Centre of Excellence for Development Impact and Learning

FEBRUARY 2021

CEDIL Methods Brief 2

Using big data for impact evaluations

Background

One key challenge in evaluating development programmes is a lack of data. This can be due to the inaccessibility of target populations, inadequate aggregation of data, data collection lag times, and data being missing. In some contexts, like pandemics, conflicts, and humanitarian emergency situations, data collection is completely impossible. Data gaps are particularly significant for populations where the need for evidence-informed policy decisions may be greatest.

Big data offers great potential to fulfil some of these data needs and to help establish which policies or interventions work, even in contexts where traditional methods of data collection are challenging. Integrating big data with traditional household surveys and administrative data can improve data availability, quality, granularity, accuracy, and frequency, and can help measure development outcomes over time and space in a number of new ways.[1] Though big data can be used with experiments, it is most commonly used with quasi-experimental methods, which require careful research designs.[2] Big data can:

- measure development outcomes that are difficult to measure using household surveys or administrative data, such as economic output, wealth, population movement, or disease spread within a given local area;
- identify comparison groups (because big data are generally available before and after programmes, and for programme areas and comparison sites);
- measure long-term programme impacts and sustainability (because some sources of big data are available for a reasonably long period);
- provide data on pre-programme trends and control variables for better statistical precision;
- provide data to evaluate the different impacts on different sub-groups;
- aid in conducting robustness analyses based on multiple comparison groups and placebo tests; and
- evaluate the impact retrospectively and at the level of individual projects, as well as at programme/portfolio level.

The United Nations Global Pulse[3] defines big data as being: digitally generated; passively produced as a by-product of digital services, transactions, and interactions; automatically collected; and geographically or temporally trackable. The amount of data that fits these criteria has exploded with recent innovations in satellites, sensors, mobile devices, call detail records (CDR), social media applications, and digital business records.

Steps in building an impact evaluation based on big data

There are well-documented best practices for preparing for an impact evaluation,[4,5] including developing a theory of change, identifying the 'treatment', and framing the evaluation questions from the theory of change. However, at the moment, few big data studies provide an adequate explanation of the programme, its theory of change, and how the evaluation questions are derived.[6] This section of the brief outlines some points researchers should consider when designing impact evaluations based on big data, although not all the steps will be relevant to every evaluation.

Put together a multidisciplinary team

Big data impact evaluations require multidisciplinary teams consisting of data experts, evaluators, subject matter professionals, and context specialists who can identify groups that might missed by the data source. Increasing availability of pre-processed satellite data that could be readily used in evaluations and ready-made algorithms that process CDR data augurs well for widespread use of big data. However, for most evaluations, the data needs would require remote sensing scientists or big data analysts to develop customised processes to collect and clean these data. Further, sector and context knowledge are required to understand the underlying processes that have generated the data (or, rather, that are missed and/or 'not explained' by the data). For example, access to technology and gender norms will influence who uses mobile phones, and therefore CDR data need to be carefully interpreted in order to make inferences about the entire population.

Consider the unit of implementation and analysis

The unit of analysis is the level at which a programme intervenes. Some programmes target individuals or households, while others target villages, districts, or larger areas. Evaluations with big data are more often suited to programmes in the latter category. One example of such an evaluation is a 2019 evaluation measuring the impact of mining activities on surrounding areas, which used 1 km square cells.[7]

Identify the treatment population or area

The target population or area can be based on whoever might be affected by a given programme. For example, in measuring the effect of newly developed road segments in the West Bank and Gaza in Palestine, one evaluation identified a 5 km buffer around each new road as the target population of interest.[8] The potential for programme spillovers into neighbouring areas should be considered when measuring such impacts.

