

# PHI Institute

## DM ROI Methods: The Good, the Bad, & the “Insufficient”

SRI -3rd Annual Disease Management Congress

Boston - August 2, 2005

Thomas Wilson, PhD, Dr. PH  
Founder & CEO

Population Health Impact Institute

[www.PHIinstitute.org](http://www.PHIinstitute.org)

513-677-8123

[twilson@PHIinstitute.org](mailto:twilson@PHIinstitute.org)

# Background

## CBO Report

“Insufficient Evidence” from the Peer-Reviewed Literature.

- *“insufficient evidence to conclude that DM programs can generally reduce overall health spending.”*



CONGRESSIONAL BUDGET OFFICE  
U.S. Congress  
Washington, DC 20515

*Douglas Holtz-Eakin, Director*

## **An Analysis of the Literature on Disease Management Programs**

October 13, 2004

# Organization

## 1) Definitions

“Good,” “Bad,” and the “Insufficient”

## 2) The PHI Institute

## 3) Five Principles of Evaluation

## 4) Illustrations of Evaluation Error (from peer-reviewed literature)

## 5) Award to PHI Institute

## 6) Future Directions: DM and the PHI Institute

# I) Definitions

*(with apologies to Sergio Leone)*

## The “Good”

- Available for public review with impact methods well-described.
  - **Transparency:** *minimal disclosure of methods and metrics necessary to replicate any impact assessment.*” ([www.PHIinstitute.org](http://www.PHIinstitute.org))

## The “Bad”

- Falsification, Fabrication, Plagiarism
  - Nature (June 2005): Very uncommon
- Secret. Impact method not available for public review

## The “Insufficient”

- Impact methods are available for partial or full review, but these are not described in sufficient detail to enable fair review.

**ALL ARTICLES REVIEWED HERE ARE “GOOD”**

# The “Good”

## *The Case of The Ivory Billed Woodpecker.*



- One group of experts (Cornell) identify a long extinct woodpecker in Arkansas, submitting evidence to prove their case
- Another group of experts dispute the findings
- Late July article in the NYT: Article and rebuttal to appear in peer-reviewed journal.
- Earlier this week: Audio recording of “song” was reported to convince the doubters.

*"We would have been disappointed if there was no close scrutiny by the scientific community," Connie Bruce [Cornell, spokesperson for Ivory Billed Woodpecker project], said "This is not surprising at all that we have controversy. It's the process. It should definitely be going through this."*

***“If it is good enough for bird watching,  
it is good enough for DM.”***

## II) The PHI Institute: A Non-Profit Organization

**Mission: Promote and conduct independent & valid evaluations of defined population health programs through ...**

### Education

- Knowledge Transfer Workshops.

### Research

- Papers and Articles based on .... Literature reviews, focus groups, surveys, health data.

### Benchmarking

- VALIDITY: of methodology (assuming transparency of methods)
- VERIFICATION: of results (using multiple methods)
  - Methods-to-Method & Program-to-Program impact Crash Tests<sup>SM</sup>

# III) Five Evaluation Principles

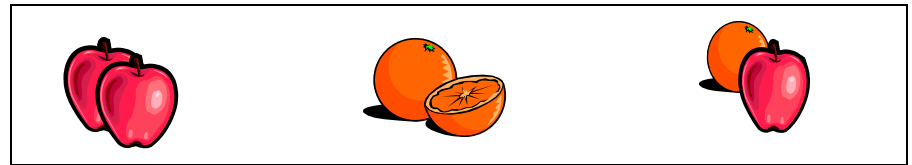
- 1) **Data Quality**
- 2) **Equivalence**
- 3) **Statistical Quality**
- 4) **Causality**
- 5) **Generalizability**

© 2005 Population Health Impact Institute. All Rights Reserved. PHI Institute encourages the private, non-commercial use of this framework, however, it reserves the exclusive right for benchmarking services.

# Evaluation Principles Visualized

## 1. Data Quality

*Measurement Error*



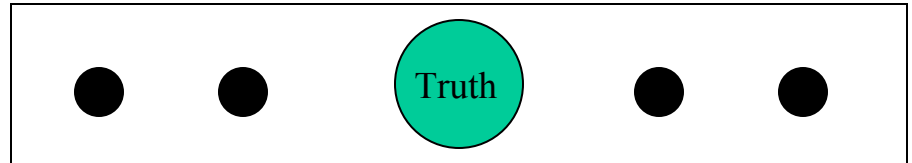
## 2. Equivalence

*Differential Error*



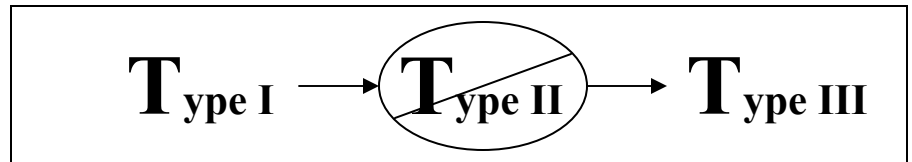
## 3. Statistical Quality

*Non-Differential Error*



## 4. Causality

*Conceptual Error*



## 5. Generalizability

*Transmission Error*



# III) Illustrations of Principles based upon “good” studies

*All had letters-to-the-editor, most had author’s responses to these letters (see reference slide at end)*

# Principle #1: Data Quality

## Potential Measurement Error

Case: Definition of population in reference and intervention groups (in this case pre and post)

Potential problem: Different inclusion and/or exclusion criteria in intervention group than in reference group. *Measurement bias*

Original Source: Wilson TW, Gruen J, Thar W, Fetterolf D, Patel M, Popiel RG, Lewis A, Nash DB. Assessing ROI of defined-population disease management interventions. *Joint Commission Journal of Quality and Safety*. 2004; 30 (11); 614-21.

