Changing The Definition Of The Kilogram

van Loo, Hanna M.; Romeijn, Jan-Willem; Kendler, Kenneth S.

Published in:
Philosophy psychiatry & psychology

DOI:
10.1353/ppp.2019.0046

Formatted: Width: 8.5", Height: 11"

CHANGING THE DEFINITION OF THE KILOGRAM

Insights For Psychiatric Disease Classification

Hanna M. van Loo, Jan-Willem Romeijn, & Kenneth S. Kendler

ABSTRACT: Attempts to improve fundamental definitions or classifications are not unique for psychiatry. In a 'hard science' such as metrology—the discipline of measurements in the natural sciences—a major definitional change has been proposed. In 2019, the kilogram will be redefined in terms of a constant of nature instead of a cylinder of platinum-iridium. In this article, we aim to better understand the reasons and procedures for changing definitions by studying the redefinition of the kilogram. In short, the case of the kilogram shows that 1) sometimes the rationale for a redefinition might be clear but the scientific discoveries are not available, and a vast research effort is needed to make progress, 2) progress can be made even in absence of gold standards by reference to the definitions' epistemic aims, and 3) definitions are unlikely to be final as future discoveries might lead to new definitions. These results support the current approach of ongoing, piecemeal revision of psychiatric disease classifications, and stress the importance of robust scientific evidence before changing definitions.

KEYWORDS: Psychiatric nosology; DSM; Metrology; Philosophy of science; Scientific progress

In psychiatry, many scientists desire to move from a classification system based on symptoms toward a system based on biological causes (Cuthbert & Insel, 2013; Kupfer, First, & Regier, 2002; Kupfer, Regier, & Kuhl, 2008). The idea is that psychiatric diseases should be redefined such that each disease would be associated with specific biological causes. This desire is intelligible because causal disease models often facilitate understanding and identification of

Deleted: -

Deleted: -

Deleted: paper

Deleted: s

new ways to intervene in disease processes (Broadbent, 2009; Carter, 2003; Kupfer et al., 2002). In its attempt to move from syndromal to specific etiological definitions, psychiatry follows the trend of general medicine (Carter, 2003).

Current psychiatric disease definitions, however, are based largely on symptoms, signs, duration, and associated disability. Very few specific causes have been identified to date. The lack of specific biological causes has led to criticisms of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM) (Cuthbert & Insel, 2013; Kendell & Jablensky, 2003; Kotov et al., 2017). Given the suspected etiological heterogeneity of current psychiatric disorders, some authors even suggest that psychiatric disorders cannot be defined in terms of symptoms and, at the same time, coincide with biological causes (Kapur, Phillips, & Insel, 2012). In this case, the DSM could hamper or even preclude the discovery of biomedical causes (Cuthbert & Insel, 2013; Kapur et al., 2012; Tabb, 2015). The criticisms have led to new classification initiatives like staging and profiling (McGorry, 2007) and dimensional alternatives based on statistically correlated symptom patterns (Kotov et al., 2017).

Dissatisfaction with fundamental definitions is not unique to psychiatry: in other sciences these are also subject to criticism and change. Radical changes are currently underway for definitions in metrology, the discipline concerned with measurement in the natural sciences. Metrology deals with definitions that are essential to many scientific disciplines and to our daily life, such as the second, meter and kilogram. For over a century, scientists have recognized the importance of using constants of nature to define these units (Quinn, 2012). Nevertheless, many units have long been defined in terms of material and changeable “artifacts,” and some still are. For instance, the kilogram has been defined as a cylinder of platinum-iridium since 1889 (Mills, Mohr, Quinn, Taylor, & Williams, 2005; Quinn, 2012). In 2019, however, the definition of the

Formatted: Font: Italic

kilogram will be altered fundamentally (Cho, 2018). The plan is to redefine it in terms of natural constants (International Committee for Weights and Measures, 2016).

The aim of this article is to shed light on redefining disorders in psychiatry by studying the redefinition of units in metrology, with an emphasis on the redefinition of the kilogram. What are the reasons for metrologists to change the definition of the kilogram, and what are their procedures? To answer these questions, we will present the history of the kilogram, draw parallels between redefinition attempts in metrology and psychiatry, and finish with a discussion including some important differences between these disciplines.

Deleted: paper

THE CHANGING KILOGRAM: FROM MATERIAL ARTIFACT TO CONSTANT OF NATURE

The Need for an International Definition of a Kilogram

Until the nineteenth century, different mass standards were used all over the world. The need for an international standard became clear around 1850: “At the great exhibition of 1851 in London, a multitude of goods, machines and devices were exhibited from all over the world. Depending on their country of origin, specifications were given in imperial, metric or other units, and this was very clear for everyone to see” (Quinn, 2012, p. 6). Increased international trade, manufacturing, and land measurement required definitions of mass, length, and time, that were constant and stable across the world. “Regardless of the field, if we wish to communicate with our colleagues to compare quantitative results on how things behave in real life, the transmission of information on such important data as how big, how much, how heavy, how fast… must be through a common language for measurement. Such a common language requires agreement on reference standards” (Quinn, 2012, pp. 4-5). In other words, a kilogram in Mexico must have the same mass as a kilogram in Australia.

Deleted: 19th

This need motivated the creation of an international committee to prepare for the general adoption of the metric system within all participating countries (Quinn, 2012). In 1869, Napoleon III signed the document to install the Metre Commission. In 1875, this led to the establishment of an international scientific laboratory, the International Bureau of Weights and Measures (Bureau International des Poids et Mesures, or BIPM) in Sevres near Paris. The BIPM's task was and still is to secure the international reference standards for mass, length, electricity, photometry, and later also ionizing radiation, time, and chemistry (Bureau International des Poids et Mesures, 2018; Quinn, 2012).

The Preference for a Definition in Terms of a 'Constant of Nature'

From very early on, even before the introduction of the first international definitions of the kilogram, it was understood that definitions of units preferably should be based on physical constants. "If ... we wish to obtain standards of length, time, mass which shall be absolutely permanent, we must seek them not in the dimensions, or the motion, or the mass of our planet [because these are not necessarily permanent], but in the wavelength, the period of vibration, and the absolute mass of these imperishable and perfectly similar molecules" (Maxwell 1870, in Quinn, 2012, p. xxvii). The reason for this is worldwide stability and availability (Milton, Williams, & Bennett, 2007; Quinn, 2012). However, since its inception, the kilogram has always been defined by a material artifact—a cylinder of 90% platinum and 10% iridium alloy, 39 mm high, 39 mm in diameter—and not in terms of a natural constant (Quinn, 2012).

