

การประเมินสถิติในบทความทางการแพทย์

สมรัตน์ เลิศมหาฤทธิ์*

Lertmaharit S. Statistical reviewing in Medical articles. Chula Med J 2002 Sep;46(9): 753 - 60

There has recently been increase in the use of statistics in medical research. Reporting results of statistical points in published paper should have been reviewed by statistician in the Editorial Board. Guidelines for statistics in critical appraisal are considered in the part of materials and methods, statistical analysis, statistical reporting, presentation of results, discussion and conclusion. The reader should know the scope of study and its appropriateness. It should be considered in the aspect of population and sample, measurement and response rate so that they can judge whether this result can be generalized to the target group. The statistical method should be appropriately used and related to the objective of the study. Statistical reporting is corresponded to the assumption, hypothesis and analysis specified in the study. In addition, discussion part should have been discussed the results in both clinical and statistical points. Finally, the reader should consider whether this result can answer the research question.

Key words: Statistical reviewing, Medical articles, Critical appraisal.

Reprint request : Lertmaharit S. Department of Preventive and Social Medicine, Faculty of Medicine, Chulalongkorn University, Bangkok 10330, Thailand.

Email : fmedslm@md2.md.chula.ac.th

Received for publication. June 15, 2002.

วัตถุประสงค์ : เพื่อต้องการให้ผู้อ่านทราบและเข้าใจถึง

1. ความสำคัญของสถิติในการอ่านบทความทางการแพทย์
2. แนวทางในการประเมินคุณค่าด้านสถิติของบทความทางการแพทย์
3. การรายงานผลด้านสถิติที่เหมาะสมและถูกต้อง
4. การนำเสนอที่เหมาะสมและข้อผิดพลาดที่พบได้บ่อยทางด้านสถิติ
5. การนำผลการวิจัยไปใช้ต่อไป

การอ่านบทความที่ตีพิมพ์ในวารสารทางการแพทย์มีความสำคัญและเป็นแนวทางนำไปสู่การคิดค้นวิจัยสิ่งใหม่ ๆ ต่อไป นอกจากนี้ยังช่วยให้ผู้อ่านได้ติดตามความก้าวหน้าในด้านนั้น ๆ อย่างไรก็ตามการอ่านบทความนั้นก็ยังคงต้องใช้วิจารณญาณในการประเมินคุณค่าของบทความนั้น ๆ ว่ามีความน่าเชื่อถือแค่ไหน การประเมินคุณค่าของบทความในด้านสถิตินั้นมีความสำคัญมาก เพราะการนำเอาผลที่ได้ไปใช้นั้นต้องมีความระมัดระวังและแน่ใจว่าถูกต้อง

การนำสถิติไปใช้ในงานวิจัยทางการแพทย์ได้มีมานานแล้ว เมื่อ 40 ปีก่อน แต่สถิติที่ใช้เป็นเพียงสถิติขั้นพื้นฐาน เช่น การใช้ t-test, Chi-square test, Correlation ฯลฯ ซึ่งไม่ได้มีความสลับซับซ้อนมากนัก แต่เมื่อได้มีการทบทวนเปรียบเทียบบทความที่ลงตีพิมพ์ในวารสารระหว่างปี ค.ศ. 1978 -1979 และปี ค.ศ. 1990 จากวารสาร The New England Journal of Medicine พบว่าได้มีการเปลี่ยนแปลงการใช้วิธีการทางสถิติจากวิธีพื้นฐานมาเป็นวิธีที่ซับซ้อนขึ้น เช่น ได้มีการนำการวิเคราะห์ด้วยวิธี Logistic Regression, Proportional Hazard Regression สำหรับ Survival data ทำให้ผลที่ได้มีความน่าเชื่อถือมากขึ้น และแนวโน้มจะมีเพิ่มขึ้นอีก⁽¹⁾

ได้มีการสำรวจการใช้สถิติในงานวิจัยที่ตีพิมพ์แล้วพบว่าประมาณ 50 % ของบทความทางการแพทย์ที่สำรวจนั้นใช้สถิติได้อย่างถูกต้องเหมาะสม เนื่องจากการทบทวนการใช้สถิติอย่างจริงจังมักจะทำเฉพาะในการวิจัยที่เป็นด้าน Clinical Trial เท่านั้น นอกจากนี้ผู้วิจัยก็ยังคงขาดความรู้พื้นฐานทางสถิติ อีกทั้งในกลุ่มทบทวนยังไม่มีนักสถิติอยู่ด้วย ดังนั้นข้อเสนอแนะสำหรับวารสารต่าง ๆ ที่ลงตีพิมพ์บทความนั้นควรจะ⁽¹⁾

- 1) มีการทบทวนด้านสถิติ
- 2) มีนักสถิติที่มีคุณภาพเป็นผู้ทบทวน
- 3) ผู้ทบทวนควรจะดูบทความที่ถูกแก้ไขแล้ว
- 4) ควรมีการบอกถึงนโยบายการตัดสินใจในแง่ของสถิติ
- 5) ควรมีการบอกถึงแนวทางการรายงานผลทางด้านสถิติด้วย

แนวทางการประเมินคุณค่าด้านสถิติ

หลังจากได้ประเมินคุณค่าของบทความในด้านการตั้งคำถามการวิจัย การตั้งสมมติฐาน ว่ามีความเหมาะสมเพียงใดแล้ว การประเมินด้านสถิติจะเริ่ม มีส่วนเกี่ยวข้องในส่วนต่าง ๆ ของบทความ ในที่นี้จะกล่าวถึงในแต่ละส่วนดังนี้