# Principle 2: Equivalence

## Potential Differential Error

Results: Positive ROI in CHF program

Case: Finding comparable reference group for op-in DM program, used non-participants and “propensity scoring” to select reference.

Potential problem: The number of people available for reference group (n=236) were considerably less than the number of participants (n=533). Many people were used more than once.

Original Source: Berg GD, Wadhwa S, Johnson AE. “A Matched-Cohort Study of Health Services Utilization and Financial Outcomes for a Heart Failure Disease-management Program in Elderly Patients.” *J Am Geriatr Soc* 52: 2004; 1–7.

# Principle 3: Statistical Quality

## Potential Non-Differential Error

Results: Positive ROI in Asthma program

Case: Statistical testing of reference group (based upon age) to show its similarity to intervention group.

Potential problem: The statistical test was stated as not-significant, when the difference was actually significant.

Original Source: Tinkelman D, Wilson S. “Asthma Disease Management: Regression to the Mean or Better?” *Am J Manag Care*. 2004; 10 (12): 948-54.

# Principle 4: Causality

## Potential “Conceptual Error”

Results: Positive ROI (8:1) in DM program

Case: ROI was based upon the difference in costs between those randomized into the intervention groups vs. the control group. A specific ROI was based upon changes among those with the actual intervention (n=961 with ROI of 68.0) compared to a like group

Potential problem: On recalculation, the ROI seems closer to 3.0 when calculating only on those who received the “care consideration” intervention, not all people randomized into the reference and control groups. Changes that occurred in the intervention population, but did not receive the intervention, were included in the 8:1 calculation  
*“False attribution” problem*

Original Source: Javitt JC, Steinberg G, Locke T, Couch JB, Jacques J, Juster I, Reisman L. "Using a Claims Data-Based Sentinel System to Improve Compliance with Clinical Guidelines: Results of a Randomized Prospective Study." *Am J Manag Care*. 2005;11:93-102.

# Principle 5: Generalizability

## *Potential “Transmission Error”*

Results: Positive results on clinical measures in CHF population, no savings reported.

Case: Study with high “internal validity” (randomized control trial)

Potential problem: Potential low “external validity” CHF patients were already following evidence based guidelines. Had the study show an effect, it may have been expanded into other areas (see the authors letter to the editor)

Original Source: Galbreath AD, Krasuski RA, Smith B, et al. “Long-Term Healthcare and Cost Outcomes of Disease Management in a Large, Randomized, Community-Based Population With Heart Failure.” *Circulation*. 2004; 110(23):3518-26.

# Observations

- All DM studies need to be “good,” (not “bad” and not “insufficient”). They need to be “transparent.”
- Publication in a peer-reviewed journal is no guarantee of high validity, however, it does allow for further review, if warranted.

# IV) Health Industry Forum Award to PHI Institute

## *Three Phases*

- 1) Peer-reviewed article on scoring system
  - Emphasis on need to improve of simple scoring based on study design (expert opinion, observational, quasi-experimental, and experimental)
  - Scoring should be based on principles, not study design.
    - *You saw a preview of that article today*
  
- 2) Delphi-panel to score 25 articles on DM in health plans--based upon framework of Five Evaluation Principles.
  
- 3) Users Guide to self-scoring of DM methods and results

# V) Future Directions

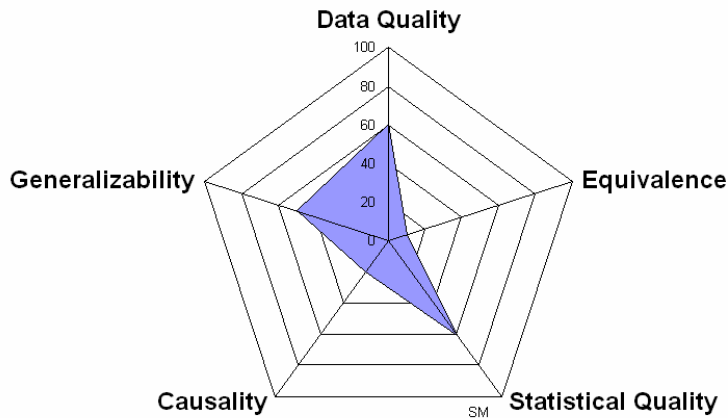
Project scoring worksite health promotion planned.

PHI Institute offers Validity and Verification scoring services to commercial “vended” population health management programs and “built” in-house defined population health programs.

