

เจริญ ตริศักดิ์*

¹ สาขาวิชาเภสัชกรรมคลินิก คณะเภสัชศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ อ.องครักษ์ จ.นครนายก 26120

* ติดต่อผู้พิมพ์: charoen@g.swu.ac.th

เสวนาสารเภสัชกรรมและบริการสุขภาพ 2557;1(2):47-52

☞ สานเสวนาพาด่านบทความวิจัย

ความเข้าใจพื้นฐานต่อ non-inferiority test ในบทความวิจัยทางคลินิก

บทคัดย่อ

การทดสอบแบบ non-inferiority test พบมากขึ้นในบทความการวิจัยทางคลินิก แนวคิดและการแปลความหมายของการทดสอบต่างจากการทดสอบทางสถิติแบบดั้งเดิมที่เป็นการทดสอบความเหนือกว่าของยาใหม่หรือการรักษาแบบใหม่ เมื่อเทียบกับยา/การรักษามาตรฐานหรือยาหลอก ในขณะที่ non-inferiority test ทดสอบว่ายาใหม่หรือการรักษาแบบใหม่อาจดีด้อยกว่ายาเดิมหรือการรักษาแบบเดิมแต่ไม่เกินไปกว่าระดับที่ยอมรับได้ บทความนี้มุ่งหวังให้ผู้อ่านสามารถเข้าใจแนวคิดเบื้องต้นและการแปลความหมายของการทดสอบความแตกต่างระหว่างกลุ่มในบทความวิจัยทางคลินิก แบบ non-inferiority test พอสังเขป โดยเน้นที่การพิจารณาผลลัพธ์ของการทดลอง ทิศทางของความต่างระหว่างกลุ่ม และเกณฑ์ในการพิจารณาความไม่ด้อยกว่า โดยจุดเน้นคือ ให้ผู้อ่านได้ฝึกหัดระบุทิศทางของการทดสอบ เน้นการไม่จำกัดทางแบบสูตรสำเร็จ โดยวิธีที่ฝึกหัดในครั้งนี้อาจวาดทิศทางของความแตกต่างเป็นรูปภาพสำหรับแต่ละการทดสอบ

คำสำคัญ: การเปรียบเทียบระหว่างกลุ่ม, non-inferiority test, การทดสอบแบบเหนือกว่า, การทดสอบแบบดั้งเดิม

Charoen Treesak *

¹ Department of Clinical Pharmacy, Faculty of Pharmacy, Srinakharinwirot University, Ongkharak, Nakhonnayok, 26120 Thailand

* Corresponding author: charoen@g.swu.ac.th

Dialogue on Pharmacy and Health Care Practice 2014;1(2):47-52

☞ Dialogue on Research Paper Reading

Basic Understanding in Non-inferiority Test in Clinical Research Articles

ABSTRACT

Non-inferiority test has been used more frequently in clinical research paper. Concept and interpretation of non-inferiority test is different from traditional statistical test which tests whether new drug or treatment is superior to standard drug/treatment, or placebo. Non-inferiority test aims to determine whether new drug/treatment is not unacceptably less efficacious than an active control treatment already in use. This article aims to help the audience understand basic concept and interpretation of non-inferiority test in clinical research paper. The interpretation on study outcomes between-groups comparison especially direction of non-inferiority test is explained, as well as non-inferiority margin. The article emphasizes the practical consideration of determining direction of group difference for non-inferiority test with cases for actual visual exercise.

Keywords: between-groups comparison, non-inferiority test, superiority test, traditional test

บทนำ

การทดสอบความต่างทางสถิติระหว่างกลุ่มการทดลองแบบ non-inferiority test พบบ่อยมากขึ้นในบทความวิจัยทางคลินิก โดยการทดสอบลักษณะนี้ต่างจากการทดสอบแบบดั้งเดิม (traditional test) ที่มีพื้นฐานแนวคิดว่าการรักษาแบบใหม่หรือยาใหม่มีประสิทธิภาพมากกว่าการรักษาแบบเดิมหรือยาที่ใช้มาก่อนหน้า ดังนั้น จึงถือว่าการทดสอบแบบดั้งเดิมเป็นการทดสอบความ “เหนือกว่าของเดิม” (superiority test) แต่ non-inferiority test มีแนวคิดพื้นฐานในการทดสอบว่า การรักษาแบบใหม่หรือยาใหม่นั้น อาจด้อย (inferior) กว่า การรักษาแบบเดิมหรือยาเดิม แต่ด้อยไม่มากไปกว่าที่ยอมรับได้ ดังนั้น เมื่ออ่านบทความวิจัยทางคลินิก ฟังระวังไว้เสมอว่า non-inferiority test ไม่ได้ทดสอบตรง ๆ ว่า การรักษาแบบใหม่หรือยาใหม่ “ไม่ด้อย” กว่า การรักษาแบบเดิมหรือยาเดิม แต่เป็นการยอมให้ด้อยกว่าได้ แต่ความด้อยไม่มากกว่าที่จะยอมรับได้