Deleted: –

Deleted: –

The First Kilogram: A Material Artifact

When the need for an international definition of the kilogram arose, there was no known constant of nature that could be used to define the unit of mass.[1] Therefore, metrologists agreed that the kilogram should be constructed out of material as dense as possible, hard, elastic, and malleable, to ensure its stability. This led to the choice to use platinum-iridium alloy to construct a new International Prototype of the kilogram and numerous national copies (Quinn, 2012). Since 1889, this platinum-iridium cylinder plus more than 80 copies, made from a single casting in London, have served to define the kilogram worldwide. However, comparisons between the International Prototype of the kilogram and its copies showed that something undesirable was going on: the ensemble of official copies and well-maintained national prototypes were drifting in mass. Over the years, working standards started to differ from the international prototype by, on average, thirty-five micrograms (Stock, Barat, Davis, Picard, & Milton, 2015) (Figure 1). Apparently, even carefully kept platinum-iridium cylinders can, over time, shed or gain atoms (Cho, 2017). This instability in mass could lead to serious problems in international science, trade, and industry.

< *Figure 1* >

The Anticipated Change of the Kilogram

A key advance in science, which opened the way to a redefinition of the kilogram into a physical constant, was the discovery of the quantum Hall effect (Quinn, 2012). In particular, Klaus von Klitzing observed that the electrical resistance of material displaying the Hall-effect is very exactly quantized, and expressible entirely as a function of the Planck constant, a fundamental constant of quantum mechanics that relates the energy of a photon to its frequency and the charge of a single electron (Klitzing, Dorda, & Pepper, 1980). This quantum Hall effect could be used in the construction of an equal arm balance to compare electrical and mechanical forces. This so-called Kibble balance, or watt balance, made it possible to weigh the kilogram in terms of the Planck constant, and thus to redefine the kilogram in terms of the Planck constant using a well-described experimental setup (Robinson & Schlamminger, 2016 **<AQ1>**).[2] Since then,

Deleted: i

Deleted: 35

Deleted: B

Formatted: Font: Bold

Formatted: Font: Bold

research groups have been determining an accurate value of the Planck constant, and the expectation is that a new definition can be realized and disseminated with a standard uncertainty of less than twenty micrograms by autumn 2018, so that it will be effective in 2019 (Cho, 2018; International Committee for Weights and Measures, 2016; Milton, Davis, & Fletcher, 2014).

Deleted: 20

Insights From Redefining Concepts in Metrology for Psychiatry

Requisite Discoveries

The story of the kilogram shows that sometimes the rationale for new definitions might be evident but scientific discoveries are needed to make progress. The goal for metrology had been clear for about 150 years—define units in constants of nature—but the way forward was not apparent. The crucial scientific discovery of the quantum Hall effect was made only in 1980. Then, it took scientists almost 40 years to measure the Planck constant with sufficient accuracy (International Committee for Weights and Measures, 2016; Milton et al., 2014). Similarly, since the rise of the etiological research program in medicine in the nineteenth century, successes of monocausal disease models such as tuberculosis, HIV/AIDS, and Down's syndrome have been clear. Disorders defined in specific causes generally improve our understanding of disease mechanisms, and identify new ways to intervene in disease processes (Broadbent, 2009; Carter, 2003; Kupfer et al., 2002).

Deleted: –

Deleted: –

Deleted: 19th

This success explains the often expressed corresponding goal for classifications in psychiatry—namely to define diseases in terms of specific causes (Kapur et al., 2012; Kendell & Jablensky, 2003; Kupfer et al., 2002, 2008). Note that a psychiatric nosology based on specific biological causes is probably an unrealistic goal for psychiatric disorders. At present, scientific evidence points into the direction that a complex developmental mix of biological, psychological and sociocultural risk factors are involved in causal pathways to psychiatric disorders as defined

Deleted: –

Deleted: al-in the DSM, as opposed to single and/or specific biological causes (Borsboom, Cramer, & Kalis, 2018; Kendler, 2014). The same is true for many other disorders, such as stroke, diabetes mellitus (type II) or heart failure (American Diabetes Association, 2012; Feigin et al., 2016; Hunt et al., 2009; van Loo, Romeijn, de Jonge, & Schoevers, 2013). Note, furthermore, that the debate on what constitutes a good psychiatric disease classification is not settled (Tabb, 2015). Disease definitions in terms of specific causes are not the only potentially useful classifications for psychiatry, but also definitions that describe patients with homogeneous course of illness patterns or treatment response could benefit mental health care (Kessler, Van Loo, et al., 2017**<AQ2>**; Maj, 2018; van Loo & Romeijn, 2015).

Formatted: Font: Bold
Deleted: Hanna M.
Formatted: Font: Bold

However, even if we take the search for psychiatric disease definitions based on causes as an example, and thus would strive for definitions of psychiatric disorders with a more homogeneous etiology, defining disorders in a multifactorial or contrastive way (Broadbent, 2009), we currently lack the requisite scientific discoveries. In preparation for the DSM-5, the American Psychiatric Association (APA) intended to change psychiatric classifications from symptom-based to etiology-based (Kupfer et al., 2002, 2008), but early expert reviews of the literature were not optimistic that the scientific progress had been sufficient to support such a move (e.g., Charney et al., 2002). When the DSM-5 work groups began to examine this issue, with the exception of those working on sleep–wake disorders, they concluded that the needed data were not yet available (Zachar, Regier, & Kendler, n.d.). Thus, although the rationale for new definitions might have been clear in both disciplines for a long time, the actual realization of redefining these concepts depends on (fundamental) scientific discoveries, and might even take centuries.

