Performance Measurement in Global Governance: Ranking and the Politics of Variability

Afshin Mehrpouya,
HEC Paris, Department of Accounting and Management Control
mehrpouya@hec.fr

Rita Samiolo
University of Innsbruck School of Management, Department of Organisation and Learning
Rita.Samiolo@uibk.ac.at

Abstract

The past thirty years have witnessed the spread of rankings, ratings and league tables as governance technologies which aim to regulate the provision of public goods by means of market pressures. This paper examines the process of company analysis underlying the production of a ranking known as the Access to Medicine Index. We conceptualize the Index as a "regulatory ranking" with the explicit mission of addressing a perceived regulatory gap and market failure: the lack of access to medicine in the Global South. The Index, which ranks the world's largest pharmaceutical companies with regards to their access to medicine policies and practices, aspires to help address the problem of access to medicine through stakeholder consultation, transparency and competition. This study unbundles the epistemic work underlying the performance measurement process leading to the creation of the Index. We trace how the goal of stakeholder consensus, the need to project objectivity and the aspiration to govern through competition shape analysts' epistemic work. We discuss how through notions such as "the good distribution" and "aspirational indicators", performance measurement and ranking become entangled in a "politics of variability" whereby company data need to be variably interpreted in order to optimise the possibilities of intervening in companies through competitive pressures, while at the same time complying with the imperatives to remain in the space of perceived stakeholder consensus and to provide a faithful representation of companies performance to inform public debates. We reflect on the challenges posed by these analysis processes for the regulatory aspirations of the ranking.

Keywords: ranking, epistemic work, professional vision, commensuration, performance measurement, regulatory capitalism.

https://doi.org/10.1016/j.aos.2016.09.001

Our manufactures now doctors sell,
And their intrinsic value meanly tell;
Nay, they discover, their spite is such
That health, than crowns more valued, costs not much;
While we must shape our conduct by these rules
To cheat as tradesmen, or starve as fools.

Garth – A London-based dispenser of medicines, February 1667[1]

Introduction

Over the past three decades, non-state regulation has grown faster than state regulation. The sources, forms, subjects and objects of regulation have multiplied – a process visible at the transnational level, but similarly traceable within national boundaries (Braithwaite, 2008; Djelic & Sahlin-Andersson, 2006; Djelic & Sahlin-Andersson, 2009; Jordana, Levi-Faur, & Marín, 2011; Levi-Faur & Jordana, 2005). "Soft" laws, best practices, expert standards of quality and performance, audits and different forms of market-based pressures including ranking, ratings, certifications and league tables are supplementing and partly replacing "hard" laws or government rules, and are increasingly being used by states themselves (Djelic & Sahlin-Andersson, 2006; Djelic & Sahlin, 2009; Mörth, 2004).

In this hybrid space of voluntary or quasi-voluntary rule-making that mobilizes market ideals in the name of public interest, various organizations devoted to intermediation between corporate actors, public authorities and concerned publics have appeared, forming "regulatory constellations that transcend the state/non state divide" (Djelic & Sahlin, 2009, p. 178). Ranging from inter-governmental organizations to multi-stakeholder initiatives and self-regulatory bodies (Fransen & Kolk, 2007), various organizations are involved in the regulation of public goods and issues that are increasingly problematized at the global level. Such issues include global warming (Kim & Lyon, 2011), farmers' conditions (Blowfield & Dolan, 2010), forestry (Bartley, 2003), mining and corruption (Wang & Rosenau, 2001), banking's social and environmental impacts[2], food security and sustainability (Reinecke, Manning, & Von Hagen, 2012), forced labour and slavery[3], malnutrition[4], labour conditions (Fransen, 2011) and access to medicine – the subject of the present study.

[1] Extracted from Bell (1843, p. 14).

[2] For example, see BankTrack (http://www.banktrack.org).

[3] For example, see the Global Slavery Index (http://www.globalslaveryindex.org).

Performance information (indicators, rankings, ratings, certifications, social labels, and the like) is mobilized to exercise competitive and reputational pressures on regulatory targets by establishing normative orders (Boli, 2006). Such performance information is increasingly produced in private or semi-private organizations dispersed throughout the polity – so-called "civil regulators" (Vogel, 2008).

In this study we draw attention to the role of rankings as regulatory instruments in the global governance space by focusing on one particular tool: the Access to Medicine Index (hereafter referred to as the Index). The Index, developed by a Dutch not-for-profit organization named the Access to Medicine Foundation (hereafter referred to as the Foundation), is comprised of a set of performance measures defined through stakeholder consultation. Using about eighty indicators, it aims to capture various aspects of the impact of pharmaceutical companies on access to medicine in the so-called Global South. The Index is issued on a bi-annual basis, and ranks the twenty largest research-based pharmaceutical companies across a set of countries and diseases. It seeks to encourage these firms to increase their efforts to provide access to needed medicines by means of the competitive pressure created through ranking. The Index, now in its fifth iteration, has become a central tool in the global health community for setting the expectations from, and measuring the performance of, the pharmaceutical companies in areas related to access to medicine.

In what follows, we show the compromises that Index analysts must make in attempting to produce performance information that occupies a neutral apolitical space based on the projected stakeholder consensus – information that "objectively" represents the pharmaceutical companies and that differentiates between them to motivate competition and benchmarking around the issue of access to medicine. The outcome is a particular form of authorized[5] public knowledge with the dual intended roles of enabling competition among firms (market-making) and facilitating firms' accountability to their imagined "stakeholders" (claim-making).

[4] For example, see the Access to Nutrition Index (www.accesstonutritionindex.org).

[5] The term authorization here refers to social processes which make a knowledge form fit for use in regulation and governance (Jasanoff, 2011; 2012). Authorization underscores the fragile network of associations and perceptions that equip knowledge with governing authority and help associate knowledge with different forms of politics (Shaw, 2004).

The epistemic work underlying the production of performance information with regulatory or quasi-regulatory purposes, increasingly prevalent in global governance (Davis, Kingsbury, & Merry, 2012; Rottenburg, Merry, Park & Mugler, 2015), has until now remained mostly a black box. Even within business organizations, performance measurement has been studied more in terms on its organisational consequences than in terms of its underlying knowledge processes and epistemic preconditions (with exceptions such as Dambrin & Robson, 2011; Chenhall, Hall & Smith, 2013; Qu & Cooper, 2011). These studies have examined performance measurement as sociomaterial processes in which textual and graphic inscriptions (Bloomfield & Vurdubakis, 1997; Busco & Quattrone, 2015; Chua, 1995; Dambrin & Robson, 2011; Ezzamel, 2004; Qu & Cooper, 2011) but also particular professional identities (Dambrin & Robson, 2011) and modes of evaluation (Chenhall, Hall & Smith, 2013) help explain the institutionalization (or lack thereof) of specific performance measurement approaches. Less attention has been devoted to the process of measurement and commensuration itself – how performance as an object of knowledge is observed and "learnt" in order to be turned into numbers for the purposes of comparison and benchmarking.

Extant research on rankings, too, has tended to focus on the consumption of rankings and their mutually constitutive relations with the objects they aim to represent (Elsbach & Kramer, 1996; Espeland & Sauder, 2007; Jeacle & Carter, 2011; Kornberger & Carter, 2010; Scott & Orlikowski, 2012), while the production side of rankings is mostly left unexplored (with exceptions such as Pollock & D'Adderio, 2012). In addition, so far, studies of "regulatory rankings", rankings mobilized to fulfil an explicit regulatory agenda vis-à-vis their objects by mobilising competitive processes, have been scarce.

Here, we attend to the epistemic work underlying one such ranking, showing how the particular regulatory mission of this ranking and its position in the polity condition the calculative practices underlying its production. In particular, we illustrate how this ranking's aspiration to regulate an industry by inciting competition around a social goal requires the analysts producing the ranking to develop a specific professional vision (Goodwin, 1994), a particular way of seeing meaningful events in company data and learning to commensurate them.

This vision, caught between the regulative ideals of mechanical objectivity and trained judgment (Daston & Galison, 2007), is geared towards capturing a "good distribution" of

scores in order to maximise the possibility of enticing competition among the regulatory targets. Rather than treating competition as the inevitable consequence of rankings, we examine and question it as a programmatic ambition (Rose & Miller, 1992) that permeates the process of ranking and directs the eye of the ranker. The possibility of competition, we argue, lies first and foremost in the specific professional vision of the ranker, trained to recognise variation in company performance and to find ways of producing and maintaining such variation by "upgrading" and "defining" (Lynch, 1985; 1988) differences at the margins. The eye of the ranker is also trained to discern mobility or immobility in companies' performance in order to establish indicators as either relative or absolute markers of performance. We discuss the "politics of variability" at play in such vision, where minimum, average and maximum scores are at once a terrain of contestation by different stakeholders and the fundamental landmarks around which a possible space of competition for improving access to medicine is made to unfold. While the possibility of market competition among the ranked organisations is actively promoted, sustained and magnified by the performance variability established in the process of ranking, agonism in the polity – conflict among stakeholders – is tamed and downplayed by seeking to measure only within the space of perceived consensus, somewhere in the middle of a polarised stakeholder base. We illustrate how analysts' calculative practices can absorb, marginalize or resurface conflicts in this contested field of global governance – in other words, their political and anti-political potentials (Didier & Tasset, 2013; Barry, 2002; see also Ferguson, 1990).

The remainder of the paper is organized as follows. First, we position this study within the literature on transnational governance and the sociology of rankings and quantification. We then describe our data sources and methods. In the empirical section, we trace the practices of performance measurement, scoring and ranking at play in the construction of the Index, and the professional vision and epistemic values driving the work of the analysts involved. In the final section, we discuss how our analysis of the Index contributes to our understanding of the use of rankings as forms of soft regulation in transnational governance, and the role of commensuration therein.

Regulatory Ranking

Levi-Faur (2005) identifies a new set of relations between capitalism and regulation – what he terms "regulatory capitalism"[6]. He emphasises how the hallmark of contemporary governance regimes is not deregulation of markets, rather it is an extensive process of re-regulation of all aspects of human societies through markets (see also Braithwaite, 2008). Regulating through markets may at times involve the establishment of fully-fledged new markets based on commoditization and exchange, such as carbon markets (Callon, 2009; Mackenzie, 2009), but can also simply mean mobilizing foundational market processes such as competition and benchmarking in order to move organisations in the direction of some variously defined collective goal (see Çalişkan and Callon, 2010 for a definition of market)[7]. Complementing or replacing traditional state regulation, regulatory capitalism is characterised by processes of privatization (*inter alia*, of regulation itself), marketization (the mobilization and organization of competitive pressures around the public goods being regulated) (Djelic, 2006)[8], and scientization (the increase in influence and authority of science and expertise) (Drori & Meyer, 2006).

Within regulatory capitalism the "privatization of the public" has its flipside in the process of "publicization of the private", whereby "public law values like transparency" (Braithwaite, 2008, p. 7) – and, one can add, pluralism and deliberation (Djelic & Sahlin, 2006) – discursively permeate the corporate world and also structure the way in which regulation is organised. Rituals of stakeholder consultation, deliberation, inclusion and consensus building are now indispensable to transnational private regulators in order to attain democratic legitimacy to govern on behalf of the "global public" (Quack, 2010). In the absence of democratic legitimacy of the nation state, and of clear boundaries around the idea of the

[6] According to Levi-Fair, regulatory capitalism is characterised by (2005, p. 27): "(1) a new division of labor between state and society (for example, privatization), (2) an increase in delegation (remaking the boundaries between the experts and the politicians), (3) proliferation of new technologies of regulation, (4) formalization of interinstitutional and intrainstitutional relations and the proliferation of mechanisms of self-regulation in the shadow of the state, and (5) the growth in the influence of experts in general and of international networks of experts in particular".

[7] As noted by Davies (2014, p. 21): "the key institution of neoliberalism is not a market as such, but particular market-based (or market-derived) forms of economization, calculation, measurement and valuation".

[8] In this study we do not analyze whether the competition that the Index aims to induce is achieved or not. The ranked companies seem to have responded in different ways to the competitive pressure the Index seeks to produce – some engaging in more detailed reporting, others engaging in better understanding of the methodology to find effective ways to improve their rank (for example by hiring former Foundation executives), some have introduced initiatives aligned with the Index methodology, and some others have mostly ignored the Index. In this paper we have chosen not to focus on these processes of reactivity (Espeland & Sauder, 2007). Rather, essential to our work is examining how the aspiration to induce competition around the delivery of access to medicine shapes the professional vision of analysts and their epistemic work. We aim to bring to the fore the contrived and programmatic nature of the competition rankings seek to harness.

global public, "civil regulators" (Vogel, 2008)[9] attain legitimacy by showing responsiveness to a variety of stakeholders and by associating themselves with legitimate national and international organizations in their respective sectors (Gulbrandsen, 2008).

In this paper, we turn to the knowledge production processes on which regulatory capitalism, with its multiplicity of sources, tools and forms of regulation, rests. More specifically, we refer here to the use of performance metrics and ranking as part of the attempt to regulate issues considered of public interest by means of competition and benchmarking and through the involvement of a variety of "stakeholders". By stressing the link between knowledge production and regulatory capitalism we aim to shed light on the regulatory or quasi-regulatory purposes of the performance information produced in this context. As illustrated below, these regulatory purposes direct the vision of the ranker, shaping the knowledge practices through which company performance emerges as an object of knowledge in the process of ranking.

