วิวัฒนาการของช่วงความเชื่อมั่นสำหรับผลต่างค่าสัดส่วนทวินาม
Evolution of Confidence Interval for the Difference Binomial Proportion

นพวรรณ รื่นแสง
Noppawan Ruensaeng
โรงพยาบาลธรรมศาสตร์เฉลิมพระเกียรติ มหาวิทยาลัยธรรมศาสตร์
Thammasat University Hospital, Thammasat University

บทคัดย่อ

นักสถิติได้พัฒนาช่วงความเชื่อมั่นสำหรับผลต่างค่าสัดส่วนทวินามอย่างต่อเนื่อง เพื่อให้ได้ช่วงความเชื่อมั่นที่มีประสิทธิภาพทั้งความน่าจะเป็นครอบคลุมและความกว้างเฉลี่ย แล้วแต่ละวิธีการที่พัฒนาขึ้นนี้ย่อมมีจุดอ่อนในบางประเด็น บทความนี้ผู้เขียนได้ทบทวนแนวคิด วิธีการพัฒนาและงานวิจัยที่เกี่ยวข้องกับช่วงความเชื่อมั่นสำหรับผลต่างค่าสัดส่วนทวินาม จากการทบทวนวรรณกรรม พบว่า ช่วงความเชื่อมั่นจะมีประสิทธิภาพดีเมื่อขนาดตัวอย่างเหมาะสม และการพัฒนาวิธีช่วงความเชื่อมั่นควรมีทฤษฎีสนับสนุนที่ชัดเจนด้วย

คำสำคัญ : ช่วงความเชื่อมั่น ค่าสัดส่วนทวินาม ความน่าจะเป็นครอบคลุม

Abstract

Statisticians have developed confidence intervals for the difference binomial proportion continuously. In order to get confidence interval have efficiency both coverage probability and average length, then these methods developed found that confidence intervals have weaknesses. This article has reviewed the concepts, development methods and related research to confidence intervals for the difference binomial proportion. From a literature review found that confidence intervals have better performance base on size appropriately. Furthermore, the development confidence interval methods should have clearly support theory.

Keyword : confidence interval, binomial proportion, coverage probability

บทนำ

ขนาดตัวอย่างในงานวิจัยเชิงทดลองทั้งทางคลินิกหรือทางระบาดวิทยาถือว่าเป็นเรื่องสำคัญมาก และส่วนใหญ่ขนาดตัวอย่างที่ใช้ในวิจัยเชิงทดลองจะมีจำนวนไม่มากนัก หากขนาดตัวอย่างที่มีจำนวนน้อยเกินไปก็อาจทำให้ขาดความเป็นตัวแทนประชากรที่ดีและผลการวิจัยไม่อาจสามารถสรุปผลอ้างอิงไปสู่ประชากรได้ และเมื่อใช้กลุ่มตัวอย่างที่มีจำนวนมากกว่าที่คำนวณได้จากสูตรหาขนาดตัวอย่างหรือโปรแกรมคำนวณขนาดตัวอย่างแล้วก็อาจส่งผลต่อระยะเวลาในการวิจัยหรืองบประมาณของงานวิจัยนั้นด้วย ซึ่งส่วนใหญ่สถิติที่ใช้วิเคราะห์ข้อมูลในงานวิจัยทางคลินิกหรือทางระบาดวิทยาจะใช้ทั้งการทดสอบความแตกต่างของสัดส่วนและการทดสอบความแตกต่างของค่าเฉลี่ย [1] นอกจากนี้งานวิจัยทางคลินิกและทางระบาดวิทยาจะใช้ช่วงความเชื่อมั่นนี้ในการอธิบายผลการวิจัยและช่วงความเชื่อมั่นจะขึ้นอยู่กับจำนวนของขนาดตัวอย่างอีกด้วย [2]

Corresponding author. E-mail : noppawan.ruen@gmail.com

ช่วงความเชื่อมั่นก็มีหลายวิธีการให้นำไปใช้กับงานวิจัย จึงทำให้นักวิจัยต้องไตร่ตรองเป็นอย่างมากเกี่ยวกับช่วงความเชื่อมั่นแต่ละวิธีว่าวิธีการใดที่จะเหมาะสมกับงานวิจัยของตน ส่วนใหญ่นักวิจัยทั้งทางคลินิกหรือทางระบาดวิทยาได้นำช่วงความเชื่อมั่นสำหรับค่าสัดส่วนทวินาม ช่วงความเชื่อมั่นสำหรับผลต่างค่าสัดส่วนทวินามหรือช่วงความเชื่อสำหรับความเสี่ยงต่างมาใช้ในการวิเคราะห์ผล [3,4] สำหรับงานวิจัยทางคลินิกหรือทางระบาดวิทยาที่ศึกษาเกี่ยวกับการเกิดของเหตุการณ์หนึ่งที่สนใจของกลุ่มตัวอย่าง 2 กลุ่ม อย่างเช่น ศึกษาผลการรักษาโรคชนิดหนึ่งด้วยยาชนิดใหม่และศึกษาผลการรักษาหลังจากที่ได้รับยาชนิดใหม่ว่าผู้ป่วยอาการมีดีขึ้นหรือไม่และนักวิจัยต้องการศึกษาสัดส่วนของอาการที่ดีขึ้นของทั้ง 2 กลุ่ม ดังนั้นนักวิจัยจะเลือกช่วงความเชื่อมั่นสำหรับผลต่างค่าสัดส่วนทวินามในการวิเคราะห์ข้อมูลลักษณะอาการที่เกิดขึ้นหลังได้รับยาชนิดใหม่แล้ว จากที่ได้อธิบายมาข้างต้นนั้นต่อไปนี้จะขอกล่าวถึงในส่วนของช่วงความเชื่อมั่นสำหรับผลต่างค่าสัดส่วนทวินาม

