

Fact-Based Policy: Improving Lives with Data and Science



How can we use data, science to improve policy and people's lives?

How do we make meaningful, measurable progress towards reducing poverty and increasing opportunity?

- State and local policy makers are tackling persistent, difficult problems with increasingly constrained resources
- Need for efficient, effective policy

Private sector is using well developed “data lakes” and top science to measure what works and innovate

- Does it save time? Does it better connect us with people that matter?
- Does it help us get our work done? Does it improve company targets?

Prototype model with Rhode Island

- Build a data lake, securely, connecting anonymized government databases across siloed agencies
 - Impact is interconnected; optimized for research
- Partner scientists with government from ideation, to measurement, to innovation
- Use data and science to impact policy and improve lives

DRAFT - NOT FOR CITATION

Data Lake: Building data resources for insights



Create a combined database of administrative records, connecting hundreds of currently siloed agency datasets to create a whole view of experiences, needs, and impact



Use proven algorithm for identifying individuals across datasets and anonymizing records which protects privacy and confidentiality, but allows aggregate insights for impact



Create derived tables to generate often-needed variables and views of individual and family needs and outcomes; standardize variable definitions for robust results across a range of projects and partners



Hosted on a secure, AWS GovCloud platform with automated auditing, logging, FedRAMP approved security; platform owned by government partner for data control

DRAFT - NOT FOR CITATION

RI 360: Example of Rhode Island data optimized for insights

Relational, anonymized and secure database for all individuals going back decades

Optimized for insights

- Birth: Natality and CDC PRAMS files, mother's hospital worksheet data, home visiting programs
- Health: Medicaid claims, hospitalizations, immunizations (EHR and PDMP coming soon)
- Education: Test scores, truancy, IEPs, college exams, college enrollment and graduation
- Labor: Wage rolls, UI, TDI, worker's compensation, labor training program enrollment
- Human Services: SNAP, TANF, SSI/DI, child care, General Public Assistance (case and benefit files)
- Criminal justice: Incarceration, in-prison records and activities, child protective services investigations and actions, police actions and calls for service
- Civic Participation: Voting, registrations, DMV records

DRAFT - NOT FOR CITATION

RIPL – \’ri-pel\ – ripple effect

Mission: We use data and science to impact policy and improve lives



Prioritize

We define policy goals with partners in policy making: governor’s team, agency directors, program managers.



Learn

We understand current programs through the lens of data and science to understand key successes and challenges.



Improve

We collaborate with policy makers to design and test improvements based on our findings, to enable efficient, effective policy which improves lives.

DRAFT - NOT FOR CITATION

Example 1: Accurate prediction of opioid dependence before prescription can save lives

National health emergency

- 80% of use disorders introduced to opioids through a prescription

Deep learning + integrated data

- Delivers accurate prediction (0.80 AUC) of who will become dependent based on data known before a prescription is given
- Empower medical experts with information
 - Evaluate cost-benefit, fairness (false positive rate)
 - Suggests using risk rating to guide prescriptions can reduce deaths and dependency

Key predictors include:

- Prior diagnoses and prescription types, age, former incarceration

Table 1: Integrated data predictive power

Agencies/Programs:	Accuracy
Human Services	0.703
Labor	0.530
Corrections	0.683
Medicaid	0.773
Police	0.622
Integrated	0.800

DRAFT - NOT FOR CITATION

Example 1: Key Predictors

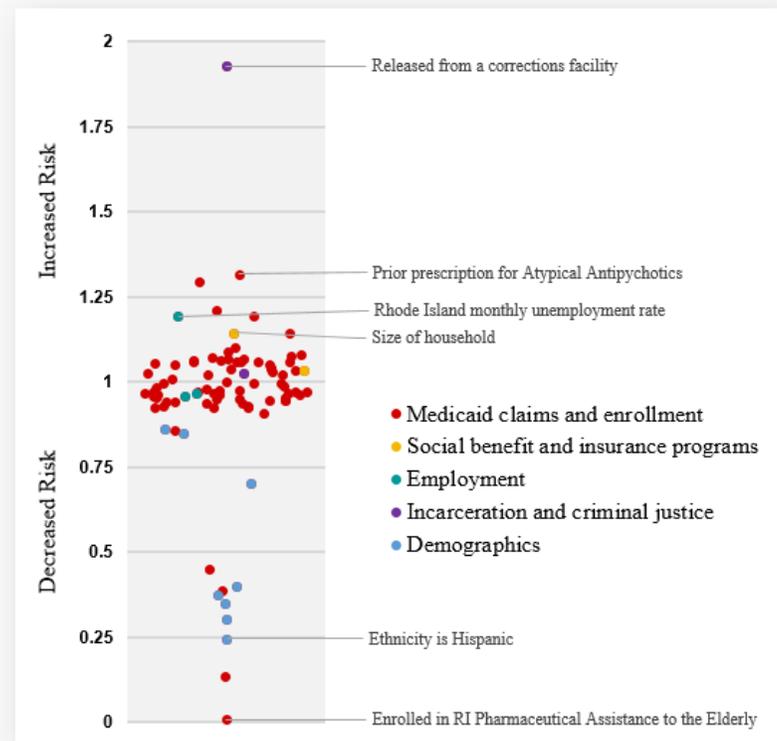
Factors that increase risk of dependency:

