

Beberapa Isu Statistik Dalam Desain, Analisis Dan Interpretasi Penelitian Epidemiologi Lingkungan

Statistical Issues in The Design, Analysis, and Interpretation of Environmental Epidemiologic Studies

Suhartono

ABSTRACT

A primary objective of many environmental epidemiologic studies is to associate potentially adverse exposures received in the community (or workplace) with potential biologic effects. Such association are strengthened considerably if exposure-response relationship can be found. Identification and characterization of exposure-response relationship in the residential community setting are made difficult by the general absence of documented data on individuals at risk and their potential exposures.

There are some limitations of epidemiological studies of communities exposed to environmental contaminants i.e. populations living in the vicinity of point source of exposure are usually small; persons living in any given area are usually heterogeneous either with respect to characteristics that can influence many health outcomes independently of exposure; actual population exposures are generally poorly defined and for many chemicals little or nothing known about toxicological effects; many of the health endpoints of interest are either rare; publicity related to the episode under study may produce or accentuate reporting bias; and the conduct of community studies is made difficult by the presence of a highly charged atmosphere of anger and fear.

There are some statistical aspects of community based investigations of health effects are i.e. exploratory vs. confirmatory study; determining sample size and statistical power; identifying and controlling systematic sources of error; assessing interaction; determining the level of the investigation; power with ecological studies; limitations of multivariate models, etc.

Key word : *Statistical Issues, Environmental Epidemiologic Studies.*

PENDAHULUAN

Semakin berkembangnya teknologi dan industri di dunia diikuti oleh semakin banyaknya limbah atau polutan yang dibuang ke lingkungan. Hal tersebut akan menimbulkan permasalahan bagi lingkungan itu sendiri, yaitu berupa pencemaran, dan bagi kesehatan masyarakat. Dalam sebuah seminar internasional tentang toksikologi lingkungan di Jakarta disebutkan bahwa setiap tahun diperkirakan terdapat seribu bahan kimia baru, dan sampai saat ini terdapat sekitar 70.000 jenis bahan kimia yang biasa digunakan, di mana 800 – 900 di antaranya telah diketahui bersifat karsinogenik. Meningkatnya jumlah kendaraan bermotor, khususnya di kota-kota besar juga menimbulkan permasalahan pencemaran udara yang tidak bisa dianggap remeh. Beberapa polutan udara telah terbukti mempunyai dampak negatif terhadap kesehatan, misalnya timah hitam (Pb) yang mengganggu sintesis hemoglobin dan menurunkan kecerdasan anak, sulfur dioksida (SO₂) yang bersifat iritan sehingga memicu timbulnya infeksi pada saluran pernafasan, karbonmonoksida (CO) yang berpengaruh

terhadap oksigenasi ke jaringan, dan sebagainya.¹⁾

Berkaitan dengan hal tersebut, maka berkembang suatu metode atau cara untuk mempelajari (membuktikan) dampak dari polutan-polutan di lingkungan terhadap kesehatan masyarakat, yaitu ilmu epidemiologi lingkungan. Metode pembuktian secara ilmiah sangat diperlukan terutama untuk menggugah kesadaran masyarakat akan bahaya atau dampak dari pencemaran lingkungan terhadap kesehatan masyarakat, dan memberikan masukan kepada pemerintah sebagai penentu kebijakan dalam penanggulangan pencemaran lingkungan.

Tujuan utama dari studi epidemiologi lingkungan adalah untuk mencari hubungan antara **pajanan** dari beberapa polutan di lingkungan, baik lingkungan kerja maupun lingkungan masyarakat (komunitas), dengan timbulnya **gangguan kesehatan** atau **efek biologis**, sehingga dapat dijelaskan adanya hubungan sebab akibat di antara keduanya. Hubungan pajanan-efek tersebut dinyatakan benar-benar kuat apabila fenomena hubungan dosis-respons bisa dibuktikan, yaitu bila peningkatan tingkat atau dosis pajanan diikuti

dr. Suhartono, M.Kes. Program Magister Kesehatan Lingkungan PPs UNDIP

Beberapa Isu Statistik

pula dengan meningkatnya frekuensi atau tingkat keparahan dari efek biologis. ^(2,3)

Diprediksi jenis dan dosis pajanan yang diterima oleh tenaga kerja. Dengan adanya data atau informasi tersebut, maka kita bisa menentukan siapa saja yang tergolong populasi yang berisiko (*population at risk*) dan memprediksi faktor-faktor yang mungkin merupakan pengganggu di dalam melihat hubungan pajanan-efek yang sebenarnya.

Mencari hubungan antara pajanan-respons lebih sulit dilakukan apabila studi dilakukan di lingkungan masyarakat. Hal ini disebabkan tidak adanya data tentang individu-individu yang berisiko dan pajanan potensial apa saja yang mungkin mereka terima. Problem lain yang mungkin timbul adalah kesulitan dalam mengukur tingkat pajanan yang sesungguhnya, karena adanya kemungkinan mendapat pajanan di tempat lain atau adanya interaksi di antara beberapa polutan, sehingga efek biologis yang timbul pun akan berbeda. Penelitian tentang kandungan Pb di dalam darah polisi lalu lintas, sopir bus, penjaja makanan, dan beberapa kelompok profesi lain yang melakukan aktivitas sehari-hari di daerah padat lalu lintas di Jakarta, menunjukkan kandungan Pb dalam darah mereka berkisar antara 0,03 – 0,09 mg/100 ml darah (nilai normal maksimum = 0,025 mg/100 ml darah). Namun, apakah jumlah kandungan Pb dalam darah sebesar itu, semata-mata hanya dari pajanan di jalan raya, masih belum bisa dipastikan.

