

นิพนธ์ต้นฉบับ

ระเบียบวิธีวิจัยและวิธีการทางสถิติในการสร้างเครื่องมือวัดคุณภาพชีวิตที่เกี่ยวข้องกับสุขภาพ ในวารสารและวิทยานิพนธ์ทางการแพทย์และสาธารณสุขของไทย

กรรณา จันทุม⁽¹⁾ และ จิราพร เขียวอยู่⁽²⁾

บทคัดย่อ

งานวิจัยนี้ศึกษาระเบียบวิธีวิจัยและสถิติที่ใช้สร้างเครื่องมือวัดคุณภาพชีวิตที่เกี่ยวข้องกับสุขภาพประเภทสร้างใหม่ จากบทความวิจัยและวิทยานิพนธ์ระดับบัณฑิตศึกษาทางการแพทย์และสาธารณสุขของไทย ที่เผยแพร่ในช่วงมกราคม 2543 ถึง ธันวาคม 2552 สืบค้นบทความวิจัยจากฐานข้อมูล E-Database และสืบค้นด้วยมือ สืบค้นวิทยานิพนธ์จากฐานข้อมูลวิทยานิพนธ์ TDC คำสำคัญในการสืบค้น ได้แก่ “health related quality of life” “illness related quality of life” “disease related quality of life” และ “disorder related quality of life”

ได้บทความวิจัยและวิทยานิพนธ์เข้าข่าย 3 บทความ และ 10 เรื่องตามลำดับ ซึ่งเมื่อพิจารณาตามมาตรฐานการสร้างสเกลที่ผู้วิจัยสร้างขึ้นโดยศึกษาจากตำรา งานวิจัยและข้อเสนอแนะของผู้เชี่ยวชาญ พบ 13 เรื่องรายงานการวิจัยนิยามตัวแปรที่จะวัด, การรวบรวมคำถาม, การกำหนดรูปแบบสเกล, และค่าความเที่ยงของสเกลด้วยสัมประสิทธิ์แอลฟา, 12 เรื่องรายงานการทดสอบสเกลเบื้องต้น, 11 เรื่องรายงานการตรวจสอบความตรงตามเนื้อหา โดย 7/11 เรื่องประเมินโดยผู้เชี่ยวชาญ 3/11 เรื่องเสนอค่า CVI, 9 เรื่องรายงานการวิเคราะห์รายการคำถาม โดย 8/9 เรื่องใช้การวิเคราะห์ปัจจัย, 6 เรื่องรายงานการนำไปใช้, 3 เรื่องรายงานการตรวจสอบความตรงตามโครงสร้าง สถิติที่ใช้คือสหสัมพันธ์อย่างง่ายและการวิเคราะห์ปัจจัยเชิงยืนยัน, 2 เรื่องรายงานการตรวจสอบความตรงตามเกณฑ์สัมพัทธ์ด้วยสหสัมพันธ์อย่างง่าย, ในทุกงานไม่พบการรายงานการประเมินความไว

แม้ขั้นตอนการสร้างสเกลของงานที่พบทวนจะครบถ้วน แต่วิธีที่ใช้ในแต่ละขั้นตอนยังต้องปรับปรุง เช่นควรใช้หลายวิธีการร่วมกันในการนิยามตัวแปร และควรประเมินความตรงตามโครงสร้างเสมอ

คำสำคัญ: ระเบียบวิธี, การสร้างสเกล, คุณภาพชีวิตที่เกี่ยวข้องกับสุขภาพ

⁽¹⁾ นักศึกษาหลักสูตรสาธารณสุขศาสตรบัณฑิต สาขาวิชาชีวสถิติ คณะสาธารณสุขศาสตร์ มหาวิทยาลัยขอนแก่น (e-mail : kc.ctrui@gmail.com)

⁽²⁾ ผู้ช่วยศาสตราจารย์ ภาควิชาชีวสถิติและประชากรศาสตร์ คณะสาธารณสุขศาสตร์ มหาวิทยาลัยขอนแก่น

Original Article

Methods and statistical applications in the construction of health related quality of life scales published in Thai medical journals and theses

Karuna Chuntum⁽¹⁾ and Jiraporn Khiewyoo⁽²⁾

Abstract

This study aimed to examine methodologies and statistical methods in constructing health related quality of life scales as reported in Thai health journals and theses from January 2000 to December 2009. Research articles were searched from E-Databases and by hand searching. Theses were search from the TDC database. Searching keywords were “health related quality of life”, “illness related quality of life”, “disease related quality of life” and “disorders related quality of life.”

Three research articles and 10 theses were identified. Standard methods for constructing scales followed suggestions of textbooks, previous studies and recommendations from experts. All 13 studies elaborated on 4 issues: 1) methods used to define the construct or definition of a variables wished to measure, 2) methods in generating items, 3) designing the scales and 4) testing the reliability of the scales with the Cronbach's alpha coefficient. Twelve out of the 13 studies pre-tested the scales. Eleven studies mentioned the methods used to assess the content validity. Seven and three studies assessed the content validity by using expert's opinions and by calculating the content validity index (CVI) respectively. Nine studies mentioned the methods used for the item analysis; 8 studies used an exploratory factor analysis. Six studies described the details of how to use the scales. Three studies described the methods for assessing the construct validity of the scales; a simple correlation coefficient was calculated and a confirmatory factor analysis was done. Two studies tested the criterion-related validity by calculating a simple correlation coefficient. None of the studies reported an evaluation of the sensitivity to change property of the scale.

Although all studies reported steps in constructing a scale covered as the standard methods suggested, but the process of each steps need to be revised. Scale developer should use several methods in defining the construct of the variable needed to measure. All studies must assess the construct validity of the scale.

Keyword: *methodology, scale development, health related quality of life*

⁽¹⁾ Master Degree in Public Health Student, Faculty of Public Health, Khon Kaen University (e-mail : kc.ctr@gmail.com)

⁽²⁾ Assistant Professor, Department of Biostatistics and Demography, Faculty of Public Health, Khon Kaen University

บทนำ

คุณภาพชีวิตที่เกี่ยวข้องกับสุขภาพ (health related quality of life) เป็นผลลัพธ์ที่ไม่สามารถมองเห็นได้จากผลทางคลินิกหรือการตรวจทางห้องปฏิบัติการ การประเมินคุณภาพชีวิตมีความสำคัญมากในการทำความเข้าใจว่าโรคและการรักษามีผลอย่างไรต่อชีวิตของผู้ป่วยจากมุมมองของผู้ป่วย ความสำคัญของการประเมินคุณภาพชีวิตที่เกี่ยวข้องกับสุขภาพมีหลายประการด้วยกัน ไม่ว่าจะเป็นการติดตามประเมินภาวะสุขภาพของผู้ป่วย การประเมินผลการรักษาหรือการดูแลทางสุขภาพรวมถึงการติดตามดูแลผู้ป่วยทางคลินิกซึ่งในทางปฏิบัติการประเมินคุณภาพชีวิตที่เกี่ยวข้องกับสุขภาพเป็นอีกวิธีการหนึ่งที่ใช้ในการประเมินภาวะสุขภาพของผู้ป่วย (Fayers and Machin, 2000) ปัจจุบันการวัดคุณภาพชีวิตที่เกี่ยวข้องกับสุขภาพมีทั้งการนำเครื่องมือที่มีอยู่แล้วจากต่างประเทศมาใช้ และสร้างขึ้นใหม่โดยนักวิจัย ในกรณีการนำเครื่องมือวัดคุณภาพชีวิตจากต่างประเทศมาใช้ ต้องผ่านกระบวนการแปลภาษาและตรวจสอบคุณสมบัติด้านความเที่ยง (reliability) ก่อนนำมาใช้ เนื่องจากเครื่องมือที่มีความเที่ยงในประชากรกลุ่มหนึ่ง ไม่ได้หมายความว่ามีความเที่ยงในประชากรอีกกลุ่มหนึ่ง โดยเฉพาะการนำเครื่องมือวัดคุณภาพชีวิตที่เกี่ยวข้องกับสุขภาพมาใช้ในประชากรที่มีวัฒนธรรมแตกต่างกัน (Fayers and Machin, 2000) สำหรับบางการวิจัยต้องการพัฒนาเครื่องมือวัดคุณภาพชีวิต เพื่อใช้ประโยชน์ในการติดตามผลการรักษาหรือติดตามผลการดำเนินงานอื่นๆ เนื่องจากเครื่องมือเดิมไม่มีประเด็นที่ครอบคลุมเรื่องที่ต้องการวัด หรือยังขาดเครื่องมือที่เหมาะสมในการทำวิจัย จึงต้องพัฒนาขึ้นใหม่ การสร้างเครื่องมือวัดตัวแปรใด ๆ เครื่องมือที่สร้างขึ้นจะต้องมีความน่าเชื่อถือโดยมักพิจารณาจากคุณสมบัติด้านจิตเมตริก (psychometric properties) ของสเกล รวมถึงความเหมาะสมในการนำเครื่องมือไปใช้ในทางปฏิบัติ การสร้างเครื่องมือวัดคุณภาพชีวิตที่เกี่ยวข้องกับสุขภาพให้มีคุณภาพดีตามหลักการของการวัดเป็นสิ่งจำเป็นและมีความสำคัญ เนื่องจากมีผลต่อความถูกต้องในการนำไปใช้ประเมินผลการรักษาผู้ป่วย จากการทบทวนงานวิจัยประเภททบทวนระเบียบวิธีการสร้างเครื่องมือประเภทสเกล ที่ปรากฏในวารสารต่าง ๆ ในต่างประเทศ ประเด็นที่นักวิจัยพิจารณา ได้แก่ วิธีการสร้างสเกลเป็นไปตามหลักการของการสร้างเครื่องมือหรือไม่ ซึ่งพบว่า มีการรายงาน

