

Penggunaan dan Penggunasalahan Statistik dalam Percobaan Klinik

Dr. Arini Setiawati

*Bagian Farmakologi, Fakultas kedokteran,
Universitas Indonesia, Jakarta.*

PENDAHULUAN

Statistik memegang peranan yang sangat penting dalam percobaan klinik, yakni dalam stadium perencanaan maupun dalam stadium penyelesaian. Dalam stadium perencanaan, kontribusi statistik adalah dalam menentukan jumlah sample, cara randomisasi, cara pengumpulan data agar dapat dianalisa dan dalam pemilihan tes statistik yang akan digunakan. Kontribusi statistik dalam stadium penyelesaian adalah dalam analisa data serta presentasi dan interpretasi hasilnya (1).

Penggunasalahan statistik merupakan hal yang sering terjadi. Para kritikus makalah kedokteran telah menemukannya pada kira-kira separuh dari artikel-artikel yang menggunakan metode statistik dan dimuat dalam majalah-majalah kedokteran yang terkemuka di dunia seperti British Medical Journal (BMJ), Circulation, Circulation Research, Annuals of Medicine, New England Journal of Medicine (NEJM), American Journal of Medicine, Archives of Internal Medicine dll (Lihat

Table 1). Hal ini disebabkan karena tidak banyak peneliti di bidang kedokteran yang telah memperoleh pendidikan formal dalam biostatistik, kebanyakan mungkin hanya memperoleh penataran statistik yang minimal. Disamping itu pemeriksaan statistik oleh staf editor pada kebanyakan jurnal bukan merupakan prosedur yang formal (2,6). Mungkin di antara staf editor tidak ada yang menguasai statistik, dan naskah yang dikirimkan secara rutin kepada ahli biostatistik untuk diperiksa hanyalah naskah-naskah yang penuh dengan penjelasan-penjelasan statistik karena salah seorang penulisnya ahli statistik, sehingga seringkali tidak memerlukan pemeriksaan statistik. Naskah-naskah yang hanya menyebutkan beberapa nilai P (probabilitas) dan tidak memberikan penjelasan-penjelasan statistik yang diperlukan adalah naskah-naskah yang paling memerlukan pemeriksaan statistik. Tetapi justru naskah-naskah ini tidak dikirimkan kepada ahli biostatistik sehingga luput dari pemeriksaan statistik (4).

Tabel 1. Contoh penggunasalahan statistik dalam majalah-majalah kedokteran yang terkemuka

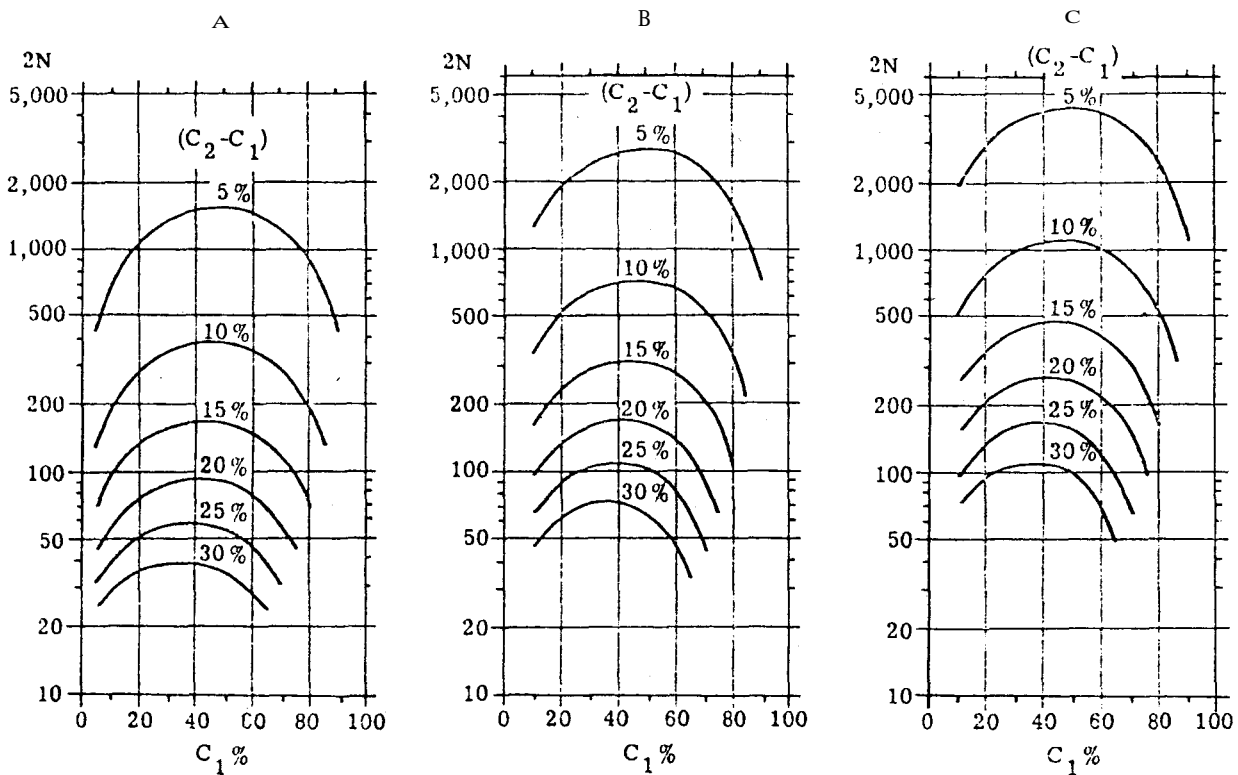
Majalah kedokteran yang disurvei		Jumlah makalah yang disurvei	Jumlah makalah tanpa statistik dan dengan penggunasalahan statistik	Keputusan
Jenis	Periode			
Circulation	Juli - Desember 1977	142 (total)	39% - tanpa statistik 27% - t - tes yang tidak benar	Glantz, 1980 (2)
Circul - Res.	Januari - Juni 1977	79 (total)	25% - tanpa statistik 46% - t - tes yang tidak benar	Glantz, 1980 (2)
BMJ	Januari - Maret 1976	77 (total)	19% - tanpa statistik 42% - statistik dengan sedikitnya satu kesalahan	Gore, 1977 (3)
10 jenis : — Ann Med — NEJM — Amer J Med — Arch Int Med — J Clin Invest — dll	Januari - Maret 1964	295 (sampel acak)	47% - metode statistik tidak benar	Schor & Karten, 1966 (4)
5 jenis : — BMJ — JAMA — NEJM — Lancet — Canad Med Assoc J	Januari - Juni 1973	1165 (total)	65% - tanpa statistik 35%-757 prosedur statistik, 20% diantaranya tidak dikenal	Feinstein 1974 (5)

Tabel 2.. Besar sampel acak yang diperlukan untuk masing-masing dari 2 kelompok pengobatan dalam suatu percobaan klinik ($\alpha = 0,05$ dan $\beta=0,05$)

Persentase kesembuhan dengan obat II **										Persentase kesembuhan dengan obat I
10	20	30	40	50	60	70	80	90	100	
53	26	17	12	9	7	—	—	—	—	0
	270	83	42	26	18	13	9	7	—	10
		402	111	53	31	20	14	9	—	20
			494	128	58	32	20	13	—	30
				539	134	58	31	18	7	40
					539	128	53	26	9	50
						494	111	42	12	60
							402	83	17	70
								270	26	80
									53	90

* Dikutip dari Rumke CL (11). Uncertainty as to the acceptance or rejection of the presence of an effect in relation to the number of observations in an experiment. Triangle 1968 ; 8 (7) : 288.

** Bila $\alpha = 0,01$ dan $\beta = 0,01$, angka-angka dalam tabel harus dikalikan 2.
 Bila $\alpha = 0,01$ dan $\beta = 0,05$, angka-angka dalam tabel harus dikalikan 1,5.
 Bila $\alpha = 0,05$ dan $\beta = 0,10$, angka-angka dalam tabel harus dikalikan 0,8.



Gambar 1. Besar sampel yang diperlukan dalam suatu percobaan klinik yang membandingkan 2 pengobatan, dimana C1 dan C2 adalah persentase kesembuhan dengan obat I dan obat II, C2 - C1 = perbedaan persentase kesembuhan dengan obat II dan obat I, N = besar sampel yang diperlukan untuk masing-masing kelompok pengobatan I dan II.

- A. $\alpha = 0,05$ dan $\beta = 0,50$
- B. $\alpha = 0,05$ dan $\beta = 0,25$
- C. $\alpha = 0,05$ dan $\beta = 0,10$

(Dikutip dari Boag JW, Haybittle JL, Fowler JF, Emery EW. The number of patients required in a clinical trial. Brit J Radiol. 1971; 44 : 123).

Akhir-akhir ini beberapa jurnal telah menyadari perlunya memperbaiki kualitas statistik dalam makalah-makalah yang dimuat dalam jurnalnya. Untuk melaksanakan hal ini, ahli biostatistik akan memeriksa aspek statistik dari semua naskah yang akan dimuat dalam jurnalnya (7 -9).