Formatted: Font: Not Italic
Deleted: -
Deleted: S
Deleted: -
Deleted: W

No Gold Standards But Conventions

For some medical conditions, there are 'gold standards' that can be used to test the performance of new definitions or measurement procedures. For instance, the gold standard test for encephalitis (inflammation of the brain parenchyma) is brain biopsy followed by pathological examination of the brain tissue (Venkatesan et al., 2013). This gold standard can be used to test the accuracy of other less invasive ways to assess encephalitis, such as investigation of the cerebrospinal fluid or neuroimaging. However, this is not the case for basic definitions such as the kilogram and major depression. There are no 'gold standards' in metrology or psychiatry: there is not one preapproved criteria to judge which definition facilitates the aims of the science in question best (Chang, 2004). In metrology, the definition of the kilogram was originally based on the mass of one cubic decimeter of pure water. There was no pre-approved definition of the kilogram to assess whether this was the optimal choice, except for the general idea of picking something that stably occurs in nature. Similarly, in psychiatry, there is no gold standard for major depression.

Instead, the choices to define a kilogram in terms of one liter of water, and major depression in terms of five out of nine depressive symptoms are best understood as 'conventions.' Conventions are well-known from the philosophy of space and time (Poincare, 1980; Reichenbach, 1958). A convention is a definition that precedes empirical measurements in a certain domain. Before investigating the mass of the earth, you need a definition of the unit of mass, like the kilogram. Before determining the properties of physical space, you need the definition of a straight line. Similarly, before investigating the course of major depression, you need a definition of major depression (van Loo & Romeijn, 2015). These conventions or

Formatted: Font: Not Italic
Formatted: Font: Not Italic
Deleted: which
Deleted: .
Deleted: Prior to
Deleted: Prior to
Deleted: prior to
Deleted: Hanna M.

definitions are not given beforehand, but once they are laid down, one can start measuring the mass of the earth, the space around heavenly bodies, or the duration of episodes of depression. For a more detailed explanation of conventionalism in psychiatry, we refer to van Loo and Romeijn (2015).

Thus, conventions are necessary conditions to organize empirical facts and serve as bases for measurement and empirical knowledge about psychiatric disorders. They are crucial for science: by associating concepts that belong to a theoretical representation with empirical reality, we make empirical reality and theoretical representations relevant for each other (Poincare, 1980; Reichenbach, 1958; van Loo & Romeijn, 2015, 2018). Because conventions are not facts, they are not true or false. Rather, they serve as bridge principles that coordinate concepts (e.g., the kilogram, or depression) and empirical reality (namely an object of platinum-iridium with a certain mass, or members of the population with a certain symptom profile). However, there is undeniably a choice element to conventions, and in the absence of gold standards it is hard to judge whether conventions are chosen optimally.

Deleted: 8; van Loo & Romeijn

Deleted: 5

Thus, in metrology and psychiatry there are no gold standards, and no independent measurement tools available to test the accuracy of definitions. However, despite the absence of gold standards and threats of definitional circularity, classifications can be improved. This is based on indirect measures, viz. successes in capturing stable empirical regularities in terms of the conventions chosen.

Choices for Definitions Based on Epistemic Aims

Certain choices of conventions are preferable over others, in that they offer a better handle on the practices that are carried out on the basis of these conventions (van Loo & Romeijn, 2015; Reichenbach, 1958). For example, in metrology more or less simultaneously two independent

Deleted: ,

discoveries were made that allowed redefinition of the kilogram using different constants of nature: the mass of silicon atoms or the Planck constant (Quinn, 2012). These discoveries evoked a debate: which constant of nature should we use to redefine the kilogram? The answer was based on the epistemic aims of the definitions—in teaching, trade, science, or elsewhere. Initially, metrologists preferred the mass of silicon atoms for a redefinition of the kilogram, because this seemed intuitive and easy to teach.

However, it appeared that there were good reasons to use the Planck constant instead. First, by fixing the Planck constant and the elementary charge for the ampere, the two quantum effects would become exact, which would benefit electrical metrology. Second, the use of the Planck constant also leads to the possibility of measuring mass in terms of the frequency and speed of light, through the equations $E = mc^2$ (Energy = Mass × Speed of light squared) and $E = hf$ (Energy = Planck constant × Frequency of photon). For these reasons, the Planck constant was chosen to redefine the kilogram (Quinn, 2012).

Also in psychiatry, a variety of research methods exists that might help in discovering more homogeneous patient classes in psychiatry as compared to the current DSM-5 definitions, such as unsupervised statistical learning methods (Kotov et al., 2017), supervised statistical learning methods (van Loo et al., 2014; Wardenaar et al., 2014; Kessler et al., 2016), and staging and profiling (McGorry, 2007), which might all underpin proposals for alternative psychiatric disease definitions. However, some definitions will benefit research and interventions more than others, and thus offer a better grip on psychiatric practice. For instance, definitions that are highly specific with respect to course of illness, familial aggregation, or molecular genetic differences are probably more successful in improving mental health care than definitions that not associated with these validators (Kendler, 2013).[3] We can test how well a convention meshes

Deleted: –

Deleted: m

Deleted: ·

Deleted: s

Deleted: ·

Deleted: f

Formatted: Font: Not Italic

Formatted: Font: Not Italic

Formatted: Font: Not Italic

with genetic data, specific neurobiological abnormalities, or course patterns.[4] There are, in short, “soft ways” to help us redefine our disorders—not gold standards.

Deleted: –

ONGOING PIECEMEAL REVISIONS BASED ON EMPIRICAL RESEARCH

Do New Definitions Outperform Current Ones?

When is a new definition good enough to replace the existing one? Progress for the new definition of the kilogram slowed considerably because it turned out to be much more difficult than had been anticipated to reach a preconceived level of accuracy in measurements of the Planck constant, which was needed to connect it to the kilogram using the Kibble Balance (Quinn, 2012). Inspired by thermometry, some metrologists proposed to proceed with larger measurement uncertainties than initially agreed on (Mills et al., 2005). Others argued that if we fix the Planck constant before measurement uncertainties are sufficiently low, there is a risk that we have to refix the Planck constant in the future. This would mean that the absolute mass of International Prototype would change, and hence every other mass standard (Milton et al., 2014, 2007). Therefore, they proposed to delay the redefinition until experiments reach lower uncertainties, comparable with estimated uncertainty in the absolute mass of the prototype. Moreover, it also had to be shown that the practical implementation of new units was feasible (Milton et al., 2014).[5] In the end, redefinition was postponed.

