

THE PROBLEM OF CAUSALITY IN INTERNATIONAL HUMAN RIGHTS LAW

DAVID MCGROGAN*

Abstract The field of human rights monitoring has become preoccupied with statistical methods for measuring performance, such as benchmarks and indicators. This is reflected within human rights scholarship, which has become increasingly ‘empirical’ in its approach. However, the relevant actors developing statistical approaches typically treat causality somewhat blithely, and this causes critical problems for such projects. This article suggests that resources—whether temporal or fiscal—may be better allocated towards improving methods for identifying violations rather than developing complicated, but ultimately ineffective, statistical methods for monitoring human rights performance.

Keywords: empirical research, human rights, international law, philosophy of science, social science.

I. INTRODUCTION

In its 2008 report to the Committee on Economic, Social and Cultural Rights (CESCR),¹ the UK informed the Committee that, amongst many other things, it had a strategy to reduce inequalities in health outcomes by 10 per cent;² that 58.5 per cent of 15-year-old school pupils achieved five or more A*–C grade GCSEs or equivalent in the period 2005–06;³ and that the number of households defined as eligible for assistance against homelessness had fallen by 43 per cent since 2006.⁴ This is by no means unusual. The international human rights system, broadly construed, is increasingly interested in aggregate outcomes—a phenomenon which sees human rights monitoring as an exercise of measuring performance across populations through statistical techniques. Notwithstanding the recent development of an individual complaints procedure for the International Covenant on Economic, Social and Cultural Rights (ICESCR), the international human rights system and its

* School of Law, Northumbria University, david.mcgrogan@northumbria.ac.uk. The author wishes to thank Lee McConnell and Rhona Smith for helpful comments on an early draft of this article.

¹ The UK’s Fifth Periodic Report, CESCR, UN Doc E/C.12/GBR/5 (31 January 2008). The UK’s most recent report, the Sixth, is available at UN Doc E/C.12/GBR/6 (25 September 2014).

² *ibid.*, paras 302–305. ‘Inequalities in health outcomes’ meant the difference in figures for infant mortality and life expectancy at birth between the fifth of areas with the worst health and deprivation indicators and the rest of the population.

³ *ibid.*, para 334. The actual report states the period as ‘2006–2006’ [sic]; it is assumed this is a typographical error for ‘2005–2006’.

⁴ *ibid.*, para 286.

academic study now concerns itself in large part with how to monitor human rights performance across groups, populations, and societies. Whether a given individual's rights have been violated in a given circumstance is, in the context of the UN treaty system, a question that is becoming almost quaint; the focus is increasingly on how far *in general* human rights are being protected in a given State, as evidenced by measured outcomes.

Critics of this phenomenon have raised concerns about the way in which human rights statistics are gathered and used,⁵ about how statistical indicators can act to obscure truth or to mask political choices,⁶ or even how the use of statistics in international governance ushers in a new era of audit and control.⁷ The author shares these concerns, and adds a more foundational epistemic one: at the heart of this development towards outcomes measurement there is a conceptual blind spot. That conceptual blind spot is causality. What *causes* statistically measured human rights outcomes?

Identifying and attributing causality—the making of ‘credible causal inferences’⁸—in human societies is fraught with difficulty. This has been known and understood since David Hume was writing, and is nowadays often tritely summarized with the maxim ‘correlation is not causation’. This means that, for instance, establishing whether the UK government actually *caused* the reported fall in the number of households defined as eligible for assistance against homelessness of 43 per cent between 2006 and 2009 is not straightforward. The fall may have been correlated with all manner of changes both in government policy and in the economic and social sphere, but identifying the spurious correlations and separating them from genuine causes is difficult—if not, indeed, impossible. By extension, therefore, the statistic does not actually demonstrate anything, on its own, about human rights performance, because unless the underlying causality is understood, or adequately and persuasively theorized, the quantitatively-measured human rights outcome of ‘number of households defined as eligible for assistance against homelessness’ cannot be attributed to specific actions of the British State. While, of course, statistical outcomes (if accurately measured and appropriately selected) may reflect the lived experiences of the right-holders in some sense, that does not necessarily permit assessment of compliance with treaty obligations.

⁵ See eg R Barsh, ‘Measuring Human Rights’ (1993) 15 *HumRtsQ* 87.

⁶ See eg A Rosga and M Satterthwaite, ‘The Trust in Indicators: Measuring Human Rights’ (2009) 27 *Berkeley Journal of International Law* 253.

⁷ See eg N Bhuta, ‘Governmentalizing Sovereignty: Indexes of State Fragility and the Calculability of Political Order’ in K Davis *et al.* (eds), *Governance by Indicators: Global Power through Quantification and Rankings* (Oxford University Press 2012) 133–61 and T Halliday, ‘Legal Yardsticks: International Financial Institutions as Diagnosticians and Designers of the Laws of Nations’, *ibid.*, 180–216.

⁸ D Ho and D Rubin, ‘Credible Causal Inference for Empirical Legal Studies’ (2011) 7 *Annual Review of Law and Social Science* 17.

This problem is elided to some degree by the notion of the obligations to respect, protect and fulfil,⁹ but, as we shall see, that elision is not particularly satisfactory if genuine improvement of human rights protection is sought. Moreover, doctrinally, while there are good reasons for arguing that the rules regarding State responsibility have certain unique characteristics in the field of human rights, the treaty texts are typically phrased in such a way that requires close assessment of the effectiveness of measures taken—which by definition requires a clear understanding of cause and effect. This is true in general, given that most of the major human rights treaties frame the obligations in terms of ‘appropriate measures’ or similar, but is especially true in the field of economic and social rights, which hinge on whether resources are being allocated appropriately. The system, in other words, is predicated on the assessment of effectiveness, but assessments of effectiveness can only be made if the underlying causality is known or can be persuasively argued.

Social scientists in various fields have in recent years increasingly begun to grapple with Hume’s ‘problem of causality’, even as a revolution in ‘Big Data’ looms on the horizon.¹⁰ It is nowadays widely recognized and understood that the problem of causality—the difficulty of making credible causal inferences—cannot simply be ignored or dismissed as nitpicking. In particular, the move towards experimental and quasi-experimental techniques which is well underway, whether in the fields of psychology,¹¹ public policy,¹² political science,¹³ or law,¹⁴ must be interpreted as a widespread rejection of the possibility that statistical measurement or econometric analysis alone can be a guide for making policy or a method for assessing its effectiveness on its own. Yet even while these approaches may hold some promise in certain fields, for human rights assessment, the complexity of the system and its actors, financial and other pragmatic concerns, and the fact that quasi-experimental techniques do not unlock the ‘black box of causality’¹⁵ mean that they are unlikely to bear any fruit for the foreseeable future.

⁹ See S Leckie and A Gallagher, *Economic, Social and Cultural Rights: A Legal Resource Guide* (University of Pennsylvania Press 2006) xx.

¹⁰ References to ‘Big Data’ are inescapable in the modern age. See eg ‘Data, Data Everywhere’, *The Economist* (25 February 2010); A Sind, *Big Data Analytics* (MC Press 2012).

¹¹ See for instance the B Nozek *et al.*, ‘Estimating the Reproducibility of Psychological Science’ (2015) *Science*; and the Reproducibility Project: Psychology (at <<https://osf.io/ezcuj/>>).

¹² See eg A Finkelstein *et al.*, ‘The Oregon Health Insurance Experiment: Evidence from the First Year’ (2011) NBER Working Paper 17190, available at <<http://www.nber.org/papers/w17190>>.

¹³ See eg J Druckman, D Green *et al.*, ‘Experimentation in Political Science’ in Druckman, Green *et al.* (eds), *The Cambridge Handbook of Experimental Political Science* (CUP 2011) 3.

¹⁴ See eg DJ Greiner, ‘Causal Inference in Civil Rights Litigation’ (2008) 11 *HarvLR* 533.

¹⁵ See K Imai *et al.*, ‘Unpacking the Black Box of Causality: Learning about Causal Mechanisms from Experimental and Observational Studies’ (2011) 105(4) *American Political Science Review* 765.

This means that the assessment of human rights performance must remain a matter for theory, politics, and, above all, narrative. The promise of statistics to provide an objective basis for assessing compliance (despite the fact that there are some uses for statistics in human rights monitoring) is a mirage. Yet this does not mean that monitoring human rights performance must be an abstract, discursive, and superficial affair. In fact, if anything, it calls for a renewed focus on the individual, because it is at the level of the individual and individuals that assessments about causality can be credibly and sensibly made. That is, the UN human rights system could much more profitably focus its attention on what can be known—whether an individual’s human rights were violated—rather than on abstract, aggregated quantitative measurement where causality cannot be plausibly attributed.

II. MONITORING THROUGH OUTCOMES

Contemporary human rights literature tends to take as a given that human rights are to be fulfilled through identifying and realizing desirable outcomes (often conceptualized through ideals about ‘human dignity’¹⁶) across populations. In this approach, the individual tends to disappear from view, to be replaced by more general, aggregated measurement. While sometimes there is an acknowledgement of the necessity to ‘disaggregate’ data by sex, ethnicity, and so forth, the unit of primary interest is the group (whether the population at large, or a ‘disaggregated’ subsection of it) rather than the individual. Arguably the roots of this phenomenon are relatively old ones, dating back to the inception of the modern UN human rights system and the creation of the major treaties. The ICESCR, for instance, places an explicit obligation on States Parties to reduce infant mortality,¹⁷ amongst other things, and elsewhere a similar approach appears through implication: Article 6 of the ICESCR requires States Parties to achieve full employment; Article 24 of the Convention on the Rights of the Child (CRC) requires States Parties to take measures to combat malnutrition; and Article 10 of the Convention on the Elimination of All Forms of Discrimination against Women (CEDAW) requires States Parties to encourage coeducation. Other examples are scattered through all of the major treaties. These obligations by their nature suggest a system of monitoring which is primarily interested in the aggregate: what is the infant mortality rate? What is the unemployment rate? What proportion of children is malnourished? And so forth.

