Editorial Manager(tm) for Nursing Research
Manuscript Draft

Manuscript Number: NRES-D-08-00245

Title: Use of the Concordance Correlation Coefficient when Examining Agreement in Dyadic Research

Article Type: Method

Corresponding Author: Post Doctoral Fellow Christina Quinn, DNS

Corresponding Author's Institution: Emory University

First Author: Christina Quinn, DNS

Order of Authors: Christina Quinn, DNS; Michael J Haber, PhD; Yi Pan, MS

Manuscript Region of Origin: UNITED STATES

Abstract: Abstract

Background:
The statistical measure used to quantify the degree of agreement or congruence between two research subjects has been the intraclass or the Pearson correlation coefficient; however, the concordance correlation coefficient (CCC) is another measure of agreement that examines agreement between two observers or raters.

Objectives:
The purpose of this article is to (i) highlight the differences among three statistical measures used to quantify the degree of agreement or congruence, (ii) demonstrate the use of the CCC in examining agreement between heart failure (HF) patients and their family members, and (iii) provide nurse researchers another method for evaluating agreement.

Method:

Symptom evaluation scores obtained from HF patients and their family members were used in the analysis of this study. In order to explain the use of the CCC in this analysis, a distinction between Pearson correlation coefficient (PCC) and ICC is discussed. The CCC calculation is then described in detail.

Results:

The HF patients in this sample were 71±9.6 years in age and 40% male, and 41.4% African American. The majority of family members were female (75%). There were several different categories of family members, however most were spouses. The CCC results indicated that no symptom achieved good agreement and 8 of 14 symptoms were in moderate agreement (0.4 ≤ CCC ≤ 0.7) within the dyads. Of the six symptoms with poor agreement (0 ≤ CCC < 0.4), HF patients and their FM's agreed least on worsening cough (CCC: 0.152, 95% CI(-.134, .413)) and bloated abdomen (CCC: 0.055, 95% CI(-.224, .325)).

Discussion:

Applying the CCC to dyadic data from HF patients and FM's, identified symptoms in which the patient and family member had the most and least agreement. The six symptoms with poor agreement were symptoms that can show HF decline and may be important when examining future nursing interventions. The CCC offered more detail such as providing a confidence interval along with the correlation. Further study using the CCC with dyadic data along with other family factors that influence agreement is needed.

Key words:

Correlations, Dyads, Agreement

November 13, 2008

Molly C. Dougherty, PhD, RN, FAAN

Editor, Nursing Research

Dear Dr. Dougherty,

Attached you will find our manuscript, Use of the Concordance Correlation Coefficient when Examining Agreement in Dyadic Research, which describes a correlation measurement that has not been widely used in nursing research.

Information presented may help guide future researchers when determining the type of analysis to use when the interest involves congruence (or agreement) between two subjects.

This is being submitted for the Methodology section of your journal. Information in this manuscript has not been published or used previously. The authors (Christina Quinn, Michael Haber, and Yi Pan) have no conflict of interest to report and have followed ethical guidelines in preparation of this manuscript.

Thank you and please notify me with any concerns or questions.

Sincerely,

Christina Quinn RN, DNS
Post Doctoral Fellow
Emory University
Nell Hodgson Woodruff School of Nursing
cquinn2@emory.edu

Use of the Concordance Correlation Coefficient when Examining Agreement in Dyadic Research

Christina Quinn, RN, DNS
Post Doctoral Fellow
Nell Hodgson Woodruff School of Nursing Emory University

Michael J. Haber, PhD.
Professor
Department of Biostatistics
Rollins School of Public Health
Emory University

Yi Pan, MS
Department of Biostatistics
Rollins School of Public Health
Emory University

Corresponding author:
Christina Quinn RN, DNS
Emory University
Nell Hodgson Woodruff School of Nursing
1599 Clifton Road
MS 1599-001-1AJ
Atlanta, GA 30322
Phone: 404 712 8478
FAX (404 727 9382)
Email: cquinn2@emory.edu

Funding: The Center for the Study of Symptoms, Symptom Interactions and Health Outcomes, NIH NINR P20 NR007798, Post Doctoral Training Grant from NIH NINR F32 NR009888-01A1 2007-2008.

