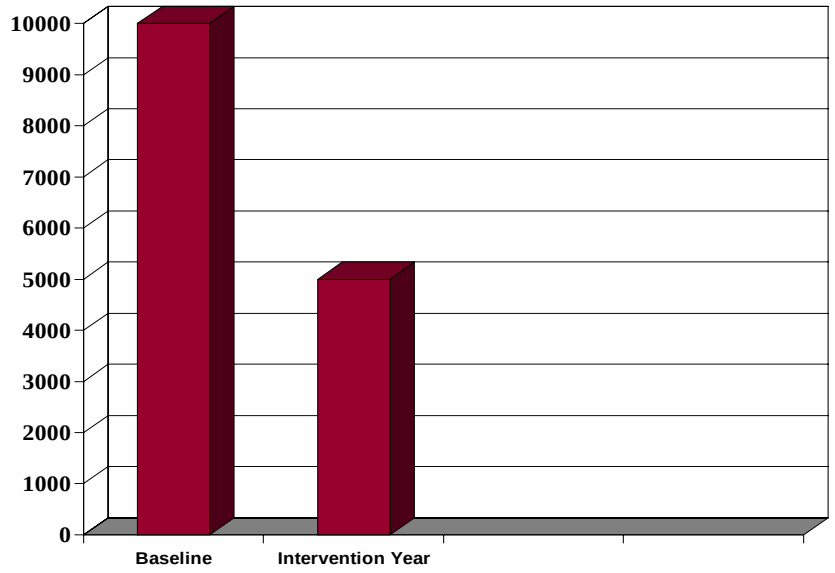
Covered in other preconference Session and at 2008 DMAA

- Plausibility Indicators
  - Specific test to ask "Did events for the disease in question fall noticeably?"
    - That would be TOTAL PLANWIDE EVENTS
  - These events have to fall by c. 20% just to break even (once again: Ariel Linden has confirmed this in citation at end of presentation)

Pre-post Analysis with a "plausibility indicator" test

- Example: Babies
- Suppose you want to reduce your plan's birth rate (now 10,000 babies a year) by instituting free contraception and family planning
- For a pre-post analysis, to find eligibles, you take everyone with a claim for giving birth during the last two years
  - That is the cohort with which you are working

Births in your 2-Year Baseline Cohort: Pre-post analysis



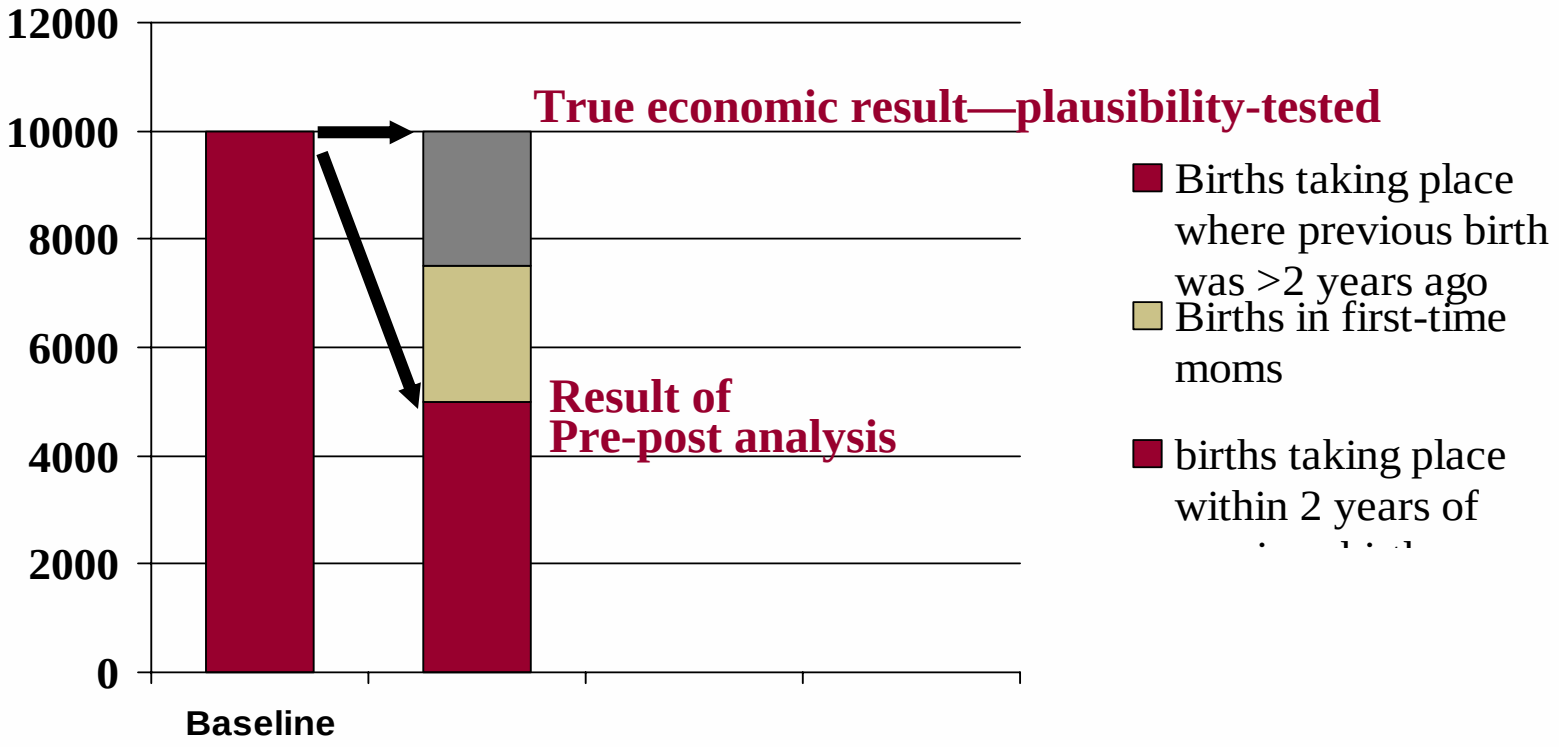
Would you say:

"We achieved a 50% reduction in births and costs of birth through our contraception and family planning programs" ?

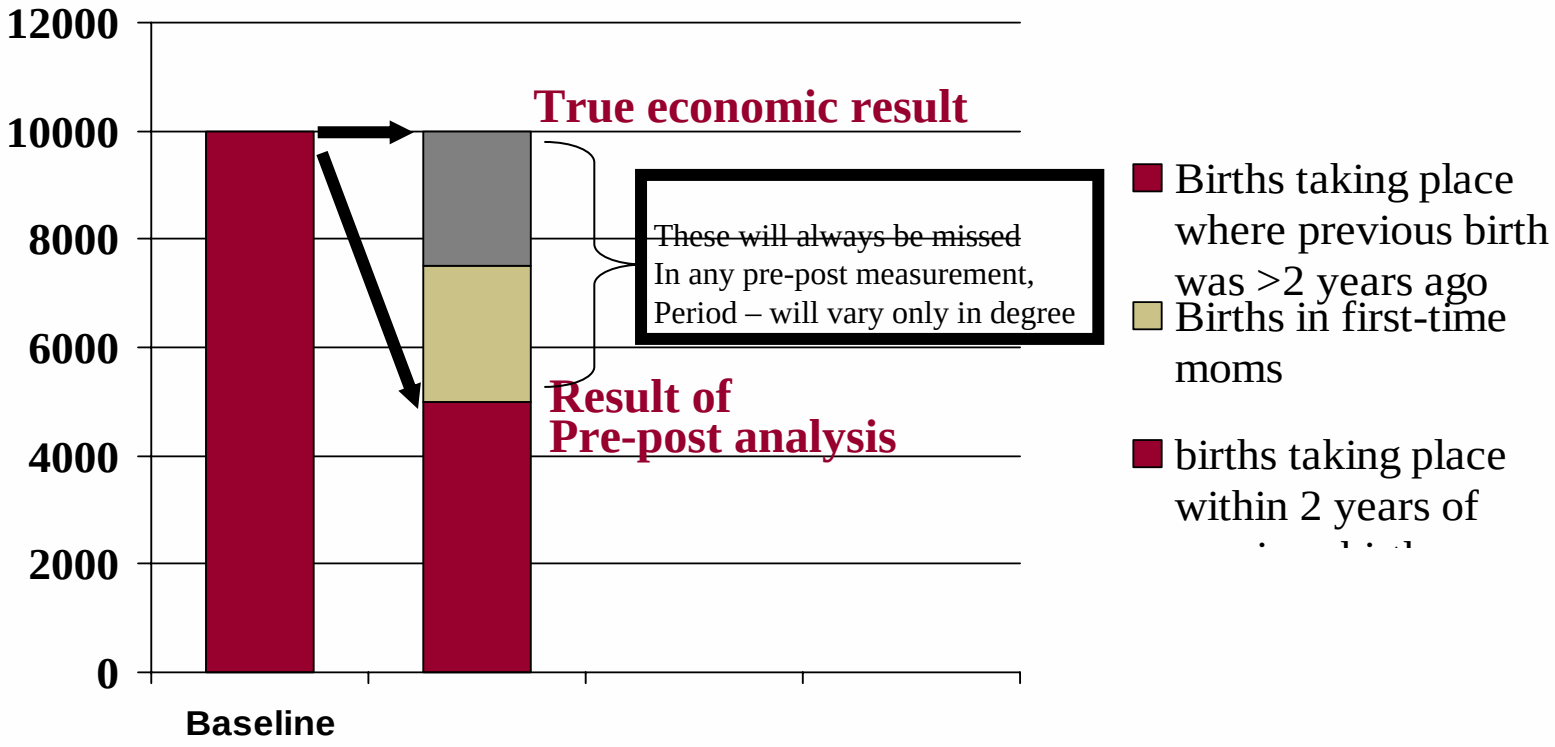
Of course not. You would say:

