DETERMINING THE CONTENT VALIDITY AND RELIABILITY OF TWED MATRIX AS A COGNITIVE DEBIASING STRATEGY IN CLINICAL DECISION MAKING IN EMERGENCY DEPARTMENT

By

DR. MOHD SHUKRI BIN MAT SAAD

A DISSERTATION SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF MEDICINE (EMERGENCY MEDICINE)

UNIVERSITI SAINS MALAYSIA

MAY 2017

AKU JANJI

Diperakui bahawa disertasi yang bertajuk DETERMINING THE CONTENT VALIDITY AND RELIABILITY OF TWED MATRIX AS A COGNITIVE DEBIASING STRATEGY IN CLINICAL DECISION MAKING IN EMERGENCY DEPARTMENT merupakan kerja dan penyelidikan yang asli dari MOHD SHUKRI BIN MAT SAAD, No kad pengenalan : 840511-04-5035, No Matrik PUM-0183/13, dari tempoh 2013 hingga 2017 adalah di bawah penyeliaan kami. Disertasi ini merupakan sebahagian daripada syarat untuk penganugerahan Sarjana Perubatan Kecemasan, segala hasil penyelidikan dan data yang diperolehi adalah hak milik terpelihara Universiti Sains Malaysia.

_____

Tandatangan Pelajar

_____

Tandatangan Penyelia Bersama

Pensyarah Pusat Pengajian Sains Perubatan

Universiti Sains Malaysia

_____

Tandatangan Penyelia Utama

Pensyarah Jabatan Kecemasan

Pensyarah Pusat Pengajian Sains Perubatan

Universiti Sains Malaysia

_____

Tandatangan Ketua Jabatan

Jabatan Kecemasan

Universiti Sains Malaysia

ACKNOWLEDGEMENT

I wish to express my sincere gratitude to the following individuals who had given their advice, guidance and expertise to make it possible for me to complete this research project:

1. Dr Andey Abd Rahman
   Main Supervisor,
   Lecturer, Emergency Department,
   School of Medical Science, Universiti Sains Malaysia

2. Dr Muhamad Saiful Bahri Bin Yusoff,
   Co-supervisor,
   Lecturer, Department of Medical Education,
   School of Medical Science, Universiti Sains Malaysia

3. Prof. Chew Keng Sheng
   Lecturer Medical Faculty and Health Science,
   Universiti Malaysia Sarawak.

4. All lecturers in the Emergency Department, School of Medical Sciences, Health Campus, Universiti Sains Malaysia.

5. My supportive classmates in the Master of Medicine (Emergency) programme 2013/17.

6. My caring wife, Bahirah Ahmad, my beloved son, Muhammad Amsyar and daughter, Nur Aisyah, my understanding parents and parents-in-law for their support and strength.

Table of Contents

Abstrak

Menentukan Kesahihan Kandungan dan Kebolehpercayaan TWED Matrix Sebagai Strategi Mengelakkan Keberatsebelahan Kognitif Dalam Membuat Keputusan Klinikal Di Jabatan Kecemasan.

Latar Belakang: Keberatsebelahan kognitif seringkali merumitkan keputusan klinikal yang dibuat di jabatan kecemasan. TWED Matrix telah dicipta sebagai satu alat menyahpincang bagi mengurangkan kesalahan diagnosis yang disebabkan oleh keberatsebelahan kognitif. Ianya dalam bentuk mnemonic untuk menggalakkan metakognisi oleh doktor untuk mengurangkan kesalahan membuat keputuan klinikal di jabatan kecemasan. Oleh kerana ianya satu tatacara yang baru, maka satu kajian telah dibuat bagi menentukan kesahihan kandungan dan kebolehpercayaan TWED Matrix bagi mengelakkan keberatsebelahan kognitif dalam membuat keputusan klinikal di jabatan kecemasan.

Metodologi: Kajian ini telah dibahagikan kepada dua peringkat; peringkat pembangunan dan peringkat penghakiman. Dalam peringkat pembangunan, 50 jenis keberatsebelahan kognitif telah dikhususkan kepada enam kelas, dan empat soalan utama merangkumi lapan item ditanya. Dalam peringkat penghakiman pula, pakar kecemasan sebagai juri pakar telah dipilih untuk menilai lapan item dalam bentuk skala Likert untuk menentukan keberkaitan dan kawakilannya. Bagi keberkaitan, indeks pengesahan kandungan (CVI) dan modifikasi Kappa dinilai dan bagi kewakilan, indeks pengesahan kandungan (CVI) dinilai. Bagi kebolehpercayaan, Cronbach Alpha digunakan untuk penilaian.

