Why Are There So Many Ways to Measure Pain? Epistemological and Professional Challenges in Medical Standardization

Hanna Grol-Prokopczyk
State University of New York at Buffalo, USA

DOI: https://doi.org/10.18778/1733-8077.21.1.03

Keywords:
Chronic Pain; Consensus Recommendations; Data Harmonization; Interdisciplinarity; Measurement; Standardization; Validity

Abstract: Pain is a profoundly subjective phenomenon, which remains largely impenetrable to the tools of biomedicine. How, then, do pain researchers—specifically, quantitative medical researchers whose work is predicated on transforming pain into numbers—measure pain in their studies? How do they select and justify specific measures, and does this process lead to measurement standardization? This article analyzes 79 published medical studies about low back pain (LBP) and 20 interviews with pain experts (including 15 with authors of the reviewed studies) to address these questions. Findings reveal that LBP researchers use an extremely diverse set of outcome measures in their studies, typically based on patient self-report. The subjectivity and interpersonal incomparability of self-reports are widely acknowledged but treated as largely unproblematic—a matter of acceptable measurement error rather than "epistemological purgatory" (Barker 2005). However, researchers frequently disagree on what constitutes a "pain measure." Many respond to the considerable challenge of treating pain intensity by redefining their work—sometimes in the face of resistance from patients—around other, putatively more treatable domains, such as disability. The diverse, arguably unstandardized approaches to measuring pain appear attributable less to pain's epistemological fragility than to its therapeutic intractability, and to the medical community's diffuse social structures and professional goals.

Hanna Grol-Prokopczyk is an Associate Professor of Sociology at the University at Buffalo, State University of New York. She is a medical sociologist who uses qualitative and quantitative methods to study the measurement of pain and other subjectively-described health conditions. She also studies social determinants of pain, and consequences of pain (such as opioid use, disability, and mortality). She received her Ph.D. from the University of Wisconsin-Madison.

email address: hgrol@buffalo.edu

To have pain is to have certainty;
to hear about pain is to have doubt.
Elaine Scarry (1985:13)

When I see a patient with arthritis coming in the front door,
I leave by the back door.
Sir William Osler,
first chief of Johns Hopkins Medical School
(late 19th century; in Graf 2010:1976)

Quantitative researchers frequently advocate for measurement standardization to facilitate the comparison and pooling of research studies. Such standardization is a key facet of the data harmonization deemed "essential" for the advancement of medical and social scientific research (Fortier et al. 2012:96). Pain researchers are no exception to this line of thinking: calls to improve standardization of pain measures date back over three decades (Institute of Medicine 1987), and recur particularly frequently with reference to outcome measures in clinical trials (e.g., Deyo et al. 1998; Turk et al. 2003). Nonetheless, measurement of pain in research studies appears—at least from some perspectives—to remain highly unstandardized, with studies often differing substantially in which pain-related domains they examine and which specific measures of those domains are used (e.g., Hjermstad et al. 2011; Kamper et al. 2011; Mulla et al. 2015).

This study seeks to document and understand the diversity of pain measures used in medical research. Why are so many measures used, and why does this pattern persist despite recurring calls for standardization? Is the "fragile factuality" (Baszanger 1992) of pain—that is, its profound subjectivity and impenetrability to the tools of biomedicine—to blame? Or is the explanation a social and institutional one, in which barriers across professional worlds prevent standardization? What role do norms of scientific justification—which would seem, *prima facie*, to support comparability—play in standardization efforts? This study explores these questions by focusing on low back pain (LBP), a common and costly pain condition, and a paradigmatic one in its etiological and therapeutic characteristics.

This topic may interest researchers who study pain or other subjective health conditions or who are themselves involved in standardization efforts. But the case of pain also elucidates theoretical issues relevant to the sociology of science and medicine. Most social scientific studies of standardization have focused on successful cases of standardization. By analyzing a case that is, at best, ambiguously successful, this study elucidates the factors that undermine scientific consensus-building. Moreover, most studies examine efforts to standardize protocols, methods, or technologies. The current focus on *measurement* standardization may reveal challenges specific to this area.

Background

Pain Standardization Efforts

Since it coalesced in the 1970s, pain medicine has been a highly international and interdisciplinary field (Whelan 2009). The International Association for the Study of Pain (IASP), founded in 1974, currently has over 6,000 members in 125 countries (IASP website 2024), who represent "the highly diverse fields of anaesthesia, neurology, psychology, general practice, psychiatry, nursing, and social work, among others" (Whelan 2009:171).

For at least three decades, pain experts across specialties have expressed dissatisfaction with "[i]nconsistencies in definitions and measurement" of pain (Institute of Medicine 1987:7). In 1992, an international group of rheumatologists responded to frustration that clinical trials were "extremely difficult to compare and combine" by founding OMERACT (Outcome Measures in Rheumatology Clinical Trials), which periodically issues recommendations for outcome measures in studies of pain-producing conditions (Tugwell et al. 2007:1). In 2002, two American psychologists founded IMMPACT (the Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials), which develops "consensus recommendations" for outcome measures in trials of chronic pain to "facilitate comparisons and pooling of data" (Turk et al. 2003:337-338). IMMPACT has thus far published at least seven highly-cited consensus recommendations focusing on outcome domains or measures (http://www.immpact.org/). Pain measures are also included in broader efforts at measurement standardization, such as the COMET Initiative (Core Outcome Measures in Effectiveness Trials) and the NIH-supported PROMIS system (Patient Reported Outcomes Measurement Information System).

Individual researchers or groups of researchers also regularly publish articles encouraging standardization. Richard A. Deyo and colleagues (1998) recommend outcome measures for studies of back pain; Alison M. Elliot, Blair H. Smith, and W. Alastair Chambers (2003) promote a measure of chronic pain severity; Bernhard Aicher and colleagues (2012) recommend pain measures for clinical trials of headache treatments; Elisabeth G. VanDenKerkhof, Madelon L. Peters, and Julie Bruce (2013) advocate outcome domains for studies of chronic pain after surgery, et cetera.

While some publications recommend certain outcome *domains*, others recommend specific *measures* of such domains (i.e., particular questionnaires). Some groups advocate for disease-, location-, or context-specific pain measures, while others, like IMMPACT, aim "to develop a consensus…that would transcend specific chronic pain syndromes" (Turk et al. 2003:338). Regardless, a majority of recommended measures are based on patient self-report. This reflects the dictum of Margo McCaffery, a nurse and founding member of IASP, that "pain is whatever the experiencing person says it is, existing whenever the experiencing person says it does" (McCaffery and Thorpe 1989:113). This idea, and the variant describing self-report as the "gold standard" of pain assessment, have diffused in the pain community to the point of being considered "conventional maxim[s]" (Schiavenato and Craig 2010:667). The following analyses consider how these features of measurement recommendations affect standardization.

Despite the proliferation of recommendations for standard measures, contemporary studies continue to demonstrate high variability in pain measurement. Steven J. Kamper and colleagues' (2011) review of 82 studies of LBP reveals a striking variety of measures in use, yielding 66 different definitions of recovery from pain. Sohail M. Mulla and colleagues (2015), examining opioid analgesic trials, find that fewer than 5% cite IMMPACT recommendations and that most IMMPACT-recommended outcome domains are omitted in a majority of studies. A 2011 Institute of Medicine report on pain reveals continued frustration with the "lack of a single, universally accepted metric," which "confounds" both clinicians' and researchers' work (2011:140). Thus far, pain standardization efforts appear not to have met with resounding success.

The Fragile Factuality of Pain

Philosophers and social scientists writing about pain show a consistent fascination with what Isabelle Baszanger (1992) calls its "fragile factuality," that is, its invisibility and interpersonal unverifiability. While some invisible health conditions achieve mediated visibility through the tools of biomedicine, pain maintains *profound invisibility*: no tool exists to prove and/or quantify its existence to others. Concerted efforts to identify objective measures of pain are ongoing, but thus far, only "*potential* pain biomarkers" have been identified (Davis et al. 2020; emphasis added). Both the original 1979 and revised 2020 definitions of pain promulgated by the International Association for the Study of Pain (IASP) explicitly acknowledge pain's subjectivity (Raja et al. 2020) and, like the new ICD-11 classification of chronic pain, note that pain is not always associated with actual or observable tissue damage (Treede et al. 2015). Consequently, those experiencing pain are vulnerable to accusations of malingering and/or drug-seeking from healthcare providers, employers, and family (Zajacova, Grol-Prokopczyk, and Zimmer 2021)—accusations that may take a high psychological toll (e.g., Glenton 2003).

And yet, when Jean Jackson (2011:378) describes "[r]elations between pain patients and health care deliverers" as "the worst in medicine," she attributes this broken relationship not only to patient-provider mistrust, but also to the sheer difficulty of treating pain, in particular chronic pain.[1] Reviews of pain treatment effectiveness are discouraging, finding that "none of the most commonly prescribed treatment regimens are, by themselves, sufficient to eliminate pain and to have a major effect on physical and emotional function in most patients" (Turk, Wilson, and Cahana 2011:2232). Physicians describe pain treatment as "frustrating and challenging" (Graf 2010:1976) and "one of the most difficult and unrewarding problems in clinical medicine" (Borkan et al. 1995:977).

In her 1992 ethnography of French pain clinics, Baszanger describes a schism between doctors who aim to *cure* chronic pain (and focus on biomedical causes and interventions) and those who aspire only to *manage* it (and typically focus on psychological mechanisms). At the time of her research, "no consensus" between the two approaches was in sight (Baszanger 1992:182). The present study may serve as an update to Baszanger's findings. Do contemporary pain specialists see chronic pain as something to cure or something to manage, and does this shape how they choose to measure it?