BESAR SAMPLE

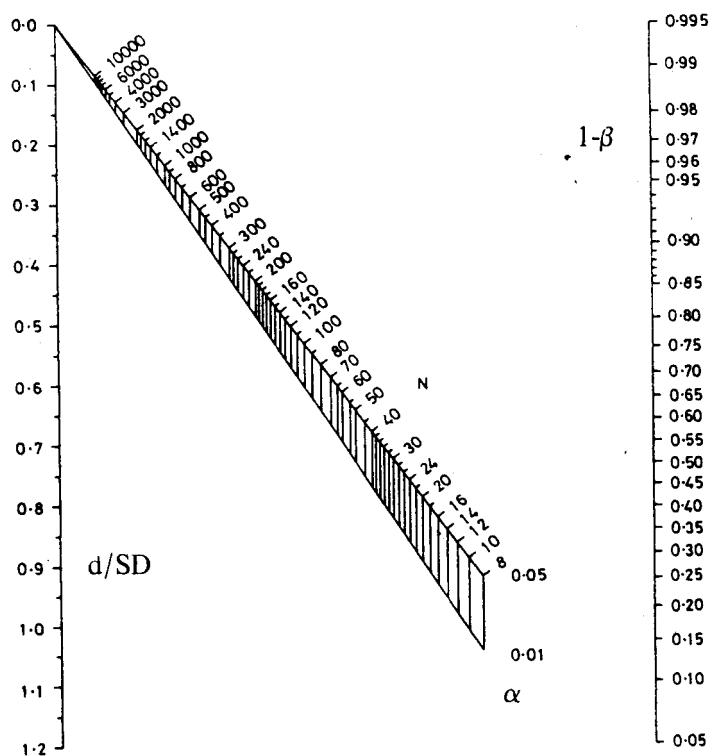
Besar sample yang diperlukan untuk suatu percobaan klinik dapat dihitung dengan rumus (10) yang dicantumkan dalam makalah dr. Armen (lihat halaman 3), atau dapat dilihat dari tabel (Tabel 2), grafik (Gambar 1) atau monogram (Gambar 2).

Arti dan telah dijelaskan dalam makalah dr Armen. Hanya kepentingan dan interpretasinya mungkin masih perlu sedikit penjelasan. Nilai $\alpha = 0,05$ berasal dari kebiasaan RA Fisher yang kemudian diikuti oleh seluruh dunia statistik sebagai nilai yang biasanya dipilih (14). Tetapi peneliti atau editor boleh saja memilih nilai yang lain, seperti 0,01 atau 0,001. Bila dari perhitungan diperoleh nilai $P > \alpha$, maka kita mengambil kesimpulan bahwa keadaan yang teramati antara hasil pengobatan I dan hasil pengobatan II secara statistik **tidak bermakna** ("not significant"). Kesimpulan tersebut benar. Kesimpulan yang salah adalah bahwa perbedaan tadi **tidak berarti** ("insignificant") sehingga kita menerima hipotesis nol. Dengan menyimpulkan bahwa perbedaan tadi tidak bermakna, berarti kita gagal untuk menolak hipotesis nol, tetapi bukan berarti bahwa kita menerima hipotesis nol! Untuk menerima hipotesis nol, masih diperlukan perhitungan, yakni kemungkinan membuat kesalahan negatif semu, yaitu kesalahan menyimpulkan tidak ada perbedaan bila sesungguhnya perbedaan itu ada. Inilah yang disebut kesalahan tipe II, dan kemungkinannya dinyatakan dengan besarnya nilai β , yakni kemungkinan kegagalan untuk menemukan perbedaan yang ada. Besarnya nilai β yang diterima untuk percobaan klinik biasanya berkisar antara 0,05 — 0,20 (13,15).

Dengan memperhitungkan untuk menentukan jumlah sampel, misalnya diambil nilai $\beta = 0,10$, maka terdapat 90% ($1 - \beta$) kemungkinan untuk menemukan perbedaan hasil pengobatan bila memang ada, pada tingkat kemaknaan $\alpha = 0,05$. Bila ternyata diperoleh nilai $P > \alpha$, kita dapat menerima hipotesis nol (bahwa tidak ada perbedaan hasil pengobatan atau bahwa perbedaannya tidak berarti) dengan kepercayaan yang cukup besar karena kemungkinan kesalahannya sudah diperhitungkan = 10% (β).

Dalam rumus untuk perhitungan besar sample, terlihat faktor Z dan Z'. Bila tidak ikut diperhitungkan, Z = 0, maka jumlah sample yang dibutuhkan menjadi jauh lebih kecil. Dalam hal ini $\beta = 0,50$, berarti bahwa kemungkinan untuk menemukan perbedaan yang sesungguhnya ada, hanya 50%. Bila nilai $\beta > 0,50$, maka Z' negatif sehingga jumlah sample yang dibutuhkan lebih kecil lagi. Jadi jelaslah bahwa makin kecil jumlah sampel makin besar risiko untuk tidak menemukan perbedaan yang ada, atau dengan perkataan lain makin kecil kemungkinan untuk dapat menemukan perbedaan yang sesungguhnya ada. Kesimpulan yang sama juga dapat ditarik dari tabel (Table 2), grafik (Gambar 1) maupun monogram (Gambar 2) yang digunakan untuk perhitungan jumlah sample.

Makin kecil nilai-nilai α dan β , makin tinggi specificitas ($1 - \beta$) dan sensitivitas ($1 - \alpha$) percobaan klinik dalam mencari



Gambar 2. Monogram untuk menentukan besar sampel yang diperlukan dalam suatu percobaan klinik yang membandingkan 2 pengobatan, dimana parameter yang diukur merupakan skala interval, d = perbedaan hasil pengobatan I dan II, SD = standar deviasi dari hasil pengobatan I (biasanya sudah diketahui dari penelitian-penelitian sebelumnya atau dari penelitian pendahulu), N = besar sampel yang diperlukan untuk kedua kelompok pengobatan (merupakan titik potong antara yang menghubungkan s/SD dan $1 - \beta$ dengan garis α). (Dikutip dari Altman DG [13]. Statistics and ethics in clinical research. III. How large a sample. Brit Med J 1980; 281 : 1337.)

ada/tidaknya perbedaan hasil pengobatan antara 2 obat yang dibandingkan, tetapi makin besar jumlah sampel yang dibutuhkan. Karena itu dalam praktek, nilai-nilai α dan β disesuaikan dengan (i) jumlah pasien yang sesungguhnya dapat diperoleh untuk percobaan klinik tersebut, dan (ii) besarnya dana yang tersedia untuk percobaan klinik tersebut (10).

Dari uraian diatas jelaslah bahwa besar sampel yang diperlukan harus diperhitungkan terlebih dahulu sewaktu merencanakan suatu percobaan klinik. Faktor-faktor yang diperlukan untuk perhitungan besar sampel tersebut, seperti nilai-nilai α dan β yang dipilih, perkiraan nilai-nilai pr, A atau d/SD, harus dilaporkan sewaktu hasilnya dipublikasi. Tetapi rupanya perhitungan besar sampel ini jarang sekali dilakukan. Di antara 172 "randomized control trials" yang dimuat dalam New England Journal of Medicine dan Lancet dari tahun 1973 s/d 1976, tidak ada satupun yang menyebutkan perkiraan besar sampel yang dibutuhkan sebelum trial dimulai, dan juga tidak ada satupun yang menetapkan besarnya perbedaan yang akan diterimanya sebagai perbedaan yang berarti dalam klinik (16). Jelaslah bahwa dalam kebanyakan trial tersebut perhitungan-perhitungan demikian tidak dilakukan.

Tidak dilakukannya perhitungan jumlah sampel yang dibutuhkan sering kali mengakibatkan sampel yang diambil terlalu

sedikit jumlahnya. Telah disebutkan bahwa sampel yang terlalu kecil meningkatkan risiko untuk tidak menemukan perbedaan-perbedaan yang sesungguhnya ada. Hal ini jelas terlihat dari hasil suatu survei mengenai 71 "randomized control trials" yang memberikan hasil "negatif" dan berasal dari 20 jenis jurnal tetapi terutama dari Lancet, NEJM dan JAMA selama periode tahun 1960 — 1977 (kebanyakan dari tahun 1970 — 1977). Bila 25% dianggap sebagai perbedaan (antara pengobatan dan kontrol) yang bermakna secara klinik dan dipilih nilai $\alpha = 0,05$, maka dari besar sampel yang digunakan, nilai n dapat dihitung. Ternyata dari 71 trial tersebut, hanya 4 yang mempunyai nilai $n \leq 0,10$. Ini berarti bahwa diantara 71 trial dengan hasil "negatif" tersebut, hanya 4 yang mempunyai jumlah sampel yang cukup sehingga hasil negatif untuk menemukan perbedaan 25% tadi cukup dapat dipercaya (kemungkinan kesalahannya $\leq 10\%$). Bila perbedaan diambil 50% dan nilai $\alpha = 0,05$, maka dari 71 trial tersebut ada 21 dengan nilai $n \leq 0,10$. Bila untuk setiap trial dihitung 90% "confidence interval" dari perbedaan antara pengobatan dan kontrol, ternyata pada 57 trial (80%) interval tersebut mencakup perbedaan 25%, dan pada 34 trial (49%) mencakup perbedaan 50%; semua ini termasuk yang mempunyai nilai $n > 0,10$. Dengan demikian dapat disimpulkan bahwa di antara 67 trial dengan nilai n (risiko untuk tidak menemukan perbedaan 25%) $> 10\%$, 57 trial (85%) mempunyai potensi untuk menemukan perbedaan 25% tersebut tetapi gagal akibat jumlah sampel terlalu kecil. Demikian juga dengan 50 trial dengan nilai n (risiko untuk tidak menemukan perbedaan 50%) $> 10\%$, 34 trial (68%) mempunyai potensi untuk menemukan perbedaan 50% tersebut tetapi gagal akibat jumlah sampelnya terlalu kecil. Sebagaimana telah disebutkan, jumlah sampel yang terlalu kecil umumnya disebabkan karena tidak dilakukan perkiraan jumlah sampel yang dibutuhkan sebelum trial dimulai. Ternyata memang hanya satu diantara 71 makalah tersebut diatas yang menyebutkan bahwa n dipertimbangkan sebelum trial dimulai, dan hanya pada 14 makalah disebutkan perlunya jumlah sampel yang lebih besar.

