International Journal of Nursing Studies 51 (2014) 488–494

Contents lists available at SciVerse ScienceDirect

International Journal of Nursing Studies

journal homepage: www.elsevier.com/ijns

On measurements and their quality: Paper 3: *Post hoc* pooling and errors of discreteness

Jason W. Beckstead *

University of South Florida College of Nursing, USA

ARTICLE INFO

Article history:
Received 5 June 2013
Accepted 5 June 2013

Keywords:
Measurement theory
Errors of discreteness
Post hoc transformations
Information theory

ABSTRACT

This is the third in a short series of papers on measurement theory and practice with particular relevance to research in nursing, midwifery, and healthcare. In this paper I demonstrate how the decisions we make regarding the *post hoc* treatment of our measurements impact the quality of our data and influence the validity of the inferences we draw from them. I address two variations of this practice, pooling data over response options found on self-report measures, and transforming measurements of continuous variables, such as age, into ranges or ordered categories. The problems inherent in this practice are examined using concepts from information theory. Pooling more precise measurements into less precise ones creates errors of discreteness that introduce unpredictable (positive or negative) bias in our results.

1. Introduction

> When we start pooling categories...we are doing something to the randomness of the sample, with unknown consequences for our inferences. The manner in which the categories are pooled can have an important effect on the inferences drawn. This practice is to be avoided if at all possible (Hays, 1981, p.552).

As good scientists we endeavor to be as precise and thorough as possible in our work. While many of our measurements are made with a certain (high) degree of precision, this precision is reduced when we group values into ranges or combine different classes of responses. Such loss of precision equates to a loss of information. Information loss limits the accuracy of our conclusions.

Health care researchers measure many continuous variables; examples include blood pressure, respiratory function, lesion size, body mass index, and age. Often these measurements are transformed into ordered categories *post hoc*, (*i.e.*, after they have been recorded) by grouping or pooling adjacent values into ranges (bins) prior to analysis and decision making (Chen et al., 2007; Naggara et al., 2011). Those who study psychosocial constructs (*e.g.*, depression, anxiety, self-efficacy, coping ability, social support, *etc.*) invest considerable time and effort to develop instruments capable of measuring individual differences. When using such instruments researchers are careful to assess and report the psychometric properties of their data implying that they believe individual differences in such measurements to be meaningful. Despite this attention to detail, many papers are published in which such precise measurements are pooled into coarser numeric values (*e.g.*, high and low, or high, medium, and low, or quartiles, *etc.*) for analysis (see MacCallum et al., 2002, for review).

Such errors of discreteness, as Cohen (1983) called them, have been shown to produce loss of information, loss of efficiency, lower statistical power, lower reliability, biased effect size estimates, and inflated Type I and Type II errors (Austin and Brunner, 2004; Beckstead and Beckie, 2011; Chen et al., 2007; Caille et al., 2012; Irwin and McClelland, 2003; MacCallum et al., 2002; Maxwell and Delaney, 1993). Although much of this methodological work has focused on the most extreme form of pooling,

* Correspondence address: University of South Florida College of Nursing, 12901 Bruce B. Downs Boulevard, MDC 22, Tampa, FL 33612, USA. Tel.: +1 813 974 7667; fax: +1 813 974 5418.
E-mail address: jbeckste@health.usf.edu.

0020-7489/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.ijnurstu.2013.06.003

dichotomizing, these problems apply where ever more precise measurements are pooled into less precise ones. In the context of clinical decision making, categorizing continuous variables has been criticized in that it does not make use of within-category information; everyone in a category is treated as equal, yet their prognosis may vary considerably (Naggara et al., 2011). Repeated warnings against the practice are to be found in various fields including consumer research (Fitzsimons, 2008), education (Kuss, 2013), psychology (Cohen, 1983), medicine (Dawson and Weiss, 2012; Royston et al., 2006), and nursing (Beckstead, 2012), yet the practice persists.

