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

Self-reported Health Status

A.

Smoker	*Poor*	*Fair*	*Good*	*Excellent*		χ^2	*df*	*p*	*bits*
yes	8	7	7	11					
no	10	7	11	39		5.915	3	.116	2.660
	18	14	18	50	100				

B.

Smoker	*Poor*	*Fair & Good*	*Excellent*		χ^2	*df*	*p*	*bits*
yes	8	14	11					
no	10	18	39		5.475	2	.065	2.346
	18	32	50	100				12%

C.

Smoker	*Poor*	*Fair, Good & Excellent*		χ^2	*df*	*p*	*bits*
yes	8	25					
no	10	57		1.300	1	.254	1.586
	18	82	100				40%

D.

Smoker	*Poor, Fair & Good*	*Excellent*		χ^2	*df*	*p*	*bits*
yes	22	11					
no	28	39		5.473	1	.019	1.875
	50	50	100				30%

E.

Smoker	*Poor & Fair*	*Good & Excellent*		χ^2	*df*	*p*	*bits*
yes	15	18					
no	17	50		4.097	1	.043	1.790
	32	68	100				33%

Fig. 1. Hypothetical data illustrating ambiguity and information loss introduced by pooling categories. Panel A shows data as recorded. Comparing bit-values of Panels B through E against A quantifies the amount (percentage) of information discarded when collapsing response options for analysis.

ways; first, the two resulting groups of respondents are of the same size, and second, the χ^2 test is now significant.

Somewhat enthusiastically, the student shares his findings with his advisor; she asks him to explain his rationale for altering the original measurements. Aside from noting that the relationship became significant, he is at a loss. The advisor points out that if data are pooled by collapsing the response options, *fair*, *good* and *excellent* and comparing them with those in the category *poor*, the χ^2 test is nowhere near significant, and that pooling data at the midpoint of the four options, *poor* and *fair* versus *good* and *excellent*, produces another significant χ^2 value (see Fig. 1, Panels C and E, respectively).

With all five analyses before them, the two researchers discuss the fact that depending on the manner in which the measurements are "rearranged" the χ^2 test is significant or it is not. The inference to be drawn about the relationship between smoking and self-reported health status is dependent on how they pool their data. How are they to describe the smoking rates among healthy and unhealthy respondents? This too depends on how the responses have been pooled.

The amount of information in the data may be calculated from the joint probabilities of the contingency tables shown in Fig. 1. The original data set (Panel A) contains 2.660 bits of information. Collapsing categories alters these joint probabilities and alters the information value of the data. Fig. 1 shows the percentage of information lost or discarded by each of the four rearrangements of the original data (Panels B through E). Unfortunately there is not a straightforward relationship between the amount of information discarded and the strength of the inferred association between the two variables; the association can increase or decrease depending on how the sample data are redistributed. The paradox in this example is now explicit; pooling data across response options discarded information which altered the relationships among the measurements and introduced uncertainty into the inference to be drawn about the association between self-reported health status and smoking behavior.

Such paradox are not limited to hypothetical settings. Kartal and Inci (2011) examined the relationship between self-reported health status and metabolic control among patients with type 2 diabetes reporting "There was found to be a statistically significant difference between self-perceived health and levels of HbA1C; 68% of participants who had high HbA1C values evaluated their health as good, whereas 63.9% of those who had poor HbA1C values evaluated their health as bad" (p. 227). In their analysis the authors combined 10 "bad" responses with 41 "fair" responses to form a category they labeled "poor" and combined 17 "very good" responses with 42 "good" responses to form a category they labeled "good" (see their Table 3). In their Table 4 the percentages of patients with good, borderline, and poor control of their HbA1C are presented as 24.5%, 44.5%, and 30.9%, respectively. In their Table 5 the authors crosstabulated their data into the two health-status categories and the three metabolic-control categories and present a significant χ^2 test. The cell frequencies and χ^2 statistic are shown here in Fig. 2, Panel A. Using the marginal distribution from their Table 3, I reconstructed the joint distribution of health status and metabolic control for comparison (see Fig. 2 Panel B).

