THE UNIVERSITY OF TENNESSEE KNOXVILLE

Teaching and Supervision in Counseling

Volume 3 | Issue 3 Article 4

September 2021

Matching Variables with the Appropriate Statistical Tests in Counseling Research

Ryan E. Flinn
New Mexico State University- Main Campus

Michael T. Kalkbrenner
New Mexico State University - Main Campus

TEACHING AND SUPERVISION IN COUNSELING
The official journal of the Southern Association for Counselor Education and Supervision

Matching Variables With the Appropriate Statistical Tests in Counseling Research

Ryan E. Flinn, Michael T. Kalkbrenner

Quantitative research literacy, including matching variables with the appropriate statistical tests, is a key element in counselor education and preparation. Counselor educators are tasked with teaching quantitative research and statistics, which counselors-in-training tend to find anxiety-producing. Authors aimed to provide a succinct overview of matching variables with appropriate statistical tests and provide strategies counselor educators can use to enhance their pedagogy.

Keywords: counselor preparation, quantitative research, research question, variables, statistics

Counselor educators are responsible for teaching classes in research methods and statistics to train their students to critically analyze empirical literature, utilize findings to inform evidence-based counseling practice, and possibly produce research that can extend the extant literature to enhance counseling practice and promote clients' well-being (Council for Accreditation of Counseling and Related Educational Programs [CACREP], 2015; International Registry of Counsellor Education Programs [IRCEP], 2015). A written research proposal in which counselors-in-training (CITs) select a topic, write a literature review, compose research questions(s), and propose a methodology is a customary assignment in counselor education programs for meeting CACREP standards that are associated with research methods and statistical analyses (e.g., CACREP, 2015, 2.F.8.f. & h) as well as the IRCEP standards for research and assessment (IRCEP, Standard V, Domain B). Counselor educators also supervise dissertations and theses for CITs who utilize quantitative designs and statistical analyses. Supporting CITs' pursuit of research literacy in quantitative methodologies, however, poses a number of challenges for counselor educators. One consistent research finding on counseling students' perceptions of statistics and quantitative methodologies is that CITs often present with statistics anxiety, low research self-efficacy, and perceived deficits in quantitative research knowledge and skills (Field, 2018; Steele & Rawls, 2015). Research training offered to CITs may also be inadequate and ineffective (Balkin, 2020; Jorgensen & Umstead, 2020), further heightening these negative perceptions of quantitative research and lack of involvement in this work among CITs at the master's-level in particular (Steele & Rawls, 2015).

While the extant literature includes textbooks and a succession of refereed journal articles that collectively present the steps for matching variables with statistical analyses (a core aspect of quantitative research literacy), these resources can be costly and overwhelming for CITs who are already anxious about quantitative research and statistics. Moreover, finding and using scholarly resources that explain statistical concepts in a clear and concise fashion is a difficult pursuit (Holmes et al., 2018; Lalayants, 2012). According to Field (2018), for example, many statistics texts "teach different tests in isolation and never really give [students] a grasp of the similarities between them," creating a sense of "unnecessary mystery" (p. xvii). While training standards from CACREP specify that master's- and doctoral-level CITs should be exposed to quantitative methodologies, no specific competency level has been operationalized stating the precise

Ryan E. Flinn, Michael T. Kalkbrenner, New Mexico State University – Main Campus. Correspondence concerning this article should be addressed to Ryan E. Flinn, flinn@nmsu.edu

skills counselors and counselor educators should possess for designing quantitative research studies (Wester & Borders, 2014). This ambiguity may contribute to the presence of errors in research produced by counselors and counselor educators. For instance, Wester et al. (2013) conducted a review of quantitative research articles published during the years 2009 and 2010 in the *Journal of Counseling & Development* and identified a number of errors related to quantitative competencies, including authors' failure to ground their studies within a theoretical framework, clearly state research questions or hypotheses, or select statistical analyses that would appropriately answer their research question.

An empirical guide for matching variables with the appropriate statistical analyses that integrates CACREP and IRCEP standards on research has potential to facilitate counselor educators' collective ability to provide instruction, supervision, and advising in support of CITs' development of quantitative research literacy. Therefore, we provide a succinct empirical overview (one-stop-shop) for writing quantitative research questions and matching variables of interest with the appropriate statistical tests. Throughout this overview, we refer readers to numerous exemplar articles to illustrate how RQs can be appropriately presented and answered using statistical tests commonly used in counseling research. This overview responds to the call to further develop a signature pedagogy in master's-level research training for CITs (Jorgensen & Umstead, 2020). Therefore, the intended audience of this article is CITs who are enrolled in graduate-level introductory research methods and statistics courses, as well as counselor educators who are looking for resources on teaching quantitative research. This article might also have utility as a primer for doctoral students and counselor educators who are conducting quantitative research.

Generate Research Question to Address a Gap in the Literature

A research question (RQ) is defined as the articulation of the specific goals of a proposed study (Creswell & Creswell, 2018). The manner in which an RQ is phrased directly impacts which methodology (e.g., quantitative, qualitative, or mixed-methods) and data analytic techniques (e.g., ANOVA, regression) should be employed. Therefore, forming an RQ is perhaps the single most important step in a research study. Given the importance of a well-considered RQ, we offer several strategies for CITs to employ as they engage in the iterative process of generating possible RQs to address a gap in the extant literature and formalizing their RQ to achieve concordance with an appropriate philosophy of science and research hypothesis.

Generating possible RQs can often be both challenging and rewarding, as this step in the research process requires CITs to draw upon numerous sources of information to make multiple decisions upfront. We recommend CITs start with identifying their overarching area of interest. In this article, we provide multiple example RQs that all relate to the overarching topic of the mental health needs of college students of color. In this example, college students of color are the population the researchers are interested in studying and college students' mental health needs are the construct of interest. CITs may identify their overarching area of interest based on many factors, including personal or professional group memberships or sociocultural experiences; community, national, or interventional events or needs; clinical interests and experiences; coursework; advisor/faculty expertise; or university partnerships and resources. Ethical and practical considerations may also constrain or create possible RQs, including CITs' access to specific (including vulnerable) populations, certain types of instruments or tools, a sufficiently large sample needed to answer different types of RQs, and funding to recruit a sample. Ultimately, all academic research is meant to improve the field's understanding of specific populations and constructs of interest, therefore, any RQ should add to the extant literature by making a novel contribution to knowledge. This requires CITs to first know what has been published in their overarching area of interest, synthesize this knowledge, consider ethical and practical concerns, and use their creativity to generate an RQ that can be practically carried out and will present novel results to a specific audience.