Keputusan: Bagi keberkaitan, semua item mempunyai nilai CVI melebihi 0.78. Purata CVI untuk keberkaitan adalah 0.89. Modifikasi Kappa telah dikira bagi membuat pelarasan bagi kebarangkalian perjanjian peluang. Bagi kewakilan pula, semua item mempunyai nilai CVI melebihi 0.78, iaitu nilai minimum yang diterima bagi CVI. Purata CVI adalah 0.94. Nilai Alpha bagi keberkaitan adalah 0.767 iaitu dalam nilai yang diterima. Kebolehpercayaan bagi kewakilan menghasilkan nilai Cronbach Alpha 0.737, nilai yang masih diterima melebihi 0.7.

Konklusi: TWED Matrix telah menunjukkan nilai yang signifikan bagi keberkaitan dan kebolehpercayaan walaupun kajian berterusan diperlukan bagi memantapkan lagi kuasa kajian ini sebelum ia boleh digunakan dengan selamat bagi mengelakkan keberatsebelahan kognitif dalam membuat keputusan klinikal di jatan kecemasan.

Kata kunci: TWED Matrix, strategi mengelakkan keberatsebelahan kognitif, kesahihan kandungan, kebolehpercayaan.

Abstract

Determining the Content Validity and Reliability of TWED Matrix as a Cognitive Debiasing Strategy in Clinical Decision Making In Emergency Department

Background: Cognitive biases always complicate clinical decision making in emergency department. TWED Matrix was invented as a tool for debiasing strategy to reduce diagnostic error caused by cognitive biases. It is in the form of mnemonic to encourage metacognition by doctors to reduce errors in clinical decision making in emergency department. As it is relatively new tool, a study was conducted to determine the content validity and reliability of TWED Matrix as cognitive debiasing strategy in clinical decision making in emergency medicine.

Methodology: This study was divided into two stages; the developmental stage and the judgment stage. In the developmental stage, we narrowed down 50 cognitive biases into six classes, and total of four vital questions covering eight items were asked. Then, in the judgment stage, emergency physicians as expert judges were chosen to evaluate eight items in the form of Likert scale for their relevance and representativeness. For relevance, CVI and modified Kappa were assessed and for representativeness, CVI was calculated. Reliability was tested using Cronbach Alpha.

Results: For relevancy, all of the items were scored of CVI more than 0.78. CVI average for relevancy was 0.89. Modified Kappa statistic was calculated to make adjustment for possibilities of chance agreement. For representativeness, all eight items produced CVI of more than 0.78 which is the minimum value for CVI acceptance. CVI

average was 0.94. Alpha value for item relevancy was 0.767 which is within acceptable value. Reliability on item representativeness produced Cronbach's Alpha of 0.737 which is still above acceptance value of 0.7.

Conclusion: TWED Matrix showed significant validity and reliability in this study although further evaluation and assessment needed to strengthen the power of the study before it could safely and widely use as a cognitive debiasing strategy in clinical decision making in emergency department.

Keywords: TWED Matrix, cognitive debiasing strategy, content validity, reliability

INTRODUCTION

Introduction and Literature Review

Making a clinical decision is part and parcel of a health practitioner's job. And making it right is vital as patients' lives depend on it. When patients arrive at the hospital, most of them will present to the emergency department first before receiving treatment and being admitted to specific wards according to specialties. Thus emergency medicine physicians have a huge responsibility to make sure that they do not miss any life threatening condition for each and every one of the patient. But working in emergency environment is not the same as working in calm and controlled environment. The unique nature of emergency medicine compared to other disciplines makes it prone for medical errors, thus compromising patient's safety(Laxmisan *et al.*, 2007). There are multiple factors that contribute either directly or indirectly towards medical errors in emergency medicine(Croskerry and Sinclair, 2001). The challenges face by emergency physicians because of their unique nature of work requires them to see all types of cases that are presented at anytime of the day by any patient. Unpredictable surge of patient load regardless of whether weekdays or weekends or during any communicable disease outbreaks is also an important factor that may lead to medical errors. Shift system which applies to almost all emergency departments healthcare providers is also one of the contributing factors which makes medical errors more likely to occur when they are working under stressful, tiring and undesirable condition.