Yet the view that compliance is something that can be measured through assessing the level of achievement of outcomes has become increasingly fixed institutionally. Relatively early on, the typology of the obligations to

¹⁶ See for instance M Nussbaum, *Frontiers of Justice* (Belknap Press 2006) 277, arguing that what is needed is a definite account of what all the world’s citizens should have, and what their dignity entitles them to.

¹⁷ Art 12(2)(c).

‘respect, protect and fulfil’ individual rights became entrenched in the methodology of the treaty bodies.¹⁸ Only the first of these, the obligation to respect, has what would be thought of as a ‘negative’ character. The others, to protect and fulfil, respectively require States to engage proactively in ensuring that rights of individuals are not deprived by private actors (or to ‘creat[e] an environment in which rights are enjoyed’¹⁹); and to strengthen the capacity for individuals to enjoy their rights.²⁰ Despite the fact that rights inhere in individuals, the obligations to protect and fulfil naturally steer the focus of the treaty bodies and States towards the aggregate—towards the way in which the State attempts to create the appropriate environment or strengthen the capacity for individuals to enjoy rights. It hardly requires pointing out that those obligations naturally also imply that measurement is required; how well a given State Party is progressing is a question which is to be at least partially answered through primarily quantitative analysis.

It follows that the aim of States Parties should be to improve human rights outcomes, and that the focus should be on State obligations and to what extent they are being fulfilled: this is sometimes called a ‘duty-bearer perspective’.²¹ The crux of this perspective is that the concern ought not simply to be with enjoyment of rights on the part of the right-holders—that is, ordinary citizens. Its emphasis is rather the efforts which the State puts into achieving those outcomes.²² In other words, the primary focus is on the measures which the State takes to improve outcomes, rather than individual violations—the interest is not so much in whether given individuals are having their rights violated, but rather whether the State is succeeding in creating an environment in which rights are enjoyed, and in strengthening the capacity for individuals to enjoy their rights.

Finally, at a practical level, the UN human rights system is not constituted—either legally or technically—to be *primarily* concerned with monitoring compliance with treaty provisions at the individual level. This is a somewhat perverse observation given that, doctrinally at least, the view remains that human rights inhere in the individual rather than the group. Yet the treaty bodies tend to hew towards an aggregate or general perspective because of their role and composition. As a matter of law, violations of given individuals’ rights are only currently for the most part relevant in the optional individual communication procedures, because of the manner in which the treaties were created, and as a technical matter the treaty bodies do not in

¹⁸ The phrase’s most prominent first appearance seems to have been in the CESCR’s General Comment No 12 (UN Doc E/C.12/1999/5, 12 May 1999) para 15.

¹⁹ F Mégret, ‘Nature of Obligations’ in D Moeckli *et al.* (eds), *International Human Rights Law* (OUP 2014) 102.

²⁰ *ibid.*

²¹ See eg S Fukuda-Parr, T Lawson-Remer and S Randolph, ‘An Index of Economic and Social Rights Fulfillment: Concept and Methodology’ (2009) 8(3) *Journal of Human Rights* 195, 197.

²² See eg R Cignarelli and D Richards, ‘Measuring Government Effort to Respect Economic Human Rights: A Peer Benchmark’ in S Hertel and L Minkler (eds), *Economic Rights: Conceptual, Measurement, and Policy Issues* (CUP 2007).

their present form have the capacity or resources to focus their attention on the detail of individual cases.²³ It is natural, then, that in the international human rights system the notion of human rights as Dworkinian protections owned by individuals so as to trump the State should be superseded by a conceptualization of rights as mechanisms for guiding policy: as tools by which to achieve improvement towards agreed outcomes. The protection of the right to health becomes a measurable phenomenon using outcomes such as immunization rates;²⁴ the right to freedom from torture or cruel, inhuman or degrading treatment becomes partly a matter of assessing improvement in the outcome, ‘proportion of women reporting forms of violence against themselves or their children’²⁵ and so forth. The nature of human rights monitoring changes accordingly.

This in turn has naturally led to an increased interest in measurement—particularly quantitative measurement—of human rights outcomes within the field in general. It manifests itself in the routine work of the UN treaty bodies, as, for instance, when we find the CESCR in its Concluding Observations on the UK’s most recent periodic report urging the UK to work towards reducing the wage gap between men and women in the private sector, to provide information on the impact of pension reform on disadvantaged and marginalized groups, and to fulfil its commitment to reducing health inequalities by 10 per cent by 2010.²⁶ It manifests itself in State reports themselves, as, for example, in the same report, which contains an entire page of extensive statistics regarding maternity and paternity work arrangements (‘the average period of maternity pay leave is now six months, up from four months in 2002 ... the proportion of dads [sic] taking more than two weeks rose from 22% to 36% in just three years ... 77% of new mothers think that fathers are confident of caring for a child’²⁷). It is a dominant theme in the supporting work of the Office of the High Commissioner for Human Rights, which has been developing, over the preceding decade, a structure and methodology for monitoring human rights performance based on the use of largely quantitative human rights indicators²⁸ that has been adopted by a number of domestic human rights institutions.²⁹ This breaks rights down into handfuls of attributes which are then further subdivided into

²³ See eg H Steiner, ‘Individual Claims in a World of Massive Violations: What Role for the Human Rights Committee?’ in P Alston and J Crawford (eds), *The Future of UN Human Rights Treaty Monitoring* (CUP 2000) 15.

²⁴ See eg OHCHR, *Human Rights Indicators: A Guide to Measurement and Implementation* (2012) 49. ²⁵ *ibid* 91.

²⁶ Committee on Economic, Social and Cultural Rights, Concluding Observations on the UK’s fourth to fifth periodic report (2009) UN Doc E/C.12/GBR/CO/5, paras 18, 23 and 32.

²⁷ *ibid*, paras 233–234.

²⁸ OHCHR (n 24).

²⁹ They include those in the UK, Brazil, Bolivia, Ecuador, Mexico, Paraguay, Portugal, the Philippines and Kenya. See the UK EHRC’s Human Rights Measurement Framework, available at <<http://www.equalityhumanrights.com/about-us/our-work/human-rights/human-rights-measurement-framework>>; and also OHCHR, 2014, at <<http://www.ohchr.org/EN/NewsEvents/Pages/HumanRightsIndicators.aspx>>.

structure, process and outcome indicators purporting to evidence commitment, effort and results respectively, all predicated on the previous UN Deputy High Commissioner for Human Rights' motto that 'If you don't count it, it won't count.'³⁰ And it is increasingly a preoccupation in the scholarship of human-rights-focused academics, with outcomes-measurement now a burgeoning field, characterized by an attempt to apply greater rigour and conceptual clarity to the notions of the duty-bearer perspective and progressive realization. Prominent examples include the Cingranelli and Richards' eponymous Human Rights Data Project (the CIRI),³¹ and the Index of Social and Economic Rights Fulfillment (SERF Index) project at NYU, which aims to develop not only a measurement tool for economic and social rights fulfilment but also a method for ranking States on the basis of the extent to which they are complying with their obligations under the ICESCR.³² A further instance is the blossoming field of human rights budget analysis, which has become fashionable not only in the academic sphere,³³ but also in the UN human rights system,³⁴ and even amongst some domestic human rights institutions.³⁵

This is undoubtedly part of a broader social-scientific movement towards greater use of 'empirical' methods³⁶ which has developed in legal scholarship over the past two decades, and must in turn surely be located as part of a wider phenomenon in the humanities and social sciences overall.³⁷ As elsewhere, a field which was once defined almost exclusively by either doctrinal argument or normative prescription has been transformed into one preoccupied with measurement. Improvement in human rights performance comes in the form of 'better' statistical outcomes which demonstrate that a right (envisioned as a kind of facet in the protection of human dignity, however that is defined) is being fulfilled in the aggregate. A higher proportion of seats in a parliament being held by women and members of

³⁰ OHCHR (n 29).

³¹ See R Cingranelli and D Richards, 'The Cingranelli and Richards (CIRI) Human Rights Data Project' (2010) 32 *HumRtsQ* 395.

³² See Fukuda-Parr, Lawson-Remer and Randolph (n 21) 195; and 'Economic and Social Rights Fulfillment Index: Country Scores and Rankings' (2010) 9(3) *Journal of Human Rights* 230.

³³ See especially A Nolan *et al.* (eds), *Human Rights and Public Finance: Budgets and the Promotion of Economic and Social Rights* (Hart 2013).

³⁴ See eg OHCHR, 'Report of the High Commissioner for Human Rights on Implementation of Economic, Social and Cultural Rights' (2009) UN Doc E/2009/90.

³⁵ Most notably South Africa—see South African Human Rights Commission and Studies in Poverty and Inequality Institute, 'How Much Are We Spending on Transforming Our Society? A Rights-Based Analysis of the 2011 Budget' (2011).

³⁶ See for instance a recent volume of the *Leiden Journal of International Law* (28(2), 2015) on the 'new legal realism'. See also G Shaffer and T Ginsburg, 'The Empirical Turn in International Legal Scholarship' (2012) 106 *American Journal of International Law* 1; D Ho and L Kramer, 'Introduction: The Empirical Revolution in Law' (2013) 65 *StanLR* 1195.

³⁷ 'Quantification is a constitutive feature of modern science and social organization', as Espeland and Stevens put it, in W Espeland and M Stevens, 'A Sociology of Quantification' (2008) 49(3) *European Journal of Sociology* 401, 402.

