



Perluakah menghitung ukuran sampel?

Is it necessary to estimate sample size?

Bhisma Murti

Department of Public Health, UNS University School of Medicine, Solo

KEYWORDS *sample size, systematic error, random error, statistical aspect.*

ABSTRACT *Sample size is an issue worth-considering but not the most essential thing to fulfil for a good research. A much more crucial cause of concern to any research is the validity of inference a research is drawing, i.e. the extent to which the research is able to control systematic error that stems from bias and confounding. Sample size refers to random error; it does not address nor correct systematic error. The larger sample size, the less random error, the more precise estimates a research can yield about difference/ association/ effect of a variable(s). Most of the assignment of values in any sample size formula is arbitrary. As such, the product of estimating sample size, regardless of the formula being used, cannot be viewed as an absolute end; the actual sample size can be larger or smaller than the estimated one. Beyond statistical aspect, several other important factors should be considered when estimating sample size, such as ethics, cost, and the amount of time available for conducting the research.*

Dalam diskusi tentang rencana penelitian atau penyajian hasil penelitian, baik di dalam kampus atau di luar kampus, kerap dipersoalkan ukuran sampel (=sample size) penelitian. Sering dijumpai pengkaji proposal penelitian, penguji dan pembimbing skripsi/tesis/disertasi menilai rendah sebuah penelitian yang tidak mencantumkan rumus ukuran sampel. Benarkah besar-kecilnya sampel merupakan hal penting untuk sebuah penelitian? Jika ya jawabnya, apa alasan rasional tentang kepentingannya? Bagaimana seharusnya pendekatan yang diambil untuk memperkirakan kebutuhan ukuran sampel suatu penelitian? Artikel ini bertujuan untuk mengupas sejumlah isu esensial tentang ukuran sampel dan menyajikan kiat untuk memperkirakan kebutuhan ukuran sampel.

"THE PRIME STUFF": KESALAHAN SISTEMATIS

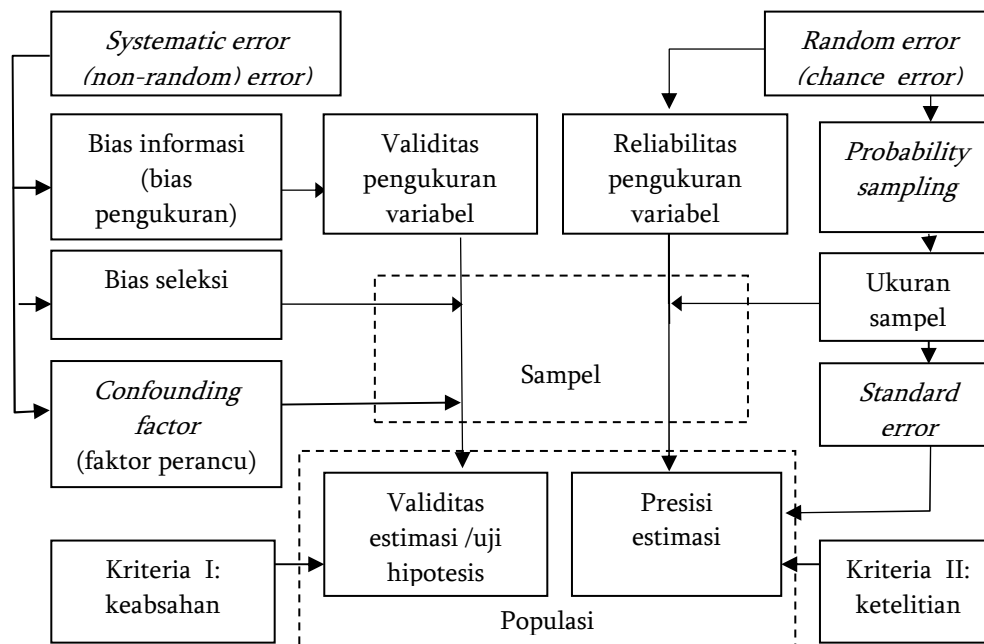
Sebelum mengupas signifikan tidaknya menghitung ukuran sampel, perlu diketahui dulu prinsip yang harus dipenuhi untuk sebuah penelitian yang baik. Penelitian yang baik adalah penelitian yang hasilnya dapat dipercaya, valid (= absah, sah, benar), dan teliti (= persis). Untuk prinsip itu, sebuah penelitian yang baik harus mampu menghindari, mengatasi, atau mengurangi seminimal mungkin dua jenis kesalahan: (1) kesalahan sistematis, dan (2) kesalahan random (Kleinbaum *et al.*, 1982; Hennekens dan Buring, 1987; Mercer, 1991; Rothman,

2002). Kesalahan sistematis, disebut juga bias (Last, 2001; Rothman, 2002), adalah deviasi hasil-hasil atau penarikan kesimpulan dari yang sesungguhnya, atau proses yang mengakibatkan deviasi itu. Kesalahan sistematis berasal dari dua sumber (Hennekens dan Buring, 1986; Mercer, 1991; Rothman, 2002): (1) bias, dan (2) kerancuan (=confounding). Gambar 1 menyajikan kesalahan sistematis dan kesalahan random, akibatnya terhadap validitas dan presisi hasil penelitian, serta peran ukuran sampel.