แนวคิดช่วงความเชื่อมั่นสำหรับผลต่างค่าสัดส่วนทวินาม

ช่วงความเชื่อมั่นสำหรับผลต่างค่าสัดส่วนทวินามเป็นสถิติพื้นฐานและงานวิจัยทางการแพทย์นิยมนำมาใช้ใน การวิเคราะห์ข้อมูล โดยช่วงความเชื่อมั่นจะมีตัวแปรสุ่ม 2 ตัวแปรที่เป็นอิสระกัน ซึ่งตัวแปรสุ่มนี้ได้มาจากประชากรที่มีการแจกแจงแบบทวินาม โดยการแจกแจงทวินามนี้จะเป็นการอธิบายลักษณะเหตุการณ์ของผลลัพธ์ที่เกิดขึ้น ผลลัพธ์ที่ได้นั่นคือ ผลลัพธ์ทั้งที่ประสบความสำเร็จ (Success) และไม่ประสบความสำเร็จ (Failure) แล้วค่าความน่าจะเป็นที่เกิดผลสำเร็จแต่ละครั้งมีค่าเท่ากับ $\hat{p} = X_i / n_i$, $i = 1,2$ เรียกว่าค่าสัดส่วนทวินาม โดย X_i เป็นตัวแปรสุ่มของจำนวนครั้งการเกิดผลลัพธ์ที่สำเร็จในการทดลองแต่ละครั้ง และช่วงความเชื่อมั่นของค่าสัดส่วนทวินามจะมีค่าระหว่าง 0 ถึง 1 และ n_i เป็นขนาดตัวอย่างกลุ่มที่ i ซึ่ง i เป็นตัวแปรสุ่มที่ศึกษาของกลุ่มที่ 1 และกลุ่มที่ 2 ตามลำดับ [5] ซึ่งช่วงความเชื่อค่าสัดส่วนทวินามมาตรฐานที่รู้จักกันทั่วไป นั่นคือ ช่วงความเชื่อมั่นวิธี Wald ซึ่งส่วนใหญ่ผู้แต่งหนังสือสถิติเบื้องต้นหรือหนังสือชีวสถิติได้นำช่วงวิธีดังกล่าวเขียนไว้ในหนังสือเนื่องจากวิธีนี้ไม่ซับซ้อนจึงเหมาะแก่ผู้เรียนในวิชาสถิติหรือชีวสถิติได้ศึกษากันในเบื้องต้นก่อน [6, 7] และจากผลการศึกษาของนักวิจัยหลายคน พบว่า ช่วงความเชื่อมั่นวิธี Wald สามารถนำไปใช้วิเคราะห์ได้ง่าย เมื่อเปรียบเทียบค่าความน่าจะเป็นครอบคลุมระหว่างวิธี Wald กับวิธีอื่น ๆ แล้ว พบว่าเมื่อขนาดตัวอย่างน้อยกว่า 30 จะทำให้ค่าความน่าจะเป็นครอบคลุมจะมีประสิทธิภาพลดลง [8-11]

ช่วงความเชื่อมั่นวิธี Wald ถือได้ว่าเป็นวิธีการดั้งเดิมสำหรับการคำนวณช่วงความเชื่อมั่นทั้งค่าสัดส่วนทวินาม $(\hat{p} = x/n)$ และผลต่างค่าสัดส่วนทวินาม $(\hat{\Delta} = \hat{p}_1 - \hat{p}_2)$ และหลักการของวิธี Wald นี้ขึ้นอยู่กับทฤษฎี Asymptotic [7, 12] สำหรับจุดอ่อนของช่วงความเชื่อมั่นวิธี Wald ที่พบคือ ถ้ากลุ่มตัวอย่างขนาดเล็กจะทำให้ช่วงความเชื่อมั่นมีประสิทธิภาพน้อยและค่าความน่าจะเป็นครอบคลุม (Coverage Probability) จะมีค่าน้อยกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดไว้ (Nominal level) [8, 9] ด้วยเหตุผลนี้จึงทำให้นักสถิติหลายคนได้พัฒนาวิธีการหาช่วงความเชื่อมั่นขึ้นเพื่อให้ได้ช่วงความเชื่อมั่นที่มีประสิทธิภาพมากยิ่งขึ้น อย่างเช่น ช่วงความเชื่อมั่นวิธี Newcombe hybrid score วิธี Agresti – Caffo และวิธีปรับความเบ้ เป็นต้น

