MILLIMAN WHITE PAPER

# Select approaches to measuring effectiveness in healthcare delivery

Katy Reed
Austin Barrington
Kit Powell

Milliman

With the rapid growth in new healthcare treatments, products, and care pathways, developing a robust and thoughtful approach to measure their effectiveness and outcomes may be a valuable component of a healthcare delivery strategy.

An observational outcomes or effectiveness study is an analytic effort designed to estimate the real-world impact of novel healthcare offerings (referred to as “programs” and/or “interventions” throughout this white paper). These studies are important when seeking to evaluate whether a healthcare program is achieving its intended goals and providing value for the resources invested. Additionally, the studies may reveal valuable insight into how to adjust the program to optimize outcomes as well as prioritize multiple programs with limited resources.

Effectiveness studies are used to measure program effect across a wide range of metrics including but not limited to: revenue, cost, utilization, clinical outcomes, and closure of gaps in care. These studies serve as valuable tools for a variety of healthcare organizations such as health tech and digital health companies, healthcare providers, payers, and government programs. This white paper discusses the benefits of a properly executed effectiveness study, a selection of the types of approaches that can be used, and the advantages and limitations of each approach.

## Benefits of effectiveness studies

Effectiveness studies play a critical role in establishing the estimated impact of programs in real-world healthcare settings, considering statistical significance of their impacts compared to random fluctuation. They support a wide range of strategic initiatives, from identifying ways to enhance program design to attracting new patients and partners. These studies vary in methodology, but the most comprehensive approaches offer a number of strategic benefits and challenges.

**Evidence-based decision making**

Effectiveness studies provide empirical evidence quantifying the estimated impact of new treatment programs on financial and clinical metrics, which can also be used for future program improvement strategy.

**Intelligence toward improving patient care and experience**

The findings from these studies can be used to refine and enhance clinical services, contributing to improved patient care and outcomes, as well as the overall patient experience. Similarly, resources expended on low-impact or negative-impact initiatives can be minimized.

**BENEFITS OF EFFECTIVENESS STUDIES**

- Evidence-based decision making
- Intelligence toward improving patient care and experience
- Market differentiation
- Enhanced credibility
- Improved financial understanding

### Market differentiation

Positive effectiveness study results can support differentiation of an organization's care model from competitors, potentially improving patient, employer, and payer loyalty. They can also be used as marketing tools to attract more patients, investors, and partners; depending on the approach used, some organizations opt to publicly report on the findings of these assessments as a way to market and grow offerings. Negative or neutral results can aid in pinpointing portions of the care or pricing model that could be improved.

### Enhanced credibility

With statistically sound methodologies, an effectiveness study can contribute to the reliability of program impact estimates and build trust among healthcare providers, patients, and other stakeholders. For startups seeking additional resources, these studies can help support ongoing investment discussions; additionally, they can help with payer negotiations by demonstrating credible evidence of program outcomes.

### Improved financial understanding

Depending on the level of analysis, these studies can provide the assessments necessary to gauge the impact of a program on healthcare utilization and costs, i.e., its estimated return on investment (ROI). These results can be useful as part of a financial forecast, contracting bid, or performance guarantee. An example case study is described later in this white paper.

Ultimately, effectiveness studies directly measure the benefits of treatments or interventions based on real-world circumstances. They allow organizations to make informed decisions about their care provisions and processes that offer tangible benefits in everyday clinical practice, not just the expected results from highly controlled trials.

## The types, benefits, and challenges of effectiveness study approaches

There are several types of effectiveness studies with various benefits and challenges. Here we focus on the most robust methodologies for studying program effectiveness: randomized controlled trials and observational matched cohort studies.

### Randomized controlled trials (RCTs)

RCTs are considered the "gold standard" for evaluating the effectiveness of programs. An RCT is an approach that aims to reduce certain sources of bias (most notably selection bias, discussed below) when testing the effectiveness of new programs, by randomly assigning research participants into two or more groups: at least one experimental group (which receives the treatment being tested) and a control group (which receives the standard treatment or placebo). A statistical, blinded comparison of outcomes between the two groups is then used to determine the causal effect of these treatments or interventions on patient outcomes.

**FIGURE 1: PROS AND CONS OF RANDOMIZED CONTROLLED TRIALS**

| Pros | Cons |
|---|---|
| Minimizes selection bias through randomized patient selection. | Complex to design and administer. |
| Demonstrates causality between an intervention and its outcome. | Expensive and time-consuming to monitor and analyze. |
| Helps ensure consistency through strict participation protocols. | Introduces ethical considerations due to withholding a potentially beneficial intervention to "control group" participants. |
| Mitigates risk of bias in the interpretation of findings. | Often challenging to monitor and enforce participation criteria. |

