

VALID AND TRANSPARENT OUTCOMES MEASURES

How to tell if your programs are working

DMPC
Disease Management
Purchasing Consortium Advisory Council

Hey, Butch, Who Are These Guys (www.dismgmt.com and Al Lewis)

- #1-ranked by Managed Healthcare Executive “most influential” rankings (2004-2008)
- Two-time winner “Most Influential” DMAA
- *New York Times*, *Wall Street Journal* quotes
- Frequent Op-Ed Contributor, *San Francisco Chronicle*
- Most frequent healthcare guest on Montel Williams
- NPR, NECN, many local radio/TV
- “Invented Disease Management” according to Google
- “Wrote the book” on disease management, available from AIS Publications (Jgutman@aispub.com)

Agenda

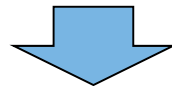
- How it is easy to make mistakes in a pre-post measurement
- How to develop event-rate driven “plausibility indicators” to benchmark your health plan’s success

Mistake #1: Why measuring on the “whole population” is not really measuring on the whole population

- This is called a “heads or tails” mistake
- Imagine everyone with a condition is a coin. “Heads” means the health plans KNOWS they have the condition. “Tails” means they don’t.

Why might a health plan not know a member has a condition?

1. Member is new to the plan
2. Member is too mild to have any claims
3. Member has claims, but not enough to trigger the algorithm (for instance, you need two 250.xx MD visits to be classified as diabetic)
4. Member is non-compliant and doesn't fill scripts
5. Member is misdiagnosed
6. Member is correctly diagnosed but the physician doesn't want to enter it in their file
7. Member does not himself or herself know he/she has the condition.



Clearly there are a lot of tails

The effect of tails on measuring savings

- Vendors (and internal DM programs for health plans) only measure the “whole population” they know about. They “flip” the heads and take credit for the 50%+ reduction in heads.
 - If it’s a 60% reduction (for example) they should only take credit for 10%
- But they don’t flip the tails and offset the 50% reduction in heads-turning-tails with the tails-turning-to-heads
- They do indeed count the heads...*in the next year’s numbers once they are already heads*



Let’s do a hypothetical and then a real example

And then fix it

© 2007-09

In this pre-post example

- Assume that “trend” is already taken into account (“trend” itself – what “would have happened absent the program” -- is also highly controversial)
- Focus on the baseline and contract period comparison

Base Case: Example from Asthma

First asthmatic has a \$1000 IP claim in 2007 (“Heads” while #2 is a tails)

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

What is the cost/asthmatic in the baseline?

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

**The cost/asthmatic in the baseline is \$1000
because only the heads are measured**

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

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 –a tail

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	\$1000	

Cost/asthmatic in contract period?