- Recent prior incarceration
- Prior prescriptions for mental health
- Prior treatment for tobacco or alcohol dependencies
- Young age

Factors which decrease risk of dependency:

- Gender is female
- Prior insurance claims for preventative health and dental care

*Results from a post-LASSO logistic regression of dependency on variables persistently selected from 1000 bootstrapped LASSO replicates to avoid random variable selection in the presence of multicollinearity



DRAFT - NOT FOR CITATION

Example 2: Early childhood investment: Spend \$4k and save \$68k

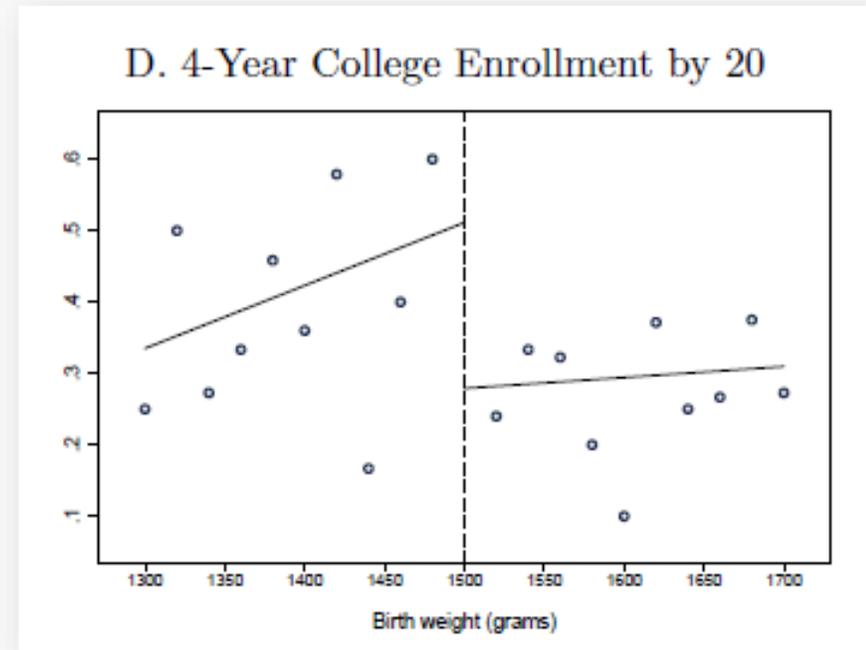
Measure the impact of hospital programs for Very Low Birthweight (VLBW) Babies

- VLBW are more likely to come from disadvantaged homes
- <1500g gets \$4,000 extra investment in health care & parental training

Regression discontinuity + Data = Policy improvement.

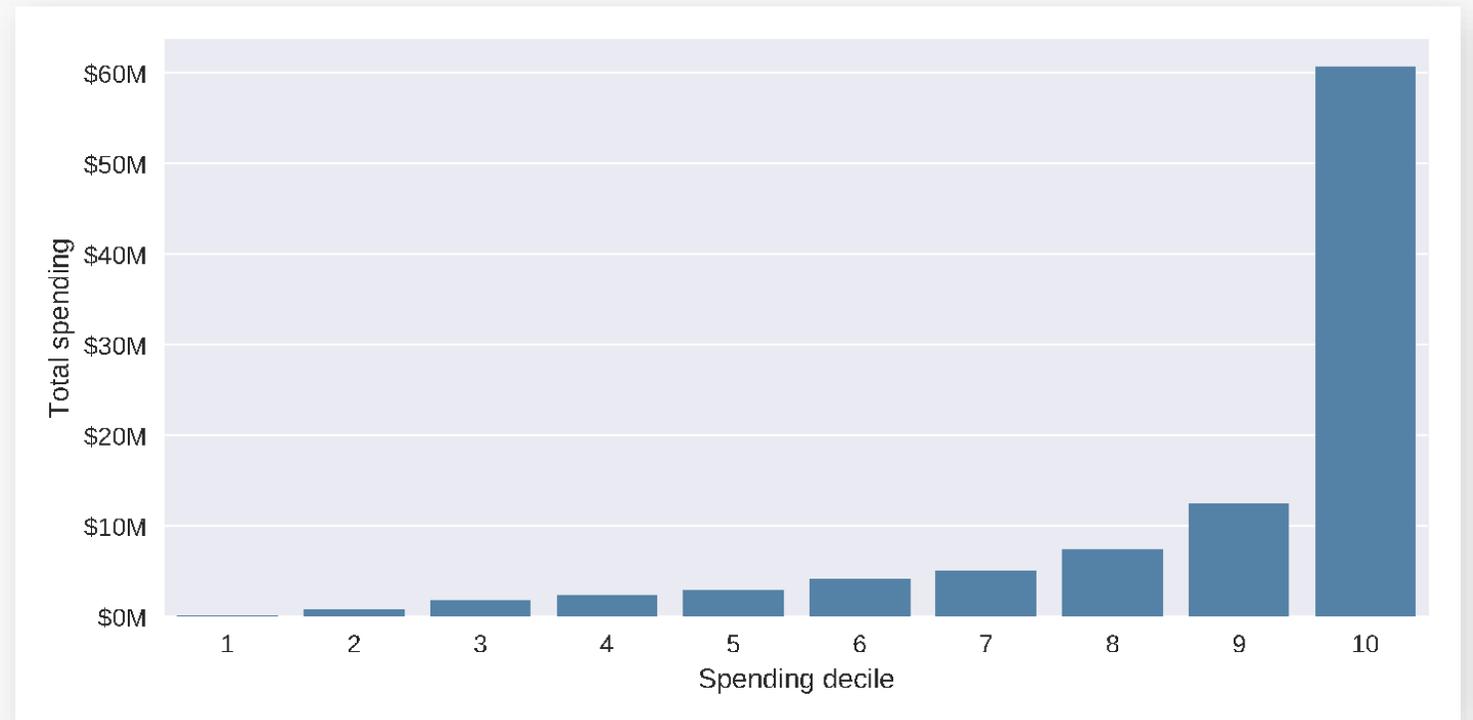
Causal effect of VLBW programs:

