Title
Evaluating diabetes health policies using natural experiments: The natural experiments for translation in diabetes study

Permalink
https://escholarship.org/uc/item/05v9k9mf

Journal
American Journal of Preventive Medicine, 48(6)

ISSN
0749-3797

Authors
Ackermann, RT
Kenrik Duru, O
Albu, JB
et al.

Publication Date
2015-06-01

DOI
10.1016/j.amepre.2014.12.010

Peer reviewed
The high prevalence and costs of type 2 diabetes makes it a rapidly evolving focus of policy action. Health systems, employers, community organizations, and public agencies have increasingly looked to translate the benefits of promising research interventions into innovative policies intended to prevent or control diabetes. Though guided by research, these health policies provide no guarantee of effectiveness and may have opportunity costs or unintended consequences. Natural experiments use pragmatic and available data sources to compare specific policies to other policy alternatives or predictions of what would likely have happened in the absence of any intervention. The Natural Experiments for Translation in Diabetes (NEXT-D) Study is a network of academic, community, industry, and policy partners, collaborating to advance the methods and practice of natural experimental research, with a shared aim of identifying and prioritizing the best policies to prevent and control diabetes. This manuscript describes the NEXT-D Study group’s multi-sector natural experiments in areas of diabetes prevention or control as case examples to illustrate the selection, design, analysis, and challenges inherent to natural experimental study approaches to inform development or evaluation of health policies.


Introduction

About 29 million Americans have diabetes, and another 86 million have prediabetes, placing them at high risk for progression to type 2 diabetes as well as additional health problems, lower quality of life, and higher medical and non-medical costs. Recent increases in the burden of type 2 diabetes have encouraged rapid policy action by both public agencies and private organizations. Optimistically, many agencies and organizations have attempted to implement interventions derived from seminal research studies, often making adaptations to enhance access, feasibility, and scalability. For example, recent policy action established a national recognition program at CDC to define standards for how community organizations should offer lifestyle-based diabetes prevention programs consistent with the Diabetes Prevention Program (DPP) clinical trial. Similarly, value-based health insurance designs are being implemented by some health systems to reduce or eliminate out-of-pocket costs for preventive services or medications that have strong evidence of benefit.

Although the use of past research to inform diabetes-related health policies is encouraging, there is no guarantee that an intervention derived from a successful clinical trial will be effective or cost effective under real-life circumstances. For example, clinical trials have demonstrated that intensive interventions to support diabetes self-management improve glycemic control...
and cardiovascular risk factors, but policies by health
payers to fund disease management interventions in
diabetes and many other areas have yielded mixed
results. Similarly, duplicative prescribing and poly-
pharmacy can increase adverse medication events, but
policies that cap the number of medications covered by
state-based public assistance programs have been linked
to lower use of recommended glucose-lowering agents,
increased hospitalization, and higher public expendi-
tures. Unintended policy consequences, such as the
potential to exacerbate rather than reduce health dispar-
ities, are also an important target for health policy
research. Ideally, when a health policy is enacted,
going strong research would establish whether the
benefits outweigh harms, and if costs are reasonable for society. Unfortunately, many health policies still
remain unevaluated or are assessed using research
designs that are vulnerable to considerable bias and are
thus potentially misleading.

Selecting Valid Research Designs for a Policy Intervention

Although randomized trials are the gold standard for clinical
efficacy research, they are often impractical or
inappropriate to test policy effectiveness. Conversely,
natural experiments use pragmatic research designs and
readily available data sources to evaluate and compare a
new or existing policy to other policy alternatives or
predictions of what may have happened in the absence of
any intervention. Natural experiments have been
applied to numerous disciplines, including political
science, education, psychology, economics, history, and
sociology. Within the context of health research, natural experiments have been particularly influential in
policies related to indoor tobacco laws, Medicaid drug
reimbursement policies, and provision of vouchers for
health and social services.

For each policy intervention, there is typically more
than one choice for a natural experimental study design,
with the goal of implementing the strongest design
feasible given the study question, implementation con-
text, and data sources available. Table 1 provides
examples of several common design options, including
how adequately they address common threats to validity.

