BUKU AJAR

BIOSTATISTIKA DAN KOMPUTER STATISTIKA

BAGI MAHASISWA KEDOKTERAN

PANDUAN BERPIKIR KRITIS KLINIS



Dari Pertanyaan Klinis | Analisis Data yang Tepat | Aplikasi Komputer Statistika | Interpretasi & Keputusan Klinis

MASALAH KLINIS → PERTANYAAN KLINIS → DESAIN & DATA → ANALISIS STATISTIK → INTERPRETASI → KEPUTUSAN KLINIS

Omat Rachmat
Rukman Abdullah
Bohari

BUKU AJAR BIOSTATISTIKA DAN KOMPUTER STATISTIKA BAGI MAHASISWA KEDOKTERAN: Panduan Berpikir Kritis Klinis

Omat Rachmat
Rukman Abdullah
Bohari

PENERBIT SALNESIA
(CV. SARANA ILMU INDONESIA)

Buku Ajar Biostatistika dan Komputer Statistika bagi Mahasiswa Kedokteran: Panduan Berpikir Kritis Klinis

Penulis:
Omat Rachmat, Rukman Abdullah, Bohari

Desain Cover:
Tim Salnesia

Tata Letak:
Tim Salnesia

Editor:
Fatimah Fitriani Mujahidah

ISBN: 978-634-05-5952-1 (PDF)

xiv+152 Halaman, Ukuran (17,6 x 25,0 cm)

Cetakan Pertama:
September, 2026

Hak Cipta 2026, Pada Penulis

Hak Cipta Dilindungi Oleh Undang-Undang

Copyright © 2026
by Penerbit Salnesia
All Right Reserved

Dilarang keras menerjemahkan, memfotokopi, atau memperbanyak sebagian atau seluruh isi buku ini tanpa izin tertulis dari Penulis dan Penerbit

PENERBIT:
SALNESIA
(CV. SARANA ILMU INDONESIA)
ANGGOTA IKAPI
Pallantikang, Maros Baru, Kabupaten Maros,
Provinsi Sulawesi Selatan
Website: penerbit.salnesia.id/index.php/press
Instagram: @salnesia

KATA PENGANTAR

Puji syukur kami panjatkan ke hadirat Tuhan Yang Maha Esa atas terselesaikannya naskah buku ajar **Biostatistika dan Komputer Statistika bagi Mahasiswa Kedokteran: Panduan Berpikir Kritis Klinis**. Buku ini disusun bukan sekadar sebagai pelengkap kurikulum akademik, melainkan sebagai sebuah jembatan konseptual yang sangat dibutuhkan untuk menghubungkan angka-angka statistik abstrak di atas kertas dengan keputusan klinis nyata di sisi tempat tidur pasien (*bedside clinical decisions*).

Selama bertahun-tahun, pembelajaran biostatistika dalam pendidikan kedokteran sering kali terjebak dalam dua ekstrem yang kurang ideal. Di satu sisi, mahasiswa dijejali dengan rumus matematika rumit dan perhitungan manual yang segera terlupakan setelah ujian usai. Di sisi lain, mahasiswa dilatih secara mekanis hanya untuk menjadi "operator perangkat lunak" yang mahir mengklik tombol SPSS, Jamovi, atau R, tanpa benar-benar memahami mengapa uji tersebut dipilih, apa asumsi di baliknya, dan bagaimana menerjemahkan tabel output komputer menjadi rekomendasi klinis yang aman bagi pasien.

Buku ini hadir untuk menawarkan paradigma ketiga, yaitu **clinical biostatistical thinking (berpikir biostatistika klinis)**. Kami menempatkan pemahaman konseptual, estimasi ukuran efek (*effect size*), penilaian rentang ketidakpastian (uncertainty melalui confidence interval), serta deteksi bias metodologis sebagai pilar utama penalaran medis. Melalui buku ini, mahasiswa akan diajak untuk tidak lagi bertanya secara naif "apakah nilai p-value kurang dari 0,05?", melainkan beralih ke pertanyaan yang jauh lebih klinis: "seberapa besar efek intervensi ini, seberapa presisi estimasinya, dan apa dampaknya bagi kualitas hidup pasien saya?"

> *"Statistik tidak menghilangkan bias atau ketidakpastian dari suatu data klinis. Analisis statistik hanya valid sejauh desain penelitian, instrumen pengukuran, data, model analisis, dan asumsi yang digunakan mendukung inferensi yang dibuat. Oleh karena itu, statistik bukanlah alat ajaib untuk menutupi kelemahan metodologi, melainkan instrumen disiplin berpikir ilmiah."*

Sajian materi dalam buku ini disusun secara sistematis dan runtut ke dalam sepuluh bab yang saling berkesinambungan:

Bab 1: Logika Pengambilan Keputusan Klinis: Membangun fondasi skeptisisme ilmiah dan kerangka berpikir probabilitas klinis dalam kedokteran.

Bab 2: Dari Pertanyaan Klinis ke Pertanyaan Statistik: Mengajarkan keterampilan menerjemahkan skenario pasien nyata menjadi pertanyaan riset yang operasional menggunakan formulasi PICOT.

Bab 3: Memahami Data Klinis: Menguraikan klasifikasi tipe variabel data medis, struktur observasi, serta manajemen data hilang (missing data) secara bertanggung jawab.

Bab 4: Membandingkan Kelompok: Menyajikan logika dasar di balik pemilihan uji komparatif parametrik dan non-parametrik serta pemahaman ukuran efek absolut (Risk Difference, NNT/NNH) dan relatif (Risk Ratio, Odds Ratio).

Bab 5: Association ≠ Causation: Melatih penalaran kausal (causal thinking) menggunakan alat bantu grafis Directed Acyclic Graph (DAG) untuk mengidentifikasi perancu (confounder), mediator, dan collider.

Bab 6: Regresi, Confounding, dan Interaction: Menyistematiskan pemodelan multivariat serta mengajarkan cara membaca dan membangun model regresi (linear, logistik, Cox) secara logis.

Bab 7: Kedokteran Diagnostik: Menyediakan panduan praktis penafsiran uji penunjang sebagai pengubah probabilitas klinis (probability modifier) menggunakan pendekatan Bayesian.

Bab 8: Survival dan Prognosis: Mengupas tuntas analisis waktu kejadian (time-to-event) menggunakan kurva Kaplan-Meier dan model bahaya proporsional Cox.

Bab 9: Telaah Kritis Jurnal Kedokteran (Critical Appraisal): Mengajarkan metode evaluasi artikel ilmiah berdasarkan tiga pilar utama: Validitas Internal, Kepentingan Klinis, dan Aplikabilitas Pasien.

Bab 10: Statistika Komputer dan Interpretasi Output: Mempersiapkan mahasiswa untuk menghadapi keluaran (output) riil perangkat lunak statistik komputer dan menginterpretasikannya secara kritis.

Penulisan buku ini tentu tidak terlepas dari bantuan dan dukungan berbagai pihak. Kami menyampaikan penghargaan dan terima kasih yang

setinggi-tingginya kepada rekan-rekan sejawat dokter spesialis, para akademisi di Fakultas Kedokteran, serta para mahasiswa kedokteran yang draf pertanyaannya telah menginspirasi lahirnya buku ini. Apresiasi khusus juga kami sampaikan kepada Penerbit Salnesia yang telah membantu merealisasikan naskah ini menjadi buku yang utuh dan profesional.

Kami menyadari bahwa perkembangan ilmu kedokteran berbasis bukti (*Evidence-Based Medicine*) berjalan sangat dinamis. Oleh karena itu, masukan, kritik, dan saran konstruktif dari pembaca sekalian sangat kami harapkan demi penyempurnaan buku ini di masa mendatang.

Akhir kata, semoga buku ajar ini dapat memberikan manfaat yang nyata bagi mahasiswa kedokteran, dokter umum, residen (PPDS) dalam memperkokoh pemahaman biostatistika, meningkatkan kualitas penelitian kedokteran di Indonesia, dan yang terpenting, meningkatkan keselamatan serta kualitas perawatan pasien kita.

September, 2026

Tim Penulis

DAFTAR ISI

KATA PENGANTAR ..... **iii**
DAFTAR ISI ..... **vii**
DAFTAR TABEL ..... **xi**
DAFTAR GAMBAR ..... **xiv**
BAB 1 MENGAPA DOKTER MEMBUTUHKAN BIOSTATISTIKA **1**
1. Pendahuluan 1
2. Capaian Pembelajaran 3
3. Populasi, Sampel, Parameter, dan Statistik Sampel 4
4. Variabilitas Dalam Data Klinis: Mengapa Pasien Tidak Pernah Identik 5
5. Estimasi, Precision, dan Ketidakpastian 6
6. Random Variation dan Bias: Dua Hal yang Berbeda 7
7. Statistical Inference dan Penerapannya pada Pasien 8
8. Statistical Significance Versus Clinical Importance 11
9. Studi Kasus Tambahan dan Latihan Berpikir Klinis 14
10. Checklist 7 Pertanyaan Kritis Sebelum Mempercayai Sebuah Angka 16
11. Rangkuman 17
12. Latihan dan Evaluasi 18
BAB 2 DARI PERTANYAAN KLINIS MENJADI PERTANYAAN STATISTIK **21**
1. Pendahuluan 21
2. Capaian Pembelajaran 22
3. Dari Masalah Klinis ke Pertanyaan Statistik 23
4. Kerangka Pico dan Picot 24
5. Population, Intervention/Exposure, dan Comparator 25
6. Outcome: Operasionalisasi dan Waktu 27
7. Enam Keluarga Pertanyaan Penelitian 30
8. Studi Kasus Lintas Bidang 31
9. Association, Prediction, dan Causation: Jangan Disamakan 33
10. Pertanyaan yang Terlalu Luas atau Terlalu Sempit 34
11. Kesalahan Konseptual yang Sering Terjadi 35
12. Checklist: Before You Choose A Test 36
13. Rangkuman 36
14. Latihan dan Evaluasi 37
BAB 3 MEMAHAMI DATA KLINIS 41
1. Pendahuluan 41
2. Capaian Pembelajaran 42

3. Pertanyaan Besar dan Alur Berpikir 42
4. Angka Bukan Satu Jenis Data 43
5. Struktur Observasi Lebih Penting Daripada Nama Variabel 45
6. Distribusi Data 46
7. Apakah Semua Data Harus Diuji Normalitas? 47
8. Outlier (Pencilan) 48
9. Missing Data (Nilai Hilang) 48
10. Data Cleaning Bukan 'Membuat Data Menjadi Bagus' 49
11. Studi Kasus dan Contoh Lintas Bidang 50
12. Mitos dan Klarifikasi 51
13. Rangkuman 52
14. Latihan dan Evaluasi 52
15. Checklist: Before You Analyse A Variable 55

BAB 4 MEMBANDINGKAN KELOMPOK 57

1. Pendahuluan 57
2. Capaian Pembelajaran 57
3. Pertanyaan Besar dan Peta Konsep Perbandingan 58
4. Filosofi Dasar: Comparison = Contrast 59
5. Menentukan Unit Perbandingan 59
6. Mengapa Pairing Mengubah Filosofi Analisis? 60
7. Dua Kelompok Independen Dengan Outcome Kontinu 60
8. Mann-Whitney: Bukan Sekadar "Uji Median" 61
9. Dua Pengukuran Berpasangan 62
10. Outcome Kategorik: Chi-Square dan Fisher Exact 62
11. Tiga Atau Lebih Kelompok: Mengapa Bukan Banyak T-Test? 64
12. Kasus Integratif Dan Studi Kasus Gizi 65
13. "Normal Atau Tidak Normal?" — Reframing 66
14. Effect Size, Confidence Interval, Dan Clinical Importance 67
15. Mitos dan Klarifikasi 67
16. Rangkuman 68
17. Latihan dan Evaluasi 69
18. Checklist: Before You Compare Groups 70

BAB 5 ASSOCIATION ≠ CAUSATION 73

1. Pendahuluan 73
2. Capaian Pembelajaran 74
3. Pertanyaan Besar dan Peta Konsep 75
4. Association, Correlation, Prediction, dan Causation 76
5. Korelasi: Apa yang Diberitahukan dan Tidak Diberitahukan 77

6. Temporal Ordering: Sebab Harus Mendahului Akibat 78
7. Confounding: *"Third Variable"* yang Bukan Sekadar Variabel Ketiga... 78
8. Crude VS Adjusted Association 79
9. Mengapa "Adjust Semua Variabel" Dapat Salah 80
10. Mediator, Collider, dan Dag 80
11. Confounder ≠ Mediator ≠ Collider 82
12. RCT dan Studi Observasional Dalam Causal Inference 82
13. Bahasa Kausal Yang Proporsional 83
14. Studi Kasus dan Mini-Case 83
15. Mitos dan Klarifikasi 85
16. Checklist 8 Pertanyaan Sebelum Menggunakan Bahasa Kausal 86
17. Rangkuman 86
18. Latihan dan Evaluasi 87

BAB 6 REGRESI, CONFOUNDING, DAN INTERACTION 91

1. Pendahuluan 91
2. Capaian Pembelajaran 91
3. Peta Konsep dan Clinical Trigger: Mengapa or Berubah? 92
4. Filosofi Regresi 93
5. Outcome Menentukan Keluarga Model 93
6. Linear Regression: Apa Arti B? 94
7. Logistic Regression: Odds Bukan Risiko 95
8. Count Outcomes: Poisson dan Negative Binomial 95
9. Cox Regression: Hazard Ratio Bukan Risk Ratio 96
10. Crude Vs Adjusted: Apa yang Berubah? 96
11. Confounder, Mediator, Dan Collider Dalam Model 97
12. Interaction dan Effect Modification: Ketika Efek Tidak Sama 97
13. Nonlinearity, Collinearity, Sparsity, dan Overfitting 98
14. Longitudinal Dan Clustered Data 98
15. Decision Framework dan Checklist Membaca Tabel Regresi 99
16. Rangkuman 100
17. Latihan dan Evaluasi 100

BAB 7 KEDOKTERAN DIAGNOSTIK (DIAGNOSTIC MEDICINE) 103

1. Pendahuluan 103
2. Capaian Pembelajaran 103
3. Peta Konsep Diagnostik dan Clinical Trigger 104
4. Tabel 2×2: Fondasi Berpikir Diagnostik 105
5. Sensitivitas dan Spesifisitas 106
6. PPV dan NPV: Membaca dari Arah Hasil Tes 107

7. Pre-Test Probability dan Likelihood Ratio 108
8. Threshold, Kurva ROC, dan AUC.....109
9. Diagnostic Accuracy ≠ Clinical Usefulness.....110
10. Bias pada Studi Diagnostic Accuracy 111
11. Studi Kasus dan Mini-Case Diagnostik.....111
12. Rangkuman 112
13. Latihan dan Evaluasi Formatif.....113
BAB 8 SURVIVAL dan PROGNOSIS.....115
1. Pengantar: Mengapa Analisis Survival Diperlukan?.....115
2. Konsep Utama: Data Tersensor (Censoring) 116
3. Estimasi Non-Parametrik: Kurva Kaplan-Meier.....117
4. Membandingkan Kurva: Uji Log-Rank 117
5. Regresi Survival: Model Cox Proportional Hazards (PH) 118
6. Penerapan Klinis: Model Prediksi Prognosis di Sisi Tempat Tidur 118
7. Evaluasi Formatif dan Latihan Mandiri 119
BAB 9 CRITICAL APPRAISAL (TELAAH KRITIS) 123
1. Pendahuluan 123
2. Struktur Dasar Telaah Kritis: Tiga Pilar EBM 123
3. Menilai Validitas Internal: Anatomi Bias dan Perancu 125
4. Menilai Kepentingan Klinis: Melampaui Nilai P.....126
5. Menilai Aplikabilitas: Menghubungkan Bukti dengan Pasien Nyata..127
6. Evaluasi Formatif: Studi Kasus Latihan Mandiri 128
BAB 10 STATISTIKA KOMPUTER DAN INTERPRETASI OUTPUT 133
1. Prinsip Utama: "Garbage In, Garbage Out" (GIGO) 133
2. Struktur dan Penyiapan Data: Tidy Data Format.....133
3. Panduan Integratif Membaca Output Multivariat 134
4. Miskonsepsi Umum dan Cara Menghindarinya.....135
5. Latihan Formatif: Membaca Output dan Studi Kasus 135
DAFTAR PUSTAKA 139
GLOSARIUM ISTILAH.....145
PROFIL PENULIS.....151

DAFTAR TABEL

Tabel 1. Variabilitas dalam data klinis: mengapa pasien tidak pernah identik 5
Tabel 2. Perbandingan presisi pada dua penelitian hipotetik dengan point estimate sama 7
Tabel 3. Perbedaan fundamental antara variasi acak (random variation) dan bias sistematis 8
Tabel 4. Informasi metodologis klinis yang wajib dijawab sebelum menyimpulkan angka mortalitas 10
Tabel 5. Dua skenario dengan perbedaan absolut yang sama tetapi jumlah informasi berbeda 11
Tabel 6. Empat komponen yang menjawab pertanyaan berbeda dalam interpretasi hasil penelitian 13
Tabel 7. Mitos umum mengenai biostatistika dan klarifikasi ilmiahnya 13
Tabel 8. Checklist 7 pertanyaan kritis sebelum mempercayai sebuah angka penelitian 16
Tabel 9. Tingkat peryantaan dan contoh rumusan pertanyaan 23
Tabel 10. Komponen PICOT 24
Tabel 11. Pertanyaan klinis spesifik 25
Tabel 12. Istilah metodologis intervention and exposure 26
Tabel 13. Jenis comparator dan pertanyaan klinis yang dijawab 27
Tabel 14. Elemen outcome yang operasional 27
Tabel 15. Klasifikasi outcome: primer, sekunder, harm, dan eksploratif 28
Tabel 16. Pengaruh faktor waktu (time): outcome sama, makna berbeda 29
Tabel 17. Enam keluarga pertanyaan penelitian 30
Tabel 18. Pertanyaan klinis dokterm pertanyaan statistik konseptual, keluarga pertanyaan penelitian 31
Tabel 19. Tahapan versi pertanyaan 31
Tabel 20. Rumusan pertanyaan penelitian gizi 32
Tabel 21. Makna kosakata 'Berguna' 33
Tabel 22. Asosiasi, prediksi, dan efek kausal 33
Tabel 23. Pertanyaan yang terlalu luas atau terlalu sempit 34
Tabel 24. Lima kesalahan konseptual klasik yang wajib dihindari oleh mahasiswa kedokteran 35
Tabel 25. Checklist pertanyaan reflektif 36
Tabel 26. Menyajikan contoh visual bagaimana variabel klinis direpresentasikan secara statistik beserta pertanyaan kritis yang

wajib diajukan oleh seorang dokter sebelum memulai analisis data 44
Tabel 27. Mengilustrasikan bagaimana variabel yang sama (HbA1c) dapat melahirkan struktur data yang berbeda dengan konsekuensi konseptual analitik yang bervariasi 46
Tabel 28. Mekanisme missing 48
Tabel 29. Dataset tunggal ini dapat melahirkan beragam pertanyaan statistik yang sangat berbeda 50
Tabel 30. Menyajikan contoh variabilitas tipe variabel dan struktur data di berbagai bidang spesialisasi klinis kedokteran dan ilmu gizi 50
Tabel 31. Merangkum beberapa mitos yang sangat populer di kalangan peneliti kedokteran beserta klarifikasi ilmiahnya 51
Tabel 32. Panduan cepat anda sebelum melakukan analisis terhadap variabel klinis 55
Tabel 33. Data hipotetik perbandingan penurunan SBP antara dua kelompok terapi 61
Tabel 34. Metode Kandidat untuk data berpasangan (paired) 62
Tabel 35. Tabel kontingensi kejadian hipoglikemia pada dua regimen insulin 64
Tabel 36. Welch ANOVA dan Kruskal-Wallis sebagai alternatif untuk perbandingan lebih dari dua kelompok 65
Tabel 37. Contoh bagaimana variabel kadar hba1c memerlukan uji berbeda sesuai pertanyaan 66
Tabel 38. Jenis ukuran efek yang relevan berdasarkan tipe outcome 67
Tabel 39. Mitos umum dalam perbandingan kelompok dan klarifikasinya. 68
Tabel 40. Checklist delapan pertanyaan sebelum membandingkan kelompok 71
Tabel 41. Contoh hipotetik perubahan estimasi odds ratio (or) sebelum dan sesudah adjustment 79
Tabel 42. Perbedaan konseptual dan konsekuensi analitik confounder, mediator, dan collider 82
Tabel 43. Satu variabel (CRP) dengan beragam pertanyaan dan analisis konseptual 83
Tabel 44. Mitos yang umum beredar mengenai kausalitas dan klarifikasinya 85
Tabel 45. Evolusi Perubahan Estimasi Kekuatan Hubungan (Odds Ratio) Hipertensi 92
Tabel 46. Panduan pemilihan keluarga model regresi berdasarkan tipe outcome 94

