

The Turn to Metrics

Malcolm Langford and Sakiko Fukuda-Parr

Malcolm Langford, Research Fellow, Norwegian Centre for Human Rights, Faculty of Law, University of Oslo and Co-Coordinator, Metrics for Human Rights, email: malcolm.langford@nchr.uio.no.

Sakiko Fukuda-Parr, Professor of International Affairs, New School, New York and Co-Coordinator, Metrics for Human Rights, email: fukudaps@newschool.edu.

I. Introduction

Quantitative methods have played a limited role in human rights practice. Beyond the generation of events-based figures (e.g., numbers of persons tortured or evicted), some high profile but controversial measures such as the Freedom House index,¹ or the monitoring of a few treaties containing quantitative commitments,² the use of numerical data has been infrequent. Instead, qualitative methods have been dominant. Information is primarily sourced and reproduced from interviews with victims and alleged perpetrators, witness accounts, audio and video footage, official documents and media reports. With a focus on human rights violations affecting specific persons, places or events, these qualitative methods seem apposite in establishing the relevant facts and narratives that can be tested against human rights standards.

Nonetheless, the field of human rights has not been immune from a global shift towards quantitative measurement in all fields of human activity. The community of activists, professionals, officials and scholars concerned with human rights has begun to explore ways in which different metrics can establish the denial of rights, reveal breaches of obligations and justify new laws or policies. This has been complemented by the use of quantitative methods in social science research and programme evaluation.³

1 Since 1972, Freedom House has issued Freedom in the World, which provides a comparative assessment of political rights and civil liberties, currently across 195 countries and 14 related and disputed territories. Freedom-House, *Freedom in the World 2012* (Freedom House, New York 2012).

2 E.g., the ILO Convention 102 on Social Security which requires states to meet different quantitative benchmarks in terms of coverage for a minimum number of pillars of social security.

3 See discussion in Section 2.

The motivations for using quantitative methods in, at least, human rights practice are many. Qualitative methods often elicit the complaint that the evidence is too *ad hoc*. It is inadequate in establishing systematic patterns of violations, identifying clear failures by duty-bearers to fulfil their positive obligations, or motivating systemic reform. At the heart of quantification is aggregation and replication. A broader swathe of events and experiences can be captured, represented and compared across time and space (e.g., localities, countries, regions etc); which can provide information on the magnitude and source of problems. Establishing seemingly neutral, scale-able and externally verifiable methods seems ideal for human rights. In an environment characterised by accountability relations, distance between actors, and mutual distrust, “audit”-like tools such as indicators can provide a mutually acceptable means for actors to assess compliance.⁴

Strategically, quantified data is a powerful tool of communication. It offers clear, comprehensible and simple snapshots of complex situations. Porter claims that the global rise of ‘numbers, graphs, and formulas’ can be traced primarily to this function.⁵ Quantification is a ‘technology of distance’ which is ‘well suited for communication that goes beyond the boundaries of locality and community’.⁶ However, his ambiguity toward another common explanation – proof of causal relations between social phenomena – is not justified. Quantification is driven by attempts to substantiate certain causative arguments and overcome resistance to them. Human rights is no exception and quantitative methods have been employed to challenge scepticism towards rights or disputes over causes of non-implementation. For instance, different actors have sought demonstrate that human rights do not generate negative externalities but instead provide verifiable benefits, whether it is peace, economic growth, improved health, democratic deliberation etc.⁷

The turn to quantitative methods has been encouraged by various UN bodies. Although, it had a very rocky start. The creation of a Human Freedom Index by the UNDP in 1992, which ranked countries across a representative range of

4 A Rosga and M Satterthwaite, ‘The Trust in Indicators: Measuring Human Rights’ (2009) 27 *Berkeley Journal of International Law* 253, 280.

5 T Porter, *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life* (Princeton University Press Princeton 1995) viii.

6 *Ibid* ix.

7 For an earlier historiography of this stream of quantitative methods in human rights, see RL Barsh, ‘Measuring Human Rights: Problems of Methodology and Purpose’ (1993) 15 *Human Rights Quarterly* 87.

rights in the Universal Declaration of Human Rights, met strong opposition from many states.⁸ The upshot was that international agencies proceeded in a diffuse and modest manner with human rights indicators, often under the cover of more innocuous terms such as governance or development rather than democracy or human rights. However, the human rights branch of the UN system has begun to re-champion the international measurement of rights, although with less emphasis on ranking.⁹ Moreover, it is not only international bodies that are driving these demands: human rights practitioners note the increasing demands by courts, media and bureaucrats amongst others for quantitative evidence. In addition, various donor programmes like the Millennium Challenge Account provide extra financial assistance for reaching the Millennium Development Goals if countries attain certain thresholds on civil and political rights and good governance.