An RCT is highly preferrable in that it has strict protocol and standardization requirements that help ensure consistency across participants and mitigate the risk of bias in interpreting results. However, an RCT should not be considered as a monolith that guarantees valid results. RCTs may also have issues with design and/or execution that impact the reproducibility and generalizability of results. In addition, RCTs are expensive and time-consuming to execute and, in certain cases, there can be ethical considerations[1] related to withholding a potentially beneficial intervention to the control group. These ethical issues can be mitigated (through mechanisms like informed consent and nonconventional study designs, for example) though doing so often increases complexity as study objectives become more nuanced.[2]

**FIGURE 2: PROS AND CONS OF OBSERVATIONAL MATCHED COHORT STUDY**

| Pros | Cons |
|---|---|
| Reduces the confounding of variables affecting treatment outcomes. | Limited to the variables used in the propensity score calculation, leaving some potential for bias. |
| Mitigates ethical considerations of RCTs. | Requires sophisticated and sometimes complex statistical techniques and expertise. |
| Applicable to existing datasets in real-world settings. | Often relies on the availability of extensive real-world claims data. |

### Observational matched cohort study

When RCTs are infeasible or impractical due to financial, ethical, or logistical implications, an observational matched cohort study is a research design that is increasingly being used as a reasonable and credible effectiveness study approach. Unlike RCTs, where participants are randomly assigned to treatment or control groups, observational studies simply observe the outcomes of naturally occurring patient exposure differences. These studies aim to estimate the effect of an intervention by comparing the change in outcomes for members who were exposed to a treatment or intervention to the change in outcomes for members who were not exposed to the treatment or intervention. Importantly, well-designed observational studies require that the individuals observed in these two groups are not self-selected, e.g., participants versus nonparticipants, because selection bias is inherent when comparing participants to nonparticipants

Selection bias represents the concern that individuals or groups within a study differ systematically from the target population being studied. This bias can be introduced in a variety of ways, including but not limited to study self-selection (as referred to above), study recruitment methods, provider-related or other decision-maker selection, geographic selection, and other measured and unmeasured clinical differences. Regarding self-selection in observational studies, a primary concern is that qualified members who opt to participate in a program (participants) are different from qualified members who do not opt to participate or decline participation (nonparticipants) on other factors that also influence the outcome (e.g., motivation to change their behavior). To mitigate against selection bias, a study may compare a group of all qualified members who were offered an intervention to a group of similarly qualified members who were not offered the intervention.

## OTHER ALTERNATIVES TO RCTS

### Observed key performance indicator (KPI) without controls models

Modeling that tracks and analyzes the KPI as observed over a set time, without attempting to isolate the effect of specific variables on performance. A straightforward but surface-level view of performance trends and outcomes, but not one that is recommended for public reporting.

### Focus group trial

Organized discussions among a selected group of stakeholders to gather detailed information, opinions, and attitudes about a specific program. A flexible, interactive approach that is relatively easy to administer, but open to bias.

### Hypothetical intervention impact model

Predictive analysis about how the program could theoretically influence certain outcomes based on assumptions drawn from research, expert opinion, and statistical frameworks. Used to estimate the effects of a potential program policy change that has not been implemented or cannot be directly tested.

1. Royall, R.M. (February 1991). Ethics and Statistics in Randomized Clinical Trials. Statist. Sci. 6 (1) 52 - 62. Retrieved November 5, 2024, from https://doi.org/10.1214/ss/1177011934.
2. Nardini, C. (January 16, 2014). The Ethics of Clinical Trials. Ecancermedicalscience;8:387. doi: 10.3332/ecancer.2014.387. PMID: 24482672; PMCID: PMC3894239.

It may also be useful to use a matching process to ensure that the treatment and control groups share similar characteristics, beyond the criteria needed to qualify for the study, that may be related to the study outcomes (e.g., demographics and/or prospective risk scores). When designing these groups for a study (that is, the individuals qualifying for an intervention versus a control group meeting the same qualification criteria), there are a variety of data sources for control groups that may be used. For example, large insurance claims-based data assets can be used to form control group data where otherwise unavailable, unless the qualification criteria are based on metrics that cannot be reasonably approximated using only claims.

Considerations other than bias must also be accounted for in observational matched cohort studies. Fundamentally, these studies require an advanced understanding and application of statistical techniques. Matching procedures can be nuanced and must be carefully designed, taking into account overall study objectives, identification and consideration of each potential confounding variable, and the availability and collection methodology of the matching variables selected. Study results can be highly technical and potentially difficult to communicate to those without a strong statistical background. It is also important to consider that a statistically significant result from a well-designed study may not translate to a clinically applicable outcome. Clinical relevance and the biological plausibility of the intended study objectives are important to consider when evaluating the results yielded by an observational study design.

Additionally, the sample size for both groups should be adequate to detect statistically significant differences in outcomes assumed to be generated by the treatment. Due to varying levels of volatility in outcomes, the sample size may be sufficient to measure statistically significant differences in some metrics but not others. Primary and secondary outcomes measures, such as mortality, cost of care, hospital readmissions, and avoidable emergency room (ER) visits may be identified as a way to gauge the program's effectiveness.

