

Improving Diagnostic Reliability in Chinese Medicine

Michael C. Popplewell

This thesis is submitted for the partial requirement for

the degree

of Doctor of Philosophy (Science)

Faculty of Science

University of Technology Sydney

December 2015

CERTIFICATE OF AUTHORSHIP/ORIGINALITY

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree except as fully acknowledged within the text.

I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

.....

Signature of Candidate

Abstract

The aim of this thesis was to assess current levels of inter-rater diagnostic agreement in the Chinese Medical (CM) profession and to propose strategies that might improve these levels.

Researchers have generally used inappropriate statistical constructs to evaluate inter-rater agreement. A more appropriate weighted chance-removed statistic is employed to determine inter-rater diagnostic agreement with ordinal data. Further, the largest number of raters which have been used in any past study was three. Similarly, no study was located which involved inter-rater diagnostic agreement with subjects drawn from an open population. This is a deficiency in understanding CM inter-rater agreement in a clinical setting.

The Diagnostic System of Oriental Medicine (DSOM) format was identified as suitable for use in CM diagnosis by practitioners. This format also enables appropriate statistics to be employed. An experiment was performed in which five experienced practitioners of CM diagnosed 42 subjects using the DSOM as the diagnostic format. Each of the sixteen diagnostic descriptors used to describe a diagnosis with the DSOM were scored 0-5. Substantial chance-removed weighted agreement of 0.60 ± 0.02 was found. The descriptors of DSOM format were edited after examining 60,000 clinical records at the UTS CM outpatient clinic to arrive at the Chinese Medicine Diagnostic Descriptor

format, the (CMDD). Conventional CM diagnostic formats can be directly mapped to CMDD, thereby making this system as subtle as conventional systems.

A second experiment was performed to evaluate inter-rater agreement with CMDD and contemporary CM diagnostic formats respectively. Groups of CM practitioners, one group utilising the CMDD and the other, the CM diagnostic formats, diagnosed 35 subjects over two days. Each of the fifteen CMDD diagnostic descriptors was scored 0-5, while three selected CM patterns were scored 1-5. The subjects were again drawn from an open population. A weighted simple agreement of only 19% was found between practitioners who employed the CM format. This is not an appropriate foundation for application or assessment of treatment. Further, chance-removed statistics or error estimates cannot be evaluated when the CM format is used with unrestricted diagnostic possibilities.

The possibility that bias was present in raters' scores was also investigated. No significant bias was present in the raters' scores. This should be used as a guide for the adoption of appropriate rater training to improve agreements. Guiding questionnaires for each descriptor whilst utilising the CMDD format, would also appear to hold potential to further improve agreement. The CMDD seems to clearly facilitate superior inter-rater agreement compared with the CM format.

The raters using the CMDD format achieved substantial chance-removed agreements of 0.67 ± 0.03 on both days. Mapping diagnoses made by raters

in the CM to the CMDD format enabled chance-removed inter-rater agreements of 0.65 ± 0.03 on day one and 0.73 ± 0.03 on day two to be calculated, significantly larger than when using the CM format. This suggests that the structure of the CMDD allows the correct inter-rater agreement to be calculated, something very difficult to achieve with the contemporary CM format. It is therefore suggested that the CMDD format be used in contemporary clinical and research settings and is also proposed that it be incorporated into the internationally recognised CONSORT and STRICTA research guidelines

Acknowledgements

I would like to take this opportunity to thank all the people who supported me, especially those who gave me support for this project.

As far as mentors go, I will always be indebted to the late Doctor Phillip Groves, who has had the greatest of influences on me. I was extremely fortunate to meet this incredible man. He showed through his own life that one could make a difference and achieve an incredible amount in a lifetime and I am humbled to have made such minor efforts in comparison to this great man.

I would like to give Associate Professor Chris Zaslowski my sincere thanks for his patience, support and understanding. The project unfolded and often changed directions in ways that all involved would not have ever predicted.

I am extremely grateful to Professor Inseon Lee for providing and processing the Diagnostic System of Oriental Medicine questionnaires, without which this work would have been much more difficult.