While all rankings and other types of performance information in the public sphere have the potential to influence the entities they represent and the broader polity through a range of intended and unintended processes of "reactivity" (Espeland & Sauder, 2007; 2015), what differentiates regulatory rankings is that they are structured around an explicit prescriptive agenda in areas where a regulatory gap is identified[10]. In other words, they have a manifest program to influence the behaviour of the entities they represent in the name of the public interest. They do so by setting prescriptive norms and standards of behaviour in the form of performance indicators whose authority is derived through stakeholder consultation rituals. Such indicators exploit the ambiguity between description and prescription, and between

[9] Civil regulators are defined as private governance organizations that attempt to regulate the social impact of business (Vogel, 2008). This literature distinguishes between first-person regulation or "self-regulation", second-person regulation and regulation attempts by adversaries and counterparts of corporations, and third-person regulation – i.e. regulation by an "independent" third party organization. The Access to Medicine Index falls under the third heading. Nonetheless, the type of work undertaken in producing the Index and the epistemic tensions driving such work may not be confined to this type of private governance organization. The aspiration to harness market forces to regulatory goals (through calculative technologies such as scores, heat maps, league tables, and the like) and the need to maintain some sort of democratic legitimacy and scientific authority can be regarded as essential features of contemporary private governance broadly defined (Dingwerth & Pattberg, 2009; Gulbrandsen, 2008).

[10] According to Levi-Faur "regulation is the promulgation of prescriptive rules as well as the monitoring and enforcement of these rules by social, business, and political actors on other social, business, and political actors." (Levi-Faur, 2010, p. 8). In this definition the term prescriptive is crucial to our analysis. The explicitness of a prescriptive agenda has significant implications for the ways in which rankings are constructed, which is the focus of our analysis of what we define as "regulatory rankings".

representation and intervention, to the extreme (Strathern, 1997; Hacking, 1983). As in the case of other rankings – such as the Magic Quadrant in the Information Technology (IT) industry (Pollock & D'Adderio, 2012) – the interventionist agenda of regulatory rankings, the need to make the ranking "useful" to various decision-makers and thus to move the market, directs the eye of the ranker. However, regulatory rankings differ in that they need to take into account and balance the decision needs and expectations of different stakeholders – they seek to move the market in order to ultimately intervene in the polity. As such, they not only need to operate with a certain conception of the market (as movable by the ranking), but need to engage with particular notions of the polity, too (as a field of clashing interests to be balanced in the act of measurement). In what follows, we illustrate how the idea of moving the market and of balancing conflicting stakeholder views animate performance measurement and shape its calculative practices.

The knowledge work of ranking

Research in the sociology of quantification and accounting is showing increasing interest in the "flood of social measures designed to evaluate the performances of individuals and organizations" (Espeland & Sauder, 2007, p. 1), and a host of studies on governance technologies such as rankings and ratings have recently appeared (Elsbach & Kramer, 1996; Espeland & Sauder, 2007; Free, Salterio, & Shearer, 2009; Jeacle & Carter, 2011; Kornberger & Carter, 2010; Pollock & D'Adderio, 2012; Sauder & Espeland, 2009; Scott & Orlikowski, 2011). Much of this scarce but growing literature tends to focus on the impact of performance measurement tools on those being evaluated and made "transparent" or on users of such technologies of transparency. However, little attention has been paid to the processes of knowledge production and validation these technologies entail or to the epistemologies they mobilize and deploy.

Espeland and Sauder's seminal work has examined the subtle ways in which rankings discipline individuals and transform organizations and inter-organizational fields by means of reactivity mechanisms (Espeland & Sauder, 2007; Sauder & Espeland, 2009). The authors illustrate how rankings can operate as self-fulfilling prophecies that change expectations, inviting behaviour which makes the ranked organizations become more like their representation in the ranking. Furthermore, rankings can change sense-making, altering "the locus and form of attention, both creating and obscuring relations among entities." (2007, p.

16). Commensuration – the process of transforming qualities into quantities by means of a common metric (Espeland and Stevens, 1998) – is at the core of how rankings act on the ranked objects and other users. As Espeland and Stevens highlight, commensuration simplifies information and creates more abstract and decontextualized knowledge. It erases differences not amenable to quantification, and restructures comparisons along the lines of what numbers reveal. It unites different objects under the banner of a common measure, but separates them along the ranks of the numerical order it generates. Commensuration also provokes reflection on "the ontological status of numbers in relation to the ontological status of the things they measure", where different attitudes to the realism of numbers may emerge (Espeland & Sauder, 2007, p. 21; Desrosières, 2001).

Commensuration, crucially, is also what is required to produce a ranking in the first place. In this study, we draw attention to commensuration not as the outcome of a ranking, but as its precondition. We examine not so much the impact of commensuration on the objects and users of rankings, but the process of analysis and the kind of professional vision (Goodwin, 1994) required from the subjects of rankings – in our case, company analysts – in order to achieve the commensuration that has been proven to be so consequential. Following Goodwin (1994, p. 606, emphasis in the original), companies' access to medicine initiatives as an object of knowledge amenable to be rated and ranked can be understood as emerging "through the interplay between a *domain of scrutiny"* (in our case, answers to a questionnaire and other mostly textual data sources) "and a set of *discursive practices"* (selecting from company data elements amenable to comparison, applying the coding scheme embedded in the scoring guidelines) "being deployed within a *specific activity*" (the scoring of companies)[11]. We highlight the discursive nature of the process of commensuration, its programmatic character (Rose & Miller, 1992), one permeated by the idea (embedded in the notion of "the good distribution") that companies can be moved along the ranks of the performance spectrum in the direction of a collective goal by means of benchmarking. We trace how analysts learn to commensurate companies by developing a specific way of seeing and interpreting company data which maximises the possibilities of benchmarking. We show the epistemic work and values at play when analysts simplify, decontextualize, establish equivalences and differences, and reflect on the correspondence between the scores they

[11] "[T]he ability to see a meaningful event is not a transparent, psychological process but instead a socially situated activity accomplished through the deployment of a range of historically constituted discursive practices" (Goodwin, 1994, p. 606).

assign and the reality these seek to capture. We characterize their attitude to the reality they research as one in which different notions of objectivity are, sometimes uneasily, combined.

One notable exception to the lack of research on the production side of rankings is Pollock's and D'Adderio's (2012) study of the specific socio-material aspects of the Magic Quadrant, a ranking device produced by analysts within the information technology industry. The authors illustrate how the specific "format" and "furniture" of the ranking contribute to shaping the competitive space within which the ranked firms move. Pollock and D'Adderio also examine the scoring practices of analysts and the pressures, constraints and normative principles informing the back-stage production of the ranking. To our knowledge, their work is one of the few that opens the black box of the scoring practices behind the production of a ranking. Our study is centred on similar issues. However, unlike Pollock and D'Adderio's work, which focuses on the market-making role of the ranking within a particular industry, our research on the Access to Medicine Index brings to the fore broader governance issues. Market-making and competition, as mobilized by the Index, are used both to regulate a public good such as health and to mediate among stakeholders with competing values, norms and agendas (Buthe, 2010; Büthe & Mattli, 2011). As a result, analysts are under pressure to produce knowledge that is acceptable to all stakeholders, seen as objective and free of political and scientific bias, while at the same time able to motivate companies to compete to improve access to medicine.

The escalating deployment of market mechanisms for regulatory purposes makes the investigation of knowledge-making at the intersection of markets and polities timely and relevant. Attending to such knowledge work is crucial for understanding how prevalent regulatory approaches based on the above-mentioned calculative instruments gain authority and stability, and thus for grasping their conditions of possibility. It also allows examining the distinctive features of different regulatory instruments as "bearers of values, fuelled by one interpretation of the social and by precise notions of the regulation envisaged" (Lascoumes and Le Galès, 2007, p. 4).

Referring to more traditional state-driven regulatory processes, Jasanoff (2011, p. 307) noted that "[e]ven the field of science and technology studies (STS), which takes as its problematic all aspects of knowledge production and use, has devoted less energy to studying knowledge-making in policy environments than in laboratories or other scientific workplaces". It may be argued that this is even truer of knowledge-making in the domain of private, voluntary

regulation, especially at transnational level, where the fast expansion of this domain represents a defining feature of regulatory capitalism (Levi-Faur, 2005). According to Jasanoff (2011, p. 307) "norms, including norms underpinning all public knowledge, are constituted at the global scale, through practices that are rarely exposed to critical scrutiny from citizens or social scientists".

Jasanoff calls for an investigation of the practices of objectivity underlying the creation of "regulatory science" (Jasanoff, 1995, 2011). According to Jasanoff (2011, p. 316):

> Objectivity in policy necessarily confronts and accommodates cultural conceptions of what counts as proper representation not only of nature but also of public interests. This double objectivity, scientific and political, is achieved through institutionalized practices whose tacit epistemological implications remain largely unrecognized by the participants in public knowledge-making.

In the German context, Jasanoff observes a prevalence of the Habermasian model of the public sphere centred on deliberative processes. In this context, "epistemic authority comes from the inclusion of all legitimate points of view", a "view from everywhere" where (unlike its British variation)[12] consensus among experts is seen as valid if all standpoints are included (p. 314). She contrasts this with the American style of objectivity, where knowledge relevant to policy must be presented as a "view from nowhere" (Nagel, 1989), "aperspectival knowledge" (Daston, 1992) relying on quantitative analysis and models. This is usually associated with "mechanical objectivity" (Porter, 1996), an understanding of objectivity that makes the latter depend on the perceived impersonality and replicability of decision rules based on numbers.

The notion of mechanical objectivity has been used also by Daston & Galison (2007) to examine the role of mechanical means of representation (e.g. photography, X-rays) in the production of scientific atlases across a variety of disciplines, in order "to capture nature with as little human intervention as possible" (2007, p. 20).[13] The authors discuss this notion of

[12] British policy culture is characterized as a pragmatic "common witnessing of demonstrable facts", usually achieved by entrusting advisory commission to elite figures whose authority however depends on their ability to "articulate a plain, common-sense vision" which anyone can attest to (Jasanoff, 2011, p. 314).

[13] This version of mechanical objectivity, unlike Porter's, is at odds with the "view from nowhere", as the individual details and idiosyncrasies of the specimen represented by mechanical means are automatically

objectivity alongside others which appeared and overlapped in the history of science, such as truth-to-nature, structural objectivity, and trained judgment. The latter is particularly relevant for the purposes of this paper, to unpack the skills analysts were required to develop in order to score and rank companies. Trained judgment, unlike the rigid impersonal protocols of mechanical objectivity, entails the ability to interpret through intuition and to enhance data in order to highlight patterns. It demands an "intuitive expert, who depends on unconscious judgment to organize experience into patterns in the very act of perception", as opposed to "the indefatigable worker" of mechanical objectivity, "whose strong will turns inward on itself to subdue the self into a passively registering machine" (Daston & Galison, 2007, p. 44).

In what follows, we show that the production of the Access to Medicine Index entailed different notions of objectivity (mechanical objectivity and trained judgment) and different styles of knowledge validation (containing elements of the view from nowhere and the view from everywhere). These different notions of objectivity, seen as regulative ideals for knowledge making, can be regarded as different modes of commensuration; they point to variations in the scientific ethos and scientific self underlying commensuration and ranking exercises. Despite being such fateful and pervasive social processes, commensuration and ranking have yet to be unpacked in their epistemic preconditions, and in how they act not only upon their objects, but their subjects, too. Our study of the Access to Medicine Index brings to the fore such issues. It draws attention to the "epistemologies of the eye" (Daston & Galison, 2007) of the ranker as the key to understanding how the market-making potential of rankings, and thus their deployment as instruments of soft regulation, is discursively and practically enacted in the process of commensuration.

Note on Sources and Methods

Empirical setting

The Access to Medicine Index is a multi-stakeholder ranking which attempts to enable dialogue among different constituencies – investors, civil society, academia, international organizations and governments – around highly contentious issues such as patents, research and the pricing of medicines in poor countries and communities. It uses the outcome of this dialogue to establish the Index methodology, revised every two years and used to measure for

transferred to the image. This is a problem for those seeking aperspectival knowledge, as in the "view from nowhere" the specific and the idiosyncratic are regarded as a particular "somewhere" to be erased (Daston & Galison, 2007, p. 123).

each company the level of fulfilment of stakeholders' expectations. The key inscription produced – a ranking of the twenty largest pharmaceutical companies – is then widely circulated in the media (Figure 1)[14].

Please insert Figure 1 about here.

The Index is produced by the Access to Medicine Foundation, a not-for-profit organization established in 2005 in Haarlem. It is financed by the Bill and Melinda Gates Foundation, the British Department for International Development (DFID) and the Dutch Ministry of Foreign Affairs. The 2010, 2012 and 2014 editions of the Index have been endorsed by the World Health Organization (WHO), DFID and other major global health actors. The Index has thus become a key source of standards and measures for the Access to Medicine performance of pharmaceutical companies (Hogerzeil, 2013).[15] The Index is now in its fifth iteration.

Several steps are involved in the bi-annual development and release of the Index: 1) conducting an online stakeholder survey to ascertain views on its scope, indicators and output; 2) holding "stakeholder roundtables" in Europe, North America and one Index Country (the latter was introduced in Index 2010); 3) organizing Expert Review Committee (ERC) meetings to discuss outstanding issues and to ratify methodology updates (this committee, which was introduced in Index 2010, includes one representative per stakeholder group); 4) finalising the updated methodology, followed by the public release of the

[14] Each Index report includes different sections for consumption by different possible users. The final ranking and an executive summary are meant for the broader "public", company report cards which highlight areas of "leading" behaviour and "areas for improvement" are largely for consumption by companies and investors, and detailed technical area chapters are to be used especially by companies, experts and academia.