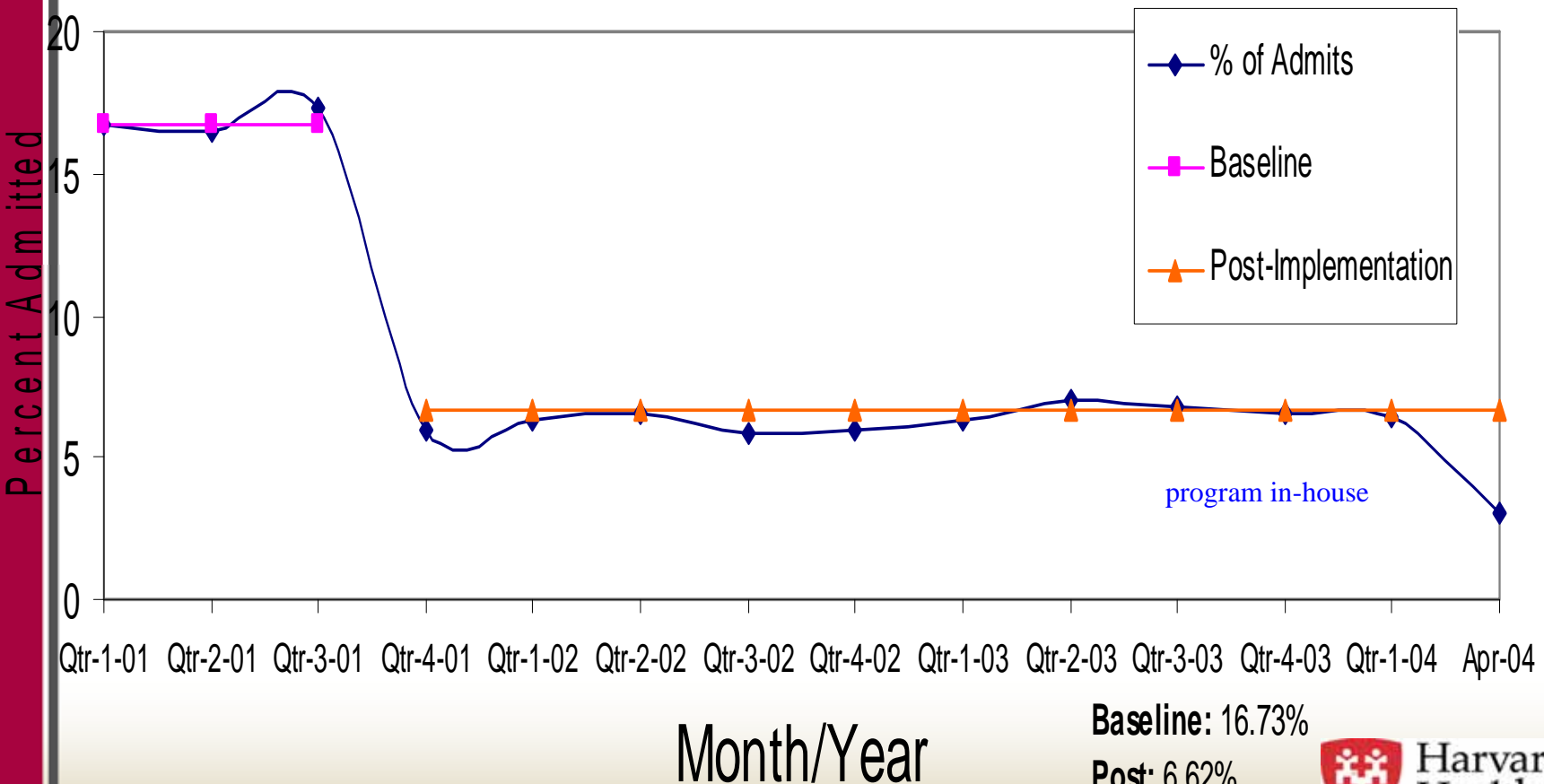
	2007 (baseline)	2008 (contract)
Asthmatic #1	1000	100
Asthmatic #2	0	1000
Cost/asthmatic	\$1000	\$550 (the increase of \$1000 in the second asthmatic's cost since 2007 didn't count – the tails turning to heads)

The asthmatic #2 was only counted –going forward -- once he flipped to heads

- It doesn't matter whether you use prospective identification or annual requalification
- Let's look at some real examples of this happening

Example #1: Measuring only the heads in the contract period yields huge decline in hospitalization rate

Health Advance Hospitalization Rate



Baseline: 16.73%

Post: 6.62%



Agenda: Now it's time to fix the problem

- Why it is easy to make mistakes using pre-post
- How to develop an event-rate driven set of “plausibility indicators” to benchmark your health plan's success

What is a plausibility test?

- You do it all the time...outside DM
- An easy way to directionally check results
- 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 next page
- Example from previous asthma hypothetical

Total event rates tracked by disease: 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 a plausibility test?

- You do it all the time...outside DM
- An easy way to directionally check results
- 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 next 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 “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. Plausibility test flunked
- Note that event rate measurement provides exactly the right answer (if drug classes are measured too)

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...

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 a plausibility benchmarking 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?**

What is a “failure point” ?