In contrast, changes in the early editions of the DSM were rarely based on systematic evidence demonstrating that new definitions outperformed older ones. But this has gradually changed in the DSM-III-R, DSM-IV, and in the latter stages of the DSM-5 under the Scientific Review Committee (Kendler, 2013). Currently, a set of validators has been introduced as criteria to judge whether new proposals will improve on current disease definitions (Kendler, 2013). If there is clear evidence that these alternative definitions outperform the current ones in terms of

Deleted: K.S.

Deleted: K.S.

reliability, validity, or clinical utility, this might lead to a change in specific diagnostic categories (First, Kendler, & Leibenluft, 2017).

Continuous Piecemeal Revision

In metrology, definitions of the second, meter and kilogram were redefined at different points in time, dependent on scientific discoveries and sufficient empirical evidence that new definitions were consistent and stable (Quinn, 2012). The current definition of the second, for instance, was introduced in 1967 after discoveries of atomic vibrations and atomic clocks, which enabled the redefinition of the second in terms of vibration of cesium atoms instead of rotation of the earth around its axis (Quinn, 2012). The current definition of the meter, namely the length of path travelled by light in a time interval of 1/299 792 458 of a second, stems from 1983, after developments in laser technology. Instead, the entire DSM has been subject to revision in 1994 and 2013 (APA, 1994, 2013). One of the downsides is that this might lead to an urge to change diagnoses without sufficient evidence to support such changes, or—conversely—not being able to adjust definitions in a timely manner (*Kendler, 2013).

In contrast to this process of entire revisions at certain time points, the APA has adopted an empirically driven continuous improvement model (First et al., 2017; Moran, 2017). This new approach of 'DSM5 as a living document' aims to incorporate changes in DSM when sufficient scientific findings have accumulated, but to prevent unnecessary changes if there is no robust evidence. The APA opened a DSM web portal to receive proposals for introduction, deletion, or revisions of diagnostic categories or subtypes. Researchers from all over the world are invited to propose changes based on sound scientific evidence, that is, a thorough review of the literature and secondary data analyses, much like the scientific review committee of the DSM-5 itself did. If there is clear evidence that these alternative definitions outperform the current ones in terms of

Deleted: a

Deleted: American Psychiatric Association

Deleted: -

Deleted: -

Deleted: i.e.

reliability, validity, or clinical utility, this might lead to a change in specific diagnostic categories (First et al., 2017). This continuous, piecemeal approach is similar to the revision procedures in metrology.

No Final Definitions

Metrology shows that it is unlikely that we will ever have a 'definitive' classification. Even regarding very basic concepts such as the second, scientific findings have led and are still leading to new definitions. In metrology, certain presumed constants of nature turned out to be subject to change after all.[6] For instance, the time for the earth to turn around its axis was presumed as a constant of nature, and used to define the second (Poincare, 1913). But the rotation of the earth around its axis appeared to get a little slower each year. By the 1950s it was known that a day was getting longer by 1.7 millisecond per century. The second was redefined into atomic transitions in 1983 (Quinn, 2012). Currently, new scientific developments to improve atomic clocks promise future redefinition of the second to improve precision (Cartlidge, 2018).

In psychiatry, we can find similar examples of scientific findings that led to new or modified definitions. For instance, Leonard's proposal to separate bipolar from unipolar depression gained prominence as lithium treatment was studied and showed to be rather specific (Baron, Gershon, Rudy, Jonas, & Buchsbaum, 1975). New definitions could also result from scientific findings that were originally not intended to find new categories. For instance, the finding of autoantibodies reacting with neural NMDA-receptors in twelve women (fourteen to forty-four years) who developed prominent psychiatric symptoms (e.g., psychosis, agitation), amnesia, seizures, and autonomic dysfunction, led to the introduction of a new category, that is, anti-NMDA receptor encephalitis (Dalmau et al., 2007). Cases formerly admitted to psychiatric units and diagnosed with acute psychosis, can now be diagnosed with this newly discovered

Deleted: 12

Deleted: 14-44

Deleted: i.e.

immune-mediated disorder, and treated more appropriately (Gibson et al., 2019). Thus, revisions in metrology and psychiatric nosology can be expected to continue as our insights into disorders continue to grow. These processes of definitional change must not be considered as pointing toward a natural end point, but taken more in the spirit of continuous improvement.

Deleted: s

Discussion

General Findings

In this article, we studied the proposed redefinition of the kilogram—from a cylinder of platinum-iridium to the Planck constant—to better understand the reasons and procedures for changing definitions in psychiatry. In short, the case of the kilogram shows that 1) sometimes the rationale to change definitions might be clear but scientific discoveries are needed to make progress, 2) definitions can be improved even in absence of gold standards by reference to their epistemic aims, and 3) definitions are unlikely to be final as future discoveries might lead to new insights. These results support the relatively novel approach in psychiatry of ongoing, piecemeal revision of definitions of psychiatric disorders when there is empirical evidence that the new categories will outperform the old categories, or indeed when there are changes in our stated epistemic aims. If the goals of a disease classification change, for example, because of changes in the practical application of psychiatric science, then this may also motivate a definitional revision.

Deleted: paper

Deleted: -

Deleted: -

Deleted: e.g.

Differences Between Metrology and Psychiatry: Strengths and Limitations of the Comparison

We chose to compare psychiatry with metrology to show that, even in a 'hard' science dealing with basic concepts in nature, classification issues arise and can be settled. The case of the change of the kilogram, and other changes in definitions in metrology raised our interest, precisely because metrology is dealing with fundamental, long-existing, and basic definitions such as the second, the meter, and the kilogram. They are quite different from the definitions of psychiatric disorders, which concern classifications of human beings with mental illnesses in less clear-cut reference classes. We observe that even in a fundamental science such as metrology, dealing with basic measurement units, the definitions are not ideal or definitive, but subject to change based on new scientific discoveries. If definitions in this 'hard' science are not ideal or definitive, we should not expect definitions and classifications in psychiatry to be ideal or definitive. Hence, the absence of ideal or definitive definitions should not be seen as undermining the scientific status of psychiatry.