การพัฒนาช่วงความเชื่อมั่นสำหรับผลต่างค่าสัดส่วนทวินาม

ในช่วงเวลา 10 ที่ผ่านมานี้ได้มีนักสถิติหลายคนได้คิดค้นและพัฒนาช่วงความเชื่อมั่นสำหรับผลต่างค่าสัดส่วนทวินามกันต่อเนื่อง เพื่อปรับแก้จุดอ่อนต่างๆ แต่ละวิธีและเพื่อให้ได้ช่วงความเชื่อมั่นที่มีประสิทธิภาพทั้งค่าความน่าจะเป็นครอบคลุม (Coverage probability) และช่วงความกว้างที่แคบและด้วยเหตุที่นักสถิติพัฒนาช่วงความเชื่อมั่นวิธีการใหม่ ๆ นี้ เนื่องจากนักสถิติแต่ละคนได้พบข้อสังเกตว่าช่วงความเชื่อมั่นที่นักสถิติอื่นพัฒนาขึ้นมายังมีจุดอ่อนบางประการ และนักสถิติที่พัฒนาวิธีการขึ้นใหม่จะมีการเปรียบเทียบระหว่างวิธีใหม่กับวิธีที่มีมาก่อนหน้านี้ เพื่อเปรียบเทียบประสิทธิภาพและข้อดีข้อเสียของแต่ละวิธี ซึ่งช่วงความเชื่อมั่นสำหรับผลต่างค่าสัดส่วนทวินามมีหลายวิธีการด้วยกัน ได้แก่ 1) ช่วงความเชื่อมั่นวิธี Wald 2) ช่วงความเชื่อมั่นวิธี Tango 3) ช่วงความเชื่อมั่นวิธี Agresti และ Coull และ 4) ช่วงความเชื่อมั่นวิธี Newcombe hybrid score เป็นต้น ในที่นี้ขอเสนอช่วงความเชื่อมั่นที่ได้พัฒนาขึ้นประกอบด้วย 1) วิธี Newcombe hybrid score (NH) 2) วิธี Agresti – Caffo (AC) 3) วิธีปรับความเบ้ (Skewness-corrected confidence interval:SC) 4) วิธีปรับแก้ของ Bonett-Price (BP) และ 5) วิธี Kang ดังนี้

1) **ช่วงความเชื่อมั่นวิธี** Newcombe hybrid score (NH)

$$\left[\left(\hat{p}_1-\hat{p}_2\right)-\sqrt{\left(\hat{p}_1-l_1\right)^2+\left(u_2-\hat{p}_2\right)}\ ,\ \left(\hat{p}_1-\hat{p}_2\right)+\sqrt{\left(\hat{p}_2-l_2\right)^2+\left(u_1-\hat{p}_1\right)}\right] \quad (1)$$

โดยที่ l_1 , u_1 เป็นสัดส่วนขนาดตัวอย่างของกลุ่มที่ 1

l_2 , u_2 เป็นสัดส่วนขนาดตัวอย่างของกลุ่มที่ 2

$\hat{p}_1$ เป็นรากที่สองของ $\left|\pi_1 - a/m\right|$ จะได้ $z\sqrt{\left\{\pi_1\left(1-\pi_1\right)/m\right\}}$

$\hat{p}_2$ เป็นรากที่สองของ $\left|\pi_2 - b/n\right|$ จะได้ $z\sqrt{\left\{\pi_2\left(1-\pi_2\right)/n\right\}}$