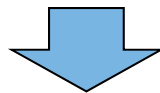
- Where people go to the ER/hospital with a preventable complication of chronic disease due to lack of control
- There is no need to finance DM or other programs to manage chronic disease members...*unless* they are out of control
 - Don't just “do disease management”
 - Instead, *focus* your efforts where they can avoid failures—people falling through the cracks and ending up in the ER/hospital with preventable complications and attacks
 - This is exactly what manufacturers do—focus improvement efforts where there are high defect rates

Key to Reading DM Plausibility Benchmarking slides

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

Benchmarking Example

- Harvard-Pilgrim Health Care (HPHC)
- National Averages
- Both in comparison
- Providence

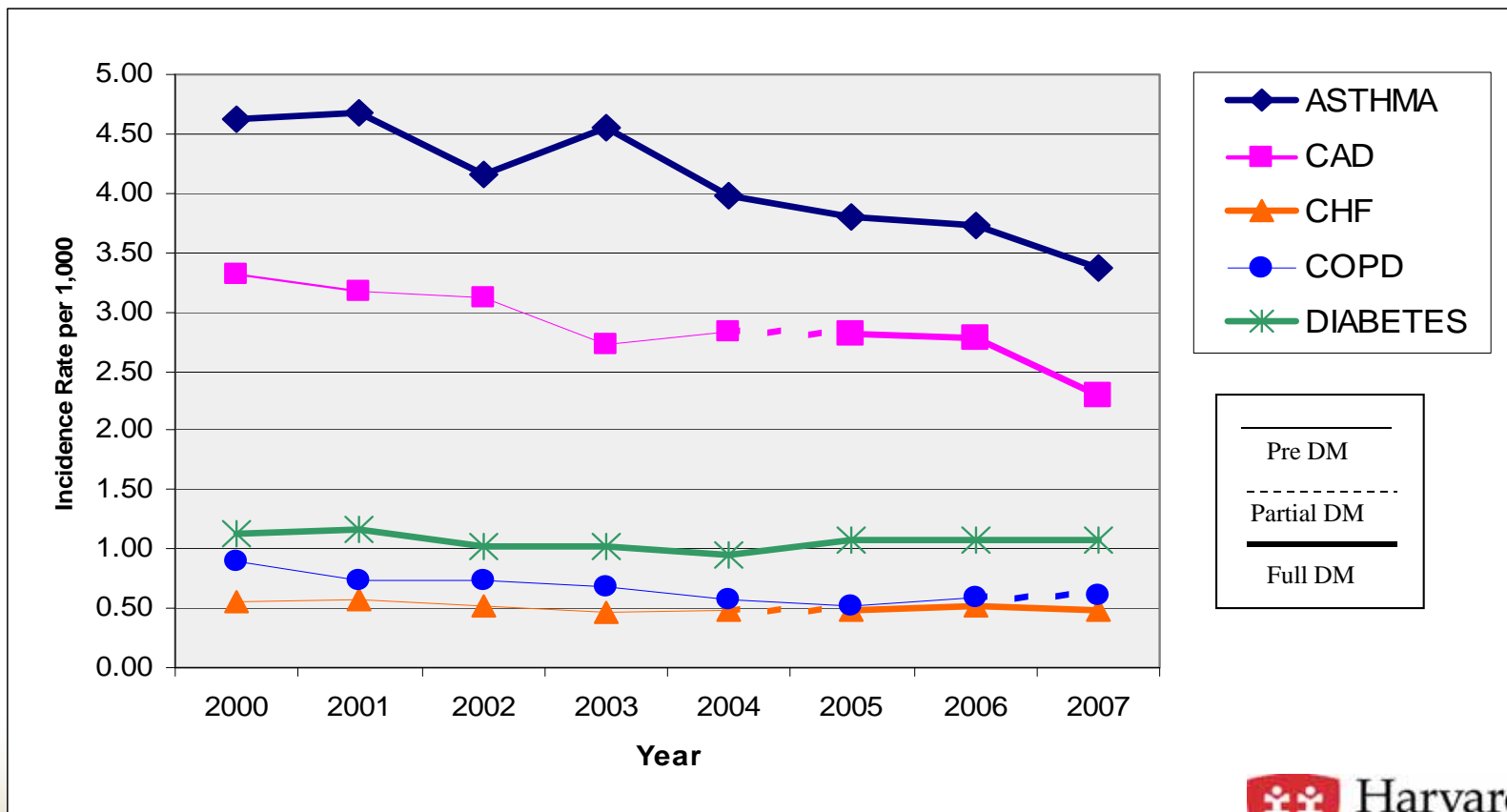


Providence and HPHC have the best outcomes in the country by many measures so don't automatically expect such good results though in general provider-owned and regional plans do better than nationals

