

Interval kepercayaan dalam analisis kemaknaan klinis

Sunarto dan Achmad Surjono

Bagian Ilmu Kesehatan Anak, Fakultas Kedokteran, Universitas Gadjah Mada, Yogyakarta

ABSTRACT

Sunarto & Achmad Surjono - *Confidence interval in the analysis of clinical significance*

In the clinical study researchers want to answer the most important question whether a new therapy is better than the old one. Many researchers continue to frame the question in terms of null hypothesis and answer the question in terms of P value. The null hypothesis itself is typically not plausible, and even in a study where the null hypothesis was plausible, our concern is typically not only limited to the issue of whether or not the treatment has any effect but we also want to know how much the impact. In the study which the researcher pressed the P value into service as an indicator of effect size, it lends itself to misinterpretation because it combines information about the magnitude of an effect with information about the precision with which that effect is estimated. By contrast, confidence intervals (CIs) focus one's attention on an estimate of a more meaningful parameter (e.g. the rate difference) and, as a separate matter, on the precision of the estimate. The CI is a range of values that is likely to cover the true but unknown value (the extremely low up to the extremely high value of e.g. rate difference, mean difference, and Odds Ratio) if we measure the value many times on samples using the same method. CIs in a clinical trial where the result is statistically significant we might find that it is of no clinical importance (in a very large sample). On the other hand, the effect of a treatment might be statistically not significant but in fact it is of clinical importance (in a small sample). CI is affected by sample size: the larger the sample size the narrower the interval of CI. The interval is also affected by the standard error and hence by standard deviation and the confidence level we claim. The higher the confidence level (90%, 95% or 99%, arbitrarily 95% CI is commonly used) the wider the confidence interval. In case that the 95% CI does not include zero value (in mean or proportion difference) or one (in Odds ratio or relative risk) it also reflects statistical significance ($p < 0.05$) with $\alpha = 5\%$. A sample size that is enough in terms of power might not be enough in terms of precision due to the confidence interval level we choose.

Key Words : P value - confidence interval - clinical importance - sample size - statistical significant

ABSTRAK

Sunarto dan Achmad Surjono - *Interval kepercayaan dalam analisis kemaknaan klinik*

Dalam penelitian klinik peneliti ingin menjawab apakah suatu terapi baru mempunyai efek lebih baik dibanding cara lama. Sejak lama dan hingga kini masih berlangsung, peneliti berusaha menjawab pertanyaan itu dengan menyusun hipotesis nol dan kemudian membuktikan berdasarkan nilai p yang ditemukan. Hipotesis nol sendiri pada hakekatnya mengandung kekurangan. Bahkan pada penelitian di mana hipotesis nol tepat untuk dibuktikan, perhatian kita pada penelitian klinik tidak terbatas sekedar membuktikan apakah suatu terapi mempunyai efek berbeda dengan terapi sebelumnya atau terapi lain. Kita juga ingin mengetahui seberapa besar dampak dari terapi. Pada penelitian klinik bila peneliti terlalu berkiblat pada nilai p sebagai dasar untuk memutuskan manfaat suatu terapi misalnya, ia dapat terbawa ke arah salah interpretasi karena nilai p menggabungkan informasi mengenai besarnya efek dan presisi. Sebaliknya, interval kepercayaan (IK) memusatkan perhatian pada estimasi terhadap parameter yang berarti dan terhadap presisi dari estimasi sebagai masalah yang terpisah. IK adalah rentang nilai yang sangat mungkin mencakup nilai sesungguhnya yang tidak diketahui (mencakup nilai rendah ekstrem sampai nilai tinggi ekstrem dari beda frekuensi, beda rerata atau *Odds ratio*) jika kita melakukan pengukuran nilai tadi terhadap sampel berulang kali dengan cara yang sama. IK dapat menunjukkan beda atau tidak beda secara klinik yang menggambarkan bermanfaat atau tidaknya secara klinik (misalnya suatu cara terapi baru). Pada penelitian klinik (bila sampel besar sekali) dapat kita temukan perbedaan statistik bermakna, tetapi beda itu secara klinik tidak ada artinya. Sebaliknya, efek dua cara terapi yang secara statistik tidak berbeda bermakna (sampel kecil), sebenarnya secara klinik mempunyai arti penting. IK dipengaruhi oleh