- Raises test scores 0.3-0.4 standard deviations
- Increases college going by 20 percentage points off a base of 30%.
- Reduces total social program expenditures by 27k by age 10, 68k by age 14



Example 3: Medicaid ED superuse costs \$\$\$

- ED claims ~ third of total Medicaid spending; 20%-25%
- Top 10% of users make up the majority of expenditures
- Can we reduce unnecessary ED cost before it occurs?



DRAFT - NOT FOR CITATION

Example 3: Data + machine learning = low cost solution

Predict ED preventable “cost-emergent” individuals

- In the lower 50% of ED spending (zero spend) in base year, top 10% following year
- Can we predict who will become a high-cost, preventable (“cost-emergent”) ED user?

Why not just see who spends a lot and then intervene?

- Wait-and-see approach is more costly than it is effective

Data-driven approach can pinpoint cost-emergent individuals, saving costs and harnessing existing programs

- We can predict cost emergence with 80-90% accuracy
- Understand drivers, which allows us to address pain points

DRAFT - NOT FOR CITATION

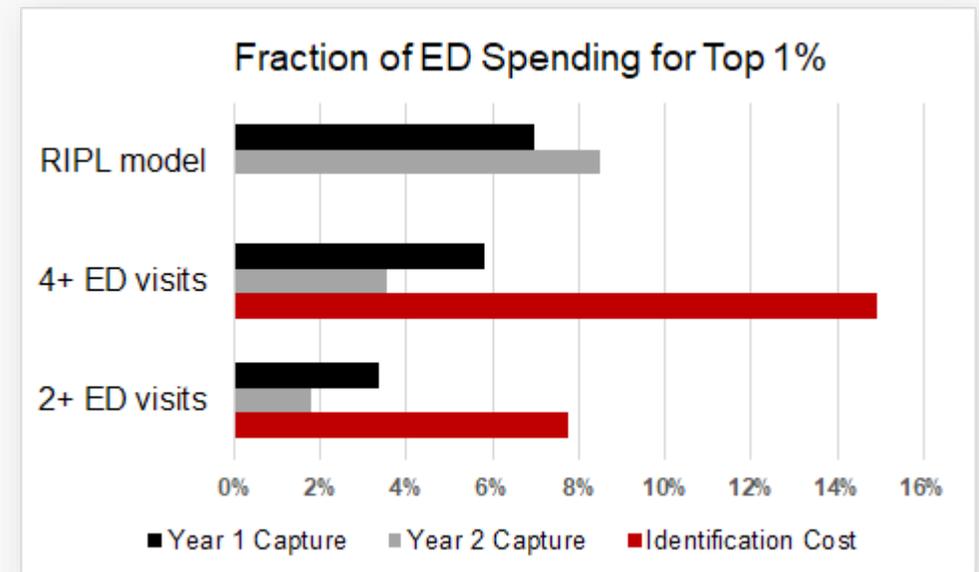
Example 3: Predictive model is a smart approach that saves \$\$

Model is accurate and captures more costs than wait-and-see methods (black and gray lines).

Saves identification costs (red line)

- Rather than waiting for costs to accrue to target “high-cost” super users, we can predict them in advance, diverting those costs before they happen

Potential savings amounts to a double-digit fraction of all annual Medicaid ED costs, even when only targeting 1% of individuals.