- "This is absurd…you would never just measure births in a cohort. You'd measure in the entire plan."
  - Measuring the entire plan is an event-based **plausibility** analysis to check the pre-post, as in this example
    - people with 0 previous findable claims are excluded from the baseline. Plausibility indicators find that

Total planwide events vs. pre-post

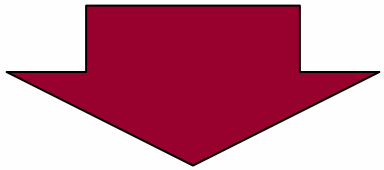


Total planwide events vs. pre-post



Babies vs. chronic disease

- "This is absurd…you would never just measure births in a cohort. You'd measure in the entire plan."



But this is precisely what you do when you measure **pre-post** for chronic disease and then track your performance vs. the baseline. Let's use a hypothetical from a chronic disease and show how
Pre-post must be confirmed by event rate plausibility and then go to some data

Let's test this dramatic cost savings…

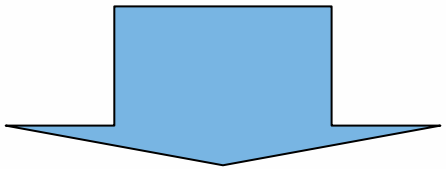
	2004 (baseline)	2005 (contract)
Asthmatic #1	1000	**100**
Asthmatic #2	0	**1000**
Cost/asthmatic	**$1000**	**$550**

…against the "event-based plausibility indicator" of total primary asthma IP codes

	2004 (baseline)	**2005** (contract)
Asthmatic #1	1000	***100***
Asthmatic #2	0	***1000***
Plausibility: Did Number of IP codes decline?	1	*1*

Results of Plausibility Analysis

- No change in # of asthma IP events planwide
  - You can't reduce asthma spending without reducing asthma events
- Plausibility analysis fails to support pre-post
  - Therefore pre-post result is invalid



This is probably the only methodology which
Produces valid measurement in long programs

Cost savings was not plausible

- Plausibility indicators are the TOTAL number of primary-coded IP events / TOTAL number of people in the plan

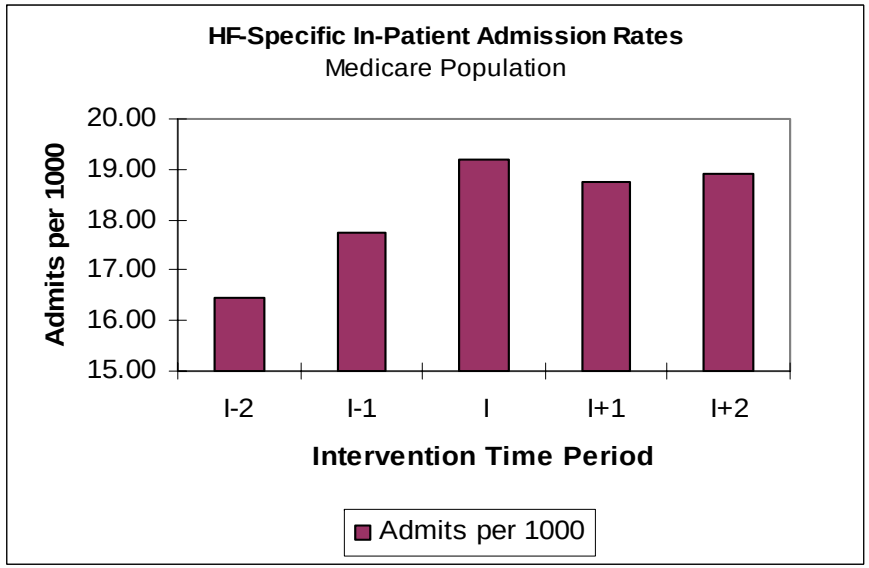
- They need to go in the same direction (down) as the spending to confirm the savings
- It didn't
  - Complete list of ICD9s by disease available free from DMPC (diseasmgmt@aol.com)

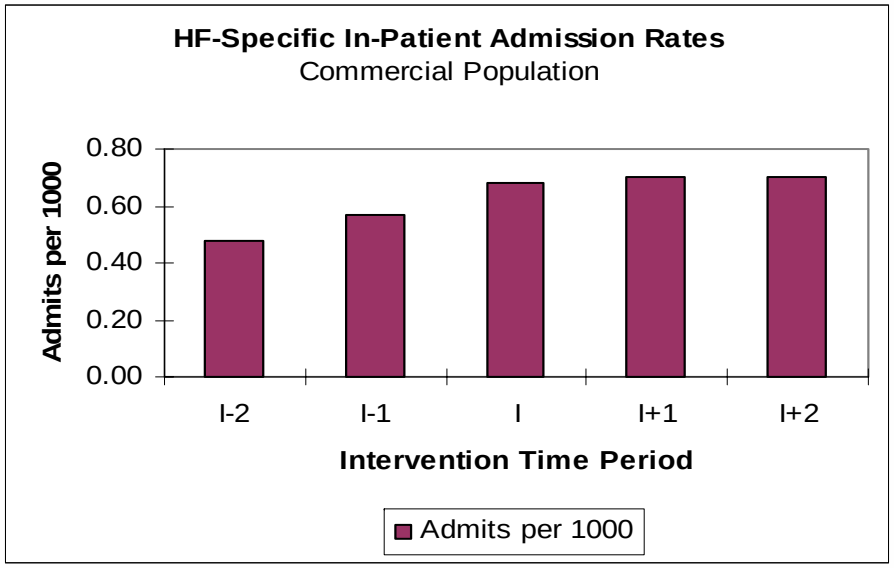
Several Examples of Plausibility Analysis

- Pacificare
- IBM
- Some which didn't turn out so well
- Plausibility-testing generally and benchmarks

PacifiCare HF Results by Alere Medical

Condition-specific utilization rates for insured population reinforces plausibility of sustained results over 3 years*