ukuran sampel (makin besar sampel makin sempit rentang IK) dan oleh *standard error*, jadi oleh *standard deviation*. IK juga dipengaruhi oleh tingkat kepercayaan yang dituntut: makin tinggi IK (90%, 95% atau 99%; arbitrer IK 95% yang sering digunakan). Bila IK tidak mencakup nilai nol (pada beda rerata atau proporsi) atau nilai satu (pada *Odds ratio*) maka IK menggambarkan juga perbedaan statistik bermakna dengan $\alpha = 5\%$ ($p < 0,05$). Ukuran sampel yang cukup besar untuk persyaratan power mungkin tidak cukup besar untuk presisi disebabkan oleh nilai IK yang kita pilih.

(B.I.Ked. Vol 29, No. 3:145-151, September 1997)

PENGANTAR

Dalam penelitian klinis mengenai terapi, pertanyaan utama yang harus dijawab adalah: apakah efek terapi secara klinis mempunyai arti atau tidak. Untuk menjawab pertanyaan itu sering ditempuh dengan cara membuat hipotesis nol dan menolak hipotesis nol atas dasar nilai p terhitung. Akar dari analisis dengan uji kemaknaan statistik ini diturunkan dari topik-topik penelitian yang menarik perhatian para peneliti pada awal abad ini yang merupakan pelopor dalam perkembangan teori statistik. Masalah penelitian mereka adalah bidang industri dan pertanian, dan yang khas menyangkut percobaan yang membentuk dasar untuk menentukan pilihan - antara dua atau lebih jalur alternatif - untuk bertindak. Percobaan itu dirancang untuk mendapatkan hasil yang memungkinkan peneliti membuat keputusan. Metode statistik yang diterapkan dimaksudkan untuk memfasilitasi pengambilan keputusan. Nilai p hanya memberi jawaban apakah suatu perlakuan mempunyai dampak atau tidak. Sedangkan pada penelitian terapi di samping ada tidaknya dampak kita juga ingin tahu seberapa besar dampak perlakuan dan seberapa tepat perkiraan efek terapi^{1,2,3}. Ini sejalan dengan tujuan utama percobaan klinik yaitu memberikan pertimbangan manfaat dan risiko (*Risk-benefit assessment*) suatu perlakuan. Masih banyak penelitian klinis membuat simpulan atas dasar nilai p , bahkan juga penelitian dengan pengukuran berulang dari waktu ke waktu (*serial measurements*) padahal sebenarnya ada uji statistik yang lebih baik⁴. Banyak peneliti yang menyatakan, bahwa penelitian klinik yang dilaporkan bermakna secara statistik di(salah)tafsirkan bermakna secara substantif padahal tidak selalu demikian^{2,5}. Bila uji kemaknaan statistik dapat menimbulkan salah interpretasi, bagaimana sebaiknya analisis kemaknaan klinik? Sejak tahun 70-an para pakar memperlakukan penggunaan nilai p pada penelitian kli-

nis, dan menganjurkan penggunaan interval kepercayaan (*IK/confidence interval*) untuk mengambil keputusan klinik¹, karena hasil penelitian klinis perlu dinyatakan sebagai suatu masalah estimasi dan IK memberi kita suatu rentang nilai yang kompatibel dengan data⁶. Menurut Poole (cit. Borenstein)² penggunaan IK dapat menghindarkan jenis-jenis salah-interpretasi di mana hipotesis nol tidak ditolak. Para penulis yang memasukkan karangan ke *British Medical Journal* sejak 1 Juli 1986 diharapkan menghitung IK jika data yang ada memungkinkan untuk itu⁷.

Dalam makalah ini akan diuraikan IK sebagai analisis untuk menyimpulkan kemaknaan klinik, dengan sedikit menyinggung masalah nilai p .

PEMBAHASAN

Nilai p

Meninjau sepintas nilai p untuk analisis data klinik, sebagai gambaran kita lihat TABEL 1 dan TABEL 2⁵.