Berikut diuraikan sebuah ilustrasi untuk menjelaskan pengertian konsep bias. Secara metodologis mudah dipahami bahwa hasil penelitian tentang hubungan atau pengaruh variabel X (=paparan, perlakuan) terhadap variabel Y (=penyakit, variabel hasil lainnya) tidak absah jika pengukuran variabel-variabel tersebut tidak dilakukan dengan benar dan akurat. Inilah contoh bias yang disebut bias pengukuran (*measurement bias*). Pengukuran variabel dengan benar dan akurat merupakan kriteria yang tidak dapat ditawar-tawar maupun ditinggalkan (*indispensible*) untuk sebuah penelitian yang valid.

Correspondence:

Dr. Bhisma Murti, MPH, MSc, PhD, Department of Public, UNS University School of Medicine, Surakarta, Jl. Ir Sutami 36A, Surakarta 57126, Telephone/Facsimile (0271) 664178. Email: murti_bhisma@yahoo.com.



Gambar 1. Kesalahan sistematis, kesalahan random, dan akibatnya terhadap validitas dan presisi hasil penelitian.

Selanjutnya satu hal yang perlu diketahui, inti metodologis dari penelitian analitik apapun adalah banding-membandingkan rata-rata (=mean) atau proporsi dari variabel hasil (=outcome variable) yang menjadi perhatian penelitian di antara sejumlah kelompok. Tidak ada yang namanya penelitian analitik dilakukan tanpa perbandingan. Dalam studi kohor, kelompok-kelompok dibandingkan berdasarkan status paparannya, yakni kelompok terpapar dan kelompok tak terpapar. Dalam studi kasus-kontrol, kelompok-kelompok dibandingkan berdasarkan status penyakitnya, yakni kelompok kasus dan kelompok kontrol. Dalam studi eksperimental, kelompok-kelompok dibandingkan berdasarkan status perlakuannya, yakni kelompok perlakuan dan kelompok kontrol. Apabila kelompok-kelompok tersebut menunjukkan perbedaan dalam mean atau proporsi variabel hasil tadi yang cukup bermakna, antara lain dengan bantuan uji statistik, maka disimpulkan terdapat perbedaan/hubungan/pengaruh variabel. Di sini terletak poin krusial, yaitu bagaimana memilih kelompok pembanding (kelompok kontrol) yang benar. Secara metodologis mudah dipahami, bahwa pemilihan subjek-subjek penelitian dari kelompok-kelompok studi tidak boleh dipengaruhi oleh kemungkinan subjek untuk mendapatkan hasil tertentu dari variabel hasil. Sebagai contoh, sebuah penelitian berminat meneliti pengaruh merokok (paparan) terhadap terjadinya infark otot jantung (penyakit). Lalu peneliti memilih subjek-subjek untuk kelompok

terpapar (perokok) berasal dari angkatan bersenjata, sedang kelompok tak terpapar (bukan perokok) berasal dari populasi umum. Dengan contoh demikian, dapatkah peneliti mengharapkan bahwa perkiraan yang ia buat tentang besarnya pengaruh merokok terhadap terjadinya infark otot jantung adalah valid alias sah? Kemungkinan besar tidak. Hampir semua orang sangat mahfum bahwa rekrutmen tentara dilakukan secara ketat antara lain melalui tes kesehatan yang seksama. Implikasinya, dengan mudah dapat dipahami bahwa rata-rata anggota angkatan bersenjata memiliki status kesehatan yang lebih baik dari pada rata-rata anggota masyarakat umumnya. Karena memiliki tingkat kesehatan yang lebih prima, maka adalah sangat logis untuk menduga bahwa risiko untuk terkena infark akibat merokok di kalangan tentara lebih kecil daripada kelompok manapun yang status kesehatannya lebih buruk. Akibatnya, dalam contoh penelitian di atas, taksiran tentang pengaruh merokok akan lebih rendah daripada yang sesungguhnya. Bias yang terjadi akibat memilih subjek-subjek penelitian dari kalangan pekerja yang rata-rata memiliki status kesehatan yang lebih baik dari pada populasi umum disebut "healthy worker bias", merupakan salah sebuah bias seleksi (Murti, 2006a),

Di samping bias, kesalahan sistematis lainnya yang harus diperhatikan bersumber dari faktor ketiga (third variable) yang disebut faktor perancu (=confounding factor). Faktor perancu adalah faktor ketiga yang merupakan faktor risiko terhadap ter-