Extensive knowledge of the literature pertaining to a specific population and/or construct takes time and perseverance to cultivate as well as strong organizational and time-management skills. As CITs read, they should begin to notice patterns, trends, and gaps, and highlight important theories, frameworks, and questions being discussed, as well as any commonalities noted in areas highlighted by authors for future research. At this point, it becomes necessary to narrow the focus of one's RQ in terms of population, constructs, or both. CITs can draw upon journalism as an example of the key features that should be encapsulated in their finalized RQ: *who, what, where,* and *when*. Referring to the example in the previous paragraph, one must decide if their RQ question will broadly pertain to college students of color (i.e., all non-White-identified college students) or only those belonging to a specific racial or ethnic group (e.g., Black or African-American-identified students). In terms of mental health needs, will the RQ pertain to levels of depression, anxiety, both, or some other mental health concern among this population? Will students from anywhere in the world be eligible to participate or only those from a specific country, region, state, city, or academic institution or institution type? Will data be gathered from these students at only one time point or at multiple points across time, and at what point in their development (e.g., age, semester in school, specific life experience) will they be invited to contribute data? CITs must also be able to answer *why* this research should take place, typically articulated in a problem and/or purpose statement (Creswell & Creswell, 2018) that offers an empirical or theory-based rationale for how answering this particular RQ will benefit the population under investigation. For example, will this research test a theory, replicate an initial empirical result, or otherwise enhance the profession's understanding of a population's experiences and needs and/or the existence or influence of a construct on those we serve? This finalized RQ must also be grounded within an appropriate philosophy of science (see the next section), which dictates one's research methodology. For example, a finalized RQ might be: *Are there statistically significant differences in depression severity by gender identity among students of color who are enrolled in a Predominately White Institution (PWI)?*

Philosophy of Science and Research Questions

Philosophies of science represent collected assumptions about the nature of reality and the appropriate ways of generating knowledge (Creswell & Creswell, 2018). While dozens of philosophies of science exist, three of the most common that inform research in the social sciences include postpositivism, critical realism, and social constructivism (Bhaskar, 1978; Creswell & Creswell, 2018). Postpositivism is a deterministic philosophy of science in which researchers seek to investigate and understand an objective reality (i.e., seek a universal truth). A postpositivist researcher, for example, might investigate which of three 20-second suicide awareness video clips are rated by 15 college students of color as the most effective for reducing the stigma associated with seeking mental health services. Consistent with a postpositivistic worldview, this researcher is seeking to uncover a universal truth by using the responses from a sample of a population (college students of color) to generate knowledge applicable (generalizable) to all members of this population who might see these clips. Contrary to postpositivism, social constructivism is centered on the notion that an infinite number of realities exist, as each independent observer constructs their own unique reality. In this same example, a social constructivist might argue that in reality 15 different sets of video clips are being watched, as each college student is experiencing the clips in a unique way based on the philosophy that each person experiences (socially constructs) the stimuli differently. Thus, a social constructivist would likely be most interested in understanding the processes by which each person experienced the stimuli (the 20-second commercials) and the meanings they generated from this experience, which would encourage a researcher grounded in this philosophy of science to use more open-ended, qualitative methods of gathering data (e.g., interviews, diary entries) rather than forced-choice, quantitative responses to survey items or scales. Finally, critical realism reflects the tenants of both postpositivism and social constructivism (Ayers, 2011). The objective or material reality consists of intransitive objects, which exist independently from the observer who engages in making sense of what is being observed (i.e., established cultural norms, values, and laws), however, each independent observer experiences the material world subjectively. Given that generalizability (the extent to which the results of analyses using data from samples of a population can be used to make conclusions about that larger population) is a core tenant of quantitative research designs (Creswell & Creswell, 2018), quantitative research questions tend to reflect postpositivist or critical realist philosophies of science. The previous example RQ is based on a critical realist philosophy of science as both objective elements (disproportionate representation of students of color at a PWI) and subjective elements of reality (one's experience of symptoms) are implied. This RQ corresponds to a critical realist philosophy of science in which each student of color experiences elements of a shared/objective reality (e.g., studying at a PWI), however, each of their particular experiences of depressive symptoms are independently constructed.

Research Hypothesis. A research hypothesis (RH) is a prediction about the results of a quantitative study. This RH (also known as the alternative hypothesis) makes a prediction about the expected results based on logic and findings in the extant literature. Building upon the sample RQ identified in the previous section, an RH corresponding to this RQ might be: *Among Black students attending a PWI, those who identify as gender nonbinary will report significantly more severe depression severity when compared to those who identify as male or female.* When written well, a RQ and RH (which are typically reported at the end of the literature review, just before the methods section) contain information that allows the reader to identify the population being studied and the variables of interest in the study. Furthermore, the manner in which the RQ and RH are written will often allow the reader to anticipate the scale at which these variables will be measured and the most appropriate statistical analysis to utilize.

Identifying Variables and Scales of Measurement

The first step in selecting the most appropriate data analytic technique or statistical analysis is identifying the variables (reflected in the RQ) and their scales of measurement. A variable refers to anything that can be measured or quantified (Field, 2018). Observed variables are rather simple to quantify and are typically appraised in a single survey question (e.g., asking research participants to specify the number of counseling sessions they have attended). In contrast, latent variables can be more challenging to quantify because they cannot be directly observed ("intelligence" or "depression"). Therefore, these inferred variables (aka *theoretical constructs*) are often measured by obtaining scores from participants (observed variables) on survey items. In quantitative psychometric research of high quality, a series of observed variables (e.g., a collection of survey items) collectively comprise a latent variable(s) that measure an underlying theoretical construct in the population of interest. In this way, observed scores participants provide in response to items that make up scales serve as proxies for latent constructs most often of interest to counselors and counselor educators (e.g., attitudes, beliefs, and personality states/traits). The Patient Health Questionnaire-9 (PHQ-9), for example, is a screening tool with rigorously validated scores for appraising depression severity (Kroenke et al., 2001). The PHQ-9 is comprised of nine items (observed variables) that collectively measure the test taker's overall level of depression severity (latent variable).