Individually randomized trials are often difficult to
implement but should still be considered in the evalua-
tion of public or organizational policies. When
individual randomization is not feasible, adaptations to
the RCT design, such as cluster-randomized parallel-
group or stepped-wedge intervention trials can provide
strong alternatives by allocating groups to interventions
at a provider, clinical practice, organizational, com-
munity, or regional level. Importantly, such designs
must carefully consider decreases in the effective sample
size, the possibility of imbalance among treatment groups
when randomized clusters are small, and the need for
special statistical models to account for correlation of
outcomes within clusters.

If no form of randomization is feasible or appropriate,
the interrupted time series design (with and without
control series), and to a lesser extent, pre–post studies
with a strongly matched comparison group provide
alternatives that can control for most common threats
to validity. These designs must carefully consider
common threats such as bias and confounding and the
role of complementary statistical approaches to address
these threats during the analysis. Examples of such
approaches include multivariable adjustment or propen-
sity matching with or without adjustment. Another
important strategy to reduce uncertainty for a particular
study design is to replicate findings by repeating the
analysis in another context (i.e., external validation) or
through sensitivity analyses that examine whether similar
results are achieved using alternative comparison groups
or study design options. Although used in many past
policy evaluations, simple cross-sectional or pre–post
designs without a comparison group are generally
considered weak for evaluating intervention effective-
ness. As reflected in Table 1, these designs are attractive
because of their simplicity but can be misleading, as they
are unable to control for common biases such as history
and selection.

The Natural Experiments for Translation in Diabetes Study

The Natural Experiments for Translation in Diabetes (NEXT-D) study is a dynamic collaboration among five
academic centers, numerous organizational and policy
partners, CDC, and the National Institute of Digestive
and Diabetes and Kidney Diseases. NEXT-D centers
include the Harvard Pilgrim Health Care Institute and
Harvard Medical School, Kaiser Permanente Northern
California, Northwestern University, New York St.
Luke’s-Roosevelt Hospital, and the University of Califor-
fornia, Los Angeles (UCLA). Each of these centers has an
evaluation partnership with one or more non-academic
terries to evaluate a public or organizational policy area
using a natural experimental study design that is prac-
tical, methodologically rigorous, and suited to the needs
and interests of each policy stakeholder. By including
policy decision makers in the network, NEXT-D seeks to
address the most relevant gaps in the scientific research
literature while developing data collection, analysis,
interpretation, and communication strategies to improve

www.ajpmonline.org
policy decision making. Partnering directly with policy stakeholders also increases the likelihood that research findings will be translated into real-life policies and practice.28

The NEXT-D study group adopted a conceptual framework to portray the relationships among NEXT-D projects. This framework was derived from existing social-ecologic heuristics29,30 and considers how each natural experiment might be affected by other interventions or contextual influences across geographic, organizational, and economic sectors. Using the framework, the study group considers how best to design natural experiments to evaluate how policies are adopted and implemented, whom they reach, and whether they are effective and sustainable. The goal for each natural experiment is a rigorously designed quasi-experimental study that carefully considers the individual and synergistic effects of influences at multiple policy levels.15,16

These interactive effects are depicted in Figure 1 and explored briefly in the text that follows.

A central focus of the framework is to understand the intended or unintended impacts of policies enacted at any level. These impacts can be related to diabetes incidence, progression, or health and economic outcomes, as well as the distribution of those outcomes (e.g., reducing or increasing health disparities) throughout the population. Policies impact these key outcomes either directly or indirectly through effects that ultimately must culminate in successful changes in the behaviors of individuals or others such as healthcare providers, public health professionals, community health workers, or peers.