Tabel 47. Contoh interpretasi koefisien β pada model linear 94
Tabel 48. Panduan interpretasi output regresi logistik 95
Tabel 49. Alasan di balik perubahan estimasi efek setelah adjustment..... 96
Tabel 50. Ringkasan peran dan tindakan kovariat dalam model regresi 97
Tabel 51. Contoh Kasus: Rasio Risiko Sama, tetapi manfaat absolut berbeda 97
Tabel 52. Checklist 10 pertanyaan kritis sebelum membaca tabel regresi .. 99
Tabel 53. Lima komponen dasar penilaian kedokteran diagnostik 104
Tabel 54. Struktur tabel kontingensi 2×2 akurasi diagnostik 105
Tabel 55. Perbandingan PPV pada dua klinik dengan prevalensi berbeda (Tes yang Sama: Se=90%, Sp=90%) 107
Tabel 56. Interpretasi praktis nilai likelihood ratio (LR) untuk keputusan klinis 109
Tabel 57. Meringkas perbedaan mendasar antara data kejadian (event) dan data tersensor (censoring) untuk memberikan landasan pemahaman sebelum melangkah ke kurva kelangsungan hidup 116
Tabel 58. Memberikan contoh perhitungan ilustratif bagaimana probabilitas kelangsungan hidup kumulatif dihitung secara interval 117
Tabel 59. Kerangka kerja tiga pilar telaah kritis jurnal kedokteran 124
Tabel 60. Contoh kasus perbandingan signifikansi statistik vs. kepentingan klinis 127
Tabel 61. Perbandingan format penyusunan data Untidy vs. Tidy 134
Tabel 62. Contoh output regresi linear multivariat (Outcome: Kadar HbA1c %) 134
Tabel 63. Contoh output regresi logistik multivariat (outcome: mortalitas 30 hari) 135

DAFTAR GAMBAR

Gambar 1. Peta alur berpikir dari asosiasi menuju kesimpulan kausal..... 76

BAB 1
MENGAPA DOKTER MEMBUTUHKAN BIOSTATISTIKA

Logika Pengambilan Keputusan Klinis di Tengah Variabilitas Biologis

1. Pendahuluan

Dokter berhadapan dengan variasi hampir setiap hari: tekanan darah pasien berubah dari waktu ke waktu, respons terhadap terapi tidak identik antarindividu, hasil pemeriksaan laboratorium tidak pernah sempurna, dan prognosis pasien dengan diagnosis yang sama pun kerap berbeda. Biostatistika menyediakan konsep dan metode yang sistematis untuk mendeskripsikan variasi tersebut, mengestimasi besarnya suatu fenomena atau efek terapi, serta menalar ketidakpastian medis dengan asumsi yang dinyatakan secara eksplisit.

Perlu ditegaskan sejak awal bahwa pembahasan pada bab ini tidak dirancang untuk mengajarkan cara menghitung statistik secara manual atau menjalankan perangkat lunak statistik seperti SPSS, R, maupun Stata. Fokus utama bab ini adalah membangun fondasi berpikir klinis yang kritis (clinical biostatistical thinking) yang mutlak diperlukan sebelum seorang dokter memilih, membaca, atau menginterpretasikan suatu analisis statistik. Pendekatan ini menempatkan estimasi, ukuran efek (effect size), ketidakpastian (uncertainty), dan interpretasi hasil yang proporsional sebagai bagian integral dari pelaporan serta penalaran statistik medis (Gardner dan Altman, 1986; Greenland et al., 2016; Lang dan Altman, 2015).

BATAS KLAIM

Statistik tidak menghilangkan bias atau ketidakpastian dari suatu data. Analisis statistik hanya valid sejauh desain penelitian, instrumen pengukuran, data, model analisis, dan asumsi yang digunakan mendukung inferensi yang dibuat. **Oleh karena itu, statistik bukanlah alat ajaib untuk menutupi kelemahan desain penelitian.**

RASIONALISASI

Kesalahan interpretasi statistik dalam literatur kedokteran jarang sekali disebabkan oleh ketidakmampuan teknis dalam menjalankan perangkat lunak komputer. Sebagian besar kesalahan krusial justru muncul saat pembaca atau peneliti mencampuradukkan populasi

> dengan sampel, menganggap ukuran sampel yang besar secara otomatis menghilangkan bias sistematis, menafsirkan p-value sebagai probabilitas bahwa hipotesis nol benar, atau secara keliru menyamakan signifikansi statistik (statistical significance) dengan kepentingan klinis (clinical importance). Literatur metodologis telah lama mendokumentasikan implikasi serius dari salah interpretasi p-value, confidence interval, dan kekuatan statistik (statistical power) terhadap keputusan klinis (Greenland et al., 2016; Wasserstein dan Lazar, 2016).

Dengan memahami materi pada bab ini, mahasiswa kedokteran diharapkan mampu:

- Membedakan variasi biologis, variasi pengukuran, variasi sampling (sampling variation), dan bias sistematis secara tepat.
- Membaca hasil penelitian klinis sebagai estimasi yang disertai ketidakpastiannya, bukan sekadar pelabelan biner "signifikan" atau "tidak signifikan".
- Menghubungkan hasil statistik dengan kualitas desain penelitian serta konteks klinis pasien yang dihadapi di dunia nyata.
- Menggunakan bahasa inferensi yang proporsional dan tidak berlebihan terhadap bukti ilmiah yang tersedia.

Pertanyaan Besar Bab Ini

Jika komputer dapat menghitung statistik dalam hitungan detik, mengapa dokter tetap perlu memahami biostatistika?

Jawaban yang diharapkan bukanlah "agar dokter dapat menghitung sendiri formulasinya", melainkan agar dokter memiliki kompetensi klinis untuk memilih pertanyaan penelitian yang tepat, menilai kualitas bukti ilmiah secara kritis, menginterpretasikan rentang ketidakpastian, mengenali potensi bias, dan pada akhirnya menerjemahkan hasil statistik tersebut ke dalam konteks perawatan pasien secara individual.

> **CATATAN PENTING**
> **Inferensi statistik tidak secara otomatis menjamin validitas inferensi klinis atau hubungan kausal.** Validitas interpretasi klinis sangat bergantung pada kualitas desain studi, seleksi peserta, presisi pengukuran, penanganan data yang hilang (missing data), ketepatan asumsi model, serta pengendalian bias dan faktor pengganggu (confounding).

STATISTIK TIDAK MENGHILANGKAN KETIDAKPASTIAN

Statistik membantu kita mendeskripsikan apa yang diamati, mengestimasi parameter atau ukuran efek, mengkuantifikasi sebagian ketidakpastian, dan membuat inferensi berdasarkan model tertentu. Karena keterbatasan ini, penggunaan bahasa yang terlampau kuat seperti "hasil penelitian ini membuktikan bahwa..." sering kali kurang tepat dan berisiko di dunia akademis. Dalam sebagian besar konteks penelitian medis, bahasa yang lebih santun, rendah hati, dan proporsional seperti **"hasil menunjukkan", "estimasi mengarah pada",** atau **"pada kondisi dan asumsi penelitian ini"** jauh lebih direkomendasikan.

RENDAH HATI SECARA STATISTIK (STATISTICAL HUMILITY)

Ketidakpastian (uncertainty) bukanlah kelemahan ilmiah yang harus disembunyikan atau ditutupi dengan retorika yang seolah-olah pasti. **Ketidakpastian adalah bagian dari informasi ilmiah berharga yang harus dilaporkan secara transparan, jujur, dan dipertimbangkan secara matang dalam proses pengambilan keputusan klinis di samping tempat tidur pasien.**

2. Capaian Pembelajaran

Setelah menyelesaikan pembelajaran pada bab ini, mahasiswa kedokteran diharapkan mampu:

a. Menjelaskan perbedaan mendasar antara populasi, sampel, parameter, dan statistik sampel dalam konteks penelitian klinis.
b. Mengidentifikasi dan menjelaskan alasan mengapa variabilitas selalu ada pada data klinis serta mengenali sumber-sumber utamanya.
c. Membedakan secara konseptual antara variasi data (data variation), variasi sampling (sampling variation), presisi (precision), ketidakpastian (uncertainty), dan bias sistematis.
d. Menjelaskan proses penarikan kesimpulan statistik (statistical inference) secara konseptual tanpa bergantung pada kalkulasi rumus matematika.
e. Menjelaskan secara rasional mengapa mutu hasil penelitian tidak cukup dan tidak boleh dinilai hanya berdasarkan nilai p-value.
f. Membedakan bukti statistik atau signifikansi statistik (statistical significance) dari kepentingan klinis yang nyata (clinical importance).
g. Mengidentifikasi informasi kritis tambahan yang masih diperlukan sebelum membuat interpretasi klinis dari angka-angka penelitian sederhana.

h. Menyatakan keterbatasan kesimpulan ilmiah secara proporsional sesuai dengan desain studi dan data yang tersedia.

Kompetensi yang Tidak Ditargetkan pada Bab Ini

- Memilih atau menghafal seluruh jenis uji statistik spesifik.
- Menghitung nilai p-value atau confidence interval secara manual menggunakan rumus matematika.
- Menjalankan perangkat lunak pengolah data statistik (SPSS, R, Stata, Jamovi, dll.).
- Memilih model analisis regresi multivariat secara mandiri atau melakukan analisis kausalitas (causal inference) secara lengkap.

> **POSISI BAB INI DALAM KURIKULUM**
> Bab 1 dirancang sebagai fondasi konseptual utama untuk melatih penalaran statistik klinis. **Keterampilan teknis lanjutan seperti pemilihan metode pengujian, analisis regresi, kedokteran diagnostik, analisis kelangsungan hidup (survival), serta penafsiran tingkat mahir akan dibangun secara bertahap pada bab-bab berikutnya.**

3. Populasi, Sampel, Parameter, dan Statistik Sampel

Sebagai ilustrasi mendasar, bayangkan sebuah skenario di mana rata-rata tekanan darah sistolik dari 260 pasien hipertensi yang berhasil direkrut dalam suatu penelitian adalah 148 mmHg. Angka 148 mmHg ini merupakan statistik sampel, yaitu nilai ringkasan yang diperoleh langsung dari data subjek yang diamati dalam studi. Angka ini dapat kita gunakan sebagai estimasi untuk memperkirakan rata-rata tekanan darah sistolik yang sesungguhnya di populasi target (yang disebut parameter populasi).

Sangat krusial untuk dipahami bahwa meskipun statistik sampel digunakan sebagai estimasi parameter populasi, keduanya secara konseptual sama sekali tidak identik. Parameter populasi adalah nilai riil yang biasanya tidak pernah diketahui secara pasti karena keterbatasan praktis untuk mengukur seluruh populasi target. Sebaliknya, statistik sampel bersifat fluktuatif dan bergantung pada subjek mana yang terpilih masuk ke dalam sampel penelitian.

ANALOGI KLINIS SAMPLING

Mengambil sampel dalam penelitian klinis memiliki analogi yang sangat mirip dengan mengambil sebagian kecil bahan makanan dari wadah besar untuk dicicipi. Kualitas rasa (kesimpulan) yang kita peroleh dari cicipan tersebut sangat bergantung pada bagaimana sampel itu diambil (apakah representatif), seberapa akurat alat indra kita dalam mengecap (akurasi alat ukur), seberapa besar variasi bahan di dalam wadah, dan ketepatan proses pengukuran itu sendiri.

4. Variabilitas Dalam Data Klinis: Mengapa Pasien Tidak Pernah Identik

a. Sumber-Sumber Variabilitas

Untuk menganalisis data medis secara objektif, kita harus memahami mengapa data tersebut bervariasi. Secara garis besar, terdapat empat sumber utama variabilitas pada data klinis yang dirangkum dalam Tabel 1.

Tabel 1. Variabilitas dalam data klinis: mengapa pasien tidak pernah identik

Jenis Variabilitas	Penjelasan Klinis dan Konseptual
Variabilitas biologis	Pasien dengan diagnosis medis yang sama dapat memiliki karakteristik klinis yang berbeda akibat faktor usia, fungsi organ, genetika, komorbiditas, respons individual terhadap terapi, dan variabel biologi lainnya.
Variabilitas intra-individu	Nilai parameter klinis pada orang yang sama dapat berubah secara dinamis menurut waktu, kondisi fisik, aktivitas, ritme sirkadian (seperti tekanan darah pagi vs malam), tingkat stres, dan faktor eksternal lainnya.
Variasi pengukuran	Instrumen yang digunakan, teknik pengukuran, subjektivitas observer (pemeriksa), waktu kalibrasi alat, serta kondisi fisik ruang pemeriksaan dapat memengaruhi akurasi dan memicu variabilitas hasil pengukuran.
Variasi sampling ***(sampling variation)***	Dua sampel acak berbeda yang diambil dari satu populasi target yang sama hampir pasti tidak akan menghasilkan nilai statistik sampel yang persis sama karena faktor kebetulan dalam pemilihan subjek.

b. Standard Deviation dan Standard Error: Pembeda Konseptual

Dua istilah statistik yang sangat sering membingungkan dan disalahgunakan di dunia kedokteran adalah Standard Deviation (SD) dan Standard Error (SE). Secara fundamental:

Standard Deviation (SD) terutama menggambarkan seberapa lebar penyebaran observasi individual di dalam data sampel yang kita amati. SD mencerminkan variabilitas biologis riil atau variasi pengukuran pada subjek penelitian Anda.

Standard Error (SE) berkaitan erat dengan variabilitas sampling dari suatu estimator (seperti nilai rata-rata sampel). SE mengukur seberapa jauh nilai rata-rata sampel kemungkinan akan berfluktuasi jika kita mengambil sampel baru secara berulang-ulang dari populasi yang sama. Oleh karena itu, SE merefleksikan tingkat presisi (precision) dari prosedur estimasi statistik kita dalam kerangka model yang diasumsikan. Perbedaan fundamental ini mencerminkan batas tegas antara aspek deskriptif (description) dan aspek inferensi (inference) (Altman dan Bland, 2005).

> **JANGAN TERTUKAR!**
> **Standard Deviation (SD)** menjawab pertanyaan klinis: **"Seberapa beragam karakteristik individual pasien di dalam sampel penelitian saya?"** Sedangkan **Standard Error (SE)** membantu menjawab pertanyaan inferensial: **"Seberapa bervariasi estimasi rata-rata yang saya peroleh akibat ketidakpastian proses sampling, dan seberapa dekat estimasi ini dengan rata-rata populasi yang sesungguhnya?"** Oleh karena itu, kedua ukuran ini memiliki tujuan konseptual yang sangat berbeda dan sekali-kali tidak boleh saling dipertukarkan atau disalahgunakan hanya untuk membuat rentang variasi data tampak lebih sempit dalam publikasi jurnal ilmiah.

5. Estimasi, Precision, dan Ketidakpastian

Dalam penelitian klinis modern, pertanyaan ilmiah yang jauh lebih informatif bagi praktik medis bukanlah sekadar "apakah ada perbedaan efek?" (pertanyaan biner dikotomis), melainkan "berapa besar perbedaan atau efek terapi tersebut?" (pertanyaan kuantitatif). Sebagai contoh, jika suatu obat baru menghasilkan perbedaan rata-rata penurunan tekanan darah sistolik sebesar −6 mmHg dibandingkan terapi standar, angka −6 mmHg ini disebut sebagai estimasi titik (point estimate).

Namun, estimasi titik ini tidak akan memiliki makna klinis yang utuh jika dilaporkan tanpa disertai ukuran ketidakpastiannya (uncertainty). Penekanan pada estimasi efek dan penggunaan interval kepercayaan (confidence interval) sebagai pelengkap utama atau bahkan alternatif terhadap fokus eksklusif pada pengujian hipotesis *(p-value)* telah lama dianjurkan secara konsisten dalam literatur metodologi medis terkemuka (Gardner dan Altman, 1986).

Tabel 2. Perbandingan presisi pada dua penelitian hipotetik dengan point estimate sama

Metrik Statistik	Penelitian A	Penelitian B
Point Estimate (Selisih Rerata)	-6 mmHg	-6 mmHg
95% Confidence Interval (95% CI)	-7 hingga -5 mmHg	-18 hingga +6 mmHg

Meskipun point estimate kedua penelitian pada Tabel 2 adalah sama persis (-6 mmHg), Penelitian B memberikan rentang kemungkinan nilai parameter yang jauh lebih luas dan tidak pasti. Secara sederhana, interval kepercayaan (CI) yang lebih lebar pada Penelitian B menunjukkan tingkat presisi (precision) yang jauh lebih rendah dalam konteks prosedur sampling dan model statistik yang digunakan.

INTERPRETASI CONFIDENCE INTERVAL YANG TEPAT

Pada pendekatan frequentist, hindari membuat pernyataan keliru bahwa setelah nilai interval kepercayaan (CI) dihitung, terdapat probabilitas sebesar 95% bahwa parameter populasi yang sebenarnya "berada di dalam interval tersebut". Pemahaman yang benar adalah bahwa interval kepercayaan harus dimaknai melalui properti prosedur pengambilan sampel berulang dan asumsi model jangka panjang. Interpretasi CI juga tidak serta-merta menghapus kemungkinan adanya bias desain yang bersifat sistematis (Greenland et al., 2016).

6. Random Variation dan Bias: Dua Hal yang Berbeda

Dalam mengevaluasi keandalan draf data klinis, kita harus tegas membedakan antara variasi acak *(random variation)* dan bias sistematis (bias). Random variation berkaitan erat dengan fluktuasi statistik tak terduga yang muncul semata karena proses sampling acak serta komponen probabilistik lain yang dimodelkan. Pada hampir seluruh situasi, peningkatan ukuran sampel *(sample size)* yang lebih besar akan meningkatkan presisi (mempersempit rentang CI). Namun, ukuran sampel yang sangat besar sekali pun tidak akan otomatis menghilangkan bias sistematis penelitian.

Tabel 3. Perbedaan fundamental antara variasi acak *(random variation)* dan bias sistematis

Variasi Acak *(Random Variation)*	Bias Sistematis (Bias)
Dapat menyebabkan hasil statistik sampel berfluktuasi secara acak dari satu pengambilan sampel ke pengambilan sampel lainnya.	Merupakan penyimpangan searah yang bersifat sistematis, dapat bersumber dari bias seleksi subjek, bias pengukuran alat, bias desain studi, kekeliruan model analisis, atau proses non-acak lainnya.
Dapat diminimalkan dan dikendalikan secara efektif dengan cara meningkatkan ukuran sampel penelitian (n).	Tidak akan pernah bisa diatasi atau diperbaiki hanya dengan menambah ukuran sampel subjek.