‘target groups’ indicates the right to participate in public affairs is being fulfilled;³⁸ a higher proportion of the population using an improved drinking water source indicates improvement regarding the right to adequate housing;³⁹ an increase in the proportion of adults with a BMI of less than 18.5 indicates failure to protect the right to adequate food;⁴⁰ an increase in the waiting list for social housing correlated with lower investment indicates that the State is failing to use its maximum available resources to protect the right to housing,⁴¹ and so forth. Human rights monitoring—whether undertaken by the treaty bodies or by academics or practitioners—is becoming increasingly sophisticated, moving away from its fairly rudimentary roots towards a technical exercise incorporating econometric and statistical methods which purport to revolutionize the manner in which compliance with human rights treaty obligations is assessed.⁴² Human rights in turn almost become conceptualized as drivers of public policy: articulations of social justice goals, progress towards which can be quantitatively measured.

III. CAUSALITY AND THE ‘OUTCOMES APPROACH’

There has been a level of criticism of this approach. Meckled-Garcia, for instance, sees in the outcomes-view a consequentialist tendency which disrupts the very notion of human rights as *rights*,⁴³ whereas Koskenniemi questions a growing managerialist tendency amongst contemporary human rights advocacy, seen most clearly in the move towards human rights ‘mainstreaming’ a critique which seems by extension to have much to say with respect to the foregoing.⁴⁴ This article acknowledges those critiques, but raises an additional epistemic concern: the question of what causes a given human rights outcome is not a trivial one.

In the first place, though, it is necessary to establish why causation matters for an outcomes or duty-bearer approach to human rights monitoring—for typically it is treated somewhat blithely in the field of human rights, where it is generally taken as a given that measured outcomes are attributable to the State. There is a certain doctrinal basis for this. To most human rights scholars, State responsibility engages when a State is in breach of an international obligation, whether through act or omission, and since the core human rights treaties all to some degree or other require State Parties to ensure, protect,

³⁸ OHCHR (n 24) 94.

³⁹ *ibid* 92.

⁴⁰ *ibid* 89.

⁴¹ See E Rooney and M Dutschke, ‘The Right to Adequate Housing: A Case Study of the Social Housing Budget in Northern Ireland’ in Nolan *et al.* (n 33) 195–217.

⁴² See for instance the work of the Human Rights Data Analysis Group at the American Association for the Advancement of Science, at <<https://hrdag.org/>>.

⁴³ S Meckled-Garcia, ‘What is the outcomes view? Contemporary consequentialist theories of human rights’ (on file with the author).

⁴⁴ M Koskenniemi, ‘Human Rights Mainstreaming as a Strategy for Institutional Power’ (2010) 1(1) *Humanity: An International Journal of Human Rights, Humanitarianism and Development* 47.

secure, or promote the rights they contain,⁴⁵ then it follows that if those rights are not being ensured, secured, etc, then a violation or violations has taken place for which the State has responsibility.⁴⁶ In contemporary human rights law, in other words, the distinction between public and private actors which the Articles on State Responsibility (2001) enshrine effectively disappears. It does not matter that, for instance, a slum clearance leaving people homeless may have been carried out by a private landlord. The State failed to create an environment in which the right to housing was protected, respected and fulfilled—through omission in failing to provide alternative social housing or appropriate legislative protections—and hence it was in violation of its obligations vis-à-vis that right.⁴⁷ State responsibility engages almost irrespective of the actor. The tripartite obligations to respect, protect and fulfil reinforce this in suggesting that a State is in violation of its obligations simply by dint of failing to create an environment in which the rights of those in its jurisdiction are protected. It follows that causation can be elided, and it can be readily established that a violation or violations have taken place on the basis of a statistical observation alone. There is objective-seeming evidence that the State is not creating an environment in which rights are enjoyed (or exercising ‘due diligence’ in preventing private acts which impact on that enjoyment⁴⁸), whether through act or omission, and hence there is a violation. It follows that States’ obligations come to be conceptualized as requirements to improve across statistical measures: the number of households eligible for assistance against homelessness has declined by 43 per cent, *ergo* the UK is performing well in terms of protecting, respecting and fulfilling the right to housing under Article 11 of the ICESCR.

This elision of, or blitheness about, causation results in both conceptual and practical problems. First, simply from a common sense perspective, it is unsatisfactory that any given measure of human rights performance should be disconnected from causal explanations. While, for instance, a fall of 43 per cent in the number of households eligible for assistance against homelessness is to be welcomed, it is surely necessary to understand why that fall took place if either the government concerned or the treaty body monitoring system has any interest at all in causing the number to drop yet further. Second, it clearly runs contrary to widely shared notions of fairness

⁴⁵ For instance, art 2 of the ICCPR requires States Parties to ‘respect and to ensure to all individuals within its territory and subject to its jurisdiction the rights recognized in the present Covenant’.

⁴⁶ See eg S Farrow, ‘State Responsibility for Human Rights Abuses by Non-State Actors’ (1988) 92 ASILProc 299.

⁴⁷ As in eg *Government of South Africa v Grootboom*, Constitutional Court of South Africa, Case CCT 11/00, 4 October 2000.

⁴⁸ Most famously the Inter-American Courts of Human Rights developed this approach in the *Velasquez-Rodriguez Case (Honduras)* 4 IACtHR (ser C) (1988), although the European Court of Human Rights has used similar if more restricted reasoning in eg *Plattform ‘Ärzte für das Leben’ v Austria* ECtHR 10126/82 (1988).

and justice to attribute liability, or assign praise, where it is not due. In the long term, it cannot be to the advantage of the UN human rights system in general to undermine its own legitimacy by relying on statistical measurement of 'outcomes' whose underlying causality may be justifiably disputed. And third, it ought to be self-evident that those engaged in the monitoring of human rights should be interested in truth for its own sake.

Yet there are also compelling legal considerations. In the first place, despite there being doctrinal arguments for holding States responsible in general for the extent to which human rights are protected in their jurisdictions, this should not be permitted to vitiate the requirement to establish causal links between State act or omission and the measurement in question. To take a paradigmatic example, the CEDAW plainly assigns responsibility to State Parties in preventing discrimination against women: it requires them to take all appropriate measures to modify or abolish not just discriminatory legislation but also customs and practices constituting discrimination, and also to take appropriate measures to modify 'social and cultural patterns of conduct of men and women' with a view to eliminating prejudices and so forth.⁴⁹ The public/private divide is clearly not applicable or relevant with respect to these provisions, and it is tempting to ignore the issue of causation in light of this: if discriminatory practices or attitudes are evidenced statistically, then by definition the State is in violation of such requirements. But this leaves unanswered the critical questions of how discriminatory customs and practices, or social and cultural patterns of conduct, can in fact be modified. What causes a discriminatory practice to develop in the first place? What causes it to continue? And what might cause it to disappear? Our interest in such questions comes not only from a concern with what might constitute best practices or what might be the best policy to implement; it also stems from the nature of States Parties' obligations under the various treaties.

This is because, while in the International Covenant on Civil and Political Rights (ICCPR) it is largely implicit,⁵⁰ all of the major human rights treaties frame State Party obligations around 'appropriate measures' or similar, in such a way that *effectiveness* of measures taken must be assessed in order to establish compliance. This is most obvious in the case of the ICESCR, which requires States Parties to take steps towards progressively realizing the rights contained in the Covenant by all appropriate means. Clearly, the question of whether the steps a State is taking *do* in fact help progressively realize the relevant rights can only be answered through understanding and assessing the effect of those steps on rights protections. Similarly, the

⁴⁹ See the CEDAW, arts 2 and 5.

⁵⁰ Only art 23(4) of the ICCPR mentions 'appropriate steps' (in relation to equality in marriage), but assessing appropriateness/effectiveness of measures seems implicit in the due diligence obligations generally interpreted to be in the Covenant.

measurement of effectiveness is immanent in the question of what is or is not 'appropriate'. This requirement is only made more acute by the requirement that States Parties use their 'maximum available resources' to realize Covenant rights. The CRC Committee, whose Covenant contains a similar obligation, has essentially expressed the view that this sort of requirement can be monitored simply by identifying which portions of a State's budget are allocated towards fulfilling rights.⁵¹ Yet, as is often explicitly or implicitly acknowledged, this is only half the story: the requirement is that States take steps towards progressively realizing rights and also spend the maximum available resources on doing so—not merely that they expend the maximum available resources on rights goals.⁵² There must be some demonstration that the resources in question are actually being expended in such a way as to progressively realize rights protections. Thus Magdalena Sepúlveda, the former Special Rapporteur on the Question of Human Rights and Extreme Poverty, considers the obligation to mean that expenditures must be shown to be efficient and effective; that corruption must be curbed; that funds assigned to ESC rights purposes must be fully expended for that purpose, and so forth.⁵³ The CESCR, meanwhile, interprets the obligation as permitting it to take into account whether a State is adopting a measure which 'least restricts Covenant rights' out of those available when assigning resources—and will only view retrogressive steps as permissible if they have been introduced after consideration of all alternatives.⁵⁴ Clearly none of this can be achieved without a mechanism for evaluating the impact of resource expenditure on actual rights protections: in other words, the extent to which a given expenditure causes a given outcome.

