Evaluation Methodologies

Seth Serxner, PhD, MPH
Mercer

Michael Cousins, PhD
CIGNA

September 2011

Part 1. Agenda

- **Measurement Landscape**
  - Point of view
  - Current state
  - Methods overview
- **Best-Practices**
  - Reporting features
  - Reporting metrics
  - Performance measures and targets
  - Methodology
- **Moving the Industry**
  - Employers/broker/consultant demands
  - Future state

Measurement Landscape

Consulting Point-of-View

- **Evaluations are to benefit the client**
- **They help the vendors document program value as well as find opportunities for improvements**
- **Recognizes the challenges of implementing ideal methods in a production environment**
  - Different program offerings and combinations
  - Variation in client sizes
  - Inconsistency in implementation of programs
  - Comprehensive reporting can be overwhelming to the client
- **Program evaluation is more than just ROI. Program evaluation includes both process and outcomes analysis**
  - Return-on-investment (ROI) analysis is a subset of outcomes analysis
  - Determining the program investment may be a complicated process
  - Process measures, including participation levels and clinical impact, add confidence to the estimated program impact used in an ROI analysis

Consulting Point-of-View

Metrics support can apply to providers as well as vendors

Broad Domains	Example Metrics
Operational excellence	– Satisfaction (Member, Stakeholder) – Timely Outreach – Timely Transfer of Data – Resources Allocation
Engagement	– Participation (overlapping and non-overlapping) – Engagement (# and duration of calls) – Retention / Average tenure – Goal setting and attainment
Health improvement	– Health Risk Change – Clinical Impact / Behavioral Impact – Quality of Life
Savings	– Health Service Utilization– Health Risk Change – Financial/ROI (Healthcare: Medical/Rx) – Financial/ROI (Productivity: Disability/Absence, Presenteeism)

Current State
Reporting on the "Value Proposition"

- **Reporting packages are still lacking in telling the entire story (i.e., leading and lagging indicators)**
  - Incoherent and technical approach to data presentation
  - Missing metrics
  - Inconsistency in operational definitions (e.g., participation, clinical)
  - Discrepancy in methods across metrics and programs
  - Reporting packages are in a constant state of change

- **Some vendors and carriers are adopting best-practice methods, but there are still many methods that are less than best-practice**
  - Adoption tends to be on a case-by-case basis within a given vendor or carrier
  - More rigorous methodologies may come at a cost to the client
  - Performance guarantee targets vary according to methodology

Current State
Reporting on the "Value Proposition"

- **Methods**
  - Financial Savings Model
  - Pre-Post Historical Control (Trend-based)
  - "Matched" Control using propensity score (matching, weighting, covariate, multi-pass no propensity score)
  - Randomized control studies

Current State: Methodologies
Financial Savings Model

Description	Strengths	Weaknesses	Recommended Enhancements to Current Methodology
• Savings modeled on some measured unit (e.g., number of members actively engaged, gaps closed, risk reduced and/or eliminated) • Savings per unit based on external book-of-business analysis or published study • Total estimated savings equals # of units X savings per unit	• Easiest to conduct -Does not require analysis of actual claims data • Quicker turnaround of financial reporting (no need for claims run-out) • Can provide more transparent reporting of active engagement if needed for model	• Highly dependent on the assumptions of the model.	• Base model on peer-reviewed literature • Make sure assumptions of model are transparent (e.g., participation) • Ensure model accounts for and adjusts to prevent double-counting of projected savings • Confirm savings are on net change (opportunity for savings and loss) • Discuss and mutually agree on all assumptions used in the model (e.g., per participant savings)