ช่วงความเชื่อมั่นวิธี Newcombe hybrid score ได้พัฒนาขึ้นในปี 1998 โดย Newcombe, R.G. (ดังสมการ 1) [9] อธิบายว่า ช่วงความเชื่อมั่นวิธี Newcombe hybrid score ภายใต้ทฤษฎี score test สำหรับที่ช่วงความเชื่อมั่นนี้มีจุดเริ่มต้นในการพัฒนาจากช่วงความเชื่อมั่นวิธีของ Wilson และช่วงความเชื่อมั่นวิธี Newcombe hybrid score ถือว่า เป็นวิธีการหนึ่งที่รู้จักกันแพร่หลาย แต่ความแปรปรวนของขอบเขตบนและขอบเขตล่างของช่วงความเชื่อมั่นนี้ไม่มีทฤษฎีใดสนับสนุน [12,13,14] ด้วยข้อบกพร่องนี้จึงทำให้นักสถิติหรือนักชีวิสถิตินำช่วงความเชื่อมั่นวิธี Newcombe hybrid score ไปใช้วิเคราะห์ข้อมูลไม่มากนัก [14] นอกจากนี้ [14] สรุปได้ว่า หากข้อมูลที่ศึกษามีขนาดตัวอย่างเล็กหรือเท่ากันแล้วนั้น เมื่อขนาดตัวอย่างทั้ง 2 กลุ่มมีขนาดเข้าใกล้ประมาณ 20 ตัวอย่างจะทำให้ช่วงความเชื่อมั่นมีค่าความน่าจะเป็นครอบคลุม (Coverage probability) น้อยกว่า 92 – 95 % ที่กำหนดไว้ ช่วงความเชื่อมั่นวิธี Newcombe hybrid score สามารถใช้ได้กับกลุ่มตัวอย่างทั้งขนาดเท่ากันและไม่เท่ากัน แต่ถ้าขนาดตัวอย่างเท่ากันแล้วค่าความน่าจะเป็นครอบคลุม (Coverage probability) เข้าใกล้ค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดและมีช่วงความกว้างเฉลี่ยแคบด้วย [10,11] นอกจากนี้ [15] ได้เสนอแนะว่า ถ้าขนาดตัวอย่างทั้ง 2 กลุ่มมีขนาดเล็ก (n_1=n_2=10) ช่วงความเชื่อมั่นวิธี Newcombe hybrid score จะมีค่าความน่าจะเป็นครอบคลุม (Coverage probability) เข้าใกล้ค่าสัมประสิทธิ์ความเชื่อมั่นกำหนดไว้ที่ 95% (Nominal level) ถือว่า ช่วงที่ได้นี้มีประสิทธิภาพดี แต่อย่างไรก็ตามยังพบข้อบกพร่องนั่นคือช่วงความเชื่อมั่นที่ได้ยังมีค่าไม่เสถียรพอ นอกจากนี้ช่วงความเชื่อมั่นวิธี Tango's score interval มีประสิทธิภาพใกล้เคียง กับช่วงความเชื่อมั่นวิธี Newcombe hybrid score [7] เมื่อผลต่างค่าสัดส่วนทวินามน้อยกว่า 0.1 ดังนั้นสรุปได้ว่าช่วงความเชื่อมั่นวิธี Newcombe hybrid score มีประสิทธิภาพดีเช่นกันแต่ผลลัพธ์ค่าความน่าจะเป็นครอบคลุมที่ได้นั้นค่อนข้างไม่คงที่กับแต่ละสถานการณ์

2) **ช่วงความเชื่อมั่นวิธีของ** Agresti – Caffo

$$\hat{p}_1-\hat{p}_2 \pm z_{\alpha/2}\sqrt{\hat{p}_1\left(1-\hat{p}_1\right)/\left(n_1+2\right)+\hat{p}_2\left(1-\hat{p}_2\right)/\left(n_2+2\right)} \quad (2)$$

โดย [11] ได้พัฒนาช่วงความเชื่อมั่นจากแนวคิดวิธี Wald (ดังสมการ 2) กำหนดให้ค่า $\hat{p}_i$ เท่ากับ $\left(X_i+1\right)/\left(n_i+2\right)$; $i=1$, 2 โดยที่ $\hat{p}_i$ เป็นสัดส่วนขนาดตัวอย่างของกลุ่มที่ 1 และกลุ่มที่ 2 ตามลำดับ และช่วงความเชื่อมั่นวิธีการใหม่นี้มีวิธีที่ง่ายกว่าช่วงความเชื่อมั่นวิธี Newcombe hybrid score นอกจากนี้ [3] ได้เสนอว่า ช่วงความเชื่อมั่นวิธีนี้จะมีค่าความน่าจะเป็นครอบคลุมที่น้อยกว่า $1-\alpha$ สำหรับกรณีที่ขนาดตัวอย่างเล็กนั้นจะทำให้ช่วงความเชื่อมั่นวิธีการ Agresti – Caffo จะมีช่วงความกว้างที่แคบกว่าช่วงแท้จริง (Exact

