Rewiring Evaluation

Approaches @ the intersection of data science and evaluation
Rewiring Evaluation Panelists

Michael Bamberger
Independent Evaluator & Author

Pete York
Principal and Chief Data Scientist
BCT Partners

Swapnil Shekhar
Co-founder and COO
Sambodhi Research and Communications, India

Veronica Olazabal
Senior Adviser and Director,
Measurement, Evaluation and Organizational Performance
Rockefeller Foundation

Jos Vaessen
Methods Adviser
World Bank Independent Evaluation Group

#gLocalEval2020
Participants by Affiliation

- Academic Institution
- Multilateral Development Institution (e.g. UN, IADB, WB)
- Non-Profit Organization (including Associations & Civil Society Organizations)
- Private Sector
- Government
- Other

#gLocalEval2020
Session one: Overview of Big Data and its Applications to Evaluation

• Welcome!
• In the chat function below the video, please share a greeting and where you are currently based.
• Throughout the session, you may ask questions in the chat function or through the following link www.pollev.com/iegnw
Rewiring Evaluation

Measuring results and impact in the age of big data: The nexus of evaluation, analytics, and digital technology

Pete York
Principal and Chief Data Scientist
BCT Partners

Michael Bamberger
Independent Evaluator & Author

Veronica Olazabal
Senior Adviser and Director, Measurement, Evaluation and Organizational Performance
Rockefeller Foundation

#gLocalEval2020
Evaluation in the Age of Big Data

Michael Bamberger
Independent consultant
1. Defining big data
2. Challenges for evaluation in a complex and transforming world
3. Potential benefits of big data to strengthen current evaluations
4. Potentially useful kinds of big data for evaluation
5. Limitations and challenges of big data
6. The benefits of big data in the age of Covid-19
7. Case studies: the application of big data in real-world evaluations
1. Defining Big Data, Data Analytics and Data Science
Big Data, Data Analytics and Data Science

Big data + Data Analytics = Data Science

- Too large to analyze on a single computer
- Generated very fast
- Always on
- Non-reactive
- Networked

- Organization, analysis and dissemination of big data + other kinds of data
- Integrated data platforms
- Predictive analytics
- Machine learning and artificial intelligence

#gLocalEval2020
2. Challenges for evaluation in a complex and transforming world
Data collection challenges

- Cost/time of data collection limits sample size and disaggregated analysis
- Difficult to collect real-time data
- Hard to collect data on large geographical areas beyond project
- Hard to collect longitudinal data
- Monitoring processes and behavioral change.
- Identifying and including difficult-to-reach groups
Complexity challenges

a. Development programs are complex but most evaluations assume "simple" linear relationships between project inputs and outcomes
b. Mapping interactions between different dimensions of complexity
c. Need to incorporate systems analysis
The intervention

Institutions and stakeholders

Causality and change

The system in which the program operates

Complexity of the evaluation

Multiple contextual factors
Analytical and dissemination challenges

a. Difficult to analyze large volumes of data
b. Integrating, and analyzing different data sets
c. Real-time analysis and dissemination of findings to help improve ongoing programs
d. Predictive modeling
Discussion

To submit questions to the panelists, go to: www.pollev.com/iegnow
If you wish, add your name and location
3. Potential benefits of big data to strengthen current evaluations
Rapid and economical data collection and analysis

a. Large volumes of data can be economically collected – covering total population not just sample

b. Including difficult to access groups
   • Geographically remote
   • Difficult to identify and contact

c. Real-time feedback to managers and policy-makers
Access to a wider range of data

a. Longitudinal data sets covering many years
   • Creating baseline
   • Tracking sustainability

b. Combining multiple sources of data into integrated data platform
   • Surveys, text, audio-visual data, satellite and drone images, social media
   • Text analytics

c. Tracking processes and behavioral change
More sophisticated kinds of analysis

a. Predictive analytics
b. Associations among multiple data sets
c. Using artificial intelligence (AI) to create quasi-experimental designs using natural experiments
Facilitates Analysis of complexity

a. Systems analysis
   • Systems mapping
   • Social network analysis
   • System dynamics

b. Contextual analysis

c. Tracking processes of change
4. Potentially useful kinds of big data
Useful kinds of big data

a. Geospatial data: satellites and drones
b. Social media: Twitter, Facebook, radio call-in programs
c. Internet searches
d. Phone call-center records
e. Mobile phone data
f. Administrative data
g. Internet of Things [IOT]
5. Challenges using Big data
Methodological challenges

a. Trust in large numbers
b. Data quality and construct validity
c. Multiple sources of bias
d. Correlation and causality
e. Over-reliance on available data – who is missing?
f. The role of theory
Strategic challenges