The purpose of this *Special Issue on Quantifying Human Rights* is threefold. It is to showcase different methods for quantitatively measuring human rights compliance, provide critical perspectives on the use of such methods, and raise opportunities and challenges for future theory and practice. It emerged from the second annual meeting of Metrics for Human Rights in New York in May 2011, which is a loose international network of scholars and practitioners working on quantitative methods and coordinated by the New School and University of Oslo. The remainder of this Introduction discusses all three aspects with reference to the broader scholarship and the six articles in this special issue.

II. Uses of Quantitative Methods in Human Rights

The rich and emerging literature and practice on quantitative methods includes a diversity of approaches with respect to: the motivation and the purposes to which data would be used (the “why”); the identification of relevant aspects of “human rights” for measurement (the “what”); and finally, the quantification method itself (the “how”). The papers in this collection contribute in all of these dimensions.

8 See discussion in *ibid* 87-90.

9 For instance, the UN High Commissioner has stated in the context of economic, social and cultural rights that, OHCHR, ‘Monitoring realization requires tools that are capable of measuring results and progress over time’ UN doc. E2009/90, para. 34.

Why

There are three broad uses for which human rights measurement tools are being developed: *monitoring* and determination of compliance; *advocacy* for improved policies and practices; and *explanation* of broader casual relationships that include human rights.

As to the first, a particular driver of recent measurement attempts has been compliance-based monitoring. For instance, in 2002 the UN Committee on Economic, Social and Cultural Rights (CESCR) called on states to cooperate with them in setting performance benchmarks for progressive realisation of various rights in the intervals between five yearly periodic reporting.¹⁰ Quantitative indicators could overcome one of the major weaknesses in monitoring processes that depends on self-reporting: They provide a consistent measurement tool that can be applied across countries and/or time and that relate to agreed standards. This interest spread and in 2006, all UN human rights treaty bodies sent a joint request to the Office of the UN High Commissioner for Human Rights to provide recommendations on how statistics and indicators could be used and developed in their monitoring function.¹¹ In response, the Office has developed a framework of indicators across all human rights.¹²

The aim to monitor compliance with human rights standards is prominent in this issue. Many authors take a point of departure in international human rights treaties although some of them give greater attention to regional or national legislative standards. Moreover, the intended users of the results are more diverse than UN bodies and include national governments, media, civil society etc. Vizard showcases the Human Rights Measurement Framework (HRMF), through which indicators are used to assist in determining compliance by British authorities, and demonstrates its application in the areas of violence, treatment of older people and child poverty. Haymann, Berrera, de Guzman, Raub and Vincent introduce the World Policy Centre Analysis (WPCA) which seeks to make available “readily-accessible comparative data on hundreds of laws and policies” and sets out their results on the right to social security. Randolph and Guyer use their Social

10 UN Committee on Economic, Social and Cultural Rights, ‘General Comment 14, The right to the highest attainable standard of health’ (Twenty-second session, 2000), U.N. Doc. E/C.12/2000/4 (2000), para. 58.

11 OHCHR, ‘Report on Indicators for Monitoring Compliance with International Human Rights Instruments – Summary’, UN Doc. HRI/MC/2006/7 (2006).

12 OHCHR, ‘Report on Indicators for Promoting and Monitoring the Implementation of Human Rights’, UN Doc. HRI/MC/2008/3 (2008).

and Economic Rights Fulfilment Index (SERF Index) data to establish historical trends in compliance by states with duties to progressively fulfil economic and social rights.

Not all quantitative monitoring though is focused on the obligations of states: other actors also carry duties. Elsewhere, quantitative methods have been applied to the performance of local governments and transnational corporations amongst others.¹³ In addition, compliance-based approaches are also not always backwards-looking. Human rights impact assessments which seek to prevent future denials of human rights are increasingly turning to indicators as a way of standardising and streamlining measurement.¹⁴

Advocacy for change for policy and practice is a second purpose. It can build on the results of monitoring work; Corkery and Way present the OPERA framework as a process using a variety of analytical tools to monitor compliance, and on this basis advocate for policy change. But the effective use of metrics for policy advocacy demands specific considerations. There is a preference for indicators that are actionable or provide a better case for spurring reforms.¹⁵ For instance, Yamin and Falb argue in this issue that greater efforts should be given to measuring Emergency Obstetric Care (EmOC) indicators rather than improving measures of maternal mortality: the former reflect policies and practices that have a proven effect in reducing maternal mortality. In other cases, data generated for monitoring purposes may be ambiguous in establishing a violation but reveal priorities for action. Vizard's regression results demonstrate that young persons, females and persons with disabilities have a higher risk of being victims of violence which 'does not provide direct evidence of the failure of duty-holders to fulfil the negative and positive duties' but highlights the 'need for more effective public policy interventions'.¹⁶

A third purpose is what could be broadly described as explanatory and is evident in academic scholarship, public policy research and human rights practice.