CONTOH NYATA BIAS VS PRESISI

Sebagai contoh klinis konkret, jika sebuah alat pengukur tekanan darah (sfigmomanometer) mengalami kerusakan kalibrasi secara sistematis sehingga selalu membaca nilai 10 mmHg terlalu tinggi, maka melakukan pemeriksaan pada 10.000 pasien sekaligus tidak akan pernah memperbaiki kesalahan kalibrasi tersebut. **Estimasi titik yang Anda peroleh dari data besar tersebut akan tampak sangat presisi (rentang CI sangat sempit), tetapi hasilnya tetap salah arah secara sistematis (tidak valid).**

PRINSIP VALIDITAS

Presisi (precision) sekali-kali tidak identik dengan validitas (validity). Interval kepercayaan (CI) yang sempit tidak dengan sendirinya membuktikan bahwa desain penelitian klinis Anda telah bebas dari bias. Presisi hanya menjelaskan satu aspek ketidakpastian statistik akibat sampling; sedangkan validitas seutuhnya membutuhkan penilaian kritis yang lebih luas terhadap seluruh cara bukti ilmiah (evidence) dihasilkan.

7. Statistical Inference dan Penerapannya pada Pasien

Secara konseptual, penarikan kesimpulan statistik (statistical inference) adalah proses komprehensif menggunakan data yang berhasil diamati pada sampel untuk membuat pernyataan ilmiah mengenai parameter populasi, model, atau populasi target yang lebih luas dengan sangat memperhatikan aspek asumsi analitis dan ukuran ketidakpastian. Jembatan kokoh dari data sampel menuju target inferensi populasi tidak akan pernah bisa dibangun hanya oleh kalkulasi rumus matematika semata; jembatan tersebut sangat bergantung pada kualitas seleksi peserta, presisi

instrumen pengukuran, desain studi, penanganan data hilang (missing data), dan validitas asumsi model analisis yang digunakan.

Pelaporan statistik medis yang bermutu tinggi wajib menjelaskan metode penelitian secara transparan serta melaporkan estimasi ukuran efek beserta ukuran ketidakpastiannya yang sesuai (seperti confidence interval), bukan hanya menyajikan deretan angka p-value yang terisolasi (Lang dan Altman, 2015).

Lebih dari itu, suatu analisis data yang dinyatakan valid untuk kelompok populasi penelitian belum otomatis menjamin bahwa intervensi klinis tersebut tepat untuk diterapkan pada seorang pasien tertentu yang sedang Anda hadapi di klinik. Seorang dokter tetap wajib mempertimbangkan tingkat kesesuaian karakteristik populasi, relevansi luaran *(outcome)*, manfaat absolut, potensi bahaya *(harms)*, preferensi nilai pasien, kelayakan *(feasibility)*, komorbiditas, serta konteks sistem layanan kesehatan lokal dengan mengajukan lima pertanyaan kritis berikut:

a. Apakah pasien yang saya hadapi saat ini cukup serupa dengan karakteristik populasi subjek di dalam penelitian?
b. Apakah luaran *(outcome)* yang dinilai dalam studi merupakan hal yang benar-benar penting dan relevan bagi kehidupan pasien saya?
c. Seberapa besar perkiraan manfaat klinis absolut yang akan diperoleh pasien, dan bagaimana perbandingannya dengan potensi bahaya *(harm)* yang mengancam?
d. Apakah efek terapi yang diamati dalam studi cukup presisi untuk mendukung keputusan klinis yang berisiko?
e. Apakah konteks dan sumber daya tempat penelitian dilakukan dapat diterapkan secara realistis pada sistem layanan klinis saya?

1) Studi Kasus Pemantik: Rumah Sakit Mana yang Lebih Baik?

Bayangkan sebuah laporan publikasi yang menyajikan draf data mortalitas rawat inap di dua rumah sakit: Rumah Sakit A sebesar 5% dan Rumah Sakit B sebesar 8%. Mendengar draf laporan tersebut, Direktur Rumah Sakit A langsung melayangkan klaim sepihak: "Data statistik ini membuktikan bahwa kualitas perawatan rumah sakit kami jauh lebih baik daripada Rumah Sakit B."

Sebagai calon dokter yang terlatih berpikir kritis, apakah Anda dapat langsung menerima kesimpulan tersebut begitu saja? Jawabannya tentu saja tidak. Sebelum tergesa-gesa menerima kesimpulan Direktur tersebut, sejumlah informasi metodologis klinis yang krusial berikut ini wajib diidentifikasi terlebih dahulu:

Tabel 4. Informasi metodologis klinis yang wajib dijawab sebelum menyimpulkan angka mortalitas

Aspek Metodologis	Pertanyaan Klinis Kritis yang Harus Dijawab
Denominator (Penyebut)	Angka mortalitas 5% dan 8% masing-masing berasal dari total berapa banyak pasien? (Apakah 5 dari 100 pasien, atau 500 dari 10.000 pasien?)
***Case Mix* (Karakteristik)**	Apakah usia rata-rata, jenis diagnosis klinis, derajat keparahan penyakit (severity), pola rujukan (referral pattern), dan komorbiditas pasien di kedua rumah sakit sebanding?
Definisi *Outcome*	Apakah definisi mortalitas yang digunakan benar-benar sama? (Apakah mortalitas selama rawat inap saja, mortalitas 30 hari pasca-masuk, atau periode lainnya?)
Durasi Observasi	Apakah durasi pemantauan (follow-up time) dan proporsi pasien yang hilang dari pemantauan (loss to follow-up) di kedua kelompok sebanding?
Kualitas Data	Apakah pencatatan rekam medis, metode verifikasi kematian, penanganan data hilang (missing data), dan kualitas pengukuran di kedua rumah sakit setara?
Estimasi Efek	Berapa perbedaan risiko absolut (risk difference), efek relatif (relative effect), beserta rentang interval kepercayaannya (confidence interval) setelah dihitung formal?
Desain Analisis	Apakah perbandingan angka mortalitas tersebut bersifat mentah (crude comparison) atau sudah mempertimbangkan penyesuaian (adjusted) terhadap perbedaan tingkat keparahan pasien (case-mix adjustment)?
Potensi Bias	Apakah terdapat bias seleksi, bias informasi rujukan, atau bentuk penyimpangan sistematis lain yang secara klinis masuk akal memengaruhi data tersebut?

TUJUAN FILOSOFIS

Angka mortalitas mentah sebesar 5% versus 8% saja sama sekali belum cukup kuat untuk menyimpulkan rumah sakit mana yang memiliki mutu pelayanan lebih baik. Interpretasi ilmiah yang valid mutlak membutuhkan pemahaman mendalam tentang denominator, aspek komparabilitas pasien, kualitas desain penelitian, ukuran estimasi efek, rentang ketidakpastian statistik, serta konteks klinis secara utuh.

2) *Same Effect, Different Uncertainty*

Untuk memperdalam pemahaman mengenai pentingnya denominator dan kuantifikasi ketidakpastian statistik, mari kita tinjau dua skenario hipotetik di bawah ini:

Tabel 5. Dua skenario dengan perbedaan absolut yang sama tetapi jumlah informasi berbeda

Skenario Analisis	Rumah Sakit A	Rumah Sakit B
Skenario 1 (Skala Kecil)	5 / 100 pasien (5%)	8 / 100 pasien (8%)
Skenario 2 (Skala Besar)	500 / 10.000 pasien (5%)	800 / 10.000 pasien (8%)

Walaupun perbedaan absolut persentase mortalitas yang diamati pada kedua skenario di atas adalah sama persis (yaitu selisih 3%), jumlah informasi dasar yang melandasinya sangatlah kontras. Pada Skenario 2, ketidakpastian sampling *(sampling uncertainty)* dari perbandingan tersebut jauh lebih kecil karena didukung oleh data 10.000 pasien.

Meskipun demikian, dokter yang kritis tetap wajib mengajukan pertanyaan lanjutan: "Bagaimana jika seluruh pasien dengan kondisi klinis paling berat sengaja dirujuk ke Rumah Sakit B karena peralatan mereka lebih lengkap?" Pertanyaan ini kembali menegaskan prinsip fundamental bahwa ukuran sampel yang sangat besar dan presisi statistik yang tinggi sekali pun tidak akan pernah otomatis mengatasi bias confounding (faktor pengganggu) atau bias seleksi desain.

8. Statistical Significance Versus Clinical Importance

Bayangkan sebuah uji klinis obat antihipertensi baru yang melibatkan 100.000 subjek penelitian dan melaporkan bahwa obat tersebut berhasil menurunkan tekanan darah sistolik rata-rata sebesar 1,0 mmHg dengan nilai p-value yang sangat kecil ($p < 0{,}0001$). Apakah hasil yang sangat signifikan secara statistik tersebut otomatis bermakna penting secara klinis (clinically important)? Jawabannya adalah tidak.

Keputusan mengenai kepentingan klinis *(clinical importance)* dari suatu efek terapi memerlukan pertimbangan konteks klinis yang komprehensif, seperti besarnya ukuran efek *(magnitude)*, risiko dasar pasien *(baseline risk)*, efek samping obat *(adverse effects)*, beban biaya (cost), ketersediaan alternatif terapi lain, preferensi nilai pasien, serta ambang batas klinis yang relevan. Nilai *p-value* murni sama sekali tidak mengukur besarnya efek ataupun signifikansi klinis dari suatu temuan (Wasserstein dan Lazar, 2016).

PERTANYAAN KLINIS UTAMA

Signifikansi Statistik (Statistical Significance) bertanya tentang tingkat kompatibilitas draf data yang diamati dengan model statistik atau hipotesis nol tertentu. Sedangkan **Kepentingan Klinis (Clinical Importance)** bertanya apakah besarnya efek terapi yang ditemukan benar-benar memberikan dampak kesehatan yang bermakna bagi keputusan hidup pasien. **Sangat dilarang mencampuradukkan atau mempertukarkan kedua pertanyaan fundamental tersebut.**

1) Non-Significant Bukan Berarti Tidak Ada Efek

Hasil analisis statistik dengan nilai p-value di atas ambang konvensional (misalnya $p > 0,05$) sama sekali tidak membuktikan secara mutlak bahwa tidak ada efek intervensi. Ketika rentang interval kepercayaan (CI) dari penelitian tersebut masih sangat luas, draf data tersebut sebenarnya masih kompatibel dengan beberapa kemungkinan arah efek yang berbeda (baik efek menguntungkan maupun merugikan). Altman dan Bland (1995) memperingatkan komunitas kedokteran melalui kredo legendaris: "Absence of evidence is not evidence of absence" (ketiadaan bukti ilmiah yang kuat bukanlah bukti bahwa efek tersebut tidak ada).

BAHASA INFERENSI YANG LEBIH AMAN DAN JUJUR

Alih-alih secara gegabah menyatakan "terapi ini tidak memiliki pengaruh klinis apa pun", biasakanlah untuk menggunakan formulasi kesimpulan ilmiah yang jauh lebih hati-hati dan jujur sesuai data, misalnya: **"Estimasi efek terapi pada penelitian ini masih tidak presisi karena rentang interval kepercayaan yang lebar mencakup berbagai kemungkinan arah efek; draf data yang tersedia belum mencukupi untuk menyimpulkan besarnya efek klinis secara definitif dengan tingkat keyakinan yang memadai."**

Pada tahap pembelajaran dasar ini, mahasiswa kedokteran cukup menanamkan prinsip kuat bahwa p-value bukanlah probabilitas bahwa hipotesis nol benar, bukanlah probabilitas bahwa hasil penelitian terjadi murni "karena kebetulan", bukanlah ukuran besarnya efek biologis, dan bukanlah tolok ukur kepentingan klinis. American Statistical Association (ASA) secara resmi menegaskan bahwa kesimpulan ilmiah dan keputusan klinis tidak seharusnya didasarkan hanya pada apakah nilai *p-value* berhasil melewati suatu ambang batas formal yang kaku (Wasserstein dan Lazar, 2016).

Tabel 6. Empat komponen yang menjawab pertanyaan berbeda dalam interpretasi hasil penelitian

Komponen Evaluasi	Pertanyaan Ilmiah yang Dijawab
Effect estimate **(Estimasi Efek)**	Seberapa besar perkiraan efek intervensi atau perbedaan hasil yang terjadi? *("How big?")*
Confidence interval **(Interval Kepercayaan)**	Seberapa presisi estimasi efek tersebut dalam kerangka model dan prosedur sampling yang digunakan? *("How uncertain?")*
p-value **(Nilai P)**	Seberapa kompatibel data yang diamati dengan model statistik atau hipotesis nol tertentu dalam prosedur pengujian? *("How credible?")*
Clinical importance **(Kepentingan Klinis)**	Apakah besarnya efek tersebut benar-benar relevan, bermanfaat, dan berharga bagi keputusan medis pasien? *("Does it matter?")*

MANTRA BERPIKIR BIOSTATISTIKA KLINIS

Setiap kali Anda membaca sebuah angka hasil penelitian klinis, biasakan untuk melafalkan empat mantra berpikir kritis berikut:

1. **How big?** (Berapa besar efeknya?)
2. **How uncertain?** (Seberapa tidak pasti estimasinya?)
3. **How believable?** (Seberapa tepercaya bukti ilmiahnya?)
4. **How clinically important?** (Apakah makna temuan tersebut benar-benar berharga bagi pasien saya?)

2) Mitos dan Klarifikasi

Untuk meluruskan berbagai kesalahpahaman konseptual yang sering beredar di kalangan akademisi kedokteran, pelajari tabel klarifikasi mitos berikut:

Tabel 7. Mitos umum mengenai biostatistika dan klarifikasi ilmiahnya

Mitos Statistik yang Keliru	Klarifikasi Ilmiah Konseptual
"Ukuran sampel yang sangat besar menjamin validitas penelitian pasti tinggi."	Salah. Ukuran sampel yang besar hanya akan meningkatkan presisi (mempersempit rentang CI) tetapi sekali-kali tidak otomatis menghapus bias sistematis desain atau kalibrasi alat.
"Nilai *p* < 0,05 membuktikan bahwa efek terapi tersebut penting secara klinis."	Salah. Besarnya ukuran efek *(effect size)* dan kepentingan klinis (clinical importance) harus dinilai secara terpisah menggunakan keahlian medis, bukan angka p-value.
"Nilai *p* > 0,05 membuktikan secara pasti bahwa terapi	Salah. Hasil tidak signifikan dapat disebabkan oleh ukuran sampel yang kecil

Mitos Statistik yang Keliru	Klarifikasi Ilmiah Konseptual
tersebut tidak memiliki efek."	(underpowered). Selalu tinjau estimasi titik, rentang interval kepercayaan (CI), dan desain studi.
"Interval kepercayaan (CI) yang sangat sempit menjamin penelitian tersebut pasti benar."	Salah. Presisi bukanlah validitas. Jika desain penelitian Anda dipenuhi bias sejak awal, estimasi Anda akan presisi tetapi salah arah secara sistematis.
"Karena komputer dapat menghitung segalanya, dokter tidak perlu memahami teori statistik."	Salah. Perangkat lunak hanya mesin kalkulasi biner. Interpretasi klinis dan metode penalaran klinis yang bertanggung jawab mutlak memerlukan keahlian metodologi seorang dokter.

9. Studi Kasus Tambahan dan Latihan Berpikir Klinis

a. Kasus 1: Analisis Lama Rawat Inap (Length of Stay)

Kasus Hipotetik: Sebuah studi klinis melaporkan bahwa Terapi Medis Baru A menghasilkan rata-rata lama rawat inap (length of stay) di rumah sakit sebesar 0,4 hari lebih pendek dibandingkan Terapi Standar B dengan signifikansi statistik $p < 0,001$.

Pertanyaan Diskusi Kelas:

- ✓ Informasi krusial apa yang sebenarnya belum kita ketahui dari draf laporan tersebut?
- ✓ Apakah selisih 0,4 hari otomatis bermakna penting secara klinis (clinical importance)?
- ✓ Informasi tambahan apa yang wajib diperoleh sebelum Anda menarik kesimpulan medis?

JAWABAN YANG DIHARAPKAN

Mahasiswa kedokteran diharapkan mampu mengidentifikasi kebutuhan data kritis lainnya, seperti: nilai interval kepercayaan (95% CI) untuk menilai presisi, distribusi data lama rawat inap (apakah miring/skewed), karakteristik populasi pasien yang diteliti, kualitas desain penelitian, potensi efek samping (adverse effects), perbedaan total biaya perawatan, serta analisis rasio efisiensi biaya nyata dari pengurangan waktu yang hanya 0,4 hari tersebut.

b. Kasus 2: Risiko Relatif dan Interval Kepercayaan Lebar

Kasus Hipotetik: Uji klinis pada obat gagal jantung baru melaporkan hasil estimasi Risiko Relatif (Relative Risk/RR) mortalitas sebesar 0,60 (95% CI: 0,28 hingga 1,30; $p = 0,18$).

Pertanyaan Klinis: Apakah tepat jika Anda langsung menyimpulkan bahwa terapi medis baru ini sama sekali tidak bermanfaat hanya karena hasilnya tidak signifikan secara statistik (p > 0,05)?

INTERPRETASI KLINIS KRITIS

Kesimpulan tersebut sama sekali tidak tepat. Meskipun tidak signifikan, nilai point estimate (RR = 0,60) mengarah pada potensi penurunan risiko kematian yang cukup besar (sebesar 40%). Namun, karena rentang interval kepercayaan (CI) yang sangat lebar (mencakup nilai 0,28 hingga 1,30), draf data tersebut masih kompatibel dengan berbagai kemungkinan arah efek — mulai dari manfaat yang sangat besar (penurunan risiko 72%) hingga potensi bahaya (peningkatan risiko 30%). **Fokus kesimpulan medis dari penelitian ini harus ditekankan pada tingkat ketidakpastian yang sangat tinggi (uncertainty), bukan membuat klaim prematur bahwa efek terapi "tidak ada".**

c. Kasus 3: Presisi Tinggi dan Pembuktian Kausalitas

Kasus Hipotetik: Sebuah studi observasional berskala raksasa yang melibatkan data 100.000 pasien menunjukkan hubungan asosiasi yang sangat erat dan presisi tinggi antara Paparan Faktor Nutrisi X dan Penurunan Risiko Penyakit Kronis Y.

Pertanyaan Kritis: Apakah hubungan asosiasi yang sangat presisi tersebut sudah pasti bersifat sebab-akibat (kausal)?

JAWABAN KONSEPTUAL

Sama sekali tidak. Ukuran sampel yang masif dan tingkat presisi statistik yang tinggi tidak akan pernah bisa mengatasi kelemahan metodologis inheren studi observasional, seperti adanya faktor pengganggu yang tidak terukur (unmeasured confounding), bias seleksi subjek, kesalahan pengukuran instrumen, kausalitas terbalik (reverse causation), atau keterbatasan desain penelitian lainnya. Pembahasan mendalam mengenai kausalitas klinis akan dikupas secara tuntas pada bab-bab selanjutnya (Bab 5).

d. Think Like a Clinician, Read Like a Statistician

Perhatikan baik-baik draf pernyataan ilmiah yang salah berikut ini: “Intervensi medis ini terbukti 100% efektif bagi pasien karena nilai p-value = 0,03.” Cobalah perbaiki kalimat cacat logika tersebut tanpa mengarang data tambahan yang tidak tersedia.

CONTOH PERBAIKAN KALIMAT INFERENSI

"Kesimpulan mengenai efektivitas suatu intervensi medis tidak sebaiknya didasarkan hanya pada nilai p-value semata. Kita wajib menilai secara terpadu besarnya estimasi efek terapi (effect size), tingkat presisi interval kepercayaan (confidence interval), mutu desain penelitian, potensi bias sistematis, serta relevansi manfaatnya bagi keputusan hidup pasien."

DISIPLIN ANTI-HALUSINASI ILMIAH

Di dalam dunia kedokteran dan sains, jika informasi kritis yang diperlukan untuk menyusun kesimpulan tidak tersedia di dalam draf laporan ilmiah, maka satu-satunya jawaban ilmiah yang jujur dan benar adalah: **"Belum dapat disimpulkan secara memadai dari draf informasi yang diberikan."** Jangan pernah mencoba mengisi celah data yang kosong dengan asumsi subyektif Anda sendiri.