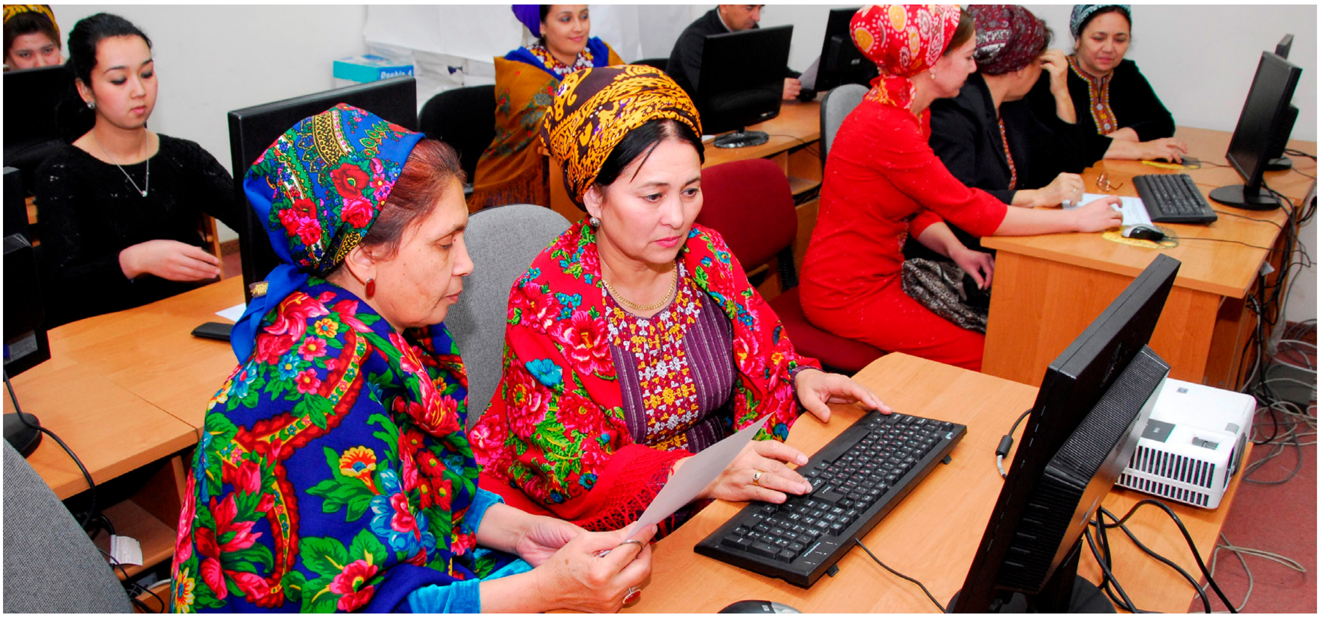
Identify the control group

One of the key value additions of big data is in identifying a valid comparison group and constructing the counterfactual. The availability of geospatial data that can be aggregated at the necessary geographical scale, such as sub-districts or 1 km by 1 km squares, allows geospatial impact evaluations to exploit geographic cut-offs between individuals who were 'treated' by a programme and those who were not. Programme placement and spatial reach will determine the potential comparison areas. Again, the possibility of programme spillovers into neighbouring areas should be considered.

Box 1: Detecting oil spills in the Amazon using drones

Measuring oil spills in remote forests is expensive and often regulators seeking to act against oil companies are handicapped by a lack of data and evidence. In order to measure oil spills in remote areas of the Ecuadorian and Peruvian Amazon, a research team[9], with the support of indigenous organisations, developed a package of digital monitoring tools for local communities to use. The package was made up of inexpensive digital tools, including mobile phones, drones, and open source apps. The custom-made drones collected aerial images of oil spills in difficult-to-access areas of the Amazon. Local youth were trained as monitors, and the technology also enabled these local community monitors to share the reports with the authorities and media, maximising the possibility that the data would be acted on.

This study employed a randomised phase-in design over approximately 24 months with 24 monitoring teams, 12 each in Ecuador and Peru. At the end of the intervention all the groups had received the monitoring package. The teams who had not yet been phased in served as a control group. The evaluation showed that the intervention increased detection and reporting of oil spills.