Four common types of quantitative variables in counseling research include independent variables (IVs), dependent variables (DVs), predictor variables, and criterion/outcome variables. IVs are comprised of levels, which are "independent" in the sense that they do not change during the study and are manipulated to determine whether they influence scores on a DV. For example, a CIT might be interested in investigating if clients' depression severity (the DV) depends on time in therapy (a within-subjects IV) that is comprised of two or more levels (e.g., [pre and post] or [pre, middle, and post]). This example lends itself to a group comparison approach, such as a repeated measures ANOVA (see later section). Predictor variables are comparable to IVs when using a regression analysis (see later section), in which one is investigating if scores on a variable can predict one's future, concurrent, or past scores on a criterion or outcome variable. For example, one might investigate the extent to which *number of counseling sessions attended during the fall semester* (predictor variable) is a significant predictor of *depression severity* among college students of color during the subsequent year (criterion variable).

Scales of Measurement

Variables are typically classified as categorical or continuous, which are further broken down into nominal, ordinal, interval, or ratio scales of measurement (Field, 2018). Identifying the scale at which variables are measured is essential given that this information, considered in tandem with the overall research design to be employed in the study (e.g., descriptive, correlational, experimental; Trusty, 2011), directly determines the selection of the appropriate statistical test (see Figure 1).

Categorical-Level Scales of Measurement. Categorical variables can be measured at the nominal or ordinal level. Nominal scales are the most basic type of categorical variable in which data are measured in discrete categories. Geographic location, for example, is measured on a nominal scale when asking research participants to specify if they live in (a) rural, (b) urban, or (c) suburban area. Ordinal scales are categorical variables with an inherent rank-order between categories. Imagine, for example, that a group of clients were asked to endorse the following statement: *attending counseling was helpful* on a Likert scale ranging from *strongly agree*, *agree*, *neutral*, *disagree*, or *strongly disagree*. Although technically Likert-type data are considered to be ordinal data (categorical-level scale of measurement), in practice, responses to Likert-type questions are sometimes inappropriately analyzed as if they were continuous-level scale data, however, deeper measurement issues are at play. Referring to the previous Likert scale, a participant who selects *strongly agree* is reporting that they found counseling more helpful than someone who selects *agree*, however, one cannot determine precisely how much more since this is an ordinal, rank-ordered scale.

Continuous-Level Scales of Measurement. Continuous-level scales of measurement include interval and ratio variables with equal distances between scale points. Interval-level variables are comprised of identical distances between measurements without an absolute zero point (i.e., without the possibility of a complete absence of the construct of measurement). The time of day, for example, measured on an interval scale as the difference between 2 p.m. and 3 p.m. is exactly the same as 7 p.m. and 8 p.m., however, the time of day is never zero (i.e., on Earth it is never 00:00 o'clock or the complete absence of time). Ratio scales of measurement include equal distances between scale points with the possibility of a true zero point. The number of counseling sessions that a person has attended, for example, is a ratio scale of measurement, as attending 20 sessions is exactly twice as many as 10 and a true zero point is possible (i.e., it is possible for someone to have never attended counseling). Subjective ratio scale scores can also be generated using a client's personal ratings and goal attainment scaling (see Ruble et al., 2012); for example, the response to *How many days did you experience the urge to use alcohol last week?* could be none.

Matching Variables With Statistical Tests

The next task is to select the most appropriate data analytic procedure (aka *statistical analysis*) to test a RH and answer the overall RQ(s). In this section, foundational information about common data analytic procedures in counseling research are presented, along with references to actual examples of these approaches utilized in recent empirical literature. Crucially, all of the parametric statistical analyses described in the current article (with the exception of the Chi-square test of independence) are based on parametric assumptions about the parameters of the sample data, therefore CITs should complete assumption checking prior to proceeding with these statistical analyses (Trusty, 2011; see Figure 2). In addition to the assumptions in Figure 2, random sampling is an assumption of most inferential statistical analyses and generalizability should be listed as a limitation when researchers use non-random sampling procedures. The statistical assumptions presented in Figure 2 are based on the recommendations from leading statisticians (Field, 2018; Tabachnick & Fidell, 2014; Warne, 2014), however, not all statisticians are in complete agreement about the necessary statistical assumptions for each analysis. Thus, counseling researchers are tasked with citing the empirical source(s) they referenced for assumption checking when conducting quantitative research. Researchers should also compute an a priori power analysis (see Balkin & Sheperis, 2011; Faul et al., 2009) to calculate the minimum sample size required for answering their research question before beginning data collection.

Figure 1

Criteria for Choosing Common Statistical Tests With Normally Distributed Data

Nature of Research Question	**Number of Independent Variables**	**Number of Dependent Variables**	**Number of Control Variables**	**Scale of Measurement**	**Statistical Test**
Group comparison	1	1	0	IV = Categorical DV = Categorical	Chi-square
Group comparison	1	1	0	IV = Categorical DV = Continuous	*t*-test or one-way ANOVA
Group comparison	1 or more	1	0	IVs = Categorical DV = Continuous	2-way or Factorial ANOVA
Group comparison	1	2 or more	0	IVs = Categorical DVs = Continuous	MANOVA
Group comparison	2	2 or more	0	IVs = Categorical DVs = Continuous	2-way MANOVA
Group comparison	3 or more	2 or more	0	IVs = Categorical DVs = Continuous	Factorial MANOVA
Group comparison	1 or more	1	1	IV = Categorical DV = Continuous CV = Continuous	ANCOVA
Group comparison	1 or more	2 or more	1 or more	IVs = Categorical DVs = Continuous CVs = Continuous	MANCOVA (1-way, 2-way, or factorial depending on # of IVs)
Related variables	1	1	0	IV = Continuous DV = Continuous	Pearson product moment correlation
Related variables	2 or more	1	0	IVs = Continuous (at least 1) DV = Continuous	Multiple regression (including HMR)
Related variables	1 or more	1	0	IV = Continuous (at least 1) DV = Categorical	Logistic regression
Related variables	1 or more	2 or more	0, 1, or more	IVs = Continuous (at least 1) DVs = Continuous	Multivariate regression or Structural Equation Modeling
Psychometric analyses	Emergent/ hypothesized	Emergent/ hypothesized	N/A	Latent	EFA & CFA

