Non-traditional Data and Technologies for Development Evaluation

Reinabelle Reyes, Ph.D. (UNDP Pintig Lab)

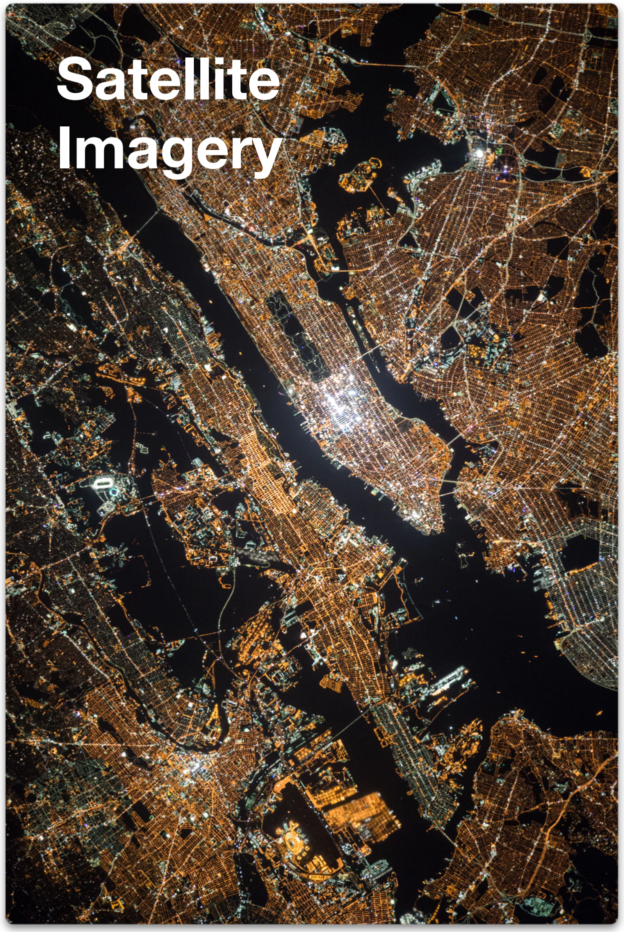
NEDA M&E Webinar Series

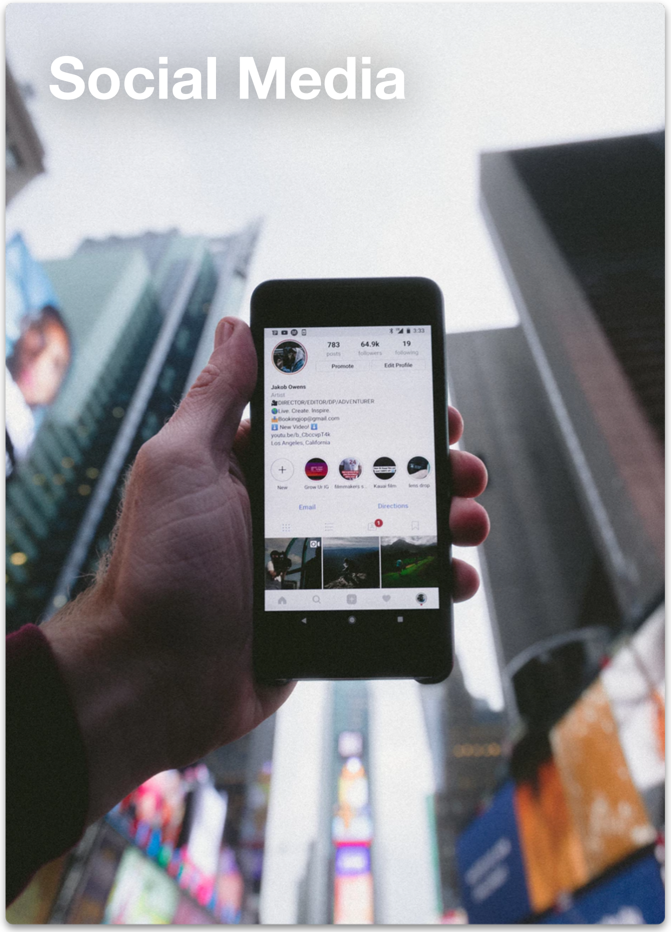
October 27, 2021

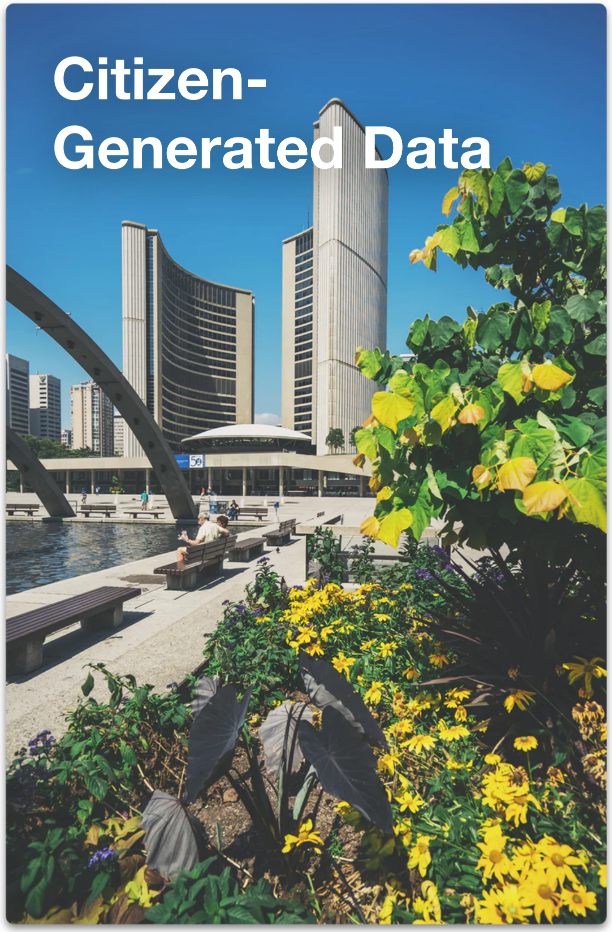
Agenda

- Context: “Data Revolution” and Emerging Technologies
- Opportunities and Challenges for Development M&E
- Use Cases: Non-traditional Data for Development M&E

New sources of real-time information about people are now available and accessible.







Opportunities for Development M&E

The data revolution for sustainable development is:

• The **integration of these new data with traditional data** to produce high–quality information that is more *detailed, timely and relevant* for many purposes and users, especially to foster and monitor sustainable development;

• The **increase in the usefulness of data** through a much greater degree of *openness and transparency*, *avoiding invasion of privacy and abuse of human rights* from misuse of data on individuals and groups; and the usefulness of data in *minimizing inequality* in production, access and use of data;

• Ultimately, **more empowered people, better policies, better decisions and greater participation and accountability** that will lead to better outcomes for people and planet.

United Nations Independent Advisory Expert Group (IAEG) on a Data Revolution, *A World that Counts*, 2014