A. As Presented

Self-reported Health Status

Control	Bad & Fair	Good & Very Good		χ^2	df	p	bits
Good	9	17					
Borderline	21	29		7.120	2	.020	2.494
Bad	22	13					24%
	52	59	111				

B. Reconstructed

Control	Bad	Fair	Good	Very Good		χ^2	df	p	bits
Good	2	7	12	5					
Borderline	4	17	21	8		5.779	3	.123	3.302
Bad	5	17	9	4					
	11	41	42	17	111				

Fig. 2. Real data illustrating ambiguity and information loss that resulted from pooling categories. Frequencies of HbA1C control and self-reported health status. Cell frequencies and χ^2 in Panel A are from Kartal and Inci (2011) Table 5. Reconstructed cell frequencies (Panel B) are based on the marginal distribution of health status responses provided in their Table 3.

Pooling health-status responses discarded 24% of the information contained in the original (reproduced) data set. Had the χ^2 test been conducted on these data their conclusion (presumably) would have been different.

In the final section of their paper the authors state "One important conclusion is that type 2 diabetes patients show a deterioration in their perceptions of their own health; this can be seen that approximately half of the participants reported 'poor' levels of self-perceived health" (Kartal and Inci, 2011, p. 233). Actually, no patients reported their health status as "poor" because "poor" was not among the response options from which they could choose. Rather, 9% indicated their health as "bad", 37% indicated their health as "fair", 38% indicated their health as "good", and 16% indicated their health as "very good", and it is most likely that these four groups of diabetic patients did not differ in terms of how well their HbA1C was controlled.

4. Grouping adjacent values into ranges or ordered categories: how old are you really?

Pooling more precise measurements into less precise ones discards information which distorts relationships. As mentioned above, healthcare researchers measure various continuous variables. In epidemiology and clinical research patient age is often viewed as a potential confounder of primary risk factors and so investigators include it in multivariate models (Chen et al., 2007; Reijneveld, 2003; Weinberg, 1995). Because of its ubiquity in research, and because its values are both intuitive and convenient to work with, I have selected age as the variable for this example.

Consider four individuals ages 20, 27, 28, and 35 years. The consecutive intervals between their ages are 7, 1, and 7, and it is obvious that the second and third individuals are closest in their ages. Things change however when we group ages into categories. For instance if age is categorized into 10-year ranges (18–27, 28–37, 38–47, 48–57, 58 or more) then the first and second individuals will have the same "score" of 1 because they fall into the first bin, and the third and fourth individuals will have scores of 2. This transformation has assigned different values to the two most similar consecutive measurements (27 becomes 1, but 28 becomes 2) and assigned the same values to the most different consecutive measurements (20 and 27 both become 1, while 28 and 35 both become 2). Alternatively if age is categorized into 5-year bins (18–22, 23–27, 28–32, 33–37, 38–42, 43–47, 48–52, 53–57, 58–62, 63 or more), the transformed scores of our four individuals become of 1, 2, 3, and 4, respectively. This transformation has imposed equal spacing onto the three consecutive intervals (7, 1, and 7 become 1, 1, and 1). Whether 5- or 10-year bins were used, this form of transformation added errors of discreteness to the original data. In other words, binning increased the uncertainty in the transformed measurements relative to the original. Increasing uncertainty equates to losing information.

The four ages in this example were deliberately chosen to highlight two distorting effects that result from binning. In practice, the distribution of age values in a given sample determine the net influence of these two distorting effects. Larger samples could be examined one value at a time to assess these effects, however this approach becomes tedious. An alternative is to examine the amount of information in the original sample and how much of this information is lost when age values are binned into consecutive ranges.

According to information theory, the average information in a continuous variable, on the assumption of a normal distribution, can be estimated as $U \approx \log_2\{\sqrt{[2\pi e(\sigma/\delta)^2]}\}$, where σ is the standard deviation of the variable and δ is a quantization of measurement precision (see Beckstead & Beckie, 2011, for details). Such estimates are relative; the amount of information increases as measurement precision increases. For example if age is a normally distributed variable with a standard deviation of 5 years and we think of it in 1-year increments ($\delta = 1$), then its information value is estimated as 4.369 bits. If instead we consider age in 5-year increments ($\delta = 5$) then its information value is estimated as 2.047 bits, and in 10-year increments ($\delta = 10$) as only 1.047 bits. Comparing these estimates we see that 53% of the information in measurements of age is expected to be lost [(4.369 − 2.047)/4.369 = 53%] when 5-year bins are used, and 76% is expected to be lost when 10-year bins are used. These estimates are theoretical and based on a known value of σ. In practice there would be sampling variability in the degree of information lost; the impact of binning values

Table 1
Potential information and information loss in simulated age measurements for various sample sizes and category widths used to pool values.

