Metrics Producing Science: The Evaluative Way in the Emergence of US Translational Science

Stephen Molldrem

Institute for Bioethics and Health Humanities, Institute for Translational Sciences, The University of Texas Medical Branch, USA/ stmolldr@utmb.edu

Jacob D. Moses

Institute for Bioethics and Health Humanities, The University of Texas Medical Branch, USA

Elise M. R. Smith

Institute for Bioethics and Health Humanities, Institute for Translational Sciences, The University of Texas Medical Branch, USA

Emma Tumilty

Institute for Bioethics and Health Humanities, Institute for Translational Sciences, The University of Texas Medical Branch, USA

Jeffrey S. Farroni

Institute for Bioethics and Health Humanities, Institute for Translational Sciences, The University of Texas Medical Branch, USA

Abstract

Translational science emerged in the 2000s, aiming to transform biomedicine by fostering strategies to move innovations more efficiently from basic research, through trials, to implementation. One aspect of translational science that has received limited critical attention is the development of evaluation practices that aim to measure its impact on biomedical innovation and society. In this article, we describe five periods in the evaluation of translational science funded by the US National Institutes of Health. Rather than simply measuring translational science's effects, we show that evaluation strategies have structured the field's emergence, produced representations of it for external actors, and helped legitimize it as a science worthy of public investment. We term this specific mode of co-production between science and metrics 'the evaluative way', and argue for the concept's utility for studying sciences that emerged in the late 20th and early 21st centuries, when evaluation became increasingly central to science policy.

Keywords: Evaluation, Translational Science, Translational Research, NCATS, CTSA

Introduction and Background

Translational science is a movement within the health sciences that emerged in the 2000s and accelerated in the United States (US) in 2005, following the establishment of the Clinical and Translational Science Award (CTSA) program within the National Institutes of Health (NIH) (Zerhouni, 2005; Austin, 2021; Aviles, 2018; NIH, 2005; Robinson, 2019). Translational science aims to transform bio-

medical discovery across therapeutic domains by helping to move innovations more efficiently from basic research, through trials, to commercialization or implementation (Austin, 2021). At its inception, this mission was often expressed as rapidly moving knowledge 'from bench to bedside' (Austin, 2018; Robinson, 2017; Zerhouni, 2005). In the US, the main vehicle for supporting translational science has been the CTSA program, which funds a network of ~60 university-based 'hubs' to support translational endeavors. Scholars in Science and Technology Studies (STS) have written both sympathetically and critically about the development of translational science and its triumphal self-narrative (Aarden et al., 2021; Aviles, 2018; Jeske, 2022; Robinson, 2017, 2019, 2020; Smith et al., 2024). However, one central aspect of translational science that has received limited critical attention is the development and influence of metrics and evaluation strategies designed to measure its impact.

Herein, we demonstrate that, rather than simply assessing translational science, efforts to develop metrics and methods to evaluate the field have also played a key role in bringing it into existence and legibility as what former National Center for Advancing Translational Science (NCATS) director Christopher Austin has described as "the new youngest science" (Austin, 2018: 456). Through an analysis of policy grey literature and key publications, we describe successive efforts to evaluate the CTSA program, NIH's flagship translational science initiative, treating it as an exemplary case in the development of translational science.[1] Since 2005, NIH has allocated approximately a half-billion dollars per year to fund CTSA hubs to catalyze and coordinate translational activities and also to train scientists in translational approaches by funding pilot grants and early-career investigators' salaries (Institute of Medicine et al., 2013; NCATS, 2023). The CTSA program has focused on developing methods in "the science of translation" to more efficiently drive innovation across therapeutic areas in a "disease agnostic," "disease independent," or "disease universal" fashion (Austin, 2021: 1631-34; Smith et al., 2024). The translational science movement and CTSA program have succeeded in moving segments of the publicly funded biosciences toward a model that encourages investigators to pursue research with the aim of trialability and end-stage implementation early in the discovery process (Llewellyn et al., 2018, 2020). The translational science model is thus more closely aligned with corporate research and development models than traditional notions of the scientific investigator existing apart from market forces (Robinson, 2017, 2019, 2020).

We describe the implementation of strategies that CTSAs and the CTSA program devised to evaluate this new science, to characterize its impact using data from evaluations, and to justify its value to sometimes-skeptical policymakers. Efforts to evaluate the CTSA program have driven the field's formation, structured its development, and played key roles in legitimizing and framing translational science as a distinctive field worthy of continued federal investment. Evaluation has taken the form of traditional quantitative or mixed methods studies by evaluation scientists, including work to create compelling stories about translational innovations to be used in public-facing narratives or with policymakers (Kaplan-Liss et al., 2022; Luke et al., 2018). However, CTSA program evaluation has also included a broad set of practices enacted by CTSA employees and funded investigators as part of routine operations (Daudelin et al., 2020; Welch et al., 2021). In sum, rather than merely measuring the development of translational science, metrics for translational science have effectively co-constructed the field by becoming part of its constitutive socio-material life and self-narrative. We argue that the drive to evaluate translational science has become so central to the field's emergence and self-understanding that the infrastructures, policy paradigms, and practices developed to achieve evaluation goals constitute a distinct *evaluative way of knowing* that permeates the field as a semi-autonomous, integrated, and scientized set of techniques that purport to measure and attest to the varied forms of value it produces for society (on 'ways of knowing,' see, Bowker et al., 2009; Dear, 1995).

Drawing on scholarship on the performativity of metrics and measurement systems (Law, 2009; MacKenzie, 2008), the co-production of science and society (Jasanoff and Kim, 2009), and valuation studies (Doganova et al., 2014; Stark,

2011), we periodize the emergence of translational science by analyzing how transformations of its evaluation paradigms enacted the field in practice (Mol, 1999, 2002). While evaluation practices for the CTSA program, individual CTSAs, and CTSA-led initiatives were initially conducted by internal and external evaluators, evaluation strategies specific to translational science eventually came to constitute a kind of science of their own and a major area of activity into which nearly all actors involved in CTSA program activities became enrolled (Callon and Law, 1982; Patel et al., 2019). This led to the emergence of pervasive cultures of evaluation within translational science, as the collective enactment of scientistic evaluation strategies became increasingly central to the everyday work of doing translational science (Trochim et al., 2013; Schneider et al., 2015; Luke et al., 2018; Patel et al., 2019; Daudelin et al., 2020; Stevens et al., 2021). We thus describe the process by which *the doing of translational science* and *its associated evaluation aims and activities* became co-productive of the field by reinforcing and performing its claims to epistemic legitimacy and value to society. This evaluative way of knowing fused increasingly scientized modes of evaluation with the basic aims of translational science, with both mutually sustaining the field and producing representations of its impact that made it legible as a science deemed worthy of sustained public investment.

Source materials and approach

The authors are faculty members in an Institute for Bioethics and Health Humanities at an academic health science center with a CTSA hub. This study was partly undertaken as part of their roles to meet their CTSA's strategic objectives, including to develop a program of research on the ethics and practice of translational science. This objective is itself an object of evaluation of their institution's CTSA and this paper will be counted toward meeting its associated metrics.

Embedded in the constructivist STS tradition, we sought to write a 'history of the present' regarding the role of evaluation in translational science and the CTSA program (Foucault, 1975: 30-31). We used documents and publications to track the development of evaluation paradigms from the establishment of the CTSA program in 2005 through 2023 (NIH, 2005, 2021). We aimed to periodize translational science through the history of CTSA metrics and evaluation strategies, tracking the development of what we came to call its 'evaluative way.'

To structure our periodization, we identified moments where empowered entities articulated shifts in translational science's evaluation priorities. We then discerned key policies, actors, evaluation methodologies, institutions, information infrastructures, and controversies in each period. We gathered documents produced by relevant actors and organizations, including peer-reviewed studies and grey literature. We met weekly for several months as we collected, read, discussed, and annotated materials. Discussions led us to construct a five-phase periodization organized around key epistemic and institutional shifts in CTSA evaluation paradigms that also attended to enduring areas of evaluative focus that cut across the periods.[2]

Historicizing evaluation in (translational) science

We approach metrics and evaluation practices in the sciences as a set of activities that do not neutrally measure fields and their impact, but also bring the evaluated fields into being, enacting them through paired practices of measurement and the dissemination of representations of fields produced by measurement strategies (Elzinga, 1988; Law, 2004, 2009; Mol, 1999, 2002). This constructivist insight has been central to scholarship in STS that describes how approaches to measurement that aim to make phenomena knowable also effectively produce the thing, subject, or object being examined (Bowker and Star, 1999; Hacking, 1995; MacKenzie, 2008). The productive effects of measurement techniques have been demonstrated regarding deviant subject-positions (Foucault, 1975, 1978), economic markets (Çalışkan and Callon, 2010; MacKenzie, 2008), sexual practices among gay men (Holt, 2013), political constituencies (Law, 2009), and diagnostic categories (Mol, 2002). MacKenzie and Law refer to the productive effects of metrics as the 'performativity' of meas-

urement (Law, 2009: 240-44; MacKenzie, 2008: 15-25; see also, Austin, 1975; Butler, 1988).

In arguing that evaluation metrics and practices in translational science have co-constructed the field, we aim to demonstrate how systems developed to measure its impact have also generated new practices within translational science and representations of it that *performed translational science into existence*. Further, when we speak of an 'evaluative way' in translational science, we refer to the broad set of measurement practices, tools, standards, and approaches that became diffused throughout the field. Together, these have come to form a historically, culturally, and cross-institutionally specific 'way of knowing' that provides part of the epistemic and infrastructural basis for a particular zone of activity in science or bureaucratic administration (see, Bowker et al., 2009; Dear, 1995, 2007; Porter, 1996; Hacking, 1982).

Metrics and evaluation practices have had utility in the governance of public science in the US since the latter half of the twentieth century (Porter, 1996; Adams, 2016; Hacking, 1982). Statistics in science are key reservoirs for documenting activities and maintaining models of innovation (Godin, 2006). While sovereigns have long derived power from their patronage of the production of knowledge, large-scale science and technology projects emerged from and helped shape modern states (Ezrahi, 1990; Jasanoff, 2005; Shapin and Shaffer, 1985). For example, the epistemic virtue of transparency, prized throughout modern science, has attained a comparably prized status as a political value in technologically advanced democracies (Daston and Galison, 2007). In the postwar US, the Administrative Procedure Act in 1946 mandated an expanded array of federal agencies to make their rules and procedures public information (Jasanoff, 2006).

Following World War II, publicly funded science grew in scale and scope. The National Science Foundation, National Aeronautics and Space Administration, and other science and technology agencies were established, and NIH expanded. These bodies were largely governed by what science policy scholars have termed 'the social contract for science' (Guston, 2000). This tacit agreement granted science significant autonomy on the basis that it was a means for securing health, economic prosperity, and national security. Prominent scientists, such as Vannevar Bush and Michael Polanyi, sought to demarcate science and politics, partly by delimiting a sphere of 'basic research' that could not be easily measured by its instrumental value (Bush, 1945; Polanyi, 1962). Well into the 1970s it was possible to seriously ask, "Can science be measured?" (Holton, 1978: 39). The case for science as a self-governing enterprise was bolstered by Robert Merton's sociological articulation of a normative structure for science (Merton, 1942).

While subsequent generations of STS scholars challenged the notion that modernist distinctions between science and society could be maintained – even conceptually (Latour, 1993) – the political feasibility of attempting such a division was met with public challenges in the 1960s and 1970s. A series of scandals cast doubt on the notion of science as self-governing, leading Congressional patrons of research to renegotiate these lines (e.g., Greenberg, 1967, 1999). Scrutiny was most directed toward areas of applied science, and especially biomedical science. One consequence of this redrawn social contract for science was endowing Congress with expanded mechanisms for monitoring and incentivizing science around forms of productivity (Guston, 2000: 5). In the biomedical sciences, the desired product was increasingly articulated as interventions that moved from the laboratory bench to the clinical bedside.