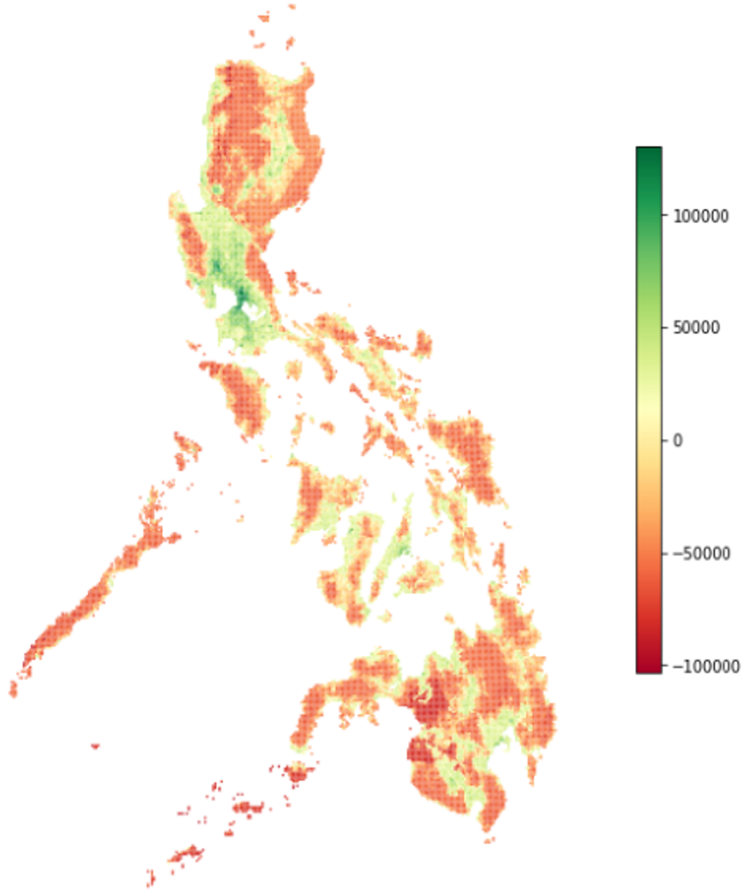
Challenges for Development M&E

- **Information is now becoming available in near real–time**, which requires new technologies for the collection, dissemination and use of this information, and new organizational processes and policies.
- **Monitoring and evaluation data typically demands high quality standards to be acceptable.** This can cause evaluators to reject or ignore new sources of data that could potentially provide valuable insights.
- **Data science and evaluation are grounded in different approaches to theory.** Data analytics has developed new approaches to impact evaluation using predictive modelling that employs an approach based on Bayesian probability analysis from the experimental methodologies generally used by development evaluators.
- There is a need for bridge building between data scientists and evaluators to allow for the development of a common language and to **identify promising areas where big data analytics can be applied in development evaluation contexts**.

UN Global Pulse, 'Integrating Big Data into the Monitoring and Evaluation of Development Programmes,' 2016

Use case for PH: Mapping poverty with AI

- *Context:* "No Poverty" is the first sustainable development goal
  - 20.8% of Filipinos live on less than $3.20/day in 2019.
  - International and humanitarian organizations are working together with government to uplift families from extreme poverty.
- *Objective*: Can we identify the locations of the most vulnerable communities?
- *Approach*: Combine accessible geospatial datasets and machine learning to generate a *high-resolution poverty map* of the Philippines