Similar reasoning applies in most other treaty contexts. The CERD, for instance, in Article 2 requires States Parties to undertaking 'appropriate means' to eliminate racial discrimination, including by 'taking effective measures' to amend or rescind regulations which create or perpetuate racial discrimination or encouraging the elimination of barriers between races. Again, immanent in those requirements are questions such as: what are the appropriate means to eliminate racial discrimination? Which regulations create or perpetuate racial discrimination, and what would be effective measures to amend them? How can barriers between races be eliminated? Establishing cause and effect is clearly crucial in answering those questions. Likewise under the CEDAW there are requirements to take all 'appropriate

⁵¹ See eg the CRC's General Comment No 5, UN Doc CRC/GC/2003/5 (2003) para 51, describing how States ought to identify the proportion of their national budgets allocated to the social sector and, within that, to children.

⁵² See *ibid.*, para 45, stating the need for predicting the impacts of imposed laws and child impact evaluation in all actions concerning children.

⁵³ See M Sepúlveda, *The Nature of the Obligations under the International Covenant on Economic, Social and Cultural Rights* (Intersentia 2003).

⁵⁴ UNCESCR, 'An Evaluation of the Obligation to Take Steps to the "Maximum of Available Resources" under an Optional Protocol to the Covenant' UN Doc E/C.12/2007/1, paras 8–12.

measures' to eliminate discrimination against women by any person, organization or enterprise; to 'modify the social and cultural patterns of conduct of men and women'; and to ensure that there are equal rights between men and women in education—amongst many other things.⁵⁵ Since these requirements are substantive as well as *de jure* in character⁵⁶ there is inescapably a need to assess the effectiveness or appropriateness of measures taken, which can only be achieved through understanding cause and effect: what, for instance, is the State doing to ensure that there is *de facto* equality in education, and is it having an impact? This will be a consideration for the vast majority of obligations throughout the core human rights treaties.

A fascinating illustration of the need for understanding cause and effect in establishing whether 'appropriate measures' have been taken is given in the CEDAW Committee's inquiry, based on Article 8 of the Optional Protocol to the CEDAW Convention, into the abduction, rape and murder of women in Northern Mexico.⁵⁷ Here, many different measures for preventing gender-based violence in the area of Ciudad Juárez are described and discussed; one example is the introduction of 700 members of the 'preventive federal police' in the city to improve security and provide community support activities to enhance social integration. But in the words of the Committee:

There is no consensus between the authorities and non-governmental organizations in their assessment of the federal presence in Ciudad Juárez. The authorities stress that progress has been made in improving security and reducing crime. The non-governmental organizations stress that the presence of the preventive federal police does more to intimidate people than to prevent crime, and that the patrols are more likely to be in areas where robberies occur than in areas where women are at risk.⁵⁸

This example neatly demonstrates the difficulty of actually translating treaty requirements into a method for assessing whether a violation has taken place, in the absence of a clear understanding of the underlying mechanisms of cause and effect. As far as the Committee is concerned, State Parties have an obligation arising under the anti-discrimination articles of the Convention to 'take appropriate and effective measures to overcome all forms of gender-based violence, whether by public or private act'.⁵⁹ Does the presence of 700 federal police members in Ciudad Juárez qualify as appropriate or effective? Without knowing the actual effects on gender-based violence of the presence of the federal police—that is, without an understanding of the causal mechanisms underlying the rate of gender-based violence in the city—it is

⁵⁵ CEDAW, arts 2(e), 5(a) and 10.

⁵⁶ See eg CEDAW, General Recommendation No 28, UN Doc CEDAW/C/2010/47/GC.2, para 16.

⁵⁷ Report on Mexico produced by the Committee on the Elimination of Discrimination against Women under art 8 of the Optional Protocol to the Convention, and reply from the Government of Mexico, CEDAW 27 January 2005, UN Doc CEDAW/C/2005/OP.8/MEXICO.

⁵⁸ *ibid*, para 184.

⁵⁹ CEDAW, General Comment No 19, para 24(a).

impossible to draw any conclusion about its appropriateness or effectiveness, and hence whether Mexico's obligations are being met. This can only be established if it can be plausibly demonstrated that the introduction of the federal police not only is reducing gender-based violence, but is also, critically, *more effective* than other possible policy measures.

Such considerations will, in essence, be true wherever there is an attempt to measure human rights quantitatively, and are of critical importance in the use of indicators: if there is no clear causal link between government policy and an indicator, then the indicator demonstrates nothing about the effectiveness of the policy. This is particularly so where indicators are categorized into structure, process and outcome. Indeed, Donabedian himself, the originator of the OHCHR's much-vaunted model for human rights indicators as it was first used in the field of health care, was quite clear about how crucial it was to establish cause and effect in structure, process and outcome rather than simply to assess them naively or in isolation. 'There must be pre-existing knowledge of the linkage between structure and process, and between process and outcome, before quality assessment can be undertaken.'⁶⁰ That is, '[t]he three-part approach to quality assessment is possible only because good structure increases the likelihood of good process, and good process increases the likelihood of a good outcome. *It is necessary, therefore, to have established such a relationship before any particular component of structure, process or outcome can be used to assess quality.* [Emphasis added.]'⁶¹ In other words, for the structure-process-outcome model to demonstrate anything at all about performance, there must be an understanding of how structural indicators—commitments—bring about better policy ('process'), and how this in turn fosters better results, or 'outcomes'.

This cannot be demonstrated without understanding the underlying causality. Taking an example from the OHCHR's *Guide* to illustrate, under the right to food an outcome indicator for the 'Nutrition' attribute is 'prevalence of underweight and stunted children under five years of age'.⁶² This is directly linked to four process indicators: the proportion of the targeted population brought above the minimum level of dietary energy consumption in the reporting period, the proportion of the population covered under public nutrition supplement programmes, the coverage of public programmes on nutrition education and awareness, and the proportion of the population with access to an improved drinking water source. It is also linked to two process indicators which are shared by all outcome indicators under the right to food: the proportion of received complaints on the right to food which are investigated by the relevant authorities, and the net official development assistance for food

⁶⁰ *ibid.*

⁶¹ A Donabedian, 'The Quality of Care: How Can it Be Assessed?' *Journal of the American Medical Association* (23/30 September 1988) 260(12) 1743, 1746.

⁶² OHCHR (n 24) 89.

security as a proportion of public expenditure on food security.⁶³ Setting aside concerns about data collection, the primary concern here must be to what extent the process indicators offered (for instance, the coverage of public programmes on nutrition education and awareness) actually result in—or cause—the outcome, ‘prevalence of underweight and stunted children under five years of age’. Without an accurate understanding of this, the process indicator demonstrates effectively nothing (either positive or negative) about performance: it has no usefulness as an assessment tool for actually monitoring the extent to which the State is living up to its obligations as a duty-bearer. It may be that 100 per cent of the population is covered by a public programme on nutrition education and awareness, but unless the effect of that programme on the prevalence of underweight and stunted children under five years of age is actually known, the figure of 100 per cent is simply a statistical observation. It may have a high or low impact on child nutrition, or none at all. (Or the impact could indeed even be negative if the educational content of the programme is erroneous.)

The requirement for understanding underlying causality is perhaps at its strongest with respect to human rights budget analysis. Here, again, in the abstract there appears to be a strong case for monitoring via resource allocation, which in practice requires a strong understanding of cause and effect. Kempf, for instance, suggests an ‘information pyramid’ approach which divides rights into three tiers—key measures, expanded indicators, and context.⁶⁴ The middle of these typically involves measuring government expenditure so as to give a ‘more in-depth understanding of the forces at work behind the key indicator’.⁶⁵ This would result, for example, in the right to education being measured through the literacy rate (Tier 1); government expenditure on education, transport and lunch programmes (Tier 2); and case studies (Tier 3).⁶⁶ Here, clearly, there is a requirement to understand how government expenditure results in the literacy rate being what it is, and how increases or decreases in government expenditure affect it; the relationship between expenditure on education and literacy must be known in order to provide a proper and accurate assessment of performance. If, for instance, expenditure is wasted on ineffective teaching (which is a perennial problem in the developing world⁶⁷), then it is unlikely on its face to contribute to improving the literacy rate. On the other hand, improvements in the literacy

⁶³ *ibid.*

⁶⁴ I Kempf, ‘How to Measure the Right to Education: Indicators and Their Potential Use by the Committee on Economic, Social and Cultural Rights’, CESCR, 19th Session (1998), UN Doc E/C.12/1998/22.

⁶⁵ *ibid.*
⁶⁶ See also C Apodaca, ‘Measuring the Progressive Realisation of Economic and Social Rights’ in Hertel and Minkler (n 22) 176–7.

⁶⁷ See eg L Pritchett, *The Rebirth of Education: Schooling Ain’t Learning* (Center for Global Development 2013).

rate may be unrelated to government expenditure where, for instance, private schools and tutors are widely used.⁶⁸

This kind of consideration will always be necessary when attempting to analyse budgets from a human rights perspective in detail; it is the matter of only brief moments of thought to generate examples of why credible causal inferences are required if monitoring is to be performed through statistical outcomes. How does government expenditure on a given programme affect the unemployment rate? How would the unemployment rate have changed if expenditure had been different? How does expenditure on a given aspect of health care improve waiting times for routine operations? What if the money had been spent in a different way? If a local education authority approves the building of a new school where the old one was growing decrepit, is this a more suitable expenditure than using the money to employ more teachers? Which option has the most impact on literacy rates? These sorts of questions are inherent in any exercise which seeks to establish whether the best alternative has been chosen, or expenditures are efficient and effective. Yet they cannot be assessed without understanding how the respective human rights outcomes are caused. This is doubly the case where analysts seek to ‘disaggregate’ expenditure for the purposes of, for instance, ‘gender-responsive budgeting’ or similar⁶⁹—which means examining, for instance, what was a given budgetary item’s impact on gender inequality or people with disabilities.⁷⁰ For such measures, a sophisticated understanding not only of the impact of funding in general but also of its impact on the disaggregated group is also required—effectively doubling the analytical workload.