Current State: Methodologies

Pre-Post Historical Control (Trend-based)*

Description	Strengths	Weaknesses	Recommended Enhancements to Current Methodology:
• Method takes the baseline and identifies the diseased population, determines total cost and per disease member per month (PDMPM) cost using the number of months as the denominator and the total cost for the population as the numerator • Pre PDMPM cost is then compared to the post PDMPM cost to see if there is a difference • Members do not need to be in both the baseline and program periods • Pre PDMPM cost is then increased by an estimated healthcare cost trend to determine the projected PDMPM. That projected PDMPM is compared to the actual program year PDMPM to determine a difference that represents savings	• Easier to conduct in a reporting environment • Does not require participation data • Simpler presentation of calculation of savings	• Does not establish causal relationship between program participation and change in costs –Measures something happened, not why • Masks the need to provide explicit participation data • Does not account for impact of other programs • Method best designed for core five chronic conditions -Not appropriate for acute conditions • No generally acceptable method for choosing trend	• Ensure methodology consistently identifies eligible members throughout all reporting periods (vs. “once in, always in”) • Use client-based trend, mutually agreed upon and adjusted for plan design and demographics • Consider 24 months of baseline data, pre-program implementation • Ensure no data gaps between program year and baseline • Confirm eligible members are enrolled in medical plan for at least six months of coverage • Confirm analysis examines the impact on total costs, not condition-specific costs • Discuss and mutually agree on exclusion criteria -Consider analysis with and without exclusions • Consider a utilization based approach to avoid the need for cost trend • Require detailed reporting of change in costs by level of program engagement

* Typically associated with Care Continuum Alliance recommendations.

Recommended Methodology
"Matched" Control Overview

Description	Strengths	Weaknesses
• Analysis "matches" program participants to similar non-participants, based on observable characteristics (e.g., demographics, risk profile, utilization) • Matching (or minimizing differences between groups) can be done by several approaches including weighting, matching, covariates, and multi-pass • Analysis then compares change in costs between baseline and program year within the participant group to the non-participant group's difference during the same time period, controlling statistically for any remaining differences between the two groups -This "difference in difference" is considered the estimated savings per participant • Savings per participant are multiplied by the number of program participants in a given period to determine program savings	• Higher causal relationship • Explicit participation data is required • Controls known and measured confounding variables • Method can account for impact of other programs • Allows a measure of confidence around the results and/or significance of the results	• More difficult to conduct • Data may not be available to conduct this level of analysis • More difficult to explain to lay audiences • Does not account for unobserved differences between participants and non-participants (biases may exist)

Recommended Methodology

Methodology Detail

- **This approach requires an individual-level database that includes detailed program participation information, demographics and outcomes data**
- **Two sets of regression analyses are used to determine savings:**
  - The first analysis matches program participants to the most similar non-participants, based on observable data available (e.g., demographics, risk profile, utilization)
  - In the second analysis, a separate regression analysis is used to compare the change in costs between baseline and program year within the participant group to the non-participant group's difference during the same time period, controlling statistically for any remaining differences between the two groups. This "difference in difference" is considered the estimated savings per participant
  - Savings per participant are multiplied by the number of program participants in a given period to determine program savings
  - A key criteria to this methodology is that the first analysis needs to be deemed successful (the two groups are equalized on critical variables such as demographics, baseline costs and health service utilization) prior to conducting the second analysis.

Best-Practice Reporting

Reporting Features
Best Practice Reporting Principles

Timely

- Timely delivery of reports following close of reporting period.
- Receipt of reports prior to presentation of results to allow for preparation.

Accurate

- Review process built into reporting production schedule.
- Accurate reports; numbers tie throughout report.

User-friendly

- Can the report stand on its own without interpretation from a “talking head”?
- Can a lay person understand the key findings?
- User-friendliness is enhanced by:
  - Judicious labeling
  - Documenting of data sources
  - Use of white space, larger fonts, graphics
  - Defining terms (providing a glossary)

Actionable

- Does the report simply “state the facts” or does it draw conclusions and provide recommendations for corrective action?
- Does the report provide both direction and meaning?
- Reports should include not only comments summarizing “what” the reader is seeing, but the “so what” as well

Reporting Features
Transparency

- Highly transparent documentation that facilitates peer review and replicability
- Assumptions and definitions working document
  - For example numerators, denominators
  - Should be consistent across all programs and analyses
- Methodologies for evaluations of all metrics (e.g., leading and lagging indicators)
- Show your work

Reporting Metrics
Key Questions: Boiling it all Down

- **Are people engaged?**
  - ✓ In the program(s)?
  - ✓ In prevention?
  - ✓ In their health care decisions?
- **Did health Improve?**
  - ✓ Self-reported health risks
  - ✓ Prevalence of chronic conditions
  - ✓ Illness burden (e.g., DCG)
- **Were there savings (aka What's the ROI)?**
  - ✓ Medical/Rx
  - ✓ Disability/Absence
  - ✓ Productivity
- **If not, who's not? (Even if so, who's not)?**
  - ✓ Engaged
  - ✓ Improving their health
  - ✓ Providing savings