10. Checklist 7 Pertanyaan Kritis Sebelum Mempercayai Sebuah Angka

Gunakanlah ketujuh pertanyaan kritis di bawah ini sebagai kerangka kerja cepat (checklist) Anda setiap kali membaca draf angka atau luaran hasil penelitian klinis:

Tabel 8. Checklist 7 pertanyaan kritis sebelum mempercayai sebuah angka penelitian

Pertanyaan Kritis	Penjelasan Klinis dan Metodologis
What? (Apa)	Apa sebenarnya yang sedang diukur, dinilai, atau dibandingkan di dalam penelitian ini? Apakah skalanya relevan?
How big? (Berapa besar)	Berapa besar perkiraan ukuran efek (effect estimate) atau selisih perbedaan hasil nyata yang diamati?
How uncertain? (Seberapa ragu)	Bagaimana tingkat presisi dan ketidakpastian statistiknya? Berapa rentang batas atas dan batas bawah interval kepercayaannya (CI)?
How generated? (Bagaimana didapat)	Bagaimana kualitas desain penelitian dan proses pengumpulan data untuk menghasilkan bukti ilmiah (evidence) tersebut?
Could bias explain it? (Mungkinkah bias)	Apakah terdapat sumber bias sistematis atau faktor pengganggu (confounding) yang klinis masuk akal menjelaskan temuan tersebut?
So what? (Lalu mengapa)	Apakah besarnya ukuran efek tersebut benar-benar penting, bermakna, dan berharga secara klinis bagi kesehatan pasien?
My patient? (Pasien saya)	Apakah populasi subjek, latar belakang setting,

Pertanyaan Kritis	Penjelasan Klinis dan Metodologis
	dan konteks sistem penelitian ini relevan untuk diterapkan pada pasien saya?

JANGAN HANYA BERTANYA P-VALUE

Jangan pernah hanya bertanya: **"Apakah nilai p < 0,05?"** Tanyakanlah dengan kritis: berapa besar ukuran efeknya, seberapa lebar rentang ketidakpastian statistiknya, seberapa kredibel bukti ilmiahnya, dan apa arti sesungguhnya bagi keselamatan jiwa pasien Anda.

11. Rangkuman

a. Nilai statistik sampel sama sekali tidak identik dengan parameter populasi.
b. Variabilitas dinamis adalah bagian alami dan inheren dari seluruh data klinis kedokteran.
c. Setiap nilai estimasi titik wajib dibaca secara integral bersama ukuran ketidakpastiannya (rentang CI).
d. Tingkat presisi statistik sekali-kali tidak identik dan tidak menjamin validitas desain penelitian.
e. Ukuran sampel penelitian yang besar tidak akan pernah otomatis menghapus bias sistematis.
f. Nilai p-value bukanlah ukuran besarnya efek biologis dan bukan tolok ukur kepentingan klinis.
g. Hasil tidak signifikan (p > 0,05) tidak boleh diartikan secara gegabah sebagai "tidak ada efek terapi".
h. Signifikansi statistik dan kepentingan klinis merupakan dua pertanyaan ilmiah yang sangat berbeda.
i. Keandalan inferensi statistik sangat bergantung pada mutu desain, seleksi data, model analisis, dan asumsi.
j. Tujuan akhir mempelajari biostatistika bukanlah menghasilkan angka-angka rumit di komputer, melainkan memahami apa yang secara objektif, ilmiah, dan wajar dapat serta tidak dapat disimpulkan dari angka-angka tersebut.

Kalimat Penutup Bab

PESAN AKHIR

"Sebelum Anda tergesa-gesa bertanya 'uji statistik apa yang harus saya gunakan?', biasakanlah untuk terlebih dahulu merenung: apa sebenarnya pertanyaan klinis saya, parameter apa yang sesungguhnya ingin saya estimasikan, seberapa besar rentang

ketidakpastiannya, dan apa arti konkret hasil angka tersebut bagi kesembuhan pasien saya?"

12. Latihan dan Evaluasi

Soal Latihan Formatif Pilihan Ganda

1) Rata-rata tekanan darah sistolik dari sampel penelitian klinis adalah 142 mmHg. Angka 142 mmHg ini merupakan:
   A. Parameter populasi secara pasti
   B. Statistik sampel
   C. Confidence interval
   D. Ukuran bias sistematis

Kunci Jawaban: B. Statistik sampel. Penjelasan: Nilai yang diperoleh langsung dari data sampel subjek yang diamati dalam studi merupakan statistik sampel, sedangkan nilai sesungguhnya di populasi target disebut parameter.

2) Sebuah penelitian medis berskala sangat besar dilaporkan memiliki potensi bias seleksi subjek yang cukup berat. Pernyataan mana di bawah ini yang paling tepat?
   A. Ukuran sampel yang besar secara otomatis menghapus bias penelitian.
   B. Potensi bias sistematis tidak penting lagi jika nilai p-value < 0,05.
   C. Ukuran sampel yang besar meningkatkan presisi, tetapi tidak otomatis menghilangkan bias sistematis.
   D. Hasil penelitian tersebut dijamin pasti valid karena melibatkan jumlah pasien yang sangat besar.

Kunci Jawaban: C. Ukuran sampel yang besar meningkatkan presisi, tetapi tidak otomatis menghilangkan bias sistematis. Penjelasan: Sesuai Tabel 3, bias sistematis tidak akan pernah bisa diperbaiki hanya dengan menambah ukuran sampel.

3) Sebuah studi klinis melaporkan nilai p-value = 0,001. Informasi tersebut dengan sendirinya menunjukkan bahwa:
   A. Ukuran efek biologis terapi pasti sangat besar.
   B. Hasil penemuan tersebut dijamin pasti penting secara klinis (clinically important).
   C. Hipotesis nol (H0) memiliki probabilitas kebenaran sebesar 0,1%.
   D. Tidak ada kesimpulan mutlak dari pilihan A, B, maupun C.

Kunci Jawaban: D. Tidak satu pun dari A–C. Penjelasan: Sesuai dengan prinsip penafsiran p-value, nilai p-value tidak mengukur besarnya efek

biologis (A), tidak mengukur kepentingan klinis (B), dan bukan merupakan probabilitas hipotesis nol benar (C).

4) Dua penelitian klinis memiliki nilai estimasi titik (point estimate) yang sama, namun salah satunya memiliki rentang interval kepercayaan (95% CI) yang jauh lebih lebar. Interpretasi mana yang paling tepat?
   A. Nilai point estimate pada penelitian dengan rentang CI lebar pasti keliru.
   B. Penelitian dengan rentang CI yang lebih lebar menunjukkan tingkat ketidakpastian (uncertainty) yang lebih besar.
   C. Luaran dari kedua hasil penelitian tersebut konseptual adalah identik.
   D. Rentang CI tidak memberikan informasi metodologis tambahan apa pun.

Kunci Jawaban: B. Penelitian dengan rentang CI yang lebih lebar menunjukkan tingkat ketidakpastian (uncertainty) yang lebih besar. Penjelasan: Sesuai dengan draf Tabel 2, rentang CI yang lebar mencerminkan rendahnya presisi dan tingginya ketidakpastian sampling.

5) Sebuah uji klinis melaporkan nilai selisih rata-rata (Mean Difference/MD) tekanan darah sebesar -7 mmHg (95% CI: -16 hingga +2 mmHg). Interpretasi mana yang paling hati-hati dan jujur?
   A. Terapi tersebut telah terbukti 100% tidak memiliki efek klinis.
   B. Terapi tersebut telah terbukti secara ilmiah sangat bermanfaat bagi pasien.
   C. Estimasi titik mengarah pada penurunan, tetapi ketidakpastian statistik masih cukup besar dan rentang interval mencakup beberapa kemungkinan efek (termasuk efek membahayakan).
   D. Terapi medis tersebut sangat berbahaya dan harus segera dilarang.

Kunci Jawaban: C. Estimasi titik mengarah pada penurunan, tetapi ketidakpastian statistik masih cukup besar dan rentang interval mencakup beberapa kemungkinan efek. Penjelasan: Karena rentang CI melewati nilai nol (-16 hingga +2), data masih kompatibel dengan kemungkinan penurunan klinis yang besar maupun sedikit peningkatan tekanan darah.

Pertanyaan Diskusi Reflektif

Soal Reflektif:

Di dalam praktik pelayanan klinis kedokteran sehari-hari, manakah yang menurut Anda jauh lebih berbahaya bagi keselamatan pasien: ketidakpastian medis yang terlihat jelas (visible uncertainty) atau

ketidakpastian medis yang sama sekali tidak disadari (unrecognized uncertainty)?

Tujuan Diskusi:
Melatih mahasiswa kedokteran agar memahami secara mendalam bahwa rentang interval kepercayaan (CI) yang lebar menunjukkan ketidakpastian statistik yang terlihat jelas dan menuntut kehati-hatian. Sebaliknya, bias sistematis (bias) yang tidak disadari pada desain penelitian dapat melahirkan rasa percaya diri yang palsu (false confidence) karena estimasi hasil tampak sangat presisi, padahal salah arah secara fatal.

> **CATATAN METODOLOGIS**
> Rubrik tingkat kompetensi di atas murni merupakan alat bantu pedagogis (pembelajaran) yang diusulkan secara internal untuk mata kuliah ini, bukan sebuah skala kompetensi statistik baku yang telah divalidasi secara eksternal.

ALUR PEMBELAJARAN BERIKUTNYA
Jembatan Menuju Bab Selanjutnya:
Bab 1 telah tuntas menjawab pertanyaan besar mengenai "Mengapa dokter membutuhkan biostatistika sebagai instrumen berpikir?" Langkah selanjutnya pada Bab 2 adalah mempelajari bagaimana cara mengubah masalah klinis yang abstrak menjadi pertanyaan statistik yang terstruktur dan dapat dianalisis secara objektif menggunakan kerangka PICOT.

BAB 2
DARI PERTANYAAN KLINIS MENJADI PERTANYAAN STATISTIK

Merumuskan Komponen PICOT, Mengidentifikasi Estimand, dan Memetakan Target Analisis

1. Pendahuluan

Dalam praktik klinis sehari-hari, seorang dokter selalu dikelilingi oleh berbagai pertanyaan yang muncul secara alami di benaknya: *"Apakah terapi baru ini lebih baik untuk pasien saya?", "Apakah obesitas menjadi penyebab utama hipertensi pada kelompok usia muda?", "Apakah biomarker baru ini cukup akurat untuk mendeteksi sepsis dini?",* atau *"Bagaimana prognosis pasien ini di masa mendatang?"*. Namun, pertanyaan-pertanyaan klinis tersebut sering kali masih bersifat sangat luas, samar, dan belum siap untuk dianalisis secara statistik. Perangkat lunak statistik tercanggih sekalipun tidak akan pernah bisa memperbaiki atau memproses pertanyaan penelitian yang tidak terdefinisi dengan baik.

Di sinilah peran biostatistika dimulai bukan saat kita memasukkan angka ke dalam program komputer, melainkan jauh sebelumnya, yaitu ketika kita merumuskan masalah klinis yang abstrak menjadi pertanyaan statistik yang konkret dan terstruktur. Kerangka kerja **PICO** (*Population, Intervention, Comparison, Outcome*) atau **PICOT** (dengan tambahan *Time*) telah lama digunakan sebagai standar emas untuk menstrukturkan pertanyaan klinis. Untuk kepentingan analisis data, komponen-komponen ini harus dioperasionalkan secara ketat: mendefinisikan kriteria populasi secara spesifik, menentukan pembanding (comparator) yang adekuat, serta mengukur outcome secara objektif baik dari segi metrik, cara agregasi, maupun jangka waktu pengamatan.

Prinsip-prinsip ini sejalan dengan pedoman pelaporan riset kedokteran modern seperti **CONSORT 2025** dan **SPIRIT 2025** untuk uji klinis acak (randomized trials), serta **STROBE** dan **STROBE-nut** untuk studi observasional dan epidemiologi gizi. Pedoman tersebut menekankan pentingnya menspesifikasi seluruh variabel secara transparan sebelum analisis dilakukan untuk mencegah bias pelaporan (selective reporting atau multiplicity). Bab ini dirancang untuk melatih calon dokter dan residen agar mampu berpikir metodologis dan menyusun pertanyaan penelitian yang kokoh sebelum melangkah ke pemilihan uji statistik.

> **TUJUAN BESAR dan FOKUS BELAJAR**
> *Mahasiswa tidak ditargetkan untuk menghafal hubungan kaku antara 'PICO = uji statistik tertentu'. Target utama bab ini adalah melatih kemampuan Anda untuk menyusun pertanyaan penelitian yang cukup jelas dan operasional, sehingga keluarga analisis statistik yang relevan dapat dikenali secara logis tanpa menebak-nebak.*

2. Capaian Pembelajaran

Setelah menyelesaikan pembelajaran pada Bab 2 ini, mahasiswa kedokteran diharapkan mampu:

a. Membedakan secara jelas tingkatan masalah klinis, pertanyaan klinis, pertanyaan penelitian, hingga pertanyaan statistik.
b. Menggunakan kerangka PICO dan PICOT secara tepat sebagai alat formulasi pertanyaan klinis tanpa menganggapnya sebagai satu-satunya metode kaku.
c. Mengidentifikasi dan merinci secara spesifik komponen Population (P), Intervention/Exposure (I/E), Comparator (C), Outcome (O), dan Time point (T) pada kasus klinis nyata.
d. Membedakan fungsi outcome primer, outcome sekunder, outcome keselamatan (harm), dan analisis eksploratif, serta menjelaskan mengapa prioritas outcome harus ditetapkan sebelum melihat data hasil penelitian.
e. Mengklasifikasikan pertanyaan penelitian ke dalam Enam Keluarga Besar: Descriptive, Comparative, Association, Prediction, Diagnostic, atau Prognostic/Time-to-event.
f. Mengubah pertanyaan klinis yang kabur atau terlalu luas menjadi pertanyaan statistik yang dapat dianalisis dan diinterpretasikan secara sahih.
g. Menunjukkan kedisiplinan ilmiah untuk menolak menebak atau memilih uji statistik apabila definisi variabel klinis dan target estimasi belum didefinisikan secara jelas.

Kompetensi yang Belum Ditargetkan pada Bab Ini:

Pada bab ini, Anda belum dituntut untuk memilih atau menghitung uji statistik spesifik secara final. Kemampuan tersebut akan dibangun secara bertahap pada bab-bab berikutnya setelah Anda memahami karakteristik tipe data, struktur observasi, asumsi statistik, dan tujuan analisis mendalam.

3. Dari Masalah Klinis ke Pertanyaan Statistik

Sering kali kesalahan fatal dalam analisis data bermula jauh sebelum kita membuka perangkat lunak pengolah data. Kesalahan tersebut berupa lompatan langsung dari *masalah klinis* ke *uji statistik* tanpa melalui proses translasi yang sistematis. Alur berpikir statistik yang benar harus bergerak secara bertahap sebagai berikut:

Masalah Klinis → Pertanyaan Klinis → Pertanyaan Penelitian → Definisi Variabel dan Waktu → Target Estimasi → Pertanyaan Statistik → Metode Analisis

Untuk memahami transisi antar-tingkat pertanyaan ini, perhatikan contoh konkret penanganan pasien hipertensi pada Tabel 9 di bawah ini:

Tabel 9. Tingkat peryantaan dan contoh rumusan pertanyaan

Tingkat Pertanyaan	Contoh Rumusan Pertanyaan	Keterbatasan Bila Berhenti di Sini
Masalah Klinis	Pasien hipertensi di klinik tetap tidak terkontrol dengan baik.	Sangat abstrak. Belum ada variabel, perbandingan, atau parameter yang dapat diteliti secara ilmiah.
Pertanyaan Klinis	Apakah intervensi diet rendah natrium dapat membantu pasien?	Kata "membantu" terlalu luas. Apakah membantu berarti menurunkan tekanan darah, mencegah strokel, atau memperbaiki kualitas hidup?
Pertanyaan Penelitian	Pada pasien dewasa dengan hipertensi primer, apakah penerapan diet rendah natrium dibandingkan diet biasa dapat menurunkan tekanan darah sistolik setelah 12 minggu?	Sudah cukup spesifik untuk merancang penelitian, namun metrik analisis statistik yang tepat untuk data tersebut belum dirumuskan.
Pertanyaan Statistik	Berapakah perbedaan rata-rata perubahan tekanan darah sistolik (SBP) dari baseline hingga minggu ke-12 antara kelompok diet rendah natrium dengan kelompok diet biasa?	Sangat konkret. Menunjukkan target estimasi (estimand) yang jelas, memandu langsung pada pemilihan metode analisis yang tepat.

PRINSIP FILOSOFIS UTAMA
Statistik tidak pernah bertugas memilihkan pertanyaan penelitian untuk Anda. Peran utama metode statistik adalah menjawab pertanyaan yang telah Anda formulasikan dengan matang secara klinis. Oleh karena itu, kualitas hasil analisis statistik Anda akan selalu dibatasi oleh kejelasan pertanyaan dan kualitas desain penelitian yang mendahuluinya.

4. Kerangka Pico dan Picot

Kerangka kerja **PICO** pertama kali diperkenalkan oleh Richardson dkk. sebagai instrumen untuk membangun pertanyaan klinis yang terstruktur dengan baik (well-built clinical question). Dalam perkembangan dunia riset kedokteran, faktor waktu (*T - Time*) sering kali ditambahkan secara eksplisit sehingga berubah menjadi **PICOT**. Penambahan komponen waktu ini sangat krusial karena indikator keberhasilan klinis seperti *"mortalitas selama rawat inap", "mortalitas 30 hari", atau "mortalitas 5 tahun"* merupakan hasil (outcome) yang sangat berbeda secara metodologis dan statistik.

Tabel 10. Komponen PICOT

Komponen PICOT	Pertanyaan Praktis untuk Panduan	Contoh Penerapan Klinis
P Population	Pada kelompok pasien atau populasi target mana pertanyaan ini berlaku?	Pasien dewasa dengan diagnosis hipertensi primer derajat 1.
I/E Intervention / Exposure	Apa intervensi aktif (uji klinis) atau paparan biologis (observasional) utama yang diteliti?	Penerapan diet rendah natrium terstruktur (asupan < 2.000 mg/hari).
C Comparator	Dibandingkan dengan standar terapi atau kondisi apa?	Diet biasa/standar tanpa pembatasan ketat (usual diet).
O Outcome	Apa indikator klinis spesifik yang diukur untuk menilai efek?	Tekanan darah sistolik (SBP, dalam satuan mmHg).
T Time point	Kapan atau selama berapa lama pengamatan outcome dilakukan?	12 minggu setelah intervensi dimulai.

PICO/PICOT Bukan Algoritma Pemilihan Uji Statistik

Meskipun kerangka PICOT sangat membantu memperjelas arah penelitian, ia tidak secara otomatis mendikte satu uji statistik tertentu. Dua studi yang memiliki elemen PICOT yang identik dapat membutuhkan pendekatan analitik yang sangat berbeda tergantung pada metrik variabel,

struktur pengukuran, dan target estimasi klinis yang dipilih. Perhatikan contoh perbandingan pada Tabel 11 berikut:

Tabel 11. Pertanyaan klinis spesifik

Pertanyaan Klinis Spesifik	Target Analisis Utama	Konsekuensi Analitik dan Uji Statistik
Apakah nilai akhir SBP berbeda antar-kelompok pada minggu ke-12?	Nilai akhir SBP pada satu titik waktu.	Perbandingan outcome kontinu pada waktu tertentu (misalnya menggunakan independent t-test atau Welch t-test).
Apakah terdapat perbedaan perubahan skor SBP dari baseline ke minggu 12?	Skor perubahan (change score / selisih nilai).	Analisis perubahan nilai kontinu. Perlu memperhatikan nilai baseline dan desain model (misalnya menggunakan uji ANCOVA atau paired t-test).
Apakah proporsi pasien yang berhasil mencapai target SBP < 140 mmHg berbeda?	Proporsi/persentase responden yang mencapai target.	Outcome berubah dari kontinu menjadi biner (mencapai target vs tidak). Memerlukan analisis uji proporsionalitas seperti Chi-Square atau regresi logistik.
Bagaimana pola pergerakan (trajectory) nilai SBP pada bulan ke- 0, 1, 3, dan 6?	Pola tren perubahan sepanjang waktu.	Struktur pengukuran berulang (repeated measures). Memerlukan analisis longitudinal seperti model efek campuran (mixed-effects model) atau GEE.