ข้อมูลระเบียบวิธีวิจัยและวิธีการทางสถิติที่ใช้สร้างสเกลค่อนข้างครบถ้วน ไม่ว่าจะเป็นการกำหนดแนวคิดหรือโครงสร้างของตัวแปรที่ต้องการสร้างสเกลเพื่อวัด การรวบรวมรายการคำถาม การกำหนดรูปแบบของสเกล การทดสอบสเกลเบื้องต้น การวิเคราะห์รายการคำถาม โดยพบการรายงานมากกว่าร้อยละ 50 สำหรับการตรวจสอบตามคุณสมบัติด้านจิตเมตริกของสเกล พบการรายงานการประเมินความตรงตามเนื้อหามากที่สุดคืออยู่ระหว่างร้อยละ 35 ถึง 100 ของบทความที่พิจารณา การประเมินความเที่ยงของสเกลประเภทความสอดคล้องภายใน พบร้อยละของบทความที่รายงานอยู่ระหว่างร้อยละ 11.9 ถึง 100 ส่วนการทดสอบคุณสมบัติความไวต่อการเปลี่ยนแปลงพบค่อนข้างน้อย โดยพบร้อยละของบทความที่รายงานอยู่ระหว่างร้อยละ 7.0 ถึง 26.9 (Hinkin, 1995; Hensley, 1999; Coste et al., 1995; Coste et al., 1997) จากการศึกษาข้อมูลเบื้องต้นโดยผู้วิจัยซึ่งทำการสำรวจบทความวิจัยที่พิมพ์เผยแพร่ในวารสารจดหมายเหตุทางการแพทย์ที่พิมพ์เผยแพร่ทั้งหมด 1,637 บทความ ระหว่าง เดือนมกราคม พ.ศ.2543 ถึงเดือนธันวาคม พ.ศ.2552 พบว่าเป็นบทความวิจัยประเภทสร้างสเกลเพื่อวัดคุณภาพชีวิตที่เกี่ยวข้องกับสุขภาพประเภทสร้างใหม่จำนวน 2 บทความ บทความแรก สร้างแบบประเมินคุณภาพชีวิตไทยแบบรูปภาพ (Phattharayuttawat et al., 2005) บทความที่สอง สร้างแบบประเมินคุณภาพชีวิตในผู้ป่วยโรคสายตาเลือนราง (Yingyong, 2007) บทความทั้งสองงานวิจัยข้างต้นเป็นการสร้างเครื่องมือเพื่อวัดตัววัดด้านสุขภาพของผู้ป่วยโรคเรื้อรังในประเทศไทยยังไม่พบงานวิจัยประเภทการทบทวนระเบียบวิธีการสร้างเครื่องมือประเภทสเกล ประกอบกับความสำคัญของคุณภาพของเครื่องมือวัดคุณภาพชีวิตที่พัฒนาใหม่ดังกล่าวแล้วข้างต้น ผู้วิจัยจึงเห็นว่า ควรจะทำการศึกษาเพื่อให้ได้ข้อมูลเชิงประจักษ์เกี่ยวกับระเบียบวิธีวิจัยและวิธีการทางสถิติในการสร้างเครื่องมือวัดคุณภาพชีวิตที่เกี่ยวข้องกับสุขภาพ ที่ปรากฏในวารสารและวิทยานิพนธ์ระดับบัณฑิตศึกษาทางการแพทย์และสาธารณสุขของไทย เพื่อนำข้อค้นพบที่ได้ ไปเป็นข้อเสนอแนะและแนวทางให้กับผู้ที่ประสงค์จะสร้างเครื่องมือวัดคุณภาพชีวิตที่เกี่ยวข้องกับสุขภาพ ให้สามารถสร้างได้อย่างมีคุณภาพต่อไป

วัตถุประสงค์การวิจัย

เพื่อศึกษาระเบียบวิธีวิจัยและวิธีการทางสถิติที่ใช้ในการสร้างเครื่องมือวัดคุณภาพชีวิตที่เกี่ยวข้องกับสุขภาพ ที่ปรากฏในบทความวิจัยที่เผยแพร่ในวารสารการแพทย์และสาธารณสุขของไทย และในวิทยานิพนธ์ระดับบัณฑิตศึกษาทางการแพทย์และสาธารณสุขของมหาวิทยาลัยในประเทศไทย

วิธีดำเนินการวิจัย

ประชากรและตัวอย่าง

ประชากรในการศึกษาค้นคว้าครั้งนี้คือ บทความวิจัยและวิทยานิพนธ์ที่มีวัตถุประสงค์เพื่อสร้างสเกลชนิดใหม่เพื่อวัดคุณภาพชีวิตที่เกี่ยวข้องกับสุขภาพ กรณีบทความวิจัยเป็นบทความที่ได้รับการตีพิมพ์เผยแพร่ในวารสารด้านการแพทย์และสาธารณสุขของเมืองไทย กรณีวิทยานิพนธ์เป็นวิทยานิพนธ์ระดับบัณฑิตศึกษาทางการแพทย์และสาธารณสุขของมหาวิทยาลัยในประเทศไทย

เกณฑ์การคัดเลือกบทความวิจัยและวิทยานิพนธ์ระดับบัณฑิตศึกษาเข้าในการศึกษา

1. เป็นบทความวิจัยหรือวิทยานิพนธ์ระดับบัณฑิตศึกษาประเภทสร้างสเกลใหม่
2. สเกลที่สร้างใหม่เป็นสเกลวัดคุณภาพชีวิตที่เกี่ยวข้องกับสุขภาพ ซึ่งมีคำสำคัญที่เกี่ยวข้องตามที่ระบุ
3. เป็นบทความวิจัยที่ปรากฏในวารสารด้านการแพทย์และสาธารณสุขของเมืองไทย โดยมีเกณฑ์คัดเลือกวารสาร ดังนี้

1) เป็นวารสารที่มีการอ้างอิงถึงมากที่สุด ตามการจัดอันดับในฐาน Journal Citation Report โดยคัดเลือกรายชื่อวารสารที่คงอยู่ในอันดับที่ 1-50 ซ้ำกัน 2 ใน 3 ปีตั้งแต่ปี พ.ศ. 2549 ถึง ปี พ.ศ.2551 จากเกณฑ์ดังกล่าวได้วารสารที่จะศึกษาทั้งหมด 15 วารสาร ได้แก่ วารสารสมาคมพยาบาลสาขากายตะวันออกเฉียงเหนือ, วารสารสุขภาพจิตแห่งประเทศไทย, วารสารเทคนิคการแพทย์และกายภาพบำบัด, จดหมายเหตุทางการแพทย์, วารสารสมาคมจิตแพทย์แห่งประเทศไทย, วารสารทันตแพทยศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย, วารสารพยาบาลศาสตร์, สงขลานครินทร์เวชสาร, วารสารพยาบาลทหารบก, The Southeast Asian Journal of Tropical Medicine and Public Health, วารสารพยาบาล, Asian Pacific Journal of

Allergy and Immunology, Walailak Journal of Science and Technology, Kasetsart Journal, และสารศิริราช

2) เป็นวารสารทางการแพทย์และสาธารณสุขของไทยที่พิมพ์เผยแพร่และปรากฏในฐานข้อมูล E-database ได้แก่ PubMed หรือ MEDLINE ที่ถือว่าเป็นวารสารที่เป็นที่ยอมรับในระดับสากล จากเกณฑ์นี้มีวารสาร 1 วารสารคือ จดหมายเหตุทางการแพทย์

เมื่อใช้เกณฑ์ทั้งสองเกณฑ์ข้างต้น มีวารสารที่เข้าข่าย 15 วารสารดังรายชื่อข้างต้น

4. เป็นวิทยานิพนธ์ระดับบัณฑิตศึกษาในสาขาวิชาทางการแพทย์และสาธารณสุขของไทย ที่ปรากฏในฐานข้อมูลวิทยานิพนธ์และการค้นคว้าอิสระของมหาวิทยาลัยไทย (Thai Digital Collection; TDC)

ตัวอย่างในการศึกษา เป็นบทความวิจัยทุกบทความในวารสารทางการแพทย์และสาธารณสุขของไทย 15 วารสารข้างต้น และเป็นวิทยานิพนธ์ระดับบัณฑิตศึกษาทางการแพทย์และสาธารณสุขของมหาวิทยาลัยในประเทศไทยในฐานข้อมูลวิทยานิพนธ์และการค้นคว้าอิสระของมหาวิทยาลัยไทย TDC ทั้งวารสารและวิทยานิพนธ์พิมพ์เผยแพร่ในช่วง เดือนมกราคม พ.ศ.2543 ถึง เดือนธันวาคม พ.ศ.2552

คำสำคัญในการสืบค้น และการสืบค้น

คำสำคัญ (keywords) ที่ใช้ในการสืบค้น ได้แก่ “health related quality of life”, “illness related quality of life”, “disease related quality of life”, “disorder related quality of life”, “scale”, “index”, “indicator”, “measure”, “measure were development”, “scale created for this study”, “development”, “new”, “construction”, “created”, “validity testing”, และ “reliability testing”

การสืบค้นบทความวิจัยแบ่งเป็น 2 กรณี กรณีแรกสืบค้นในฐานข้อมูล E-database ได้แก่ PubMed หรือ MEDLINE โดยใช้คำสำคัญที่ใช้ในการสืบค้นที่กำหนดข้างต้น กรณีที่สอง สืบค้นบทความวิจัยที่ปรากฏในวารสารที่ไม่มีการตีพิมพ์เผยแพร่ในฐานข้อมูล E-database โดยการค้นหาด้วยมือ (Hand searching) ซึ่งสืบค้นจากห้องสมุดคณะแพทยศาสตร์ สำนักวิทยบริการ มหาวิทยาลัยขอนแก่น สำนักวิทยบริการ มหาวิทยาลัยมหาสารคาม และสำนักวิทยบริการ มหาวิทยาลัยบูรพา ตามคำสำคัญที่ใช้ในการสืบค้น

การสืบค้นวิทยานิพนธ์ระดับบัณฑิตศึกษาในสาขาวิชาทางการแพทย์และสาธารณสุขของไทย สืบค้นในฐานข้อมูลวิทยานิพนธ์การค้นคว้าอิสระมหาวิทยาลัยของไทย TDC ตามคำสำคัญที่ใช้ในการสืบค้น

ผลการสืบค้น ได้งานวิจัยเข้าข่ายทั้งหมด 13 เรื่อง เป็นบทความจากวารสาร 3 เรื่อง วิทยานิพนธ์ 10 เรื่อง โดยทั้ง 13 เรื่องนี้มาจากการใช้คำสำคัญสี่คำแรก คำสำคัญอื่น ๆ ไม่ให้ผลการสืบค้นใด ๆ