Most patients who make their first visit to emergency department are incapable of giving proper medical history and sometimes even their particulars to their primary healthcare providers because of the severity of their illness. Despite having minimal

information, emergency healthcare providers still need to give their best point of care towards patients.

There are many types of medical errors, but one of the most important errors is cognitive error(Croskerry, 2005). Cognitive error means error in the process of thinking. In emergency medicine, any error in the process of thinking towards making the right diagnosis or decision may lead to catastrophic outcome. Patients may end up with morbidity or mortality, emergency physician may even be sued by the patients (Studdert *et al.*, 2006). Allegations of substandard diagnostic care in emergency department was as much as 28% (Kachalia *et al.*, 2007). Diagnostic errors in emergency medicine ranged between 1%- 12% and 95 % of it were from cognitive errors. Besides cognitive error, lack of knowledge and skills, misinterpretation of lab results, junior healthcare provider and multitasking also might contribute to medical and diagnostic errors(Kachalia *et al.*, 2007). Croskerry described as many as 42 cognitive biases which lead to cognitive error contributing to medical errors(Croskerry, 2005). These were further categorized into classification scheme which divided them into Cognitive Disposition to Respond (CDR) and Affective Disposition to Respond (ADR)(Croskerry, 2005). Each type of cognitive biases can influence decision making thus risk in having medical errors.

There are many strategies which have been invented as cognitive debiasing strategies to reduce errors in clinical decision making but none of them are perfect as there are many types and subtypes of cognitive errors upon making a decision in daily works in emergency department(Croskerry *et al.*, 2013). To improve clinical decision making in emergency department which affected by cognitive errors, Chew et al 2015 (Chew *et al.*, 2015)have created a new cognitive debiasing strategy to reduce errors in clinical decision making in emergency department known as; TWED Matrix.

The TWED Matrix is a mnemonic where ‘T’ for ‘Threat’ (is there any limb threat that I need to rule out in this patient, ‘W’ for ‘what else’ (What if I am wrong?, what else could it be?), ‘E’ for ‘evidences’ (Do I have sufficient evidences to support my diagnosis?) and ‘D’ for ‘dispositional influences’ (Is there any dispositional influence that affects my decision?). The dispositional influence can further be divided into 2 subcategory of ‘E’’s. First ‘E’ is for environmental factors such as stressful clinical setting and second ‘E’ is for Emotional factors such as anger.

The TWED matrix although it is still new, it might reduce major cognitive errors as it covers a wide spectrum of common errors during clinical decision making in emergency department. The purpose of this validation and reliability study is to show whether the TWED matrix is a valid tool aimed to reduce cognitive errors in clinical decision making in emergency department.

OBJECTIVES OF THE STUDY

General Objectives

- To Validate TWED matrix as cognitive debiasing as a strategy to reduce cognitive errors in clinical decision making in emergency department

Specific Objectives

1. To determine content validity index (CVI) of relevance of classes of cognitive biases covered in TWED matrix debiasing strategy.
2. To determine CVI of representativeness of each class covered in TWED debiasing strategy.
3. To determine the reliability of TWED matrix as a strategy to reduce cognitive errors in clinical decision making in Emergency Department

MANUSCRIPT

Manuscript Title Page

DETERMINING THE CONTENT VALIDITY AND RELIABILITY OF TWED MATRIX AS A COGNITIVE DEBIASING STRATEGY IN CLINICAL DECISION MAKING IN EMERGENCY DEPARTMENT

Author:

Dr Mohd Shukri bin MAT SAAD

Emergency Department,

Hospital Universiti Sains Malaysia,

Kubang Kerian, 16150 Kelantan.

Corresponding Author:

Dr Mohd Shukri Bin Mat Saad

Emergency Department,

Hospital Universiti Sains Malaysia,

Kubang Kerian, 16150 Kelantan.