13 See for example J Dugard, M Langford and E Anderson, 'Determining Progress on Access to Water and Sanitation: Law and Political Economy in South Africa' in M Langford and A Russell (eds), *The Right to Water: Theory, Practice and Prospects* (Cambridge University Press, Cambridge 2013); Danish Centre for Human Rights, *Human Rights Compliance Assessment*, Human Rights and Business Department, available at <http://www.humanrightsbusiness.org/compliance+assessment>.

14 See *Human Rights Impact Assessments for Trade and Investment Agreements*, Report of the Expert Seminar, 23–24 June 2010, Geneva, Switzerland.

15 For a discussion, see: M Orkin, C Naval, JR Suesser and RS de Miguel, 'Towards the Democratic Monitoring of Governance: the Metagora Experience' in UNDP (ed) *Making the State Responsive: Experience with Democratic Governance Assessments* (UNDP, New York 2011).

16 Corkey and Way, this Special Issue, 255.

Quantitative data can permit us to better understand the relationship between human rights and other social and political processes. Currently, there is a major interest in understanding the impact or effects of human rights approaches. Do they work? The rise of the paradigm of rights has seen a broad array of treaties, laws, policies, development programmes, civil society strategies which all seek to advance various human rights objectives. In analysing the extent to which these “interventions” have reached their goals, quantification has become an important tool, particularly when research is carried out by social scientists or consultants.¹⁷ Promoters and critics are interested to learn whether the intervention has worked, how it could be improved and whether alternatives are available. This use of quantitative methods is not the focus of this issue but Heymman and her fellow authors demonstrate for example that infant mortality is negatively correlated with maternity leave laws.

More traditionally, the focus has been on understanding and explaining the nature and causes of behaviour by different actors.¹⁸ In this issue, Jung and Rosevear outline the Toronto Initiative for Economic and Social Rights (TIESR), which creates a dataset that measures the presence, absence, and justiciability of seventeen separate economic and social rights in 136 constitutions. It is motivated by a desire to understand how and where constitutionalisation of these rights is spreading but also establish the basis for determining their effects and influence. Randolph and Guyer also use the Social and Economic Rights Fulfilment Index (SERF Index) to explore the relationship between rights fulfilment and economic growth.

What

Each quantitative tool draws on a particular conceptual model of human rights and focus on specific concerns with human rights. An important advance in

17 See, e.g. E Hafner Burton and J Ron, 'Human Rights Institutions: Rhetoric and Efficacy' (2007) 4 *Journal of Peace Research* 379; V Gauri, 'Public interest litigation in India: overreaching or underachieving?' (2011) 1 *Indian Journal of Law and Economics*; D Pope and C Bamba, 'Has the Disability Discrimination Act closed the Employment Gap?' (2005) 27 *Disability and Rehabilitation* 1261; JJ Donohue, M Stein, C Griffin and S Becker, 'Assessing Post-ADA Employment: Some Econometric Evidence and Policy Considerations' (2011) 8 *Journal of Empirical Legal Studies* 477.

18 See, e.g., B Simmons, *Mobilizing for Human Rights. International Law in Domestic Politics* (Cambridge University Press, New York 2009) (in relation to treaty ratification); J Foweraker and T Landman, *Citizenship Rights and Social Movements: A Comparative and Statistical Analysis* (Oxford University Press, Oxford 1997); PA Goff, J Eberhardt, M Williams and MC Jackson, 'Not yet human: Implicit knowledge, historical dehumanization, and contemporary consequences' (2008) 94 *Journal of Personality and Social Psychology* 292.

recent thinking about human rights measurement has been to move beyond the mere achievement of certain outcomes or results. Human rights standards often contain obligations of conduct or a mixture of conduct and result and not just results. They also contain various exceptions, defences and conditions which need to be taken into account for any finding of violation. The result is that a measurement of outcomes focuses on the perspective of the rights holder and may only tell us that human rights have been denied; not whether any duty bearer has breached an obligation or could have taken remedial or corrective action.