ประเด็นที่ศึกษาและเครื่องมือที่ใช้ในการศึกษา

เครื่องมือเก็บข้อมูลเป็นแบบบันทึก ผู้วิจัยสร้างแบบบันทึกโดยศึกษาจากเอกสาร ตำราและบทความวิจัยเกี่ยวกับระเบียบวิธีมาตรฐานในการสร้างสเกล แล้วนำไปให้ผู้เชี่ยวชาญตรวจสอบคุณภาพ แบบบันทึกมี 2 ส่วน ส่วนแรกเป็นลักษณะทั่วไปของบทความวิจัยตัวอย่าง ส่วนที่สองเป็นขั้นตอนมาตรฐานต่าง ๆ ในการสร้างสเกลวัดคุณภาพชีวิตด้านสุขภาพ โดยในแต่ละขั้นตอนจะพิจารณาว่ามีการรายงานไว้ในบทความหรือไม่ รายละเอียดที่ต้องรายงานมีดังต่อไปนี้

1. การกำหนดแนวคิด คำนิยามหรือโครงสร้างของตัวแปรที่ต้องการสร้างสเกลเพื่อวัด

การรายงาน คือ บทความวิจัยให้รายละเอียดของการกำหนดโครงสร้างเชิงเนื้อหาของสเกล เช่น 1) การทบทวนวรรณกรรม 2) แนวคิดหรือเนื้อหาทางทฤษฎีที่เกี่ยวข้อง 3) งานวิจัยที่ผ่านมาหรือ 4) สเกลที่ปรากฏในปัจจุบันที่เกี่ยวข้องกับสิ่งที่นักวิจัยต้องการสร้างสเกลเพื่อวัด

2. การรวบรวมข้อคำถาม

การรายงาน คือ บทความวิจัยให้รายละเอียดของวิธีการรวบรวมรายการคำถาม เช่น (1) ปรับปรุงจากงานวิจัยที่มีผู้ศึกษาไว้ (2) เขียนขึ้นใหม่โดยอาศัยแนวคิดหรือทฤษฎีที่เกี่ยวข้อง (3) สัมภาษณ์จากผู้เชี่ยวชาญและการระดมสมอง หรือ (4) ความเห็นและข้อเสนอแนะที่ได้จากการสนทนากลุ่มจากบุคคลที่มีส่วนเกี่ยวข้องกับเรื่องที่จะสร้างสเกล

3. การกำหนดประเภทและรูปแบบสเกล

การรายงาน คือ บทความวิจัยให้รายละเอียดของสเกลและรูปแบบที่ใช้

4. การประเมินความตรงตามเนื้อหา (content validity)

การรายงาน คือ รายงานวิจัยให้รายละเอียดของการประเมินความตรงเชิงเนื้อหา เช่น (1) ตรวจสอบเชิง

คุณภาพโดยผู้เชี่ยวชาญ (2) ตรวจสอบเชิงปริมาณโดยหาค่าดัชนีความตรงเชิงเนื้อหา (content validity index: CVI) หรือ (3) ตรวจสอบเชิงปริมาณโดยนำไปทดสอบในกลุ่มเป้าหมาย

5. การทดสอบสเกลเบื้องต้น

การรายงาน คือ บทความวิจัยให้รายละเอียดของวิธีการทดสอบสเกลเบื้องต้น

6. การวิเคราะห์รายการคำถาม

การรายงาน คือ บทความวิจัยให้รายละเอียดของการวิเคราะห์รายการคำถาม ข้อใดข้อหนึ่งหรือทั้งหมด ดังนี้ การกระจายคำตอบของคำถามแต่ละข้อ, การวิเคราะห์ความสัมพันธ์ระหว่างแต่ละข้อคำถาม, การวิเคราะห์ความสัมพันธ์ระหว่างคะแนนของแต่ละข้อคำถามกับคะแนนรวม, การวิเคราะห์ความสอดคล้องภายในในระหว่างข้อคำถาม, และการวิเคราะห์ปัจจัย

7. การประเมินความตรงเชิงโครงสร้าง (construct validity)

การรายงาน คือ บทความวิจัยให้รายละเอียดของประเมินความตรงเชิงโครงสร้าง เช่น ประเมินความตรงเชิงลู่เข้าหรือเชิงจำแนกและสถิติที่ใช้ การวิเคราะห์ปัจจัยและการทดสอบในกลุ่มที่ทราบความแตกต่าง (known group method) และสถิติที่ใช้

8. การประเมินความตรงตามเกณฑ์สัมพัทธ์ (criterion-related validity)

การรายงาน คือ บทความวิจัยให้รายละเอียดของการประเมินความตรงตามเกณฑ์สัมพัทธ์ เช่น ความตรงร่วมสมัยและระบุค่าสถิติที่ใช้ หรือ ความตรงเชิงทำนายและระบุค่าสถิติที่ใช้

9. การประเมินความเที่ยง (reliability test)

การรายงาน คือ บทความวิจัยให้รายละเอียดของการประเมินความเที่ยงในข้อใดข้อหนึ่งหรือทั้งหมด เช่น ความสอดคล้องภายใน ความคงที่หรือความเสถียรของการวัด หรือ ความเที่ยงคู่ขนานและสถิติที่ใช้ในการประเมินความเที่ยงแต่ละกรณี

10. การทดสอบคุณสมบัติความไวต่อการเปลี่ยนแปลง (sensitivity to change)

การรายงาน คือ บทความวิจัยให้รายละเอียดของการประเมินความไว โดยใช้สถิติใดสถิติหนึ่งต่อไปนี้

Cohen's effect size, Guyatt's responsiveness, The standardize response mean (SRM), หรือ t-test comparison

11. การรายงานการนำไปใช้ แบ่งเป็น

รายงาน คือ บทความวิจัยให้รายละเอียดของการนำไปใช้

ในแต่ละประเด็นที่พิจารณาข้างต้น ในกรณีที่มีการรายงาน ผู้วิจัยจะบันทึกรายละเอียดต่าง ๆ ทั้งหมดที่พบจากการรายงาน เพราะถือว่าเป็นข้อมูลที่สำคัญ

การเก็บรวบรวมข้อมูลและตรวจสอบคุณภาพ

เก็บข้อมูลด้วยนักวิจัยสองคน โดยนักวิจัยทั้งสองคนอ่านบทความและบันทึกข้อมูลอย่างเป็นอิสระกัน แล้วนำผลที่ได้มาพิจารณาร่วมกัน หากพบขัดแย้งกัน ก็จะหาข้อสรุป และทำงานกระทั่งได้ผลการบันทึกข้อมูลที่ตรงกัน

การประมวลผลและวิเคราะห์ข้อมูล

ผู้วิจัยนำข้อมูลที่รวบรวมได้มาลงรหัสและบันทึกเข้าเครื่องคอมพิวเตอร์ด้วยโปรแกรม EpiData โดยผู้นำเข้า 2 คน โดยแยกแฟ้มข้อมูล แล้วตรวจสอบความตรงกันของการนำเข้าข้อมูลและแก้ไขให้ถูกต้อง นำข้อมูลที่ถูกต้องไปวิเคราะห์ต่อไป และการวิเคราะห์ข้อมูลต่าง ๆ ทั้งหมด ใช้สถิติพรรณนา โดยนำเสนอเป็นความถี่และร้อยละ และบรรยายรายละเอียดของระเบียบวิธีวิจัยและวิธีการทางสถิติในการสร้างสเกลที่พบในบทความวิจัยหรือวิทยานิพนธ์ที่ศึกษา

ผลการวิจัย

ข้อมูลทั่วไปของบทความวิจัยหรือวิทยานิพนธ์

งานวิจัยที่ทบทวนมีทั้งหมด 13 เรื่อง เป็นบทความวิจัย 3 บทความ วิทยานิพนธ์ 10 เรื่อง พบว่า 9 เรื่อง เป็นงานวิจัยที่มีวัตถุประสงค์หลักเพื่อสร้างสเกลวัดคุณภาพชีวิตที่เกี่ยวข้องกับสุขภาพ และ 4 เรื่อง เป็นงานวิจัยที่มีวัตถุประสงค์หลักเพื่อประเมินคุณภาพชีวิตแต่ผู้วิจัยสร้างสเกลวัดคุณภาพชีวิตขึ้นใหม่ (ตาราง 1) ซึ่งพบว่า 9 เรื่อง สร้างสเกลเพื่อวัดตัววัดสุขภาพเกี่ยวกับโรคเรื้อรัง เช่น คุณภาพชีวิตผู้มีปัญหาสายตา คุณภาพชีวิตผู้ป่วยโรคหลอดเลือดหัวใจตีบ คุณภาพชีวิตผู้ป่วยโรคหัวใจพิการแต่กำเนิด ที่เหลือเป็นตัววัดสุขภาพในประชากรเฉพาะกลุ่ม ได้แก่ คุณภาพชีวิตสตรีวัยหมดประจำเดือน ผู้เลิกบุหรี่ และผู้ป่วยเด็ก

ระเบียบวิธีวิจัยและวิธีการทางสถิติในการสร้างสเกลที่ปรากฏในบทความวิจัยหรือวิทยานิพนธ์

ตาราง 2 แสดงการพิจารณาระเบียบวิธีวิจัยและวิธีการทางสถิติที่ใช้โดยอิงตามมาตรฐานการสร้างสเกล พบว่าขั้นตอนหรือประเด็นที่มีการรายงานในทุกงานวิจัย คือ วิธีการนิยามตัวแปรที่จะวัด, การรวบรวมคำถาม, การกำหนดรูปแบบสเกล, และการประเมินความเที่ยงของสเกลโดยใช้สัมประสิทธิ์แอลฟาของครอนบาค (Cronbach's alpha coefficient) ประเด็นหรือขั้นตอนอื่น ๆ ที่มีการรายงาน เมื่อเรียงลำดับจากพบมากไปน้อย เป็นดังนี้ การทดสอบสเกลเบื้องต้น พร้อมให้รายละเอียดของวิธีการทดสอบและขนาดตัวอย่าง พบ 12/13 เรื่อง, การตรวจสอบความตรงตามเนื้อหา พบ 11/13 เรื่อง โดย 7/11 เรื่องประเมินโดยผู้เชี่ยวชาญ 3/11 เรื่องเสนอค่าดัชนีความตรงเชิงเนื้อหา (content validity index; CVI), การวิเคราะห์รายการคำถาม พบ 9/13 เรื่อง โดย 8/9 เรื่องใช้การวิเคราะห์ปัจจัย, การนำไปใช้พบ 6/13 เรื่อง, การตรวจสอบความตรงตามโครงสร้าง โดยสถิติที่ใช้ เช่น สหสัมพันธ์อย่างง่าย และการวิเคราะห์ปัจจัยเชิงยืนยัน พบ 3/13 เรื่อง, การตรวจสอบความตรงตามเกณฑ์สัมพัทธ์ด้วยสหสัมพันธ์อย่างง่าย พบ 2/13 เรื่อง, และปรากฏว่าในทุกงานไม่พบการรายงานการประเมินความไว (sensitivity to change)