รู้จัก non-inferiority test ในเบื้องต้น

ในบทความการวิจัยหนึ่ง มีวัตถุประสงค์เพื่อทดสอบประสิทธิภาพของ boosted protease inhibitor (PI) ร่วมกับ

integrase inhibitor หนึ่งตัว (lopinavir/ritonavir [LPV/r] + raltegravir [RAL] เทียบกับ boosted PI + 2 NRTIs (LPV/r + tenofovir/emtricitabine [TDF/FTC] ในผู้ป่วย HIV ผู้ใหญ่ที่ยังไม่เคยได้รับยาต้านไวรัสมาก่อน (antiretroviral (ARV)-naive adults infected with HIV-1) ที่พิจารณาแล้วว่าสูตรยา boosted PI + 2 NRTIs อาจไม่ใช่ทางเลือกที่ดีที่สุด (Reynes et al, 2011)¹ โดยผู้วิจัยหวังที่จะพิสูจน์ว่า LPV/r + RAL อาจด้อยกว่า LPV/r + TDF/FTC ในการกด HIV virus แต่ไม่ด้อยเกินกว่าจะยอมรับได้

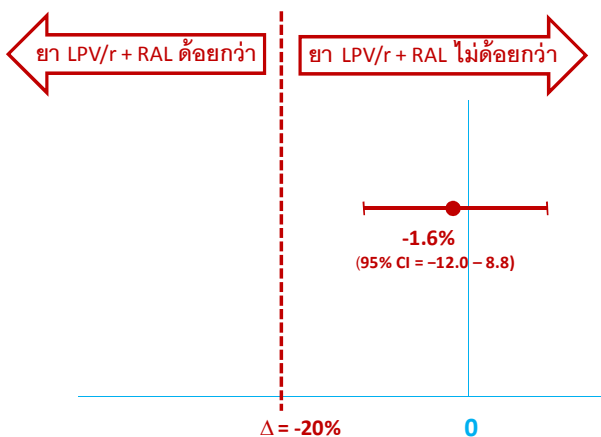
ในการศึกษานี้ ตัวแปรที่ใช้ทดสอบความต่างทางสถิติคือ % ของผู้ป่วยที่มี HIV RNA < 40 copies ต่อ มล. ที่ 48 สัปดาห์ ซึ่งหากยาสามารถกด HIV virus ได้ผล ค่า % ของผู้ป่วยที่มี HIV RNA < 40 copies ต่อ มล. ที่ 48 สัปดาห์ควรมีค่าสูง ดังนั้น ถ้า สมมติโดยทั่วไปว่า หาก LPV/r + RAL มีประสิทธิภาพด้อยกว่า LPV/r + TDF/FTC แล้ว ยา LPV/r + RAL อาจทำให้ได้ % ของผู้ป่วยที่มี HIV RNA < 40 copies ต่อ มล. ที่ 48 สัปดาห์ เพียง 10% ในขณะที่ LPV/r + TDF/FTC ซึ่งมีประสิทธิภาพสูงกว่า อาจได้ถึง 20% และเมื่อพิจารณาความต่างโดยตั้งค่า LPV/r + RAL แล้วลบด้วยค่าจาก LPV/r + TDF/FTC ก็จะได้ค่าเป็นลบดังนี้ 10%

- 20% = -10% ซึ่งสามารถแปลความจากตัวอย่างสมมติข้างต้นนี้ได้ว่า ยา LPV/r + RAL ต้องกว่ายา LPV/r + TDF/FTC ในสัดส่วนของผู้ป่วยที่มี HIV RNA < 40 copies ต่อ มล. ที่ 48 สัปดาห์อยู่ -10%

แต่ในความจริงแล้ว ในการศึกษาโดย Reynes et al (2011)¹ นี้ ผู้วิจัยคาดว่ายา LPV/r + RAL ต้องกว่ายา LPV/r + TDF/FTC แต่ผู้วิจัยได้ตัดสินใจกำหนดว่า คงไม่ต้องมากไปกว่า -20% โดยค่า -20% นี้ ถือเป็น non-inferiority margin (เส้นประในรูปที่ 1) ซึ่งผู้วิจัยกำหนดค่าความต่าง -20% นี้จากผลการศึกษาก่อนหน้าหลาย ๆ การศึกษา ดังนั้นผู้วิจัยจึงตั้งสมมติฐานว่า จะสรุปว่ายา LPV/r + RAL ต้องกว่ายา LPV/r + TDF/FTC ถ้าพบว่าขอบล่างของ 95% confidence interval (95% CI) ของค่าที่ลบกันนั้นทับหรือคร่อมเลข -20% ไปด้านซ้าย แต่ถ้าไม่ทับหรือไม่คร่อมก็แสดงว่าไม่ต้องกว่า LPV/r + TDF/FTC

ทั้งนี้ บทความวิจัยดังกล่าวได้แสดงเนื้อความที่ระบุเกณฑ์ non-inferiority margin ขนาด -20% สำหรับการทดสอบดังกล่าวไว้ในส่วน Statistical Analysis ว่า:

To test for noninferiority of LPV/r + RAL treatment compared with LPV/r + TDF/FTC treatment, the exact 95% confidence interval (CI) for the difference in response rates ([LPV/r + RAL] - [LPV/r + TDF/FTC]) was determined and compared to the pre-specified noninferiority margin of -20%.