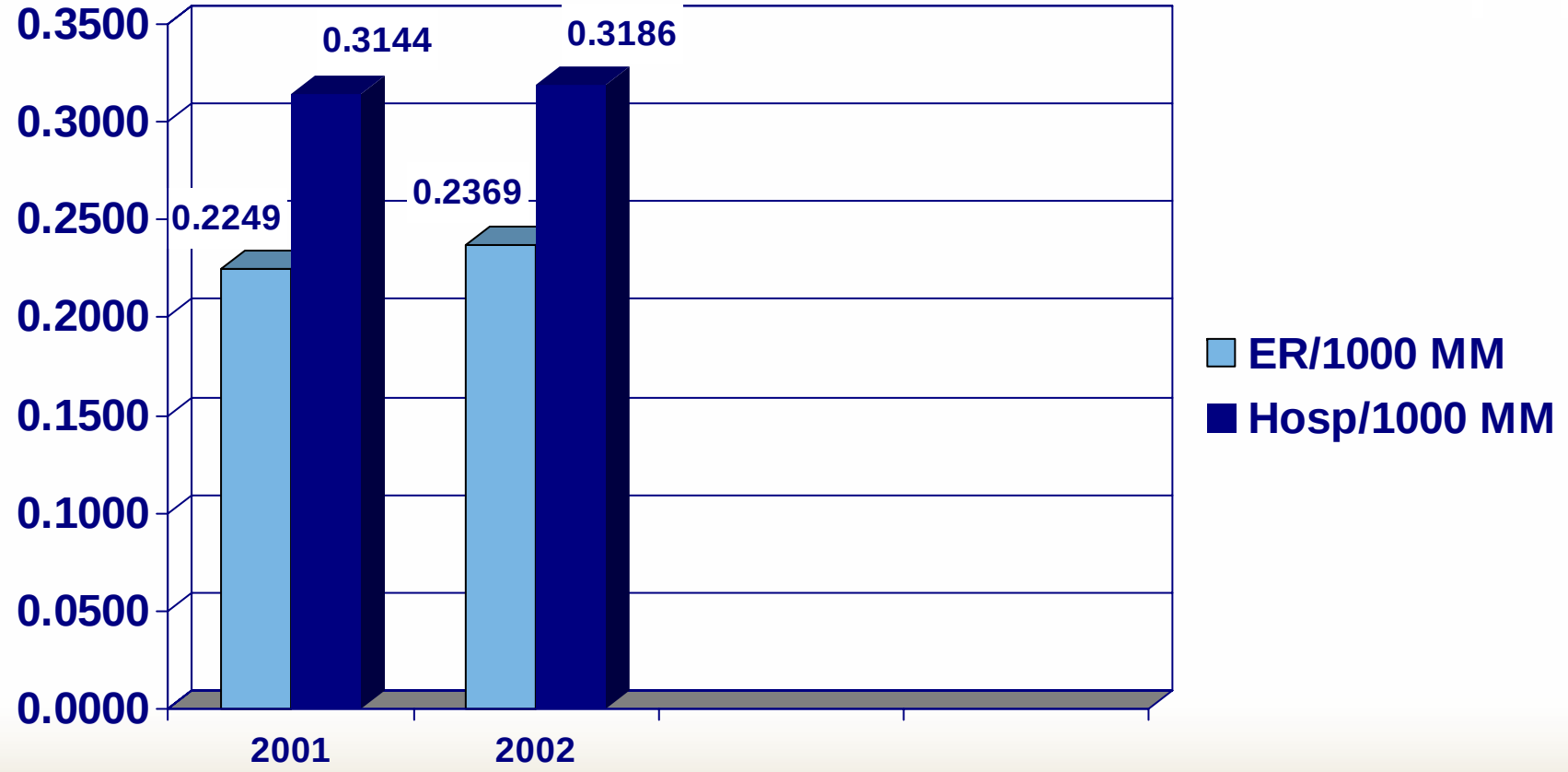
"I" = First DM Implementation Year

* National and local data show no decline in utilization trend for HF admissions; in fact, rising HF prevalence drives rising trend in absence of DM even with other MM in place

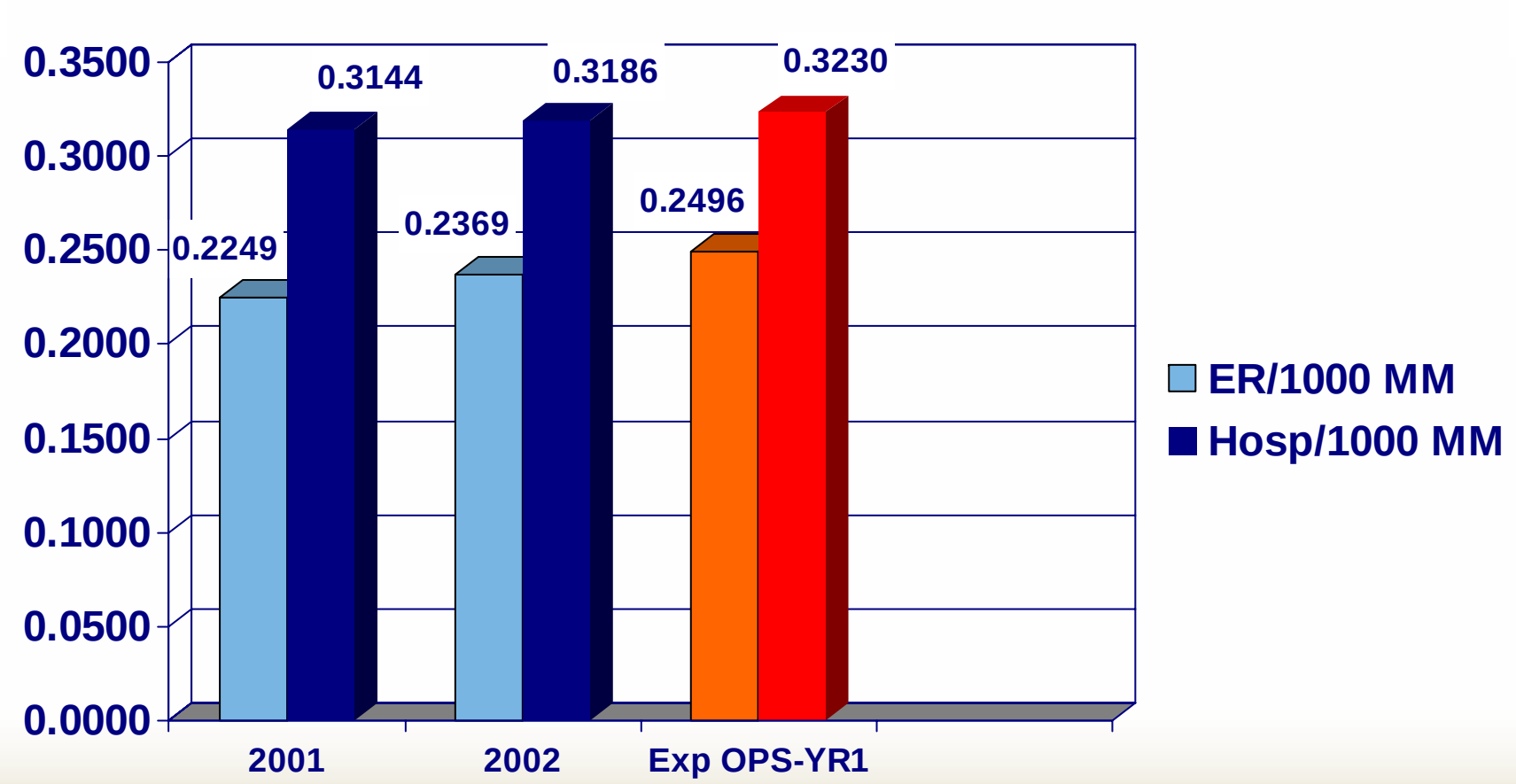
IBM examples

- Two years of trend to establish baseline
- “Expected” based on trend (in red)
- Actual experience in study year