Dengan hanya melihat harga p saja, kita akan berkesimpulan bahwa obat A (TABEL 1) memberi efek yang berbeda sangat bermakna dibandingkan dengan obat lama sehingga obat A akan digunakan sebagai alternatif terpilih. Di lain pihak, obat B (TABEL 2) tidak memberi efek berbeda dengan obat lama sehingga kita abaikan sebagai alternatif. Dengan memperbesar sampel pada TABEL 2 ternyata justru hal sebaliknya yang dapat disimpulkan (lihat TABEL 3). Harus diingat, bahwa nilai p yang sangat bermakna dapat terjadi pada perbedaan yang dapat diabaikan, jika sampel sangat besar; dan nilai p yang tidak mengesankan dapat terjadi pada hubungan kuat jika sampel kecil^{1,5}. Nilai-nilai (0,001 pada TABEL 1 dan 0,09 pada TABEL 2) juga tidak mampu menunjukkan penemuan deskriptif, yaitu beda proporsi 8% (beda keberhasilan obat baru dibandingkan obat lama pada TABEL 1), dan beda 24% pada TABEL 2. Masalah lain menge-

nilai p adalah harus ditarik garis pemisah antara "bermakna" dan "tidak bermakna" untuk menentukan pilihan obat. Garis pemisah demikian tidak ada dan biasanya bersifat arbitrer secara konvensi. Pemisah itu ditaruh pada tingkat 0,05. Satu-satunya keuntungan dari pilihan angka ini hanyalah bahwa dia berkaitan erat dengan deviasi dalam distribusi normal yang setara dengan dua simpang baku (SB). Jadi, batas pemisah tadi mengaitkan diri pada ilmu hitung. Konvensi itu tidak merugikan hanya apabila dapat dihindari anggapan bahwa hasil 0,04 berarti terbukti kuat dan anggapan bahwa tingkat kemaknaan 0,06 berarti berlabel tidak bermakna dan dapat diabaikan sama sekali⁸.

TABEL 1. - Efek obat baru A dibandingkan obat lama

Obat yang diberikan	Ada efek	Tak ada efek	Jumlah
Obat A	480 (60%)	320	800
Obat lama	416 (52%)	384	800

$p = 0,001$

TABEL 2. - Efek obat baru B dibanding obat lama

Obat yang diberikan	Ada efek	Tak ada efek	Jumlah
Obat B	15 (60%)	10	25
Obat lama	9 (39%)	16	25

$p = 0,09$

Interval Kepercayaan

Kalau nilai p merupakan satu titik arbitrer, nilai IK tidak terikat pada satu nilai seperti rerata, beda rerata, taksiran titik dan sebagainya tetapi merupakan suatu rentang nilai. Lebar IK tergantung pada SE (*standard error*) dan karena itu tergantung kepada SB (*SD/standard deviation*) dan pada ukuran sampel (*sampel size*). IK juga tergantung pada tingkat "kepercayaan" yang kita tetapkan untuk dikaitkan dengan hasil penelitian⁹.

IK dapat digambarkan sebagai berikut^{2,10}: misalkan, kita ambil sampel dari suatu populasi (misalnya bayi baru lahir di suatu Rumah Sakit dalam periode tertentu) kemudian diadakan pengukuran terhadap variabel tertentu (misalnya berat lahirnya) kita akan mendapat suatu nilai rerata berat bayi baru lahir dan simpang bakunya. Kalau pengukuran pada populasi itu dilakukan berulang kali (teoretis sampai tak terhingga kali) dengan teknik sampling yang sama, maka akan ditemukan sejumlah nilai rerata. Mudah dimengerti bah-

wa nilai-nilai rerata yang kita peroleh dari pengukuran sejumlah sampel tersebut akan tidak sama satu dengan yang lain. Nilai-nilai rerata itu, dari nilai rerata ekstrem rendah sampai ekstrem tinggi, merupakan *interval* (rentang) dan disebut IK.

Nilai IK dinyatakan bersama tingkat kepercayaan yang kita tuntut, misalnya 99%, 95%, 90%, atau lebih rendah lagi yang jarang dipakai. Jika misalnya, pada penimbangan bayi baru lahir tersebut di atas ditemukan IK 95% 3115,5-3304,5 g, ini dapat diartikan: bila kita lakukan penelitian berulang kali dengan teknik sampling yang sama pada populasi tersebut, kita akan mempunyai kepercayaan sebesar 95% bahwa rerata dari penelitian-penelitian itu akan terletak dalam interval 3115,5 sampai 3304,5 g dan nilai sesungguhnya yang di luar nilai ekstrem tadi hanya 5% dari kemungkinan^{3,9,10}.