jadinya penyakit atau variabel hasil yang menjadi perhatian penelitian itu, berhubungan dengan paparan atau perlakuan, dan bukan merupakan variabel antara dalam mekanisme kausal paparan-penyakit. Jika peneliti gagal mengendalikan pengaruh faktor perancu, maka kesimpulan peneliti tentang perbedaan/hubungan/pengaruh variabel juga tidak absah alias tidak valid. Contoh klasik, andaikata seorang peneliti berminat meneliti pengaruh kebiasaan membawa korek api terhadap terjadinya kanker paru. Hasil analisis data menunjukkan bahwa proporsi (insidensi) kanker paru di antara kelompok pembawa korek api jauh lebih tinggi dan secara statistik bermakna ketimbang kelompok bukan pembawa korek api (sebut saja $p=0.002$). Validkah jika peneliti menarik kesimpulan bahwa membawa korek api merupakan faktor risiko terjadinya kanker paru? Tidak. Meskipun $p=0.002$ mengandung arti bahwa kesimpulan itu mengandung kesalahan sangat rendah, yaitu 2 kesalahan di antara 1000 kesempatan, tetapi argumen itu hanya dilandasi premis statistik, alias keputusan yang dibuat berdasarkan "permainan angka" semata. Secara konseptual tidak ada teori yang masuk akal yang dapat menjelaskan mengapa membawa korek api bisa meningkatkan kejadian kanker paru. Karena epidemiologi merupakan penelitian kesehatan, maka yang dimaksudkan dengan "teori" di sini tentu saja teori yang masuk akal secara biologis (*biological plausibility*). Selain itu, hampir semua pembawa korek api adalah perokok dan hampir semua bukan pembawa korek api adalah bukan perokok. Jadi secara metodologis membawa korek api berhubungan kuat dengan kebiasaan merokok, dan karena itu sangat mungkin bahwa kebiasaan merokok merupakan penjelasan alternatif terhadap hasil yang menemukan bahwa proporsi (insidensi) kanker paru di antara kelompok pembawa korek api lebih tinggi daripada kelompok bukan pembawa korek api. Jadi, kesimpulan yang menyatakan terdapat hubungan statistik yang sangat bermakna antara membawa korek api dan kejadian kanker paru ($p=0.002$) dan karena itu membawa korek api merupakan faktor risiko terjadinya kanker paru adalah kesimpulan yang "secara statistik" benar tetapi "secara metodologis" salah, karena gagal mengendalikan pengaruh kebiasaan merokok sebagai faktor perancu.

"THE SECONDARY STUFF": KESALAHAN RANDOM

Kesalahan random adalah kesalahan penarikan kesimpulan dalam suatu penelitian yang bersifat

random (acak). Berbeda dengan kesalahan sistematis yang terjadi akibat kegagalan peneliti dalam memilih subjek penelitian yang tepat, mengukur variabel-variabel dengan betul, atau mengendalikan pengaruh faktor perancu, kesalahan random terjadi karena peristiwa yang "*unpredictable*", tidak terkait dengan karakteristik peristiwa-peristiwa itu. Jika sebuah penelitian dilakukan berulang-ulang, maka ada kemungkinan hasilnya bervariasi, meskipun peneliti tidak mengubah cara memilih subjek penelitian, mengukur variabel, maupun mengendalikan kerancuan. Variasi tersebut terjadi hanya karena merupakan konsekuensi logis dari proses "*sampling*" (=mencuplik sebagian dari populasi), karena itu kesalahan random disebut juga variasi random (*sampling variability*), dan salah satu indikator variasi random adalah *standard error* (Vogt, 1993).

Intinya, dalam setiap penelitian, kesalahan sistematis membawa akibat lebih serius daripada kesalahan random (Mercer, 1991; Rothman, 2002). Kesalahan sistematis menentukan validitas (=keabsahan, kesahihan) penelitian; kesalahan random menentukan presisi (=ketelitian) penaksiran (=estimasi) yang dilakukan sebuah penelitian. "*There is no point to discuss precision when there is no validity*". Tidak ada gunanya secara prematur mendiskusikan presisi penelitian jika temuan penelitian tidak valid. Ibaratnya, tidak ada gunanya mempersoalkan mana yang lebih baik, menembak dengan persis mengenai lensa mata kanan atau meleset mengenai alis mata kanan seekor kerbau, jika sasaran yang diinginkan sesungguhnya adalah mata kiri kerbau, bukan mata kanan. Sebaliknya, "*the presence of validity does not guarantee precision*". Artinya, keberadaan validitas tidak dengan sendirinya menjamin adanya ketelitian.

PERLUKAH MENGHITUNG UKURAN SAMPEL?

Intinya, ukuran sampel merupakan salah sebuah isu yang perlu dipertimbangkan dalam sebuah penelitian kuantitatif. Ukuran sampel yang ideal tidak terlalu kecil, tidak pula terlalu besar. Sampel terlalu kecil mengakibatkan hasil penelitian memiliki presisi (=ketelitian) rendah dalam membuat estimasi kekuatan hubungan/pengaruh variabel. Sampel terlalu kecil juga mengakibatkan kemampuan penelitian rendah dalam menunjukkan hubungan/pengaruh variabel ketika hubungan/pengaruh memang ada. Sebaliknya, sampel terlalu besar berarti membuang-buang waktu, sumber daya, dan uang, yang tidak perlu. Selain itu, setiap penelitian, lebih-lebih menyangkut subjek manusia, selalu berkaitan dengan etika. Makin besar sampel, makin besar pula

sebuah penelitian bersinggungan dengan masalah etika. Sebagai contoh, pada eksperimen, terlalu banyak subjek yang mendapatkan terapi baru berarti terlalu banyak subjek yang dirugikan oleh terapi yang belum tentu menguntungkan (jika sudah pasti menguntungkan tentu tidak perlu diteliti).

