

กรณีศึกษา : ความผิดพลาดในการคำนวณสถิติ จากค่าลอการิทึม

สุธน วงษ์ชัย* กิตติพงษ์ เกิดฤทธิ์** สุรศักดิ์ หมั่นพล* มะลิวัลย์ หอมจันทร์*
และพรหมภัสสร ดิษสระ*

*สำนักมาตรฐานห้องปฏิบัติการ กรมวิทยาศาสตร์การแพทย์ ถนนติวานนท์ นนทบุรี 11000

**สำนักโรคติดต่อทางเดินหายใจ กรมควบคุมโรค ถนนติวานนท์ นนทบุรี 11000

บทคัดย่อ ได้ทบทวนค่าสถิติของผลงานวิจัยการตรวจปริมาณไวรัส ซัยโตเมกาโลไวรัส (Cytomegalo virus, CMV) ที่นำเสนอในวารสารระดับนานาชาติ (J Mol Diagn. Mar 2009; 11(2): 87-92.) พบความผิดพลาดจากการคำนวณที่ละเอียดถี่ถ้วน การคำนวณตัวเลขลอการิทึม ทำให้ค่าเฉลี่ยคณิตศาสตร์ต่ำกว่าค่าที่ถูกต้อง ค่าเบี่ยงเบนมาตรฐานที่คำนวณจากค่าลอการิทึมโดยตรงมีค่าผิดพลาดในทางต่ำกว่าค่าที่ถูกต้อง ทำให้ได้ข้อสรุปที่ผิดพลาดว่า ผลเฉลี่ยจากการวิเคราะห์จากห้องปฏิบัติการ 23 แห่ง ได้ผลไม่แตกต่างจากค่าอ้างอิง และส่งผลให้สถิติทั้งหมดที่ใช้ค่าเบี่ยงเบนมาตรฐานผิดพลาดทั้งหมดคือ Pair t-test, 95% Confidence Intervals, Coefficient of Variance, Correlation Coefficient และส่งผลต่อข้อสรุปของผู้ประพันธ์ที่ว่าผลตรวจวิเคราะห์ปริมาณ CMV ระหว่างห้องปฏิบัติการทั้ง 4 ระดับนั้น มีค่าไม่แตกต่างกัน ทำให้ผู้ประพันธ์ลดความสำคัญเรื่องความผันแปรของค่า CMV load ในการเปรียบเทียบระหว่างห้องปฏิบัติการ แต่ไปให้ความสำคัญกับเชื้อมาตรฐานที่ใช้ในการเปรียบเทียบ

บทนำ

ลอการิทึมเป็นวิธีการนำเสนอและคำนวณข้อมูลตัวเลขยกกำลังให้สะดวกขึ้น ส่วนประกอบของตัวเลขลอการิทึมมี 2 ส่วน คือ ตัวเลขหน้าจุดทศนิยม (Characteristic) ระบุระดับเลขยกกำลัง ตัวเลขหลังจุดทศนิยม (Mantissa) แสดงตัวเลขนัยสำคัญของการวัด ในบทความนี้จะใช้ลอการิทึมฐาน 10 โดยละค่าฐานไว้ในที่เข้าใจ ลอการิทึม ประดิษฐ์ขึ้นโดย John Napier of Merchiston นักคณิตศาสตร์ นักฟิสิกส์ นักดาราศาสตร์ และ นักวิทยาศาสตร์ ชาวสกอตผู้มีชีวิตอยู่ ค.ศ. 1550-1617 ได้ตีพิมพ์ลอการิทึมครั้งแรกเมื่อ ค.ศ. 1614 ที่เมืองเอดินเบิร์ก เป็นภาษาละติน ชื่อ Merifici Logarithmorum Descriptio และแปลเป็นภาษาอังกฤษใน ค.ศ. 1616

กฎพื้นฐานในการคำนวณลอการิทึมที่เกี่ยวข้องใช้แค่ การบวกและการลบเลขลอการิทึมให้เป็นการคูณและการหาร ตามลำดับ ดังตัวอย่างข้างล่าง

การบวก	$\log M + \log N = \log (M*N)$
ตัวอย่างการคำนวณ	$\log 5 + \log 2 = \log (5*2) = \log 10 = 1$
ขณะที่	$\log 5 = 0.699, \log 2 = 0.301$
	หากบวกค่าลอการิทึมโดยตรงจะได้ $\log 7$ ซึ่งมีค่า $\text{antilog } 7 = 0.845$
การลบ	$\log M - \log N = \log (M/N)$
ตัวอย่างการคำนวณ	$\log 5 - \log 2 = \log (5/2) = \log 2.5 = 0.398$
	หากนำค่าลอการิทึมมาลบโดยตรงจะได้ $\log 3$ ซึ่งมีค่า $\text{antilog } 3 = 0.477$