Performance Measures & Targets

Performance Monitoring and Improvement

- Identify broad domains and available metrics
- Mutually agree on methodology for calculating metrics
- Establish targets (both minimum standards and targets for success)
- Determine fees at risk and distribution for performance guarantees

Broad Domains	Available Metrics	Targets	Fees at Risk
Operational excellence	– Satisfaction (Member, Stakeholder) – Timely Outreach – Timely Transfer of Data	Establish a Range: • Minimum = Performance Standard (with or without fees at risk) • Maximum = targets for Success	To be determined <u>after</u> methodologies and targets are agreed upon
Engagement	– Participation – Engagement (# of calls) – Retention		
Health improvement	– Risk Change – Clinical Impact		
Savings	– ROI (Medical/Rx, Disability/Absence, Presenteeism)		

Moving the Industry

Employer / Broker / Consultant Demands
Moving the Industry

- **Clients are pushing vendors to perform more rigorous ROI analytic methodologies**
  - More important than the amount of fees at risk is the methodology
  - Clients would rather have accurate, actionable information with realistic targets than poor methodology that hits non-credible targets.
- **A number of vendors accepted and implemented a rigorous ROI analytic methodology and put fees at risk for the outcomes**
- **Vendors asking for Interesting / creative caveats**
  - HERO Scorecard target
  - Requirements for:
    - Incentive levels
    - Engagement levels
    - Communication activity
    - Minimum sample size

Employer / Broker / Consultant Demands
Key Demands

- **Raised expectations of what they would like to receive from vendors**
  - Mutually agreed on a reporting package
    - Reporting calendar and regular quarterly reporting meetings focused on key performance measures
  - Established targets (both minimum standards and stretch goals)
  - Enhanced the rigor of the financial savings methodology
- **Clients will no longer accept excuses for not receiving reporting**
  - Excuses include, it is “too difficult” or “not statistically significance”, but these can be addressed through mutual partnership with client and vendor
- **Integrated analyses that demonstrate that incremental value of each program component**
- **Mutually agreed upon methodologies before setting performance targets**
- **Transparent, timely, accurate, user-friendly, and actionable reporting**
  - Best-in-class reporting provides information needed to evaluate performance of the programs and to articulate the value to key stakeholders

Future State

- Demonstration of the full-value proposition (i.e., leading and lagging indicators)
- More vendors implementing best-practice ROI methodology and putting associated fees at risk
- As the evidence continues to mount and new methods emerge, clients will continue to push for more rigor in how vendors measure and guarantee the impact of their interventions
- More transparent, accurate, timely, user-friendly, and actionable reporting
- Methodology that validates specific intervention approaches or models such as the number of days until outreach for case management

Part 2. Agenda

- Describe the question and challenge
- List our viable options
- Highlight our choice
- Share lessons learned

Two interesting questions:

1. **Speed to contact** – is that important?
   We studied it to determine the effect speed had on results.
2. For those we contacted, **did we have an impact?**