1. ด้านวัสดุและวิธีการ (Materials and Methods)

ในส่วนนี้ประกอบด้วย

1.1 การกำหนดประชากร ในบทความนั้นควรจะต้องระบุให้ชัดเจนว่าประชากรเป้าหมาย (Target population) นั้นคืออะไร และประชากรที่จะสุ่มตัวอย่างนั้น (Sampled population) มีความเหมาะสมแค่ไหน เพื่อจะได้พิจารณาว่าการศึกษานั้นสามารถเอาไปขยายผลได้ดีแค่ไหน คือ มี Generalizability นั้นเอง

1.2 วิธีการสุ่มตัวอย่าง (Sampling Technique) ควรพิจารณาว่าในบทความนั้นมีการสุ่มตัวอย่างด้วยวิธีใด และวิธีที่ใช้นั้นมีอคติในการสุ่ม (Sampling bias) หรือไม่ ถ้าวิธีการสุ่มไม่เหมาะสมหรือมีอคติจะทำให้ผลที่ได้ก็ไม่เหมาะสมที่จะนำไปสรุปผล

1.3 การคำนวณขนาดตัวอย่าง (Sample size calculation) ในบางบทความผู้วิจัยไม่ได้กล่าวถึงว่าทำไมใช้ขนาดตัวอย่างเท่านี้ บทความที่ตีพิมพ์ควรมีการกล่าวถึงการคำนวณขนาดตัวอย่างอย่างนั้น ๆ โดยอาจจะไปถึงสูตรที่ใช้ในเอกสารอ้างอิง และบอกถึงค่าที่สำคัญที่ใช้ในการคำนวณ เช่น ตัววัดที่ใช้คืออะไร กำหนดขนาดความแตกต่างไว้เท่าไร และยอมให้มีอำนาจ (Power) ของการทดสอบเป็นเท่าไร คือ กำหนดว่าโอกาสที่พบความแตกต่างนั้นได้เป็นเท่าไร จะช่วยให้ผู้อ่านประเมินได้ว่า การศึกษานั้นมีความน่าเชื่อถือแค่ไหนและจำนวนตัวอย่างเพียงพอที่จะค้นหาความแตกต่างหรือไม่

1.4 การเก็บรวบรวมข้อมูล (Data Collection) ควรมีการระบุว่ามีแหล่งที่มาของข้อมูลอย่างไร เช่น ถ้าใช้แบบสอบถามก็ควรระบุว่า ใช้ถามแบบตัวต่อตัว หรือให้ตอบด้วยตนเอง ถ้าแหล่งข้อมูลเป็นแบบทุติยภูมิ ก็ควรบอกว่า ได้มาจากไหน เช่น เก็บจากเวชระเบียนของผู้ป่วย เป็นต้น

1.5 การวัด (Measurement) ผู้วิจัยควรมีการรายงานที่ใช้เครื่องมือชนิดใดในการเก็บข้อมูล และได้มีการระบุหรือไม่ได้มีการทดสอบ Reliability และ Validity ของเครื่องมือแล้ว นอกจากนี้ตัววัดต่าง ๆ ควรบอกให้ชัดเจน เช่น บอกว่าอะไรเป็น Study factor และอะไรเป็น Outcome factor นอกจากนี้การวัดนั้น ๆ มีอคติเกิดขึ้นหรือไม่

1.6 อัตราการตอบกลับ (Response rate) ถ้าการศึกษานั้นต้องมีการติดตามผล เช่น ให้ส่งแบบสอบถามคืนมา อัตราการตอบกลับควรจะมีมากพอ โดยทั่วไปไม่ควรต่ำกว่า 80 % มิฉะนั้นจะทำให้มีอคติจากการตอบกลับนี้ได้ หรือถ้าการศึกษาที่มีการทดลองและต้องติดตามผล จะต้องคำนวณขนาดตัวอย่างเมื่อไว้ในตอนเริ่มต้นเพื่อจะได้มีครบในตอนสิ้นสุดการศึกษา

1.7 ค่าที่สูญหาย (Missing value) ในการวิเคราะห์ข้อมูลนั้น บางรายอาจมีข้อมูลไม่ครบทุกตัวแปร ในกรณีทีวิเคราะห์ทีละตัวแปร (Univariate analysis) อาจไม่มีปัญหามากนักแต่ถ้าวิเคราะห์เชิงพหุ (Multivariable analysis) แล้ว อาจเกิดปัญหาคือจะได้จำนวน Case น้อยลง เพราะต้องเลือกเฉพาะ Case ที่มีข้อมูลครบทุกตัวแปรที่จะวิเคราะห์ ถ้าตัวแปรที่เป็น Outcome มีค่าสูญหายไม่มากก็อาจจะไม่มีผลในการวิเคราะห์ แต่ถ้า outcome มีค่าสูญหายมากก็อาจมีผลในการวิเคราะห์ด้วย จึงต้องมีการตรวจสอบว่ากลุ่มที่มี Outcome กับกลุ่มที่ไม่ทราบ Outcome มีความแตกต่างกันหรือไม่ ถ้าไม่แตกต่างกันก็แสดงว่าการสูญหายนั้นเป็นไปอย่างสุ่ม (Missing at random) การมีค่าสูญหายมากเกินไปทำให้การศึกษานั้นขาดความน่าเชื่อถือทางสถิติได้

2. การวิเคราะห์ทางสถิติ (Statistical analysis)