Berikut ini adalah beberapa keterbatasan dan kendala yang sering dihadapi di dalam penelitian epidemiologi lingkungan adalah: ²⁾

1. Populasi yang tinggal di sekitar sumber pajanan (misalnya tempat pembuangan limbah toksik) biasanya tidak banyak, sehingga variasi maupun jumlah *outcome* (efek) menjadi sangat terbatas
2. Orang-orang yang tinggal di suatu lokasi pada umumnya mempunyai karakteristik yang berbeda-beda, misalnya dalam hal umur, ras, status sosial ekonomi, pekerjaan, kebiasaan merokok, dan sebagainya. Jenis, tingkat, durasi atau saat pajanan juga bisa sangat berbeda antara satu individu dengan individu lainnya. Hal lain yang juga harus dipertimbangkan adalah migrasi (masuk atau pun keluar area) dan mobilitas masyarakat dalam area atau wilayah tersebut
3. Populasi yang benar-benar terpajan pada umumnya tidak bisa ditetapkan (didefinisikan) dengan baik dan banyak bahan kimia maupun senyawanya yang belum ada informasi tentang efek toksiknya
4. Beberapa dampak kesehatan akibat pajanan terhadap bahan kimia kejadiannya sangat jarang (misal: malformasi kongenital), bersifat kronis dan membutuhkan pajanan

dalam waktu panjang (misal: kanker), atau dalam banyak hal seringkali tidak tercatat (diingat), seperti aborsi spontan. Hal tersebut menyebabkan instrumen yang akan dipakai untuk mengukur dampak kesehatan tersebut (misal: kuesioner) menjadi tidak peka (sensitif)

5. Publikasi ilmiah yang berhubungan dengan dampak kesehatan akibat polutan di lingkungan kemungkinan menghasilkan atau menonjolkan adanya bias laporan, yaitu adanya kecenderungan bahwa suatu penelitian yang dilakukan berdasarkan adanya laporan tentang suatu dampak (efek) kesehatan di masyarakat, secara metodologis maupun statistik telah diarahkan untuk membuktikan adanya dampak atau efek tersebut
6. Kemungkinan timbulnya rasa takut atau keresahan di masyarakat apabila hasil penelitian bisa menunjukkan adanya efek negatif akibat pajanan suatu polutan di lingkungan

Adanya beberapa keterbatasan di dalam penelitian epidemiologi lingkungan tersebut merupakan kendala untuk melakukan analisis hubungan antara pajanan dari bahan toksik di lingkungan dengan efek kesehatan yang ditimbulkannya, di samping merupakan penghambat dalam upaya generalisasinya. Pada makalah ini akan dibahas bagaimana keterbatasan-keterbatasan metodologis dalam penelitian epidemiologi lingkungan akan berdampak terhadap aspek statistik dari penelitian tersebut.

ISU-ISU STATISTIK DALAM PENELITIAN EPIDEMIOLOGI LINGKUNGAN

Berikut ini adalah bahasan tentang isu-isu statistik di dalam penelitian yang berkaitan dengan desain, analisis, dan interpretasi data dari penelitian epidemiologi lingkungan.

Penelitian Eksplorasi vs. Penelitian Konfirmasi

Sebelum interpretasi yang tepat atas data hasil penelitian dibuat, maka kita perlu mengetahui dan memahami tentang metode penelitian yang dilakukan. Dalam pendahuluan telah disampaikan bahwa terdapat beberapa keterbatasan di dalam penelitian epidemiologi lingkungan, misalnya sulitnya mengidentifikasi kelompok terpajan, efek atau *outcome* yang jarang, bersifat kronis dan membutuhkan pajanan dalam waktu lama, dan sebagainya. Berkaitan dengan hal tersebut, maka penelitian epidemiologi lingkungan seringkali masih bersifat eksplorasi (penjajagan), sehingga bersifat terbuka dan masih mencari-cari (konfirmasi). Hal ini disebabkan karena masih terbatasnya informasi atau pengetahuan pendukung tentang masalah yang diteliti. ⁴⁾ Sehubungan dengan hal tersebut, maka upaya generalisasi dari hasil penelitian eksplorasi sulit untuk dilakukan.

Sebaliknya, di dalam penelitian yang bersifat konfirmasi atau eksplanatori (penjelasan), harus sudah ada data atau informasi pendukung yang melatarbelakangi dilakukannya penelitian tersebut, sehingga bisa dilakukan suatu uji hipotesis untuk melihat hubungan kausalnya.²⁾ Dalam upaya generalisasi hasil penelitian, selain hasil uji hipotesis, perlu juga dipertimbangkan hasil-hasil penelitian sebelumnya, besar sampel, dan kekuatan (*power*) penelitian.

Besar Sampel dan Kekuatan (*Power*) Penelitian

Dalam merancang dan menginterpretasikan hasil penelitian, besar sampel dan *power* penelitian merupakan bahan pertimbangan yang penting, karena membantu dalam menentukan desain penelitian dan memberikan landasan obyektif untuk interpretasinya.^{2, 5,6)}

Nilai *power* yang sering digunakan dalam penelitian adalah 80% atau 90%. Sebuah

penelitian mempunyai *power* 80%, artinya adalah penelitian itu mempunyai peluang sebesar 80% untuk mendeteksi perbedaan efek (*outcome*) bila perbedaan tersebut dalam populasi memang ada.⁵⁾ *Power* penelitian merupakan fungsi dari beberapa parameter penelitian, yaitu:

1. *Jumlah sampel dan kelompok kontrol.* Pada umumnya, *power* bertambah bila jumlah sampel bertambah
2. *Rate (angka kejadian) dari efek atau pajanan yang diteliti.* Pada penelitian kohort, *power* berhubungan dengan insidens rate dari efek di antara individu yang tidak terpajan dalam populasi target. Dalam penelitian kasus kontrol, *power* berhubungan dengan frekuensi relatif pajanan di antara kelompok kontrol pada populasi target. Pada tabel 1 disajikan beberapa efek kesehatan, frekuensi, dan unit analisisnya.