IK juga berlaku untuk lebih dari satu jenis sampel. Misalnya pada penelitian efek terapi baru dibandingkan dengan terapi lama. Beda hasil terapi antara obat baru dibandingkan dengan obat lama (misalnya proporsi atau rerata) merupakan taksiran (*estimate*) dari efek obat yang kita sebut taksiran titik (*point estimate*). Beda yang sesungguhnya terletak di sekitar titik taksiran³ yang akan membentuk suatu rentang nilai, jika dimisalkan percobaan itu dilakukan berulang kali dengan cara yang sama. IK dapat diterapkan pada beda mean, beda proporsi, juga risiko relatif dan *Odds Ratio*.

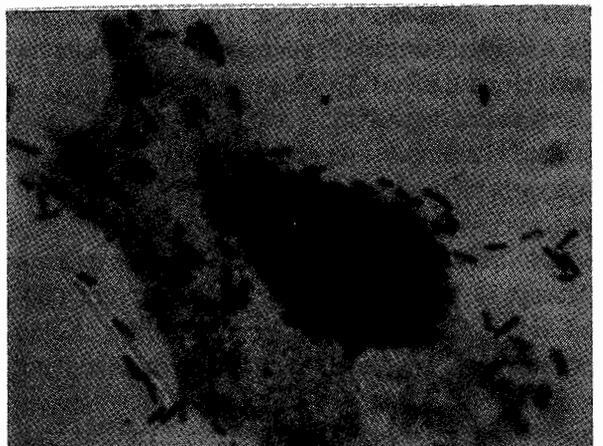
Untuk beda efek dua perlakuan, Borenstein² membuat rumusan umum:

$$\frac{\text{Nilai Atas}}{\text{Nilai Bawah}} = (\text{Besarnya Efek} \pm \text{Nilai Kritis} \times \text{SE})$$

Besarnya Efek adalah beda mean ($x_1 - x_2$) atau beda proporsi ($p_1 - p_2$). Nilai Kritis adalah $t_{1-\alpha/2}$ untuk data rasio dan $N_{1-\alpha/2}$ untuk proporsi (diperoleh dari tabel); α adalah kesalahan tipe 1. SE (*standard error*) untuk dua sampel adalah SE_{beda} .

Gardner dan Altman (1986)⁹ memerikan rumus dan contoh perhitungan IK untuk mean dan proporsi.

- Sampel tunggal, IK untuk rerata populasi: $x - (t_{1-\alpha/2} \times SE)$ sampai $x + (t_{1-\alpha/2} \times SE)$. (x = mean, α = kesalahan tipe 1, SE = *standard error*, dan nilai $t_{1-\alpha/2}$ sesuai dengan nilai dari



GAMBAR 1. - Pelekatan pada sel Hcp-2 galur kontrol positif dengan pengecatan Giemsa 10% pada perbesaran 1000x



GAMBAR 2. - Pelekatan pada sel Hcp-2 galur kontrol negatif dengan pengecatan Giemsa 10% pada perbesaran 1000x

tersebut juga dapat dilihat bahwa tingkat pelekatan isolat EPEC pada sel Hcp-2 tidak ada hubungannya dengan serotipe. Seperti yang terlihat pada galur-galur D14.10, D15.10 dan D11.4 semuanya berserotipe O128 tetapi memiliki tingkat pelekatan yang berlainan, yaitu $9,86 \pm 2,39$, $12,77 \pm 2,10$, dan $54,7 \pm 7,88$ (TABEL 3). GAMBAR 3 adalah pengamatan pelekatan EPEC 0011.2 pada sel Hcp-2.