รูปที่ 1 การพิจารณาความไม่ด้อยกว่า (non-inferiority) ของยาสูตร LPV/r + RAL เทียบกับยาสูตร LPV/r + TDF/FTC ในการลด HIV virus ที่ 48 สัปดาห์ (HIV RNA < 40 copies ต่อ มล.) โดยแกนนอนแสดงความต่างระหว่าง (% ผู้ป่วยในกลุ่ม LPV/r + RAL ที่มีค่า HIV RNA < 40 copies ต่อ มล.) ลบด้วย (% ผู้ป่วยในกลุ่ม LPV/r + TDF/FTC ที่มีค่า HIV RNA < 40 copies ต่อ มล.)

ผลการศึกษานี้แสดงให้เห็นว่ากลุ่ม LPV/r + RAL มี % ผู้ป่วยที่มี HIV RNA < 40 copies ต่อ มล. ที่ 48 สัปดาห์ เป็น 83.2% ซึ่งดูเหมือนด้อยกว่ากลุ่ม LPV/r + TDF/FTC ซึ่งพบ 84.8% โดยความต่างเป็น -1.6% และ 95% CI เป็น (-12.0% - 8.8%) เห็นได้ว่าขอบล่างของ 95%CI ไม่ทับและไม่คร่อมค่า -20% ไปด้านซ้าย

ดังนั้นความไม่ด้อยกว่านี้มีนัยสำคัญทางสถิติ จึงสรุปว่า สูตรยา LPV/r + RAL ไม่ต้องน้อยกว่า (non-inferior) ยาสูตร LPV/r + TDF/FTC เมื่อพิจารณาประสิทธิภาพการลด HIV virus จากค่า % ผู้ป่วยที่มี HIV RNA < 40 copies ต่อ มล. ที่ 48 สัปดาห์ โดยเทียบกับเกณฑ์ non-inferiority margin ขนาด -20% (ดูรูปที่ 1 ประกอบ)

ทิศทางของ non-inferiority test

ในตัวอย่างข้างต้น ผลลัพธ์ที่เปรียบเทียบกันระหว่างยาสูตรใหม่กับยาสูตรมาตรฐานนั้น เป็นผลลัพธ์ที่ค่าสูงหมายถึงสิ่งที่ดี ซึ่งหากยาสูตรใหม่ที่ถูกต้องทดสอบนั้นคาดว่าจะต้องกว่ายาสูตรมาตรฐานแล้ว ค่าความต่างก็จะเป็นลบ และค่าเกณฑ์ non-inferiority margin ที่ผู้วิจัยกำหนดก็เป็นค่าลบเช่นกัน ค่าถามคือ การศึกษาอื่น ๆ จะมีทิศทางในการพิจารณา non-inferiority แบบเดียวกันหรือไม่ หรือว่ามีทิศทางต่างกันได้ และต่างอย่างไร ตัวอย่างต่อไปนี้จะช่วยให้ผู้อ่านบทความวิจัยสามารถพิจารณาผลการทดสอบแบบ non-inferiority ได้ชัดเจนยิ่งขึ้น

ในบทความการวิจัยหนึ่ง มีวัตถุประสงค์เพื่อประเมินประสิทธิภาพของยา linagliptin (5 mg วันละครั้ง) ในการลดระดับ HbA1C ว่าไม่ต้องน้อยกว่า gliclazide (1 - 4 mg วันละครั้ง) โดยวัด HbA1C ที่ baseline แล้วให้ยานาน 104 สัปดาห์ แล้ววัด HbA1C อีกครั้ง (Gallwitz et al, 2012)² ทั้งนี้ ผู้วิจัยหวังที่จะพิสูจน์ว่า ยา linagliptin คงไม่ต้องน้อยกว่ายา gliclazide แต่จะใช้ความเด่นของ linagliptin ที่ทำให้เกิดอาการระดับน้ำตาลต่ำ (hypoglycemia) ได้น้อยกว่า gliclazide มาเป็นจุดสนับสนุนการใช้ linagliptin แทน gliclazide ในผู้ป่วยที่มักเกิด hypoglycemia บ่อย

ในการศึกษานี้ ตัวแปรที่ใช้ทดสอบความต่างทางสถิติคือ ค่า HbA1C ที่เปลี่ยนแปลงจาก baseline (change from baseline) ซึ่งหากยาสามารถลดระดับน้ำตาลในเลือดได้ผล ค่า HbA1C (ค่าเป็น %) ควรลดลง ทำให้ค่า change from baseline เป็นค่าติดลบ ดังนั้น ถ้าสมมติโดยทั่วไปว่า หาก linagliptin มีประสิทธิภาพด้อยกว่า gliclazide แล้ว ยา linagliptin อาจลด HbA1C ได้เพียง -0.20% ในขณะที่ gliclazide ซึ่งมีประสิทธิภาพสูงกว่า อาจลดได้ถึง -0.45% และเมื่อพิจารณาความต่างโดยตั้งค่า linagliptin แล้วลบด้วยค่าจาก gliclazide ก็จะได้ค่าเป็นบวก ดังนี้ -0.20% - (-0.45%) = +0.25% ซึ่งสามารถแปลความจากตัวอย่างสมมติข้างต้นนี้ได้ว่า ยา linagliptin ต้องกว่ายา gliclazide ในการลด HbA1C อยู่ +0.25%