Phone: +609-7676721 Fax: +609- 7673219

Email:drshukrisaad@gmail.com

ACKNOWLEDGEMENT

I wish to express my sincere gratitude to the following individuals who had given their advice, guidance and expertise to make it possible for me to complete this research project:

1. Dr Andey Abd Rahman
   Main Supervisor,
   Lecturer, Emergency Department,
   School of Medical Science, Universiti Sains Malaysia

2. Dr Muhamad Saiful Bahri Bin Yusoff,
   Co-supervisor,
   Lecturer, Department of Medical Education,
   School of Medical Science, Universiti Sains Malaysia

3. Prof. Chew Keng Sheng
   Lecturer Medical Faculty and Health Science,
   Universiti Malaysia Sarawak.

4. All lecturers in the Emergency Department, School of Medical Sciences, Health Campus, Universiti Sains Malaysia.

5. My supportive classmates in the Master of Medicine (Emergency) programme 2013/17.

6. My caring wife, Bahirah Ahmad, my beloved son, Muhammad Amsyar and daughter, Nur Aisyah, my understanding parents and parents-in-law for their support and strength.

Abstract

Determining the Content Validity and Reliability of TWED Matrix as a Cognitive Debiasing Strategy in Clinical Decision Making In Emergency Department

Background: Cognitive biases always complicate clinical decision making in emergency department. TWED Matrix was invented as a tool for debiasing strategy to reduce diagnostic error caused by cognitive biases. It is in the form of mnemonic to encourage metacognition by doctors to reduce errors in clinical decision making in emergency department. As it is relatively new tool, a study was conducted to determine the content validity and reliability of TWED Matrix as cognitive debiasing strategy in clinical decision making in emergency medicine.

Methodology: This study was divided into two stages; the developmental stage and the judgment stage. In the developmental stage, we narrowed down 50 cognitive biases into six classes, and total of four vital questions covering eight items were asked. Then, in the judgment stage, emergency physicians as expert judges were chosen to evaluate eight items in the form of Likert scale for their relevance and representativeness. For relevance, CVI and modified Kappa were assessed and for representativeness, CVI was calculated. Reliability was tested using Cronbach Alpha.

Results: For relevancy, all of the items were scored of CVI more than 0.78. CVI average for relevancy was 0.89. Modified Kappa statistic was calculated to make adjustment for possibilities of chance agreement. For representativeness, all eight items produced CVI of more than 0.78 which is the minimum value for CVI acceptance. CVI

average was 0.94. Alpha value for item relevancy was 0.767 which is within acceptable value. Reliability on item representativeness produced Cronbach's Alpha of 0.737 which is still above acceptance value of 0.7.

Conclusion: TWED Matrix showed significant validity and reliability in this study although further evaluation and assessment needed to strengthen the power of the study before it could safely and widely use as a cognitive debiasing strategy in clinical decision making in emergency department.

Keywords: TWED Matrix, cognitive debiasing strategy, content validity, reliability

Introduction and Literature Review

Making a clinical decision is part and parcel of a health practitioner's job. And making it right is vital as patients' lives depend on it. When patients arrive at the hospital, most of them will present to the emergency department first before receiving treatment and being admitted to specific wards according to specialties. Thus emergency medicine physicians have a huge responsibility to make sure that they do not miss any life threatening condition for each and every one of the patient. But working in emergency environment is not the same as working in calm and controlled environment. The unique nature of emergency medicine compared to other disciplines makes it prone for medical errors, thus compromising patient's safety(Laxmisan *et al.*, 2007). There are multiple factors that contribute either directly or indirectly towards medical errors in emergency medicine(Croskerry and Sinclair, 2001). The challenges face by emergency physicians because of their unique nature of work requires them to see all types of cases that are presented at anytime of the day by any patient. Unpredictable surge of patient load regardless of whether weekdays or weekends or during any communicable disease outbreaks is also an important factor that may lead to medical errors. Shift system which applies to almost all emergency departments healthcare providers is also one of the contributing factors which makes medical errors more likely to occur when they are working under stressful, tiring and undesirable condition.

Most patients who make their first visit to emergency department are incapable of giving proper medical history and sometimes even their particulars to their primary healthcare providers because of the severity of their illness. Despite having minimal information, emergency healthcare providers still need to give their best point of care towards patients.