Interval) แล้วช่วงความเชื่อมั่นวิธีนี้มีประสิทธิภาพทั้งความกว้างเฉลี่ย (Expected length) และความแปรปรวนของความกว้าง (Variance of length) ที่ใกล้เคียงกับวิธี Newcombe hybrid score นอกจากนี้ [8] ได้อธิบายเพิ่มเติมว่า ช่วงความเชื่อมั่นวิธี Agresti – Caffo มีจุดอ่อนด้วยกัน 2 ประเด็น ได้แก่ 1) ผลจากการศึกษาที่ได้จากการจำลองข้อมูลจะมีค่าความน่าจะเป็นครอบคลุม (Coverage probability) ที่มีความแม่นยำสูงแต่ความแม่นยำนี้ยังไม่มีทฤษฎีใดที่จะมา สนับสนุนได้ชัดเจน 2) การพัฒนาช่วงความเชื่อมั่น เมื่อกำหนดค่าระดับความเชื่อมั่นไว้ที่ 95 % ซึ่งการพัฒนานี้มีวัตถุประสงค์เพื่อเพิ่มประสิทธิภาพความแม่นยำของค่าความน่าจะเป็นครอบคลุม โดยมีวิธีการบวกขนาดตัวอย่างกลุ่มประสบความสำเร็จและกลุ่มที่ล้มเหลวเพิ่มด้วย 2 ทั้ง ($n_i=n_i+2$; $i=1,2$) เรียกวิธีนี้ว่าวิธี Adding 4 แต่เมื่อนำช่วงความเชื่อมั่นวิธี Agresti – Caffo ไปใช้กับค่าที่กำหนดระดับความเชื่อมั่นค่าอื่น ๆ (Nominal level) แล้วจะได้ว่า ค่าความน่าจะเป็นครอบคลุม (Coverage probability) มีความแม่นยำลดลง [4] ดังนั้นช่วงความเชื่อมั่นนี้จะมีประสิทธิภาพดีเมื่อนักวิจัยนำมาใช้และกำหนดระดับความเชื่อมั่นที่ 95 % และจากการเปรียบเทียบประสิทธิภาพของช่วงความเชื่อมั่นวิธีการนี้ของ [10] พบว่า ช่วงความเชื่อมั่นจะประสิทธิภาพดีก็ต่อเมื่อผลต่างค่าสัดส่วนทวินามไม่เท่ากับ 0 ทั้งขนาดตัวอย่าง $n_1=n_2 \leq 25$; $n_1=n_2>25$ และขนาดตัวอย่าง $n_1 \neq n_2 \leq 25$; $n_1 \neq n_2>25$

3) ช่วงความเชื่อมั่นวิธีปรับความเบ้

การพัฒนาช่วงความเชื่อมั่นวิธี Agresti – Caffo มีวัตถุประสงค์เพื่อให้ช่วงความเชื่อมั่นมีประสิทธิภาพสูงขึ้น แต่การพัฒนานี้ยังพบประเด็นที่สำคัญ นั่นคือ ลักษณะการแจกแจงตัวอย่างสำหรับผลต่างค่าสัดส่วนทวินามมีความเบ้สูง (Zou, A. & Donner, A., 2004) จากข้อบกพร่องดังกล่าว Zhou, X.-H. et al. (2004) ได้พัฒนาช่วงความเชื่อมั่นขึ้นเพื่อปรับความเบ้ (Skewness-corrected confidence interval) โดยใช้วิธีการ Edgeworth expansion ตามแนวคิดทฤษฎีของ Hall ในปี 1982 มาประยุกต์ในการพัฒนาช่วงความเชื่อมั่นดังสมการ 3

$$I_{1\alpha}=\begin{bmatrix}\hat{p}-\left(\dfrac{\hat{p}_1\hat{q}_1}{n_1}+\dfrac{\hat{p}_0\hat{q}_0}{n_2}\right)^{1/2}\left(z_{1-\alpha/2}-\dfrac{1}{\sqrt{n}}\hat{Q}\left(z_{1-\alpha/2}\right)\right),\\ \hat{p}-\left(\dfrac{\hat{p}_1\hat{q}_1}{n_1}+\dfrac{\hat{p}_0\hat{q}_0}{n_2}\right)^{1/2}\left(z_{\alpha/2}-\dfrac{1}{\sqrt{n}}\hat{Q}\left(z_{\alpha/2}\right)\right)\end{bmatrix} \tag{3}$$

โดยที่ $\hat{Q}(t)=\acute{\sigma}^{-1}\left(\hat{a}+\hat{b}t^2\right)$; $\hat{\sigma}=\left(\dfrac{n}{n_1}p_1q_1+\dfrac{n}{n_0}p_0q_0\right)^{1/2}$, $\hat{a}=\dfrac{\delta}{6\sigma^2}$, $\hat{b}=\dfrac{n\left(1-2p_1\right)}{2n_1}-\dfrac{\delta}{6\sigma^2}$,

$$\delta=\left(\frac{n}{n_1}\right)^2 p_1q_1\left(1-2p_1\right)-\left(\frac{n}{n_0}\right)^2 p_0q_0\left(1-2p_0\right)\ ,\ n=n_0+n_1\ ,\ \hat{p}=\hat{p}_1-\hat{p}_0$$

$\hat{p}_i=X_i/n_i$ เป็นสัดส่วนขนาดตัวอย่างของกลุ่มที่ 1 $(i=1)$ และกลุ่มที่ 2 $(i=0)$ ตามลำดับ