**Note*: IV = independent variable, DV = dependent variable, CV = control variable.

Figure 2

Assumptions of Common Statistical Tests (Univariate and Multivariate)

Assumption	**Description**	**Strategies for Assumption Testing**	**Chi-square**	***t*-test or one-way ANOVA**	**2-way or Factorial ANOVA**	**MANOVA**	**2-way MANOVA**	**Factorial MANOVA**	**ANCOVA**	**MANCOVA**	**Correlation**	**Regression (HR &HMR)**	**Logistic Regression**
Absence of outliers	The data are lacking extreme scores (i.e., *z*-score > 3 or 3.29).	Histograms, boxplots, *z*-scores		X	X	X	X	X	X	X	X	X	X
Homogeneity of covariance	The multivariate counterpart to homogeneity of variance (see below).	Box's M test				X	X	X		X			
Homogeneity of regression slopes	The association between the outcome variable and control variable does not deviate statistically across the levels of the independent variable.	Construct general linear model to test for an interaction effect between the IV and control variable. $p > .05$ indicates that this assumption is met.							X	X		X	
Homogeneity of variance (AKA homoscedasticity)	The variance in values of the independent variable are equal ($p > .05$) across different values of the dependent variable.	Levene's Test; graphing *zpred* vs. *zresid*		X	X				X			X	X
Independence of observations	Data from each participant is only counted once (single observation).	Data collection procedure	X	X	X	X	X	X	X	X	X	X	X
Linearity	Presence of a fixed relationship between the predictor variables is (i.e., graphically represented as a straight line).	Graphing *zpred vs. zresid;* scatterplots									X	X	X
Absence of Multicollinearity	Variables (independent or dependent) are measuring separate traits (i.e., not too highly correlated with one another to be considered independent constructs).	Pearson correlation ($r \geq .9$) indicates the absence of multicollinearity and/or the largest VIF < 10 (or the tolerance is below .1)				X	X	X	X	X		X	X
Normality (univariate and/or multivariate)	The parametric properties of sample data are symmetrical and consistent with the population data, in which the majority of scores cluster around the mean with increasingly fewer outliers (extreme scores) at the tails of the distribution.	Frequency distributions; P-P & Q-Q plots; kurtosis and skewness values; Kolmogorov-Smirnov; Shapiro-Wilk, Mahalanobis distances (multivariate normality)		X	X	X	X	X	X	X	X	X	X

**Note*: The statistical assumptions presented in the table above are based on the recommendations from leading statisticians (see citations in the Matching Variables with Statistical Analyses section), however, the descriptions are not exhaustive and not all statisticians are in absolute agreement about the necessary statistical assumptions for each analysis. Thus, counseling researchers are tasked with citing the empirical source(s) they referred to for assumption checking when conducting quantitative research.

Group Comparisons

Group comparison analyses encompass a family of statistical tests centered on investigating mean differences between or within groups. Some of the most common group comparison analyses in counseling research include: (a) Chi-square test of independence, (b) *t*-test, (c) univariate analysis of variance, (d) multivariate analysis of variance, and (e) analysis of covariance.

Chi-Square Test of Independence. A Chi-square test of independence (aka *the Pearson Chi-square test*) has utility for comparing two categorical-level variables. Prior to computing a Chi-square test of independence, researchers should complete the statistical assumption procedures depicted in Figure 2 and ensure that data are measured as counts or frequencies in discrete categories (McHugh, 2013). Building on the previous example RQ, a Chi-square test of independence would be appropriate if a CIT wished to test for *significant differences in binary versus nonbinary gender identity (1 = binary identity [male, female] or 2 = nonbinary identity [gender nonbinary, agender]) and White versus non-White racial identity (1 = White or 2 = non-White racial identity) among college students*. Turner et al. (2017) provide an example of a Chi-square test by examining race (1 = White students or 2 = students of color) and help-seeking history (past use of mental health services), 1 = none or 2 = any; the test returned a significant result showing a higher representation of White students reporting previous use compared to students of color (see p. 303 for details). An inherent limitation associated with the use of this test in counseling research is the possibility of obscuring differences among individuals placed in each category by analyzing count data measured in discrete categories.

T-Test. A *t*-test (see Figure 2 for assumptions) is a group comparison analysis with utility for comparing two mean scores (i.e., one categorical-level IV with only two levels and one continuous-level DV). A *t*-test could answer the following RQ: *Are there significant differences in depression severity by gender identity (1 = binary or 2 = nonbinary)*? The categorical-level IV (gender identity) includes only two levels (1 = binary and 2 = nonbinary) and the DV (depression severity measured by the PHQ-9) is appraised on an interval-level scale. In an example from the extant literature concerning college students of color, Mushonga and Henneberger (2020) used independent-samples *t*-tests to examine positive mental health (e.g., DVs = self-esteem, spirituality, racial identity, social support) among traditional (group 1: ages 18–24) and nontraditional (group 2: students aged ≥ 25) Black college students (see p. 152 for details). The same limitation discussed previously for Chi-square applies to this test.

Analysis of Variance: Univariate. Analysis of variance (ANOVA) is a group comparison analysis (see Figure 2 for assumptions) for investigating mean differences between two or more IVs (with two or more levels) across a single continuous-level DV. Depending on the number of IVs, ANOVA is either one-way (one IV), two-way (2 IVs), or factorial (three or more IVs). Essentially, a *t*-test is just the most basic case of an ANOVA (one IV with just two levels). In an example of a one-way ANOVA, Turner et al. (2017) investigated differences in fears about therapy (DV) between students who had attended therapy in the past and students who had not (one IV, with two levels; level 1 = have attended therapy in the past; level 2 = never attended). In support of their hypothesis, their results showed that those who had attended therapy in the past reported less therapy fears, on average, than those who had not ever attended therapy (see p. 303).