Tabel 1: Beberapa efek kesehatan, frekuensi dan unit analisisnya

Hasil Kesehatan	Frekuensi	Satuan Analisis
Efek reproduksi ^a		
Azospermia	1 x 10 ⁻²	Laki-laki
Berat lahir < 2500 g	7 x 10 ⁻²	Lahir hidup
Aborsi spontan		
Setelah 8 – 28 minggu kehamilan	1–2 x 10 ⁻¹	Kehamilan
Anomali kromosom di antara konsepsi teraborsi secara spontan	3-4 x 10 ⁻¹	Aborsi spontan
Cacat lahir	2-3 x 10 ⁻²	Lahir hidup
Cacat tabung neural	1 x 10 ⁻⁴ – 1 x 10 ⁻²	Lahir masih/hidup
Insidens kanker ^b		
Semua tempat	3,2 x 10 ⁻³	Individual
Perut	9,8 x 10 ⁻⁵	Individual
Kolon	3,3 x 10 ⁻⁴	Individual
Paru dan Bronchus	4,5 x 10 ⁻⁴	Individual
Kandung kemih	1,5 x 10 ⁻⁴	Individual
Ginjal	6,4 x 10 ⁻⁵	Individual
Lymphoma	1,2 x 10 ⁻⁴	Individual
Leukemia	9,5 x 10 ⁻⁵	Individual
Mortalitas ^c		
Semua sebab	9,5 x 10 ⁻³	Individual
Semua sisi kanker	1,6 x 10 ⁻³	Individual
Cirrhosis hati	1,5 x 10 ⁻⁴	Individual
Anomali congenital	8,4 x 10 ⁻⁵	Individual

^a Dari Bloom A.D., Ed., *Guidelines for Studies of Human Populations Exposed to Montagen and Reproductive Hazards*, March of Dimes Birth Defect Foundation, White Plains, NY, 1981

^b Average annual age-adjusted (1970) incidence rates, all S.E.E.R. sites

^c 1970 U.S. mortality rates

3. *Angka kesalahan tipe I.* Dengan parameter penelitian lain telah ditetapkan, maka *power* (1- β) berhubungan dengan angka kesalahan tipe I (lihat Gambar 1)
4. *Jumlah asosiasi yang diuji.* Semakin banyak pajanan (faktor risiko, variabel bebas) atau efek (variabel terikat) yang diteliti, maka *power* penelitian akan menurun. Oleh karena

- itu, di dalam analisis statistik, sebaiknya dilakukan pengendalian terhadap variabel pengganggu dengan cara membatasi jumlah variabel pajanan dan atau variabel efeknya.
5. *Besar hubungan yang diduga antara pajanan dan efek.* Dengan parameter penelitian lain telah ditetapkan, maka *power* berhubungan dengan angka ini. Berkaitan dengan hal itu,

Beberapa Isu Statistik

maka sebelum menghitung besar sampel, peneliti harus lebih dahulu menentukan berapa angka minimal dari efek yang dianggap bermakna (secara klinis, kesehatan masyarakat, atau biologis). Angka besaran hubungan antara pajanan-efek, kesalahan tipe-I dan kesalahan tipe II, akan menentukan jumlah sampel yang diperlukan

6. *Desain penelitian dan teknik analisis statistik yang digunakan.* Terdapat beberapa desain khusus dan teknik analisis statistik yang bisa digunakan untuk meningkatkan power, antara lain: melakukan seleksi secara ketat tentang

riwayat pajanan untuk menghindari bias karena salah klasifikasi, mendefinisikan secara ketat dan benar variabel efek untuk membedakan dengan efek lain yang mirip (koheren), meningkatkan jumlah sampel dengan cara mengintensifkan penemuan kasus, membuat variabel komposit/gabungan (baik untuk variabel pajanan maupun variabel efek), memakai variabel dengan skala kontinyu untuk variabel efek, melakukan pengukuran berulang pada setiap sampel, stratifikasi, *matching*, dan teknik kluster.

		KEADAAN DALAM POPULASI	
		Berbeda	Tidak berbeda
UJI HIPOTESIS	Ho ditolak	Positif Benar ($1-\beta$; Power)	Kesalahan tipe I (α)
	Ho tidak ditolak	Kesalahan tipe II (β)	Negatif Benar

Gambar 1: Kesalahan pada uji hipotesis

Identifikasi dan Pengendalian Sumber-sumber Kesalahan Sistematis

Tujuan akhir dari suatu penelitian epidemiologi lingkungan adalah membuktikan adanya hubungan antara pajanan (faktor risiko; polutan di lingkungan) dengan terjadinya efek (gangguan kesehatan tertentu). Untuk bisa membuktikan bahwa terjadinya efek semata-mata hanya karena adanya pajanan tersebut, maka faktor-faktor penyebab atau faktor-faktor lain yang diduga berhubungan dengan terjadinya efek (variabel pengganggu; konfounding), harus bisa disingkirkan (dikendalikan). Kegagalan dalam mengendalikan variabel-variabel pengganggu akan mengakibatkan terjadinya bias di dalam estimasi hubungan pajanan-efek.

Berikut ini adalah beberapa jenis bias yang sering terjadi di dalam penelitian epidemiologi lingkungan di populasi masyarakat:^{2,5)}

1. Bias Pengganggu (*konfounding*)

Bias konfounding atau sering hanya disebut sebagai 'konfounding' saja, terjadi dari kegagalan untuk mengendalikan pengaruh dari faktor-faktor lain yang berhubungan dengan pajanan dan efek (*outcome*). Sebuah variabel dikatakan sebagai konfonder apabila: (1) variabel tersebut merupakan faktor risiko bagi terjadinya efek; (2) variabel tersebut berhubungan dengan pajanan; dan (3) variabel tersebut bukan merupakan variabel antara pada jalur hubungan kausalitas antara pajanan dengan efek. Konfounding dapat dikendalikan dalam tahap desain penelitian (misalnya dengan *matching* pada penelitian kasus-

kontrol), dan pada tahap analisis statistik (misalnya dengan stratifikasi), atau keduanya.