The Business Question

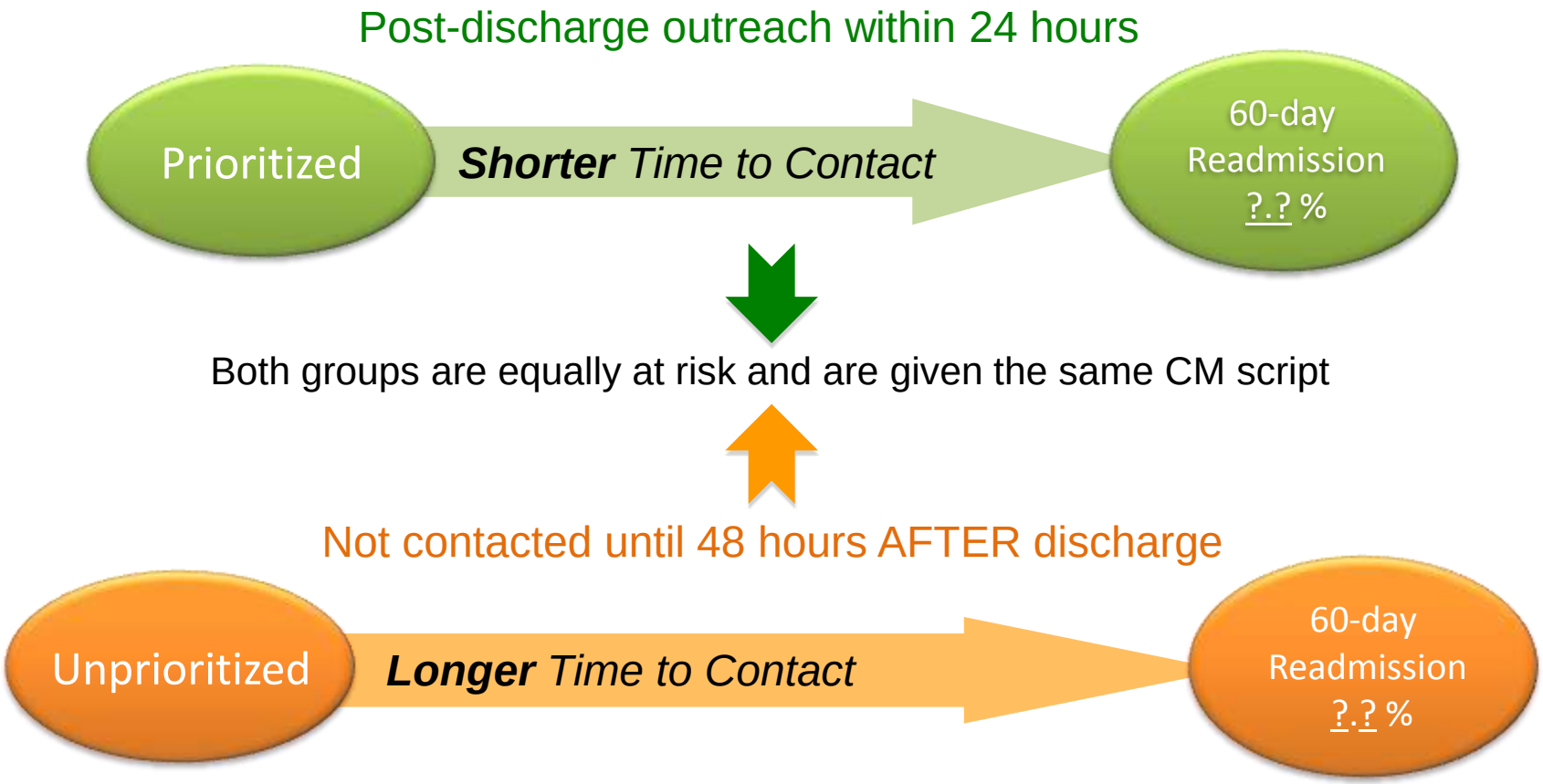
- Our case management program has changed over the years, and there have been questions internally and externally about program effectiveness, so we wanted to know: is post-hospitalization case management still an effective way of reducing urgent, emergent re-admissions?
- **Challenge:** Credibly answer this question without disrupting case management operations

Options

#	Options	Pros	Cons
1	Retrospective, Actuarially adjusted historical control (AKA pre/post)	• Relatively simple • Industry workhorse	• Little insight into drivers of trend • Long time to wait
2	Retrospective, matched case control	• Relatively simple • Better insights into drivers of trend	• Strong but not ironclad • Long time to wait
3	Prospective, randomized managed vs controls	• Clear insights into drivers of trend • Live dashboard available early	• Breaks promise to employer clients
4	Prospective, randomized prioritization of managed	• Clear insights into drivers of trend • Live dashboard available early	• Control group receives intervention

- Based on a review of the options, we proceeded with #4

Pilot Design



Does earlier post-discharge contact reduce members' likelihood of an urgent, emergent readmission?

Participant Waterfall

Eligible for Randomization.
Diagnosed with GI, Heart or LR and had LOS 3=<

1,994 in Prioritized Group:

Matched 1-to-1 with an unprioritized group member based on age, gender, diagnosis, discharge hospital and ERG risk score

48%
Contacted by Case Managers (961/1,994)*

1,994 in Unprioritized Group:

Matched 1-to-1 with a prioritized group member based on age, gender, diagnosis, discharge hospital and ERG risk score

40%
Successfully Contacted by Case Managers (801/1,994)*

*Contacted is defined as initial post-discharge phone contact between CM and customer when health advocacy is provided.