I would also like to thank Associate Professor Peter Meier for his generous support in supplying data from the UTS outpatient clinic, which greatly assisted this project.

I would above all like to thank Professor John Reizes, without whose friendship, input, guidance and incredible support this project would not have been possible. John is more like a father than a supervisor to me and it is a privilege to work with and be educated by this wonderful, generous man.

Table of Contents

Abstract	i
Acknowledgements	iv
List of Figures	xiii
List of Tables	xiv
Nomenclature	xix
Glossary and Definition of Terms	xxii
Chapter 1 Introduction	1
1.2 Statement of problem	2
1.3 Aims and objectives	5
1.4 Contents of thesis.	6
Chapter 2 Literature Review	9
2.1 The diagnostic procedure in CM	9
2.2 The Definition of Agreement	10
2.2.1 Simple Agreement	12
2.2.2 Kappa Statistics	14
2.2.3 The AC1 Statistic	19
2.2.4 Weighted Statistics	20
2.2.5 Types of Weighting for Ordinal Data	21
2.2.6 Interpretion of Agreement with Agreement by Chance Removed	25
2.2.7 Standard Error and Confidence Intervals	26

2.2.8 Determining differences between two chance-removed inter-rater agreement statistic values	28
2.3 Chinese Medicine Literature on inter-rater agreement	29
2.3.1 Chinese Medical Diagnostic Agreement reported with Simple Agreement and Kappa statistics	30
2.3.1.1 Comments on papers that reported agreement with average simple agreement and Fleiss' Kappa	38
2.3.1.2 Generalised observations regarding Fleiss' Kappa and average Simple Agreement studies	42
2.3.2 Chinese Medical inter-rater agreement evaluated with the AC1 statistic	45
2.3.3 Weighted Statistics	55
2.4 Validation exercise using the AC2 software	59
2.5 Strategies that lead to higher inter-rater agreement	64
2.5.1 Rater Training	64
2.5.2 Simplifying a diagnostic system	64
2.5.3 Questionnaires	66
2.6 The Diagnostic System of Oriental Medicine	70
2.7 Summary of Literature review	75
Chapter 3 Investigation of DSOM Diagnostic Reliability	80
3.1 Data Collection Details	80
3.3.1 Setting	82
3.3.2 Subjects	82
3.3.3 Practitioners	83

3.3.4 Data Processing to Determine Agreement	83
Chapter 4 The DSOM Data Collection	87
4.1 Data Recorded by the Practitioners	87
4.2 Simple and AC2 Agreement	89
4.3 Total Patient Pathogenic Score and Agreement	92
4.4 Practitioner Agreement in the Three Wellness Groups	94
4.5 Fleiss' Kappa revisited	95
4.6 Agreement in the Individual Descriptors	99
4.7 Individual Descriptor Agreement and Total Descriptor Score	105
4.8 DSOM Questionnaire Agreement with the Practitioner Diagnoses	111
4.9 Discussion of results of DSOM data collection	114
Chapter 5 UTS Outpatient Clinic Data	118
5.1 UTS Student Clinic Data Mapped to the DSOM Descriptors	123
5.2 DSOM Study and UTS Outpatient Clinic Data Compared	125
Chapter 6 The Chinese Medicine Diagnostic Descriptor	128
6.1 Mapping CM Diagnoses to the CMDD	134
6.2 CMDD Scoring Example	137
6.3 CMDD Agreement Calculation	139
Chapter 7 CMDD and CM Diagnostic Agreement Investigation	142
7.1 CMDD and CM Diagnostic Data Collection Details	142
7.1.1 Subjects	142
7.1.2 Practitioners	143