DRAFT - NOT FOR CITATION

Example 4: Child maltreatment

Recent increases in child deaths from abuse and maltreatment

- Family home visiting programs which reach out to “high risk” families
 - Budget shortfalls
- Could we do better in prevention?

Once we investigate, are we implementing the right treatment?

- Is removal or staying at home the right decision?

Use data, economics, data science to understand the current state of programs

- Measure how well we can predict maltreatment, and inform new policy for defining “risk,” proactively connecting support services to families who would benefit the most.
- Measuring the causal impact of removing a child from their home on short- and long-term educational and life outcomes.

DRAFT - NOT FOR CITATION

Example 4: Key Findings - Prediction

Construct a data set of detailed characteristics known in a base year

- Focus on social services population – Medicaid, SNAP, TANF
- Focus on children before school age
- How well can we predict substantiated investigations in the next year?
- What are key predictors?
 - Permanent and time varying features of social program use, law enforcement actions, detailed Medicaid claims, family structure changes

Results from initial models

- AUC-ROC from LASSO is 0.79
- Key positive predictors
 - Positive (increased risk): Medicaid diagnosis of mother drug abuse, mental disorder, contusions; new male in household; new infant in household; arrests for mother or father; number of household IDs associated with a child in prior year
 - Negative (decreased risk): Hispanic, Asian, Native American; married; child age; prior removal; parental education

Next steps: results integrated into a report on revamping targeted support services

DRAFT - NOT FOR CITATION

Example 4: Key Findings - Removal

Use quasi-random assignment of caseworkers to cases to create random variation in removal probability

Measure the causal impact of removal on the marginal child using case worker tendency for removal as an instrument for removal

Findings:

- Girls: Removal causes large and significant increases in test scores: 0.8 to 1.1 standard deviations
 - Impact appears in 3rd grade and persists
 - Significantly lowers rates of IEP by nearly 100%
- Boys: Removal causes large but insignificant decreases in academic outcomes: -0.30 to -0.40 standard deviations
 - Significantly raises IEP rates by nearly 50%
- Differential impacts not caused by differential pathways post-removal

Policy implications are unclear on removal, but perhaps can be identified on support services

DRAFT - NOT FOR CITATION

Example 5: Brightening the future for low-income families

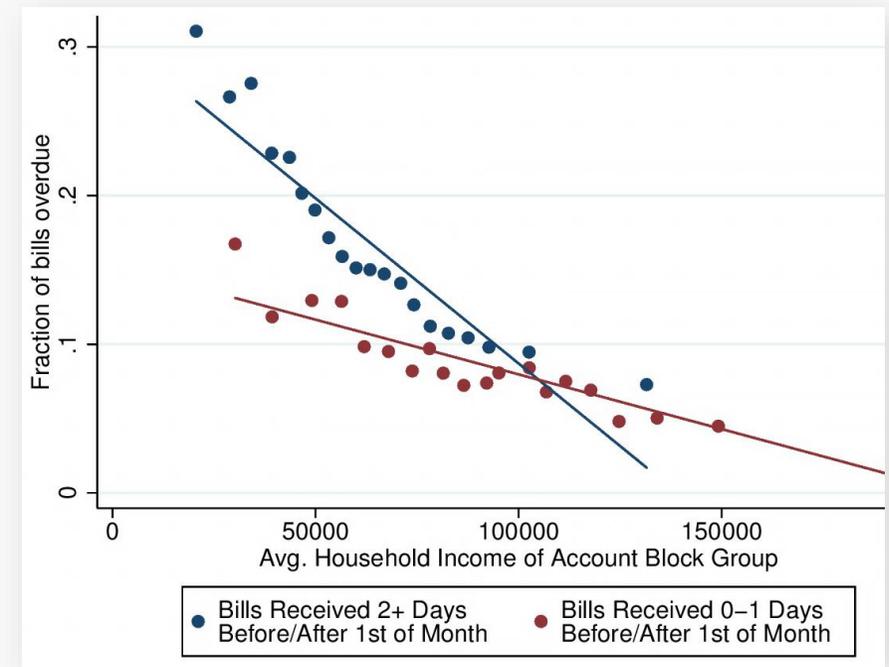
Many low-income families struggle to pay utilities

- Small changes can make a big difference
- Families in high poverty neighborhoods are more likely to pay their bill when they receive it on the first few days of the benefit month

RIPL is partnering with major electricity providers to help families maintain financial stability

If all low-income families received bills at the beginning of the month:

- 64% of utility shut-offs avoided
- Late payments reduced by 42%
- Total overdue payment amounts reduced by 38%



Average fraction of electricity bills over a month overdue by bill receipt timing and block group income level

DRAFT - NOT FOR CITATION

Example 6: SNAP, spending, nutrition

Goal: Can the Supplemental Nutrition Assistance Program better meet needs, improve nutrition?

