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 Zaslawski 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 priveledge to work with and be educated by this wonderful, generous man.

Table of Contents

List of Figures

List of Tables

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))
$\boldsymbol{F}$	Normalising factor (Equations (10.4, 10.8))
$\boldsymbol{F'}$	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))
$\bar{\boldsymbol{P}}$	Average simple agreement for more than two raters (equation (2.4))
$\bar{\boldsymbol{P}}_{\boldsymbol{e}}$	Average agreement by chance for more than two raters (Equation (2.5))
$\boldsymbol{p_0}$	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

$\bar{}$ Mean value

Subscripts

Coh	Cohen
Fle	Fleiss
i, j, k	Indices
l	Linear
q	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.