7.1.3 Location	144
7.2 The CM and CMDD forms used in the data collection	144
7.3 Agreement Calculation in Chinese Medicine and the Chinese Medicine Diagnostic Descriptor Formats	146
Chapter 8 Chinese Medicine Format Agreement	148
8.1 Initial Observations	148
8.1.1 CM Patterns Chosen by Practitioners	148
8.1.2 Data Collected Compared with UTS Clinic Data	148
8.2 Agreement calculation in the CM diagnostic format	155
8.3 Chance-removed Agreement calculation and the CM diagnostic format.	157
8.4 Agreement Results with the CM diagnostic format	157
8.5 Agreement with the CM diagnostic format Conclusion	162
Chapter 9 Chinese Medicine Diagnostic Descriptor Agreement	164
9.1. Data Recorded by the Practitioners in the CMDD study	164
9.2 CMDD and UTS Student Clinic Data Compared	166
9.3 CMDD Agreement Results	168
9.4 CM diagnoses mapped to the CMDD Format	170
9.5 CMDD Agreement in the Three Wellness Groups	171
9.6 Agreement in the CMDD's Individual Descriptors	172
9.7 CMDD Descriptor Agreement and TPS	174
9.8 Discussion of the results of the CMDD and CM data collections	179

Chapter 10 Normalisation of data: attempt at partial removal of practitioner bias	184
10.1 Utilisation of the DSOM data instead of the CMDD data	186
10.2 Bias Measurement Theory	187
10.3 Score Bias and Descriptor Bias Approaches to Normalisation	188
10.4 Score Bias Normalisation by “Trip factor”	191
10.4.1 Calculation of Score Bias Normalisation by Trip Factor	191
10.4.2 Implementation of Score Bias Normalisation by Trip Factor	192
10.4.3 Score Changes after Score Normalisation by Trip Factor	196
10.4.4 Inter-Rater Agreement with Score Bias Normalisation by Trip Factor	199
10.5 Score Bias Normalisation by a “Score Factor”	200
10.5.1 Score Bias Normalisation by Score Value Calculation Equation	201
10.5.2 Score Bias Normalisation by Score Value Calculation	202
10.5.3 Score changes after Score Normalisation by Score Factor approach	204
10.5.4 Inter-Rater Agreement after Score Factor Normalisation	206
10.6 Descriptor Bias Normalisation by “Trip Factor”	207
10.6.1 Descriptor Bias Normalisation by Trip Factor Calculation Equation	208
10.6.2 Descriptor Bias Normalisation by Trip Factor Calculation	209
10.6.3 Scores and Score changes after Descriptor Normalisation by the Trip Factor approach	211
10.6.4 Results of Descriptor Bias Normalisation by Trip Factor	215
10.7 Descriptor Bias Normalisation by Score Factor	219

10.7.1 Equation for Descriptor Bias Normalisation by Score Factor Calculation	219
10.7.2 Values calculated for Descriptor Bias Normalisation by Score Factors	220
10.7.3 Score Changes after Descriptor Normalisation by “Score Factor”	222
10.7.4 Agreements after Descriptor Normalisation by Score Factor	223
10.8 Normalisation attempts conclusions	228
Chapter 11 Conclusion	229
References	240
Appendices	247
<i>Appendix 1 The DSOM Questionnaire</i>	247
<i>Appendix 2 Subject Information Statement for participants of the DSOM inter-rater study</i>	263
<i>Appendix 3 DSOM Subject Consent Form</i>	267
<i>Appendix 4 Diagnostic form for recording the DSOM diagnosis</i>	269
<i>Appendix 5 Subject Information Statement for Participants of the CMDD study</i>	272
<i>Appendix 6 Consent Form CMDD and CM study</i>	275
<i>Appendix 7 Form for recording the Contemporary Chinese Medical Diagnosis</i>	277
<i>CM Diagnosis form</i>	277
<i>Appendix 8 Form for recording the CMDD diagnosis</i>	279
<i>Appendix 9a. Score Normalisation by Trip Factor Calculation Table, day one</i>	280
<i>Appendix 9b. Score Normalisation by Trip Factor Calculation Table, day two</i>	280
<i>Appendix 9c. Score Normalisation by Trip Factor Calculation Table, day three</i>	281
<i>Appendix 10.a Rater-Descriptor combination scores after Descriptor Normalisation by Trip Factor application, day one</i>	281
<i>Appendix 10.b Rater-Descriptor combination scores after Descriptor Normalisation by Trip Factor application, day two</i>	282