[15] There are several indications that companies and other important constituencies are reacting to the Index, though in different ways (a report on the Index's impact was recently published to substantiate this: see Technopolis, 2016). Some companies publicly disclose access to medicine related information based on Index criteria in their annual report. Several others issue press releases to announce their position in the Index. Constituencies such as the WHO and DFID provide links to the Index on their websites, while prestigious reports on access to medicine such as the "MDG Gap Task Force" report on the United Nations Millennium Development Goals refer to the Index (UN, 2011). Index data have been made available to investors through Bloomberg terminals since 2014 (ATMF, 2014). Furthermore, the Index has been used by both companies and NGOs in recent contentious cases related to access. In addition, several recently created Regulatory Rankings such as the Access to Nutrition Index (https://www.accesstonutrition.org), the Responsible Mining Index (http://responsibleminingindex.org), or the Access to Seeds Index (http://www.accesstoseeds.org) utilize the Access to Medicine Index as their template. Taken together, these examples suggest that the Index is having diverse intended and unintended effects both on companies and across the polity.

Stakeholder and Methodology Report; 5) collecting data, analysing and scoring companies; 6) launching the Index report. Since 2012, four technical sub-committees (TSCs), comprising mostly of "independent experts and academics", have provided further advice to the Index team on more detailed technical issues, with the aim of separating the political deliberation role envisaged for the ERC from the independent and neutral technical advisory role of the TSCs.

The Methodology and Stakeholder Report defines the disease, geographical and company scope of the Index, and illustrates the indicators used (29 for Index 2008, 107 for Index 2010, 102 for Index 2012, 95 for Index 2014, 83 for Index 2016) and their weights. The scope of the Index is a key area of contention and is debated within the ERC based on the analyst team's proposals. The International Classification of Diseases[16] and the Global Burden of Disease[17] databases of the WHO are used to set the disease scope (diseases for which there is an access to medicine need), the United Nations Human Development Index[18] and the World Bank country income classifications[19] are relied upon to set the country scope, while Thomson Reuters databases help set company scope, covering those companies whose largest revenues are sourced from pharmaceutical activities. Key contentions in the process of "territorializing" (Mennicken & Miller, 2012) and setting the boundaries around access to medicine include primarily companies pushing for coverage of countries or diseases where they have more activities. The Index's measurement territory is also influenced by changes in the problematisation of access to medicine introduced by legitimate and central organizations such as the WHO. For example, the WHO's increasing focus on the rise of non-communicable diseases in poor countries[20] led, from Index 2012 onwards, to an expanded coverage of cancer and several other such diseases.

The indicators are organized under seven Technical Areas (listed below) each weighted for importance and analysed across four Strategic Pillars (Commitments, Transparency, Performance, Innovation):

- General Access to Medicine Management

[16] See http://www.who.int/classifications/icd/en.

[17] See http://www.who.int/topics/global_burden_of_disease/en.

[18] See http://hdr.undp.org/en/content/human-development-index-hdi.

[19] See http://data.worldbank.org/news/2015-country-classifications.

[20] See http://www.who.int/chp/ncd_global_status_report/en/.

- Public Policy and Market Influence
- Research & Development
- Equitable Pricing, Manufacturing and Distribution
- Patents & Licensing
- Capability Advancement in Product Development and Distribution
- Product Donations and Philanthropic Activities.

Performance indicators – predominantly based on qualitative information – are developed within each Technical Area (performance is assessed on a scale of zero to five depending on the answer to a specific qualitative question). Certain indicators are based on quantitative information (e.g. a performance indicator in Research & Development: "Portion of financial R&D investments dedicated to Index diseases out of the company's total R&D expenditures"). Such indicators are also evaluated on a scale of zero to five. Each Technical Area, Strategic Pillar and indicator is assigned a weight based on a combination of the perceived contribution of each to access to medicine and also the perceived "quality of indicators"[21]. The ERC discusses Technical Area and Strategic Pillar weights, while indicator level weights are left to the discretion of the analyst team, who are since Index 2012 advised by the TSCs. Any indicator whose data quality is considered "insufficient and unreliable" for the majority of companies (ATMI, 2010a, p. 13) is converted into an "experimental indicator" with a zero weight.[22] A questionnaire to collect data from the pharmaceutical companies is then devised. A team of analysts verifies and analyses data from diverse sources including those received from the companies, scoring them accordingly.

Up until 2012, the Foundation comprised of its founder Wim Leereveld and at different times one to two administrative employees, and it outsourced all analysis work. All research for Index 2008 was conducted by a project manager and two analysts of the sub-contractor (Innovest Strategic Value Advisors – a Canadian Environmental, Social and Governance (ESG) research firm). For Index 2010 a team of 4-5 analysts organized by the subcontractor (MSCI Inc. which acquired RiskMetric in 2009 which had in turn acquired Innovest in 2008) conducted company analysis and scoring, a senior analyst and a project manager conducted methodology adjustments, quality checks, coordination and managing the relationship with

[21] The quality of an indicator is defined by analysts as based on the perceived quality and accessibility of the information needed to score that indicator, and on the comparability of this information across companies.

[22] Indicators and their weights are available from Index reports at www.accesstomedicineindex.org.

the Foundation and stakeholders. Each analyst was in charge of analysing about five companies. This included managing the relationship with those companies, collecting data about them and analysing it, quantifying the data and producing a company score for each indicator. For Index 2012 a similar team at MSCI conducted stakeholder consultation, methodology update, data collection and analysis, whereas a team of two external consultants at the Foundation undertook the writing of the report. For Index 2014, methodology update, data collection and analysis were outsourced to another ESG research firm named Sustainalytics, while additional analysis and the writing of the report were undertaken by the Foundation's in-house team of four analysts and research director. As of August 2016, the Foundation has fifteen employees working on different rankings, including the Access to Medicine Index but also the Access to Generics and Access to Vaccines Indices. The team includes five analysts and two research program managers. Since October 2015 the Founder, Wim Leereveld has retired from being the Chief Executive Officer and has joined the Foundation's supervisory board.

Data sources

This study draws upon diverse sources of qualitative material covering the development of the Access to Medicine Index from 2005 to 2015. Data sources include documentary evidence, participant observation and interviews. Documentary evidence includes press releases, news and media articles, minutes of meetings and outputs from stakeholder consultations (online stakeholder survey results in 2009 and 2011, four stakeholder meetings in 2009 and 2010, one meeting with NGOs in Nairobi, Kenya in 2010, two stakeholder meetings, two meetings with the pharmaceutical industry and two meetings with investors in 2011). Public media coverage of the Index from 2007 to 2015 includes approximately 112 news articles plus 12 reports and academic commentaries in journals such as The Lancet, PLoS and The British Medical Journal.

A relatively high level of access was possible since one of the authors (Mehrpouya) was employed as a reviewer for the 2008 Index and subsequently as the Index Project Manager from November 2009 to June 2011 at the subcontracted firm, Innovest (later MSCI Group). In his role as project manager, he was responsible for organizing stakeholder consultation, methodology design, data collection, analysis and writing of the report for Index 2010 and the stakeholder consultation process for Index 2012. From June 2011 up until the launch of the

third iteration of the Index in November 2012, he was a member of the editorial board and steering committee for Index 2012. In this latter capacity, he advised the Index team on methodology and reviewed the project's documentary output. Since September 2012, when the current research project around the Index started, he has been conducting participant observation (Jorgensen, 1989), collecting documentary evidence and taking notes of his engagements with the Foundation.

We also conducted twenty-two interviews (average of sixty minutes) with a range of actors involved in the development and/or use of the Index. Interviewees included the Foundation's founder, two former CEOs, two pharmaceutical company representatives, two NGO representatives, a UK Department for International Development (DFID - major funder of the Foundation) program manager, a WHO director present in the ERC and Index stakeholder roundtables from Index 2008 onwards, chair of ERC for Index 2010 and 2012, two former managers of Innovest Strategic Value Advisors, the project managers for Index 2008, Index 2012 and Index 2014, and members of the Index 2008, 2010, 2012 and 2014 analyst teams (list of interviews provided in Appendix 1). In addition, earlier drafts of this paper have been discussed with the Foundation's founder and its current Executive Director/former Research Director.

The Access to Medicine Problem

As the poem in the epitaph illustrates, pricing medicine and reconciling the "intrinsic value" of health with the market value of pharmaceutical knowledge has been a contentious issue for centuries. The WHO played an important role in bringing this issue to the transnational space by introducing the "Essential Drugs List" (EDL) in 1977 (Greene, 2011). Defining a category of mostly off-patent medicines as "essential" to health and by extension as "a right" faced a dramatic backlash from the pharmaceutical companies. According to the latter, the distinction implied that other medicines were to be regarded as unessential, as mere market commodities (Siddiqi, 1995). At the WHO general assembly of 1986 the US government voiced concerns about the "alarming prospects of politicizing health" (Greene, 2011, p. 22), and ultimately decided to cut US funding to the WHO for two years. By 1988, when funding was re-established, the WHO had a new general director and the EDL program had been downgraded in the WHO hierarchy from general director's office to a sub-division of the pharmaceutical sector division.

The year 1994 marked a turning point for medicine in emerging markets, as negotiations at the World Trade Organization (WTO) led to the ratification of the Trade-Related aspects of Intellectual Property Rights (TRIPS). The TRIPS agreement placed obligations on all member countries to create regulatory platforms and enforcement mechanisms to protect Intellectual Property (IP) through patents. Consequently, pharmaceutical companies began paying increasing attention to emerging markets (T'Hoen et al., 2003). The agreement also meant that pharmaceutical products still under patent would remain under the exclusive control of international pharmaceutical firms, leading to significant price increases and the decreased supply of those products to emerging market countries (Grace, 2004). This was concurrent with the emergence of HIV/AIDS as a key global health concern, especially in poor countries.

The introduction of TRIPS and its impact on access to HIV/AIDS drugs, in particular, prompted a rapid rise in civil society activism, spearheaded by the "Campaign for Access to Essential Medicines" (Greene, 2011). The United Nations has been a key source for formulating the Access to Medicine "need". The introduction of the United Nations Millennium Development Goals (MDG) in 2000 has been a crucial milestone in defining the access to medicine problem setting out a role for the pharmaceutical companies to help address this "gap" (MDG, 2000)[23]. A transnational Access to Medicine community (Djelic & Quack, 2010) has since emerged, gathering hundreds of NGOs and experts whose aim is to devise or use a constellation of norms, standards, knowledge forms and compliance pressures to push the pharmaceutical industry to improve access to medicine.

While consensus about the Access to Medicine need and its scale has been reached, contention about its causes and possible solutions continues. Pharmaceutical companies highlight their role in developing new medicines and in providing access to them in poor countries, while shifting the blame to ineffective healthcare systems and weak national regulations. In contrast, NGOs, generics firms and Southern governments frequently blame "Big Pharma" for its patenting practices and anti-competitive behaviour (Morris, 2010). Besides patents and generics competition concerns, a number of other issues have become increasingly central to the access to medicine debates. Such issues include: the pharmaceutical industry's role in researching "neglected diseases" which occur almost

[23] MDG Target 8.E: "In cooperation with pharmaceutical companies, provide access to affordable essential drugs in developing countries" (http://www.un.org/millenniumgoals/global.shtml, accessed August 2016).

exclusively in poor countries (where pharmaceutical markets are frequently unviable), their role in providing affordable prices, and their marketing and lobbying ethics in low and medium-income countries (Morris, 2010). The Access to Medicine Index was launched within this field of broken dialogue, contention and litigation.

A ranking to restore the market mechanism

Wim Leereveld, the Foundation's founder, had previously worked as an executive at a company that sold market data to the pharmaceutical sector, a profession that introduced him to the "Big Pharma" world. Leereveld believed that ranking offered a potent mechanism for changing access-related practices, because of the competitive pressure it would put on pharmaceutical companies (Interview – Founder and former CEO of the Foundation):

> Companies are interested in what you think, or what I think, or what governments think. But they're mostly interested in what their peers think […]. They compete every day.

The Index tries to push companies to introduce innovative business models to bring markets to areas of perceived weak regulation and market failure by motivating competition around the access to medicine issue (Interview – Founder and former CEO of the Foundation):

> What we do with this index is make a list of competitiveness and a market, because the market isn't working for access to medicine. Today, the Index means we restore the market mechanism.

The Foundation treads carefully in defining the territory of its measurement, trying to avoid overlaps with areas where markets and/or effective state regulation are present. Thus, in order to be considered within the Index's measurement territory, companies' initiatives have first of all to go beyond compliance with national regulatory requirements (excluding countries where regulatory enforcement is considered weak – such as those of Sub-Saharan Africa). Second, they have to be in areas where there are no established markets. For example, most research efforts for non-communicable diseases such as diabetes and cancer are not counted in the scoring process, because lucrative Western markets are seen as the primary motives for such research. Only research initiatives targeting the specific needs of Index countries are

considered. Among such initiatives considered beyond companies' regulatory mandate and within areas of market failure, those that appear purely charitable are valued, but those that could lead to new market development (e.g. new business models tailored to the specificities of Index countries) are much more valued. In other words, if the social case is also backed up by a business case then the initiative in question can be deemed sustainable[24]. Third, the potential for impact is assessed. If an initiative has some potential for impact on access, but such impact is not a clear company objective, it is normally given a low score. If evidence can be found that the potential for impact is translated into an explicit target by the company, a higher score is assigned, with top scores in those cases where the achievement of the target can be proved. Pushing for "sustainable" and "innovative business models" (ATMF, 2015, p. 31) basically means promoting new market development as a solution to the issue of access to medicine – while emphasizing the substantiation of social impact.

In the early days of the Index it became apparent that the legitimacy of the Index and the agenda it attempted to set for the industry depended on broadening its stakeholder base, in pursuit of the "balanced views" on access to medicine invoked by the United Nations. In Leereveld's words (Interview - Founder of the Foundation and its former CEO):

> […] you should have a report for the whole world because of course we all want to agree on these kind of things and I knew the United Nations […] were interested in balanced stakeholder views on Big Pharma.