---

**Table 1.** Hierarchy of Natural Experimental Study Designs and their Capacity to Address Threats to Internal Validity

<table>
<thead>
<tr>
<th>Study design</th>
<th>Strength of design</th>
<th>Key strengths and weaknesses in ability to control for threats to validity†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Randomized controlled trial (experiment) Treatment: Randomize O₁ X O₂ Comparison: Randomize O₁ − O₂</td>
<td>Strong</td>
<td>• Gold standard of evidence • Controls for major validity threats* (history; maturation; selection; instrumentation) • Frequently not feasible to randomize individuals • Randomization of groups (e.g., clinics) is an option but can introduce bias; can lead to type 2 error (inadequate sample size); and requires statistical models for intercorrelation of data within groups • Often requires exclusion criteria that limits external generalizability</td>
</tr>
<tr>
<td>Interrupted time series with a comparison series† Treatment: O₁ O₂ O₃ …… O₁₂ X O₁₃ O₁₄ O₁₅ …… O₂₆ Comparison: O₁ O₂ O₃ …… O₁₂ − O₁₃ O₁₄ O₁₅ …… O₂₆</td>
<td>• Multiple baseline observations allows control for major validity threats* (history; maturation; selection; instrumentation)</td>
<td></td>
</tr>
<tr>
<td>Single interrupted time series‡ Treatment: O₁ O₂ O₃ …… O₁₂ X O₁₃ O₁₄ O₁₅ …… O₂₆</td>
<td>• Baseline trends allow visible effects/control for most biases‡</td>
<td></td>
</tr>
<tr>
<td>Pre-post with comparison group Treatment: O₁ X O₂ Comparison: O₁ − O₂</td>
<td>• Comparability of baseline trend unknown • Strong risk of history, maturation, selection bias, regression to mean*</td>
<td></td>
</tr>
<tr>
<td>Pre-post without comparison Treatment: O₁ X O₂</td>
<td>• Single observation before and after with no baseline trends • Weak design for assessing effectiveness • Strong risk of history, maturation, instrumentation bias, regression to mean*</td>
<td></td>
</tr>
<tr>
<td>Cross-sectional designs Treatment: X O₂ Comparison: O₁</td>
<td>Weak</td>
<td>• Weak design for assessing effectiveness • Simple correlation, no baseline, no measure of change • Very strong risk for most major threats to validity*</td>
</tr>
</tbody>
</table>

Note: Readers are referred to Shadish, Cook, and Campbell (17) for a more extensive list of threats to validity.
X: represents intervention.
O: represents outcome measurement point.
†Threats to Validity: History: an event occurring between pre- and post-tests when the event is not the intervention of research interest. Maturation: Subjects growing older, wiser, healthier (or sicker), etc. between pre- and post-test, & changes are not the intervention of interest. Instrumentation: a change in the measuring instrument (either humans or test). Selection: pre-intervention differences between people in one experimental group vs another; Selection biases can interact with maturation, history, and instrumentation biases, in which participants or groups selected as part of the intervention have different levels of maturation, history, or instrumentation than those in the control condition. Statistical regression to the mean: the tendency that when a measurement is extreme a second measurement will be closer to the mean.
‡Time Series: In practice, a minimum of 12-24 observations (e.g., monthly rates) are usually required to measure discontinuities in outcomes while controlling for baseline trends (the counterfactual).
and services to support those behaviors. Examples include the design of new coverage benefits, special programs to support prevention and care management, and restructuring payments to incentivize behaviors that lead to better health and health care.\textsuperscript{33–36}

Individuals and populations also regularly interact with community resources from religious, civic, socially driven, or commercial entities. Community resources are a crucial influence on health-related behaviors and health status. For example, faith-based organizations play an increasing role in social support, self-management, and even clinical services for people with diabetes or prediabetes.\textsuperscript{37}

Finally, the public policy environment may play a strong role in influencing health through actions such as the development of laws and regulations to shape the economic, physical, social, and cultural environments in which people live, work, and recreate. Public policies may result in both intended and unintended consequences through direct regulatory impacts or via interactions with other interventions that are deployed simultaneously across other levels.

Examples of Existing Natural Experiments Supported by NEXT-D

Diverse study designs are employed in NEXT-D to determine the impacts of policies enacted at multiple levels. Below, we describe very briefly the existing NEXT-D projects as case examples for the selection, design, analysis, and challenges of using robust natural experimental research designs to inform the evolution of policies in public health and healthcare contexts.