Satu prinsip yang harus diketahui, ukuran sampel berkaitan dengan kesalahan random tetapi tidak berkaitan dengan kesalahan sistematis. Artinya, ukuran sampel yang tepat dapat mengurangi kesalahan random, tetapi ia tidak ada hubungannya dan tidak bisa mengoreksi kesalahan yang lebih serius dan merusak kredibilitas penelitian yang disebut kesalahan sistematis. Contoh, meskipun kesalahan random secara teoretis dapat dikurangi dengan cara meningkatkan ukuran sampel seratus kali dari 30 subjek menjadi 3000 subjek penelitian, tetapi apabila variabel-variabel penelitian diukur dengan instrumen yang salah, maka tetap saja penelitian menghasilkan kesimpulan salah alias tidak valid. Demikian juga, andaikata sebuah eksperimen dengan ukuran sampel 15 subjek menunjukkan terdapat pengaruh perlakuan yang secara statistik bermakna, maka jangan pula terburu-buru mengatakan bahwa penelitian itu hebat hanya karena dengan ukuran sampel kecil saja bisa dihasilkan kesimpulan yang secara statistik bermakna. Hasil penelitian dengan ukuran sampel kecil atau besar yang menunjukkan kemaknaan statistik tetap saja merupakan limbah ilmiah jika penelitian itu gagal mengendalikan sebagian besar dari kesalahan sistematis.

Dalam literatur dikenal sejumlah rumus standar untuk menghitung ukuran sampel (misalnya, Cochran, 1977; Rothman dan Boyce, 1982; Kleinbaum *et al.*, 1982; Kelsey, 1986; Harris, 1985; WHO, 1986; Kothari, 1990; Lemeshow *et al.*, 1990; Lwanga dan Lemeshow, 1997; Tabachnick dan Fidell, 1996; Hair *et al.*, 1998; Lohr, 1999; Thabane, 2005), yang jika diteruskan referensinya bisa sepanjang "The Long And Winding Road" The Beatles. Tetapi satu hal perlu diketahui adalah bahwa sesungguhnya semua produk rumus ukuran sampel tersebut bukan merupakan harga mati. Sebab nilai dari hampir semua faktor-faktor yang diperhitungkan dalam rumus ukuran sampel, baik yang terletak pada pembilang (numerator) maupun penyebut (denominator) ditentukan menurut kemauan peneliti. Dengan kata lain, angka yang dimasukkan ke dalam rumus ukuran sampel "suka-suka" peneliti. Istilah akademik untuk bahasa gaul "suka-suka" adalah *arbitrary* (Rothman, 1986; Greenhalgh, 1997). Implikasinya, meskipun rumus ukuran sampel yang digunakan peneliti sudah tepat untuk masalah penelitiannya, tidak ada

satupun orang yang bisa mengatakan bahwa hasil perhitungan ukuran sampel merupakan harga mati. Hanya satu prinsip yang bisa digunakan sebagai pedoman umum, bahwa makin besar sampel makin teliti (persis) hasil penaksiran yang dapat diharapkan dari sebuah penelitian, karena penelitian tersebut lebih mampu menyingkirkan kesalahan random, *ceteris paribus* (=lain-lain aspek sama). Implikasi lain argumentasi di atas adalah bahwa tidak pada tempatnya pengkaji proposal penelitian, penguji dan pembimbing skripsi/tesis/disertasi mendewakan (=mengkultus individualkan) ukuran sampel maupun rumus ukuran sampel, seperti yang kerap terjadi selama ini. Selain aspek statistik, ada beberapa aspek lain yang juga perlu diperhatikan dalam menaksir kebutuhan ukuran sampel, yaitu aspek etika, biaya, dan waktu untuk melakukan penelitian (Murti, 2006b)

MENAKSIR UKURAN SAMPEL UNTUK MENGUJI HIPOTESIS DUA MEAN POPULASI

Berikut disajikan contoh menaksir kebutuhan ukuran sampel untuk menguji hipotesis satu sisi $H_0: \mu_1 = \mu_2$ versus $H_0: \mu_1 > \mu_2$. Tujuannya adalah untuk memeragakan bahwa ukuran sampel sebagaimana yang dihitung dari rumus ukuran sampel bukan harga mati yang bersifat absolut. Contoh: Bulliyya (2002) meneliti pengaruh konsumsi ikan laut terhadap distribusi fraksi lipoprotein kolesterol serum. Dengan sampel 500 subjek konsumen ikan laut dan 500 subjek non-konsumen ikan laut, uji t menemukan bahwa rata-rata nilai LDL-C serum (faktor risiko aterogenik) secara signifikan lebih rendah dan rata-rata nilai HDL-C serum secara signifikan lebih tinggi di antara konsumen ikan ketimbang non-konsumen ikan. Bagaimana menaksir kebutuhan ukuran sampel untuk kasus seperti contoh penelitian di atas?