Estimated Wealth Index using AI Model

ZERO EXTREME POVERTYPH 2030

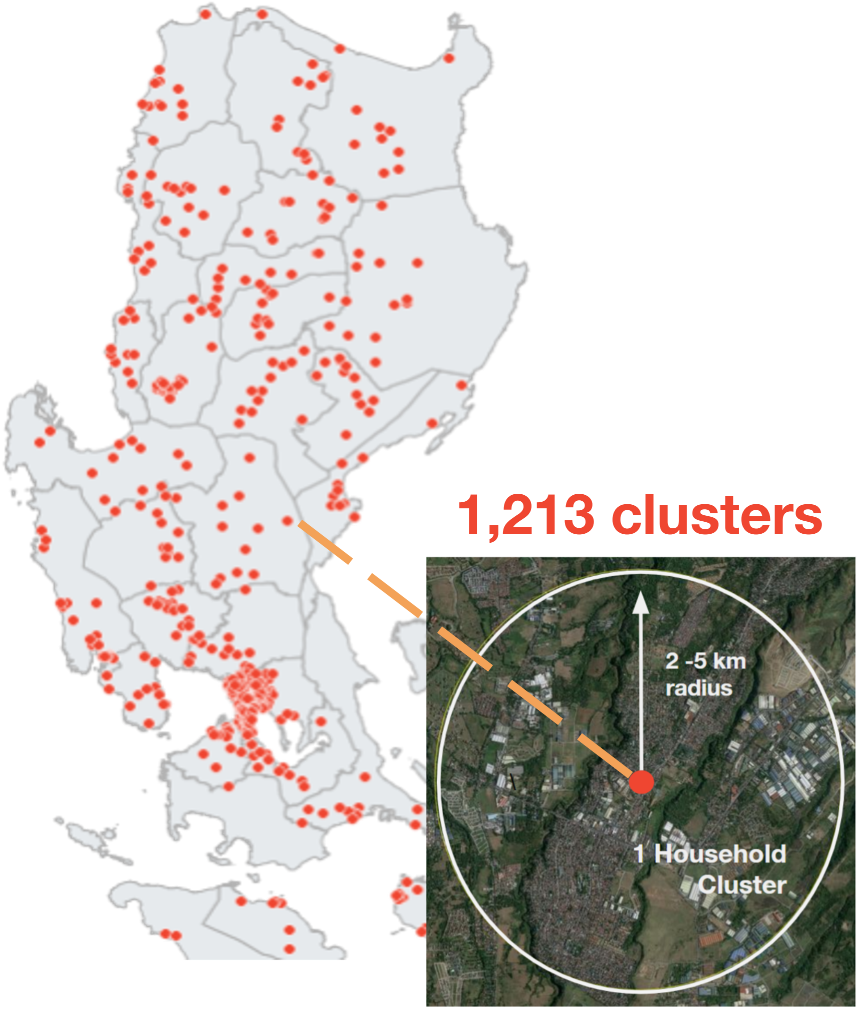
Thinking Machines

2017 Demographic and Health Survey (DHS)

Contains a number of socioeconomic indicators related to Asset Ownership, Education, Health, Sanitation and Hygiene

Wealth Index - primary indicator of socioeconomic well-being (see DHS Wealth Construction)

2017 DHS Data



Methodology: Geospatial datasets were combined to develop a machine learning model

Remote sensing

FB Data

POIs

Random Forest Regression Model → Socioeconomic Indicators

Nighttime Lights, NDVI, and Land Surface Temperature: Summary statistics for each cluster

Facebook Marketing API: Percentage of households with 4G/3G/2G/WiFi connection, Apple / Android device owners, consumer preferences

OpenStreetMap, CheckMySchool, and Department of Health Registry data: Number of banks, schools, hospitals, golf courses, cinemas, bars, hotels, parks, etc. in each cluster

Results: Model performance on wealth index

R-squared Results for Wealth Estimation

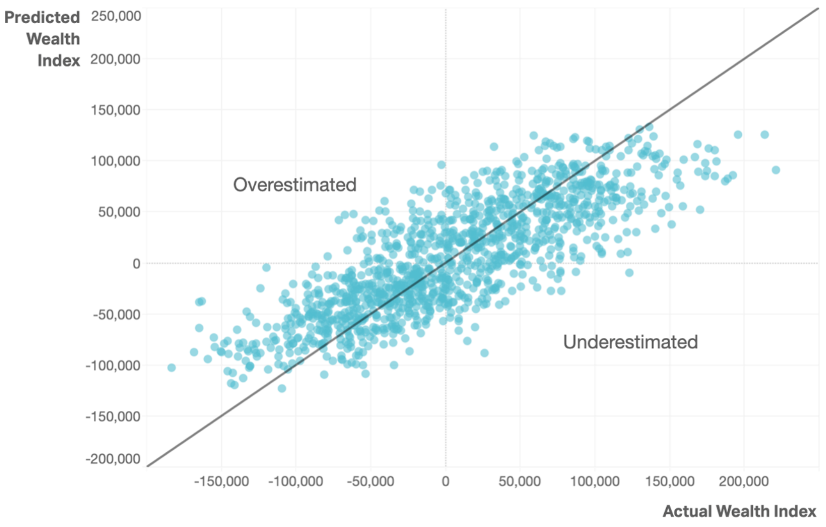
Wealth Index: The best model is able to explain **66%** of the variance.

Models	R^2
POI Data Only*	0.49
Remote Sensing Data Only*	0.59
Social Media Data Only*	0.55
Social Media Only (Fatehkia, et al.)	0.63
High Resolution Images and Deep Learning (Tingzon, et al.)	0.63
Our approach*	**0.66**

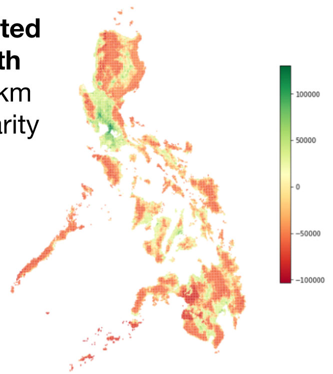
*Using a Random Forest Regression model which was evaluated using 5-fold cross validation

Our wealth model can explain up to 66% of the variance using low-cost, accessible datasets

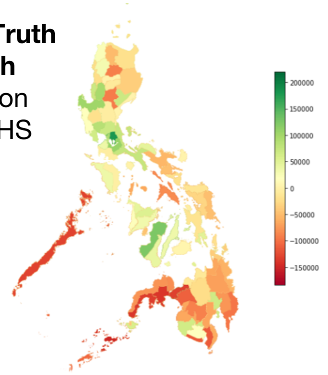
The chart below compares the actual vs. predicted average household wealth index for each of the 1200+ clusters surveyed in the 2017 Demographic and Health Survey.



