# Research Designs For Evaluating Disease Management Program Effectiveness

Disease Management Colloquium
June 27-30, 2004

**Ariel Linden, Dr.P.H., M.S.**
**President, Linden Consulting Group**

# What's the Plan?

- Discuss "threats to validity"
- Provide methods to reduce those threats using currently-used evaluation designs
- Offer additional designs that may be suitable alternatives or supplements to the current methods used to assess DM program  effectiveness

Measurement Error

Treatment Interference

Seasonality

Loss to Attrition

Hawthorne Effect

New Technology


Maturation

Benefit Design

Access

Reimbursement

Selection Bias

Unit Cost Increases

Case-mix

Regression to the Mean

Secular Trends

# Selection Bias

Definition: Participants are not representative of the population from which they were drawn:

- Motivation
- Severity or acuteness of symptoms
- Specifically targeted for enrollment

# Selection Bias (cont')

- Fix #1: Randomization
- How: Distributes the "Observable" and "Unobservable" variation equally between both groups
- Limitations: costly, difficult to implement, intent to treat, not always possible

# Selection Bias (cont')

- Pretest-posttest Control Group:

| | | | |
|---|---|---|---|
| R | $O_1$ | X | $O_2$ |
| R | $O_3$ | | $O_4$ |

- Solomon 4-Group Design:

| | | | |
|---|---|---|---|
| R | O | X | O |
| R | O | | O |
| R | | X | O |
| R | | | O |

# Selection Bias (cont')

- Fix #2: Standardized Rates
- How: Direct/indirect adjustment enables comparisons over time or across populations by weighting frequency of events
- Limitations: does not control for "unobservable" variation

# Age-adjusted Program Results

| Age Group | Pre-Program (rate/1000) | r X *P* | Program (rate/1000) | r X *P* | Proportion (*P*) of Population |
|---|---|---|---|---|---|
| 20 – 29 | 7.3 | 0.9 | 10.2 | 1.2 | 0.1189 |
| 30 – 39 | 65.2 | 5.7 | 79.9 | 6.9 | 0.0868 |
| 40 – 49 | 190.8 | 13.4 | 173.6 | 12.2 | 0.0703 |
| 50 – 59 | 277.9 | 21.3 | 226.1 | 17.4 | 0.0768 |
| 60 - 69 | 408.4 | 25.2 | 287.8 | 17.7 | 0.0616 |
| 70 - 79 | 475.8 | 17.7 | 368.8 | 13.8 | 0.0373 |
| 80 + | 422.2 | 8.4 | 356.0 | 7.0 | 0.0198 |
| Adjusted rate | | 92.6 | | 76.2 | |

# Tenure-adjusted Program Results

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

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

# Selection Bias (cont')

- Fix #3: Propensity Scoring
- What?: Logistic regression score for likelihood of being in intervention
- How: Controls for "Observable" variation
- Limitations: does not control for "unobservable" variation