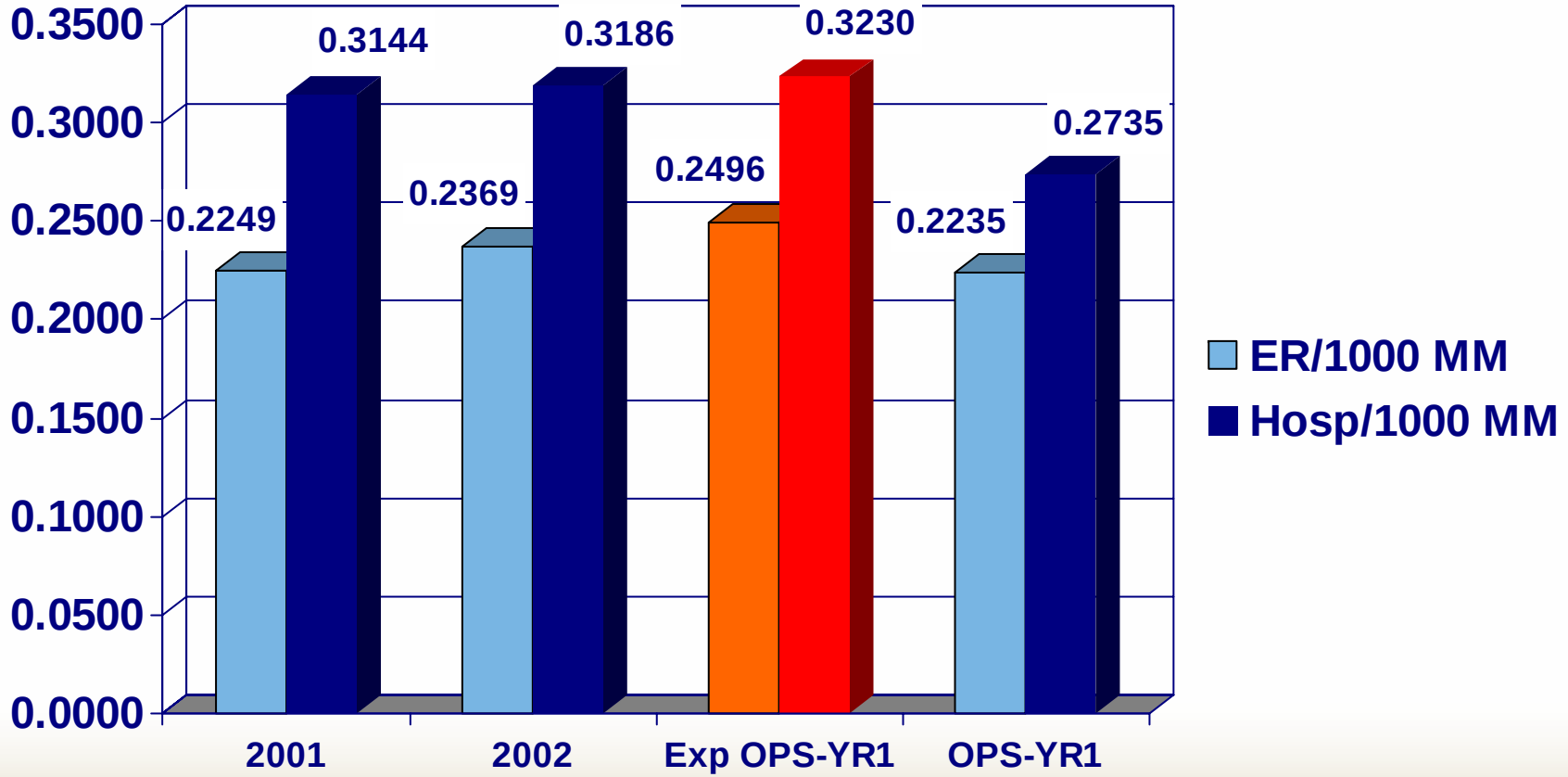
IBM -- Diabetes
Hosp Admits ICD-9 '250.xx' per 1000 MM



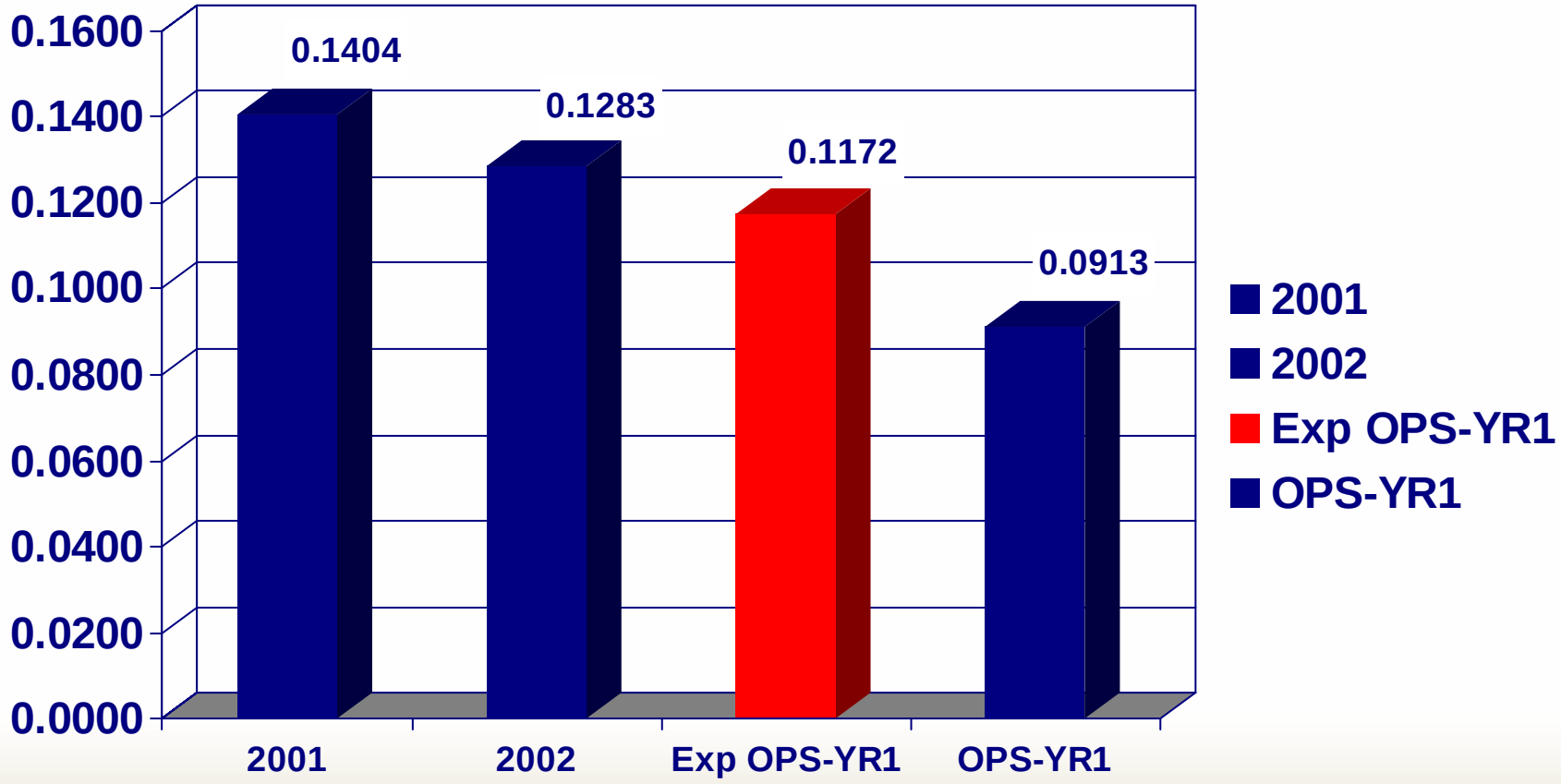
IBM -- Diabetes
Hosp Admits ICD-9 '250.xx' per 1000 MM