Dari survei tersebut diatas jelaslah bahwa untuk semua trial dengan hasil negatif, dalam laporannya perlu dicantumkan "confidence interval" dari perbedaan yang diamatinya. (17,18). Letak dan lebarnya confidence interval ini dapat memberikan gambaran tentang kemungkinan adanya perbedaan dan jumlah sampel yang dibutuhkan untuk dapat menemukan perbedaan tersebut (15). Dengan demikian trial-trial dengan hasil negatif tidak dibuang begitu saja, tetapi masih dapat dilihat kemungkinannya bahwa hasil negatif tersebut semu. Sayangnya para editor majalah-majalah kedokteran tidak mengharuskan para penulis makalah untuk melaporkan "confidence interval" dalam trial-trial dengan hasil negatif (Rose, 1980). Tampaknya hanya satu jurnal, yakni British Journal of Surgery, yang mengharuskan "confidence interval" dicantumkan dalam makalah-makalah hasil trial yang akan dimuatnya (17). Disamping itu masih ada masalah etik. Melakukan suatu trial yang kecil kemungkinannya untuk mendapatkan hasil negatif, misalnya untuk dapat menemukan suatu efek pengobatan, jelas tidak etis, apalagi bila prosedur trial menimbulkan risiko dan rasa tidak enak pada pasien (13,19).

RANDOMISASI

Proses randomisasi dalam uji klinik digunakan untuk alokasi penderita kedalam kelompok-kelompok yang sebanding (1,20). Penggunaan lain dari proses randomisasi adalah untuk memilih sampel dari populasi, dengan maksud untuk memperoleh sampel yang representatif (21). Dalam uji klinik, sampel penderita tidak pernah dipilih secara random dari populasi penderita yang memenuhi persyaratan trial dan bersedia ikut serta dalam trial tersebut, melainkan diambil semuanya sampai tercapai jumlah yang dikehendaki. Dengan demikian sampel uji klinik sebenarnya tidak mewakili populasi penderita dengan penyakit yang sama, melainkan mewakili diri mereka sendiri (20). Dasar penggunaan tes statistik adalah untuk dapat mengekstrapolasikan hasil yang diperoleh pada sampel kepada populasinya. Itulah sebabnya untuk menggunakan tes statistik disyaratkan pemilihan sampel yang dapat mewakili populasinya, yakni sampel yang dipilih secara random dari populasinya. Akibatnya, hasil suatu uji klinik hanya dapat digunakan/diekstrapolasikan pada penderita-penderita yang kondisinya persis sama dengan kondisi penderita-penderita yang diobati dalam trial tersebut (20). Dengan demikian kondisi masing-masing penderita dalam trial harus dijelaskan beserta hasil pengobatan masing-masing. Hal ini dapat dilaksanakan dengan membagi penderita dalam strata (subkelompok-subkelompok) prognostik, seperti misalnya prognosis baik, prognosis sedang ("fair") dan prognosis buruk. Bila stratifikasi ini dapat ditetapkan sebelum trial dimulai, randomisasi dilakukan dalam setiap stratum, sehingga akan diperoleh kelompok-kelompok pengobatan yang komposisi prognostiknyabenar-benar sebanding. Bila prestratifikasi tersebut tidak dapat dijalankan, randomisasi dilakukan secara keseluruhan, dan stratifikasi dilakukan belakangan. Meskipun komposisi prognostik dari kelompok-kelompok pengobatan benar-benar sebanding, hasil pengobatan harus dianalisa dalam masing-masing stratum (22). Hasil pengobatan dari masing-masing stratum inilah yang dapat diekstrapolasikan pada penderita-penderita prognosis yang sama.

Dalam kebanyakan uji klinik, sampel tidak digunakan untuk memperkirakan parameter populasi (dengan mengekstrapolasikan hasil yang diperoleh pada sampel), melainkan untuk membandingkan 2 atau lebih jenis pengobatan (21). Si peneliti jarang peduli akan parameter populasinya yang hipotetik dan biasanya hanya memperhatikan hasil pada sampelnya. Karena itu dalam pemilihan sampel harus diperoleh kelompok-kelompok pengobatan yang sebanding, sehingga bila diperoleh hasil yang berbeda, perbedaan tersebut bukan disebabkan oleh perbedaan dalam komposisi kelompok melainkan disebabkan oleh perbedaan dalam pengobatan atau oleh faktor kebetulan. Besarnya kemungkinan (probabilitas) bahwa perbedaan tersebut akibat faktor kebetulan dapat diperhitungkan dengan **tes statistik** yang sesuai. Besarnya risiko yang kita bersedia tanggung untuk membuat kekeliruan dengan menerima faktor kebetulan sebagai perbedaan yang sebenarnya, kita tetapkan sebagai nilai α (biasanya 0,05). Bila kemungkinan perbedaan tadi akibat faktor kebetulan (dari perhitungan statistik) lebih kecil dari risiko yang kita bersedia tanggung, dikatakan bahwa perbedaan tersebut **bermakna secara statistik**. Jadi kebanyakan tes statistik hanya boleh digu-

nakan untuk membandingkan hasil dari 2' atau lebih jenis pengobatan, bila kelompok-kelompok pengobatannya sebanding, terutama dalam komposisi prognostiknya. Sebagai kekecualian adalah berbagai tes permutasi, atau disebut juga tes randomisasi, yang tidak memerlukan persyaratan tersebut (22). Untuk memperoleh kelompok-kelompok pengobatan yang sebanding, maka alokasi penderita harus dilakukan secara tidak "bias", yakni dilakukan secara random (20). Hal ini berlaku terutama untuk disain perbandingan kelompok (group comparison), yakni disain yang paling sering digunakan dalam uji klinik. Untuk disain menyilang ("cross-over"), alokasi random juga diperlukan karena urutan atau waktu pemberian obat mungkin juga mempengaruhi hasilnya. Untuk disain pasangan serasi ("matched—pair"), alokasi random ini tentu saja tidak diperlukan, tetapi disain ini jarang sekali dilakukan karena kesulitan dalam pelaksanaannya.

Kesalahan yang sering kali dilakukan sehubungan dengan randomisasi ini ialah bahwa peneliti hanya menyatakan "alokasi penderita dalam kelompok-kelompok pengobatan dilakukan secara random", tanpa memberikan penjelasan tentang cara yang dipilih dan bagaimana melaksanakannya (3). Selain itu masih banyak percobaan klinik yang dilaksanakan tanpa dilakukan randomisasi, dalam hal ini perlu diberikan penjelasan tentang alasan-alasan mengapa dilakukan alokasi non-random, untuk menunjukkan pada para pembacanya bahwa alokasi yang bias tidak terjadi (3).

PENGUMPULAN DATA

Setelah randomisasi, penderita telah resmi diikutsertakan dalam trial. Sekali diikutsertakan dalam trial, penderita harus di "follow up" dengan ketat, baik penderita yang mendapat pengobatan baru maupun penderita yang mendapat plasebo/kontrol. Bila setelah randomisasi ternyata diagnosis penderita salah, penderita tersebut di "drop" dan tidak diperhitungkan dalam analisa statistik karena memang tidak memenuhi kriteria pemilihan penderita. Penderita yang tidak kembali untuk "follow up" harus dihubungi (via telpon, surat menyurat, atau bila perlu dengan kunjungan rumah) untuk mengetahui nasibnya. Hal ini juga berlaku bagi penderita yang pindah ke luar negeri yang di "drop" dan diperhitungkan statistik berlaku sampai saat penderita tersebut berangkat (23).

Penderita-penderita yang menyimpang dari protokol trial, seperti penderita-penderita yang menolak untuk meneruskan partisipasinya dalam trial, yang tidak mematuhi ketentuan-ketentuan trial, yang pengobatannya dihentikan atau dialihkan dari pengobatan yang satu ke pengobatan yang lain dalam trial tersebut atau ke pengobatan yang lain sama sekali (diluar trial), atau yang melakukan penyimpangan apapun juga, harus tetap dimasukkan dalam perhitungan statistik sampai akhir trial (tidak boleh hanya sampai saat penyimpangan terjadi) (23). Menyajikan hasil antara penderita-penderita yang mematuhi protokol saja yang dapat dilakukan, tetapi tidak dapat dilakukan perbandingan dengan menggunakan tes statistik, karena materi penderitanya tidak lagi merupakan sampel acak yang menjadi syarat digunakannya tes statistik. Disamping itu dalam mengeluarkan penderita-penderita yang menyimpang dari protokol trial, mungkin masuk faktor bias.