As Emma Whelan (2003) notes, scholarly examinations of pain's epistemological elusiveness have focused on its implications for people with pain and for their interactions with others, especially healthcare providers. Less research has examined implications for the biomedical *science* of pain—a claim that remains true over 20 years since Whelan first made it. Whelan's (2003; 2009) work analyzes lay- and expert-designed forms for assessing endometriosis pain to compare the goals and "knowledge concerns" of the respective communities; and elsewhere documents the difficulties of generating a classification of pain in the highly multidisciplinary and international pain research community. However, neither the classification nor the endometriosis forms were designed to be outcome measures in quanti-

[1] The present article focuses largely, but not exclusively, on chronic pain since acute pain may be assessed with similar measures and since some reviewed studies examine both acute and chronic pain patients.

tative studies. Indeed, even the gynecologist-developed form "carefully avoids any explicit suggestion that the patient's experience of pain can be reduced to a numerical value" (Whelan 2003:468).

While many of the researchers interviewed in the present study are clinician-researchers, the current focus is on how they measure pain in their quantitative research, where the demands for quantification are higher than in clinical settings. These researchers *must* represent pain via numbers. How, precisely, do they do so? What measures do they select, and how do they justify their use? Marion V. Smith (2008:993) matter-of-factly writes that "[s]urvey researchers' purposes are satisfied by statistical tests that demonstrate adequate levels of validity and reliability [of survey questions]." However, "validity and reliability," when examined closely, are not such straightforward goals—and may not, in fact, be the sole drivers of researchers' decisions about pain measurement.

Pain researchers operate not as isolated individuals but as members of professional social worlds—communities that can shape professional goals and scientific norms, including norms of measurement. As Whelan (2003:463) writes, "systems of pain measurement…must be seen as the products of epistemological communities with particular interests, aims and methods which affect the construction of pain." The present study builds on this insight in at least two related ways: by evaluating whether and how much pain researchers have been influenced by recent measurement standardization efforts in the field and by examining to what extent pain researchers (in particular, low back pain researchers) constitute an integrated epistemological community versus representing multiple communities with multiple norms.

The Sociology of Standardization

As Stefan Timmermans and Marc Berg (1997:273) write, "[u]niversality through standardization is at the heart of medical and scientific practice." In their interrogations of how standardization happens, Timmermans and Berg highlight two points of high relevance to the present study. First, standards rarely result from a single individual's efforts, but instead involve the interplay of many actors working across social networks. The role of networks, specifically pre-existing networks, is key: new standards are "plugged into a physical and cultural infrastructure that allows [them] to function" (Timmermans and Berg 1997:283). Second, "[u]niversality is always local universality" (Timmermans and Berg 1997:297), that is, standards must be adapted and integrated into local circumstances. A corollary of this is that "loose" standards may diffuse more successfully than "rigid" ones, as they are more easily molded to local requirements (Timmermans and Epstein 2010:81).

These characteristics are indeed evident in many examples of successful standardization (e.g., Fujimura 1996; Skloot 2010). Little research, however, explores cases of *unsuccessful* standardization. While Stefan Timmermans and Steven Epstein (2010:81) acknowledge that "[s]tandards may fail implementation for countless reasons," they provide no detailed examples. One may ask, then, what specific features of professional social networks—or what other factors altogether—undermine standardization. For example, if a field is highly international and interdisciplinary, does this help or hinder the adoption of standards?

Another question inviting exploration is how much flexibility in standards is optimal. Too little, and stan-

dards will be rejected for not meeting local needs, but too much, and their status as "standards" may be undermined. Timmermans and Epstein (2010:81) are explicit that standards that have been modified, "whether slightly or fundamentally," should not be considered failures, since "a standard's flexibility is often key to its success." But how, then, does one distinguish standardization from fragmentation? To explore this issue, I took an emic perspective in my interviews, asking pain experts directly whether they considered pain measures to be standardized and why.

Data and Methods

To shed light on how pain researchers measure pain, and to identify factors supporting or hindering measurement standardization, this study relies primarily on two sources of data. The first is 79 peer-reviewed medical research articles on LBP published between 1999 and 2008. These articles were originally identified by Kamper and colleagues (2011) in a systematic review of definitions of recovery from LBP.[2] While my research focus is different, the articles are well-suited for this study for several reasons.

First, LBP is not only extremely common and costly (with a lifetime prevalence in Western countries of approximately 80% [Brötz et al. 2003]), but it is also a paradigmatic pain condition in terms of its fragile factuality. That is, LBP is characterized by a surprisingly weak correlation between observable physical pathologies and experienced pain. Many individuals have spinal conditions such as herniated disks but report no pain, while for those who do report pain, "there is no correlation between the severity of the abnormality"—if one is observed at all—"and the degree of pain" (Cassar-Pullicino 1998:218). Next, the date range of the Kamper articles (1999-2008) conveniently straddles the publication of the first IMMPACT recommendations in late 2003 (Turk et al. 2003), permitting a comparison of outcome measures used before and after.

While Kamper and colleagues (2011) reviewed 82 articles, I excluded three because they were not in English or could not be located. My documentary analysis of how pain is measured in research publications is based on the remaining 79 articles, henceforth referred to as "the Kamper articles." Citations for all articles may be found in Kamper and colleagues (2011), and are available upon request.

Table 1 summarizes the characteristics of these articles and confirms the disciplinary and national diversity of the LBP research community. The articles were published in general medical, spine, pain, and other specialty journals, including physiotherapy, epidemiology, and rheumatology journals. Most journals were highly ranked, including *The New England Journal of Medicine*, *BMJ*, *The Lancet*, *Spine*, and *Pain*. Researchers with European affiliations authored 62% of the articles, those with US or Canadian affiliations authored 25%, and researchers in other regions authored 19% (see Table 1 note). 63% of the articles were published after 2003, that is, after IMMPACT's initial publication recommending specific pain outcome domains.

[2] Full inclusion criteria for the Kamper articles were: peer-reviewed articles accessible in medical electronic databases; published inclusively between 1999-2008; focusing on non-specific LBP; designed as "prospective, longitudinal stud[ies], including randomised controlled trials;" including references to "recovery" or "resolution" in their abstract, methods, or results sections; excluding studies of surgical management of LBP; and written in English or in a language where translation could be easily arranged (Kamper et al. 2011:10). Most Kamper articles focused on chronic LBP, although some focused on acute cases or included both (e.g., to examine which acute cases would become chronic).

Table 1. Characteristics of 79 analyzed studies of low back pain

	Number of articles	Percentage of articles
Study type		
Prospective	39	49%
Clinical trial	32	41%
Economic evaluation	3	4%
Other	5	6%
Journal type		
Spine	35	44%
Pain	10	13%
General medicine	9	11%
Other specialty	25	32%
Year of publication		
1999-2003	29	37%
2004-2008	50	63%
Region of authors' institutional affiliation (see note below)		
Europe	49	62%
US or Canada	20	25%
Other	15	19%

Note: Numbers by region do not add to 79 (100%) because five articles were co-authored by researchers with affiliations in multiple regions.

Source: Self-elaboration.

My second main source of data was interviews I conducted with 20 pain experts: 15 who were lead authors of one or more Kamper articles and 5 who had other valuable experience with pain measurement or standardization. The 15 lead authors had collectively authored 25 (32%) of the 79 Kamper articles and represented a range of specialties and geographic regions (with 6 affiliated with US or Canadian institutions, 7 with European ones, and 2 with institutions from other regions). The five additional pain experts included a US researcher involved with IMMPACT since its inception; an Australian researcher involved in efforts to standardize definitions in LBP research; a US clinical nurse specialist whose pain assessment questionnaires are used in over a dozen countries; and two heads of US pain clinics. For brevity, I refer to researchers who authored Kamper articles as "authors," to the other five interviewees as "specialists," and to interviewees collectively as "pain experts."

This research was approved by the University of Wisconsin-Madison IRB. Interviews were conducted in person, by telephone, or by Skype in 2013, with occasional email follow-ups. Reflecting the regimented schedules of physicians and academics, interviews were typically scheduled in 30- or 60-minute blocks and averaged 45 minutes. For confidentiality, interviewees are referred to by pseudonyms, with identifying geographic and professional details omitted. Interviews were semi-structured. I first asked respondents about their professional training, conference attendance, and what journals they frequently read to better understand their professional affiliations and identities. Thereafter, questions focused on what pain measures respondents used in their work and why, whether they were familiar with organizational efforts to standardize pain measurement, and what they saw as their main research and/or clinical goals. I also read many other publications about pain measurement and standardization and attended academic pain conferences and IMMPACT meetings. I conducted an inductive, grounded theoretical analysis of all texts, including interview transcripts (Glaser and Strauss 1967). The main findings are presented in eight subsections below.

Findings

The Diversity of Pain Measures in LBP Research

One aim of this study was to document and classify all pain outcome measures used in the Kamper articles. However, as discussed below, pain researchers disagree about what constitutes a pain measure. For the initial classification, I adopted a broad view of the term, including measures both directly and indirectly presented as pain measures. For example, if a study treated reduction in a particular score as a mark of improvement in pain, I considered that score a pain measure.

The diversity of pain-related outcome measures in the articles is shown in Table 2, where they are classified into 12 domains. No domain was represented in all studies. The most common domains were pain-related disability (appearing in 73% of articles) and numeric pain intensity (in 63%); all others appeared in fewer than half of articles. There was a large variation in the number and permutations of domains included in individual articles: some included a single outcome measure (e.g., Tubach, Beauté, and Leclerc 2004), while others included up to seven domains (e.g., Smeets et al. 2006). The selected domains showed no clear association with each other, year of publication, or authors' institutional location.