The fact that metrology is such a different science than psychiatry also limits the comparison. In many respects, there are important differences. First, the challenges in psychiatry are more complex as the 'reference class' is less unequivocal. In metrology, the aim is to approximate a certain historically defined unit (e.g., the cylinder of platinum iridium) as closely as possible to maintain consistency of units over time. The concept of mass is on a relatively firm footing, and the metrologist's task is mostly to find the procedures to access this concept empirically. In psychiatry, the task of redefining disorders is more challenging because the conceptual foundations are less settled. The main problem with the classification is that it does not ensure sufficient homogeneity with respect to important clinical characteristics (Baumeister & Parker, 2012; Kessler, van Loo, et al., 2017**<AQ3>**). Although there are historical precedents in psychiatry for specific diagnostic syndromes (e.g., melancholia/depression) (Kendler, 2017), they are not as cleanly defined as the meter or kilogram. There is often disagreement on disease boundaries: do current definitions capture the right part of the population, or are our definitions too broad or too narrow? (e.g., Zachar, First, & Kendler, 2017)

Formatted: Font: Bold

Formatted: Font: Bold

Deleted: While

Deleted: K.S.

Deleted: *et al.*

Formatted: Font: Not Italic

Second, the situation in psychiatry is also different from metrology, as psychiatry has different, less clear-cut, epistemic aims, and there are more criteria that may lead to valid classifications (Kendler, 2013). The ultimate goal for definitions in metrology is that they are stable, consistent, and reproducible standards. The goals for definitions in psychiatry are that they are stable, consistent, and reproducible classifications but, moreover, that they improve mental health care, guide the clinical treatment of patients and aid in our understanding of etiology. The exact goals of psychiatric disease classification, and the question which criteria will help achieve these goals (the relevant 'validators') is subject to debate (Maj, 2018; Tabb, 2015). Our aim here was not to argue that psychiatric disease classifications should be exclusively based on causes, or that the currently proposed validators for DSM-5 (Kendler, 2013) are optimally chosen. Rather, we want to emphasize the following analogy between metrology and psychiatry—what constitutes a useful definition depends on the classification's epistemic aims, and no gold standards are required to make progress.

Third, different forces are at play in deciding on classifications. Next to scientific reasons, many other forces influence definitions in psychiatry, and they arguably differ in nature from the forces that drive metrology. Examples of these forces are societal norms, and historical contingencies. For instance, premenstrual dysphoric disorder was excluded from past editions of the DSM in part because of strong concerns about the possible social abuses of the diagnosis, leading to negative impact on women in divorce proceedings or running for political office (Zachar & Kendler, 2014). Ego-syntonic homosexuality was included in the DSM as a psychiatric disorder until it was eliminated from the manual in 1973, for a complex mix of reasons including changing cultural attitudes, new empirical data and indeed a new definition of mental illness (Zachar & Kendler, 2012). The addition of post-traumatic stress disorder was

Deleted: K.S.

Deleted: K.S.

Deleted: –

greatly influenced by the experiences of U.S. military veterans of the Vietnam war (Scott, 1990). Although decisions in metrology might also be subject to forces besides epistemic ones (e.g., the needs of industry, national pride) (Quinn, 2012), we suspect that these forces are different and possibly less strong than in psychiatry.

Deleted:)

Fourth, metrology and psychiatry impact on different practices. In both cases, classification changes will have some practical consequences, and thus there is a cost associated with every change. For instance, the new definition of the kilogram needs to be disseminated, taught, and measurement procedures will have to change and possibly related machinery (Milton et al., 2014; Quinn, 2012). Similarly, changes in the DSM need to be implemented in clinical practice, clinicians trained, and findings of scientific studies need to be transposed into the new vocabulary. A major difference, however, is that metrology's redefinitions concern units in the natural sciences, whereas redefinitions in psychiatry concern human beings. To be classified as having a mental disorder might have important consequences for the individual's self-image, treatment and participation in society, and access to healthcare (Hacking, 1995, 2014), which is obviously a major difference with a change of the kilogram.

CONCLUSIONS

The story of the kilogram illustrates that redefinition issues are inherent to science and are certainly not unique for psychiatry. It also suggests that, if we eventually want to define psychiatric disorders more in terms of causes than in terms of signs and symptoms, we are dependent on fundamental scientific discoveries. Moreover, the shape of our nosological revisions will be determined not just by the availability of these crucial scientific findings, but also by the epistemic goals that we set ourselves, i.e., the precise ways in which we aim to improve mental health care. The novel approach of the APA with respect to the revision process of the DSM (the DSM as a 'living document') is in line with the ongoing, piecemeal, evidence-based redefinition procedures in metrology. Finally, because there are fundamental differences in redefinition issues in metrology and psychiatry, future studies of other medical disciplines, biology, or ecology, might offer additional insights relevant for psychiatry.

Deleted: since

NOTES

1. Instead, the predecessor of the international kilogram—the old Archives prototype of the kilogram—was a cylinder with a mass based on measurements of the weight of one cubic decimeter of water and had been constructed by sintering powder at high temperatures. The interstices were full of gas that would have been removed in vacuum, as well as the surface contamination. This would alter the weight of the Archives kilogram, and made it inadequate to use as an international reference (Quinn, 2012).

Deleted: -

Deleted: -

2. The moving-coil watt or Kibble balance was invented by Bryan Kibble in 1975, and relates virtual mechanical and electrical power (Robinson, 2016 **<AQ4>**). The Kibble balance is an equal arm balance, consisting of one arm with a mass (e.g., one kilogram), and a second arm with a coil of wire placed in a strong magnetic field. This enables the Kibble balance to relate macroscopic mass to the Planck constant, which links the amount of energy a photon carries with the frequency of its electromagnetic wave.

Formatted: Font: Bold

Formatted: Font: Bold

3. In this article, we refer to the set of validators that are currently proposed as criteria to judge new proposals for changes in the DSM-5 psychiatric disease classification scheme (Kendler, 2013). We use this set of validators as an example of criteria to assess whether new classifications benefit the epistemic goals of psychiatry. Note that our aim is not to demonstrate that these criteria are optimally chosen.

Deleted: manuscript

Deleted: which

4. For an example of an empirical validation of new data-driven subtypes of depression with respect to the specificity of course of illness, we refer to (van Loo et al., 2014; Wardenaar et al., 2014; Kessler et al., 2016).

5. In 2010, the Consultative Committee for Mass and related Quantities agreed on a set of technical criteria to be met before redefinition of the kg in order to ensure that, after redefinition, one kilogram can be realized and disseminated with a standard uncertainty of twenty micrograms or less (Milton et al., 2014) (List 1).