2. Bias Seleksi

Bias ini terjadi bila terdapat sebuah perbedaan sistematis antara karakteristik-karakteristik dari individu-individu yang diseleksi untuk sebuah penelitian dan karakteristik-karakteristik dari mereka yang tidak diseleksi untuk penelitian yang sama. Sumber terjadinya bias seleksi ini adalah: kelemahan dari desain penelitian, terutama adalah saat penentuan anggota kelompok yang akan dibandingkan (untuk semua jenis penelitian); pemilihan sampling frame yang tidak tepat (terutama pada penelitian kasus-kontrol dan *cross-sectional*); hilang dari pengamatan atau tidak memberi respons selama pengumpulan data (dalam penelitian kohort); dan *selective survival*, yaitu individu-individu yang terpilih untuk subyek penelitian adalah individu-individu yang mempunyai karakteristik tertentu, misalnya status kesehatannya atau survivalnya (dalam penelitian kasus kontrol dan *cross-sectional*).

3. Bias Informasi

Istilah lain dari bias informasi adalah 'misklasifikasi'. Bias ini berkaitan dengan instrumen dan teknik yang digunakan dalam pengumpulan data, baik data pajanan, efek, maupun data faktor-faktor penelitian yang lain. Bias informasi terjadi bila pengukuran-pengukuran secara individual, klasifikasi penyakit (efek, *outcome*) atau klasifikasi pajanan dilakukan dengan tidak akurat (tidak mengukur apa yang seharusnya diukur). Selain itu, terdapat banyak sekali sumber-sumber bias dalam pengukuran dan efek-efeknya mempunyai variasi kepentingan yang berbeda-beda. Sebagai contoh, pengukuran-

pengukuran biokimia atau fisiologis itu tidak pernah akurat secara lengkap dan seringkali laboratorium yang berbeda memberikan hasil yang berbeda-beda pula untuk spesimen yang sama. Bias ingatan (*recall bias*) juga merupakan salah satu bentuk bias informasi yang mempunyai peranan penting dalam penelitian kasus-kontrol retrospektif. Sebagai contoh: individu-individu dalam kelompok kasus mungkin mempunyai kecenderungan untuk mengingat pajanan yang pernah diterimanya (dibanding individu-individu pada kelompok kontrol), terutama bila telah banyak diketahui bahwa pajanan tersebut berhubungan dengan penyakit yang sedang diteliti.

4. Bias Pelaporan

Terdapat beberapa kesulitan yang berkaitan dengan investigasi dalam suatu populasi yang telah diketahui terpajan oleh polutan karena bertempat tinggal di sekitar sumber pencemaran lingkungan. Hal ini akan meluas atau berdampak terhadap isu metodologi maupun statistik yang akan digunakan. Contoh: bila mana penelitian dilakukan berdasarkan laporan dari masyarakat tentang kejadian yang merugikan, maka kepentingan statistik, dalam hal ini hipotesis, telah diarahkan atas dasar data laporan tersebut. Uji hipotesis yang biasa digunakan mungkin tidak sesuai untuk dipakai karena kita akan cenderung untuk menguji hipotesis berdasarkan data (efek) yang telah dilaporkan. Sebaiknya, observasi dilakukan di tempat lain, yang memiliki karakteristik pajanan yang serupa dengan tempat yang dilaporkan tersebut. Namun, hal tersebut seringkali tidak mungkin dilakukan karena adanya karakteristik khusus, yang mungkin hanya ada di tempat yang dilaporkan tersebut. Bila dihadapkan pada dilema ini, maka diperlukan pendekatan yang memberikan kompensasi untuk terjadinya bias positif tanpa memasukkan bias negatif. Hal ini bisa dilakukan dengan cara mengeluarkan suatu *cluster* yang dilaporkan dapat memberikan bias negatif pada hubungan pajanan-efek, jika masyarakatnya kecil, hasilnya jarang, dan sebagian besar kasus pada saat ini telah diketemukan. Penelitian kohort pada kelompok sampel yang benar-benar independen bisa dilakukan bila pajanan masih berlangsung; namun demikian desain ini pada umumnya tidak mudah/feasible untuk meneliti efek yang jarang.

Penilaian Interaksi

Individu atau subyek dalam sebagian besar penelitian epidemiologi lingkungan terpajan oleh campuran atau kombinasi dari beberapa zat kimia yang beracun. Bila hal ini terjadi, maka selain menilai hubungan antara masing-masing pajanan tersebut dengan efek, harus diperhitungkan pula adanya interaksi atau hubungan di antara beberapa pajanan (zat kimia) tersebut.²⁾ Hal ini bisa dimengerti, karena di antara beberapa zat kimia

bisa terjadi interaksi (reaksi kimia), yang bisa berupa efek aditif (efek saling menambah), sinergis (efek saling menguatkan), potensi (zat kimia yang satu meningkatkan potensi zat kimia lainnya), atau antagonis (efek meniadakan, mengurangi; zat kimia yang satu mempunyai efek mengurangi potensi dari zat kimia lainnya).⁷⁾

Beberapa penulis telah mempertimbangkan isu statistik/epidemiologi dari interaksi sebagaimana penerapannya pada pengaruh kombinasi dari dua buah pajanan atau lebih. Walaupun metoda analitik untuk menilai interaksi telah dikembangkan, penerapannya pada studi lingkungan mungkin terbatas, karena analisis yang paling berguna dan dapat ditinterpretasikan dari interaksi membutuhkan aplikasi teknik statistik multivariate. Namun, secara tipikal data yang diperoleh dari studi epidemiologi lingkungan pada umumnya lemah dan tidak sempurna, sehingga hasil dari analisis multivariat tidak dapat dipertanggungjawabkan. Beberapa jenis model statistik juga telah tersedia untuk menilai interaksi antar variabel ini, namun masih dipertanyakan mana yang paling sesuai. Model statistik yang paling relevan untuk menggambarkan hubungan faktor biologis di lingkungan dengan penyakit masih belum bisa dipastikan, sampai bisa diketahui lebih banyak bagaimana beragamnya paparan zat kimia menghasilkan keadaan sakit atau pengaruh biologi.