Table 2. Pain-related outcome measures in 79 studies on low back pain

Outcome measure type	Example	Number of articles	Percentage of all articles
Binary pain outcome	Had LBP during study period (yes/no)	11	14%
Numeric pain intensity	Current pain intensity on 100mm VAS (Visual Analogue Scale)	50	63%
Other directly pain-related outcomes	Pain frequency (1-6, no pain to constant pain)	17	22%
Disability/function	Roland-Morris Disability Questionnaire	58	73%
Work	Ability to return to work	26	33%
Global perceived effect (GPE)	5-point self-rating scale, from "completely recovered" to "much worse"	34	43%
Satisfaction	Patient satisfaction on a 3-point scale	12	15%
Measured outcome (not self-report)	Range of motion (fingertip-to-ground distance)	21	27%
Healthcare or medication utilization	Number of pills taken daily	15	19%
Time	Time until disability claim closure	12	15%
Mental health or affect	Beck Depression Inventory	6	8%
Other (less directly pain-related) outcome	Tampa Scale of Kinesiophobia	16	20%

Source: Self-elaboration.

Moreover, *within* each of the 12 domains, there were usually many different measures in use. Pain intensity, for example, was often measured with a 0-10 Numeric Rating Scale (NRS) or a 100-mm Visual Analogue Scale (VAS; a line scored by measuring the distance from 0 to the patient's "X"). However, many studies used other measures: the pain intensity section of the McGill Pain Questionnaire (e.g., Burton et al. 2004); the Short Form 36 Health Survey's (SF-36's) pain subscale (McGuirk et al. 2001); a 0-10 pain relief scale (Collins, Evans, and Grundy 2006); et cetera.

Furthermore, not all NRS- or VAS-based measures were identical. They often varied in the recall period specified, which could be current pain, daytime pain, pain over the last week, pain over the last three months, et cetera—or could remain unspecified. One study constructed a measure averaging NRS scores of current pain, usual pain during the past two weeks, and least pain during the past two weeks (Dunn, Jordan, and Croft 2006); another used a similar procedure but asked for *worst* pain (Jensen et al. 2007); another used worst pain but specified the past *four* weeks (Skillgate, Vingård, and Alfredsson 2007). Nor were labels for VAS and NRS endpoints consistent: the right-most point could be designated "very severe pain" (Van der Roer et al. 2008:446), "unbearable" pain (Ozturk et al. 2006:623), "the worst pain ever" (Peul et al. 2008:181), et cetera. Existing reviews of non-site-specific pain research reveal similar inconsistencies. For example, Marianne Hjermstad and colleagues (2011) found 41 different versions of NRSs among 54 articles (see also Smith et al. 2015).

The pain-related disability category (sometimes referred to as "function") also comprised a diverse set of measures. Those appearing repeatedly included the Roland-Morris Disability Questionnaire (RDQ), the Oswestry Disability Index (ODI), the Quebec Back Pain Disability Scale (QBPDS), and 10-point Patient-Specific Functional Scales. Many studies used modified versions of measures: Federico Balagué and colleagues (1999) excluded one question from the ODI; Kate M. Dunn and colleagues (2006) did the same with the RDQ; Eva Skillgate and colleagues (2007) modified the Hoving Whiplash Disability Questionnaire to refer to back or neck pain instead of whiplash; et cetera.

Overall, the articles showed high variability in terms of chosen domains, chosen measures, and specific implementations of measures. Direct comparison or pooling of findings across studies would often be challenging or impossible. Before fully exploring the causes of this variability, however, I examine whether researchers consider these measures to be measures of *pain*.

What Counts as a Measure of Pain?

The Kamper articles provide initial evidence that researchers disagree about what constitutes a measure of pain. For instance, some studies explicitly presented the Oswestry Disability Index (ODI) as a measure of disability and treated it as a separate concept from pain (as in Unlu et al.'s discussion of "reductions in pain *and* disability" [2008:191; emphasis added]). However, others introduced the scale specifically as a measure of "back pain" (e.g., Giles and Muller 2003:1494). Some trials did not directly classify the ODI or other disability measures, but included no other more direct measures of pain among their outcomes (e.g., Rattanatharn et al. 2004). One can thus assume that the disability measure was considered a measure of pain; otherwise, the study would be evaluating a treatment for pain without ever assessing pain.

Responses to my initial interview question about pain—"How do you measure pain in your work?"—also revealed a range of understandings of what it means to measure pain. A minority of interviewees interpreted this narrowly, volunteering only information about pain intensity measures. Dr. Nussbaum, for example, mentioned the VAS and NRS as his primary pain measures. When I later asked about other measures in his publications, he responded, "[T]he Quebec Back Pain Disability Questionnaire is actually a measure of disability… Not really a pain measurement. And neither, by the way, is the EuroQol…[T]hese are not measures of pain or pain intensity but disability." Some authors took an intermediate position, limiting "pain" to pain intensity, but describing other domains as essential to studies of pain nonetheless: "They're not pain scales *per se*. They're sort of functional impairment. But… with back pain, you really have to deal with both" (Udovitch interview).

The majority of authors I interviewed, however, spontaneously mentioned domains other than pain intensity before I brought them up, presenting them as measures of pain. Indeed, several explicitly rejected the notion of pain as synonymous with pain intensity. Dr. Legac, for example, explained that he assessed pain using measures of intensity, disability, and work capacity because "[s]ome people consider that measuring pain is more or less equivalent to measuring pain intensity…but… you need several domains to be more or less comfortable with what you are doing."

Pain experts with a clinical focus appeared more likely to argue for broad conceptions of pain. [Dr. Nussbaum, who earlier limited "pain" to pain intensity, was an epidemiologist with no clinical experience.] Moreover, several interviewees made a distinction between acute and chronic pain, explaining that measures of pain intensity might suffice when treating the former but not the latter. As one specialist explained, acute pain settings are "fairly transactional situation[s]," in which patients provide a pain intensity number and caregivers respond with a medication or prescription. In contrast,

> [P]eople like myself, who deal with these folks [chronic pain patients], don't really care about the pain rating scale…Typically what we're doing when we assess outcome in chronic pain states is… to assess disability, function, depression, anxiety, anger, work satisfaction, et cetera, et cetera. And so that's why you see this plethora of standardized questionnaires, that are desperately trying to say, what is the pain experience? The chronic pain experience. [Nadeau interview]

Dr. Ostergaard similarly dismissed the importance of pain intensity: "[T]he most interesting thing about doing chronic pain work, the most surprising thing to me, is pain itself is not a vital outcome measure." The very feature of pain that most laypeople see as its essence—how much it hurts—is here presented as *in*essential.

In short, to measure pain may mean to assess pain intensity, to assess intensity along with other domains, or to assess *primarily* other domains. The diversity of pain measures in use reflects, in part, disagreement on what counts as a pain measure to begin with. An explanation for this lack of consensus is presented in the final Findings section.

How Are Pain Measures Justified?

How do pain experts justify their choice of pain measures? Among the Kamper articles, a small proportion included no justification of chosen mea-

sures, even indirect justification through citations (e.g., Brötz et al. 2003). The majority, however, directly justified their measures in one or both of two ways.

First, selected measures were very frequently described as "valid," "reliable," and/or "responsive." For brevity, I sometimes abbreviate this set of characteristics as "validity" below. Rob J. E. M. Smeets and colleagues (2006:5), for example, described their various outcome measures as "a valid and reliable instrument," "a reliable measure of pain intensity," "a reliable, valid…questionnaire," "[having] fairly good validity and reliability," et cetera. Second, pain authors often highlighted that their chosen measures were "from the literature" (e.g., Balagué et al. 1999:2518), were "similar to the outcomes used in other prognostic studies" (Bekkering et al. 2005:1882), or, simply, "ha[ve] been used previously" (Mehling et al. 2005:46). Footnotes were typically provided after each measure, signaling these same attributes: prior validation and/or prior use. A prototypical justification, then, might resemble this: "All these questionnaires have been validated in the literature with the references cited above" (Ferguson et al. 2001:59).

These same themes emerged clearly in my interviews. Indeed, 17 of 20 interviewees spontaneously mentioned "validity and reliability" (Oilman interview), "multiple validations" (Udovitch interview), "validated scales" (Fow interview), et cetera in explaining their choice of measures. Several indicated that they conducted thorough literature reviews at the onset of every project to identify the most valid measures (e.g., Dr. Heddy). I explore what researchers mean by "validity" in the next section, but here note that interviewees often presented validity as a property *inherent in a measure*: a measure was validated or not, and this was context-independent.

Thirteen of my interviewees volunteered that their selected measures came from "the literature." In a minority of cases, "the literature" was seen to represent the expertise of individual or organizational authorities. Dr. Xanthos chose measures "by looking at the literature. Mostly by reading some papers by Deyo… He's kind of the papa of back pain research." Dr. Staab volunteered that he chose his measures based on IMMPACT recommendations. More commonly, however, authors appeared indifferent as to what particular literature was being invoked, mentioning neither specific authors, journals, or organizations. Instead, the most relevant feature of "the literature" was typically its *quantity*. Describing questionnaires as "widely used in the literature" (Nicolson interview), "all over the place" (Ostergaard interview), or "all over the literature" (Udovitch interview) was high praise. "The literature," then, served primarily as a sign of collective justification. Researchers' nearly universal reliance on existing measures (or minor variations thereof) further underscores their desire to conform to established measurement norms.