There are many types of medical errors, but one of the most important errors is cognitive error(Croskerry, 2005). Cognitive error means error in the process of thinking. In emergency medicine, any error in the process of thinking towards making the right diagnosis or decision may lead to catastrophic outcome. Patients may end up with morbidity or mortality, emergency physician may even be sued by the patients (Studdert *et al.,* 2006). Allegations of substandard diagnostic care in emergency department was as much as 28% (Kachalia *et al.,* 2007). Diagnostic errors in emergency medicine ranged between 1%- 12% and 95 % of it were from cognitive errors. Besides cognitive error, lack of knowledge and skills, misinterpretation of lab results, junior healthcare provider and multitasking also might contribute to medical and diagnostic errors(Kachalia *et al.,* 2007). Croskerry described as many as 42 cognitive biases which lead to cognitive error contributing to medical errors(Croskerry, 2005). These were further categorized into classification scheme which divided them into Cognitive Disposition to Respond (CDR) and Affective Disposition to Respond (ADR)(Croskerry, 2005). Each type of cognitive biases can influence decision making thus risk in having medical errors.

There are many strategies which have been invented as cognitive debiasing strategies to reduce errors in clinical decision making but none of them are perfect as there are many types and subtypes of cognitive errors upon making a decision in daily works in emergency department(Croskerry *et al.,* 2013). To improve clinical decision making in emergency department which affected by cognitive errors, Chew et al 2015 (Chew *et al.,* 2015)have created a new cognitive debiasing strategy to reduce errors in clinical decision making in emergency department known as; TWED Matrix.

The TWED Matrix is a mnemonic where ‘T’ for ‘Threat’ (is there any limb threat that I need to rule out in this patient, ‘W’ for ‘what else’ (What if I am wrong?,

what else could it be?), ‘E’ for ‘evidences’ (Do I have sufficient evidences to support my diagnosis?) and ‘D’ for ‘dispositional influences’ (Is there any dispositional influence that affects my decision?). The dispositional influence can further be divided into 2 subcategory of ‘E’’s. First ‘E’ is for environmental factors such as stressful clinical setting and second ‘E’ is for Emotional factors such as anger.

The TWED matrix although it is still new, it might reduce major cognitive errors as it covers a wide spectrum of common errors during clinical decision making in emergency department. The purpose of this validation and reliability study is to show whether the TWED matrix is a valid tool aimed to reduce cognitive errors in clinical decision making in emergency department.

Methodology

Materials

TWED matrix (Figure 1) is a checklist in the form of a mnemonic, aimed to encourage metacognition by the doctor so as not to miss important diagnoses as well as to be convinced that he has sufficient evidence for his decisions. Each quadrant represents a reflective question. The letter 'T' stands for 'Threat' ('is there any life or limb threat that I need to rule out in this patient?'), 'W' for 'Wrong/What else' ('What if I am wrong? What else could it be?'), 'E' for 'evidences' ('do I have sufficient evidences to support or exclude this diagnosis?') and 'D' for 'dispositional factors' ('is there any dispositional factor that influences my decision'; and the two dispositional factors can be represented by 2 'E's: the 'environmental' factors and 'emotional' factors' on the part of the doctor and the patient. The purpose of this validation study is to ascertain the content validity of the classes of cognitive biases represented in the four quadrants of the TWED matrix.

Procedure

Stage 1 – The development stage

To develop the TWED matrix, a two-stage approach was employed. First, in the development stage, a literature search on cognitive biases in clinical decision making was conducted. Although over 50 cognitive biases have been identified, about 30 of these biases are relevant in the context of clinical medicine (Croskerry 2008). These biases are also coined as cognitive predispositions to respond (CDRs) as they steer a clinician's cognitive process towards a certain direction in his clinical decisions which