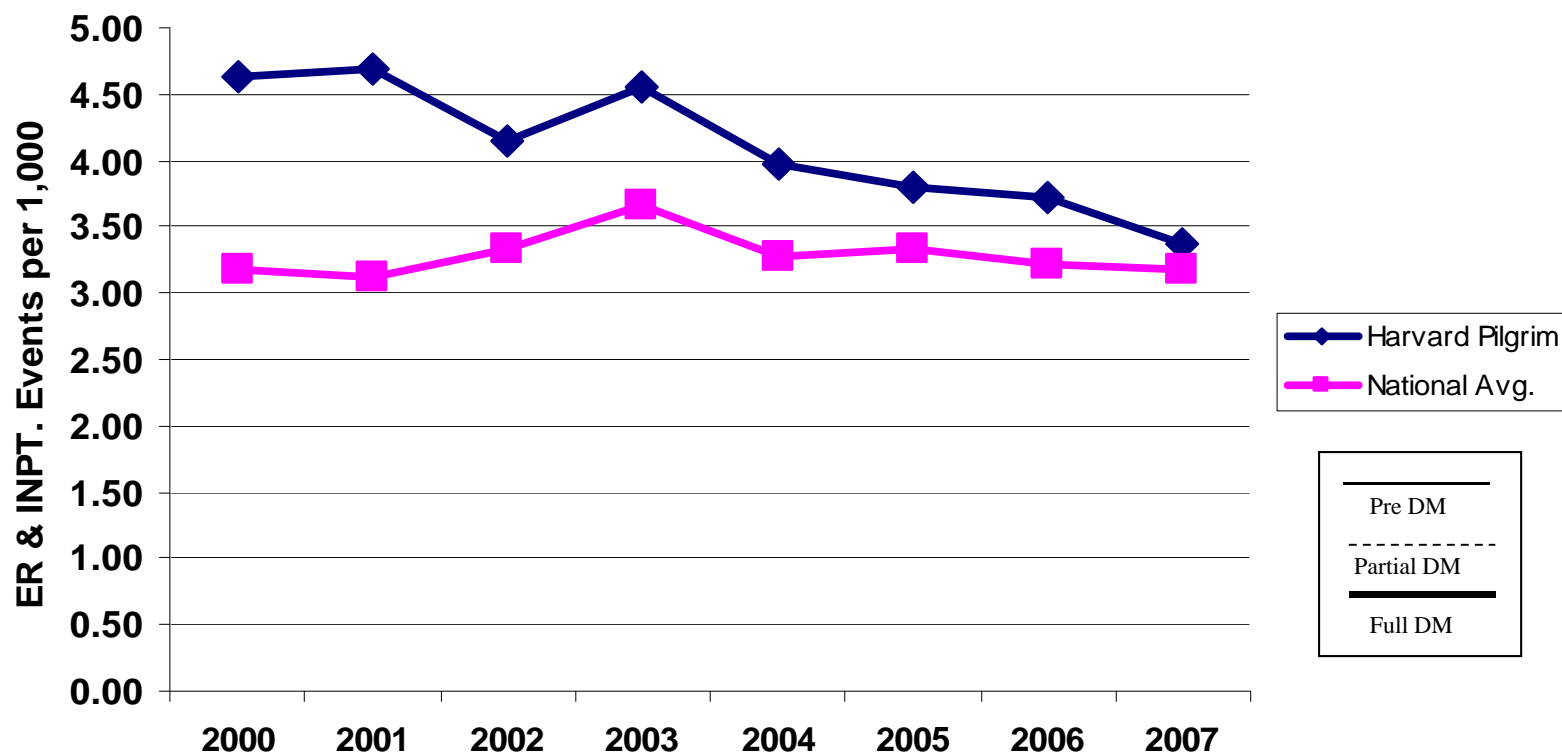
Key to Reading DM Benchmarking slides

- Harvard Pilgrim Disease Management
 - Thin lines are pre-program
 - Dotted lines are periods in which program was partially in place
 - Thick lines are program fully implemented
- National Average
 - Based on 25+ commercial health plans