5. Population, Intervention/Exposure, dan Comparator

a. Population: Lebih dari Sekadar Diagnosis

Populasi penelitian klinis (study population) tidak pernah hanya didefinisikan oleh nama penyakitnya saja. Sebagai contoh, merumuskan populasi sebagai *"pasien sepsis"* tidaklah adekuat. Kita harus merinci kriteria tersebut secara ketat: Sepsis di Unit Gawat Darurat (UGD) atau di ICU? Apakah pasien mengalami syok septik (septic shock)? Apakah populasi anak-anak atau dewasa? Apakah mereka telah menerima resusitasi cairan awal? Detail tersebut sangat memengaruhi reliabilitas hasil. Pedoman **STROBE** mengharuskan kriteria eligibilitas, latar tempat (setting), serta metode pemilihan subjek dijelaskan secara transparan agar pembaca dapat menilai tingkat generalisasi hasil ke pasien mereka sendiri.

b. Intervention dan Exposure: Jangan Dipertukarkan

Dalam statistika kedokteran, sangat penting untuk tidak mencampuradukkan peran konseptual setiap variabel dalam model analisis. Istilah-istilah seperti intervensi, paparan (exposure), prediktor, faktor pengganggu (confounder), dan modifikator efek memiliki arti yang berbeda, seperti yang dirangkum pada Tabel 12 di bawah ini:

Tabel 12. Istilah metodologis intervention and exposure

Istilah Metodologis	Konteks dan Peran dalam Analisis	Contoh Kasus Klinis
Intervention (Intervensi)	Tindakan atau terapi yang sengaja dialokasikan oleh peneliti kepada subjek dalam studi eksperimental (uji klinis).	Pemberian obat baru antihipertensi, regimen diet natrium khusus, atau keikutsertaan dalam program olahraga.
Exposure (Paparan)	Karakteristik, perilaku, atau kondisi biologis yang dimiliki atau dialami subjek secara alami (studi observasional).	Riwayat merokok selama 10 tahun, obesitas, asupan natrium tinggi harian, atau paparan polusi udara.
Predictor (Prediktor)	Variabel bebas yang digunakan dalam model statistik untuk memperkirakan atau memproyeksikan probabilitas terjadinya outcome.	Skor SOFA untuk memprediksi mortalitas ICU, kadar troponin untuk prognosis infark miokard.
Confounder (Faktor Perancu)	Variabel ketiga yang secara kausal berhubungan dengan paparan dan outcome, yang berpotensi mendistorsi estimasi efek murni.	Faktor usia pada hubungan antara obesitas dengan risiko hipertensi, karena usia memengaruhi berat badan sekaligus tekanan darah.
Effect Modifier (Modifikator Efek)	Variabel yang mengubah kekuatan atau arah hubungan antara paparan dengan outcome antar-strata subjek yang berbeda.	Sebuah terapi obat baru mungkin memiliki efektivitas sangat tinggi pada anak-anak, namun tidak menunjukkan manfaat signifikan pada lansia.

c. Comparator: Menentukan Makna Efek

Sebuah klaim klinis yang menyatakan bahwa *"Terapi A menurunkan mortalitas"* tidak memiliki arti ilmiah apa pun jika kita tidak mendefinisikan pembandingnya (comparator). Besarnya efek (effect size) suatu intervensi selalu bersifat relatif terhadap apa ia dibandingkan. Sebagai contoh, efektivitas terapi A akan tampak sangat berbeda jika dibandingkan dengan placebo, dibandingkan dengan terapi standar (usual care), atau dibandingkan dengan terapi aktif golongan lain (terapi B).

Tabel 13. Jenis comparator dan pertanyaan klinis yang dijawab

Jenis Comparator (Pembanding)	Pertanyaan Klinis yang Dijawab
Placebo (Plasebo)	Berapakah efek murni dari komponen aktif obat baru ini dibandingkan dengan kondisi tanpa pengobatan (tetapi mendapat sugesti setara)?
Usual Care / Standard Care	Berapakah nilai tambah atau manfaat ekstra yang diberikan oleh intervensi baru ini di atas standar pelayanan medis rutin saat ini?
Active Comparator (Terapi Aktif Lain)	Bagaimanakah efektivitas dan tingkat keamanan terapi baru (Strategi A) jika diadu langsung dengan terapi pilihan utama saat ini (Strategi B)?
No Exposure / Low Exposure	Bagaimanakah perbedaan profil status kesehatan atau kejadian penyakit pada kelompok subjek yang tidak terpapar (atau terpapar sangat rendah) dibandingkan kelompok terpapar tinggi?

PERINGATAN METODOLOGIS (PITFALL)
Mengubah atau memilih-milih kelompok pembanding (comparator) setelah hasil data penelitian terlihat demi mendapatkan nilai p-value yang paling signifikan (< 0,05) merupakan tindakan pelanggaran integritas ilmiah (data dredging). Comparator wajib ditetapkan sejak awal perencanaan riset berdasarkan kebutuhan klinis riil dan logika desain penelitian.

6. Outcome: Operasionalisasi dan Waktu

a. Mendefinisikan Outcome secara Operasional

Dalam draf penelitian, kata-kata deskriptif seperti *"pasien membaik", "terapi berhasil", atau "prognosis baik"* bukanlah definisi outcome yang valid untuk analisis statistik. Outcome harus dioperasionalkan secara terperinci. Pedoman uji klinis modern (**CONSORT 2025** dan **SPIRIT 2025**) mewajibkan setiap outcome (baik primer maupun sekunder) didefinisikan secara eksplisit menggunakan empat elemen operasional, seperti yang ditunjukkan pada Tabel 14:

Tabel 14. Elemen outcome yang operasional

Elemen Outcome yang Operasional	Penerapan dan Contoh pada Kasus Hipertensi
Variabel Pengukuran Spesifik	Variabel fisik atau biologis yang diukur, lengkap dengan satuannya. Contoh: Tekanan darah sistolik (Systolic Blood Pressure - SBP, dalam satuan mmHg).
Metrik Analisis Tingkat	Metode penghitungan nilai untuk setiap subjek

Elemen Outcome yang Operasional	Penerapan dan Contoh pada Kasus Hipertensi
Individu	penelitian. Contoh: Selisih/perubahan nilai SBP dari kondisi awal (baseline) hingga akhir penelitian (nilai minggu ke-12 dikurangi nilai minggu ke-0).
Cara Agregasi Tingkat Kelompok	Metode statistik yang digunakan untuk meringkas nilai dari seluruh subjek di kelompok tersebut. Contoh: Nilai rata-rata (mean) perubahan SBP per kelompok.
Titik Waktu Pengukuran (Time Point)	Jadwal waktu spesifik dilakukannya pengukuran hasil. Contoh: Tepat pada 12 minggu setelah proses randomisasi subjek.

b. Klasifikasi Outcome: Primer, Sekunder, Harm, dan Eksploratif

Dalam draf protokol penelitian, kita harus menetapkan prioritas analisis agar tidak terjebak dalam masalah multiplisitas (multiplicity bias / terlalu banyak menguji variabel hingga menemukan hasil yang signifikan secara tidak sengaja). Pembagian kategori outcome tersebut dirinci pada Tabel 15.

Tabel 15. Klasifikasi outcome: primer, sekunder, harm, dan eksploratif

Kategori Outcome	Fungsi dan Posisi dalam Desain	Contoh Penerapan Klinis
Primary Outcome (Outcome Primer)	Outcome tunggal yang dianggap paling penting secara klinis, digunakan sebagai dasar utama penarikan kesimpulan studi dan penentuan besar sampel (sample size).	Rata-rata perubahan tekanan darah sistolik (SBP) pada minggu ke-12.
Secondary Outcome (Outcome Sekunder)	Hasil tambahan yang direncanakan sejak awal untuk memperkuat bukti, memberikan kedalaman klinis, atau menjawab pertanyaan sekunder.	Rata-rata perubahan tekanan darah diastolik (DBP), proporsi pasien mencapai target SBP < 140 mmHg, dan skor kualitas hidup pasien.
Harm / Safety Outcome (Efek Samping)	Outcome wajib untuk mendokumentasikan konsekuensi negatif yang tidak diinginkan, komplikasi, atau efek samping demi keselamatan pasien.	Kejadian hipotensi simptomatik, gangguan elektrolit darah, atau gagal ginjal akut selama masa terapi harian.
Exploratory Outcome (Outcome Eksploratif)	Digunakan untuk mengeksplorasi hipotesis baru atau analisis subkelompok. Hasilnya bernilai akademis tetapi tidak boleh disimpulkan sebagai bukti konklusif.	Evaluasi perubahan biomarker inflamasi baru (seperti IL-6) setelah pemberian intervensi.

c. Pengaruh Faktor Waktu (Time): Outcome Sama, Makna Berbeda

Bahkan jika kita mengukur indikator biologis yang persis sama, perbedaan titik waktu (time point) pengamatan akan mengubah total pertanyaan klinis yang dijawab. Hal ini dirangkum pada Tabel 16 di bawah ini:

Tabel 16. Pengaruh faktor waktu (time): outcome sama, makna berbeda

Indikator Klinis (Outcome)	Titik Waktu Pengamatan	Pertanyaan Klinis Spesifik yang Dijawab
Mortalitas Pasien Sepsis	Selama Rawat Inap (In-hospital)	Apakah pasien berhasil bertahan hidup selama episode akut perawatan kritis di rumah sakit?
Mortalitas Pasien Sepsis	30 Hari Pasca-Masuk Rumah Sakit	Apakah intervensi memberikan dampak proteksi jangka pendek terhadap risiko kematian?
Mortalitas Pasien Sepsis	1 Tahun Pasca-Pemulangan	Bagaimanakah tingkat kelangsungan hidup (survival) jangka panjang pasien setelah fase sepsis terlewati?
Kadar HbA1c Pasien Diabetes	3 Bulan Pasca-Intervensi Diet	Apakah terjadi perbaikan kontrol glikemik atau efisiensi metabolik awal?
Kadar HbA1c Pasien Diabetes	12 Bulan Pasca-Intervensi Diet	Apakah efek positif dari intervensi diet tersebut bersifat jangka panjang dan berkelanjutan (sustained effect)?

d. Dari Outcome ke Estimand (Preview Konseptual)

Dalam metodologi uji klinis modern, panduan **ICH E9(R1)** memperkenalkan konsep *estimand*. Estimand adalah definisi yang sangat presisi tentang efek terapi yang ingin kita ketahui, yang mencakup bagaimana kita memperlakukan kejadian-kejadian pasca-acak (intercurrent events) yang mengacaukan pembacaan hasil (seperti penghentian terapi akibat efek samping atau penggunaan obat penyelamat/rescue medication). Sebagai contoh: Apakah kita ingin mengestimasi efek terapi seandainya semua pasien patuh menggunakan obat? Ataukah kita ingin mengestimasi efek riil di dunia nyata di mana sebagian pasien pasti akan menghentikan pengobatan? Pertanyaan yang berbeda ini membutuhkan metode pengumpulan data dan analisis statistik yang berbeda pula.

7. Enam Keluarga Pertanyaan Penelitian

Setiap pertanyaan riset kedokteran yang diajukan oleh dokter spesialis atau akademisi klinis pasti masuk ke dalam salah satu dari **Enam Keluarga Besar Pertanyaan Penelitian**. Pemahaman konseptual ini sangat penting agar kita tidak salah memilih model analisis data, seperti yang dirinci pada Tabel 17:

Tabel 17. Enam keluarga pertanyaan penelitian

Keluarga Pertanyaan	Pertanyaan Inti yang Dijawab	Contoh Penerapan Klinis Nyata
Descriptive (Deskriptif)	Bagaimanakah profil, gambaran, atau distribusi karakteristik klinis dari kelompok pasien saya?	Berapakah nilai median usia, persentase status malnutrisi, dan profil komorbiditas pasien gagal jantung di rumah sakit ini?
Comparative (Perbandingan)	Apakah nilai outcome klinis tertentu menunjukkan perbedaan nyata di antara dua atau lebih kelompok/intervensi?	Apakah terdapat perbedaan rata-rata kadar HbA1c antara pasien yang mendapat Terapi A dengan yang mendapat Terapi B setelah 6 bulan?
Association (Asosiasi/Hubungan)	Apakah terdapat hubungan atau pola pergerakan bersama antara Faktor X (paparan) dengan Faktor Y (outcome)?	Apakah tingkat asupan serat harian berhubungan dengan kadar HbA1c pada pasien diabetes tipe 2?
Prediction (Prediksi)	Seberapa akuratkah kombinasi dari beberapa variabel klinis dalam memperkirakan kejadian outcome pada pasien?	Seberapa baik kombinasi variabel usia, kadar laktat darah, dan skor SOFA dapat memprediksi probabilitas mortalitas pasien di ICU?
Diagnostic (Akurasi Diagnostik)	Seberapa andalkah sebuah metode pemeriksaan baru dalam membedakan antara orang yang sakit dengan yang sehat?	Seberapa baik akurasi pemeriksaan kadar Biomarker X dalam mendeteksi kejadian sepsis akut dibandingkan dengan kultur darah standar?
Prognostic / Time-to-event	Berapakah probabilitas terjadinya suatu kejadian klinis seiring berjalannya waktu, dan kapan kejadian itu umumnya terjadi?	Apakah pemberian terapi adjuvan baru dikaitkan dengan durasi kelangsungan hidup (overall survival) yang lebih panjang pada pasien kanker payudara?

Untuk menekankan bahwa uji statistik tidak ditentukan oleh *"nama variabel"*, perhatikan bagaimana satu variabel yang sama (misalnya kadar HbA1c) dapat melahirkan uji statistik yang sangat bervariasi tergantung pada keluarga pertanyaan penelitian yang dipilih (Tabel 18):

Tabel 18. Pertanyaan klinis dokterm pertanyaan statistik konseptual, keluarga pertanyaan penelitian

Pertanyaan Klinis Dokter	Pertanyaan Statistik Konseptual	Keluarga Pertanyaan Penelitian
Bagaimanakah profil kontrol glikemik pasien diabetes melitus di poliklinik kami?	Berapakah nilai rata-rata, median, standar deviasi, dan rentang nilai HbA1c di populasi sampel?	Descriptive
Apakah Terapi A memberikan penurunan kadar HbA1c yang lebih baik dibanding Terapi B setelah 6 bulan?	Apakah terdapat perbedaan rata-rata nilai HbA1c antara Kelompok A dan Kelompok B pada bulan ke-6?	Comparative
Apakah kebiasaan asupan serat kasar harian berhubungan dengan nilai HbA1c pasien?	Bagaimanakah kekuatan dan arah hubungan (koefisien korelasi) asupan serat terhadap kadar HbA1c?	Association
Dapatkah nilai indeks massa tubuh (BMI), usia, dan durasi diabetes memprediksi pasien yang akan mengalami kegagalan terapi (HbA1c > 9%)?	Berapakah akurasi (sensitivitas, spesifisitas, area di bawah kurva ROC) model regresi multivariabel dalam mendeteksi status HbA1c > 9%?	Prediction
Bagaimanakah tren perubahan nilai HbA1c pasien pada bulan ke-0, 3, 6, dan 12?	Bagaimanakah bentuk lintasan (trajectory) dan interaksi waktu terhadap nilai HbA1c pasien?	Longitudinal Comparative / Modeling

8. Studi Kasus Lintas Bidang

a. Kasus Kedokteran: Evaluasi Obat Antihipertensi

Perhatikan evolusi perumusan pertanyaan dari yang awalnya sangat kabur hingga menjadi pertanyaan statistik yang presisi pada Tabel 19:

Tabel 19. Tahapan versi pertanyaan

Tahapan Versi Pertanyaan	Keterbatasan dan Catatan Perbaikan Metodologis
"Obat hipertensi mana yang lebih baik?"	Istilah "lebih baik" tidak didefinisikan secara ilmiah. Siapa populasinya, apa indikator keberhasilannya, dan berapa lama durasi penilaiannya masih menjadi misteri.
"Apakah Obat A dapat menurunkan tekanan darah?"	Tidak ada kelompok pembanding (comparator) yang jelas. Kita tidak tahu apakah penurunan disebabkan efek obat atau hanya variasi alami biologis (regression to the mean).
"Pada pasien dewasa dengan	Sangat operasional dan siap diteliti. Komponen

hipertensi primer tidak terkontrol, apakah pemberian Obat A dibandingkan dengan Obat B memberikan penurunan rata-rata tekanan darah sistolik (SBP) yang lebih besar pada minggu ke-12?"	PICOT (P, I, C, O, T) telah lengkap dan mengarahkan langsung pada target analisis statistik: menghitung perbedaan rata-rata perubahan SBP antar-kelompok.

b. Kasus Gizi: Asupan Natrium dan Tekanan Darah

Dalam ilmu gizi klinik, perumusan pertanyaan riset harus sangat hati-hati dalam memilih kosakata kesimpulan kausal vs asosiatif, seperti yang ditunjukkan pada Tabel 20.

Tabel 20. Rumusan pertanyaan penelitian gizi

Rumusan Pertanyaan Penelitian Gizi	Konsekuensi Desain dan Batasan Klaim Statistik
"Apakah asupan natrium yang tinggi menyebabkan hipertensi?"	Menuntut pembuktian kausalitas (causal effect). Klaim ini tidak boleh diajukan atau disimpulkan jika desain penelitian yang digunakan hanya bersifat potong lintang (cross-sectional).
"Apakah asupan natrium harian berasosiasi secara linear dengan nilai SBP?"	Sesuai untuk studi observasional (seperti studi kohort atau cross-sectional). Target statistik adalah menghitung kekuatan hubungan (correlation coefficient atau koefisien regresi) dengan tetap mengontrol faktor pengganggu (confounders).
"Apakah intervensi pembatasan natrium diet terstruktur menurunkan tekanan darah dibandingkan diet biasa?"	Sesuai untuk studi eksperimental / uji klinis acak (RCT). Metode analisis difokuskan untuk mengukur perbedaan efek intervensi langsung.

Pedoman epidemiologi gizi (**STROBE-nut**) secara ketat mengingatkan peneliti untuk mendefinisikan eksposur zat gizi secara detail (apakah berupa bahan makanan, zat gizi makro/mikro, atau pola diet umum) serta menjelaskan alat ukur asupan (dietary assessment) yang dipakai. Kesalahan pengukuran asupan (measurement error) gizi yang sangat sering terjadi dapat melemahkan atau mengacaukan kekuatan asosiasi statistik yang dilaporkan.

c. Kasus Diagnostik: Menilai Kegunaan Biomarker X

Kalimat *"Apakah biomarker X berguna untuk kasus sepsis?"* masih terlalu kabur. Kata "berguna" dapat dioperasionalkan menjadi empat pertanyaan yang sangat berbeda berikut ini (Tabel 21):

Tabel 21. Makna kosakata 'Berguna'

Makna Kosakata 'Berguna'	Pertanyaan Spesifik yang Lebih Operasional	Uji / Analisis Statistik Relevan
Association (Asosiasi)	Apakah nilai rata-rata konsentrasi Biomarker X berbeda nyata antara pasien sepsis dengan pasien non-sepsis?	Uji perbandingan mean (independent t-test atau Mann-Whitney).
Diagnostic Accuracy (Akurasi)	Seberapa andalkah Biomarker X dalam mengklasifikasikan status pasien sepsis vs non-sepsis?	Analisis Kurva ROC, nilai AUC, penentuan nilai sensitivitas, spesifisitas, dan Likelihood Ratio (LR).
Clinical Utility (Manfaat Klinis)	Apakah keputusan klinis berbasis biomarker X memperbaiki luaran kesembuhan pasien dibanding terapi tanpa biomarker?	Uji perbandingan superioritas klinis menggunakan desain uji klinis acak (RCT).
Prognosis (Prognosis)	Apakah kadar awal Biomarker X dapat memprediksi risiko mortalitas 30 hari pada pasien sepsis?	Analisis kelangsungan hidup (survival analysis) seperti kurva Kaplan-Meier atau model regresi Cox Proportional Hazards.

