Science in Service of Cities.
Using Predictive Analytics to Improve Inspection Targeting

Michael Greenstone
Faculty Director
University of Chicago Energy & Environment Lab
Agenda

1) Background on UChicago Energy & Environment Lab

2) Using Predictive Analytics to Improve RCRA Inspection Targeting

3) Improving Inspection Targeting and Efficiency in Your Program
Agenda

1) Background on UChicago Energy & Environment Lab

2) Using Predictive Analytics to Improve RCRA Inspection Targeting

3) Improving Inspection Targeting and Efficiency in Your Program
What is Predictive Analytics?

UChicago E&E Lab Leadership

Michael Greenstone
Faculty Director

Olga Rostapshova
Executive Director

Mark Templeton
Regulatory Advisor
OUR APPROACH

WE PARTNER WITH POLICYMAKERS TO:

IDENTIFY
Promising solutions to environmental challenges

TEST
The most promising policies and programs

SCALE UP
The most effective and cost-efficient policies and programs
OECA and the University of Chicago Energy & Environment Lab began our partnership in early 2015.

**Goal:** Collaborate to develop and evaluate the effectiveness of novel techniques for supporting compliance assurance efforts.

UChicago Research Fellow joined EPA HQ in June 2017 (embedding is part of our model).
Our Partners

We partner with policymakers at the local, state and federal levels
Agenda

1) Background on UChicago Energy & Environment Lab

2) Using Predictive Analytics to Improve RCRA Inspection Targeting

3) Improving Inspection Targeting and Efficiency in Your Program
What is Predictive Analytics?

**Predictive analytics** is the practice of using data to determine patterns and predict future probabilities. Its appeal is that it can discern subtle relationships in the data.

**Also known as:** Machine Learning, Artificial Intelligence, Data Science, Big Data, etc.
Private companies have been using predictive analytics to inform decisions and improve performance for years

- Detects potential credit card fraud
- Product recommendations at Amazon and Netflix

This revolution is coming to the public policy sphere, where government agencies are learning how to use data to provide better services and increase their cost-effectiveness

- New York Fire Department prioritizes building inspections by predicting which buildings are more likely to catch fire
- Chicago Department of Public Health predicts children at high risk for lead poisoning for screening
This wave of technological change is making the world much more efficient.

Yet these methods are not yet widely applied in environmental regulation – leaving potentially substantial efficiency gains on the table.

Predictive analytics has the potential to transform inspection targeting – and enforcement and compliance more broadly – in environmental regulation by achieving greater efficiency in an increasingly resource-constrained environment.
The Challenge: How to Improve Inspection Targeting

All levels of government enforce laws that protect the environment, health, food, worker safety, and much more, by conducting inspections.

EPA has resources to inspect a small fraction of active hazardous waste facilities each year – how can they find more violators with fewer resources?
Proposed Solution: Identifying high-risk facilities

Use predictive analytics to identify facilities most likely to violate RCRA regulations

1. Generate predictors
2. Train model on historical data
3. Predict severe violations

Built model based on 15 years of historical data, including tens of thousands of variables such as:

- Facility characteristics (location, industry, etc.)
- Historical enforcement and compliance data – RCRA and other regulations such as the Clean Air Act
Use predictive analytics to identify facilities *most likely to violate* RCRA regulations.

**Proposed Solution: Identifying high-risk facilities**

We evaluate model performance by:

1. Training the model on historical data (2000-2013)
2. Predicting facility risk level in the recent past (2014)
3. Comparing these results to whether the facility actually violated regulations (in 2014)
Proposed Solution: Identifying high-risk facilities

Use predictive analytics to identify facilities **most likely to violate** RCRA regulations.

- Generate predictors
- Train model on historical data
- Predict severe violations

The model generates a **risk score** for each facility quantifying the likelihood that an inspection of the facility will find a severe violation of RCRA regulations.
Performance on Historical Data

The model achieves a 50% increase in the detection of severe violations, compared to status quo targeting.

**Conclusion:** Predictive analytics works for inspection targeting for environmental enforcement when using historical data.
Before incorporating the model into a targeting tool to be used by inspectors, we worked with Regional offices to conduct additional testing of the model in the field.

This on-the-ground testing serves to demonstrate the model’s performance in the real world.

The field test also helps build inspector confidence in the targeting tool before it is rolled out nationwide.
All large quantity generators (LQGs) eligible to be inspected in a given year.
Both the model and the EPA choose their top facilities to inspect.
Results

Preliminary results indicate that detection rate for:

1. EPA chosen inspections is 22%
2. Model chosen inspection is 40%

→ That is an **80% increase over the current hit rate for severe violations**

→ Field evidence also indicates that predictive analytics can **substantially** improve efficiency of environmental inspections programs.
Increased Efficiency

In FY17, the EPA conducted ~1,190 RCRA inspections

If the RCRA predictive analytics model was used nationwide, the EPA:

• Could have **found 214 additional severe violators** with the same inspection resources

-OR-

• Could have **conducted 535 fewer inspections** while finding the same number of severe violators
Taking Predictive Analytics to Scale

Using this model has great potential to increase the effectiveness of RCRA inspection targeting without increasing spending for Regional offices and state agencies.

At the conclusion of the field test, OC plans to implement the targeting model as a component of ECHO Gov so that states can leverage it when developing their inspection plans.
Agenda

1) Background on UChicago Energy & Environment Lab

2) Using Predictive Analytics to Improve RCRA Inspection Targeting

3) Improving Inspection Targeting and Efficiency in Your Program
Other possible applications could include:

• Inspection targeting for other areas such as Clean Water Act, Clean Air Act, etc.

• Identifying high-risk facilities to help prioritize compliance assistance

• Identifying potentially un-permitted or otherwise non-compliant facilities (e.g., utilizing satellite data and image classification)
What else do we do?

The UChicago Energy & Environment Lab also:

- Provides **technical assistance** to regulators and state agencies
- Conducts **impact evaluations** to determine causal impact of programs and policies using
  - Randomized control trials (RCTs)
  - Quasi-experimental methods

Please **reach out to us** to discuss how we can help you design and test effective and efficient policies, regulations and enforcement strategies.
Interested in working with us?

Contact us!

**Michael Greenstone**
Faculty Director, UChicago Energy & Environment Lab
mgreenst@uchicago.edu

**Olga Rostapshova**
Executive Director, UChicago Energy & Environment Lab
oro@uchicago.edu
Thank you.