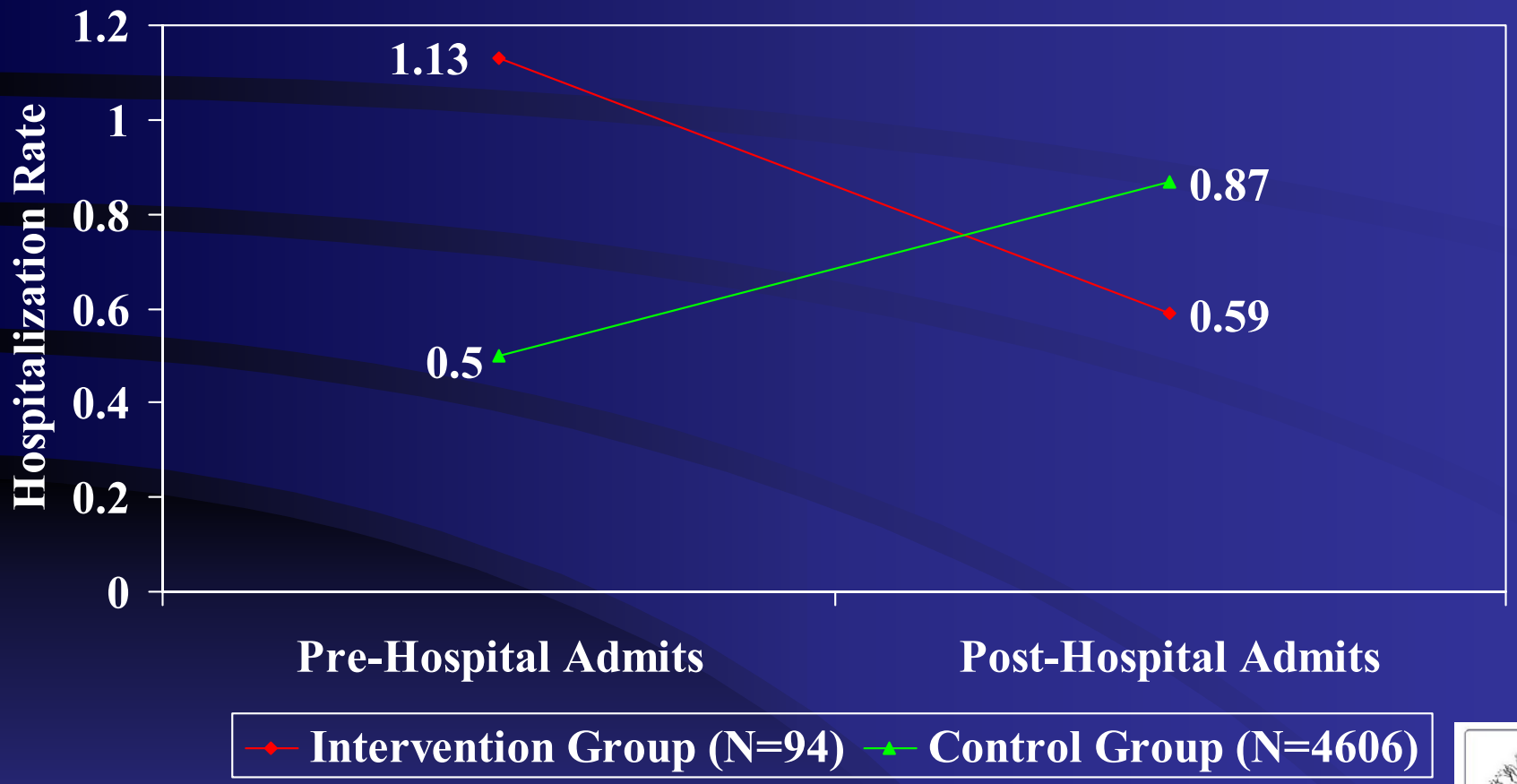


LCG

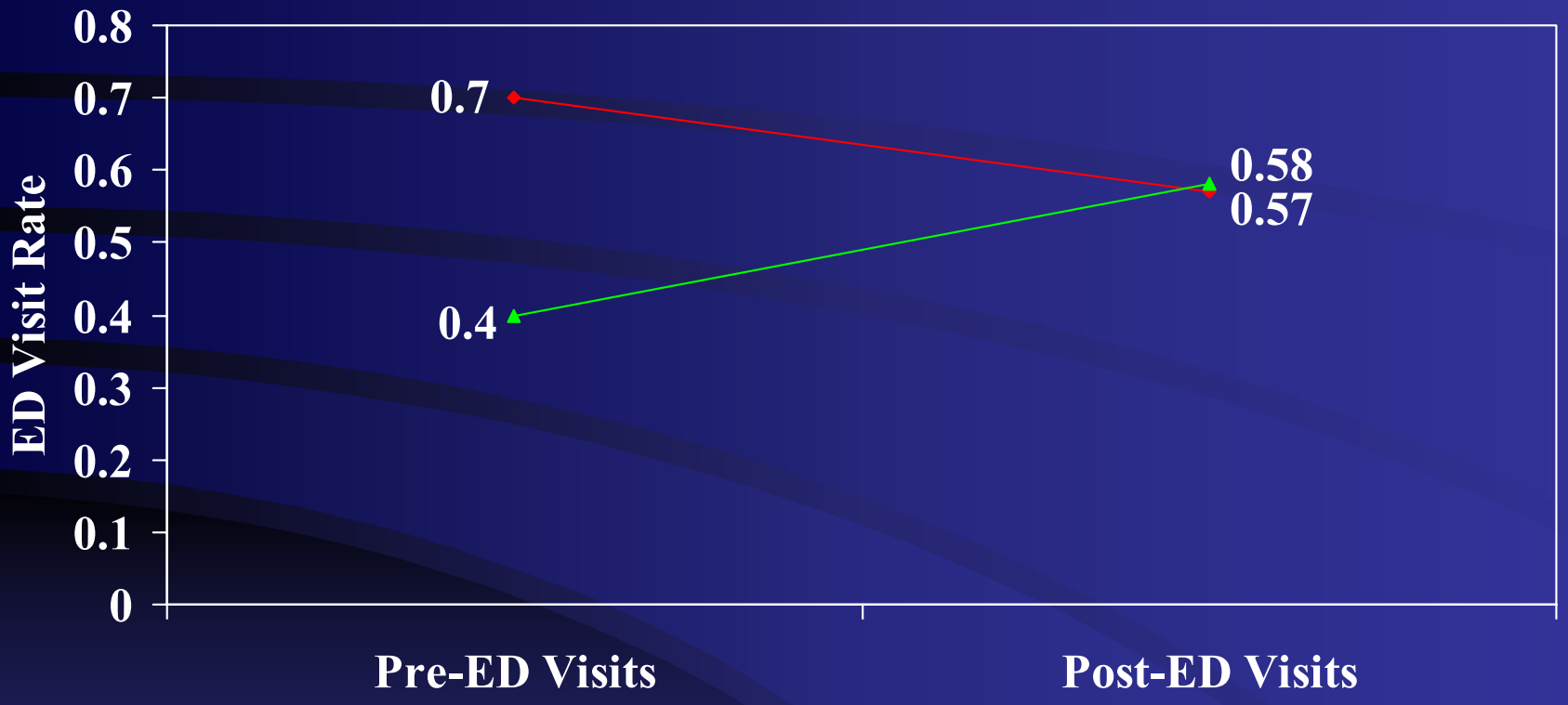
# 1st Year CHF Program Results

| | Intervention (N=94) | Control Group (N=4606) | P(T<=t) two-tail |
|---|---|---|---|
| Age | 77.4 | 76.6 | NS |
| % Female | 0.51 | 0.56 | NS |
| % Portland | 0.17 | 0.69 | p<0.0001 |
| Pre-Hospitalization | 1.13 | 0.5 | p<0.0001 |
| Pre-ED | 0.7 | 0.4 | p=0.003 |
| Pre-Costs | $18,287 | $8,974 | p<0.0001 |
| Post-Hospitalization | 0.59 | 0.87 | p=0.008 |
| Post-ED | 0.57 | 0.58 | NS |
| Post-Costs | $11,874 | $16,036 | p=0.005 |