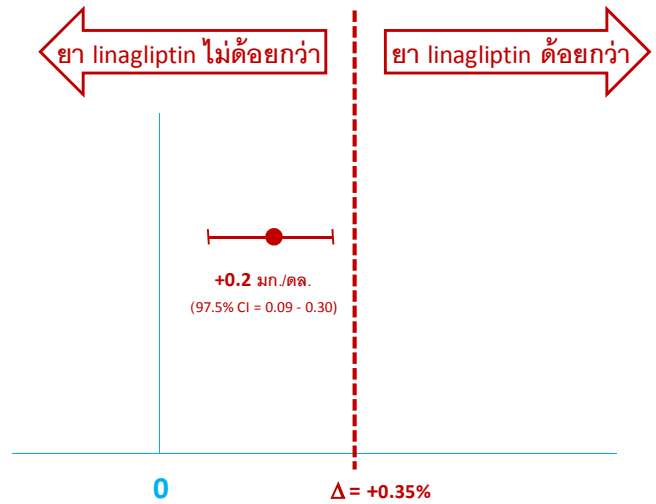
ซึ่งในความจริงแล้ว ในการศึกษาโดย Gallwitz et al (2012)² นี้ ผู้วิจัยคาดว่ายา linagliptin ต้องกว่ายา gliclazide แต่ผู้วิจัยได้ตัดสินใจกำหนดว่า คงไม่ต้องมากไปกว่า +0.35% โดยค่า +0.35% นี้ ถือเป็น non-inferiority margin (เส้นประในรูปที่ 2) ซึ่งผู้วิจัยกำหนดค่าความต่าง +0.35% นี้จากผลการศึกษาก่อนหน้าหลาย ๆ การศึกษา ดังนั้นผู้วิจัยจึงตั้งสมมติฐานว่า จะสรุปว่ายา

linagliptin ดีน้อยกว่ายา glimepiride ถ้าพบว่าขอบบนของ 97.5% confidence interval (97.5% CI) ของค่าที่ลบกันนั้นทับหรือคร่อมเลข +0.35% ไปด้านขวา แต่ถ้าไม่ทับหรือไม่คร่อมก็แสดงว่าไม่ดีกว่า glimepiride (ทั้งนี้ การตั้งค่า % confidence interval จะเป็น 95% หรือ 97.5% นั้นอาจแตกต่างกันในแต่ละบทความวิจัย แล้วแต่ความเคร่งครัดของการทดสอบ)

บทความวิจัยดังกล่าวได้แสดงเนื้อหาความที่ระบุเกณฑ์ non-inferiority margin ขนาด +0.35% สำหรับการทดสอบดังกล่าว ไว้ในส่วน Statistical Analysis โดยเฉพาะในส่วนที่ขีดเส้นใต้ว่า:

.... On the assumption of an SD of change in HbA1c from baseline of 1.3%, a sample size of 707 participants per treatment group was needed for 90% power to show non-inferiority through a 97.5% CI for treatment difference in the adjusted mean change from baseline to endpoint of less than 0.35% HbA1c at the level of $\alpha = 0.0125$ (one-sided).

โดยการทดสอบความแตกต่างทางสถิติ นั้น เป็นการทดสอบแบบ one-sided เนื่องจากมีทิศทางที่ชัดเจนว่า จะถือว่า ยา linagliptin ไม่ดีกว่า glimepiride หากความต่างนั้นไม่มากกว่า +0.35% ไม่ใช่เพียงแค่ทดสอบว่า “แตกต่างกันหรือไม่”



รูปที่ 2 การพิจารณาความไม่ดีกว่า (non-inferiority) ของยา linagliptin เทียบกับยามาตรฐาน glimepiride ในการลดระดับ HbA1C จาก baseline โดยเกณฑ์แสดงความต่างระหว่าง (ค่า HbA1C ที่เปลี่ยนแปลงจาก baseline ในกลุ่ม linagliptin) ลบด้วย (ค่า HbA1C ที่เปลี่ยนแปลงจาก baseline ในกลุ่ม glimepiride)

จากผลการศึกษาในตารางที่ 2 ที่ปรากฏในบทความดังกล่าว (ดูตารางที่เน้นตัวเลขในรูปที่ 3) พบว่ายา linagliptin ทำให้ระดับ