6. Even in the natural sciences, the idea that you would ever get a theory of everything is flawed (cf. Cartwright, 1983). There is always a messy story when you transition from theory to practice.

REFERENCES**

American Diabetes Association. (2012). Diagnosis and classification of diabetes mellitus. *Diabetes Care*, *35* (Suppl 1), s64–s71.

American Psychiatric Association (APA). (1994). *Diagnostic and statistical manual of mental disorders: DSM-IV*. *Diagnostic and statistical manual of mental disorders* (4th ed.). Arlington, VA: American Psychiatric Publishing.

American Psychiatric Association (APA). (2013). *Diagnostic and statistical manual of mental disorders, fifth edition: DSM-5* (5th ed.). Washington, DC: American Psychiatric Publishing.

Baron, M., Gershon, E. S., Rudy, V., Jonas, W. Z., & Buchsbaum, M. (1975). Lithium carbonate response in depression. *Archives of General Psychiatry*, *32* (9), 1107.

Baumeister, H., & Parker, G. (2012). Meta-review of depressive subtyping models. *Journal of Affective Disorders*, *139* (2), 126–140.

Formatted: Font: Not Italic

Formatted: Font: Not Italic

Formatted: Font: Not Italic

Deleted: (CCM)

Deleted: 1 kg

Deleted: ≤ 20

Deleted: ement

Deleted: *(4th ed.).*

Deleted: US

Deleted: , Inc

Deleted: Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=1994-97698-000&site=ehost-live&scope=site

Deleted: https://doi.org/10.1001/archpsyc.1975.01760270039003

Deleted: https://doi.org/10.1016/j.jad.2011.07.015; 10.1016/j.jad.2011.07.015

Borsboom, D., Cramer, A., & Kalis, A. (2018). Brain disorders? Not really… Why network structures block reductionism in psychopathology research. *Behavioral and Brain Sciences*, 1–54.

Broadbent, A. (2009). Causation and models of disease in epidemiology. *Studies in History and Philosophy of Biological and Biomedical Sciences*, *40* (4), 302–311.

Bureau International des Poids et Mesures. (2018). About the BIPM. Available: www.bipm.org/en/about-us/. Accessed November 10, 2018.

Carter, K. C. (2003). *The rise of causal concepts of disease: Case histories*. Aldershot: Ashgate.

Cartlidge, E. (2018). Better atomic clocks herald new era of timekeeping. *Science*, *359* (6379), 968.

Cartwright, N. (1983). *How the laws of physics lie*. Oxford: Oxford University Press.

Chang, H. (2004). *Inventing temperature: Measurement and scientific progress*. New York: Oxford University Press.

Charney, D. S., Barlow, D. H., Botteron, K., Cohen, J. D., Goldman, D., Gur, R. E., … Zalcman, S. J. (2002). Neuroscience research agenda to guide development of a pathophysiologically based classification system. In D. J. Kupfer, M. B. First, & D. A. Regier (Eds.), *A research agenda for DSM--V*. Washington, DC: American Psychiatric Association.

Cho, A. (2017). Plot to redefine the kilogram nears climax. *Science (New York, N.Y.)*, *356* (6339), 670–671.

Cho, A. (2018). World poised to adopt new metric units. *Science*, *362* (6415), 625–626.

Cuthbert, B. N., & Insel, T. R. (2013). Toward the future of psychiatric diagnosis: The seven pillars of RDoC. *BMC Medicine*, *11*, 126.

Deleted: https://doi.org/10.1017/S0140525X17002266

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

Deleted: *Stud Hist Philos Biol Biomed Sci*

Deleted: https://doi.org/10.1016/j.shpsc.2009.09.006

Deleted: Retrieved November 10, 2018, from https://

Deleted: *c*

Deleted: *(New York, N.Y.)*

Deleted: https://doi.org/10.1126/science.359.6379.968

Deleted: *m*

Deleted: *;*

Deleted: ,

Deleted: D.C.

Deleted: https://doi.org/10.1126/science.356.6339.670

Deleted: *(New York, N.Y.)*

Deleted: https://doi.org/10.1126/science.362.6415.625

Deleted: t

Deleted: https://doi.org/10.1186/1741-7015-11-126 [doi]

Dalmau, J., Tüzün, E., Wu, H., Masjuan, J., Rossi, J. E., Voloschin, A., … Lynch, D. R. (2007). Paraneoplastic anti-N-methyl-D-aspartate receptor encephalitis associated with ovarian teratoma. *Annals of Neurology*, *61* (1), 25–36.

Deleted: https://doi.org/10.1002/ana.21050

Feigin, V. L., Roth, G. A., Naghavi, M., Parmar, P., Krishnamurthi, R., Chugh, S., … Global Burden of Diseases, Injuries and Risk Factors Study 2013 and Stroke Experts Writing Group. (2016). Global burden of stroke and risk factors in 188 countries, during 1990–2013: A systematic analysis for the Global Burden of Disease Study 2013. *Lancet Neurology*, *15* (9), 913–924.

Deleted: a

Deleted: *The*

Deleted: https://doi.org/10.1016/S1474-4422(16)30073-4

First, M. B., Kendler, K. S., & Leibenluft, E. (2017). The future of the DSM implementing a continuous improvement model. *JAMA Psychiatry*, *74* (2), 115.

Deleted: https://doi.org/10.1001/jamapsychiatry.2016.3004

Gibson, L. L., Pollak, T. A., Blackman, G., Thornton, M., Moran, N., & David, A. S. (2019). The psychiatric phenotype of anti-NMDA receptor encephalitis. *Journal of Neuropsychiatry and Clinical Neurosciences*, *31* (1), 70–79.

Deleted: *The*

Deleted: https://doi.org/10.1176/appi.neuropsych.17120343

Hacking, I. (1995). *Rewriting the soul: Multiple personality and the sciences of memory*. Princeton, NJ: Princeton University Press.

Deleted: *m*

Hacking, I. (2014). On the ratio of science to activism in the shaping of autism. In K. S. Kendler & J. Parnas (Eds.) *Philosophical issues in psychiatry III: The nature and sources of historical change* (pp. 326–339). Oxford University Press.