Percobaan *in vitro* telah membantu mempelajari mekanisme pelekatan EPEC pada sel inang⁹. Interaksi EPEC dengan sel inang terjadi melalui tiga tahap. Tahap pertama adalah asosiasi bakteri dengan sel inang dalam ikatan tak intim yang diperantarai oleh pili. Ikatan ini terjadi secara *irreversible*. Pada tahap kedua pelekatan

GAMBAR 3. - Pelekatan EPEC0011.2 dengan pengecatan Giemsa 10% pada perbesaran 1000x



bakteri pada sel inang memicu terjadinya sinyal transduksi yang mengaktivasi tirosin kinase inang dan mengakibatkan kenaikan kalsium intraselular inang. Tahap ketiga terjadi pelekatan intim yaitu bakteri berasosiasi lebih dekat dengan inang dan mempengaruhi aktin di sekitar bakteri¹⁵. Protein membran luar dari bakteri dilaporkan berperan pada pelekatan intim. Ekspresi sifat protein membran luar EPEC dipengaruhi oleh kondisi lingkungan pertumbuhan bakteri. Bagaimana karakteristikasi ekspresi protein membran luar EPEC isolat Indonesia masih dalam penelitian.

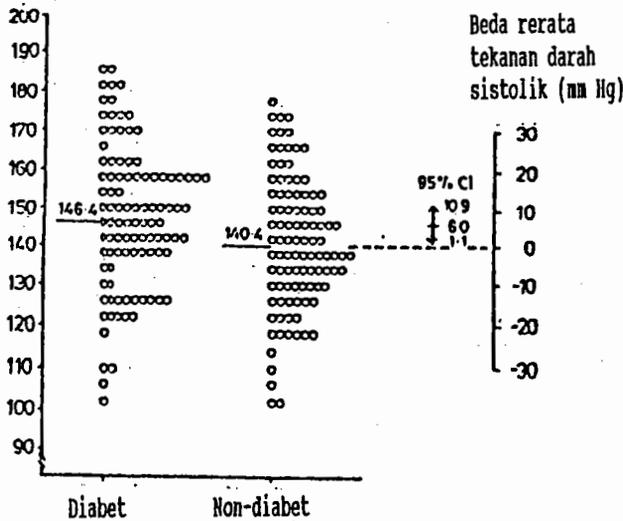
KESIMPULAN

Dari penelitian ini dapat disimpulkan bahwa bakteri EPEC didapatkan pada 55% penderita diare. Isolat-isolat EPEC memiliki serotipe beragam pada satu daerah dengan daerah lain. Serotipe O111 dan O86 didapatkan di seluruh daerah uji. Seluruh isolat EPEC melakukan pelekatan pada sel Hcp-2 dengan berbagai tingkat. Tingkat pelekatan tidak ditentukan oleh jenis serotipe.

UCAPAN TERIMA KASIH

Penelitian ini didbiayai oleh proyek Hibah Bersaing Perguruan Tinggi dengan nomor kontrak 008/P4M/DPPM/

Tekanan darah sistolik (mm Hg)

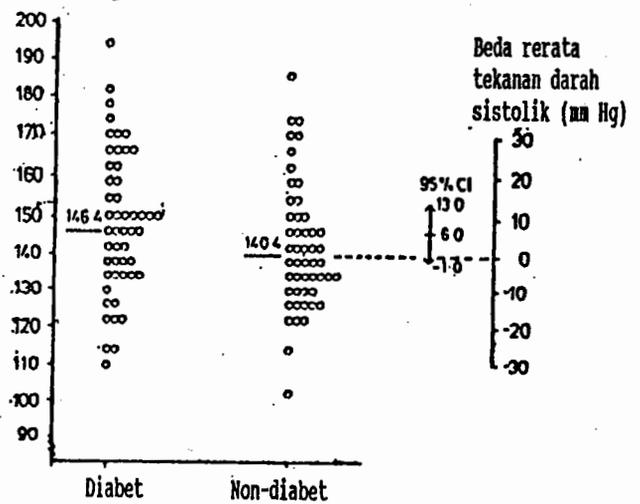


GAMBAR 1a. - Tekanan darah sistolik pada 100 pasien DM dan non-DM dengan nilai rerata berturut-turut 146,4 dan 140,4 mm Hg. Beda antara kedua rerata sampel 6,0 mm Hg ditunjukkan di sebelah kanan bersama-sama dengan interval IK 95% 1,1 sampai 10,9 mm Hg⁷

dianggap penting sebagai batas untuk menyatakan bermakna klinis atau tidak.

