

8 • WHEN NUMBERS AND STORIES COLLIDE Randomized

Controlled Trials and the Search for Ethnographic Fidelity in
the Veterans Administration CAROLYN SMITH-MORRIS

Assuming both are medically stable, a fifty-three-year-old American computer programmer with tetraplegia from an injury sustained in fighting in Vietnam typically has a lot more job options than a recently injured American thirty-eight-year-old roofer with the same SCI. But they look about the same in randomized controlled trial (RCT) data. Why? Because, in the stripped-down language of clinical trials, facts about people are homogenized in order to foster clear outcomes in relation to a small set of isolated factors. The U.S. medical system's reliance on these unembellished metrics and the austere research designs that support them has far-reaching implications not only for health (especially for persons with SCI) but for entire health care infrastructures. Here, at what some might call the center and others might call the periphery of the global health community, we find trends in the use of metrics that are simultaneously specific and global in significance.

Drawing from research within the largest integrated health care system in the United States, the Veterans Health Administration (VHA), I examine some of the challenges of bringing qualitative data and research methodologies to bear within the politically charged, outcome-driven context of a U.S. government institution. My experience as a researcher in one RCT among veterans with spinal cord injury serves as a case study for this chapter. Increasingly all forms of care and all types of healers and care providers are pressed by market-driven biomedical ideologies to funnel all outcomes evidence into the RCT mold. The state colludes with this pressure through organizations like the VHA and asserts new levels of conformity in both patient and professional populations. But using these models to decipher outcomes and determine optimal modes of care can be misleading, sometimes promoting metric quantification over case-specific quality, which, after all, is the ultimate outcome of care. I will document a project that illustrates how both the work of the RCT and the kinds of information that the RCT fails to acknowledge reveal and create gaps even while trying to ensure that these disparate metrics are integrated. I wish to show how numbers and stories can be used collaboratively but also in ways that are fraught in large institutions of health care. The VHA, which meets the needs and rights of U.S. veterans for not only health care but for the best that scientific research can supply, exposes the collision of numeracy and ethnographic data in global health metrics.

The Quality Revolution(s)

At the beginning of my career in the late 1980s, hospital care was monitored through systems of quality assurance, or what was also called *quality improvement* (based on the hyperproductive and efficient Japanese companies from which the Deming model and team concepts emerged). Quality monitoring, then and now, includes everything from reducing errors in medical record transcription and drug delivery to utilization review (nurses with insurance or actuarial expertise) and having agreements in place for the transportation of psychotic patients by bus, using physical and chemical restraints if necessary. One of my first positions was assisting, and eventually leading, the Quality Improvement Office and accreditation review cycles of a mental health hospital. So my sensitivity to this language of quality, measurement, monitoring, and evidence has been long in the making.

By the 1990s “quality” health care was infused with customer service messaging, lessons in “total quality management,” and the need to foster cultures of “continuous quality improvement” so that every worker and every patient was empowered to identify, and then act, to correct or improve upon a problem (Lammers et al. 1996).¹ My position was an irritant to many of the hospital staff and a necessary evil in the world of increasingly “managed” and “accredited” health care that had to be done in ways that demonstrated both good quality and a continuous effort to measure and improve.

The early 1990s was also a time during which the VHA—the national health system designed for research, health care, and health insurance for the honorably discharged veterans of the U.S. armed forces—had a reputation for *poor* quality care (Oliver 2007). In the era of total quality management and continuous quality improvement, the VHA was under tremendous pressure to transform itself using Deming models or any other means necessary. The secretary of the VHA at the time was widely hailed as revolutionary. His campaigns included a major reorganization of the lines of authority and decision making across this vast system, increased emphasis on primary care over inpatient and acute care, implementation of an electronic health record, and the development of “a larger sustained effort to systematically study and enhance VHA clinical programs, including their quality, processes, and outcomes” (McQueen et al. 2004: 340).² Part of this transformation was QUERI, the VHA Quality Enhancement Research Initiative,³ so by the mid-1990s the era of “quality” at the VHA had begun. Over the next decade there was tremendous growth in the QUERI program, and by the late 2000s quality in VHA care was reported by some to have been not only completely transformed but better than private care available in the United States (Kizer et al. 2008; Kupersmith et al. 2007; Oliver 2007; Stetler et al. 2008). Even so, yet another swing of the pendulum would bring wait times, limited prescription benefits, and hospital error to the Veterans Administration (VA) system, and to the news media as well. In 2009 the appointment of a new secretary of the VA was intended to produce another necessary massive transformation in addressing quality of care problems (Demakis et al. 2000; Kizer et al. 2008; McQueen et al. 2004), and this led to sixteen key initiatives, seven of which explicitly targeted the VHA and health benefits. The 2014 media frenzy over wait times and other “quality” issues (e.g., veteran homelessness, poor quality mental health care) led to this secretary’s resignation. McCullough et al. (2013: 10) wrote prophetically

cally, “VA leadership . . . [is] under public scrutiny and any adverse event could trigger ripples of political consequences that could potentially go all the way up to Capitol Hill.” These types of political drama, in which heads of organizations take the fall for the most egregious—or most public—errors, is neither new nor noteworthy. But the relevance of this particular history for the VA, and specifically for the VHA and its dual treatment-research role in the nation’s largest health care system, is disquieting.

