

## What's the evidence?

An APPP perspective on quality issues in governance policy research

David Booth<sup>1</sup>

**T**he Africa Power and Politics (APPP) synthesis report<sup>2</sup> argues for a shift in approaches to African governance. The quality of the evidence used to make this argument matters for three main reasons:

- as researchers, we are committed to meeting scientific standards
- we need to be able to show that we have solid evidence to back up what we say, and
- our funders want to assess the robustness of our findings.

The obvious question that arises is: what is meant by quality of evidence or robustness of findings? The answer may appear simple but, unfortunately, it is not. This policy brief explains why, drawing examples from the larger field of policy research on governance for development as well as from APPP studies in particular.

It is a good moment to be considering these issues. The textbook literature on social science research approaches has grown explosively over recent decades, and the fresh insights and clearer standards this has brought to the fore are not yet widely appreciated. This is an additional reason for adding some methodological reflections to APPP's final reporting.

### Evidence quality and research design

Quality of evidence has been a central concern of APPP from the outset. The first phase of our research included an extended exploration of the approaches to causal inference and theory-building that would be appropriate, given the questions we were asking.<sup>3</sup> This was based on the view that research designs need to be driven by research questions, not the other way round.



APPP synthesis report © APPP

The best governance research is empirical, aiming to uncover causal processes and formulate testable hypotheses by inference from evidence or experience. Modes of inference differ, depending on whether there is an existing body of relevant theory and whether the relevant data are quantifiable or not. However, a core set of principles applies to social science as science.<sup>4</sup>

As a rule, the robustness of an inference depends on two factors: the quality of the data and the use that is made of the data in generating and testing hypotheses. The latter is partly about observing sound logic and partly about making the best possible use of an appropriate research design. In applied research, inference quality also includes aspects of the process by which findings get adopted, or not, into guidelines that shape policy.

It might be thought obvious that data quality is the more important of these issues. In fact, quality of inference is no less significant as a source of scientific weakness in policy research. We take the two topics in reverse order therefore.

## Facts don't speak for themselves

Quality of inference starts with sound logic and ends with good research design. It might seem that the first can be taken for granted and that the second is normal. Unfortunately, that is not so.

### Five pitfalls

At least five failings have been relatively common in recent policy-oriented theorising on state-building, governance reform and development:

- tautological argument, with insufficient distinction between causal claims and propositions that are true by definition – as when claims about 'good' governance mix up evidence on effects with normative claims about what is desirable
- teleological reasoning, where an institutional form is justified by reference to the functions it performs or its assumed historical end-point – as when an ineffective institutional arrangement is defended as a precondition for state-building or democratisation
- inferences of causality based only on evidence of statistical association or the regular co-occurrence of events or factors – such as inferring that reductions in agricultural market distortions are caused by multi-party electoral competition, when a third factor such as donor influence may be responsible for both better macroeconomic management and multi-partyism
- claims about impact that are based on simple before and after evidence, with no effort to address the counter-factual situation – as in some classic forms of NGO advocacy around the social impacts of economic liberalisation
- propositions suggesting simple causal relationships when, based on the totality of the evidence collected, multiple or some other form of complex causation may be suspected – as in the studies most often cited on the power of information and budget transparency to improve public service provision.

### Avoiding the pitfalls

Much of the reasoning with which we take issue in the APPP synthesis report falls squarely into one or more of these traps. Of course, it is only possible to make judgments about robustness of inference with reference to particular propositions, not a whole body of work. However, we would claim

**“ quality of inference is no less significant as a source of scientific weakness.”**

that the APPP findings and analytical conclusions stand up reasonably well in terms of avoiding these classic pitfalls.

Generally, we have taken care to formulate our research questions and tentative hypotheses in strictly causal terms, avoiding both tautology and teleology. We have made efforts to tailor our research designs to the questions being asked by setting up the most telling feasible comparative investigations, with due consideration of the possibility of complex causal pathways.

In some research streams, we studied and compared countries or time-periods which displayed different combinations of the variables of interest. In others, we investigated localities or sectors of activity which differed in terms of possible causal factors or key outcomes or both. In all cases, we made as much as possible of opportunities to follow chains of events directly and draw inferences about mechanisms and not just directions of change. In drawing conclusions from actual cases, we considered the relevant counterfactuals – the alternative situations that were imaginable but not available for empirical inspection.