increases the vulnerability of the clinician to medical error (Campbell 2007, Croskerry 2008). Due to their considerable overlaps and interactions, Campbell et al (2008) has classified these cognitive biases into 6 classes. These categories are 1) errors of over attachment to a particular diagnosis (examples of cognitive biases in this class include anchoring and confirmation bias), 2) errors due to failure to consider alternative diagnoses (example of cognitive biases: search satisficing), 3) errors due to inheriting someone else's thinking (examples of cognitive biases: diagnostic momentum and framing effect), 4) errors in prevalence perception or estimation (examples of cognitive biases: availability bias, gambler's fallacy and posterior probability error) 5) errors involving patient characteristics or presentation context (examples of cognitive biases: fundamental attribution error, gender bias) and 6) errors associated with physician affect or personality (examples of cognitive biases: visceral bias and sunk cost fallacy). Personal communications were held with Professor Patrick Croskerry, a content expert on cognitive biases in clinical decision-making and patient safety in Dalhousie University, Halifax, Canada. The objective of these personal discussions was to identify concise vital questions aimed to encourage metacognition as a form of cognitive debiasing strategy. To cover for the six classes of cognitive biases suggested by Campbell et al (2007), three basic questions are identified: Question 1: 'Do I have sufficient evidences to support or exclude this diagnosis?' This question covers two classes of cognitive biases, viz., cognitive biases due to inheriting someone else's thinking such as triage cueing and diagnostic momentum as well as cognitive biases due to erroneous estimation or perception of prevalence (such as availability bias and gambler fallacy); Question 2: 'What if I am wrong? What else could it be?' This question covers another two classes of cognitive biases, viz., cognitive biases due to overattachment to a particular diagnosis (such as anchoring and confirmation bias) as

well as cognitive biases due to failure to consider alternative diagnoses (such as diagnosis momentum) and 3) 'is there any dispositional influence that affects my decision?' This question covers last two classes of cognitive biases, viz., cognitive biases associated with patient characteristics or presentation context (such as fundamental attribution bias) and cognitive biases associated with doctor's affect or personality (such as ego bias and sunk cost fallacy). Besides the emotive impact of the patient and the doctor that predispose a doctor to fall into the trap of cognitive biases, the impact of the workplace environment or milieu where the clinical decision is made is also identified as an important factor. A chaotic or overcrowded environment often pressurizes a doctor to increased risk-taking behavior such as making faster decisions when the laboratory results are not yet available (Croskerry CJEM 2013). In addition, a separate question, Question no. 4, 'is there any life or limb threat that I need to rule out in this patient?' represents the rule-out-worst-case-scenario (ROWS) heuristic as this is almost pathognomonic of any clinical decision made in the emergency department. (Croskerry 2002). ROWS is vital because cognitive biases due to failure to consider these worst-case scenarios can be disastrous in an emergency department. Pulling all these important reflective questions together and tagging a letter to each of these questions as a form of mnemonic for easy remembering, the TWED matrix is created (Figure 1). The classes of cognitive biases covered in each quadrant of the TWED matrix is given in Table 1.

Stage 2 The judgment stage

A group of emergency physicians from the emergency department of Hospital Universiti Sains Malaysia was invited to judge the content validity of the TWED matrix as well as the underlying classes of cognitive biases that are covered in each of the quadrant. The content validity of the matrix measures the degree to which the matrix has the relevant items or reflective questions to represent the purported aim of the matrix, viz., as a cognitive screening tool to minimize the risk of committing cognitive biases. To do so, the content validity index (CVI) is determined. The content validity is assessed on a four-point Likert scale for its representativeness (where 1 = not representative of the quadrant, 2 = somewhat representative of the quadrant, 3 = quite representative of the quadrant, 4 = highly representative of the quadrant) and for relevance (where 1 = not relevant at all, 2 = somewhat relevant, 3 = quite relevant and 4 = highly relevant). The CVI for relevance and representativeness is defined as the proportion of the judges (in this case, the emergency doctors) who rate the item as a 3 or 4 on that four-point Likert scale. The content validity for the entire TWED matrix is calculated by averaging the CVIs of each individual items (Lynn 1986; DeVon 2007). However, as expounded by Polit et al (2007), as an inter-rater agreement index, CVI does not make adjustment to the possibility of chance agreement. To account for chance agreement on items that are relevant only (minus chance agreement on items that are not relevant), the modified kappa statistics for each item is also calculated (Polit et al 2007). The evaluation criteria for modified kappa (k*) is based on the guidelines described in Cicchetti & Sparrow (1981) and Fleiss (1981) where k* = 0.40 – 0.59 is considered as fair, k* = 0.60 – 0.74 as good and k* >0.74 as excellent (Polit et al 2007).