Contoh kasus di atas menguji beda rata-rata LDL-C dan HDL-C dua kelompok, maka rumus ukuran sampel untuk menguji hipotesis satu sisi beda dua *mean* populasi (Lemeshow *et al.*, 1990) adalah:

$$n = \frac{2\sigma^2 \left[Z_{1-\alpha} + Z_{1-\beta} \right]^2}{(\mu_1 - \mu_2)^2} \quad (1)$$

di mana $\mu_1 - \mu_2$ merupakan beda *mean* yang diperkirakan. Varians σ^2 merupakan varians populasi yang tidak diketahui nilainya, tetapi dapat diper-

kirakan dari studi awal menggunakan s_p^2 , dengan cara menggabungkan (*pooling*) varians dua sampel, s_1^2 dan s_2^2 , dengan rumus sebagai berikut (Lemeshow *et al.*, 1990; Pagano dan Gauvreau, 2000):

$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{(n_1-1) + (n_2-1)} \quad (2)$$

di mana n_1 dan n_2 adalah ukuran-ukuran sampel dalam studi awal.

Contoh: Sebuah studi ingin menguji hipotesis tentang pengaruh diet rendah garam terhadap tekanan darah sistolik (TDS). Dari studi awal diperoleh simpang baku TDS di antara komunitas dengan diet tinggi sodium 12 mmHg, di antara komunitas dengan rendah sodium 10.3 mmHg. Jika $\alpha=0.10$ dan $\beta=0.10$, berapa besar ukuran sampel diperlukan untuk dapat mendeteksi perbedaan *mean* TDS sebesar 5 mmHg antara kedua komunitas tersebut?

Jawab: Varians dari masing-masing kelompok digabung menggunakan rumus (2):

$s_p^2 = \frac{s_1^2 + s_2^2}{2} = \frac{144.0 + 106.1}{2} = 125.1$. Dengan formula (1) kemudian dihitung ukuran sampel, sebagai berikut: $n = 2(125.1)(1.64 + 1.28)^2 / 5^2 = 85.3$. Jadi diperlukan sampel 86 subjek untuk masing-masing kelompok.

Pendekatan serupa dapat digunakan untuk memperkirakan ukuran sampel untuk uji hipotesis dua sisi tentang dua *mean* dari dua populasi: $H_0: \mu_1 = \mu_2$ versus $H_a: \mu_1 \neq \mu_2$. Uji hipotesis dua sisi merupakan pilihan konservatif, dipilih ketika arah pengaruh faktor penelitian terhadap variabel hasil secara teoretis belum jelas, bisa meningkatkan, bisa menurunkan, atau tidak berpengaruh terhadap variabel hasil. Karena terdapat dua hipotesis alternatif, maka nilai ambang α untuk menolak atau tidak menolak H_0 dibagi dua, terletak di sisi kanan dan kiri distribusi normal teoretis Gauss. Rumus ukuran sampel untuk uji hipotesis dua sisi perbedaan dua *mean* dari dua populasi (Lemeshow *et al.*, 1990) adalah:

$$n = \frac{2\sigma^2 \left[Z_{1-\alpha/2}^2 + Z_{1-\beta}^2 \right]}{(\mu_1 - \mu_2)^2} \quad (3)$$

Contoh: Peneliti berminat melakukan uji hipotesis dua sisi terhadap perbedaan *mean* TDS antara sampel

dengan diet tinggi sodium dan sampel dengan diet rendah sodium. Dengan asumsi parameter lainnya sama (*ceteris paribus*), berapa besar sampel yang dibutuhkan?

Jawab: Dengan menggunakan rumus (3) dapat dihitung: $n = 2(125.1)(1.96 + 1.28)^2 / 5^2 = 105.1$. Jadi diperlukan sampel 106 subjek untuk masing-masing kelompok.

Dari rumus (1) dan (3) dapat ditarik pelajaran sebagai berikut. Pertama, jika penelitian menyangkut data kontinu seperti contoh kasus di atas, maka ukuran sampel tergantung perbedaan *mean* variabel hasil pada kelompok-kelompok populasi yang dibandingkan. Makin kecil perbedaan *mean* variabel hasil antara kelompok-kelompok yang dibandingkan terpapar, makin besar kebutuhan ukuran sampel. Demikian pula ukuran sampel tergantung varians nilai masing-masing subjek terhadap *mean*. Makin besar varians, makin besar kebutuhan sampel. Persoalannya, tidak ada angka yang pasti tentang perbedaan *mean* maupun varians tersebut (sebab kalau ada angka pasti, tentu penelitian yang sedang direncanakan tidak perlu dilakukan!). Jadi angka-angka tersebut diperkirakan oleh peneliti, berdasarkan penelitian awal (*pilot study*) atau penelitian serupa di tempat/populasi lain yang serupa (Lenth, 2001).

Kedua, dalam memperkirakan ukuran sampel untuk pengujian hipotesis, baik kesalahan tipe I (α) maupun kesalahan tipe II (β) harus diperhitungkan. Sebaliknya jika tujuan penelitian bukan menguji hipotesis, melainkan menaksir (=estimasi) besarnya perbedaan/hubungan/pengaruh, maka ukuran sampel cukup memperhitungkan kesalahan tipe I, yaitu seberapa besar peran peluang yang dapat diterima ketika menaksir besarnya perbedaan/hubungan/pengaruh. Dalam hal ini peneliti bebas untuk memilih $\alpha=0.01$, 0.05 , dan sebagainya. Meskipun bebas, harus disadari implikasinya terhadap pengambilan keputusan tentang hipotesis atau taksiran parameter. Pemilihan α berkaitan dengan kesalahan tipe I, yaitu besarnya kesalahan menyimpulkan terdapat perbedaan/hubungan/pengaruh, ketika sesungguhnya tidak ada. Makin kecil α , makin kecil kesalahan tipe I yang bisa ditoleransi, makin besar statistik uji Z yang dibutuhkan untuk menyimpulkan terdapat perbedaan/hubungan/pengaruh yang bermakna, makin besar kebutuhan ukuran sampel. Kemudian, bedakan antara ukuran sampel untuk uji hipotesis satu sisi dan dua sisi. Jika hipotesis bersifat satu sisi, maka gunakan $Z_{1-\alpha}$. Sebaliknya, jika hipotesis dua sisi, maka gunakan $Z_{1-\alpha/2}$. Demikian