The most immediate precursor to CTSAs were General Clinical Research Centers (GCRCs), some of which successfully applied to become CTSAs (Nathan and Nathan, 2016; Senate Committee on Appropriations, 1959). Rather than focusing on a specific disease or organ system, the GCRCs were designed to provide general infrastructure for investigator-driven clinical research in hospitals and eventually out-patient settings (Nathan and Nathan, 2016; NIH, 2009a). When he assumed the role of NIH Director in 1975, Donald S. Fredrickson noted there was a "translational gap" in delivering clinical results from biomedical studies, stating a desire to "extend the continuum of biomedical research across preceivable [sic] gaps in translation" (Fredrickson, 1975: 6; also quoted in Guston,

2000: 127). In 1978, the Office of Medical Applications of Research was established to influence clinical practice (Olszewski, 2018).

Policy efforts, including technology transfer practices and policies epitomized by the Bayh-Dole Act of 1980, sought to dissolve lines between university and market. The law incentivized federally-funded research undertaken at universities to be transferred into commercial products (Berman, 2012). New political and commercial charges for assessment and faster discovery transformed the biosciences (Clarke et al., 2003; Parthasarathy, 2007; Rabinow and Bennett, 2012; Rose, 2006; Sunder Rajan, 2006). Economic metrics became principal measures of productivity and were explicitly incentivized. For example, the Federal Technology Transfer Act of 1986 required federally supported labs to "ensure that efforts to transfer technology [to the commercial market] are considered positively in laboratory job descriptions, employee promotion policies, and evaluation in job performance of scientists and engineers" (15 US Code § 3710, cited in Guston, 2000: 125). 'Boundary organizations' designed to reduce 'gaps' or 'lags' in translation, such as the Office of Technology Transfer at NIH, were established at the border of science and politics in furtherance of this revised vision (Guston, 2000). Relatedly, evaluation science was professionalized as a new form of technical expertise during this period (Madaus et al., 1983).

Evaluation science demands forming normative value judgments, often to identify gaps and achieve greater efficiency. Similar to projects of standardization, evaluation is both an epistemic and political undertaking not only because it is often directed toward judging outcomes of policies and providing evidence for policymakers but because the approach itself adopts and enacts normative, normalizing, and political stances (Elzinga, 1988; Timmermans and Epstein, 2010). As Carol Weiss has noted, evaluation science "makes implicit political statements about such issues as the problematic nature of some programs and the unchallengeability of others, the legitimacy of program goals and program strategies, the utility of strategies of incremental reform, and even the appropriate role of the social scientist in policy and program formation" (Weiss, 1973: 94). In 1970, Congress amended the Public Health Service Act to include 1% of funds for program evaluation (Guston, 2000: 79); at NIH, these activities were administered by the Program Evaluation Branch in the Office of the Director (US Office of Technology Assessment, 1986). As evaluation science developed, evaluations of federally funded programs were also becoming standard practice and a feature of science policy.

Audit cultures present in business, government, and academia magnified as these spheres overlapped in areas of science that explicitly sought to translate scientific work to applications in society (Strathern, 2000). As an ideology of the market took hold in science policy as in other areas of governance in the 1970s and 1980s, pressure grew to assess the desired productivity of science against economic outcomes. Financialization emerged as one dominant approach for managing increased public scrutiny over the dividends science could deliver (Robinson, 2019; Murphy, 2017; Hallonsten, 2021; Whitacre, 2024). Links to the financialization of biomedicine, particularly following the 2008 financial crisis, grew as translational science emerged and framed itself as an effort to center the risk of early-stage research on universities, thus 'de-risking' discovery prior to later-stage commercialization (Robinson, 2019). Increasingly, commercialization became a key metric for translational research, thus shifting the responsibilities of investigators (Jeske, 2022; Lander and Atkinson-Grosjean, 2011; Robinson, 2019).

Translational science was thus a product of a particular historical conjuncture marked by key themes in 20th and 21st century science policy. Following Stuart Hall, we understand a historical conjuncture as "not a slice of time, but…defined by the accumulation/condensation of contradictions, the fusion or merger – to use Lenin's terms – of 'different currents and circumstances'" that define an era (Hall, 1988: 130). Key historical currents that shaped translation science have included: (1) the expanded role of publicly funded science after World War II; (2) growing political pressures to evaluate federal programs; and (3) a renegotiated 'social contract for science' that emphasized measuring returns on public investment. Although translational policies at NIH span

back to the 1950s and intensified in the 1970s, the financialization of health research around the turn of the 21st century marked a distinctive shift in not just seeking to bring discoveries to market but embodying the logic of markets in the doing of science and making of technologies, including in publicly-funded scientific research (Robinson, 2019; Whitacre, 2024). In what follows, we track how metrics sought to embody a new kind of 'evaluative way' that staked the value of translational science on its ability to measure, assess, and represent its impacts and 'returns on investment' in ways that reflexively and scientistically attested to its value on terms legible to markets, policymakers, and others. As David Guston notes regarding previous shifts in how science is evaluated, "Getting used to oversight, measurement, and incentives is a difficult task for both politicians and scientists. It requires different institutional forms, different types of information, and different kinds of behaviors" (Guston 2000: 6).

The evaluative way and its milieu

We characterize the centrality of metrics and the formation of cultures of evaluation in translational science as the emergence of an 'evaluative way' that drove the field's development and helped establish its legitimacy. This evaluative way ultimately helped bring translational science into the world through fostering everyday socio-technical practices and representations of translational science generated using evaluation data. The evaluative way names the epistemic underpinnings, policy paradigms, institutionalized practices, and socio-technical infrastructures that became embedded in initiatives to measure the impact of the CTSA program, which often stood as a proxy for assessing the overall success of translational science.

The evaluative way in translational science is what Geoffrey C. Bowker and colleagues (2009) call a 'way of knowing' particular to the socio-technical milieu, epistemic habitus, and political economy from whence it sprung (see also, Dear, 1995). Translational science emerged in the late 2000s and early 2010s – a heady time when the informational, organizational, and policymaking infrastructures that made science possible were also undergoing substantive transformation at vast scales. These transformations tracked along other major changes in how science and society were co-produced through information systems and neoliberal forms of economization (Çalışkan and Callon, 2009, 2010; Castells, 2010; Jasanoff, 2004; Peterson and Panofsky, 2023; Ribes, 2014; Ribes and Bowker, 2008; Ribes and Finholt, 2009). In the biosciences, this milieu involved developments ranging from the emergence of 'big data' systems and paradigms; the rise of genomics; pushes at NIH to fund multi-institutional collaborations and public-private partnerships; a generalized *ethos* of 'innovation' and 'disruption' which marked a potent orientation toward technology and capital; the growing influence of the open science movement; and newly routinized uses of technologies designed to facilitate collaboration such as videoconferencing software, the 'webinar' as a genre of event, and document-sharing platforms (Ribes, 2014).

The historical conjuncture translational science emerged from also involved shifts in policy requirements pertaining to the evaluation of federal programs, including through the diffusion of 'continuous quality improvement' practices and the passage of laws connected to evaluation requirements such as the NIH Reform Act of 2006 and the Foundations for Evidence-Based Policymaking Act of 2018. These developments reflected the rise of managerialism in the sciences as well as what feminist critics have characterized as the 'corporate university' (Deem, 2001; Greene, 2022; Nash and Owens, 2016: vii-xi). We thus understand the evaluative way in translational science as part of a more general trend related to notions of efficiency, economization, continuous improvement, evidence-based practice, datafication, and related developments that were becoming pervasive at this point in history. The concept is thus relevant for scholars studying myriad aspects of late 20th and early 21st century developments in state, industry, and academia beyond translational science.

The evaluative way in translational science involved developing and disseminating a shifting set of strategies, guidance documents, software packages, and technical assistance materials for actors within the CTSA hub network to plan,

gather data, and self-organize to meet metrics. By enabling actors to generate representations that attested to the impact of translational science, fulfilling evaluation metrics became central to cementing it as a legitimate science on the terms of 21st century science policy, rather than a fad or as merely epiphenomenal to widespread innovation 'hype' (see, Vinsel and Russell, 2020).

Key features of the evaluative way include its (1) *Epistemic underpinnings* (e.g., foundational scientific concepts, motivating ideas, and related notions); (2) *Supporting policy paradigms* (e.g., key pieces of legislation and policy frameworks); (3) *Institutionalized practices of evaluation* (e.g., methods and cross-CTSA knowledge-sharing strategies); (4) *Socio-technical infrastructures to support evaluation* (e.g., evaluation 'clearinghouse' websites, conferences, trainings, and interest groups within professional societies); and (5) *Representations of translational science desired by proponents* (e.g., as innovative, as collaborative, as providing worthwhile returns on investment). Together, these five features of evaluation paradigms developed for translational science through the CTSA program led to metrics becoming increasingly central to representations of translational science's impact and to the field's social reproduction. Supplementary Table 1 (available at: https://perma.cc/R9BY-2X46) presents key features of the evaluative way along with demonstrative examples of supportive paradigms.

In elucidating how the evaluative way took shape, we also argue that STS scholars studying the emergence of new sciences in the late 20th and early 21st centuries should focus on the increasing importance of evaluation practices in those fields' development. This was a period when myriad features of the global economy, administrative policy, audit cultures, data-intensive practices, information-technology infrastructures, and US federal policies requiring evaluation intensified. This historical conjuncture created conditions of possibility for new ways of knowing to emerge, including ones which positioned evaluation as a legitimizing force through its ability to demonstrate the 'value' of an enterprise (Doganova et al., 2014; Vatin, 2013). Translational science – and the CTSA program particularly – is one area where these potent and consequential styles of scientized evaluation took center stage, deeply shaping the field's ambitions to transform processes of discovery, dissemination, and commercialization in the health sciences (Austin, 2018; Robinson, 2019; Zerhouni, 2005).

Periodizing the evaluative way

We present a five-phase periodization in which the evaluative way in translational science developed through the CTSA program. The periodization is reviewed in brief below, with each period being described in further detail in later sections. We also represent the periodization in Supplementary Table 2 (available at: https://perma.cc/68CM-D3SE), which shows the shifting role that metrics played in the production of translational science along with the forces driving shifts in evaluation practices.

Several periods overlap, corresponding with moments of controversy described below. The first lasted from 2005-2011, when the CTSA program was funded and overseen by the National Center for Research Resources (NCRR) within NIH. During this time, evaluation metrics were not standardized for individual CTSA hubs and NCRR mainly required CTSA awardees to engage in unspecified forms of self-evaluation and to participate in a national evaluation of the CTSA program contracted to the consulting firm Westat (NIH, 2005, 2007, 2009b; Frechtling et al., 2012). The second period extended from 2010-2012. In 2010, a controversy was set in motion over the planned closure of NCRR and the shift of the CTSA program to a new center for translational science within NIH, which would occur when NCATS was established in 2011. In 2011 – in response to critics of the new NCATS – Congress commissioned an Institute of Medicine (IOM) report to build a new system of metrics for the CTSAs "to ensure the benefits of this investment are maintained," reflecting the market-driven logic of 21st century science policy (US Congress, 2011: 1137).

The third period, 2012-2014, was a time of flux constituted mainly by fallout from this controversy and the standardization of CTSA metrics. The IOM produced its report recommending "common metrics" for the CTSA program, and in 2014, NCATS announced the new "Common

Metrics Initiative" (IOM, 2013: 5-6, 82-89; Patel et al.: 61, 2019; see also, Rubio, 2013; Rubio et al., 2015) This led to the fourth period, from 2015-2021, when NCATS established and implemented the Common Metrics Initiative across the CTSA hub network. As we describe further below, implementing the CTSA Common Metrics Initiative involved a federated approach that utilized extensive training and created new cultures of sharing evaluation strategies across the CTSA hub network and within individual hubs. The Common Metrics Initiative implementation process thus led to a diffusion of both localized and cross-hub practices that resulted in a shared set of social norms, technical approaches, and routinized practices for CTSAs to self-evaluate that were emic to the institutional and organizational structures of individual CTSAs as well as etic and determined by national Common Metrics Initiative requirements set by NCATS.

