

Robert A. Pollak
January 27, 2010

Standardized Measurement

Standardized measurement is a consequence of successful science, not a cause. That is, standardized measurement is associated with successful science because successful science has implications for measurement. I do not argue against standardization, but against premature standardization: top down imposition of standardized measurement before it is warranted by the science is unlikely to lead to success.

Of course standardized measurement is desirable when the alternative measurements are equivalent. For example, metric units (e.g., meters) English units (e.g., feet) are equivalent. The coexistence of equivalent systems of measurement has no advantages and some disadvantages. The loss of the Mars Climate Orbiter, which was launched in December 1998, illustrates the disadvantages. One NASA team used metric units and another used English units; "mistakes were made" because someone failed to make the necessary conversion. The Phase I Report of the "Mishap Investigation Board" (1999) identifies the problem clearly: "The MCO [Mars Climate Orbiter] MIB [Mishap Investigation Board] has determined that the root cause for the loss of the MCO spacecraft was the failure to use metric units in the coding of a ground software file, 'Small Forces,' used in trajectory models.... [P]rocessing of the data from AMD [Angular Momentum Desaturation] file by the navigation software algorithm therefore, underestimated the effect on the spacecraft trajectory by a factor of 4.45, which is the required conversion factor from force in pounds to Newtons. An erroneous trajectory was computed using this incorrect data" (p. 16).

Metric units and English units are equivalent in the sense that they both convey the same information (e.g., about distance). We can convert metric units into English units and vice versa by multiplying by appropriate conversion factors. Fahrenheit and Celsius (Centigrade) are equivalent systems for expressing temperature: we can convert one into the other and they convey the same information.

My concern is with standardization when the alternative measurements are not equivalent: we cannot convert one into the other and they convey different information. When alternative measures are not equivalent, the choice among them should be guided by science.

A recent National Research Council "summary of a workshop," *Improving the Measurement of Late-Life Disability in Population Surveys: Beyond ADLs and IADLs, Summary of a Workshop*, (NRC, 2009)

appears to take a different view: "The lack of consistency in the concept and definition of disability, and analysis of trends in the prevalence of disability in the population, has been a real barrier to moving forward in the area of measurement of disability. Because of different views about what is meant by the term disability, over the years researchers have used a variety of measures (p. 5)" The report goes on to quote with approval Jette and Badley (2002, pp. 183-184): "Achieving a commonly accepted conceptual language is one of the primary challenges facing the field of disability research..." Three paragraphs later, the report states: "How disability is defined will, explicitly or implicitly, set the goals for how society chooses to deal with disability. Different definitions will suggest different kinds of solutions and indicate different targets for interventions and action. It is important to reach some clarification on what is meant by the term 'disability'."

I agree with the authors of *Improving...* that standardized measurement would affect public perceptions and government policy.

Whether standardized measurement would lead to better policy or to worse policy can only be discussed in terms of a particular standardization of measurement *and* a particular view of what constitutes better policy. Those who believe that government now does too little for the disabled would presumably prefer definitions of disability showing greater "unmet needs," while those who believe that government is already doing too much would presumably prefer definitions showing less unmet needs. Defining disability with an eye on policy implications also provides an opportunity to frame a definition that gives greater weight to disabilities that the framers consider more deserving of policy attention (e.g., cognitive impairment vs mobility impairment.)

The authors of *Improving...* also appear to believe that standardizing measurement would have positive effects on scientific research. I'm dubious.

In economics, theory has implications for measurement, and economists regard "measurement without theory" with skepticism and suspicion. The phrase "measurement without theory" is the title of an influential article by the Nobel Prize winning economist Tjalling Koopmans (1947) in which he reviewed a book on the measurement of business cycles by Arthur F. Burns and Wesley C. Mitchell. Because I know very little about business cycles, I'll talk about price indexes instead.

Price indexes especially the Consumer Price Index (CPI), illustrate the role of economic theory in the construction and interpretation of standardized measures. After discussing the measurement of consumer prices, I turn to the measurement of disability.