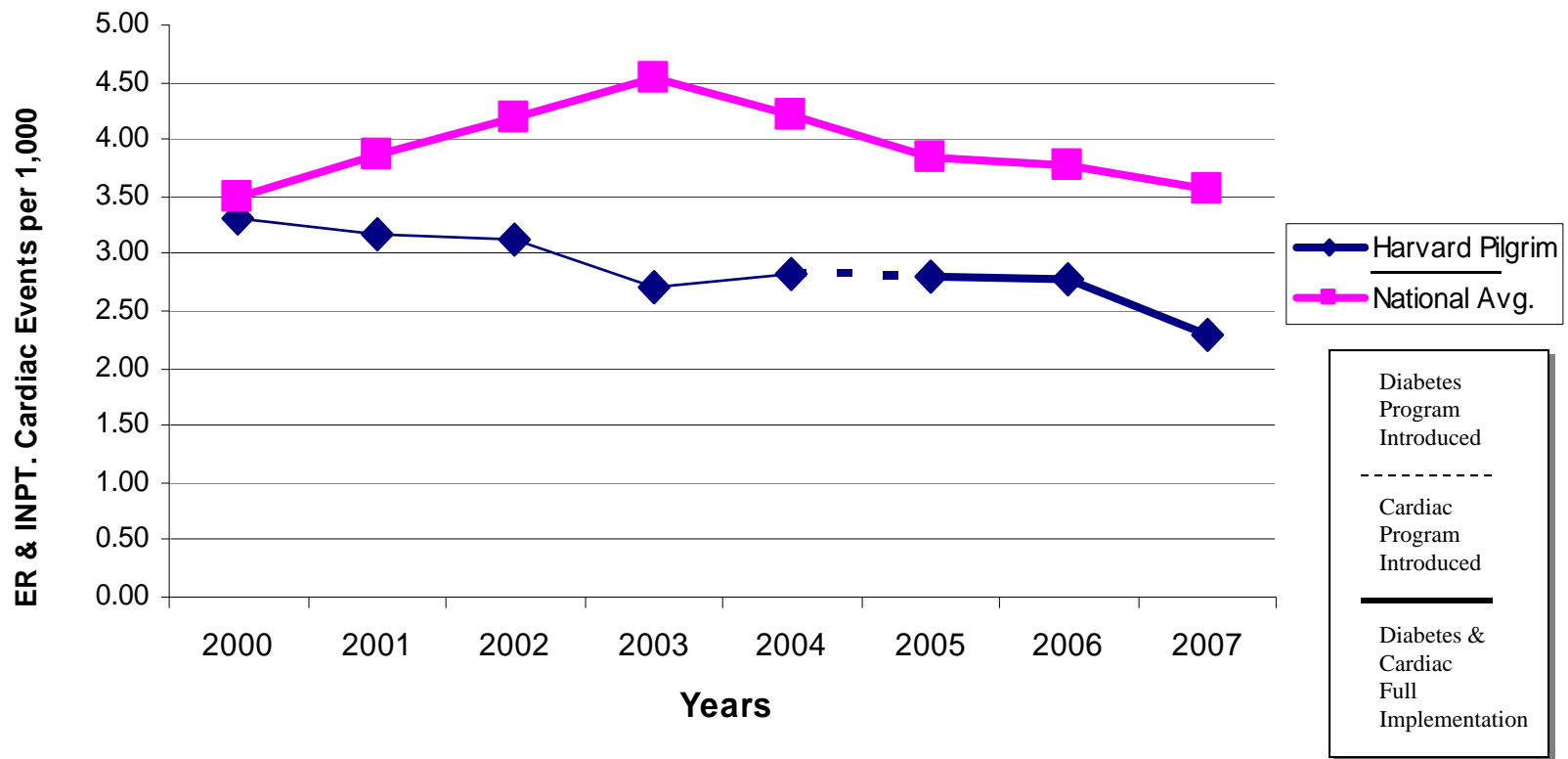
ER and Inpatient Events Per 1,000 Commercial Members All Chronic Conditions