Deleted: ,

Deleted: https://doi.org/10.1093/med/9780198725978.003.0038

Hunt, S. A., Abraham, W. T., Chin, M. H., Feldman, A. M., Francis, G. S., Ganiats, T. G., … Yancy, C. W. (2009). 2009 focused update incorporated into the ACC/AHA 2005 Guidelines for the Diagnosis and Management of Heart Failure in Adults: A report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines: Develope. *Circulation*, *119* (14), e391-479.

Deleted: a

Deleted: d

Deleted: https://doi.org/10.1161/CIRCULATIONAHA.109.192065

International Committee for Weights and Measures. (2016). *CIPM: Report of the 105th meeting*. Available: www.bipm.org/utils/en/pdf/CIPM/CIPM2016-EN.pdf.

Deleted: Retrieved from https

Deleted: //

Kapur, S., Phillips, A. G., & Insel, T. R. (2012). Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? *Molecular Psychiatry*, *17* (12), 1174–1179.

Deleted: https://doi.org/10.1038/mp.2012.105

Kendell, R., & Jablensky, A. (2003). Distinguishing between the validity and utility of psychiatric diagnoses. *American Journal of Psychiatry*, *160* (1), 4–12.

Deleted: https://doi.org/10.1176/appi.ajp.160.1.4

Kendler, K. S. (2013). A history of the DSM-5 scientific review committee. *Psychological Medicine*, *43* (9), 1793–1800.

Deleted: https://doi.org/10.1017/S0033291713001578

Kendler, K. S. (2014). The structure of psychiatric science. *American Journal of Psychiatry*, *171*(9), 931–938.

Deleted: *The*

Deleted: https://doi.org/10.1176/appi.ajp.2014.13111539 [doi]

Kendler, K. S. (2017). The genealogy of major depression: Symptoms and signs of melancholia from 1880 to 1900. *Molecular Psychiatry*, *22* (11), 1539–1553.

Deleted: https://doi.org/10.1038/mp.2017.148

Kessler, R. C., van Loo, H. M., Wardenaar, K. J., Bossarte, R. M., Brenner, L. A., Cai, T., … Zaslavsky, A. M. (2016). Testing a machine-learning algorithm to predict the persistence and severity of major depressive disorder from baseline self-reports. *Molecular Psychiatry*, *21* (10), 1366–1371.

Deleted: https://doi.org/10.1038/mp.2015.198

Kessler, R. C., van Loo, H. M., Wardenaar, K. J., Bossarte, R. M., Brenner, L. A., Ebert, D. D., … Zaslavsky, A. M. (2017). Using patient self-reports to study heterogeneity of treatment effects in major depressive disorder. *Epidemiology and Psychiatric Sciences*, *26* (01), 22–36.

Deleted: https://doi.org/10.1017/S2045796016000020

Kessler, R. C., van Loo, H. M., Wardenaar, K. J., Bossarte, R. M., Brenner, L. A., Ebert, D. D., … Zaslavsky, A. M. (2017). Using patient self-reports to study heterogeneity of treatment

Deleted: V

effects in major depressive disorder. *Epidemiology and Psychiatric Sciences*, *26* (1).

Deleted: https://doi.org/10.1017/S2045796016000020

Klitzing, K. V., Dorda, G., & Pepper, M. (1980). New method for high-accuracy determination of the fine-structure constant based on quantized Hall resistance. *Physical Review Letters*, *45*, 494.

Deleted: https://doi.org/10.1103/PhysRevLett.45.494

Kotov, R., Waszczuk, M. A., Krueger, R. F., Forbes, M. K., Watson, D., Clark, L. A., … Zimmerman, M. (2017). The hierarchical taxonomy of psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *Journal of Abnormal Psychology*, *126* (4), 454–477.

Deleted: https://doi.org/10.1037/abn0000258

Kupfer, D. J., First, M. B., & Regier, D. A. (Eds.). (2002). *A research agenda for DSM-V*. Washington, DC: American Psychiatric Association.

Deleted: (First)

Deleted: D.C.

Kupfer, D. J., Regier, D. A., & Kuhl, E. A. (2008). On the road to DSM-V and ICD-11. *European Archives of Psychiatry and Clinical Neuroscience*, *258*, 2–6.

Deleted: https://doi.org/10.1007/s00406-008-5002-6

Maj, M. (2018). Why the clinical utility of diagnostic categories in psychiatry is intrinsically limited and how we can use new approaches to complement them. *World Psychiatry*, *17*(2), 121–122.

Deleted: https://doi.org/10.1002/wps.20512

McGorry, P. D. (2007). Issues for DSM-V: Clinical staging: A heuristic pathway to valid nosology and safer, more effective treatment in psychiatry. *American Journal of Psychiatry*, *164* (6), 859–860.

Deleted: https://doi.org/10.1176/appi.ajp.164.6.859

Mills, I. M., Mohr, P. J., Quinn, T. J., Taylor, B. N., & Williams, E. R. (2005). Redefinition of the kilogram: A decision whose time has come. *Metrologia*, *42* (2), 71–80.

Deleted: a

Deleted: https://doi.org/10.1088/0026-1394/42/2/001

Milton, M. J. T., Davis, R., & Fletcher, N. (2014). Towards a new SI: A review of progress made since 2011. *Metrologia*, *51* (3), R21–R30.

Deleted: a

Deleted: https://doi.org/10.1088/0026-1394/51/3/R21

Milton, M. J. T., Williams, J. M., & Bennett, S. J. (2007). Modernizing the SI: Towards an

Deleted: t

improved, accessible and enduring system. *Metrologia*, *44* (5), 356–364.

Moran, M. (2017). Portal now open for making DSM a 'living document.' *Psychiatric News*, *52* (1), 1–1.

Poincare, H. (1913). The measure of time. In: *The Foundations of Science* (pp. 222–234). New York: Science Press.

Poincare, H. (1980). *Science and hypothesis*. New York: Dover.

Quinn, T. (2012). *From artefacts to atoms*. New York: Oxford University Press.

Reichenbach, H. (1958). *The philosophy of space & time*. New York: Dover.

Scott, W. J. (1990). PTSD in DSM-III: A case in the politics of diagnosis and disease. *Social Problems*, *37* (3), 294–310.