Ada hubungan erat antara IK dan uji hipotesis dua arah⁹. Dalam rumus IK tercantum nilai α , sehingga adanya hubungan itu mudah dimengerti. GAMBAR 1a memperlihatkan contoh: tekanan sistolik 100 pasien diabetes (DM) dengan rerata 146,4 mm Hg dan 100 non-DM dengan rerata 140,4 mm Hg; terdapat beda rerata sebesar 6 mm Hg. Dari perhitungan ditemukan IK 95% adalah 1,1 mm sampai 10,9 mm Hg. GAMBAR 1b memerikan tekanan sistolik 50 pasien DM dan non-DM dengan nilai rerata seperti pada GAMBAR 1a, tetapi ditemukan IK 95% -1 sampai 13,0 mm Hg. Pada skala sebelah kanan dari gambar 1a dan 1b tergambar titik nol, yaitu titik tidak ada perbedaan antara rerata tekanan darah sistolik kelompok DM dan non-DM. Perbedaan nol antara rerata kedua kelompok adalah yang diuji di bawah hipotesis nol; pada GAMBAR 1a rentang IK 95% terletak di atas titik nol. Titik nol terletak di luar rentang IK 95% itu menunjukkan adanya perbedaan statistik yang bermakna antara tekanan darah kedua kelompok pada $\alpha=5\%$ bila kita

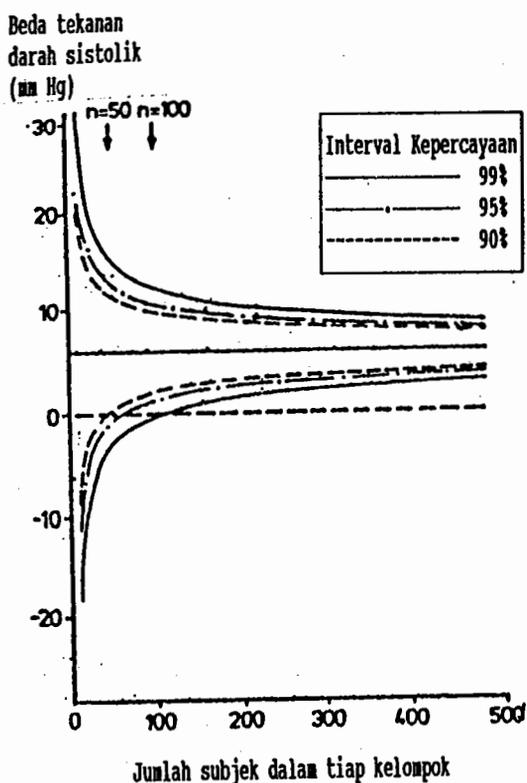
Tekanan darah sistolik (mm Hg)



GAMBAR 1b. - Seperti gambar 1a menunjukkan tekanan darah dari dua sampel dengan ukuran setengah (50 tiap kelompok). Nilai rerata dan simpang baku seperti pada GAMBAR 1a, tetapi interval Ik 95% lebih lebar, dari -1 sampai 13,9 mm Hg⁷.

terapkan uji t tak berpasangan. Sebaliknya, pada GAMBAR 1b titik nol terletak di dalam rentang IK 95%, ini berarti dengan sampel sebesar 50 pada masing-masing kelompok terbukti tidak ada perbedaan bermakna secara statistik. Hubungan IK dengan uji hipotesis juga tampak pada TABEL 3. Pada penelitian 1 dan 4 nilai IK 95% semuanya terletak di atas 0%; ini memberi petunjuk adanya perbedaan bermakna secara statistik. Pada penelitian 2 dan 3 rentang nilai IK 95% memotong garis nol, ini menunjukkan tidak adanya perbedaan statistik bermakna.

Pada GAMBAR 1a, IK 95% 1,1 - 10,9 mm Hg mencerminkan juga perbedaan bermakna secara statistik karena nilai terkecil (1,1 mm Hg) terletak di atas nol. Di samping itu dapat juga dibaca (tidak hanya bahwa beda mean tekanan sistolik sebesar 6 mm adalah kecil dan mungkin tidak mempunyai relevansi praktis), tetapi juga, kemungkinan bahwa beda tekanan sistolik yang sesungguhnya dalam populasi antara DM dan non-DM (sebesar 10,9 mm Hg atau lebih) hanya mempunyai peluang 2,5%⁷.