The fifth period began in 2021, when the Common Metrics Initiative ended. Evaluation functions were largely devolved to CTSAs and translational science teams. Documents announcing this transition spoke positively about the culture of evaluation that had been created during the Common Metrics Initiative era. Building this culture had enabled CTSAs to transition from a set of limited national 'common metrics' to evaluation strategies grounded in local practices of 'continuous quality improvement' and other evaluation models developed specifically to assess innovations in translational science. Varied forms of continuous and ongoing evaluation had become part of the fabric of doing translational science, supporting translational science activities, and representing the field to external parties (NIH, 2021). In becoming central to the sociomaterial life of translational science, evaluation practices and evaluative ways of knowing came to constitute and enact it.

Enduring areas of evaluative focus and key examples

Despite changes over time in how the CTSA program pursued evaluation, the field's overarching areas of evaluative focus remained somewhat consistent across time. Changes in evaluation strategies can thus be understood as both responsive to external pressures and reflective of debates among evaluation stakeholders within the CTSA program about how to best measure these enduring areas of evaluative focus. This difficult task yielded radically different overall approaches to CTSA evaluation across the five periods.

We identified nine key enduring areas of evaluative focus across the CTSA program's history. They were: (1) The overall impact of translational science on biomedical innovation; (2) The scientific career in translational research; (3) Research productivity among translational researchers; (4) Lessening regulatory burden for conducting translational research; (5) Collaboration to support translational research; (6) Community engagement; (7) Demographic diversity in the translational science workforce; (8) Social and economic impact of translational science; and (9) Establishing a unified metric or 'meta-metric' that could be used to evaluate a range of translational science initiatives and attest to the field's impact. We also present each enduring area of evaluative focus in Supplementary Table 3 (available at: https://perma.cc/7TFJ-4FMC), where the middle column shows how each was treated during the Common Metrics Initiative era, when evaluation activities were most centralized and standardized.

The evaluation of the CTSA program began as a set of localized practices and a unified national evaluation study conducted by outside evaluators. CTSA evaluation then became standardized and federated through the Common Metrics Initiative. What ultimately emerged after the Common Metrics era was a transformed field of translational science that centrally emphasized the further development of the very styles of evaluation that had helped bring it into existence and legibility as a science. However, across all five periods, these nine enduring areas of evaluative focus remained core preoccupations. Underlying these areas of focus was the belief that accurately measuring them would generate representations and understandings of translational science's impacts on biomedical innovation, health, and society.

We now describe three key examples of major innovations developed by NCATS that were promoted as exemplifying the potential impact of the CTSA hub network on the basis that each led to improvements across several of the enduring

areas of evaluative focus described above. Further discussion of these examples is presented in Supplementary Table 4 (available at: https://perma.cc/P7RW-GGX3), with the left column providing additional information about each case and the right column showing corresponding enduring areas of evaluative focus described in the previous section and Supplementary Table 3.

The first example is the Streamlined, Multisite, Accelerated Resources for Trials (SMART) IRB initiative. NCATS described SMART IRB as a "platform" to implement the NIH single IRB (sIRB) policy, which went into effect in 2018 (NIH, 2018; Vardeny et al., 2019). sIRB aimed to centralize the ethics review for non-exempt multisite clinical trials which would streamline the IRB process. Under sIRB, one institution's IRB would serve as the "IRB of record," enabling other trial performance sites' IRBs to rely on its determinations instead of conducting their own reviews. SMART IRB provided a master agreement and resources for institutions to implement sIRB. SMART IRB was distributed through the CTSA hub network and evaluated by CTSA-affiliated evaluators, who noted its role in smoothing the implementation of sIRB and hastening the pace of multisite trials research (Green et al., 2023). As of 2025, over 1300 institutions were signatories to the master SMART IRB agreement – "including all CTSA hubs" (SMART IRB, n.d.).

The second example is the Trial Innovation Network (TIN), established in 2016 (NCATS, 2024b; Shah et al., 2017). TIN was launched as an NCATS and CTSA program-led initiative that sought to create new approaches for the design and conduct of multicenter randomized control trials (mRCTs). As of 2025, TIN was housed at two CTSA-funded universities – Johns Hopkins and Vanderbilt (Palm, Edwards, et al., 2023). TIN served to disseminate knowledge and expertise about innovative trial designs through trainings and hosting a resource bank (Harris et al., 2023; Trial Innovation Network, n.d.). The TIN also provided resources for trainees and was evaluated on how it impacted a range of factors related to the efficient conduct of trials and its utilization of the CTSA hub network to accomplish its goals (Palm, Edwards, et al., 2023; Palm, Thompson, et al., 2023).

The third example is the National COVID Cohort Collaborative (N3C), which was started in 2020 and grew out of an earlier NCATS-supported initiative called the National Center for Data to Health (CD2H), established in 2017 (CD2H, n.d.[a]). From its onset, N3C has been anchored by CTSA-funded institutions. It aggregated electronic health record (EHR) data about people diagnosed with COVID-19 to facilitate large-scale research utilizing "real-world" data (Haendel et al., 2021). N3C was evaluated mainly on the quality, speed and volume of its research outputs, and for its success as an example of a new kind of data-sharing initiative in translational science that was facilitated by the CTSA hub network (Sidky et al., 2023). N3C was regarded as so successful that, in late 2024, it was renamed that National Clinical Cohort Collaborative (retaining the "N3C" acronym), with COVID-19 becoming one "data enclave" alongside pilot cancer and renal enclaves that aimed to replicate N3C's success with COVID-19 across other clinical domains (NCATS, 2024a).

Each of these examples shows how the CTSA program has been promoted by NCATS as a vehicle for driving innovation in translational science, and how innovations have been evaluated in alignment with the major areas of evaluative focus presented in the preceding section and Supplementary Table 3. The examples show translational science's evaluative way in action. Each is a translational science innovation pursued through the CTSA program that was evaluated for its impact on the conduct of scientific research and only then deemed worthy of continued investment.

The emergence of the evaluative way

We now offer detailed accounts of events that led to our periodization. We focus on key moments of controversy, continuity, and change in the history of evaluation in the CTSA program, while also emphasizing enduring areas of evaluative focus across the periods.

Period 1: Metrics-in-formation: Pre-NCATS metrics (2005-2011)

From 2005-2011, when the CTSA program was administered by NCRR, awardees were required to engage in self-evaluation and to participate in a national evaluation study contracted to Westat (NIH, 2005, 2007, 2009b, 2012; Trochim et al., 2013). This structure led CTSAs to establish "evaluation cores" as one of several "core" and "key function" formations along with ethics, community engagement, biomedical informatics, and others (NIH, 2005; Patel et al., 2019: 60; Trochim et al., 2013: 304-6). Trochim et al. (2013) approvingly stated that, "unlike most other large initiatives of the NIH, the CTSA initiative included evaluation efforts at its outset" (Trochim et al., 2013: 304). The centrality of evaluation was thus written into the CTSA program from its inception, making the practice critical to the new enterprise of translational science while also laying the groundwork for the emergence of new styles of evaluation particular to it. From the earliest phases of the CTSA program, evaluation efforts centered on assessing the value of federal investment in this new area of the health sciences, which made bold promises about transforming biomedical innovation (Robinson, 2019; Zerhouni, 2005). Those claims would later be met with skepticism, prompting shifts in CTSA evaluation practices.

The initial report of the national CTSA evaluation by Westat was released in 2012 (Frechtling et al., 2012). It noted that "From its beginnings, evaluation has been an integral feature of the CTSA program. Evaluation occurs at multiple levels in multiple formats that include local institutional data collection and information tracking, monitoring by NCRR and other NIH program officers (project collaborators), and program-level evaluation by external evaluation experts" (Frechtling et al., 2012: 2). The report spoke positively about the program's achievements, while also noting that its evaluation was preliminary and "designed both to provide baseline data for tracking CTSA program accomplishments over time and to lay the groundwork for continued evaluation" (Frechtling et al., 2012: 2). One of the report's main recommendations was for CTSAs to "Conduct longer term evaluation," guidance under which included directives to "Work with local evaluators to develop metrics and measures and carry out evaluation studies" and to "Focus on establishing standard metrics, shared data collections, and coordinated research studies that can be aggregated to create program-wide information" (Frechtling et al., 2012: xii). Under this recommendation, the report also encouraged the CTSA program to "Establish an evaluation advisory group that includes representatives from the PIs, local evaluators, IC leadership, professional associations, community groups, and outside evaluation experts" (Frechtling et al., 2012: xii). This would later come to pass.

Notably, the 2012 report stressed that its evaluation did not bear directly on whether the CTSA program improved health outcomes but rather spoke to accomplishments and challenges in how the translational science "movement" sought to transform the system of health research (Frechtling et al., 2012: 30). The report aimed to offer "an early assessment of whether, and in what ways, the program has succeeded in enhancing institutional capacity for conducting clinical and translational research and advancing scientific knowledge (Frechtling et al., 2012: vii, emphasis in the original)". Westat framed the complexity of such an undertaking as being beyond its capacities as an evaluator, noting that "evaluating the CTSA program is a challenge, not only because of its ambitious goals but also because of its many moving parts and the flexibility afforded the participating institutions" (Frechtling et al., 2012: vii). The diversity of implementation sites prevented the creation of shared outcome measures from the onset. Therefore, the authors emphasized the importance of understanding local practices developed by the CTSAs, rather than tracking standardized and predetermined metrics, noting that "many of the changes associated with the CTSA program were not *initiated* by the program, but rather were *encouraged* and *facilitated* by the infrastructure it provides" (Frechtling et al., 2012: vii). This tension between top-down and bottom-up evaluation strategies would structure subsequent developments.

Period 2: Renewed demands for metrics: The contested establishment of NCATS (2010-2012)

From 2010-2012, the transition of the CTSA program from NCRR to the proposed new NCATS produced substantial controversy that was caught up in broader efforts to transform NIH in the wake of the 2005 *NIH Roadmap* and NIH Reform Act of 2006 (Zerhouni, 2003, 2005). This controversy was organized around determining what the central aims of NCATS ought to be, and whether it should wholly replace NCRR (Vastag, 2011). Conflicts played out among actors in Congress, regulatory panels at NIH, science journalism, and peer-reviewed literature. The outcome led to shifts in the structure and focus of CTSA evaluation activities through the creation of initiatives that aimed to demonstrate its value in the form of 'return on investment' to taxpayers. Resolution was reached in the form of an Institute of Medicine (IOM) report commissioned by Congress in 2011 and published in 2013 (IOM et al., 2013). This process ultimately further ensconced the construction of increasingly scientistic metrics to assess the impact of the CTSA program as being central to the ongoing development, validity, and success of translational science.

The conflict over the transition of the CTSA program to NCATS bore on three questions. The first was whether NCATS was established too hastily and non-transparently in an aura of hype. The second was whether a new center dedicated to translational medicine was worthy of massive federal investment, given that its promises to hasten innovation had yet to be demonstrated. The third was whether establishing NCATS should necessarily mean closing NCRR, which some saw as a vital resource that NCATS would not fully replace. The main result was the closure of NCRR and the transition of the CTSA program to NCATS.[3] However, to carry out these decisions – and in part to manage fallout among those who were displeased – Congress commissioned the 2011 IOM report to help resolve questions about the value of the CTSA program.[4] One of IOM's primary charges to NCATS was to develop evaluation strategies to measure the impact and return on investment of the CTSA program. We now describe this highly consequential controversy in detail.

Wheels had been set in motion to close NCRR and transition its functions to NCATS in July 2010, when an entity called the Translational Medicine and Therapeutics Working Group (TMAT) was established (NIH Scientific Management Review Board, 2010b). TMAT was created as part of a review initiated by an entity within NIH called the Scientific Management Review Board (SMRB). SMRB had been established by the NIH Reform Act of 2006 and was charged with evaluation functions, of entities within NIH including "[r]eviewing the research portfolio of the [NIH] in order to determine the progress and effectiveness and value of the portfolio and the allocation among the portfolio activities of the resources of NIH" (United States Congress, 2007). In addition to fostering greater efficiency in resource allocation, SMRB was charged with supporting transparency, robust decision-making processes, and the evaluation of key decisions during NIH reform.