The new national (and global) obsession with accountability for quality health care—for veterans through the VHA, for the uninsured through the Affordable Care Act, or “Health for All” through the World Health Organization—suggests how performance metrics, outcome quantification, and the quantification of quality have been drawn deeply into the global public consciousness. In the United States responsibility for these practices fell squarely on the government. The past twenty years of VHA history are a microcosm in the progression of quality metrics and the types of evidence-based medicine that have been promoted through private and public sector health policies. The continuing cycles of quality decline, followed by heroic and revolutionary change, play back like repeating cycles in a daytime television drama. One wonders whether the same volatility is evident at the granular and local level and whether the contentious marriage of state policy with market-driven cost-effective health care and medical research outside of the VHA is equally volatile.

Producing Qualitative Data in a Clinical Trial

The VHA is staffed by thousands of clinical and health science researchers, 60 percent of whom are also clinicians. And there are at least twenty-five anthropologists in the VA (Gemmae Fix, personal communication, January 8, 2015). As in other teaching hospitals, VHA research institutions are a blend of academic and clinical space, merging these two pursuits into complex and dynamic work relationships with varying degrees of success. Research is conducted for multiple reasons in the VHA, but the most public of these reasons is “to fulfill VHA’s core research missions through a diverse portfolio of medical (basic and clinical), rehabilitation, and health services research” (McQueen et al. 2004: 339). In its various websites, strategic plans, and literature, the VHA takes pride in combining research with clinical care as a way to “identify the direct needs of patients at chair and bedside, and to find discoveries and innovations directly in-step with these

needs—keeping the Veteran at the center of health care from the very beginning.”⁴

The VHA manages research through its Office of Research and Development. The budget request for this vast national research program for 2015 was over \$589 million.⁵ The dual responsibilities of the VHA as provider of clinical care for a politically influential and highly symbolic segment of the population (i.e., to care for “him who shall have borne the battle and for his widow, and his orphan,” according to Abraham Lincoln, who supplied the original motto of the VA) and as a competitive research center with direct ties to and corresponding obligations under federal funding, make the VHA an intriguing incubator for policy reverberation, feedback on fiscal priorities, and scientific echo.⁶

The VHA trial for which I was hired as a contract medical anthropologist was designed to promote quality of life for veterans with spinal cord injury (SCI), one of ten focal areas for which a QUERI Coordinating Center was established. The project was integrally part of the VHA agenda for quality monitoring, research producing evidence of best practice, and the granting of research dollars through the VA system to specialized VHA sites across the nation. My role was to provide supplemental qualitative data about a trial that was already in full swing by the time I arrived. This role gave me insights on the quality and evidence-based medicine agenda within the trial and, more generally, on how these forces impact researchers who design and conduct trials.⁷

The project, called Spinal Cord Injury Vocational Integration Program: Implementations and Outcomes (SCI-VIP), was a multicenter clinical trial evaluating the impact of evidence-based supported employment for veterans with SCI. It was mostly organized around the performance of specific services and the collection of quantitative data on outcomes, but the purpose of the qualitative research (in which I was involved) was to describe the universe of experiences and opinions represented in the illness narratives of participants and staff. Research methods included semistructured and open-ended interviews with staff and participants, observations of provider-participant encounters, and participant observation in clinical and service activities as well as veterans’ homes and communities. We also observed several staff meetings on a semiregular basis.⁸ Ethnographic interviews with veterans occurred exclusively in community and home settings. In all interviews and conversations, key foci were (1) experiences of the veteran within the project or in the VA system of care generally, (2) the vet-

eran's story of and perspective on the injury, (3) the meaning of work and the veteran's goals related to work, and (4) experiences and feelings about being a veteran and receiving care through the VHA. Not surprisingly these data were not being captured in the trial's original protocol.

Turning Support into a "Model"

Although my discussion in this chapter focuses on the outcome metrics of the trial, a brief introduction to the unique target of this trial is necessary. Were it not for an arguably extreme version of patient-centeredness in this model and the trial's commitment of fidelity to that model, I surely would not have two forms of evidence to discuss. As late as just one decade ago, patients recovering from SCI were discouraged from thinking about goals such as paid employment (Hammell 2007). Rehabilitation focused solely on physical rehabilitation, and critics noted that therapy was "treating the SCI but not the person who had sustained the SCI" (269). The Supported Employment model outlines an integrated approach to helping people obtain and maintain community-based competitive employment in their chosen occupations (Becker et al. 2006; Bond et al. 2001; Bond 2004). In the model a vocational rehabilitation counselor is seen as an integrated member of the existing interdisciplinary SCI team. The counselor is given more structured access to and is able to utilize and call upon the treatment team and its resources to enhance vocational rehabilitation services. This aspect of the model is not only distinctive and new but also crucial. That is, the counselor, rather than the patient, is now responsible for integrating and coordinating all vocational rehabilitation services, including job-finding efforts (which I describe later). In short the counselor is designed to serve a broader scope of duties in managing treatment for the patient and for collaborating with clinical and employment professionals.