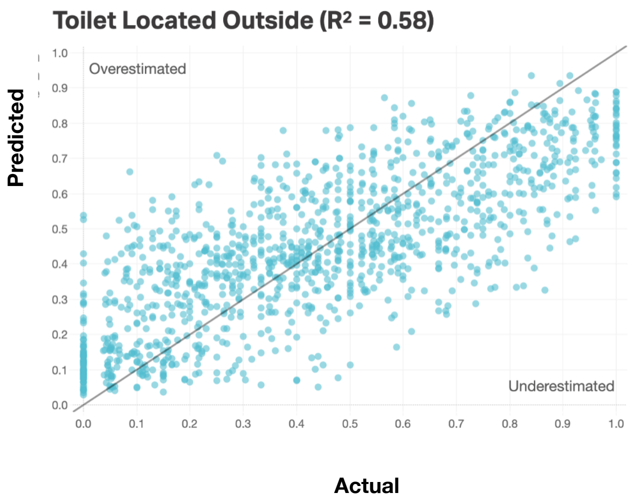
Estimated Wealth
18 sq km granularity

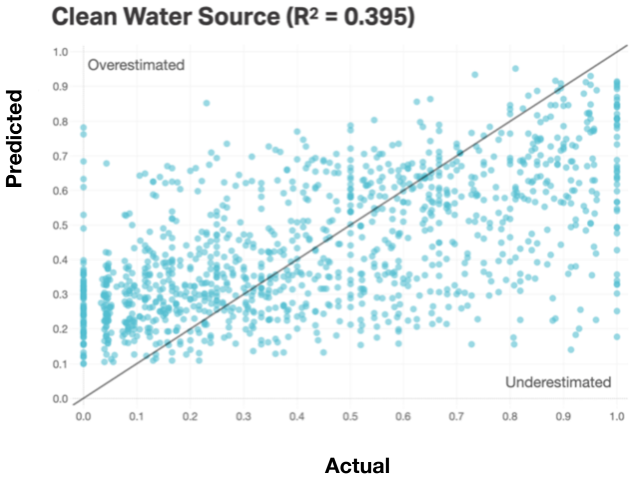


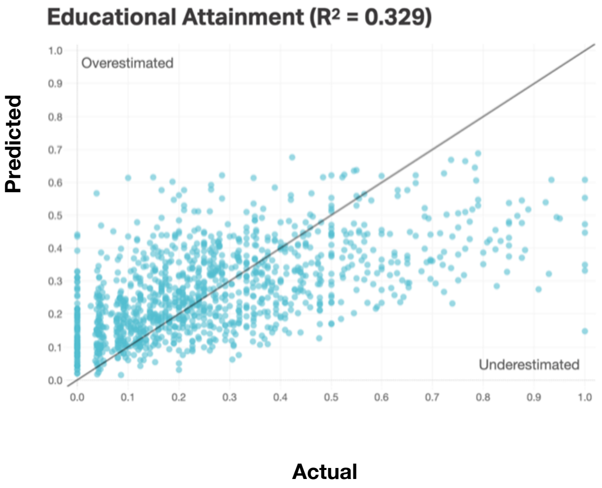
Ground Truth Wealth
Based on 2017 DHS



Results: Model performance on other socioeconomic indicators



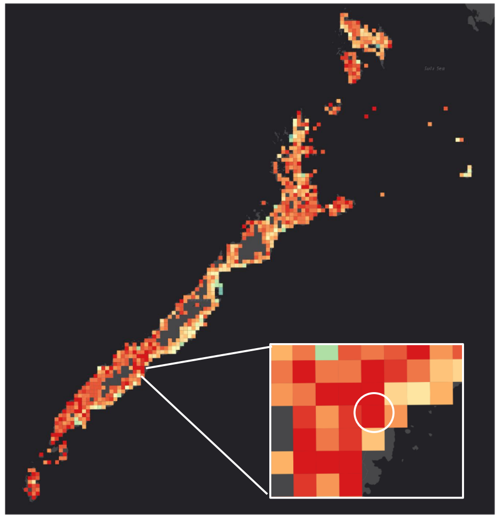


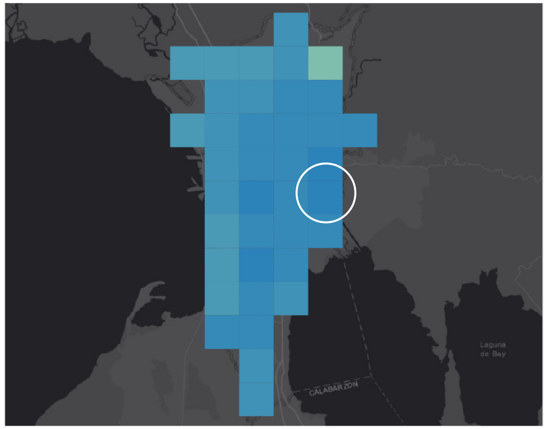


Note: Each data point represents the proportion of households with access to toilet, clean water, and higher education in each graph, respectively.