On December 7th, 2010, SMRB voted 12-1 to accept the recommendation to transition the CTSA program to a new entity – which would become NCATS – and to close NCRR (SMRB, 2010a). However, as described in a December 2010 SMRB report, the board desired better metrics that would provide evidence about the effectiveness of translational science. SMRB thus "endorsed and supported the NIH's commitment to undertake a more extensive and detailed analysis through a transparent process to evaluate the impact of the new Center on other relevant extant programs at NIH, including NCRR" (SMRB, 2010a: 2). The lone dissenter in the SMRB vote to establish NCATS was Jeremy Berg, then-director of the National Institute for General Medical Science (NIGMS) (Vastag, 2011). According to an article in *Science* that characterized Berg as an actor at NIH whose "role switched from model citizen to dissident," he "rushed from the room after the 12-1 vote, visibly upset" and later resigned from his positions at NIH in the wake of what he characterized as a rushed decision to close NCRR (Kaiser, 2011b: 533). Berg, a biochemist by training, left NIH shortly thereafter. He indicated in an interview with *Science* that "This is the first time in the history of NIH that a center has been abolished, and to do it in this manner is a very bad precedent" (Kaiser, 2011b: 533). While providing a full account of this contro-

versy is not possible here, the shuttering of NCRR also generated bipartisan pushback in the Senate and House (Basken, 2011; Kaiser, 2011a, 2011c, 2011d; Wadman, 2011).

Regardless of the merits of how NCATS was established, its creation was inseparable from discourses of innovation and disruption that were ubiquitous in the 2000s and early 2010s, which were driven partly by wide-scale transformations in the digital economy, genomics, and related areas that shaped what has been characterized as the neoliberalization of science (Bowker et al., 2009; Hedgecoe and Martin, 2003; Robinson, 2019; Rose, 2006). Maintaining momentum around NCATS was also a priority of NIH leadership and was framed within a broader national strategy of keeping the US a global leader in biotechnology. In 2012, then-NIH Director Francis Collins gave testimony to a committee of the US House of Representatives that addressed the speed at which NCATS had been created. He invoked laudatory remarks about biomedical innovation made by then-recently deceased Apple CEO Steve Jobs, as the latter was experiencing terminal liver failure. Collins admonished committee members to "please bear Jobs' words in mind: today technological advances are driving science" (Collins, 2012: 4). In a Senate Appropriations Committee hearing in 2011, Collins argued for the country to invest in translational science via NCATS to maintain its competitive edge, noting that the US "still is the world leader in science and engineering research. But that leadership role is being challenged by China, India, and other nations as they recognize the economic, health, and social benefits of investing in R&D" (Senate Hearing, 112th Congress, 2011: 28). Collins, a geneticist by training, had previously been head of the National Human Genome Research Institute (NHGRI) during the Human Genome Project.

The controversy over the decision to close NCRR and transition the CTSA program to the new NCATS led Congress to include language in a 2011 spending bill, which noted that Congress was "also aware of concerns that the NIH process for evaluating the merits of the NCATS reorganization did not comply with the NIH Reform Act of 2006 with respect to the role of the [SMRB])" (US Congress, House of Representatives., 2011: 1134). The bill called for IOM to generate a report about the CTSA program that included developing plans to more rigorously evaluate it. Invoking the idea of return on investment for taxpayers and the need for better metrics, the bill noted that "CTSAs now represent an investment of half a decade of innovation in translational research. To ensure the benefits of this investment are maintained, the conferees urge NIH to support a study by the IOM that would evaluate the CTSA program and recommend whether changes to the current mission are needed" (US Congress, House of Representatives., 2011: 1137).

The controversy was thus resolved by a political demand for the swift development of more robust impact metrics to ensure that taxpayer dollars allocated to translational science were determined by evaluators to be worthy financial investments. This outcome created urgency around the task of not only applying evaluation strategies to the CTSA program, but for the development of evaluation practices specific to CTSAs and translational science generally. The stakes of evaluating translational science were thus an existential matter for the field.

Period 3: The drive toward common metrics (2012-2014)

2012-2014 was a period of transition marked by efforts to consolidate CTSA evaluation. Major areas of controversy over how to measure the value of translational science were addressed through deliberative processes whereby key institutions coalesced around shared evaluation goals. This focalization of metrics proceeded through three overlapping processes: (1) the release of a series of CTSA evaluation reports by Westat covering 2005-2012; (2), the 2013 release of the Congressionally commissioned IOM report; and (3) NCATS's response to the IOM report in 2014. These initiatives focused on how the impact of the CTSA program ought to be evaluated in the long-term and would result in the creation of the Common Metrics Initiative in 2015.

IOM's report was assembled by a panel of stakeholders from the public sector, private sector, academia, and advocacy. It was designed to create an enforceable consensus about the future of the CTSA program. Results stemmed from meetings

of the panel, community and stakeholder engagement sessions, as well as reviews of the scientific literature, Westat reports, and other materials (IOM et al., 2013: 18-19).

Evaluation practices were a major focus, particularly the need for CTSA program leadership to develop "a set of common metrics that reflect the program's mission and strategic goals" (IOM et al., 2013: 86). A major recommendation was to "Formalize and Standardize Evaluation Processes for Individual CTSAs and the CTSA Program" (IOM et al., 2013: 8). This was envisioned as "a formal, programwide [sic] evaluation process… to measure progress toward NCATS's vision," while noting that "[e]lements that could be considered and been identified by NCATS staff…and by the IOM committee" (IOM et al., 2013: 88). The recommended potential common metrics reflect the nine enduring areas of evaluative focus shown in Supplementary Table 3 (IOM et al., 2013: 88). Finally, the IOM report recommended the development of "a more formalized evaluation process" that would build upon mechanisms already utilized by CTSAs (IOM et al., 2013: 89).

In its 2014 response to IOM, NCATS moved to implement the recommendations (NCATS Advisory Council Working Group on the IOM Report, 2014). Shortly thereafter, NCATS launched the Common Metrics Initiative. The 2014 CTSA funding call's section on "Evaluation and Continuous Improvement," noted that "NCATS will work with the successful applicants to develop specific operational plans to implement a set of common metrics across the hubs, building on existing efforts in the CTSA program" (NCATS, 2014). This was the beginning of a new period of centralization in CTSA evaluation. However, the Common Metrics Initiative would proceed in a federated manner whereby individual CTSAs would be trained by contracted entities and peer-to-peer learning infrastructures to collect data and meet common metrics in ways suited to their specific institutional contexts.

This federated approach to making evaluation part of the day-to-day work of translational science bears many hallmarks of 21st century US science policy, NIH reform, and the rise of marketization and managerialism in science, specifically though the need for continuous evaluation attesting to the "value" of programs to taxpayers (Grazier et al., 2013: 478-91). Implementing the Common Metrics Initiative required nearly every staff person or scientist who was involved in or funded by a CTSA to contribute to the project of setting goals and collecting data to meet common metrics. This ultimately led to the development of a new culture of evaluation across the CTSA hub network, involving thousands of personnel enacting new evaluation paradigms. As we describe next, the Common Metrics Initiative period is when 'the evaluative way' in translational science transmuted and solidified from being a specialized activity into a shared set of practices that spanned the field.

Period 4: Enacting common metrics: New standardized cultures of evaluation (2015-2022)

Implementing the Common Metrics Initiative from 2015-2022 was critical in solidifying the centrality of evaluation practices in the everyday work of translational science. The process drove the development of a culture of evaluation across the CTSA network that would make evaluation central to many CTSA activities, ranging from planning to execution and then monitoring progress toward goals.

NCATS established three national common metrics: (1) IRB Review Duration; (2) Pilot Funding Publications; and (3) Careers in Clinical and Translational Science. Each of these related to a core mission of the CTSA, which was to make the biosciences more efficient and to support scientists in pursuing translational careers. Other common metrics were piloted, but only these three were ever fully implemented.

CTSAs were required to participate in training programs and to refashion their work and priorities by evaluating and building plans to improve their existing local activities to meet the common metrics (Daudelin et al., 2020; Tufts, 2018). While national common metrics were set by NCATS, they were enacted through a federated process whereby their implementation was guided and overseen by NCATS-funded entities and hired consultants, but wherein individual CTSAs were able to construct specific implementation plans suited to their particular institutional needs. One

set of Common Metrics Initiative evaluators at the Tufts CTSA hub likened the national CTSA network to the US-style federalism, noting that "the decentralized organization of federal public health programs mirrors the CTSA Consortium. Although all CTSA hubs strive toward the same mission of catalyzing the clinical and translational research enterprise, each hub has autonomy to develop the approach and processes that are most effective in its local context" (Welch et al., 2021: 1-2). Because each CTSA engaged in different kinds of translational initiatives that were built around local strengths and needs, each required evaluative flexibility. Therefore, while national common metrics could be set, a uniform approach to fulfilling them could not be. However, all CTSAs were required to use the same method of evaluation.

The methodological framework used by NCATS to implement the Common Metrics Initiative was called "Results-Based Accountability" (RBA). RBA was derived from a 2005 book by consultant Mark Friedman titled *Trying Hard is Not Good Enough: How to Produce Measurable Improvements for Customers and Communities* (Friedman, 2005). RBA was exclusively licensed in the US to a firm called Clear Impact. Clear Impact assisted NCATS and its contracted Common Metrics Initiative evaluators and trainers at the Tufts University and University of Rochester CTSA hubs in implementing the Common Metrics Initiative and evaluating the national implementation process. Indeed, the Common Metrics Initiative – a program designed to facilitate evaluation – was itself also an object of evaluation (Lee, 2016). The use of RBA in the Common Metrics Initiative is a key example of how the CTSA program sought to transform biomedical research through marketization practices, chiefly by introducing approaches from business research and development into publicly funded science and its evaluation.

Implementing the Common Metrics Initiative required essentially all CTSA personnel or individuals funded by a CTSA to actively engage in the work of evaluating their CTSAs, to become trained in RBA approaches and software tools, and to coordinate with other CTSAs and national entities set up to assist in gathering data to support evaluation processes aligned with the national common metrics. According to a 2018 report by the Tufts CTSA, "this initiative also was seen as an opportunity to develop, demonstrate, and disseminate methods of improving 'the science of doing science,'" positing evaluation of translational science as being part-and-parcel of its improvement (Tufts, 2018: 2). Thus, at the end of this period, a science and culture of evaluation emerged that was emic to the field of translational science and the infrastructures provided by the network of CTSA hubs. A set of processes were enacted beyond the simple application of evaluation practices developed in other domains, such as RBA, to translational science. Additionally, a new way of doing evaluation was created specifically to support, measure, and generate representations of translational science within the culture and infrastructure of the national network of CTSAs (Daudelin et al., 2020; Patel et al., 2019).

Enacting the Common Metrics Initiative was thus an extremely complex undertaking that induced major shifts in the culture and practice of evaluation across the CTSA network that made *evaluating translational science* more central to *doing translational science*. Evaluation was no longer an external practice or an activity done by 'cores' comprised of a subset of evaluation experts within CTSAs. Rather, meeting common metrics became a network-wide enterprise that all CTSA stakeholders had to contribute to within the RBA framework. According to slides presented to CTSA leaders by the Tufts CTSA, "We are implementing the Common Metrics collaboratively. Together we will learn how to ensure the usefulness of the Common Metrics for maximizing the impact of the CTSA Program" (Lee, 2016: 8). Thus, while the main three common metrics were straightforward, the processes by which hubs would need to collect data to report on them for their institutions was so complicated that it required the construction of a new culture of evaluation specific to the national CTSA network, as well as fusing the RBA model with more longstanding evaluation practices recognized within the field.

In a model combining ideologies of neoliberal efficiency, scientific managerialism, and continuous improvement derived from business supply chain and research and development practices, RBA requires organizations to set goals based

on their current trajectories and then to develop 'turn-the-curve' plans. 'Turn-the-curve' plans aim to change the course of a particular area of activity toward more favorable outcomes, thus 'turning' the trajectory of a projected downward curve upward, toward improved 'return on investment' (See Figure 1.).