Appendix 10.c Rater-Descriptor combination scores after Descriptor Normalisation by Trip Factor application, day three	282
Appendix 11a. Differences of Descriptor-rater totals from the mean in Raw and Descriptor Normalised by Trip factor method data, day one	283
Appendix 11b. Differences from the mean of Descriptor-rater totals in Raw and Descriptor Normalised by Trip factor method data, day two	284
Appendix 11c. Differences from the mean of Descriptor-rater totals in Raw and Descriptor Normalised by Trip factor method data, day three	285
Appendix 12a. Non-zero scores of each rater-Descriptor combination for calculation of Descriptor Normalisation by Score Factors, day one	285
Appendix 12b. Non-zero scores of each rater-Descriptor combination for calculation of Descriptor Normalisation by Score Factors, day two	286
Appendix 12c Non-zero scores of each rater-Descriptor combination for calculation of Descriptor Normalisation by Score Factors, day three	286
Appendix 13.a Score differences of each rater-Descriptor group from the mean after Descriptor normalisation by Score Factor, day one	287
Appendix 13.b Score differences of each rater-Descriptor group from the mean after Descriptor normalisation by Score Factor, day two	288
Appendix 13.c Score differences of each rater-Descriptor group from the mean after Descriptor normalisation by Score Factor, day three	288
Appendix 14a Descriptor Normalisation Score Factors raw and rounded and capped, day one	289
Appendix 14b Descriptor Normalisation Score Factors raw and rounded and capped, day two	290
Appendix 14c. Descriptor Normalisation Score Factors raw and rounded and capped, day three	290

Appendix 15.a <i>Descriptor Normalising Factors applied in the two methods used day one</i>	291
Appendix 15.b <i>Descriptor Normalising Factors applied in the two methods used day two</i>	292
Appendix 15.c <i>Descriptor Normalising Factors applied in the two methods used day three</i>	292
Appendix 16.a <i>Differences from the mean of Descriptor-rater totals in Raw and Descriptor Normalised by Score Factor data, day one</i>	293
Appendix 16.b <i>Differences from the mean of Descriptor-rater totals in Raw and Descriptor Normalised by Score Factor data, day two</i>	294
Appendix 16.c <i>Differences from the mean of Descriptor-rater totals in Raw and Descriptor Normalised by Score Factor data, day three</i>	294

List of Figures

Figure 1.2.1 Chinese Medicine treatment effectiveness flowchart.....	4
Figure 2.2.5.1 Graph of weightings applied as a function of score difference for (a) Quadratic, (b) Linear and (c) Radical weightings	21
Figure 2.3.2.1 Simple agreement and AC1 statistics from table 2.3.2.1 sorted according to simple agreement	51
Figure 2.3.2.2: Graph of β statistic against simple agreement for various values of agreement by chance.....	53
Figure 2.3.2.3: Graph of the difference between simple agreement and β as a function of simple agreement	54
Figure 2.3.2.4: Graph of the ratio of agreement with chance removed to simple agreement against Simple agreement	54
Figure 4.5.1 Average agreement of three wellness groups using Linearly weighted Gwet's AC2 and Fleiss' Kappa.....	97
Figure 6.0.1 The Chinese Medicine Diagnostic Descriptor	129
Figure 6.2.1 CMDD pattern mapping example.....	139