9. Association, Prediction, dan Causation: Jangan Disamakan

Salah satu jebakan terbesar bagi dokter saat membaca atau menulis artikel ilmiah adalah mencampuradukkan konsep **asosiasi (association), prediksi (prediction),** dan **efek kausal (causal effect)**. Ketiganya menjawab pertanyaan yang berbeda secara fundamental, seperti yang dirangkum pada Tabel 22 di bawah ini.

Tabel 22. Asosiasi, prediksi, dan efek kausal

Konsep Metodologis	Pertanyaan Inti yang Dijawab	Fokus Utama Analisis Data
Association / Correlation	Apakah variabel X dan variabel Y cenderung bergerak bersama-sama di dalam populasi sampel?	Mengukur kekuatan, arah, dan signifikansi hubungan linear/monoton (seperti uji korelasi Pearson atau Spearman).
Prediction	Seberapa baik kombinasi variabel independen dapat memperkirakan nilai outcome klinis pada tingkat individu?	Mengukur akurasi model, kalibrasi, dan diskriminasi (misalnya nilai R-squared, kurva ROC/AUC, atau sensitivitas/spesifisitas).
Causal Effect	Apakah perubahan paksa (intervensi) pada variabel X secara langsung mengakibatkan perubahan	Memerlukan kontrol ketat terhadap bias dan perancu melalui randomisasi atau asumsi kausalitas yang kuat

	pada variabel Y?	(seperti diagram kausal DAG).

Sebagai contoh nyata: Sebuah biomarker kadar CRP darah dapat menjadi *prediktor* yang sangat baik untuk mendeteksi mortalitas pasien di rumah sakit, tanpa harus menjadi *penyebab (causal)* langsung dari kematian tersebut (CRP hanya menandakan adanya inflamasi berat di tubuh). Sebaliknya, mengintervensi atau menurunkan kadar CRP secara paksa belum tentu menurunkan risiko mortalitas pasien.

CATATAN PENTING UNTUK RESIDEN

Penerapan model regresi multivariabel yang 'disesuaikan' (adjusted) dalam studi observasional tidak secara otomatis mengubah hubungan asosiasi menjadi hubungan sebab-akibat (kausal). Interpretasi kausalitas yang valid tetap bersandar pada kualitas desain penelitian, kontrol bias, ketepatan pengukuran, dan ketepatan asumsi model.

10. Pertanyaan yang Terlalu Luas atau Terlalu Sempit

Sebuah pertanyaan penelitian harus dirumuskan dengan tingkat spesifikasi yang seimbang. Pertanyaan yang terlalu luas (too broad) akan membingungkan langkah operasional, sedangkan pertanyaan yang terlalu sempit (too narrow) sering kali lahir dari proses pencarian hasil signifikan pasca-data dikumpulkan (post-hoc subgroup analysis). Perhatikan perbandingan pada Tabel 23:

Tabel 23. Pertanyaan yang terlalu luas atau terlalu sempit

Pertanyaan Kabur (Too Broad)	Masalah Utama	Versi Perbaikan (Terstruktur)
"Apakah vitamin D baik untuk kesehatan?"	Variabel populasi, dosis, pembanding (comparator), indikator keberhasilan (outcome), dan waktu pengamatan sama sekali tidak jelas.	Pada lansia dengan defisiensi vitamin D, apakah pemberian suplementasi Vitamin D3 dosis 2.000 IU/hari dibandingkan plasebo dapat mengurangi risiko kejadian jatuh (falls) selama 12 bulan pengamatan?
"Apakah kondisi obesitas berbahaya?"	Outcome klinis yang dinilai dan horizon waktu pengamatan tidak dispesifikasikan secara jelas.	Pada subjek dewasa tanpa riwayat penyakit kardiovaskular (CVD) awal, apakah kondisi obesitas (BMI >= 27,5 kg/m2) pada baseline berasosiasi dengan peningkatan insiden kejadian CVD mayor

		(MACE) selama 10 tahun masa follow-up?
"Apakah manajemen terapi ICU kami efektif?"	Intervensi yang diteliti dan indikator luaran yang dinilai terlalu luas dan umum.	Pada pasien dewasa dengan syok septik (septic shock) di ICU, apakah penerapan strategi cairan restriktif dibandingkan dengan strategi cairan liberal dapat menurunkan angka mortalitas 28 hari?

11. Kesalahan Konseptual yang Sering Terjadi

Dalam menyusun pertanyaan klinis dan statistik, terdapat lima kesalahan konseptual klasik yang wajib dihindari oleh mahasiswa kedokteran (Tabel 24).

Tabel 24. Lima kesalahan konseptual klasik yang wajib dihindari oleh mahasiswa kedokteran

Kesalahan Konseptual Klasik	Koreksi dan Solusi Konseptual
"Saya sudah memiliki data di excel, uji statistik apa yang harus saya gunakan?"	Uji statistik tidak ditentukan oleh isi excel. Analisis wajib dimulai dengan mendefinisikan pertanyaan klinis, menentukan variabel, menetapkan target estimasi (estimand), barulah memilih metode.
"Outcome penelitian saya adalah keberhasilan terapi."	Kata "keberhasilan terapi" tidak dapat dipahami komputer. Operasionalkan indikatornya (misal: penurunan SBP dalam mmHg), skala pengukuran, batas ambang (threshold), dan waktu pengukurannya.
"Seluruh faktor risiko (X) dalam penelitian ini saya sebut sebagai variabel independen."	Sangat keliru. Anda harus membedakan secara konseptual peran masing-masing variabel: mana yang merupakan paparan utama (exposure), prediktor, perancu (confounder), perantara (mediator), atau modifikator efek (modifier).
"Karena variabel X memiliki korelasi signifikan dengan Y, maka X adalah faktor risiko penyebab Y."	Korelasi (association) tidak sama dengan hubungan sebab-akibat (causation). Diperlukan evaluasi terhadap urutan waktu kejadian (temporal ordering) dan kontrol bias sebelum mengklaim faktor risiko.
"Saya akan melakukan uji korelasi pada semua kolom data, mana yang nilai $p < 0,05$ akan saya jadikan tujuan utama penelitian saya."	Tujuan utama dan outcome primer wajib diprespesifikasi secara transparan pada fase perencanaan (registrasi protokol). Membentuk tujuan riset berdasarkan data yang sudah terlihat (p-hacking) dapat merusak validitas kesimpulan.

12. Checklist: Before You Choose A Test

Sebelum Anda memutuskan untuk memilih metode analisis atau berkonsultasi dengan biostatistisian, gunakan delapan pertanyaan kritis berikut ini sebagai panduan cepat (Tabel 25).

Tabel 25. Checklist pertanyaan reflektif

Checklist Pertanyaan Reflektif	Deskripsi Detail yang Harus Anda Jawab
Who? (Siapa populasi targetnya?)	Sebutkan kriteria eligibilitas subjek secara spesifik (usia, derajat keparahan penyakit, lokasi perawatan, komorbiditas).
What? (Apa intervensi atau eksposurnya?)	Jelaskan dosis, intensitas, cara pemberian intervensi (uji klinis) atau detail paparan klinis (studi observasional).
Compared with? (Dibandingkan dengan apa?)	Tetapkan kelompok pembanding yang relevan secara klinis (plasebo, terapi standar medis saat ini, kelompok tidak terpapar).
Outcome? (Apa outcome klinisnya?)	Operasionalkan variabel hasil (nama variabel spesifik, metrik analisis individu, cara agregasi kelompok).
When? (Kapan outcome dinilai?)	Tentukan titik waktu pengamatan yang presisi (misalnya jam ke-24, hari ke-30, atau tahun ke-5 harian).
6Purpose? (Apa tujuan utamanya?)	Klasifikasikan tujuan riset Anda: deskriptif, perbandingan kelompok, mencari kekuatan hubungan, akurasi alat diagnosis, prediksi, atau prognosis klinis?
Target? (Apa target estimasi statistik?)	Apakah Anda ingin mengetahui perbedaan rata-rata (mean difference), perbedaan risiko (risk difference), rasio risiko (risk ratio/odds ratio), atau probabilitas kelangsungan hidup?
What is missing? (Informasi apa yang belum lengkap?)	Tinjau kembali apakah ada variabel penting atau data yang hilang yang dapat mengacaukan analisis atau interpretasi klinis Anda.

13. Rangkuman

a. Pertanyaan analisis statistik yang tajam hanya bisa lahir dari proses penyempurnaan pertanyaan klinis yang matang.
b. Kerangka PICO dan PICOT merupakan instrumen terstruktur untuk memandu penulisan draf protokol riset, bukan sebuah algoritma otomatis untuk memilih uji statistik.
c. Outcome klinis wajib dioperasionalkan secara ketat, mencakup variabel spesifik, metrik penilaian tingkat individu, metode agregasi kelompok, serta jangka waktu pengamatan.

d. Istilah intervensi, eksposur (exposure), prediktor, perancu (confounder), dan modifikator efek tidak boleh dipertukarkan karena memiliki fungsi konseptual yang sangat berbeda.
e. Kelompok pembanding (comparator) adalah kunci penentu makna klinis dari besarnya ukuran efek (effect size) yang dilaporkan.
f. Variabel yang sama (seperti kadar gula darah atau tekanan darah) dapat membutuhkan metode analisis statistik yang sepenuhnya berbeda jika struktur data atau pertanyaan risetnya berubah.
g. Hubungan asosiasi, model prediksi diagnostik, model prognosis, dan pembuktian efek kausalitas memiliki batasan interpretasi ilmiah yang berbeda dan tidak boleh disamakan.
h. Outcome primer dan analisis utama wajib didefinisikan sebelumnya secara transparan (pre-specified) pada registrasi protokol sebelum data mulai dikumpulkan.
i. Pemilihan uji statistik dilarang keras dilakukan hanya karena iseng mencoba-coba seluruh uji pada komputer hingga menemukan hasil p-value yang signifikan.
j. Tujuan akhir biostatistika kedokteran bukanlah untuk menghasilkan angka p-value semata, melainkan untuk memahami secara logis apa yang layak dan tidak layak disimpulkan dari angka penelitian demi kepentingan perawatan pasien.

KALIMAT PENUTUP BAB 2:
"Analisis statistik tidak pernah dimulai saat Anda membuka perangkat lunak di komputer. Analisis statistik dimulai saat Anda menetapkan pertanyaan penelitian apa yang layak dijawab, pada populasi pasien yang mana, dengan intervensi apa, dibandingkan dengan standar pelayanan medis apa, dan pada waktu kapan."

14. Latihan dan Evaluasi

Latihan Formatif Pilihan Ganda

1) Pertanyaan penelitian klinis yang berbunyi: "Apakah pemberian Obat A efektif?" dianggap paling bermasalah dari sudut pandang metodologi statistik karena:
   A. Tidak menyertakan hasil perhitungan nilai p-value.
   B. Tidak menggunakan perangkat lunak analisis data mutakhir.
   C. Komponen populasi target, comparator (pembanding), definisi outcome yang operasional, dan waktu penilaian tidak dijelaskan secara spesifik.
   D. Data belum dilakukan pemeriksaan uji asumsi normalitas.

Jawaban: C. Pertanyaan tersebut masih terlalu luas dan kabur (tidak operasional). Agar dapat diteliti, harus diurai berdasarkan komponen PICOT secara spesifik.

2) Di antara pilihan berikut, manakah formulasi yang paling tepat disebut sebagai outcome klinis yang operasional?
   A. Kondisi klinis pasien sepsis membaik secara signifikan
   B. Pemberian terapi kombinasi antibiotik dinyatakan berhasil
   C. Perubahan nilai rata-rata tekanan darah sistolik (SBP) dari baseline hingga minggu ke-12 pengamatan
   D. Prognosis klinis jangka panjang pasien dinyatakan baik

Jawaban: C. Formulasi ini memiliki variabel spesifik (SBP), metrik individu (perubahan dari baseline), agregasi kelompok (rata-rata), dan titik waktu (minggu ke-12).

3) Pertanyaan penelitian yang berbunyi: "Apakah skor SOFA pada hari pertama perawatan dapat digunakan untuk memprediksi risiko mortalitas pasien selama di ICU?" paling tepat diklasifikasikan ke dalam keluarga:
   A. Descriptive
   B. Prediction / Prognostic
   C. Diagnostic Accuracy
   D. Comparative

Jawaban: B. Pertanyaan ini bertujuan untuk memproyeksikan kejadian di masa depan (mortalitas) berdasarkan nilai prediktor awal (skor SOFA).

4) Dalam konteks analisis data penelitian kedokteran observasional, hubungan istilah eksposur (exposure) dengan faktor perancu (confounder) adalah:
   A. Kedua istilah tersebut selalu memiliki arti yang sama secara konseptual.
   B. Tidak perlu dibedakan sama sekali dalam model statistik regresi.
   C. Memiliki fungsi konseptual yang sangat berbeda di mana confounder berpotensi merusak estimasi murni efek dari exposure.
   D. Peran keduanya hanya ditentukan berdasarkan nilai signifikansi p-value dari uji bivariat.

Jawaban: C. Exposure adalah variabel bebas utama yang diteliti hubungannya, sedangkan confounder adalah faktor luar yang dapat merusak/mengacaukan estimasi hubungan murni tersebut.

5) Melakukan perubahan atau penggantian parameter outcome primer riset uji klinis setelah melihat data hasil penelitian menunjukkan signifikansi statistik yang tinggi adalah tindakan yang:
   A. Sangat ideal demi efisiensi hasil laporan publikasi
   B. Dapat menimbulkan bias metodologis yang berat serta menurunkan transparansi dan integritas laporan
   C. Selalu dilarang secara mutlak oleh algoritma matematika komputer
   D. Tidak memberikan pengaruh apa pun terhadap validitas internal penelitian

Jawaban: B. Tindakan ini merupakan pelanggaran metodologi berat (p-hacking / selective outcome reporting) yang merusak kepercayaan terhadap hasil penelitian.

Soal Diskusi dan Refleksi Klinis:
"Manakah yang lebih berbahaya dalam praktik kedokteran nyata: memilih metode uji statistik yang rumit tanpa memahami pertanyaan klinis yang ingin dijawab, atau memilih pertanyaan penelitian yang salah namun dianalisis dengan metode statistik yang sangat presisi?"

Tujuan diskusi:
Mendorong mahasiswa kedokteran untuk menyadari bahwa metodologi riset kedokteran yang andal selalu mengutamakan kejelasan pertanyaan klinis dan kualitas desain penelitian di atas kehebatan rumus matematika statistik.

BAB 3
MEMAHAMI DATA KLINIS

Representasi Angka, Struktur Observasi, dan Tata Kelola Kualitas Data

1. Pendahuluan

Dokter berhadapan dengan beragam bentuk data klinis setiap hari: tekanan darah, kadar kreatinin, stadium kanker, status kelangsungan hidup (hidup-mati), jumlah episode eksaserbasi asma, skor nyeri, waktu sampai terjadinya event, hingga hasil pengukuran berulang pada pasien yang sama. Meskipun seluruh informasi ini dapat disimpan sebagai "angka" di dalam lembar kerja (spreadsheet), makna statistik dan interpretasi klinis setiap variabel tidaklah sama.

Bab ini didesain khusus untuk membangun kemampuan berpikir kritis mahasiswa kedokteran dalam mengenali jenis variabel dan struktur observasi sebelum melompat pada pemilihan metode analisis statistik. Penekanan utama diletakkan pada hubungan intuitif antara makna klinis variabel, cara data dikumpulkan, target estimasi, dan asumsi dasar analisis statistik. Seperti yang ditekankan oleh Altman dan Bland (2009), analisis data kontinu tidak semata ditentukan oleh keputusan mekanis seperti label "normal" atau "tidak normal", melainkan oleh tujuan analisis yang mendasarinya serta asumsi dari metode yang digunakan.

> **PRINSIP INTI**
> *Nama variabel tidak pernah cukup untuk menentukan uji statistik. Sebagai contoh, tekanan darah dapat bertindak sebagai outcome kontinu, perubahan berpasangan (paired change), pengukuran berulang (repeated measure), prediktor/kovariat, atau bahkan variabel kategorik (misal: hipertensi vs normal) — tergantung pada pertanyaan klinis Anda dan bagaimana data tersebut digunakan.*

Kesalahan pemilihan analisis sering kali bermula jauh sebelum perangkat lunak statistik dibuka. Beberapa kesalahan umum yang sering dijumpai meliputi: memperlakukan skor ordinal sebagai data kontinu tanpa pembenaran metodologis, mengabaikan korelasi pada pengukuran berulang, menghapus nilai hilang (missing data) tanpa menilai mekanisme hilangnya data, serta membuang pencilan (outlier) semata karena nilainya tampak ekstrem. Kesalahan-kesalahan konseptual ini secara langsung akan

mendistorsi estimasi, mempersempit presisi, dan merusak validitas inferensi klinis.

2. Capaian Pembelajaran

Setelah mempelajari bab ini dengan saksama, mahasiswa diharapkan mampu:

- Mengklasifikasikan variabel klinis sebagai nominal, ordinal, kontinu, count, binary, atau time-to-event berdasarkan makna biologis dan skala pengukurannya secara tepat.
- Membedakan data independen, paired (berpasangan), repeated measures, clustered (berkelompok), dan longitudinal secara konseptual.
- Memilih ringkasan deskriptif (seperti mean, median, standard deviation [SD], atau interquartile range [IQR]) yang sesuai dengan bentuk distribusi data dan tujuan pelaporan klinis.
- Menjelaskan secara konseptual mengapa uji formal normalitas (seperti Shapiro-Wilk atau Kolmogorov-Smirnov) tidak boleh dijadikan satu-satunya pintu gerbang penentu metode analisis.
- Mengidentifikasi pencilan (outlier) serta menjelaskan secara metodologis mengapa nilai ekstrem tersebut tidak boleh otomatis dihapus dari dataset.
- Mengenali missing data sebagai masalah metodologis yang memerlukan penilaian kuantitas, pola, mekanisme penyebab (MCAR, MAR, MNAR), dan asumsi penanganannya.
- Menghubungkan jenis outcome dan struktur data dengan keluarga metode statistik yang akan dipelajari pada bab-bab berikutnya.

3. Pertanyaan Besar dan Alur Berpikir

> **PERTANYAAN PEMANTIK**
> *Jika semua variabel klinis sudah tersimpan rapi berupa angka di spreadsheet, mengapa kita masih perlu memahami jenis dan struktur data secara konseptual sebelum melakukan analisis?*

Jawaban klinisnya adalah: karena angka hanyalah representasi matematis dari realitas biologis pasien yang sangat kompleks. Angka "1" dan "2" pada variabel jenis kelamin (misal Laki-laki dan Perempuan) tidak memiliki makna jarak matematis, sedangkan angka "1" dan "2" pada variabel kadar kreatinin mewakili perbedaan klinis yang masif. Informasi statistik yang valid ditentukan oleh makna klinis variabel, skala pengukuran,

hubungan antar-observasi, bentuk distribusi, proses pengukuran, dan pola data yang tidak lengkap.