ANOVA offers advantages over Chi-square and *t*-test in that researchers can impute more variables with multiple levels into their models of between- and within-group differences. This allows for the testing of statistical models that likely better reflect the complex, nuanced nature of social reality affecting the daily lives of our counseling clients. For instance, the following RQ uses a two-way ANOVA to build upon the previous RQ example for the *t*-test by adding a second IV (with multiple levels): *Are there significant differences in depression severity by gender identity (1 = binary or 2 = nonbinary) and racial identity (1 = White or 2 = Black or 3 = Latinx)?* A two-way ANOVA is appropriate for answering this RQ as it reflects two categorical-level IVs, including gender identity comprised of the two following levels: (a) binary or (b) nonbinary. Racial identity is a categorical-level IV comprised of three levels in this RQ: (a) White, (b) Black, or (c) Latinx. The continuous-level DV is the clients' interval-level score on the PHQ-9 (depression severity). If the ANOVA indicates a statistically significant difference in depression severity exists as a main effect of gender identity (an IV with only two levels), no follow-up analyses in this example would be necessary; a CIT could visually inspect the mean depression severity scores for binary-identified versus non-binary identified students. Planned post hoc analyses would be useful if the ANOVA detects a main effect of racial identity on depression severity, as this IV has three levels and it is therefore possible for a main effect to signal that each group (represented by the levels) has a significantly different mean depression severity score from the other groups (e.g., White < Latinx < Black), or that only one group significantly differs in depression severity from the other two (Latinx ≈ Black > White). In models with two or more IVs, ANOVA allows for testing of both main effects and interaction effects (where the effect of one IV on a DV depends on the level of another IV).

Analysis of Variance With Repeated Measures. A repeated measures ANOVA is appropriate when one is employing a within-subjects design, in which data are collected from the same participants on two or more different occasions. In addition to the assumptions for ANOVA listed in Figure 2, the data should meet the assumption of sphericity for an ANOVA with repeated measures. Tests of sphericity (e.g., Mauchly's test of sphericity [W]) examine if the difference between all pairs of means is equal enough for statistical analysis. Sphericity replaces the assumption of independence for within-subjects analyses (e.g., dependent samples *t*-tests and repeated measures analysis of variance). For instance, a CIT might seek to investigate the following: *To what extent, if any, are there statistically significant differences over time in depression severity among nonbinary college students of color in the semester before, the semester during, and the semester after "bathroom bill" legislation was being considered in their state of residence?* These students' depression severity (scores on the PHQ-9) is the DV, and the IV, time of assessment, is comprised of three levels including (a) before, (b), during, and (c) after legislation affecting transgender rights are being discussed in their state legislature. Repeated measures can be added to any of the ANOVA/MANOVA analyses that are described in the following sections and are often utilized in quasi- or true-experimental designs. Hussey and Bisconti (2010), for example, employed repeated measures ANOVA to test the effectiveness of two different interventions to reduce sexual minority stigma among members of sororities on college campuses. Data on the DVs were gathered from all participants before and after the interventions. In a series of repeated measures ANOVAs (DV in each ANOVA was a different interval-level scale or subscale measuring attitudes and behaviors toward gay- and lesbian-identified people), the type of intervention (two levels: video and discussion intervention or panel discussion intervention) was the between-subjects factor whereas time of assessment (two levels: pre- or post-intervention) was the within-subjects factor.

Analysis of Variance: Multivariate. The fundamental difference between univariate and multivariate analyses is the number of DVs: univariate analyses include only one DV and multivariate analyses contain two or more DVs (Warne, 2014). Thus, a multivariate analysis of variance (MANOVA) is a group comparison analysis (see Figure 2 for assumptions) with categorical-level IV(s) and two or more continuous-level DVs. MANOVA should be computed when evidence from the extant literature suggests that the DVs are correlated (Trusty, 2011), as MANOVA aggregates the DVs into a linear variate or latent variable (Warne, 2014). For example, a CIT might pose the following research question: *Are there significant differences in depression severity and anxiety severity* (by *gender identity (1 = binary or 2 = nonbinary*) *and racial identity* (*1 = White or 2 =Black or 3 = Latinx)?* A two-way MANOVA is appropriate for answering this RQ, as there are two categorical-level IVs and two interval-level DVs including anxiety severity and depression severity. Similar to ANOVA, post hoc tests are completed for statistically significant findings in MANOVA. Computing a series of univariate ANOVAs is the most commonly used post hoc test for MANOVA, however, a discriminant analysis (DA) is a more appropriate follow-up test (Warne, 2014). A central underlying premise of MANOVA is that the DVs are correlated, however, each DV is investigated separately in a univariate ANOVA whereas a DA keeps the analysis in the multivariate realm by reversing the MANOVA to determine which of the DVs is contributing the most to group separation between the levels of the statistically significant IV. Kalkbrenner et al. (2020), for example, computed a factorial MANOVA with three categorical-level IVs, gender (female or male), ethnicity (White or non-White), and help-seeking history (sought personal counseling in the past or had not attended counseling in the past), to uncover differences across these groups in community college students' mental health literacy. The DVs, mental health literacy, were comprised of participants' scores on three composite scales (established surveys), each of which appraised a type of mental health literacy. Kalkbrenner et al. (2020) utilized a discriminant analysis (DA) as a post hoc test for significant MANOVA results (see p. 178).

Analysis of Covariance: Univariate. Analysis of covariance (ANCOVA) allows researchers to enter a continuous-level covariate (aka *control variable*) into the model to investigate mean differences between two or more IVs across a single DV while holding the covariate constant. In other words, ANCOVA is simply an ANOVA with a covariate added. Consider if a CIT posed the following RQ: *Are there significant differences in level of depression severity by gender identity (1 = binary or 2 = nonbinary*) *after controlling for the number of counseling sessions students have attended?* The ANCOVA would control for a potentially confounding variable (number of counseling sessions) by holding this variable constant (i.e., as if all participants attended the same number of counseling sessions), which will allow a CIT to more precisely investigate potential group differences in depression severity by generational status. Alif et al. (2020), for example, utilized ANCOVAs to compare mean scores on various DVs (e.g., fear of deportation for self, fear of deportation for family members, psychological distress, self-esteem, and academic performance) for community college students of color who self-identified as having one of three immigration statuses (one IV, with three levels: level 1 = stable; level 2 = temporary; level 3 = at-risk), while holding constant the following covariates: age, sex, region of origin, hours of paid work per day, hours of sleep per day, and hours spent socializing per day.

Analysis of Covariance: Multivariate. Analogous to the differences between ANOVA and ANCOVA, MANCOVA is simply a MANOVA that includes one or more control variables. (Recall that MANOVA is a multivariate test, i.e., there are at least 2 DVs.) A CIT might build on the previous RQ by asking the following: *Are there significant differences in depression severity and anxiety severity by gender identity (1 = binary or 2 = nonbinary*) *after controlling for the number of counseling sessions students have attended and their GPA?* Extending the example RQ for ANCOVA, the present RQ includes a second DV (anxiety severity) as well as an additional covariate (GPA). An example of MANCOVA is provided in Kam et al. (2019) who employed a 4 (Ethnic Group) X 2 (Gender) MANCOVA to test for differences in six help-seeking variables while holding constant age and sexual orientation.