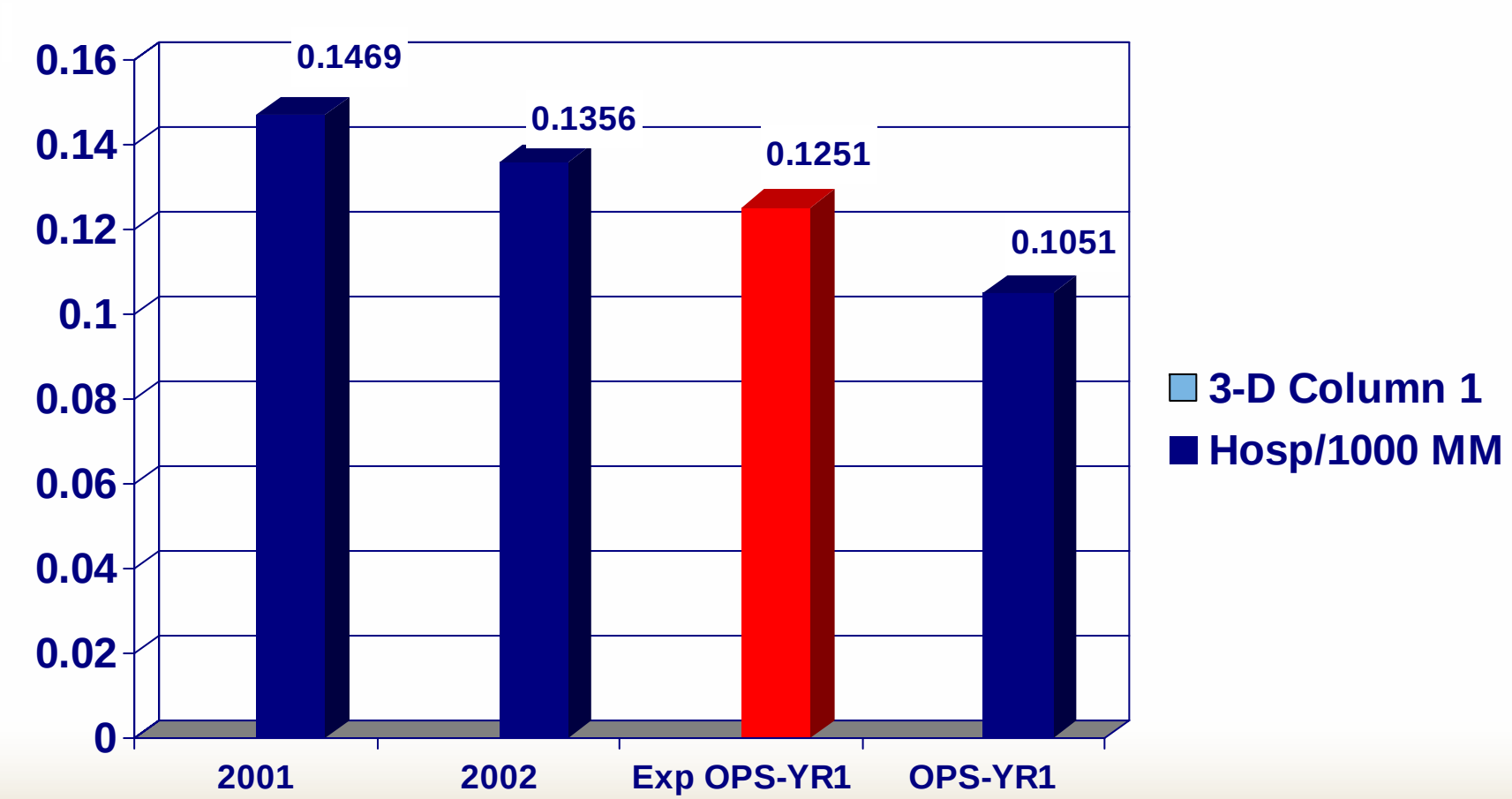
IBM -- Diabetes
Hosp Admits ICD-9 '250.xx' per 1000 MM



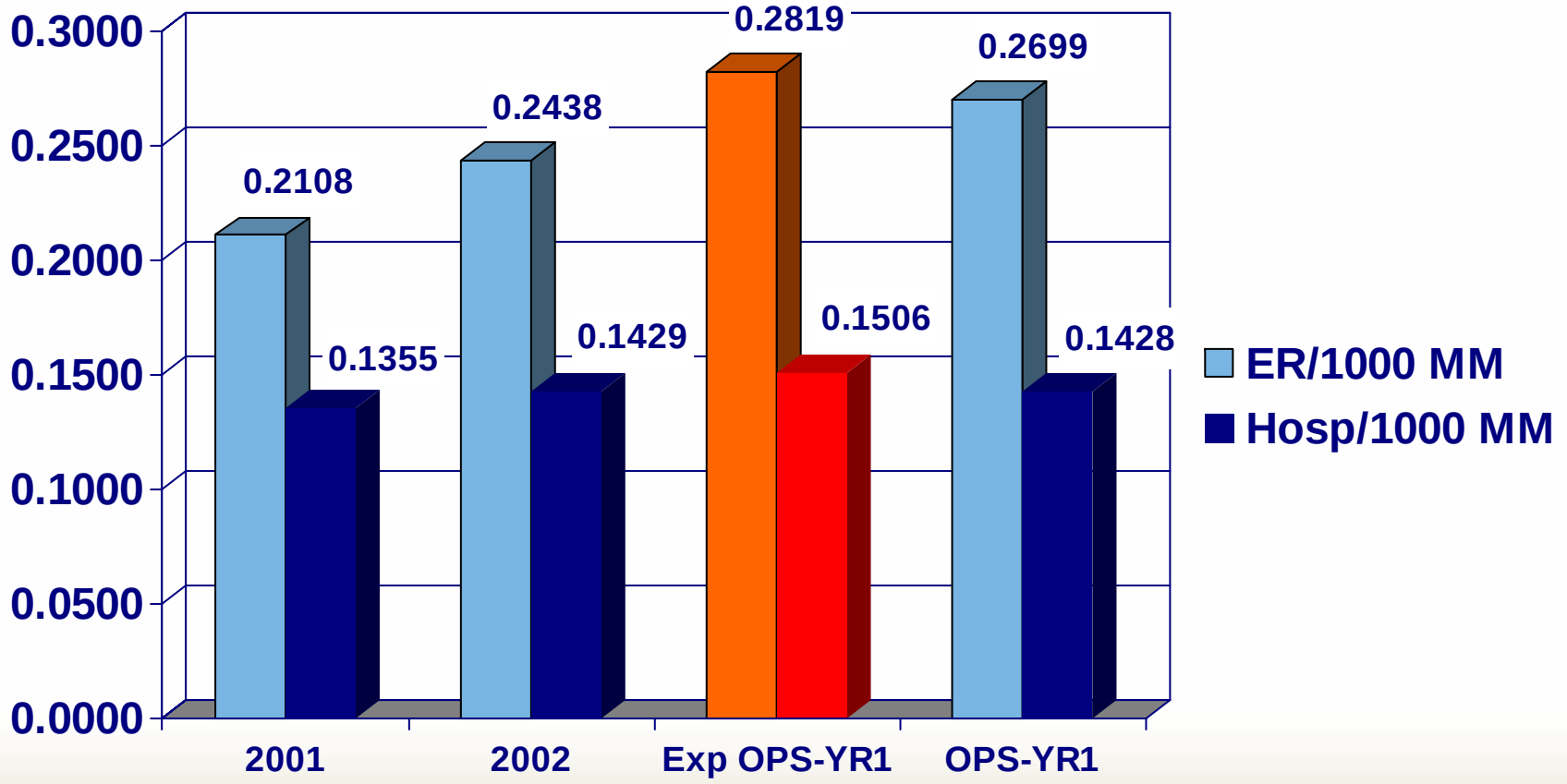
Angina Pectoris
Hosp Admits ICD-9 '413.xx' per 1000 MM