The problem in determining compliance with civil rights by merely using general outcome indicators is well known. When is violence committed by private actors attributable to a state: i.e., how far do positive obligations to protect extend in this situation? When is state-sponsored killing an “extrajudicial killing”: i.e., does the killing fall into a particular exception? Yet another aspect of the ill-fitting indicator is the neglect of the obligations of progressive realisation for economic, social and cultural rights¹⁹ but also equality rights and some civil and political rights.²⁰ Implementation to develop the necessary institutions and systems requires time, administrative capacity and resources.²¹ Inherent in this idea of “progressive realization”, therefore, is the principle that countries with greater economic resources have a correspondingly greater duty to ensure equitable and widespread enjoyment of these rights. In the case of economic, social and cultural rights this was made very explicit in international standards: states have obligations to fulfil these rights subject to maximum available resources. Incorporating such state capacity into the measurement of how well a country is doing in meeting its obligations under international law is essential: Countries cannot be simply compared on the same scale as they are in the Millennium Development Goals.²²

There have been various attempts at overcoming this problem. The first is to typologise: to disaggregate obligations and connect them with relevant indicators.

19 See UN CESCR, ‘General Comment 3, The Nature of States Parties Obligations (Art. 2, para. 1)’, UN Doc. E/1991/23 (1990), Annex III; Charter of the United Nations (San Francisco, 26 June 1945), 3 Bevans 1153, 59 Stat. 1031, T.S. No. 993, *entered into force* 24 Oct. 1945, Art. 1, para. 3

20 UN CCPR, ‘General Comment 31, The Nature of the General Legal Obligation Imposed on States Parties to the Covenant (Art. 2)’, UN Doc. CCPR/C/74/CRP.4/Rev.6 (2004).

21 According to the principle of progressive realisation, states must strive to fulfil economic and social rights obligations to the maximum extent possible in the face of economic resource constraints *Id.*

22 See discussion in S Fukuda-Parr and J Greenstein, ‘How Should MDG Implementation Be Measured: Faster Progress or Meeting Targets?’, *World Development* (forthcoming); M Langford, ‘A Poverty of Rights: Six Ways to Fix the MDGs’ (2010) 41 IDS Bulletin 83.

This alerts us to the multitude of duties that require measurement. This can be done by categorising obligations according to the trichotomy of respect, protect and fulfil and populating the categories with relevant indicators.²³ Similarly, but from a more statistical perspective, human rights are increasingly being divided up according to whether they refer to the enjoyment of rights (outcome); institutional arrangements made by governments such as constitutional provisions (structure); and the taking of steps and implementation of policy measures (conduct).²⁴ For instance, the TISER (Jung and Rosevear) and WPCA (Heymann and others) focus on measuring “structural” indicators.

The disadvantage of this approach is that it may overlook the relationship between these categories (the progressive realisation obligations is properly characterised as conduct-result) and the indicators may not always incorporate exceptions, defences and conditions. An alternative approach is to use different indicators in a structured fashion. In this issue, Corkery and Way present the OPERA framework which provides an overarching analytical framework within which multiple tools and techniques, including both conduct indicators and result indicators can be eclectically integrated to determine compliance. An alternative is to seek to bridge the gaps through various quantitative techniques. A focus on the duty bearer and differentiated resource constraints is the chief motivation for the creation of the SERF Index.²⁵ By using the achievement possibilities frontier, state performance is judged not by the level of outcome achieved but by the shortfall to what could have been achieved with the resources available. Other approaches focus though on the relationship of a country’s performance to the average performance at the same resource capacity (and sometimes with a broader range of resource indicators) – the country whose performance is worse than average then warrants further examination.²⁶ Randolph and Guyer in this issue

23 V Roaf, A Khalfan and M Langford, *Indicators for the Right to Water* (Heinrich Boell Foundation, Concept Paper No. 13, Berlin 2005).

24 See, e.g., OHCHR, ‘Report on Indicators for Promoting and Monitoring the Implementation of Human Rights’, UN Doc. HRI/MC/2008/3 (2008).

25 The SERF index used by these authors reflects the highest feasible performance at any per capita income level. See generally S Randolph, S Fukuda-Parr and T Lawson-Remer, ‘Economic and Social Rights Fulfillment Index: Country Scores and Rankings’ (2010) 9 *Journal of Human Rights* 230.

26 See E Anderson and M Langford, ‘A Distorted Metric: The MDGs, Human Rights and Maximum Available Resources’, Working Paper, 2012. In this paper, Anderson and Langford find that the country rankings with the SERF and outlier approaches disagree, with the latter more favourable to lower-resourced countries. They also note that the maximalist approach of SERF may be closer to the normative benchmark but may equally provide an unrealistic benchmark.

use the SERF index to develop empirical evidence of how compliance has improved or regressed over time, showing that in some cases there has been regress even if the actual achievement level has increased.

How

Numerous initiatives in the last decade have introduced innovations in methods for quantitative analysis with distinctive approaches to data source, indicator selection, level of aggregation and data use.