การประเมินว่าสถิติที่ใช้ในการวิเคราะห์นั้นเหมาะสมหรือไม่นั้น จะขึ้นอยู่กับวัตถุประสงค์ของการวิจัย วิธีวิเคราะห์จะต้องสามารถตอบวัตถุประสงค์ได้ ดังนั้นผู้อ่านควรจะทราบถึงหลักวิธีการเลือกใช้สถิติเพื่อจะได้ประเมินว่าผู้วิจัยใช้ได้เหมาะสมถูกต้องหรือไม่ ในที่นี้จะกล่าวถึงสถิติที่ใช้ในการเขียนบทความให้เหมาะสม จะ

กล่าวโดยย่อ ๆ เพื่อผู้อ่านจะได้นำไปประเมินคุณค่าบทความนั้น

แนวทางในการใช้สถิติที่จะกล่าวถึงนี้ จะช่วยให้ผู้ที่เขียนบทความทราบว่าอะไรที่สำคัญในทางสถิติและจะนำเสนออย่างไร นอกจากนี้จะทำให้ผู้อ่านได้ทราบว่าการวิเคราะห์ผลนั้นถูกต้องเหมาะสมแล้วหรือไม่ วิธีการต่าง ๆ ที่ใช้ในการวิเคราะห์จะต้องแจ่มแจ้งให้ชัดเจนว่าใช้วิธีอะไรบ้าง วิธีการที่ใช้กันทั่วไป เช่น Chi-Square test, Correlation, etc. นั้นไม่จำเป็นต้องบอกรายละเอียดแต่วิธีการที่มีให้เลือกมากกว่าหนึ่งแบบเช่น t-test นั้นต้องระบุว่าเป็นแบบ Paired t-test หรือ Unpaired t-test การวิเคราะห์ที่มีความซับซ้อนต้องมีการอธิบายให้ทราบและควรมีการบอกเอกสารอ้างอิงที่เชื่อถือได้ การบอกที่มาจะทำให้ผู้อ่านสามารถไปค้นคว้าศึกษาเพิ่มเติมได้ การวิเคราะห์ข้อมูลด้วยคอมพิวเตอร์นั้นผู้เขียนควรระบุด้วยว่าใช้โปรแกรมสำเร็จรูปชนิดใดในการวิเคราะห์

จากบทความนี้จะทำให้เราได้ทราบว่า การศึกษานี้เป็นการศึกษาชนิดใด เก็บข้อมูลที่ไหนมีการเลือกตัวอย่างเข้ามาศึกษาอย่างไร นอกจากนี้ยังได้บอกถึงตัววัดผลชนิดหลักและรองการคำนวณขนาดตัวอย่างบอกอย่างสั้น ๆ เพื่อให้ผู้อ่านได้ทราบขนาดของความแตกต่างระหว่างผลที่ได้และ Power ที่ใช้ในการคำนวณ พร้อมทั้งเอกสารอ้างอิงในการนำเสนอใช้คำนวณซึ่งผู้อ่านสามารถไปค้นคว้าเพิ่มเติมได้ และสุดท้ายได้บอกถึงวิธีการทางสถิติที่ใช้ในการวิเคราะห์นี้ด้วย

3. การรายงานผลทางสถิติ (Statistical reporting)⁽²⁾

3.1 รายละเอียดที่บรรยาย (Descriptive information) การบรรยายลักษณะของข้อมูลที่ศึกษานั้นจะมาก่อนเริ่มทำการวิเคราะห์ โดยบรรยายบอกให้ได้รายละเอียดมากที่สุด ซึ่งอาจนำเสนอในรูปของกราฟหรือค่าสถิติ เช่น ข้อมูลชนิดต่อเนื่องก็จะสรุปในรูปของค่าเฉลี่ย, SD หรือมัธมฐาน, เปอร์เซ็นไทล์, interquartile range ($Q_3 - Q_1$) แต่ถ้าข้อมูลเป็นแบบลำดับมาตรา ควรนำเสนอเป็นสัดส่วน นอกจากนี้ควรมีการเปรียบเทียบการแจกแจงของข้อมูลพื้นฐานในแต่ละกลุ่มที่ศึกษา

ตัวอย่างที่ 1 เป็นตัวอย่างของการเขียนในส่วนวิธีการศึกษา โดยขีดบางตอนจากบทความของ Kamolratanakul P., Sawert H., Lertmaharit S., et.al. *Randomized controlled trial of directly observed treatment (DOT) for patients with pulmonary tuberculosis in Thailand. Transactions of the royal society of tropical Medicine and Hygiene (1999) 93, 552-557.*

Methods

Protocol

This study was a randomized controlled trial conducted at 8 district hospitals, 3 provincial hospitals and 4 referral centres ('Zonal TB Centres') of the government health system in Thailand. To obtain a representative sample of all tuberculosis patients in Thailand, health care facilities at all levels and in all areas of the country were included in the study. Four referral centres ('Zonal TB Centres') were randomly selected from a total of 12 centres, which were stratified to cover the country's 4 geographical regions (Central, Southern, Northern and Northeastern). Similarly, 4 provincial hospitals in each of the country's 4 geographical regions. One of the 4 initially selected provincial hospitals was later excluded from the study because patient allocation procedures were not followed according to the study protocol. In each province where a selected provincial hospital was located, 2 district hospitals were randomly selected among an average number of 10 district hospitals per province.

The primary outcome measure was the cure rate, defined as the percentage of patients who completed 6 months of treatment and had 2 negative sputum examinations during treatment, of which 1 was at the end of treatment. Prior to the study, the estimate of the achievable cure rate under DOT was 85%. Sample size calculations were based on a desired ability to detect a difference in cure rates of 10% between study arms with a power of 90%. The required sample size was 336 patients in both arms (LEMESHOW et al., 1990). The size of the actual groups enrolled in the study was greater.