Checking Comparability

- It is critical that the 2 groups are similar and that the only difference between the groups is the time-until-contact.
- We checked:
  - ✓ Demographics
  - ✓ Prior ERG risk and utilization
  - ✓ Benefit plan enrollment
  - ✓ Discharge diagnosis
  - ✓ Hospital facility

Baseline Characteristics of Ideal Patient Pilot Population		
	Prioritized (n=1,994)	Unprioritized (n=1,994)
Demographics	mean (sd) med	mean (sd) med
Male*	51%	51%
Age	50 (16) 54	50 (15) 53
ERG Risk Score*	15 (6) 17	15 (7) 17
Had a Prior Admission Within 12 Months	36%	36%
Number of Admissions	0.82 (1.9) 0	0.82 (1.8) 0.0
Enrolled in PPO	70%	70%
Enrolled in Flex	17%	16%
Methodist Germantown – TN Discharges	2%	2%
St. Francis – FL Discharges	1%	1%
Orlando Hospital – FL Discharges	1%	1%
Diagnosis at Initial Discharge		
% Lower Respiratory	13%	13%
% Heart/Circulatory	39%	39%
% GI	45%	45%
% Endocrine	1%	1%
% Skin	<1%	<1%
Application of Intervention		
Percent Contacted by Case Manager	**48%**	**40%**
Days to Post-discharge Contact	**10 (18) 5**	**15 (27) 6**

