

# Statsconsultancy Ltd

Freelance Statistical Consultancy

**Address:** 40 Longwood Lane, Amersham, Bucks. HP7 9EN

**Email:** [paul@statsconsultancy.co.uk](mailto:paul@statsconsultancy.co.uk)

**Phone:** 07905 530446

**Web:** [www.statsconsultancy.co.uk](http://www.statsconsultancy.co.uk)

## Statistical Report

**Author:** Paul Bassett

**Client:** Natasha Lindsay

**Date:** 27<sup>th</sup> September 2019

**Report Number:** 1

**Description:** Inter-rater agreement for the NCAAD  
spotlight dataset

## Introduction

The National Clinical Audit of Anxiety and Depression performed a Spotlight audit relating to Psychological Therapies. As part of the audit, the data for some patients were collected twice by two different raters. The objective of the analysis in this report is to examine the inter-rater agreement for the data collected.

## Statistical methods

All analysis examined the strength of the inter-rater agreement for a number of the audit measures. The method of analysis was dependent on the nature of the data collected. Continuous variables were analysed using the intra-class correlation (ICC), whilst categorical variables were analysed using the kappa method. In addition to the calculated values, the uncertainty in each measurement was indicated by calculating a corresponding confidence interval.

The ICC method divides the total variation in the measurements from both raters combined into two sources, that due to variation between patients, and that due to variation within patients (due to repeat measurements by the two raters). The ICC is the proportion of the total variation that is due to variation between patients. If the agreement is good there should be little within-patient variation, with most variation between-patients, resulting in an ICC value close to 1.

For the categorical variables, the kappa statistic measures the agreement over and above that which would be expected due to chance. This is measured on a scale ranging up to a maximum agreement of 1.

Although the kappa and ICC methods both score agreement on a scale up to 1, the interpretation of the values produced is different. A suggested interpretation of the strength of agreement based on the kappa and ICC values is suggested in Table 1. The kappa interpretation is fairly well documented, whilst there is less agreement on how to interpret the ICC values.

*Table 1: Suggested interpretation of Kappa and ICC values*

Strength of agreement	Kappa	ICC
Poor	< 0.20	<0.60
Fair	0.21 – 0.40	0.61 – 0.70
Moderate	0.41 – 0.60	0.71 – 0.80
Good	0.61 – 0.80	0.81 – 0.90
Very Good	0.81 – 1.00	0.91 – 1.00

## Results

### a) Demographic variables

The first set of analyses examined the agreement between raters for the demographic variables.

The agreement for the continuous variables is summarised in Table 2. The figures reported are the number of subjects in the analysis, along with the intra-class correlation and its corresponding confidence interval.

*Table 2: Continuous demographic variables*

Question	n	ICC (95% CI)
Age (Q1)	792	0.97 (0.97, 0.98)

The data suggested good agreement between raters in terms of the age of the patient.

Analyses were also performed for the demographic measures that were assessed on a categorical scale. A summary of the results is shown in Table 3. The figures shown are the number of subjects in the analysis, and the calculated kappa values along with their corresponding confidence intervals.

*Table 3: Categorical demographic variables*

Variable	n	Kappa (95% CI)
Gender (Q2)	790	0.95 (0.88, 1.00)
Ethnicity (Q3)	702	0.84 (0.80, 0.89)
Sexual Orientation (Q4)	355	0.88 (0.79, 0.96)
Employment status (Q5)	587	0.79 (0.76, 0.83)
Accommodation status (Q6)	631	0.69 (0.64, 0.73)
No disability (Q7)	401	0.71 (0.62, 0.81)
Behaviour and emotional (Q7)	401	0.71 (0.61, 0.80)
Hearing (Q7)	401	0.44 (0.35, 0.53)
Manuel dexterity (Q7)	401	0.80 (0.70, 0.89)
Memory / ability concentrate (Q7)	401	0.85 (0.75, 0.95)
Mobility and gross motor (Q7)	401	0.66 (0.57, 0.76)
Perception danger (Q7)	401	0.67 (0.57, 0.76)
Personal, self-care (Q7)	401	0.73 (0.64, 0.83)
Progressive conditions (Q7)	401	0.57 (0.47, 0.67)
Sight (Q7)	401	0.71 (0.61, 0.81)
Speech (Q7)	401	1.00 (#)

Other disability (Q7)	401	0.67 (0.57, 0.77)
Problem therapy offered for (Q9)	784	0.65 (0.62, 0.68)

(#) Unable to calculate confidence interval as no occurrences of speech disability for either measurement

The specific level of agreement varied for the different variables. However, the majority of kappa values were over 0.6, suggesting good agreement between the raters.