Secondary outcome measures were treatment completion and sputum conversion rates. The treatment completion rate was defined as the percentage of patients who completed 6 months of treatment but did not receive 2 sputum examinations during treatment. The sputum conversion rate was defined as the percentage of patients whose sputum was negative for AFB at the end of the 3rd month of treatment. Further secondary outcome measures were the proportion of patients who defaulted from treatment, died, or were transferred to different treatment locations. In addition, a case holding proportion was defined as the percentage of patients who completed treatment divided by the total number of patients excluding those who died or were transferred to other locations during treatment.

Statistical analyses were performed on an intention-to-treat basis, using the χ^2 test to assess differences between outcome proportions in the 2 study groups.

3.2 ข้อสมมติ (Underlying assumptions) การวิเคราะห์แต่ละวิธีการจะทำภายใต้ข้อสมมติของการวิเคราะห์นั้น ๆ เช่น การทดสอบวิธีต่าง ๆ นั้นมักจะมีข้อสมมติว่าข้อมูลนั้นมาจากประชากรที่แจกแจงแบบปกติ (Normal distribution) ถ้าไม่เป็นไปตามนั้นจะต้องใช้วิธีทางสถิติชนิด Nonparametric แทน โดยทั่วไปข้อสมมตินี้ไม่ได้เข้มงวดมากนักอาจไม่ต้องทำการทดสอบให้ยุ่งยาก แต่ในบางกรณีที่มีข้อมูลมีจำนวนไม่มากนักก็อาจต้องทดสอบว่าเป็นไปตามข้อสมมตินั้นหรือไม่

3.3 การทดสอบสมมติฐาน (Hypothesis tests) การทดสอบสมมติฐานที่มักจะใช้เป็นแบบสองทาง (two-sided test) ถ้ามีการทดสอบแบบทางเดียวก็จะต้องมีการบอกถึงเหตุผลและความเหมาะสม ส่วนผลการทดสอบนอกจากจะรายงานค่า p-value แล้วควรจะมีการแสดงค่าช่วงความเชื่อมั่น (Confidence interval) ไว้ด้วย^(3,4) เพื่อให้ผู้อ่านได้ทราบถึงผลของความแตกต่างที่แท้จริงและแสดงถึงความแม่นยำของการประมาณค่านั้นด้วยเพราะช่วงที่กว้างเกินไปทำให้ความเชื่อถือน้อยลง นอกจากนี้ผลการทดสอบแบบสองทางจะให้ผลสอดคล้องกับค่าแสดงช่วงความเชื่อมั่นด้วย ดังนั้นการศึกษาที่มีการเปรียบเทียบผลนั้นควรจะรายงานค่าช่วงความเชื่อมั่นของผลต่างที่ได้ไว้ด้วย

3.4 การวัดค่าสังเกต (Observations) ในการเปรียบเทียบแต่ละกลุ่มนั้นต้องประเมินว่าผู้ศึกษาได้ทำการศึกษาข้อมูลในแต่ละคนอย่างไร ถ้ามีการวัดผลในคนคนเดียวกัน ข้อมูลจะเกี่ยวข้องกันจึงจัดเป็น Paired observations การทดสอบจึงต้องใช้วิธีการทางสถิติที่เหมาะสมเช่น Paired t-test หรือ Wilcoxon-Signed Rank test หรือ Mc Nemar's test แล้วแต่กรณี ส่วนการวัดซ้ำที่มีตั้งแต่ 2 ครั้งขึ้นไปในคนคนเดียวกันจะต้องนำมาพิจารณาและวิเคราะห์เป็นแบบวัดซ้ำกัน (Repeated measurement) ถ้ามีการวัดซ้ำกันแต่ผู้วิจัยใช้ข้อมูลที่จุดใดจุดหนึ่ง เช่น วัดค่าความดันโลหิตหลายช่วงเวลาแต่เอาค่าสูงสุดหรือต่ำสุดมาวิเคราะห์จะทำให้ผลการศึกษามีผิดพลาดได้

3.5 การแปลงข้อมูล (Data transformation) ข้อมูลทางด้านการแพทย์ส่วนมากมักจะเบี่ยงไปทางขวาเนื่องจากมักจะมีค่าที่สูงมากเกินไปการแปลงข้อมูลที่มักจะใช้กันบ่อย คือ การทำเป็นค่า logarithm ดังนั้นเมื่อคำนวณค่าเฉลี่ยของค่าข้อมูลที่เป็นค่า log แล้วควรจะแปลงค่ากลับเป็นข้อมูลเดิม คือหาค่า Antilog นั้นเอง ส่วนค่า SD, SE ไม่จำเป็นต้องทำ แต่ถ้าจะนำเสนอเป็นค่า CI (Confidence Interval) ก็ควรหาค่า Antilog ของค่าที่ได้นั้นก่อนเพื่อจะได้ค่าข้อมูลจริงทั้งช่วง การแปลงข้อมูลควรจะทำจริง ๆ ว่าจะทำให้ข้อมูลนั้นแจกแจงเป็นแบบปกติ

3.6 ค่าที่อยู่นอกกลุ่ม (Outliers) ค่าที่มีค่ามากเกินไปหรือต่ำเกินไป ซึ่งจะตรวจพบได้จากกราฟ Box-plot (จะไม่กล่าวรายละเอียดในที่นี้) ไม่ควรจะต้องตัดออกจากการวิเคราะห์นอกจากว่ามีเหตุผลที่เพียงพอที่จะไม่เชื่อถือได้ก็ควรตัดออกไป ดังนั้นอาจต้องทำการวิเคราะห์ดูทั้งกรณีที่มีและไม่มีค่านั้นอยู่ แล้วประเมินผลที่ได้ว่าควรเลือกอย่างไร