Penentuan Tingkat Investigasi

Investigasi pengaruh lingkungan terhadap kesehatan dapat diklasifikasikan ke dalam tiga tingkat. Penelitian tingkat I didasarkan pada data laporan pajanan dan efek (penyakit) yang telah ada, diambil secara rutin, dan mudah didapatkan. Penelitian tingkat I dapat mengambil data dari statistik vital (misalnya surat kelahiran, dan sebagainya), registrasi khusus untuk tumor atau malformasi untuk melihat berat badan lahir, mortalitas perinatal, insidens kanker atau mortalitas, atau adanya faktor keturunan. Power pada penelitian tingkat I sangat rendah, antara lain karena keterbatasan di dalam pengendalian variabel pengganggu (konfunder).

Penelitian tingkat II meliputi studi epidemiologi jangka pendek, seperti *cross-sectional*, kasus kontrol, atau kohort jangka pendek, yang membutuhkan pengumpulan data tentang karakteristik individu, data pajanan dan dampak kesehatan (penyakit), serta data tentang faktor pengganggu yang lebih cermat. Beberapa alasan untuk melanjutkan penelitian tingkat I ke penelitian tingkat II antara lain adalah:²⁾

1. Ditemukan hubungan yang bermakna antara pajanan dengan efek pada studi tingkat I, namun diperlukan adanya informasi yang lebih mendalam tentang hubungan tersebut,

Beberapa Isu Statistik

misalnya untuk mengeksplorasi spektrum yang lebih luas dari hasil dan/atau untuk menspesifikasikan hubungan pajanan-efek dengan lebih teliti (misalnya ambang dan komponen dosis-respons dari paparan).

2. Tujuan akhir dari penelitian (adanya dampak kesehatan), atau dampak lain yang masuk akal (mirip;koheren) akibat pajanan, tidak bisa didapatkan pada penelitian tingkat I
3. Hasil dari penelitian tingkat I belum bisa disimpulkan

Dalam studi tingkat II, pertimbangan statistik atas power, bias, dan interaksi dapat diterapkan untuk memilih desain penelitian, sehingga memungkinkan peneliti untuk memanfaatkan secara maksimal jumlah subyek yang sedikit, kejadian yang jarang, dan sumber-sumber informasi yang tidak tentu. Studi tingkat II mempunyai suatu jangkauan luas dari titik akhir (dampak kesehatan), dan dapat meliputi hasil yang diperoleh dari catatan medis (aborsi spontan, malformasi, gangguan perilaku atau psikologi), melalui wawancara dengan subyek studi (pada kejadian aborsi spontan, disfungsi seksual, gejala atau tanda rash, kelumpuhan, iritasi mata), atau melalui studi biologi dari subyek (biokimia, imunologi, dan penentuan seks khromosom, dan kecepatan konduksi saraf).

Penelitian tingkat III meliputi penelitian yang dirancang dengan baik dan dalam waktu yang cukup panjang, misalnya studi kohort di masyarakat yang akan membandingkan terjadinya dampak kesehatan pada sekelompok penduduk yang terpajan oleh suatu polutan (yang ada di lingkungan tempat tinggalnya, misalnya masyarakat yang tinggal di bawah listrik tegangan tinggi) dengan sekelompok penduduk yang tidak terpajan oleh polutan tersebut. Karena desain ini sesuai sekali untuk penyakit-penyakit dengan masa laten yang panjang, maka sangat bermanfaat untuk menemukan adanya faktor risiko atau pajanan yang bersifat karsinogenik. Penelitian tingkat III ini sangat memerlukan adanya data dari pencatatan dan pelaporan yang akurat, terpusat, dan mudah diakses, misalnya data kelahiran, kematian dan penyakit.

Permasalahan Dari Penelitian Ekologi ^{2,5)}

Penelitian ekologi adalah suatu penelitian yang datanya diambil dari data di populasi yang besar, atau dengan kata lain unit-unit analisisnya adalah populasi-populasi atau kelompok-kelompok orang, bukan individu-individu.⁵⁾ Sebagai contoh, ada yang menghubungkan konsumsi garam dengan kanker oesophagus di Cina. Dari data terlihat bahwa semakin banyak konsumsi garam per orang per tahun, semakin banyak pula penderita kanker oesophagus. Namun demikian, hasil penelitian tersebut tidak bisa menjelaskan adanya hubungan sebab-akibat, dan tidak dapat menjelaskan adanya faktor risiko lain yang

kemungkinan juga berperan terhadap kejadian kanker oesophagus.³⁾

Karena sifatnya yang sederhana dan praktis, maka penelitian ekologi sangat cocok untuk penelitian pendahuluan atau penelitian yang bersifat eksploratif. Namun demikian, banyak keterbatasan dalam interpretasi data yang dihasilkannya, antara lain karena: ²⁾