A "cost-of-living index" compares two "price situations," where a price situation is a complete list of the prices of all consumer goods and services at a particular time and place. (Mathematically, a price situation is a price vector.) The two price situations to be compared are called "reference prices" and "comparison prices." The cost-of-living index is defined as the ratio of the minimum expenditure required to attain a "base indifference curve" at reference prices to the minimum expenditure required to attain the same indifference curve at comparison prices. To construct a cost-of-living index requires knowledge or an estimate of the consumer's indifference map and the selection of a base indifference curve from that map. Economic theory implies that, faced with a price increase, a consumer given just enough additional income to remain on his or her original indifference curve will reduce consumption of the good whose price has risen and increase consumption of other goods; technically, the relevant demand functions are "compensated" (i.e., "Hicksian") demand functions. Knowledge of the base indifference curve implies knowledge of the strength of the "substitution effect" (i.e., the "marginal rate of substitution" which quantifies the willingness of the consumer to trade off one good against another). The information required to construct a cost-of-living index for an individual consumer is daunting, but the construction and interpretation of the cost-of-living index for an individual consumer has deep roots in economic theory of consumer behavior. To construct a cost-of-living index for a group of consumers, whether a single household or an entire society, requires additional theory; Pollak (1981) filled this lacuna by introducing the "social cost-of-living index." As this paragraph has suggested, the theory of consumer behavior provides a theoretical framework for the construction and interpretation of cost-of-living indexes.

Economic theory also provides a theoretical framework for constructing and interpreting two fixed weight indexes, the Laspeyres and the Paasche. A fixed weight index compares the cost of a fixed collection or "market basket" of goods and services at reference prices with its cost at comparison prices. The fixed weights in the Laspeyres index correspond to the basket of goods and services purchased in the reference price-expenditure situation. It is easy to show that the Laspeyres index is an upper bound on the cost-of-living index, where the base indifference curve is the curve attained in the reference price-expenditure situation. The fixed weights in the Paasche index correspond to the basket of goods and services purchased in the comparison price-expenditure situation. It is also easy to show that the Paasche index is a lower bound on the cost-of-living

index where the base indifference curve is the curve attained in the comparison price-expenditure situation.

The Laspeyres and the Paasche are specific fixed-weight price indexes. Because the weights are the expenditure shares in the reference or in the comparison price-expenditure situation, constructing a Laspeyres or a Paasche index requires conducting a survey of consumer expenditures. The U.S. Bureau of Labor Statistics (BLS) uses the Consumer Expenditure Survey (CEX) to calculate the weights for the Consumer Price Index (CPI). The CEX records how much consumers spend on broad categories of goods and services (e.g., "food," "clothing," etc.), on subcategories (e.g., "fruits and vegetables," "meat," etc.), and on narrowly-defined goods (e.g., apples, bananas, etc.). A Laspeyres index is then constructed by calculating the cost of the reference basket of goods and services at reference prices and at comparison prices; the index is the ratio of these costs. (The story is actually more complicated as the following two paragraphs indicates. One complication is that instead of pricing every item in the price vector, probability sampling is used to select the items that are priced.)

The take home point is that economic theory implies a procedure for constructing two fixed weight indexes, the Laspeyres and Paasche. It identifies the goods and services to be priced and the appropriate weight for each good and service. Economic theory and the results of the CEX determines which goods and services are included in the index and their weights.

The theory of the cost-of-living index not only determines which goods to include in the index and their weights, but also provides guidance for resolving many of the practical problems that arise in index number construction. In a recent National Bureau of Economic Research conference volume, Reinsdorf and Triplett (2009) write: "The economic approach to index numbers is much more than a framework for determining how index number components should be aggregated (the index number formula). It provides a framework that can be used to analyze the domain of the index (the components that are included) and to analyze how index components should be measured." Referring to the alternative "test" approach to price indexes, they add: "The test approach is completely silent on these essential matters (p. 21)."