เมื่อผลศึกษาการจำลองข้อมูลเพื่อเปรียบเทียบประสิทธิภาพค่าความน่าจะเป็นครอบคลุม (Coverage probability) และความกว้างเฉลี่ย (Average length) ของแต่ละระดับความเชื่อมั่นที่กำหนด (Nominal level) ประกอบด้วย ระดับความเชื่อมั่นที่ 90% 95% และ 99% โดยจำแนกตามขนาดตัวอย่างทั้งที่มีขนาดกลุ่มตัวอย่างเท่ากัน (Balance) และขนาดกลุ่มตัวอย่างไม่เท่ากัน (Unbalance) พบว่า ช่วงความเชื่อมั่นที่พัฒนาขึ้นนี้มีค่าความน่าจะเป็นครอบคลุม (Coverage probability) ที่มีความแม่นยำสูงและความกว้างเฉลี่ยที่แคบ (Average length) เมื่อขนาดตัวอย่างทั้ง 2 กลุ่ม มีขนาดเท่ากับ 30 หากเปรียบเทียบค่าความน่าจะเป็นครอบคลุมระหว่างช่วงความเชื่อมั่นวิธีที่พัฒนาขึ้นกับช่วงความเชื่อวิธี Newcombe hybrid score และวิธี Agresti – Caffo ปรากฏว่า ค่าความน่าจะเป็นครอบคลุม (Coverage probability) ของทั้ง 3 วิธีมีความแม่นยำใกล้เคียงกัน แต่ความกว้างเฉลี่ย (Average length) ของช่วงความเชื่อมั่นวิธี Wald มีความกว้างเฉลี่ยแคบกว่าช่วงความเชื่อมั่นวิธีปรับความเบ้นี้ และ [8] พบว่า ช่วงความเชื่อมั่นวิธีปรับความเบ้ที่ได้พัฒนาขึ้นมานี้มีข้อจำกัดในการนำไปใช้ คือ ช่วงความเชื่อมั่นไม่เหมาะสมกับค่าสัดส่วนทวินาม $\hat{p}_0$ และ $\hat{p}_1$ มีค่าเข้าใกล้ 0 หรือ 1 (Boundary points)

4) **ช่วงความเชื่อมั่นวิธี** Kang [3]

จากประเด็นปัญหาที่ผ่านมาของช่วงความเชื่อมั่นวิธีอื่น คือ เมื่อขนาดตัวอย่างเล็กจะทำให้ค่าความน่าจะเป็นครอบคลุมมีค่าน้อยกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดไว้ (Nominal level) ด้วยประเด็นนี้เอง [3] ซึ่งได้ศึกษาการประมาณช่วงความเชื่อมั่นผลต่างระหว่างค่าสัดส่วน 2 กลุ่มที่มีความสัมพันธ์กันโดยจะพิจารณากลุ่มตัวอย่างที่มีขนาดเล็ก และพัฒนาช่วงความเชื่อมั่นจากทฤษฎี Edgeworth expansion และได้นำหลักการของ [8] มาเป็นแนวทางในการพัฒนาช่วงความเชื่อสำหรับผลต่างสัดส่วนของตัวอย่างข้อมูล binary ที่มีความสัมพันธ์กัน โดยมีการแจกแจงแบบ beta-binominal และกำหนดขนาดตัวอย่างทั้ง 2 กลุ่มเท่ากับ 10 ซึ่ง Edgeworth expansion มีบทบาทที่สำคัญในสถิติยุคใหม่และได้มีงานวิจัยสาขาสถิติหลายงานที่นำมาใช้ในการพัฒนาช่วงความเชื่อมั่น ซึ่งวิธีการที่ใช้มีความเหมาะสมที่นำมาใช้ในการปรับความเบ้ของข้อมูล [6,8,15] และค่าความน่าจะเป็นครอบคลุมของช่วงความเชื่อมั่นวิธี Kang นี้มีประสิทธิภาพมากกว่าช่วงความเชื่อมั่นวิธีเชิงเส้นกำกับ (Asymptotic) เมื่อผลต่างค่าสัดส่วนเท่ากับ 0.2 และ 0.3 และนอกจากนี้มีช่วงความกว้างที่แคบกว่าช่วงความเชื่อมั่นวิธีเชิงเส้นกำกับด้วย [3]

5) **ช่วงความเชื่อมั่นวิธีปรับแก้** Wald (Adjustment Wald confidence interval)

ช่วงความเชื่อมั่นวิธี Wald ที่ได้กล่าวไปในข้างต้นนั้น เป็นวิธีการที่นักวิจัยสามารถนำมาใช้วิเคราะห์ข้อมูลได้ง่าย แต่ช่วงความเชื่อมั่นสำหรับผลต่างค่าสัดส่วนทวินามนี้ก็ยังพบข้อบกพร่องหลัก ๆ นั่นคือ ช่วงความเชื่อมั่นวิธี Wald มีประสิทธิภาพยังไม่ดีและช่วงความเชื่อมั่นที่ได้ไม่มีความชัดเจนเมื่อขนาดตัวอย่างเล็ก [7] และ [17] ได้พัฒนาช่วงความเชื่อมั่นวิธี Wald ในรูปแบบฟังก์ชันเชิงเส้นตรงและวิธีการที่ได้นี้มีประสิทธิภาพดีกว่าช่วงความเชื่อมั่นวิธี Wald แต่อย่างไรก็ตาม [7] ก็ได้พัฒนาช่วงความเชื่อมั่นวิธี Wald อีกครั้ง โดยกำหนดให้ค่า $\boldsymbol{\pi}_{ij}$ เท่ากับ $(f_j+1)/(n+2)$ โดยที่ f_{ij} เป็นความถี่ของตารางการจรณ์แบบ 2x2 ดังสมการ (4)

$$\hat{\pi}_{21}-\hat{\pi}_{12} \pm z_{\alpha/2}\sqrt{\left[\hat{\pi}_{21}+\hat{\pi}_{12}-\left(\hat{\pi}_{21}-\hat{\pi}_{12}\right)^2\right]/(n+2)} \qquad (4)$$