Finally, it bears emphasizing that if States have obligations to protect, respect and fulfil rights to the extent that the treaty bodies have generally argued, and especially where the text of a treaty provision suggests that there is no distinction to be made between public and private actors in terms of State responsibility, then much of the above discussion also holds true with respect to causality and the role of non-State actors. What roles private actors play in causing measured outcomes—and to what extent the actions of private actors are in turn ‘caused’ or contributed to by the State—are, of course, questions giving rise to a similar set of considerations, and this creates yet another layer of complexity and further requirements to demonstrate and understand cause and effect.

It is not just from practical and conceptual perspectives, then, that a failure to properly address matters of causation is problematic: it poses critical problems for the legal questions of whether a State is enacting appropriate or effective

⁶⁸ See eg J Tooley, *The Beautiful Tree: A Personal Journey into How the World’s Poorest People Are Educating Themselves* (Cato Institute 2009).

⁶⁹ See eg S Quinn, ‘Equality-Proofing the Budget: Lessons from the Experiences of Gender-Budgeting?’ in Nolan *et al.* (n 33) 163.

⁷⁰ *ibid* 174.

measures to achieve human rights protections. And as we shall now see, the apparent blitheness about causation serves to mask a host of difficulties associated with an outcomes-approach to human rights monitoring.

IV. THE PROBLEM OF CAUSALITY AND CREDIBLE CAUSAL INFERENCE

In recent decades, there has been a strong movement in econometrics, policy studies, and related fields, away from what might be called a naïve regression-based view of causation. This naïve view was perhaps most prominently and succinctly expressed by Leamer in his famous article ‘Let’s Take the Con out of Econometrics’.⁷¹ Leamer used an illustrative analogy of a comparison between an agricultural experimenter and an econometrician. The agricultural experimenter divides a farm into smaller plots of land and randomly selects which he will fertilize; if some plots are fertilized but some not, the difference in mean yield between the fertilized and the non-fertilized plots will be a measure of the effect of fertilizer on agricultural yields. This is the way econometricians like to think of themselves, according to Leamer, but in fact this is ‘grossly misleading’. Rather:

The applied econometrician is like a farmer who notices that the yield is somewhat higher under trees where birds roost, and he uses this as evidence that bird droppings increase yields. However, when he presents this finding at the annual meeting of the American Ecological Association, another farmer in the audience objects that he used the same data but came up with the conclusion that moderate amounts of shade increase yields. A bright chap in the back of the room then observes that these two hypotheses are indistinguishable, given the available data. He mentions the phrase ‘identification problem’, which, though no one knows quite what he means, is said with such authority that it is totally convincing.⁷²

The econometricians, in other words, do not understand that it is generally impossible to know or demonstrate convincingly what causes a statistical pattern through analysis of data that is not the product of a controlled experiment. The agricultural experimenter uses the nearest thing possible to a laboratory experiment, and his inferences about the effect of fertilization on crop yields are convincing. The econometrician attempts to infer causation from noticing a statistical pattern, but other econometricians infer different causal mechanisms and there is no way to distinguish between their competing

⁷¹ E Leamer, ‘Let’s Take the Con out of Econometrics’ (1983) 73(1) *American Economic Review* 31–43. See also eg R Berk, *Regression Analysis: A Constructive Critique* (Sage 2004); J Donohue and J Wolfers, ‘Uses and Abuses of Empirical Evidence in the Death Penalty Debate’ (2006) 58 *StanLRev* 791; J Pfaff, ‘A Plea for More Aggregation: The Looming Threat to Empirical Legal Scholarship’ SSRN Working Paper (2010), available at <http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1641435>. For an interesting technical exposition of some of the problems with using conventional regression analysis as a tool for inferring causation, see Ho and Rubin (n 8).

⁷² Leamer (n 71) 31.

causal claims. A similar process takes place in the monitoring of human rights by statistics: a fall in the number of households requiring assistance against homelessness is observed. Different observers may, however, infer different causal mechanisms, and there is no objective method to prefer one to another.

This is, of course, essentially a restatement of what David Hume had demonstrated philosophically in the mid-eighteenth century, which is that we can never ‘by our utmost scrutiny discover any thing but one event following another’.⁷³ That is, causality can never be proven, because there may always be hidden or unmeasurable conditionals affecting a given outcome. The laboratory experiment, which allows the measurement of known variables through holding others constant, is a suitable and practical method of diminishing the problem, but beyond the laboratory making causal inferences is fraught with problems.⁷⁴

Without straying too far into technical detail, regression analysis is often used as a tool for solving the problems social scientists encounter when attempting to isolate the effect of a variable. In layman’s terms, a regression analysis is a method of investigating relationships between variables, but typically it means seeking to ascertain causal effects, such as the effect of price on demand.⁷⁵ An example might be a model which attempts to measure the relationship between unemployment and the suicide rate; typically this would take the form of a ‘multiple regression’ which aimed to control for other independent variables than unemployment (eg sex, age, etc) in an attempt to determine how unemployment impacts on the suicide rate in isolation from other factors. It is, in other words, an attempt by a statistician to move away from the position of the farmer who observes the correlation between roosting birds and high crop yields, and towards the position of the agricultural experimenter who manipulates one variable—fertilization—while holding the others constant.

At root, the use of multiple regression analysis as a tool for inferring causation is predicated on measuring the effect of one variable while controlling for other variables—purely through statistical manipulation. The endeavour is always confounded, then, by the problem that not all other variables are necessarily known: indeed, it is not logically possible to be sure that all variables have been identified. This results in two insurmountable barriers to making credible causal inferences through statistical analysis alone.

The first of these is the problem of omitted variable bias: since controlling for all other relevant variables cannot be done—or at least, the statistician can never be sure that all other relevant variables have been controlled for—the results of the regression analysis could always potentially have been biased by the fact

⁷³ From D Hume, *An Inquiry Concerning Human Understanding*, section VII.

⁷⁴ Of course, Karl Popper used this as a dividing line between science and non-science. See K Popper, *The Logic of Scientific Discovery* ([1934] 1959 Hutchinson & Co).

⁷⁵ See eg D Montgomery *et al.*, *Introduction to Linear Regression Analysis* (5th edn, Wiley-Blackwell 2012); J Miles and M Shevlin, *Applying Regression & Correlation* (Sage 2001).

that there is a hidden conditional affecting the outcome. An illustrative example given by King and Keohane is a hypothetical study of sub-Saharan African States which finds that coups d'état appear more frequently where regimes are repressive. It is plausible, however, that high unemployment may be associated with an increased probability of both coups d'état and political repression.⁷⁶ Such a study would therefore need to control for unemployment, but it would not be possible to do this if accurate unemployment figures were unavailable. Even if those figures were available and the unemployment variable controlled for, however, the researchers may have overlooked the effect of another variable that might plausibly have an effect on the frequency of coups d'état: the independence of the military. They may find some way to control for that variable also, but then overlook the level of salary that soldiers could expect; dissatisfaction amongst soldiers may also have an effect on the likelihood of a coup d'état occurring. And so forth. The list of omitted potential variables may go on ad infinitum. And second, since the list of omitted potential variables may go on indefinitely, the results of a naïve regression-based analysis can always be disputed—as Leamer so aptly demonstrated: another scholar can always examine the same set of data and come up with a competing interpretation, with no means of deciding whose interpretation is preferable. This is largely the reason why so many perennial and widespread social debates have never been resolved, despite huge arsenals of statistical ‘evidence’ arrayed on either side: Pfaff gives the American-centric examples of whether the death penalty deters crime or whether gun ownership increases violence;⁷⁷ other examples might be whether abortion has any effect on the crime rate,⁷⁸ whether the minimum wage affects employment,⁷⁹ or whether microfinance actually helps the very poor.⁸⁰ Both ‘sides’ in such debates find it straightforward to identify omitted variables in each other’s data, and to identify their own correlations which confirm their existing biases, so neither is ever in a position to cede defeat.⁸¹

⁷⁶ See G King *et al.*, *Designing Social Inquiry* (Princeton University Press 1994) 170–1.

⁷⁷ See Pfaff (n 71) 4.

⁷⁸ See J Donohue and S Levitt, ‘The Impact of Legalised Abortion on Crime’ 116(2) *QJ Econ* (2001) 379, versus C Foote and C Goetz, ‘The Impact of Legalised Abortion on Crime: Comment’ (2008) 123(1) *QJ Econ* 407.

⁷⁹ See D Card and A Krueger, *Myth and Measurement: The New Economics of the Minimum Wage* (Princeton University Press 1995) versus R Burkhauser *et al.*, ‘Who Gets What from Minimum Wage Hikes: A Re-Estimation of Card and Krueger’s Distributional Analysis in *Myth and Measurement: The New Economics of the Minimum Wage*’ (1996) 49(3) *Industrial and Labor Relations Review* 547.

⁸⁰ See M Pitt and S Khandker ‘The impact of group-based credit programs on poor households in Bangladesh: Does the gender of participants matter?’ (1998) 106 *Journal of Political Economy* 958 versus A Banerjee *et al.*, ‘The miracle of microfinance? Evidence from a randomized evaluation’ (2015) 7(1) *American Journal of Economics: Applied Economics* 22.

⁸¹ See eg D Braman and D Kahan, ‘Cultural Cognition and Public Policy’ 24 *Yale L&Pol’y Rev* (2006) 147.