One challenge of Supported Employment interventions was in creating an evidence base that would support its use. Evidence-based approaches had not been widely used or clinically tested among persons with physical disabilities, including SCI. Using the Supported Employment approach was already hard to accomplish and even harder to quantify. Unemployment for veterans with spinal injuries is a serious and common problem, contributing to lower reported quality of life, life satisfaction, and adjustment (Chapin and Kewman 2001; Krause 1990, 1992; Trieschmann 1980; Westgren and Levi 1998; Yerxa and Baum 1986), and our trial was meant to

remedy precisely this set of problems. Thus one of the goals of the research was to find ways to acknowledge the veteran with SCI as a person in context, requiring coherence among the physical, emotional, and vocational aspects of his or her life. Still it was not entirely clear how to do an evidence-based version of RCT research that would provide data to show this.

The clinical researchers had to demonstrate the effectiveness of this program in ways the QUERI and the clinical and funding worlds would recognize. The research therefore adopted the gold standard trial methodology and the guidelines for reporting results of an RCT, including measures of *fidelity* to an evidence-based model of care, in this case supported employment. What these RCT efforts suggest is a rather large chasm between the qualitative data that make sense of these interventions and the quantitative requirements used in accounting for their outcomes.

Two Views of the Same Trial

Two of the RCT's published products reveal some of the tensions between the kinds of data QUERI needed (Ottomanelli, 2012, Effectiveness of Supported Employment for Veterans with spinal cord injuries: Results from a randomized multisite study). One (Ottomanelli et al. 2012) was written for a clinical journal conforming to RCT-driven models called the Consolidated Standards for Reporting Trials, or CONSORT. The original CONSORT statement (Begg et al. 1996; Moher et al. 2001) was written in 1996, with revisions in 2001 and 2010, to improve the quality of reporting so that readers of a published report would have access to "complete, clear, and transparent information on its methodologies and findings" (Schulz et al. 2010: 1). Using CONSORT was important because only certain forms and venues of publication (e.g., CONSORT publications for RCTs or PubMed-indexed journals) were (and are) valued within the VHA (McCullough et al. 2013).

The other published products were written for an anthropological journal and an edited volume (Smith-Morris et al. 2013, 2014). The differences between traditional RCT outcome data and traditional ethnographic data are great, but both studies were needed to adequately account for the trial's design, performance, and achievements. In the end the differences between these forms of evidence, or data, reveal some of the challenges in defining "quality" when it comes to outcomes, along with challenges involved in information gained or lost in this federally funded and government-affiliated research. My intention in making this contrast is not to disparage or ap-

plaud either perspective but rather to consider the spectacular voids created by CONSORT guidelines for RCT publication and the comparative messiness of the ethnographic accounts.

The CONSORT guidelines include a checklist of twenty-five items of information to include when reporting an RCT. Among these are methods for randomization, allocation, and blinding; details on participant flow, recruitment, outcomes, and estimation; and discussion of limitations, generalizability, and interpretations (Schulz et al. 2010). Most PubMed-indexed journals require that authors follow these guidelines. In fact, however, it became quickly apparent that what the RCT CONSORT guidelines called for is *fidelity* and *outcome* statistics, not necessarily empirical or complete details about which aspects of the program worked and which did not.

Fidelity

The term *fidelity* refers to whether a trial's intervention was conducted consistently and completely as outlined in the model's protocol. Clinical trials that adopt the fidelity approach rigorously monitor, coach, and supervise the services performed to ensure ideal adherence to the intervention model (Bond et al. 2008a, 2008b). Values associated with the fidelity approach include replicability of data and the reliability with which researchers can attest that the model, as designed, was indeed what occurred during research. Fidelity to the treatment model is considered one of the requisite metrics of scientific rigor in an RCT.⁹ Fidelity was described in the clinical journal *Archives of Physical Medicine and Rehabilitation* this way: "The 3 [project] sites providing [Supported Employment] were evaluated every 6 months using the 15-item IPS Fidelity Scale. No differences in Fidelity scores over time or between sites in these biannual visits were observed during the study period ($F_{7,18} = 0.8$, $p < .619$; range, 59–68). A review of each site's fidelity ratings suggests significant achievement toward good [Supported Employment] implementation. As a whole, site fidelity scores averaged 63.4 ± 2.5 , which falls within the upper portion of the 'fair' range. Fidelity scores for staffing (12.8 ± 0.5), organization (11.6 ± 1.6), and services (38.9 ± 1.5) demonstrated little change over the study period" (Ottomanelli et al. 2012: 743–44). In this reporting of fidelity scores, the emphasis is on how closely the intervention adhered to its procedural (and paperwork) protocols. But these numbers do not (and do not claim to) report the empirical outcomes of the intervention. In other words, even if the study had

been conducted perfectly according to the model's protocols, the outcome could have been poor. Fidelity and outcome are not the same.

The quotation from Ottomanelli et al. (2012) reports an overall fidelity score for the model as a whole, then adds others for staffing, organization, and services rendered. Adhering to individual principles in the published supported employment model (e.g., measures of how well the vocational rehabilitation [VR] counselor was integrated into the clinical team or the degree to which care was multidisciplinary and how that actually worked) are *not*, however, part of this scoring system. Indeed fidelity metrics give only a broadly aggregated sense of how well the model was followed and which aspects of the model might have been more challenging.

