

Title: Using data differently and using different data

Authors: Ingvild Almas, Orazio Attanasio, Francisco Oteiza

This paper will address the challenges of data and measurement in evaluation. Problems of lack of adequate measures to capture relevant factors and measurement error in existing measures often constitute the main impediment to robust evaluation. One way in which this has been addressed in the field of evaluation is through the introduction of Randomised Controlled Trial methodology. In some instances, this approach allows to overcome concerns about the omission of unmeasured or unmeasurable relevant factors in the estimation of impacts of interest to a greater degree than other available methods. Indeed, if one is interested exclusively in measuring the impact of an intervention on a certain outcome, within a properly executed RCT the only measures necessary are those of the outcomes of interest. However, RCTs do not necessarily help with (a) adequately measuring outcomes of interest; (b) measuring factors which relate to the mechanisms of estimated impacts; and (c) conducting robust evaluation in areas where the RCT methodology is not feasible. Therefore, innovation in the field of measurement and, more generally, considering several data sources and available data remain critical to the advancement of rigour and breadth of questions that can be answered through evaluation.

In this paper, we will discuss four categories of related approaches to innovation in data and measurements relevant for evaluation, motivating the relevance of each of the categories, providing an overview of the key issues demonstrated by a few detailed examples and finishing with an outline of promising directions for future work.

1. We start with a discussion of advancements in and potential for *construction of new measures*. Evaluation of impacts of policies and interventions is almost always hindered by lack of appropriate measures to capture core parameters. For example, expectations of key actors about returns to an investment or probability of an adverse event happening are an important determinant of how they respond to a policy which aims to boost that investment or offer insurance against that adverse event. At the same time, investment choices will depend on the beliefs of the relevant actors, about their usefulness, beliefs that are not necessarily correct. Questions about who within the household should be the recipient of a cash transfer require an understanding of how spending and investment decisions are made within the household. Often appropriate

measures of these beliefs, expectations, intra-household bargaining positions and many other factors are missing, however. There is large scope for more work on development of these. Such measures would not only improve the estimates of impact evaluation but would be key in shedding light on the mechanisms behind such impacts. The challenge is to develop measures that meet the needs of specific models that we want to estimate both in terms of the theoretical construct we are interested in as well as the empirical requirements for application of appropriate estimation methods.

2. The second category of approaches relates to the *use of multiple measures*. Even in areas where measures do exist they often only partially capture the underlying construct of interest. For example, while many different IQ tests exist, no one would argue that there is one that truly and completely captures human intelligence. Similarly, there is a growing number of satellite data products that can be used to measure and predict crop growth. To capture appropriately the weather conditions over a growing season, one requires a combination of measures, such as the total precipitation, the degree of humidity or the number degree-days, as well as a model of how these affect crop developments. Leaving any of these components out of the estimation can severely bias our results.

The first best approach in the face of imperfect measurement is therefore to use multiple incomplete/flawed measures constructed from the growing volume of data available, including Big Data, and appropriate statistical techniques such as machine learning and factor analysis to combine these to construct more accurate measures of the variables of interest.

3. There is potential in *integrating different sources of data* (e.g. survey, administrative, satellite) to create a resource suitable for addressing a more nuanced set of evaluation questions. Even in developing countries, the possibility of using administrative data are growing and they can be usefully combined with survey data to assess both the processes of implementing a scaled-up intervention and its impacts. In some contexts, it is possible to link tax return data to firm and household surveys. Similar considerations apply to alternative data sources, including Big Data. For instance, satellite data on land use, roofing materials and night lights can be used to infer variation in incomes across space or over time, where appropriate measurements are missing.

The combination of large amounts of data and supervised machine learning algorithms can therefore help overcome the problem of missing measures, in many contexts.

4. Finally, we will talk of the potential and the usefulness of *integrating quantitative and qualitative data sources*. The use of qualitative evidence, coming from focus groups or similar approaches, is often neglected and in any case often considered separate from the use of quantitative data. We will stress how each approach can benefit from the other.