	Linagliptin	Glimepiride	Difference (linagliptin-glimepiride)		
			Adjusted* mean (SE, %)	CI	p value
HbA_{1c} in full analysis set (LOCF)					
n	764	755
Mean at baseline (SE, %)	7.69% (0.03)	7.69% (0.03)
Change from baseline					
Mean (SE, %)	-0.21% (0.03)	-0.41% (0.03)
Adjusted* mean (SE, %)	-0.16% (0.03)	-0.36% (0.03)	0.20% (0.05)	0.09-0.30†	0.0004‡
HbA_{1c} in PPS completers (OC)					
n	477	458
Mean at baseline (SE, %)	7.43% (0.04)	7.53% (0.04)
Change from baseline					
Mean (SE, %)	-0.37% (0.04)	-0.61% (0.04)
Adjusted* mean (SE, %)	-0.35% (0.04)	-0.53% (0.04)	0.17% (0.05)	0.07-0.28†	0.0001‡
HbA_{1c} in the completers cohort					
n	233	271
Mean at baseline (SE, %)	7.17% (0.04)	7.31% (0.04)
Change from baseline					
Adjusted* mean (SE, %)	-0.56% (0.03)	-0.63% (0.03)	0.08% (0.04)	0.00-0.15‡	0.0468¶

HbA_{1c} =glycated haemoglobin A_{1c}. LOCF=last observation carried forward. PPS=per-protocol set. OC=observed cases.
*Model includes treatment, baseline HbA_{1c}, and number of previous oral antidiabetic drugs. †97.5% CI. ‡p<0.0125, one-sided. §95% CI. ¶p<0.05, two-sided.

Table 2: Change in HbA_{1c} from baseline to week 104

รูปที่ 3 เป็น Table 2 จากบทความวิจัยของ Gallwitz et al (2012)² ที่ดัดแปลงเพื่อแสดงค่า change from baseline ของค่า HbA1C ในกลุ่มที่ใช้ยา linagliptin และยา glimepiride และความแตกต่างของค่าดังกล่าวพร้อมค่า 97.5% confidence interval ตามที่ล้อมรอบด้วยกรอบเส้นประ

HbA1C ที่ 104 สัปดาห์ลดลงจาก baseline ได้เพียง -0.16% ในขณะที่ glimepiride ลดลงได้ถึง -0.36% คิดเป็นความต่างเท่ากับ $-0.16\% - (-0.36\%) = +0.20\%$ หากพิจารณาคร่าว ๆ ก็เหมือนว่า linagliptin ดีกว่า glimepiride แต่เมื่อพิจารณาว่าขอบบนของ 97.5%CI ซึ่งเท่ากับ +0.30% ก็จะเห็นว่าไม่ทับหรือไม่คร่อมค่า +0.35% ที่ผู้วิจัยกำหนดให้เป็นเกณฑ์ non-inferiority margin (พิจารณารูปที่ 2 ร่วมด้วย) ดังนั้นการลดระดับ HbA1C โดย linagliptin นั้น ไม่ดีไปกว่า (non-inferior) ยา glimepiride อย่างมีนัยสำคัญทางสถิติ

ทั้งนี้ ผู้วิจัยได้แสดงเนื้อหาความไม่บทความที่ระบุถึงการทดสอบ non-inferiority ดังกล่าวดังนี้:

After 2 years of treatment, linagliptin was non-inferior to glimepiride in reducing HbA1c (table 2). At week 104, adjusted mean changes in HbA1c from a baseline of 7.7% were -0.16% with linagliptin and -0.36% with glimepiride in the full analysis set; the difference between treatment groups met the non-inferiority criterion and was 0.20% (97.5% CI 0.09–0.30; $p = 0.0004$, < 0.0125 [one-sided]).

จะแปลความกรณีที่ยาใหม่บางตัวดูเหมือนไม่ดีกว่ายามาตรฐานอย่างไร

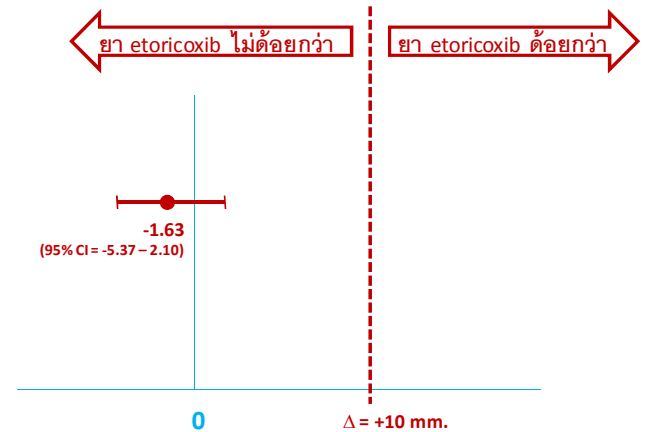
จากตัวอย่างข้างต้น ยา linagliptin ทำให้ระดับ HbA1C ลดลงจาก baseline ได้น้อยกว่ายา glimepiride ซึ่งถือเป็นยามาตรฐาน กรณีเช่นนี้ถือเป็นตัวอย่างการเปรียบเทียบที่ตรงไปตรงมาที่สุดสำหรับ non-inferiority test อย่างไรก็ตาม ในบางการศึกษาที่ตั้งใจทดสอบแบบ non-inferiority test นั้น ยาใหม่บางตัวให้ประสิทธิภาพที่สูงกว่ายามาตรฐาน กรณีเช่นนี้จะแปลความหมายของ non-inferiority test อย่างไร