Consider the construct validity of the outcome variable

In order to use big data in an impact evaluation, a source of data, such as satellite images, CDR, or social media data, must contain information relevant to measuring the variable of interest. This can be ensured by calibrating and testing the big data-based measure against a traditional data source, to ensure it is measuring the outcome variable well. For example, satellite data on the brightness of lights at night is believed to measure economic activity based on human-generated light. The validity of using nightlights as a proxy for economic activity can be tested by considering the GDP at the sub-national level calculated based on the conventional measures. Similarly, a research team[10] used anonymised transaction data from a large electronic payment technology company to estimate the impact of hurricanes on consumer spending. In order to test the validity of this measure, they compared the new data with a conventional consumer spending index based on the US Census Bureau's monthly retail trade survey. They showed that their measure based on big data was similar to the official measure, but included higher frequencies and sub-national detail, and was better suited to studying localised or short-lived economic shocks.

Box 2: Is night light a valid measure for economic activity?

Night light data are often used as a measure of local economic activity at the sub-regional level. This measure is particularly valuable in difficult-to-reach areas and smaller administrative units, such as villages, counties, and municipalities. Several studies have used night lights to evaluate the impact of newly developed infrastructure or foreign direct investment on local economic activity. Night light can also be a good proxy for human development at the local level.[11]

However, there are a few technical challenges in using night light data to measure economic activity. Highly-lit areas could seep into the near-by cells, thus biasing the results. Growth may not be evident in highly-lit or poorly-lit areas because of over-saturation or lack of measurement precision.[12] Temporary, sharp spikes and drops in night light may not necessarily reflect changing economic activities, but could be the result of technical errors. Using annual averages without correcting for these spikes and drops could be misleading. Further, a study[13] showed that night light estimates as a proxy for economic activity in a specific context cannot automatically be generalised to other contexts. Hence, this new measure is promising, but researchers should exercise caution in handling, interpreting, and generalising the findings based on it.

Check for baseline balance and pre-programme parallel trends

Quasi-experimental research designs require pre-programme balance between the treatment and the comparison groups across numerous measures. Further, the timing of data collection is of particular importance in impact evaluations – pre-programme data should be collected immediately before a programme begins, and post-programme data should be timed to capture a programme's potential effects. For example, the impact of a new canal irrigation programme can be measured only after a growing season.

Satellite imagery and other forms of big data can be helpful here, because several periods of pre- and post-programme data can often be collected cheaply. For example, in an evaluation of a deforestation-prevention programme, the primary analysis was conducted with images collected immediately before the programme (in 2011) and after the programme (in 2013). To improve statistical precision, the researchers also used images from 1990 to 2010.[14]

Control for covariates

Big data provides an opportunity to measure and control for a number of localised, time-varying factors, such as temperature, precipitation, or vegetation cover, that may affect the outcomes of interest. Introducing appropriate control variables can increase the precision of impact estimates, and big data sources allow for the inclusion of covariates that cannot easily be measured in other ways.

Analyse the heterogeneity of effects

Any given treatment may have different effects on sub-groups. We often overlook these variations due to data limitations, but big data can help overcome these limitations. For example, satellite data could help assess programmes' different effects on areas that were previously wealthier (based on night light) or further from a new road.[15] Data availability on the entire treatment and comparison areas, greater granularity, and limited additional costs in collecting these data help in conducting sub-group analyses.

Conduct robustness analyses

Big data allows researchers to use multiple comparison groups and outcome measures to assess the robustness of the primary research design. Such additional robustness checks could be very expensive, or impossible, using traditional data sources like surveys.

Assess implementation fidelity and contamination of comparison group

For any impact evaluation, understanding all aspects of the programme, including the quality of implementation, is critical when interpreting the evaluation findings. The development community has long experience in studying implementation fidelity, and big data impact evaluations should draw on this experience. One particular implementation concern that big data evaluations should check for is contamination of the control group, which is when a programme inadvertently affects the control group, even though it did not target them.

Explore long-term impact

In most cases, big data will allow for the the collection of data at several different points in time, both pre- and post-programme. The ability to collect additional data without undertaking field visits enables measurement of the long-term programme impacts.