Jadi untuk dapat membandingkan secara statistik, harus diperhitungkan data dari semua penderita, termasuk penderita-penderita yang menyimpang dari protokol trial, yang dalam perhitungan statistik tetap dimasukkan dalam kelompok dimana mereka dialokasikan secara random pada awal trial. Trial demikian membandingkan "policy" pengobatan yang berbeda (23).

Pelanggaran terhadap prinsip tersebut di atas, seperti yang terjadi pada studi Anturane untuk mencegah "sudden death" setelah infark miokard (24), dimana penderita-penderita yang tidak memakan obatnya selama 7 hari dikeluarkan dari analisa statistik, menjadi salah satu alasan yang menyebabkan studi tersebut diragukan validitasnya sehingga ditolak oleh FDA (25). Paul Meier, ahli statistik pada University of Chicago mengatakan bahwa : "Kebanyakan trial lainnya, dan semua trial yang dilakukan oleh NHLBI (National Heart, Lung, and Blood Institute di Amerika) tetap memperhitungkan penderita-penderita yang tidak mematuhi aturan pengobatannya" (25). Salah satu contoh adalah trial yang membandingkan pengobatan antara 2 jenis beta-blocker untuk infark miokard yang "suspected" (26), dimana penderita-penderita yang dihentikan pengobatannya karena alasan efek samping, tetap diperhitungkan dalam analisa statistiknya.

Kesalahan lain yang sering dilakukan peneliti dalam proses pengumpulan data adalah membuang hasil observasi yang tampaknya jauh berbeda dari data lainnya. Dalam hal ini hasil observasi tersebut harus diperiksa kembali untuk melihat apakah ada kesalahan pencatatan. Bila ternyata tidak ada bukti salah catat, dan nilai tersebut memang mungkin maka tidak boleh dikeluarkan dari analisa statistik. Janganlah sekali-kali membuang nilai-nilai demikian hanya atas dasar bahwa nilai tersebut merupakan angka terbesar atau terkecil (27).

Kesalahan lain lagi dalam pengumpulan data adalah kealpaan mencatat data, kesalahan pencatatan dB yang biasanya tidak banyak sehingga tidak banyak mempengaruhi hasilnya. Disamping itu masih ada data yang tidak dapat dicatat karena nilainya berada dibawah sensitivitas alat ukur yang digunakan (misalnya kadar obat dalam plasma), atau karena penelitian telah dihentikan sebelum peristiwa yang harus dicatat terjadi (misalnya kematian). Dalam hal ini mungkin datanya perlu dianalisa secara non-parametrik (27).

Kesalahan-kesalahan dalam pengumpulan data sering kali akibat kegagalan pada stadium pencernaan dalam memperkirakan masalah-masalah yang akan timbul dalam pelaksanaannya. Karena itu untuk trial-trial yang besar perlu dilakukan suatu percobaan pendahuluan untuk dapat menemukan kekurangan-kekurangan yang utama. Karena kita tidak dapat meramalkan gejala-gejalanya yang mungkin relevan, maka penting sekali untuk melakukan randomisasi dan berpegang teguh padanya (27).

ANALISA DATA

Ada 3 jenis data berdasarkan skala pengukurannya, yakni *skala nominal atau klasifikasi* (misalnya mati/tidak - mati, sembuh/tidak sembuh, berhasil/gagal, jenis-jenis golongan darah dll), *skala ordinal atau "ranked"* (misalnya sakit sekali-sakit—sakit sedikit—tidak sakit, yang dapat diberi skor 3+, 2+,

1+, dan 0) dan *skala interval atau numerik* (misalnya nilai Hb, nilai tekanan darah, nilai kadar kolesterol dll). Jenis data ini menentukan prosedur statistik yang akan digunakan.

1. Nilai Sentral dan Dispersi

Dispersi atau variasi data dalam penelitian kedokteran paling sering dinyatakan sebagai *SD* ("*standard deviation*") atau *SEM* ("*standard error of the mean*") (5,28), tetapi rupanya pengertian akan SD dan SEM ini masih belum difahami sepenuhnya (2,28,29,30).

SD dan SEM hanya dapat digunakan untuk data numerik dan tidak boleh digunakan untuk data yang berupa skor (kecuali bila nilai skor merupakan fungsi linier sehingga menyerupai nilai interval). SD menunjukkan variabilitas dari data yang diperoleh, sedangkan SEM menunjukkan presisi dari nilai rata-rata sampel terhadap nilai rata-rata yang sesungguhnya (nilai rata-rata dari populasi). Pada umumnya peneliti harus memberikan gambaran tentang data yang diperolehnya. Untuk maksud ini peneliti harus melaporkan nilai rata-rata dan SD. Nilai rata-rata ± 2 SD menunjukkan "range" yang meliputi 95% dari data, dengan batas-batas yang sama jauh dari nilai rata-rata. Gambaran data demikian tentu saja hanya berlaku bila datanya mempunyai distribusi yang kurang lebih normal. Bila distribusi data menyimpang jauh dari normal (dapat diperkirakan misalnya dari nilai median yang jauh berbeda dari nilai rata-rata, atau dari nilai 2 SD yang jauh lebih besar dari nilai rata-rata sehingga batas bawah dari range data bernilai negatif suatu hal yang tidak mungkin), SD tidak dapat menggambarkan penyebaran data. Dalam hal ini untuk memberikan gambaran tentang datanya, peneliti harus melaporkan nilai median dan nilai *persentil* (10% dan 90% atau 5% dan 95% untuk sampel yang lebih besar). Nilai persentil ini sekarang banyak dianjurkan untuk menggambarkan penyebaran data karena dianggap lebih realistis, tidak tergantung pada asumsi tentang distribusi data, dan dapat digunakan juga untuk data ordinal. Penyebaran data dapat juga dinyatakan dalam "range" tetapi nilai-nilainya sangat tergantung pada nilai-nilai ekstrim atau nilai-nilai yang menyimpang sehingga kurang representatif.

Meskipun SEM dihitung dari SD ($SEM = \frac{SD}{\sqrt{n}}$, dimana n = jumlah sampel), tetapi SEM sama sekali tidak menggambarkan penyebaran data melainkan menunjukkan besarnya penyimpangan nilai rata-rata sampel dari nilai rata-rata yang sesungguhnya. Dengan demikian SEM ini diperlukan dalam perhitungan-perhitungan tes-tes statistik dan perhitungan "confidence interval". Nilai rata-rata ± 2 SEM menunjukkan "range" dimana kita percaya 95% terletak nilai-nilai rata-rata yang sesungguhnya, "range" ini disebut juga "95% confidence interval". Berbeda dengan "range" nilai rata-rata ± 2 SD, yang hanya berlaku untuk menggambarkan data bila distribusi datanya kurang lebih normal, "range" nilai rata-rata ± 2 SEM tetap berlaku untuk memperkirakan letak sesungguhnya meskipun distribusi datanya tidak normal (30,31).

Sejalan dengan paling seringnya SD dan SEM digunakan untuk menyatakan dispersi, dapat diperkirakan (meskipun tidak disebutkan) bahwa nilai sentral paling sering dinyatakan sebagai nilai rata-rata ("mean"). Analog dengan SD dan SEM, nilai rata-rata hanya boleh digunakan untuk data numerik

dan tidak boleh untuk data skor kecuali bila nilai skornya merupakan fungsi linier. Untuk data ordinal, nilai sentral dapat dinyatakan sebagai median atau "mode", sedangkan untuk data nominal hanya dapat digunakan "mode" (32).

Kesalahan-kesalahan yang sering terjadi dalam penggunaan maupun presentasi nilai sentral atau dispersi adalah :

- Nilai rata-rata digunakan untuk menghitung nilai sentral dari data berupa skor yang bukan merupakan fungsi linier.
- SD digunakan untuk menyatakan variabilitas data yang distribusinya menyimpang jauh dari normal.
- SEM dinyatakan sebagai indeks dispersi data.
- Data dituliskan misalnya : "tekanan darah distolik 150 \pm 20 mmHg" tanpa menyebutkan apa yang dimaksudkan. Nilai 150 dapat diperkirakan nilai rata-rata, tetapi nilai 20 dapat berarti SD, atau SEM, atau 2 SD, atau 2 SEM dll.

Untuk menghindari kesalahpahaman mengenai pengertian SD dan SEM, Bunce dkk (28) mengusulkan agar selalu melaporkan SD (bersama nilai rata-rata dan jumlah sampel) untuk menggambarkan variabilitas data pasien, dan bila SEM dilaporkan agar dinyatakan bahwa SEM ini tidak menggambarkan variabilitas data pasien.

2. Tes Statistik

Pemilihan tes statistik dalam uji klinik didasarkan atas:

- jenis skala pengukuran : nominal, ordinal atau numerik.
- jumlah kelompok/jenis pengobatan : 2 atau lebih.
- kelompok berkaitan ("related") atau tidak ("independent").
 - dan (c) tergantung pada disain uji klinik :
 - perbandingan kelompok ("group comparison")
 - disain menyilang ("cross-over design")
 - pasangan serasi ("matched pairs")
 - "randomized block design"
 - "Latin square design"
- besar sampel

Jenis-jenis tes statistik yang digunakan dalam uji klinik serta syarat-syarat penggunaannya dapat dilihat pada Table 3.