Note:

The formula for modified kappa statistic (k*) is as follow:

$k^* = (CVI – pc)/(1 – pc)$

where pc represents probability of a chance occurrence

the formula for pc is: $(N!/A!\ (N-A)!)*0.5^N$

N = the number of judges, A = the number agreeing on good relevance, CVI-

Sample Size

After inclusion and exclusion criteria applied, the numbers of judges selected are 10(Polit and Beck, 2006) which are the Emergency Medicine specialists.

Sampling Method

Ten senior most emergency physicians were selected as expert judges to answer the questionnaire.

Data Collection

Data collection was done by giving the research participants a one page questionnaire to be completed during weekly CME session in Emergency Department. Each judge estimated to complete the questionnaire within 10 minutes after the initial briefing regarding the study.

Validity And Realiability of the Measurements Tools

The Content Validity Index (CVI) for relevance and representativeness is defined as the proportion of the judges (in this case, the emergency doctors) who rate the item as a 3 or 4 on that four-point Likert scale. The content validity for the entire TWED matrix was calculated by averaging the CVIs of each individual item (Lynn 1986; DeVon 2007). Lynn 1986 recommended that values of no lower than 0.78 is acceptable. Coefficient Alpha (Cronbach's Alpha) and Intraclass Correlation Coefficient were used to analyze

the reliability of each item represented in this study. For Cronbach's Alpha, values of 0.7 – 0.8 are regarded as satisfactory (Bland and Altman, 1997).

Definition of Operational Terms

The evaluation criteria for modified kappa (k*) is based on the guidelines described in Cicchetti & Sparrow (1981) and Fleiss (1981) where k* = 0.40 – 0.59 is considered as fair, k* = 0.60 – 0.74 as good and k* >0.74 as excellent (Polit et al 2007)

Intended Statistical Analysis

For validation study, Content Validity Index was used; and for Reliability, Cronbach's Alpha and Intraclass Correlation Coefficient were used. Data were analyzed using SPSS version 22 and Microsoft Excel.

Results

Ten senior most emergency physicians in Emergency Department HUSM were selected as expert judges to determine the relevance and representativeness of all 8 items representing 4 quadrants of TWED matrix as part of cognitive debiasing strategy in clinical decision making in emergency department.

CVI were calculated for each item by summing the number of judges who rated 3 or 4 on 4-point Likert scale with total number of judges which are 10. For relevancy (Table 2), all of the items were scored of CVI more than 0.78 which was the minimum value for acceptance in CVI except item six which scored 0.7. CVI average for relevancy was 0.89 and CVI universal agreement was 0.4. Modified Kappa statistic was calculated to make adjustment for possibilities of chance agreement. Minimum score was for item six with 0.7 and maximum score of 1.0 for item 1, 3, 7 and 8.

For representativeness (Table 3), all 8 items produced a significant CVI value of representativeness with minimum of 0.8 for item 4 and 6 and maximum CVI for item 1, 2, 5, 7 and 8. CVI average was 0.94 and CVI universal agreement was 0.5.

Reliability of the item relevancy and representativeness were calculated via Cronbach's Alpha. For relevancy, Cronbach's Alpha was 0.767 (Table 4). For each item rated by expert judges, mean score ranged from 3.2 to 3.9 for all 8 items on 4-point Likert scale (Table 5). Minimum standard deviation was for item 1 with 0.32 and maximum of 1.1 for item 4. Cronbach's Alpha if item deleted increased for item 8 with 0.784 and reduced to 0.695 and 0.698 for item 2 and 4 respectively (Table 6).

Cronbach's Alpha for representativeness was also shown a significant value of 0.737 (Table 4). Mean score distribution ranges from 3.2 to 3.9 for 8 items analyzed (Table 8). Standard deviation for item 1 showed minimum value of 0.32 and maximum

for item 4 with 1.03. Cronbach's Alpha if item deleted increased for item 5 and 8 with 0.76 and 0.75 respectively. Alpha value fall as low as 0.65 if item 2 deleted (Table 9).

Discussion

The process to determine the content validity and reliability of TWED Matrix as a tool for cognitive debiasing strategy in clinical decision making in emergency department was crucial. Traditionally, medical education has focused on content rather than clinical decision making (Kovacs and Croskerry, 1999), thus making a quick and accurate clinical decision a challenge especially in the emergency department where inadequate physician training, high acuity and high volume of patient contribute to the preventable emergency department error (Croskerry and Sinclair, 2001).