GAMBAR 2. - Interval Kepercayaan dari rerata-rerata dan simpang baku yang sama seperti pada GAMBAR 1, tetapi menunjukkan pengaruh ukuran-ukuran sampel sampai 500 subjek dalam tiap kelompok terhadap IK. Garis horizontal menunjukkan: ----- beda nol antara kedua rerata; _____ beda rerata hasil penelitian (6 mm Hg). Panah menunjukkan IK yang ditunjukkan dalam GAMBAR 1a dan 1b untuk sampel 100 dan 50 pada kelompok⁹.

Untuk menarik kesimpulan kemaknaan klinis, nilai batas yang ditentukan peneliti merupakan kunci. Pada TABEL 3, kalau nilai batas kemaknaan klinis tidak diambil 15% tetapi misalnya 20%, maka kesimpulan kemaknaan klinis akan lain. Menentukan batas ini mempunyai kepentingan praktis. IK dapat diterapkan untuk berbagai data seperti rerata, beda rerata, beda proporsi, regresi, *Odds ratio*, dan risiko relatif.

Pada TABEL 3 juga tampak bahwa IK 95% pada penelitian 2 dan 3 masing-masing mempunyai rentang nilai lebih lebar dibandingkan pada penelitian 1 dan 4, meskipun beda proporsi pada penelitian 1 sama dengan pada penelitian 2 (8%) dan pada 3 sama dengan pada 4 (24%). Ini disebabkan oleh ukuran sampel yang kecil. Makin besar ukuran sampel, makin sempit rentang nilai IK. GAMBAR 2 memperlihatkan pengaruh ukur-

an sampel. Dengan sampel sebesar 50 pada masing-masing kelompok, garis nol terletak dalam rentang nilai IK 95% (-1 sampai 13 mm Hg) berarti $p > 0,05$. Dengan memperbesar sampel menjadi 100 atau lebih (GAMBAR 1b), secara statistik perbedaan menjadi bermakna (garis nol terletak di luar rentang nilai IK 95%); dan rentang nilai IK menjadi lebih sempit yang menuntun ke kesimpulan dengan presisi yang lebih besar^{2,9}.

Pengaruh ukuran sampel juga terlihat pada rumus IK, yakni makin besar ukuran sampel makin kecil SE, makin sempit nilai IK terhitung. Secara logika dapat dimengerti bahwa makin besar ukuran sampel makin besar kemungkinan terambil subjek-subjek dengan nilai ekstrem besar sebanding dengan yang ekstrem kecil dan kedua nilai ekstrem tadi saling meniadakan sehingga rentang nilai menjadi sempit. Sebaliknya makin kecil sampel makin besar kemungkinan terambil hanya subjek dari kelompok ekstrem rendah atau tinggi saja, dan kemungkinan untuk meniadakan satu sama lain makin kecil, karena itu IK menjadi lebih lebar.

IK juga dipengaruhi oleh tingkat kepercayaan yang kita inginkan, makin tinggi tingkat kepercayaan yang kita inginkan makin lebar rentang nilai IK (GAMBAR 2)⁹. IK 99%, IK 95% dan IK 90% pada GAMBAR 1 berturut-turut adalah -0,5-13 mmHg, 1,1-10,9 mmHg, 1,9-10,1 mmHg. Secara intuitif mudah dimengerti, tidak mungkin kita akan percaya bahwa beda sesungguhnya tekanan sistolik DM-nonDM dalam populasi adalah tepat = 6 mmHg. Lebih besar kepercayaan kita bahwa beda itu adalah $(6 \pm a)$ mm Hg. Makin dekat nilai ekstrem rendah maupun nilai ekstrem tinggi dengan 6 mmHg tersebut (dengan lain perkataan makin sempit interval antara nilai ekstrem rendah dan ekstrem tinggi) makin kecil kepercayaan kita. Biasanya digunakan IK 95%, tetapi tidak ada keharusan demikian. IK 90% cenderung digunakan dalam penelitian-penelitian kecil untuk membatasi lebar IK, dan IK 99% pada penelitian besar yang menuntut kemaknaan statistik pada tingkat = 1%⁶. Dalam kaitan dengan ukuran sampel, IK diperlukan untuk menghitung besarnya sampel yang diperlukan agar sesuai dengan presisi (lebar IK) yang kita inginkan. Sampel dengan ukuran yang secara statistik cukup untuk *power* yang kita kehendaki (misalnya untuk *power* 80% diperlukan sampel untuk kelompok tindakan dan

kelompok kontrol masing-masing sebesar n subyek), mungkin untuk IK yang kita kehendaki ukuran sampel tadi tidak cukup besar dan sampel yang diperlukan dapat sampai dua kali lipat².