From the outset, the Index was projected as representing stakeholder consensus, seen as the precondition to lead the pharmaceutical industry in the direction of greater access to medicine (Leereveld, 2014):

> When I first started work on the Access to Medicine Index, I knew that pharmaceutical companies were willing to do more, but that they didn't know which path to follow. And that was because there was no consensus among the various stakeholders on what the companies should be doing or where they should focus. So we knew that the first

[24] This is how sustainability is defined: "[t]he pharmaceutical industry license to operate is based on development and delivery of affordable, accessible and high-quality pharmaceutical products. To ensure long-term sustainability, the pharmaceutical industry must balance economic objectives with fulfilling its role in society. In this context, it is important to note that an active and innovative approach to the Index Country markets is also backed by a more immediate economic rationale" (ATMIa, 2010, p. 49).

> step would be to build this consensus, by talking to NGOs, the World Health Organisation, patient groups, investors and to the companies themselves.

The Index, in short, is based on the assumption of a market "movable" towards greater access to medicine (companies "willing to do more") but only once the polity is re-ordered by building consensus about the direction for such movement. The polity is conceived as a polarised space to be balanced, and in that balancing act performance measurement would become possible: "we can only measure what's part of the consensus, but if we have no consensus we have no sense" (Interview - Founder and former CEO of the Foundation).

However, in practice, views on issues such as patents, pricing and competition with generics varied widely amongst ERC members and stakeholders. According to the Index 2010 and 2012 Chair of the ERC:

> Whether these issues are resolvable – and I don't think they are …it becomes more a question of […] managing the inherent tension rather than trying to resolve it by reaching consensus on an issue – Personally, I don't think this is possible.

If stakeholders failed to reach consensus, decisions were shifted from the ERC to the Foundation, and by extension, to the analyst team. The team then updated the methodology in accordance with the stakeholder input received, attempting to represent the "common view", and presented it back to the ERC.

The "balance" of views captured by the Index, especially in its early days, was often in question. In a letter to *The Lancet*, the head of Health Action International (HAI), a key international NGO in driving the Access to Medicine debates, observed (Reed, 2008, p. 890):

> The weakness of the study lies in its "on-paper" approach to measuring the effectiveness of corporate social responsibility programmes […]. The index is based on information made available by pharmaceutical companies and a few non-governmental organisations, with the largest input coming from surveys of the pharmaceutical industry itself.

The letter further emphasizes that (Reed, 2008, p. 890):

> The aspiration to produce a quantifiable study of corporate social responsibility standards that promotes access to medicines is laudable. However, measuring access to essential medicines is best achieved by looking at both pharmaceutical policy and the view from the ground. Without accompanying data on the effectiveness of pharmaceutical corporate social responsibility policy, the value of the index is, at best, debatable. Having the rules is not the same as playing by the rules. What we need is a comprehensive analysis of industry performance on access to essential medicines that rewards impact rather than rhetoric and promotes health outcomes rather than investment incentives.

This juxtaposing of health outcomes with investment incentives challenges the principle that the former can be achieved primarily through the latter – that health can be regulated via market mechanisms. Yet this critical voice did seem to agree on the desirability and possibility to measure companies' impact on access to medicine. Faith in quantification united views and actors otherwise radically divided, though some NGOs were more ready to credit information collected "on the ground" and inclusive of local views, whereas other constituencies such the WHO and academics tended to primarily value globally comparable, replicable and thus more often quantitative information.

For example, at a panel organized by the Index team at one of the largest global access-related summits named "The International Conference for Improving Use of Medicines" (ICIUM), one panel member, Goran Tomson, Professor of International Health Systems Research at Karolinska Institute, argued (Burril Report, 2011):

> The Index's methodology cannot be reproduced, hence it cannot be considered statistically valid.

Similarly, the WHO representative in the ERC described his first meeting with Wim Leereveld in 2006 as centred on the need to construct a solid scientific base for the Index and its performance indicators, which should feature, *inter alia*, "replicability and reliability" (Interview - WHO member of the ERC). Scientificity was crucial in order to be perceived as neutral and to avoid being accused of political agenda or "opinion" (Interview - CEO of the Foundation 2009-2010).

> I am the type of person who also likes to express an opinion about whether a company is doing right or good. But it was really a board decision not to express any opinion about whatsoever company. Really, [we were] data driven, facts driven.

To sum up, the goals of promoting competition among pharmaceutical companies, offering a balanced view of access to medicine able to mediate among stakeholders, and producing a scientifically solid quantitative representation of the industry drove the production of the Index. The following section shows how these pressures conditioned the process of scoring and ranking.

Epistemic work underlying the construction of the Access to Medicine Index

The majority of analysts involved in data collection, analysis and scoring involved in all iterations of the Index are women, between 24 and 30 years old, educated in Europe, US, Canada and, in the case of Index 2012 only, India (when data collection was conducted at a low-cost site of MSCI Group in India). They ranged from recent graduates to more senior analysts with several years of experience in Environmental, Social and Governance (ESG) analysis or public health (see Appendix 1). Despite differences in terms of geographical location and educational, national, professional and personal backgrounds, analysts were confronted with common pressures dictated by the particular position of the Access to Medicine Foundation in the polity and the programmatic ambitions of the Index. These analysts had to produce knowledge aiming to intervene in a highly contentious field and to influence large multinational corporations whilst remaining legitimate in the eyes of publics such as funders, global health authorities such as the WHO, and active and outspoken NGOs in the field of global health and development.

Analysing as a "robot" and a "stakeholder collector" – measuring in the middle

Analysts were encouraged by the Foundation to consider themselves neutral stakeholder collectors: "you are a stakeholder collector and you are a person: so what can you do about that?" (Interview - Foundation's Research Director since 2013 and Executive Director since late 2015). Within the Foundation the mediation between different views and the taming of analysts' political selves have become an explicit organizational value, sought through rituals

such as dedicated lunch meetings and discussion sessions (Interview - Foundation's Research Director since 2013 and Executive Director since late 2015):

> We have heated discussions upstairs. It's almost like a debate where you have to argue both sides in order to settle on something in the middle, and then when you're in the middle you say "okay, what do we actually want to measure here?"

These mediation efforts aimed at remaining equidistant in a polarized space, are described in terms of taming one's emotions (Interview - Foundation's Research Director since 2013 and Executive Director since late 2015):

> being objective is very difficult especially in the space we are [in], where pharma is the bad guy, and NGO is the good guy, and I come from a world of seeing the best and the worst of people. So the first thing I had to say [was] "okay, put everybody's emotions outside of the office when you come in here, that's what you are: you are a robot and you have to find a way to develop it [the Index] as a robot. And of course that's not possible, no one could live like that.

This effacement of one's emotions – attempting to act as a robot – is presented as a "test run" to promote self-reflection and self-awareness (Interview - Foundation's Research Director since 2013 and Executive Director since late 2015):

> so that was the initial test run, and then people started reflecting: "what am I leaving outside, what are all the things I need to put outside in order to develop the Index"? So all these things we put outside. Then I said: "bring them in", because those are the things that [are] going to drive you and drive the Index, those are the most valuable parts of it. If you are just a robot it has no meaning whatsoever. So we are objective but we have to use all those things that have made us who we are, that's our skill set, our expertise and we need to bring those two together to talk about it.

Analysts' self-awareness entailed reflexivity about any positive or negative bias towards companies (Interview - Index 2010 analyst III):

> I expect that the most difficult version [of bias] to eliminate would be the nature of the relationship to the company. For example, the woman at [*company name withheld*] was quite a difficult character, so it's the personal relationships between the analyst and the data provider within the company. […] [M]aybe on a subconscious level, however professional and "by the book" you try to be, potentially the nature of the relationship … this may be most difficult to eliminate through scoring.

Analysts combined data obtained directly from the companies via the questionnaire with information derived from other publicly-available sources such as Factiva or Lexis Nexis databases, government data, information published by organizations such as the WHO, the United Nations Development Program (UNDP), the World Bank (WB), the Organization for Economic Cooperation and Development (OECD) and NGOs (see Appendix 2). For analysts to feel comfortable with their information sources, the latter had to be framed as data speaking for themselves, independent of the analyst standpoint. Being "close to the data" was seen as way to leave one's opinions and feelings outside the analysis (Interview - Index 2010 analyst I):

> I can only speak for myself, but I looked at every single source, especially if it was a contentious issue. […] You do that enough and you get a fuller picture of what's going on. The key thing for me was presenting it in a manner that was legally accurate and neutral. Whenever you are so close to the data, you can't really sort out your personal opinions or how you feel about something. It's about what the data is telling you.

The aspiration to multiply information sources reflected the need to collect views from all stakeholders, especially when these were polarized. Analysts seeking to act as stakeholder collectors were pursuing a "view from everywhere".

Analysts considered quantified information to be more scientific (Interview – Index 2008 Project Manager):

> My way was clearly not particularly scientific as it was mostly based on qualitative [research methods]. I did use some indicators that had been validated through the methodology process, but it was very qualitative compared to what [Index 2010] did,

> which was a much [more], you know, robust process. Clearly the data I had was not very strong, so it meant lots of qualitative assessment.

Yet quantitative information was often the most difficult to work with. Indicators based on quantitative information (e.g. kilograms of active ingredients, volume of sales, number of molecules in the pipeline, amount of investments etc.) were problematic because of differences in measurement units and the lack of comparability across companies. However, as specified by one analyst (Interview – Index 2010 analyst III):

> I think we just made a very implicit assumption that quantitative data were more frequently associated with better legitimacy of claims or statements, or were more dependable.

Collective discussions, sharing views, comparing companies and contrasting analyses were seen as ways to root out possible biases. This amounted to an exhaustive and (for some) nerve-racking process of quality assurance (Interview - Index 2010 analyst I):

> We did so many quality checks and had so much collaboration between team members. There was one point when [*analyst name withheld*] and I were on the phone for ten hours. […]. You have to continuously assess the data, again and again, with different people, thinking about it from different angles. You would occasionally change your mind about something because more data would come in, or someone would make a really good point and you'd think, "ok, that actually makes sense". We really tried to get to the point where people from our team were on the same page about everything that we did, and there was constant email communication on the technical areas.

These attempts to get everyone "on the same page" sought to make the scoring of companies as impersonal and replicable as possible. Mechanical objectivity was being pursued. Indeed, the idea to leave the scoring of companies to software rather than to analysts had been explored by the Foundation in the early days of the Index (Interview – Foundation CEO 2009-2010). Furthermore, at least some of the analysts perceived their job as quite automatic: "data is data, we scored along the scoring guidelines" (Interview - Index 2012 Analyst Team Head). Most analysts however were comfortable with the idea that judgment was required to interpret the scoring guidelines, and that scoring entailed a notion of analysts as something more than

"data monkeys" – a sarcastic expression used by some analysts in the 2010 team to claim the right to contribute more to the definition of the scoring guidelines, rather than apply them passively.

Most analysts accepted that intuition could have a role to play in data analysis. The "truth" of company performance was perceived as something analysts develop a "feel for" (Interview - Index 2012 Analyst)[25]:

> You start to develop a feeling, or a sense, about what's going on. We try to find out more, so hopefully, when we look at the ranking, it shouldn't come as a shock […]. It should flow with the sense you got from your qualitative research and the final quantitative scoring. That's what, for me, feels good.

As stated by another analyst (Interview - Index 2010 analyst IV):

> As the spreadsheets were finalized and the scoring completed, I remember that we all had a final look and were given the green light to move things around if, by this point [...] we felt that [something] wasn't correct. Especially for the middle section of the Index compared to each extreme. [...] We'd been working on this data for a long time and it was almost as if we had a gut feeling. It wasn't a gut feeling, really, it was... you know, you've been working on it for a long time and you can tell if it feels like the truth or not.

The persona of the analyst moved between a certain idea of mechanical objectivity – the analyst as a robot and a passive stakeholder collector – and that of trained judgment – the analyst as endowed with the power of interpretation and intuition.

[25] While several analysts used the term "feeling" to describe the process of getting comfortable with the results of their research, this process was not particularly emotionally charged. "Feeling the truth" was associated with the ability to anticipate the results of one's analysis without the need to re-trace all its steps, as knowledge of company performance progressively sedimented in the mind of analysts in ways that were both conscious and unconscious. In this respect, "feeling" the truth of company performance can be equated with having an intuition about it.

Scoring as market-making – differentiating at the margins

What the data collected "revealed" about companies was not self-evident and did not automatically match the criteria set in the scoring guidelines[26]. Analysts had to collectively develop a specific "professional vision", i.e. the "ability to see meaningful events" (Goodwin, 1994, p. 606) in the data obtained from companies, so as to be able to commensurate and rank the latter on a scale of zero to five. A variety of different policies, initiatives or other "behaviour" had to be assessed in terms whether they matched the scoring guidelines in a similar way and to the same extent. Rather than the "blind sight" that would be at play if scoring guidelines were self-evident and perfectly matching the data collected, analysts had to adopt a kind of "physiognomic sight" (Daston and Galison, 2007, p. 314) to identify similarities and differences across companies' initiatives, recognize patterns in behaviour and cluster companies at different ranks.

The process was a largely comparative one where each company's performance was defined in relation to the performance of everyone else, rather than against absolute benchmarks (ATMI, 2010a, p. 52):

> Note that the Access to Medicine Index is a relative index. A score of five on a scale of zero to five for the majority of the indicators simply signifies the leading practice among the companies under coverage. The Index does not evaluate companies against aspirational best practices. Instead, it enables a comparison of current ATM practices of the researched pharmaceutical companies.

Scoring required analysts to identify a "normal" or average level of performance, distinguishing between under-performers and top-performers in comparison with the average, usually associated with a score of 2.5. This score corresponded to the most frequent level of performance predicted and/or observed. In this way, a baseline (i.e. what most companies were expected to be doing or appeared to be doing with respect to the specific issues captured by each indicator) had to be agreed upon. This was then assigned a score of 2.5 out of five and considered "the norm". As far as possible, a score of five (the highest) and a score of zero

[26] The relevant dimensions of access captured by the Index are largely selected in consultation with ERC members during the methodology development and review process, and are coded in the set of indicators which make up the Index. Scoring guidelines are then agreed in order to drive the process of scoring companies based on their answers to the questionnaire.