Abstract

Background: The statistical measure used to quantify the degree of agreement or congruence between two research subjects has been the intraclass or the Pearson correlation coefficient; however, the concordance correlation coefficient (CCC) is another measure of agreement that examines agreement between two observers or raters.

Objectives: The purpose of this article is to (i) highlight the differences among three statistical measures used to quantify the degree of agreement or congruence, (ii) demonstrate the use of the CCC in examining agreement between heart failure (HF) patients and their family members, and (iii) provide nurse researchers another method for evaluating agreement.

Method: Symptom evaluation scores obtained from HF patients and their family members were used in the analysis of this study. In order to explain the use of the CCC in this analysis, a distinction between Pearson correlation coefficient (PCC) and ICC is discussed. The CCC calculation is then described in detail.

Results: The HF patients in this sample were 71±9.6 years in age and 40% male, and 41.4% African American. The majority of family members were female (75%). There were several different categories of family members, however most were spouses. The CCC results indicated that no symptom achieved good agreement and 8 of 14 symptoms were in moderate agreement ($0.4 \leq CCC \leq 0.7$) within the dyads. Of the six symptoms with poor agreement (0

$\leq$ CCC $<$ 0.4), HF patients and their FM’s agreed least on worsening cough (CCC: 0.152, 95% CI(-.134, .413)) and bloated abdomen (CCC: 0.055, 95% CI(-.224, .325)).

Discussion: Applying the CCC to dyadic data from HF patients and FM’s, identified symptoms in which the patient and family member had the most and least agreement. The six symptoms with poor agreement were symptoms that can show HF decline and may be important when examining future nursing interventions. The CCC offered more detail such as providing a confidence interval along with the correlation. Further study using the CCC with dyadic data along with other family factors that influence agreement is needed.

Key words: Correlations, Dyads, Agreement

Background/Introduction

Examining congruence or agreement between patient and family member ratings of some aspect of health have increased in frequency in studies of patients with different chronic disease states. Early studies of agreement with oncology patients focused on understanding how family caregivers' interpreted symptom distress as compared to the patients' own perception of the same symptoms (Lobchuk, Kristjanson, Degner, Blood, & Sloan, 1997; O'Brien & Francis, 1988). Cancer pain management and the congruence between patient and family member have also been studied in depth (Lin, 2001; McMillan & Moody, 2003; Riley-Doucet, 2005). Recently, congruence of depressive symptoms between persons with stroke and their caregivers has been studied (Klinedinst, Clark, Blanton, & Wolf, 2007). Other recent examples of interest in dyadic congruence can be found in the cardiology literature with a study examining if patient and spouse's perceptions of myocardial infarction predict recovery (Figueiras & Weinman, 2003) and, a study on heart failure symptom congruence between patient and family member (Quinn, 2007).

In most of these studies, the statistical measure used to quantify the degree of agreement or congruence between two research subjects has been the intraclass correlation (ICC) or the Pearson's R correlation. In addition to the ICC, some studies use mean directional difference scores between patient and family member and analyze this with a paired t-test (McPherson, Wilson, Lobchuk, & Brajtman, 2008).

While the ICC and Pearson's statistics are useful to examine if the measurements or ratings by two people vary between one another, these tests are examining if the measures vary in the same direction. These coefficients often attain high values even when there are substantial differences between the measurements made by the observers within the dyad. The concordance correlation coefficient (CCC) is another measure of agreement that examines how closely the actual measurements of two observers agree with one another (Lin, 1989, Barnhart, Haber, & Lin, 2007). The CCC has been developed and tested in the biopharmaceutical industry to assess agreement between methods of measurement, and can also be applied to nursing research studies. When examining the CCC, values ranging from 0.7 – 1.0 are considered as good agreement, 0.4 - 0.7 as moderate agreement, and 0.0 – 0.4 as poor agreement.