At face value, either validity or comparability with "the literature" could encourage movement toward standardization. If "validity" is inherent in a measure, then researchers might independently recognize the superior validity of a certain measure, and eventually all come to select the same one. Alternately, the desire to emulate professional peers could, with time, lead to growing comparability of measures as researchers drop rarer measures in favor of those "all over the literature." To assess the plausibility of such processes, I now scrutinize the concepts of "validity" and "comparability."

Scrutinizing Validity

When asked what they meant by "validity," my interviewees often began by providing textbook definitions of the term and/or enumerating key types of validity: "[T]hings like face validity. Do they measure what both physicians and patients would generally perceive to be the correct items, say around back pain? Do they have construct validity?...Criterion validity against standard measures. That's the sort of thing I'm talking about" (Hannigan interview). Other aspects of validity mentioned included "reliability" (e.g., Nussbaum interview) and "sensitivity and specificity" (e.g., Heddy interview). Pain experts expressed concern not only with the validity of specific measures but also with validity at the level of the study. Thus, validity encompassed designing a study with a limited number of outcome measures and sufficient statistical power (to avoid false positive findings due to multiple hypothesis testing) (e.g., Staab and Eusanio interviews). Study-level validity also entailed avoiding patient fatigue, confusion, or non-response, as from an excessive number of questions (e.g., Cata and Hannigan interviews). Finally, multiple researchers mentioned that the validity of scientific findings could be improved by combining studies in a meta-analysis, which depended on comparability of measures (e.g., Hannigan interview); comparability thus bolstered validity. In short, "validity" was a complex and multidimensional concept.

Not all interviewees mentioned all these facets of validity, but no interviewee rejected any of them. That is, I found no evidence of incommensurable beliefs about—or less than universal desire for—validity *in the abstract*. Nonetheless, my interviewees often came to highly divergent evaluations of specific measures. For example, some chose the Roland-Morris Disability Questionnaire over the Oswestry Disability Index because it "showed better reliability and validity" (Maquet interview); others strongly preferred the ODI, arguing that "we should get rid of the Roland Morris" due to its lack of construct validity (Staab interview); and still others rejected both in favor of other scales (Cata and Nicolson interviews), also citing concerns about validity.

How can researchers agree on definitions of validity, yet come to such different conclusions about the validity of specific measures? The answer emerging from my data is that validity was not, in practice, inherent in measures; instead, it was contextual. Local circumstances determined or negated a measure's validity. Moreover, because of its multifaceted nature, two or more facets of validity could come into conflict even in a specific context.

Perhaps the most easily predicted example of this is that validity is culture- and language-dependent. The Kamper articles indicated this, as when Balagué and colleagues (1999:2518) explained that the Oswestry question about sexual function showed "low acceptance" among local participants, and thus was excluded; or when studies specified that they used questionnaires validated in local languages. One interviewee noted that even relatively subtle differences between American and British English could undermine measure validity (Hannigan interview).

Most examples of the locality of validity, however, were based on study setting, design, and goals. Clinicians often described themselves as working under tremendous time pressure— "[W]e have one patient every ten minutes" (Cata interview)—and thus prized measures for their brevity. One clinician desired "measures that can be incorporated into electronic medical records" and which patients could

answer by computer before appointments (Udovitch interview). Researchers conducting telephone interviews had their priorities. Dr. Ehrling rejected the multiple-choice Oswestry scale because it could not be easily used by phone; Dr. Nicolson avoided visual analog scales for the same reason. Dr. Hannigan, who conducted large mail surveys, prized "instrument[s] that people find easy to complete." Several authors limited the number of outcome measures in their studies due to time or space constraints (e.g., Udovitch interview), or due to small sample size (to avoid "too much [statistical] testing," [Staab interview]).

The severity, type, or location of the pain under examination also affected evaluations of a measure's validity. Dr. Udovitch described the Oswestry scale as "most sensitive at the high end of impairment," while the "Roland scale is a bit more sensitive at the middle and lower ends," and chose his measure accordingly. Dr. Xanthos explained that in post-surgical settings, "the Oswestry is very good. But in primary care, it's the Roland Morris score." Dr. Staab, who previously used the McGill Pain Questionnaire, vowed "never to use it again. Because it's really not telling you anything more for people with no specific pain. It might be more interesting with neuropathic pain." As it happens, a specialist examining neuropathic pain considered the McGill questionnaire and found it had too *few* relevant questions; he extended it with additional ones (Eusanio interview). One LBP researcher found certain LBP-specific disability questionnaires sufficiently valid, but switched to a general pain disability scale for comparability with his colleagues, who studied many different pain conditions (Staab interview).

As such examples suggest, researchers frequently found themselves contending with competing notions of validity, especially when defined broadly to include comparability with other studies. Dr. Legac had foreign workers among his patients, so generally used the ODI rather than the RDQ because he could find the former in five relevant languages. However, he used the RDQ in his studies of teenagers, for comparability with "studies in the UK… we tried to reproduce the same protocol." Desire for comparability with others' studies competed with and undermined his desire for consistency across his studies. Using both scales simultaneously was rejected as causing excessive respondent burden. Dr. Udovitch praised the SF-36 disability measure's "good validation and responsiveness," but nonetheless used a back-pain-specific measure because it had "good benchmarking" and because it "relates to the back. The patient's going to like it. So, how's that for rationale?" Comparability with his previous research and patient approval competed with and ultimately trumped the desirable properties of the SF-36. That Dr. Udovitch recognized a degree of arbitrariness in this outcome is suggested by his closing rhetorical question.

The local and multifaceted nature of validity helps explain the earlier example of four authors who came to divergent conclusions about pain-related disability scales. Dr. Maquet worked in primary care—and thus, sharing Dr. Xanthos's evaluation of the RDQ as more valid in such settings than the ODI, preferred the former. Dr. Staab came to the opposite conclusion, rejecting the RDQ for its poor construct validity, as it includes items about sleep and other topics that are not strictly "disability." Dr. Cata, whose research was based on mailed questionnaires, dismissed both measures, reasoning that their length would suppress response rates; he preferred the shorter Chronic Pain Grade. Dr. Nicolson preferred the Patient-Specific Func-

tional Scale for its customizable content and the resulting benefits to content validity and responsiveness. All four authors, then, justified their chosen measures by invoking "validity"—but they were different validities, leading to different choices of measures.

Timmermans and Berg (1997:273) argue that "universality is always 'local universality.'" The present data support a parallel argument, namely, that validity is always local validity—constrained by circumstances including language, study mode and setting, research topic, choice of reference study, et cetera. Simultaneously, such constraint is often incomplete: even in specific contexts, researchers may face a proliferation of competing validities, due partially to the multifacetedness of validity itself.

Scrutinizing Comparability

Most interviewees readily acknowledged that self-reports of pain are subjective and ultimately incomparable: "My six isn't the same as your six, and no two sixes will ever be the same" (Washington interview); "I don't think we can ever really be certain that ratings across respondents on self-report measures are comparable or mean the same thing" (Ehrling interview). Nonetheless, many interviewees saw a lack of interpersonal comparability as posing little problem for their research. What mattered, interviewees repeatedly told me, is "[n]ot actually what the number is, it's whether I can change it" (Bembery interview). For example, if a subject's self-reported pain declined from six to four, that indicated improvement, regardless of how that six corresponded to another person's six (Eusanio interview). In such contexts, it was not interpersonal but *intra*personal comparability that was essential—and that was assumed to be present.

Some pain experts did suggest that for certain types of research, lack of interpersonal comparability could be problematic: "From a perspective of doing epidemiological research, ideally yes[, you'd have interpersonal comparability]" (Zahar interview); "that's a problem when you want to compare groups, yeah" (Staab interview). And yet, even researchers specializing in pain epidemiology sometimes denied that interpersonal incomparability posed a problem (e.g., Dr. Udovitch) or resisted its characterization as "a problem": "[I]n epidemiological studies, you have to just take what you get and assume that seven equals seven, irrespective. [Interviewer: So do you think this poses a problem in your research?] You can call it a problem if you like. It's definitely a factor. But again, we have to go by what the patient says…'Pain is what the patient says it is'" (Cata interview).

As this invocation of McCaffery's dictum suggests, most interviewees supported the idea that "the gold standard" of pain measurement "is patient report" (Udovitch interview)—some quite adamantly (e.g., Dr. Hannigan). Others were more circumspect, defining self-reports as the best *available* rather than the best possible measures (Oilman interview). Overall, however, most researchers did not express a desire for more objective measures of pain.

In her presentation of "a sociology of measurement," Linda Derksen (2002:803) argues that "credible scientific knowledge is produced through the systematic erasure of uncertainty and random variation." All measurement, she notes, involves error, but how much error is deemed "reasonable" is socially negotiated (Derksen 2002:805). For the researchers I interviewed, the error introduced by the subjectivity of self-reported pain *was* considered reasonable. This is shown by a reluctance to even deem it a "prob-

lem," as in the Cata quote above, and also more directly, as when Dr. Hannigan stated, "There is definitely noise in the system. But despite that noise, we are able to pick up some quite strong signals of predictors for onset or outcome of back pain." Interpersonal incomparability of pain is here depicted as "noise" in a system characterized by a high signal-to-noise ratio—that is, something undesired, but forgivable.

In contrast—and as predicted by the valorization of "the literature" described earlier— interviewees highly valued comparability *across studies*. Dr. Xanthos declared inter-study comparability "very important. On a scale from zero to ten, ten." As Dr. Nicolson explained, consistency of measures is "the only thing that really enables us to compare findings across studies. Generally, we don't believe anything these days from one study. So we want to have multiple studies and…we want to be able to pool data." Such views were widespread.