Stock, M. (2018). The Revision of the SI—Towards an international system of units based on defining constants. *Measurement Techniques*, *60* (12), 1169–1177. **<AQ5>**

Stock, M., Barat, P., Davis, R. S., Picard, A., & Milton, M. J. T. (2015). Calibration Campaign against the international prototype of the kilogram in anticipation of the redefinition of the kilogram part I: Comparison of the international prototype with its official copies. *Metrologia*, *52*, 310–316.

Tabb, K. (2015). Psychiatric progress and the assumption of diagnostic discrimination. *Philosophy of Science*, *82* (5), 1047–1058.

van Loo, H. M., Cai, T., Gruber, M. J., Li, J., de Jonge, P., Petukhova, M., … Kessler, R. C. (2014). Major depressive disorder subtypes to predict long-term course. *Depression and Anxiety*, *31* (9), 765–777.

van Loo, H. M., & Romeijn, J. W. (2015). Psychiatric comorbidity: Fact or artifact? *Theoretical Medicine and Bioethics*, *36* (1), 41–60.

Deleted: https://doi.org/10.1088/0026-1394/44/5/012

Deleted: https://doi.org/10.1176/appi.pn.2017.1a6

Deleted: Publications

Deleted: https://doi.org/10.2307/800744

Formatted: Indent: Left: 0", First line: 0", Widow/Orphan control, Adjust space between Latin and Asian text, Adjust space between Asian text and numbers

Deleted: –

Formatted: Font: Bold

Deleted: https://doi.org/10.1007/s11018-018-1336-2

Formatted: Font: Bold

Formatted: Font: (Default) +Body (Calibri), English (UK)

Deleted: c

Deleted: https://doi.org/10.1088/0026-1394/52/2/310

Deleted: https://doi.org/10.1086/683439 LK - https://rug.on.worldcat.org/oclc/5966430721

Deleted: https://doi.org/10.1002/da.22233 [doi]

Deleted: f

Deleted: https://doi.org/10.1007/s11017-015-9321-0

van Loo, H. M., & Romeijn, J. W. (2018). Letter to the Editor: Measuring and defining: The double role of the DSM criteria for psychiatric disorders. *Psychological Medicine*, *48* (5), 872-873

Deleted: V

Deleted: https://doi.org/10.1017/S0033291717001799

van Loo, H. M., Romeijn, J. W., de Jonge, P., & Schoevers, R. A. (2013). Psychiatric comorbidity and causal disease models. *Preventive Medicine*, *57* (6), 748–752

Deleted: https://doi.org/10.1016/j.ypmed.2012.10.018 [doi]

Venkatesan, A., Tunkel, A. R., Bloch, K. C., Lauring, A. S., Sejvar, J., Bitnun, A., … Yoder, J. (2013). Case definitions, diagnostic algorithms, and priorities in encephalitis: Consensus statement of the International Encephalitis Consortium. *Clinical Infectious Diseases*, *57* (8), 1114–1128

Deleted: https://doi.org/10.1093/cid/cit458

Wardenaar, K. J., van Loo, H. M., Cai, T., Fava, M., Gruber, M. J., Li, J., … Kessler, R. C. (2014). The effects of co-morbidity in defining major depression subtypes associated with long-term course and severity. *Psychological Medicine*, *44* (15), 3289–3302.

Zachar, P., First, M. B., & Kendler, K. S. (2017). The Bereavement Exclusion Debate in the DSM-5: A History. *Clinical Psychological Science*, *5* (5), 890–906

Deleted: https://doi.org/10.1177/2167702617711284 LK - https://rug.on.worldcat.org/oclc/7129371070

Zachar, P., & Kendler, K. S. (2012). The removal of Pluto from the class of planets and homosexuality from the class of psychiatric disorders: A comparison. *Philosophy, Ethics, and Humanities in Medicine*, *7*, 4

Deleted: a

Deleted: https://doi.org/doi:10.1186/1747-5341-7-4

Zachar, P., & Kendler, K. S. (2014). A Diagnostic and Statistical Manual of Mental Disorders History of Premenstrual Dysphoric Disorder. *Journal of Nervous and Mental Disease*, *202* (4), 346–352

Deleted: *The*

Deleted: https://doi.org/10.1097/NMD.0000000000000128

Zachar, P., Regier, D. A., & Kendler, K. S. (n.d.). The aspirations for a paradigm shift in DSM-5: An oral history. *Submitted*. **<AQ6>**

Deleted: a

Formatted: Font: Bold

Formatted: Font: Bold

[1] Instead, the predecessor of the international kilogram - the old Archives prototype of the kilogram - was a cylinder with a mass based on measurements of the weight of one cubic decimeter of water and had been constructed by sintering powder at high temperatures. The interstices were full of gas that would have been removed in vacuum, as well as the surface contamination. This would alter the weight of the Archives kilogram, and made it inadequate to use as an international reference (Quinn, 2012).

[2] The moving-coil watt or Kibble balance was invented by Bryan Kibble in 1975, and relates virtual mechanical and electrical power (Robinson, 2016). The Kibble balance is an equal arm balance, consisting of one arm with a mass (e.g. one kilogram), and a second arm with a coil of wire placed in a strong magnetic field. This enables the Kibble balance to relate macroscopic mass to the Planck constant, which links the amount of energy a photon carries with the frequency of its electromagnetic wave.

[3] In this manuscript, we refer to the set of validators which are currently proposed as criteria to judge new proposals for changes in the DSM-5 psychiatric disease classification scheme (Kendler, 2013). We use this set of validators as an example of criteria to assess whether new classifications benefit the epistemic goals of psychiatry. Note that our aim is not to demonstrate that these criteria are optimally chosen.

[4] For an example of an empirical validation of new data-driven subtypes of depression with respect to the specificity of course of illness, we refer to (van Loo *et al.,* 2014; Wardenaar *et al.,* 2014; Kessler *et al.,* 2016).

[5] In 2010, the Consultative Committee for Mass and related Quantities (CCM) agreed on a set of technical criteria to be met before redefinition of the kg in order to ensure that, after redefinition,

1 kg can be realized and disseminated with a standard uncertainty ≤ 20 microgram (Milton et al., 2014)(List 1).

[6] Even in the natural sciences, the idea that you would ever get a theory of everything is flawed (cf. Cartwright, 1983). There is always a messy story when you transition from theory to practice.