จะเห็นว่าการคำนวณที่ใช้ค่าลอการิทึมจำเป็นต้องใช้กฎการคำนวณของลอการิทึม ไม่ใช่การบวกหรือลบตัวเลขลอการิทึมโดยตรง ทางเลือกสำหรับผู้ไม่สันทัด อาจแปลงค่าลอการิทึมกลับเป็นตัวเลขยกกำลัง ซึ่งเป็นค่าสัมบูรณ์ที่ใช้คำนวณโดยตรงได้

วิธีการ

ในบทความนี้จะใช้ข้อมูลระดับไวรัส CMV จากการเปรียบเทียบระหว่าง 23 ห้องปฏิบัติการ ที่รายงานในวารสาร J Mol Diagn. Mar 2009; 11(2): 87-92 เรื่อง **Multi-Site PCR-Based CMV Viral Load Assessment-Assays Demonstrate Linearity and Precision, but Lack Numeric Standardization**. A Report of the Association for Molecular Pathology ซึ่งได้รับอนุญาตให้ใช้ข้อมูลจากเจ้าของลิขสิทธิ์ บริษัท Elsevier Inc., ใช้โปรแกรมสถิติของเอกเซล แต่เปลี่ยนข้อมูลลอการิทึมเป็นค่าสัมบูรณ์ก่อนคำนวณตัวสถิติทั้งหมด แล้วแปลงค่าที่คำนวณได้เป็นลอการิทึม (ตารางที่ 1 และ 2)

ตารางที่ 1 ข้อมูลระดับ CMV load, copies/ml จาก 23 ห้องปฏิบัติการ ในค่าลอการิทึม
* ค่าที่ไม่แสดงในตารางเป็นค่าที่ต่ำกว่า 50 copies/ml

ค่าเฉลี่ยระดับ CMV load จากสมาชิก 23 ห้องปฏิบัติการ				
Laboratory	CMV Viral 2.70 copies/ml	DNA 3.70 copies/ml	Panels 4.70 copies/ml	5.70 copies/ml
1	4.12	4.37	5.08	5.97
2	*	*	4	4.78
3	3.27	4.04	5.18	5.74
4	2.92	3.89	5.01	5.75
5	2.54	3.17	4.47	5.02
6	3.02	3.64	4.79	5.94
7	3.53	4.72	5.77	6.58
8	3.06	3.98	4.97	6.02
9	2.79	3.59	4.76	5.7
10	3.36	4.31	5.27	6.25
11	2.77	3.84	4.92	5.93
12	2.98	3.76	4.63	5.68
13	2.59	3.8	4.71	5.66
14	4.08	5.2	6.1	7.25
15	2.04	3.2	4.26	5.56
16	2.71	3.68	4.73	5.67
17	*	3.45	4.14	4.75
18	2.51	3.5	4.62	5.63
19	*	3.29	4.3	4.97
20	2.6	3.68	4.72	5.68
21	2.53	3.54	4.46	5.47
22	2.3	2.95	3.74	4.83
23	3.41	4.3	5.04	5.45
Mean	2.96	3.81	4.77	5.66

(ที่มา: ประยุกต์จากตารางที่ 2 ของเอกสารอ้างอิง 2 ได้รับอนุญาตให้ใช้ได้ จากเจ้าของลิขสิทธิ์ บริษัท Elsevier Inc.,)

ผล

ค่าของตัวสถิติจากวารสารที่ศึกษาที่คำนวณจากค่าลอการิทึมโดยตรง และไม่ทำตามกฎการคำนวณบวกหรือลบค่าลอการิทึมเปรียบเทียบกับค่าตัวสถิติที่คำนวณจากค่าสัมบูรณ์ได้แสดงรายละเอียดในตารางที่ 2 ดังนี้

1. ค่าเฉลี่ยทางคณิตศาสตร์ที่แสดงในวารสารเป็นค่าที่คำนวณผิดอย่างชัดเจน เนื่องจาก

- 1.1 คำนวณค่าเฉลี่ยทางคณิตศาสตร์จากค่าลอการิทึมโดยไม่ทำตามกฎการบวกค่าลอการิทึม
- 1.2 ค่าที่ต่ำกว่าขีดจำกัดในการตรวจที่ 50 copies/ml พบ 3 ค่า เมื่อตรวจตัวอย่างอ้างอิงที่ 500 copies/ml และค่าที่ต่ำกว่าขีดจำกัดในการตรวจที่ 50 copies/ml พบ 1 ค่า เมื่อตรวจตัวอย่างอ้างอิงที่ 5000 copies/ml ถูกเฉลยไม่นำมาคำนวณด้วย ทั้งๆ ที่อาจจะมีค่า ตั้งแต่ 0 - 50 copies/ml ทำให้ค่าเฉลี่ยทางคณิตศาสตร์สูงกว่าค่าที่ควรจะเป็น