(the lowest) were each assigned to at least one company, to ensure that the Index captured the whole range of relevant performance displayed within the industry. The aim was to "showcase those companies that were best performers and the top performers and then those companies that really weren't doing much" (Interview – Index 2010 analyst I). The assumption was that good practice should be rewarded, so as to create a "race to the top".

A distribution of companies' performance would thus emerge for each indicator. The judgment of analysts consisted in the ability to define and recognize average behaviour, to distribute companies around the mean, and to "read" statistical patterns in the often confused and idiosyncratic company data[27]. This reading, crucially, was a "flexible" one. The vision of analysts could be deepened so as to adjust the distribution. The scoring guidelines were redefined and made more detailed whenever scores appeared too close. Analysts adjusted indicators and scoring guidelines to exclude information that was considered not comparable across companies and to maximize the extraction of information value – the "variability" of companies' performance. This variability was crucial to the definition of a "good distribution" (Interview – Index 2010 analyst I):

> We had to continuously adjust our scoring indicators to make sure we were catching the full range of company activities and to ensure we had a good distribution for that indicator. … Some companies were really close but then you wanted to make sure… if they didn't fit into a certain group, we wanted to make sure they were either higher or lower.

Each indicator score was, to borrow Lynch's words, a "geometrical space" – a five-tier scale – into which companies were mapped. The exercise of judgment in applying the scoring guidelines made companies' behaviour "progressively less recalcitrant to the textual devices of describing, comparing, casually accounting, mapping, measuring" (Lynch 1988, p. 216). More specifically, for companies to be "squeezed" into the same rank, differences with other companies at the same score level had to be "filtered" (Lynch 1988, p. 209). For example, "commitments" indicator C.I.3 for Index 2010 was defined as: "The company commits to make available for free the products in the countries where the clinical trials for those

[27] See Daston and Galison (2007, p. 370) for a discussion of pattern detection through trained judgment.

products were carried out, consistent with codes such as the Helsinki Code for Clinical Trials". The scoring guideline for indicator C.I.3 had initially been formulated as follows:

> 5- The company commits to the Declaration of Helsinki for international drug trials
> 2.5- The company has an alternative set of internal or global codes in this area.
> 0- No commitments in this area

It was only after data collection and in the process of analysis that the scoring guidelines could be updated as follows:

> 5- Company has a specific, detailed approach to post-trial access – for trials conducted in Index Countries – which assures access to patient benefits in a large variety of different circumstances.
> 3.5- Company has a specific approach to post-trial access – for trials conducted in Index Countries – which does not assure access to patient benefits in all likely circumstances.
> 1.5- Company does not specifically refer to post-trial access other than through its stated compliance with the Declaration of Helsinki OR no evidence of any clinical trial conduct in Index Countries (Neutral).
> 0- No commitments.

As most companies did mention their commitment to the Declaration of Helsinki or similar codes, the original scoring guideline would have led to a skewed distribution where most companies would be clustered at the high end of the performance spectrum – not a "good" distribution. Performance was then further segmented and specified across dimensions emerging as relevant for comparison from the filtering of data. Variability in companies' performance, and thus the information value and market-making potential of the Index, emerged as a function of the analysts' ability to map and distribute companies in the five-tier scale, and to adjust their depth of vision whenever companies could not be sufficiently differentiated.

The contours of post-trial access to medicines, and the specific terms thereof, emerged only after data were received and compared. Other data concerning firms' policies on clinical

trials, including ambiguous wording that could not be used for comparison (such as the mention of a "compassionate use program" which evoked the possibility of post-trial access but without any specific detail or other clarification) was filtered out. The filtering happened in two stages. First, information deemed relevant to the indicator in question was selected by analysts, who produced a short interpretive account of the answer received from companies and information obtained from other sources, through which the score given would be justified in the light of the scoring guidelines. Second, the textual elements that could be compared (often colour coded or highlighted by formatting key sentences in bold and adding exclamation marks) were assigned a score[28]. This filtering process, driven by the imperatives of comparability and commensuration, did not amount to mere simplification or loss of information. It was, similarly to what Lynch observed for the visual documentation of objects in the life sciences, transformative (1988, p. 205). Filtering was essential to the (re)definition of each indicator, whose meaning became clear only after the data were observed and compared. In the case of indicator C.I.3, the specific terms of post-trial access that came to be seen as the key differentiating factors between companies below the average and those above it, and thus the meaning of the indicator, emerged only in the process of filtering.

Once assigned a score and placed on the same numerical scale, different companies' initiatives and policies were thus made more "uniform" (Lynch, 1988), comparable by virtue of being textually represented next to each other (with relevant parts of the text highlighted and thus more easily matched across companies), and distinguishable by uniform degrees of separation in the discrete quantitative order of the zero-five scale.

For companies to be ranked at different levels, the differences identified through filtering had to be "upgraded" and "defined" (Lynch 1988, p. 209). As mentioned, this was done by making the scoring guidelines for an indicator more and more detailed, in order to separate companies initially appearing too similar and thus obtain a good distribution. Once a score was settled, the "definitiveness" with which the scoring guidelines were documented made any "cognitive and perceptual uncertainties" disappear (see Goodwin, 1994, p. 609 who made this point with respect to the work of archaeologists). This further contributed to "solidifying" the meaning of a score.

[28] See Goodwin (1994, pp. 609-611) for a discussion of "highlighting" as a way to make "the perceptual field […] enhanced in a work-relevant way by human action on it" – a process very similar to the one described by Lynch.

Analysts had to learn to "synthetize, highlight, and grasp relationships in ways that were not reducible to mechanical procedure" (Daston & Galison, 2007, p. 314). Their vision entailed a judgment of similarity which required an element of tacit knowledge and intuition, almost a "connoisseurship" – "the knowledge of innumerable different blends" of access relevant initiatives accumulated "from example to example" (Shapin, 2012, p. 179). Analysts had to learn to see and interpret company data in such a way that equivalences between companies' initiatives could be established (through filtering and uniforming), while differences had to be made sufficiently clear to justify the assignment of different scores and the positioning of companies in different ranks (through upgrading and defining). In this way, differences at the margin were amplified so as to create the space within which companies are expected to compete for their access to medicine performance. As found in other studies of rankings (Pollock and D'Adderio, 2012; Espeland and Sauder, 2007), producing an Access to Medicine Index depends on the possibility to meaningfully differentiate the participating organizations. Unlike the case of the Magic Quadrant (Pollock and D'Adderio, 2012), the number of companies in the Index is fixed and could not be limited at the discretion of the analyst. To maintain meaningful distinctions, analysts had to progressively "mine" their data through the process of "zooming in" allowed by the refinement of the scoring guidelines.

As in the case of indicator C.I.3 illustrated above, when most companies initially appeared clustered at the high end of the performance spectrum, refining the scoring guidelines meant making them more demanding. When most companies appeared clustered towards the bottom of the performance spectrum, refining the scoring guidelines meant trying to take the definition of a five closer to the best observed performance. In both cases, and especially for companies falling in the middle of the ranking, the differences eventually captured after refining the guidelines tended to become less tangible (Interview – Index 2010 analyst IV):

> at the extremes you could always, to some extent, know why you would give a five to [*company name withheld*] or why you would give a one to [*company name withheld*] or [*company name withheld*] or whoever came down. [At the] extremes it becomes quite obvious. […] If you go into the middle of the Index and [observe] the way we classify them, you'll see that the differences become very, very subtle.

The same analyst believed that redistributing companies at different score levels based on adjusted scoring criteria to maximize differentiation, and assigning a value of 2.5 to average performance and a value of five to the best-observed performance (in order to capture the whole range of observed performance and obtain a good distribution) clashed with ideals of accuracy and representational faithfulness – a reason for personal frustration (Interview - Index 2010 analyst IV):

> So we ended up taking the best case scenario among the data that we had, and giving [that company] the top score as an absolute best [...] you can imagine that we ended up giving quite high scores for…you know, [a company doing] half of the job.

This tension between faithful representation and intervention, between the aspiration to be objective and the desire to move companies by means of competitive pressures, confirms the broader concern expressed by Strathern (1997, p. 308) that "when a measure becomes a target, it ceases to be a good measure". The need to "intervene" in the market by setting the right incentives though the "good distribution" thus shaped the way of representing the same market via quantitative tools, bending representation in the direction of amplifying differences. Here, the mission of the analyst as an objective scientist striving for accuracy and representational faithfulness clashed with the mission of the analyst as a market-maker, who had to represent the market in such a way that competition was made possible and a "race to the top" incentivized.

Similar tensions existed between the stakeholder collector and market-making roles of analysts. In the case of indicators where performance was considered unsatisfactory for all companies, analysts made an exception and did not redistribute the companies' scores from zero to five, leading to all companies receiving a score of zero (for example the indicator on disclosure of policy on competitive behaviour vis-à-vis generics companies B.II.4)[29]. The analysts for Index 2010 called these indicators "aspirational indicators" (Interview – Index 2010 analyst I). In the Index 2010 report, this is how such aspirational indicators are described (ATMI, 2010a, p. 236):

[29] B.II.4 is defined as: "The company publicly discloses its policies related to competition in areas such as data exclusivity, patent extensions etc. in the Index Countries" (ATMI, 2010b, p. 34).

> The scoring guidelines for all the indicators have been defined in such a way that, relative to all other companies in the peer group, the top performing company receives a score of five and the lowest performing company receives a score of zero. An important exception to this approach was when all the companies scored zero on an indicator, where achieving good distribution is not feasible.

While justified by the inability to obtain a "good distribution", the use of aspirational indicators was also motivated by the need to send a strong signal to the industry. In these cases, analysts gave primacy to the role of the Index as an instrument of stakeholder accountability rather than a tool for competition. Crucially, the use of aspirational indicators was limited to the bottom of the performance spectrum, where all companies appeared to be doing very little, and thus a good distribution was considered impossible. This was in part a discretionary choice, as the distribution of scores assigned for each indicator is itself, as shown, the result of trained judgment and does not descend automatically from the scoring guidelines. Some indicators were made aspirational whereas others were not; not all "bad distributions" resulted in the switch from relative to absolute measurement. At the very bottom of the performance spectrum (areas in which the industry as a whole was deemed immobile) making the Index an absolute ranking was acceptable, whereas at the top of the spectrum the imperative remained to reward top performers (so as to set a benchmark for the "laggards").

Relative measurement was essentially about rendering visible and highlighting differences between companies, while by extension making less visible the difference between each company's performance and a set of absolute demands. This was contested by some influential NGOs, who preferred the Index to be a tool for accountability vis-à-vis absolute benchmarks; a tool that exposed those companies "blocking access to medicine" (Interview - Head of international NGO – member of ERC for Index 2012).

However, from one iteration to the next the Index tends to become more demanding. According to the Foundation's research director, the Index gradually moves towards including an increasing number of "absolute indicators" in which more companies receive a low score and the best observed performance falls short of a score of five. This has made the maximum score (GlaxoSmithKline) decrease from 3.7 in Index 2010 to 3.3 in Index 2014. As the legitimacy of the Index in the governance space grows, the ability to make the

methodology more demanding increases. For example, in the Index 2016 methodology report, Leereveld emphasizes that: "The methodology is more demanding, guiding and ambitious than ever before: performance is even more important now." (ATMI, 2015, p. 3).

Discussion: the professional vision of commensuration and the politics of variability

This study documents the knowledge work underlying a ranking mobilized towards voluntary regulation of the global pharmaceutical industry in the area of access to medicines. The professional vision (Goodwin, 1994) required to produce this ranking is geared towards stimulating competition and inducing companies to self-manage and improve their "performance" in terms of the accessibility of their products. This vision entails the exercise of a form of trained judgment (Daston and Galison, 2007) allowing analysts to see equivalences and differences in ways that optimise the possibilities of benchmarking, while at the same time adopting a "view from everywhere" (Jasanoff, 2011) that seeks to remain equidistant within a polarized stakeholder base, as well as committing to represent the "view from nowhere" through the assumed replicability, impersonality and scientificity of quantification.

Unlike other studies of rankings centred on the consequences of commensuration, the case examined here brings to the fore some of the conditions under which commensuration and ranking can happen, which lead to different modes of commensuration. While "[c]ommensuration makes possible more mechanized decision-making" (Espeland and Stevens, p. 316), our study speaks of the different notions of objectivity that can be at play in processes of commensuration – different ways of seeing in order to compare and measure, and make commensuration and thus ranking possible in the first place. It shows commensuration as requiring the development of a specific expert vision of its objects, where such vision, as a regulative ideal for knowledge making, can move along the continuum that connects more standardized criteria of comparison and measurement with more subjective and ad-hoc ones, the "blind sight" of mechanical objectivity with the more sophisticated and discerning eye of trained judgment[30].

[30] As the Index methodology becomes more established, the form of trained judgment at play in the scoring of companies may be expected to become more standardised. However, scoring still requires the type of physiognomic sight detailed above rather than the blind sight of mechanical objectivity.

A ranking produced under the regulative ideal of mechanical objectivity relies on some agreed measurement protocol and rule of aggregation of relatively uncontroversial, standardized data. If the data and the aggregation rules are treated as "given", the process can largely appear as impersonal and mechanical. The common metric required to commensurate can be taken for granted, embedded in the measurement unit of the underlying data. If the raw data are not readily quantified and/or standardized, if they remain open to interpretation and require validation, the resulting process will appear less mechanical and more discretionary. In our case, the basic numbers which are aggregated in the ranking are not "pre-given" and readily comparable quantitative data, but are a set of often idiosyncratic qualitative and quantitative company data. The common metric and the scale within which it is assigned are also not given, but need to be derived by analysts in the process of scoring, which itself results from the interpretation and flexible use of the scoring guidelines.