3.7 สหสัมพันธ์ (Correlation) ส่วนมากจะมีการแสดงกราฟเป็น Scatter plot ให้ดูก่อน และเสนอค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient) ถ้ามีตัวแปรหลาย ๆ ตัวควรมีการแสดงค่าสปส. สหสัมพันธ์ในแต่ละคู่ของตัวแปร ในรูปของตารางซึ่งอาจเป็น Correlation matrix ข้อผิดพลาดที่พบได้บ่อย คือ มีการใช้ค่าสปส. สหสัมพันธ์กับข้อมูลที่มีการวัดค่าสังเกตมากกว่าหนึ่งครั้งในแต่ละ subjects ซึ่งการวัดแต่ละครั้งจะไม่เป็นอิสระกันจึงไม่สมควรใช้วิธีนี้ อีกประการหนึ่งมักจะใช้ค่าสหสัมพันธ์ในการเปรียบเทียบการวัดที่ต่างวิธีกันเพื่อดูความสอดคล้อง (Agreement) ของทั้ง 2 วิธีซึ่งไม่ถูกต้อง เพราะสหสัมพันธ์เพียงแต่ใช้ประเมินดูความสัมพันธ์ (Association) เท่านั้น ไม่ได้ดูความสอดคล้องกัน นอกจากนี้การวิเคราะห์ความถดถอย (Regression) ก็ใช้ในวัตถุประสงค์ต่างกับสหสัมพันธ์ จึงไม่จำเป็นต้องทำคู่กันเสมอไป ดังจะกล่าวในหัวข้อต่อไป

3.8 ความถดถอย (Regression) การวิเคราะห์วิธีนี้จะช่วยบอกถึงสมการทำนายที่เหมาะสมกับข้อมูล

แต่การบอกจะมีประโยชน์มากขึ้นถ้าได้บอกค่าความชัน (Slope) และ SE ของความชันไว้ด้วย การวิเคราะห์ด้วยวิธีนี้ถ้ามีกลุ่มย่อย ๆ (Subgroups) หลาย ๆ กลุ่ม ซึ่งมีความแตกต่างกันในค่าเฉลี่ยของตัวแปรตาม (Y) ควรจะวิเคราะห์โดยวิธีของ Analysis of Covariance (ANCOVA) จึงจะเหมาะสมกว่า

3.9 ข้อมูลการอยู่รอด (Survival data) ข้อมูลชนิดนี้ควรรายงานในรูปของกราฟหรือตารางชีพ (Life table) และบอกรายละเอียดของจำนวนผู้ป่วยที่เหลืออยู่ในแต่ละช่วงที่ติดตามผล (follow up time) การศึกษาข้อมูลชนิดนี้จะมีประสิทธิภาพกับข้อมูลที่หายไป (Censored data) การนำเสนอค่า Survival time ไม่เหมาะที่จะใช้ค่า mean เพราะ ข้อมูลมักจะเบี่ยงไปทางขวา จึงควรใช้ค่า Median

4. การนำเสนอผลการวิเคราะห์ (Presentation of results)

4.1 การสรุปค่าสถิติ (Summary statistics) การนำเสนอค่าสถิติต่าง ๆ เช่น ค่าเฉลี่ย ควรจะมีค่า SD, SE กำกับไว้ด้วย แต่จะต้องทราบถึงความแตกต่างของ 2 ค่านี้ ค่า SD จะบอกถึงความผันแปรในระหว่างหน่วยที่ศึกษา แต่ค่า SE จะบอกถึงความแม่นยำของค่า Sample mean การใช้สัญลักษณ์ + นั้นจะทำให้สับสนได้ จึงควรใช้วงเล็บบอกไว้ เป็น Mean (SD...) หรือ Mean (SE...) นอกจากนี้การที่ผู้วิจัยกำกับด้วยช่วงความเชื่อมั่น (CI) จะทำให้ผู้อ่านประเมินถึงความแม่นยำถูกต้องของการเอาไปใช้ด้วย การนำเสนอในรูปของเปอร์เซ็นต์ควรจะมีค่าของจำนวนที่เป็นตัวหาร (Denominator) กำกับไว้ แต่ถ้าจำนวนตัวอย่งมีน้อย ๆ ก็ไม่จำเป็นต้องบอกเป็นเปอร์เซ็นต์ ควรบอกไว้เป็นค่าสังเกตที่นับได้จริง ๆ จะได้ประโยชน์มากกว่า

4.2 ผลการทดสอบสมมติฐาน (Results of hypothesis tests) ในบทความนั้นควรจะต้องรายงานค่า t, z, χ^2) และค่า p-value การรายงานค่า p-value ก็ควรบอกค่าจริงไว้เลย โดยใช้ทศนิยมเพียง 2 หลักก็พอ ดังนั้นการที่รายงานผลว่าจะมีนัยสำคัญหรือไม่มีนัยสำคัญนั้นยังไม่พอ ควรบอกขนาดของความแตกต่างกำกับด้วย CI ไว้ด้วย เพื่อรายงานค่า p ในตารางแล้วก็ไม่จำเป็นต้อง