จากตารางที่ 2 ค่าอ้างอิงของเชื้อมาตรฐานที่ใช้มีค่าลอการิทึมเป็น 2.7, 3.7, 4.7 และ 5.7 หรือ 500, 5000, 50000, 500000 copies/ml ขณะที่ค่าเฉลี่ยทางคณิตศาสตร์ในวารสารเป็น 2.96, 3.81, 4.77, 5.66 หรือ 912, 6457, 58884 และ 4570888 copies/ml ตามลำดับ เมื่อทดสอบทางสถิติด้วย Paired t-test, two tail, $t = 0.345$, $p = 0.376$ สรุปว่าค่าที่ได้จาก 23 ห้องปฏิบัติการ ไม่แตกต่างจากค่าอ้างอิง

หากคำนวณค่าเฉลี่ยทางคณิตศาสตร์โดยใช้กฎการบวกของลอการิทึมจะได้ค่าเฉลี่ยลอการิทึมเป็น 3.33, 3.81, 4.77, 5.66 หรือ 2156, 16147, 138298 และ 1404755 copies/ml ตามลำดับ เมื่อทดสอบทางสถิติด้วย Paired t-test, two tail, $t = 25.8855$, $p = 0.0008$ สรุปว่าค่าที่ได้จาก 23 ห้องปฏิบัติการ แตกต่างจากค่าอ้างอิงอย่างมีนัยสำคัญทางสถิติ

2. ค่าเบี่ยงเบนมาตรฐานและสถิติอื่นๆ ที่ใช้ค่าเบี่ยงเบนมาตรฐานร่วมในการคำนวณที่แสดงในวารสารเป็นค่าที่คำนวณผิดอย่างชัดเจนเพราะไม่ทำตามกฎการลบค่าลอการิทึมรวมถึงความผิดพลาดในข้อ 1.1 และ 1.2

- 2.1 ค่าเบี่ยงเบนมาตรฐานในวารสารเป็น 0.54, 0.53, 0.53, 0.58 คิดเป็น 3.5, 3.4, 3.4, 3.8 copies/ml ตามลำดับ ซึ่งเป็นไปไม่ได้ หากการขีดจำกัดในการตรวจหาอยู่ที่ 50 copies/ml

หากคำนวณค่าเบี่ยงเบนมาตรฐานโดยใช้กฎการลบของลอการิทึมจะได้ค่าเบี่ยงเบนมาตรฐานลอการิทึมเป็น 3.57, 4.53, 5.43 และ 6.56 หรือ 3684, 33816, 271659, 3656718 copies/ml ตามลำดับ

- 2.2 ค่าสัมประสิทธิ์ความแปรปรวน (Coefficient of Variance). ในวารสารเป็น 18.40%, 13.86%, 11.17% และ 10.20% ขณะที่คำนวณจากค่าสัมบูรณ์เป็น 171%, 209%, 196%, 256% ตามลำดับ ทำให้ข้อสรุปของผู้ประพันธ์ผิดพลาด ที่ว่า ค่าที่เปลี่ยนแปลงในระดับ 3-5 เท่า มีนัยสำคัญต่อการรักษา

- 2.3 ค่าระดับความเชื่อมั่นร้อยละ 95 จากวารสารที่คำนวณเป็นค่าลอการิทึม 0.24, 0.22, 0.22, 0.24 หรือประมาณ 1.7 copies/ml ตามลำดับ ซึ่งเป็นไปไม่ได้หากการขีดจำกัดในการตรวจหาอยู่ที่ 50 copies/ml

หากคำนวณค่าระดับความเชื่อมั่นร้อยละ 95 โดยใช้กฎการลบของลอการิทึมจะได้ค่าระดับความเชื่อมั่นร้อยละ 95 ลอการิทึม เป็น 1.56, 1.89, 2.22, 2.68 หรือ 1615, 14130, 119058, 1602598 copies/ml ตามลำดับ