Kesalahan yang terjadi umumnya adalah kesalahan dalam pemilihan tes statistik yang tepat karena kebanyakan peneliti tidak mengetahui/memperhatikan syarat-syarat penggunaannya, serta tidak menyadari akibat-akibatnya. Kebanyakan peneliti (termasuk para editor!) juga tidak menyadari bahwa tes statistik tidak selalu diperlukan. Analisa statistik dimaksudkan untuk membantu presentasi dan interpretasi data. Informasi terletak didalam datanya sendiri. Bila percobaan direncanakan dengan baik, biasanya interpretasi datanya jelas. Bila demikian halnya, analisa statistik (meskipun tes statistiknya tepat) tidak diperlukan karena tidak menambah sesuatu yang berarti pada presentasinya. Manfaat statistik yang sesungguhnya terletak bukan pada analisisnya tetapi pada perencanaannya (38).

Selain itu sering terjadi pada penulis makalah hanya menulis nilai p (probabilitas), tanpa menyebutkan tes statistik apa yang digunakannya. Feinstein (5) menunjukkan bahwa dari 389 makalah kedokteran yang menggunakan tes statistik, 128 (33%) diantaranya tidak menyebutkan tes statistik apa yang digunakannya, sehingga pembaca harus memperkirakan sendiri dari presentasi datanya.

Tabel 3. Berbagai tes statistik yang digunakan dalam uji klinik serta syarat-syarat penggunaannya, disamping syarat bahwa alokasi penderita ke dalam kelompok-kelompok pengobatan harus dilakukan secara random (31 - 37).

Jenis Skala	JUMLAH KELOMPOK/JENIS PENGOBATAN			
	2 Kelompok/Jenis Pengobatan		> 2 (m) Kelompok/Jenis Pengobatan	
	Tidak Berkaitan ("Independent")	Berkaitan ("Related")	Tidak Berkaitan ("Independent")	Berkaitan ("Related")
NOMINAL N = total jumlah sampel. E = "Expected value" n ₁ dan n ₂ = jumlah sampel masing-masing kelompok. n = jumlah sampel per kelompok p = proporsi/persentase SD = deviasi standar	Tes X² (tabel 2 x 2) - data : frekuensi dalam kategori nominal - N > 40 - gunakan koreksi Yates - bila N ≥ 100 dan semua E ≥ 10 boleh tanpa koreksi Yates. Tes X² (tabel 2 x k) - data : idem - semua E ≥ 1 dan E < 5 hanya pada <20 % kotak tanpa koreksi Yates Tes eksak Fisher (tabel 2 x 2) - data : idem - N < 20 - N 20-40 dan E < 5 Perbedaan proporsi - data : proporsi/persentase - n ₁ dan n ₂ masing-masing > 30. - dengan/tanpa koreksi Yates - p gabungan/masing-masing	Tes McNemar (Tabel 2 x 2) = tes X ² yang berkaitan ("Paired X ² test") = "Paired alternatives" - data : frekuensi dalam katagorei nominal - E ≥ 5 - gunakan koreksi Yates	Tes X² (tabel m x k) - data : frekuensi dalam kategori nominal - semua E ≥ 1 dan E < 5 hanya pada ≤ 20 % kotak - tanpa koreksi Yates	Tes Cochran - data : nilai nominal - n jumlah sampel per kelompok tidak terlalu kecil (Cochran tidak menyebutkan beberapa minimal).
ORDINAL	Tes Kolmogorov-Smirnov (tabel 2 x k) - data : frekuensi dalam kategori ordinal - N kecil maupun besar Tes Mann - Whitney - data : nilai skor - N kecil maupun besar	Tes " Sign " - data : nilai skor - N kecil maupun besar Tes Wilcoxon - data : nilai skor - N kecil maupun besar	Tes Kruskal - Wallis - data : nilai skor - N kecil maupun besar	Tes Friedman - data : nilai skor - N kecil maupun besar
INTERVAL - NUMERIK	Tes t - data : nilai numerik yang distribusinya kurang lebih normal - n ₁ dan n ₂ masing-masing ≤ 30 Varians ₁ dan varians ₂ tidak berbeda bermakna - SD gabungan Tes Z - data : idem - n ₁ dan n ₂ masing-masing > 30 - SD masing-masing	Tes t yang berkaitan ("Paired t test") - data : nilai numerik yang distribusinya kurang lebih normal - n kecil maupun besar	Analisa Varians - data : nilai numerik yang distribusinya kurang lebih normal. - varians per kelompok tidak berbeda bermakna.	Analisa Varians - data : nilai numerik yang distribusinya kurang lebih normal. - varians per kelompok tidak berbeda bermakna
Disain Uji Klinik	* Perbandingan kelompok	* Pasangan serasi * Disain menyilang	* Perbandingan kelompok	* " Randomized block design" * "Latin Square design"

Dalam makalah ini hanya akan dibahas kesalahan-kesalahan yang sering kali menyertai penggunaan tes t dan tes χ^2 , yakni 2 tes statistik yang paling banyak digunakan dalam penelitian-penelitian kedokteran (5).

2.1 *Tes t (dari Student)*. Ini adalah tes statistik yang paling populer dalam penelitian kedokteran (5). Syarat-syarat penggunaannya dapat dilihat dalam Tabel 3. Kesalahan-kesalahan yang biasanya dibuat serta akibat-akibatnya ialah :

- (a) Alokasi penderita tidak dilakukan secara random. Bila alokasi random benar-benar diperlukan (seperti pada perbandingan kelompok) tetapi tidak dilakukan, maka akan masuk faktor "bias" dan diperoleh kelompok-kelompok pengobatan yang tidak sebanding, sehingga bila diperoleh hasil yang berbeda, perbedaan tersebut mungkin disebabkan oleh kelompoknya yang berbeda dan bukan oleh pengobatannya.
- (b) Distribusi data tidak normal. Dalam praktek tidak ada data yang benar-benar normal, dan meskipun beberapa jauh data menyimpang dari distribusi normal dapat diuji, misalnya dengan menggunakan kertas probabilitas normal (31), atau lebih sederhana dengan melihat perbedaan antara nilai rata-rata dan median, atau dengan melihat "scatter diagram" dari datanya (3) tetapi dalam praktek peneliti lebih sering hanya menggunakan pertimbangannya saja (39). Bila distribusi data menyimpang jauh dari normal, harus dilakukan transformasi ke bentuk logaritmanya, yang sering kali menghasilkan distribusi yang mendekati normal (3).
- (c) Tes digunakan untuk nilai skor yang bukan merupakan fungsi linier. Untuk ini tes t tidak boleh digunakan sama sekali.
- (d) Varians kedua kelompok berbeda bermakna (diuji dengan tes F). Bila diperoleh hasil yang berbeda, perbedaan tersebut mungkin akibat variansnya yang berbeda dan bukan karena nilai rata-ratanya (yakni pengobatannya) yang berbeda. Dalam hal ini tidak boleh digunakan SD gabungan, tetapi harus SD masing-masing, dan df-nya harus dikurangi (3,31).
- (e) Tes digunakan untuk data yang berpasangan (berkaitan). Dalam hal ini harus digunakan tes t yang berkaitan ("paired t test") karena tes t yang biasa (tes t untuk 2 sampel) kurang sensitif untuk menemukan perbedaan pada data yang berpasangan.
- (f) Tes t digunakan untuk membandingkan lebih dari 2 kelompok, misalnya beberapa kelompok pengobatan

yang berbeda atau beberapa kelompok respons pada waktu yang berbeda. Dalam hal ini harus digunakan analisa varians, karena bila tes t digunakan untuk membandingkan kelompok-kelompok tersebut sepasang-sepasang, maka nilai p (probabilitas) yang besarnya kira-kira = penjumlahan nilai p dari tiap pasangan kelompok. Misalnya bila tes t digunakan untuk membandingkan 3 kelompok pengobatan (misalnya A, B, dan C), maka ada 3 tes t yang dapat dilakukan (antara A dengan B, B dengan C, dan A dengan C). Bila dalam hal ini dilaporkan bahwa ada perbedaan dengan nilai $p < 0,05$, maka nilai p yang sebenarnya kira-kira = $3 \times 0,05 = 0,15$ (tepatnya 0,13) (2).

Dalam trial suatu pengobatan baru, etik merupakan salah satu pertimbangan utama; dimana harus diusahakan agar penderita yang mendapat pengobatan yang inferior seminimal mungkin jumlahnya maupun lamanya pengobatan. Hal ini seringkali menyebabkan peneliti tergoda untuk menganalisa datanya berulang-ulang sebelum jumlah sampel yang telah diperhitungkan tercapai, dan menghentikan trialnya pada saat hasilnya memberikan perbedaan dengan $p < 0,05$. Tindakan ini tidak benar karena dari trial dengan jumlah sampel yang telah ditentukan lebih dahulu, hasil trial direncanakan hanya untuk dianalisa satu kali, yakni setelah jumlah sampel tercapai. Bila dianalisa lebih dari satu kali, maka probabilitas untuk mencapai nilai (batas kemaknaan) tidak lagi = 0,05, tapi makin meningkat dengan makin seringnya tes statistik dilakukan (Lihat Table 4) (40).

Dari Table 4 dapat dilihat bahwa nilai yang 5% bila tes statistik dilakukan satu kali setelah jumlah sampel tercapai, akan meningkat menjadi 19,3% bila tes tersebut diulang untuk ke-10 kalinya.