Example 1: Kaiser Permanente Northern California

Telephonic coaching to encourage healthy lifestyle choices is a population-based approach to wellness and diabetes prevention that is being explored by many health plans and purchasers. As gestational diabetes mellitus (GDM) has increased in incidence, more health plans are also implementing strategies to encourage glucose screening for earlier identification and treatment of both type 2 diabetes and prediabetes postpartum. Kaiser Permanente Northern California (KPNC) has partnered with a large healthcare purchaser to assess the impact of providing population-based telephonic wellness coaching, as well as a postpartum diabetes risk-reduction intervention targeting women who have had GDM.\textsuperscript{38} These studies use multiple data sources and interrupted time series with control series, as well as
as pre–post longitudinal control group designs to assess patient satisfaction, perceptions of the programs, BMI, smoking-cessation rates, hemoglobin A1c and fasting plasma glucose levels, and onset of diabetes in populations. The studies will also examine potential impacts on health care and health disparities, as well as the efficiency and programmatic costs to help stakeholders understand the barriers and facilitators to expanding these programs within the health plan if they are found to be effective.

Example #2: University of California, Los Angeles
At UCLA, investigators are partnering with the Innovations Group at United Healthcare, the largest commercial health insurance company in the U.S., to evaluate the health and economic impacts of a new diabetes-specific health insurance benefits package. The Diabetes Health Plan (DHP), which is now being offered by several dozen large- and medium-sized employers throughout the U.S., providing chronic condition management, financial incentives, and improved access to value-based services (e.g., established preventive medications, treatments, and services at very low or no cost) for patients with prediabetes and diabetes. The study uses a longitudinal quasi-experimental design with a propensity-matched comparison group of employers who do not offer the DHP to evaluate the effect on adherence to evidence-based medications, use of emergency department or hospital inpatient services, rates of progression from prediabetes to diabetes, and total costs of care. UCLA meets regularly with United’s Innovations Group to ensure that the study’s findings will help United make decisions about further DHP refinements and its replication. This study also will determine what employers should expect when implementing this type of disease-specific plan in real-world settings, where program uptake and employee engagement can vary.

Example #3: St. Luke’s-Roosevelt Health System
Investigators at St. Luke’s-Roosevelt are collaborating with six primary care clinics in New York City to study pragmatic interventions at the health system/provider level that address challenges to early diabetes detection and prevention in primary care practice. Interventions use chronic care model principles and involve electronic health record modifications and integrated health care systems/community linkages that are rolled out across each clinic in a stepped fashion, allowing for comparisons over time as well as across implementing and non-implementing clinics. Using electronic health record data sources and a time series study design, this study will compare rates of test completion and detection of diabetes or prediabetes. A pre–post with propensity-matched comparison group study design will be used to determine the effects of the enhanced management interventions on clinical intermediate outcomes (e.g., hemoglobin A1c), for about 1,500 adults who are identified as having prediabetes or type 2 diabetes who receive care at a practice location implementing (versus not implementing) the enhanced interventions.80 Evaluation results will inform the continuing implementation and wider replication of enhanced management intervention strategies to other clinics in the network.

Example #4: Harvard Pilgrim Health Care Institute
More than 40% of workers now have high-deductible health plans (HDHPs),14 and recent policies are expected to spur further increases because of coverage mandates, incentives for plans with lower actuarial value (non-“Cadillac” plans), and greater upfront affordability.6 It is not yet known whether HDHPs will have their intended effect of creating “activated healthcare consumers” who change their behaviors and seek higher-value services to lower their own health expenditures or, rather, who simply reduce utilization, even for care that is considered essential. Harvard partnered with OptumInsight Life Sciences, Inc., to develop a national data set involving a 13-year rolling sample (2001–2013) of 1.3 million adults with diabetes from all 50 states who are health plan members of an HDHP or more traditional plans (controls). The study is employing an interrupted time series with comparison series study design to assess the impact of HDHPs on diabetes control, risk factors, preventive care, and health outcomes such as emergency department visits and preventable hospitalizations; utilization, including outpatient visits; and costs.43–45 Results of this work could be used to design health plan features that promote more equitable high-quality care and better outcomes among patients with diabetes.