List of Tables

Table 2.2.5.1 Agreement weights applied with radical weighting distribution as a function of score difference	23
Table 2.2.6.1 Scale for discussing agreements proposed by Landis and Koch	26
Table 2.3.1.1 Chinese Medical diagnostic agreement studies reporting Simple agreement and Kappa values	32
Table 2.3.1.2.1 Number of options made available to raters by researchers using Fleiss' Kappa and simple agreement.....	42
Table 2.3.2.1 Papers reporting percentage, Kappa Number and AC1 agreement summarised	46
Table 2.3.3.1 Weighted simple agreement and Kappa	57
Table 2.4.1 Scores for AC2 exercise	60
Table 2.4.2 Results with un-weighted simple agreement and Gwet's AC1 statistics.....	60
Table 2.4.3 Weighting matrix for linearly distributed weights calculate	61
weighted agreement between two raters	61
Table 2.4.5 Weighted and Un-weighted agreement in each variable	62
Table 2.4.5 Results with linearly weighted simple agreement and Gwet's AC2 statistics....	63
Table 2.6.1.1 Question categories and number of questions in the DSOM questionnaire ...	72
Table 3.3.4.1 Data processing format for the calculation of agreements using the DSOMf.	84
Table 4.1.1 Score selections of the practitioners with the DSOM data set.	87
Table 4.1.2 Descriptor Score selections	88
Table 4.2.1 DSOM agreement using all 16 descriptors.....	90
Table 4.3.1 Average Total Practitioner Score and percentage scored one or above for each Descriptor.....	93
Table 4.4.1 Agreements in each Wellness group.....	94
Table 4.5.1 Score selections by the practitioners in the three wellness groups.....	95
Table 4.5.2 Linearly weighted Fleiss' Kappa and AC2 agreement and Differences in the three Wellness Groups.....	96

Table 4.6.1 Individual Descriptor Agreement expressed in linearly weighted percentage and AC2, and the Average TPS and frequency of occasions the descriptor was scored one or greater.....	100
Table 4.6.2 (a) Day One Descriptor Raw Scores and Average Scores.....	102
Table 4.6.2.1 (b) Day Two Descriptor Raw Scores and Average Scores.....	102
Table 4.7.1 Number of subjects that were included in the intra-descriptor subgroups....	107
Table 4.7.2 Linearly weighted simple agreement in the intra-descriptor subgroups.....	108
Table 4.7.3 Linearly weighted AC2 agreement in the intra-descriptor subgroups	109
Table 4.8.1 Agreements in each Wellness group	112
Table 4.8.2 Agreements between the five practitioners with DSOM questionnaire data .	112
Table 4.8.3 Changes in agreements between the five practitioners after the inclusion of DSOM questionnaire data.	113
Table 5.0.1 Percentage of total diagnoses according to order of choice.....	119
Table 5.0.2 Practitioners top 56 Diagnostic Pattern selections as a percentage of all Diagnostic Patterns	119
Table 5.0.3 The top 56 patterns selected at UTS outpatient clinic.....	120
Table 5.1.1 Mapping of Descriptors from the most popular 56 patterns selected at UTS CM Outpatient Clinic	124
Table 5.2.1 Descriptor choices in the DSOM and UTS outpatient clinic data	125
Table 6.1.1 CM diagnoses.....	135
Table 6.1.2 CM mapped to CMDD - descriptors selected by raters in <i>bold italics</i>	136
Table 6.2.1 CM to CMDD mapping example	138
Table 6.3.1 Agreement Calculation Data.....	140
Table 8.1.2.1 Comparison between the diagnoses selected at UTS outpatient clinic and the present study.....	149
Table 8.2.1 Linear Weights Utilised.....	155
Table 8.4.1 Chinese Medical format Agreements	158
Table 8.4.2 Patterns that attracted matches and the number of selections.....	161
Table 9.1.1 Scores selected in the CMDD data set	164
Table 9.1.6 Differences in percentage selection in the DSOM and CMDD data sets.....	165