Use new data to understand how SNAP is spent

- Understand models of human behavior and understand implications for policy

Data sources

- Scanner data from a major grocery retailer
- Administrative data from the State of Rhode Island in collaboration with USDA

Use changes in SNAP enrollment to measure causal impact of SNAP on:

- Food expenditures (Marginal Propensity to Consume Food out of SNAP)
- Shopping effort
- Nutrition

DRAFT - NOT FOR CITATION

Example 6: Designing a smarter SNAP

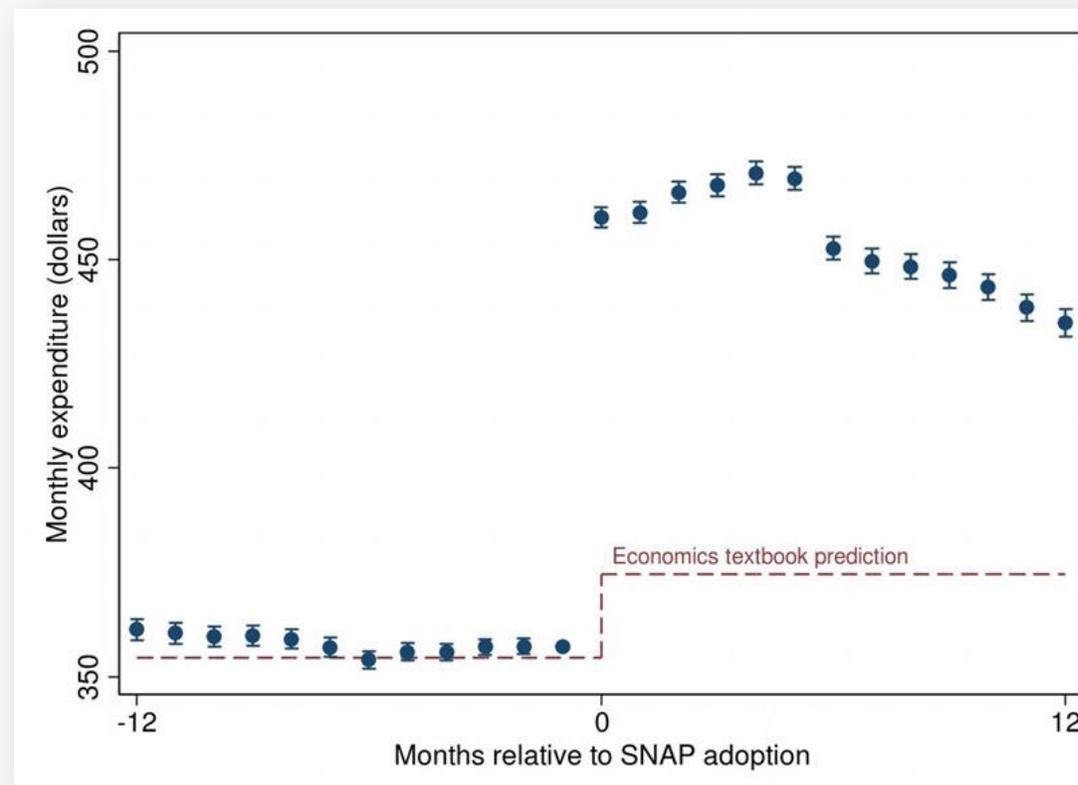
Data on 6 billion grocery purchases

SNAP should be the same as cash:

- At SNAP adoption, food purchases increase by 30%
- Non-food closer to 10%
- Response to cash is 10%
- People clip fewer coupons, buy too much food

One explanation is “mental accounting”:

- Evaluate SNAP relative to food budget
- Feel food wealth, shopping effort declines