- Unrecognized sources of bias
- Exaggerated claims for artificial intelligence
- The dangers of automated decision-making – algorithms
- Top-down decision-making – social exclusion
- The divide between research and evaluative thinking
6. The benefits of big data in the age of Covid-19
Addressing the challenges of remote data collection

a. Multiple remote data collection tools
b. Contextual analysis – understanding the big picture
c. Real-time data generation, analysis and dissemination
d. Using satellites, drones and geo-tagging to identify virus “hot-spots”
e. Monitoring social distance and contact tracing with smart-phones
f. Facilitating citizen feedback + promoting social solidarity
g. Integrating health, economic and social data into a single platform
Discussion

To submit questions to the panelists, go to: www.pollev.com/iegnow
If you wish, add your name and location
A Big Data Science Approach to Evaluation

Pete York
Principal and Chief Data Scientist
BCT Partners
Topics Covered

- A big data science approach to evaluation
- Potential benefits of a big data science approach to evaluation
- Limitations and challenges
- The data, technology and skills required to get started
A big data science approach to evaluation

Methodological Overview
A big data science approach to evaluation requires two types of collaborators.
Methodological Overview

1. Find and access relevant data sets
2. Prepare the data for evaluation
3. Conduct quasi-experimental modeling
4. Produce the results
Step 1: Find and access relevant data sets

PROGRAM
ADMINISTRATIVE DATA

GOVERNMENT DATA

BIG DATA
Step 1: Requirements

Evaluator requirements

- Research skills
- Permission
- Subject matter expertise

Data scientist requirements

- Access
- Assess connection requirements
Step 2: Prepare the data for evaluation

CONNECT, EXTRACT AND SET UP THE DATA
CLEAN THE DATA
IDENTIFY AND CREATE EVALUATION VARIABLES
CONDUCT LOGIC MODEL LABELING OF THE VARIABLES
Step 2: Requirements

Evaluator requirements

- Evaluation framework
- Subject matter expertise
- Analysis software

Data scientist requirements

- Extraction, transformation and loading
- Analysis software
- Python, R
Step 3: Conduct quasi-experimental modeling

- Train machine learning algorithms to find matched comparison groups
- Train machine learning algorithms to find and assess counterfactual experiences
- Run statistical inferential analyses to test hypotheses
Step 3: Requirements

Evaluator requirements

- Data analysis using structured and unstructured data, applying statistics and machine learning
- Analysis software
- Subject matter expertise

Data scientist requirements

- Analysis software
- Machine learning
- Python, R
Step 4: Produce the results

- Produce evaluation reports
- Produce program management dashboards
- Produce front line decision-making tools
Step 4: Requirements

**Evaluator requirements**
- Synthesis and writing
- User-centered design
- Data visualization software

**Data scientist requirements**
- Back-end development
- Application development
- Data visualization
Example of a Big Data Science Approach to Evaluation

Does the right level of access to nonprofit direct services improve community well-being?
No-Cost Toolset
Step 1: Find and access relevant data sets

IRS 990 DATA
N=325,000 DIRECT SERVICE PROVIDERS ACROSS 18 SECTORS

AMERICAN COMMUNITY SURVEY DATA
N=70,832 CENSUS TRACTS
Step 2: Prepare the data for evaluation

---

**Context**
- Population density
- Local commute times
- Local transit
- Socioeconomic status

**Baseline**
- Area Deprivation Index (2015)

**Strategy**
- Ideal amount of locally-accessible direct services (by sector)

**Outcome**
- Area Deprivation Index (2018)

---

*Does the right level of access to nonprofit direct services improve community well-being?*

---

*Ideal level of access was derived by using predictive geospatial algorithms, controlling for context*
Step 3: Conduct quasi-experimental modeling