บทสรุปและอภิปรายผล

ผลจากการพิจารณาระเบียบวิธีวิจัยและวิธีการทางสถิติที่ใช้ในการสร้างเครื่องมือวัดคุณภาพชีวิตที่เกี่ยวข้องกับสุขภาพตามมาตรฐานการสร้างสเกลที่ปรากฏในงานวิจัยที่ทบทวน มีข้อสรุปและการอภิปรายผลดังนี้

1. การกำหนดแนวคิดหรือโครงสร้างของตัวแปรที่ต้องการสร้างสเกลเพื่อวัด เมื่อพิจารณาจากบทความวิจัยและวิทยานิพนธ์ระดับบัณฑิตศึกษา ที่ทบทวนทั้งหมด 13 เรื่อง ทุกงานรายงานว่าทำขั้นตอนนี้ โดยมี 1 งาน ใช้ทุกวิธีการต่อไปนี้ คือ ทบทวนแนวคิด/ทฤษฎี และศึกษาจากรายงานวิจัยที่ผ่านมา และนำบางส่วนของสเกลเดิมที่มีอยู่แล้วที่เกี่ยวข้องกับสเกลที่ต้องการสร้างเพื่อวัดคุณภาพชีวิตมี 9 งาน ใช้การทบทวนแนวคิด/ทฤษฎี ร่วมกับการศึกษาจากรายงานวิจัยที่ผ่านมา, มี 2 งาน ใช้การทบทวนแนวคิด/ทฤษฎี ร่วมกับการสัมภาษณ์เชิงลึกจากผู้เชี่ยวชาญหรือทำการสนทนากลุ่มจากกลุ่มเป้าหมาย และมี 1 งาน ใช้การทบทวนแนวคิด/

ทฤษฎี ร่วมกับการนำบางส่วนของสเกลเดิมที่มีอยู่แล้วที่เกี่ยวข้องกับสเกลที่ต้องการสร้างเพื่อวัดคุณภาพชีวิต จะเห็นวิธีการที่ปรากฏในงานที่ทบทวน สอดคล้องกับคำแนะนำของ Streiner and Norman (2003) และ Spector (1992) แนะนำไว้ว่า ในการกำหนดโครงสร้างของตัวแปรเพื่อสร้างสเกลเพื่อวัดนั้น ควรใช้หลายวิธีประกอบกัน ได้แก่ การทบทวนแนวคิดทฤษฎีที่เกี่ยวข้อง การศึกษาจากงานวิจัยที่ผ่านมา การนำบางส่วน ของงานในทำนองเดียวกันมาใช้ ตลอดจนการสอบถามความเห็นจากผู้เชี่ยวชาญและกลุ่มเป้าหมายด้วยวิธีการ สัมภาษณ์เชิงลึกและหรือการสนทนากลุ่ม และเป็นไปใน ทำนองเดียวกับการศึกษาของ Hinkin (1995) ซึ่งทบทวน บทความ 75 บทความ พบ ร้อยละ 83 (62/75) ของบทความที่ ทบทวนมีการกำหนดนิยามของสิ่งที่ต้องการวัดเพื่อนำมา กำหนดโครงสร้างเชิงเนื้อหาของตัวแปรที่ต้องการสร้างสเกล เพื่อวัด โดยใช้วิธีการทบทวนแนวคิด/ทฤษฎี และศึกษาจาก รายงานวิจัยที่ผ่านมา, ร้อยละ 11 (8/75) ของบทความที่ ทบทวนใช้วิธีการสนทนากลุ่มจากกลุ่มเป้าหมาย, และ Hensley (1999) ทบทวนบทความ 6 บทความ พบว่า 6/6 ของบทความที่ ทบทวน รายงานการกำหนดโครงสร้างเชิงเนื้อหาของตัวแปรที่ ต้องการสร้างสเกลเพื่อวัด โดยใช้วิธีการทบทวนแนวคิด/ ทฤษฎี และศึกษาจากรายงานวิจัยที่ผ่านมา อย่างไรก็ตาม ใน งานที่ทบทวนนี้มีสิ่งที่พบแตกต่างกัน คือ บางงานใช้หลายวิธี ร่วมกัน แต่บางงานใช้เพียงหนึ่งหรือสองวิธีข้างต้น

2. การรวบรวมรายการคำถาม เมื่อพิจารณาจาก บทความวิจัย และวิทยานิพนธ์ระดับบัณฑิตศึกษา ที่ทบทวน ทั้งหมด 13 เรื่อง ทุกงานรายงานว่าทำขั้นตอนนี้ โดยมี 2 งาน ใช้ทุกวิธีการต่อไปนี้ คือ ทบทวนแนวคิด/ทฤษฎี และศึกษา จากรายงานวิจัยที่ผ่านมา และ/หรือสัมภาษณ์เชิงลึกโดยระดม สมองจากผู้เชี่ยวชาญหรือการสนทนากลุ่มจากกลุ่มเป้าหมาย และมี 3 งาน ใช้การทบทวนแนวคิด/ทฤษฎี ร่วมกับการ สัมภาษณ์เชิงลึกจากผู้เชี่ยวชาญหรือทำการสนทนากลุ่มจาก กลุ่มเป้าหมาย วิธีการที่การศึกษา 5 งานนี้ใช้ได้รับการยอมรับ ว่าเป็นวิธีการที่เหมาะสมที่สุด (Streiner and Norman, 2003) มี 8 งาน ใช้การทบทวนแนวคิด/ทฤษฎี ร่วมกับการศึกษาจาก รายงานวิจัยที่ผ่านมา วิธีการที่ 8 งานนี้ใช้เป็นวิธีที่นักวิชาการ ด้านการสร้างสเกลหลายคนแนะนำ (Spector, 1992; Devellis, 2003; Streiner and Norman, 2003; Fayers and Machin, 2000) ดังนั้นในประเด็นของการรวบรวมรายการคำถามนี้ จึงสรุปได้

ว่า งานวิจัยที่ทบทวนใช้วิธีการที่หลากหลายประกอบกัน ซึ่งถือ ว่าสอดคล้องกับหลักการสร้างเครื่องมือ นั่นคือในขั้นต้น ผู้สร้างกำหนดรายการคำถาม(item pool) ให้มีปริมาณมากที่สุด เพื่อครอบคลุมเนื้อหาของสิ่งที่ต้องการวัด และเพิ่มความเที่ยง ของเครื่องมือ (Spector, 1992; Devellis, 2003; Streiner and Norman, 2003; Fayers and Machin, 2000)

3. การกำหนดประเภทและรูปแบบสเกล เมื่อ พิจารณาจากบทความวิจัย และวิทยานิพนธ์ระดับบัณฑิตศึกษา ที่ทบทวนทั้งหมด 13 เรื่อง พบว่า ทั้งหมดใช้สเกลของลิเคิร์ต (Likert scale) โดย 11/13 เรื่อง กำหนดตัวเลือกของคำตอบ เป็น 5 ระดับ ซึ่งสอดคล้องกับ Hinkin (1995) ที่ทบทวน บทความ 75 บทความ แล้วพบ 36/75 ของบทความที่ทบทวน กำหนดประเภทสเกลและรูปแบบคำตอบเป็นเช่นนี้ ส่วนงาน ของ Hensley (1999) พบว่า 5/6 ของบทความที่ทบทวนใช้สเกล ของลิเคิร์ตเช่นกัน แต่กำหนดคำตอบเป็น 5-7 ระดับ ส่วนบทความวิจัยและวิทยานิพนธ์ที่เหลืออีก 2 งาน ใช้สเกล ของลิเคิร์ตเช่นกัน แต่กำหนดคำตอบเป็น 3 ระดับ

สเกลของลิเคิร์ตมีลักษณะเฉพาะคือ ประกอบด้วย รายการคำถามหลายข้อ แต่ละข้อวัดออกมาเป็นเชิงปริมาณได้ ไม่มีคำตอบที่ถูกหรือผิดแต่จะให้ผู้ตอบประเมินค่า และแต่ละ รายการคำถามจะมีความสัมพันธ์กันเนื่องจากทุกรายการ คำถามมุ่งวัดสิ่งเดียวกัน ดังนั้นการเลือกใช้สเกลประเภทนี้ จึงมีส่วนช่วยให้เครื่องมือที่สร้างขึ้นมีความตรง (validity) และ ความเที่ยง (reliability) สูง (จิราพร เขียวอยู่, 2552)

การกำหนดระดับของตัวเลือกของคำตอบมีตั้งแต่ 3-7 ระดับ ขึ้นอยู่กับผู้สร้าง ที่พบบ่อยคือ 4-5 ระดับ (Fayers and Machin, 2000)การกำหนดตัวเลือกของคำตอบจำนวน น้อย เช่น 3 ระดับ เครื่องมือจะมีความเที่ยงต่ำกว่ากรณี กำหนดตัวเลือกของคำตอบจำนวนมากเช่น 5 ระดับ ดังนั้น ควรกำหนดประมาณ 5-9 ระดับ (Spector, 1992)

การกำหนดรูปแบบสเกลเป็นสเกลของลิเคิร์ตนี้ มักเป็นทางเลือกในการสร้างเครื่องมือ ที่มีวัตถุประสงค์เพื่อ การประเมิน (evaluative instrument) หรือเพื่อใช้วัดการ เปลี่ยนแปลงของบุคคลหรือกลุ่มตัวอย่างในช่วงเวลาหนึ่ง (Fayers and Machin, 2000)