Box 3: A railway's long-term impact

To measure the long-term effects of a new railway segment in India, a research team[16] combined satellite imagery with census data to conduct an impact evaluation. The railway was built between 1991 and 1998 and included 760 km of new track. The study measured effects on economic growth, livelihood opportunities, regional integration, environmental degradation, displacement, and losses of livelihoods. It used a mixed-method and quasi-experimental approach, considering areas further away from the railroad as a control group. Satellite data were used for measuring the land use pattern and census data were used to measure socioeconomic outcomes, such as changes in livelihoods and the composition of the work force. In order to measure the long-term impact, the study compared satellite images from 1991, 2001, and 2011.

The study found significant environmental degradation in the short run, when the construction happened, but little degradation in forest cover close to the railway project area over the period from 2001 to 2011. This study is an excellent example of combining census data with big data sources to evaluate a programme's impact over a long period retrospectively.

Report on data quality

Most big data studies – 87% – in the International Initiative for Impact Evaluation's (3ie) systematic map of big data studies do not report on data quality issues, representativeness, construct validity, or generalisability.[17] This gap points to the need for better-quality reporting standards.

Provide transparent data and code for replication

Publishing all relevant materials, including data and code, allows for independent verification. However, there are two challenges here for research drawing on big data. First, some of the major sources of big data, such as CDR from phone companies, are proprietary and may not be allowed to be shared beyond the closed group of researchers. In those cases, researchers should at least share the code and algorithms in the public domain for verification. Second, the data must be de-identified before it can be shared, to preserve individuals' anonymity, but studies have shown that re-identification of subjects is a real possibility. Therefore, it is crucial to check whether there are combinations of variables that could be used to re-identify research subjects.

Report on ethical challenges

Ethical challenges, such as consent, data privacy, and data security, are well documented with respect to impact evaluations. However, the use of big data may present new challenges, such as difficulty in obtaining explicit permission from the phone users, cyber security risks, and inadvertent exclusion of certain sections of the population.[18] Aside from satellite data, most big data sources that use human-generated data without individuals' explicit consent for secondary use should be reviewed by institutional review boards.

Report on the cost of collecting, analysing, and storing big data

There is very little publicly available information on the cost of collecting, analysing, and reporting big data. There can be multiple hidden costs, such as staff time or the costs of necessary computing infrastructure, like data storage. In several big data studies, the actual data collection costs for satellite or CDR data were nominal, but the costs of analysing and reporting the data (in staff time and the computing resources) accounted for 35% to 50% of the total cost of the study.[19]

Conclusion

Big data can contribute to the evidence base in development sectors where evaluations are often infeasible due to data issues. Given the rapidly increasing availability of big data and improving computation capacity, there is a great potential for using big data in future impact evaluations. Big data can also contribute to evaluations through providing new ways to identify control groups and establish counterfactuals, and can strengthen the analysis with data on pre-programme trends, covariates, and sub-groups, as well as enabling better robustness analyses. However, there are several analytical, ethical, and logistical challenges that may hinder the use of big data in impact evaluations. Standards should be set for the reporting of data quality issues, data representativeness, and data transparency. More interaction is needed between big data analysts, remote sensing scientists, and evaluators.

About this brief

This brief is based on

Rathinam, F., Khatua, S., Siddiqui, Z., Malik, M., Duggal, P., Watson, S, and Vollenweider, X. 2020. 'Using big data for evaluating development outcomes: a systematic map'. *CEDIL Methods Working Paper 2.* Oxford: Centre of Excellence for Development Impact and Learning (CEDIL). Available at: https://doi.org/10.51744/CMWP2

Suggested citation

Rathinam, F., Thissen, P. and Gaarder, M. 2021. 'Using big data for impact evaluation', *CEDIL Methods Working Brief 2*. Oxford: Centre of Excellence for Development Impact and Learning (CEDIL). Available at: https://doi.org/10.51744/CMB2

Endnotes

[1] Rathinam, F., Goldblatt, R., Glandon, D., and Araya, G. (2020) *Taking a Closer Look at Impact from Hundreds of Kilometers Above the Earth*, GIM International.