Sebaliknya, peneliti yang ingin menganalisa datanya lebih dari satu kali sehingga dapat menghentikan trialnya lebih awal, dapat menggunakan Table 5 dimana tercantum analogi (batas kemaknaan) untuk lebih dari satu kali analisa dibandingkan dengan nilai α untuk satu kali analisa (40). Jadi, Table 5 menunjukkan bahwa peneliti yang sebelum trial dimulai telah memutuskan untuk menganalisa datanya secara statistik sebanyak 5 kali, harus mencapai nilai $p \leq 0,0159$ pada salah satu dari ke-5 tes yang dilakukan agar hasilnya benar-benar bermakna pada $p < 0,05$. Yang penting di sini adalah berapa kali data akan dianalisa dan berapa nilai α nya telah ditetapkan sebelum trial dimulai. Bila kedua hal tersebut belum diputuskan, peneliti kadang-kadang tergoda untuk setiap saat melakukan analisa statistik, sedangkan makin sering dia melakukan analisa statistik makin kecil nilai p yang harus diperoleh untuk mencapai kemaknaan.

Tabel 4 .Probabilitas untuk mencapai nilai (batas kemaknaan) setelah suatu tes statistik diulang n kali, bila tidak ada perbedaan efek antara kedua pengobatan (dinyatakan dalam %) * (40%).

Nilai (Batas Kemaknaan) (%)	Jumlah Pengulangan Tes Statistik (n)									
	1	2	3	4	5	10	25	50	200	
1	1	1,8	2,4	2,9	3,3	4,7	7,0	8,8	12,6	
5	5	8,3	10,7	12,6	14,2	19,3	26,6	32,0	42,4	
10	10	16,0	20,2	23,4	26,0	34,2	44,9	52,4	65,2	

* Dikutip dari McPherson K. Statistics : the problem of examining accumulating data more than once. N Engl J Med. 1974; 290:502.

Tabel 5. Nilai (batas kemaknaan) yang diperlukan untuk mencapai nilai yang sebenarnya, bila tes statistik diulang n kali (dinyatakan dalam %) * (40).

Nilai Sebenarnya (%)	Jumlah Pengulangan Tes Statistik (n)													
	1	2	3	4	5	6	7	8	9	10	15	20	100	
1	1	0,56	0,41	0,33	0,28	0,25	0,23	0,21	0,20	0,19	0,15	0,13	0,06	
5	5	2,96	2,21	1,83	1,59	1,42	1,30	1,20	1,13	1,07	0,86	0,75	0,32	
10	10	6,01	4,62	3,85	3,37	3,04	2,80	2,60	2,45	2,32	1,88	1,66	0,72	

* Dikutip dari McPherson K. Statistics : the problem of examining accumulating data more than once. N Engl J Med. 1974 ; 290 : 502.

2.2. *Tes χ^2* . Tes ini menempati urutan kedua dalam popularitasnya di kalangan para peneliti kedokteran (5). Syarat-syarat penggunaannya tercantum dalam Tabel 3. Kesalahan-kesalahan yang sering kali dilakukan pada penggunaan twe ini adalah :

- (a) Alokasi penderita ke dalam kelompok-kelompok pengobatan tidak dilakukan secara random. Bila diperoleh hasil pengobatan yang berbeda, mungkin akibat perbedaan dalam kelompoknya dan bukan akibat pengobatannya (Lihat keterangan pada 2.1a).
- (b) Tes digunakan pada penelitian dengan jumlah sampel (N) dan/atau "Ekspected value" (E) yang terlalu kecil. Dalam hal ini, untuk tabel 2 x 2 harus digunakan tes eksak Fisher (Lihat Tabel 3 untuk batas-batas N dan E). Untuk tabel 2 x k, nilai E dapat diperbesar dengan menggabungkan kategori-kategori yang berdampingan, asalkan penggabungan tersebut tidak menyebabkan data menjadi tidak berguna. Hal ini dapat dihindarkan dengan perencanaan jumlah sampel yang cukup untuk masing-masing kategori (35). Dalam perhitungan χ^2 , nilai E menjadi penyebut, sehingga nilai E yang terlalu kecil akan memperbesar nilai χ^2 dan dengan demikian akan memperbesar kemungkinan membuat kesalahan positif semu (menyatakan ada perbedaan, bila sebenarnya tidak ada).
- (c) Koreksi Yates tidak digunakan bila seharusnya digunakan. Koreksi Yates **harus** digunakan pada tes χ^2 tabel 2 x 2 yang tidak berkaitan maupun yang berkaitan, kecuali bila $N \geq 100$ dan semua $E \geq 10$ **boleh** tanpa koreksi Yates (lihat Tabel 3). Koreksi Yates memperkecil nilai-nilai χ^2 sehingga memperkecil kemungkinan membuat kesalahan positif semu pada sampel-sampel yang relatif kecil.
- (d) Tes χ^2 yang tidak berkaitan digunakan untuk data yang berpasangan. Dalam hal ini harus digunakan tes χ^2 yang berkaitan karena tes χ^2 yang tidak berkaitan kurang sensitif untuk menemukan perbedaan pada kata yang berpasangan.
- Tes χ^2 digunakan untuk data ordinal (frekuensi dalam kategori ordinal). Untuk ini lebih baik digunakan tes Kolmogorov - Smirnov.

3. Korelasi dan regresi

Meskipun korelasi dan regresi cukup sering digunakan untuk menyatakan asosiasi dalam penelitian kedokteran (5), tetapi rupanya perbedaan keduanya masih belum jelas bagi banyak peneliti (41).

Koefisien korelasi (r), menunjukkan derajat hubungan linier antara 2 variabel numerik. Syarat penggunaannya adalah bahwa variabel tersebut mempunyai distribusi kurang lebih normal, sehingga "scatter diagram"-nya kira-kira berbentuk elips. Bila tidak demikian halnya, tes kemaknaan dari *r* tidak valid.

Regresi menunjukkan ketergantungan dari satu variabel pada satu/lebih variabel lain. Pada regresi linier kita hitung persamaan garis lurus yang menghubungkan variabel yang "dependent" (*y*) terhadap variabel yang "independent" (*x*). Syarat penggunaannya adalah bahwa variabel *y* mempunyai distribusi yang kurang lebih normal dengan varians yang sama untuk tiap harga *x*. Penyimpangan dari kondisi ini biasanya dapat dilihat dari "scatter plot" nya. Persamaan garis regresi dapat digunakan untuk memperkirakan nilai variabel *y* dari nilai variabel *x*.

Kesalahan-kesalahan pada penggunaan korelasi adalah (41) :

- (a) Koefisien korelasi *r* dihitung untuk variabel yang distribusinya menyimpang jauh dari normal. Dalam hal ini harus dilakukan transformasi data, misalnya salah satu variabel dijadikan bentuk logaritmanya, yang biasanya akan menyebabkan distribusi menjadi kurang lebih normal. Bila tidak, dihitung *r_s* (koefisien korelasi "rank" dari Spearman) yang tidak memerlukan distribusi yang normal.
- (b) Tanpa "melihat scatter diagram" datanya, koefisien korelasi tidak berarti apa-apa karena *r* dapat rendah semu bila hubungan antara kedua variabel berupa garis lengkung, atau *r* dapat tinggi semu bila ada beberapa observasi yang sangat menyimpang dari kebanyakan data lainnya.
- (c) Korelasi digunakan untuk menghubungkan berbagai kelompok subyek yang mempunyai sifat-sifat yang berbeda, maka akan diperoleh *r* yang tinggi semu karena misalnya kelompok 1 mempunyai nilai-nilai yang rata-rata tinggi untuk kedua variabel, sedangkan kelompok 2 mempunyai nilai-nilai yang rata-rata rendah untuk kedua variabel.
- (d) Korelasi sering kali digunakan secara berlebihan, mungkin karena mudahnya menghitung koefisien korelasi. Seharusnya korelasi digunakan terutama untuk merintis penelitian selanjutnya.
- (e) Korelasi digunakan untuk membandingkan 2 metode pengukuran atau 2 jenis alat ukur. Dalam hal ini seharusnya digunakan tes *t* yang berkaitan.

Penggunasalahan persamaan regresi adalah (41) :

- (a) Persamaan regresi digunakan untuk memperkirakan nilai *y* dari nilai-nilai *x* diluar batas-batas dari aslinya (melakukan ekstrapolasi).

- (b) Menghitung persamaan regresi linier untuk data yang berupa garis lengkung.
 - (c) Menggunakan regresi sederhana untuk berbagai sub-kelompok yang heterogen.
- Dalam hal ini seharusnya digunakan analisa kovarians.
- (d) Persamaan regresi y terhadap x digunakan untuk memperkirakan nilai x dari nilai y (kecuali dalam keadaan-keadaan tertentu).

Kebanyakan penggunasalahan tersebut diatas mungkin tidak terlihat dalam makalah bila "scatter plot" datanya tidak diberikan.

Data yang dianalisa pada prinsipnya adalah data yang dapat menjawab pertanyaan-pertanyaan yang diajukan sebelum trial dimulai. Trial didisain untuk menjawab pertanyaan-pertanyaan tersebut, maka analisa data harus dilakukan sesuai dengan hipotesis dan disain trial aslinya. Data tambahan yang ditemukan secara kebetulan tidak menjawab pertanyaan tetapi justru menimbulkan pertanyaan baru, sehingga menjadi petunjuk untuk penelitian lebih lanjut (38,41).