เขียนซ้ำอีกในข้อความที่บรรยายนั้น

4.3 การนำเสนอด้วยกราฟ (Graphical presentation) การนำเสนอด้วยกราฟในบทความจะช่วยให้ผู้อ่านได้รับรายละเอียดเกี่ยวกับการศึกษามากขึ้น การนำเสนอก็มีหลายแบบตามความเหมาะสมของลักษณะข้อมูล ข้อควรสังเกต คือ ผู้วิจัยมักนำเสนอกราฟเป็น Error bar ซึ่งบอกค่า ± 1 SE ซึ่งหมายถึง แสดงค่า 67% CI แต่ผู้อ่านมักจะเข้าใจเป็นค่าที่ 95% CI ดังนั้นผู้อ่านต้องมีความเข้าใจในเรื่องนี้ด้วย เพื่อจะได้ไม่แปลผลผิดพลาด

4.4 การนำเสนอในรูปของตารางและตัวเลข (Presentation of tables and numeric) การนำเสนอผลเป็นตัวเลขนั้นควรเสนอในแนวเป็นคอลัมน์ ซึ่งจะทำให้ผู้อ่านมองได้ง่ายกว่าในแนวแถว นอกจากนี้ควรกำกับค่าสังเกตทั้งหมด (n) ในตารางแต่ละครั้งด้วย ตารางที่นำเสนอข้อมูลต่าง ๆ ที่บอกถึงลักษณะของตัวอย่างที่ศึกษานั้น การจัดเรียงตามระดับของตัวแปรต่าง ๆ จะทำให้อ่านง่ายขึ้น ส่วนการนำเสนอค่าเฉลี่ยนั้นควรแสดงเป็นค่าทศนิยมที่มากกว่าค่าข้อมูลดิบเพียง 1 ตำแหน่ง และค่า SD ก็แสดงเป็นค่าทศนิยมที่มากกว่าค่าเฉลี่ยอยู่ 1 ตำแหน่งก็พอ ถ้ามีการปิดเศษเกิดขึ้น ก็ยังไม่ควรทำในระหว่างการวิเคราะห์เพราะอาจทำให้ผลคาดเคลื่อนได้ควรทำหลังจากเสร็จการวิเคราะห์แล้ว ค่าสถิติต่าง ๆ เช่น t, χ^2 นั้นนำเสนอเพียง 2 ตำแหน่งทศนิยมก็เพียงพอ

5. การวิจารณ์ผล (Discussion)

ในส่วนนี้จะกล่าวถึงการเอาผลทางสถิติไปเกี่ยวข้องกับการวิจารณ์ผล ซึ่งจะแบ่งเป็นประเด็นต่าง ๆ ดังนี้

5.1 ผลการทดสอบที่มี/ไม่มีนัยสำคัญทางสถิติ (Statistical Significance) ผู้วิจัยจะต้องไม่ให้ความสำคัญทางสถิติมากเกินไป ต้องพิจารณาถึงความมีนัยสำคัญทางคลินิกด้วย (Clinical Significance) ดังนั้นผู้อ่านควรใช้วิจารณญาณในการพิจารณาผลการศึกษาให้ถี่ถ้วนที่จะนำไปใช้

5.2 ผลการวิเคราะห์นั้นสัมพันธ์กับวัตถุประสงค์ของการศึกษาหรือไม่ ควรพิจารณาว่าผู้วิจัยได้ทำการเสนอผลการวิเคราะห์ได้ตามวัตถุประสงค์ที่ตั้งไว้หรือไม่

ไม่ และผลนั้นเมื่อเปรียบเทียบกับ การทบทวนวรรณกรรม ที่เคยมีผู้ศึกษาไว้ก่อนหน้านี้เป็นอย่างไรอย่างไหนมีความ น่าเชื่อถือได้ดีกว่าและการศึกษาครั้งนี้มีข้อดีกว่าครั้งก่อน หรือไม่

5.3 ข้อจำกัดของการวิจัย ผู้วิจัยได้บอกถึงข้อ จำกัดของการศึกษาครั้งนี้ไว้หรือไม่ เพื่อที่ผู้อ่านจะได้ทราบ และนำไปใช้ได้ถูกต้อง ข้อจำกัดที่ว่าได้แก่ ตัวอย่างที่ศึกษา เพียงพอแค่ไหนในการเอาผลไปขยายใช้ในประชากรเป้าหมาย นั่นคือดูว่ามี Generalizability แค่ไหน

6. การสรุปผล (Conclusion)

ผู้อ่านควรพิจารณาว่าผู้วิจัยได้สรุปผลซึ่งสามารถ ตอบคำถามการวิจัยได้จริงหรือไม่ นอกจากนี้ผู้วิจัยได้ เสนอสมมติฐานอันใหม่ไว้หรือไม่ สุดท้ายผู้อ่านเห็นด้วย หรือไม่กับการสรุปผลของผู้วิจัยในครั้งนี้

ข้อผิดพลาดที่พบได้บ่อยในบทความทางการแพทย์

Altman DG. (1998)⁽¹⁾ ได้ประเมินคุณค่าทางสถิติ ของบทความทางการแพทย์ที่ตีพิมพ์แล้วพบว่ายังมี ข้อผิดพลาดทางด้านสถิติซึ่งเกี่ยวข้องกันในวิธีการศึกษา, การวิเคราะห์, การนำเสนอและการแปลผล ดังนี้

1. วิธีการศึกษา มีข้อผิดพลาดที่พบดังนี้

- 1.1 ไม่มีการใช้วิธีการของ Randomization ในการ วิจัยทางคลินิก
- 1.2 ใช้กลุ่มควบคุมไม่เหมาะสม
- 1.3 ใช้วิธีการศึกษาแบบ Crossover design กับ การศึกษาภาวะบางอย่างที่รักษาให้หายได้
- 1.4 ไม่ได้คำนึงถึงขนาดตัวอย่างว่าเพียงพอหรือไม่
- 1.5 ไม่ได้พิจารณาถึงอัตราการตอบกลับ (Response Rate) ว่าเพียงพอแล้วหรือไม่