Sample Size	Potential information			Percent of information lost	
	Measured in years	5-year Ranges	10-year Ranges	5-year Ranges	10-year Ranges
30	3.721 (0.183)	1.965 (0.182)	1.183 (0.190)	47.2% (3.8%)	68.3% (4.5%)
50	3.954 (0.146)	2.016 (0.145)	1.200 (0.150)	49.1% (2.6%)	69.7% (3.2%)
100	4.149 (0.101)	2.057 (0.101)	1.220 (0.104)	50.4% (1.6%)	70.6% (2.1%)

Note. Potential information values are in bits. Entries indicate means (and standard deviations) of values over 4000 random samples. Samples were drawn from a normal distribution of integers with mean of 42 and standard deviation of 5.

would vary from sample to sample. To examine this issue in more detail I ran a small-scale Monte Carlo simulation.

This simulation was conducted to address two questions. The first question deals with how much information we find in age measurements for samples of various sizes, and how much of this information is lost through *post hoc* pooling of values into ranges. The second question deals with how information loss effects correlations. Binning values into ranges is expected to attenuate correlations (Pearson, 1913). But, as we saw in Fig. 1, information loss can either strengthen or weaken the observed association between two variables depending on how the sample data are redistributed. This raises an interesting question, could the information lost due to binning cause a sample correlation to increase, simply because of sampling error? This point is relevant because researchers sometimes justify binning because it yields a higher correlation (MacCallum et al., 2002).

For illustration let say we are interested in knowing the value of the correlation between the ages of cohabiting couples. This value might be important for studies of social support or caregiver-recipient relationships. Let X be the age of the first partner and Y be the age of the second partner in each couple. Samples of integer values were generated from bivariate normal distributions with means of 42 and standard deviations of 5 years to simulate the distribution of ages among adults (18–64 years old). The design of this simulation used three sample sizes ($N = 30$, 50, and 100) and four values of the population correlation between X and Y ($\rho_{XY} = .10$, .30, .50, and .70). These two design variables were arranged in a factorial (3×4) design yielding 12 conditions. For each condition 1000 random samples were drawn to produce a total of 12,000 samples for analysis. After each sample was drawn, both variables (X and Y) were transformed into 5-year bins (18–22, 23–27, 28–32, 33–37, 38–42, 43–47, 48–52, 53–57, 58–62, 63 or more) and also into 10-year bins (18–27, 28–37, 38–47, 48–57, 58 or more). The resulting variables are X_5, Y_5, X_{10} and Y_{10}, respectively, where the subscripts indicate the widths of the bins.

The information contained in the X measurements of each sample was calculated from the relative frequency distribution as $U_X = \sum[p_i \times \log_2(1/p_i)]$. This formula was then applied to the relative frequency distributions of X_5 and X_{10} for comparison. The same calculations were performed on Y, Y_5, and Y_{10} but because both X and Y were sampled from populations with the same means and standard deviations, the calculated values for Y were virtually identical to those of X in each sample. Therefore only the analysis of X is presented. The correlations between X and Y, between X_5 and Y_5, and between X_{10} and Y_{10} were also calculated for each sample. The results of the simulation are summarized in Table 1 and Table 2.

Table 1 shows that the amount of information contained in the samples increased with sample size. While this is to be expected, this simulation quantifies by how much the information increased when sample size was increased. Age measurements in samples of size 30 contained on average 3.721 bits of information, or about 85.2% of the 4.369 bits estimated to exist in the population distribution of age values. For samples of size of 50 this information increased to 3.954 bits (90.5%), and age measurements in samples of size 100 contained on average 4.149 bits (95.0%). Examining the standard deviations shows that the amount of information contained in the measurements varied more from sample to sample for the smaller sample sizes. In other words, the amount of information we can expect to obtain from our measurements is more unpredictable for smaller samples due to sampling error.

Pooling adjacent age values into ranges reduced the amount of information in the samples and the larger the bin size the more severe was the information loss (see Table 1). Approximately 50% of the information contained in each sample of age measurements was lost when values were binned into 5-year ranges and about 70% of the information was lost when values were binned into 10-year ranges. Although the percentage of information lost was somewhat less for the smaller samples, these tended to contain less information to begin with.

