

David Hadorn

Wellington, New Zealand

The Scientific Measurement of Subjective Improvement

The scientific measurement of subjective qualities, such as degree of pain or an increase or decrease in pain, can be considered on at least two levels. First is the question of whether uncontrolled, individual experiences can be considered as valid at all for purposes of assessing health and health outcomes. Many proponents of randomized controlled trials would answer 'no' to this question.

A good example of this kind of thinking can be found in a report prepared by the U.S. Drug Enforcement Administration (DEA) in over-ruling its own administrative law judge, Frances Young, who in 1988 had concluded that cannabis is a safe and effective medicine for several conditions. Much of the evidence upon which this judgement was based came in the form of patient self-reports, which the DEA dismissed as valueless.

Why do scientists consider stories from patients and their doctors to be unreliable? First, sick people are not objective scientific observers, especially when it comes to their own health. We have all heard of the placebo effect. . . . Any mind-altering drug that produces euphoria can make a sick person think he feels better.

But is it really possible for a sick person to 'think he feels better', when in fact he does not actually feel better? In my judgement, most philosophers would answer 'no' to this question, just as they would conclude that it is not possible for someone to be wrong in believing that he is happy (or sad), or that he seems to see red, or has an itch, or that a joke is (seems) funny, and so forth. The immunity to error of such private sensations flows directly from the cogito doctrine in Descartes' Meditations – roughly, that we cannot be mistaken about the reality and existence of our own cognitive activities.

Moreover, to suggest that sick people cannot provide valid information concerning the extent of pain or disability they experience and any changes experienced over time, seems equally misguided. Indeed, patients are widely considered to be the most valid source of information on the health outcomes brought about by treatments. Certainly clinicians obtain and act on on such information every day. For these reasons, dismissing patients' reports of improvement as 'mere placebo effects' is fundamentally incoherent.

Even granted the validity of personal subjective experience, however, a second level of skepticism is still possible, namely whether such experiences can be measured.

Measuring Subjective States

Measurement is the hallmark of science. Defined as a process by which numbers are assigned to objects or situations based on the extent to which certain attributes are perceived to be present in those objects or situations, measurement is accepted as valid if independent observers obtain the same (or similar) numbers when applying the process to the same (or similar) objects or situations. Thus, the temperature of a liquid can be measured with validity because multiple independent observers regularly obtain approximately the same value when reading a thermometer. (Of course, the thermometer reading can be in error for any number of other reasons, e. g., if the mercury column is leaking.)

Can the degree of pain, disability, or depression experienced by patients also be measured? Can the degree of improvement (or worsening) in these attributes be measured? Certainly the large number of health outcome studies conducted over the past decade would suggest that the answer to these questions is 'yes.' In such studies, questionnaires completed by patients measure (changes in) patients' health just as thermometers measure (changes in) the temperature of liquids.

But something seems fundamentally different about the two situations. Most starkly, there is no way to check the numbers obtained from patient-completed questionnaires, unlike a thermometer which is accessible to all – and which can be checked with a separate thermometer. Because there is no way to get inside patients' heads to check the level of pain, etc., external validation will forever remain unavailable for subjective sensations. One could conceivably design experiments to check the correlations between reported pain intensity and heart rate, blood pressure, galvanic skin response, and the like. However, none of these quantities is pain, and the results of such experiments would likely remain inconclusive.

Because the patient-centered process used to assign numbers to health and health outcomes is not amenable to inter-observer reliability testing, health outcomes research can never aspire to the level of objectivity seen in chemistry and physics. However, more than a century of psychophysics research has shown that people are easily able to quantify subjective sensations such as the relative brightness of lights or heaviness of objects. Moreover, dozens if not hundreds of studies conducted over the past few decades have shown that self-reported measurements of patients' health are reliable in two other important senses (i.e., besides the inter-observer sense): (1) when patients re-take questionnaires a short time later (before their health status has time to change substantially), similar answers are generally observed (test-retest design) and (2) answers to questions covering a single dimension of health (e.g. pain) generally correlate more highly than questions that purport to measure different dimensions (e.g., pain versus disability). Indeed, the extent of correlation observed in these settings compares favourably to the inter-observer reliability observed for many 'hard' clinical tests, including x-rays and blood pressure.

Another potentially serious problem with measuring subjective states is the observed variance across patients with respect to tolerance for pain and disability, including the use of medication and mobility aids, and in how they report pain (and pain relief). Thus, some patients might report a certain level of pain as 'mild', whereas others might grade the 'same level of pain' (if this term can be used meaningfully) as 'moderate.' Similarly, some patients might assign substantially different values to the 'same' level of pain relief or to the improvement in mobility afforded by a cane.

Although there is no way to expunge this problem, at least three approaches might reduce its practical impact. First and foremost is to seek safety of numbers. Thus, by collecting large amounts of data we can smooth out variations in (say) pain thresholds and pain-reporting tactics, permitting the analysis to report what the average patient (of a certain clinical kind) reported concerning the effect of a particular treatment. A second possible avenue to standardizing responses concerning subjective sensations is to supplement the various possible questionnaire responses with descriptions or examples of each level. For example, mild pain might be described as pain which is sufficiently bothersome to report, but which generally does not limit daily activities and which generally responds to over-the-counter pain medicine. Finally, inter-patient variability might be reduced with respect to the reporting of health states (and changes in health status) by fostering a sense of duty to report in ways that are generally accepted by society. Thus, patients might be asked to adjust their responses for any known personal idiosyncrasies and to keep in mind that their responses will be used to assess the value of medical treatments from a societal perspective.