But with *which* studies should researchers pursue comparability? For several pain experts, comparability with their prior work was paramount: "You want to be able to…look at your body of research over a period of time in a way that's somehow comparable" (Zahar interview). Sometimes a single study captured a researcher's attention and became the basis for comparison (e.g., Nicolson interview). Many interviewees desired broad comparability with others, but had to choose precisely *which* others to prioritize. Dr. Staab, as noted, measured pain-related disability with a general rather than LBP-specific scale, to match local colleagues who studied many types and sites of pain. In the process, however, he reduced his work's comparability with that of other LBP researchers. Other interviewees, in contrast, committed to comparability with a specific field, as when Dr. Nussbaum rejected a general measure in favor of measures "commonly used in the field of back pain research." Sometimes measure choice appeared shaped by broad geographical norms: "Scandinavian countries [are] very much used to us[ing] the ODI…England and the United States are more fond of the RDQ" (Staab interview). It should be noted that *every* measurement decision involved a comparability trade-off: increasing comparability with one reference group necessarily decreased comparability with others.

Overall, inter-study comparability, like validity, was multiple and underdetermined: researchers had many reasonable choices for which studies to emulate in terms of measurement outcomes. I next explore how the informational structure of the LBP research community constrained—or failed to constrain—this plethora of choices.

Informational Push and Pull, and Organizational Recommendations

During interviews, I asked pain experts what professional journals they subscribed to or frequently read. Responses were often minimally revealing of professional identity, as many interviewees listed numerous journals from many disciplines, or declined to name *any* journals, noting that they searched in electronic databases rather than restricting themselves to specific periodicals.

Reference lists in the Kamper articles confirm that, when it comes to pulling information for literature reviews, LBP researchers put up few disciplinary or national barriers.[3] Balagué and colleagues (1999),

[3] Most cited articles were authored by researchers in the US, Canada, Europe, or Australia—which, if the Kamper articles are indicative, is precisely where most LBP research is conducted.

for example, included citations from spine, pain, physiotherapy, neurology, rehabilitation medicine, orthopedic, rheumatology, internal medicine, and general medicine journals and monographs, published by authors in no fewer than nine countries. Such citational heterogeneity was typical.

However, when asked what conferences they most frequently attended, pain authors' responses paint a picture of a less integrated field. Six respondents regularly attended an LBP-specific conference, with several of these also attending other pain conferences. These respondents often directly identified as back pain researchers. Dr. Nicolson, for example, though trained as a physiotherapist, explained, "[M]ost of the research I do now I would describe as back pain research across the disciplines rather than physiotherapy as such." Other authors, in contrast, embraced specifically disciplinary identities, as when Dr. Legac—despite prolific publication on low back pain—asserted that he is not "a specialist in pain. I am a rheumatologist." Consequently, the other conferences listed by interviewees, while numerous (e.g., primary care, physiotherapy, rheumatology, epidemiology, complementary and alternative medicine, work disability prevention, health economics, etc.), were rarely listed by more than two interviewees each. Indeed, three authors' preferred conferences did not overlap with those of any other respondent. Geography also affected conference attendance, with researchers most likely to attend conferences on their home continent.

These data suggest that the LBP research community consists of a minority of "insiders" (who attend LBP conferences and see themselves as LBP researchers) along with a majority for whom "low back pain researcher" is a second-order identity, or perhaps not an identity at all. Members of the latter group move in professional circles that are somewhat or even highly isolated from other LBP researchers.

Additional evidence that interviewees often occupied minimally overlapping professional worlds is that they were frequently aware of very different recommendations for measurement standardization. During interviews, I asked if respondents were familiar with the recommendations of OMERACT (the longest-standing group focused on pain-related outcome standards) or IMMPACT (whose recommendations have been highly cited, and which has aimed to publicize its recommendations through "lots of word of mouth, lots of visibility, lots of publications" [Eusanio interview]). Approximately half of the pain authors had heard of one or both groups. ["Approximately" because some gave equivocal answers, for example, "Yeah… I think there's a couple groups" (Udovitch interview).] Five said that they had, or in the future likely would, incorporate IMMPACT's or OMERACT's recommendations in their research. Many interviewees, however, immediately pointed me to *other* groups working on measurement standardization—typically groups reflecting their professional identity or location. Dr. Nicolson, who identified as an LBP researcher, mentioned recommendations made by LBP researchers; Dr. Udovitch, who attended the National Center for Complementary and Alternative Medicine (NCCAM) conference, volunteered, "I know the NCCAM group;" Dr. Heddy, located in northern Europe, suggested "maybe you want to go to a website called COSMIN…It's an effort to improve measurements, and it's run by people in the Netherlands;" et cetera. There were exceptions—as when Dr. Hannigan, located in Europe, mentioned the US-based PROMIS system—but overall, information about specific measurement recommendations appeared to concentrate in disciplinary or geographic pockets, leav-

ing different researchers familiar with different recommendations.

Just how interdisciplinary and international the LBP research community is, then, depends on which informational flows are being examined. The "pull" of information characterizing literature reviews indeed appears largely unimpeded by disciplinary or national borders. In contrast, the "push" of information—as when researchers or organizations attempt to disseminate and promote standard measures—appears at least partially blocked at such borders. This difference may reflect the fact that, thanks to Internet-based search engines, informational pull can be easily achieved without leaving one's office. Publicizing recommendations, however, may require more direct interpersonal interactions, facilitated by professional organizations and conferences. This thesis is supported by Dr. Nussbaum's explanation of the two measures' relative popularity: "[T]he person who developed the Quebec Back Pain Disability…actually didn't travel around the world to go into scientific conferences to promote his instrument. And the Roland-Morris…has [been] widely promoted and therefore more often used." Failure to broadly publicize recommended standards appears common. At the 2014 American Pain Society meeting, I attended a presentation on the NIH Pain Consortium's new recommendations for measuring LBP. When I asked whether the recommendations were publicized outside the US, I was told they were not: because the initiative was NIH-sponsored, it was seen as an American endeavor.

A final point about recommendations for standardizing pain measurement is that they are often not particularly specific. Some list many outcome domains that researchers are invited to select among (e.g., Turk et al. 2008a). Others recommend a specific construct but provide multiple options for measuring it—sometimes not particularly intercomparable options (e.g., Deyo et al. 1998; Dworkin et al. 2005). While flexible or multiple-choice recommendations may appeal to a larger number of researchers, they may also ultimately undermine study comparability.

Are Pain Measures Currently Standardized?

When asked whether pain measures are well-standardized in pain research, interviewees provided a wide range of responses, ranging from "No, definitely not. There are way too many measures" (Ehrling interview) at one end to "that [i.e., standardization] has been quite successful" (Nicolson interview) at the other. Most responses were equivocal, for example, "I think they're not that bad" (Udovitch interview) or "Some are and some are not" (Nussbaum interview). What explains such a variety of opinions?

How standardization is evaluated often depends on what is meant by a "pain measure." Dr. Oilman was largely positive in his evaluation of standardization efforts: "[T]he research community is coming toward some sort of consensus as to which are the best outcome measures to use." When asked to specify those "best outcome measures," he named "a Visual Analogue Scale or Numerical Rating Scale"—that is, he took "pain measures" to refer to pain *intensity* measures. Indeed, in the Kamper articles, the pain intensity measures were relatively small in number and relatively comparable, apart from the problems of inconsistent recall periods and endpoint labels. However, researchers who defined pain broadly generally took a more critical view of the state of pain measurement. Dr. Washington, who espoused a highly multidimensional view of pain ("It's phys-

ical and emotional and functional interference"), described the state of standardization as "very disappointing": "It's really hard to compare all of this research. Everybody's kind of out there doing their own thing." Disagreement about what constitutes a pain measure led to disagreement about whether such measures are well-standardized.

Researchers also varied in whether they focused on standardization of *outcome domains* or of specific measures, and in whether they evaluated standardization in pain research as a whole or in specific areas such as LBP or fibromyalgia research (e.g., Hannigan and Udovitch interviews). Opinions about standardization were more sanguine when aspirations were lower, that is, when domain-related or cause-specific standardization was considered sufficient. Variant measures of a single measure were accepted as consistent with standardization more easily than varying permutations of concepts.

Attempts to harmonize pain measures may be hampered by disagreement on what counts as a pain measure, what counts as standardization, and at what scale standardization efforts should be evaluated. To better understand the first of these, I now explore why researchers have such varied understandings of pain.

The Intractability of Pain

A recurring theme in the Kamper articles is the difficulty of predicting or alleviating pain. Studies seeking to identify prognostic factors report null findings with striking frequency. Eva Vingård and colleagues' (2002:2159) key conclusion—"No predictive factors for recovery were found"—accurately summarizes the results of many Kamper articles. A similar pattern is observed among clinical trials: here, Pieter H. Helmhout and colleagues' (2008:1675) finding that, "No significant differences between the 2 groups were shown for any of the outcome measures, at any time" is emblematic. Given the likelihood of publication bias favoring positive findings, the high proportion of negative findings is all the more remarkable—and gives a strong sense of the intractability of LBP (cf. Turk et al. 2011).

While such findings are clearly discouraging for those who suffer from LBP, they also pose a problem for those who study and treat it. Kristin Barker (2005:17-19) describes the field of rheumatology as struggling with low prestige and a "gloomy mood" due to its lack of effective treatments. Quotes from doctors specializing in pain attest to the disheartening effects of treatment failure: "[H]onestly, it can be a thankless task working with chronic pain patients. Who wants to be confronted with failure every day?" (Kenny 2004:302). One may wonder how LBP researchers (or chronic pain researchers more broadly) contend with their frequent failures. My data provide one answer to this question: LBP researchers (and clinicians) focus on facets of "the pain experience" that appear amenable to improvement or prediction.