Table 2 summarizes the correlations calculated on each sample. As expected the variability in sample correlations around their expected values decreased as sample size increased (see standard deviations). For all conditions simulated, the average size of the sample correlation was reduced by binning the original measurements. These attenuations were uniformly more severe with larger bin sizes. This effect was slightly more pronounced for correlations in the $\rho_{XY} = .50$ condition.

Table 2
The effects of categorizing simulated age measurements into 5- and 10-year bins on sample correlations for various population correlations (ρ_{XY}) and sample sizes.

Population ρ_{XY}	Sample Size	Sample correlations			Frequencies of $r_{XY\text{bin}} > r_{XY}$	
		r_{XY}	$r_{\text{5-year bins}}$	$r_{\text{10-year bins}}$	5-year bins	10-year bins
.10	30	.087 (.190)	.083 (.190)	.072 (.192)	476	448
	50	.099 (.143)	.090 (.141)	.072 (.141)	425	409
	100	.103 (.101)	.096 (.102)	.076 (.102)	424	356
.30	30	.287 (.174)	.267 (.175)	.220 (.185)	393	294
	50	.300 (.130)	.279 (.128)	.230 (.133)	340	238
	100	.300 (.086)	.276 (.088)	.221 (.089)	263	132
.50	30	.482 (.146)	.444 (.157)	.362 (.166)	268	162
	50	.492 (.112)	.458 (.116)	.372 (.127)	237	100
	100	.495 (.077)	.456 (.080)	.367 (.088)	122	18
.70	30	.688 (.101)	.635 (.114)	.520 (.151)	170	56
	50	.690 (.077)	.640 (.087)	.522 (.115)	108	18
	100	.696 (.053)	.645 (.059)	.528 (.078)	32	4

Note. Entries indicate means (and standard deviations) of r values over 1000 random samples. Frequencies indicate the number of samples out of 1000 in which $r_{X\text{bin}Y\text{bin}} > r_{XY}$. Samples were drawn from bivariate-normal distributions of integers with means of 42 and standard deviations of 5.

Table 2 also shows the number of samples in each condition where binning actually increased the sample correlation coefficient. When sample size and ρ_{XY} were relatively small, it was not uncommon to find that binning produced an increase in the sample correlation simply due to sampling error. In other words, even though binning values of a normally distributed variable must cause the population correlation to decrease, the practice can result in an increase in the sample correlation depending on the manner in which original values are redistributed into bins.

The likelihood of obtaining an increased sample correlation coefficient was greater when age was binned into 5-year ranges than into 10-year ranges (3258 and 2235 of 12,000 samples, respectively). This may be explained by considering the number of bins produced by each transformation. Pearson (1913) showed that the probability of obtaining a larger correlation increases as a function of the number of bins. Transforming age into 10-year ranges typically produced only three bins ($M = 3.283$, SD = 0.458) per sample, but when 5-year ranges were used the number of resulting bins was nearly doubled ($M = 5.790$, SD = 0.732). When transformed into 10-year ranges (*i.e.*, into three bins) the numbers of samples in which sampling error led to an increase in the sample correlation coefficient were similar to those reported by MacCallum et al. (2002) where values were dichotomized into two bins. Both the transformations examined here discarded information and produced biased correlations, however the degree of information loss occurring with 5-year ranges seems to have a more unpredictable biasing effect.

5. Summary and recommendations

Because measurements are so central to our research, it is important that we make informed choices when deciding among measurement options (Beckstead, 2013a). Pooling more precise measurements to form less precise ones distorts relationships among measurements and between variables. In this paper I have addressed two variations of this practice, pooling data over response options found on self-report measures, and binning measurements of continuous variables into ranges or ordered categories. The problems inherent in this practice were examined using concepts from information theory. When information is discarded it is replaced by uncertainty; errors of discreteness introduced through such transformations increase the uncertainty in our data. Unlike random measurement errors that can only attenuate observed relationships among variables (Beckstead, 2013b), the bias introduced by errors of discreteness is more unpredictable; relationships may appear strengthened or weakened. The examples in this paper demonstrate how the decisions we make regarding which response options to pool or which bin widths to impose over adjacent values can affect the quality of our data and compromise the integrity of the inferences we draw from them. The best choice we can make regarding *post hoc* pooling is to avoid this practice altogether.

References

Alexopoulos, E.C., Geitona, M., 2009. Self-rated health: inequalities and potential determinants. International Journal of Environmental Research in Public Health 6, 2456–2469.