FIND MATCHED COMPARISON GROUPS

Urban, working poor census tracts where >50% use public transit for a 30-minute commute to work
Step 3: Conduct quasi-experimental modeling

FIND COUNTERFACTUAL EXPERIENCES

Ideal Level of Access to Direct Services vs Less Than Ideal Access to Direct Services

Urban, working poor census tracts where >50% use public transit for a 20-minute commute to work
Step 3: Conduct quasi-experimental modeling

RUN INFERENTIAL ANALYSES

Ideal Level of Access to Direct Services vs Less Than Ideal Access to Direct Services

Urban, working poor census tracts where >50% use public transit for a 20-minute commute to work
Step 4: Produce the Results

Accessible services make a significant difference on Community Well-Being*  

*$p<.01$ for all sectors except community building ($p=.07$)
Step 4: Produce the Results

Percent of Communities with the Right Amount of Access to Services

Order: Greatest Inequality to Least Inequality

- More White
- More People of Color

*p<.05*
Step 4: Produce the Results
Discussion

To submit questions to the panelists, go to: www.pollev.com/iegnow
If you wish, add your name and location
Benefits of a Big Data Approach to Evaluation

**RAPID**

**FLEXIBLE**

**COST-EFFECTIVE**

**AUTOMATED**

**USEFUL**

**THE EXAMPLE’S BOTTOM LINE**

- Setup & training
  - 4 weeks
  - Re-train w/new variables
  - 1 week

- Used
  - quantitative & qualitative data, together

- Evaluator
  - 120 hours
  - Data Scientist
  - 120 hours

- Analyze new data and produce results, instantly

- On-demand reports, dashboards, visualizations & decision-making tools
Limitations and Challenges

- Can require a big data architecture
- The collaborative technical learning curves of evaluators and data scientists
- Reconciling competing philosophies of evaluators and data scientists
- Requires integrating functions of evaluation and data management
To submit questions to the panelists, go to: www.pollev.com/iegnow
If you wish, add your name and location
Additional Resources

- For additional resources, visit [https://ieg.worldbank.org/event/datascience-and-evaluation](https://ieg.worldbank.org/event/datascience-and-evaluation)
Different Organizational Perspectives on Big Data and Evaluation

Coming up @ 10:30am EDT

Swapnil Shekhar
Co-founder and COO
Sambodhi Research and Communications, India

Veronica Olazabal
Senior Adviser and Director,
Measurement, Evaluation and Organizational Performance
Rockefeller Foundation

Jos Vaessen
Methods Adviser
World Bank Independent Evaluation Group

#gLocalEval2020
Participants by Affiliation

- Academic Institution
- Multilateral Development Institution (e.g. UN, IADB, WB)
- Non-Profit Organization (including Associations & Civil Society Organizations)
- Private Sector
- Government
- Other

#gLocalEval2020
Session two: Different Organizational Perspectives on Big Data and Evaluation

• Welcome!
• In the chat function below the video, please share a greeting and where you are currently based.
• Throughout the session, you may ask questions in the chat function or through the following link www.pollev.com/iegnow
Rewiring Evaluation

Measuring results and impact in the age of big data:
The nexus of evaluation, analytics, and digital technology

Pete York
Principal and Chief Data Scientist
BCT Partners

Michael Bamberger
Independent Evaluator & Author

Veronica Olazabal
Senior Adviser and Director, Measurement, Evaluation and Organizational Performance
Rockefeller Foundation

#gLocalEval2020
Perspectives on using Data Science for Evaluation: Learnings from India

Swapnil Shekhar
Co-founder and COO, Sambodhi Research and Communications, India

gLOCAL Evaluation Week Webinar
We are in the midst of a Data Revolution.

With the increase in digital penetration in India, there has been an explosion of information on difficult to access populations.

This has groundbreaking implications for Evaluations and data-driven decision making for SDGs.

Applications of Data Science in India

Administrative MIS Data
- Data collected by government and public institutions (like primary health centres and by frontline workers).
- Serve as large reserves of primary, demand-side data that is updated near real-time.