As well as the issue of variables being unknown is the question of how variables interact. JS Mill referred to this problem as the ‘intermixture of effects’,⁸² although it is more commonly referred to in the modern day as the problem of endogeneity. Put briefly, what Mill observed was that, when confronted with complexity, there is a tendency to attempt to single out ‘from the multitude of antecedent circumstances’ one condition as a potential cause, and then to measure it.⁸³ In fact, however, ‘causes’ may interfere with one another; they are not discrete, but intermingled. Manzi uses the example of attempting to assess the impact of brand difference on sales in shops, holding all other factors equal. A possible variable likely to affect sales is the presence of an ATM in a shop, and this therefore needs to be held constant if we are interested in measuring the impact of brand difference alone. But this may have different effects in different contexts: in a large shop, having an ATM may drive sales because it draws in customers, but in a small shop, having an ATM may reduce sales because it increases crowding near the cash register that discourages customers. Yet ‘holding the presence of an ATM constant’ in a typical regression equation only allows either a positive or negative coefficient for that variable—either an ATM is present in a shop or not. This does not capture the way the variable changes according to context. This problem is remedied by adding further interaction terms: replacing ‘ATM in shop’ with other variables such as ‘ATM in shop AND shop is large’ and ‘ATM in shop AND shop is small’, and so forth. But interactions-with-interactions can quickly become myriad: an ATM may increase net sales in large shops, but not when at a highway rest stop (motorway services, in British parlance)—so there would need to be further interaction terms: ‘ATM in shop AND shop is large AND shop is in highway rest stop’, and so forth. Interaction effects always tend to proliferate, and to do so exponentially.⁸⁴ For a typical example of how extreme these effects can become, Ho and Rubin discuss how introducing covariates for sentence length by month; and age, employment status, sex, prior strikes and marital status of prisoner, result in 69 million different parameters when attempting to measure the effect of prisoner classification status on misconduct.⁸⁵ The problem of endogeneity gives the lie to the notion, sometimes advanced in the literature, that the issue is one of counterfactuals: if only there was some way to know what *would* have happened had circumstances been different, causality could be observed.⁸⁶ The truth is even more complex: since variables interact, the mere act of controlling one variable may bias others.

⁸² JS Mill, *A System of Logic* (1843) ch X.

⁸³ This has elsewhere been called the ‘myth of monocausality’; see Miles and Shevlin (n 75) 28.

⁸⁴ J Manzi, *Uncontrolled* (Basic Books 2012) 136.

⁸⁵ Ho and Rubin (n 8) 26.

⁸⁶ See eg L Epstein and G King, ‘The Rules of Inference’ (2002) 69(1) UChiLRev 1, 36.

These and similar problems⁸⁷ have led to widespread acceptance in various disciplines that the ‘age of regression’ is over.⁸⁸ The kind of naïve use of regression analysis that sees scholars attempting to isolate and measure the effects of variables in a data set is no longer generally viewed as being a credible way to draw causal inferences except in limited cases. Instead, there has been a proliferation in past decades of what are often referred to as ‘quasi-experimental techniques’: better methods for replicating, or approximating, what goes on in the laboratory or the agricultural experimenter’s field.⁸⁹ The most widely known of these is the ‘gold standard’ of the randomized field trial, which is essentially what Leamer’s agricultural experimenter was performing, and which is used to some effect in the fields of medicine and public health: here, a group of like subjects are identified and randomly separated into a test group and a control group, with the test group having one variable manipulated so as to isolate its effects. This has not changed in principle since the experiments of James Lind to discover the effect of citrus juice on combating scurvy. While the randomized field trial is by no means perfect even in the field of medicine,⁹⁰ through widespread, continuous and rigorous replication it can ultimately persuasively demonstrate causality. This is because, with a large enough initial group which is then randomly assigned into test and control groups, and with good experimental design, it can be assumed that differences between individuals even out and the test and control groups are comparable in all other respects than the variable of interest, which is being manipulated for the test group.

In the social sciences, however, randomized field trials tend to be difficult to perform—usually because costs are prohibitive (although there are increasingly innovative ways of carrying out such experiments⁹¹). Where trials cannot take place, experimenters use various methods to attempt to replicate something approaching a randomized field trial through intervening in the data. One prominent method is what is called ‘regression discontinuity analysis’, which takes advantage of a natural break or discontinuity in the data to measure effects around it. Perhaps the most famous and frequently cited example of

⁸⁷ Other examples include unwarranted assumptions about linearity and homogeneity.

⁸⁸ See eg S Morgan and C Winship, *Counterfactuals and Causal Inference: Methods and Principles for Social Research* (CUP 2007); C Manski, *Identification Problems in the Social Sciences* (Harvard University Press 1995).

⁸⁹ See eg J Druckman *et al.*, ‘The Growth and Development of Experimental Research in Political Science’ (2006) 100 *American Political Science Review* 627; DJ Greiner, ‘Causal Inference in Civil Rights Litigation’ (2008) 11 *HarvLRev* 533; A Chilton and D Tingley, ‘Why the Study of International Law Needs Experiments’ 52 *ColumJTransnatlL* (2013) 173.

⁹⁰ As Manzi aptly notes, in 1998, 100, 000 people died in US hospitals due to drugs that had been approved by the Food and Drug Administration, presumably after extensive randomized trials, and which had been correctly administered. See Manzi (n 84) 68.

⁹¹ See eg M Frankel *et al.*, ‘Impact of Legal Counsel in Outcomes for Poor Tenants in New York City’s Housing Court: Results of a Randomized Experiment’ (2001) 35(2) *Law and Society Review* 419.

this is Angrist and Lavy's study of class sizes in Israeli schools.⁹² In the Israeli public education system there was a strict cap on classroom sizes at 40 students, meaning that if in a given year there was an enrolment of 41 or greater at a school, the students would be split into two classes—for instance of 20 and 21. If on the other hand there was an enrolment of 39, the students would remain in one class. Since it is plausible that abilities of students do not greatly vary on average, year on year, and it is plausible that a cohort of 41 students will have similar average ability to a cohort of 39, it is credible that measuring the academic achievements of classes of 20 versus classes of 39 will demonstrate the effect of class size on academic achievement. And, indeed, it seemed that students in smaller class sizes tended to perform better than those in larger ones. Since nowadays there is simply vastly more data available than there once was, discovering discontinuities and taking advantage of them to measure their effects is becoming more easily achieved.

This increasing use of experimental and quasi-experimental data has led some to claim that there is a 'credibility revolution' taking place in empirical economics and related fields⁹³—although it is important to note that there remains a strong level of scepticism.⁹⁴ Yet this same credibility revolution does not yet seem to have crept into the field of international law in general or international human rights monitoring in particular, where naïve statistical observations and regression analysis are typically unquestioningly treated as demonstrative of causality (if causality is addressed at all).⁹⁵ There is usually scant attention paid to issues such as the identification problem or omitted variables bias in the literature, and indeed correlations are very often presented as *prima facie* indicative of causation. This is most evident in the State reports, and indeed the UK's 2009 report to the CESCR is an illustration *par excellence* of this: a mirage of meretricious statistical observations provided so as to create a spurious sense of compliance. Yet it is also in general true of the academic work, which remains rooted in the 'age of regression', left behind by developments in other fields. And, indeed, there are persuasive reasons for arguing that, except perhaps in the very long term, there are no reasons to assume that a 'credibility revolution' can ever in fact

⁹² J Angrist and V Lavy, 'Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement' (1999) 114 *QJEcon* 533.

⁹³ J Angrist and J-S Pischke, 'The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con out of Econometrics' (2010) 24(2) *Journal of Economic Perspectives* 3.

⁹⁴ See eg E Leamer, 'Tantalus on the Road to Asymptopia' (2010) 24(2) *Journal of Economic Perspectives* 31; also CA Sims, 'Comment on Angrist and Pischke' (Princeton University Press 2010) (available at <<http://sims.princeton.edu/yftp/AngristPischkeJEP/AngristPischkeComment.pdf>>; and, for a somewhat older perspective, J Heckman and J Smith, 'Assessing the Case for Social Experiments' (1995) 9 *Journal of Economic Perspectives* 85.

⁹⁵ For an illustration, see Apodaca: '[T]he real problem in collecting economic and social data is political, not methodological or even conceptual.' See (n 66) 179.

take place in the arena of human rights monitoring. Let us now turn to addressing why this should be the case.

V. NO 'CREDIBILITY REVOLUTION' IN HUMAN RIGHTS MEASUREMENT?

There are two core reasons, or groups of reasons, for having severe doubts about the applicability of quasi-experimental techniques as a method of resolving the problems associated with quantitative human rights measurement. These are complexities arising from the continuing 'black box' nature of causality,⁹⁶ and the connected problem of good research design.

First, it is well acknowledged that even where robust results are generated by experimental or quasi-experimental techniques, the causal mechanism does not simply emerge by default. Very often, the results lead to murky conclusions, or no conclusions at all. A classic example of this problem is cited by Manzi, who describes a 2009 study which tested the effect of free primary medical care for a sample of 1,300 test patients versus a randomized control group in Ghana.⁹⁷ The results indicated that adult guardians of patients in the test group reported in diaries that they brought their children to more formal health care visits, but relied less on informal, traditional healers. Yet there was no statistically significant improvement in health outcomes for the test group versus the control group. How to interpret these results? Why did free primary medical care apparently not cause any improvement in health? Manzi lists four possible theories: the marginal value of increased health care spending has very little effect (a common observation made in developed economies); traditional healing remedies are undervalued (the test group used traditional healers less, so the results may indicate there is no difference between Western medicine and traditional health care methods in the area); standards of care in Ghanaian clinics are very poor (so attending a clinic has no or little value); and that parents lied when filling in diaries in order to demonstrate they were doing something socially reliable, but were not actually taking their children to formal health care visits in the frequency suggested (indicating free primary care was not a sufficient incentive to attend). There are undoubtedly more. The results, in other words, provide no basis for conclusions about the impact of free primary medical care, and no evidence on which to formulate health care policy or assign funding, without theoretical explanations—but there are competing theoretical explanations which are in large part dependent on pre-existing biases and which are all to some degree or other plausible.⁹⁸ Most tellingly, the results of the study do not even provide us with evidence about the most fundamental matter of all—whether spending on

⁹⁶ See eg Imai *et al.* (n 15).