The purpose of this article is to (i) highlight the differences among the three coefficients mentioned above, (ii) demonstrate the use of the CCC in examining agreement between heart failure (HF) patients and their family members, and (iii) provide nurse researchers another method for evaluating agreement.

Approach

Data Collection

The data were collected as part of a study of HF patients and their family members which examined congruence between 14 symptoms common to heart failure patients. Data were collected from 70 dyads all receiving home health care. The study was approved by the University institutional review board (IRB) and permission was attained from the respective home health agencies. Informed consent was obtained from both patient and family member prior to any data collection.

The Heart Failure Symptom Survey (HFSS) was used as the primary measurement tool for symptom assessment. This tool consists of 14 symptoms common to HF such as: shortness of breath with activity; fatigue; extremity swelling; depression; and others (Pozehl, Duncan, & Hertzog, 2006). Each symptom is assessed using an eleven-point scale with scores ranging from 0-10. A score of zero indicates that the person did not experience the symptom; with 10 indicating very severe or frequent. A mean of two of the items (frequency and severity) for each symptom is computed to provide a mean symptom evaluation score for each symptom. Dyads completed measures of perceived frequency and severity of these HF symptoms experienced by the patient over a seven-day period. During the data collection process, the dyad completed the instruments without collaborating or discussing responses.

Methodology

Data analysis

The CCC analysis was computed using a SAS (SAS Institute, Cary, NC) macro that is available at the Rollins School of Public Health at Emory

University's web site: http://www.sph.emory.edu/observeragreement/. The CCC computation is also described in detail using two biomedical examples published by Crawford et al. (Crawford, Kosinski, Lin, Williamson, & Barnhart, 2007).

Symptom evaluation scores for each dyad were used in the estimation of the CCC in this study. In order to explain the use of the CCC in analysis, a distinction between Pearson correlation coefficient (PCC), the ICC and the CCC will be presented.

Agreement between two observers is usually evaluated via the PCC or an ICC. Let's denote the observations made on subjects i by observer 1 and 2 as X_i and Y_i, respectively. Then the PCC is defined as:

$$PCC = \frac{\sum_i (X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_i (X_i - \overline{X})^2 \sum_i (Y_i - \overline{Y})^2}},$$

where $\overline{X}$ and $\overline{Y}$ are the sample means. The PCC is designed to detect any linear association between the observers' readings. Thus, if the readings satisfy:

$$Y_i = a + b \cdot X_i$$

for *any* constants a and b then the PCC will attain the value 1. The observations in Example 1 (Table 1 and Figure 1) satisfy this equation with $a = 1$ and $b = 2$.

Intraclass Correlation Coefficients (ICC's) are usually defined in the context of an analysis of variance (ANOVA) model. This model expresses the total variance as a sum of three components: the between-subject variance, σ_S^2, the between-observers variance, σ_O^2, and the error (residual) variance, σ_E^2. Using the terminology and notation of McGraw and Wong (McGraw & Wong, 1996), there are at least two types of ICC's: the ICC for consistency, ICC(C,1), and the ICC for agreement, ICC(A,1), which are defined as follows:

$$\text{ICC(C,1)} = \frac{\sigma_S^2}{\sigma_S^2 + \sigma_E^2},$$

$$\text{ICC(A,1)} = \frac{\sigma_S^2}{\sigma_S^2 + \sigma_O^2 + \sigma_E^2}.$$

The ICC(C,1) is designed to detect a linear association of the type given in equation (1) with $b = 1$. In other words, this coefficient attains the value 1 whenever there is a fixed difference between X and Y. Example 2 (Table 1 and Figure 1) presents data of this kind with $a = 3$.