The Sweater Experiment

The principal issues involved in accepting uncontrolled observations as evidence of causal efficacy occur regularly in all aspects of life. Consider the following dialogue:

Common-sensical Man: It's cold in here. I think I'll put on my sweater.

Die-hard RCT Advocate: Wait a minute. You don't know that putting on a sweater will make you feel warmer, or more generally, that putting on sweaters makes people feel warmer.

CM: What do you mean?

DRCTA: There's been no prospective, blinded, randomized trial. All we have is a bunch of uncontrolled personal subjective experiences in which people say that they think they feel warmer after they put a sweater on.

CM: What's wrong with that?

DRCTA: That sort of experience is subject to all sorts of biases. For example, people who put sweaters on expect to start feeling warmer. They know they are supposed to feel warmer when they put a sweater on, so when they put a sweater on they feel warmer. It's called the placebo effect.

CM: But if someone were to put a sweater on me while I wasn't paying attention or while I was asleep, I'd still feel warmer than I would otherwise when I awoke.

DRCTA: You can't know that for sure. There's no control. To obtain valid causal knowledge about the role of sweaters in making people feel warmer you would need to run an experiment with at least 40 or 50 people along the following lines:

1. Randomize and blindfold study participants on cold day
2. Place sweater on randomly selected half of participants and placebo garment (very thin with large holes) on other half
3. Have participants fill out 'coldness questionnaire'
4. Conduct statistical analysis of results.

CM: That seems like a lot of trouble.

DRCTA: Well, there's no getting around it if you want to avoid bias. And by the way, the 'coldness questionnaire' would first have to be validated in a separate study sample. For example, let's say people were asked how cold they felt on a scale of 0 to 100, where 0 meant that they were not cold at all (perhaps just starting to feel a little warm) and 100 meant they were about to pass out, with shivering and teeth chattering. You'd have to test the questionnaire to make sure that the numbers given by respondents corresponded to more objective measures, like their body temperatures, degree of shivering, and the number of goose-bumps.

CM: All that work is needed just to prove that wearing a sweater makes people feel warmer?

DRCTA: I understand your skepticism. But consider for a moment an old, threadbare, hole-laden sweater that a widower keeps for

sentimental reasons and which he puts on when it's mildly cold outside. He thinks it makes him feel warmer. But is it not possible that such a sweater acts primarily through the placebo effect? And would not a randomized trial identify this effect?

CM: Perhaps so. But that's not the normal situation.

DRCTA: Indeed not, but by studying those cases in which we lead into error we can hope to understand more generally how to avoid error. And just as bulkier sweaters grade off into threadbare ones, so do truly effective medical treatments grade off into sham ones. RCTs are powerful tools for distinguishing the one from the other.

CM: Again, perhaps so, but I'm not sure an RCT is really necessary in my case. I know my sweater makes me feel warm. (Puts on sweater and recalls a previous engagement.)

CM quite rightly considers his personal subjective experiences as sufficient warrant for believing that sweaters keep him warm (and, by extension keep people warm in general). Similarly, most of our knowledge about the way the world works – including what works in the way of medical treatments – comes from such personal experiences. To deny the reality of this knowledge would be pedantic in the extreme.

Nonetheless, there is an important element of truth to DRCTA's concerns. Specifically, although it is wrong to think that people might be mistaken about thinking they feel warmer, or that randomised trials are required to prove the fact, it is nonetheless highly appropriate to try to measure those feelings. Doing so could provide a more powerful method than unformalised subjective reporting ever could for such scientific purposes as attempting to understand how, in the case at hand insulating properties depend on the kind of material used, thickness of the fabric, type of external coatings or treatments, and the like. We could also study how people differ in their sensations of cold and warmth, with potentially useful results in terms of understanding the sensori-neural pathways responsible for feeling cold or warm.

One common problem is that our infallible subjective sensations can at times conflict with objective measurements. For example, imagine that tests of cold-feeling were performed and it was determined that subjects regularly reported feeling warmer in a room

with air held at 20°C than in another room with an air temperature of 18°C (say), or while wearing sweaters that objectively did not insulate as well as others. Who would be 'right'? Well, if a sweater manufacturer wished to make a profit he would likely decide that subjective sensations 'rule' (as my teenagers might say) and choose to manufacture the sweaters reported to provide greatest warmth – irrespective of objective measurements. The proof of the pudding is in the eating, and in this case feeling warm is the proof of the sweater.

These same considerations apply to health care, of course. If forced to choose between health care services that improve objective measures of health, such as range of motion in an arthritic hip, as opposed to those that relieve (subjectively reported) pain and mobility in the same hip, it is obvious which we should prefer. Subjective improvements in pain and function are the end goal of health services, and to the extent we can measure those improvements we will have more confidence in our ability to deliver them.

Conclusion

The take-home messages from this discussion are (1) that people cannot be wrong in thinking they feel warmer (or happier, in less pain, etc.), (2) we can increase the power of our efforts to discover how to bring about such desirable changes in our subjective sensations by subjecting them to measurement, and (3) such measurements are potentially as valid as those of blood pressure, heart function, and similar clinical processes and functions.

Not every subjective feeling or sensation is measurable, however, and we must take care to distinguish those that are from those that aren't. In this regard, measurement procedures must pass tests of inter-observer or test-retest reliability, as appropriate. Results of such tests should be presented in reports of studies using instruments to measure subjective experience.

It is this connection between subjective and objective – of feelings and the measurement of feelings – that is both so problematic and yet (often) so fruitful in science and medicine. To the extent we can understand and manipulate that connection we will be well on our way toward developing more effective health services – and warmer sweaters.