จากการเปรียบเทียบประสิทธิภาพ ได้ว่า ช่วงความเชื่อมั่นวิธีปรับแก้ Wald มาตรฐานนี้จะมีช่วงความกว้างที่แคบเมื่อขนาดตัวอย่างเล็ก ($n_i<30, i=1,2$) และเมื่อระดับความเชื่อมั่น 99% ช่วงความกว้างเฉลี่ยของวิธีนี้น้อยกว่าช่วงความเชื่อมั่นวิธี Tango's score มากกว่าร้อยละ 5 แต่ถ้าระดับความเชื่อมั่น 90% ช่วงความเชื่อมั่นวิธี Tango's score มีความกว้างเฉลี่ยที่กว้างกว่าวิธีปรับแก้ Wald ถ้าหากกล่าวถึงระดับความเชื่อมั่นที่กำหนดไว้ 95% แล้วนั้น จะได้ว่า ช่วงความเชื่อมั่นวิธีปรับแก้ Wald กับช่วงความเชื่อมั่นวิธี Tango's score จะมีความกว้างเฉลี่ยที่ใกล้เคียงกัน

การพิจารณาประสิทธิภาพของช่วงความเชื่อ มั่นสำหรับผลต่างค่าสัดส่วนทวินาม

ช่วงความเชื่อมั่นวิธีต่าง ๆ ที่กล่าวมาประกอบด้วย วิธี Newcombe hybrid score วิธีปรับความเบ้ วิธีปรับแก้ของ Wald และวิธี Kang จะมีค่าความน่าจะเป็นครอบคลุมและความกว้างเฉลี่ยที่มีประสิทธิภาพ เมื่อขนาดตัวอย่างทั้ง 2 กลุ่มมีขนาดตัวอย่างเท่ากับ 30 หรือน้อยกว่า 30 และช่วงความเชื่อมั่นอีกวิธีหนึ่ง นั่นคือ วิธี Agresti – Caffo ที่มีความต่างจากวิธีอื่น คือ ค่าความน่าจะเป็นครอบคลุมและความกว้างเฉลี่ยมีประสิทธิภาพดี เมื่อขนาดตัวอย่างเท่ากันทั้ง 2 กลุ่มและมีขนาดตัวอย่างน้อยกว่า 30 ส่วนช่วงความเชื่อมั่นวิธี Kang จะมีความเหมาะสมกับงานวิจัยที่ศึกษากับกลุ่มตัวอย่างทั้ง 2 กลุ่มมีขนาดเท่ากับ 10 รายละเอียดจาก Table 1

Table 1. The efficiency of the coverage probability and expected length of confidence interval from diffent methods

ช่วงความเชื่อมั่น	ประสิทธิภาพ	
	ค่าความน่าจะเป็นครอบคลุม (Coverage probability)	**ความกว้างเฉลี่ย** (Average length)
วิธี Newcombe hybrid score (1998)	ค่าความน่าจะเป็นครอบคลุมเข้าใกล้ค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด เมื่อ $n_1=n_2 < 30$	ช่วงความกว้างเฉลี่ยจะแคบเมื่อ $n_1=n_2 < 30$
วิธี Agresti – Caffo [16]	ค่าความน่าจะเป็นครอบคลุมมีประสิทธิภาพดีเมื่อขนาดตัวอย่างคือ $n_1=n_2 \leq 25$; $n_1=n_2 > 25$ และ $n_1 \neq n_2 \leq 25$; $n_1 \neq n_2 > 25$ แต่ขนาดตัวอย่างไม่เกิน 30	ขนาดตัวอย่าง $n_1=n_2 = 30$ จะมีความกว้างเฉลี่ยแคบกว่าเมื่อขนาดตัวอย่าง $n_1 = 15, n_2 = 10$
วิธีปรับความเบ้ของ Zhou, X.H. [8]	ค่าความน่าจะเป็นครอบคลุมมีความแม่นยำน้อยเมื่อค่าสัดส่วนทวินามทั้ง 2 กลุ่มมีค่าเข้าใกล้ 0 หรือ 1 แต่จะมีความแม่นยำมากยิ่งขึ้นเมื่อ $n_1=n_2 = 30$	ความกว้างเฉลี่ยจะแคบเมื่อ $n_1=n_2 = 30$
วิธี Kang [3]	ค่าความน่าจะเป็นครอบคลุมที่ดีเมื่อ $n_1=n_2 = 10$ และค่าสัดส่วนทวินาม $\hat{p}_1 > 0.3$ และ $\hat{p}_2 > 0.2$	ช่วงความเชื่อมั่นจะมีความกว้างที่แคบกว่าวิธีเชิงเส้นกำกับ เมื่อ $n_1=n_2 = 10$
วิธีปรับแก้ Wald [7]	ค่าความน่าจะเป็นครอบคลุมจะเข้าใกล้ค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด 95% เมื่อ $n_1=n_2 < 30$	ช่วงความกว้างแคบเมื่อค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด 95% และ $n_1=n_2 < 25$