Table 9.1.3 Scores selections by all practitioners in each Descriptor	165
Table 9.2.1 Descriptor Rankings with the CMDD data and the UTS outpatient clinic	167
Table 9.3.1 Linearly weighted agreements calculated in the CM and CMDD formats.....	169
Table 9.5.1 Agreement in wellness groups CMDD data collection day one	171
Table 9.5.2 Agreement in wellness groups CMDD data collection day two	171
Table 9.6.1a Individual Descriptor Agreement expressed in linearly weighted percentage and AC2, and the Average TPS and % of occasions the Descriptor was scored one or greater in day 1 data	172
Table 9.6.1b Individual Descriptor Agreement expressed in linearly weighted percentage and AC2, and the Average TPS and % of occasions the Descriptor was scored one or greater in day 2 data	173
Table 9.7.1a Intra-descriptor wellness groups Simple agreement day one	175
Table 9.7.1b Intra-descriptor wellness groups Simple agreement day two.....	176
Table 9.7.2a Intra-descriptor wellness groups AC2 agreement day one	176
Table 9.7.2b Intra-descriptor wellness groups AC2 agreement day two	177
Table 10.4.2.1 Sum of Raw scores for each Rater on each day	193
Table 10.4.2.2 Normalising Factors of the raters on each day.....	193
Table 10.4.2.3 Trip Factor Normalisation Values applied to the Raw Data on each day (these are not normalising factors)	195
Table 10.4.3.1 Total from average on each day before and after Score Normalisation by Trip Factor approach implementation	197
Table 10.4.3.2 Total absolute differences from the mean in the raw and Score Normalised by Trip value data	198
Table 10.4.4.1 Linearly Weighted simple and AC2 agreement in the Trip Factor Normalised data.....	199
Table 10.4.4.2 Differences in linearly weighted and AC2 agreements between the raw DSOM and Trip Factor normalised data	199
Table 10.5.2.1a Score Normalisation by Score Value calculation, day one	202
Table 10.5.2.1b Score Normalisation by Score Value calculation, day two	203
Table 10.5.2.1c Score Normalisation by Score Value calculation, day three.....	203

Table 10.5.2.2 Trip and Score Factors for Score Normalisation for each day and each rater	204
Table 10.5.3.1 Score Differences from average on each day before and after Score Normalisation by Score Factor approach implementation	205
Table 10.5.3.2 Total absolute differences from the mean in the raw and Score Normalised by Score value data	206
Table 10.5.4.1 Linearly weighted simple and AC2 agreement after Score Normalisation by Score Factor in the three Wellness Groups and All Groups	206
Table 10.5.3.2 Differences between Score Normalised by Score Factor and Raw data agreement	207
Table 10.6.3.1 Total absolute differences from the mean in raw and normalised data on each day	214
Table 10.6.3.1a Linearly Weighted Simple Agreement after Descriptor Normalisation by Trip Factor approach	215
Table 10.6.3.1b Linearly Weighted AC2 Agreement after Descriptor Normalisation by Trip Factor approach	216
Table 10.6.3.2a Changes in Linearly Weighted Simple Agreement and standard errors after Descriptor Normalisation by Trip Factor approach	217
Table 10.6.3.2b Changes in Linearly Weighted AC2 Agreement after Descriptor Normalisation by Trip Factor approach	218
Table 10.7.2.1 Summary of differences between trip factor and score factor normalisation values applied to implement Descriptor Normalisation	222
Table 10.7.2.2 Non-normalisations that occurred each day and overall in the Trip and Score Factor approaches	222
Table 10.7.3.1 Absolute total score differences from the mean of the raw and Descriptor normalised by Score Factor data	223
Table 10.7.4.1 Linearly Weighted Simple Agreement after Score Factor Normalisation Implementation	224
Table 10.7.4.2 Linearly Weighted AC2 Agreement after Score Factor Normalisation Implementation	225

Table 10.7.4.3 Changes in Linearly Weighted Simple Agreement after Score Factor	
Normalisation Implementation.....	226
Table 10.7.4.4 Changes in Linearly Weighted AC2 Agreement after Score Factor	
Normalisation Implementation.....	227