juga peneliti bebas untuk memilih β , misalnya 0.10, 0.15, 0.20, dan sebagainya. Pemilihan β berkaitan dengan kesalahan tipe II, yaitu besarnya kesalahan menyimpulkan tidak terdapat perbedaan/hubungan/pengaruh, ketika sesungguhnya ada. Makin kecil β , makin kecil kesalahan tipe II yang dapat ditoleransi, makin besar statistik uji Z , makin besar kebutuhan perkiraan ukuran sampel. Kuasa statistik (= *statistical power*) penelitian adalah $1-\beta$, misalnya 0.80, 0.85, 0.90, dan sebagainya. Pemilihan $1-\beta$ berkaitan dengan seberapa besar peluang yang diharapkan untuk mendeteksi perbedaan/hubungan/pengaruh, ketika pengaruh itu memang ada. Makin besar kuasa statistik yang diinginkan, makin besar ukuran sampel yang dibutuhkan.

Ketiga, ukuran sampel sebagaimana yang dihitung dari rumus ukuran sampel bukan merupakan harga mati. Nilai dari hampir semua faktor-faktor yang diperhitungkan dalam rumus ukuran sampel, baik yang terletak pada pembilang (numerator) maupun penyebut (denominator) ditentukan menurut kemauan peneliti, bersifat *arbitrary* (Rothman, 1986; Greenhalgh, 1997). Implikasinya, meskipun rumus ukuran sampel yang digunakan peneliti sudah tepat untuk masalah penelitiannya, tidak ada satupun orang yang bisa mengatakan bahwa hasil perhitungan ukuran sampel merupakan angka mutlak yang harus dipenuhi. Hasil penaksiran ukuran sampel dengan rumus harus dipandang sebagai "ancar-ancar" dan tidak harus diikuti dengan otak fanatik. Fakta ini umumnya tidak dipahami banyak orang yang tidak mengerti hakikat statistik. Statistik bukanlah ilmu pasti maupun ilmu dewa.

STRATEGI UNTUK MENENTUKAN KEBUTUHAN UKURAN SAMPEL

Ditegaskan di muka bahwa dalam literatur dikenal aneka rumus ukuran sampel. Pertanyaannya, bagaimana memilih rumus yang tepat? Dalam banyak kasus, peneliti kerap memerlukan lebih dari sebuah pendekatan menghitung kebutuhan ukuran sampel. Berikut disajikan strategi untuk menentukan kebutuhan ukuran sampel (Murti, 2006b): (1) Sesuaikan rumus ukuran sampel yang dipilih dengan tujuan dan desain penelitian; (2) Sesuaikan rumus ukuran sampel dengan skala pengukuran variabel dependen; (3) Pilih tingkat kesalahan tipe I, tipe II, dan presisi yang diinginkan; (4) Mulailah memperkirakan ukuran sampel dengan pendekatan yang sederhana meskipun untuk problem yang kompleks; (5) Sesuaikan hasil taksiran ukuran sampel dengan keadaan-keadaan yang lebih kompleks sebagaimana

yang diantisipasi oleh peneliti; (6) Pertimbangkan aspek etika, biaya, dan waktu yang tersedia untuk melakukan sebuah penelitian.

Pertama, formula ukuran sampel harus menyesuaikan tujuan dan desain penelitian. Contoh, jika tujuan penelitian adalah menguji hipotesis perbedaan pengaruh penurunan gula darah oleh suatu obat anti-diabetik oral, maka tentu saja rumus yang relevan adalah rumus ukuran sampel yang dibutuhkan untuk menguji hipotesis perbedaan *mean* antara dua atau lebih populasi, bukan rumus untuk menaksir (= *estimate*) besarnya perbedaan itu. Sebaliknya, jika tujuan penelitian adalah menaksir kadar nikotin dalam darah para perokok, atau menaksir kadar timah hitam dalam darah kelompok masyarakat yang bekerja di sekitar ruas-ruas jalan di Jakarta yang padat kendaraan bermotor, maka rumus yang sesuai adalah rumus ukuran sampel untuk menaksir *mean* sebuah populasi.

Kedua, skala pengukuran variabel dependen menentukan pemilihan rumus ukuran sampel, yakni apakah rumus ukuran sampel untuk membandingkan perbedaan proporsi atau perbedaan *mean* (Thabane, 2005). Contoh, jika variabel dependennya adalah status merokok, yang terbagi ke dalam dua kategori: perokok kini, dan bukan perokok/mantan perokok, maka rumus yang sesuai adalah rumus ukuran sampel yang digunakan untuk menaksir proporsi, bukan *mean* dalam suatu populasi.