ICC(A,1) will attain the value 1 only if the linear association (equation (1)) between the observers is such that $a = 0$ and $b = 1$, i.e., only if $X_i = Y_i$ for every subject i. Example 3 in Table 1 and figure 1 illustrates this situation: (see Table1 and Figure 1).

In 1989 Lin (Lin, 1989) introduced the CCC as a measure of agreement. This coefficient is based on comparing the mean squared difference between the reading of the observers, $(X - Y)^2$, to its expected value in the case where the observers are independent of each other. The CCC is not based upon an ANOVA model, and it can be estimated directly from the means and variances of the joint distribution of X and Y:

$$CCC = \frac{2 \cdot S_{XY}}{(\bar{X} - \bar{Y})^2 + S_X^2 + S_Y^2}$$

where S_X^2 and S_Y^2 are the sample variances and S_{XY} is the sample covariance. Barnhart et al. (Barnhart, Haber, & Song, 2002) showed that if the assumptions of the above-mentioned ANOVA model are satisfied then the CCC coincides with the agreement ICC. Since the CCC is easier to calculate and it does not require the ANOVA assumptions, we will now refer to ICC(A,1) as the CCC.

So far we introduced three coefficients for evaluating inter-observer agreement: PCC, ICC-C and CCC. In order to decide which of the coefficients should be used in a specific case, one must first realize that the three coefficients measure different kinds of agreement. The PCC is a measure of *linear agreement*, where the points (X_i, Y_i) just need to be close to a straight line as in Example 1. There is no requirement that the differences $X_i - Y_i$ will be small or fixed across subjects. The ICC-C is a measure of *additive agreement*, hence it indicates the extent to which the differences $X_i - Y_i$ are fixed across subjects, but there is no requirement that this fixed difference will be small (Example 2). The CCC measures *absolute agreement*, i.e., it requires the differences $X_i - Y_i$ to be small (Example 3). These concepts of linear, additive and absolute agreement are special cases of *relational agreement* (Fagot, 1993). The choice of the appropriate coefficient depends on the type of agreement that one wants to assess. In most cases one wants to check whether the observed values, X_i and Y_i are close to each other for all (or most) subjects. In this case the coefficient of absolute agreement, i.e., the CCC, is the one that should be used. However, there are situations where one only wants to check whether the observations (X_i, Y_i) follow a certain pattern. If one only wants to know whether X and Y usually vary in the same direction, i.e. small X's go with small Y's and large X's go with large Y's, then the coefficient of linear agreement, i.e., the PCC, is an appropriate measure of agreement. In fact one may argue that correlation is a special case of observer agreement where the observers are not penalized if their readings are different as long as the reading of one of them is a linear function of those of the other observer, as in Example 1. Similarly, the coefficient of additive agreement, ICC-C, is an acceptable measure of observer agreement if observers' readings are allowed to differ by a fixed (though maybe large) value.

In the current example we are interested in the actual differences between the symptom severity scores assigned by the patient and the family member. In other words, we are interested in *absolute* agreement, and therefore we use the concordance correlation coefficient (CCC).

Results

The heart failure (HF) patients in this sample of 70 dyads were 71±9.6 years in age, 40% male, 44.3% married, and 41.4% African American and 57.1% Caucasian. The majority of family members were female (75%) and younger than the patient (57±15 years). There were several different categories of family members, however most were spouses. Table 2 summarizes the demographic variables of interest in this study.

Agreement between HF patients and their FM's was assessed using the CCC as well as its 95% confidence interval (Table 3). Within the entire 14 symptoms, no symptom achieved good agreement; however, HF patients and their FM's experienced moderate agreement on 8 symptoms within the dyads. The symptom with highest agreement was extremity swelling (CCC= .620, 95% CI(.406 - .767). Of the six symptoms with poor agreement, HF patients and their FM agreed least on worsening cough (CCC= .152, 95% CI(-.134 - .413) and bloated abdomen (CCC= .055, 95% CI(-.224 - .325).