Alur Berpikir

Alur berpikir yang dianjurkan dalam memahami data klinis bergerak secara runtut melalui tahapan berikut:

- Makna Klinis: Apa fenomena biologis yang sedang diukur?
- Jenis Variabel: Apakah berupa kategori, urutan, nilai riil, jumlah kejadian, atau waktu?
- Struktur Observasi: Bagaimana hubungan antarsubjek (independen, berpasangan, longitudinal, atau berkelompok)?
- Distribusi Data: Bagaimana bentuk sebaran data klinis tersebut?
- Kualitas Data: Apakah terdapat pencilan (outlier) atau nilai yang hilang (missing data)?
- Ringkasan Deskriptif: Bagaimana cara terbaik merangkum informasi ini untuk dokter lain?
- Keluarga Analisis: Model statistik mana yang menghormati struktur data dan menjawab pertanyaan klinis?

4. Angka Bukan Satu Jenis Data

Dalam biostatistika kedokteran, kita membagi data ke dalam enam keluarga utama berdasarkan sifat matematis dan makna klinisnya:

a. Nominal

Variabel nominal mengekspresikan kategori tanpa adanya urutan intrinsik di antara kategori tersebut. Contoh klasiknya adalah golongan darah (A, B, AB, O), jenis mikroorganisme penyebab infeksi, atau status merokok (Ya/Tidak). Variabel nominal diringkas menggunakan frekuensi dan proporsi (persentase).

b. Ordinal

Variabel ordinal memiliki tingkatan atau urutan yang logis, namun jarak (interval) antar-kategori tidak dapat diasumsikan setara secara matematis. Contoh klinisnya meliputi stadium kanker (I, II, III, IV), derajat edema, atau kategori rasa nyeri (Ringan, Sedang, Berat).

c. Kontinu

Variabel kontinu secara konseptual memiliki nilai rill tak terbatas dalam suatu rentang pengukuran. Contoh medis meliputi tekanan darah sistolik (mmHg), kadar gula darah (mg/dL), berat badan (kg), atau volume ekspirasi paksa (FEV1). Sifat kontinunya memungkinkan kita mengukur perbedaan yang sangat halus antar-pasien.

d. Count (Hitungan)

Count adalah jumlah kejadian diskrit yang terjadi dalam ruang atau waktu tertentu. Contohnya adalah frekuensi eksaserbasi asma dalam satu tahun, jumlah rawat inap, atau jumlah episode hipoglikemia. Variabel count bernilai bulat positif (0, 1, 2, ...) dan memiliki sifat distribusi khusus (seperti Poisson atau Negative Binomial).

e. Binary (Biner)

Outcome biner adalah variabel kategorik yang hanya memiliki dua kemungkinan keadaan yang saling lepas dalam definisi analitik tertentu. Contoh utamanya adalah status kelangsungan hidup (hidup/meninggal), kejadian infeksi luka operasi (infeksi/tidak), atau status remisi penyakit (remisi/tidak).

f. Time-to-Event

Pada variabel time-to-event, informasi utama bukan sekadar apakah suatu kejadian terjadi, melainkan kapan kejadian tersebut terjadi dan berapa lama waktu yang dibutuhkan subjek untuk mencapainya. Ini melibatkan penanganan subjek yang tidak mengalami event hingga akhir studi *(censoring)*. Variabel ini merupakan dasar dari Analisis Survival.

Tabel 26. Menyajikan contoh visual bagaimana variabel klinis direpresentasikan secara statistik beserta pertanyaan kritis yang wajib diajukan oleh seorang dokter sebelum memulai analisis data

Variabel Klinis	Representasi Statistik	Pertanyaan Kritis yang Wajib Diajukan
Tekanan darah sistolik	Kontinu / Rill	Apakah nilai tunggal, perubahan dari baseline, atau pengukuran berulang sepanjang waktu?
Stadium kanker	Ordinal / Tingkatan	Apakah urutan kategorinya telah didefinisikan secara klinis dengan konsisten?
Mortalitas 30 hari	Binary / Biner	Apakah pembagi (denominator) dan horizon waktu pengamatan didefinisikan dengan jelas?
Jumlah eksaserbasi/tahun	Count / Hitungan	Apakah periode pengamatan (follow-up) setiap pasien sama panjang?
Waktu sampai relaps	Time-to-Event	Apakah terdapat penyensoran (censoring) pada subjek yang hilang dari pengamatan?
Skor nyeri (VAS) 0-10	Ordinal / Score	Apa asumsi dan target analisis ketika skor ini

		diperlakukan sebagai numerik kontinu?

> **HATI-HATI**
> *Sebuah skor klinis yang ditulis dengan angka (seperti skor depresi atau skor nyeri 0-10) tidak otomatis memiliki sifat matematis riil yang sama dengan kadar hemoglobin atau berat badan. Selisih skor nyeri dari 2 ke 3 belum tentu mencerminkan perbedaan intensitas nyeri biologis yang sama dengan selisih dari 8 ke 9. Keputusan memperlakukan skor ordinal sebagai variabel kontinu harus dipertanggungjawabkan berdasarkan konstruk teoretis, bentuk distribusi, dan tujuan penelitian klinis.*

5. Struktur Observasi Lebih Penting Daripada Nama Variabel

Dalam dunia nyata, keputusan memilih metode analisis tidak hanya didasarkan pada jenis data (nominal, ordinal, atau kontinu), melainkan sangat bergantung pada bagaimana pengamatan (observasi) tersebut terstruktur secara metodologis:

a. Data Independen:
Independent Observations: Ketika informasi dari satu unit pasien tidak terkait secara struktural dengan pasien lain. Misalnya, mengukur kadar HbA1c dari 100 pasien yang berbeda di dua kelompok terapi.

b. Data Berpasangan:
Paired Data: Timbul secara alami ketika dua pengamatan saling berkaitan erat atau berpasangan. Contoh paling klasiknya adalah pengukuran tekanan darah sistolik sebelum dan sesudah terapi pada pasien yang sama, atau desain kasus-kontrol yang dicocokkan (matched case-control).

c. Pengukuran Berulang:
Repeated Measures / Longitudinal Data: Terjadi ketika subjek yang sama diukur berkali-kali sepanjang waktu (misalnya HbA1c pada bulan 0, 3, 6, dan 12). Karena pengukuran berada dalam pasien yang sama, data cenderung saling berkorelasi erat. Mengabaikan korelasi ini akan menghasilkan kesalahan estimasi standard error yang fatal.

d. Data Berkelompok:
Clustered Data: Terjadi ketika pasien dikelompokkan dalam unit tertentu (seperti ruang rawat, dokter yang merawat, atau rumah sakit). Pasien-pasien di dalam satu rumah sakit cenderung memiliki karakteristik yang lebih mirip dibandingkan pasien dari rumah sakit lain, menciptakan korelasi intra-cluster.

Tabel 27. Mengilustrasikan bagaimana variabel yang sama (HbA1c) dapat melahirkan struktur data yang berbeda dengan konsekuensi konseptual analitik yang bervariasi

Skenario Penelitian	Struktur Data	Konsekuensi Konseptual dan Analitik
HbA1c Kelompok A vs Kelompok B pada pasien yang berbeda	Independent (Independen)	Perbandingan dilakukan antar-pasien; unit observasi tidak saling memengaruhi.
Kadar HbA1c sebelum vs sesudah terapi pada pasien yang sama	Paired (Berpasangan)	Fokus analisis bergeser pada perubahan (selisih) nilai di dalam individu itu sendiri.
Pengukuran HbA1c pada bulan 0, 3, 6, dan 12 pada pasien yang sama	Repeated / Longitudinal	Wajib memperhitungkan korelasi pengukuran berulang (misal: Mixed-Effects atau GEE).
Pasien diabetes yang direkrut dari 12 rumah sakit berbeda	Clustered (Berkelompok)	Evaluasi efek cluster harus dilakukan karena pasien dalam satu RS tidak sepenuhnya independen.

CLINICAL REASONING

Outcome klinis yang identik dapat membutuhkan strategi analisis yang sama sekali berbeda ketika struktur observasinya berubah. Inilah alasan mendasar mengapa pemahaman jenis variabel saja tidak akan pernah cukup untuk memilih uji statistik secara valid.

6. Distribusi Data

Distribusi menggambarkan pola penyebaran nilai suatu variabel dalam sampel kita. Untuk memahami distribusi secara utuh, kita menggabungkan visualisasi grafis (seperti histogram, boxplot, dan Q-Q plot) dengan ringkasan deskriptif numerik (mean, median, SD, dan IQR).

a. Distribusi Normal Bukan Standar Moral

Istilah "Normal" dalam statistika murni merujuk pada distribusi probabilitas kontinu berbentuk lonceng simetris (Gaussian). Altman dan Bland (1995) menegaskan bahwa konsep "Normality" dalam statistik tidak memiliki kaitan dengan istilah "normal" dalam pengertian klinis sehari-hari (yaitu sehat atau wajar). Data yang berdistribusi "tidak normal" tidak berarti data tersebut cacat atau memiliki kesalahan klinis; itu hanyalah cerminan dari variabilitas alami biologi manusia.

b. Mean dan Standard Deviation (SD)

Mean (rata-rata) sangat bermanfaat untuk menggambarkan pusat data jika sebarannya simetris. Namun, mean sangat sensitif terhadap nilai

ekstrem. Sementara itu, Standard Deviation (SD) menggambarkan seberapa jauh nilai-nilai individu pasien menyebar di sekitar mean sampel. SD adalah potret variabilitas biologis riil dari pasien Anda.

c. Median dan Interquartile Range (IQR)

Median (nilai tengah) adalah nilai yang membagi data menjadi dua bagian sama besar. Median sangat tangguh (robust) terhadap pengaruh outlier atau distribusi yang sangat miring (skewed). IQR menggambarkan rentang 50% data di bagian tengah (persentil ke-25 hingga ke-75) dan merupakan pendamping terbaik untuk median.

d. Skewness (Kemiringan)

Bentuk distribusi sebaiknya tidak dinilai hanya berdasarkan angka ringkas tunggal. Altman dan Bland (1996) menekankan pentingnya penggunaan grafik visual seperti histogram atau scatter diagram untuk mendeteksi skewness, bimodalitas, atau keberadaan pencilan secara langsung.

7. Apakah Semua Data Harus Diuji Normalitas?

Pendekatan mekanis tradisional yang mengajarkan alur kaku: "uji normalitas formal (Shapiro-Wilk/Kolmogorov-Smirnov) -> jika p > 0,05 t-test; jika p < 0,05 Mann-Whitney" adalah simplifikasi yang menyesatkan dalam praktik kedokteran.

Uji normalitas formal memiliki kelemahan yang nyata berkaitan dengan ukuran sampel:

Pada sampel besar (misal n > 150), uji formal memiliki kekuatan (power) yang sangat tinggi. Penyimpangan kecil dari distribusi normal yang secara klinis tidak penting akan dideteksi sebagai "tidak normal" (p < 0,05). Akibatnya, peneliti secara keliru menghindari t-test, padahal t-test sangat tangguh (robust) pada sampel besar berkat Central Limit Theorem.

Pada sampel kecil (misal n < 30), uji formal memiliki kekuatan yang sangat rendah. Uji ini sering kali gagal mendeteksi penyimpangan normalitas yang parah (p > 0,05). Akibatnya, peneliti merasa aman menggunakan t-test padahal asumsi parametriknya telah dilanggar secara serius.

Oleh karena itu, penilaian normalitas harus merupakan bagian dari proses model checking yang menggabungkan inspeksi visual (histogram/Q-Q plot), ukuran sampel, sifat alami variabel klinis tersebut, dan ketangguhan dari metode yang akan dijalankan.

PESAN UNTUK PPDS

Asesmen normalitas data adalah bagian dari pemeriksaan kelayakan model (model checking) bukan gerbang otomatis biner yang memutuskan seluruh keputusan analisis statistik Anda.

8. Outlier (Pencilan)

Outlier adalah nilai observasi yang tampak sangat menyimpang atau jauh dari sebagian besar data lainnya dalam konteks penelitian. Nilai ekstrem ini dapat bersumber dari: kesalahan input data, kegagalan kalibrasi instrumen, variasi biologis nyata pasien (misalnya kadar laktat yang sangat tinggi pada pasien septic shock), atau keberadaan subpopulasi yang berbeda. Mengapa outlier tidak boleh langsung dihapus dari spreadsheet?

- Outlier biologis nyata mungkin membawa informasi klinis terpenting dalam penelitian tersebut.
- Menghapus data semata-mata karena nilainya ekstrem atau tidak sesuai dengan hasil yang diinginkan peneliti akan memicu bias seleksi yang parah.
- Outlier dapat menjadi petunjuk adanya heterogenitas klinis yang perlu dieksplorasi lebih mendalam.

Dokter harus menerapkan algoritma penanganan outlier yang bertanggung jawab:

Temukan nilai ekstrem -> Verifikasi sumber data/rekam medis -> Tinjau konteks klinis biologis subjek -> Analisis pengaruh outlier terhadap model (lakukan sensitivity analysis dengan dan tanpa outlier) -> Dokumentasikan keputusan secara transparan dalam laporan penelitian.

9. Missing Data (Nilai Hilang)

Missing data bukan sekadar sel kosong yang dapat diabaikan begitu saja di spreadsheet. Dampaknya terhadap validitas riset klinis bergantung pada seberapa banyak data yang hilang, variabel mana yang hilang, dan mekanisme hilangnya data tersebut. Little dan Rubin (2019) membagi mekanisme hilangnya data klinis ke dalam tiga kategori konseptual penting yang disajikan pada Tabel 28.

Tabel 28. Mekanisme missing

Mekanisme Missing Data	Makna Konseptual Sederhana	Catatan Metodologis dan Konsekuensi
MCAR (Missing Completely At Random)	Hilangnya data benar-benar acak; tidak bergantung pada variabel apa pun (baik yang diamati maupun	Merupakan asumsi yang sangat kuat dan jarang terpenuhi secara murni di dunia medis klinis.

Mekanisme Missing Data	Makna Konseptual Sederhana	Catatan Metodologis dan Konsekuensi
	tidak).	
MAR (Missing At Random)	Setelah mempertimbangkan data yang terdokumentasi, hilangnya data tidak bergantung lagi pada nilai yang hilang tersebut.	Sering digunakan sebagai basis teoretis untuk metode imputasi modern atau model regresi model-based.
MNAR (Missing Not At Random)	Hilangnya data berkaitan langsung dengan nilai yang hilang itu sendiri (bahkan setelah melihat data yang ada).	Memerlukan analisis sensitivitas (sensitivity analysis) atau model statistik khusus untuk mencegah bias seleksi.

JANGAN LAKUKAN SECARA OTOMATIS

Menghapus baris pasien yang memiliki nilai kosong (complete-case analysis / listwise deletion) secara otomatis oleh perangkat lunak dapat merusak kekuatan uji statistik (menurunkan precision) dan memperkenalkan bias seleksi yang berat jika mekanisme hilangnya data bersifat MAR atau MNAR.

10. Data Cleaning Bukan 'Membuat Data Menjadi Bagus'

Data cleaning (pembersihan data) bertujuan untuk mengidentifikasi ketidakkonsistenan administrasi dan memastikan bahwa data mencerminkan observasi klinis yang sesungguhnya. Tujuannya sama sekali bukan mengubah nilai agar menghasilkan p-value yang signifikan secara statistik. Checklist praktis pembersihan data klinis meliputi:

- Periksa rentang nilai yang mustahil secara klinis (misal: SBP 400 mmHg atau tinggi badan 300 cm).
- Periksa konsistensi unit pengukuran (misal: mg/dL vs mmol/L, atau kilogram vs gram).
- Periksa duplikasi entri data pasien yang tidak sengaja terinput ganda.
- Periksa logika tanggal (misal: tanggal pulang rumah sakit tercatat mendahului tanggal masuk).
- Periksa kesalahan penulisan kode kategori (misal: penulisan 'L', 'l', 'Laki', 'Laki-laki' pada variabel yang sama).
- Dokumentasikan setiap koreksi data dan aturan eksklusi secara transparan.

11. Studi Kasus dan Contoh Lintas Bidang

a. Kasus Pemantik: Data Sama, Pertanyaan Berbeda

Bayangkan Anda memiliki satu lembar dataset yang berisi informasi dari 120 pasien diabetes melitus tipe 2: kadar HbA1c pada bulan ke-0, 3, dan 6; pembagian kelompok intervensi vs kontrol; usia; jenis kelamin; penggunaan terapi insulin; dan frekuensi kejadian hipoglikemia. Tabel 29 mengilustrasikan bagaimana dataset tunggal ini dapat melahirkan beragam pertanyaan statistik yang sangat berbeda.

Tabel 29. Dataset tunggal ini dapat melahirkan beragam pertanyaan statistik yang sangat berbeda

Pertanyaan Klinis	Jenis Outcome	Struktur Observasi	Keluarga Analisis Statistik
Berapa kadar HbA1c baseline pasien di klinik?	Kontinu	Deskriptif tunggal	Rerata, SD, Median, atau IQR
Apakah kadar HbA1c baseline berbeda antarkelompok?	Kontinu	Dua kelompok independen	Independent t-test atau Mann-Whitney
Apakah HbA1c menurun dari bulan ke-0 ke bulan ke-6?	Kontinu	Berpasangan (paired)	Paired t-test atau Wilcoxon signed-rank
Apakah pola perubahan HbA1c 0-3-6 bulan berbeda antarkelompok?	Kontinu	Repeated / Longitudinal	Mixed-Effects Model atau GEE
Apakah intervensi nutrisi meningkatkan kejadian hipoglikemia?	Binary atau Count	Dua kelompok independen	Chi-Square, Fisher Exact, atau Poisson

b. Contoh Kasus Lintas Bidang Kedokteran dan Gizi

Tabel 30. Menyajikan contoh variabilitas tipe variabel dan struktur data di berbagai bidang spesialisasi klinis kedokteran dan ilmu gizi

Bidang Spesialisasi	Variabel Utama	Jenis dan Struktur Data	Pertanyaan Klinis Utama
Kardiologi	Tekanan Darah Sistolik	Kontinu, Repeated measures	Bagaimana tren perubahan SBP selama setahun follow-up?
Onkologi	Stadium Kanker	Ordinal, Independen	Apakah angka kesembuhan berbeda menurut stadium diagnosis?

Pulmonologi	Frekuensi Eksaserbasi	Count (Hitungan), Independen	Apakah terapi inhaler baru menurunkan frekuensi serangan asma?
Gizi Klinik	Asupan Energi Harian	Kontinu (Skewed), Independen	Bagaimana distribusi asupan energi pada pasien geriatri?
Gizi Masyarakat	Status Anemia	Binary (Biner), Independen	Apakah proporsi anemia berbeda menurut tingkat paparan polusi?

12. Mitos dan Klarifikasi

Tabel 31. Merangkum beberapa mitos yang sangat populer di kalangan peneliti kedokteran beserta klarifikasi ilmiahnya

Mitos yang Umum Beredar	Klarifikasi Ilmiah yang Benar
"Jika variabel diukur dengan skala angka 0-10, maka data tersebut otomatis kontinu."	Tidak otomatis. Kita harus menilai makna biologis skala tersebut, ketiadaan desimal alami, dan kesesuaian asumsi analisis numerik.
"Setiap data numerik wajib diuji normalitasnya secara formal."	Tidak selalu. Kebutuhan uji normalitas sangat bergantung pada jenis uji statistik yang akan dijalankan, ukuran sampel, dan target parameter estimasi.
"Hasil p-value > 0,05 pada uji normalitas membuktikan bahwa data berdistribusi normal sempurna."	Salah. Hasil tersebut hanya menyatakan gagal menolak hipotesis nol (H0), bukan membuktikan bahwa H0 pasti benar, terutama pada sampel kecil.
"Pencilan (outlier) harus segera dibuang dari dataset agar hasil penelitian bagus."	Tidak boleh. Pencilan biologis harus diverifikasi, dipahami mekanismenya, dan dievaluasi dampaknya lewat sensitivity analysis.
"Missing data di bawah 5% di spreadsheet pasti aman untuk diabaikan."	Tidak ada ambang batas persentase yang aman jika mekanisme hilangnya data (seperti MAR atau MNAR) memperkenalkan bias seleksi sistematis.
"Data pengukuran berulang (repeated measures) pada pasien yang sama dapat dianalisis seolah-olah mereka berasal dari pasien yang berbeda."	Sangat dilarang. Struktur korelasi di dalam diri pasien (intra-subject correlation) wajib dipertahankan dalam model analisis.