Technology-enabled Big Data
- Data collected automatically through technology or digital devices.
- Includes data stored on mobile phones, social media, drones, satellite devices, IoT, ATM card payments, geospatial mapping.
- Government programs such as Aadhar, UPI ID, Aarogya Setu etc. create an enabling environment
- Philanthropic focus on creating grassroot data science infrastructure is a stimulant
However, the Indian context poses unique challenges to Data Science...

- **Without contextual human-level information, data science offers only part of the story**
  
  Essentially, Data Science offers digital footprints of human behavior, which in a country like India is highly complex due to the demographic diversity. Data Science may not reflect the extent of factors that influence human decisions.

- **Weak data culture**
  
  The infrastructure and cultural mindset to collect data and abide by data-driven processes is weak. For example, frontline workers might be reluctant to carry out data collection and management for fear of punitive action and being held accountable.

- **Disparate institutional receptivity**
  
  Decision-makers across levels, sectors, and states have varying levels of receptivity to data science. Public institutions are moving towards data and analytics-driven processes for decision-making. Consciousness on its potential for evaluations needs to be strengthened.

- **Lacks representativeness**
  
  While Data Science offers vast amount of insights, it might lack sampling rigor and geographical representativeness. This makes it difficult to offer valid insights at the district, state, or national level for policy making.

- **Can supplement, not completely replace, primary, human-level indicators**
  
  Primary, demographic indicators are ultimately key in any policy, development or philanthropic decision-making. Should think more in terms of how Data Science being integrated with primary survey data processes instead of as a proxy indicator.

Technology, innovation, and new data collection practices which are cheaper and faster can help overcome some of these gaps in India.
Designing solutions for India

- **Revisiting Data Collection**
  - **Extending Data Science Application to demand-side data**
    Not limiting Data Science application to tech-enabled big data sources can enhance its use value. Integrating primary survey data with big data can help application fit the Indian context.
    [https://www.thehindubusinessline.com/opinion/collecting-the-right-data-is-crucial-for-good-decision-making/article29896962.ece](https://www.thehindubusinessline.com/opinion/collecting-the-right-data-is-crucial-for-good-decision-making/article29896962.ece)
  - **Longitudinal Panel Data**
    Panels are a viable source of primary, demand-side data which can be collected remotely and at lowered costs.
  - **Reinventing the survey**
    Post-COVID, surveys might have to become shorter and easier to record remotely through telephone or on text message. Sambodhi recently surveyed 5,000+ rural households telephonically to assess food consumption patterns during the pandemic-induced national lockdown.

- **Cultural shift towards data usage and appreciation**
  The mindset for data-driven processes will have to be fostered in India. Important to sustain the momentum generated by public and philanthropic efforts for strengthening infrastructure to harness data.
  [https://academyhealth.confex.com/academyhealth/2015di/meetingapp.cgi/Paper/7712](https://academyhealth.confex.com/academyhealth/2015di/meetingapp.cgi/Paper/7712)

- **Multi-sectoral partnerships**
  The private sector has access to vast amounts of big data (payment behavior, usage information, satellite geospatial data etc.) and data science potential which can benefit SDG efforts. Data Philanthropy - endorsed by Rockefeller Foundation and Mastercard Center for Inclusive Growth is a step in this direction.
Discussion

To submit questions to the panelists, go to: www.pollev.com/iegnow
If you wish, add your name and location
Evaluation in the Age of Big Data: Prospects and Challenges for Independent Evaluation

Jos Vaessen, PhD
Methods Adviser
World Bank Independent Evaluation Group
Outline

• ‘embedded’ versus independent evaluation

• examples of how data science applications can improve evaluation practice in the framework of independent evaluation

• prospects and challenges going forward
Introduction

‘embedded’ evaluation

• reporting to management/operations
• learning, program improvement, accountability
• retrospective evaluation, real-time evaluation, developmental evaluation
• ‘small’ evaluands: activity, project, program
• ‘simple’ evaluation design
• small teams

Independent evaluation

• independent from management/operations
• accountability and (strategic) learning
• retrospective evaluation
• ‘large’ evaluands: global / regional / country / sector program, thematic area of work
• ‘complicated’ evaluation design: multi-level, multi-site evaluation
• large teams
1. Enhancing efficiency (and quality): identifying the ‘evaluand’ (*nutrition evaluation*)