2. การวิเคราะห์ข้อมูล มีข้อผิดพลาดที่พบดังนี้

- 2.1 มักจะใช้วิธีของ Unpaired กับข้อมูลชนิด Paired data
- 2.2 ใช้วิธีของ t-test ในการเปรียบเทียบเวลาการอยู่รอด (Survival times) ซึ่งบางครั้งมีข้อมูลชนิด censored ด้วย
- 2.3 ใช้วิธีการของ Correlation กับ การดูการเปลี่ยนแปลง

แปลงเมื่อเทียบกับค่าเริ่มต้น

- 2.4 เปรียบเทียบค่า p-values กันซึ่งไม่ถูกต้อง
- 2.5 การใช้วิธีการของ Parametric นั้น บางครั้งใช้ กับข้อมูลที่ไม่ได้แจกแจงแบบปกติ
- 2.6 การจัดกลุ่มข้อมูล (Categorization) โดยปราศ จากเหตุผลที่เพียงพอ
- 2.7 ผู้วิจัยไม่ได้ปรับตัวแปรกวน (Confounding factors) ในการวิเคราะห์
- 2.8 บางครั้งผู้วิจัยใช้ชื่อของการทดสอบผิดพลาด เช่น ใช้ Variance analysis ซึ่งอันที่จริง คือ Analysis of Variance (ANOVA)

3. การนำเสนอ มักจะพบข้อผิดพลาดดังนี้

- 3.1 ใช้ค่า SE แทน SD ในการพรรณนาข้อมูล
- 3.2 ใช้กราฟวงกลมในการแสดงข้อมูลต่อเนื่อง
- 3.3 ผลที่แสดงมีแต่ค่า p-value
- 3.4 มีการแบ่งสเกลในการนำเสนอด้วยฮิสโตแกรม
- 3.5 การนำเสนอด้วย Scatter diagrams ไม่แสดงข้อมูลครบทุกคู่
- 3.6 มักจะนำเสนอตัวเลขค่าต่าง ๆ เป็นทศนิยมมากเกินไป
- 3.7 การนำเสนอค่า p-value บางครั้งแสดงค่า p เป็น $p = NS$, $p = 0.0000$ ซึ่งควรแสดงเป็นค่าจริง ๆ หรือเขียน เป็น $P < .0001$ จะดีกว่า

4. การแปลผล มักจะพบข้อผิดพลาดดังนี้

- 4.1 ไม่ได้พิจารณาถึงค่า CI เมื่อพบว่าผลการศึกษา ไม่มีนัยสำคัญทางสถิติ
- 4.2 ทำการสรุปถึงสาเหตุ (Causation) จากผลการ ศึกษาความสัมพันธ์โดยขาดหลักฐานสนับสนุน
- 4.3 ผู้วิจัยมักจะไม่ค่อยวิจารณ์ผลการวิเคราะห์ไว้ใน บทความ
- 4.4 ผู้วิจัยมักจะสรุปว่าการศึกษานั้นดีทั้ง ๆ ที่มีข้อ บกพร่องมากมาย

โดยสรุปจะเห็นได้ว่าการอ่านบทความทางการแพทย์ นั้นผู้อ่านควรใช้วิจารณญาณในการอ่าน ตลอดจนการสรุป ผลไปใช้จะทำให้การอ่านบทความมีความน่าเชื่อถือยิ่งขึ้น

อ้างอิง

1. Altman DG. Statistical reviewing for medical journals. *Stat. Med* 1998 Dec 15; 17(23): 2661-74
2. Bailar JC 3 rd. Mosteller F. Guidelines for statistical reporting in articles for medical journals. Amplifications and Explanations. *Ann Inter Med* 1988 Feb; 108(2): 266 - 73
3. Altman DG, Machin D, Bryant TN, Gardner MJ, eds. *Statistics with Confidence, 2nd ed* BMJ Books, British Library Cataloguing in Publication Data, JW Arrowsmith Ltd, Bristol 2000
4. Walter SD. Methods of reporting statistical results from Medical research studies. *Am J Epidemiol* 1995 Vol. 141(10): 896-906

ศูนย์บริการ
สุขภาพกรีนมหาวิทยาลัย

กิจกรรมการศึกษาต่อเนื่องสำหรับแพทย์

ท่านสามารถได้รับการรับรองอย่างเป็นทางการสำหรับกิจกรรมการศึกษาต่อเนื่องสำหรับแพทย์ กลุ่มที่ 3 ประเภทที่ 23 (ศึกษาด้วยตนเอง) โดยศูนย์การศึกษาต่อเนื่องของแพทย์ จุฬาลงกรณ์มหาวิทยาลัย ตามเกณฑ์ของศูนย์การศึกษาต่อเนื่องของแพทย์แห่งแพทยสภา (ศนพ.) จากการอ่านบทความเรื่อง "การประเมินสถิติในบทความทางการแพทย์" โดยตอบคำถามข้างล่างนี้ พร้อมกับส่งคำตอบที่ท่านคิดว่า ถูกต้องโดยใช้แบบฟอร์มคำตอบท้ายคำถาม แล้วใส่ซองพร้อมซองเปล่า (ไม่ต้องติดแสตมป์) จำนวนของถึง ตัวท่าน ส่งถึง