Discussion

Agreement between observers when examining symptoms of a disease process can be assessed using several methods. The purpose of this article has been to demonstrate the use of the CCC in examining agreement between HF patients and their FM when asked to rate the same symptoms in a specified time frame. Although other methods such as PCC and ICC are frequently used, this article has illuminated an alternative method for assessing agreement. Advantages of the CCC are the ability to categorize (good, moderate, and poor) agreement based on CCC obtained and the ability to examine a confidence interval (CI). From a statistical perspective, a limitation of the CCC is its dependence on the variability between subjects (Atkinson G. & Nevill, 1997). For example, the CCC will attain a much higher value if it is calculated from a group of patients that include less severe and more severe cases than if it was calculated from a group of very severe cases only. This will happen regardless of the magnitude of the differences between scores assigned by the observers to the same patient. In other words, one has to be very careful when comparing values of the CCC obtained from different studies, or even from the same study conducted by the same observers at different locations. Recently, Haber and Barnhart (Haber & Barnhart, 2008) introduced a new family of coefficients of agreement that do not share this limitation. However, these new *coefficients of individual agreement* require that each observer makes at least two readings on each study subject.

After obtaining the CCC, other statistical measures should be used to examining agreement in more detail. A graphical approach that can be used to explore the data is by the Bland-Altman method (Bland & Altman, 1986). This is a useful tool for plotting data between two observers. Graphics add a visual component to the analysis of agreement.

Applying the CCC to dyadic data from HF patients and their FM, identified symptoms in which the patient and FM had the most and least agreement. The six symptoms with poor agreement were symptoms that can show HF decline: fatigue, shortness of breath (SOB) at rest, SOB when lying in bed, SOB waking up at night, worsening cough, and bloated abdomen. Analysis of these data with Spearman Rho resulted in five of these same symptoms having poor agreement (Quinn, Dunbar, & Higgins, 2008). Thus, the

CCC offered more detail and refinement to this HF symptom data. Further study is required to examine the usefulness of the CCC with dyadic data and other factors that influence agreement such as family functioning, communication, and family support.

References

Atkinson G., & Nevill, A. B.-. (1997). Comment on the use of concordance correlation to assess the agreement between two variables. *Biometrics, 53*, 775-777.

Barnhart, H. X., Haber, M., & Song, J. (2002). Overall concordance correlation coefficient for evaluating agreement among multiple observers. *Biometrics, 58*(4), 1020-1027.

Barnhart, H. X., Haber, M. J., & Lin, L. I. (2007). An overview on assessing agreement with continuous measurements. *Journal of Biopharmaceutical Statistics, 17*(4), 529-569.

Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet*, 307-311.

Crawford, S. B., Kosinski, A. S., Lin, H.-M., Williamson, J. M., & Barnhart, H. X. (2007). Computer programs for the concordance correlation coefficient. *Computer Methods & Programs in Biomedicine, 88*(1), 62-74.

Fagot, R. F. (1993). A generalized family of coefficients of relational agreement for numerical scales. *Psychometrika 58*, 357-370.

Figueiras, M. J., & Weinman, J. (2003). Do similar patient and spouse perceptions of myocardial infarction predict recovery? *Psychology & Health Vol 18(2) 2003, 201-216*.

Haber, M. J., & Barnhart, H. X. (2008). A general approach to evaluating agreement between two observers or methods of measurement from quantitative data with replicated measurements. . *Statistical Methods in Medical Research, 17*, 151-169.

Klinedinst, J., Clark, P. C., Blanton, S., & Wolf, S. L. (2007). Congruence of depressive symptom appraisal between persons with stroke and their caregivers. *Rehabilitation Psychology, 52*(2), 215-225.

Lin, C. (2001). Congruity of cancer pain perceptions between Taiwanese patients and family caregivers: relationship to patients' concerns about reporting pain and using analgesics. *Journal of Pain and Symptom Management 21*(1), 18-26.