1. Data tentang faktor risiko luar (seperti umur, seks, pekerjaan, kebiasaan pribadi, dan lain-lain) mungkin tidak bisa didapatkan pada tingkat ekologi. Hal ini akan menyulitkan dalam upaya pengendalian variabel konfonding. Sebagai contoh: adanya variasi mortalitas menurut kondisi geografis mungkin lebih disebabkan adanya perbedaan-perbedaan dalam status sosioekonomi, konsumsi alkohol dan tembakau, atau diet daripada adanya perbedaan pajanan terhadap suatu polutan, terutama bila pajanan itu tidak sangat intens atau sangat beracun (toksik)
2. Populasi yang berisiko (*population at risk*) sering telah meninggalkan studi sebelum investigasi dimulai, sehingga penelitian dengan pendekatan *cross-sectional* yang memakai metode analisis ekologi, dapat mengarah pada bias seleksi
3. Data tahunan yang menggambarkan kecenderungan atau pola dari suatu kejadian di populasi biasanya sangat terbatas (pada umumnya hanya data kematian yang tersedia). Data morbiditas kadangkala bisa didapatkan, namun sulit untuk diinterpretasi, karena adanya ketidakpastian di dalam perubahan dan definisi penyakit. Selain itu, data trend tahunan, seringkali dipengaruhi oleh faktor luar, seperti angka kejadian dari penyakit kronis di populasi (suatu indikator dari kerentanan efek akibat pajanan zat toksik di lingkungan), dan beberapa faktor lain, seperti cuaca, infeksi saluran pernafasan, dan adanya bencana alam
4. Dalam analisis ekologi, terdapat kemungkinan untuk terjadinya bias dalam bentuk modifikasi efek, yang dikenal sebagai kekeliruan ekologis (*ecological fallacy*). Hal ini terjadi bila karakteristik yang berkaitan dengan faktor yang diteliti pada masing-masing kelompok tidak homogen. Secara teoritik, bias ini bisa membuat suatu asosiasi menjadi tampak lebih kuat atau sebaliknya menjadi tampak lebih lemah, dibanding yang ada pada tingkat individu; namun dalam praktik bias ini umumnya melampaui besaran suatu asosiasi yang benar, jika memang terdapat
5. Dengan analisis ekologi, variabel prediktor tertentu (khususnya variabel-variabel sosiodemografi dan lingkungan) cenderung lebih saling berkorelasi satu dengan yang lain

daripada pada tingkat individu, suatu fenomena yang dinamakan 'multikolinearitas'. Adanya korelasi di antara variabel-variabel prediktor akan menyulitkan di dalam upaya melihat pengaruh dari masing-masing prediktor terhadap terjadinya dampak kesehatan. Multikolinieritas merupakan masalah yang paling sering dihadapi di dalam penelitian ekologi, selain unit analisis secara geografis yang lebih luas atau lebih sempit.

Namun masalah ini dan masalah analisis ekologi yang lain, seperti kesalahan pengukuran dan ketidakjelasan hubungan sebab akibat, dapat diminimalkan melalui pendekatan-pendekatan berikut: (1) penggunaan metode regresi, tidak sekedar korelasi, dengan memasukkan ke dalam model faktor risiko sebanyak mungkin, (2) penggunaan data yang dikelompokkan ke dalam unit geografi sekecil mungkin (sebagai unit analisis, dan (3) berusaha untuk memastikan bagaimana kelompok-kelompok dibentuk dan menganalisis dengan menggunakan semua variabel yang diperkirakan berhubungan dengan proses pengelompokan.

Jumlah Sampel Tambahan dan Pertimbangan Power

Terdapat sejumlah faktor-faktor lain yang kurang begitu dikenal yang berpengaruh terhadap power statistik atau jumlah sampel dalam penelitian epidemiologi lingkungan. McKeown-Eyssen dan Thomas, seperti dikutip oleh Marsh, menelaah pengaruh distribusi pajanan pada perhitungan jumlah sampel dan menunjukkan bahwa akan terjadi estimasi yang berlebihan (*overestimate*) pada perhitungan jumlah sampel, apabila skala variabel pajanan yang berskala kontinu dijadikan dikotom. Hasil tersebut sangat relevan diterapkan pada distribusi pajanan dengan variabilitas yang sangat tinggi atau sangat rendah atau yang distribusinya tidak normal (*skewed*), seperti paparan radon, yang umumnya berdistribusi secara logaritma normal dalam populasi berisiko.

Sementara itu, Smith dan Day telah menghitung seberapa besar pengaruh efek konfounding terhadap jumlah sampel yang dibutuhkan pada penelitian kasus kontrol. Menurut perhitungan mereka, diperlukan penambahan jumlah sampel sebanyak 20 persen sampai 60 persen untuk bisa mendeteksi peningkatan risiko sebanyak dua kali dibanding hasil perhitungan risiko awal, setelah dilakukan pengendalian terhadap variabel pengganggu (konfounding). Hasil ini bisa dicapai apabila terdapat hubungan yang kuat antara pajanan dengan konfonder.

Semua metode tersebut dapat digunakan untuk menentukan power untuk menguji efek-efek interaksi. Metode-metode ini menunjukkan bahwa jika tujuan suatu penelitian adalah untuk

mendeteksi interaksi-interaksi penting di antara variabel pajanan, maka jumlah sampel paling tidak harus empat kali lebih besar dibanding bila tujuan dibatasi hanya untuk mendeteksi besar pengaruh dari satu variabel pajanan utama saja.

Keterbatasan dari Model-model Multivariat^{2,7)}

Dalam penelitian-penelitian lingkungan yang mencakup sejumlah besar variabel pajanan dan variabel pengganggu, metode deskriptif konvensional dan metode stratifikasi dalam analisis statistik mungkin tidak layak (*feasible*) untuk dilakukan, khususnya jika variabel-variabel studi berkaitan dengan waktu atau ada kepentingan untuk menilai interaksi di antara variabel-variabel pajanan dan variabel-variabel pengganggu. Akhir-akhir ini, telah terjadi kemajuan pesat dalam pengembangan model-model matematik multivariat, khususnya perangkat lunak komputer yang terkait, untuk menganalisis data yang demikian kompleks. Salah satu kelas model yang khususnya dipakai untuk analisis data prevalensi atau insiden adalah model multiplikasi. Model-model multiplikasi yang umumnya digunakan mencakup regresi logistik, regresi Poisson, dan model Cox proportional hazards. Jika digunakan dengan tepat, model-model ini menyajikan suatu kerangka kerja yang berguna untuk uji hipotesis, pengendalian variabel pengganggu, meningkatkan power statistik, dan mengeksplorasi mekanisme biologis dari suatu dampak kesehatan akibat pajanan lingkungan.