Correlational/Predictive Analyses

Correlational/predictive analyses are used to measure the relationship or association between variables. Pearson product–moment correlation, regression analyses, and to a lesser extent, psychometric analyses are three common correlational/predictive analyses in counseling research.

Pearson Product–Moment Correlation. A Pearson product–moment correlation (see Figure 2 for assumption checking) allows one to investigate the association between two continuous-level variables (Swank & Mullen, 2017). Pearson's *r* is discussed in the present article, as it is the most commonly reported correlation coefficient in counseling research, however, a number of other correlational analyses exist, including point–biserial correlations for examining the association between one categorical-level variable and one continuous variable (see Bonett, 2019). Pearson's *r* ranges from -1 to +1 with absolute values closer to one denoting a stronger correlation. Negative values signify indirect relationships (increases in the level of one variable are associated with decreases in the level of the other) and positive values denote a direct relationship in which increases in the level of one variable are associated with increases in the level of the other. For example, a Pearson product–moment correlation would be appropriate if a researcher posed the following RQ: *To what extent, if any at all, is there an association between college students' grade point average (GPA) and their depression severity*? If, however, data fail to meet assumptions specified in Figure 2 or if variables are measured on an ordinal scale with a small sample, Spearman's rank correlation coefficient should be utilized (see Mukaka & Mukaka, 2012). Dueñas and Gloria (2020) utilized a Pearson product–moment correlation to identify associations existing among their primary study variables in a sample of Latinx undergraduates in the Midwestern United States. Three of these variables (motivation, belonging, and congruity) were explicitly grounded in a psychosociocultural framework meant to highlight the experiences of Latinx students in higher education (see pp. 104–105 for details).

Regression: Multiple Regression and Logistic Regression. Regression refers to a family of analyses in which predictor variables (similar to IVs in group comparison analyses, typically denoted as *X*) are used to predict (or regress) scores on a criterion variable (similar to DVs in group comparison analyses, typically denoted as *Y*). By predicting (regressing) *Y* on *X*, researchers can model the average value of *Y* as a function of *X*. This allows researchers to predict (with some degree of error) how the average value of *Y* will change as *X* changes. Simple regression is analogous to a correlation, as the analysis includes one continuous-level predictor variable (X_1) and one continuous-level criterion variable (Y_1). However, unlike a correlational analysis, simple regression computes an r^2 value or the coefficient of determination, which represents the shared variance between variables. In the context of simple regression, this allows a researcher to estimate the proportion of variance in *Y* explained by *X* (or the proportion of variance in *X* explained by *Y*, since *X* and *Y* are interchangeable in simple regression — just as they are in correlation).

Multiple Regression. Multiple regression is an extension of simple regression and allows one to test the extent to which multiple continuous-level predictor variables are significant predictors of one continuous-level criterion variable. For example, multiple regression would be the most appropriate statistical test if a CIT posed the following RQ: *Are the number of personal counseling sessions and the weekly average number hours of sleep significant predictors of college students' depression severity?* Turner et al. (2017), for example, used multiple regression to test a model in which past psychotherapy use (measured continuously) was predicted by students' ethnicity (measured categorically), therapy fears (measured continuously), and symptoms of psychological distress (measured continuously). Their overall model was significant, with therapy fears and psychological distress both explaining unique variance in past psychotherapy use. Specifically, as students' fears increased, the model predicted a reduction in past service use; in contrast, as students' levels of psychological distress increased, the model predicted increased use of past services. This example highlights how both continuous and categorical predictor variables can be used in the same multiple (or hierarchical multiple) regression model, as long as at least one predictor is measured continuously. In contrast to r^2 used in simple regression, R^2 is computed in multiple regression to represent the coefficient of multiple determination, which estimates the proportion of variance in the DV (*Y)* explained by the set of IVs ($X_1, X_2 \ldots, X_n$). A multivariate regression analysis or a path analysis based on structural equation modeling allows one to investigate the capacity of multiple continuous-level variables to predict scores on two or more continuous-level outcome variables. Outlining the details of these multivariate regression and path analysis extends beyond the scope of this article, however, readers can refer to Kline (2016) for more information if they are attempting to answer an RQ involving multiple criterion variables.

Hierarchical Multiple Regression. Hierarchical multiple regression (HMR) extends the regression model to allow CITs to examine if adding an additional predictor variable(s) to the analysis (aka *a second regression block*) significantly improves the overall predictive capacity of the model. HMR is typically most appropriate when variables have a priori relationships specified in the literature, often within a theoretical framework. HMR, for example, would allow a CIT to answer the following RQ: *Does adding the number of counseling sessions that* *college students attend improve the prediction of depression severity above age (measured in years) alone?* Age (X_1) would be entered into the first regression block and tested as a significant predictor variable of depression severity. The number of counseling sessions that college students attend would be entered as a predictor variable (X_2) into the second regression block and the change in value of R^2 will reveal if adding this variable improves the model's capacity to predict depression severity. If it does, that suggests X_2 (number of counseling sessions attended) explains unique variance not previously explained in Y by X_1 (age). An example of HMR is found in Dueñas and Gloria (2020), who used a four-step hierarchical regression to clarify which of three correlated IVs (motivation, belonging, and congruity) were significant predictors of Latinx undergraduates' sense of mattering (DV; see p. 105).

Logistic Regression. A binary logistic regression (LR) analysis allows one to test a categorical (dichotomous) criterion variable using continuous predictor variable(s). Specifically, LR tests the extent to which scores on at least one continuous-level predictor variable predict group membership in the levels of the categorical-level criterion variable. For example, a CIT could pose the following RQ: *Are college students' number of personal counseling sessions attended a significant predictor of whether they graduate?* The dichotomous criterion variable, graduation, is comprised of two categorical levels, including 1 = graduated from college or 2 = did not graduate from college. An example of LR is found in Goodwill and Zhou (2020), who found that perceived public stigma of receiving mental health treatment predicted suicidal ideation among college students of color (see pp. 3–4).