The numbers claim fidelity as one success of the RCT; that is, the treatment model was implemented successfully, according to its terms and principles. However, these metrics offer very little "evidence" of any of the ethnographically significant findings relevant to the performance of the model. For instance, one principle of the model emerged over others in importance, and some principles required tremendous intensity of effort that was completely invisible in trial outcome or even fidelity metrics. The importance of these themes emerged only through analysis of the narrative accounts we gathered. Grounded analyses of multiple recorded and transcribed interviews with staff (with whom excellent rapport and trust had been developed over time) involved two independent rounds of coding of interview transcripts and analysis toward consensus of the relative importance of emergent themes—in other words, the *qualitative* research methodologies and the *qualitative* evidence. Still, for the CONSORT guidelines, the authors had to render this material legible as a metric of fidelity.

To be sure, fidelity monitoring helped the trial by ensuring that the principles were given appropriate attention and support during the trial. But the fidelity scores did *not* reveal the relative importance of things like integration or intensity of effort to the rest of the Supported Employment model's principles and to the trial's overall success in this site. Thus the fidelity report in clinical format, while affirming that the trial met this CONSORT requirement for trial evidence, is a limited data set.

Integration

The integration of the VR counselor into the clinical team's meetings, activities, and planning was itself somewhat revolutionary. A crucial and dis-

tinctive part of the Supported Employment model, integration envisions employment (historically an element of patients' lives that was considered part of their social context, and therefore largely irrelevant to clinical care) as part of all health care. VR counselor integration was discussed in the clinical publication this way: "Given the complexity of issues and barriers faced by individuals with SCI, there may be many reasons why [Supported Employment] has benefits for this population over conventional methods of VR. For one, this model involves integrating vocational services within the medical rehabilitation health care setting. This integration meant that the interdisciplinary health care team, which included the VR specialist, would identify and address barriers with the Veterans as they began a job search, as well as after the job was obtained and new issues emerged" (Ottomanelli et al. 2012: 746).

Given the inability of fidelity statistics to say much about how integration actually occurred, we do not learn how the VR counselor accomplished the very clear agenda laid out in the goals of the project. How, on the other hand, could it be shown that the VR counselor in the SCI-VRP did so much more than her non-Supported Employment counterparts? And how could all of these contextual barriers to positive outcomes become data points for reporting of outcomes? What I am drawing attention to is that even though the RCT reported fidelity, the very nuanced information about what made this intervention work better than previous models formed an evidence base that did not show up in these reports.

In contrast the ethnographic publications considered this process in detail. Things that were invisible in fidelity reporting became immediately obvious in the ethnographic research: a portion of the VR counselor's success was attributable to a smaller caseload; another portion to her longevity in similar work and familiarity with the local job markets; still another to her positive and constructive relationships with colleagues both in and outside of the trial—and even the VA—who shared cultural and social capital with her relevant to this work. However, as ethnographers learned when we duplicated the VR counselor's driving and home and employer visit schedule, the time savings of the smaller caseload were quickly and completely overrun by travel time to homes and communities within a hundred-mile radius of the office.

Ethnographers also watched and recorded as the VR counselor accessed and utilized the treatment team's resources, including clinical professionals and knowledge central to SCI rehabilitation. Persons with these injuries

have more frequent medical complications, such as urinary tract infections, skin issues, or depression. The counselor communicated directly with the veterans' physicians, sped up medical attention when necessary, and offered input to treatment planning and healing that enabled quicker returns to work. The counselor also conducted job finding and utilized techniques like job carving, job coaching, and follow-along with employers to ensure the veterans' medical status met or exceeded their workplace demands.

As staff talked about the Supported Employment principles and the challenges of meeting these guidelines, integration of care (principle #1) was named far more often than any other as critical to their success.¹⁰ At team meetings, we heard these sorts of comments: "[In the inpatient team meetings] we talk about cases—things that have happened with their medical [condition], things that have happened with people and how that has interfered or helped, how that impacted. . . . Or maybe good things, like so-and-so had a medical issue that came up and [the VR counselor] rapidly referred him to me and we took care of it and avoided a problem. . . . [The VR counselor] came to our [inpatient department] meeting [to tell us about this problem], and those are things that help." This staff member is describing a novel feature of the SCI-VRP: the counselor's regular access to the inpatient clinical team—the same team of clinicians who cared for the client when he or she first received the SCI. Research staff members augmented the importance of this sort of evidence by saying things like the following:

When the [inpatient] staff meeting occurs, we try to give an update on [SCI-VRP] enrollment. . . . We try to kind of keep it in the forefront of [the clinical staff's] minds. We've been trying to kind of translate the success stories back to the team, so that they can see kind of the fruits of their labor in terms of what actually is happening and try to facilitate some of that mutual learning process between the vocational clinician and the SCI clinician that didn't previously have any exposure to what voc rehab meant. So, you know, the vocational clinician maybe doesn't know what an [occupational therapist] does and so we've been sort of trying to cultivate this cross, the integration.

Cultivating communication and mutual support between the VR counselor and the other experts on the clinical team is a key part of integration of care and is named by multiple types of staff as a key to the success of

the Supported Employment model. By educating the clinical team on the principles and methods of the Supported Employment model, the trial expanded the normal purview of that team well beyond the clinic and into the communities and work settings of their patients. Likewise clinical integration of the counselor promoted her greater knowledge of and sensitivity to the medical needs of veterans, yielding greater success in her work.

Not surprisingly none of these kinds of data is available in the clinical publication. This is partly because so little room is available after CONSORT requirements are met. Thus, despite the fact that my collaborating researchers *did* mention several of the ethnographic findings in the report, the standard reporting format dictated that the ethnographic detail be relegated to a different article and a different journal. In other words, the CONSORT requirements did not account for what we considered to be the most important kinds of evidence for why the program worked. We do not see this as simply a problem of publication conventions. Rather it is a problem of what kinds of evidence count in the reporting of success and to whom these data count.