บทสรุป

จากการทบทวนงานวิจัยที่ผ่านมาทำให้ทราบว่าช่วงความเชื่อมั่นจะขึ้นกับขนาดตัวอย่าง ถ้าขนาดตัวอย่างมีความเหมาะสมจะมีส่วนที่ทำให้ช่วงความเชื่อมั่นสำหรับผลต่างค่าสัดส่วนทวินามมีประสิทธภาพดีด้วย สำหรับช่วงความเชื่อมั่นวิธี Newcombe hybrid score เป็นอีกวิธีการหนึ่งที่มีประสิทธิภาพดีแต่ในบางสถานการณ์ และอาจมีสาเหตุมาจากช่วงความเชื่อมั่นวิธีนี้ไม่มีทฤษฎีมาสนับสนุนอย่างชัดเจน แม้ว่ายังไม่สามารถอธิบายได้ชัดเจนว่าทฤษฎีใดที่สามารถทำให้ช่วงความเชื่อมั่นวิธี Newcombe hybrid score มีประสิทธิภาพดี ดังนั้นถ้ามีทฤษฎีมาสนับสนุนอย่างชัดเจนอาจทำให้ช่วงความเชื่อมั่นมีความน่าเชื่อถือในการวิเคราะห์ข้อมูล แล้วค่าความน่าจะเป็นครอบคลุม (Coverage probability) มีค่าเข้าใกล้ตามค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด (Nominal level) และความกว้างเฉลี่ยจะมีประสิทธิภาพมากยิ่งขึ้น

เอกสารอ้างอิง

1. กรมควบคุมโรค. การวิเคราะห์ความเสี่ยงทางสุขภาพสำหรับเจ้าหน้าที่สาธารณสุข. [online] เข้าถึงได้จาก http://beid.ddc.moph.go.th/th_old/images/doc_interest/doc_in_11.pdf. 2552.
2. Gebski, V.J. & Keech, A.C. 2003. Statistical methods in clinical trials. The Medical Journal of Australia. 178: 182-184.
3. Kang, S.-H., 2008. New confidence interval for the difference between two proportions in two-sample correlated binary data. Journal of the Korean Statistical Society. 37: 175-183.
4. Klingenberg, B. 2014. A new and improved confidence interval for the Mantel–Haenszel risk difference. Statist. Med. 33: 2968–2983.
5. Riffenburgh, R.H. 2012. confidence interval. In Academic Press (3rd ed.), Statistics in Medicine (pp. 137-155). doi:10.1016/B978-0-12-384864-2.00007-X.
6. Zhou, X.-H., Li, C.M., & Yang, Z. 2008. Improving interval estimation of binomial proportion. Philosphical Trancation of the Royal Society. 366: 2405-2418.
7. Bonett, DG., & Price, RM. 2012. Adjusted Wald confidence intervals for a difference of binomial proportions based on paired data. Journal of Educational and Behavioral Statistics. 37: 479-488.
8. Zhou, X.-H., Tsao, M., & Qin G. 2004. New intervals for the difference between two independent binomial proportions. Journal of Statistical Planning and Inference. 123: 97-115.
9. Zou, G. & Donner, A. 2004. A simple alternative confidence interval for the difference between two proportions. Controlled Clinical Trials. 25: 3-12.
10. Broun, L., & Li, X. 2005. Confidence intervals for two sample binomial distribution. Journal of Statistical Planning and Inference. 130: 359-375.
11. Reed III, JF. 2009. Improved confidence intervals for the difference between two proportions. Journal of Modern Applied Statistical Methods. 8: 207-214.
12. Zhou, X.-H., & Qin, G. 2005. A New confidence interval for the difference between two binomial proportions of paired data. Journal of Statistical Planning and Inference. 128: 527-542.
13. Fagerland, M.W., Lydersen, S. & Laake, P. 2011. Recommended confidence intervals for two independent binomial proportions. Statistical Methods in Medical Research. 0: 1-31.
14. Prendergast, L.A., & Staudte, R.G. 2014. Better than you think: interval estimators of the difference of binomial proportions. Journal of Statistical Planning and Inference. 148: 38-48.
15. Cojbasic, V., & Loncar, D. 2010. One-sided confidence intervals for population variances of skewed distributions. Journal of Statistical Planning and Inference. 141: 1667-1672.
16. Agresti, A., & Caffo, B. 2000. Simple and Effective Confidence Intervals for Proportions and Differences of Proportions Result from Adding Two Successes and Two Failures. American Statistical Association. 54: 280-288.
17. Price, R.M., & Bonett, D.G. 2004. An improved confidence interval for a linear function of binomial proportions. Computational Statistics & Data Analysis. 45: 449-456.