According to Daston and Galison (2007) the emergence of trained judgment as an epistemic virtue was partly the outcome of frustration with the scientific objects produced under mechanical objectivity. These objects – epitomized by the photograph – produced faithful but often idiosyncratic representations, accurate in detail but frequently difficult to relate to each other. Claiming the authority to compare and interpret was the hallmark of the arrival of "trained judgment" – associated with objects such as stellar spectra or electroencephalograms, which were regarded as not speaking for themselves. Trained judgment was a response to the "fear of paralysis" (p. 370) that many scientists in the twentieth century felt vis-a`-vis mechanized representation, as similarity relations and thus classification could not always be determined by means of standard criteria. The desire to learn to see patterns and family resemblances in the objects of science was driven by the need to "classify and manipulate" (p. 322), to discern "normal" cases from abnormal, unexpected, unusual or novel cases inviting scientific intervention. Trained judgment was invoked to make the image useful, to make it readable as something that could be acted upon. Similarly, in the case of the Access to Medicine Index trained judgment is linked to the interventional character of a ranking aspiring to move the market.

This analogy between scientific atlases and ranking is not without limitations. After all, a ranking produced under the regulative idea of mechanical objectivity retains potential for acting upon its objects as much as one produced under the aegis of trained judgment. The common metric provides the comparability which the images of mechanical objectivity

examined by Daston and Galison may have lacked. The difference between mechanical objectivity and trained judgment in ranking lies in the extent to which the interventional character of the ranking is explicitly acknowledged and consciously planned, ingrained in the professional vision of the ranker. This has important implications.

In the first place, at stake in the distinction between mechanical objectivity and trained judgment is what is expected from the ranker's scientific self. Different notions of objectivity can be associated to different epistemic virtues, each embracing a different ethos (Daston & Galison, 2007). As different types of objectivity become superimposed, such ethos can be a contradictory one. In our case, analysts were cultivating the virtue of self-effacement that is the hallmark of mechanical objectivity, but the more confident self associated with trained judgment, while rarely foregrounded, was absolutely needed in order to apply the scoring guidelines and make the ranking possible. Behind the supposedly mechanical process of scoring, judgment was alive and kicking, but had to be tamed. Analysts engaged in exhausting data checks and collective peer review in order to be "on the same page" and reintroduce an element of impersonality where subjective judgment was resorted to but feared. The self-effacement practices adopted by analysts to reflect on their possible biases and "do a measurement completely objectively" (Interview - Foundation's Research Director since 2013 and Executive Director since late 2015), exhaustive data checks and peer review, speak of how commensuration and ranking are "fateful" not only for their objects (Espeland and Sauder, 2007), but for their subjects, too.

Secondly, at stake in the distinction is the extent to which performance variability can be managed in ways regarded as optimal to fulfil the aspiration to "move the market" so as to ultimately intervene in the polity. Ranking under mechanical objectivity means taking the common metric, and the variability of individual instances under such metric, for granted. It means running the regulatory task at hand on automatic pilot. Ranking under trained judgment means that a specific sight needs to be developed, negotiated and authorized in order to capture and interpret differences that are not automatically given. It means not taking variability for granted, but questioning it, evaluating it, and – as in the case of the adjustment of scoring guidelines or the use of aspirational indicators – deciding whether variability should be made visible by adjusting the depth of vision, or left to disappear into the background. Variability (or its lack) becomes an instrument of explicit intervention, aimed at producing a narrative of innovation and improvement (associated with the good distribution

resulting from the relative use of indicators) or one of immobility and under-performance (associated with the use aspirational indicators as absolute benchmarks of stakeholder expectations). The specific notion of objectivity embraced is thus related to what the ranker wants from its knowledge objects[31]. Understanding rankings through the lenses of trained judgment helps us bring to the fore the politics of variability associated with the interventionist agenda of the ranking, and the underlying "fear of paralysis" which a ranking that aspires to move the market so as to intervene in the polity needs to come to terms with.

Regulatory ranking and balancing the polity

A key mission of the Index was to achieve stakeholder consensus or "balanced views" on the direction the industry should be taking to tackle the problem of access to medicine in a field characterized by pervasive conflict. Throughout the construction of the Index, different perceived and/or expressed stakeholder views and interests were in principle given equal weight, which was seen as a precondition for them to be made commensurable through the Index (the need to measure only "what is in the middle"). This invited a notion of the analyst as a neutral stakeholder collector, transparent vis-a`-vis the preferences expressed by the stakeholder base and able to combine all perspectives in a view from everywhere. In as far as stakeholder preferences are seen as automatically transferrable into the analysis process, and the analyst is expected to simply "collect" them as a sort of "robot", this ideal process has much in common with mechanical objectivity (Daston & Galison, 2007; Porter, 1996)[32]. The Foundation, a civil regulator whose resources, alliances and authoritativeness in the highly contested field of access to medicine have had to be slowly built over the years, could find an "epistemic shelter" from accusation of political bias by embracing mechanical objectivity.

[31] Unlike credit ratings (Carruthers, 2013; Guseva & Rona-Tas, 2001; Poon, 2007), which seek to predict behaviour based on the law of large numbers and relatively "mechanized" statistical approaches, the scores given to achieve a "good distribution" in the case of the Access to Medicine Index are not based on formal statistical analysis but on a complex set of criteria ultimately resting on the assumption that companies can be moved, a movement which the eye of the ranker is trained to detect and make possible.

[32] In this respect, maintaining a view from everywhere is not dissimilar to the kind of erasure of the subject involved in developing the view from nowhere and aperspectival knowledge. However, while the view from nowhere is associated with authoritative science – it provides legitimacy and authority by virtue of its perceived impersonality and detachment from the sphere of interests and politics – the view from everywhere requires to hold together and balance diverse and, in our case, conflicting views. It confers legitimacy by virtue of its multi-stakeholder nature, but for the same reason remains more prone to breaking-up and collapsing into incompatible "somewheres".

Equidistance from stakeholders was paramount. As noted by Espeland and Stevens (1998, p. 336):

> When political disputes are framed as a contest over interests, parties are granted a formal, categorical equality among those with a political stake. Interest-group politics portrays outcomes as if differences were a matter of magnitude – of how much something matters, or of whose interests were served – rather than as disparate modes of investment in the decision.

To what extent and in what ways was the formal equality of stakeholders (epitomized by the ideal of measuring only "in the middle") achieved through performance measurement and ranking, if at all? How were qualitative differences among stakeholders and their different "modes of investment" in the access to medicine problem combined in the commensuration efforts of analysts? In the polarized space in which the Index operates, the commensuration of interests around the table was indeed a "formidable achievement" (Espeland and Stevens, 1998, p. 336). In practice, in certain areas views were and still are divided. The unresolved conflicts in the Index stakeholder base were partly absorbed, in a flexible and ad-hoc manner (technical area by technical area, indicator by indicator) by analysts and the Foundation. The use of absolute benchmarks and "aspirational indicators" to send a strong signal to the industry can be regarded as a way to deal with such tensions, and to adjudicate some of the issues at stake in favour of stakeholders such as some NGOs and the WHO when the industry appeared to be particularly immobile.

Thus, indicators, while seeking to capture variability, were themselves used "variably", sometimes as absolute sometimes as relative markers of performance, in a balancing act between stakeholder views. Underlying this variable use was the capacity to detect in the data signals of the industry moving (a reason to reward top performers) or being immobile (chronically lacking variability, a reason to send a strong signal). In this classification of mobility versus immobility and the related use of indicators as relative or absolute is a politics of variability which subordinates the use of the Index as a tool to expose the industry for doing too little (or even "blocking" access to medicine) to the benchmarking possibilities embedded in the idea of the "good distribution". Aspirational indicators were used only in extreme cases, when a good distribution was deemed not possible – though over time the number of such cases increases, and with it the use of absolute benchmarks and aspirational indicators.

By keeping benchmarks relative, the Foundation pushes for incremental change between each two iterations of the Index, but by increasing the number of aspirational indicators and absolute benchmarks over time, it seeks to create the conditions for more radical change in the long term. In contrast, some NGOs are weary of a ranking that has a company with a high score on top. Even though the "relative" nature of the ranking is made explicit, such stakeholders raise their concerns about the Index being a "triumphant output" (Interview – Head of International NGO) for the industry and would expect to see the average score decrease, in order to project a gap in companies' performance that can be the basis for claim-making.

Thus, despite all the attempts to render the Index an apolitical space and cultivate trust in the objectivity of numbers, the process of quantification and commensuration itself remained inevitably politicized. Our analysis of the scoring process underlying this regulatory ranking shows that commensuration can play an ambivalent role for contestation and claim-making in regulatory fields. In our case, the programmatic aspiration is to expunge political conflict from the ranking exercise and use ranking as an "anti-political" technology (Barry, 2002; Ferguson, 1990) aimed at managing contestation through the market-based organizing of political views (through consensus building) and practices (through ranking). In practice, however, contestation and conflict re-emerge in the production of numbers, which are kept by analysts (variably) open to become "ammunition machines" (Burchell et al., 1980) for different stakeholders and still carry a promise of transparency and democratization (Espeland, 2000; Porter, 1996). Our study has taken further steps towards detailing the processes through which commensuration and benchmarking can not only absorb and hide, but also preserve the possibility of politics in regulatory fields (Didier & Tasset, 2013). If neoliberal forms of rule have been seen as "the pursuit of the disenchantment of politics by economics" through the substitution of "political judgment with economic evaluation" (Davies, 2014, p. 3), the economic evaluation examined here illustrates that the flip side of such processes is a (silent) re-enchanting of economics through its absorption of the political.

Regulatory ranking and moving the market

In order to make companies commensurable for the purposes of incentivizing competition, analysts had to "intervene" (Hacking, 1983) in the market through the fine-tuning of scoring

guidelines, which allowed marginal differences between companies to be amplified in order to create a variable geometry within which competition could be enticed (the so-called "good distribution"). The level of detail sought in the data, analysts' depth of vision and thus the granularity of the differences between the companies captured by the Index were a function of this need, which required analysts to exercise a form of trained judgment by upgrading and defining some differences while filtering and uniforming others (Lynch, 1985; 1988).

Especially in the middle of the ranking, differences between companies tended to become more and more subtle – a sign that the change invited by the Index remains potentially marginal. Espeland and Sauder (2007) show that when organizational performance is placed in the discrete order created by ranking, immaterial differences between organisations can be projected as major differences in the final ranking, and such projections of difference can act as self-fulfilling prophecies by virtue of different reactivity processes. In the case studied here, the amplification of marginal differences was an explicit policy dictated by the regulatory mission of this ranking and its goal to entice competition. It required analysts to follow the scoring guidelines while keeping an eye on the resulting distribution, and to go back and forth between the two in order to find ways to improve the latter when it was not considered "good" for the purposes of benchmarking. If all rankings entail an assumption of difference and, more generally, neoliberal forms of government tend to be based on the "idea that differences are a fact" (Davies, 2014, p. 51), such facticity, as we have shown, depends on a particular professional vision and related epistemic practices of variability.

Pollock and D'Adderio (2012) discuss the market-making potential of the two-by-two graphs used in the IT industry as based on having an optimal number of dots to maximize the variability of company representations and hence improve the legibility of the graph for its intended audience. The "beautiful picture" of company performance is "the result of a negotiated, devised and contrived intervention" (p. 582) aimed at improving the information value of the ranking and thus its commercial value. In the case of the Index, in contrast, the "good distribution" of company scores is entangled in "politics of variability" where the minimum, average and maximum scores are allowed to vary and become at once ways to "fix the market mechanism" (by keeping the scoring relative and amplifying differences at the margin), objects of contestation (because relative rather than absolute measurement is seen by some NGOs as inadequate in exposing the industry) and means to manage those stakeholder conflicts in areas where the industry as a whole is perceived as under-performing (by

switching to absolute benchmarks and aspirational indicators). In other words, indicators, scores and mundane performance measurement processes become spaces for compromise and contention, where a more muted politics can resurface in numerical disguise in the form of the "good distribution" and "aspirational indicators".

A regulatory regime that depends on observed differences among targets is limited by the scale of those differences (and may have to resort to learn to see difference where none or little is found). Even if the ranking may succeed in inciting competition among some companies, the relative benchmarking imperative limits the scale of change invited in the regulatory targets. The Index however aspires to push for more ambitious regulatory targets across iterations by making scoring guidelines and indicators more demanding. Soft regulation as embodied by the Index can thus be seen as a form of "piecemeal regulation", whereby maximum variability and marginal differences are pursued in each iteration, while scoring guidelines are progressively made more demanding over the long term, making average scores decrease. This gradual phasing in of more stringent claims on companies becomes possible as the Foundation and the Index methodology become more established and less concerned about challenge from the industry.

Thus, the Index achieves a difficult balance between intervention and non-intervention. As an instrument that seeks to govern through competition, its ordering of the market needs to maintain "uncertainty of outcome" (Davies, 2014, p. 37). That is, it needs to leave the specific solutions to the access to medicine problem captured by each indicator open-ended, a function of the variability of company performance over time along the trajectory of improvement traced by the Index, rather than fixing them in a set of absolute benchmarks. The Index lets the assumed dynamism of the industry, cautiously enticed by the good distribution and aspirational indicators, identify best practice and lead the way to change. It seeks to intervene enough to make the potential for competition visible by amplifying marginal differences in performance and sending a signal to the industry as a whole when companies across the spectrum are deemed too immobile. But these interventions are meant to stimulate innovation and invite best practice, not to replace the market as a source of solutions. Each indicator acts as an empty frame to be filled by companies' initiatives. Variability in company performance vis-a`-vis competitors and over time becomes the axiom based on which the ranking is able to regulate through markets, by means of soft rather than hard measures. The politics of this

axiom rests precisely in the creation of a carefully controlled potential to intervene, both in the market and the polity, as an invisible hand.