Intensity of Effort

Central to the purpose of the RCT was demonstration that the Supported Employment model of care was better than the traditional vocational rehabilitation model of care. That is, consistent with RCT designs, we used comparisons to establish impact effect. But a major factor in explaining the failure of the VR model was its generally lower likelihood of actually providing services (i.e., not that their actual services were less effective). The clinical publication had this to say: "It has been our observation that when Veterans are referred to providers or agencies outside the SCI center, relatively few of them actually receive any VR services. This is supported by our study finding that Veterans in the [Supported Employment] group had twice as many VR visits than those who were referred outside the SCI center for vocational care" (Ottomanelli et al. 2012: 746).

The Supported Employment model being tested addressed this problem (nonreceipt of services) explicitly. The model calls for "rapid engagement," "zero exclusion," and "ongoing support," making an exceptionally intensive job for the VR counselor. We saw similar trends in relation to what researchers called "intensity of effort": how much effort was put into support services for each veteran.

In the traditional model clients come to the VR counselor, but in the Supported Employment model VR involves numerous home visits and intensive support. One VR counselor, for example, had a client who lived at the near maximum distance of the catchment area (one hundred miles). The drives to see him were long, and his reluctance to visit with the counselor (even at his own home) was exacerbated by his drug use. In the traditional VR model, this man might never have received support for employment because those supports were available only in a professional office. The trial's VR counselor made this problem clear: "I had a guy in Fort Worth, who had an obvious drug problem and didn't want anyone in his house. He hated company. He admitted to me he signed on [to the SCI-VIP] just to get the fifty dollars [research stipend] every month. So I continued to go out there, you know. Then he started to get angry, because he didn't really want me to visit . . . but I continued to go out anyway. Just kept going, kept going, kept trying to talk to this guy. And he and I started to get along pretty well" (Smith-Morris et al. 2014: 148).

Through persistence and a flexible approach to goal setting, the VR counselor established rapport and was able to help this client get two job interviews prior to close of the trial, a much better result than traditional methods of VR might have achieved. The VR counselor reflected on the difficulty of achieving the same outcome (i.e., paid employment) with clients who start in such different states of readiness: "That's difficult, trying to find people at that right place. And spinal cord cases are delicate, and some people take a long time. You know, I've had cases that didn't get a job until that twelfth month. I have people who have never found a job. You know, to think it's going to be just a blanket fix, you know, it's not going to happen" (Smith-Morris et al. 2014: 148).

Identifying good job targets and then providing job leads or setting up interviews for participants was not enough to reach a positive employment outcome. The VR counselor had to discover participants' interests, maximize their readiness for job interviews and work, expand their capabilities through mechanical, technical, and clinical supports, and then discover possible sources for work to which the participant might apply. The process was invariably and intensively social, and its success depended on something that was hard to calibrate on a quantitative index: intensity of effort. This counselor was simultaneously a support counselor for the patient and an administrative assistant for setting up, and sometimes driving the veteran to, job interviews.

However, this absolutely radical approach to patient support shown by VR counselors in the SCI-VIP was sadly missed in the fidelity reporting that became the primary source of clinical information about what worked and did not work. Very little of this sort of ethnographic detail would find its way into the clinical publication of outcomes that were relied on to evaluate and promote this program with the government.

Competitive Employment

As a last example of the way qualitative evidence was pushed to the side in the evaluation exercises of the RCT and its publications, I turn to the case of competitive employment. Here the clinical publication offers a statement about the central metric of the trial's outcomes:

Our primary outcome variable, employment, was defined as competitive employment obtained after the baseline interview. . . . Among the 201 subjects, 35 subjects (17.4%) accounted for 90 total jobs. The rate of employment for [Supported Employment] subjects was significantly greater (29.6%; 95% CI, 20.8–40.4) than that of either the TAU-IS¹¹ group (11.8%; 95% CI, 4.6–19.1; $p < .003$) or the TAU-OS¹² group (4.8%; 95% CI, 0.5–16.7; $p < .002$). When employment was restricted to competitive employment only, [Supported Employment] subjects significantly accounted for 50 (69.4%) of 72 jobs and were significantly more likely to achieve employment (25.9%; 95% CI, 17.6–36.5) compared with either TAU-IS subjects (10.5%; 95% CI, 3.6–17.4; $p < .008$) or TAU-OS subjects (2.3%; 95% CI, 0.0–12.9; $p < .002$). (Ottomanelli et al. 2012: 744)

The reduction of employment facts to a set of numerical figures about employment accomplishes a great deal for policymakers who want to show bottom lines in relation to employment statistics. Still we might ask if this sort of reporting also occludes the kinds of information that are of key importance to *understanding* how this program was able to achieve these indicators of program success in the first place.