Lin, L. I. (1989). A concordance correlation coefficient to evaluate reproducibility. *Biometrics, 45*(1), 255-268.

Lobchuk, M. M., Kristjanson, L., Degner, L., Blood, P., & Sloan, J. A. (1997). Perceptions of symptom distress in lung cancer patients: I. Congruence between patients and primary family caregivers. *Journal of Pain & Symptom Management, 14*(3), 136.

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1*, 30-46.

McMillan, S. C., & Moody, L. E. (2003). Hospice patient and caregiver congruence in reporting patients' symptom intensity. *Cancer Nursing 26*(2), 113-118.

McPherson, C. J., Wilson, K. G., Lobchuk, M. M., & Brajtman, S. (2008). Family caregivers' assessment of symptoms in patients with advanced cancer: concordance with patients and factors affecting accuracy. *Journal of Pain and Symptom Management, 35*(1), 70-82.

O'Brien, J., & Francis, A. (1988). The use of next-of-kin to estimate pain in cancer patients. *Pain, 35*(2), 171-178.

Pozehl, B., Duncan, K., & Hertzog, M. (2006). *Heart failure symptom survey: validity and reliability*. Paper presented at the Heart Failure Society of America 10th Annual Scientific Meeting.

Quinn, C., Dunbar, S.B., Higgins, M. (2008). *Can the family member serve as proxy in heart failure symptom assessment?* Journal of Cardiovascular Nursing (under review).

Quinn, C. (2007). *Symptom Management: Heart failure patients and family caregivers. NIH NINR F32 NR009888-01A1 2007-2008* .

Riley-Doucet, C. (2005). Beliefs about the controllability of pain: congruence between older adults with cancer and their family caregivers. *Journal of Family Nursing 11*(2), 225-241.

Table 1

Three examples of readings (X_i, Y_i) of two observers on four subjects

	Example 1		Example 2		Example 3	
Subject (i)	X_i	Y_i	X_i	Y_i	X_i	Y_i
1	1.0	4.0	1.0	5.0	1.0	1.0
2	2.0	6.0	2.0	6.0	2.0	2.0
3	3.0	8.0	3.0	7.0	3.0	3.0
4	4.0	10.0	4.0	8.0	4.0	4.0
PCC	1.00		1.00		1.00	
ICC (C,1)	0.80		1.00		1.00	
CCC	0.67		0.72		1.00	

Table 2

Demographic Variables: heart failure patients and family caregivers

	Patient	Family Member
Age	71 ± 9.6 years	56.7 ± 15 years
Gender	40% male	25% male
	60% female	75% female
Ethnicity	41.4% African American	Same as patient
	57% Caucasian	
	1.4% Asian	
Marital Status	44.3% Married	60% Married
	7.1% Single	40% Single
	48.6% Widowed/Divorced	
Relationship of		43% Spouse
family member		32% Adult Children
to patient		5% Adult Sibling
		20% Other

N= 70 dyads

Table 3
Heart Failure Symptom Survey:
Concordance Correlation Coefficient (CCC) with 95% Confidence Interval

Symptom	Moderate Agreement	Poor Agreement
Extremity Swelling	.620 (.406 - .767)	
Forgetfulness	.488 (.243 - .674)	
Dizziness	.480 (.233 - .668)	
Irregular Heart Beat	.455 (.208 - .648)	
SOB with Activity	.446 (.192 - .644)	
Chest Pressure	.443 (.187 - .641)	
Difficulty Sleeping	.414 (.154 - .620)	
Depressed	.404 (.151 - .609)	
Fatigue		.398 (.137 - .608)
SOB at rest		.391 (.127 - .603)
SOB lying down		.330 (.056 - .557)
SOB wakeup at night		.305 (.037 - .533)
Worsening Cough		.152 (-.134 - .413)
Bloated Abdomen		.055 (-.224 - .325)

SOB=shortness of breath

Figure

Figure 1

Three examples of readings (X_i, Y_i) *of two observers on four subjects*