DRAFT - NOT FOR CITATION

Example 6: Findings on nutrition

SNAP causes almost no change in measures of nutrition

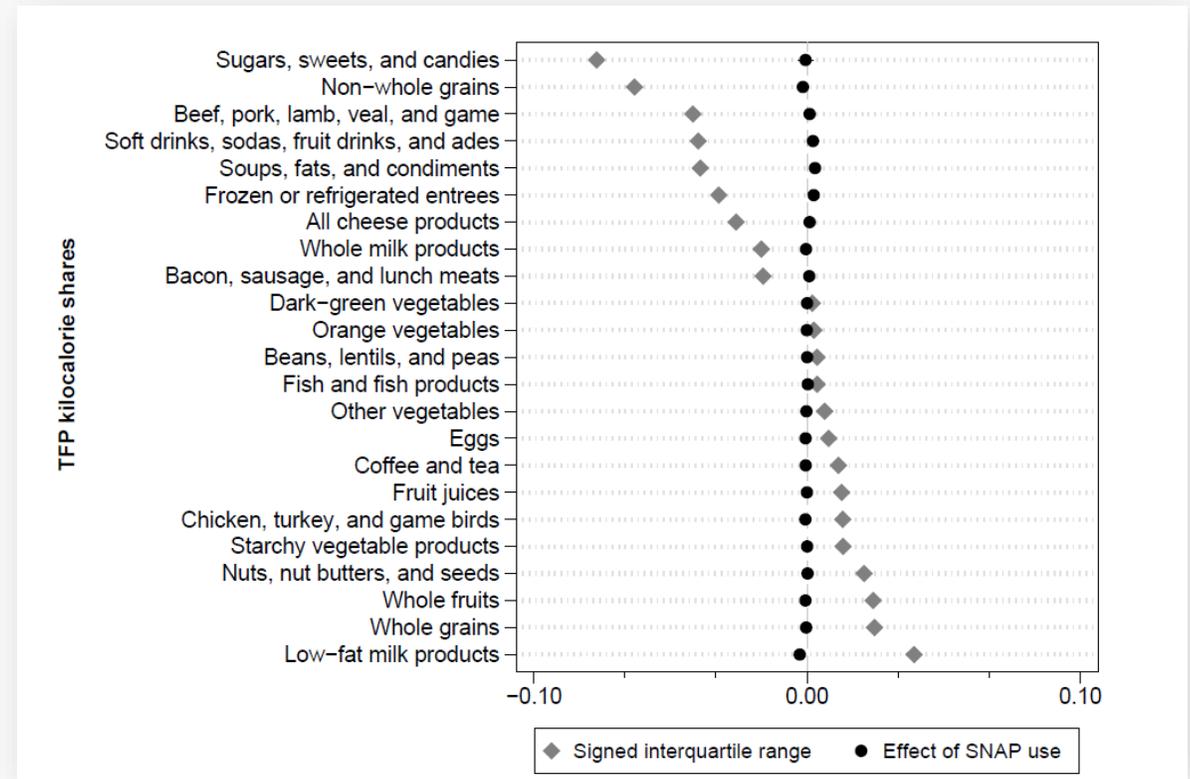
- Despite increasing expenditures by ~30%, SNAP has no impact on healthy eating

Splitting up payments could help families budget, increase nutrition

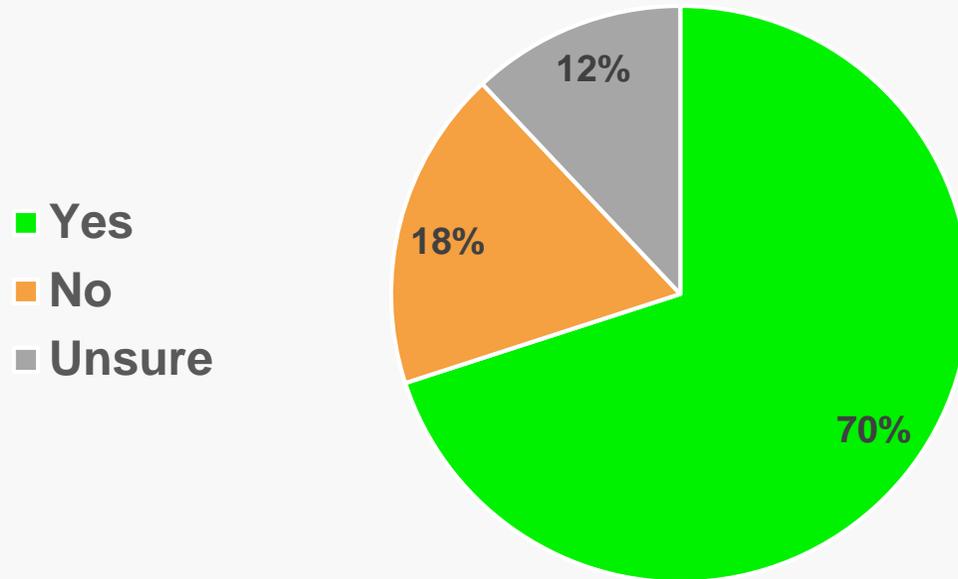
- ~70% say they prefer split payments

We developed a demonstration pilot with USDA & RI to split payments to twice a month

- Costless to implement



Example 6: What do people say?



"If I go shopping on the first of the month due to need, then I end up paying more for my groceries."

"I don't like receiving my SNAP all at once. Split it like welfare [TANF] so I can budget more easily."

"It would be good. It would help me budget. When I get my benefits, my eyes are bigger than my stomach... I buy what I want, not what I need."

"Split Issuance would be perfect. It would force me to buy smarter. I get things I don't really need on the first of the month."