With respect to data source, several organisations and researchers have attempted to *create data* that is relevant to human rights priorities, ranging from empirical “situation testing” such as ILO studies of discrimination in employment²⁷ to field surveys such as individual/victim perception in Haiti,²⁸ and the coding of qualitative data. Coding has been widely used in better known international civil and political rights indices such as the Freedom House freedom scores, Political Terror Scale, Cingranelli and Richards Indices (CIRI) and others.²⁹ These approaches code qualitative reports that evaluate human rights situations. The coding is often done according to rigorous standards but relies on subjective rather than objective information. Some approaches avoid this dilemma and have focused on coding objective information, such as constitutions. TIESR (Jung and Rosevar) presented in this issue is an example of such an effort.

In contrast, another strand of work has attempted to make better use of existing survey and sometimes events-based data, notably economic and social data that are in the public domain, in most cases published by official statistical bodies and harmonised in international statistical processes. The indicator sets of the OHCHR are largely comprised of these data. In this issue, OPERA (Corkery and Way), HRMF (Vizard), and SERF (Randolph and Guyer) rely on these sources.

With respect to indicator selection, an important innovation in recent years has been to employ participatory and dialogical processes. Stakeholders, sometimes including rights holders, reach a consensus on a relevant indicator as well as

27 E.g., P Arriijn, S Feld and A Nayer, 'Discrimination in access to employment on grounds of foreign origin: the case of Belgium' (ILO, Geneva 1998).

28 E.g., M Satterthwaite, N Reisch, F Brody and J Simeone, *Yon Je Louvri: Reducing Vulnerability to Sexual Violence in Haiti's IDP Camps* (CHRGJ, New York 2012).

29 See an overview in T Landman, 'Measuring Human Rights: Principles, Practice and Policy ' (2004) 26 Human Rights Quarterly 906 and D Cingranelli and D Richards, 'The Cingranelli and Richards (CIRI) Human Rights Data Propect' (2010) 32 Human Rights Quarterly 395.

the relevant normative benchmark for that indicator. Examples include the IBSA approach of the UN CESCR, HRMF (Vizard in this issue), treaties such as ILO Convention 102 and court orders/settlements.

Another methodological divide concerns the level of aggregation. This concerns the units of analysis, covering both the scope of cases (from a single case or individual to all individuals affected) and the categorical unit—expanding to include subnational entities, countries or the world. The motivation behind much of the quantitative work of recent years has been to move beyond the individual case and situation that has been the cornerstone of human rights investigative methodology. The papers in this issue include those that aggregate to the country level (OPERA, HRMF) and others that provide global data sets (SERF, TISER, WPCA) where country data can be compared. Note that some of these national (HRMF, OPERA) and global (SERF) data sets can be disaggregated to sub-national levels, which is a high priority requirement of human rights analysis.

Another type of aggregation is across different rights. Much of the recent quantitative work has focused on the use of specific indicators and has eschewed the earlier emphasis on aggregation across rights and attempts to develop composite indices.³⁰ This is reflected in all the papers in this issue with the exception of the SERF Index (Randolph and others). Composite approaches may however play a role, particularly in communication: different types of data serve different analytical purposes. The lower the level of aggregation, the richer the detail but less the information on overall magnitude and trends. Instead, the thrust of many in the human rights community has been to focus on disaggregation of existing indicators, particularly to see whether human rights are being enjoyed by women and groups that face discrimination or social exclusion.

Finally, there are several approaches to the ways that data are put to use. In some instances, *descriptive* methods are used e.g. counting, disaggregation, simple probabilities, ratios, allocations, benefits or cost incidence. Others use more *comparative* approaches that build in common denominators to establish trends, rankings, indexes and outliers. Yet others seek to identify *causal* relationships, such as by use of bivariate and multivariate regression, which may also inform comparative approaches. Very early examples includes Becker's method for

30 For an overview of different global indices developed on human rights and governance in the 1990s, see A Wilde, 'The Democratization of Governance Assessments' in UNDP (ed) *Making the State Responsive: Experience with Democratic Governance Assessments* (UNDP, New York 2011), 29-48.

decomposing labour market gaps (e.g., for ethnic minorities, women and recently persons with disabilities) to determine whether employers were engaged in discriminatory practices.³¹ Vizard and Randolph and Guyer use regression analysis to respectively establish probabilities and rankings. However, Corkery and Way also show how simple methods can be deployed to raise a *prima facie* case for non-compliance or feasibility of a measure.³²

III. Limitations

The customary concerns over the use of quantitative approaches are not foreign to the field of human rights. However, there is a danger that each new wave of metrics in human rights takes insufficient heed of its inherent limitations and dangers. Whereas there is much to be gained by the turn to numbers, it is our view that the human rights community must develop the practice with a spirit of self-criticism and modesty, an openness to peer review, and an awareness of the political environment in which data is constructed and received.