ดังนั้น จึงกล่าวสรุปได้ว่า การกำหนดประเภทและ รูปแบบสเกลของงานที่ทบทวนทั้งหมดเหมาะสม ซึ่งถึงแม้จะมี 2 งานที่กำหนดตัวเลือกของคำตอบเป็น 3 ระดับ ก็ไม่จัดว่าไม่

เหมาะสม เพราะใน 2 การศึกษานั้น งานหนึ่งศึกษาในผู้ป่วย หลังผ่าตัดเปิดกะโหลกศีรษะ อีกงานหนึ่งศึกษาในผู้ป่วยเด็ก ซึ่งกลุ่มเป้าหมายของทั้งสองงานนี้ หากใช้ตัวเลือกของคำตอบ มากกว่า 3 ระดับ ผู้ตอบก็อาจไม่สามารถแยกแยะคำตอบได้

4. การประเมินความตรงตามเนื้อหา (content validity) การประเมินความตรงตามเนื้อหาเป็นการพิจารณาว่าเครื่องมือมีเนื้อหาครอบคลุมทุกประเด็นของคุณสมบัติของตัวแปรที่ต้องการวัดหรือไม่ โดยเป็นเนื้อหาที่สอดคล้องกับแนวคิดและเป็นไปตามค่านิยมเชิงปฏิบัติการของตัวแปรนั้น จากการศึกษาครั้งนี้พบว่า มี 2/13 เรื่อง ไม่รายงานการประเมินความตรงด้านเนื้อหา ที่เหลือ 11/13 รายงานว่า ประเมินความตรงด้านเนื้อหา แต่รายละเอียดของวิธีการที่ใช้แตกต่างกัน โดยพบดังนี้ 3/13 เรื่อง รายงานว่าประเมินความตรงด้านเนื้อหา โดยนำเครื่องมือที่สร้างขึ้น ไปสัมภาษณ์กลุ่มเป้าหมาย เพื่อพิจารณาความเข้าใจภาษา หากกลุ่มเป้าหมายไม่เข้าใจข้อความใดก็จะตัดข้อความนั้นออกไป และพิจารณาจากความถี่ของการเลือกคำตอบของแต่ละข้อความ ถ้าตัวเลือกใดมีความถี่มากกว่าร้อยละ 90 จะตัดข้อความนั้นออกไป ซึ่งการทำเช่นนี้น่าจะเรียกว่าการทดสอบสเกลเบื้องต้น มากกว่าการประเมินความตรงด้านเนื้อหา, 8/13 เรื่อง รายงานว่าประเมินความตรงด้านเนื้อหา โดยใช้วิธีขอความเห็นจากผู้เชี่ยวชาญ กรณีเน้นวิธีการเชิงคุณภาพ คือ ให้ผู้เชี่ยวชาญให้ความเห็นในส่วนต่าง ๆ ของเครื่องมือ แล้วผู้วิจัยนำความเห็นของผู้เชี่ยวชาญมาปรับแก้เครื่องมือ การประเมินที่พบในลักษณะนี้ สอดคล้องกับผลการศึกษาของ Coste et al. (1997) ที่ทบทวนบทความจำนวน 42 บทความ พบว่า 2/42 ของบทความที่ทบทวน รายงานการประเมินความตรงตามเนื้อหาด้วยวิธีขอคำแนะนำจากผู้เชี่ยวชาญ ส่วนงานของ Hensley (1999) และ Coste et al. (1995) ก็พบว่าใช้วิธีการตรวจสอบเชิงคุณภาพโดยผู้เชี่ยวชาญเช่นกัน, และมีงานที่ทบทวน 3/13 เรื่อง รายงานว่าประเมินความตรงด้านเนื้อหา โดยใช้วิธีขอความเห็นจากผู้เชี่ยวชาญร่วมกับการคำนวณค่าดัชนีความตรงด้านเนื้อหา (content validity index: CVI) แล้วสรุปว่าค่า CVI เป็นไปตามเกณฑ์ของ Lynn เกณฑ์ของ Lynn (1986, อ้างใน Streiner and Norman, 2003) กำหนดดังนี้ เมื่อกำหนดระดับความเห็นด้านความตรงเชิงเนื้อหาเป็น 4 ระดับ 4 คือเห็นด้วยมากที่สุด 1 คือเห็นด้วยน้อยที่สุด หากให้ผู้เชี่ยวชาญ 5 คน ประเมินความตรงด้านเนื้อหา ผู้เชี่ยวชาญ

ทุกคนต้องให้ระดับการประเมินเป็น 3 หรือ 4 นั่นคือ CVI เท่ากับ 1 แต่ถ้าใช้ผู้เชี่ยวชาญ 6-10 คนประเมิน จำนวนผู้เชี่ยวชาญที่ประเมินระดับต่ำกว่า 3 มีได้เพียง 1 คน เท่านั้น นั่นคือ กรณีนี้ CVI มีค่าระหว่าง 0.83-0.90

สำหรับจำนวนผู้เชี่ยวชาญที่ใช้ในการตรวจสอบความตรงด้านเนื้อหา พบว่า มีจำนวนระหว่าง 3-9 คน โดยส่วนใหญ่ใช้ผู้เชี่ยวชาญ 5 คน

แม้จะพบว่างานที่ทบทวนส่วนใหญ่ มีการประเมินความตรงตามเนื้อหา แต่ก็มี 2 บทความวิจัยที่ไม่มีการรายงานในส่วนนี้ ซึ่งผู้ศึกษาไม่แน่ใจว่า การไม่ได้รายงาน เป็นเพราะไม่ได้ทำ หรือทำแต่ไม่ได้เขียนรายละเอียดลงในบทความวิจัย อย่างไรก็ตาม หากยึดตามสิ่งที่ปรากฏในบทความวิจัย ก็ถือว่าส่วนหนึ่งของงานที่ทบทวนยังไม่ได้ใช้ระเบียบวิธีมาตรฐานในการสร้างเครื่องมือ เพราะขั้นตอนนี้เป็นสิ่งที่จำเป็นต้องทำ

5. การทดสอบสเกลเบื้องต้น การทดสอบสเกลเบื้องต้นมีเป้าหมายเพื่อตรวจสอบความเข้าใจคำถามของกลุ่มเป้าหมาย (Fayers and Machin, 2000) และนำข้อมูลที่ได้ไปวิเคราะห์รายการคำถาม จากบทความวิจัยและวิทยานิพนธ์ที่ทบทวนทั้งหมด 13 เรื่อง พบ 12/13 เรื่อง รายงานการทดสอบสเกลเบื้องต้นจำนวน โดยนำไปทดสอบกับกลุ่มคนที่มีลักษณะเหมือนประชากรเป้าหมาย สอดคล้องกับการศึกษาที่ผ่านมาของ Hensley (1999) ที่ทบทวนบทความ 6 บทความ พบว่า 6/6 ของบทความที่ทบทวน รายงานการทดสอบสเกลเบื้องต้นกับกลุ่มตัวอย่างที่มีลักษณะคล้ายกับประชากรเป้าหมาย จาก 12 เรื่องของงานที่ทบทวนจากการวิจัยครั้งนี้ ที่รายงานการทดสอบสเกลเบื้องต้น มี 4/12 เรื่อง ใช้การสัมภาษณ์จากตัวอย่างของประชากรที่มีลักษณะเหมือนประชากรเป้าหมาย ขนาดตัวอย่างที่ใช้มีจำนวน 30, 47, 52 และ 60 คน ตามลำดับ และ 3/12 เรื่อง ใช้การตอบด้วยตนเอง ขนาดตัวอย่างที่ใช้มีจำนวน 6, 20, 30 คน ตามลำดับ ส่วนงานที่เหลือไม่ระบุวิธีการ

การทดสอบสเกลขั้นต้นที่มีเป้าหมายเพื่อตรวจสอบความเข้าใจภาษาในรายการคำถามของกลุ่มเป้าหมาย ต้องการขนาดตัวอย่างจำนวนไม่มาก คือประมาณ 10 – 30 คน (Fayers and Machin, 2000) พิจารณาเฉพาะส่วนนี้สรุปได้ว่างานส่วนใหญ่ (7/12 เรื่อง) ที่ทบทวนในการศึกษานี้ ใช้วิธีการที่เหมาะสม

6. การวิเคราะห์รายการคำถาม การวิเคราะห์รายการคำถามมีเป้าหมายเพื่อ ลดข้อคำถามให้เหลือเฉพาะที่จำเป็นเท่านั้น คือ ครอบคลุมเนื้อหา และไม่มากเกินไปจนต้องใช้เวลาในการตอบคำถาม การวิเคราะห์รายการคำถาม ทำให้หลายวิธีดังต่อไปนี้ (จิราพร เขียวอยู่, 2548)

(1) การพิจารณาการกระจายคำตอบของแต่ละข้อคำถาม (distribution of response) (2) วิเคราะห์ความสัมพันธ์ระหว่างแต่ละรายการคำถาม (inter-item correlation) (3) วิเคราะห์ความสัมพันธ์ระหว่างคะแนนของแต่ละข้อคำถามกับคะแนนรวม (item-total correlation) (4) วิเคราะห์ความสอดคล้องภายในระหว่างแต่ละรายการคำถาม (inter-item consistency) และ 5) การวิเคราะห์ปัจจัยเชิงสำรวจ (exploratory factor analysis) โดยวิธีที่ (2) และ (3) นั้นอาจจะไม่ดำเนินการก็ได้ ข้อสรุปเกี่ยวกับวิธีการวิเคราะห์รายการคำถามข้างต้น เป็นไปตามข้อเสนอแนะจากเอกสารตำราหลายรายการ ได้แก่ Streiner and Norman (2003) Spector (1992) Devellis (2003) และ Nunnally (1978)