Agreement between raters was assessed for the patient diagnoses, with the results summarised in Table 4. Each patient had only one primary diagnosis, so only one analysis was performed for this variable. However, patients could have multiple secondary diagnoses, and so separate analyses was performed for each of these.

*Table 4: Primary and secondary diagnosis*

Diagnosis	n	Kappa (95% CI)
Primary diagnosis	724	0.71 (0.69, 0.74)
<u>Secondary diagnoses</u>		
Bipolar affective disorder	745	0.54 (0.47, 0.61)
Mild depressive episode	745	0.36 (0.29, 0.43)
Moderate depressive episode	745	0.35 (0.28, 0.42)
Sever depressive episode	745	0.24 (0.17, 0.31)
Other depressive episode	745	0.28 (0.21, 0.35)
Recurrent depressive disorder	745	0.51 (0.44, 0.58)
Other mood disorders	745	0.28 (0.22, 0.34)
Agoraphobia	745	0.71 (0.64, 0.78)
Social phobias	745	0.40 (0.33, 0.46)
Specific isolated phobias	745	1.00 (#)
Other phobic anxiety disorder	745	0.57 (0.50, 0.63)
Panic disorder	745	0.70 (0.63, 0.76)
Generalised anxiety disorder	745	0.61 (0.53, 0.68)
Mild anxiety / depression	745	0.49 (0.42, 0.56)
Other anxiety disorders	745	0.52 (0.44, 0.59)
Obsessive-compulsive disorder	745	0.80 (0.73, 0.87)
Post-traumatic stress disorder	745	0.64 (0.57, 0.71)
Other reaction to stress	745	0.49 (0.42, 0.56)
Organic	745	0.44 (0.37, 0.51)
Psychoactive substance use	745	0.56 (0.49, 0.63)
Physiological disturbances	745	0.57 (0.50, 0.64)
Adult personality and behaviour	745	0.53 (0.46, 0.60)
Other diagnosis not listed	745	0.54 (0.46, 0.61)

(#) Unable to calculate confidence interval as no occurrences of specific isolated phobias for either measurement

### b) Appointment dates and therapy attendance

The next set of analyses examined the inter-rater agreement for the appointment dates and attendance to receive different types of therapy

Within this section, only one variable, the date of referral, was measurement a continuous scale. The results, including the ICC value are summarised in Table 5.

*Table 5: Date of referral to therapy*

Question	n	ICC (95% CI)
Date of referral (Q10)	762	0.75 (0.972, 0.78)

The data suggested only moderate agreement between the observers for the date of referral.

Agreement was also assessed for the therapy received. Initially this was assessed for the type of therapy (individual, couples etc.). Patients could receive more than one type of this therapy, and so a separate analysis was performed for each type. Additionally, analyses were performed for whether the patient received each individual type of therapy or not. For these analyses, a single analysis was performed irrespective of the manner in which this was delivered. In other words, the therapy could be individual, group or couples/family. The analysis results are summarised in Table 6.

Table 6: Therapy received

Therapy	n	Kappa (95% CI)
Individual	792	0.81 (0.74, 0.88)
Group	792	0.87 (0.80, 0.94)
Couples / Family	792	0.74 (0.67, 0.81)
Acceptance and commitment	792	0.66 (0.59, 0.73)
Applied relaxation	792	0.53 (0.57, 0.60)
Arts psychotherapies	792	0.86 (0.78, 0.93)
Behaviour activation	792	0.48 (0.41, 0.55)
Behaviour couples therapy	792	0.00 (-0.07, 0.06)
Cognitive analytic therapy	792	0.82 (0.75, 0.89)
Cognitive behaviour therapy	792	0.70 (0.63, 0.77)
Compassion focused therapy	792	0.69 (0.62, 0.76)
Counselling	792	0.53 (0.46, 0.60)
Dialectical behaviour therapy	792	0.68 (0.61, 0.75)
Dynamic interpersonal therapy	792	0.67 (0.60, 0.73)
Eye movement reprocessing	792	0.87 (0.80, 0.94)
Facilitated cognitive therapy	792	0.63 (0.56, 0.70)
Guided / supported self-help	792	0.62 (0.56, 0.69)
Humanistic therapy	792	0.51 (0.44, 0.57)
Integrative psychotherapy	792	0.56 (0.49, 0.63)
Interpersonal psychotherapy	792	0.34 (0.27, 0.41)
Long-term psychotherapy	792	0.60 (0.53, 0.67)
Mentalisation based therapy	792	0.83 (0.76, 0.90)
Mindfulness	792	0.54 (0.47, 0.61)
Mindfulness cognitive therapy	792	0.49 (0.42, 0.56)
Narrative exposure therapy	792	0.89 (0.82, 0.96)
Non-facilitated self-help	792	0.33 (0.27, 0.40)
Problem solving therapy	792	0.55 (0.48, 0.62)
Psycho-education	792	0.51 (0.44, 0.58)
Short-term psychotherapy	792	0.46 (0.40, 0.53)
Signposting	792	0.58 (0.51, 0.64)
Solution focused therapy	792	0.80 (0.73, 0.87)
Structured exercise	792	0.80 (0.73, 0.87)
Support for psychotropic meds	792	0.22 (0.15, 0.29)
Systemic / family therapy	792	0.77 (0.70, 0.84)
Other therapy	792	0.63 (0.56, 0.70)