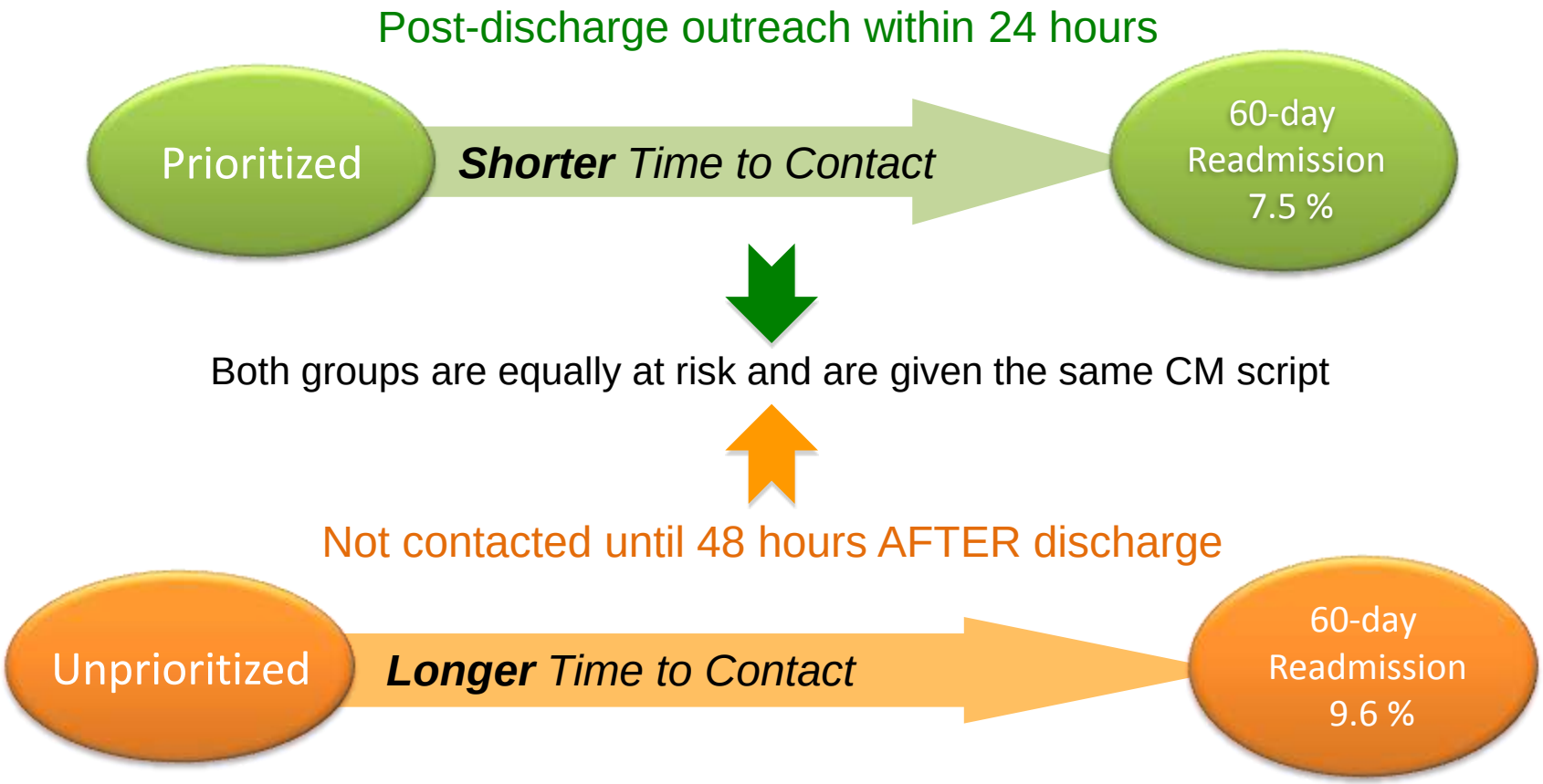
Baseline Characteristics After Matching

A — Demographics, utilization, and health risk are similar

B — No one hospital made up a disproportionate number of discharges

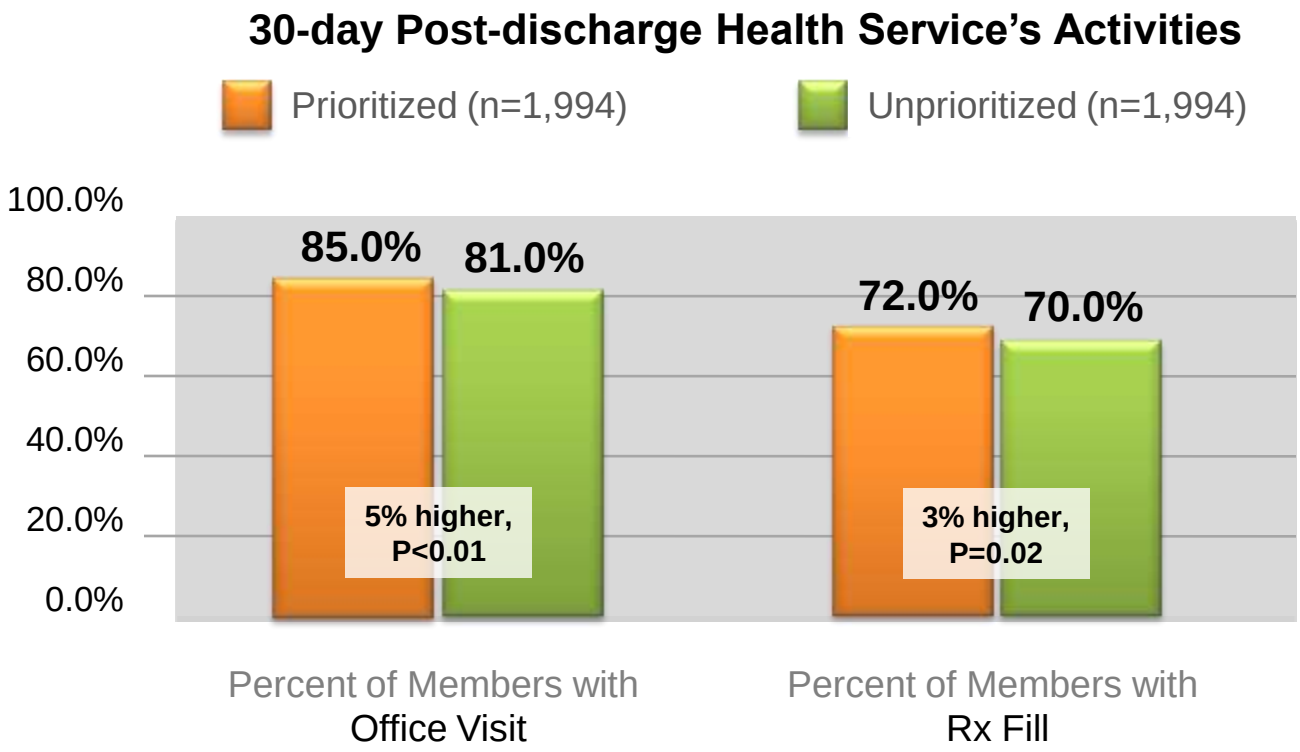
C — Non-parametric, two-sided test ($p<0.01$)

Pilot Design



Does earlier post-discharge contact reduce members' likelihood of an urgent, emergent readmission?

Mechanism of Action



A greater percentage of the prioritized population had an office visit and/or drug fill within 30 days of discharge.

Conclusions

Methodology

- Randomized controlled pilot is viable and does not have to adversely impact case management operations
- Prospective randomization useful to reduce the N and speed up the results
- Stratified random sampling (versus just simple randomization) can also reduce the N and speed up the results