Example #5: Northwestern University Feinberg School of Medicine
To address the growing incidence of type 2 diabetes in the U.S., UnitedHealth Group (UHG), Y-USA, and CDC have partnered to bring a group-based adaptation of the DPP lifestyle intervention to a national scale. This initiative combines actions at the purchaser, community resource, and public policy levels and leverages efforts by CDC and Y-USA to build the capacity of community
organizations nationwide to deliver a standardized lifestyle intervention program derived from the DPP. These interventions ultimately target several million eligible adults with prediabetes in almost 50 large geographic markets throughout the U.S. The study will use an interrupted time series with propensity-matched comparison series design to determine the relative effects of UHG outreach/engagement activities and local efforts by YMCAs to enroll participants within different demographic subgroups. By integrating additional data from Y-USA, the project will examine how implementation varies across regions, and using national UHG claims data sources, effectiveness will be evaluated through changes in prescription expenditures; obesity-related risk factors such as total cholesterol and blood glucose; healthcare utilization; and total healthcare expenditures.46

Discussion

New public and organizational policies are being enacted every day to improve health and behavior, but we continually miss important opportunities to evaluate those policies using valid research designs. Natural experiments are forms of pragmatic research that use readily available data sources to evaluate and compare the intended and unintended outcomes of new or existing policies, rather than simply assume they are a wise use of limited resources. No study design is without limitation; thus, the process of selecting a natural experimental approach must carefully consider the trade-offs between feasibility and scientific validity.

Ideally, the design of a natural experiment should be deliberated within the context of a partnership that includes capable evaluators, policy decision-makers, and other stakeholders. Together, these partners must consider important choices and ultimately select a design that will be feasible to implement while providing organizational leaders and other individuals with valid and actionable outcomes, such as who can be reached by new initiatives; whether all target groups have equitable access; if benefits vary across different vulnerable population subgroups; if there are unintended consequences; and the amount and types of expenditures that are needed to implement those initiatives “at scale.” Such research can be considered one badly needed form of implementation science, which seeks to elucidate how best to replicate and scale evidence-based programs across diverse populations and differing settings. Partners may also collaborate to devise creative strategies for overcoming common pragmatic research challenges, such as the paucity of information available in administrative data systems to determine the characteristics of target populations (e.g., race/ethnicity or SES). Overcoming these challenges might require linkages with other data sources, such as neighborhood-level data regarding socioeconomic indicators, or the use of individual surname analysis done prior to de-identifying the data for analysis. These are but a few of the examples of the common challenges that can be solved collaboratively through research partnerships, or even networks of partnerships, as is the case in NEXT-D.

With an eye on quantifying the intended and unintended impacts of population-targeted diabetes prevention and control initiatives, the NEXT-D network is positioned to provide empirical and pragmatic information to guide the ongoing evolution of policies and practices to improve diabetes care and prevention. Through the expanded use of strong natural experimental methods, NEXT-D evaluates public health and healthcare policies while those policies are being implemented “unperturbed,” in a fashion that truly reflects health impacts for real-world populations and at natural levels of “exposure” (or “dosage”). Although the individual natural experimental methods adopted by NEXT-D sites are not entirely novel, the support of a network of collaborating academic, health system, public health, commercial, and community stakeholders to advance the evolution of health policies in chronic disease care and prevention is quite unique. In addition, although NEXT-D focuses on diabetes, its collaborative approach for conceptualizing the policy environment, engaging policy partners in the design and application of natural experiments, and transferring findings back to key stakeholders in ways that promote continual progress toward more efficient and effective population health initiatives is not only relevant but of paramount importance to the entire policy community.

The Natural Experiments for Translation in Diabetes (NEXT-D) Study is funded jointly by CDC and the NIH/ National Institute of Diabetes and Digestive and Kidney Diseases under grant numbers U58-DP002717, DP002718, DP002719, DP002721, and DP002722. The findings and conclusions in this report are those of the authors and do not necessarily represent the views of CDC or the NIH.

The NEXT-D Study Group: Harvard Medical School and Harvard Pilgrim Healthcare Institute: Frank Wharam, MB, BCH, BAO MPH (Co–Principal Investigator [PI]), Steve Soumerai, ScD, MSPH, Connie Trinacty, PhD, Emma Eggleston, MD, MPH, Fang Zhang, PhD, Christine Lu, PhD, Robert LeCates, MA, Claire Canning, MA, Dennis Ross-Degnan, ScD, MSPH (Co-PI); Kaiser Permanente Northern California: Julie A. Schmittdie1, PhD (PI), Assiamira Ferrara, MD, PhD, Romain S. Neugebauer, PhD, Sara Adams; Northwestern

www.ajpmonline.org
References


17. Sargent RP, Shepard RM, Glantz SA. Reduced incidence of admissions for myocardial infarction associated with public smoking ban: before June 2015