Concluding Remarks

From a specialist and internally-oriented sub-field of accounting, performance management, which underlies hundreds of credit ratings, ESG ratings, profitability rankings, employee rankings, consumer rankings, stock market indices, development indicators and a host of regulatory rankings, is becoming a foundational ordering principle of private and public governance (Davis, Kingsbury, & Merry, 2012; Humphrey & Owen, 2000; Espeland & Sauder, 2007). This paper uncovers some of the knowledge production processes and tensions at play in the ascending role of performance measurement within the "audit society" (Power, 1999) or the "performance measurement society" (Humphrey & Owen, 2000) in a transnational context.

Our study of the epistemic work needed to measure performance for the purposes of a "regulatory ranking" points to the need to engage in a more systematic analysis of different performance measurement technologies, taking into account the specific mission they embrace and the position they carve for themselves in the polity, as well as their modes of commensuration and the particular professional vision they embed. We have shown that the epistemic work involved in pursuing the creation of a market for access to medicine by means of performance measurement and ranking requires at once an ethos of self-effacement and a particular way of seeing the market based on disciplined interpretation, where different attitudes to objectivity and thus different modes of commensuration, are at play.

While commensuration is deeply ingrained in an increasing set of social relations, it is not an invariable achievement, the almost unintended consequence of the spread of quantification. Our study draws attention to the specific professional vision underlying commensuration exercises, a vision that is variable, that needs to be cultivated, that can incur into epistemic tensions, and that can be at times effaced as a way to manage those tensions. Such tensions, in turn, can only be understood by taking into account the particular aspirations and programmatic ambitions (Rose & Miller, 1992) animating specific commensuration exercises, as these can come to constitute commensuration's modes of operation and the particular ways of seeing the objects it seeks to compare and measure, and thus intervene upon.

If “all epistemology begins in fear” (Daston & Galison, 2007, p. 373), then underlying the construction of a regulatory ranking in today’s regulatory capitalism is the fear of the political self, of unmanageable conflict and of unmovable markets, where conflict that cannot find resolution elsewhere is absorbed – reproduced but also muted – in the commensuration exercise. Our case shows that the explosion of quantification in spaces of governance does not coincide with depoliticization tout-court. Rather, we painted a more complex picture where numbers can be flexible carriers of both depoliticising and re-politicising aspirations, though the latter tend to remain invisible – a slow but persistent movement under the surface of marketization processes. Politics is enacted both in the creation of an invisible hand entrusted with the regulation of a social good like health, and in the gradual resurfacing of stakeholder demands within the commensuration process, in what we have termed “piecemeal regulation”.

Our study illustrates how the tension between the political and anti-political potentials (Didier & Tasset, 2013; Barry, 2002; see also Ferguson, 1990) of performance measurement exercises can unfold in the work of rankers, inviting a closer scrutiny of the mutually constitutive relations between calculative technologies and the polity. While the shift from state-based regulation to soft modes of governance through performance measurement and ranking tends to “vanquish a vocabulary of power and hence power’s visibility from the lives and venues that governance organizes and directs” (Brown, 2015, p. 129), rather than discounting all such technologies as “hostile to politics” (Brown, 2015, p. 127), it becomes even more important to investigate the conditions under which both depoliticization and the channelling and resurfacing of politics through numbers can occur. Such analyses can help bring attention to how the immanence of commensuration in global governance can retain emancipatory potential (Espeland & Stevens, 1998) – and thus to the conditions for enhancing such potential.

Hundreds of analysts throughout the polity are struggling everyday with these competing aspirations, producing a host of performance measures, rankings and other market-based regulatory technologies comparing organizational performances in the delivery of public goods across a wide range of social issues of concern to millions of people. We are hopeful that our analysis can serve as a starting point to foreground epistemic processes of political mediation, knowledge authorization and commensuration that underlie the fast expanding role

of regulatory rankings as a core technology of soft governance in regulatory capitalism. These processes, as we have shown, have significant implications for the three fundamental promises of such rankings – political mediation by restructuring a space of conflict into one of performance measurement and benchmarking, moving regulatory targets through competitive pressures, and providing a faithful representation of those same targets to inform public debates. We are hoping that this study will lead to a more systematic analysis of regulatory rankings and their complex interventions across the polity.

Acknowledgments

Earlier drafts of this manuscript have been presented at the London School of Economics and Political Science, Leicester University School of Management, ESSEC Business School, and at a variety of workshops and conferences. We wish to thank the participants to all these events and especially our discussants for the constructive comments and critical views they offered on our work in progress. In addition, we are grateful to Diane-Laure Arjaliès, Sebastian Becker, Matthias Benzer, Yasmine Chahed, Silvia Jordan, Daniel Martinez and Keith Robson for reading and providing feedback on earlier drafts of this paper. The paper has greatly benefited from the observations and suggestions of the editor and two anonymous reviewers. We express sincere appreciation to the Access to Medicine Foundation and especially Wim Leereveld and Jayasree Iyer for their support and useful comments on our on-going research and on this paper in particular. Finally, we thank all the analysts and other access to medicine experts who shared with us their views on the Index and the access to medicine problem more generally.

References

ATMI. (2010a). Access to Medicine Index 2010. Retrieved March 3, 2013, from: http://www.accesstomedicineindex.org/sites/www.accesstomedicineindex.org/files/2010-access-to-medicine-index_0.pdf.

ATMI. (2010b). Access to Medicine Index 2010 Methodology and Stakeholder Report. Retrieved March 3, 2013, from: http://www.accesstomedicineindex.org/sites/www.accesstomedicineindex.org/files/2010_access_to_medicine_index_methodology_stakeholder_review.pdf.

ATMI. (2012a). Access to Medicine Index 2012. Retrieved March 3, 2013, from http://www.accesstomedicineindex.org/sites/www.accesstomedicineindex.org/files/2012-access-to-medicine-index-full-report-clickable.pdf.

ATMI. (2012b). Methodology Report 2012 Stakeholder Review. Retrieved from http://www.accesstomedicineindex.org/sites/www.accesstomedicineindex.org/files/2012_Methodology_Report_-_Stakeholder_Review_-_May_2012.pdf.

ATMI. (2015). The Access to Medicine Index Methodology 2015. Retrieved May 11, 2015 from: http://www.accesstomedicineindex.org/sites/2015.atmindex.org/files/2015methodology_2016accesstomedicineindex_accesstomedicinefoundation.pdf.

ATMF. (2014). Annual Report 2014. Retrieved from http://www.accesstomedicineindex.org/sites/2015.atmindex.org/files/access_to_medicine_foundation_2014_annual_report.pdf.

Barry, A. (2002). The anti-political economy. *Economy and Society*, *31*(2), 268–284.

Bartley, T. (2003). Certifying forests and factories: states, social movements, and the rise of private regulation in the apparel and forest products fields. *Politics & Society*, *31*(3), 433.

Bell, J. (1843). *A concise historical sketch of the progress of pharmacy in Great Britain*. London, UK: John Churchill.

Bloomfield, B., & Vurdubakis, T. (1997). Visions of organization and organizations of vision: The representational practices of information systems development. *Accounting, Organizations and Society*, *22*(7), 639–668.

Blowfield, M., & Dolan, C. (2010). Outsourcing governance: Fairtrade's message for C21 global governance. *Corporate Governance*, *10*(4), 484–499.

Boli, J. (2006). The rationalization of virtue and virtuosity in world society. In Djelic & K. Sahlin-Andersson (Eds.), *Transnational Governance - Institutional Dynamics of Regulation* (pp. 95–118). Cambridge, UK: Cambridge University Press.

Braithwaite, J. (2008). *Regulatory capitalism: how it works, ideas for making it work better*. London, UK: Edward Elgar Publishing.

Brown, W. (2015). *Undoing the Demos*. Durham, NC: Duke University Press.

Burchell, S., Clubb, C., Hopwood, A., Hughes, J., & Nahapiet, J. (1980). The roles of accounting in organizations and society. *Accounting, Organizations and Society*, 5(1), 5–27.

Burril Report. (2011). Talking Turkey About Access to Medicines - 11/22/2011 - The Burrill Report. Retrieved March 4, 2013, from http://www.burrillreport.com/article-talking_turkey_about_access_to_medicines.html.

Busco C., & Quattrone, P. (2015). Exploring how the balanced scorecard engages and unfolds: Articulating the visual power of accounting inscriptions. *Contemporary Accounting Research*. 32(3), 1236-1262.

Buthe, T. (2010). Special Issue: Private Regulation in the Global Economy. *Business and Politics*, *12*(3).

Büthe, T., & Mattli, W. (2011). *The new global rulers: The privatization of regulation in the world economy*. Princeton, NJ: Princeton University Press.

Çalışkan, K., & Callon, M. (2010). Economization, part 2: a research programme for the study of markets. *Economy and Society*, *39*(1), pp. 1-32.

Callon, M. (2009). Civilizing markets: Carbon trading between in vitro and in vivo experiments. *Accounting, Organizations and Society*, *34*(3-4), 535-548.

Carruthers, B. G. (2013). From uncertainty toward risk: the case of credit ratings. *Socio-Economic Review*, *11*(3), 525–551.

Chenhall, R. H., Hall, M. R., & Smith, D. (2013). Performance measurement, modes of evaluation and the development of compromising accounts. *Accounting, Organizations and Society*, *38*(4), 268 - 287.

Chua, W. F. (1995). Experts, networks and inscriptions in the fabrication of accounting images: A story of the representation of three public hospitals. *Accounting, Organizations and Society*, *20*(2/3), 111–145.

Dambrin, C., & Robson, K. (2011). Tracing performance in the pharmaceutical industry: Ambivalence, opacity and the performativity of flawed measures. *Accounting, Organizations and Society*, *36*, 428–455.

Daston, L. (1992). Objectivity and the Escape from Perspective. *Social Studies of Science, 22*(4), 597-618.

Daston, L., & Galison, P. (2007). *Objectivity*. Brooklyn, N.Y.: Zone Books.

Davis, K. E., Kingsbury, B., & Merry, S. E. (2012). Indicators as a technology of global governance. *Law & Society Review*, *46*(1), 71–104.

Davies, W. (2014). The limits of neoliberalism: Authority, sovereignty and the logic of competition. London, UK: Sage Publications.

Desrosieres, A. (2001). How Real Are Statistics? Four Possible Attitudes. *Social Research 68(2),* 339-355.

Didier, E., & Tasset, C. (2013). Pour un statactivisme. La quantification comme instrument d'ouverture du possible. *Tracés*, 24(1), 123–140.

Dingwerth, K., & Pattberg, P. (2009). World politics and organizational fields: The case of transnational sustainability governance. *European Journal of International Relations*, *15*(4), 707–743.

Djelic, (2006). Marketization: From intellectual agenda to global policy-making. In Djelic & K. Sahlin-Andersson (Eds.), *Transnational governance. Institutional dynamics of regulation*. Cambridge, UK: Cambridge University Press.

Djelic, M. L., & Quack, S. (2010). *Transnational Communities: Shaping Global Economic Governance.* Cambridge, UK: Cambridge University Press.

Djelic, M. L., & Sahlin-Andersson, K. (2006). *Transnational Governance: Institutional Dynamics of Regulation*. Cambridge, UK: Cambridge University Press.

Djelic, M.-L., & Sahlin, K. (2009). Governance and Its Transnational Dynamics: Towards a Reordering of our World? In Christopher S. Chapman, David J. Cooper, and Peter B. Miller (Eds.), *Accounting, Organizations, and Institutions Essays in Honour of Anthony Hopwood*. Oxford, UK: Oxford University Press.

Drori, G. S., & Meyer, J. W. (2006). Scientization: making a world safe for organizing. In Djelic M.L., Sahlin-Andersson, K. (Eds.), *Transnational Governance. Institutional Dynamics of Regulation.* Cambridge, UK: Cambridge University Press.

Elsbach, K. D., & Kramer, R. M. (1996). Members' responses to organizational identity threats: Encountering and countering the Business Week rankings. *Administrative Science Quarterly*, 442–476.

Espeland, W. N. (2000). Bureaucratizing democracy, democratizing bureaucracy. *Law & Social Inquiry*, 25(4), 1077–1109.

Espeland, W. N., & Stevens, M. L. (1998). Commensuration as a social process. *Annual Review of Sociology*, 313–343.

Espeland, W. N., & Sauder, M. (2007). Rankings and Reactivity: How Public Measures Recreate Social Worlds. *American Journal of Sociology*, *113*(1), 1–40.

Espeland, W., & Sauder, M. (2015). *How rankings re-organize education.* Unpublished book manuscript.

Ezzamel, M. (2004). Accounting representation and the road to commercial salvation. *Accounting, Organizations and Society, 29*, 783–813.

Ferguson, J. (1990), The anti-politics machine: "Development", depoliticization, and bureaucratic power in Lesotho. Cambridge; New York: Cambridge University Press.

Fransen, L. (2011). Why do private governance organizations not converge? A political–institutional analysis of transnational labor standards regulation. *Governance*, *24*(2), 359-387.

Fransen, L. W., & Kolk, A. (2007). Global rule-setting for business: A critical analysis of multi-stakeholder standards. *Organization*, *14*(5), 667–684.

Free, C., Salterio, S. E., & Shearer, T. (2009). The construction of auditability: MBA rankings and assurance in practice. *Accounting, Organizations and Society*, *34*(1), 119–140.

Goodwin, C. (1994). Professional vision. *American Anthropologist*, *96*(3), 606–633.

Grace, C. (2004). The effect of changing intellectual property on pharmaceutical industry prospects in India and China. *DFID Health Systems Resource Centre, London*.

Greene, J. A. (2011). Making medicines essential: The emergent centrality of pharmaceuticals in global health. *BioSocieties*, *6*(1), 10–33.