ตารางที่ 2 เปรียบเทียบค่าตัวสถิติที่คำนวณจากข้อมูลลอการิทึมโดยตรง กับค่าตัวสถิติที่คำนวณจากค่าสัมบูรณ์ โดย CMV load ของตัวอย่างอ้างอิงเป็น copies/ml อยู่ในค่าลอการิทึมและค่าสัมบูรณ์ แสดงในวงเล็บ

	2.70 (500) copies/ml	3.70 (5000) copies/ml	4.70 (50000) copies/ml	5.70 (500000) copies/ml
1. ค่าที่คำนวณจากลอการิทึมโดยตรง				
1.1 ค่าเฉลี่ยคณิตศาสตร์	2.96 (912)	3.81 (6457)	4.77 (58884)	5.66 (457088)
1.2 ค่าเบี่ยงเบนมาตรฐาน	0.54 (3.5)	0.53 (3.4)	0.53 (3.4)	0.58 (3.8)
1.3 ค่าสัมประสิทธิ์ความแปรปรวน (%)	18.40	13.86	11.17	10.20
1.4 ค่าระดับความเชื่อมั่นร้อยละ 95 (+/-)	0.24 (1.7)	0.22 (1.7)	0.22 (1.7)	0.24 (1.7)
2. ค่าที่คำนวณจากค่าสัมบูรณ์หรือใช้กฎการบวกหรือลบค่าลอการิทึมแล้วแปลงกลับเป็นค่าลอการิทึม				
2.1 ค่าเฉลี่ยคณิตศาสตร์	3.33 (2156)	4.21 (16147)	5.14 (138298)	6.15 (1404755)
2.2 ค่าเบี่ยงเบนมาตรฐาน	3.57 (3680)	4.53 (33800)	5.43 (272000)	6.56 (3660000)
2.3 ค่าสัมประสิทธิ์ความแปรปรวน (%)	171	209	196	256
2.4 ค่าระดับความเชื่อมั่นร้อยละ 95 (+/-)	3.57 (1610)	4.53 (14100)	5.53 (111000)	6.56 (1602598)

วิจารณ์

สรุปความผิดพลาดจากการละเลยกฎการคำนวณค่าลอการิทึมและผลกระทบในการคำนวณค่าสถิติต่าง ๆ ที่นำเสนอในวารสาร J Mol Diagn. Mar 2009; 11(2): 87-92 เรื่อง Multi-Site PCR-Based CMV Viral Load Assessment-Assays Demonstrate Linearity and Precision, but Lack Numeric Standardization. A Report of the Association for Molecular Pathology. ดังนี้

1. ค่าเฉลี่ยทางคณิตศาสตร์ที่แสดงในวารสารเป็นค่าที่คำนวณผิด ได้ค่าต่ำกว่าค่าที่เป็นจริงตามการทดลอง ส่งผลให้ผู้รายงานสรุปว่า ค่า CMV load ที่ได้จากห้องปฏิบัติการ 23 แห่ง ไม่แตกต่างจากค่าอ้างอิง ทั้ง ๆ ที่ค่าดังกล่าวแตกต่างจากค่าอ้างอิงอย่างมีนัยสำคัญทางสถิติ หากผู้วิจัยได้คำนวณอย่างถูกต้อง จะต้องยุติการศึกษาในขั้นต่อไป แล้วเริ่มต้นวิเคราะห์หาสาเหตุแห่งความผิดพลาดนี้

2. ค่าเบี่ยงเบนมาตรฐานและสถิติอื่น ๆ ที่ใช้ค่าเบี่ยงเบนมาตรฐานร่วมในการคำนวณที่แสดงในวารสารเป็นค่าที่ต่ำกว่าค่าที่เป็นจริง ทำให้สรุปว่าความแม่นยำระหว่างห้องปฏิบัติการ อยู่ในระดับที่เปรียบเทียบกันได้ และได้แสดงค่าความแม่นยำที่ต่ำกว่าจะเป็นจริงได้ อาทิ ค่าเบี่ยงเบนมาตรฐานในวารสารเป็นค่าลอการิทึมเป็น 0.54, 0.53, 0.53,

0.58 คิดเป็น 3.5, 3.4, 3.4, 3.8 copies/ml ตามลำดับ ซึ่งเป็นไปไม่ได้ หากขีดจำกัดในการตรวจหาอยู่ที่ 50 copies/ml