In analysis of large and medium datasets, some of the most important design choices are about sampling and margins of statistical error. This applies to APPP in so far as we draw inferences from sample surveys. However, our main propositions are based on comparative case studies, where both the sample and the universe of potential cases are far smaller than in statistical analysis.

As discussed in the literature,<sup>5</sup> the principles governing sampling rigour and the avoidance of researcher bias are the same in all fields of research. However, their implications in case study research are not the same as in statistical analysis.

For example, representativeness is not achieved with a random sample but with a set of cases that captures as well as possible the range of relevant variation in the universe of possible cases. The relevant form of observer bias is the deliberate or careless avoidance of cases that might contradict the emerging hypothesis of the study.

### Comparative designs

Ideally, especially in the approach favoured by Ragin,<sup>6</sup> comparative designs are based on considerable advance knowledge of the distribution of dependent and independent variables of interest across available cases. However, as emphasised in some of the more recent literature, this level of advance knowledge is a relatively rare luxury, especially in research that is pushing forward the frontiers of understanding of a particular topic.

Study design, including choices about case selection, often has to proceed in an iterative way as initial hunches get refined into hypotheses and knowledge of the range of variation is built up. As Gerring puts it, ‘One cannot construct the perfect research design until (a) one has a specific hypothesis and (b) one is reasonably certain about what one is going to find “out there” in the empirical world ... In short, the perfect case study research design is usually apparent only *ex post facto*’.<sup>7</sup>

Given its non-standard research questions, APPP was in exactly this position. Although we were initially inspired by Ragin’s views, our most interesting causal propositions emerged during empirical research, in the manner suggested by Gerring. They then shaped the way we conducted further research, including any remaining decisions about case selection.

This was a feature of our comparative analysis of political regimes, to be reported in full in the forthcoming book by Kelsall and associates.<sup>8</sup> It applied even more clearly to APPP’s comparative studies of the local governance of public goods provision.

In the local governance research, the key bottlenecks in provision, which became our principal outcome variable, needed to be identified by observation during the fieldwork – that is, after the selection of countries and fieldwork sites. The explanatory factors to which we ultimately gave most attention – policy incoherence, top-down disciplines and room for local problem-solving – were not predicted in advance on the basis of existing theory. They emerged in the course of the empirical work. Comparative analysis and further case selection were conducted iteratively as the issues were progressively sharpened.

## Achieving optimal data quality

While facts do not speak for themselves, data quality is important in obvious and not-so-obvious ways. We discuss three major topics here: reliability, validity, triangulation and the value of fieldwork.

### Reliability and validity

It is conventional to distinguish two variable features of a dataset or measurement approach: *reliability* and *validity*. The reliability of an observation is the extent to which further observations would produce the same result. Validity, in contrast, concerns whether the indicator or descriptor that is used measures or captures what it is supposed to.

Reliability and validity are relevant to the whole range of empirical methods available to social scientists. Just as the data in a particular quantitative set may be judged to be of poor quality – failing to meet these two criteria – because of the way they have been



***possibilities for triangulation are maximised when data collection is extended over time.***

generated, so there can be bad empirical data about a case or cases investigated with qualitative methods. However, typical threats to data quality differ along the spectrum of methodological approaches. The techniques that researchers use in grappling with these challenges also vary.

The data drawn upon in APPP research were at the qualitative end of the spectrum. The research questions asked were those left unanswered by 20 years of research using the available large quantitative datasets. It was essential to make heavy use of comparative and case-study methods. In this broad field, the key technique for ensuring reliability and validity of data is called triangulation.

### Triangulation

Triangulation refers to the cross-checking of information from different sources to assess its reliability and validity. The analogy is with the traveller who is able to situate herself with confidence on a map by taking no less than three compass readings against visible landmarks.

For example, in carrying out APPP case studies, researchers compared factual claims made in one interview with claims on the same or a similar subject by other interviewees occupying different vantage points. Reports of events or episodes were assessed by means of repeated direct observation of equivalent events, as well as by testimony from other quarters. Fieldworkers recorded in their notebooks both basic information and their emerging interpretations of situations and patterns of social interaction. These records were subsequently interrogated, and if necessary challenged, by other members of the field team.