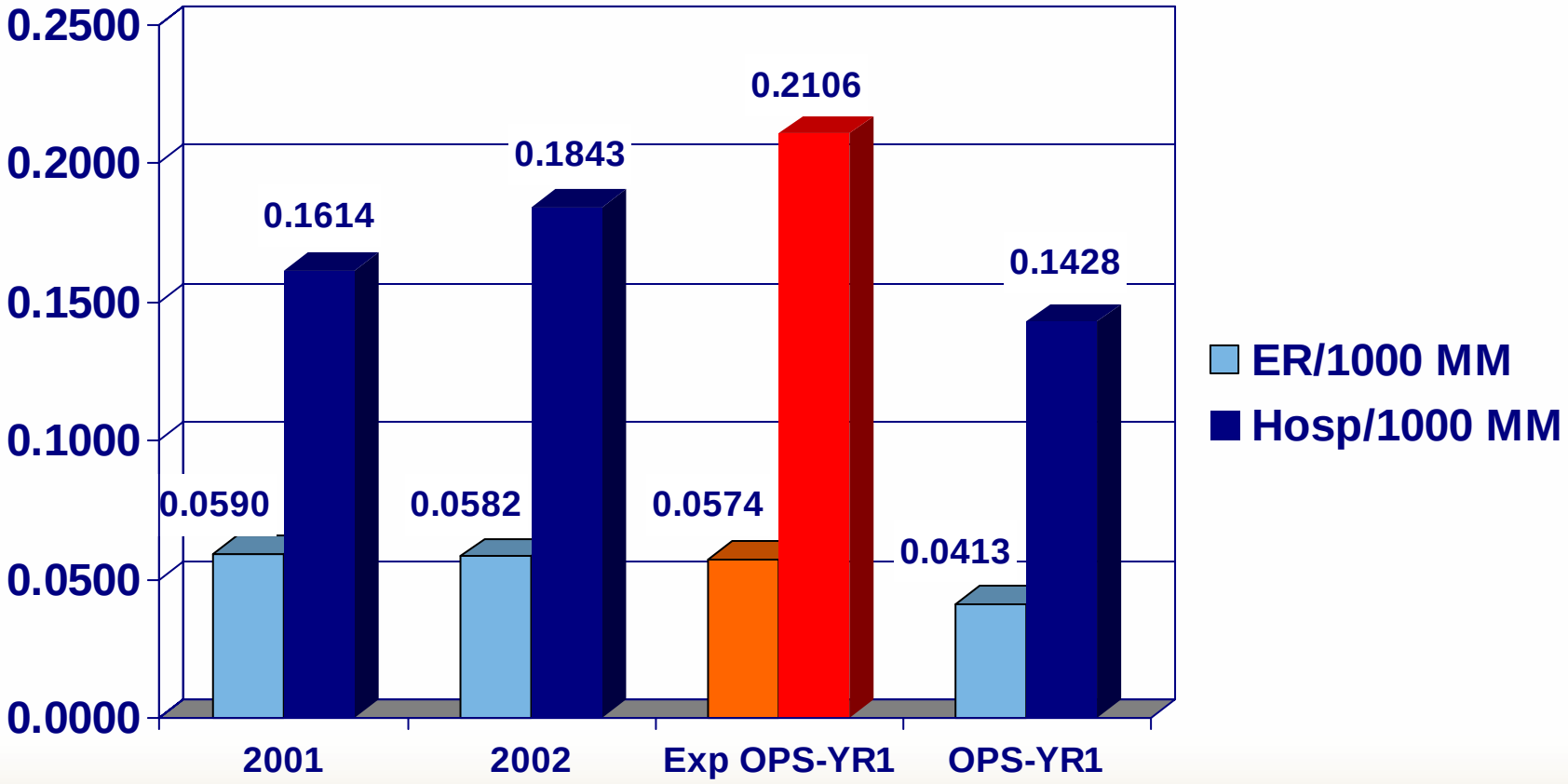
Acute MI
Hosp Admits ICD-9 '410.xx' per 1000 MM



Asthma
Hosp Admits ICD-9 '493.xx' per 1000 MM



Heart Failure
Hosp Admits ICD-9 '428.xx' per 1000 MM

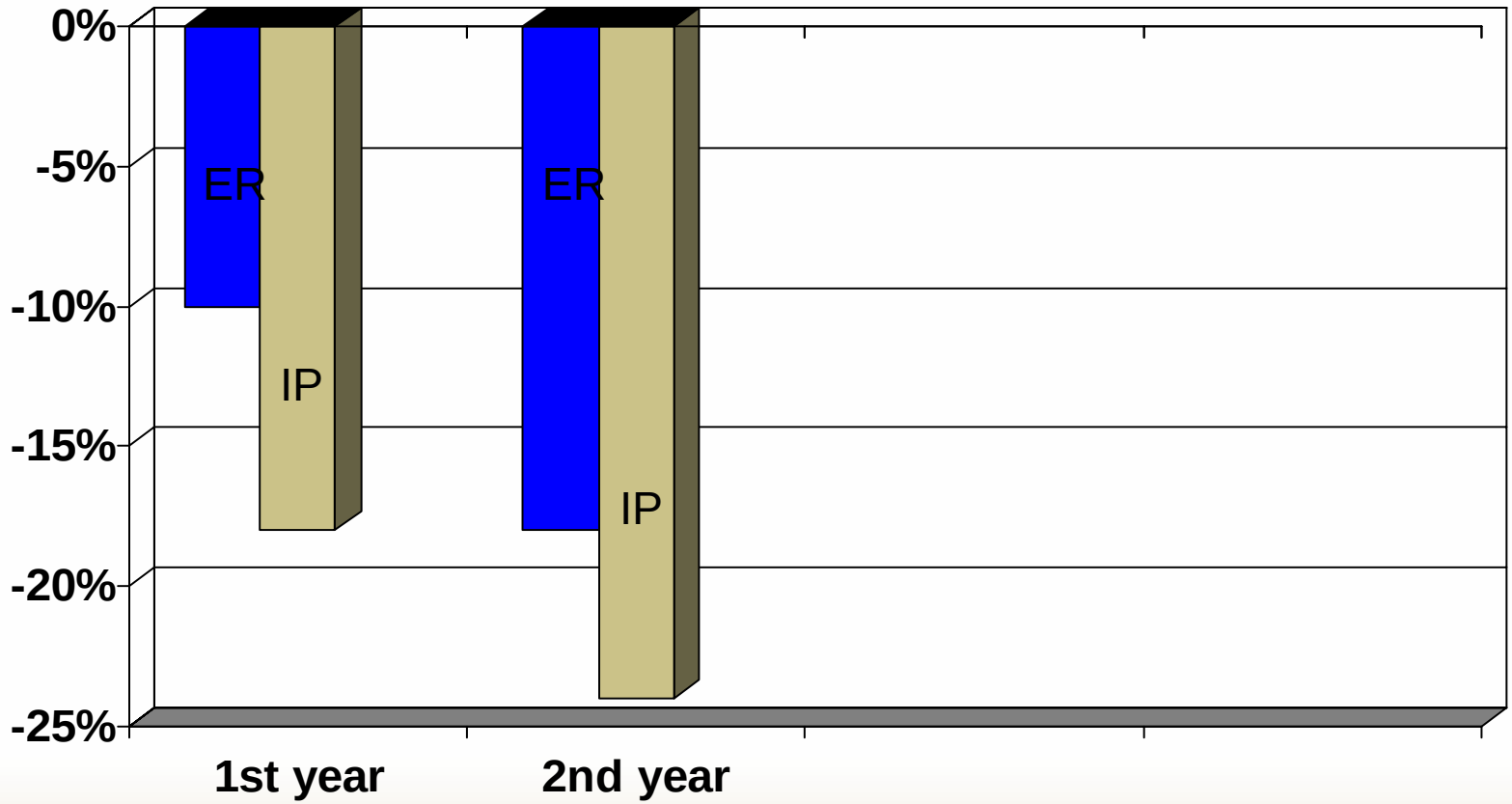


Where are the claims from *previously undiagnosed* asthmatics?