การวิจัยหนึ่งซึ่งผู้วิจัยต้องการทดสอบเปรียบเทียบประสิทธิภาพของยา etoricoxib (30 mg วันละครั้ง) ในการลดปวดใน osteoarthritis ว่าไม่ดีกว่า celecoxib (200 mg วันละครั้ง) โดยวัดความปวดด้วย WOMAC VA 3.0 Pain Subscale ซึ่งเป็น visual analogue scale (VAS) ยาว 100 มิลลิเมตร โดยประเมินที่ baseline แล้วให้ยา 12 สัปดาห์แล้ววัดความปวดอีกครั้ง (Yoo et al, 2014)³

ในการศึกษาที่ ตัวแปรที่ใช้ทดสอบความต่างทางสถิติคือ ค่าความปวด WOMAC Pain VAS ที่เปลี่ยนแปลงจาก baseline (change from baseline) ซึ่งหากยาสามารถลดความปวดได้ ค่า WOMAC Pain VAS ควรลดลง (ค่าเป็น มม. โดยค่ามากแสดงว่าปวดมาก) ทำให้ค่า change from baseline เป็นค่าติดลบ ดังนั้นหากคิดว่า etoricoxib ดีกว่า celecoxib ก็จะต้องพบว่า ค่า change from baseline ในกลุ่ม etoricoxib ลดด้วย ค่า change from

baseline ในกลุ่ม celecoxib ก็เป็นค่าบวก (+) เช่นเดียวกับที่พบในตัวอย่างก่อนหน้านี้

ในการศึกษาที่ ผู้วิจัยคาดว่า etoricoxib ดีกว่า celecoxib แต่ผู้วิจัยได้กำหนดว่า คงไม่ดีกว่าไปกว่า +10 มม. โดยค่า +10 มม. นี้ ถือเป็น non-inferiority margin (เส้นประในรูปที่ 4) ซึ่งผู้วิจัยกำหนดค่าความต่าง +10 มม. นี้ จากผลการศึกษาก่อนหน้าหลาย ๆ การศึกษา ดังนั้น ผู้วิจัยจึงตั้งสมมติฐานว่า จะสรุปว่ายา etoricoxib ดีกว่ายา celecoxib ถ้าพบว่าขอบบนของ 95% CI ของค่าที่ลบกันนั้นทับหรือคร่อมเลข +10 มม. ไปด้านขวา แต่ถ้าไม่ทับหรือไม่คร่อมก็แสดงว่าไม่ดีกว่า celecoxib



รูปที่ 4 การพิจารณาความไม่ดีกว่า (non-inferiority) ของยา etoricoxib เทียบกับยามาตรฐาน celecoxib ในการลดความปวดจาก baseline (วัดโดย WOMAC Pain VAS) โดยแกนนอนแสดงความต่างระหว่าง (ค่า WOMAC Pain VAS ที่เปลี่ยนแปลงจาก baseline ในกลุ่ม etoricoxib) ลบด้วย (ค่า WOMAC Pain VAS ที่เปลี่ยนแปลงจาก baseline ในกลุ่ม celecoxib)

ทั้งนี้ บทความวิจัยดังกล่าวได้แสดงเนื้อหาความไม่บทความที่ระบุเกณฑ์ non-inferiority margin ขนาด +10 มม. สำหรับการทดสอบดังกล่าว ไว้ในส่วน Statistical Analysis ว่า:

Etoricoxib 30 mg was considered non-inferior to celecoxib 200 mg if the upper bound of the two-sided 95% confidence interval (CI) of the between-treatment difference (etoricoxib minus celecoxib) in least-squares (LS) mean change from baseline for this endpoint was no greater than 10 mm VAS (noninferiority margin).

และเช่นเดิม การทดสอบความแตกต่างทางสถิติสำหรับ non-inferiority test นั้น ใช้แนวคิดของการทดสอบแบบ one-sided เนื่องจากมีทิศทางที่ชัดเจนว่า จะถือว่า ยา etoricoxib ไม่ดีกว่า celecoxib หากความต่างนั้นไม่มากกว่า +10 มม. ดังที่กล่าวไว้ใน Statistical Analysis โดยเฉพาะส่วนที่ขีดเส้นใต้ ดังนี้:

One hundred and five evaluable patients per treatment group would achieve approximately 90% power to demonstrate that etoricoxib is non-inferior to celecoxib with

respect to the primary endpoint, the time-weighted average change from baseline over 12 weeks in WOMAC Pain Subscale (VAS), at a one-sided, 2.5% alpha level.