DRAFT - NOT FOR CITATION

Example 7: Rhode to College

Priority: Governor Raimondo identified college access, particularly among low-income students, as a high priority – with the goal of achieving 70% postsecondary attainment in RI by 2025

- As of February 2018, only 47% of Rhode Islanders have a postsecondary degree or credential
- Governor Raimondo reached out to RIPL to design an effective and efficient way to improve college access

Partners: Governor’s Office, RIDE, College Board, RIPL, First Data Corporation

Funding: Chan Zuckerberg Initiative

Solution: Rhode to College is designed using proven, cost-effective methods to increase college-going rates among low-income, high-achieving Rhode Island students

- Based on behavioral economics: “If you want people to do something, remove the obstacles. Make it easy.” – Richard Thaler
- We will measure the success of the program along the road

DRAFT - NOT FOR CITATION

Example 7: How it works

Rhode to College pays students a small amount for completing milestones now, with a larger amount saved payment at college enrollment

Example milestones:

- Completing and updating college plans
- Studying for the SAT using official prep programs
- Completing the FAFSA
- Completing the Common App

Throughout the program, students can communicate via text 24/7 with AI chatbot named “Rhody”:

- Answers a wide array of questions
- Informs students of their progress and total funds earned
- Reminds them of upcoming milestones and deadlines, and provides encouragement

Figure 1: Screenshots from the Rhody AI chatbot

Hi Jonelle! Our record shows that you have finished answering this month's questions. Great job completing Milestone 4! 🎉👏

You've earned \$20 CFC and \$20 in Savings 😊 Be on the lookout for a text from Gyft with your \$20 Gyft card!

Your total R2C earnings so far are \$140! \$45 in Cash-for-Completion and \$95 in Savings Funds. Great work! 🙌

Don't forget to keep studying for the SAT! A good 🎯 study goal 🎯 is to accomplish 40 practice problems each week from your personalized Khan Academy study plan.

Can I do the College Search on the computer?

You can use College Search on any internet-enabled device: a phone, a tablet, laptop 🖥️ or desktop computer 🖨️. Just type the link I gave you into a browser and use your PIN to get searching! Any internet browser should work.

How do I log in to the College Search?

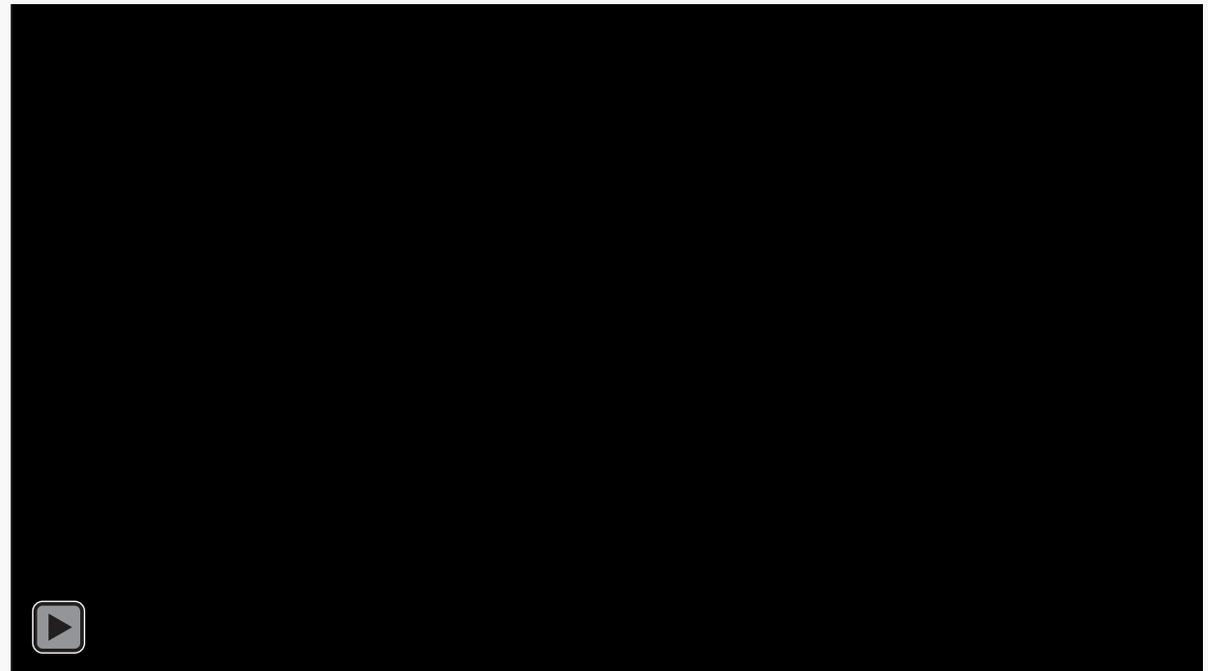
Have you tried using the personalized link and PIN I sent you? (Scroll up in our convo!) If you're still having trouble, email info@rhode2college.org with your issue!