Penerapan yang tepat dari model-model statistik ini untuk data yang kompleks yang juga memerlukan serangkaian asumsi matematis yang lebih kompleks akan cukup memuaskan bagi peneliti, dan analisis yang didasarkan pada model-model ini dapat sangat sensitif terhadap adanya penyimpangan (*deviasi*) dari asumsi ini. Sebagai contoh, model logistik yang mencakup sejumlah variabel pajanan yang kuat, khususnya eksponensialitas dari hubungan pajanan-respon untuk variabel-variabel berskala kontinu, dan perkalian dari interaksi di antara beberapa variabel. Jika asumsi-asumsi seperti itu tidak bisa dipenuhi, maka model-model parametrik alternatif harus dipertimbangkan, misalnya dengan menggunakan metoda '*general relative risks models*' model-model risiko relatif umum" dengan menyusun hubungan pajanan-respon nonparametrik dan menguji '*goodness-of fit*' untuk variabel berskala kontinu.

Identifikasi Variabel Pengganggu

Dalam permodelan multivariat dari hubungan-hubungan pajanan-respons perhatian khusus harus diberikan pada adanya '*overadjusting*' untuk efek konfonder. *Overadjustment* dapat terjadi dari: (1) memasukkan suatu variabel yang merupakan variabel antara pada jalur hubungan pajanan-efek (penyakit); (2) pencakupan suatu variabel yang

Beberapa Isu Statistik

mempunyai korelasi sangat kuat dengan variabel pajanan, namun tidak mempunyai hubungan kausalitas dengan variabel efek/penyakit (dalam studi kasus kontrol, hal ini akan memperlebar varians *odds ratio* namun tidak akan mengintroduksi bias), dan (3) hanya memasukkan variabel-variabel yang mengurangi *odd ratio* dari hubungan variabel pajanan-efek yang diteliti.

Untuk mengetahui apakah suatu variabel dapat dikatakan sebagai variabel pengganggu dapat dilakukan analisis kovarians. Analisis kovarians dilakukan berdasarkan pertimbangan bahwa dalam kenyataannya ada variabel tertentu yang tidak dapat dikendalikan tetapi sangat mempengaruhi atau sangat berkorelasi dengan variabel efek yang diamati.⁸⁾ Namun demikian, di dalam menentukan apakah suatu variabel adalah benar-benar suatu pengganggu memerlukan pertimbangan epidemiologis dan pertimbangan kredibilitas biologis, dan tidak dapat hanya semata-mata diselesaikan dengan analisis statistik dari asosiasi yang tampak dalam data.

Permasalahan Dalam Menarik Kesimpulan (Inferensi) Ganda²⁾

Karena sebagian besar studi lingkungan mencakup hipotesis ganda, risiko kesalahan tipe I perlu dipertimbangkan. Ada dua aspek untuk masalah ini, yaitu hipotesis biologi yang multipel (ganda) dan banyaknya uji statistik yang bisa dipakai untuk menguji hipotesis yang diajukan. Salah satu cara (pendekatan) yang bisa dipakai untuk mengurangi masalah inferensi ganda adalah dengan cara membuat variabel komposit, yaitu suatu variabel yang merupakan gabungan dari satu atau lebih variabel-variabel yang diteliti. Contoh: membuat variabel komposit 'jumlah pajanan terhadap polutan udara' dengan cara menjumlahkan skor pajanan dari beberapa indikator (variabel) polutan udara, seperti skor pajanan terhadap SO₂, CO, NO_x, dan debu. Variabel komposit hasil bentukan baru ini kemudian bisa dimasukkan ke dalam model multivariat, yang hasilnya akan lebih baik bila dibanding masing-masing variabel secara individual yang dimasukkan ke dalam model tersebut. Modifikasi dengan cara ini akan memberikan nilai hubungan pajanan-efek yang lebih kuat dan lebih sederhana.

Dengan memperhatikan masalah hipotesis biologi ganda, terdapat sejumlah prosedur statistik yang biasa dipakai untuk mengatur (*adjusting*) perbandingan ganda, misalnya dengan cara mengalikan nilai-*p* dengan jumlah uji statistik, yang sesuai.

Manfaat Meta-analisis

Istilah meta-analisis (secara harfiah, analisis terhadap analisis-analisis) umumnya digunakan untuk menunjukkan suatu metode kuantitatif yang dikembangkan untuk memfasilitasi sintesis hasil

dari bermacam-macam penelitian dengan hipotesis yang saling berkaitan. Hal ini merupakan masalah utama ketika kita berusaha melakukan kaji ulang (*review*) dan menginterpretasi suatu badan ilmu pengetahuan (*body of knowledge*) dari hasil penelitian yang dipublikasikan. Secara tipikal, badan pengetahuan ini didasarkan pada sejumlah penelitian yang berbeda dalam desain, analisis, validitas, reliabilitas dan kesimpulan.

Walaupun terdapat sejumlah rangkaian kegiatan dalam metode meta-analisis yang khusus, kebanyakan adalah mencakup suatu simpulan kuantitatif dari beberapa hasil studi yang melakukan investigasi terhadap pertanyaan penelitian yang bersifat umum. Penerapan meta-analisis dalam epidemiologi lingkungan antara lain mencakup efek dari merokok pasif, hubungan antara pajanan asbes dan kanker usus/gastrointestinal, dan hubungan antara klorinasi air dengan kanker tertentu.