It is worth noting that triangulation does *not* refer particularly to cross-checking between qualitative and quantitative data sources. There is an important argument for combined methods in social research. However, it is based on the fact that different methods are good for addressing different questions – e.g. how poor a household is versus how it copes with its situation. That is about complementarity, whereas triangulation is about comparing data on the same or similar questions.

## The value of fieldwork

Possibilities for triangulation are maximised when data collection is extended over time and concentrated on limited geographical areas or organisational settings. That is why ethnographers place such importance on length of fieldwork and on residing in the fieldwork area. The robustness as well as the empirical range and richness of the APPP case studies reflects the attention we gave to these criteria.

There were two main types of APPP case studies:

- interview- and document-based studies of whole countries or particular sectors using detective-style investigative techniques; and
- field studies carried out in particular local areas over extended periods using ethnographic methods (conversational interviewing plus direct, semi-participant, observation).

One sub-set of field studies was based on observation and informal interviews over time. The Malawi Local Governance research, for example, was based on 17 weeks of fieldwork by a mixed team in 2009-10 and nine months in 2010-11. In Rwanda, four researchers lived for 11 months in two contrasting rural districts. The research by the Laboratoire d'Etudes et Recherches sur les Dynamiques Sociales et le Développement Local (LASDEL) in Niger took five field researchers to three previously studied urban/rural municipalities for a month in 2009 and again for several weeks in 2010.

Other field studies included systematic social surveys using structured questionnaires but complemented this with observational work. For example, APPP's Local Justice researchers undertook six months of daily observation in courts and dispute-resolution sessions. The work on Cotton Sector Reforms included fieldwork in villages over several months in 2009 and in 2010 as well as a large set of semi-structured interviews with key stakeholders. These combinations of data-collection methods generated rich opportunities for triangulation.

## Conclusions

We have said that evidence quality matters for three reasons, and posed the question of what counts as quality. We have suggested that this is, in part, about data quality but equally about modes of inference from data. Illustrations have been drawn from APPP research and the wider field of policy debate about governance for development. We have, in effect, laid out the criteria by which we expect the robustness of research in our field to be assessed.

A broader implication for development agencies is that assessing the evidence-base for a particular policy position is a complex business. Single-stranded measures of quality will never provide satisfactory answers to the question 'what's the evidence?' Headcounts of studies that approach different research problems, with methods tailored to answering their specific questions, are unlikely to be helpful. Nonetheless, as we have seen, criteria for distinguishing between sound and unsound inference from data do exist. With due care and attention, it is possible to make judgements about how well evidence speaks to policy.

## References

1. Director of APPP and Research Fellow, Politics and Governance Programme, Overseas Development Institute.
2. D. Booth, *Development as a Collective Action Problem: Addressing the Real Challenges of African Governance*, ODI, 2012.
3. G. Hyden, *Institutions, Power and Policy Outcomes in Africa*, APPP DP 2, 2008; D. Booth, *A Research Design Fit for Purpose*, APPP DP 3, 2008; J.-P. Olivier de Sardan, *Researching the Practical Norms of Real Governance in Africa*, APPP DP 5, 2008; T. Kelsall, *Game-theoretic Models, Social Mechanisms and Public Goods in Africa*, APPP DP 7, 2009.
4. J. Gerring, *Social Science Methodology: A Criterial Framework*, CUP, 2004.
5. See, e.g., C.C. Ragin, *The Comparative Method*, California U.P., 1987; A.L. George and A. Bennett, *Case Studies and Theory Development in the Social Sciences*, MIT Press, 2005; J. Gerring, *Case Study Research*, CUP, 2007.
6. Ragin, *Comparative Method*.
7. Gerring, *Case Study Research*, p. 149.
8. T. Kelsall et al., *Business, Politics and the State in Africa*, Zed Books, 2013.



The Africa Power and Politics Programme is a consortium research programme funded by the UK Department for International Development (DFID) and Irish Aid for the benefit of developing countries. The views expressed in this publication are those of the author and should not be attributed to DFID or Irish Aid, or any of APPP's member organisations.

Africa Power and Politics  
Overseas Development Institute  
203 Blackfriars Road  
London SE1 8NJ, UK  
T: +44 (0) 20 7922 0300  
E: info@institutions-africa.org

[www.institutions-africa.org](http://www.institutions-africa.org)