# Example of Evaluation Scoring

## Article 13

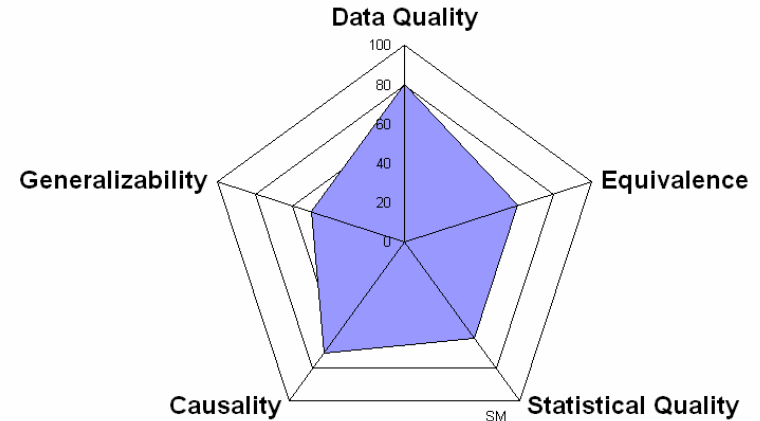
### VALIDITY PENTAGON<sup>SM</sup>



© 2005 Population Health Impact Institute, Cincinnati, Ohio. For private use only. Commercial use prohibited. All other rights reserved.

## Article 25

### VALIDITY PENTAGON<sup>SM</sup>



© 2005 Population Health Impact Institute, Cincinnati, Ohio. For private use only. Commercial use prohibited. All other rights reserved.

© 2005 Population Health Impact Institute. All Rights Reserved. PHI Institute encourages the private, non-commercial use of this framework, however, it reserves the exclusive right for benchmarking services.

# Conclusion:

## Future of DM Measurement

- Compare and contrast, openly
  - *Depends on transparency*
- Compare and contrast, credibly.
  - *Depends on evaluation principles*

# “Good” Bibliography

#1 DATA QUALITY: Wilson TW, Gruen J, Thar W, Fetterolf D, Patel M, Popiel RG, Lewis A, Nash DB. . Assessing ROI of defined-population disease management interventions. *Joint Commission Journal of Quality and Safety*. 2004; 30 (11); 614-21.

Commentary: Duncan I. "Accuracy in the Assessment of Return on Investment of Defined Population Interventions" *Joint Commission Journal of Quality and Safety*. 2005; 31 (6); 357. Wilson T, Thar W, Gruen J. "Reply." *Joint Commission Journal of Quality and Safety*. 2005; 31 (6); 358.

#2 EQUIVALENCE: Berg GD, Wadhwa S, Johnson AE. “A Matched-Cohort Study of Health Services Utilization and Financial Outcomes for a Heart Failure Disease-management Program in Elderly Patients.” *J Am Geriatr Soc* 52: 2004; 1–7.

Commentary: Wilson TW, Linden A. “Potential Bias in ‘Controls’ Used in a Heart Failure Disease Management Program.” *J Am Geriatr Soc* (in press).

#3 STATISTICAL QUALITY: Tinkelman D, Wilson S. “Asthma Disease Management: Regression to the Mean or Better?” *Am J Manag Care*. 2004; 10 (12): 948-54.

Commentary: Levin-Scherz, J. "To the Editor." *Am J Manag Care*. 2005; 11 (3): 134,136. Wilson T, Linden A. “To the Editor.” *Am J Manag Care*. 2005; 11 (3): 136. Wilson S, Tinkelman D. "The Authors Respond." *Am J Manag Care*. 2005; 11 (3): 136-37.

#4 CAUSALITY: Javitt JC, Steinberg G, Locke T, Couch JB, Jacques J, Juster I, Reisman L. "Using a Claims Data-Based Sentinel System to Improve Compliance with Clinical Guidelines: Results of a Randomized Prospective Study." *Am J Manag Care*. 2005;11:93-102.

Commentary: Wilson TW, Linden A. "To the Editors." *Am J Manag Care*. 2005;11: 999. Javitt JC. "In Reply." *Am J Manag Care*. 2005;11: 999.

#5 GENERALIZABILITY: Galbreath AD, Krasuski RA, Smith B, et al. “Long-Term Healthcare and Cost Outcomes of Disease Management in a Large, Randomized, Community-Based Population With Heart Failure.” *Circulation*. 2004; 110(23):3518-26.

Commentary: Linden A, Wilson T. "Letter." *Circulation*. 2005; 112: e11. Galbreath A, Freeman GL, Smith B, Krasuski RA, Stajduhar KC, Kwan MD, Ellis R. "Response." *Circulation*. 2005; 112: e11.

# Contact Information

Thomas Wilson, PhD, Dr. PH

Founder & CEO

Population Health Impact Institute

10663 Loveland-Madeira Rd, #210B

Loveland, Ohio 45140

email: [twilson@PHIinstitute.org](mailto:twilson@PHIinstitute.org)

phone: 513-677-8123

web: [www.PHIinstitute.org](http://www.PHIinstitute.org)