Importance of Study Design in Causal Inference in Policy Analyses Using Real World Data

Dennis Ross-Degnan, ScD

Many real-world natural experiments intended to improve diabetes prevention or treatment are implemented in contexts that are not conducive to randomization, so causal inference about their impacts is methodologically challenging. Many of these natural experiments involve changes in policies or programs within health delivery or insurance systems, in which an intervention is implemented either in an entire system or within non-randomly selected subgroups. In such situations, routine data collected as part of health care delivery or payment can be used to design stronger evaluations. Interrupted time series with comparison series, one of the strongest quasi-experimental study designs, depends on the availability of routine data from before and after an intervention for measurement of changes in outcomes. Trends in outcome data prior to the intervention permit estimation of a counterfactual, or what would likely have occurred in the absence of the intervention. In addition, routine baseline data can be used in several ways to identify matched comparison groups in order to minimize selection effects, balance differences between intervention and comparison groups, and increase the strength of causal inference. This presentation will provide an overview of these issues and illustrate their application in a study conducted as part of the Natural Experiments in Diabetes Translation (Next-D) initiative examining the impact of high-deductible health plans on out-of-pocket costs, utilization of care, and clinical outcomes among patients with diabetes.
Importance of Study Design in Causal Inference in Policy Analyses Using Real World Data

Dennis Ross-Degnan, ScD
Harvard Medical School and Harvard Pilgrim Health Care Institute
Presenter Disclosure Information

In compliance with the accrediting board policies, the American Diabetes Association requires the following disclosure to the participants:

Dennis Ross-Degnan, ScD

Disclosed no conflict of interest.
Overview

➢ Opportunities and challenges of real-world data
➢ Quasi-experimental designs for causal inference
  ■ Scope of policy implementation
  ■ Interrupted time series with comparison series
➢ Matching
Considerations for Policy Evaluations

- Retrospective vs. prospective
  - Degree of control over implementation

- Structure of implementation
  - Individuals
  - Clusters (clinics, worksites, communities, etc.)
  - Whole system at once

- Type and timing of expected effects
- Availability of routine data
Types of Routine Health System Data

- Clinical data (e.g., EMR)
  - Generated during process of care
  - Rich clinical detail but difficult to standardize

- Administrative data (e.g., insurance claims)
  - Derived from payment or supply systems
  - More standardized but less detail

- Medical technologies (e.g., IQVIA)
  - Drugs or devices approved, sold, or used

- Regular standardized surveys
  - BRFSS, MCBS
Example: FDA Sentinel Distributed Database

- 19 data partners
- Common data model
- Well-defined person-time
  - Medical & pharmacy coverage
  - 66.9m current members
  - 292.5m cumulative members 2000-2017
  - 14.4b pharmacy fills
  - 13.3b medical encounters
  - 45.6m members with 1+ lab result

https://www.sentinelinitiative.org/
Example: Data Quality Assurance in Sentinel

Data Quality Review and Characterization Programs v4.1.0

**Project Title**: Data Quality Review and Characterization Programs v4.1.0

**Description**: The Sentinel Data Quality Review and Characterization Center (SOCC) for data quality review and characterization. To create the SDD, each Data Partner transforms their Data Model (SDMI) format. The SOCC created a set of data quality review and characterization standards that the SDD data meets expectations needed for the validation and improvement of data quality.

**Link**: Sentinel Data Quality Review and Characterization Center

---

**Table 3. Definitions and examples of enrollment date range relationships by PatID**

<table>
<thead>
<tr>
<th>Definition</th>
<th>Illustration</th>
<th>Enr_Start</th>
<th>Enr_End</th>
<th>DMQA Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disjoint: No conflict or overlap</td>
<td></td>
<td>01/01/2015</td>
<td>03/31/2015</td>
<td>Pass</td>
</tr>
<tr>
<td>Consecutive: Two date ranges are consecutive</td>
<td></td>
<td>01/01/2015</td>
<td>03/31/2015</td>
<td>Warn</td>
</tr>
<tr>
<td>Overlap: Two date ranges overlap over a range</td>
<td></td>
<td>01/01/2015</td>
<td>03/31/2015</td>
<td>Fail</td>
</tr>
<tr>
<td>Duplicate: Two date ranges are identical</td>
<td></td>
<td>01/01/2015</td>
<td>03/31/2015</td>
<td>Fail</td>
</tr>
<tr>
<td>Subset: One date range is a subset of another</td>
<td></td>
<td>01/01/2015</td>
<td>03/31/2015</td>
<td>Fail</td>
</tr>
</tbody>
</table>

---

https://www.sentinelinitiative.org/sentinel/surveillance-tools/validations-lit-review
RCTs: Gold Standard in Study Design but Rare in Natural Experiments or Policy Evaluations

Randomized Controlled Trial

- Intervention Group
- Control Group

\[ X = \text{policy intervention} \quad O_t = \text{Measurement at time } t \]

- Baseline comparability maximized by randomization
- Typically Difference in Difference (DiD) analysis approach
Feasibility of Randomization in Policy Evaluations

- Structural constraints
  - Impossible to prevent exposure

- Political constraints
  - Assignment to controls unacceptable
  - Desire to roll out quickly