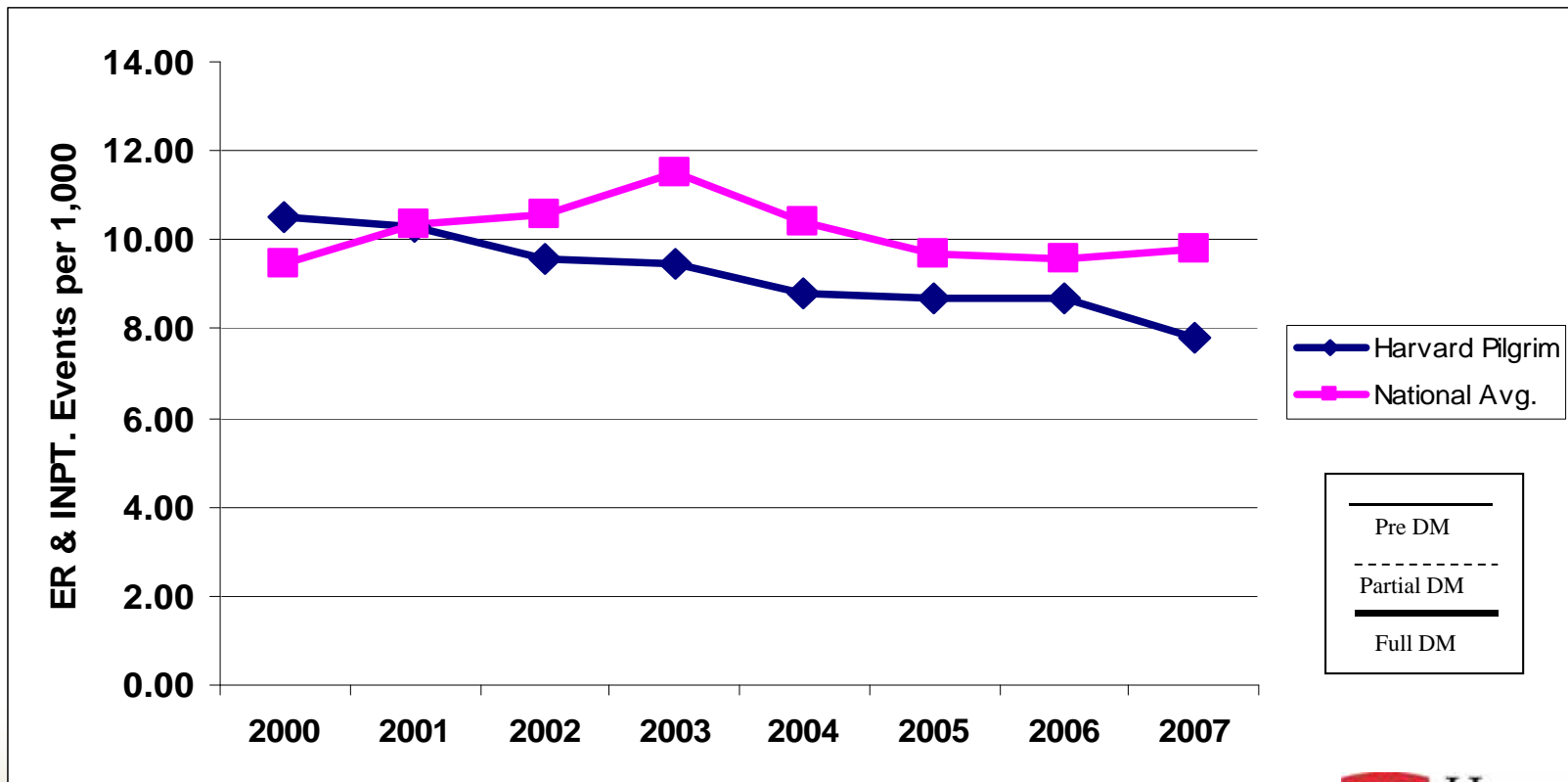
ER and Inpatient Events Per 1,000 Commercial Members Your Plan Compared To National Average Asthma



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



*ER and Inpatient Events Per 1,000 Commercial Members
Harvard Pilgrim Compared To National Average
Total Events For All 5 Common Chronic Conditions*



Other advantages of measuring event rates

- Requirement for DMPC Certification
- Requirement to win a HIRC award as a “best payor in disease management”
- Can be done much faster than actuarial methodologies
- Can be done with ingredients you already have in your kitchen
- And also...

Allows you to generate a valid ROI (10,000 or more covered lives)

- DMPC has a tool to do this, transparently and validly
- Beyond the scope and time allotment for this presentation
- Write or call 781 856 3962 to set up a conference call to do this with your own numbers
 - Will require tallying your event rates first (using our extraction algorithm, which we can also send)

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