Psychometrics: Validity and Reliability Evidence of Scores

CITs using established scales that generate continuous-level data to measure a construct in counseling research must demonstrate that scores on the scale are appropriate for use in their population of interest. The process of creating a psychometrically-validated scale to measure a theoretical construct in a specific population is a rigorous, multistep, empirical process (explained in detail by Kalkbrenner, 2021) and typically involves conducting Exploratory and Confirmatory Factor Analyses (EFA and CFA, respectively) to establish construct validity. Even for CITs who do not wish to engage in scale construction research, it is crucial that all CITs understand the definition of validity and reliability, as well as benchmarks they can utilize to evaluate these qualities in different scales they are considering for use in their research or clinical work. Construct validity involves the degree to which scores on a test (such as a scale) measure the construct that the test was designed to measure and reliability refers to the consistency of scores on a test (Kalkbrenner, 2021). For example, the construct of depression severity is often measured using the PHQ-9 (described previously). The PHQ-9 tends to be considered a valid measure of depression severity with multiple populations since the construct validity of scores on the scale have been established (EFA) and confirmed (CFA) in a number of populations. Scores on the PHQ-9 also tend to correlate in expected ways with other measures of functioning and symptom impact (Kroenke et al., 2001), further supporting its construct validity. Scores on the PHQ-9 were also found to be reliable in that a measure of internal consistency reliability was within acceptable limits. These are some of the psychometric features CITs must consider when selecting scales to use in their counseling research or clinical practice. See Kalkbrenner (2021) for an overview of validity and reliability evidence.

Implications for Counselor Education

The present article has a number of implications for enhancing counselor education considering the CACREP standards associated with research methods and statistical analyses (CACREP, 2015, 2.F.8.f. & h) coupled with frequent errors in counseling research in regards to selecting the appropriate statistical analyses to answer the stated RQ (Wester et al., 2013). To this end, counselor educators can recommend this article to CITs to help them overcome common stumbling blocks identified in the extant literature pertaining to enhancing their quantitative research literacy and understanding of statistics, such as anxiety, lack of research self-efficacy, difficulty finding and using scholarly resources that explain statistical concepts in a clear and concise fashion, and gaps in master's-level quantitative research training (Holmes et al., 2018; Jorgensen & Umstead, 2020; Lalayants, 2012). The intended audience of this article is CITs who are enrolled in graduate-level introductory research methods and statistics courses as well as counselor educators who are looking for resources on teaching quantitative research. To this end, the authors provide implications for both CITs and counselor educators in the following sections.

Implications for Counselors-In-Training

The present authors aimed to demystify the quantitative research process by providing a general overview of writing quantitative research questions and matching variables with a number of commonly-used statistical tests in counseling research, as well as providing references to exemplar articles for each statistical test, delivered in a single and accessible article. We hope this accessibility and comprehensibility will improve CITs' perceptions of the research process and increase involvement of CITs in quantitative research (Steele & Rawls, 2015). The present empirical guide for matching variables with the appropriate statistical analyses is based, in part, on the research-based CACREP standards (e.g., CACREP, 2015, 2.F.8.f. & h) and has potential to facilitate CITs quantitative research literacy. The present article has pragmatic utility for CITs working on quantitative research proposals or theses, as they can refer to this resource (one-stop-shop) when matching their variables of interest with the most appropriate statistical test to answer their research question (see Figure 1).

The two figures in this article provide a concise resource for matching variables with statistical analyses (see Figure 1) as well as an outline of statistical assumptions and corresponding analyses (see Figure 2), which CITs can use as a reference for evaluating the rigor and utility of research findings for potential use with clients. Specifically, CITs can compare the methodology and statistical analyses in a research study to the guidelines for matching RQs, variables, and statistical analyses in this article as one way to evaluate the rigor and potential generalizability of research findings for informing their work with clients. This allows CITs to approach empirical literature as informed consumers on behalf of their clients as they consult this literature to determine what evidence exists to support the validity and reliability of scores on instruments they might use in evaluating their clients' concerns or to ascertain which interventions are evidence-based (Dukic, 2015) for particular client populations. This article could also be used by CITs to diagnose gaps in their understanding of the quantitative research process or evaluate quantitative research competencies at various points in graduate training.

Implications for Counselor Educators

The present article has several uses for structuring course content in counselor education. Counselor educators, for example, can include this article as required or recommended reading in graduate classes such as counseling research, testing and assessment, and other classes that include coursework in statistics and quantitative research methods. Faculty can refer to the present article to structure class lectures, discussion, and assignments. Counselor educators can use the two figures in this article as handouts or educational tools for teaching CITs how to match variables with statistical analyses (see Figure 1) and when teaching about statistical assumptions and corresponding analyses (see Figure 2). This article can also be utilized during experiential class activities. For example, counselor educators can randomly assign CITs into breakout groups and designate each group a particular statistical analysis. With the support of the instructor, students can work together to create a RQ and explain their statistical analysis to the rest of the class. Counselor educators might also invite CITs to quiz one another in identifying the appropriate statistical test if the nature of the RQ, the number of IVs, the number of DVs, the number of control variables, or the scale at which any of these variables were measured were changed in some manner (as often happens in actual research practice).

Conclusion

Ultimately, we hope that reading this article will support CITs in refining the skills necessary to articulate specific quantitative research questions and testable hypotheses, select appropriate statistical procedures, and make defensible claims about their research findings, thus contributing to the knowledge base within counselor education and supervision. The present article offers counselor educators and their students a one-stop-shop, or single scholarly source, for accessing: (a) a succinct overview of common statistical tests; (b) criteria for matching variables with statistical analyses and recognizing the assumptions underlying these approaches; and (c) numerous exemplars of these approaches found in refereed journal articles.

References

Alif, A., Nelson, B. S., Stefancic, A., Ahmed, R., & Okazaki, S. (2020). Documentation status and psychological distress among New York City community college students. *Cultural Diversity and Ethnic Minority Psychology, 26*(1), 11–21. https://doi.org/10.1037/cdp0000290

Ayers, D. F. (2011). A critical realist orientation to learner needs. *Adult Education Quarterly, 61*(4), 341–357. https://doi.org/10.1177/0741713610392769

Balkin, R. S. (2020). The challenge of counseling research in developing a signature pedagogy for quantitative methods. *Teaching and Supervision in Counseling, 2*(2), 47–52. https://doi.org/10.7290/tsc020206

Balkin, R. S., & Sheperis, C. (2011). Evaluating and reporting statistical power in counseling research. *Journal of Counseling & Development, 89*(3), 268–272. https://doi.org/10.1002/j.1556-6678.2011.tb00088.x

Bhaskar, R. (1978). *A realist theory of science* (2nd ed.). Harvester Press.