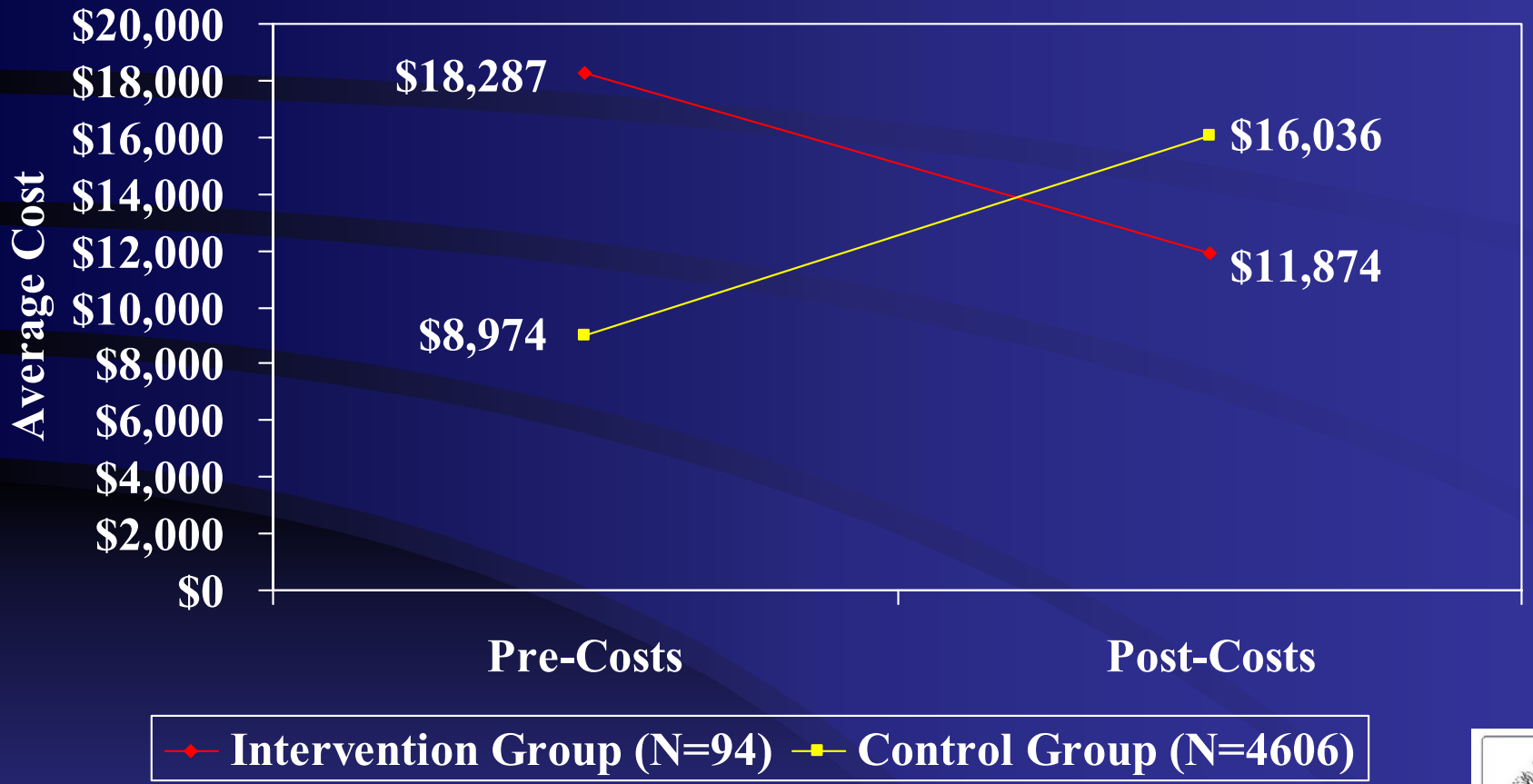
# 1st Year CHF Program Results Admits



# 1st Year CHF Program Results
# ER Visits



# 1st Year CHF Program Results Costs

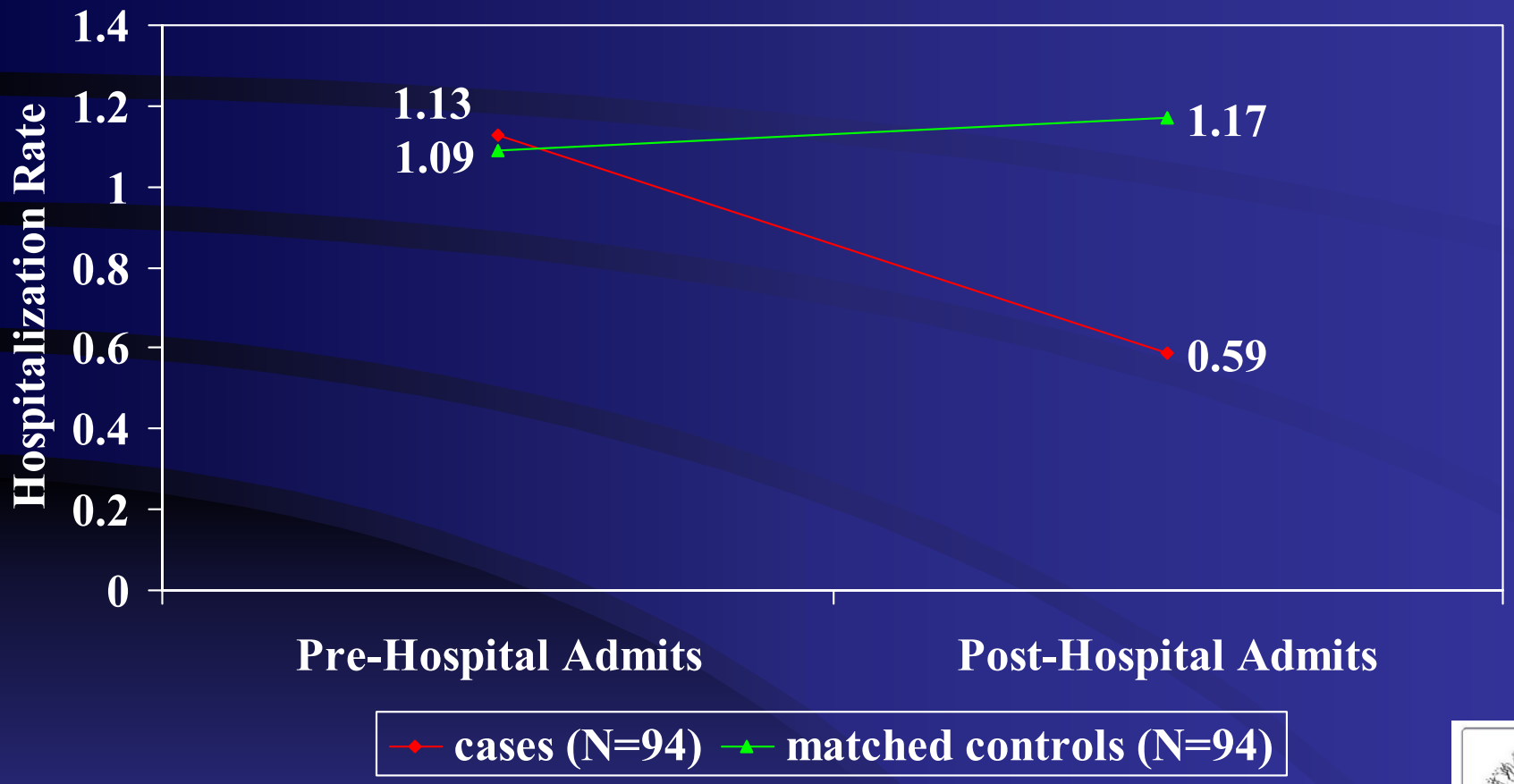


# 1st Year CHF Program Results
## Propensity Scoring Method

| | Cases (N=94) | Matched Controls (N=94) | P(T<=t) two-tail |
|---|---|---|---|
| Propensity Score | 0.061 | 0.062 | NS |
| Age | 77.4 | 78.2 | NS |
| % Female | 0.51 | 0.51 | NS |
| % Portland | 0.17 | 0.17 | NS |
| Pre-Hospitalization | 1.13 | 1.09 | NS |
| Pre-ED | 0.70 | 0.67 | NS |
| Pre-Costs | $18,287 | $17,001 | NS |
| Post-Hospitalization | 0.59 | 1.17 | 0.005 |
| Post-ED | 0.57 | 0.77 | 0.026 |
| Post-Costs | $11,874 | $24,085 | 0.003 |