Nomenclature

A_0	Number of agreements between two raters observed in an experiment (Equation (2.1));
A_p	Number of possible agreements between two raters (Equation (2.1));
CI	Confidence interval (equation (2.11))
d	Score difference (Equation (2/8))
F	Normalising factor (Equations (10.4, 10.8))
F'	Rounded normalising score (Equation (10.9))
J	Number of subjects (Equation (10.2))
K	Number of practitioners (equation (10.3))
k	Number of correct answers
L	Number of categories for assessment; (Equation (2.4))
m	m^{th} answer to multiple answer question
M	Number of non-zero scores (Equation (10.8))
N	Number of raters per subject (Equation 2.4))
n	Number of possible ratings (Equation (2.4))
n_o	Number of observations (equation (2.10))
\bar{P}	Average simple agreement for more than two raters (equation (2.4))
\bar{P}_e	Average agreement by chance for more than two raters (Equation (2.5))
P_0	Simple agreement between two raters (Equation (2.1))

p_e	Agreement by chance for two raters (Equation (2.1))
SE	Standard error (equation (2.10))
s	Score (equation (10.1))
T	Total score (Equation (10.6))
W	Weighting
X	Difference between average simple agreement and average agreement by chance (Equation (2.15))
Y	Difference between perfect agreement and agreement by chance (Equation (2.16))

Greek Symbols

α	Ratio of β to the average simple agreement (Equation (2.17))
β	Generic definition of agreement with agreement by chance removed (Equation (2.14))
Δ	Difference between average simple agreement and the κ_{Fle} statistic (Equation (2.12))
Δ_1	Difference between AC1 and κ_{Fle} statistics (Equation 2.13))
κ	$\kappa = \frac{P_0 - p_e}{1 - p_e}$ for two raters or $\kappa = \frac{\bar{P} - \bar{p}_e}{1 - \bar{p}_e}$ for more than two raters

Superscript

–	Mean value
---	------------

Subscripts

<i>Coh</i>	Cohen
<i>Fle</i>	Fleiss
<i>i, j, k</i>	Indices
<i>l</i>	Linear
<i>q</i>	Quadratic

Glossary and Definition of Terms

Chance-removed agreement (Chapter 2, p. 38): inter-rater agreement from which chance has been removed. For instance if two people were attempting to predict the outcome of a coin toss, there would be a 50% chance that they agreed. The removal of chance agreement is aimed at estimating the 'true agreement' between raters that is not inflated by the presence of chance agreement.

Chinese Medicine (CM): the contemporary style of classic Chinese medicine that was developed in China is practiced in China and Australia. There are other styles of Chinese medicine; Japanese and Korean acupuncture are two that are used by significant numbers of practitioners.

Diagnostic Agreement: diagnostic agreement between practitioners. Diagnostic agreement relies upon the exactly the same terms being used in each practitioner's diagnosis.

Descriptors (Chapter 6 p. 129): are defined as the terms that have specific Chinese medical meaning and form part of the nomenclature of Chinese diagnosis. The Descriptors are the key CM diagnostic attributes used to define Chinese Medical Diagnoses in the DSOM and CMDD. The Descriptors utilised within the CMDD are referenced to the World Health Organisation's International Standard Terminologies of Traditional Medicine in the Western Pacific Region.

DSOM (Chapter 2 p. 70): the Diagnostic System of Oriental Medicine developed by Inseon Lee of South Korea.

DSOMf (Chapter 2 p. 74): The format of the DSOM used for the presentation of diagnostic results of a subject or patient.

DSOM Questionnaire (Chapter 2 p. 72): the diagnostic questionnaire filled in by patients or subjects and used to determine the Descriptors scores of the DSOMf and other data.

Intra-Descriptor Wellness Groups (Chapter 4 p. 104) are defined as groups of subjects allocated to a wellness cohort within a Descriptor according to the scoring characterized by the Total Descriptor Score of the diagnosing raters for that Descriptor.

Simple Agreement (Chapter 2 p 38) is defined as agreement calculated as the number of agreements divided by the number of agreements possible. It is the basis upon all other more complex agreement calculations are made.

Total Descriptor Score (TDS) (Chapter 2 p 38) is the total score allocated by all practitioners to a Descriptor for a particular subject. TDS is used to form the Intra-Descriptor Wellness Groups.

Total Patient Pathogenic Score (TPS) (Chapter 4 p. 116) is defined as the total of all scores allocated to a subject by the diagnosing raters. TPS are used to define Wellness groups.

Wellness Groups are defined as groups of subjects included according to the characteristics of their TPS.