Bonett, D. (2019). Point-biserial correlation: Interval estimation, hypothesis testing, meta-analysis, and sample size determination. *British Journal of Mathematical & Statistical Psychology, 73*(1), 113–144. https://doi.org/10.1111/bmsp.12189

Council for Accreditation of Counseling and Related Educational Programs. (2015). *2016 CACREP standards*. https://www.cacrep.org/for-programs/2016-cacrep-standards/

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE.

Dueñas, M., & Gloria, A. M. (2020). ¡Pertenecemos y tenemos importancia aquí! Exploring sense of belonging and mattering for first-generation and continuing-generation Latinx undergraduates. *Hispanic Journal of Behavioral Sciences, 42*(1), 95–116. https://doi.org/10.1177%2F0739986319899734

Dukic, M. S. (2015). The research training environment and its potential influence on graduate level counseling students' attitudes toward and interest in research. *Vistas Online, 12*, 1–11.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*(4), 1149–1160.

Field, A. P. (2018). *Discovering statistics using IBM SPSS Statistics* (5th ed.). SAGE.

Goodwill, J. R., & Zhou, S. (2020). Association between perceived public stigma and suicidal behaviors among college students of color in the US. *Journal of Affective Disorders, 262*, 1–7. https://doi.org/10.1016/j.jad.2019.10.019

Holmes, B., Waterbury, T., Baltrinic, E., & Davis, A. (2018). Angst about academic writing: Graduate students at the brink. *Contemporary Issues in Education Research (CIER), 11*(2), 67–72.

Hussey, H. D., & Bisconti, T. L. (2010). Interventions to reduce sexual minority stigma in sororities. *Journal of Homosexuality, 57*(4), 566–587. https://doi.org/10.1080/00918361003609184

International Registry of Counsellor Education Programs. (2015). *The IRCEP standards*. https://www.ircep.org/

Jorgensen, M. F., & Umstead, L. K. (2020). Developing a signature pedagogy in master's research training. *Teaching and Supervision in Counseling, 2*(2), 53–61. https://doi.org/10.7290/tsc020207

Kalkbrenner, M.T. (2021). A practical guide to instrument development and score validation in the social sciences: The MEASURE Approach. *Practical Assessment, Research & Evaluation. 26*, Article 1. https://doi.org/10.7275/svg4-e671

Kalkbrenner, M.T., Sink, C.A., & Smith, J.L. (2020). Mental health literacy and peer-to-peer counseling referrals among community college students. *Journal of Counseling and Development. 98*, 172–182. https://doi.org/10.1002/jcad.12311

Kam, B., Mendoza, H., & Masuda, A. (2019). Mental health help-seeking experience and attitudes in Latina/o American, Asian American, Black American, and White American college students. *International Journal for the Advancement of Counselling, 41*(4), 492–508. https://doi.org/10.1007/s10447-018-9365-8

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). Guilford Press.

Kroenke, K., Spitzer, R., & Williams, J. (2001). The PHQ-9. *Journal of General Internal Medicine, 16*(9), 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Lalayants, M. (2012). Overcoming graduate students' negative perceptions of statistics. *Journal of Teaching in Social Work, 32*(4), 356–375. https://doi.org/10.1080/08841233.2012.705259

McHugh, M. (2013). The chi-square test of independence. *Biochemia Medica, 23*(2), 143–149.

Mukaka, M., & Mukaka, M. (2012). Statistics corner: A guide to appropriate use of correlation coefficient in medical research. *Malawi Medical Journal: The Journal of Medical Association of Malawi, 24*(3), 69–71.

Mushonga, D. R., & Henneberger, A. K. (2020). Protective factors associated with positive mental health in traditional and nontraditional Black students. *American Journal of Orthopsychiatry, 90*(1), 147–160. https://doi.org/10.1037/ort0000409

Ruble, L., McGrew, J., & Toland, M. (2012). Goal attainment scaling as an outcome measure in randomized controlled trials of psychosocial interventions in autism. *Journal of Autism and Developmental Disorders, 42*(9), 1974–1983. https://doi.org/10.1007/s10803-012-1446-7

Steele, J. M., & Rawls, G. J. (2015). Quantitative research attitudes and research training perceptions among master's-level students. *Counselor Education and Supervision, 54*(2), 134–146. https://doi.org/10.1002/ceas.12010

Swank, J. M., & Mullen, P. R. (2017). Evaluating evidence for conceptually related constructs using bivariate correlations. *Measurement and Evaluation in Counseling and Development, 50*(4), 270–274. https://doi.org/10.1080/07481756.2017.1339562

Tabachnick, B. G., & Fidell, L. S. (2014). *Using multivariate statistics* (6th ed.). Pearson Education, Inc.

Trusty, J. (2011). Quantitative articles: Developing studies for publication in counseling journals. *Journal of Counseling & Development, 89*(3), 261–267. https://doi.org/10.1002/j.1556-6678.2011.tb00087.x

Turner, E. A., Camarillo, J., Daniel, S., Otero, J., & Parker, A. (2017). Correlates of psychotherapy use among ethnically diverse college students. *Journal of College Student Development, 58*(2), 300–307. https://psycnet.apa.org/doi/10.1353/csd.2017.0022

Warne, R. T. (2014). A primer on multivariate analysis of variance (MANOVA) for behavioral scientists. *Practical Assessment, Research & Evaluation, 19*(1), 17. https://doi.org/10.7275/sm63-7h70

Wester, K. L., & Borders, L. D. (2014). Research competencies in counseling: A Delphi study. *Journal of Counseling & Development*, *92*(4), 447–458. https://doi.org/10.1002/j.1556-6676.2014.00171.x

Wester, K. L., Borders, L. D., Boul, S., & Horton, E. (2013). Research quality: Critique of quantitative articles in the *Journal of Counseling & Development*. *Journal of Counseling & Development*, *91*(3), 280–290. https://doi.org/10.1002/j.1556-6676.2013.00096.x