There was good agreement between raters in the therapy setting (individual, group etc.). The agreement between raters varied for the individual therapy types. For some this was very good (e.g. arts psychotherapies, narrative exposure therapy), whilst this was only poor to moderate for others (e.g. support for psychotropic medication).

### c) Individual therapy

The next analyses focussed on the delivery of individual therapy.

The results for the continuous measures are summarised in Table 7. Due to the positively skewed distribution of the number of sessions, this variable was analysed on the log scale.

*Table 7: Continuous individual therapy variables*

Variable	n	ICC (95% CI)
Date first appointment (Q12)	638	0.48 (0.42, 0.54)
Date of first treatment (Q14)	567	0.83 (0.81, 0.86)
Date of last treatment (Q15)	566	0.12 (0.03, 0.20)
Number of sessions (Q16) (*)	648	0.90 (0.89, 0.92)

(\*) Variable analysed on the log scale

The results suggested good agreement between raters for the number of sessions and the date of the first treatment. However, there was poor agreement for the date of the first appointment and the date of last treatment.

The results for the categorical measures are summarised in Table 8.

*Table 8: Categorical individual therapy variables*

Variable	n	Kappa (95% CI)
Reason for appointment (Q13)	625	0.56 (0.48, 0.63)
Reason for ending therapy (Q17)	624	0.69 (0.65, 0.74)

There was moderate agreement between raters for the reason for the appointment, but good agreement for the reason for ending therapy.

### d) Group therapy

A similar set of analyses were performed relating to group therapy.

The results for the continuous group therapy measures are summarised in Table 9.

Table 9: Continuous group therapy variables

Variable	n	ICC (95% CI)
Date first appointment (Q18)	151	0.75 (0.67, 0.81)
Date of first treatment (Q20)	145	0.81 (0.75, 0.86)
Date of last treatment (Q21)	146	0.86 (0.81, 0.90)
Number of sessions (Q22) (*)	153	0.93 (0.90, 0.95)

(\*) Variable analysed on the log scale

There was very good agreement between raters for the number of sessions, and good agreement for the dates of first and last treatment. Agreement was only moderate for the date of the first appointment.

The results for the categorical group therapy measures are summarised in Table 10.

Table 10: Categorical group therapy variables

Variable	n	Kappa (95% CI)
Reason for appointment (Q19)	152	0.59 (0.47, 0.70)
Reason for ending therapy (Q23)	142	0.70 (0.59, 0.81)

There was good agreement for the reason for ending therapy, but only moderate agreement for the reason for the appointment.

#### e) Family/couples therapy and other outcomes

The final analyses examined the agreement between raters for measures relating to family / couples therapy. The results for the continuous measures are summarised in Table 11.

Table 11: Continuous family/couples therapy variables

Variable	n	ICC (95% CI)
Date first appointment (Q24)	24	0.72 (0.46, 0.87)
Date of first treatment (Q26)	23	0.82 (0.62, 0.92)
Date of last treatment (Q27)	23	0.98 (0.95, 0.99)
Number of sessions (Q28) (*)	24	0.90 (0.78, 0.95)

(\*) Variable analysed on the log scale



There was very good agreement for the number of sessions and also the date of the last treatment. Agreement was not quite as strong for the date of the first appointment.

The results for the categorical therapy measures, along for those regarding the measurement of other outcomes are summarised in Table 12.

*Table 12: Categorical family/couples therapy variables & other outcomes*

Variable	n	Kappa (95% CI)
Reason for appointment (Q25)	23	0.58 (0.27, 0.88)
Reason for ending therapy (Q29)	21	0.60 (0.36, 0.83)
Other outcomes measured (Q30)	784	0.76 (0.69, 0.83)

There was moderate to good agreement for the reason for the appointment and for ending therapy, with good agreement between raters for whether any other outcomes were measured.