Results: Interpretable high-resolution poverty maps at 18 sq km resolution

SOFRONIO ESPAÑOLA, PALAWAN		MARIKINA CITY, METRO MANILA
0	No. of Banks	5
0	No. of Restaurants	14
0	Proportion of Population with 4G Access	39%
25%	Proportion of Population with 3G Access	<1%
25%	Proportion of Population with WiFi Access	14%
0.03	Nighttime Luminosity	11.72



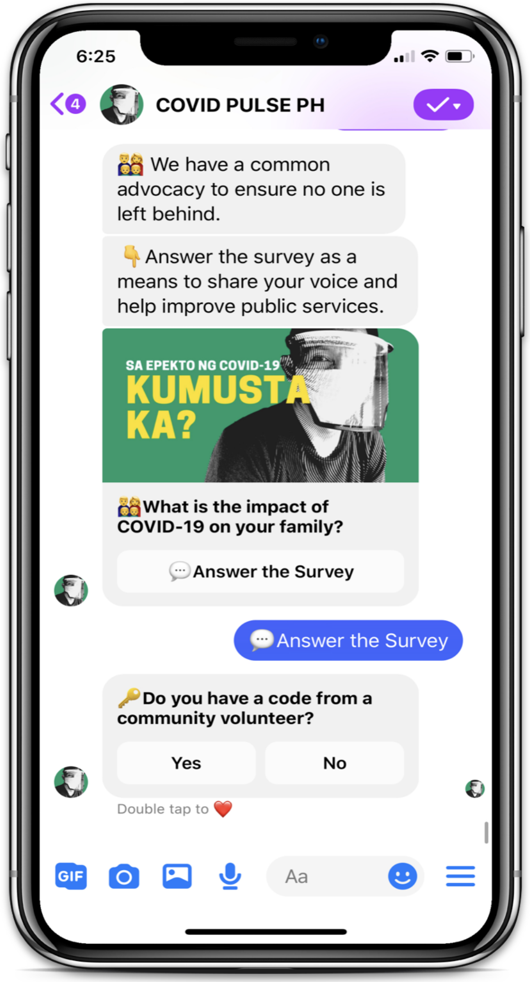


Legend

Lower Wealth Index — Higher Wealth Index

[1] For reference, the average *barangay* (neighborhood) size in the Philippines is ~7 square kilometers.