Ketiga, dari berbagai macam rumus ukuran sampel, intinya dapat disimpulkan pola umum sebagai berikut. Rumus untuk menaksir beda *mean* maupun beda proporsi selalu mengandung komponen d yang menunjukkan presisi taksiran yang diinginkan. Komponen d tersebut diletakkan pada bagian penyebut dari rumus. Pada saat yang sama, rumus untuk menaksir beda *mean* mengandung komponen σ^2 (=varians) yang diletakkan pada bagian pembilang dari rumus. Karena masalahnya adalah menaksir sejauh mana terdapat beda *mean* atau beda proporsi, maka kesalahan pengambilan keputusan yang relevan adalah kesalahan tipe I ($=\alpha$), yakni menyimpulkan ada perbedaan/hubungan/pengaruh variabel, padahal sesungguhnya tidak ada perbedaan itu. Jadi hanya kesalahan tipe I yang perlu diperhitungkan dan dimasukkan ke dalam rumus untuk menaksir beda *mean* atau beda proporsi itu, yakni diletakkan pada bagian pembilang. Sedang rumus untuk menguji hipotesis beda *mean* atau proporsi selalu mengandung perkiraan beda *mean* atau beda proporsi itu yang diletakkan pada bagian penyebut dari rumus. Karena masalahnya adalah menguji apakah terdapat atau tidak terdapat beda *mean*

ataupun beda proporsi, maka kesalahan pengambilan keputusan yang relevan adalah kesalahan yang mengakomodasi dua kemungkinan tersebut, yaitu kesalahan tipe I maupun II ($=\beta$). Kesalahan tipe II adalah menyimpulkan tidak ada perbedaan/hubungan/pengaruh variabel, padahal sesungguhnya terdapat perbedaan itu.

Keempat, dianjurkan untuk mulai dengan pendekatan perhitungan ukuran sampel yang sederhana sebelum dikembangkan kepada problem yang kompleks (Thabane, 2005). Sebagai contoh, jika tujuan penelitian adalah menaksir hubungan sejumlah variabel independen seperti "*social capital*", *income*, tingkat pendidikan, usia, status asuransi, dan tempat tinggal, dengan keputusan orang untuk merokok, maka salah satu pendekatan adalah menggunakan taksiran kekuatan hubungan variabel-variabel dalam populasi di dalam rumus ukuran sampel. Tetapi pendekatan seperti itu tentu menjadi tidak praktis karena terdapat banyak variabel independen, yang berarti terdapat sejumlah kemungkinan hubungan variabel-variabel yang saling bersaing untuk digunakan sebagai ancang-ancang dalam rumus ukuran sampel. Pendekatan yang lebih sederhana adalah menggunakan variabel dependen keputusan merokok yang terukur dalam skala dikotomi sebagai titik awal (*starting point*) dalam memperkirakan rumus ukuran sampel. Dalam contoh ini, informasi tentang prevalensi merokok dapat digunakan sebagai "ancang-ancang" proporsi merokok dalam rumus ukuran sampel untuk menaksir proporsi merokok pada sebuah populasi (Murti, 2005).

Kelima, setelah mendapatkan taksiran ukuran sampel melalui suatu pendekatan rumus yang sederhana, maka ukuran sampel tersebut perlu disesuaikan dengan keadaan yang lebih kompleks sebagaimana yang diantisipasi oleh peneliti. Sebagai contoh, teknik pencuplikan probabilistik yang paling mendasar sebagai langkah awal adalah "*simple random sampling*" (SRS). Apabila desain pencuplikan yang dipilih peneliti adalah "*stratified random sampling*", maka peneliti perlu memperhitungkan "*design effect*" sebagai akibat menggunakan skema yang bukan "*simple random sampling*" (*non-SRS*), agar dapat diperoleh taksiran dengan presisi yang kurang-lebih sama dengan taksiran SRS (Lohr, 1999). Demikian pula penyesuaian lebih lanjut perlu dilakukan untuk memperhitungkan kemungkinan berkurangnya ukuran sampel karena berbagai sebab: (1) "*non-response*", baik *unit-non response* maupun *item-non response* yang mengakibatkan "*missing values*", (2) kesalahan pengukuran dan *data entry* yang mengakibatkan "*extreme values*" atau "*outliers*"

sehingga sebagian data tidak dapat dianalisis; (3) berbagai sebab lainnya yang tidak terduga tetapi mengakibatkan ukuran sampel menjadi berkurang. Jadi, seperti disebutkan di muka, pada banyak kasus, peneliti perlu menggunakan sejumlah rumus ukuran sampel untuk menentukan kebutuhan ukuran sampel.

Keenam, di samping aspek statistik, penaksiran ukuran sampel perlu mempertimbangkan aspek etika, biaya dan waktu. Setiap penelitian yang menggunakan subjek manusia perlu memperhatikan implikasi etis dari penelitian. Jika suatu perlakuan diperkirakan banyak memberikan manfaat tetapi juga banyak memberikan kerugian, maka skala eksperimen tentu saja terbatas. Peneliti tidak dapat banyak merekrut subjek penelitian untuk diberi perlakuan. Demikian pula, karena anggaran dan sumber daya selalu terbatas, maka peneliti perlu berusaha mendapatkan sampel yang efisien, dalam arti mampu memberikan informasi sebanyak-banyaknya. Karena untuk mendapatkan subjek sakit atau subjek terpapar lebih sulit dan mahal ketimbang subjek kontrol, maka perbandingan yang efisien adalah jumlah kasus atau terpapar relatif lebih kecil dibandingkan dengan jumlah kontrol atau tak terpapar.