**Nutrition Portfolio Identification Strategy**

1. **Search criteria**
   - **Evaluation Period**: FY08-FY19

2. **Identification criteria**
   - **Limited to...**
     - Sector codes
     - Agriculture
     - Education
     - Health
     - Social Protection
     - Public Admin.
     - Water
   - **Theme codes**
     - Private Sector
     - Public Sector
     - SD, HD and Gender
     - Rural Development
     - Environment
   - **High child stunting**
     - 88 countries with stunting rate $\geq$ 20% in evaluation period

3. **Inclusion criteria**
   - **Database D**
     - Extracted project components from Project documents in Image Bank
   - **Database C**
     - 4,617 projects
   - **Key words in project title, objectives, indicators, components**
   - **Clustering score**
     - above threshold for key concepts in the conceptual framework in objectives, indicators and components
   - **Saliency scores**
     - above threshold for key concepts in the conceptual framework in components

- **Included**
  - (357 projects)

- **Included Additional Financing**
  - Parent project met inclusion criteria

- **Nutrition portfolio**
  - (390 projects in 69 countries)

#gLocalEval2020
...process using machine learning
(nutrition evaluation)
2. Enhancing quality (and efficiency): knowledge extraction from existing (evaluative) evidence (*project insights*)

- Lessons from IEG’s private sector project-level evaluation reports used to be identified manually

- Given the fairly standardized reporting, an automated process for lessons classification was developed
... platform *(project insights)*
3. Broadening the range of questions: geospatial impact evaluation using satellite imagery data *(urban spatial growth evaluation)*
...comparison over time (urban spatial growth evaluation)
Prospects and challenges going forward

• Incentives and unit of analysis
• Building capacities
• Interaction between evaluator and data scientist
• Data-driven versus questions-driven evaluation

#gLocalEval2020
Discussion

To submit questions to the panelists, go to: www.pollev.com/iegnow
If you wish, add your name and location
Additional Resources

- For additional resources, visit https://ieg.worldbank.org/event/datascience-and-evaluation
Thank you for joining
Additional slides from Michael Bamberger
6. Case studies illustrating evaluation applications of big data
A. Using mobile phone data to assess integration of refugees in Turkey

• **Evaluation question:** How successful with government and UN programs in integrating refugees from Syria into Turkey.

• **Data sources:** Phone call-center records (Call detail records = CDRs).

• **Evaluation design:** Phone call records of refugees in Turkey were used to estimate trends in segregation, isolation and homophily (preference for people with similar characteristics to oneself). Communication and mobility patterns also provided insight into social integration.
B. Using satellites and drones to evaluate forest protection programs

- **Evaluation questions:** How effective were GEF programs to protect forest growth in protected forest areas

- **Data sources:** Satellites and drones combined traditional data collection

- **Evaluation design:** Collecting a wide range of indicators (distance to roads and settlements, slope, moisture content, forest activities etc) which permitted to the use of propensity score matching to create a quasi-experimental design
  - Pretest-posttest comparison group design with control
  - Using longitudinal data to measure trends over a period of years before project began and after it ended.
C. Social media analysis of the effectiveness of programs to increase women’s voter registration

- **Evaluation questions:** Effectiveness of UN Women’s programs in increasing women’s registration to vote in Mexico and Pakistan?
- **Data sources:** Analysis of Twitter (Mexico) and Facebook and radio call-in programs (Pakistan)
- **Evaluation design:** Several months were spent identifying tweets and posts relating to elections and voting. An analysis was also made of themes covered in radio call-in programs. The analysis calculated trends in the frequency of different kinds of references plus sentiment analysis to assess whether posts were favorable or unfavorable to voting.
D. Evaluating the effectiveness of a public child-welfare system in the US

- **Evaluation question:** The effectiveness of a management system to improve performance of a child welfare system.
- **Data sources:** Integrated data platform combining social worker action reports, agency records, and socio-economic surveys.
- **Evaluation design:** Artificial intelligence (AI) was used to construct natural experiments comparing cases where different social workers had made different recommendations on how to treat identical cases (domestic abuse, violent behavior, absenteeism from school etc). The analysis of the best action in each situation was used to recommend how future cases should be treated.