Ethnographic materials that were gathered in the process of producing these numbers offer many more narrative terms and contextual detail, but they have great importance for the eventual replication of the trial's outcomes in clinical care, which is key to the VHA mission of blended research and clinical care. Ethnographic reports emphasize how the SCI-VIP

adapted to each veteran's needs and readiness for employment. In fact despite their appearance of being similar when presented in numerical form, the specific services offered by the VR counselor were rarely consistent across veterans. Instead the VR counselor managed a caseload with a wide range of needs and readiness: securing interviews and job fitting for some, focusing on motivational interviewing with others. These varied widely from patient to patient, counselor to counselor. Take these kinds of statements from clients about what kinds of work would be meaningful or how they might obtain these jobs:

We need purpose. And I know that I like feeling useful, not just to myself, I like to feel that I'm useful to society, to the country. That I can contribute in some small way, it's all of us working together to make the whole thing work. And of course, I feel useful. (Smith-Morris et al. 2013: 154)

[This project] can't be made any better. It's just that they're reaching the people that they can reach. It's just up to them whether or not they want to better themselves by going back to work. (Smith-Morris unpublished research)

The veterans with spinal cord injuries became employed and the kinds of employment they found matched these sensibilities about what mattered in their relationships with VR counselors. Two examples of employed vets offer insight on this. One, Tim, was an online researcher and a payroll clerk. Another, Joe, found employment with a land title research company. Joe explained that his job consists of Internet research to inform land disputes and litigation. His work hours are flexible, and he works from home with minimal supervision and oversight. Joe states that he does some work in the morning or late at night, as he has ready access to the Internet. He considers the job ideal since the work does not interfere with his medical treatments and appointments. His notion of what mattered here (and why he could keep his job) were things that the VR counselor had to understand and take into consideration in her intervention efforts, and they reflected more fundamental concerns about how to be valuable to society.

Working in a similar environment, Tim is a payroll clerk for a company based in New Jersey (hundreds of miles away from his home). He explained that his work puts him in charge of verifying employee hours, calculating wages, and reporting to corporate offices—and he does all this

from his home via the Internet. Much like Joe's, Tim's work schedule is fairly flexible, and he considers it the perfect fit for his life and needs. Again the VR counselor's ability to be attentive to diverse set of concerns, specific to each vet, was what made this employment trajectory successful. What is missed, then, in the streamlining and reducing of this information for CONSORT and fidelity reporting, in the turning of these stories into numbers?

The ethnographic details about what kinds of jobs work, how to find them, and how to manage diverse caseloads that are attentive to individual client views of meaningful employment are integral to the program's success. Recognizing that these are precisely the kinds of details that are made invisible in reports that count, however, begs the question of not just fidelity (as in what kinds of things are evidence of successful program implementation) but reproducibility as well. Could other programs in other places really reproduce this success without knowing in detail *how* the VR counselor accomplished her goals? I turn now to consider what these evidentiary contrasts mean for the nation's most complex health care and research system.

Solemn Duty and the Politics of New Metrics

It would be far too simple to argue that quality monitoring and improvement are simple technologies of state domination or market exploitation. That is, quality monitoring and improvement are undoubtedly sites where the state enacts policies and where the market drives agendas (particularly cost-effectiveness). Given the depths to which quality and evidence-based rhetoric have penetrated and permeated the VHA system, we might consider how these practices work and what their broader effects are. As the VHA faces aggressive and cyclical political pressure to use taxpayer dollars wisely and to fulfill its "solemn duty" to care for those who "have borne the battle," we cannot underestimate the impact of the techniques for accounting and measuring what these programs accomplish.

The QUERI system is the VHA's "ambitious attempt to develop a data-driven national quality improvement program . . . to ensure excellence in all places where VHA provides health care services" (Demakis et al. 2000). The system is both a communicative system, conveying referential information, and an ideological system, deploying ideologies, political priorities, and administrative agendas through "collaborative links between researchers, clinicians, managers, and policy makers throughout different divisions with

the organization" (118). Its *modus operandi*, however, is research. The eight steps of the QUERI process focus on the identification, definition, implementation, and documentation of "best practices," and it is in this "identification of best practices" (Step 2) that definitions of "quality" achieve hegemonic status. Through the trope of quality, biases and information choices are made invisible. Thus in Demakis et al., "quality" is aligned with a bodyless authority except to invoke the internationally powerful Institute of Medicine as corroborative in guideline decisions: "All [quality] guidelines and other evidence-based documents were assessed for quality (especially the rigor of the methods used during development and the degree to which the documentation explicitly states the strength of evidence underlying each recommendation) and also assessed for applicability to Veterans. The criteria developed by the Institute of Medicine were useful for this purpose" (122). Nowhere in this document is the concept of evidence explained or defined. Rather it is assumed to be found primarily in the numbers, through presumptions about scientific omniscience and objectivity via mechanisms like the CONSORT statement and the gold standard of the RCT.

In fact any reader familiar with clinical and health services literature of the past two decades knows well that the reference to quality implies little more than fidelity to the stage model for clinical outcome trials.¹³ Here research modeled on the gold standard RCT offers the highest and best form of evidence for what works, what counts (Smith-Morris et al. 2014). For those of us doing ethnographic research, the question then remains: What happens to all that other evidence that falls outside the RCT model? For the VHA's QUERI system, this problem is summed up in a dismissive phrase: "Although process and structural data are also needed, outcome measurement is prioritized" (Demakis et al. 2000: 123). Indeed the transition of the VHA toward evidence-based clinical care and clinical trial research is described as a "maturation of the health services research field" (123). But what kind of maturation is this, and what kinds of obfuscations are entailed in making these measurements possible?