Gulbrandsen, L. H. (2008). Accountability arrangements in non-state standards organizations: instrumental design and imitation. *Organization*, *15*(4), 563–583.

Guseva, A., & Rona-Tas, A. (2001). Uncertainty, risk, and trust: Russian and American credit card markets compared. *American Sociological Review*, *66*(5), 623–646.

Hacking, I. (1983). *Representing and intervening: Introductory topics in the philosophy of natural science*. Cambridge, UK: Cambridge University Press.

Hogerzeil, H. V. (2013). Big Pharma and social responsibility--the Access to Medicine Index. *The New England Journal of Medicine*, *369*(10), 896–899.

Humphrey, C., & Owen, D. (2000). Debating the "power" of audit. *International Journal of Auditing*, *4*(1), 29–50.

Jasanoff, S. (1995). Procedural choices in regulatory science. *Technology in Society*, *17*(3), 279-293.

Jasanoff, S. (2011). The practices of objectivity in regulatory science. In Camic C, Gross N, & Lamont M, *Social knowledge in the making.* University of Chicago Press, pp. 307–338. Chicago, IL.

Jasanoff, S. (2012). *Science and public reason*. New York, NY: Routledge.

Jeacle, I., & Carter, C. (2011). In TripAdvisor we trust: Rankings, calculative regimes and abstract systems. *Accounting, Organizations and Society*, *36*(4), 293–309.

Jordana, J., Levi-Faur, D., & i Marín, X. F. (2011). The global diffusion of regulatory agencies: channels of transfer and stages of diffusion. *Comparative Political Studies*, *44(10)*, 1343–1369.

Jorgensen, D. L. (1989). *Participant observation, A Methodology for Human Studies*. London, UK: Sage Publications.

Kim, E. H., & Lyon, T. P. (2011). Carbon Disclosure Project. In T. Hale & D. Held (Eds.), *Handbook of Transnational Governance: Institutions and Innovations*, Cambridge, UK: Polity Press.

Kornberger, M., & Carter, C., (2010), Manufacturing competition: how accounting practices shape strategy making in cities. *Accounting, Auditing & Accountability Journal 23* (3), 325-349.

Lascoumes, P., & Le Gales, P. (2007). Introduction: understanding public policy through its instruments. *Governance*, *20*(1), 1–21.

Leereveld, W. (2014). How do you get an industry moving in the right direction? The Access to Medicine Index. *LSE Blog*. Retrieved from: http://blogs.lse.ac.uk/businesshumanrights/2014/04/25/wim-leereveld-how-do-you-get-an-industry-moving-in-the-right-direction-the-access-to-medicine-index/.

Levi-Faur, D. (2005). The global diffusion of regulatory capitalism. *The Annals of the American Academy of Political and Social Science*, 598, 12-32.

Levi-Faur, D., & Jordana, J. (2005). The rise of regulatory capitalism: The global diffusion of a new order. *The Annals of the American Academy of Political and Social Science*, *598*(1), 200–217.

Levi-Faur, D. (2010). Regulation &Regulatory Governance. *Jerusalem Papers in Regulation & Governance Working Papers Series*. Retrieved from: http://levifaur.wiki.huji.ac.il/images/Reg.pdf.

Lynch, M. (1985). Discipline and the material form of images: An analysis of scientific visibility. *Social Studies of Science*, *15*(1), 37–66.

Lynch, M. (1988). The externalized retina: Selection and mathematization in the visual documentation of objects in the life sciences. *Human Studies*, *11*(2), 201-234.

MacKenzie, D. (2009). Making things the same: Gases, emission rights and the politics of carbon markets. *Accounting, Organizations and Society*, *34*(3-4), 440–455.

MDG. (2000). Millenuim Development Goals. Retrieved August 8, 2015 from http://www.un.org/millenniumgoals/global.shtml.

Mennicken, A., & Miller, P. (2012). Accounting, territorialization and power. *Foucault Studies*, *13*, 4–24.

Morris, K. (2010). Bridging the gap in access to medicines. *Lancet*, 10(8), 514-515.

Mörth, U. (2004). *Soft law in governance and regulation: an interdisciplinary analysis*. Cheltenham, UK: Edward Elgar Publishing.

Nagel, T. (1986). *The view from nowhere*. Oxford, UK: Oxford University Press.

Pollock, N., & D'Adderio, L. (2012). Give me a two-by-two matrix and I will create the market: Rankings, graphic visualisations and sociomateriality. *Accounting, Organizations and Society*, 37(8), 565–586.

Poon, M. (2007). Scorecards as devices for consumer credit: the case of Fair, Isaac & Company Incorporated. *The Sociological Review*, *55*(s2), 284–306.

Porter, T. M. (1996). *Trust in numbers: The pursuit of objectivity in science and public life*. Princeton, NJ: Princeton University Press.

Power, M. (1999). *The Audit Society: Rituals of Verification*. Oxford, UK: Oxford University Press.

Power, M. (2004). Counting, control and calculation: Reflections on measuring and management. *Human Relations*, *57*(6): 765–783.

Qu, S., & Cooper, D. (2011). The role of inscriptions in producing a balanced scorecard. *Accounting, Organizations and Society, 36,* 344–362.

Quack, S. (2010). Law, expertise and legitimacy in transnational economic governance: an introduction. *Socio-Economic Review*, *8,* 3-16.

Reed, T. (2008). The new Access to Medicine Index. *The Lancet*, *372*(9642): 890-1.

Reinecke, J., Manning, S., & Von Hagen, O. (2012). The emergence of a standards market: Multiplicity of sustainability standards in the global coffee industry. *Organization Studies*, *33*(5-6), 791–814.

Rottenburg, R., Merry, S. E., Park, S.-J., & Mugler, J. (2015). *The World of Indicators: The Making of Governmental Knowledge Through Quantification*. Cambridge, UK: Cambridge University Press.

Rose, N., & Miller, P. (1992). Political power beyond the state: problematics of government. *British Journal of Sociology*, *43*(2), 173–205.

Sauder, M., & Espeland, W. N. (2009). The discipline of rankings: Tight coupling and organizational change. *American Sociological Review*, *74*(1), 63–82.

Scott, S. V., & Orlikowski, W. J. (2011). Reconfiguring relations of accountability: Materialization of social media in the travel sector. *Accounting, Organizations and Society*. *37*(1), 26-40.

Shapin, S. (2012). The sciences of subjectivity. *Social Studies of Science, 42*(2), 170-184.

Shaw, K. (2004). Knowledge, foundations, politics. *International Studies Review*, *6*(4), 7–20.

Siddiqi, J. (1995). *World health and world politics: the World Health Organization and the UN System*. University of South Carolina Press.

Strathern, M. (1997). "Improving ratings": audit in the British University system. *European Review*, *5*(03), 305–321.

Technopolis. (2016). Independent evaluation of the Access to Medicine Index. Retrieved March 15, 2016, from: http://www.accesstomedicineindex.org/sites/2015.atmindex.org/files/publication/executive_summary_independent_evaluation_of_the_atm_index_by_technopolis_group_.pdf.

T Hoen, E., Moatti, J., Coriat, B., Souteyrand, Y., Barnett, T., Dumoulin, J., & Flori, Y. (2003). TRIPS, pharmaceutical patents and access to essential medicines: Seattle, Doha and beyond. *Economics of AIDS and Access to HIV/AIDS Care in Developing Countries: Issues and Challenges. Paris: Agence Nationale de Recherches Sur Le Sida*, 39–67.

UN (2011). Millenium Development Goal 8: The Global Partnership for Development: Time to Deliver. Retrieved March 15, 2016, from: http://www.un.org/en/development/desa/policy/mdg_gap/mdg_gap2011/mdg8report2011_engw.pdf.

Vogel, D. (2008). Private global business regulation. *Annual. Review of. Political. Science, 11*, 261–282.

Wang, H., & Rosenau, J. N. (2001). Transparency international and corruption as an issue of global governance. *Global Governance, 7(1)*, 25-49.

Appendix 1: List of interviewees

Organization	Position held	Role at the Foundation
Access to Medicine Foundation	Founder and CEO from 2012 until late 2015 – current member of ATMF advisory board	Founded the Foundation in 2005 and since has been involved in various executive and chair positions at the Foundation
	Chief Operating Officer since 2011	Supervised the production of Index 2012 and 2014 and other operations of the Foundation
	Research Director since 2013 and Executive Director since late 2015	In charge of managing analyst team and production of Index 2012 and 2014
	CEO from November 2008 to December 2010	Supervised Foundation operations during the development of Index 2010
MSCI Group	Analyst I for Index 2010	Were hired following the finalization of methodology for the analysis/ranking phase for Index 2010 – from November 2009 to May 2010. Each analyst was responsible for analysing and scoring five companies and also to specialize in 1-2 technical areas to check the quality of the other analysts' work in this areas.
	Analyst II for Index 2010	
	Analyst III for Index 2010	
	Analyst IV for Index 2010	
	Senior analyst	Headed the analyst team at MSCI for Index 2012
MSCI Group – Access to Medicine Foundation	Former analyst at MSCI Group for ATMI 2012 – Analyst at the Foundation since 2012	Analyst for Index 2012 and 2014
Innovest Strategic Value Advisors (later acquired by MSCI Group)	Senior analyst	Project manager and head of analyst team Index 2008
	Former CEO	Involved in discussion leading to the development of Index 2008 and 2010
	Former research Director – 2011-2013 consultant at the Foundation	Liaised with the Foundation and other stakeholders and supervised the development of Index 2008 – from 2010 to 2012 – responsible for investor advocacy at the Foundation
Sustainalytics	Analyst for Index 2014	Involved in and company analysis and quality checks
Health Action International	Executive director	ERC member for ATMI 2012 representing NGOS
HiVOS	Former Program Manager	NGO involved in funding of Index 2008 and 2010 and present in stakeholder meetings for those editions of the Index
Department for International Development	Consultant for the Foundation in 2012 and then Program Manager	Consultant for Index 2012 at Foundation in charge of leading the Foundation's in-house team and the write-up of the report

(DFID)	at DFID	- in charge of managing relationship/reporting with the Foundation at DFID since 2013
Novartis	Head of Global Policy	In charge of company response to Index data request and present in the Foundation's company events
Merck	Former Head of CSR	Member of stakeholder roundtable for Index 2008 and ERC for Index 2010
World Health Organization	Former Team Coordinator - Medicines Information and Evidence for Policy Unit	Member of ERC for Index 2008, 2010, 2012 and 2014 – reviewer for Index 2010, 2012, 2014 and 2016
DFID	Former economic adviser	Member of ERC for Index 2008, 2010, 2012
Meteos	Co-founder and partner - formerly SustainAbility	Member of stakeholder roundtable for Index 2008 and ERC Chair for Index 2010 and 2012

Appendix 2: Sources of information used in the analysis process (Index Methodology Report 2012, p.27)

Journals/ Articles

The Lancet, E-drug, ICIUM-3 conference abstracts, PLoS, British Medical Journal.

Corporate documents

Annual reports, environmental and CSR reports, securities filings, 10k and other, websites.

Government and multilateral organisation data

Publications, databases and interviews with governmental officials, e.g. the Center for Responsive Politics (Public Policy Influence & Advocacy), the US National Institutes of Health (R&D and Clinical Trials), PubMed (drug indications), FDA (drug quality and promotion), EMEA, WHO (Prequalification, registration, patents, pricing).

Website content and reports

clinicaltrials.gov, WTO(Compliance with TRIPS), ICH-GCP (Research Ethics), ANDI Network, WIPO Re:Search, UN (UNHR - clinical trial conduct, UNDP - IP Rights, UNITAID, UNCTAD, UNGRI, UNICEF), OECD (anti-corruption), WBI, CHAI, and World Economic Forum (PACI).

Online news databases & MSCI Search Engines

LexisNexis, Impact Monitor.

Other online databases

Policy Cures ('G-Finder'), globalhealthprogress.org (Capacity building, philanthropy & R&D), mims.com, REPRISK.

Industry sources

Pharmaceutical industry publications and reports, e.g. IFPMA, ABPI, PhRMA, EFPIA, NEFARMA, LEEM, Industry journals, e.g. BioExecutive, PharmaFocus, Pharmaceutical Executive, and Pharmatimes.

Not for Profit and Civil Society

Either reports from and/or interviews with Non-Governmental Organizations familiar with the companies' operations have taken place with the following entities: African Medicines Regulatory Harmonisation Initiative (AMRHI, NEPAD), Campaign for Global Development (CG Dev), CARE, Center for Political Accountability, Centre for Research on Multinational Corporations (SOMO), Competition Authorities reports, Concept Foundation, Elizabeth Glazer Pediatric AIDS Foundation, European-Developing Country Clinical Trials Program (EDCTP), Global Fund for AIDS, TB and Malaria Price Reporting Mechanism, Health Action International (HAI), Institute for One World Health (iOWH), Médecins Sans Frontières (MSF), Nuffield Council on Bioethics, South Centre, United Nations Industrial Development Organization [(UNIDO) /GTZ partnership project for local production in

Africa], Wellcome Trust.

Multilateral agencies

European Commission, European Parliament Directorate General for External Policies Policy Department, United Nations (UN), World Health Organization (WHO), World Intellectual Property Organization (WIPO), World Trade Organization (WTO).

Government agencies

DFID

Other third-party sources

International agreements and codes of conduct were consulted, including the Partnering Against Corruption Initiative (PACI) of the World Economic Forum and Declaration of Helsinki. Reports and interviews with the stakeholders we consulted during the develop- ment of the Index framework including investors, consultants and academics.

Only specific information is sought from company representatives where there are gaps in data or inconsistencies among the above-mentioned sources. The companies are the primary source of information in areas such as research pipeline details and products port- folio details. They respond to a detailed information collection package, including an online questionnaire and email and phone communications, which covers primarily these areas and also other areas where our analysts need additional information from the companies.