Example 7: Success so far

First ten milestones delivered:

- Over 60% enrollment; higher for minorities and females
- Over 70% average milestone completion rate
- Significantly increased the quality of colleges students are planning to apply to as measured by labor market and employment returns on investment
- Quadrupled the amount of students studying the College Board-recommended amount of 20 hours or more for the SAT, and doubled the amount of time that students studied on average

What Rhode 2 College means to students – click to play video:



[View video on the web](#)

DRAFT - NOT FOR CITATION

Examples - Helping policymakers use their data

Provided RIPL database and software to the Executive Office of Health and Human Services (EOHHS) to build and support their data ecosystem

- Supported report on child protection and Medicaid reporting not before possible

Streamlined data access across agencies to enhance program operation

- Department of Labor and Training (DLT) and EOHHS
- DLT and vital records

Helped DLT use data (instead of paper) to identify workers who can receive federal funds

- Saves state training \$\$ (see [Governing Magazine Article](#))

Identified challenges with reporting with the police

- Supported them in implementing a solution for more accurate reporting

Worked with the Departments of Corrections and Health and Human Services to design and implement streamlined connection of prison releases to social service benefits to lower recidivism

Identified high-value in-prison training programs which reduced recidivism, redirected resources

Helped EOHHS develop targeted policy to support families at risk of child neglect and/or maltreatment

DRAFT - NOT FOR CITATION

Scaling...

We are working with state and local governments to:

- Scale the research data lake infrastructure to their locality
 - Using a scalable GovCloud solution with AWS
 - Cloud offers high security and transparency, and preserves ownership
 - Easy to use: QuickStart environment and ETL code with GUI interface to easily customize and extend standard data lake
- Scale projects to their locality to inform policy and increase learning
 - Common structure means that code developed in one place can be adopted in others
- Equip partners with partnership templates, guidance, and computing infrastructure to produce their own insights and partner with local researchers going forward

DRAFT - NOT FOR CITATION

From some of our partners....

Eric J. Beane, Former Secretary of the Executive Office of Health and Human Services, Former Deputy Chief of Staff to the Governor of Rhode Island:

- *"Investing in data-driven and fact-based policy is vital to helping people in need. We need to be innovative and proactive in government to support our communities and families to the best of our ability. Our partnership with RIPL helps us use objective research and data at high speed so we can do just that. RIPL's approach to data and science for social good could benefit local and state government throughout the country."*

Scott Jensen, Director of the Rhode Island Department of Labor and Training:

- *"RIPL is a responsive, skilled and trusted partner that shares our commitment to using data to inform policymaking. The invaluable insight and technical assistance provided by RIPL helped the Department fully leverage our data to make better use of federal funds, understand how best to help workers re-enter the workforce, and measure the impact of workforce development programs. We look forward to future collaboration with RIPL as we work to prepare all Rhode Islanders for in-demand jobs."*

DRAFT - NOT FOR CITATION

From some of our partners....

Macky McCleary, Administrator of the Rhode Island Division of Public Utilities and Carriers

- *“RIPL is making it possible for my agency solve problems with data that we weren’t able to solve before. Their commitment to bringing innovative data solutions to the table means that I’m supported in understanding how policies in my agency and those in other agencies impact or interfere with each other. I am able to make informed policy decisions in ways I wasn’t able to before the RIPL partnership.”*

Trista Piccola, Director of the Rhode Island Department of Children, Youth and Families:

- *“RIPL helps us prove a concept—that we can measure what happens to children DCYF touches and use those facts to transform family policy for the better. RIPL and DCYF proved that not only could we measure these outcomes short- and long-term, but we could measure the impact of our own policies retrospectively and make informed decisions for children and families in need. RIPL researchers have the expertise and ability to aggregate data across multiple systems, which gave us a comprehensive analysis of our state’s children and youth and the effectiveness of particular services. RIPL gave us the support we needed to build a better juvenile justice system for the state and strengthen our home and community-based service array.”*

DRAFT - NOT FOR CITATION

From some of our partners....

Arthur Nevins, Education Policy Advisor at the Office of the Governor:

- *“RIPL has become a trusted partner we can engage to help us design new ways to improve lives for more students meaningfully and measurably. RIPL always get the job done efficiently, effectively, and with a focus on our core needs. They consistently prioritize how they can deliver the most value for the students and families we serve.”*

Kimberly Paull, Director of Data and Analytics for the Rhode Island Executive Office of Health and Human Services:

- *“Through partnership, insights, and integrated data, RIPL gave our team the jump start we needed to improve how our state serves whole people, families and communities. RIPL’s ability to quickly gather, clean, connect, and curate data for analysis – and then apply machine learning in collaboration with our teams’ context – gave us a new sense of the possible for data-driven policy-making, academic partnerships, and continuous improvement. Because of this jump start and the credibility it helped us build through our first major integrated data project, we have now built the capacity to conduct our own integration and analysis in the future, expanding our ability to serve our communities and help families in need.”*

DRAFT - NOT FOR CITATION

From some of our partners....

Chief Sid Wordell, Executive Director of the Rhode Island Police Chiefs' Association:

- *“RIPL brings people together around data to solve problems. Working with RIPL over the last couple years has given RI Law Enforcement the opportunity to use RIPL’s integrity and data mining expertise to foster partnerships with a broad range of stakeholders to review often challenging policies. Their professional approach to using data has built trust and promoted transparency without judgement, which has been a refreshing approach for us. I look forward to future partnerships with RIPL as we develop more processes that improve policy with facts.”*

DRAFT - NOT FOR CITATION