Among the most difficult CPI issues confronting the BLS are the treatments of new goods, quality change in existing goods, financial services, medical services, consumer durables, and owner-occupied housing. All of these measurement issues are illuminated by the theory of the cost-of-living index. When the ideal treatment is not feasible, the theory often provides guidance in choosing among the feasible alternatives.

Without an accepted theoretical framework, we would lack a touchstone for resolving any of the practical problems that arise in index number construction. Without a theory, we can evaluate alternatives in terms of their implications for the behavior of the index or their implications for the cost of constructing the index. For example, each alternative resolution has implications that will make the index either higher or lower than it would otherwise be; without a theoretical framework, we can prefer one resolution over another on the basis of its consequences (i.e., whether it would make the index higher or lower). For example, if the CPI is higher, then social security benefits, which linked to the CPI, will also be higher. If you think social security recipients deserve more, you presumably favor a resolution that implies a higher index, but if you think social security recipients already receive more than they deserve, you presumably favor a resolution that implies a lower index. Proposed resolutions also are likely to imply different costs of index number construction (e.g., because one resolution is less expensive for BLS to implement than the other); we can favor one resolution over another on the basis of its cost. Constructing an index without a theoretical framework is, to appropriate the famous Robert Frost quip about free verse, "like playing tennis without a net." Measurement without theory is measurement without principles, and without principles there can be no principled resolution of measurement problems.

Recent discussions of trends in age-specific disability among the elderly in the United States provide an instructive example and an opportunity to evaluate calls for standardized measurement.

As the subtitle of *Improving . . . , (Beyond ADLs and IADLs)* suggests, the usual approach to measuring disability relies on ADLs (basic "activities of daily living") and IADLs ("instrumental activities of daily living"). (The ADLs are: transferring, dressing, bathing, toileting, eating, and walking across a room. The IADLs are: grocery shopping, preparing meals, taking medications, using a telephone, and managing household finances.) I focus on ADLs, treating them as an index of disability.

Viewing ADLs as an index of disability, an economist would begin by asking how the activities/items included in the index were chosen and how their weights were chosen. (ADLs give equal weight to each of the 6 included activities/items.) An economist would ask whether the activities/items on the list are "more important" than those not on the list. An economist would then ask about the weights, that is, why are all of the activities/items on the list assigned equal weight?. (The "why" question calls for a substantive answer, not an historical one. In the price index context, for example, in response to a question about the

Laspeyres index, it would not suffice to say that this is the index Étienne Laspeyres proposed in 1871.)

Suppose a critic were to claim that the list of ADLs gives too little weight to cognitive impairments relative to other impairments. (Of course cognitive impairment may affect some of the ADLs on the list, but our critic understands this and claims that cognitive impairment "deserves" greater weight. Suppose another critic were to claim that some ADLs (e.g., toileting) "deserve" greater weight than the others. And suppose a third critic wants to know why erectile dysfunction is not an ADL. Without a conceptual framework for measuring disability, these critics cannot be answered.

I do not argue that the preference/ utility-based framework of the cost-of-living index is the appropriate framework for measuring disability, although I do believe it is an approach worth exploring. An alternative is to define disability as an index that best predicts nursing home entry within some specified period, say one year. Another alternative is to define disability as an index that best predicts home care costs or nursing home costs within some specified period.

During the 1990s different surveys purporting to measure disability reported different trends among older Americans. A "technical working group" was convened to report on the sources of these differences. Although discussing the particulars of each survey analyzed by the working group would take us far afield, a paragraph from their report suggests some of the issues:

"One of the most widely used scales to measure personal-care disability in old age, the Activities of Daily Living scale...was originally designed to assess the rehabilitative potential of older institutionalized residents. Since that time, surveys of older American have adopted various approaches to asking respondents whether they have *difficulty* with an activity (generally but not always *without* help or equipment); others have asked about the *need* for help or whether the respondents are unable to complete the activity without help; and still others have identified persons with disabilities through their reports about the *use of accommodations*, that is, whether they reported getting help, being supervised, or using special equipment to carry out the basic activities. These differences in wording are not only important conceptually, but may correspond loosely with the extent of fit between an individual and his or her home environment and therefore result in different prevalence levels...Moreover, getting help reflects not only the level of severity but also the availability of help, whether from family members, friends, or paid caregivers" (p. 419; italics in original).