Interviewees were often forthcoming about the relationship between choice of outcome measures and expectations of therapeutic efficacy. As Dr. Nussbaum explained, "We measured these outcomes because we expected an effect of the intervention on these outcomes." Dr. Hannigan gave a similar rationale:

> [W]hen we were doing this study on chronic widespread pain, there was a lot of discussion amongst the investigators on what was it that we thought would improve. Did we think that people's pain would im-

> prove? Or did we think that people's pain would stay the same but they could just manage it better? Or would it be that their sleep would improve? Or their fatigue?

Dr. Ostergaard further noted that outcomes should register change before a study's completion. In the Kamper articles, preference for measures indicating success was also apparent: studies often highlighted outcomes improved by the intervention and deprioritized others. Pim Luijsterburg and colleagues (2008:509), for example, found that only one of five measures responded to the intervention but presented precisely this one as the "primary outcome."[4] Pain experts desire measurable success in their work and choose outcome measures to support this goal.

To explore interviewees' therapeutic aspirations, I asked 19 of the pain experts whether they saw pain as something to manage or something to cure. Thirteen described it as something to manage, including several who likened chronic pain to a lifelong condition such as diabetes (e.g., Ostergaard and Washington interviews). Four volunteered that while acute pain can be cured, chronic pain can only be managed. A mere two expressed hope for curing chronic pain. The Kamper articles also repeatedly argued for a management-based model (e.g., Burton et al. 2004). While Baszanger (1992) could not, 25 years ago, predict whether cure-based or management-based approaches would come to dominate in pain medicine, the answer is increasingly clear: most contemporary pain experts view chronic pain as something requiring potentially lifelong management; something that typically cannot be cured.

[4] Such practices can occur even when outcomes are pre-registered: a systematic review found that 30% of registered trials showed primary outcome discrepancies, including treating registered primary outcomes as non-primary in publications (Smith et al. 2013).

One reason LBP studies include many outcome domains besides pain intensity, then, is because researchers believe they have little chance of eliminating or reducing pain, but that they can achieve improvements in other domains. As Dr. Ehrling stated, "Pain is really a secondary outcome of interest since there is no guarantee that pain can be cured or reduced, but we know we can help people improve functionally despite the pain." Dr. Ostergaard's claim that "pain itself is not a vital outcome measure" reflects similar reasoning.

Notably, professionals' framings of pain as largely intractable were often at odds with patients' understandings and hopes. In multiple interviews, clinicians described how substantial "expectations management" (Udovitch interview) was required to "talk people out of the 'cure' belief" (Ehrling interview)—and noted that "many patients will not accept that initially" (Hannigan interview). The multidimensional view of pain, in which professionals aim to improve diverse aspects of "the pain experience" but not necessarily pain intensity itself, often appeared more acceptable to professionals than to patients.

Patient desires notwithstanding, persistent pain is not, at present, a "doable problem" (Fujimura 1996) for clinicians and researchers. Pain experts thus reconfigure it, through the use of varied outcome measures, to make it more "doable," that is, treatable or predictable. This contributes to the diversity of pain measures in use.

Discussion

Despite long-standing calls to standardize pain measurement, especially in the context of clinical trials, contemporary pain research continues to

feature a wide variety of pain-related outcome domains and specific measures of those domains. The present study confirms this finding in the case of LBP research, and uses qualitative methods (including interviews with centrally-positioned pain researchers) to better understand *why* standardization efforts have been only partially successful.

From the attention paid by scholars to pain's "fragile factuality" (Baszanger 1992), one might suspect that epistemological challenges explain the difficulty of settling on standard measures of this profoundly invisible, self-reported condition. However, while LBP researchers do acknowledge the subjectivity of self-reports of pain, they do not see this subjectivity as an epistemological roadblock. In some study designs, interpersonal incomparability of self-reports is deemed irrelevant; in others, it is treated as a form of acceptable measurement error. The "systematic erasure of uncertainty" (Derksen 2002:803) required for scientific knowledge production is achieved without noticeable tension. Bruno Latour's (1987:99) argument that "Nature" does not itself explain scientific closure is supported here: Nature (pain) is admittedly inaccessible to the technoscientific gaze, and yet inspires minimal scientific doubt.

What factors, then, do undermine measurement standardization? Previous research suggests that the desire to avoid type I errors, doubt that specific outcomes will change during the trial period, and concerns about patient burden (Mulla et al. 2015; Turk et al. 2008b) contribute to the rejection of recommended outcomes. These concerns were, indeed, salient to my interviewees, but this study identifies several additional factors undermining standardization, summarized below. Each theoretical point is accompanied by a practical corollary, of potential relevance to participants in standardization processes.

Validity and Comparability Are Local and Multiple

In explaining their choice of pain measures, LBP researchers typically highlight their desires for validity (broadly defined) and for comparability with existing literature. However, the pursuit of these attributes does not lead straightforwardly to standardization, and may, indeed, lead away from it. Validity is contextual, and pain researchers work in many linguistic, cultural, conceptual, and study-design-related contexts; it is also multidimensional and multiple, meaning that multiple valid measures are generally available in any circumstance. Pursuing comparability also has unpredictable results because there are so many answers to the question, "Comparable with *what or whom*?" With one's prior work? With specific studies to be replicated? With geographically local colleagues or disciplinary colleagues? With organizational recommendations (and if so, which organization's)? Overall, researchers face a surfeit of legitimate scientific justifications, which leaves measurement choices underdetermined.

Practical Corollary

Researchers could advocate more effectively for preferred measures if they acknowledged the locality and multifacetedness of validity. That is, rather than simply proclaiming a measure "valid," they might establish and advertise its local validities: can it be administered by phone, mail, and iPad? Will it work across disease types, demographic and linguistic groups, different types of research (e.g., clinical and survey-based), et cetera? Advocates of standardiza-

tion could also constrain measurement choice by acknowledging and adjudicating among competing facets of validity. For example, if one measure has better construct validity but another has higher responsiveness, which should be prioritized? Targets for comparability could be constrained through better diffusion of organizational recommendations, as discussed in the next section.

Barriers to Informational Push Are Greater Than Barriers to Informational Pull

In LBP research, and likely in many other scholarly fields, there is a disjuncture between how knowledge is pulled and how it is pushed. Informational pull—the deliberate seeking out of information, as for a literature review—appears minimally impeded by national or disciplinary boundaries. In contrast, informational push—publicizing information to recipients who are not actively seeking such information—reveals clear effects of such boundaries. Organizational recommendations for standard measures, which depend on informational push, are thus often unknown to researchers outside of specific disciplinary or geographic pockets. This limits the recommendations' potential to harmonize pain measurement. In addition, recommendations sometimes present very general or multiple-choice options; these may constrain measurement choices weakly even if they reach their intended audience.

Practical Corollary

Campaigners for standardization would do well to acknowledge the substantial effort required to push recommendations to a diverse research community (see Timmermans and Epstein [2010"81] on standardization as an "active, time- and resource-intensive process"). Advocates may need to physically cross disciplinary and national borders, as by publicizing recommendations at conferences in fields and countries beyond those of the standards' originators. Doing so may lead to a greater number of competing recommendations; if so, the recommendations themselves may need to be harmonized. Ideally, this would not occur through multiple-option lists: suggestions for specific, individual measures are likely to enhance standardization more than menu-style ("pick one from column A and one from column B") recommendations. Finally, institutional *requirements* would be more effective than recommendations at constraining the choice of measures; for example, if a large funding agency began requiring the use of specific pain measures, standardization would be hastened.

Researchers Select Outcome Measures to Enhance Their Professional Credibility and Success

LBP researchers disagree about what, precisely, is meant by "pain." While for some, pain refers to pain intensity, for others, pain encompasses multiple, varied domains (e.g., "disability, function, depression, anxiety, anger, work satisfaction, etc., etc." [Nadeau interview]). A key reason many researchers take a broad view of pain is that this enables them to focus, in their clinical and academic work, on outcomes less resistant to improvement and prediction than pain intensity. That is, researchers redefine the problem of pain to improve their chances of professional success. In particular, many LBP experts embrace a multidimensional, management-based approach to pain, which treats the reduction of pain intensity as an unlikely or secondary goal. However, patients frequently resist this approach (cf. Kamper et al. 2010 on the salience of pain elimination to patients). The idea that pain intensity "is not a vital

outcome measure" (Ostergaard interview) is a hard sell to make to pain sufferers and may contribute to their poor relations with healthcare providers (Jackson 2011). One may ask whether researchers' interest in multiple facets of pain should be lauded for its holism or critiqued for its negation of patient perspectives.

Practical Corollary

Researchers should include standard measure(s) of pain intensity in every study, even if other outcome domains are also represented. This way there is at least one measure that is constant across studies—and one that reflects patient priorities. More broadly, reflection on how definitions and measures of pain contribute to the nonalignment of expert and patient goals could improve relations between the two.

Standardization May Be More Appealing as an Ideal Than in Practice

In publications and interviews, LBP researchers were enthusiastic supporters of measurement standardization—in the abstract. When asked about measurement choice in specific scenarios, however, they often had good reasons to reject standardization (including those pertaining to validity, comparability, and professional success described above).