ในการศึกษาครั้งนี้ เมื่อพิจารณาบทความวิจัยและวิทยานิพนธ์ ที่ทบทวนทั้งหมด 13 เรื่อง พบว่า 4/13 เรื่อง ไม่รายงาน ที่เหลือ 9/13 เรื่องรายงาน ใน 9 เรื่องที่รายงาน มีรายละเอียดดังนี้ 1 เรื่อง วิเคราะห์รายการคำถามทุกวิธีดังเสนอข้างต้น, 2 เรื่อง วิเคราะห์รายการคำถามทุกวิธีดังเสนอข้างต้น ยกเว้น การวิเคราะห์ความสัมพันธ์ระหว่างคะแนนของแต่ละข้อคำถามกับคะแนนรวม, 1 เรื่อง วิเคราะห์รายการคำถามทุกวิธีดังเสนอข้างต้น ยกเว้น การวิเคราะห์ความสัมพันธ์ระหว่างแต่ละรายการคำถาม, 1 เรื่อง การพิจารณาการกระจายคำตอบของแต่ละข้อคำถาม และ การวิเคราะห์ปัจจัยเชิงสำรวจ, 1 เรื่อง วิเคราะห์ความสอดคล้องภายในระหว่างแต่ละรายการคำถามและ การวิเคราะห์ปัจจัยเชิงสำรวจ, 2 เรื่อง การวิเคราะห์ปัจจัยเชิงสำรวจ, 1 เรื่อง ไม่รายงานวิธีการที่ใช้ จะเห็นว่า กรณีการวิเคราะห์ปัจจัย พบ 8/9 เรื่องใช้วิธีการนี้ และ 5/9 เรื่องใช้การวิเคราะห์ปัจจัยกับการวิเคราะห์ความสอดคล้องภายใน ซึ่งถือว่าการใช้ระเบียบวิธีที่เหมาะสม สถิติที่ใช้ประเมินความสอดคล้องภายใน คือ สัมประสิทธิ์แอลฟาของครอนบาค ค่าที่ได้อยู่ระหว่าง 0.74-0.98 สำหรับการพิจารณาการกระจายของคำตอบซึ่งเป็นกระบวนการเริ่มต้นของการวิเคราะห์รายการ

คำถามที่ทบทวนครานั้น มี 5/9 งานที่ดำเนินการ และพบว่า ใน 5 งานนี้ทำการวิเคราะห์ปัจจัยด้วย ดังนั้นจึงกล่าวได้ว่า งานที่ทบทวนในการศึกษาครั้งนี้ ใช้วิธีที่เหมาะสมสำหรับการวิเคราะห์รายการคำถาม อย่างไรก็ตามในที่นี้ผู้ศึกษาไม่ได้พิจารณาในประเด็นความถูกต้องของวิธีการวิเคราะห์ปัจจัย ผลการศึกษานี้สอดคล้องกับการศึกษาของ Hinkin (1995) ที่พบว่า วิธีการวิเคราะห์รายการคำถามที่ใช้ได้แก่ การประเมินความสอดคล้องภายในด้วยสถิติสัมประสิทธิ์แอลฟาของครอนบาค (ร้อยละ 97 (73/75) ของบทความที่ทบทวน) และการวิเคราะห์ปัจจัยเชิงยืนยัน (ร้อยละ 33 (25/75) ของบทความที่ทบทวน) และการศึกษาของ Hensley (1999) ที่พบว่า 6/6 บทความ รายงานการประเมินความสอดคล้องภายใน ด้วยสถิติสัมประสิทธิ์แอลฟาของครอนบาค และ 3/6 บทความ รายงานว่าใช้การวิเคราะห์ปัจจัยเชิงยืนยัน

7. การประเมินความตรงเชิงโครงสร้าง จากบทความวิจัยและวิทยานิพนธ์ที่ทบทวนทั้งหมด 13 เรื่อง พบเพียง 3/13 เรื่องที่รายงานความตรงเชิงโครงสร้าง เรื่องแรก ประเมินความตรงเชิงโครงสร้างทั้งประเภทความตรงเชิงลู่เข้า ความตรงเชิงจำแนก สถิติที่ใช้ได้แก่ สัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน และการวิเคราะห์ปัจจัยเชิงยืนยัน เรื่องที่สอง ประเมินความตรงเชิงโครงสร้างประเภทความตรงเชิงลู่เข้าและความตรงเชิงจำแนก แต่ไม่ระบุสถิติที่ใช้ เรื่องที่สาม ประเมินความตรงเชิงโครงสร้างประเภทความตรงเชิงลู่เข้า ประเมินโดยใช้เทคนิคการทดสอบในกลุ่มที่ทราบความแตกต่าง (Known Group method) สถิติที่ใช้คือ Mann-Whitney U test และใช้การวิเคราะห์ปัจจัยเชิงยืนยัน วิธีการที่ใช้ประเมินความตรงเชิงโครงสร้างในงานที่ทบทวนครั้งนี้ สอดคล้องกับวิธีการที่พบในการศึกษาของ Hinkin (1995), Hensley (1999), Coste et al. (1995) และ Coste et al. (1997) โดยในงานที่ Hinkin (1995) ทบทวน พบ 7/75 เรื่อง และ 6/75 เรื่อง ประเมินความตรงเชิงโครงสร้างประเภทความตรงเชิงลู่เข้าและความตรงเชิงจำแนก ตามลำดับ งานที่ Hensley (1999) ทบทวน พบอย่างละ 1/6 เรื่อง เท่ากัน ประเมินความตรงเชิงโครงสร้างประเภทความตรงเชิงลู่เข้าและ ความตรงเชิงจำแนก สำหรับการศึกษาของ Coste et al. (1995) และ Coste et al. (1997) พบการรายงานการประเมินความตรงเชิงโครงสร้างประเภทความตรงเชิงลู่เข้าในภาพรวม ดังนั้นจึงกล่าวได้ว่า ในงานที่ทบทวนครั้งนี้ในส่วนของการประเมินความตรงเชิงโครงสร้าง

นั้น ได้ใช้วิธีการที่เหมาะสมเพราะสอดคล้องกับข้อเสนอแนะของ Spector (1992)

8. การประเมินความตรงตามเกณฑ์สัมพัทธ์ ความตรงตามเกณฑ์สัมพัทธ์หมายถึง เครื่องมือวัดได้ตรงกับเกณฑ์หรือเครื่องมือวัดมาตรฐาน (gold standard) เมื่อพิจารณาจากบทความวิจัยและวิทยานิพนธ์ ที่ทบทวนทั้งหมด 13 เรื่อง พบว่ามีการรายงานประเด็นการประเมินความตรงตามเกณฑ์สัมพัทธ์ 2/13 เรื่อง ใน 2 เรื่องนี้รายงานการประเมินความตรงร่วมสมัย (concurrent validity) โดยมีเครื่องมือมาตรฐานในการเปรียบเทียบทั้ง 2 เรื่อง สถิติที่ใช้คือสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน สำหรับงานที่เหลือไม่ได้ประเมินความตรงตามเกณฑ์สัมพัทธ์ ทั้งนี้ น่าจะเป็นเพราะว่าเครื่องมือประเมินคุณภาพชีวิตโดยส่วนใหญ่ ไม่มีเครื่องมือมาตรฐานสำหรับการเปรียบเทียบ ดังนั้นจึงเป็นข้อจำกัดในการทดสอบความตรงตามเกณฑ์สัมพัทธ์ของเครื่องมือประเมินคุณภาพชีวิตที่เกี่ยวข้องกับสุขภาพที่สร้างใหม่ดังที่ Coons (2002) กล่าวไว้

9. การประเมินความเที่ยงของสเกล เครื่องมือวัดคุณภาพชีวิตจะประกอบคำถามหลาย ๆ ข้อ ดังนั้น ความเที่ยงของเครื่องมือในกรณีนี้ หมายถึง ความเที่ยงประเภทความสอดคล้องภายใน (internal consistency) จากบทความวิจัยและวิทยานิพนธ์ ที่ทบทวนทั้งหมด 13 เรื่อง พบว่า ทุกงานมีการรายงานค่าความเที่ยง ด้วยสัมประสิทธิ์แอลฟาของครอนบาค ซึ่งสอดคล้องกับการศึกษาของ Hinkin (1995), Hensley (1999), Coste et al. (1995) และ Coste et al. (1997) การรายงานค่าสัมประสิทธิ์แอลฟาของครอนบาคในงานที่ทบทวนครั้งนี้ มีการรายงานค่าในภาพรวม และรายด้านจำนวน 11/13 เรื่อง และ 9/13 เรื่อง ตามลำดับ ซึ่งเป็นไปตามคำแนะนำของ Spector (1992) ที่กล่าวไว้ว่าการหาความเที่ยงของเครื่องมือโดยคำนวณค่าสัมประสิทธิ์แอลฟา ควรคำนวณรวมทุกด้าน และรายด้าน ค่าสัมประสิทธิ์แอลฟาที่รายงานในงานที่ทบทวนครั้งนี้มีค่าระหว่าง 0.59 – 0.97 โดยมีสองเรื่องที่มีค่าสัมประสิทธิ์แอลฟาต่ำกว่า 0.7 ดังนั้นจึงกล่าวได้ว่า งานที่ทบทวนทั้งหมดได้ดำเนินการขั้นตอนนี้ซึ่งเป็นขั้นตอนที่จำเป็นต้องทำ และงานส่วนใหญ่ (11/13) ที่ทบทวนมีค่าความเที่ยงอยู่ในระดับที่ยอมรับ นั่นคือ ค่าสัมประสิทธิ์แอลฟาไม่ควรต่ำกว่า 0.7 แต่ไม่เกิน 0.9 (Nunnally, 1978 อ้างใน Streiner and Norman (2003)

10. การทดสอบคุณสมบัติความไวต่อการเปลี่ยนแปลง ความไวต่อการเปลี่ยนแปลงของเครื่องมือ (sensitive to change) คือความสามารถของเครื่องมือในการตรวจจับการเปลี่ยนแปลงของภาวะทางสุขภาพที่เกิดขึ้นในตัวผู้ถูกวัด (Maurice et al., 1998; Fayers and Machin, 2000) การประเมินคุณสมบัติประการนี้ของเครื่องมือ ยังระบุไม่ได้ว่าวิธีการคำนวณหรือสถิติใดดีที่สุด ดังนั้นจึงควรใช้หลายวิธีเพื่อยืนยันผล อีกทั้งควรพิจารณาความสัมพันธ์ของค่าความไวที่คำนวณได้กับมาตรฐานการรักษาอื่นๆ เช่น การเปลี่ยนแปลงทางคลินิก แล้วจึงพิจารณาความสามารถของเครื่องมือในการวัดความแตกต่างระหว่างกลุ่ม (Fayers and Machin, 2000; Streiner and Norman, 2003) อย่างไรก็ตามในหลายสถานการณ์ที่เป็นจริงของคนไข้ หรือด้านสุขภาพ ไม่เอื้อต่อการประเมินนี้ ดังนั้นจึงมักพบว่า ในการสร้างเครื่องมือวัดคุณภาพชีวิตที่เกี่ยวข้องกับสุขภาพ ไม่มีการรายงานการประเมินคุณสมบัตินี้ ซึ่งจากบทความวิจัยและวิทยานิพนธ์ที่ทบทวนทั้งหมด 13 เรื่อง ไม่พบการรายงานในประเด็นดังกล่าวเช่นกัน