Without belaboring the details, the quoted paragraph suggest that some of the differences in the reported trends in disability reflect differences in the questions asked. (They also reflect differences in the populations studied, the recall period, the handling of item nonresponse, etc., but I focus on the questions themselves.) If, in the 1970s, experts at the Office of Management and Budget (OMB) had decided to standardize on one of these measures and require use of that measure in all subsequent surveys, the problem of inconsistent trends would not have arisen.

We would observe only the trend in the standardized measure that OMB had chosen, and we would not have known that, had OMB chosen a different standardized measure, we would have observed a different trend. "Ignorance is bliss." This poses no problem if OMB had chosen the "right" measure (i.e., the one I prefer), but what is at stake is a substantive decision about what counts as disability, not a harmless measurement convention like choosing between metric and English units.

To reiterate: my concern is with non-trivial standardization, with choices between measurement protocols that convey different information. Such choices should depend on the purpose of measurement. I have argued that the theory of consumer behavior and the theory of the cost-of-living index provide a framework for the CPI. As Reinsdorf and Triplett (2009) argue, such a foundation not only legitimates the choice of activities/items and their weights in the CPI, but also provides guidance in resolving other problems that arise in constructing and interpreting the CPI.

Analogous issues arise in educational testing. For example, we might construct a test for 7th grade students that predicts their success in school (e.g., high school graduation). Alternatively, we might construct a test that predicts their labor force participation at age 30, or one that predicts their wage rate at age 30.

Except in cases in which the alternative measurements are equivalent (e.g., choosing between metric and English units), choosing a standardized measure is tantamount to deciding what is important and what is not. Those making decisions about which standardized measure to promulgate are likely to appeal to theory, but the appeal is likely to be implicit rather than explicit. In practice, what looks at first glance like measurement without theory often turns out to be measurement with implicit theory. Science is better done in the open.

References

Diewert, W. Erwin, John S. Greenlees, and Charles R. Hulten, *Price Index Concepts and Measurement*, Chicago: University of Chicago Press, 2009.

Freedman, Vicki A. et al., "Resolving Inconsistencies in Trends in Old-Age Disability: Report from a Technical Working Group," *Demography*, Vol. 41, No. 3, (August 2004), 417-441.

Koopmans, Tjalling C., "Measurement without Theory," *Review of Economics and Statistics*, Vol. 29, No. 3, (August 1947), 161-172.

Mars Climate Orbiter, Mishap Investigation Board, Phrase I Report, November 10, 1999.

Pollak, Robert A., "The Social Cost of Living Index," *Journal of Public Economics*, Vol. 15, No. 3, (June 1981), 311-336.

Pollak, Robert A., *The Theory of the Cost of Living Index*, New York: Oxford University Press, 1989.

Pollak, Robert A., "The Consumer Price Index: A Research Agenda and Three Proposals," *Journal of Economic Perspectives*, Vol. 12, No. 1, (Winter 1998), 69-78.

Reinsdorf, Marshall and Jack E. Triplett, "A Review of Reviews: Ninety Years of Professional Thinking About the Consumer Price Index," in Diewert, Greenlees and Hulten (2009), 17-83.

Schoeni, Robert F., Vicki A. Freedman, and Linda G. Martin, "Why is Late-Life Disability Declining?" *Milbank Quarterly*, Vol. 86, No. 1, (March 2008), 47-89.

Stigler, George J., "Imperfections in the Capital Market," *Journal of Political Economy*, Vol. 75, No. 3, (June 1967), 287-292.