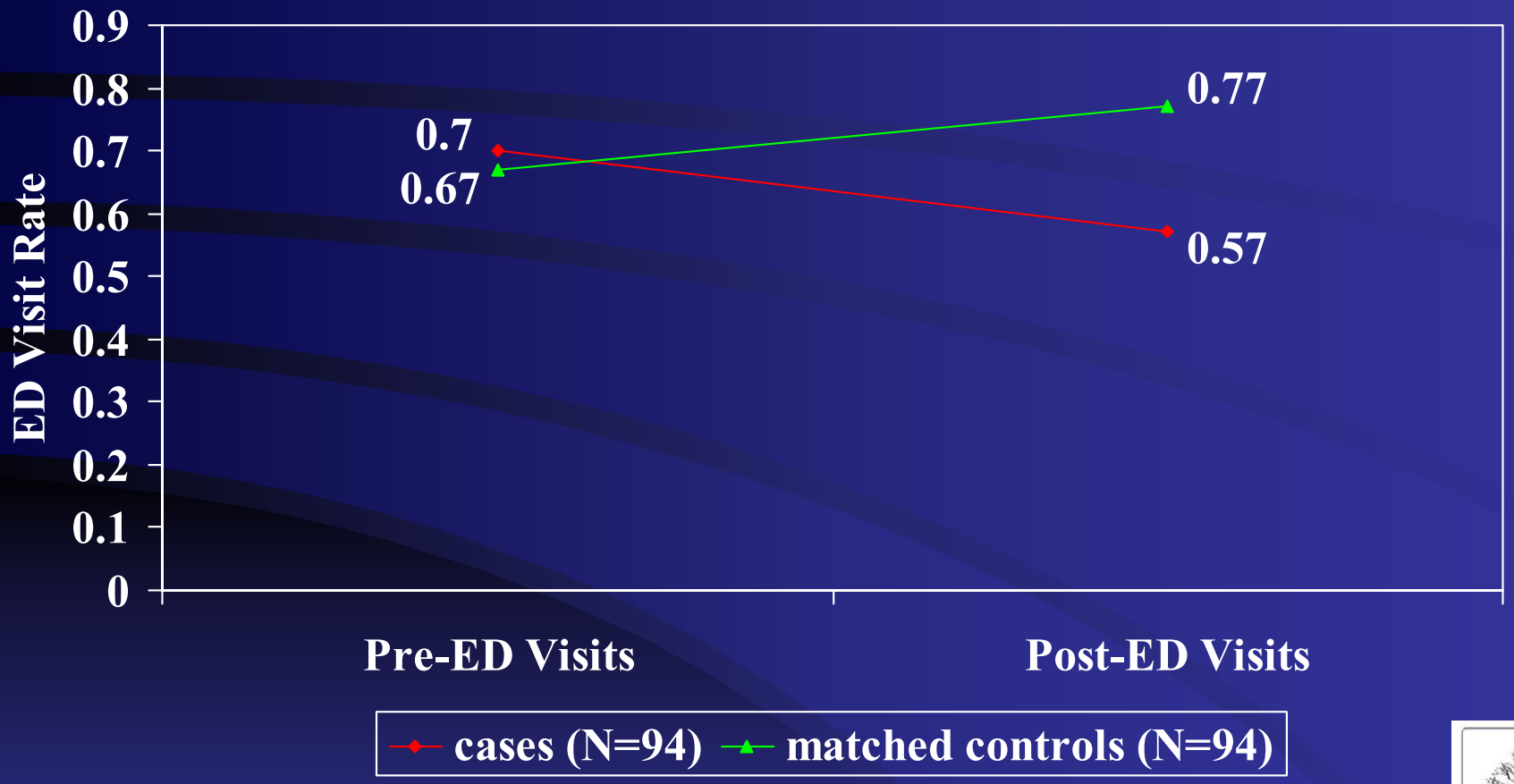
# 1st Year CHF Program Results

## Propensity Scoring Method - Admits



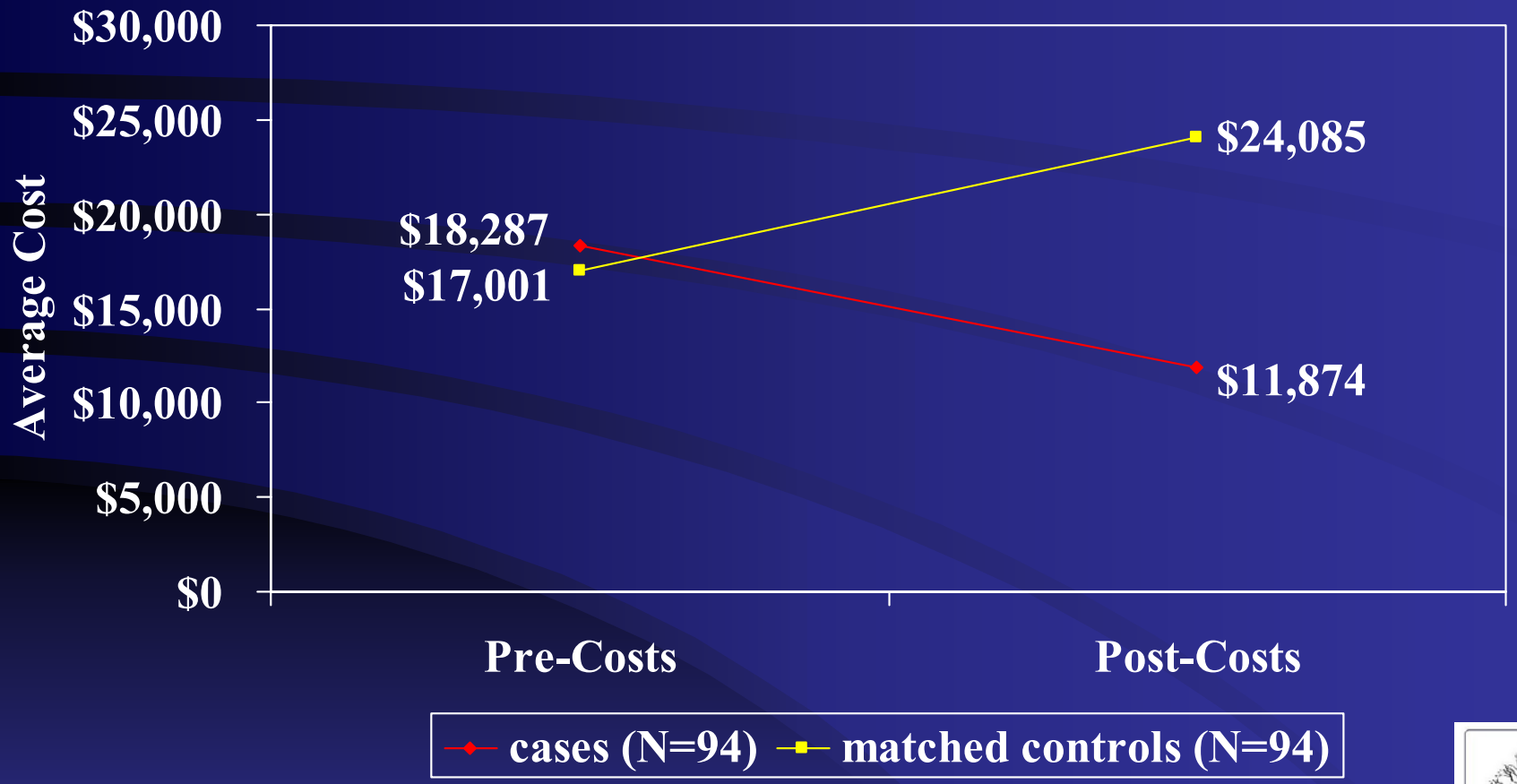
# 1st Year CHF Program Results

## Propensity Scoring Method – ED Visits



# 1st Year CHF Program Results

## Propensity Scoring Method – Costs

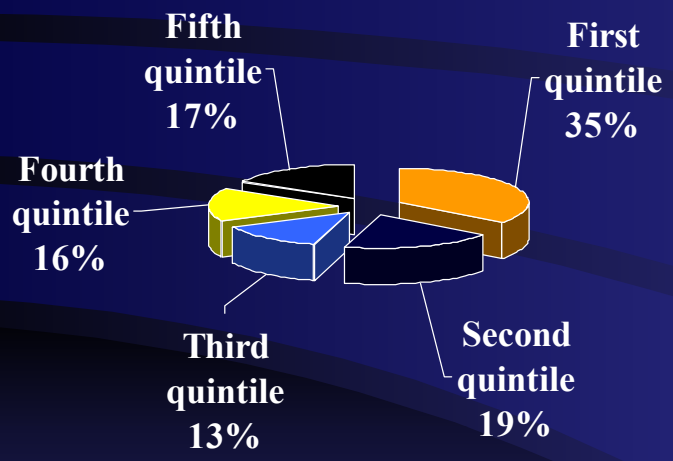


# Regression to the Mean

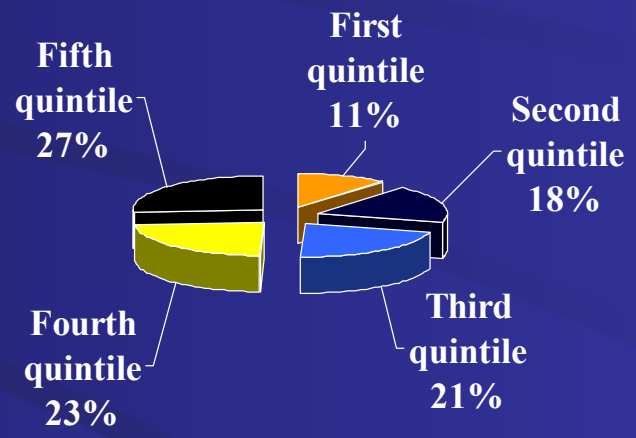
Definition: After the first of two related measurements has been made, the second is expected to be closer to the mean than the first.