I am eager to point out that the VHA and its QUERI system have at times relied on ethnographic data (what Demakis et al. 2000 called "structural and process data") to do the work of measuring the impact of trials and outcome research. Volume 37 of the *Annals of Anthropological Practice* was dedicated to ethnographic work within the VA system, and is a robust testimony to the relevance of ethnographic data for clinical trials. But it is also

clear that these “structural and process data” find limited room in policy and clinical publications. They are not part of CONSORT requirements. Journals carrying these data are less valorized (lower impact factors, lower science indices), and the social scientists producing them are not typically invited to working groups that design both research and treatment guidelines (Adams 2002, 2005; Smith-Morris 2015). Hammell (2007: 271) confirmed this in finding that “surprisingly little research has explicitly sought to explore the experience of rehabilitation following SCI.”

The limited room for qualitative data is important and notable, but it is not the only problem. Simple inclusion of qualitative data in or alongside quantitative reports is insufficient since the decision-making power, resource allocation, and professional hierarchies of clinical work profoundly impact care *and* the way that research is conducted. The truths about program effectiveness are found in critical examination not only of the trial and its interventions and outcomes but also in the dynamic relationships among humans and within institutions through which RCTs are enacted (Nguyen 2010; Petryna 2009). Thus although anthropological tools and methods can be attached to outcomes research, it is clear that quantitative research has a better chance of being recognized and influential in the places that matter.

The VHA example parallels the large-scale, state investments in middle- and low-income country health care and allows us to extend this conversation of “which metrics matter” into global health. Nguyen’s (2010: 91) “therapeutic citizenship” encompasses the activism, knowledge about potentially better treatments, and “a sense of political engagement” that gay activists in South Africa developed over access to HIV medications. Demanding first access to participate in RCTs, and later to the medications themselves, this generation of activist patients conceptualized their targets as *rights* and the duty of the state to ensure them. The VHA as a government body is vulnerable to the same politicized pressures—although this pressure does not always emerge from patient groups. Instead the VHA responds to superior political and governmental offices, to external stakeholders and competitors for federal funds (\$589 million requested for 2015), to their own veteran members, and ultimately to the voting citizenry.

Nguyen’s (2010) therapeutic citizens and Petryna’s (2009) biological citizens emerged as influential only after they became organized, informed, and politicized. In both of these cases organization cohered around a par-

ticular diagnosis or related syndrome of problems. Even in these diagnostically related groups, however, membership in the groups is heterogeneous. On their way to heterodox status, groups advocating health *rights* claim converts from the margins of orthodox medicine. Similarly Davis and Nichter (2015) address how biomedical doctors who are friendly to the calls for holistic and chronic attention to Lyme disease add legitimacy and force to therapeutic citizenship. Indeed social scientists who work alongside RCTs (as in my work) or in “health partnerships” to support clinicians (e.g., in Haiti; see Minn, this volume) may offer one way of intersecting contextualized data and metrics into global health.

But, as Taussig (1980) and others have warned, these collaborations are fraught. Political aims in a variety of global contexts are regularly legitimized through quantitative metrics. I return to Demakis et al.’s (2000: 123) comment about the VHA, that the “maturation” of a health science research “field” is tied to “evidence-based clinical care and clinical trial research.” This strategic phraseology envisions a “field” that is first about methods of quantification and objectification, not about clear and full understanding of the results of interventions. A “field” built on metrics that include only quantities and ignore scientific evidence that resists quantification is a house of cards: a thin-walled, unstable, and hollow structure. While the benefits of RCT methodology and evidence gathering are not disputed, the gap between outcomes and the contextual evidence that explains those outcomes is wide and growing (Adams 2002, 2005; Adams et al. 2005; Smith-Morris et al. 2014). My fear is that qualitative research seems to have lost the battle for health care and federal-funding relevance.

The legitimization of political aims through scientific metrics allows a variety of actors to claim both moral and scientific authority. Better global health will require both better metrics and ethnographic engagement, social scientific evidence, and a tolerance for the small scale—not always seeking economies of scale. In their comprehensive discussion of poverty alleviation, Banerjee and Duflo (2011) argue that to influence health care, one must understand health-seeking behaviors; that top-down health policies (e.g., mandated prevention behaviors) are a costly waste of time in some environments; and that there are “deep reasons” (154) and “time inconsistencies” (194) that de-link truly effective health care from any single strategy, no matter how well evidenced and supported by RCT metrics.¹⁴

The greater dilemma for global health that an obsession with metrics

will create is what David Armstrong (1995) calls “surveillance medicine.” Just as colonial medicine controlled the brown and black hordes through pathologization of their bodies and identities, surveillance medicine problematizes normality, redrawing the relationships among symptom, sign, and illness and using medicine’s newest tools—its evidentiary forms, including the RCT and CONSORT guidelines—to define risk and moralize conformity. Biehl and Petryna’s (2013) more recent volume does an excellent job of exposing this new species of health surveillance in the name of better governance.

With all this winning, maturation, and quality at our fingertips, it is a wonder that VHA secretaries still rotate in and out with regular swings of the political pendulum. Arguably this is because the environment is shifting as well as the citizenry. The VHA as an institution claims to serve the *rights* of veterans to receive health care and to enjoy the benefits of well-funded, scientific investigatory efforts. The VHA therefore harnesses some of the same tropes of therapeutic citizenship as the people it serves! It is beyond this discussion to dissect the hierarchy of power and political influence that yields any given VHA budget. Suffice it to say that the QUERI program represents one successful director’s agenda: with the ideologies, symbols, goals, targets, methods, promises, and strategies that were not only institutionally achievable but were also popular in the media-informed public. The evidence-based agenda will continue to propel a diversity of agendas because it is the trope that works.