- IRVING, Texas--(BUSINESS WIRE)--Nov. 18, 2003--A pediatric asthma disease management program offered by AdvancePCS 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

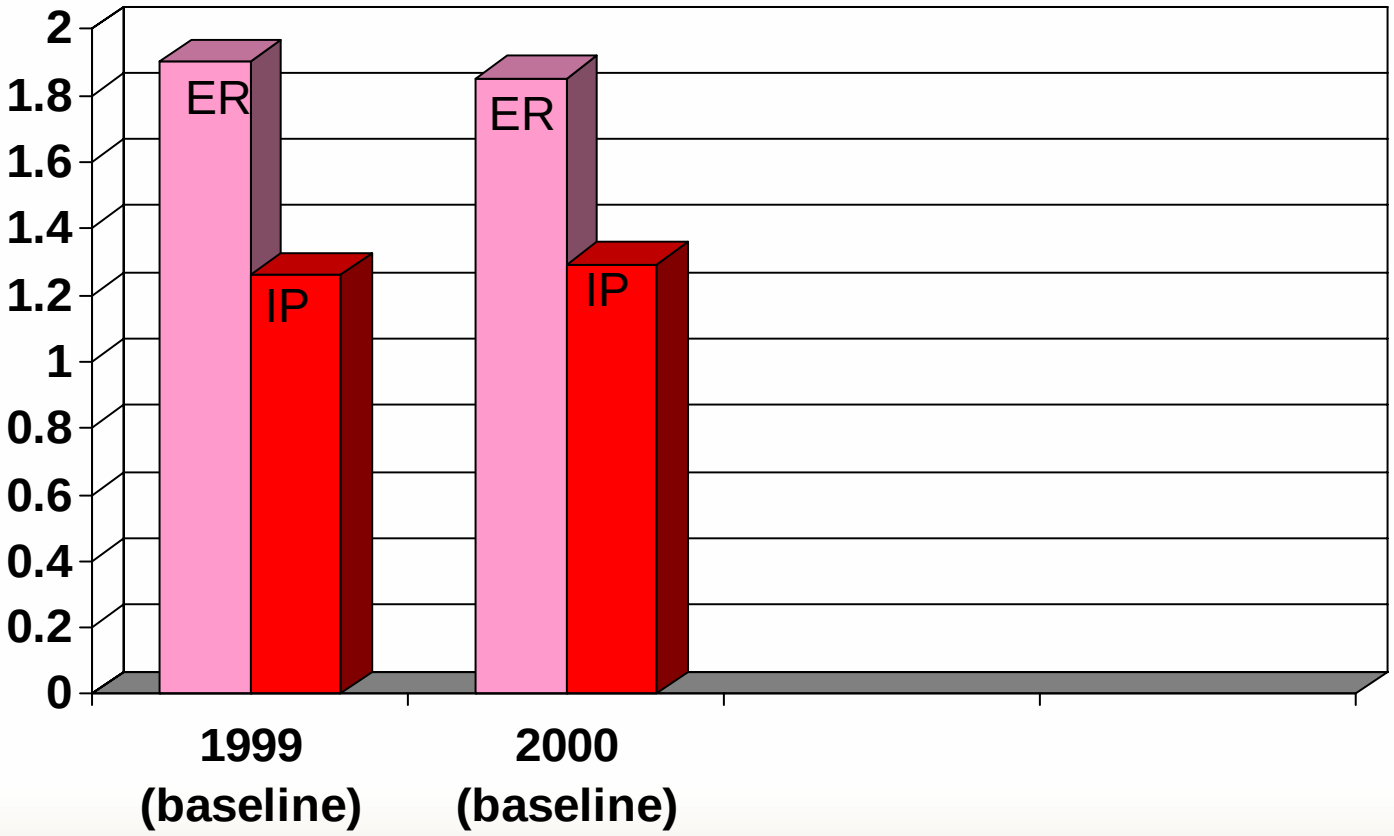


What we did…

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

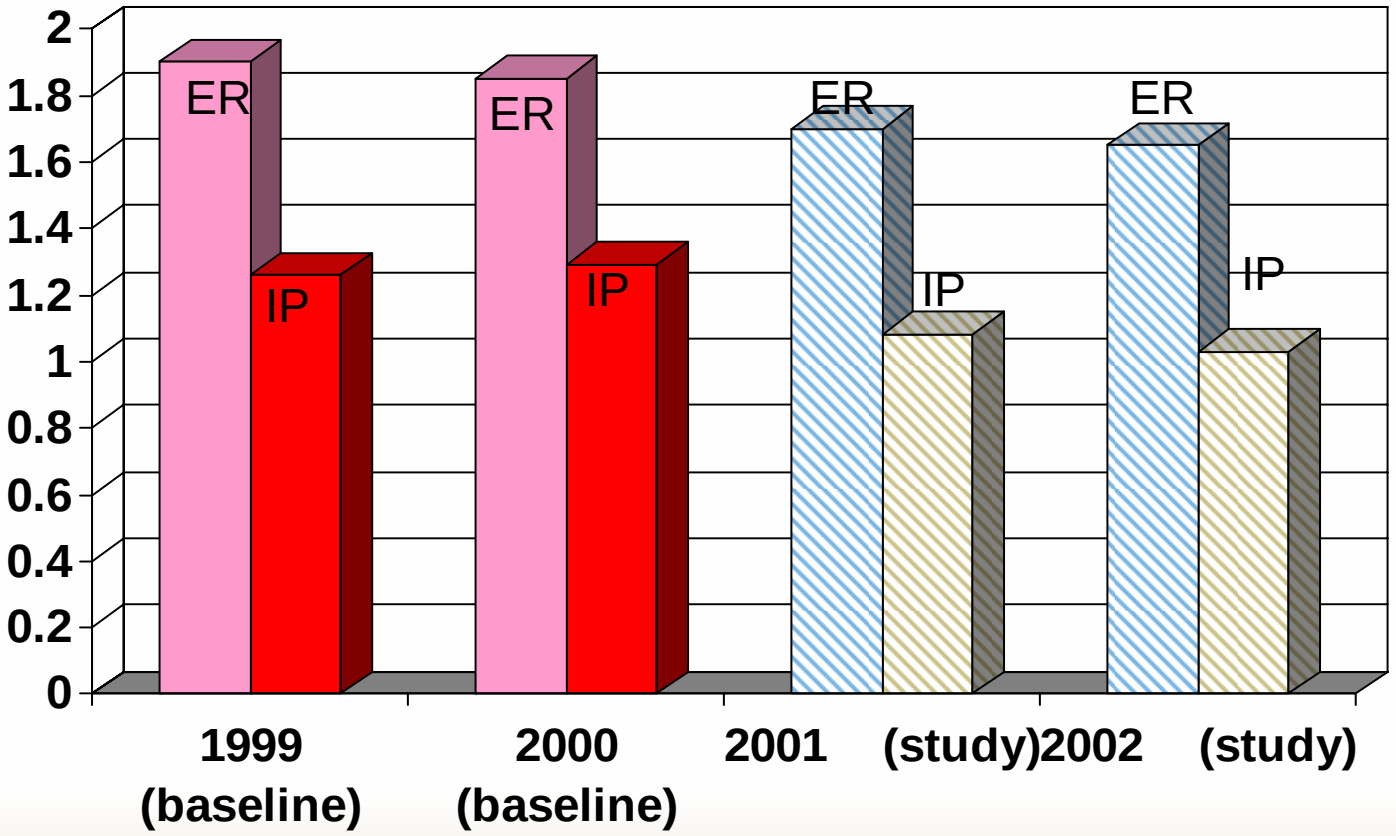
Baseline trend for asthma ER and IP Utilization