Practical Corollary

Arguably, before any other considerations, researchers may wish to acknowledge the downsides of measurement standardization: It can undermine local validity and comparability, and it may reduce the appearance of professional success. Before asking, "How do we achieve standardization?," researchers should ask, "Do we *want* standardization?"

These findings emerge from a close examination of LBP research but are likely relevant to many other scientific fields, including social scientific ones. Desires for valid and comparable measures, and for measurable professional success, are widespread. So, too, is the asymmetry between relatively unconstricted informational pull and more narrowly channeled informational push. Such factors may undermine measurement standardization in many contexts.

This study confirms that standardization is a distributed activity occurring across social networks (Timmermans and Berg 1997; Timmermans and Epstein 2010) by highlighting how standardization is impeded when networks are constricted, as when recommendations fail to diffuse across professional subgroups. In this case, the "resistance" stage in the "life course of standards" (Timmermans and Epstein 2010:74) takes an indirect or accidental form: resistance results from sheer ignorance of the standard's existence. At the same time, integrated networks *also* present challenges for standardization: researchers plugged into multiple professional circles are particularly likely to encounter multiple competing standards with no clear basis for adjudicating among them. Timmermans and Epstein (2010:79) note that voluntary standards may not catch on without "built-in incentives [to] promote compliance." Indeed, strong incentives for adoption are typically absent for recommended pain measures, while competing incentives (e.g., pursuing local validity or comparability, building professional credibility, etc.) are numerous. This study also shows that standardization can be a difficult phenomenon to demarcate. LBP researchers varied

widely in whether they considered pain measurement to be well standardized, with their opinions shaped by their definitions of pain and of standardization itself.

Cases of successful standardization feature many of the factors undermining pain measurement standardization, but in reverse. Joan Fujimura (1996) describes the unification of 1980s cancer research around an oncogene-based "theory-methods package." A large part of this package's appeal was that it suited scientists' professional goals, generating "publications, and academic career boosts, along with applied products…and financial profits" (Fujimura 1996:69). Indeed, because technologies in this package could be (and were) commercialized, corporations had financial incentive to engage in informational push, as when Du Pont advertised its "OncoMouse" in multiple science journals (Fujimura 1996:7-8). For this reason and others, news of the oncogene approach diffused thoroughly in the interdisciplinary cancer research community. Similarly, the successfully standardized protocols described by Timmermans and Berg (1997:286, 289) helped advance professionals' goals—and when they did not, were resisted. They were also effectively diffused by a host of institutions and organizations. Because few studies examine *measurement* standardization, it is difficult to comparatively assess my findings about measurement validity and comparability, but the importance of standards both supporting professional goals and being forcefully diffused is supported by these examples.

This study is limited by focusing on articles on low back pain published before 2009 and on interviews conducted in 2013. Research using more recent articles in other substantive areas, as well as my readings of newer articles, suggest that the lack of standardization described here is restricted neither to pre-2009 articles nor to LBP studies (Mulla et al. 2015), but further research formally analyzing more recent publications and measurement practices is warranted. Though desires for valid and comparable measures, and for measurable professional success, are likely common in many regions and fields, further qualitative work could build on and clarify the generalizability of these findings to other scientific domains. Another limitation is that I lacked data on patient characteristics and how they may shape pain measurement. This could also be an important direction for future research, given evidence that stereotypes—such as of "stoic men" and "hysterical women" (Samulowitz et al. 2018), or of racial/ethnic minorities as less sensitive to pain than whites (Hoffman et al. 2016)—shape clinicians' assessment of pain. This, in turn, shapes the treatment of pain, for better or worse. Nearly half a million Americans have died from opioid overdoses since the turn of the century (Zajacova et al. 2021), highlighting how determinations of pain may have life-or-death consequences. The role of pharmaceutical companies in developing markets for pain medications is outside the scope of this article (and was not mentioned as a factor shaping pain measurement by interviewees) but is, nonetheless, an important topic for research.

Scientists continue to search for objective measures of pain (e.g., Wager et al. 2013; Davis et al. 2020). Would achieving this goal end the difficulties of standardizing pain measurement? An objective pain measure—if valid for clinical use, for assessing intermittent pain, et cetera—could revolutionize patient-provider interactions, since accusations of malingering or drug-seeking could be confirmed or denied. However, my findings suggest that such a measure would have only moderate effects on research study design since many pain researchers to-

day seek not more *objective* pain measures but more *treatable* ones.

Thus—perhaps counterintuitively—the development of successful treatments for pain has greater potential to standardize pain measurement than the development of objective pain measures. If pain intensity comes to be seen as something highly treatable, or if chronic pain becomes a curable condition, then pain studies may come to prioritize pain intensity over the diverse sets of other domains currently in use. Reducing these alternate measures would enhance standardization. Standardization of pain measures—which is pursued so that researchers may better advance their knowledge of pain, including how to treat it—may thus ultimately be a *result* of therapeutic advances as well as, ideally, their *cause*. We may hope that this cycle progresses rapidly, to sooner reduce the suffering caused by pain.

References

Aicher, Bernhard et al. 2012. "Pain Measurement: Visual Analogue Scale (VAS) and Verbal Rating Scale (VRS) in Clinical Trials with OTC Analgesics in Headache." *Cephalalgia* 32(3):185-197.

Balagué, Federico et al. 1999. "Recovery of Severe Sciatica." *Spine* 24(23):2516-2524.

Barker, Kristin. 2005. *The Fibromyalgia Story: Medical Authority & Women's Worlds of Pain*. Philadelphia: Temple University Press.

Baszanger, Isabelle. 1992. "Deciphering Chronic Pain." *Sociology of Health & Illness* 14(2):181-215.

Bekkering, Geertruida E. et al. 2005. "Prognostic Factors for Low Back Pain in Patients Referred for Physiotherapy: Comparing Outcomes and Varying Modeling Techniques." *Spine* 30(16):1881-1886.

Borkan, Jeffrey et al. 1995. "Talking about the Pain: A Patient-Centered Study of Low Back Pain in Primary Care." *Social Science & Medicine* 40(7):977-988.

Brötz, Doris et al. 2003. "A Prospective Trial of Mechanical Physiotherapy for Lumbar Disk Prolapse." *Journal of Neurology* 250(6):746-749.

Burton, A. Kim et al. 2004. "Long-Term Follow-Up of Patients with Low Back Pain Attending for Manipulative Care: Outcomes and Predictors." *Manual Therapy* 9(1):30-35.

Cassar-Pullicino, Victor. 1998. "MRI of the Ageing and Herniating Intervertebral Disc." *European Journal of Radiology* 27(3):214-228.

Collins, Daniel L., Joseph M. Evans, and Reed H. Grundy. 2006. "The Efficiency of Multiple Impulse Therapy for Musculoskeletal Complaints." *Journal of Manipulative and Physiological Therapeutics* 29(2):162e1-162e9.

Davis, Karen D. et al. 2020. "Discovery and Validation of Biomarkers to Aid the Development of Safe and Effective Pain Therapeutics: Challenges and Opportunities." *Nature Reviews Neurology* 16(7):381-400.

Derksen, Linda. 2002. "Towards a Sociology of Measurement: The Meaning of Measurement Error in the Case of DNA Profiling." *Social Studies of Science* 30(6):803-845.

Deyo, Richard A. et al. 1998. "Outcome Measures for Low Back Pain Research: A Proposal for Standardized Use." *Spine* 3(18):2003-2013.

Dunn, Kate M., Kelvin Jordan, and Peter R. Croft. 2006. "Characterizing the Course of Low Back Pain: A Latent Class Analysis." *American Journal of Epidemiology* 163(8):754-761.

Dworkin, Robert H. et al. 2005. "Core Outcome Measures for Chronic Pain Clinical Trials: IMMPACT Recommendations." *Pain* 113(1):9-19.

Elliott, Alison M., Blair H. Smith, and W. Alastair Chambers. 2003. "Measuring the Severity of Chronic Pain: A Research Perspective." *Expert Review of Neurotherapeutics* 3(5):581-590.

Ferguson, Sue A. et al. 2001. "Predicting Recovery Using Continuous Low Back Pain Outcome Measures." *Spine Journal* 1(1):57-65.

Fortier, Isabel et al. 2012. "Harmonizing Data for Collaborative Research on Aging: Why Should We Foster Such an Agenda?" *Canadian Journal on Aging* 31(1):95-99.

Fujimura, Joan. 1996. *Crafting Science: A Sociohistory of the Quest for the Genetics of Cancer*. Cambridge, MA: Harvard University Press.

Giles, Lynton G. F. and Reinhold Muller. 2003. "Chronic Spinal Pain: A Randomized Clinical Trial Comparing Medication, Acupuncture, and Spinal Manipulation." *Spine* 28(14):1490-1502.

Glaser, Barney and Anselm L. Strauss. 1967. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. New York: Aldine Publishing Company.

Glenton, Claire. 2003. "Chronic Back Pain Sufferers—Striving for the Sick Role." *Social Science & Medicine* 57(11):2243-2252.

Graf, Jonathan. 2010. "Analgesic Use in the Elderly: The 'Pain' and Simple Truth." *Archives of Internal Medicine* 170(22):1976-1978.

Helmhout, Pieter H. et al. 2008. "Isolated Lumbar Extensor Strengthening versus Regular Physical Therapy in an Army Working Population with Nonacute Low Back Pain: A Randomized Controlled Trial." *Archives of Physical Medicine and Rehabilitation* 89(9):1675-1685.

Hjermstad, Marianne Jensen et al. 2011. "Studies Comparing Numerical Rating Scales, Verbal Rating Scales, and Visual Analogue Scales for Assessment of Pain Intensity in Adults: A Systematic Literature Review." *Journal of Pain and Symptom Management* 41(6):1073-1093.