Item Relevancy

For content validity, CVI for relevancy was calculated by the proportion of judges who rated three or four on four- point Likert scale for each item. For item relevancy, all items were scored more than 0.78 which is the minimum acceptable value when there are six or more judges as mentioned by Lynn 1986 (Polit and Beck, 2006) except for item six which scored 0.7. Item six explored about pre-dispositional factor and the question asked about errors associated with patient characteristics or presentation context. The lower score contributed by three judges who only rated item six as somewhat relevant only (score 2). Although CVI UA for relevancy is low with 0.4, but the CVI average is significant with 0.89. It is difficult to get a universal agreement for all the expert judges for each item as TWED Matrix is relatively new and covers variant numbers of cognitive biases. But with the significant CVI average, all of the items representing four quadrants of TWED Matrix are consider being relevant. With some modification and adjustment to question representing item six, the score might be improved. To adjust CVI for chance agreement Modified Kappa (k^*) was applied. The index is called a modified kappa because it is an index of agreement of

certain type, namely agreement among the judges that the item is relevant (Polit *et al.*, 2007). Item six scored a minimum but good modified kappa index as it has the lowest CVI and the remaining items have an excellent score of more than 0.74.

Item Representativeness

Eight item questions representing four quadrants in TWED Matrix examined for their representativeness by calculating their CVI. All eight items produced CVI of more than 0.78 which is the minimum value for CVI acceptance. CVI average was 0.94 and CVI UA was 0.5. The questions for each item showed a significant representativeness while covering four quadrants of TWED matrix despite showing lower score in CVI UA as total agreement was impossible to achieve with a newly developed TWED Matrix and high number of expert judges.

Reliability

Cronbach's Alpha was used to determine the reliability of item each for relevancy and representativeness. It is vital to determine the internal consistency of the item to make sure it measures the same concept or construct (Tavakol and Dennick, 2011). Alpha value for item relevancy was 0.767 which is within acceptable value of 0.70 to 0.95, although in clinical application a higher value of 0.9 to 0.95 is desirable (Bland and Altman, 1997). For item total statistic for relevancy, all items are worthy of retention, the greatest increase in Alpha would come by removal of item eight. Therefore, we should not remove any of these items especially item two and four as value will fall below 0.7. Removal of item eight will lead to small improvement of Cronbach's Alpha (0.784), and we can also see that the corrected item total correlation was low (0.138). This might give us an idea whether we should remove this item or not (Gliem and Gliem, 2003).

Reliability on item representativeness produced Cronbach's Alpha of 0.737 which is still above acceptance value of 0.7. We can comfortably say that these items do indeed tap into underlying construct of TWED Matrix. Cronbach's Alpha for relevancy revealed that all item still produced a reliable Alpha value if item deleted except for item two, four and six where the value fall below 0.7 (0.647, 0.687, 0.657) as the respective item deleted.

This validation study on TWED Matrix as a cognitive debiasing strategy in clinical decision making in emergency department involving ten senior most emergency physicians who are currently practicing in emergency department, Hospital Universiti Sains Malaysia. They are considered as expert judges based on their vast experience working and managing emergency cases throughout their exciting career. Overall, TWED Matrix can be considered a valid and reliable tool for Cognitive debiasing strategy in clinical decision making in emergency department. However, as this is relatively new tool produced to reduced cognitive errors, some modification have to be done on certain item to improve their validity and reliability.

Although the results are promising, they only reflect a single centre validation study. The original author of TWED Matrix who was previously practicing in Hospital Universiti Sains Malaysia could also affect the results of this study.

For future recommendation, we would like to take this tool for further evaluation of their validity with face validation to see their feasibility. Furthermore, the validation could be projected to another emergency medicine department to determine whether the results are reproducible or not. In a bigger scale, if the results continue to show that the TWED Matrix is relevant, then it might be possible to include TWED Matrix as part of training module for the undergraduates and postgraduates student in emergency

department. Hopefully with the development of this tool, it would help emergency doctors to improve their clinical decision making without missing any serious and life threatening condition supported with good evidence for their diagnosis (Croskerry, 2002; Guly, 2001).