3. ค่าสถิติที่ใช้ค่าเบี่ยงเบนมาตรฐานในการคำนวณผิดพลาดทั้งหมด เช่น ค่าสัมประสิทธิ์ความแปรปรวน และค่าระดับความเชื่อมั่นร้อยละ 95 ได้ค่าต่ำเกินจะเป็นจริงในทางปฏิบัติได้ โดยแสดงค่าสัมประสิทธิ์ความแปรปรวนเพียง 10-20% ส่วนค่าระดับความเชื่อมั่นร้อยละ 95 บวกลบเพียง 2 copies/ml ทำให้ผู้วิจัยสรุปผิดว่าผลตรวจที่เปลี่ยนแปลงในระดับ 3-5 เท่า มีนัยสำคัญต่อการรักษา

สรุป

การคำนวณชุดข้อมูลที่เป็นลอการิทึมต้องทำตามกฎของลอการิทึม หากคำนวณโดยเพิกเฉยต่อกฎทำแบบตัวเลขปกติจะได้ผลลัพธ์ทางคณิตศาสตร์ที่ผิดพลาด ผลการวิเคราะห์ทางสถิติผิดพลาด ได้ผลสรุปที่ผิดพลาด และอาจนำไปสู่การใช้งานทางคลินิกที่ผิดพลาด หากเป็นผลงานวิจัยวิทยาศาสตร์การแพทย์ ในบทความที่ใช้เป็นกรณีศึกษาครั้งนี้ สรุปว่าผลตรวจ CMV load ระหว่าง 23 ห้องปฏิบัติการ มีค่าไม่แตกต่างกันหรือ เปรียบเทียบกันได้นั้น เป็นการสรุปที่ผิดพลาดเนื่องจาก ได้ละเลยการใช้กฎการคำนวณลอการิทึม ซึ่งแท้ที่จริงแล้ว ผลตรวจ CMV load ระหว่าง 23 ห้องปฏิบัติการ มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติ และผู้วิจัยต้องค้นหาสาเหตุของความแตกต่างกัน ก่อนที่จะเปรียบเทียบค่า CMV load ระหว่างห้องปฏิบัติการต่อไป ผู้เขียนบทความได้แสดงวิธีการและข้อควรระวังไว้ในกลุ่มเฟสบุ๊ค อี เลินนี่ <https://www.facebook.com/groups/labqualitycenter> เรื่อง คำนวณลอการิทึม ระวังพลีลือก ตั้งแต่ พ.ศ. 2555 แต่ไม่มีประจักษ์พยานที่เป็นวารสารตีพิมพ์ ข้อมูลจากวารสารฉบับนี้จึงเป็นเพียงตัวอย่างการคำนวณที่ละเลยความหมายและกฎของลอการิทึม

เอกสารอ้างอิง

1. John Napier. [online]. 2014 [cited 2014 Oct 16]: [6 screens]. Available from: URL: http://en.wikipedia.org/wiki/John_Napier
2. Wolff DJ, Heaney DL, Neuwald PD, Stellrecht KA, Press RD. Multi-site PCR-based CMV viral load assessment—assays demonstrate linearity and precision, but lack numeric standardization: a report of the association for molecular pathology. *J Mol Diagn* 2009; 11(2): 87-92.
3. การคำนวณค่าลอการิทึม. [ออนไลน์]. 2558; [สืบค้น 28 ม.ค. 2558]; [3 หน้า]. เข้าถึงได้จาก: URL: <https://www.facebook.com/groups/labqualitycenter>

Failure in Calculating The Statistics from Logarithm Values: Case Study

Suthon Vongsheree* Kittipoong Keardrit** Surasak Muenphon*
Maliwan Homchan* and Prompussorn Dissara*

**Bureau of Laboratory Quality Standards, Department of Medical Sciences, Tiwanond Road, Nontaburi 11000. Thailand. **Bureau of the Vector-Borne Diseases, Department of Diseases Control, Tiwanond Road, Nontaburi 11000. Thailand.*

ABSTRACT The statistics of an article on multi-center evaluation of the viral loads (Cytomegalo virus, CMV) presented in international journals (J Mol Diagn. Mar 2009; 11 (2): 87-92., were reviewed to check the validity of statistical calculation by using logarithmic rules. There were gross errors in log calculation of the means CMV loads which lead to missed conclusion that CMV loads amongst 23 laboratories were not different from the reference values, statistically. The Standard deviations (SD) was missed calculated from the log values directly, resulting in lower SD, compared to those calculated by using the absolute values. All statistics those using SD were miss-calculated : the pair t-test, 95% confidence intervals, Coefficient of Variance, Correlation Coefficient and affected the conclusions of the author that there was no difference in CMV load determination amongst these 23 laboratories. Hence, the significant CMV load variation amongst 23 laboratories were neglected and the need for reference materials were over-emphasized.

Key words: miss calculation, Logarithmic data, error means and SD