Hoffman, Kelly M. et al. 2016. "Racial Bias in Pain Assessment and Treatment Recommendations, and False Beliefs about Biological Differences between Blacks and Whites." *Proceedings of the National Academy of Sciences of the United States of America* 113(16):4296-4301.

IASP [International Association for the Study of Pain] website. Retrieved December 30, 2024 (https://www.iasp-pain.org/membership/).

Institute of Medicine; Committee on Pain, Disability, and Chronic Illness Behavior. 1987. *Pain and Disability: Clinical, Behavioral, and Public Policy Perspectives*. Washington, DC: National Academies Press.

Institute of Medicine; Committee on Advancing Pain Research, Care, and Education. 2011. *Relieving Pain in America: A Blueprint for Transforming Prevention, Care, Education, and Research*. Washington, DC: National Academies Press.

Jackson, Jean. 2011. "Pain: Pain and Bodies." Pp. 370-387 in *A Companion to the Anthropology of the Body and Embodiment*, edited by F. Mascia-Lees. West Sussex: Blackwell Publishing.

Jensen, Tue Secher et al. 2007. "Magnetic Resonance Imaging Findings as Predictors of Clinical Outcome in Patients with Sciatica Receiving Active Conservative Treatment." *Journal of Manipulative and Physiological Therapeutics* 30(2):98-108.

Kamper, Steven J. et al. 2010. "How Little Pain and Disability Do Patients with Low Back Pain Have to Experience to Feel That They Have Recovered?" *European Spine Journal* 19(9):1495-1501.

Kamper, Steven J. et al. 2011. "How Is Recovery from Low Back Pain Measured? A Systematic Review of the Literature." *European Spine Journal* 20(1):9-18.

Kenny, Diana T. 2004. "Constructions of Chronic Pain in Doctor-Patient Relationships: Bridging the Communication Chasm." *Patient Education and Counseling* 52(3):297-305.

Latour, Bruno. 1987. *Science in Action*: *How to Follow Scientists and Engineers through Society*. Cambridge, MA: Harvard University Press.

Luijsterburg, Pim et al. 2008. "Physical Therapy Plus General Practitioners' Care versus General Practitioners' Care Alone for Sciatica: A Randomised Clinical Trial with a 12-Month Follow-Up." *European Spine Journal* 17(4):509-517.

McCaffery, Margo and D. M. Thorpe. 1989. "Differences in Perception of Pain and the Development of Adversarial Relationships among Health Care Providers." Pp. 113-125 in *Advances in Pain Research and Therapy, Volume 11*, edited by Ch. S. Hill and W. S. Fields. New York: Raven.

McGuirk, Brian et al. 2001. "Safety, Efficacy, and Cost Effectiveness of Evidence-Based Guidelines for the Management of Acute Low Back Pain in Primary Care." *Spine* 26(23):2615-2622.

Mehling, Wolf E. et al. 2005. "Randomized, Controlled Trial of Breath Therapy for Patients with Chronic Low-Back Pain." *Alternative Therapies* 11(4):44-52.

Mulla, Sohail M. et al. 2015. "Reporting of IMMPACT-Recommended Core Outcome Domains among Trials Assessing Opioids for Chronic Non-Cancer Pain." *Pain* 156(9):1615-1619.

Ozturk, Bulent et al. 2006. "Effect of Continuous Lumbar Traction on the Size of Herniated Disc Material in Lumbar Disc Herniation." *Rheumatology International* 26(7):622-626.

Peul, Wilco C. et al. 2008. "Influence of Gender and other Prognostic Factors on Outcome of Sciatica." *Pain* 138(1):180-191.

Raja, Srinivasa N. et al. 2020. "The Revised International Association for the Study of Pain Definition of Pain: Concepts, Challenges, and Compromises." *Pain* 161(9):1976-1982.

Rattanatharn, Rattana et al. 2004. "Effectiveness of Lumbar Traction with Routine Conservative Treatment in Acute Herniated Disc Syndrome." *Journal of the Medical Association of Thailand* 87(Suppl 2):S272-S277.

Samulowitz, Anke et al. 2018. "'Brave Men' and 'Emotional Women': A Theory-Guided Literature Review on Gender Bias in Health Care and Gendered Norms towards Patients with Chronic Pain." *Pain Research and Management* 2018:6358624.

Scarry, Elaine. 1985. *The Body in Pain*. New York, Oxford: Oxford University Press.

Schiavenato, Martin and Kenneth Craig. 2010. "Pain Assessment as a Social Transaction: Beyond the 'Gold Standard.'" *Clinical Journal of Pain* 26(8):667-676.

Skillgate, Eva, Eva Vingård, and Lars Alfredsson. 2007. "Naprapathic Manual Therapy or Evidence-Based Care for Back and Neck Pain: A Randomized, Controlled Trial." *Clinical Journal of Pain* 23(5):431-439.

Skloot, Rebecca. 2010. *The Immortal Life of Henrietta Lacks*. New York: Crown Publishers.

Smeets, Rob J. E. M. et al. 2006. "Active Rehabilitation for Chronic Low Back Pain: Cognitive-Behavioral, Physical, or Both? First Direct Post-Treatment Results from a Randomized Controlled Trial." *BMC Musculoskeletal Disorders* 7:5.

Smith, Marion V. 2008. "Pain Experience and the Imagined Researcher." *Sociology of Health & Illness* 30(7):992-1006.

Smith, Shannon M. et al. 2013. "Discrepancies between Registered and Published Primary Outcome Specifications in Analgesic Trials: ACTTION Systematic Review and Recommendations." *Pain* 154(12):2769-2774.

Smith, Shannon M. et al. 2015. "Quality of Pain Intensity Assessment Reporting: ACTTION Systematic Review and Recommendations." *The Journal of Pain* 16(4):299-305.

Timmermans, Stefan and Marc Berg. 1997. "Standardization in Action: Achieving Local Universality through Medical Protocols." *Social Studies of Science* 27(2):273-305.

Timmermans, Stefan and Steven Epstein. 2010. "A World of Standards but Not a Standard World: Toward a Sociology of Standards and Standardization." *Annual Review of Sociology* 36:69-89.

Treede, Rolf-Detlef et al. 2015. "A Classification of Chronic Pain for ICD-11." *Pain* 156(6):1003-1007.

Tubach, Florence, Julien Beauté, and Annette Leclerc. 2004. "Natural History and Prognostic Indicators of Sciatica." *Journal of Clinical Epidemiology* 57(2):174-179.

Tugwell, Peter et al. 2007. "OMERACT: An International Initiative to Improve Outcome Measurement in Rheumatology." *Trials* 8:38.

Turk, Dennis C. et al. 2003. "Core Outcome Domains for Chronic Pain Clinical Trials: IMMPACT Recommendations." *Pain* 106(3):337-345.

Turk, Dennis C. et al. 2008a. "Identifying Important Outcome Domains for Chronic Pain Clinical Trials: An IMMPACT Survey of People with Pain." *Pain* 137(2):276-285.

Turk, Dennis C. et al. 2008b. "Analyzing Multiple Endpoints in Clinical Trials of Pain Treatments: IMMPACT Recommendations." *PAIN* 139(3):485-493.

Turk, Dennis C., Hilary D. Wilson, and Alex Cahana. 2011. "Treatment of Chronic Non-Cancer Pain." *Lancet* 377(9784):2226-2235.

Unlu, Zeliha et al. 2008. "Comparison of 3 Physical Therapy Modalities for Acute Pain in Lumbar Disc Herniation Measured by Clinical Evaluation and Magnetic Resonance Imaging." *Journal of Manipulative and Physiological Therapeutics* 31(3):191-198.

VanDenKerkhof, Elizabeth G., Madelon L. Peters, and Julie Bruce. 2013. "Chronic Pain after Surgery: Time for Standardization? A Framework to Establish Core Risk Factor and Outcome Domains for Epidemiological Studies." *Clinical Journal of Pain* 29(1):2-8.

Van der Roer, Nicole et al. 2008. "Economic Evaluation of an Intensive Group Training Protocol Compared with Usual Care Physiotherapy in Patients with Chronic Low Back Pain." *Spine* 33(4):445-451.

Vingård, Eva et al. 2002. "Seeking Care for Low Back Pain in the General Population: A Two-Year Follow-Up Study: Results from the MUSIC-Norrtalje Study." *Spine* 27(19):2159-2165.

Wager, Tor D. et al. 2013. "An fMRI-Based Neurologic Signature of Physical Pain." *New England Journal of Medicine* 368(15):1388-1397.

Whelan, Emma. 2003. "Putting Pain to Paper: Endometriosis and the Documentation of Suffering." *Health* 7(4):463-482.

Whelan, Emma. 2009. "How Classification Works, or Doesn't: The Case of Chronic Pain." Pp. 169-182 in *The SAGE Handbook of Case-Based Methods*, edited by D. Byrne and Ch. Ragin. London: Sage.

Zajacova, Anna, Hanna Grol-Prokopczyk, and Zachary Zimmer. 2021. "Sociology of Chronic Pain." *Journal of Health and Social Behavior* 62(3):302-317.

Citation

Grol-Prokopczyk, Hanna. 2025. "Why Are There So Many Ways to Measure Pain? Epistemological and Professional Challenges in Medical Standardization." *Qualitative Sociology Review* 21(1):46-72. Retrieved Month, Year (http://www.qualitativesociologyreview.org/ENG/archive_eng.php). DOI: https://doi.org/10.18778/1733-8077.21.1.03