Use case for PH: Pulse surveys w/chatbots



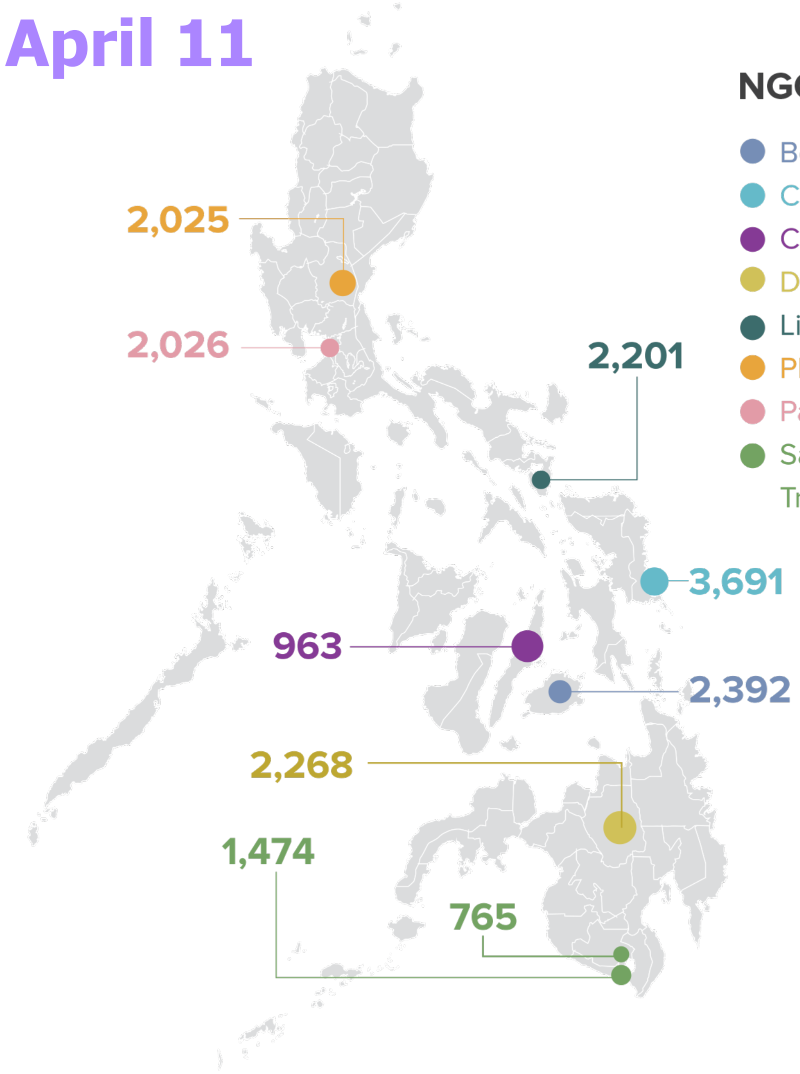
COVID PULSE PH

- Assess the impacts of COVID on poor households and find signals that will be relevant to recovery programming
- Test & scale an accessible tool to surpass mobility restrictions & reach across the digital divide
- Provide a safe space for poor households to give feedback
- Check change over time through data collection waves
- Phases 1 & 2 conducted in 2020 in Metro Manila; Granular lockdown survey conducted in Sept-Oct 2021

powered by: AI4GOV CODE NGO PRRM PHILSSA CENVISNET BIDEF Caritas Borongan Del Monte FOUNDATION SPECTRUM ICM

Survey Coverage

March 03 - April 11



NGO Partners

- Bohol Integrated Development Foundation
- Caritas Borongan & PRRM Eastern Samar
- Central Visayas Network of NGOs
- Del Monte Foundation (with DSWD Region 10)
- Lingap para sa Kalusugan ng Sambayanan
- PRRM Nueva Ecija
- Partnership of Philippine Support Service Agencies
- Sarangani Province Empowerment and Community Transformation Forum

Total effective sample size:

17,805 households

(for 18,872 total households reached)

Impact and Recovery

Impact of COVID19 in 2020

75% Decreased Income

47%
Loss of job or closed business

41%
Experienced hunger

36%
Difficulty in accessing health services

10%
Children had to stop formal schooling

Did Income recover in 2021?