KESIMPULAN

Ukuran sampel merupakan salah sebuah isu yang perlu dipertimbangkan tetapi bukan yang terpenting dari sebuah penelitian. Persoalan yang jauh lebih krusial dan perlu dikaji dengan kritis dari penelitian apapun adalah validitas (= keabsahan) penelitian. Validitas penelitian merujuk kepada sejauh mana suatu penelitian mampu menyingkirkan atau mengatasi kesalahan sistematis, baik yang bersumber dari bias maupun kerancuan (= *confounding*). Ukuran sampel merujuk kepada kesalahan random, ia tidak ada hubungannya dan tidak dapat memperbaiki kesalahan sistematis. Makin besar ukuran sampel, makin kecil kesalahan random, makin teliti hasil penaksiran yang dilakukan penelitian tentang perbedaan/hubungan/pengaruh variabel. Hasil perhitungan ukuran sampel dengan rumus apapun bukan merupakan harga mati, artinya boleh lebih dan boleh kurang dari taksiran yang dihasilkan dari rumus. Selain aspek statistik, sejumlah aspek lainnya perlu dipertimbangkan dalam menaksir ukuran sampel, yakni etika, biaya, dan waktu yang tersedia untuk melakukan penelitian.

KEPUSTAKAAN

- Bulliyya G 2002. Influence of fish consumption on the distribution of serum cholesterol in lipoprotein fractions: comparative study among fish-consuming and non-fish-consuming populations *Asia Pacific J Clin Nutr* 11(2): 104-111
- Cochran WG 1977. *Sampling technique*. New York: John Wiley & Sons.
- Greenhalgh T 1997. How to read a paper: Statistics for the non-statistician. II: "Significant" relations and their pitfalls. *BMJ*, 315:422-425 (16 August)
- Hair JE, Jr, Anderson RE, Tatham RL, Black WC 1998. *Multivariate data analysis*. Upper Saddle River, NJ: Prentice Hall.
- Harris RJ 1985. *A primer of multivariate statistics* (2nd ed). New York: Academic Press.
- Hennekens CH, Buring JE 1987. *Epidemiology in medicine*. Boston: Little, Brown and Company.
- Kelsey JL, Thompson WD dan Evans AS 1986. *Methods on observational epidemiology*. New York: Oxford University Press.
- Kleinbaum DG, Kupper LL dan Morgenstern H 1982. *Epidemiologic research: Principles and quantitative methods*. New York: Van Nostrand Reinhold.
- Kothari CR 1990. *Research methodology: methods and techniques*. New Delhi: Wiley Eastern Limited.
- Last JM 2001. *A dictionary of epidemiology*. New York: Oxford University Press, Inc.
- Lemeshow S, Hosmer Jr DW, Klar J, Lwanga SK 1990. Adequacy of sample size in health studies. New York: John Wiley & Sons.
- Lenth RV 2001. *Some practical guidelines for effective sample-size determination*. Department of Statistics, University of Iowa.
- Lohr SL 1999. *Sampling: Design and analysis*. Pacific Grove, CA: Duxbury Press.
- Lwanga SK, Lemeshow S 1997. *Sample size determination in health studies*. Geneva: World Health Organization.
- Mercer D 1991. *Intermediate epidemiology (Coursework)*. New Orleans, LA: Tulane School of Public Health and Tropical Medicine.
- Murti B 2005. The family as health producer in Indonesia: An examination using Grossman's model and its extensions (Disertasi). Centre for Clinical Epidemiology and Biostatistics, University of Newcastle, Australia.
- _____ 2006a. *Prinsip dan metode riset epidemiologi*. Edisi ke 3. Yogyakarta: Gadjah Mada University Press.
- _____ 2006b. *Desain dan ukuran sampel untuk penelitian kuantitatif dan kualitatif di bidang kesehatan*. Yogyakarta: Gadjah Mada University Press.
- Pagano M, Gauvreau K 2000. *Principles of biostatistics*. Pacific Grove, CA: Duxbury
- Rothman KJ dan Boyce JD 1982. *Epidemiologic Analysis with a Programmable Calculator* (2nd ed.). Brookline, MA: Epidemiology Resources Inc.
- Rothman KJ 1986. *Modern epidemiology*. Boston: Little, Brown, and Company.
- _____. 2002. *Epidemiology: An introduction*. New York: Oxford University Press.
- Tabachnick BG, Fidell LS 1996. *Using multivariate statistics (3rd ed)*. New York: Harper-Collins.
- Thabane L 2005. *Sample size determination in clinical trials*. Center for Evaluation of Medicine. Hamilton, ON. [http:// www.lehanthabane.com](http://www.lehanthabane.com).
- WHO (World Health Organization) 1986. *Sample size determination: A user's manual*. Geneva: Epidemiological and Statistical Methodology Unit, World Health Organization.