Ethnographic and qualitative data must catalyze their own therapeutic citizenship. In defense and promotion of qualitative data, a number of authors (Biehl and Petryna 2013; Sobo 2009) remind us of the power of the idiosyncratic story to capture a reading public’s imagination. We might also ask how this sort of work bolsters our sensibilities to what counts, to what is reliable, valid, and analytically powerful. Qualitative researchers will continue to narrate local and idiosyncratic lessons from the field, but we must also claim relevance in the highly structured world of intervention trials and evidence-based care. It is the job of social critics to be mindful of the myriad and sometimes unquantifiable forces acting in “evidence” production, especially in places that are under the rationed funding and bureaucratic supervision of governments.

Notes

1. There are even studies on the quality of the Total Quality Management approach and of programs that purport to employ the approach, in a disorienting circular system of reference (e.g., Lammers et al. 1996).
2. Indeed the transformation was actually based on a “value equation, in which value is considered to be a function of technical quality, access to care, patient functional status, and services satisfaction all divided by the cost or price of the care” (Kizer and Dudley 2008: 18.5).
3. It should not be surprising that the man who by many accounts accomplished the needed transformation was described as “a dynamic policy entrepreneur” (Oliver 2007: 15).
4. U.S. Department of Veterans Affairs, “Funding,” accessed November 2, 2014, <http://www.research.va.gov/funding>.
5. The VHA Office of Research and Development is divided into four parts: Biomedical Laboratory R&D, Clinical Services R&D, Health Services R&D, and Rehabilitation R&D. The \$589 million budget requested is labeled “medical and prosthetic research,” while another \$59 billion is requested for “medical care” for eligible veterans.
6. See Gary Marcus and Ernest Davis, op-ed “Eight (No Nine!) Problems With Big Data,” *New York Times*, April 6, 2014.
7. My role as a limited-time contractor speaks to the marginalized market position of many qualitative researchers within the VHA system. McCullough et al. (2013) have written about the tenuousness of this type of contract for professionals (especially those without university affiliations) whose scholarly expertise is drained by a contract employment system because it does not allow time for, or even necessarily see the value of, publication and engagement within the larger anthropological community. The contract labor format facilitates an agile, relatively inexpensive, and constantly fresh supply of expert researchers in any given location. But the format is unstable, professionally draining, and intellectually parasitic for scholars whose continuing expertise depends in part on an ongoing nurturing of methodological, theoretical, professional, and topical learning exclusive to their degree-granting field, something categorically difficult to obtain in contract work (McCullough et al. 2013; Bowman et al. 2008).
8. Including the weekly meetings of all vocational rehabilitation (VR) counselors, the VIP VR counselor conference calls, the VIP site coordinator conference calls, and the Inter-Disciplinary Team meetings.
9. Fidelity to the Supported Employment model was measured in a number of ways. The program records of SCI-VIP as well as participant medical records are reviewed biannually to assess the degree to which the program (documented in these notes) adheres to the principles of evidence-based Supported Employment described by Bond 2004. During site visits consultant trainers gather data on program fidelity through semistructured interviews with consumers, employers, clinicians, program managers, health care team members, administrative management and leadership; chart reviews; and observation of team meetings. A fidelity scale for Supported Employment, which has been widely used and is consistently linked with employment outcomes among

programs that serve persons with mental illness, was also used in SCI-VIP (Becker et al. 2001; Bond et al. 2008a).

10. There were eight principles. Please see Bond et al. (2001) for details.
11. Therapy As Usual — Interventional Site.
12. Therapy As Usual — Observational Site.
13. U.S. Food and Drug Administration, “The FDA’s Drug Review Process: Ensuring Drugs Are Safe and Effective,” November 6, 2014, <http://www.fda.gov/drugs/resourcesforyou/consumers/ucm143534.htm>.
14. That is, the psychological difficulty of matching our actions today with our vision for the future.

9 • THE TYRANNY OF THE WIDGET An American Medical Aid Organization’s Struggles with Quantification PIERRE MINN

In 1939 the General Motors Corporation created a six-minute animated film to explain economic processes to children. Entitled *Round and Round* the film focuses on a factory that produces “widgets.” The widgets appear as nondescript cubes, and the narrator tells us, “A widget might be a radio, a refrigerator, a musical instrument, or a motor car. A widget, you know, is just a symbol for any manufactured product that people use.” As the story unfolds, the widget maker sells his widgets to a farmer, a coal miner, a steel manufacturer, and a lumberman. These individuals in turn sell their raw materials to the widget maker. Money circulates in one direction, widgets in the other. The result, according to the film, is a harmonious circle of exchanges. By using “widgets,” the filmmakers direct their young viewers’ attention away from the specificities, intentions, or results of exchanges. The focus is on the process of exchanges rather than their particularities or contents.¹

CRITICAL GLOBAL HEALTH · *Evidence, Efficacy, Ethnography*
A series edited by Vincanne Adams and João Biehl

METRICS

What Counts in Global Health

VINCANNE ADAMS, EDITOR

Duke University Press · Durham London · 2016