6 out of 10
Reported Income worsened this March 2021 compared to last year

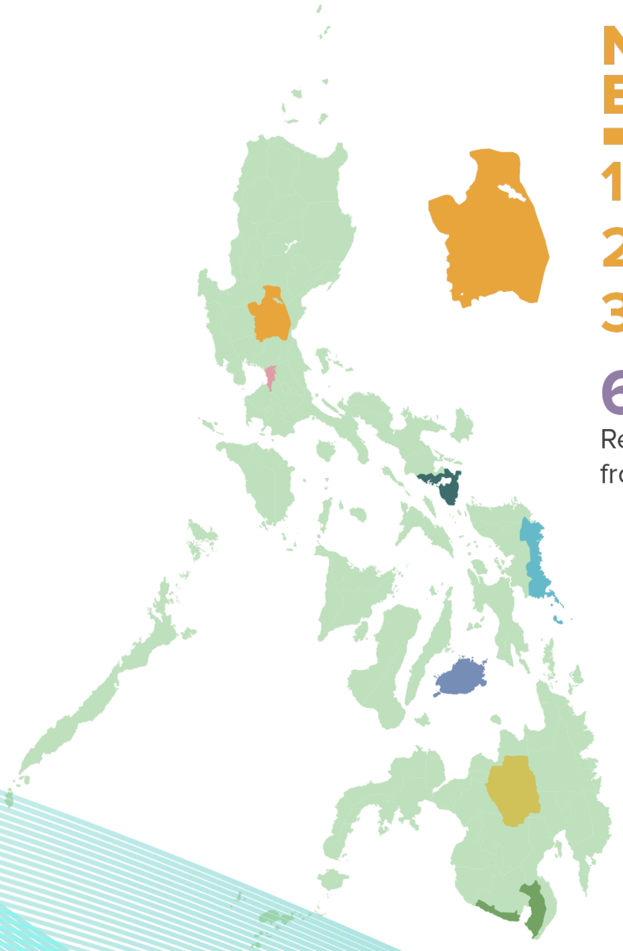
58.2% Worsened

29.6% The same

12.1% Improved

Factors for Recovery

Among those who were better off compared to 2020, how did they cope?



NUEVA ECIJA

1. Received support from government
2. Sideline work / business
3. Received support from family or friends

69% Received support from government

METRO MANILA

1. Received support from government
2. Sideline work / business
3. Received support from family or friends

70% Received support from government

BOHOL

1. Received support from government
2. Im back to my previous job
3. Sideline work / business

71% Received support from government

EASTERN SAMAR

1. Sideline work / business
2. Received support from government
3. Received support from family or friends

58% Received support from government

SARANGANI

1. Received support from government
2. Sideline work / business
3. Im back to my previous job

63% Received support from government

BUKIDNON

1. Sideline work / business
2. Received support from government
3. Received support from family or friends

71% Received support from government

SORSOGON

1. Received support from government
2. Sideline work / business
3. Received support from family or friends

68% Received support from government

Emerging Personas

*Summary of significant drivers for each segment based on **household characteristics***

Persona X
Poor

📍 **High proportion of respondents from Bukidnon, Cebu, Nueva Ecija**

⬆️ **High proportion of reported decrease in income**

🌐 **With higher proportion of daily internet access than the average**

🌟 **Higher proportion of Human Capital vs the average**

🔒 **Have savings/capital as safety net**

🔺 **% for Access to Market Services:**
- Microfinance institutions
- Street lenders
- Rural or other Banks

Persona Y
Extremely Poor

📍 **High proportion of respondents from Bohol, Bukidnon, Eastern Samar, Saranggani and GenSan**

⬆️ **High proportion of reported loss of jobs/business & experience of hunger**

🌐 **With lower proportion of daily internet access than the average**

🌟 **Higher proportion of Social and Physical/Economic Capital vs the average**

🔒 **Relied on friends/family & government for safety nets**

🔺 **% for Access to Market Services:**
Microfinance institutions

Persona Z
Better Off

📍 **High proportion of respondents from Cebu, Metro Manila, Sorsogon**

⬆️ **High proportion of those who did not experience a negative effect due to COVID**

🌐 **With high proportion of daily internet access than the average**

🌟 **Higher proportion of Human Capital vs the average**

🔒 **Have savings/capital, insurance and company aid as safety nets**

🔺 **% for Access to Market Services:**
- Rural or other banks
- Government & private techvoc training

Methodology: Cluster analysis then used Chi-square test of independence to check which criterion in the clusters have an association

Summary and Takeaways

- The availability and accessibility of new sources of real–time information provides opportunities for development program M&E to **integrate these new data with traditional data** to produce high–quality information that is more *detailed, timely and relevant*

- This digital and data revolution are linked to new data-driven techniques (machine learning, AI, data analytics) that can help to **identify development needs, plan, implement and evaluate development programs**. Future M&E systems are likely to be more closely linked to broader systems encompassing program identification, design and management.

- Many development and government agencies are still in the process of defining their policies on the use of non-traditional data and digital technologies. Open dialogue and collaboration between M&E practitioners and data scientists need to continue to bridge the two practices, address existing challenges, and **identify promising areas where new data sources and techniques can be applied in development evaluation contexts**.

Contact us:

UNDP Pintig Lab

c/o Dr. Reinabelle Reyes
Chief Technical Adviser
Pintig Lab, UNDP Philippines

Email:
reinabelle.reyes@undp.org