Each CTSA hub was required to create teams with five roles dedicated to implementing the common metrics: (1) Project champion; (2) Project Leader; (3) RBA Framework Lead; (4) Scorecard Software Lead; and (5) Metrics Topic Lead (Daudelin et al., 2020: 17). These personnel and other hub leaders were also required to participate in national Common Metrics Initiative working groups, to become trained in specialized 'scorecard' software developed for recording progress toward common metrics, and also to work with national consultants contracted by NCATS to guide implementation at their sites (Daudelin et al., 2020). This process, implemented differently at each CTSA hub with national oversight, created infrastructures to enable new forms of evaluation and a culture of evaluation in translational science that marked a clear break from the pre-Common Metrics Initiative era. In the earlier period, evaluation was mainly a function of largely autonomous evaluation cores within CTSAs and of external evaluators such as Westat. In contrast, the introduction of RBA and the process of constructing and fulfilling individualized 'turn-the-curve' plans at specific CTSAs as part of the Common Metrics Initiative marked the arrival of a new kind of evaluative way into the everyday socio-material life of translational science.

One major cross-hub evaluation initiative called *Insights to Inspire* (I2I) exemplifies the federated process by which the Common Metrics Initiative was implemented. Common Metrics Initiative implementation consultants at the University of Rochester CTSA led the I2I program in 2020 to promote the RBA framework and offer guidance to CTSAs. I2I was designed to disseminate results of the implementation process for the career development common metric through blog posts, webinars, and other means, to promote best practices (Ruiz, et al., 2022a). Per a report on I2I, the initiative's main result was a framework for constructing what they characterized as a 'community of practice' in the evaluation of translational science initiatives (Ruiz et al., 2022a: 6). These authors noted that "This cross-institution collaboration could result in changes of internal processes or policies thus potentially improving the metric quantitative data" (Ruiz et al., 2022a: 6). They further described this process as creating a "framework for communicating and disseminating actionable intelligence" which they claimed "has the potential of replicability to other national evaluation and research networks," and said that "These efforts are in alignment with the NIH-NCATS recently released award, PAR-21-293, focused on the use of continuous quality improvement" (Ruiz et al., 2022a: 6).

The impact of the Common Metrics Initiative in fostering new cultures and infrastructures of

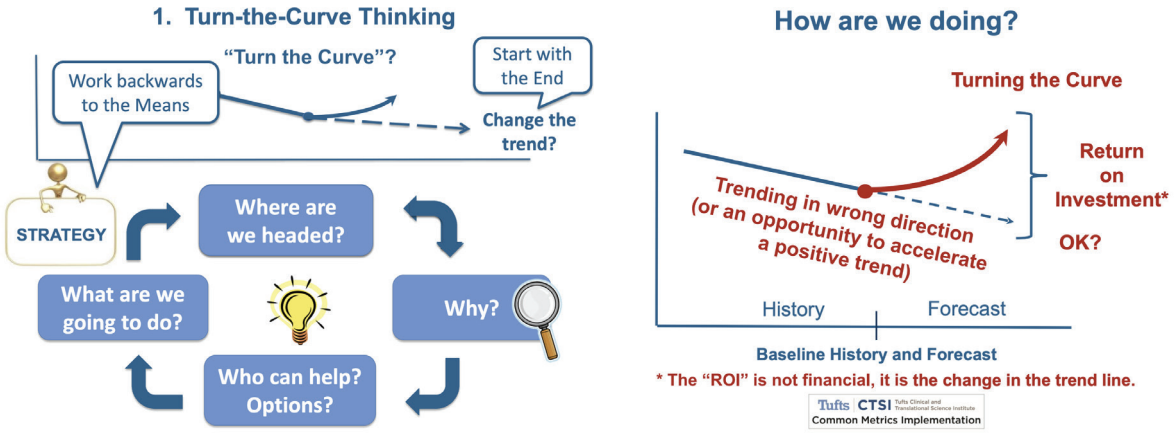

Figure 1. Two excerpts from a Common Metrics Initiative "Turn-the-Curve" training slide deck. Slides excerpted from Lee (2016).

evaluation across the CSTA program was echoed in the 2021 announcement from NCATS that the Common Metrics Initiative would end in 2022. NCATS described the program's implementation as a success because it "creat[ed] a culture across the [CTSA] Program of using data to leverage change and sharing unique approaches and best practices to enhance metrics" (Kurilla, 2021: 1). The perceived success of the Common Metrics Initiative was thus characterized partly through the claim that creating new cultures of evaluation across the CTSA network was critical to the foundation for building sustainable evaluation strategies for translational science that could operate autonomously without centralized coordination around meeting specific national metrics. The future approaches envisioned would be capable of ensuring adequate 'returns' on federal 'investments' in translational science, which were defined using frameworks that fused business approaches with ones from traditional biomedical research in ways that reflected late 20th and early 21st century US science policy.

Period 5: Emic evaluation cultures after common metrics (2022-)

The period after 2022 was marked by a defederation and devolution of evaluative functions to individual CTSAs in a structure that somewhat resembled evaluation as practiced during the first years of the CTSA program. There would be no more national common metrics, partly on the basis that CTSAs were better positioned to self-evaluate after having been capacitated to do so through the multi-year Common Metrics Initiative. To this point, the socio-technical infrastructures and knowledge about evaluation practices that were created during the Common Metrics Initiative era would be leveraged to support a new evaluative regime. CTSAs were required to use 'continuous quality improvement' and related approaches for self-evaluation as well as other evaluation strategies that were suited to their contexts and justified in CTSA grant renewal applications (NIH, 2021).

The central paradox of this phase of evaluation in US translational science is that there was a broad consensus that the field still did not know how to evaluate its overall impact on biomedicine and society. Leaders were constantly attesting to this, while also praising progress made. One paper that reflected on metrics in the translational enterprise and reported the results of a workshop at an event called "Translate!2021" captured this sentiment. The authors simultaneously lamented the lack of quality metrics for translational science, lauded gains in evaluation techniques, and made an aspirational call for better impact metrics across scales, from benefits to individual patients to evaluating the whole of society:

> Simply installing new, translationally oriented metrics in an academic institution for measuring success may be insufficient and counterproductive…Translate!2021 participants supported patient and societal benefit as the ultimate impact against which translational success should be gauged. More precise details for establishing and asserting success metrics for these important goals must now be considered at the level of individual academic institutions. However, this definition also requires broader societal discussion among academic institutions and the diverse societal stakeholders that provide, pay for, and benefit from the products of translational research…This requires academic institutions to readjust expectations and measures of translational success (Duda et al., 2023: 6).

The call for better metrics was also reflected in proposals such as the Translational Science Benefits Model (TSBM), framed as a flexible "framework designed to support institutional assessment of clinical and translational research outcomes to measure clinical and community health impacts beyond bibliometric measures" that "includes 30 specific and potentially measurable indicators that reflect benefits that accrue from clinical and translational science research such as products, system characteristics, or activities" (Luke et al., 2018: 77). However, exact TSBM metrics are not pre-defined; rather, the model and others like it instead call for using it to further refine evaluation strategies to develop both translational science and the science of evaluating translational science using a variety of methodologies. This includes qualitative approaches to evaluation that can generate accounts of the value of specific initiatives in policymaker-facing and public-facing nar-

ratives, stories, and case studies (Huebschmann et al., 2024; Institute of Clinical and Translational Sciences, n.d.).

In sum, the end of the most successful metrics initiative in the history of the CTSA program – the Common Metrics Initiative – coalesced around yet another round of calls for more and better metrics to assess the impact of translational science on biomedical innovation and society. However, this was not simply a Freudian 'return of the repressed' or Nietzschean 'recurrence of the same.' It was a call for another renewal of the evaluative spirit, complete with evangelizing overtones. In this period, evaluators called for new or revised metrics that reflected and built upon changes in the underlying evaluation cultures and infrastructures within the CTSA program. Rather than suggesting an eventual endpoint where impact metrics and evaluation strategies would be settled, the nature and scope of evaluation metrics were set to continually develop and expand along with the field of translational science, helping to drive its development. The science of evaluation and the science of translation had become further yoked.

The TSBM, for example, was framed as building directly on the Common Metrics Initiative, and arguments about evaluation needing to play a role in demonstrating the value of translational science to taxpayers and varied other stakeholders are invoked throughout the framework. The TSBM authors note that it was designed to measure the benefits of translational science in "four categories. i) Clinical and Medical Benefits… ii) Community and Public Health Benefits…iii) Economic Benefits… [and] iv) Policy and Legislative Benefits" (Luke et al., 2018: 79).

The TSBM was just one such framework proposed and disseminated before, during, and after the Common Metrics Initiative era; others included the Translational Research Impact Scale (TRIS) and the use of 'evaluation scorecard' systems customized for translational science (Dembe et al., 2014; Kotarba and Wooten, 2017). These and other proposals reflected, on the one hand, the desire to utilize the new culture of evaluation within CTSAs to strengthen evaluation practices and, on the other hand, the longstanding aim of the CTSA program to fund "detailed self-evaluation plan[s] to assess implementation of the short-term and long-term CTSA goals, including implementing program activities and tracking trainees and scholars and their mentors, their pilot projects, and their involvement with multidisciplinary team research" (NIH, 2005).

From 2005-2022, the desire for metrics to measure the impact of translational science had not waned, nor had it been fulfilled. Indeed, the demand to use evaluation strategies to justify the existence and value of the field to its own practitioners, funders, and other external stakeholders had escalated and intensified. During this period, evaluation became increasingly scientized and coextensive with undertaking translational research endeavors. Intervening years had seen fundamental transformations in evaluation practices that had produced an evaluative way in translational science that was central to the field and involved a set of enduring infrastructures, practices, and areas of evaluative focus. There would apparently be no end to evaluation or the need to further develop evaluation strategies; the goal, rather, was to continue evaluating while also enacting better evaluative ways to enable the performance of translational science.

Conclusion

In the mid-2020s, fears emerged that NCATS could become a victim of its own success. In 2023, a coalition of current and former NCATS leaders, industry groups, civil society organizations, and patient advocates formed the political influence group "NCATS Alliance" to advocate for continued federal investment in the CTSA program and NCATS (NCATS Alliance, n.d.; Owermohle, 2023). This was done at least partly as a reaction to the creation of the multi-billion dollar Advanced Research Projects Agency for Health (ARPA-H) in 2022, which had been described supportively by then-NIH Director Francis Collins as "like NCATS on steroids," in reference to its goal of accelerating the development of translational innovations with broad applicability across domains (AAMC, 2021). Evidently, the translational science paradigm had become so dominant in biomedical research that key actors in translational science's core public institutions felt that the specific functions of NCATS and the CTSA program to be in

need of defending in policy fora and public-facing advocacy (Owermohle, 2023). As of early 2025, APRA-H was still in the process of fully standing up its functions; however, the evaluation of APRA-H-funded innovations through 'independent verification and validation' (IV&V) and 'return on investment'-focused economic impact studies paired with projects was central to several major calls for funding (ARPA-H, 2024; Sanghavi, 2024). In addition to promoting translational approaches in biomedical research, APRA-H has enacted and extended translational science's evaluative way.

We have described how attempts to evaluate translational science have been central in producing the field, attesting to its value, and making it legible as a science. The CTSA program is an exemplary case that demonstrates how translational science and its evaluation have become hand-in-glove enterprises that mutually develop one another. The role of evaluation in translational science has in fact been so central to its emergence that the practice of evaluating it became its own kind of science, fusing the 'science of translation' and the 'science of evaluation.' This state-of-affairs arose partly through the development of cultures of evaluations that underwent several transmutations during the CTSA program's history. The emergence of this 'evaluative way' conjoined the work of *doing* translational science with the work of *measuring its impact* and of producing accounts of the 'returns on investment' it yielded from federal funding.