Keuntungan dari meta-analisis antara lain adalah untuk meningkatkan ketelitian (presisi) dalam estimasi risiko, sebagai review pustaka yang lebih sempurna dan sistematis, dan sebagai suatu penilaian yang lebih formal dari konsistensi temuan-temuan dari berbagai penelitian. Kekurangan yang utama dari meta-analisis dalam epidemiologi lingkungan antara lain adalah seringkali menghapus studi-studi 'negatif' yang tidak dipublikasikan, ketidakmampuan untuk mengevaluasi kualitas penelitian individual yang ditelaah dalam review, adanya kesulitan dalam menggabungkan penelitian-penelitian yang berbeda desainnya, lemahnya metoda untuk mengendalikan konfounding, dan adanya keterbatasan dalam pendekatan untuk penilaian/penaksiran pajanan.

KESIMPULAN

Dari beberapa uraian yang telah disampaikan dapat diambil beberapa simpulan, sebagai berikut:

1. Tujuan dari penelitian epidemiologi lingkungan adalah untuk mencari hubungan antara pajanan terhadap polutan yang ada di lingkungan dengan timbulnya gangguan kesehatan, sehingga dapat dijelaskan adanya hubungan sebab-akibat di antara keduanya
2. Terdapat beberapa keterbatasan di dalam pelaksanaan penelitian epidemiologi lingkungan, yaitu populasi berisiko yang tidak banyak; karakteristik yang berbeda di antara populasi yang berisiko; informasi tentang efek pajanan yang masih sangat terbatas; dampak atau efek kesehatan yang timbul pada umumnya jarang, bersifat kronis, dan membutuhkan pajanan dalam jangka panjang sampai timbulnya efek; adanya bias laporan; dan adanya rasa takut bila ditemukan efek kesehatan yang bermakna

3. Beberapa isu statistik yang harus diperhatikan dalam penelitian epidemiologi lingkungan, adalah:
 - a. Diperlukan suatu penelitian eksplorasi pada pada keadaan di mana informasi dan pengetahuan tentang masalah yang akan diteliti masih sangat terbatas
 - b. Besarnya power dalam penelitian epidemiologi lingkungan ditentukan oleh jumlah sampel dan kelompok kontrol; rate dari efek atau pajanan, angka kesalahan tipe I; jumlah asosiasi yang akan diuji; besar hubungan antara pajanan dengan efek; dan desain penelitian serta teknik analisis statistik yang digunakan
 - c. Diperlukan identifikasi dan pengendalian terhadap sumber-sumber kesalahan sistematis, yakni: bias pengganggu (*confounding*), bias seleksi, bias informasi dan bias pelaporan
 - d. Diperlukan penilaian atas interaksi yang bisa terjadi di antara variabel-variabel penelitian
 - e. Diperlukan tingkat investigasi yang cukup tinggi agar suatu penelitian epidemiologi lingkungan bisa memberikan penjelasan tentang hubungan sebab-akibat antara pajanan dengan efek
4. Terdapat beberapa isu statistik lain yang bersifat khusus, yaitu:
 - a. Penelitian ekologi mempunyai banyak keterbatasan untuk interpretasi data yang dihasilkannya, karena: ketidaklengkapan data yang tersedia; populasi yang berisiko meninggalkan studi; banyaknya faktor luar yang berpengaruh terhadap data tahunan yang diamati; kemungkinan terjadi *ecological fallacy*; dan kemungkinan yang sangat besar untuk terjadinya multikolinearitas di antara variabel-variabel
 - b. Diperlukan tambahan sampel (20% sampai 60% dari perhitungan sampel awal) agar penelitian bisa mendeteksi adanya faktor risiko dan power penelitian tetap terjaga, setelah dilakukan pengendalian terhadap variabel pengganggu
 - c. Diperlukan pemilihan model-model statistik multivariat yang tepat untuk menjelaskan pengaruh dari sejumlah besar variabel pajanan, yang juga saling berinteraksi, terhadap variabel efek (*outcome*)
 - d. Bisa terjadi *overadjusting* pada permodelan multivariat dari hubungan-hubungan pajanan-respons
 - e. Untuk memudahkan (menyederhankan) di dalam pembuatan kesimpulan, perlu dipertimbangkan pembentukan variabel komposit yang merupakan gabungan dari beberapa variabel pajanan
 - f. Beberapa manfaat dari penelitian meta-analisis dalam epidemiologi lingkungan, yaitu untuk meningkatkan ketelitian (presisi) dalam estimasi risiko, sebagai review pustaka yang lebih sempurna dan sistematis, dan sebagai suatu penilaian yang lebih formal dari konsistensi temuan-temuan berbagai penelitian

DAFTAR PUSTAKA

1. Philp RB. Environmental hazards & human health. New York: Lewis Publishers, 1995:95-8.
2. Talbott EO, Craun GF. Introduction to environmental epidemiology. New York: Lewis Publishers, 1995:47-62.
3. Soemirat J. Epidemiologi lingkungan. Yogyakarta: Gadjah Mada University Press, 2000:156.
4. Sutomo AH, Kusnanto H. Dasar-dasar epidemiologi. Yogyakarta: Gadjah Mada University Press, 1997:192-213.
5. Sastroasmoro S, Ismael S. Dasar-dasar metodologi penelitian klinis. Jakarta: CV Sagung Seto, 2002:262-3.
6. Lemeshow S, Hosmer Jr DW, Klar J, and Lwanga SK. Adequacy of sample size in health studies. Chicester: John Wiley & Son Ltd, 1993:64-6.
7. Tabachnick BG, Fidel LS. Using multivariate statistics. 4th ed Boston: Allyn & Bacon, 2001:115-22.