⁹⁷ EK Ansah and S Narh-Bana *et al.*, 'Effect of Removing Direct Payment for Health Care on Utilisation and Health Outcomes in Ghanaian Children: A Randomised Controlled Trial' (2009) 6 (1) PLoS Medicine e1000007.

⁹⁸ See Manzi (n 84) 86–8.

health care has any impact on health outcomes or not. If anybody wished to assess, therefore, whether Ghana had enacted appropriate or effective measures regarding the right to health, to the maximum of its available resources, the results of this study would provide no resolution whatsoever.

Similarly, the Angrist and Lavy study relies on an understanding of the Israeli education system combined with a relatively straightforward and plausible theoretical proposition: that in a smaller class, individual students tend to receive more attention and hence perform better on average. Its results alone do not suggest a causal mechanism: causality must be theorized. When similar studies take place in other jurisdictions, where conditions are different, other results may appear which need to be theorized in turn. A similar project to Angrist and Lavy's took place in Chile almost a decade later; its different results were plausibly suggested by the authors as being due to the fact that in the Chilean school system wealthier parents have opportunities to send their children to schools which they know will have smaller class sizes—an 'enrolment manipulation' phenomenon which contaminates the findings.⁹⁹ But again, this observation came from familiarity with the Chilean school system itself, combined with a theoretical explanation—it did not simply emerge magically from the data.

What this suggests, of course, is that there is no substitute for substantive, deep and expert knowledge of the subject matter at hand—especially when it comes to interpreting data. Contrary to the claims of, for instance, the OHCHR that quantitative measurements provide objective, transparent and credible methods for monitoring human rights performance,¹⁰⁰ in actual fact it is typically the familiarity of the researcher with the subject matter at hand, combined with a plausible theoretical explanation of causation, which makes a statistically-based claim credible. In the absence of a persuasive *theoretical* causal explanation—an answer to the question, 'How?'—then an observation remains at best only a proposition about correlation.¹⁰¹

It also suggests that, as is well understood in the field of public health, in order for experimental or quasi-experimental techniques to provide robust evidence for cause and effect, there must be consistent, and repeated, replication in a variety of contexts. Otherwise results which may appear initially convincing could be due to environmental factors whose effects are not observed. Angrist and Lavy's study may allow credible, or at least plausible, inference of causality, in a narrow context, but a naïve conclusion drawn from it (small class sizes result in better academic achievement) may be limited to the social, cultural and temporal context in which it takes place. While the Chilean study in a sense supports the Angrist and Lavy study (it indicates that, intuitively at least, parents prefer their

⁹⁹ See M Urquiola and E Verhoogen, 'Class-Size Caps, Sorting, and the Regression-Discontinuity Design' (2009) 99(1) *American Economic Review* 179.

¹⁰⁰ OHCHR (n 24) 17.

¹⁰¹ R Mayntz, 'Mechanisms in the Analysis of Social Macro-Phenomena' (2004) 34 *Philosophy of the Social Sciences* 237, 241.

children to be in smaller classes—presumably because they ‘know’ it makes for better academic achievement), it may not always and everywhere be true. Different educational systems have different characteristics and different methods of teaching.¹⁰² Repeating the experiment in a variety of different contexts makes the conclusion more robust if similar results are discovered elsewhere. This is doubly necessary where there simply is no agreed theoretical explanation for the results, as in the case of free primary medical care in Ghana. Only widespread, consistent, repeated experimental or quasi-experimental results which seem to indicate persistent correlations between a policy measure and a certain effect will prove to be credible.

What this means is that, even if human rights scholars and the UN system were to move away from naïve statistical tools, they would be unlikely to receive any benefits from this putative ‘credibility revolution’ except perhaps on an ad hoc basis. It is an extremely complex task to identify causal mechanisms in a credible fashion in fairly narrow contexts—let alone across a scope as large as that of an international human rights treaty. And this in turn means that using the results of experimental or quasi-experimental studies as bases for measuring human rights performance is fiendishly difficult when considered in detail.

To continue with the class-size example, the notion that smaller class sizes tend to result in better academic achievement, all else being equal, may have been plausibly demonstrated to be true in the Israeli education system at least. Yet this does not make, for instance, ‘average primary school class size’ a suitable indicator of performance against the right to education: in a jurisdiction such as Chile, such an indicator would not capture the fact that small class sizes could primarily be composed of students from wealthier backgrounds. It would therefore not suggest a great deal about protection of the right to education; wealthier students tending to end up with a better educational experience is not, one would suggest, of interest regarding that particular right. It is also, naturally, contingent on teacher quality, which can be assumed to be relatively high in Israel, but much less so in other environments.¹⁰³ If such difficulties of conceptualization can occur with such a relatively straightforward-seeming measure, one can imagine the complexities surrounding the measurement of Ghana’s performance regarding the right to health if the apparently obvious-seeming ‘availability of free primary medical care’ was selected as a measure or indicator. Put simply, nobody knows whether making free primary medical care available in Ghana improves health outcomes for children—at least based on the available study.

¹⁰² And, indeed, extensive literature reviews have suggested that there is no evidence that class size has a systematic effect on student achievement. See E Hanushek, ‘The Failure of Input-Based Schooling Policies’ (2003) 113 *The Economic Journal* 64.

¹⁰³ See Pritchett (n 67).

But perhaps above all, this fundamental complexity mitigates against accurate statistical human rights measurement because of the expense in time and monetary resources necessary to generate robust and reliable results on which to base it—especially given that reliable results require extensive and widespread replication. The treaty bodies have limited time to investigate the statistics and studies laid before them by States Parties and NGOs (and indeed, generally do not currently see this as their role) and human rights scholars with the necessary training and skills to *critique* the research design of others are few and far between. States Parties clearly do not have strong incentives to fund or conduct robust human-rights-specific research. The idea that appropriate and effective measures for the protection of human rights could be guided by extensive use of experimental and quasi-experimental techniques is therefore simply not realistic in the short or medium term.

The prospect of statistical measures and econometric tools revolutionizing the practice of human rights monitoring, then, is a mirage. Yet it is not merely a harmless illusion, for two important reasons.

The first of these is straightforward: there is an opportunity cost, in time and other resources, associated with the move towards statistical analysis. Time spent running regressions is time lost investigating human rights violations, promoting human rights, better theorizing or conceptualizing human rights, or engaging in deep study of social phenomena. This may seem a trite observation, but it is one which is not sufficiently frequently made.

The second of these is more pernicious. As has already been alluded to, States Parties to human rights treaties have every incentive to make it appear as though they are in compliance with their treaty obligations, and the more that human rights performance becomes seen as quantitatively measurable, the more States will rely on statistical ‘evidence’ to demonstrate improved performance. Yet, as this article has sought to demonstrate, and as social scientists are increasingly willing to acknowledge, statistical ‘evidence’ in the social sphere is often bogus (usually being comprised of correlations without a credible causal explanation), and this has two particularly dangerous consequences for human rights monitoring. On the one hand, reliance on statistical measures allows States to game the system by using apparently neutral and objective-seeming veils of numbers to demonstrate compliance—a particular problem where, as in the UN treaty system, States Parties are encouraged to develop their own sets of indicators and cite their own statistics. It hardly needs explaining why this might result in the undesirable situation that States Parties simply select the measures that appear to show improvement, irrespective of cause. The intellectual dishonesty of the UK’s State representative to the CESCR claiming credit for a larger number of men taking paternity leave is a typical example of this. As well as having little to do with the actual performance of States Parties, such manipulation hardly contributes to a sense that human rights monitoring is a legitimate and robust exercise. The fact that ‘good governance’ and, by extension, evidence of good human rights

performance is so frequently a stated or implied consideration of donors regarding the provision of aid clearly also has the potential to affect the incentives of developing States engaged in that process.¹⁰⁴ And on the other hand, the fundamentally contingent and complex nature of attempting to ascribe causality makes it fairly straightforward to undermine or dispute statistical measurements, on the basis of omitted or intermixed variables, or for other flaws in research design. This makes it simple for States Parties to simply explain away measurements which appear to demonstrate lack of compliance. In other words, naïve use of statistical measurement makes it easy for States Parties to muddy the waters of the reporting procedures, whether by using statistics to ‘buffer away’ close monitoring, or by exploiting the contingent nature of statistical measurement of performance to undermine the monitoring process entirely.¹⁰⁵ The ‘manufacturing of uncertainty’ is hardly unknown in the field of regulation, and it would be naïve to expect that matters should be different in that of human rights monitoring.¹⁰⁶

VI. CONCLUSION: REFOCUSING ON THE INDIVIDUAL?

What are the lessons, then, for human rights monitoring? First, there must be a stronger emphasis placed on good fieldwork, and on the expert fieldworker. If developments in the social sciences in recent decades have taught us anything, it is that even the most robust, well-designed and widely replicated studies do not generate meaningful results without an appropriate interpretation from an expert or experts with deep knowledge of the subject at hand. Simply put, there is no substitute for embedded local knowledge giving a plausible theory about causality. The reason why, for instance, free primary medical care in a region of Ghana appears to have no effect on health outcomes—opaque to those engaged in carrying out the experiment—may be clear to the fieldworker whose familiarity with the social context permits them to give plausible interpretations of the results. This means that, contrary to the implied rejections of ‘subjective’ or narrative expert assessments present in much of the work on statistical human rights measurement,¹⁰⁷ the reality is that if statistical measurement of human rights performance is to be attempted, then typically *only* experts with (subjectively generated) explanatory theories can

¹⁰⁴ The EU, for instance, provides ‘budget support’ as an ‘implicit recognition that the partner country’s overall policy stance and political governance is on track’. See European Commission, ‘The Future Approach to EU Budget Support to Third Countries’ COM(2011) 638 Final (13/10/2011) para 2.1.1.