We will not discuss exhaustively all of the challenges but name four that we consider significant. The first is the definition and choice of indicators – construct validity. The very strengths of quantification – simplification and abstraction in applying a single measurable definition across different contexts – are its Achilles heel. The criteria for creating an indicator may not match the relevant human rights standard. To take an example: The commonly-used measure of “improved water” by the WHO and UNICEF is often used as a stand-alone proxy for the basic minimum for the right to water. It represents approximately 20 litres of water per person per day.³³ However, the indicator is questionable as a human rights standard. It does not really tell us whether this minimum has been realised: it is arguably too permissive in its criteria³⁴ and allows water supply to being ticked as “improved” even if it is irregular or not

31 G Becker, *The Economics of Discrimination* (Chicago University Press, Chicago 1971).

32 See also CESR and ICEFI, *Rights or Privileges? Fiscal commitment to the rights to health, education and food in Guatemala* (CESR, Madrid 2009). See also E Felner, 'Closing the 'Escape Hatch': A Toolkit to Monitor the Progressive Realization of Economic, Social, and Cultural Rights' (2009) 1 *Journal of Human Rights Practice* 402.

33 In the practice of the UN CESCR it is the dominant indicator.

34 J Bartram, 'Improving on haves and have-nots' (2008) 452 *Nature* 283. Interestingly, he notes that the standard for sanitation may, to the contrary, be too strict.

portable,³⁵ unaffordable³⁶ or culturally unacceptable³⁷ - all constituent elements of the right to water in international law and jurisprudence.³⁸ Indeed, some of these elements, such as cultural acceptability, raise the intractable problem of whether some human rights are quantifiable at all: the inclusion of local contextual factors makes them inherently difficult to accommodate in a universally applicable quantitative definition.

In this issue Vizard shows how some of these dangers might be minimised. In building up a quantitative evidence database in England, Scotland and Wales on human rights, the HRMF project began with OHCHR's international framework and "illustrative indicators". However, they engaged in an extensive consultation with national human rights, policy and statistical actors that mapped in more detail the legal requirements and available data sources and thus sought to identify better-fitting indicators. This demonstrates the possibility of developing indicators with higher construct validity within a community or a country, but its limitation is that it cannot be extended to develop cross-country comparable data.

The problem of ill-fitting proxies does not, however, stop at the phase of constructing, processing and interpreting data. It is also a reflexive process whereby an indicator reshapes its parent norm. As Davis, Kingsbury and Merry put it, indicators embody a 'theoretical claim about the appropriate standards for evaluating actors' conduct'.³⁹ If an indicator is loosely matched with a standard or simply achieves prominence, it can quickly take on a normative life of its own. One pertinent example is income poverty. In 2000, the Millennium Declaration elevated the \$1 US dollar a day indicator from being one marker of extreme income poverty to being *the* standard of income poverty itself.⁴⁰ Disturbingly, it is far

35 G Mboup, 'Existing Indicators in the Water and Sanitation Sector: Indicators for Accessibility, Affordability and Non-Discrimination' (Indicators for the Right to Water, Concept Paper No. 13, 2005).

36 OECD, *Social Issues in the Provision and Pricing of Water Services* (OECD, Paris 2003); H Smets, *De l'eau potable à un prix abordable* (Johanet, Paris 2009).

37 N Singh, 'Socio-Cultural Norms, Human Rights and Access to Water and Sanitation' in M Langford and A Russell (eds), *The Right to Water: Theory, Practice and Prospects* (Cambridge University Press, Cambridge 2013).

38 See, e.g., Committee on Economic, Social and Cultural Rights, 'General Comment 15, The right to water', U.N. Doc. E/C.12/2002/11 (2003); Dublin Statement on Water and Sustainable Development, International Conference on Water and the Environment: Development Issues for the 21st Century, UN Doc. A/CONF.151/PC/112 (1992), Principle 4.

39 K Davis, B Kingsbury and SE Merry (eds), *Governance by Indicators: Global Power through Classification and Rankings* (Oxford University Press, Oxford 2012) 9.

40 In the Millennium Declaration and the subsequent Millennium Development Goals, the targets are formulated in such a way as to provide space for different indicators.

from clear that this represents a minimum core of the right to an adequate standard of living given the prevailing prices of basic goods. Indeed, the indicator may conveniently blind us from the fact that the world's poorest would 'grow' in number from approximately a billion to approximately 2.5 billion if we used two dollars a day as a yardstick, and even more if we included health and education costs in actual measurement.⁴¹ This risk can be even greater for process or conduct-oriented indicators – it can be assumed that they embody a one-size-fits-all policy prescription. In moving the quantitative agenda forward, the human rights community needs to consider what are the normative implications of the use of a particular indicator.