Our findings reinforce a basic premise of STS, which is that a scientific field cannot emerge or endure without the communal and material infrastructures to support its existence and sociotechnical reproduction. The emergence of the evaluative way thus affirms Robinson's argument about "the limits of a predominantly epistemic historiography" in the history of translational science (Robinson, 2020: 33). The history of translational science should be written in terms of its material practices, infrastructures, and the historical conjuncture it emerged within – not solely based on its purportedly novel epistemic groundings. However, the evaluative way in translational science also suggests that evaluation practices in the field should not only be understood as activities that shape it, but also as *productive of* its constitutive epistemology. Evaluation is not just a way of knowing in translational science; engaging in the work of evaluation is also part of practicing and performing translational science. Metrics have, in short, produced translational science, insofar as they have been scientific outputs of translational science, a major area of activity within the field, and generative of representations of its varied impacts.

Therefore, the search for 'better' metrics and evaluation strategies to assess translational science continues. There will likely never be a set way of measuring the field's impact that practitioners agree on, because any metrics that are developed will shortly thereafter be replaced by other "better" metrics or ones suited to local contexts or specific translational initiatives. Ever-shifting metrics and strategies of evaluation are thus inherent to the very idea of translational science. Evaluation practices are best treated as an element of scientific ways of knowing and modes of technoscientific life in the translational sciences that co-develop and change with it and other fields that arose during the late 20th and early 21st centuries.

Indeed, we have described how the emergence of translational science in the US has been driven by the concomitant development of systems to measure its impacts. The construction of successive evaluation strategies organized around several deep enduring areas of evaluative focus in effect produced practices and cultures of evaluation that would become constitutive elements of translational science. This is how 'the evaluative way' emerged and solidified.

The emphasis on developing evaluation practices for translational science came from both within the field and its historical conjuncture. The drive to evaluate translational science was also motivated by external pressures from policymakers in Congress who demanded 'returns on investment' for taxpayers. This emphasis on evaluation was also shaped by federal policies implemented in the 2000s that required new programs to be evaluated for efficiency and by the broader rise of evaluation paradigms such as 'continuous quality improvement' and 'dissemination and implementation' research in the health sciences. The role of digital technologies to facilitate new

forms of training and collaboration also played key functions in enabling the emergence of translational science and specific styles of evaluation that produced it. It follows that the evaluative way in translational science can only be fully understood as a kind of assemblage-effect stemming from numerous forces internal and external to the field. These include logics of neoliberal governmentality, scientific managerialism, marketization, and efficiency that dominated this period of federal policy and transnational developments in the capitalist mode of production (Robinson, 2019).

NCATS, the CTSA program, and the field of translational science have also clearly influenced the general development of the health sciences. Impacts have included supporting the emergence of new professional associations with annual conferences such as the Association for Clinical and Translational Science (founded in 2009), new journals such as *Science Translational Medicine* (started in 2009) and the *Journal of Clinical and Translational Science* (started in 2017), which are dominated by CTSA-affiliated actors. Translational science has also driven the establishment of translational funding priorities and strategies within NIH beyond NCATS. Examples include a translational research division at the National Heart, Lung, and Blood Institute (NHLBI) and the National Institute of Allergy and Infectious Diseases (NIAID) orienting its post-COVID-19 pandemic preparedness strategy using a translational framing (NHLBI, n.d.; NIAID, 2021). NCATS and the CTSA program have also been principal drivers of the rise of the 'real-world' evidence movement and the development of large-scale informatics projects in biomedical research which unite the concerns of biomedical research and data science (Morrato et al., 2023; Sidky et al., 2023). The evaluative way in translational science thus emerged from, and then partially remade, the political economy of biomedicine.

We conclude with a question to guide future inquiry: what might the evaluative way in translational science and its effects indicate about the possible emergence of evaluative forms of life in the sciences? Future studies of the role and impacts of evaluation practices in science should attend to the forms of scientific subjectivity that develop around the desire for evaluation strategies that can attest to various forms of value created by specific fields, particularly for new or emergent sciences. The history of translational science suggests that other novel sciences' particular styles of self-evaluation will be as central to their emergence as their claims to epistemic novelty.

Acknowledgements

This paper was supported by a Clinical and Translational Science Award (UL1TR001439) from the National Center for Advancing Translational Sciences (NCATS), National Institutes of Health (NIH), through the University of Texas Medical Branch (UTMB) Institute for Translational Sciences (ITS). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. The funder had no role in study conceptualization, design, or execution. The authors thank their colleagues at UTMB in the Institute for Bioethics and Health Humanities and participants in a workshop of this paper, particularly Alberto Aparicio, Lisa Campo-Engelstein, and Richard Gibson. Finally, the authors thank their colleagues in the UTMB ITS who engaged with this work, particularly Kevin Wooten and Kelly Grace.

References

AAMC (2021) Appropriators Voice Support for Increased NIH Investments in FY22 Budget Hearings. Available at: https://web.archive.org/web/20250123190524/https://www.aamc.org/advocacy-policy/washington-highlights/appropriators-voice-support-increased-nih-investments-fy22-budget-hearings (accessed 23 January 2025) .

Aarden E, Marelli L and Blasimme A (2021) The translational lag narrative in policy discourse in the United States and the European Union: a comparative study. *Humanities and Social Sciences Communications* 8(1): 1–9.

Adams V (ed.) (2016) *Metrics: What Counts in Global Health*. Critical global health. Durham London: Duke University Press.

ARPA-H (2024) BREATHE Questions & Answers. Available at: https://web.archive.org/web/20250123185500/https://arpa-h.gov/explore-funding/programs/breathe/faqs.

Austin CP (2018) Translating translation. *Nature Reviews Drug Discovery* 17(7). 7. Nature Publishing Group: 455–456.

Austin CP (2021) Opportunities and challenges in translational science. *Clinical and Translational Science* 14(5): 1629–1647.

Austin JL (1975) *How To Do Things With Words*. 2nd ed. Cambridge: Harvard University Press.

Aviles NB (2018) Situated Practice and the Emergence of Ethical Research: HPV Vaccine Development and Organizational Cultures of Translation at the National Cancer Institute. *Science, Technology, & Human Values* 43(5): 810–833.

Basken P (2011) NIH Pushes Ahead on Research Division to Speed Drug Development, Despite Protests From Scientists and Others. *The Chronicle of Higher Education*, 23 February. Available at: https://www.chronicle.com/article/nih-pushes-ahead-on-research-division-to-speed-drug-development-despite-protests-from-scientists-and-others/ (accessed 1 February 2025).

Berman EP (2012) *Creating the Market University: How Academic Science Became an Economic Engine*. Princeton: Princeton University Press.

Bernard GR, Harris PA, Pulley JM, et al. (2018) A collaborative, academic approach to optimizing the national clinical research infrastructure: The first year of the Trial Innovation Network. *Journal of Clinical and Translational Science* 2(4): 187–192.

Boulton R, Sandall J and Sevdalis N (2020) The Cultural Politics of 'Implementation Science'. *Journal of Medical Humanities* 41(3): 379–394.

Boutzoukas AE, Olson R, Sellers MA, et al. (2023) Mechanisms to expedite pediatric clinical trial site activation: The DOSE trial experience. *Contemporary Clinical Trials* 125: 107067.

Bowker GC and Star SL (1999) *Sorting Things Out: Classification and Its Consequences*. Cambridge: MIT Press.

Bowker GC, Baker K, Millerand F, et al. (2009) Toward Information Infrastructure Studies: Ways of Knowing in a Networked Environment. In: Hunsinger J, Klastrup L, and Allen M (eds) *International Handbook of Internet Research*. Dordrecht: Springer Netherlands, pp. 97–117.

Bush V (1945) *Science, the Endless Frontier*. Washington, DC: National Science Foundation.

Butler J (1988) Performative Acts and Gender Constitution: An Essay in Phenomenology and Feminist Theory. *Theatre Journal* 40(4): 519.

Çalışkan K and Callon M (2009) Economization, part 1: shifting attention from the economy towards processes of economization. *Economy and Society* 38(3): 369–398.

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

Callon M and Law J (1982) On Interests and their Transformation: Enrolment and Counter-Enrolment. *Social Studies of Science* 12(4): 615–625.

Castells M (2010) *The Rise of the Network Society: The Information Age: Economy, Society, and Culture Vol. 1*. Malden: Wiley-Blackwell.

Clarke A, Mamo L, Fishman JR, et al. (2003) Biomedicalization: Technoscientific Transformations of Health, Illness, and U.S. Biomedicine. *American Sociological Review* 66(2): 161–194.

Collins FS (2012) *Testimony to the House Subcommittee on Labor – HHS – Education Appropriations*. 20 March. Washington, DC. Available at: https://www.nih.gov/sites/default/files/institutes/olpa/20120320-house-testimony-collins.pdf (accessed 16 May 2024).

Daston L and Galison P (2007) *Objectivity*. New York: Zone Books

Daudelin DH, Peterson LE, Welch LC, et al. (2020) Implementing Common Metrics across the NIH Clinical and Translational Science Awards (CTSA) consortium. *Journal of Clinical and Translational Science* 4(1): 16–21.

Daudelin DH, Peterson LE and Selker HP (2021) Pilot test of an accrual Common Metric for the NIH Clinical and Translational Science Awards (CTSA) Consortium: Metric feasibility and data quality. *Journal of Clinical and Translational Science* 5(1): e44.

Dear P (1995) *Discipline & Experience: The Mathematical Way in the Scientific Revolution*. Chicago: University of Chicago Press.

Dear PR (2007) *The Intelligibility of Nature: How Science Makes Sense of the World*. Paperback ed. Chicago: University of Chicago Press.

Deem R (2001) Globalisation, New Managerialism, Academic Capitalism and Entrepreneurialism in Universities: Is the Local Dimension Still Important? *Comparative Education* 37(1): 7–20.

Dembe AE, Lynch MS, Gugiu PC, et al. (2014) The translational research impact scale: development, construct validity, and reliability testing. *Evaluation & the Health Professions* 37(1): 50–70.

Doganova L, Giraudeau M, Helgesson C-F, et al. (2014) Valuation Studies and the Critique of Valuation. *Valuation Studies* 2(2): 87–96.

Duda GN, Grainger DW, Guldberg RE, et al. (2023) Measuring translational research impact requires reaching beyond current metrics. *Science Translational Medicine* 15(707): eabp8258.

Elzinga A (1988) The Consequences of Evaluation for Academic Research. *Science & Technology Studies* 1(1): 5-14.

Ezrahi Y (1990) *The Descent of Icarus: Science and the Transformation of Contemporary Democracy*. Cambridge: Harvard University Press.

Foucault M (1975) *Discipline and Punish: The Birth of the Prison*. 2nd Vintage Books ed. New York: Vintage Books.

Foucault M (1978) *The History of Sexuality, Vol. 1*. New York, NY: Vintage Books.

Fourth International Conference on Grey Literature (1999) What is Grey Literature? The New York Academy of Medicine. Available at: http://www.greylit.org/about (accessed 12 May 2023).

Frechtling J, Raue K, Michie J, et al. (2012) *The CTSA National Evaluation Final Report*. 3 April. Washington, DC: Westat. Available at: https://dpcpsi.nih.gov/sites/default/files/CTSA%20National%20Eval%20Final%20Report%2004-16-2012%20(09-2005-R%20NCRR)%2004-05-13_0.pdf (accessed 2 June 2023).

Fredrickson DS (1975) *The Purposes of the National Institutes of Health: 1. On the Translation Gap*.

Friedman M (2005) *Trying Hard Is Not Good Enough: How to Produce Measurable Improvements for Customers and Communities*. Bloomington: Trafford Publishing.

Godin B (2006) The Linear Model of Innovation: The Historical Construction of an Analytical Framework. *Science, Technology, & Human Values* 31(6): 639–667.

Grazier KL, Trochim WM, Dilts DM, et al. (2013) Estimating Return on Investment in Translational Research: Methods and Protocols. *Evaluation & the Health Professions* 36(4): 478–491.

Green JM, Goodman P, Kirby A, et al. (2023) Implementation of single IRB review for multisite human subjects research: Persistent challenges and possible solutions. *Journal of Clinical and Translational Science* 7(1): e99.

Greenberg DS (1967) NIH: Fountain Committee Issues Bitter Attack on Programs. *Science* 158(3801): 611–614.