¹⁰⁵ See M Power, *The Audit Society* (OUP 1997).

¹⁰⁶ See D Michaels, *Doubt is Their Product: How Industry’s Assault on Science Threatens Your Health* (OUP 2008); and D Michaels and C Monforton, ‘Scientific Evidence in the Regulatory System: Manufacturing Uncertainty and the Demise of the Formal Regulatory System’ (2005) 13 *JL&Pol’y* 17–42.

¹⁰⁷ See eg G de Beco, ‘Human Rights Indicators for Assessing State Compliance with International Human Rights’ (2008) 77 *Nordic Journal of International Law* 23.

offer plausible interpretations of the results. What this also means is that apparently ‘judgement-proof’ methods such as human rights indicators and statistical measures, which might appear to allow objective measurement which bypasses the need for time-consuming and unreliable subjective expert judgement, in reality offer very little.

The second lesson is that there is a need for a renewed focus on individual human rights violations, rather than outcomes. There is a temptation to conclude that, since quantitative measurement of human rights performance is concrete and objective, the alternative is for human rights monitoring to simply descend into a morass of subjective and hence opaque and unreliable judgement-making based on narrative accounts.¹⁰⁸ Yet there is no need for this to be the case: in fact, since purportedly ‘objective’ quantitative measurement is itself so unreliable, a retreat from it may have the effect of making human rights monitoring more robust. As long ago as 1996 Chapman was making the observation that, given the difficulties of statistical measurement of economic and social rights performance, it was both more practical and more moral to concentrate on individual violations rather than to pursue the quixotic goal of monitoring ‘progressive realisation’ (or what may be thought of as the modern ‘outcomes’ approach¹⁰⁹). Despite 20 years having passed since the article was published, most of Chapman’s comments regarding measurement of development, as we have seen, remain true: it is ‘unrealistic and impossible to handle’¹¹⁰ due to the difficulties and costs of analysing the available data. At that time the treaty bodies were still relying on physical records with almost no computerization, of course, but as this article has sought to demonstrate, the problems run much deeper than a mere lack of computational speed—and Chapman’s conclusions remain trenchant.

The first of these conclusions was that since identification of violations was much more straightforward than assessing performance through the use of statistics, it was simply a more effective method for evaluation. Chapman herself eventually retreated from this position¹¹¹ and, indeed, what came to be known as ‘the violations approach’ ultimately resulted in a perpetuation of many of the problems identified in this article: a focus on statistical measures and a naïve understanding of causality. The Maastricht Guidelines, which stemmed from the original article, assume, for instance, that it is possible to tell what ‘appropriate steps’ are, and seek to make the failure to develop and

¹⁰⁸ It is common to encounter statements such as that of J Welling, in ‘International Indicators and Economic, Social and Cultural Rights’ (2008) 30 HRC 933 at 958, that without statistical indicators the international community would be ‘uninformed’ about economic, social and cultural rights performance.

¹⁰⁹ A Chapman, ‘A “Violations Approach” for Monitoring the International Covenant on Economic, Social and Cultural Rights’ (1996) 18 HumRtsQ 23. ¹¹⁰ *ibid* 34.

¹¹¹ See eg AR Chapman, ‘The Status of Efforts to Monitor Economic, Social and Cultural Rights’ in Hertel and Minkler (n 22) 143.

apply human rights indicators a violation in and of itself.¹¹² However, the original core argument—which is, in essence, that one should focus on what is possible to know, rather than what is impossible to know—is persuasive. Establishing, in particular, whether an individual’s rights have been violated in a specific instance is something which courts do as a matter of routine—violations can be defined and identified, if not simply, then in a fashion which is well practised and understood.¹¹³ The reader will of course be familiar with the manner in which courts, both international and domestic, achieve this. And, while they are not courts, the UN treaty bodies are able to perform a quasi-judicial function in assessing whether a violation has taken place, and currently, of course, do so through the (albeit under-resourced) individual communications procedures. Different treaty bodies have, for instance, found violations where a State failed to exercise due diligence in preventing a woman from being killed by her estranged husband,¹¹⁴ where a State ordered its civil servants not to reply to written or oral communication in a minority language;¹¹⁵ and where a State failed to prosecute a perpetrator of hate speech.¹¹⁶ And, similarly, NGOs, activists, academics and practitioners can engage relatively straightforwardly in identifying instances of what may amount to individual violations. To put the matter somewhat crassly, monitoring a State Party’s performance under the CEDAW vis-à-vis discriminatory violence cannot be done through simply counting the number of incidents and checking whether it is rising or falling, because changes in that statistic cannot be attributed to a set of policies, nor ‘appropriate measures’ identified, due to the problems of causality already outlined. But if a woman is murdered by her estranged husband because the police fail to exercise due diligence, and if this is proved, then a breach of an international obligation has clearly taken place and a remedy must be provided. It is clear which of these techniques is more reliable and useful.

However, this focus on individual violations is not only to be recommended for its conceptual clarity. Chapman was also at pains to stake out a moral claim for its importance: as she put it, ‘the goal of any approach to human rights is to enhance the enjoyment of rights of individual subjects and to bring them some form of redress when the [*sic*] rights are violated, not to abstractly assess the degree to which a government has improved its level of development on a range of statistical indicators’.¹¹⁷ In other words, human rights law is a ‘tangible’ domain.¹¹⁸ It concerns individual people who find themselves at

¹¹² See The Maastricht Guidelines on Violations of Economic, Social and Cultural Rights (1997) paras 14 and 15.

¹¹³ Chapman (n 109) 38.

¹¹⁴ *Goekce v Austria*, CEDAW, Communication No 5/2005, UN Doc CEDAW/C/39/D/5/2005.

¹¹⁵ *Diergaardt et al v Namibia*, HRC, Communication No 760/1997, UN Doc CCPR/C/69/D/760/1996.

¹¹⁶ *The Jewish Community of Oslo et al v Norway*, CERD, Communication No 30/2003, UN Doc CERD/C/67/D/30/2003.

¹¹⁷ Chapman (n 109) 38. ¹¹⁸ *ibid.*

the whim of the oppressive State, and it attempts to provide them with a remedy when they suffer at its hands. In abstracting human rights to the realm of data and econometric technique—in subsuming individual human interests into aggregated statistical measures such as ‘the literacy rate’—the moral importance of the individual and his or her right to education, with all that it brings, becomes lost or ignored. And this, correspondingly, removes moral responsibility from the State: the language of outcomes is the language of management and of technical expertise (how best to improve measured performance); the language of violations is appropriately accusatory and shaming—a weapon.¹¹⁹ While there are compelling practical and theoretical reasons for avoiding econometric approaches to the monitoring of human rights, then, there are also important moral reasons which should not be ignored.

This also has resource implications that must be acknowledged. A consistent theme in this article has been availability of resources. On the one hand, academics are focusing more time, energy and financial resources on the development of statistical tools for measuring human rights performance. On the other, there is a lack of resources available for the treaty bodies to engage in quasi-judicial activities and in the kind of fact-finding necessary to identify violations. It may be suggested that what little resources are available, be they temporal or financial, could be more productively spent by improving the individual communications procedures and widening knowledge about them, and by improving the fact-finding capacities of the treaty bodies when engaging in analysing State Reports, than by directing those rewards towards fruitless attempts at quantification.

Statistical measurement does have its uses in the field of human rights. It is, of course, important to use statistics to identify problems. For instance, it is undoubtedly useful to know, from a public policy perspective, that the labour rate amongst women in a given ethnic group is much lower than the national average, or that poor white boys perform worst in school.¹²⁰ But there are extremely good practical and theoretical reasons for avoiding the use of statistics and statistical techniques in the assessment of human rights performance, or compliance with human rights treaty obligations. In summary, these reasons are as follows. First, and most importantly, statistical measurement alone simply provides correlations, at best, and correlations do not amount to plausible demonstration of causality and hence do not permit analysis of the appropriateness or effectiveness of policy. This makes statistical measurement unsuitable, on its face, for establishing whether States are acting appropriately or effectively to protect the rights of individuals in their jurisdictions. Second, over-reliance on statistics is a boon for States Parties to

¹¹⁹ *ibid.*

¹²⁰ See CEDAW, The UK’s Response to the List of Issues for Its Fifth and Sixth Periodic Reports (2008) UN Doc CEDAW/C/UK/Q/6/Add.1, para 51; Equality and Human Rights Commission, ‘Is Britain Fairer?’ (2015) available at <<http://www.equalityhumanrights.com/about-us/our-work/key-projects/britain-fairer-0>>.

human rights treaties, because it easily allows them to produce bogus ‘evidence’ of improved performance based on meretricious ‘objective’-seeming data, which the treaty bodies have little time or inclination to critically analyse—and in turn to problematize evidence of compliance gaps. And third, blitheness about the complexities of human rights protection undoubtedly has a serious and large opportunity cost, as academics, practitioners and activists focus their attentions on the production of statistical measurements and econometric analyses, and correspondingly neglect other—possibly more effective—approaches. The monitoring of human rights has become increasingly quantitative, and all trends indicate that it is likely to become more so. Yet it would behove those engaged in the process to consider developments outside of the field and ask whether, in fact, that trend is leading towards a cul-de-sac from which economists and other social scientists have retreated.