ศ. นพ. สุทธิพร จิตต์มิตรภาพ
บรรณารักษารุฬาลงกรณ์เวชสาร
และประธานคณะกรรมการการศึกษาต่อเนื่อง
คณะแพทยศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
หน่วยจุฬาลงกรณ์เวชสาร
ตึกอบรบวิชาการ ชั้นล่าง
เขตปทุมวัน กทม. 10330

จุฬาลงกรณ์เวชสารขอสงวนสิทธิ์ที่จะส่งเฉลยคำตอบพร้อมหนังสือรับรองกิจกรรมการศึกษา ต่อเนื่องอย่างเป็นทางการ ดังกล่าวแล้วข้างต้นสำหรับท่านที่เป็นสมาชิกจุฬาลงกรณ์เวชสารเท่านั้น สำหรับ ท่านที่ยังไม่เป็นสมาชิกแต่ถ้าท่านสมัครเข้าเป็นสมาชิกจุฬาลงกรณ์เวชสารสำหรับวารสารปี 2545 (เพียง 200 บาทต่อปี) ทางจุฬาลงกรณ์เวชสารยินดีดำเนินการส่งเฉลยคำตอบจากการอ่านบทความให้ตั้งแต่ฉบับ เดือนมกราคม 2545 จนถึงฉบับเดือนธันวาคม 2545 โดยสามารถส่งคำตอบได้ไม่เกินเดือนมีนาคม 2546 และจะส่งหนังสือรับรองชนิดสรุปเป็นรายปีว่าท่านสมาชิกได้เข้าร่วมกิจกรรมการศึกษาต่อเนื่องที่จัดโดย จุฬาลงกรณ์เวชสาร จำนวนกี่เครดิตในปีที่ผ่านมา โดยจะส่งให้ในเดือนเมษายน 2546

คำถาม - คำตอบ

- ค่าของข้อมูลที่สูญหาย (Missing Value) คืออะไร ?
 - ค่าที่วัดแล้วไม่สามารถบอกค่าโดยละเอียดได้
 - ค่าที่ผู้ตอบไม่ยอมให้ข้อมูล
 - ค่าที่ผู้ตอบไม่ทราบว่าจะตอบอย่างไร
 - ค่าที่ไม่สามารถนำมาวิเคราะห์ได้จึงต้องตัดทิ้งไป

คำตอบ สำหรับบทความเรื่อง "การประเมินสถิติในบทความทางการแพทย์"

จุฬาลงกรณ์เวชสาร ปีที่ 46 ฉบับที่ 9 เดือนกันยายน พ.ศ. 2545

รหัสสื่อการศึกษาต่อเนื่อง 3-15-201-2000/0209-(1019)

ชื่อ - นามสกุลผู้ขอ CME credit เลขที่ใบประกอบวิชาชีพเวชกรรม.....
ที่อยู่.....

- (ก) (ข) (ค) (ง)
- (ก) (ข) (ค) (ง)
- (ก) (ข) (ค) (ง)
- (ก) (ข) (ค) (ง)
- (ก) (ข) (ค) (ง)

2. อำนาจของการทดสอบ (Power of the test) หมายถึงอะไร ?
 - ก. โอกาสที่ทดสอบได้ถูกต้อง
 - ข. โอกาสที่จะผิดพลาด
 - ค. โอกาสที่จะพบความแตกต่างได้ถ้าความแตกต่างนั้นมีจริง
 - ง. โอกาสที่มีนัยสำคัญทางคลินิก
3. เกี่ยวกับการนำเสนอด้วยค่า SD, SE ข้อใดที่ไม่ถูกต้อง
 - ก. ค่า SD จะบอกถึงความผันแปรของข้อมูลที่น่าเสนอ
 - ข. ค่า SE จะบอกถึงการกระจายของข้อมูลที่มีอยู่
 - ค. ค่า SE จะมีค่าน้อยกว่า SD
 - ง. จำนวนตัวอย่างที่มากจะทำให้ได้ค่า SE น้อย
4. ข้อใดไม่ถูกต้องเกี่ยวกับค่า p-value และค่า CI
 - ก. การรายงานผลเพียงค่า p-value อย่างเดียวไม่พอควรมีค่า CI กำกับไว้ด้วย
 - ข. ค่า CI จะให้ผลสอดคล้องกับการทดสอบแบบ two-tailed test
 - ค. การรายงานค่า CI มักจะรายงานที่ 95 % CI
 - ง. ค่า CI จะช่วยสรุปผลทั้งในแง่ความสำคัญทางสถิติและความสำคัญทางคลินิก
5. ในการอ่านบทความทางการแพทย์นั้น สิ่งที่ควรพิจารณาความเหมาะสมคือ
 - ก. มีการเลือกตัวอย่างไว้อย่างเหมาะสมหรือไม่
 - ข. วัตถุประสงค์ที่ตั้งไว้ผลตามที่รายงานหรือไม่
 - ค. มีอัตราการตอบกลับสูงพอหรือไม่
 - ง. ถูกทุกข้อ

ท่านที่ประสงค์จะได้รับเครดิตการศึกษาต่อเนื่อง (CME credit)
กรุณาส่งคำตอบพร้อมรายละเอียดของท่านตามแบบฟอร์มด้านล่าง

ศาสตราจารย์นายแพทย์สุทธิพร จิตต์มิตรภาพ
ประธานคณะกรรมการการศึกษาต่อเนื่อง
คณะแพทยศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
หน่วยจุฬาลงกรณ์เวชสาร ตึกอบรมวิชาการ ชั้นล่าง
คณะแพทยศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
เขตปทุมวัน กทม. 10330