493.xx ER visits and IP stays/1000 planwide



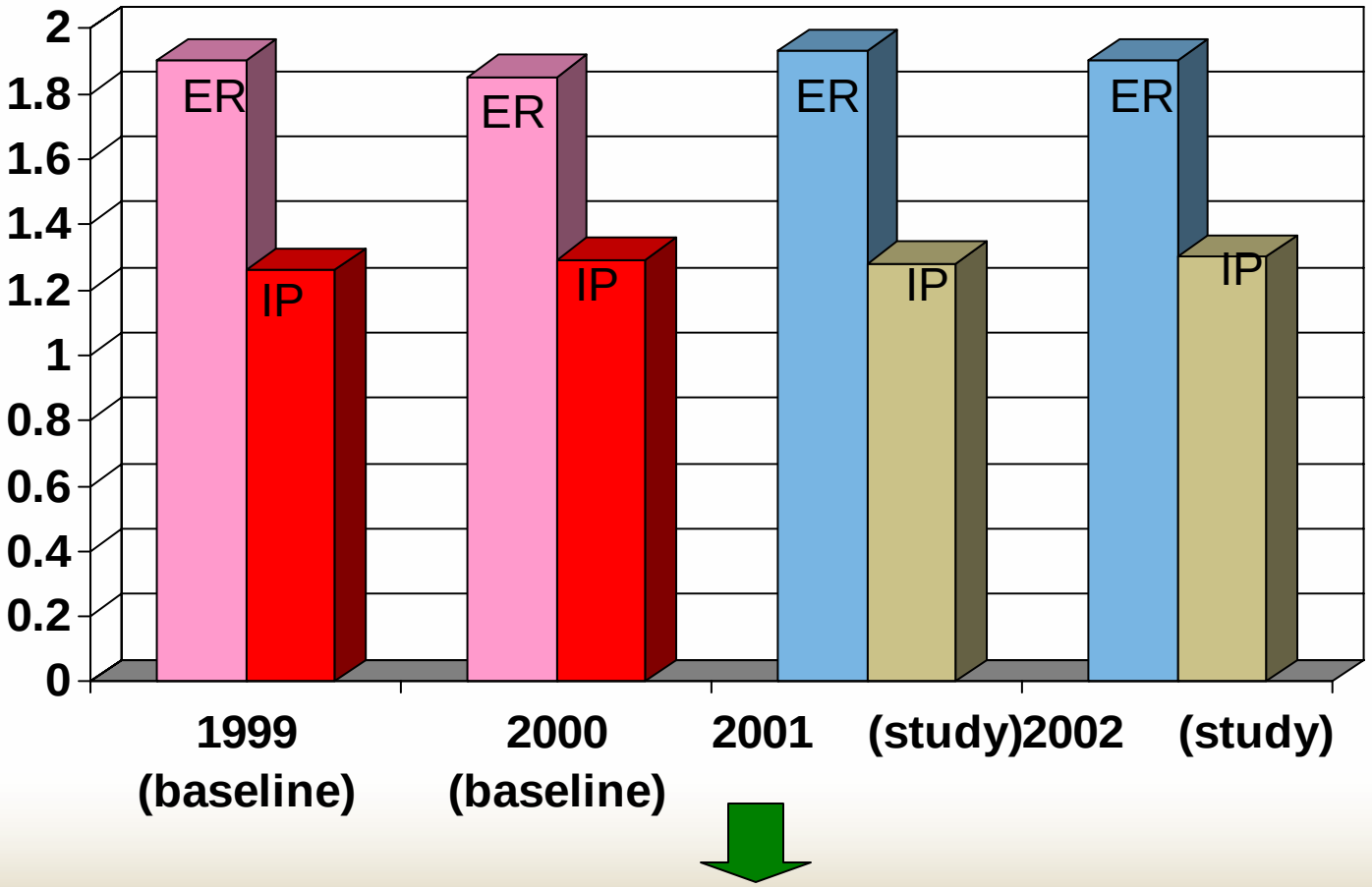
Expectation is something like…

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

493.xx ER visits and IP stays/1000 planwide

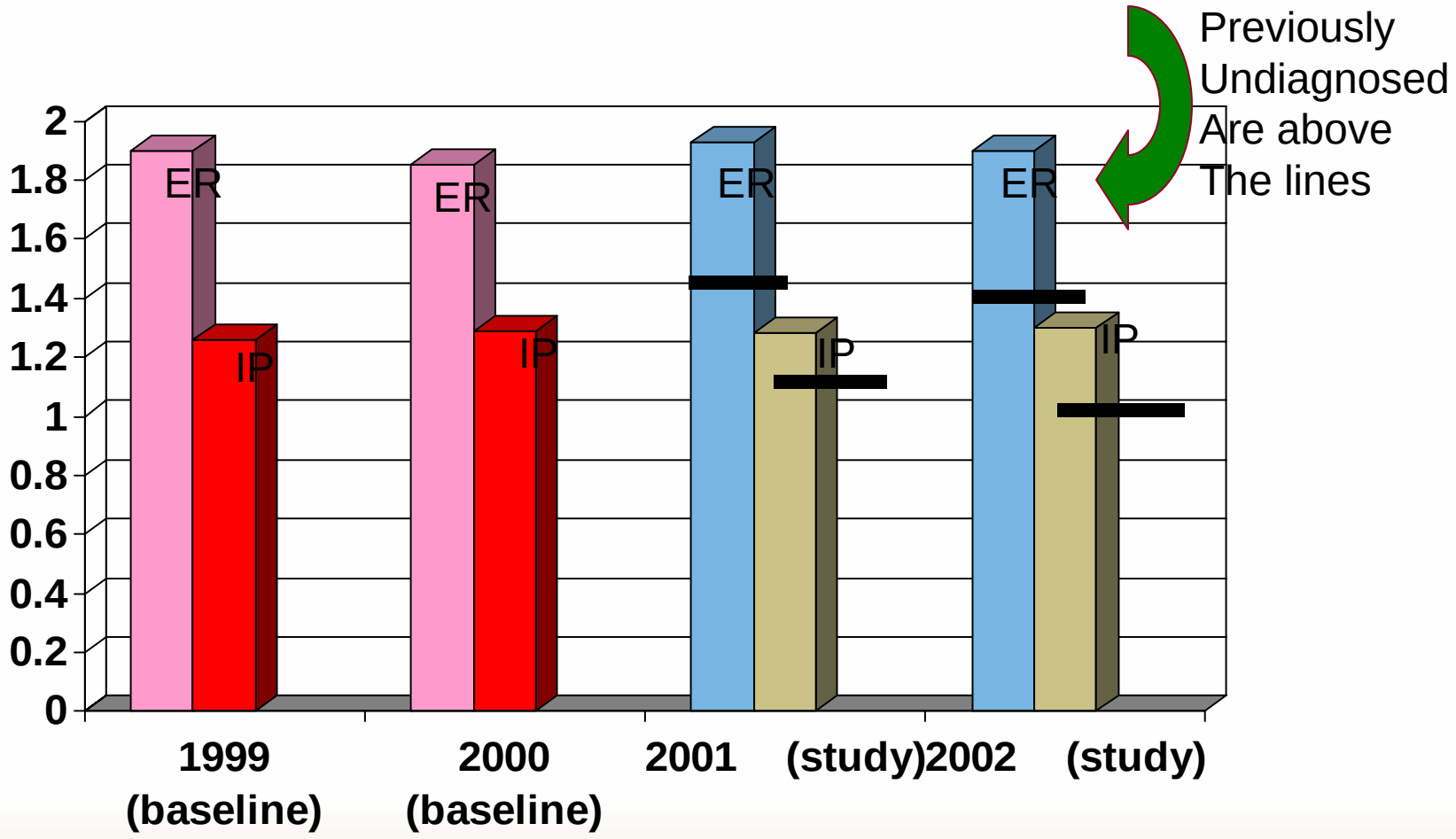


How could the vendor's methodology have been so far off?

We then went back and looked…

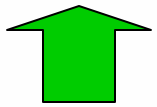
- …at *which* claims the vendor included in the analysis…

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



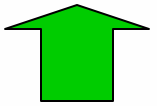
Possible impact of testing and then validating with plausibility

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



Impact

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



- Credibility of ROI from DM: Priceless

Bibliography – info from the leading academic researcher to support this presentation

Regression to the mean citation is:

- Linden A. Estimating the effect of regression to the mean in health management programs. *Dis Manage and Healt Outc*. 2007;15(1):7-12.

The citation for the plausibility indicators is:

- Linden A. Use of the total population approach to measure U.S. disease management industry's cost savings: issues and implications. *Dis Manage and Healt Outc*. 2007;15(1):13-18.

Questions sent in advance (from earlier registrants)

- Do I support the DMAA guidelines?
- Can you give specifics on applying the plausibility test? Is it every event in the disease-eligible population
- Do we need to be a DMPC member and use plausibility to get the DMPC Certification for Savings Measurement?
- Do vendors support plausibility-testing?
- Does the DMAA support plausibility-testing?

Disease Management Outcomes Conclusion

Yes, Virginia, the DMAA guidelines DO work…if they pass the tests and are confirmed with plausibility

DMPC
Disease Management
Purchasing Consortium Advisory Council