# Regression to the Mean
## CAD

Where the 1st Quintile (N=749) Went In Year 2

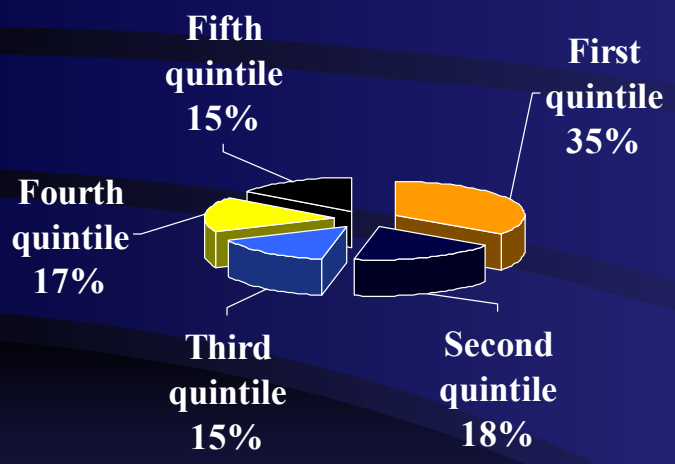



Where the 5th Quintile (N=748) Went In Year 2

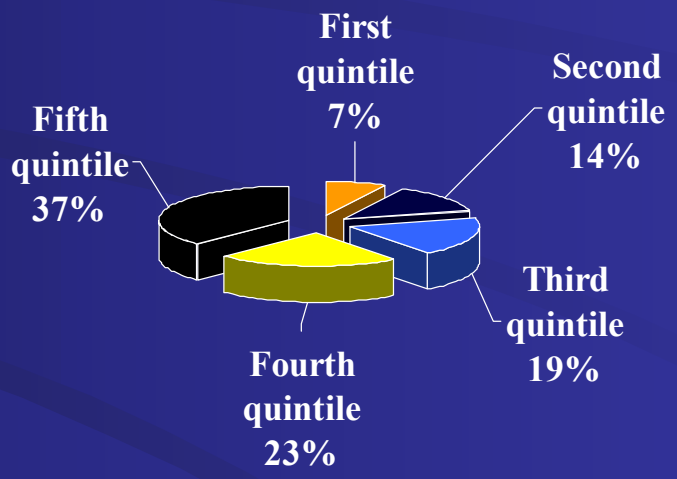


# Regression to the Mean
## CHF

Where the 1st Quintile (N=523) Went In Year 2



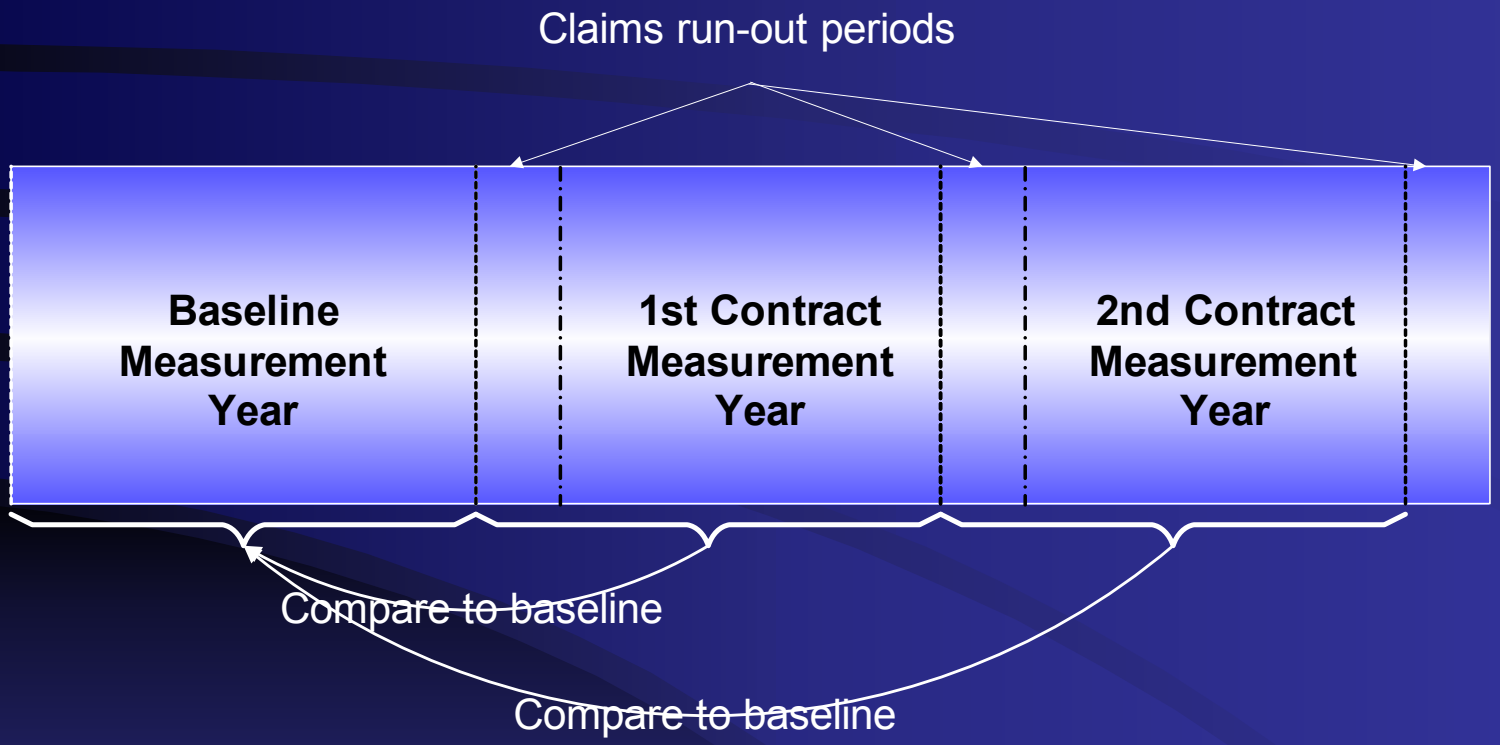

Where the 5th Quintile (N=537) Went In Year 2



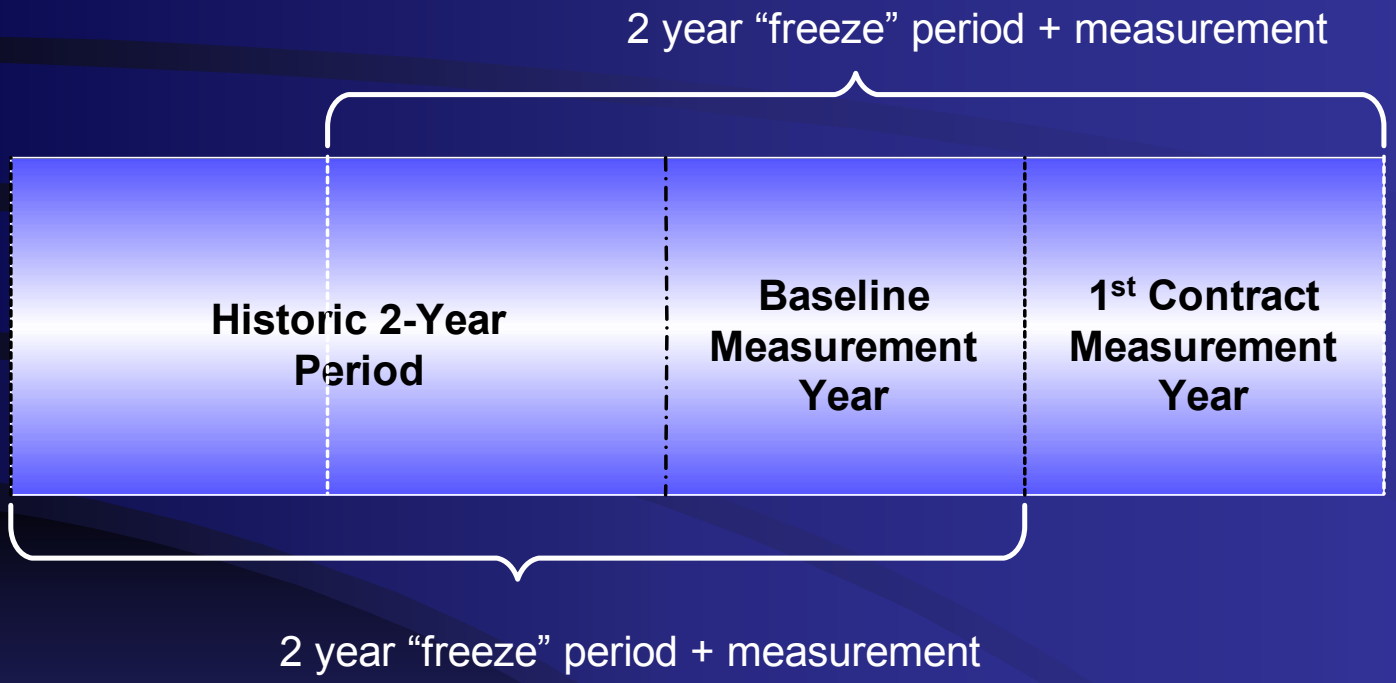
# Regression to the Mean (cont')

- Fix #1: Increase length of measurement periods
- How: Controls for movement toward the mean across periods
- Limitations: periods may not be long enough, availability of historic data

# Regression to the Mean (cont')
## Currently-Used Method



# Regression to the Mean (cont')
## Valid Method (from Lewis presentation)



# Regression to the Mean (cont')

- Fix #2: Time Series Analysis
- How: Controls for movement across many periods (preferably > 50 observations)
- Limitations: availability of historic data, change in collection methods

# Measurement Error

Definition: Measurements of the same quantity on the same group of subjects will not always elicit the same results. This may be because of natural variation in the subject (or group), variation in the measurement process, or both (random vs. systematic error).

# Measurement Error (cont')

- Fix #1: Use all suitables in the analysis (to adjust for the "zeroes")
- Fix #2: Use identical data methods pre and post (like unit claims-to-claims comparison)
- Fix #3: Use utilization and quality measures instead of cost.