To promote transparency and increase trust, organizations can include any statistically significant negative effects as well. It is critical to demonstrate the full picture of the program's effect; for example, a program may reduce ER visits in exchange for increased primary care office visits or stronger medication adherence at the cost of additional pharmacy spend.

## Other considerations for conducting effectiveness studies

**Program-specific costs**

There are a wide range of costs associated with the program whose effectiveness is being measured that should be considered in overall program impact estimates, including:

- Expenses associated with the provision of treatment or intervention
- Program setup expenses
- Costs of partnership contracting
- Implications of provider/member abrasion

**Opportunity costs**

- Regardless of the financial viability or cost savings of certain programs, patients, providers, and payers have come to expect certain service or care management offerings
- Loss of member participation in an organization's other care programs

**Statistical implications**

- **Causal interference**: Depending on the methodology, study results can be confounded by interactions with other interventions, solutions, and events. This can complicate organizations' ability to determine true program impacts.
- **Reversion to the mean**: This is a phenomenon that occurs when a program impact is extreme on its first measurement due to measuring the difference from a period when an acute event occurred, but subsequent measurements tend to be closer to the population average. This is a common concern for effectiveness studies and should be carefully considered and accounted for. Matching the study and control populations on baseline levels of key outcome variables may help avoid reversion to the mean.

### Data availability

- Some healthcare organizations are in the early stages of building their programs and have limited available data, only have data on the self-selected treatment group, or only have data within a few healthcare service categories.
- In these instances, the focus should be on either:
  - Developing a framework for program performance estimation that can be evaluated once more patient data becomes available.
  - Conducting a benchmarking exercise for the target population, for those organizations with at least some relevant data, comparing it to a robust set of claims-based benchmarks.
- In some situations, it may be possible to perform an outcomes study by matching and comparing to a control group based on market benchmark data, as discussed above.

### Other factors

Force majeure events, such as a pandemic or other external forces, can present unavoidable anomalies in the result findings.

## Case study

### USING AN OBSERVATIONAL MATCHED COHORT STUDY TO DETERMINE A PROGRAM'S ROI

### Client

A national virtual care company (called "Wellness Health" for purposes of this case study) focused on patient wellness, inclusive of chronic conditions like obesity, diabetes, and hypertension.

### Objective

To conduct an effectiveness study measuring the change in spending and utilization for a population under Wellness Health's care model compared to a control group comprised of a mix of employer group populations not managed by Wellness Health. Results are ultimately intended to support an ROI calculation for the Wellness Health program.

### Approach

Milliman calculated the change in spending and utilization for Wellness Health's client population from a baseline period (i.e., the period prior to becoming a Wellness Health client) to the post period (i.e., the period after becoming a Wellness Health client). These changes were compared to a control group comprised of employer group claims sourced from a reference dataset, generally using interventions other than Wellness Health. The Wellness Health analysis was based on all members meeting the qualification criteria for being offered Wellness Health's intervention (i.e., it was not limited to only those using the intervention). The control group was also limited to members meeting the same qualification criteria but not offered the intervention. Each member from the Wellness Health population was matched with a control group member based on additional criteria expected to be related to study outcomes (e.g., risk score). Additionally, our analysis relied on other clinical and financial results from Wellness Health's program to estimate potential savings and ultimately calculate ROI relative to program fees.

### Result

Milliman's analysis focused on patient members who met qualification criteria for Wellness Health's care program for at least six months to ensure adequate time for outcomes to emerge. We tracked program costs for acute non-maternity inpatient (IP) admits, ER visits, specialty visits, primary care visits, and prescription drug claims incurred across a two-and-a-half-year period. The study identified all outcomes, focusing on statistically significant improvements in both utilization and per member per month (PMPM) cost savings. Key findings included:

- Allowed PMPM for the Wellness Health client(s) had a smaller increase than the control group in allowed PMPM from the base period to the study period. The difference was statistically significant.
  - The difference was primarily driven by a larger reduction in the rate of acute non-maternity inpatient admissions for Wellness Health compared to the control group.
- The rate of ambulatory care sensitive condition inpatient admissions dropped more for Wellness Health clients than for the control group.

- The rate of primary care office and specialty visits increased more for Wellness Health clients than for the control group. This increase in professional services was an intended effect of the program but did result in higher spending in these categories.
- The rate of diabetes-related complications for members with diabetes during the baseline period increased less for Wellness Health clients than it did for matched members from the control group.

## Conclusion

Effectiveness studies are evidence-based methods for healthcare organizations to monitor and fine-tune innovative patient care and healthcare delivery processes. They may play a significant role in creating new ways to approach patient care and ensuring that resources are effectively used for care management. With proper design and execution, effectiveness studies can highlight the clinical and financial impact of these programs and identify opportunities to improve financial performance and clinical care. Study findings are typically compiled into a report to share with relevant stakeholders, including healthcare professionals, policymakers, patients, prospective partners, and potentially the public, to educate the community on new treatment and intervention opportunities.



**CONTACT**

Katy Reed
katy.reed@milliman.com

Austin Barrington
austin.barrington@milliman.com

Kit Powell
kit.powell@milliman.com