In this, the third paper in this series on measurement quality (Norman and Griffiths, 2013), I address two variations of this practice, pooling data over response options found on self-report measures, and transforming measurements of continuous variables into ranges or ordered categories. My objective is to demonstrate how the decisions we make regarding the *post hoc* treatment of our measurements impact the quality of our data and influence the validity of our conclusions. The issue is not so much when such decisions are made, but that we are deciding to alter our measurements *after we have obtained them*. The problems inherent in pooling of responses from different categories or binning adjacent values will be examined using concepts from information theory.

2. Information theory

In the 1940s Claude Shannon developed the means to quantify the amount of information in a given set of data (Shannon and Weaver, 1949). Central to the theory are the quantities information and uncertainty. When we complete a history and physical examination of a patient, read a book, or simply have a question answered, we (presumably) reduce our uncertainty and acquire information about the world. In a technical sense, the amount of information we get *via* any of these acts has no relevance to whether the information is correct, incorrect, useful, or useless. Shannon's theory deals only with quantifying the amount of information, not with its meaning or importance. Shannon proposed various mathematical functions relating information and uncertainty to probability. When uncertainty is reduced it becomes information; when information is discarded it is replaced by uncertainty.

Information and uncertainty are quantified in binary digits or "bits". A bit is the amount of information necessary to reduce the number of possible elements in a given set by half. Each particular element, i, in a set of mutually exclusive elements has a probability, p_i, and its information value in bits is defined as $\log_2(1/p_i)$. When we consider the entire set of elements (a discrete probability distribution), the average information or uncertainty, U, is computed by determining the information associated with each element separately and then obtaining a weighted average. The weights are the respective probabilities; thus, $U = \sum[p_i \times \log_2(1/p_i)]$.

As measurements are taken with greater and greater precision, the amount of information they contain increases; when measurements are taken with a certain precision and this precision is then reduced by collapsing values into bins or ranges information is lost. Beckstead and Beckie (2011) used information theory to show that dichotomizing multiple clinical measurements into a binary indicator (metabolic syndrome: present, absent) led to discarding 98% of the information contained in a set of continuous measurements taken on a sample of patients and how such information loss can have serious consequences for statistical power and the validity of inferences drawn in medical research. Information theory has also been used to explicate the computational shortcomings of the so called content validity index (Beckstead, 2009).

3. Pooling data over response options: self-reported health status

For a number of years epidemiologists and clinical researchers have a studied people's self-reported health status using single questions such as: "How do you rate your current state of health?" (Kaplan et al., 1996), "How would you rate your state of health in general?" (Eller et al., 2008), "How do you rate your general state of health?" (Alexopoulos and Geitona, 2009), or "How do you rate your state of health in general?" (Kartal and Inci, 2011). The longstanding use of this approach may be due to its ease of use and because early on Wannamethee and Shaper (1991) described this type of measure as a good overall indicator of health status, comprising the perception of symptoms, diagnoses and health behaviors.

Individuals typically answer these questions by selecting a response from response-option sets including: extremely good, good, average, bad, extremely bad, (Kaplan et al., 1996), very good, good, moderate, poor, very poor, (Alexopoulos and Geitona, 2009), very good, good, satisfactory, less than good, poor, (Eller et al., 2008), or, very good, good, fair, bad, (Kartal and Inci, 2011). The effects of such subtle variations in question wording and in the verbal anchors chosen for response options will be addressed in a subsequent paper in this series. Here I address a problem common to all these studies and to many others, namely, the practice of pooling data by collapsing over different response options.

To illustrate the consequences of this practice, consider a graduate student who is working on analysis of some survey data collected by his advisor. The student decides to examine the relationship between two variables, self-reported health status and whether or not the respondent is a smoker. For the sake of illustration, let us say he has available data from 100 respondents. The data are shown in Panel A at the top of Fig. 1. When these data are analyzed using a χ^2 test, the result indicates that there is no significant relationship between the two variables ($\chi^2 = 5.915$, $df = 3$, $p = 116$). Disappointed by this result, the student decides to pool data from adjacent response options and rerun the test. In his first attempt, he collapses the two middle options (*fair* and *good*) and the χ^2 test approaches significance (see Fig. 1 Panel B). Encouraged by this result, he decides to regroup the data again, this time collapsing over the options, *poor*, *fair*, and *good* and compares them against responses in the category *excellent* (see Fig. 1 Panel D). He finds this result appealing in two