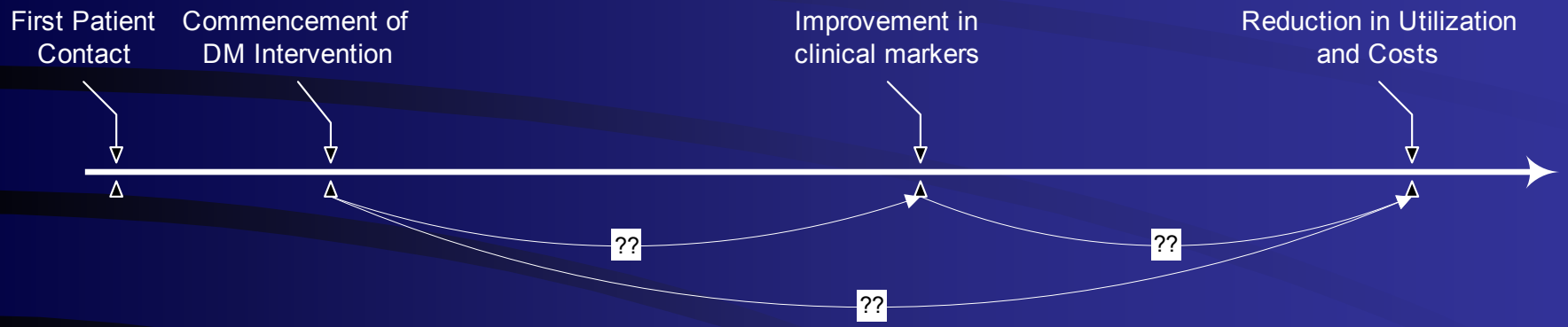
# Alternative Designs

- Survival Analysis
- Time Series Analysis
- Time-dependent Regression

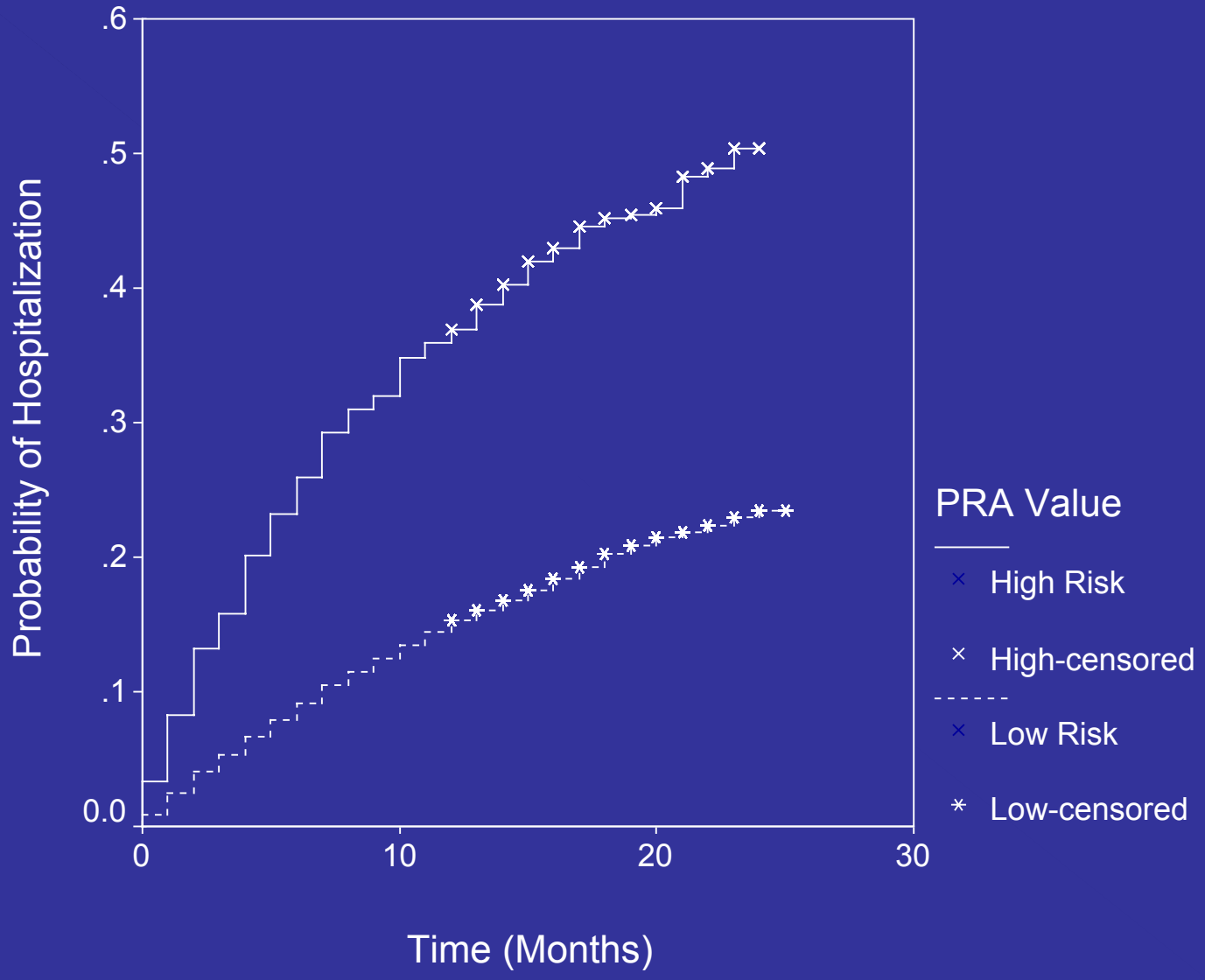
# Survival Analysis

- Features:
  - Time to event analysis – longitudinal
  - Censoring
  - Allows for varying enrollment points