PRESENTASI HASIL

1. Grafik

Grafik dimaksudkan untuk memberikan efek visual, karena itu skalanya tidak boleh ditransformasikan karena akan memberi efek visual yang berbeda sehingga menimbulkan interpretasi yang berlainan. Kecuali bila hendak dilakukan analisa terhadap data yang telah ditransformasi misalnya menjadi bentuk logaritmanya, maka data lebih baik digambarkan dalam bentuk "scatter diagram" dengan skala logaritma untuk menunjukkan bahwa data yang telah ditransformasi tersebut memenuhi syarat distribusi normal (30).

2. "Scatter Diagram"

Untuk data yang sederhana, gambaran data dalam bentuk "scatter diagram" sangat membantu pembaca dalam mengevaluasi analisa yang dilakukan. Titik-titik yang berimpit harus diperhatikan. Bila terdapat berbagai sub-kelompok, maka harus ditunjukkan dengan simbol yang berbeda-beda, agar dapat dievaluasi tepat tidaknya melakukan analisa data secara keseluruhan atau untuk masing-masing sub-kelompok secara terpisah (30).

Menggambarkan "scatter diagram" bukan berarti harus menghitung koefisien korelasi dan menggambarkan garis regresi, karena belum tentu salah satu atau keduanya relevan atau dapat dilakukan (memenuhi syarat-syarat penggunaannya) (30).

3. Garis Regresi

Persamaan garis regresi barulah ada gunanya dihitung dan regresinya digambarkan bila hubungan antara kedua variabelnya cukup kuat, yakni bila koefisien korelasinya bermakna (30,32).

Bila persamaan garis regresi ($y = a + bx$) diberikan, maka harus diberikan juga "standard error" dari "slope" (SEb), jumlah observasi, dan SD "residual" ("residual mean square"). SEb memberikan gambaran tentang kemaknaan "slope". SD "residual" menunjukkan variabilitas dari perbedaan antara nilai observasi dan nilai yang diperkirakan dari garis regresi, dengan demikian menunjukkan berapa dekat garis regresi dari datanya.

Garis regresi tidak boleh digambarkan melewati "range" datanya, karena diluar "range" datanya belum tentu garis regresi tersebut berlaku (30).

4. Garis SEM ("error bar")

Garis SEM merupakan cara yang populer untuk menggambarkan nilai rata-rata \pm SEM secara visual. Garis SEM ini, seperti halnya dengan SEM sendiri, tidak dapat digunakan untuk menunjukkan kemaknaan perbedaan nilai rata-ratanya karena :

- (a) Garis SEM, seperti halnya dengan SEM dari masing-masing nilai rata-rata, hanya berguna untuk menunjukkan presisi nilai rata-rata, dari kelompok-kelompok yang tidak berkaitan, tetapi tidak berguna untuk kelompok-kelompok yang berpasangan, yang kemaknaan perbedaannya tidak dihitung dari masing-masing SEM tersebut.
- (b) Garis SEM yang tidak "overlap" belum tentu menunjukkan bahwa perbedaan nilai rata-ratanya bermakna. Hanya garis SEM yang "overlap" yang pasti menunjukkan bahwa perbedaan nilai rata-ratanya tidak bermakna. Karena itu kemaknaan perbedaan nilai rata-rata harus dilaporkan tersendiri, atau bila hendak digambarkan secara visual, garis SEM tersebut diganti dengan garis "confidence interval" (30).

5. Angka-angka

Angka-angka untuk nilai rata-rata, SD, SEM dll biasanya cukup diberikan dalam 3 digit, terutama untuk tabel, agar terlihat lebih jelas.

Pada prinsipnya, dalam presentasi hasil harus jelas disebutkan :

- apa satuannya (mmHg, SEM, mg, ug/ml dll)
- apa yang dimaksud (SD, SEM, 2 SD, 2 SEM dll)
- tes statistik yang digunakan
- informasi minimal yang diperlukan agar pembaca dapat melakukan perhitungan statistik yang digunakan, misalnya nilai rata-rata, SD, jumlah observasi dll (38).

INTERPRETASI DATA

Berbagai jenis kesalahan yang telah disebutkan sebelumnya dapat menimbulkan hasil yang salah dan akibatnya interpretasi/konklusi yang salah. Disini akan dibicarakan kesalahan-kesalahan yang khusus terdapat pada interpretasi datanya sendiri.

1. Bermakna dan tidak bermakna

Arti interpretasi ini telah dijelaskan dalam pembicaraan "Besarnya Sampel", dan juga telah disinggung dalam "Randomisasi". Dari keterangan-keterangan tersebut jelaslah bahwa interpretasi tersebut berdasarkan probabilitas atau kemungkinan, bukannya kepastian. Makin kecil nilai p, makin kecil kemungkinan bahwa hipotesis nol (H_0), yang menyatakan tidak ada perbedaan antara kedua pengobatan, adalah benar. Tetapi berapapun kecilnya nilai p, kemungkinan bahwa H_0 benar selalu ada. Sebab itu tidak semua hasil yang dinyatakan bermakna adalah benar. Karena batas kemaknaan 0,05 berarti 1 diantara 20 kali trial yang sama akan memberikan hasil yang bermakna meskipun sebenarnya tidak ada perbedaan antara kedua pengobatan yang dibandingkan. Demikian juga tidak semua hasil yang dinyatakan tidak bermakna berarti hasilnya tidak ada atau tidak penting. Karena hasil yang tidak bermakna hanya

berarti bahwa hasilnya tidak cukup kuat untuk menolak hipotesis nol. Hal ini dapat disebabkan oleh jumlah sampel yang terlalu kecil (lihat pembahasan pada "Besar Sampel"). Dalam hal-hal demikian, hasil yang tidak bermakna tersebut sebaiknya dianggap sebagai tidak konklusif, dan diperlukan pengumpulan data lebih lanjut.

Karena anggapan yang salah bahwa hasil yang tidak bermakna berarti hasilnya tidak ada atau tidak penting, maka banyak peneliti tidak suka melaporkan trial dengan hasil yang tidak bermakna, dan cenderung untuk hanya melaporkan hasil-hasil yang bermakna. Hal ini dapat menyebabkan suatu obat yang sebenarnya tidak berguna dianggap berguna, berdasarkan adanya beberapa publikasi yang menunjukkan hasil yang bermakna, serta karena tidak adanya/jarangnya publikasi yang menunjukkan hasil yang tidak bermakna, meskipun sebenarnya banyak trial mengenai obat tersebut telah dilakukan dengan hasil negatif tetapi trial-trial tersebut tidak/ sedikit sekali yang dipublikasi. Telah disebutkan bahwa meskipun suatu obat tidak berguna, ada kemungkinan 1 kali dalam 20 kali trial akan memberikan hasil yang bermakna. Karena itu sangat dianjurkan untuk mempublikasi trial-trial dengan hasil negatif ($p > 0,05$) dan untuk mengulang trial-trial yang sama serta melaporkan hasilnya (38).

2. Korelasi dan Hubungan Kausal

Adanya korelasi tidak otomatis menunjukkan adanya hubungan kausal. Banyak contoh menunjukkan adanya korelasi antara 2 variabel tanpa adanya hubungan kausal, misalnya bila kedua variabel tersebut mempunyai penyebab yang sama, sebagai contoh adalah berkurangnya nafsu makan dan sukarnya tidur pada perokok. Seringkali dilupakan bahwa tidak ada tes statistik yang dapat menguji adanya hubungan kausal; tes statistik yang ada hanya dapat menguji adanya korelasi. Tapi kenyataannya, Schoolman (38) menemukan pada 72 (36%) diantara 202 makalah penelitian kedokteran yang dimuat dalam *Journal of Laboratory and Clinical Medicine* selama 1 tahun, penggunaan beberapa tes statistik yang menyimpulkan kemungkinan adanya hubungan kausal.

3. Prediksi

Dalam pembahasan tentang "Randomisasi" telah disebutkan bahwa sampel dalam uji klinik mewakili diri mereka sendiri dan bahwa hasil pengobatan dalam trial hanya dapat digunakan untuk melakukan prediksi pada penderita-penderita dengan kondisi yang persis sama dengan penderita-penderita dalam trial. Demikian juga dengan persamaan regresi hanya dapat digunakan untuk prediksi pada penderita-penderita dengan kondisi yang persis sama dengan penderita-penderita yang digunakan untuk membuat persamaan regresi tersebut (18).

4. Kemaknaan Statistik vs Kemaknaan Klinik

Kemaknaan statistik tidak identik dengan kemaknaan klinik. Kemaknaan statistik barulah ada artinya bila disertai dengan kemaknaan klinik. Karena itulah dalam perencanaan suatu uji klinik harus ditentukan lebih dulu berapa besar perbedaan yang dianggap berarti secara klinik antara obat lama dengan obat baru yang akan diuji. Setelah batas kemaknaan klinik ditetapkan, barulah dihitung besar sampel yang diper-

lukan untuk dapat mencapai kemaknaan klinik tersebut secara statistik, yakni dengan memperhitungkan juga batas-batas kemaknaan statistik yakni nilai-nilai α dan β . Cara perhitungan besar sampel ini telah dibahas dalam bab "Besar Sampel". Dengan demikian bila kemudian diperoleh hasil yang perbedaannya bermakna secara statistik, perbedaan tersebut juga bermakna secara klinik.