Dalam mengambil kesimpulan dari IK perlu diperhatikan bukan hanya satu nilai ekstrem saja yang ditinjau melainkan kedua nilai, baik nilai ekstrem atas maupun ekstrem bawah⁶. Bila IK tidak dicantumkan, dari nilai p kita dapat menemukan tiga cara pendekatan³. Pertama, bila $p=0,05$, ini berarti nilai bawah IK 95% adalah nol untuk beda efek (satu untuk risiko relatif), sehingga kita tidak dapat menyingkirkan kemungkinan perlakuan tidak punya efek. Bila nilai p menurun di bawah 0,05, ini berarti nilai bawah IK 95% meningkat di atas 0 (untuk efek) atau 1 (untuk risiko relatif). Kedua, dengan sedikit perhitungan menggunakan rumus IK bila tercantum nilai SE dari beda efek (atau risiko relatif) karena $IK = \text{beda efek} \pm \text{nilai kritis (nilai kritis dihitung kasar 2 kali SE)}$. Ketiga, dengan perhitungan statistik. Gardner⁹ menganjurkan untuk menuliskan secara lengkap nilai p dan IK bersama nilai lain: rerata dan SB (bila ada), dan db (derajat kebebasan). Sekiranya antara IK dan p harus dipilih mana yang dicantumkan, maka lebih dipilih nilai p yang tidak dicantumkan.

KESIMPULAN

Pada penelitian klinis, untuk membuat kesimpulan, khususnya dikaitkan dengan kepentingan praktis, tidak cukup hanya didasarkan atas nilai P ; simpulan demikian dapat menyesatkan (*misinterpretation*). IK lebih baik dalam memberi gambaran kualitatif mengenai data yang diperoleh dari percobaan klinis. Di samping kemaknaan klinis IK juga dapat memberi gambaran kemaknaan statistik. Rentang IK dipengaruhi oleh ting-

kat kepercayaan yang kita inginkan. IK juga dipengaruhi oleh ukuran sampel seperti halnya juga nilai p . Dalam hal ini ukuran sampel untuk kemaknaan statistik tidak selalu sama dengan untuk kemaknaan klinis. Beberapa pakar berpendapat, sekiranya peneliti klinis harus memilih, mencantumkan IK lebih dipilih daripada mencantumkan nilai p .

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kepada Drs. Zulaela, Fakultas MIPA UGM, yang telah sudi membaca, memberi kritik dan koreksi atas tulisan ini.

KEPUSTAKAAN

1. Rothman KJ. A show of confidence. *New Engl J Med* 1978;299(24):1362-63.
2. Borenstein M, The case for confidence intervals in controlled clinical trials. *Contr Clin Trials* 1994; 15:411-28.
3. Guyatt GH, Sackett DL & Cook DJ. User's guide to the medical literature: II. How to use an article about therapy or prevention. B. What were the resultants and will they help me in caring for my patients? *JAMA* 1994;271: 59-64.
4. Matthews JNS, Altman DG, Campbell MJ, Royston P. Analysis of serial measurements in medical research. *BMJ* 1990;300:230-35.
5. Braitman LE. Confidence interval assess both clinical significance and statistical significance. *Annals Int Med* 1991;114:515-17.
6. Bulpitt CJ. Confidence intervals. *Lancet*, Februari 28, 1987;494-97.
7. Langman MJS. Towards estimation and confidence intervals. *BMJ* 1986;292:716.
8. Healy MJR. Significance tests. *Arch Dis Child* 1991; 66:1457-58.
9. Gardner MJ, Altman DG. Confidence interval rather than P values: Estimation rather than hypothesis testing. *BMJ* 1986;292:746-50.
10. Sastroasmoro S. Memahami makna p dan interval kepercayaan dalam studi analitik. *Sari Pediatri* 1994;1: 169-78.