# Survival Analysis



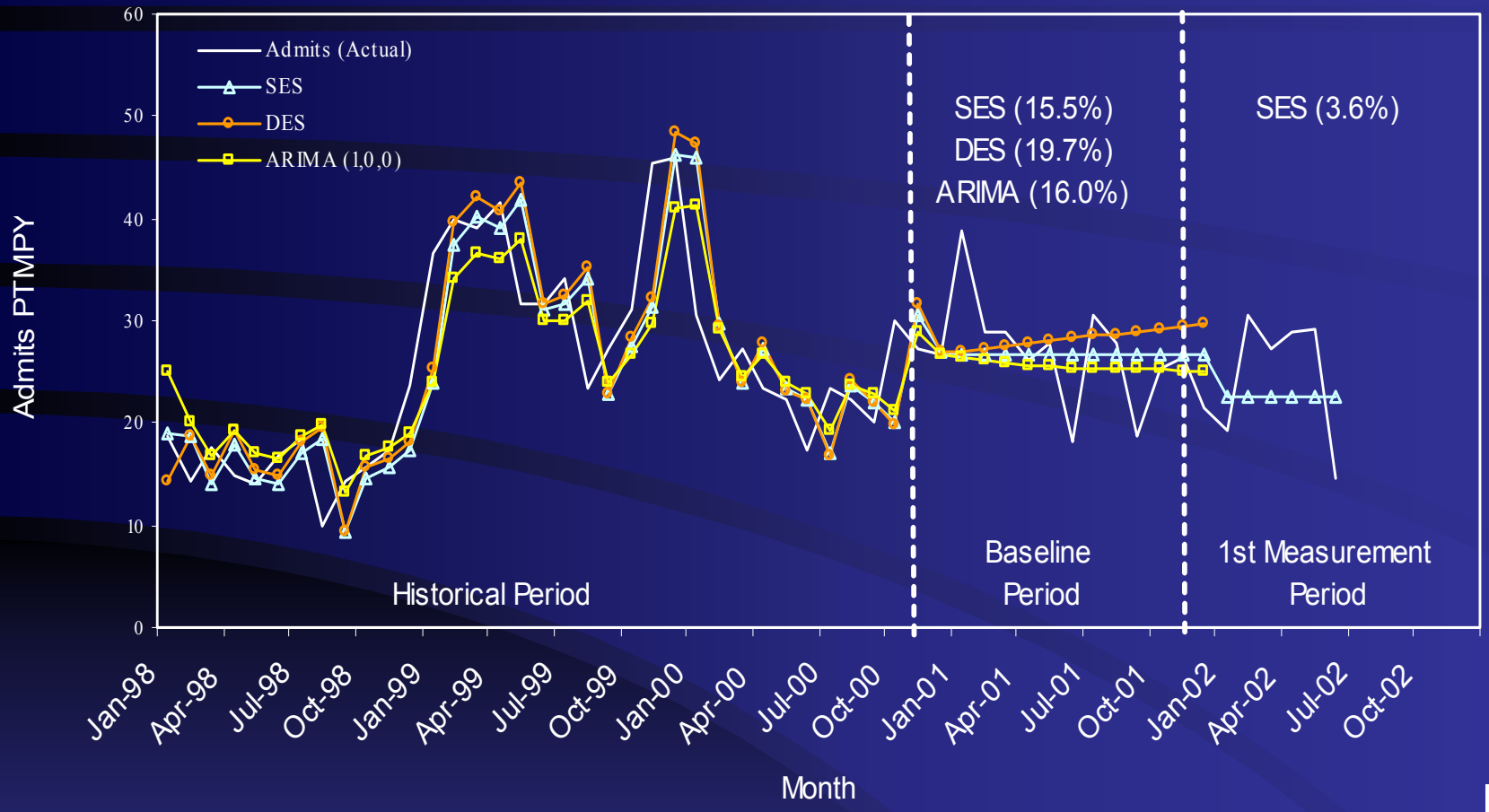
# Survival Analysis



# Time Series Analysis

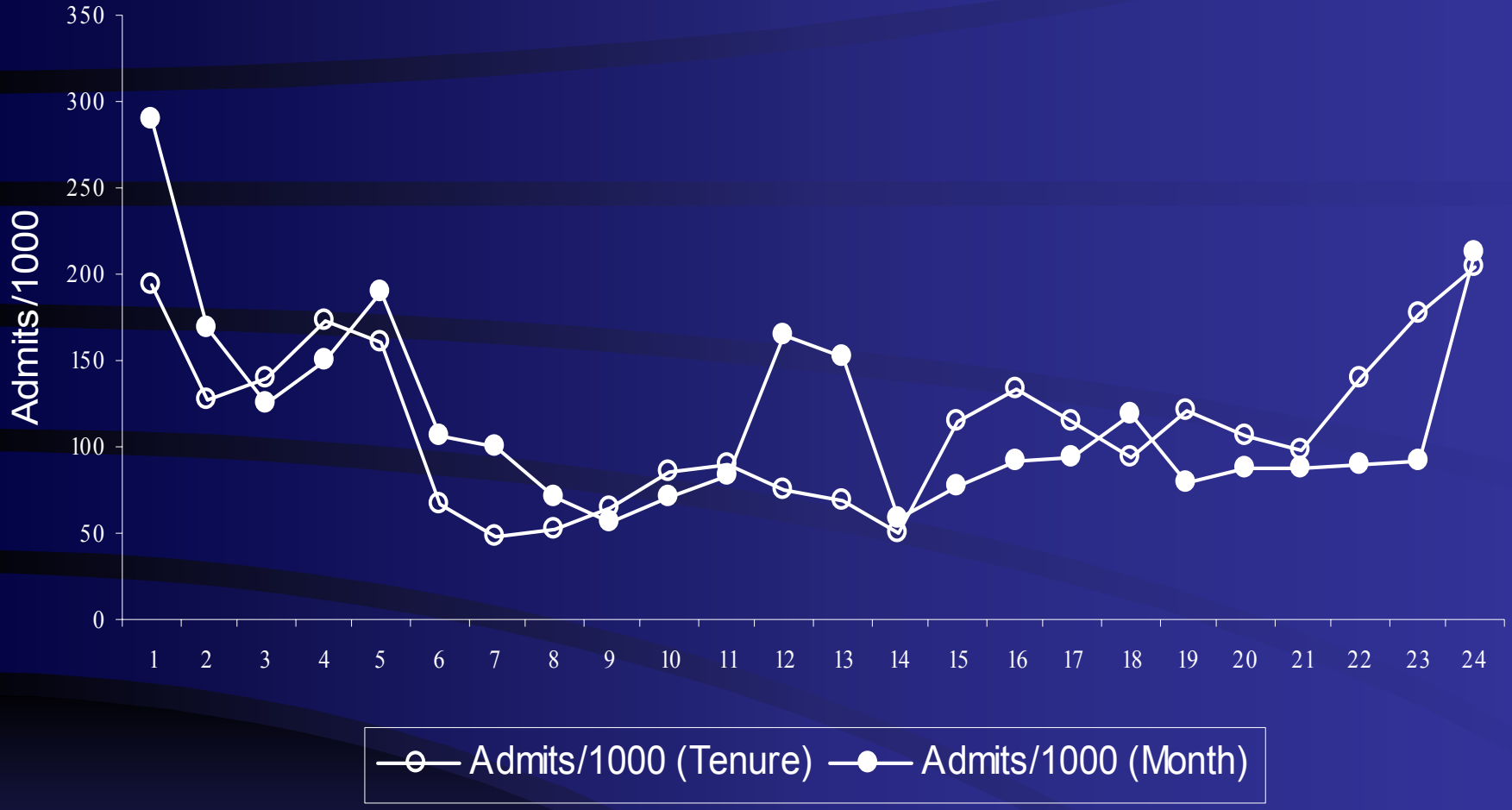
- Features:
  - Longitudinal analysis
  - Serial Dependency (autocorrelation)
  - Does not require explanatory variables
  - Controls for trend and seasonality
  - Can be used for forecasting

# Time Series Analysis (cont')



# Time-dependent Regression

- Combines important elements of other models to create a new method, including variables such as:
  - Program tenure (censuring)
  - Seasonality (important for Medicare)
  - Can be used for forecasting




Simulated hospital admissions per thousand members based on program tenure and month-of-year (months 1-12 represent Jan – Dec of program year 1, and months 13-24 represent Jan – Dec of program year 2).

# Conclusions

- Identify potential threats to validity before determining evaluation method
- Choose outcome variables that mitigate measurement bias (e.g. all identified members vs those with costs)
- There is no panacea! Use more than one design to validate results.

# How does this presentation differ from what you just saw?

- Lewis approach is the only valid pre-post population-based design in use today
- But valid ≠ accurate. "Valid" just means adjustment for *systematic* error
- These methods reduce chances of *non-systematic error* to increase accuracy

# References (1)


1. Linden A, Adams J, Roberts N. An assessment of the total population approach for evaluating disease management program effectiveness. *Disease Management* 2003;6(2): 93-102.

2. Linden A, Adams J, Roberts N. Using propensity scores to construct comparable control groups for disease management program evaluation. *Disease Management and Health Outcomes Journal* (in print).

3. Linden A, Adams J, Roberts N. Evaluating disease management program effectiveness: An introduction to time series analysis. *Disease Management* 2003;6(4):243-255.