จากผลการศึกษาในตารางที่ 2 ที่ปรากฏในบทความดังกล่าว (ดูตารางที่เน้นตัวเลขในรูปที่ 5) พบว่ายา etoricoxib ทำให้ความปวด (วัดโดยค่า WOMAC Pain VAS) ที่ 12 สัปดาห์ลดลงค่อนข้างมาก คือมีค่า mean change from baseline เป็น -21.40 มม. ซึ่งดูเหมือนจะดีกว่า celecoxib เสียด้วยซ้ำ (ความปวดลดลง -19.76 มม.) คิดเป็นความต่างเท่ากับ -21.40 มม. - (-19.76 มม.) = -1.63 มม. (โดยมีค่า 95% CI เท่ากับ -5.37 - 2.10) จะเห็นว่า etoricoxib ไม่ด้อยกว่า celecoxib เพราะขอบบนของ 95% CI เป็นค่า +2.10 ซึ่งไม่คร่อมหรือไม่ทับ +10.0 มม. ซึ่งเป็นเกณฑ์ของ non-inferiority margin ที่กำหนดในการศึกษานี้ (ดูรูปที่ 4 ประกอบด้วย)

ทั้งนี้ ดูเหมือนว่า etoricoxib อาจจะเหนือกว่า celecoxib ด้วยซ้ำ เพราะลดความปวดได้มากกว่า celecoxib แต่การทดสอบนัยสำคัญทางสถิติแบบดั้งเดิม (traditional test หรือ superiority test) ไม่ใช่วัตถุประสงค์หลักของการศึกษานี้ ดังนั้นจึงแปลความหมายทางสถิติได้เพียงว่า etoricoxib ไม่ด้อยกว่า celecoxib ในการลดความปวดในผู้ป่วย osteoarthritis ชาวจีน ตามที่วัดความปวดด้วย WOMAC Pain VAS ในการใช้ยา 12 สัปดาห์ ไม่ควรแปลว่า ยา etoricoxib บรรเทาอาการปวดได้เหนือกว่า celecoxib

กล่าวโดยสรุป ในการอ่านบทความวิจัยที่ทดสอบ non-inferiority test นั้น ไม่ควรพิจารณาทิศทางของความต่างของผลลัพธ์ระหว่างกลุ่มโดยการจำ แต่ควรพิจารณาทิศทางเป็นกรณี ๆ ไป และการวาดภาพเป็นวิธีหนึ่งที่จะช่วยให้เข้าใจทิศทางความแตกต่างได้ง่ายขึ้น

Table 2. Summary of efficacy endpoints.

	Etoricoxib 30 mg Mean (95% CI)	Celecoxib 200 mg Mean (95% CI)	Difference in LS Mean Change (95% CI)
LS Mean Time-weighted Average Change from Baseline over 12 Weeks in WOMAC VA 3.0 Pain Subscale (VAS; mm)	-21.40 (-24.06, -18.74)	-19.76 (-22.41, -17.12)	-1.63 (-5.37, 2.10)
Time-weighted Average Change from Baseline over 12 Weeks in WOMAC VA 3.0 Physical Function Subscale (VAS; mm)	-17.78 (-20.31, -15.25)	-16.46 (-18.98, -13.93)	-1.32 (-4.88, 2.23)
Time-weighted Average Change from Baseline over 12 Weeks in Patient Global Assessment of Disease Status (VAS; mm)	-24.71 (-27.82, -21.60)	-23.61 (-26.73, -20.49)	-1.09 (-5.84, 3.30)
Time-weighted Average Response over 12 Weeks in Patient Global Assessment of Response to Therapy (Likert)	1.48 (1.35, 1.61)	1.56 (1.43, 1.69)	-0.08 (-0.26, 0.10)
Time-weighted Average Change from Baseline over 12 Weeks in Investigator Global Assessment of Disease Status (Likert)	-1.71 (-1.83, -1.59)	-1.62 (-1.74, -1.50)	-0.09 (-0.25, 0.08)
Time-weighted Average Change from Baseline over 12 Weeks in WOMAC VA 3.0 Stiffness Subscale (VAS; mm)	-18.41 (-21.33, -15.50)	-17.55 (-20.46, -14.65)	-0.86 (-4.96, 3.24)
Time-weighted Average Response over 12 Weeks in Investigator Global Assessment of Response to Therapy (Likert)	1.39 (1.25, 1.53)	1.49 (1.35, 1.64)	-0.10 (-0.30, 0.10)
Responder Rates (% of patients with 50% improvement) Based on the Time-weighted Average Change in WOMAC VA 3.0 Pain Subscale (VAS; mm)	32.8% (39/119)	24.4% (29/119)	8.40 (-3.07, 19.79)
Responder Rates (% of patients with 50% improvement) Based on the Time-weighted Average Change in WOMAC VA 3.0 Physical Function Subscale (VAS; mm)	25.2% (30/119)	23.5% (28/119)	1.68 (-9.19, 12.58)
Responder Rates Based on the Time Weighted Average Percent Change from Baseline over 12 Weeks in Patient Global Assessment of Disease Status (VAS; mm)	35.3% (42/119)	26.1% (31/119)	9.24 (-2.52, 20.82)

LS = Least Square; VAS = Visual Analog Scale; WOMAC VA 3.0 = Western Ontario and McMaster Universities Arthritis Index Version 3.0.