11. การรายงานการนำไปใช้ จากบทความวิจัยและวิทยานิพนธ์ ที่ทบทวนทั้งหมด 13 เรื่อง พบ 6/13 เรื่อง รายงานการนำไปใช้จำนวน รายละเอียดที่ระบุคือวิธีคิดคะแนนและการแปลผล โดยพบเฉพาะในวิทยานิพนธ์เท่านั้น ซึ่ง Spector (1992) และพรรณทิพา ตักดีทอง (2550) กล่าวว่าเครื่องมือที่สร้างใหม่ควรมีคู่มือการใช้ มีวิธีการคิดคะแนนและการแปลผล และมีรายละเอียดของหน่วยงานหรือบุคคลที่สามารถติดต่อเพื่อขอเช่าเครื่องมือไปใช้ ดังนั้น จึงสรุปว่าในงานที่ทบทวนยังดำเนินงานส่วนนี้ไม่เป็นไปตามระเบียบวิธีมาตรฐานของการสร้างเครื่องมือ

ข้อเสนอแนะจากการวิจัย

ในการพิจารณาระเบียบวิธีวิจัยและวิธีการทางสถิติในการสร้างเครื่องมือวัดคุณภาพชีวิตที่เกี่ยวข้องกับสุขภาพของบทความวิจัยและวิทยานิพนธ์ด้านการแพทย์และสาธารณสุขของไทย ว่าเป็นไปตามมาตรฐานของการสร้างสเกลหรือไม่ พบว่า ในงานที่ทบทวนทั้งหมด 13 งาน ทุกงานมีขั้นตอนการสร้างเครื่องมือเป็นไปตามมาตรฐาน ยกเว้นกรณีการประเมินความไวของเครื่องมือ ซึ่งไม่พบในทุกงาน แต่กรณีนี้มักพบได้เสมอ เพราะมีข้อจำกัดสูง สำหรับในขั้นตอนอื่น ๆ ที่งานที่ทบทวนได้ดำเนินการ ความแตกต่าง

ของแต่ละงานอยู่ที่รายละเอียดของวิธีการที่ใช้ในแต่ละขั้นตอน บางงานดำเนินการอย่างครบถ้วน บางงานก็ดำเนินการเพียงบางส่วน ผู้วิจัยจึงมีข้อเสนอแนะแก่ผู้ต้องการสร้างเครื่องมือวัดคุณภาพชีวิตในอนาคตดังนี้

1. ในขั้นตอนของการกำหนดนิยามตัวแปรที่จะสร้างสเกลเพื่อวัด ไม่ควรใช้เพียงวิธีใดวิธีหนึ่ง แต่ควรใช้หลายวิธีร่วมกัน ได้แก่ การทบทวนวรรณกรรมประเภทแนวคิดทฤษฎีที่เกี่ยวข้องและงานวิจัยที่ผ่านมา ร่วมกับการสัมภาษณ์เชิงลึกจากผู้เชี่ยวชาญและการสัมภาษณ์เชิงลึกหรือสนทนากลุ่มจากกลุ่มเป้าหมาย

2. ควรมีการประเมินความตรงเชิงโครงสร้างเสมอ ซึ่งมีวิธีสถิติหลายวิธีที่สามารถใช้ได้ ไม่จำเป็นต้องใช้การวิเคราะห์ปัจจัยเชิงยืนยันเท่านั้น หากจะใช้การวิเคราะห์ปัจจัยเชิงยืนยันจำเป็นต้องคำนึงถึงขนาดตัวอย่างที่เพียงพอด้วย

3. กรณีประเมินความตรงตามเกณฑ์สัมพัทธ์ จำเป็นต้องมั่นใจว่า มีตัววัดที่เป็นเกณฑ์มาตรฐานในการเปรียบเทียบ

4. ในการเผยแพร่การพัฒนาเครื่องมือ ควรนำเสนอรายละเอียดของระเบียบวิธีและวิธีการทางสถิติที่ใช้ให้ครบถ้วนและชัดเจน เพื่อให้ผู้อ่านมีความมั่นใจในการนำไปใช้ต่อไป

กิตติกรรมประกาศ

ขอขอบคุณ รองศาสตราจารย์ พญ.สุวรรณา อรุณพงศ์ไพศาล รองศาสตราจารย์ ดร.นุจรี ประทีปะวนิช จอห์นส และผู้ช่วยศาสตราจารย์นิคม ถนอมเสียง ที่กรุณาตรวจสอบเครื่องมือ และให้ข้อเสนอแนะให้งานวิจัยมีความสมบูรณ์ยิ่งขึ้น

เอกสารอ้างอิง

- จิราพร เขียวอยู่. (2548). การออกแบบและการทดสอบเครื่องมือ. ขอนแก่น: ภาควิชาชีวสถิติและประชากรศาสตร์ คณะสาธารณสุขศาสตร์ มหาวิทยาลัยขอนแก่น.
- จิราพร เขียวอยู่. (2552). สัมประสิทธิ์แอลฟา. จดหมายข่าวชีวสถิติ มข. 1(1): 2-3.
- จิระภา รัตนกิจ. (2540). คุณภาพชีวิตของสตรีวัยหมดประจำเดือนในภาคใต้. วิทยานิพนธ์พยาบาลศาสตรมหาบัณฑิต สาขาการพยาบาลแม่และเด็ก บัณฑิตวิทยาลัย มหาวิทยาลัยมหิดล.
- พรรณทิพา ศักดิ์ทอง. (2550). คุณภาพชีวิตที่เกี่ยวกับสุขภาพ. ไทยเภสัชสาร, 2(3): 327-337.
- Bergner, M. (1989). Quality of life. Health status and clinical research. **Medical care**, 25(3): 148-156.
- Bowling, A. (2005). **Measuring health : a review of quality of life measurement scales**. 3rded. Maidenhead: Open University Press.
- Coste, J. and Fermanian, J., et al. (1995). Methodological and statistical problems in the construction of composite measurement scales: a survey of six medical and epidemiological journals. **Statistics in Medicine**, 14(4): 331-45.
- Coste, J. and Guillemin, F., et al. (1997). Methodological approaches to shortening composite measurement scales. **Journal of Clinical Epidemiology**, 50(3): 247-52.
- Coons, S.J.(2002). **Health outcome and quality of life**. 5thed. New York: McGraw-Hill.
- DeVellis, R.F. (2003). **Scale development: theory and application**. 2nded. Newbury Park, CA: Sage Publication.
- Fayers, P.M. and Machin, D. (2000). **Quality of Life: Assessment, analysis and interpretation**. West Sussex: John Wiley and Sons.
- Hensley, R.L. (1999). A review of operations management studies using scale development techniques. **Journal of Operations Management**. 17: 343-358.
- Hinkin, T.R. (1995). A review of scale development practices in the study of organizations. **Journal of Management**. 21(5): 967-988.
- Meeberg, G.A. (1993) Quality of life: A concept analysis. **Journal of Advance Nursing**. 18 (1) 32-38.
- Phattharayuttawat, S., Ngamthipwattana, T., and Pitiyawanun, B. (2005). The development of the Pictorial quality of life. **Journal of the Medical Association of Thailand**, 88 (11): 1605-18.
- Spector, P.E. (1992). **Summated rating scale construction: An introduction**. Sage University Paper Series on Quantitative Applications in the Social Sciences, Series No. 07-082. Newbury Park, CA: Sage Publication.

- Staquet, M.J., Hays R.D., and Fayers P.M. (1998) **Quality of life assessment in clinical trials: methods and practice**. Oxford: Oxford University: Press.
- Streiner, D.L. and Norman, G.R. (2003). **Health Measurement Scales. A Practical Guide to their Development and Use**. Oxford: Oxford University Press.
- Stromberg, M.F. (1984). Selective and instrument of measure quality of life. **Oncology Nursing Forum**, 11 (5):88-91.
- Wilkin D., Hallam L., and Doggett M.A. (1992). **Measure of need and outcome for primary health care**. New York: oxford University Press.
- World Health Organization (1993). **WHOQoL study protocol**. WHO (MNH7PSF/93.3).
- Yingyong, P. (2007). Evaluation of the Thai Low Vision Quality of Life Questionnaire (LVQOL). **Journal of the Medical Association of Thailand**, 90 (12): 2658-61.
- Zhan, L. (1992). **Quality of life: conceptual and measurement issues**. *Journal of Advanced Nursing*, 17(7): 795-800.