4. Linden A, Adams J, Roberts N. Evaluating disease management program effectiveness: An introduction to survival analysis. *Disease Management* 2004;7(2):XX-XX.




LCG

# References (2)


5. Linden A, Adams J, Roberts N. Evaluation methods in disease management: determining program effectiveness. Position Paper for the Disease Management Association of America (DMAA). October 2003.

6. Linden A, Adams J, Roberts N. Using an empirical method for establishing clinical outcome targets in disease management programs. *Disease Management.* 2004;7(2):93-101.

7. Linden A, Roberts N. Disease management interventions: What's in the black box? *Disease Management.* 2004;7(4):XX-XX.

8. Linden A, Adams J, Roberts N. Evaluating disease management program effectiveness: An introduction to the bootstrap technique. *Disease Management and Health Outcomes Journal* (under review).

# References (3)


9. Linden A, Adams J, Roberts N. Generalizability of disease management program results: getting from here to there. *Managed Care Interface* 2004;(July):38-45.

10. Linden A, Roberts N, Keck K. The complete "how to" guide for selecting a disease management vendor. *Disease Management.* 2003;6(1):21-26.

11. Linden A, Adams J, Roberts N. Evaluating disease management program effectiveness adjusting for enrollment (tenure) and seasonality. *Research in Healthcare Financial Management*. 2004;9(1): XX-XX.

12. Linden A, Adams J, Roberts N. Strengthening the case for disease management effectiveness: unhiding the hidden bias. *J Clin Outcomes Manage* (under review).




LCG

# Software for DM Analyses

- The analyses in this presentation used XLStat for Excel. This is an Excel add-in, similar to the data analysis package that comes built-in to the program.
- Therefore, users familiar with Excel will find this program easy to use without much instruction.

# Questions?

Ariel Linden, DrPH, MS
ariellinden@yahoo.com