Austin, P.C., Brunner, L., 2004. Inflation of type I error rate when a continuous confounding variable is categorized in logistic regression analysis. Statistics in Medicine 23, 1159–1178.

Beckstead, J.W., 2009. Content validity is naught. International Journal of Nursing Studies 46, 1274–1283.

Beckstead, J.W., 2012. To be, or to be two, the question of dichotomizing variables. International Journal of Nursing Studies 49, 635–636.

Beckstead, J.W., 2013a. On measurements and their quality. Paper 1. Reliability – history, issues and procedures. International Journal of Nursing Studies 50 (7) 968–973.

Beckstead, J.W., 2013b. On measurements and their quality. Paper 2. Random measurement error and the power of statistical tests. International Journal of Nursing Studies 50 (10) , http://dx.doi.org/10.1016/j.ijnurstu.2013.04.010.

Beckstead, J.W., Beckie, T., 2011. How much information can the metabolic syndrome provide? An application of information theory. Medical Decision Making 31, 79–92.

Caille, A., Leyrat, C., Giraudeau, B., 2012. Dichotomizing a continuous outcome in cluster randomized trials: Impact on power. Statistics in Medicine 31, 2822–2832.

Chen, H., Cohen, P., Chen, S., 2007. Biased odds ratios from dichotomization of age. Statistics in Medicine 26, 3487–3497.

Cohen, J., 1983. The cost of dichotomization. Applied Psychological Measurement 7, 247–253.

Dawson, N.V., Weiss, R., 2012. Dichotomizing continuous variables in statistical analysis: a practice to avoid. Medical Decision Making 32, 225–226.

Eller, M., Holle, R., Landgraf, R., Mielck, A., 2008. Social network effect on self-rated health in type 2 diabetic patients: results from a longitudinal population-based study. International Journal of Public Health 53, 188–194.

Fitzsimons, G.J., 2008. Death to dichotomization. Journal of Consumer Research 35, 5–8.

Hays, W.L., 1981. Statistics, 3rd ed. Holt Rinehart and Winston, Inc., New York.

Irwin, J.R., McClelland, G.H., 2003. Negative consequences of dichotomizing continuous predictor variables. Journal of Marketing Research 40, 366–371.

Kaplan, G.A., Goldberg, D.E., Everson, S.A., Cohen, R.D., Salonen, R., Tuomilehto, J., Salonen, J., 1996. Perceived health status and morbidity and mortality: evidence from the Kuopio Ischaemic Heart Disease Risk Factor Study. International Journal of Epidemiology 25, 259–265.

Kartal, A., Inci, F.H., 2011. A cross-sectional survey of self-perceived health status and metabolic control values in patients with type 2 diabetes. International Journal of Nursing Studies 48, 227–234.

Kuss, O., 2013. The danger of dichotomizing continuous variables: a visualization. Teaching Statistics 35, 78–79.

MacCallum, R.C., Zhang, S., Preacher, K.J., Rucker, D.D., 2002. On the practice of dichotomization of quantitative variables. Psychological Methods 7, 19–40.

Maxwell, S.E., Delaney, H.D., 1993. Bivariate median splits and spurious statistical significance. Psychological Bulletin 113 .

Naggara, O., Raymond, J., Guilbert, F., Roy, D., Weill, A., Altman, D.G., 2011. Analysis by categorizing or dichotomizing continuous variables is inadvisable: an example form the natural history of unruptured aneurysms. American Journal of Neuroradiology 32, 437–440.

Norman, I., Griffiths, P., 2013. Measurement decisions in nursing and midwifery research. International Journal of Nursing Studies 50 (7) .

Pearson, K., 1913. On the measurement of the influence of "broad categories" on correlation. Biometrika 9, 116–139.

Reijneveld, S.A., 2003. Age in epidemiological analysis. Journal of Epidemiology and Community Health 57, 397.

Royston, P., Altman, D.G., Sauerbrei, W., 2006. Dichotomizing continuous predictors in multiple regression: a bad idea. Statistics in Medicine 25 .

Shannon, C.E., Weaver, W., 1949. The Mathematical Theory of Communication. University of Illinois Press, Urbana.

Wannamethee, G., Shaper, A., 1991. Self-assessment of health status and mortality in middle-aged British men. International Journal of Epidemiology 20, 239–245.

Weinberg, C.R., 1995. How bad is categorization? Epidemiology 6, 345–347.