[2] Rathinam, F., Khatua, S., Siddiqui, Z., Malik, M., Duggal, P., Watson, S, and Vollenweider, X. 2020. 'Using big data for evaluating development outcomes: a systematic map'. *CEDIL Methods Working Paper 2*. Oxford: Centre of Excellence for Development Impact and Learning (CEDIL). Available at: https://doi.org/10.51744/CMWP2

[3] United Nations Global Pulse (2013) 'Big data for development: A primer', United Nations Global Pulse.

[4] White, H. (2011) 'Achieving high quality impact evaluation design through mixed methods: the case of infrastructure', *Journal of Development Effectiveness* 3(1), pp. 131–144.

[5] White, H. (2013) 'An introduction to the use of randomised control trials to evaluate development interventions', *Journal of Development Effectiveness* 5(1), pp. 30–49, DOI: 10.1080/19439342.2013.764652

[6] Rathinam, Khatua *et. al.* (2020) 'Using big data'. Available at: https://doi.org/10.51744/CMWP2

[7] Pande, R. and Sudarshan, A. (2019) 'Harnessing transparency initiatives to improve India's environmental clearance process for the mineral mining sector', *3ie Impact Evaluation Report 92*, 3ie, New Delhi. Available at: https://doi.org/10.23846/TW8IE92

[8] BenYishay, A., Trichler, R., Runfola, D. and Goodman, S. (2018) 'Final report: Evaluation of the Infrastructure Needs Program II', AidData at William & Mary, Williamsburg, VA.

[9] Pellegrini, L (2019) 'Community monitoring of socio-environmental liabilities with advanced technologies in the Ecuadorian and Peruvian Amazon', *3ie Grantee Final Report,* 3ie, New Delhi.

[10] Aladangady, A., Aron-Dine, S., Dunn, W., Feiveson, L., Lengermann, P., & Sahm, C. (2019). 'From transactions data to economic statistics: Constructing real-time, high-frequency, geographic measures of consumer spending' (No. w26253). *National Bureau of Economic Research*.

[11] Bruederle, A., and Hodler, R. (2018). 'Nighttime lights as a proxy for human development at the local level', *PloS one*, *13*(9), e0202231.

[12] Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., & Ermon, S. (2016). 'Combining satellite imagery and machine learning to predict poverty', *Science*, *353*(6301), 790-794.

[13] Head, A., Manguin, M., Tran, N., & Blumenstock, J. E. (2017, November). 'Can human development be measured with satellite imagery?', In *Ictd* (pp. 8-1).

[14] Jayachandran, S. et al (2017) 'Cash for carbon: A randomized trial of payments for ecosystem services to reduce deforestation', Science 357(6348), pp. 267–273.

[15] BenYishay *et al*. (2018) 'Final report'.

[16] Jaiswal, S., Bensch, G., Navalkar, A., Jayaraman, T., Murari, K. and Patnaik, U. (2020) 'Evaluating the impact of infrastructure development: Case study of the Konkan Railway in India', *3ie Impact Evaluation Report 114*, 3ie, New Delhi.