รูปที่ 5 เป็น Table 2 จากบทความวิจัยของ (Yoo et al, 2014)³ ที่ดัดแปลงเพื่อแสดงค่า change from baseline ของค่า WOMAC Pain VAS ในกลุ่มที่ใช้ยา etoricoxib และยา celecoxib และความแตกต่างของค่าดังกล่าวพร้อมค่า 97.5% confidence interval ตามที่ล้อมรอบด้วยกรอบเส้นประ

ทำไมผู้วิจัยจึงใช้ non-inferiority test

มีกรณีใดบ้างที่นักวิจัยตัดสินใจใช้ non-inferiority test แทนที่จะใช้ traditional test (หรือ superiority test) ที่ใช้กันอย่างกว้างขวาง กรณีต่อไปนี้อาจเป็นเหตุผลในการทดสอบแบบ non-inferiority test⁵⁻⁸ เช่น การทดลองที่ใช้ยาหลอก (placebo) เป็นสิ่งที่ไม่ควรทางจริยธรรม ดังนั้นต้องเปรียบเทียบยาใหม่กับยามาตรฐานที่ใช้กันอย่างแพร่หลาย แต่การเปรียบเทียบกับยามาตรฐานอาจไม่ทำให้เห็นความต่าง และผู้วิจัยอาจทราบอยู่แล้วว่ายาใหม่ไม่ดีกว่ายามาตรฐานมากนัก จึงเปรียบเทียบเพื่อพิสูจน์ว่าไม่ด้อยกว่า ทั้งนี้ ยาใหม่อาจมีจุดเด่น เช่น มีความปลอดภัยสูงกว่า ยามาตรฐาน ราคาถูกกว่า วิธีการใช้ยาสะดวกกว่า หรือความร่วมมือในการใช้ยาดีกว่า เป็นต้น เหตุผลที่สอง คือ ยาใหม่อาจไม่ได้ดีกว่ายามาตรฐานในผลลัพธ์หลัก (primary outcome) แต่อาจดีกว่าในผลลัพธ์รอง (secondary outcome) ดังนั้น จึงทดสอบแบบ non-inferiority test สำหรับผลลัพธ์หลัก และทดสอบแบบ superiority test สำหรับผลลัพธ์รอง

สรุป

บทความนี้มุ่งหวังให้ผู้อ่านสามารถเข้าใจการแปลความหมายของการทดสอบความแตกต่างระหว่างกลุ่มในบทความวิจัยทางคลินิก แบบ non-inferiority test พอสังเขป โดยเน้นที่การพิจารณาผลลัพธ์ของการทดลอง ทิศทางของความต่างระหว่างกลุ่ม และเกณฑ์ในการพิจารณาความไม่ด้อยกว่า โดยจุดเน้นคือให้ผู้อ่านได้ฝึกหัดระบุทิศทางของการทดสอบ เน้นการไม่จำกัดทิศทางแบบสูตรสำเร็จ โดยวิธีที่ฝึกหัดในครั้งนี้คือการวาดทิศทางของความแตกต่างเป็นรูปภาพสำหรับแต่ละการทดสอบ

สำหรับประเด็นอื่น ๆ เช่น การพิจารณาความสมเหตุสมผลในการทดสอบแบบ non-inferiority test กลุ่มประชากรสำหรับการทดสอบทางสถิติที่เหมาะสมสำหรับ non-inferiority test (เช่น

intention to treat population หรือ per protocol population) ขนาดกลุ่มตัวอย่างที่เหมาะสม ความเหมาะสมของการเลือก non-inferiority margin และการทดสอบทั้ง non-inferiority และ superiority ในผลลัพธ์ตัวเดียวกัน และการทดสอบแบบ equivalence test ฯลฯ จะได้กล่าวถึงในโอกาสต่อไป

References

1. Reynes J, Lawal A, Pulido F, et al. Examination of noninferiority, safety, and tolerability of lopinavir/ritonavir and raltegravir compared with lopinavir/ritonavir and tenofovir/emtricitabine in antiretroviral-naïve subjects: The PROGRESS study, 48-week results. *HIV Clin Trials* 2011;12(5):255-267.
2. Gallwitz B, Rosenstock J, Rauch T, et al. 2-year efficacy and safety of linagliptin compared with glimepiride in patients with type 2 diabetes inadequately controlled on metformin: a randomised, double-blind, non-inferiority trial. *Lancet* 2012;380(9840):475-483.
3. Yoo MC1, Yoo WH, Kang SB, et al. Etoricoxib in the treatment of Korean patients with osteoarthritis in a double-blind, randomized controlled trial. *Curr Med Res Opin* 2014;30(12):2399-2408.
4. Schumi J, Wittes JT. Through the looking glass: understanding non-inferiority. *Trials* 2011;12:106. (Accessed on Jun. 5, 2014, at <http://www.trialsjournal.com/content/12/1/106>)
5. Hahn S. understanding noninferiority trials. *Korean J Pediatr* 2012;55(11):403-407.
6. Gupta SK. Non-inferiority clinical trials: Practical issues and current regulatory perspective. *Indian J Pharmacol* 2011;43(4):371-374.
7. Walker E, Nowacki AS. Understanding equivalence and noninferiority testing. *J Gen Intern Med* 2011;26(2):192-196.
8. da Silva GT, Logan BR, Klein JP. Methods for equivalence and noninferiority testing. *Biol Blood Marrow Transplant* 2008;15(1 suppl):120-127.