5. Kesimpulan yang tidak didukung oleh Data

Gore (3) menemukan pada 5 (85) diantara 62 makalah penelitian kedokteran yang dimuat dalam *British Medical Journal* selama 3 bulan (Januari s/d Maret 1976), kesimpulan/ pernyataan yang tidak didukung oleh datanya, kesimpulan demikian tentunya tidak valid.

PENUTUP

Penggunasalahan statistik dapat menimbulkan kesimpulan dan jawaban yang salah atas permasalahan yang diteliti. Dengan demikian penelitian tersebut tidak berguna sehingga tidak etis untuk dilakukan karena (18,42) :

- (a) menggunasalahkan penderita dengan menimbulkan risiko dan rasa tidak enak tanpa ada gunanya.
- (b) menghamburkan berbagai sumber, termasuk waktu penelitian, secara sia-sia.
- (c) publikasi dari hasil/kesimpulan yang salah dapat menyesatkan praktek kedokteran dan penelitian selanjutnya.

KEPUSTAKAAN

1. Breslow N. Perspectives on the statistician's role in cooperative clinical research. *Cancer* 1978; 41 : 326—332.
2. Glantz SA. Biostatistics : How to detect , correct and prevent errors in the medical literatur. *Circulation* 1980; 61 : 1—7.
3. Gore SM, Jones 16, Rytter EC. Misuse of statical methods : critical assessment of articals in *BMJ* from January to March 1976. *Brit Med J* 1977; i : 85—87.
4. Schor S, Karten J. Statistical evaluation of medical journal manuscripts. *JAMA* 1966; 45 : 1123—1128.
5. Feinstein AR. Clinical biostatistics : XXV A survey of the statistical prosedures in general medical journals. *Clin Pharmacol Ther* 1974; 15 : 97—107.
6. Altman DG. Statistics and etics in medical research : VIII Improving the quality of statistics in medical journal. *Brit Med J* 1981; 282 : 44—47.
7. Shuster JJ, Brnion J, Moxley, J et al. Stastical review process : Recommended procedures for Biomedical research articles. *JAMA* 1976; 235 : 334—5.
8. Rosen MR, Hoffman. BF. Statisticals, biomedical Scientist, and circulation Research. *Ciro Res* 1978; 42 : 739.
9. Rennie D. Vive la diffence ($P < 0,05$). *N. Engl J Med* 1978;299 : 828-9.
10. Feinstein AR. Clinical biostatistics : XXXIV The other side of statistical significance : alpha, beta, delta and the calculation of sample size. *Clin Pharmacol Ther* 1975; 18 (4),491—505.
11. Rumke CL. Uncertainty as to the acceptance or rejection of the presence of an affect in relation to the number of observations in an experiment. *Triangle* 1968; 8 (7) : 284—9.
12. Boag JW, Haybittle JL, Fowler JF, Emery EW. The number of patients required in clinical trial. *Brit J Radiol*-1971; 44 122—125.
13. Altman DG. Statistics and ethics in medical research : III How large a sample ? *Brit Med J* 1980; 28J : 1336—1338.
14. Feinstein AR. Clinical biostatistics : XXXII Biologic dependency, 'hypothesis testing' , unilateral probabilities, and other assues in scientific direction vs statistical duplexity. *Clin Pharmacol Ther* 1975; 17 : 449—513.

15. Freiman JA, Chalmers TC, Smith H Jr, Kuebles RR. The importance of beta, the type II error and sample size in the design and interpretation of the randomized control trial. *N Engl J Med* 1978; 299: 690-694.
16. Ambros A, Chalmers TC, Smith H, Schroeder B, Freiman JA, Shareck EP. Deficiencies of randomized control trials. *Clin Res* 1978; 26 : 280 A. *Dalam* : Kepust no 14.
17. Anonim. Interpreting clinical trials. *Brit Med J* 1978; ii : 1318.
18. Altman DG. Statistics and ethics in medical research : VII Interpreting result. *Brit Med J* 1980; 281 : 1612—4.
19. Newell DJ. Type II errors and ethics. *Brit Med J* 1978; IV : 1789.
20. Feinstein AR. Clinical biostatistics : XXIII The role of randomization in sampling, testing, allocation, and credulous idolatry (Part 3). *Clin Pharmacol Ther* 1973; 14 (6) : 1035—1051.
21. Feinstein AR. Clinical biostatistics : XXII The role of randomization in sampling, testing, allocation, and credulous idolatry (Part 1). *Clin Pharmacol Ther* 1973; 14 (4) : 601-615.
22. Feinstein AR. Clinical biostatistics : XXIII The role of randomization in sampling, testing, allocation, and credulous idolatry (Part 2). *Clin Pharmacol Ther* 1973; 14 (5) : 898—915.
23. Peto R, Rike MC, Armitoge P, Breslow NE, Cox DR, Howard SV, Mantel N, McPherson K, Peto J, Smith PG. Design and analysis of randomized clinical trials requiring prolonged observation of each patient: 1. Introduction and design. *Brit J Cancer* 1976; 34 :585-612.
24. The Anturane Reinfarction Trial Research Group. Sulfinpyrasone in the prevention of sudden death after myocardial infarction. *N Engl J Med* 1980; 302: 250—256.
25. Kolata GB. FDA says no to Anturane. *Science* 1980;208:1130—1132.
26. Wilcox RG, Roland JM, Banks DC, Hampton JR, Mitchell JRA. Randomized trial comparing propranolol with a enolol in immediate treatment of suspected myocardial infarction. *Brit Med J* 1980; 280 : 885-888.
27. Altman DG. Statistics and ethics in medical research : collecting and screening data. *Brit Med J* 1980; 281 : 1399-1401.
28. Bunce H, Hokanson JA, Weiss GB. Avoiding ambiguity when reporting variability in biomedical data (editorials). *Amer J Med* 1980; 69 : 8-9.
29. Gardner JM. Understanding and presenting variation. *Lancet* 1975; i : 230-231.
30. Altman DG. Statistics and ethics in medical research : VI Presentation of results. *Brit Med J* 1980; 281 : 1542—1544.
31. Dixon WJ, Massey FJ Jr. Introduction to statistical Analysis. New York McGrawhill, 1969.
32. Meddis R. Statistical Handbook for Non-statisticians. London : McGraw-Hill, 1975.
33. Colguhoun D. Lectures on Biostatistics. London : Oxford University Press, 1971.
34. Swinscow TDV. Statistics at Square One- British Medical Association, London, 1976.
35. Siegel S. Nonparametric Statistics : For the Behavioral Sciences. Tokyo : McGraw-Hill Kogakusha, 1956.
36. Talogo RW. Statistik nonparametrik. *Dalam* : Naskah Lengkap Penataran Farmakologi Klinik, Suherman SK & Syamsudin HU (editor), Jakarta : Bagian Farmakologi FKUI, 1980; hal. 134—146.
37. Basuki B. Membandingkan proporsi dan "mean". *Dalam* : Metodologi Penelitian Bidang Kedokteran, Tjokronegoro A & Purwanto SL (editor), Jakarta : Komisi Pengembangan Riset dan Perpustakaan FKUI, 1979; hal 178—183.
38. Schoolman HM, Becktel JM, Best WR, Johnson AF. Statistics in medical research : Principles versus practices. *J Lap Clin Med* 1968; 71 (3) : 357—367.
39. Anonim. Statistical errors. *Brit Med J* 1977; i : 66.
40. McPherson K. Statistics and ethics in medical research : collecting data more than once. *N Engl J Med* 1974; 290 (9) : 501—502.
41. Altman DG. Statistics and ethics in medical research : V Analysing data. *Brit Med J* 1980; 281 : 1473—1475.
42. Altman DG. Statistics and ethics in medical research : I Misuse of statistics in unethical. *Brit Med J* 1980; 281 : 1182—1184.

MELATIH KERA SEBAGAI PEMBANTU PARA CACAT JASMANI

Sejenis kera kecil, yang sering terlihat sebagai pemungut uang dari pemain-pemain organ putar di pinggir-pinggir jalan di luar negeri, dalam waktu yang tak lama lagi akan memainkan peranan yang lebih penting sebagai pembantu beribu-ribu orang lumpuh yang terdapat di dunia ini.

Jenis kera ini, capuchin (*Cebus capucinus*), yang berasal dari Amerika Tengah dan Selatan, terbukti memiliki ketrampilan, kecerdasan dan loyalitas seperti anjing-anjing penuntun para tunanetra, bahkan melebihinya.

Kera-kera ini dapat membuka lemari es, membuka atau menutup pintu dengan kunci, bahkan dapat dilatih sampai dapat mengambil piringan hitam dari album dan memasangnya di atas alat pemutarnya.

Keuntungan dibanding dengan anjing penuntun ialah biaya yang harus dikeluarkan untuk melatih kera ini kurang lebih sama besar, tetapi kera ini dapat hidup lebih kurang 30 tahun dalam pemeliharaan (lebih kurang 3 X umur anjing).

Sesungguhnya beberapa jenis kera di Indonesia juga sudah dapat dilatih untuk maksud-maksud tertentu, seperti beruk pemetik kelapa dan sebagai penari dalam ronggeng monyet.

Apakah kera-kera Indonesia tidak dapat juga dimanfaatkan untuk tugas yang lebih mulia?

International Exchange News Winter 1980 — 1981.