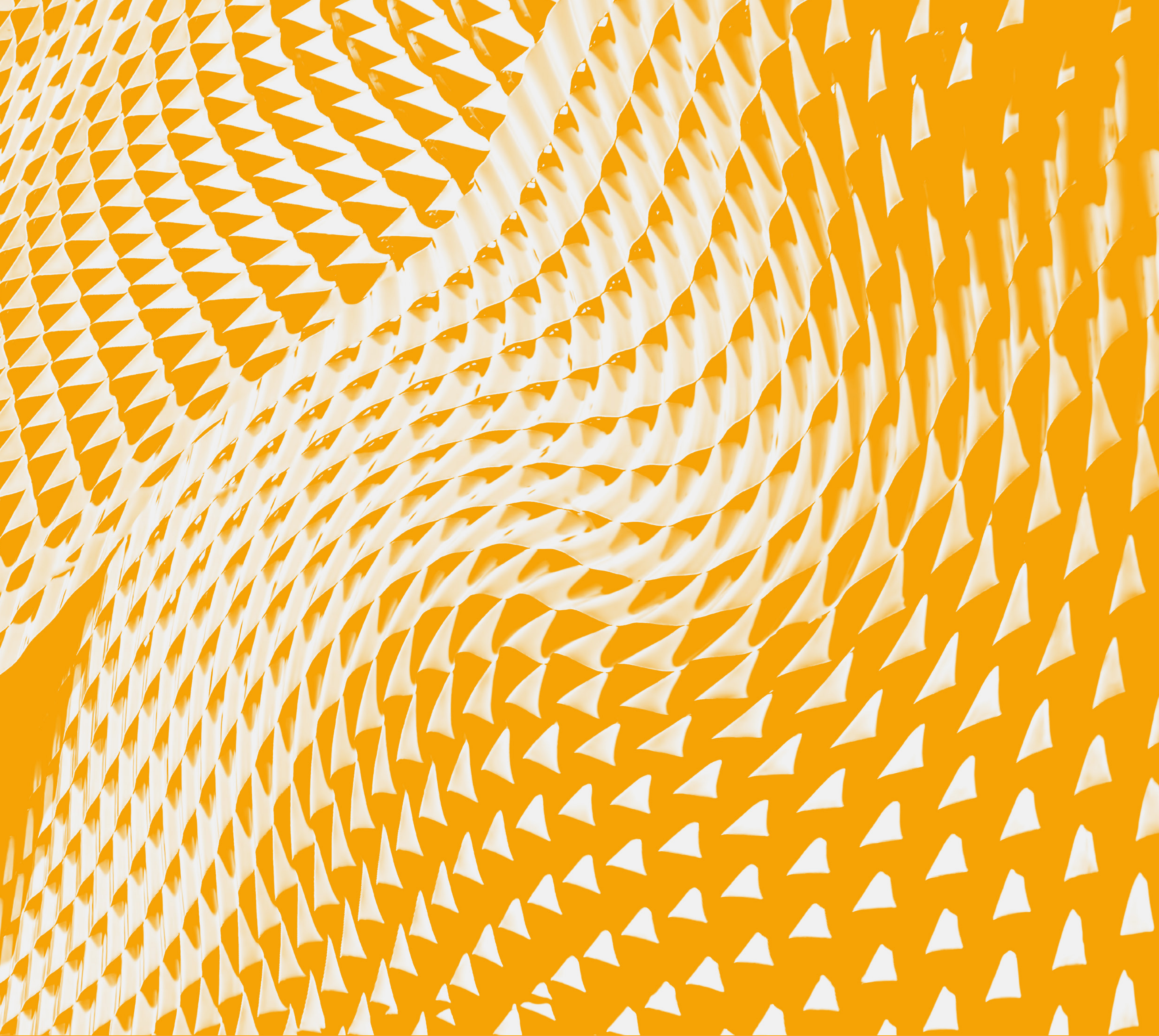
[17] Rathinam, Khatua *et. al.* (2020) 'Using big data'. Available at: https://doi.org/10.51744/CMWP2

[18] *Ibid.*

[19] *Ibid.*

Photo credits

p.1: Bao-Ping Zhu, CDC Global, p.3: World Bank, p.4: Daniel Bachhuber p.6: Sergio Hola



@CEDILProgramme

Centre of Excellence for Development Impact

Centre of Excellence for Development Impact and Learning

UKaid from the British people

International Initiative for Impact Evaluation

About CEDIL

The Centre of Excellence for Development Impact and Learning (CEDIL) is an academic consortium supported by the UK Government through UKaid. The mission of the centre is to test innovative methodologies in evaluation and evidence synthesis and to promote evidence-informed development. CEDIL-supported projects fall into three programmes of work: evaluating complex interventions, enhancing evidence transferability, and increasing evidence use.

For more information on CEDIL, contact us at **cedil@opml.co.uk**
or visit our website **www.cedilprogramme.org**