The second challenge is the reliability and validity of data. Recorded observations may not be an accurate reflection of the reality that a measuring instrument is trying to capture. In the process of data creation, subjectivity enters: classifying an event as a violation, coding qualitative information according to a scale, or conducting surveys in different cultural or linguistic contexts may bias responses. Even data that is meant to capture subjectivity – such as perception/ barometer surveys – needs to be used cautiously: an individual's response may not correspond to their behaviour or even their attitudes.⁴² Moreover, there are the practical challenges of missing data and technical dilemmas, such as weightings given to respondent groups or indicators in composite indexes.

The full implications of using unreliable data are demonstrated by Yamin and Falb in this issue. The reduction of maternal mortality by two-thirds was established as a Millennium Development Goals but the prevailing maternal mortality ratios are 'notoriously unreliable due to a number of factors including the quality of the underlying source data regarding the number of maternal deaths, and varying specifications of statistical models including, but not limited to the selection of covariates and handling of HIV-related deaths'.⁴³ Statistical modelling exercises have sought to overcome the problems but the result is contradictory data sets and data that is largely meaningless. The authors give the example of Afghanistan where estimated maternal deaths in one study dropped slightly from 1640 to 1,575 deaths per 100,000 live births between 1980 and

41 See generally T Pogge, 'Millions Killed by Clever Dilution of Our Promise' (CROP Poverty Brief, August, 2010). See also A Fischer, 'The Political within the Depoliticised: Poverty Measurement, Implicit Agendas and the MDGs' in M Langford, A Sumner and AE Yamin (eds), *The MDGs and Human Rights: Past, Present and Future* (Cambridge University Press, Cambridge 2013).

42 The framing and phrasing of the questions, the nature of the survey instrument and the context in which the questions are being asked can be critical in shaping the responses.

43 Yamin and Falb, this Special Issue, 353.

2008. However, the (enormous) confidence interval was largely unchanged (632 – 3,527 deaths), which means that it is ‘simply impossible to say whether and to what extent levels of maternal mortality have declined in Afghanistan over the past few decades’.⁴⁴

Reducing systemic error can be achieved through various methods such as repetition, using complementary indicators and testing the measuring instruments under different conditions. For instance in her paper, Vizard describes how the HRMF project in the United Kingdom sought to overcome the unreliability of official sources on domestic abuse and sexual violence (which understates prevalence) by also drawing on a general population survey which includes questions on experiences in a self-completion module.

The third challenge is aggregation. Higher levels of aggregation are valuable because it provides an overall – summary – picture of the magnitude of achievement and deficits, progress and regress. This makes it possible to show broad trends and highlight major areas of concern. By the same token, they do not provide adequate detail and differentiation. For instance, data truncation is a particular problem with global data sets.⁴⁵ Highly diverse situations are grouped together in a single category, such as in the Freedom House Index, where a large number of countries are given a score of 1 for political rights yet this masks important differences between these countries which include all Western countries, most East European countries and Israel. Diverse countries can also be bunched together at the end of a continuous scale.

The SERF index presented by Randolph and Guyer in this issue has sought to avoid this problem by creating a separate index for high income countries and differentiates with greater detail with scores ranging from 1 to 100 incorporating decimal points. The index has been used to advocate poor performance of the US when the country was reviewed at the UN as they ranked at the bottom of the high income country scores. These poor scores reflect the exclusion of disadvantaged groups and inadequacy of social protection measures. Yet these global rankings could be misused to deflect criticism or encourage complacency if these higher scores at the upper end of the scale are contrasted with much lower levels. The same occurs when even the best-performing high-income countries are not called to account for unjustified denial of rights and exclusion of significant but smaller groups, e.g., prisoners, ethnic minorities,

44 Ibid 356.

45 For a discussion in the context of human rights, see Barsh (n 7), 102-3; Landman (n 29), 923.

homeless persons. These highly aggregated indices should be used in conjunction with more detailed studies.⁴⁶

The final challenge is interpretation and use of quantitative measurement: what indicators ‘actually communicate, and to whom, may not be what their producers and promulgators sought to communicate’.⁴⁷ Allusion has already been made to the misuse of data. One particular problem for human rights practice is that some countries may be judged too lightly or harshly because of problems in the data or method rather than reality. This risk is particularly prominent in rankings methods but exists in any approach that seeks to arrive at a normative conclusion. As in any other area, qualitative and cross-checking methods are needed for interpretation and awareness is needed as to how data will be used in the public sphere.