ตารางที่ 1 ลักษณะทั่วไปของบทความวิจัยและวิทยานิพนธ์ที่ศึกษา

ลักษณะทั่วไป	บทความวิจัย	วิทยานิพนธ์ระดับบัณฑิตศึกษา	รวม
	(n=3)	(n=10)	(n=13)
	จำนวน	จำนวน	จำนวน
วัตถุประสงค์การศึกษา			
- มีวัตถุประสงค์หลักเพื่อสร้างสเกลเพื่อวัดคุณภาพชีวิต	3	6	9
- มีวัตถุประสงค์หลักเพื่อประเมินคุณภาพชีวิตที่เกี่ยวข้องกับสุขภาพ แต่สร้างสเกลขึ้นใหม่	-	4	4
ตัววัดสุขภาพและประชากรเป้าหมาย			
- คุณภาพชีวิตผู้มีปัญหาด้านสายตา	1	1	2
- คุณภาพชีวิตผู้มีปัญหาทางจิต	1	-	1
- คุณภาพชีวิตผู้ป่วยโรคหลอดเลือดหัวใจตีบ	-	1	1
- คุณภาพชีวิตผู้ป่วยโรคหัวใจพิการแต่กำเนิด	-	1	1
- คุณภาพชีวิตผู้ป่วยหลังผ่าตัดเปิดกะโหลก	-	1	1
- คุณภาพชีวิตผู้ป่วยมะเร็งระบบสืบพันธุ์	-	1	1
- คุณภาพชีวิตผู้ป่วยเบาหวาน	-	1	1
- คุณภาพชีวิตผู้ป่วยไตวายเรื้อรัง	-	1	1
- คุณภาพชีวิตหญิงวัยหมดประจำเดือนที่มีภาวะกระดูกพรุน	-	1	1
- คุณภาพชีวิตในคนไทยที่เลิกบุหรี่	-	1	1
- คุณภาพชีวิตในผู้ป่วยเด็ก	-	1	1
- คุณภาพชีวิตในผู้ให้บริการทางเพศ	1	-	1

ตารางที่ 2 ระเบียบวิธีวิจัยและวิธีการทางสถิติที่พิจารณาที่ปรากฏในบทความวิจัยและวิทยานิพนธ์ที่ศึกษา

ระเบียบวิธีการสร้างสเกล/รายละเอียดที่พิจารณา	บทความวิจัย	วิทยานิพนธ์ระดับบัณฑิตศึกษา	รวม
	(n=3)	(n=10)	(n=13)
	จำนวน	จำนวน	จำนวน
การนิยามตัวแปรหรือกำหนดโครงสร้างเชิงเนื้อหาของสเกล			
ไม่รายงาน	-	-	-
รายงาน	3	10	13
- เขียนขึ้นใหม่โดยอาศัยแนวคิดหรือทฤษฎีที่เกี่ยวข้อง, ปรับปรุงจากงานวิจัยที่มีผู้ศึกษาไว้, นำบางส่วนของสเกลเดิมที่เกี่ยวข้องกับสเกลที่ต้องการสร้างเพื่อวัดคุณภาพชีวิต	-	1	1
- เขียนขึ้นใหม่โดยอาศัยแนวคิดหรือทฤษฎีที่เกี่ยวข้อง, ปรับปรุงจากงานวิจัยที่มีผู้ศึกษาไว้	2	7	9
- เขียนขึ้นใหม่โดยอาศัยแนวคิดหรือทฤษฎีที่เกี่ยวข้อง, สัมภาษณ์เชิงลึกโดยระดมสมองจากผู้เชี่ยวชาญหรือการสนทนากลุ่มจากกลุ่มเป้าหมาย	1	1	2
- เขียนขึ้นใหม่โดยอาศัยแนวคิดหรือทฤษฎีที่เกี่ยวข้อง, นำบางส่วนของสเกลเดิมที่เกี่ยวข้องกับสเกลที่ต้องการสร้างเพื่อวัดคุณภาพชีวิต	-	1	1
การรวบรวมข้อคำถาม			
ไม่รายงาน	-	-	-
รายงาน	3	10	13
- เขียนขึ้นใหม่โดยอาศัยแนวคิดหรือทฤษฎีที่เกี่ยวข้อง, ปรับปรุงจากงานวิจัยที่มีผู้ศึกษาไว้, และ/หรือสัมภาษณ์เชิงลึกโดยระดมสมองจากผู้เชี่ยวชาญหรือการสนทนากลุ่มจากกลุ่มเป้าหมาย	-	2	2
- เขียนขึ้นใหม่โดยอาศัยแนวคิดหรือทฤษฎีที่เกี่ยวข้อง, ปรับปรุงจากงานวิจัยที่มีผู้ศึกษาไว้	2	6	8
- เขียนขึ้นใหม่โดยอาศัยแนวคิดหรือทฤษฎีที่เกี่ยวข้อง, และ/หรือสัมภาษณ์เชิงลึกโดยระดมสมองจากผู้เชี่ยวชาญหรือการสนทนากลุ่มจากกลุ่มเป้าหมาย	1	2	3
จำนวนกลุ่มตัวอย่าง/ผู้ให้ข้อมูล (กลุ่ม)			
- 1 กลุ่ม	3	8	11
ขนาดตัวอย่าง/จำนวนผู้ให้ข้อมูล (ค่าต่ำสุด – ค่าสูงสุด)	30-160	10-28	10-160
-วิธีการเก็บข้อมูล			
-ให้กลุ่มเป้าหมายตอบเอง	1	-	1
-ให้กลุ่มเป้าหมายตอบเองและสัมภาษณ์โดยตรง	-	1	1
-สัมภาษณ์ความรู้สึกนึกคิดของผู้ป่วยโดยตรง	-	1	1
-สัมภาษณ์แบบเจาะลึก	-	2	2
-รวบรวมจากเครื่องมือที่มีอยู่เดิมแล้วเลือกคำถามบางข้อ	-	1	1
-จำนวนคำถามเริ่มต้น (ค่าต่ำสุด – ค่าสูงสุด)	18-30	22-187	18-187
-จำนวนคำถามเมื่อสิ้นสุด (ค่าต่ำสุด – ค่าสูงสุด)	18-22	35-83	18-83
-จำนวนองค์ประกอบ (ค่าต่ำสุด – ค่าสูงสุด)	4-6	4-7	4-7

ระเบียบวิธีการสร้างสเกล/รายละเอียดที่พิจารณา	บทความวิจัย (n=3)	วิทยานิพนธ์ระดับบัณฑิตศึกษา (n=10)	รวม (n=13)
	จำนวน	จำนวน	จำนวน
การกำหนดประเภทและรูปแบบสเกล			
ไม่รายงาน	-	-	-
รายงาน	3	10	13
-ลิเคิร์ต สเกล (Likert scale)	3	10	13
-คำตอบ 3 ระดับ	-	2	2
-คำตอบ 5 ระดับ	3	8	11
การประเมินความตรงตามเนื้อหา			
ไม่รายงาน	2	-	2
รายงาน	1	10	11
-ตรวจสอบจากผู้เชี่ยวชาญ	1	7	8
-ดัชนีความตรงเชิงเนื้อหา (CVI)	-	3	3
-นำไปทดสอบในกลุ่มเป้าหมาย	-	3	3
การทดสอบสเกลเบื้องต้น			
ไม่รายงาน	1	-	1
รายงาน	2	10	12
-วิธีการทดสอบ			
-สัมภาษณ์	1	4	5
-ตอบด้วยตนเอง	-	3	3
-ลักษณะกลุ่มตัวอย่างในการทดสอบ			
-กลุ่มตัวอย่างคล้ายกับประชากรเป้าหมาย	2	10	12
-ขนาดตัวอย่างในการทดสอบ (ค่าต่ำสุด – ค่าสูงสุด)	20-25	6-48	6-48
การวิเคราะห์รายการคำถาม			
ไม่รายงาน	-	4	4
รายงาน	3	6	9
- การกระจายคำตอบของตัวเลือก (distribution of response)	-	5	5
-ความสัมพันธ์ระหว่างข้อคำถาม (inter-item correlation)	-	3	3
-วิธีสถิติสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน	-	3	3
-ความสัมพันธ์ระหว่างข้อคำถามกับสเกล (item-total correlation)	-	2	2
-วิธีสถิติสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน	-	2	2
-ความสอดคล้องภายใน (internal consistency)	1	4	5
-รายงานสัมประสิทธิ์แอลฟาของครอนบาค			
-ภาพรวม	1	4	5
-รายด้าน	1	1	2
-การวิเคราะห์ปัจจัย (factor analysis)	3	5	8
- การวิเคราะห์ปัจจัยเชิงสำรวจ (Exploratory factor analysis)	3	5	8

ระเบียบวิธีการสร้างสเกล/รายละเอียดที่พิจารณา	บทความวิจัย	วิทยานิพนธ์ระดับบัณฑิตศึกษา	รวม
	(n=3)	(n=10)	(n=13)
	จำนวน	จำนวน	จำนวน
การประเมินความตรงตามโครงสร้าง			
ไม่รายงาน	1	7	8
รายงาน	2	3	5
-ความตรงเชิงลู่เข้า	-	2	2
-วิธีสถิติสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน	-	2	2
-ความตรงเชิงจำแนก	-	4	4
-วิธีสถิติสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน	2	2	4
-การวิเคราะห์ปัจจัย	-	1	1
-การวิเคราะห์ปัจจัยเชิงยืนยัน (Confirmatory factor analysis)	-	1	1
-การทดสอบในกลุ่มที่ทราบความแตกต่าง	-	1	1
- Mann-Whitney U test	-	1	1
การประเมินความตรงตามเกณฑ์สัมพัทธ์			
ไม่รายงาน	1	10	11
รายงาน	2	-	2
-ความตรงร่วมสมัย	2	-	2
-วิธีสถิติสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน	2	-	-
รายละเอียดรายงานความตรงร่วมสมัย			
-เครื่องมือมาตรฐานเพื่อเปรียบเทียบ			
-ไม่มี	-	-	-
- มี เช่น PTQL (เครื่องมือ) และ WHOQOL, เครื่องมือที่สร้างขึ้นกับ Health related self report scale (HRSR)	2	-	2
-ความตรงเชิงทำนาย	-	-	-
-วิธีสถิติ ใช้การวิเคราะห์ถดถอย	-	-	-
การประเมินความเที่ยง			
ไม่รายงาน	-	-	-
รายงาน	3	10	13
-ความสอดคล้องภายใน	3	10	13
-วิธีการแบ่งครึ่ง	-	-	-
-วิธีสัมประสิทธิ์แอลฟา	3	10	13
รายงานค่าสัมประสิทธิ์แอลฟา			
-ภาพรวม	2	9	11
-รายด้าน	3	6	9
-ความคงที่ของการวัด	-	1	1
-วิธีสถิติสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน	-	1	1

ระเบียบวิธีการสร้างสเกล/รายละเอียดที่พิจารณา	บทความวิจัย (n=3)	วิทยานิพนธ์ระดับบัณฑิตศึกษา (n=10)	รวม (n=13)
	จำนวน	จำนวน	จำนวน
การทดสอบคุณสมบัติความไวต่อการเปลี่ยนแปลง			
ไม่รายงาน	3	10	13
รายงาน	-	-	-
การรายงานการนำไปใช้			
ไม่รายงาน	3	4	7
รายงาน	-	6	6
-คู่มือการใช้	-	-	-
-วิธีการคิดคะแนนและการแปลผล	-	6	6
-รายละเอียดของหน่วยงานหรือบุคคลที่ใช้ติดต่อเพื่อขอ นำเครื่องมือไปใช้	-	-	-