- Ethical constraints
  - Withholding beneficial program

**Best practice:** Retain as many elements of randomization in study design as possible and test for group differences
"Quasi-Experimental" Study Design Is Typical in Natural Experiments or Policy Evaluations

Non-Random Control Group Design

\[ O_1 \quad X \quad O_2 \]

\[ O_1 \quad O_2 \]

\( X = \) policy intervention  \( O_t = \) Measurement at time \( t \)

- Difference in Difference (DiD) analysis approach
- Baseline comparability is a key threat to validity
Interventions in Subgroups: Parallel Cluster Design

- Ideally, match and randomize clusters to group
- Collect key group-level data for adjustment
- Use difference in difference or time series analytic framework

Clusters exposed to intervention at same time

Adapted from Hemming K et al. BMJ 2015
Interventions in Subgroups: Stepped Wedge Design

- Can address political or logistical constraints
- Ideally, clusters randomized to start times
- Analyses control for group differences & effect of time

Clusters exposed to intervention in planned way at different times

Adapted from Hemming K et al. BMJ 2015
Other Analysis Issues Important to Policy Evaluations

- **Uptake**
  - % offered intervention who participate

- **Fidelity**
  - Same intervention in all groups, all individuals

- **Magnitude of exposure**
  - % of planned intervention experienced

- **Chronological timing**
  - When exposure occurs during intervention period

- **Nature of observed change**
  - Statistical significance vs. clinical importance
Interrupted Time Series: Robust Quasi-Experimental Design for Policy Interventions

![Diagram showing time series analysis with experimental and comparison groups with time points O1 to O6]

- Baseline trend provides built-in “control”
- Stronger if includes a comparison group
- Evidence about possible external events

- Typically uses segmented regression analysis approach at aggregate or individual levels
- Can also use segmented survival models (time until outcome before and after intervention)
Parameters Estimated by Interrupted Time Series Segmented Regression Model

Assumption: Baseline trend reflects what would have happened without intervention (counterfactual)
Using Routine Data to Improve Comparability Between Non-random Treatment and Comparison Groups

- Choice of comparison
  - “Comparable” unexposed group
  - Outcomes in exposed group not expected to change

- Matching
  - Match on observed characteristics
  - Potential bias due to unobserved factors
  - Stronger if groups matched on baseline trend

- Use rich set of covariates to adjust for selection bias
Propensity Score Matching

- Estimate likelihood of being a case using observed baseline characteristics
- Exclude cases and controls outside region of common support
- Restrict analysis to matched members or use inverse propensity weights in analysis

https://www.summitllc.us/propensity-score-matching
Coarsened Exact Matching (A Type of Monotonic Imbalance Bounding)

- Mimics sampling process of a stratified RCT
- Matching characteristics selected based on knowledge about likely causality
- Cases and controls with same bin signature are matched and then weighted in analysis to maintain population balance
NEXT-D Study of the Impacts of Transitions to High-Deductible Health Plans among Diabetes Patients

- Study period: 2003-2014, rolling cohort
- Participants: commercially-insured members in Optum health insurance claims dataset
  - Age 12-64 with predominantly type 2 diabetes mellitus
  - Continuously enrolled for ≥2 years
  - HDHP members (≥$1000) matched to controls who had ≤$500 deductibles in all years
NEXT-D Study Examining Effects of Employer Selection of Employees into High Deductible Health Plans
Key Strategies for Maximizing Causal Inference in Next-D HDHP Study

- Include only “full-replacement” employers
- 2-level (employer and member) matching on:
  - Baseline demographic and clinical characteristics
  - Baseline outcome and cost trends
  - Hybrid matching methods
- Appropriate statistical models
  - Segmented regression
  - DiD models
  - Survival models
## Matching Level 1: Employer Selection

- Calculate propensity of employer to switch to HDHP based on: index month, size, % women, age strata, % low SES, race strata, region, median costs/ cost trend, average ACG score, outpatient copay
- Stratify member-level propensity score matching by employer-level propensity score quartile

<table>
<thead>
<tr>
<th>Employer propensity</th>
<th>High-deductible</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mid-high</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low-mid</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Matching Level 2: Member Level

- Patient-level propensity model within employer-level propensity quartile
  - Patient demographic characteristics
  - Baseline OOP costs

- Coarsened exact matching
  - Employer size, region, quarterly outcome trends, employer-level propensity, patient-level propensity
Additional Related Issues in Next-D HDHP Study Design

- Selection of controls
  - Best control for HDHP member might be a low-deductible member who switches in a later year
  - Available pool of employers choosing traditional plans becomes increasingly less comparable

- Possible approaches
  - Randomly assign member to only one time period in which they qualify (either HDHP switch or one control period)
  - Could conduct annual analyses of HDHP switches and then conduct meta-analysis of findings
Summary

- Routine data offer potential for robust evaluation of natural experiments
- Quasi-experimental study designs can accommodate interventions rolled out in different ways in a system
- Robust interrupted time series (ITS) designs depend on availability of routine data
  - Estimate counterfactual
  - Balance intervention and comparison groups
- Next-D study of high deductible health plans illustrates a hybrid application of these methods