IV. Moving Forward with an Agenda

In developing the field of human rights measurement, data creation and testing needs to take a prominent place. It is obviously important to identify how existing data can be used. The papers by Vizard and Randolph and Guyer demonstrate how this can be done in a rigorous but also consultative manner. But this is not adequate: many of the most pressing human rights issues are simply not measured by official agencies. Whether it is forced evictions, political participation, freedom of speech, affordability of certain social rights or corruption, statistical agencies provide few meaningful, reliable or comparable measures.⁴⁸ The result is that academics and NGOs collect, analyse and report most of this data

46 For example, in another paper, Randolph has disaggregated the US score by racial groups and states to highlight disparities. Randolph, Susan, Michelle Prairie, John Stewart (2009) ‘Economic Rights in the Land of Plenty: Monitoring State Fulfillment of Economic and Social Rights Obligations in the United States’ <http://www.serfindex.org/wp-content/uploads/2010/12/Research-Economic-Rights-in-the-Land-of-Plenty_WithoutAnnexes.pdf>. They can be used in complementarity with other more detailed analysis that disaggregates by right, such as Corkery and Way in this issue that arrives at the same conclusion regarding the poor performance of the US in comparison with other OECD countries.

47 *Davis, Kingsbury and Merry* (n 39) 10.

48 For in-depth discussion of this problem, see C Naval, S Walter and R Suarez de Miguel (eds.), ‘Measuring Human Rights and Democratic Governance: Experiences and Lessons from Metagora – Special Issue’ (2008) 9 *OECD Journal on Development*. Note that the Metegora project has had success in motivating statistical agencies to work on data collection projects concerning what were traditionally considered sensitive human rights issues.

but this approach faces its own challenges of bias, reliability, validity, and the lack of repetition. Efforts such as WPCA and TIESR extend the frontiers of quantitative methods but face these ongoing challenges. Greater effort needs to be made in pushing agencies to include data that matters as well as support non-government actors to expand their capacity to create data given that this will be the only solution in the medium-term. But human rights groups need to consider carefully the limitations in collecting new data (given the demands on agencies) and whether compliance or a policy-actionable data should be prioritised.

A second area is to better coordinate collection, analysis and use of data. This will be helped by making data available, free of charge, online as far as possible, when it does not compromise the sources of data, and providing spaces for networking and peer review. This should minimise duplication of resources but also ensure comparable development of data sets and methods.

A third is cooperation with social scientists. Many academics and researchers use human rights data sets and methods but are asking different questions. A clear example is Jung and Rosevear in this special issue: their analysis of the constitutionalisation of social rights seeks to answer questions about patterns of state behaviour and the effects of constitutionalisation. However, their database also provides a tool for measuring a state's normative commitments to judicialise economic and social rights or incorporate the ICESCR.⁴⁹ Social science explanatory and operational methods can be used to justify recommendations that stem from compliance findings. The findings in this issue on the contribution of rights to economic growth and declines in infant mortality are examples. However, further engagement may also help build the skills of human rights practitioners and foster collaboration with scholars whose work may support compliance, advocacy or impact assessment efforts.

The collection of papers in this issue demonstrates the promise and perils of human rights measurement. The strength of some approaches are the flaws of others. Aggregation to the country as the unit of analysis and inter-country comparability can provide an overview of global trends in human rights and an evaluation of performance on a common standard taking account of both positive and

49 For instance, the Committee on Economic, Social and Cultural Rights in General Comment 9 asked states to justify the absence of judicial remedies and incorporation and its concluding observations regularly address the subject now. M Langford and JA King, 'Committee on Economic, Social and Cultural Rights: Past, Present and Future' in M Langford (ed) *Social Rights Jurisprudence: Emerging Trends in International and Comparative Law* (Cambridge University Press, Cambridge 2008). Moreover, the issue will be raised in cases taken under the Optional Protocol to ICESCR: i.e., whether a state has domestic remedies that must be exhausted.

negative obligations. But it faces the challenges of validity, reliability and equivalence over widely varying contexts. The development of country or locality-specific indicators provides measures that may reflect more accurately the contextual application of universal principles and be more actionable in practice. But they cannot be compared with other countries which help in establishing what would be reasonable performance and, to a certain extent, exploring broader causal relationships. In practice both approaches are needed.

A common theme that emerges in the issue is the need for a *broad* evidence base. Qualitative methods tend to be critical at all stages in the use of quantitative methods for human rights, providing (1) the basis for framing a hypothesis for metrics or motivating data creation; (2) the source of data creation in many cases (3) a constant check on data creation techniques, indicator matching or methods; (4) and complementary evidence for interpreting quantitative results. Both Vizard and Corkery and Way propose systematic methods for the integration of quantitative and qualitative methods. Indeed, quantitative methods will usually not do more than create a *prima facie* argument – the rest is often left to qualitative methods. One needs to avoid the danger of turning exercises of judgment into ones of measurement.