Greenberg DS (1999) *The Politics of Pure Science*. New ed. Chicago: University of Chicago Press.

Greene M (2022) Experts, Managerialism, and Democratic Theory. *Spontaneous Generations* 10(1): 1-21.

Gregor C, Melvin A and Goss C (2019) 3411 Maximizing the Value of Your Trial Innovation Network Hub Liaison Team. *Journal of Clinical and Translational Science* 3(s1): 48–48.

Guston DH (2000) *Between Politics and Science: Assuring the Integrity and Productivity of Research*. 1st ed. Cambridge: Cambridge University Press. Available at: https://www.cambridge.org/core/product/identifier/9780511571480/type/book (accessed 8 July 2023).

Hacking I (1982) Biopower and the Avalanche of Printed Numbers. *Humanities in Society* 5: 279–95.

Hacking I (1995) The Looping Effects of Human Kinds. In: Sperber D, Premack D, and Premack AJ (eds) *Causal Cognition: A Multi-Disciplinary Debate*. New York: Oxford University Press, pp. 351–383.

Haendel MA, Chute CG, Bennett TD, et al. (2021) The National COVID Cohort Collaborative (N3C): Rationale, design, infrastructure, and deployment. *Journal of the American Medical Informatics Association* 28(3): 427–443.

Hall S (1988) *The Hard Road to Renewal: Thatcherism and the Crisis of the Left*. London: Verso.

Hallonsten O (2021) Stop evaluating science: A historical-sociological argument. *Social Science Information* 60(1): 7–26.

Harris PA, Dunsmore SE, Atkinson JC, et al. (2023) Leveraging the Expertise of the CTSA Program to Increase the Impact and Efficiency of Clinical Trials. *JAMA Network Open* 6(10): e2336470–e2336470.

Hedgecoe A and Martin P (2003) The Drugs Don't Work: Expectations and the Shaping of Pharmacogenetics. *Social Studies of Science* 33(3): 327–364.

Hilgartner S (2000) *Science on Stage: Expert Advice as Public Drama*. Stanford: Stanford University Press.

Holmes K (2017) CTSA Program Common Metric for Informatics Solutions (presentation to the CTSA Program Steering Committee). Northwestern University. Available at: https://www.tuftsctsi.org/wp-content/uploads/2017/07/Informatics-Metric-SC-Meeting-6.12.2017-FINAL_website.pdf (accessed 13 May 2024).

Holt M (2013) Enacting and Imagining Gay Men: The Looping Effects of Behavioural HIV Surveillance in Australia. *Critical Public Health* 23(4): 404–417.

Holton G (1978) Can Science Be Measured? In: Elkana Y, Center for Advanced Study in the Behavioral Sciences (Stanford, Calif.), and Social Science Research Council (U.S.) (eds) *Toward a Metric of Science: The Advent of Science Indicators*. Science, culture, and society. New York: Wiley.

Huebschmann AG, Brega AG, Stotz SA, et al. (2024) Gaps and opportunities for measuring equity with the Translational Science Benefits Model: Recommendations from the Center for American Indian and Alaska Native Diabetes Translation Research. *Journal of Clinical and Translational Science* 8(1): e206.

Institute of Clinical and Translational Sciences (n.d.) Case Studies – Translational Science Benefit Model. Available at: https://translationalsciencebenefits.wustl.edu/case-study/ (accessed 11 January 2025).

Institute of Medicine (U.S.), Leshner AI, Terry S, et al. (eds) (2013) *The CTSA Program at NIH: Opportunities for Advancing Clinical and Translational Research*. Washington: National Academies Press.

Jasanoff S (2004) The Idiom of Co-Production. In: Jasanoff S (ed) *States of Knowledge: The Co-Production of Science and Social Order*. London: Routledge, pp. 1–12.

Jasanoff S (2005) *Designs on Nature: Science and Democracy in Europe and the United States*. Princeton: Princeton University Press.

Jasanoff S (2006) Transparency in Public Science: Purposes, Reasons, Limits. *Law and Contemporary Problems* 69(3): 21–45.

Jasanoff S and Kim S-H (2009) Containing the Atom: Sociotechnical Imaginaries and Nuclear Power in the United States and South Korea. *Minerva* 47(2): 119–146.

Jeske M (2022) "Conflict of interest" or simply "interest"? Shifting values in translational medicine. In: Hauray B, Boullier H, and Gaudillière J-P (eds) *Conflict of Interest and Medicine: Knowledge, Practices, and Mobilizations*. Routledge studies in the sociology of health and illness. London: Routledge.

Kaiser J (2011a) Collins Sparks Furor With Proposed NIH Reshuffling. *Science* 331(6016): 386–386.

Kaiser J (2011b) Jeremy Berg: An Independent Scientist Departs NIH's Ranks. *Science* 332(6029): 533–533.

Kaiser J (2011c) NIH Reorganization Draws Scrutiny From Congress. *Science*, 27 November. Available at: https://www.science.org/content/article/nih-reorganization-draws-scrutiny-congress (accessed 11 January 2025).

Kaiser J (2011d) Senators Oppose Plan to Dismantle NIH Resources Center. *Science*, 11 February. Available at: https://www.science.org/content/article/senators-oppose-plan-dismantle-nih-resources-center (accessed 11 January 2025).

Kaplan-Liss E, Mitchell L, Crossno C, et al. (2022) Finding the Story. *Journal of Clinical and Translational Science* 6(1): e25.

Kotarba JA and Wooten K (2017) The innovation scorecard for continuous improvement applied to translational science. *Journal of Clinical and Translational Science* 1(5): 296–300.

Kurilla MG (2021) *CTSA Program Updates: Sunsetting of the Common Metrics Initiative in Summer 2022*. 22 September. Bethesda: National Center for Advancing Translational Science.

Lander B and Atkinson-Grosjean J (2011) Translational science and the hidden research system in universities and academic hospitals: A case study. *Social Science & Medicine* 72(4): 537–544.

Latour B (1993) *We Have Never Been Modern*. Cambridge: Harvard University Press.

Law J (2004) *After Method: Mess in Social Science Research*. London: Routledge.

Law J (2009) Seeing Like a Survey. *Cultural Sociology* 3(2): 239–256.

Lee PL (2016) CTSA Program Common Metrics Initiative PI Overview: Maximizing CTSA Program Impact. Available at: https://www.tuftsctsi.org/wp-content/uploads/2016/06/PI-RBA-Overview_HANDOUT_VERSION.pdf (accessed 12 July 2023).

Llewellyn N, Carter DR, Rollins L, et al. (2018) Charting the Publication and Citation Impact of the NIH Clinical and Translational Science Awards (CTSA) Program From 2006 Through 2016: *Academic Medicine* 93(8): 1162–1170.

Llewellyn N, Carter DR, DiazGranados D, et al. (2020) Scope, Influence, and Interdisciplinary Collaboration: The Publication Portfolio of the NIH Clinical and Translational Science Awards (CTSA) Program From 2006 Through 2017. *Evaluation & the Health Professions* 43(3): 169–179.

Luke DA, Sarli CC, Suiter AM, et al. (2018) The Translational Science Benefits Model: A New Framework for Assessing the Health and Societal Benefits of Clinical and Translational Sciences. *Clinical and Translational Science* 11(1): 77–84.

MacKenzie DA (2008) *An Engine, Not a Camera: How Financial Models Shape Markets*. First MPI Press paperback ed. Inside technology. Cambridge: MIT Press.

Madaus GF, Stufflebeam D and Scriven MS (1983) Program Evaluation. In: *Evaluation Models*. Dordrecht: Springer Netherlands, pp. 3–22. Available at: http://link.springer.com/10.1007/978-94-009-6669-7_1 (accessed 26 June 2023).

Mehta TG, Mahoney J, Leppin AL, et al. (2021) Integrating dissemination and implementation sciences within Clinical and Translational Science Award programs to advance translational research: Recommendations to national and local leaders. *Journal of Clinical and Translational Science* 5(1): e151.

Merton RK (1942) The Normative Structure of Science. In: Storer NW (ed.) *The Sociology of Science: Theoretical and Empirical Investigations*. Chicago: University of Chicago Press, pp. 267–278.

Mol A (1999) Ontological Politics. A Word and Some Questions. *The Sociological Review* 47(1_suppl): 74–89.

Mol A (2002) *The Body Multiple: Ontology in Medical Practice*. Durham: Duke University Press.

Morrato EH, Lennox LA, Dearing JW, et al. (2023) The Evolve to Next-Gen ACT Network: An evolving open-access, real-world data resource primed for real-world evidence research across the Clinical and Translational Science Award Consortium. *Journal of Clinical and Translational Science* 7(1). 2023/09/29 ed. Cambridge University Press: e224.

Murphy M (2017) *The Economization of Life*. Durham: Duke University Press.

Nash JC and Owens EA (2016) Introduction: Institutional Feelings: Practicing Women's Studies in the Corporate University. *Feminist Formations* 27(3): vii–xi.

Nathan DG and Nathan DM (2016) Eulogy for the clinical research center. *Journal of Clinical Investigation* 126(7): 2388–2391.

National Center for Advancing Translational Sciences (2014) *Clinical and Translational Science Award (U54) (RFA-TR-14-009)*. 12 September. Bethesda: National Institutes of Health. Available at: https://grants.nih.gov/grants/guide/rfa-files/RFA-TR-14-009.html (accessed 20 April 2024).

National Center for Advancing Translational Sciences (2023) *National Center for Advancing Translational Sciences Congressional Justification FY 2023*. Bethesda, MD. Available at: https://ncats.nih.gov/sites/default/files/NCATS-FY-2023-CJ-508.pdf (accessed 13 May 2024).

National Center for Advancing Translational Sciences (2024a) National Clinical Cohort Collaborative (N3C). Available at: https://web.archive.org/web/20250112051932/https://ncats.nih.gov/research/research-activities/n3c/overview (accessed 11 January 2025).

National Center for Advancing Translational Sciences (2024b) Trial Innovation Network. Available at: https://web.archive.org/web/20241226015738/https://ncats.nih.gov/research/research-activities/ctsa/projects/tin (accessed 11 January 2025).

National Center for Data to Health (CD2H) (n.d.[a]) CD2H: Harmonizing the Informatics Community. Available at: https://web.archive.org/web/20241219070347/https://cd2h.org/ (accessed 11 January 2025).

National Center for Data to Health (CD2H) (n.d.[b]) What is N3Covid? Available at: https://web.archive.org/web/20250104065606/https://covid.cd2h.org/about/ (accessed 11 January 2025b).

National Heart, Lung, and Blood Institute (n.d.) Center for Translation Research and Implementation Science. Available at: https://www.nhlbi.nih.gov/about/divisions/center-translation-research-and-implementation-science.

National Institute of Allergy and Infectious Diseases (2021) *NIAID Pandemic Preparedness Plan*. December. Available at: https://www.niaid.nih.gov/research/pandemic-preparedness.

National Institutes of Health (2018) Single IRB for Multi-Site or Cooperative Research. Available at: https://grants.nih.gov/policy-and-compliance/policy-topics/human-subjects/single-irb-policy-multi-site-research (accessed 11 January 2025).

NCATS Advisory Council Working Group on the IOM Report (2014) *NCATS Advisory Council Working Group on the IOM Report: The CTSA Program at NIH*. 16 May. Available at: https://ncats.nih.gov/files/CTSA-IOM-WG-report-5-2014.pdf (accessed 12 July 2023).

NCATS Alliance (n.d.) NCATS Alliance. Available at: https://ncatsalliance.org/ (accessed 23 January 2024).

NIH (2005) RFA-RM-06-002: Institutional Clinical and Translational Science Award. Available at: https://grants.nih.gov/grants/guide/rfa-files/RFA-RM-06-002.html (accessed 2 June 2023).

NIH (2007) RFA-RM-07-002: Institutional Clinical and Translational Science Award (U54). Available at: https://grants.nih.gov/grants/guide/rfa-files/RFA-RM-07-002.html (accessed 2 June 2023).

NIH (2009a) Important Events in the Division of Research Resources* (DRR) History (*Predecessor to NCRR). Available at: https://web.archive.org/web/20090225165257/http://www.nih.gov/about/almanac/organization/NCRR.htm (accessed 30 April 2024).

NIH (2009b) RFA-RM-09-019: Institutional Clinical and Translational Science Award (U54). Available at: https://grants.nih.gov/grants/guide/rfa-files/RFA-RM-06-002.html (accessed 2 June 2023).

NIH (2012) RFA-TR-12-006: Institutional Clinical and Translational Science Award (U54). Available at: https://grants.nih.gov/grants/guide/rfa-files/rfa-tr-12-006.html (accessed 12 July 2023).

NIH (2021) PAR-21-293: Clinical and Translational Science Award (UM1 Clinical Trial Optional). Available at: https://grants.nih.gov/grants/guide/pa-files/par-21-293.html (accessed 2 June 2023).

NIH Scientific Management Review Board (2010a) *Report on Translational Medicine and Therapeutics*. December. Available at: https://smrb.od.nih.gov/documents/meetings/TMAT/report_12072010.pdf (accessed 12 July 2023).

NIH Scientific Management Review Board (2010b) *Translational Medicine and Therapeutics Working Group*. 26 October. Available at: https://osp.od.nih.gov/wp-content/uploads/SMRB_Jul_2010_TMAT_WG_Charge.pdf (accessed 12 July 2023).

Olszewski TM (2018) Between Bench and Bedside: Building Clinical Consensus at the NIH, 1977–2013. *Journal of the History of Medicine and Allied Sciences* 73(4): 464–500.

Owermohle S (2023) A new lobby shop looks to remind Congress about NIH's translational research arm. *STAT News*, 24 July. Available at: https://www.statnews.com/2023/07/24/nih-translational-research-lobbyist/ (accessed 1 January 2025).

Palm ME, Edwards TL, Wieber C, et al. (2023) Development, implementation, and dissemination of operational innovations across the trial innovation network. *Journal of Clinical and Translational Science* 7(1): e251.

Palm ME, Thompson DD, Edwards T, et al. (2023) The Trial Innovation Network Liaison Team: building a national clinical and translational community of practice. *Journal of Clinical and Translational Science* 7(1): e249.

Parthasarathy S (2007) *Building Genetic Medicine: Breast Cancer, Technology, and the Comparative Politics of Health Care*. Inside technology. Cambridge: MIT Press.

Patel T, Rainwater J, Trochim WM, et al. (2019) Opportunities for strengthening CTSA evaluation. *Journal of Clinical and Translational Science* 3(2–3): 59–64.

Peterson D and Panofsky A (2023) Metascience as a Scientific Social Movement. *Minerva* 61(2): 147–174.

Polanyi M (1962) The Republic of Science: Its Political and Economic Theory. *Minerva* 1(1): 54–73.

Porter TM (1996) *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton: Princeton University Press.

Rabinow P and Bennett G (2012) *Designing Human Practices: An Experiment with Synthetic Biology*. Chicago: The University of Chicago Press.

Ribes D (2014) Ethnography of scaling, or, how to a fit a national research infrastructure in the room. In: *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, Baltimore Maryland USA, 15 February 2014, pp. 158–170. ACM. Available at: https://dl.acm.org/doi/10.1145/2531602.2531624 (accessed 31 May 2023).

Ribes D and Bowker GC (2008) Organizing for Multidisciplinary Collaboration: The Case of the Geosciences Network. In: Olson GM, Zimmerman A, and Bos N (eds) *Scientific Collaboration on the Internet*. Cambridge: MIT Press, pp. 310–330.

Ribes D and Finholt TA (2009) The Long Now of Technology Infrastructure: Articulating Tensions in Development. *Journal of the Association for Information Systems* 10(5): 447–473.

Robinson M (2017) Translational Medicine: Science, risk and an emergent political economy of biomedical innovation. In: Tyfield D, Lave R, Randalls S, et al. (eds) *The Routledge Handbook of the Political Economy of Science*. New York: Routledge, pp. 249–262.

Robinson M (2020) Why translational medicine is, in fact, "new," why this matters, and the limits of a predominantly epistemic historiography. *History and Philosophy of the Life Sciences* 42(3): 34.

Robinson MD (2019) *The Market in Mind: How Financialization Is Shaping Neuroscience, Translational Medicine, and Innovation in Biotechnology*. Cambridge: MIT Press.

Rose N (2006) *The Politics of Life Itself: Biomedicine, Power, and Subjectivity in the Twenty-First Century*. Princeton: Princeton University Press.

Rubio DM (2013) Common Metrics to Assess the Efficiency of Clinical Research. *Evaluation & the health professions* 36(4): 432–446.

Rubio DM, Blank AE, Dozier A, et al. (2015) Developing Common Metrics for the Clinical and Translational Science Awards (CTSAs): Lessons Learned. *Clinical and Translational Science* 8(5): 451–459.

Ruiz R, Vitale A, Schwartz A, et al. (2022a) From insights to action: Enriching the clinical and translational research workforce by developing diverse and inclusive career programs. *Journal of Clinical and Translational Science* 6(1): e104.

Ruiz R, Schwartz A, Orlando E, et al. (2022b) Moving beyond annual data reports: A blueprint for communicating and disseminating actionable intelligence. *Journal of Clinical and Translational Science* 6(1): e70.

Sanghavi D (2024) HEROES: HEalth care Rewards to Achieve Improved OutcomES Public Slide Deck. Available at: https://web.archive.org/web/20250117222127/https://arpa-h.gov/sites/default/files/2024-01/HEROES%20Public%20Deck%20Focused%201-12-24_FINAL.pdf (accessed 23 January 2025).

Schneider M, Guerrero L, Jones LB, et al. (2015) Developing the Translational Research Workforce: A Pilot Study of Common Metrics for Evaluating the Clinical and Translational Award KL2 Program: A Study of Common Metrics for the CTSA KL2 Program. *Clinical and Translational Science* 8(6): 662–667.

Senate Committee on Appropriations, 86th Congress, First Session (1959) *Metabolic and Therapeutic Facilities, Report 425*, 23 June.

Senate Hearing, 112th Congress - Departments of Labor, Health and Human Services, and Education, and Related Agencies Appropriations for Fiscal Year 2012 (2011) United States Senate. Available at: https://www.govinfo.gov/content/pkg/CHRG-112shrg19104451/pdf/CHRG-112shrg19104451.pdf (accessed 16 May 2024).

Shah M, Culp M, Gersing K, et al. (2017) Early Vision for the CTSA Program Trial Innovation Network: A Perspective from the National Center for Advancing Translational Sciences. *Clinical and Translational Science* 10(5): 311–313.

Shapin S and Shaffer S (1985) *Leviathan and the Air-Pump: Hobbes, Boyle, and the Experimental Life*. Princeton classics. Princeton: Princeton University Press.

Sidky H, Young JC, Girvin AT, et al. (2023) Data quality considerations for evaluating COVID-19 treatments using real world data: learnings from the National COVID Cohort Collaborative (N3C). *BMC Medical Research Methodology* 23(1): 46.

SMART IRB (n.d.) Supporting single IRB review: Advancing collaborative research. Available at: https://smartirb.org/ (accessed 11 January 2025).

Smith EMR, Molldrem S, Farroni JS, et al. (2024) Articulating the social responsibilities of translational science. *Humanities and Social Sciences Communications* 11(1): 85.

Stark D (2011) *The Sense of Dissonance: Accounts of Worth in Economic Life*. Princeton: Princeton University Press.

Stevens KR, De La Rosa E, Ferrer RL, et al. (2021) Bootstrapping implementation research training: A successful approach for academic health centers. *Journal of Clinical and Translational Science* 5(1): e168.

Strathern M (ed) (2000) *Audit Cultures: Anthropological Studies in Accountability, Ethics, and the Academy*. European Association of Social Anthropologists. London: Routledge.

Sunder Rajan K (2006) *Biocapital: The Constitution of Postgenomic Life*. Durham: Duke University Press.

Timmermans S and Epstein S (2010) A World of Standards but not a Standard World: Toward a Sociology of Standards and Standardization. *Annual Review of Sociology* 36(1): 69–89.

Trial Innovation Network (n.d.) TIN Toolbox. Available at: https://trialinnovationnetwork.org/tin-toolbox/ (accessed 11 January 2025).

Trochim WM, Rubio DM, Thomas VG, et al. (2013) Evaluation Guidelines for the Clinical and Translational Science Awards (CTSAs). *Clinical and Translational Science* 6(4): 303–309.

Tufts CTSI (2018) *Report by Tufts CTSI on Common Metrics Implementation Program and Evaluation Study*. September. Boston, MA. Available at: http://tuftsctsi.s3.amazonaws.com/d2m2/TuftsCTSI_CommonMetrics_Report_FullReport.pdf.

United States Congress (2007) National Institutes of Health Reform Act of 2006. Available at: https://smrb.od.nih.gov/documents/reformact_2006.pdf (accessed 12 July 2023).

U.S. Congress, House of Representatives (2011) Military Constructions and Veterans Affairs and Related Agencies Appropriations Act: Conference report to accompany HR 2055. Available at: http://www.gpo.gov/fdsys/pkg/CRPT-112hrpt331/pdf/CRPT-112hrpt331.pdf (accessed 5 July 2023).

U.S. Office of Technology Assessment (1986) *Research Funding as an Investment Can We Measure the Returns?: Report*. Science policy study. Background report.no. 12. Washington: U.S. G.P.O. Available at: https://catalog.hathitrust.org/Record/002769565 (accessed 26 June 2023).

Vardeny O, Hernandez AF, Cohen LW, et al. (2019) Transitioning to the National Institutes of Health single institutional review board model: Piloting the use of the Streamlined, Multi-site, Accelerated Resources for Trials IRB Reliance. *Clinical Trials* 16(3): 290–296.

Vastag B (2011) Panel backs new NIH center devoted to translational medicine. *Nature Medicine* 17(1): 5–5.

Vatin F (2013) Valuation as Evaluating and Valorizing. *Valuation Studies* 1(1): 31–50.

Vinsel L and Russell AL (2020) *The Innovation Delusion*. New York: Currency.

Wadman M (2011) NIH centre faces spell in limbo. *Nature* 477(7363): 141–142.

Weiss CH (1973) Where politics and evaluation research meet. *Evaluation: The International Journal of Theory, Research and Practice* 1(3): 37–45.

Welch LC, Tomoaia-Cotisel A, Noubary F, et al. (2021) Evaluation of initial progress to implement Common Metrics across the NIH Clinical and Translational Science Awards (CTSA) Consortium. *Journal of Clinical and Translational Science* 5(1): e25.

Whitacre R (2024) 'Financial fallout' in the US biopharmaceutical industry: Maximizing shareholder value, regulatory capture, and the consequences for patients. *Social Science & Medicine* 344: 116598.

Zerhouni EA (2003) The NIH Roadmap. *Science* 302(5642): 63–72.

Zerhouni EA (2005) Translational and Clinical Science — Time for a New Vision. *New England Journal of Medicine* 353(15): 1621–1623.

Notes

1 "Grey literature" is defined as "[T]hat which is produced on all levels of government, academics, business and industry in print and electronic formats, but which is not controlled by commercial publishers" (Fourth International Conference on Grey Literature, 1999).

2 Our study focused on the history of CTSA program evaluation and the evaluation of translational initiatives led by, or conducted through, the CTSA program. We do not focus on the emergence of implementation science, a field which emerged partly from communities of practice in evaluation science that included many CTSA-affiliated actors. Implementation science activities led by CTSAs starting in the 2010s focused on building capacities for institutions to conduct implementation science research about evidence-based health interventions, not on the evaluation of CTSAs or the field of translational science (Boulton et al., 2020; Mehta et al., 2021). The emergence of implementation science in relation to translational science should be the subject of future STS research.

3 Describing the full functions, history, and afterlives of the NCRR are beyond the scope of this paper.

4 Reports issued by IOM and its successor – the National Academy of Medicine – have long been utilized in this manner to manage tensions in science policy (see, Hilgartner, 2000).