13. Rangkuman

a. Representasi angka di spreadsheet hanyalah simbol; makna klinis biologi pasienlah yang menentukan jenis data yang sesungguhnya.
b. Jenis outcome dan struktur observasi adalah dua entitas yang berbeda, namun memiliki peran yang sama penting dalam mendikte pilihan analisis.
c. Data paired (berpasangan), repeated measures, dan clustered tidak boleh dianalisis menggunakan metode yang mengasumsikan independensi observasi.
d. Distribusi data wajib dipahami secara komprehensif melalui inspeksi visual grafis (histogram, boxplot, Q-Q plot) dan ringkasan deskriptif numerik — bukan hanya bersandar pada satu uji formal normalitas.
e. Istilah "Normal" dalam statistika murni merupakan nama model probabilitas matematis, bukan standar moral klinis yang membedakan kondisi "sehat" dan "sakit".
f. Pencilan (outlier) adalah sinyal ilmiah untuk penyelidikan metodologis lebih dalam, bukan perintah otomatis untuk menghapus baris data.
g. Missing data (nilai hilang) membawa asumsi analitis tertentu; metode listwise deletion dapat secara serius merusak presisi estimasi dan memicu bias.
h. Data cleaning (pembersihan data) bertujuan meluruskan representasi observasi agar akurat, bukan memanipulasi data demi mendapatkan signifikansi p-value.
i. Sebelum melompat pada pemilihan uji statistik, tanyakan: Apa jenis outcome saya? Bagaimana observasi saling berhubungan? Dan apa kuantitas klinis yang ingin saya estimasikan?
j. Memahami jenis dan struktur data pada bab ini akan menjadi fondasi mutlak untuk melangkah pada filosofi perbandingan kelompok pada Bab 4.

14. Latihan dan Evaluasi

Latihan Formatif (Pilihan Ganda)

1) Kadar HbA1c diukur pada sekelompok pasien diabetes melitus tipe 2 yang sama sesaat sebelum terapi dimulai dan kembali diukur setelah 3 bulan terapi. Struktur data utama pada skenario ini adalah:
   a. Independen
   b. Paired (Berpasangan)
   c. Clustered (Berkelompok)
   d. Time-to-Event

Kunci dan Pembahasan: Kunci Jawaban: b. Paired (Berpasangan)

Rasional: Karena pengukuran kadar HbA1c dilakukan sebanyak dua kali pada subjek pasien yang sama (sebelum dan sesudah intervensi), yang berarti kedua observasi tersebut memiliki korelasi biologis alami (dependen) di dalam diri subjek. Analisis harus berfokus pada selisih (perubahan) nilai intra-individu.

2. Seorang peneliti menemukan adanya satu nilai ekstrem yang sangat tinggi pada variabel kadar bilirubin pasien geriatri. Tindakan yang paling tepat secara metodologis adalah:

A. Langsung menghapus baris data tersebut agar sebaran data menjadi normal.
B. Mengganti nilai ekstrem tersebut dengan nilai rata-rata (mean) sampel.
C. Melakukan verifikasi sumber data rekam medis, meninjau konteks klinis biologis pasien, serta mengevaluasi pengaruhnya lewat sensitivity analysis.
D. Mengabaikan pencilan tersebut karena ukuran sampel penelitian sudah sangat besar.

Kunci dan Pembahasan: Kunci Jawaban: c. Melakukan verifikasi sumber data rekam medis, meninjau konteks klinis biologis pasien, serta mengevaluasi pengaruhnya lewat sensitivity analysis.
Rasional: Outlier tidak boleh langsung dibuang tanpa justifikasi metodologis karena nilai ekstrem tersebut bisa mencerminkan realitas biologis pasien kritis yang sangat penting. Analisis sensitivitas (sensitivity analysis) dijalankan dengan dan tanpa pencilan untuk melihat stabilitas kesimpulan klinis.

3. Pernyataan manakah yang paling akurat mengenai penilaian normalitas data dalam riset klinis?

A. Uji formal normalitas (seperti Shapiro-Wilk) adalah satu-satunya penentu mutlak metode analisis.
B. Uji normalitas formal sangat kuat (powerful) pada sampel kecil sehingga sangat bisa diandalkan.
C. Inspeksi visual grafis (histogram/Q-Q plot) dikombinasikan dengan ukuran sampel, sifat variabel, dan ketangguhan metode analitik adalah cara terbaik menilai sebaran data.
D. Semua data yang miring (skewed) wajib ditransformasi secara matematis sebelum dianalisis.

Kunci Jawaban: C. Inspeksi visual grafis (histogram/Q-Q plot) dikombinasikan dengan ukuran sampel, sifat variabel, dan ketangguhan metode analitik adalah cara terbaik menilai sebaran data.

Rasional: Uji normalitas formal memiliki bias kekuatan sampel (terlalu sensitif pada sampel besar, dan tidak bertenaga pada sampel kecil). Penilaian sebaran data harus diletakkan dalam konteks model checking yang memadukan visualisasi grafis dan pertimbangan klinis.

4. Mengapa mekanisme hilangnya data (missing data) perlu dinilai secara kritis sebelum analisis statistik dilakukan?

A. Karena missing data dapat dengan mudah diabaikan jika software statistik tetap berhasil memproses data.
B. Karena hilangnya data dapat merusak presisi estimasi dan memicu bias seleksi sistematis tergantung pada mekanismenya (MCAR, MAR, atau MNAR).
C. Karena hilangnya data hanya menjadi masalah metodologis jika persentase sel kosong di spreadsheet melebihi 50%.
D. Karena setiap missing data cukup diisi (diimputasi) menggunakan nilai rata-rata (mean) kelompok.

Kunci dan Pembahasan: Kunci Jawaban: B. Karena hilangnya data dapat merusak presisi estimasi dan memicu bias seleksi sistematis tergantung pada mekanismenya (MCAR, MAR, atau MNAR).
Rasional: Penghapusan baris data kosong secara otomatis (complete-case analysis) akan mempersempit ukuran sampel efektif (presisi turun) dan berisiko tinggi memicu bias estimasi jika hilangnya data berkaitan dengan karakteristik pasien (MAR atau MNAR).

5. Kadar gula darah sewaktu dari pasien yang sama diukur pada 4 titik waktu yang berbeda. Mengapa pengamatan ini tidak boleh diperlakukan sebagai 4 observasi pasien yang berbeda?

A. Karena datanya berskala nominal.
B. Karena pengukuran-pengukuran di dalam individu pasien yang sama cenderung berkorelasi erat (intra-subject correlation) sehingga melanggar asumsi independensi.
C. Karena seluruh data repeated measures wajib dianalisis menggunakan uji nonparametrik.
D. Karena variabilitas biologis intra-individu tidak memiliki makna klinis.

Kunci dan Pembahasan: Kunci Jawaban: B. Karena pengukuran-pengukuran di dalam individu pasien yang sama cenderung berkorelasi erat (intra-subject correlation) sehingga melanggar asumsi independensi.
Rasional: Mengabaikan struktur korelasi data longitudinal/pengukuran berulang dengan memperlakukannya sebagai observasi independen akan

merusak nilai standard error, mendistorsi p-value, dan menghasilkan kesimpulan yang menyesatkan.

Soal Reflektif untuk Diskusi Kelas

"Mana yang lebih berbahaya dalam praktik klinis kedokteran sehari-hari: ketidakpastian yang terlihat (seperti rentang interval kepercayaan yang lebar akibat ukuran sampel kecil) atau ketidakpastian yang tidak disadari (seperti bias sistematis akibat pengabaian outlier biologis atau penanganan missing data yang salah)?"

Panduan Diskusi Pedagogik: Mahasiswa diharapkan mampu menalar bahwa ketidakpastian yang terlihat (CI lebar) memberikan kejujuran ilmiah mengenai keterbatasan data, sehingga dokter akan bersikap lebih hati-hati saat menerapkannya pada pasien. Sebaliknya, ketidakpastian yang tidak disadari (bias tersembunyi akibat manipulasi data, membuang outlier biologis tanpa laporan, atau mengabaikan missing data) dapat menghasilkan angka estimasi yang tampak sangat presisi (CI sempit) namun salah arah secara sistematis. Hal ini akan menimbulkan rasa percaya diri palsu (false confidence) pada dokter, yang berisiko fatal membahayakan keselamatan pasien.

15. Checklist: Before You Analyse A Variable

Gunakan keenam pertanyaan kunci pada Tabel 32 sebagai panduan cepat Anda sebelum melakukan analisis terhadap variabel klinis apa pun:

Tabel 32. Panduan cepat anda sebelum melakukan analisis terhadap variabel klinis

Pertanyaan Kritis Sebelum Analisis	Penjelasan Praktis dan Tindakan Klinis
What does it mean? (Apa maknanya?)	Apakah variabel ini mewakili kategori murni, tingkatan urutan, kadar kontinu biologis, jumlah hitungan, biner, atau waktu sampai event?
Who or what is the unit? (Apa unit observasinya?)	Apakah setiap baris data berasal dari pasien yang berbeda, berpasangan (paired), pengukuran berulang (repeated), atau cluster rumah sakit?
What does the distribution look like? (Bagaimana sebarannya?)	Periksa sebaran data menggunakan visualisasi histogram/Q-Q plot. Perhatikan nilai rata-rata, median, standar deviasi, dan skewness data.
Is anything missing? (Adakah data yang hilang?)	Berapa banyak data yang hilang? Pada variabel mana? Di kelompok mana? Serta apa kemungkinan alasan hilangnya data tersebut (MCAR, MAR, MNAR)?

Pertanyaan Kritis Sebelum Analisis	Penjelasan Praktis dan Tindakan Klinis
Is the extreme value real? (Apakah pencilan itu nyata?)	Verifikasi keaslian pencilan melalui rekam medis sebelum memutuskan menghapus atau mempertahankan nilai tersebut dalam analisis sensitivitas.
What am I trying to estimate? (Apa yang ingin diestimasi?)	Apakah tujuan Anda mendeskripsikan data, membandingkan kelompok, mencari hubungan asosiasi, memprediksi risiko, mendiagnosis, atau melihat prognosis?

DARI BAB 3 KE BAB 4:

Jembatan Menuju Bab Selanjutnya

Setelah Anda memahami sepenuhnya bentuk dan struktur data pada Bab 3 ini, langkah logis berikutnya adalah mempelajari filosofi perbandingan kelompok (Bab 4). Kita akan membedah bagaimana membandingkan kelompok klinis secara tepat, mengapa mean, median, dan proporsi memerlukan pendekatan yang berbeda, serta bagaimana hubungan independen vs berpasangan mengubah keseluruhan kalkulasi statistik.

BAB 4
MEMBANDINGKAN KELOMPOK

Menilai Bukti secara Valid, Mengidentifikasi Perbedaan Kelompok, dan Menolak Penalaran Statistik Mekanis

1. Pendahuluan

Variabilitas adalah bagian alami dari data medis. Pasien dengan diagnosis yang sama sering kali menunjukkan respons terapi yang berbeda, memiliki profil biokimiawi yang bervariasi, dan menunjukkan luaran (outcome) klinis yang tidak identik. Ketika seorang klinisi atau peneliti ingin menguji apakah suatu intervensi baru lebih efektif dibanding terapi standar, atau apakah kadar biomarker tertentu berbeda antara pasien sehat dan pasien dengan kondisi kritis, ia sedang berhadapan dengan masalah perbandingan kelompok.

Dalam pendidikan biostatistika kedokteran tradisional, topik ini kerap kali direduksi menjadi serangkaian algoritma mekanis yang kaku, misalnya: 'jika data normal, gunakan uji parametrik (t-test); jika tidak normal, gunakan uji non-parametrik (Mann-Whitney)'. Pendekatan semacam ini berpotensi membahayakan penalaran medis mahasiswa. Klinisi dilatih untuk menjadi operator software statistika, alih-alih pemikir statistik yang kritis.

Fokus pembahasan pada Bab 4 ini adalah mendudukkan kembali filosofi dasar perbandingan kelompok ke tempatnya yang benar: di atas kerangka berpikir klinis. Sebelum bertanya 'uji statistik apa yang harus saya jalankan?', seorang calon dokter harus terlebih dahulu mengklarifikasi pertanyaan klinisnya, memahami unit observasi dan struktur datanya, merumuskan target estimasi (estimand) yang relevan, serta menyadari asumsi-asumsi teoretis dan metodologis yang membatasi kesimpulannya. Melalui pendekatan ini, statistik diletakkan sebagai instrumen penalaran kritis untuk menerjemahkan angka di atas kertas menjadi keputusan medis di samping tempat tidur pasien.

2. Capaian Pembelajaran

Setelah menyelesaikan bab ini, mahasiswa diharapkan mampu:

a. Menyusun dan merumuskan pertanyaan perbandingan klinis yang kabur menjadi target estimasi (estimand) statistik yang konkret dan operasional.

b. Membedakan dengan tegas antara struktur data independen (independent observations) dan data berpasangan (paired/matched data) serta konsekuensi metodologisnya terhadap pemilihan metode analisis.
c. Menjelaskan landasan filosofis, kegunaan, asumsi utama, dan keterbatasan di balik independent t-test, Welch t-test, paired t-test, Mann-Whitney U, Wilcoxon signed-rank, chi-square, Fisher exact, ANOVA, Welch ANOVA, Kruskal-Wallis, serta prosedur komparasi berganda (post-hoc).
d. Menjelaskan mengapa uji normalitas formal (seperti Shapiro-Wilk) tidak boleh diperlakukan sebagai gerbang tunggal yang kaku dalam pemilihan metode analisis.
e. Membedakan pengujian hipotesis (p-value) dari estimasi besaran efek (effect size) dan kepentingan klinis (clinical importance) secara mandiri.
f. Mengenali dan mengidentifikasi kesalahan konseptual yang lazim terjadi dalam praktik komparasi, seperti mengabaikan struktur pairing, melakukan t-test berulang pada tiga kelompok atau lebih, dan menyimpulkan kesamaan mutlak dari p-value yang tidak signifikan.
g. Menginterpretasikan hasil perbandingan secara proporsional, menyadari keterbatasan bukti, dan menghindari overclaim kausalitas yang melampaui batas desain penelitian yang digunakan.

> **PRINSIP EMAS**
> *Pemilihan uji statistik bukanlah langkah pertama. Langkah pertama adalah merumuskan pertanyaan klinis Anda secara presisi: Siapa yang dibandingkan, dalam kondisi apa, apa outcome-nya, apa target estimasinya, dan bagaimana struktur observasinya?*

3. Pertanyaan Besar dan Peta Konsep Perbandingan

Setiap kali seorang dokter dihadapkan pada tabel atau data yang membandingkan kelompok, sebelum ia melihat ke arah p-value, ia harus menanyakan serangkaian pertanyaan kritis berikut:

a. Outcome apa yang sebenarnya sedang dibandingkan? Apakah berupa variabel kontinu (seperti kadar kolesterol atau tekanan darah), variabel biner (seperti hidup/meninggal), variabel ordinal (seperti derajat nyeri atau stadium kanker), atau variabel waktu-ke-kejadian (time-to-event)?
b. Apa target estimasi (estimand) kita? Apakah kita ingin mengetahui selisih rata-rata (mean difference), perbedaan proporsi/risiko (risk difference), rasio risiko (risk ratio), atau pergeseran distribusi peringkat (rank-based contrast)?

c. Berapa banyak kelompok yang kita miliki? Dua kelompok atau lebih dari dua kelompok?
d. Bagaimana hubungan antarsubjek di dalam kelompok tersebut? Apakah mereka sepenuhnya independen, berpasangan (paired), berulang (longitudinal), atau berkelompok (clustered)?
e. Bagaimana sifat variabilitas, keberadaan pencilan (outlier), dan heterogenitas varians di dalam data?
f. Apakah perbedaan besaran efek yang teramati ini benar-benar penting dan bermakna bagi pasien?

4. Filosofi Dasar: Comparison = Contrast

Secara filosofis, melakukan perbandingan statistik berarti mengevaluasi suatu kontras (contrast). Kontras ini mencerminkan perbedaan nyata antar-intervensi atau kondisi klinis yang dievaluasi. Uji signifikansi hipotesis nol (null hypothesis significance testing) dan nilai p (p-value) hanyalah alat bantu untuk melihat seberapa kompatibel data kita dengan model teoretis tanpa efek (hipotesis nol).

Klinisi harus menyadari bahwa signifikansi statistik ($p < 0,05$) tidak menjamin adanya manfaat klinis yang nyata bagi pasien. Oleh karena itu, standar pelaporan artikel medis modern (seperti pedoman CONSORT 2025 untuk uji acak terkendali) mewajibkan peneliti untuk melaporkan hasil ringkasan per kelompok secara transparan, lengkap dengan ukuran efek (effect estimate) dan interval kepercayaannya (confidence interval) yang merepresentasikan ketidakpastian sampling.

5. Menentukan Unit Perbandingan

Sebelum memilih metode, kita harus mendefinisikan unit perbandingan secara presisi. Desain penelitian klinis menentukan struktur hubungan antar-observasi. Terdapat dua kategori utama unit perbandingan:

a. Observasi Independen (Independent Observations):

Terjadi ketika subjek di Kelompok A sepenuhnya berbeda dan tidak memiliki kaitan struktural dengan subjek di Kelompok B. Contoh: Uji klinis acak yang menguji obat antihipertensi baru pada 60 pasien (Kelompok A) dan obat plasebo pada 60 pasien lainnya (Kelompok B).

b. Observasi Berpasangan (Paired / Matched Observations):

Terjadi ketika setiap observasi di Kelompok A berpasangan secara alami dengan observasi di Kelompok B. Contoh paling klasik adalah desain sebelum-dan-sesudah (pre-post design) di mana kadar kolesterol pasien yang sama diukur sebelum terapi (Kelompok A) dan setelah terapi selama

12 minggu (Kelompok B). Contoh lainnya adalah desain kasus-kontrol yang dicocokkan (matched case-control) berdasarkan karakteristik klinis tertentu (seperti usia dan komorbiditas).

6. Mengapa Pairing Mengubah Filosofi Analisis?

Desain berpasangan secara konseptual mengubah pertanyaan penelitian. Kita tidak lagi bertanya 'apakah rata-rata Kelompok A secara keseluruhan berbeda dengan rata-rata Kelompok B?'. Pertanyaan yang diajukan adalah 'bagaimana distribusi selisih atau perubahan individual di dalam setiap pasangan?'

Dengan menggunakan setiap subjek sebagai kontrol bagi dirinya sendiri, kita mengeliminasi variabilitas biologis antarindividu. Hal ini membuat kekuatan statistik (statistical power) uji berpasangan jauh lebih kuat dibandingkan uji independen untuk ukuran sampel yang sama. Jika kita mengabaikan struktur pairing ini dan menganalisis data pre-post menggunakan uji independen, kita melakukan kesalahan konseptual berat yang membuang informasi berpasangan, memperbesar kesalahan standar (standard error), dan dapat menghasilkan kesimpulan false negative (Kegagalan mendeteksi efek terapi yang sebenarnya ada).

7. Dua Kelompok Independen Dengan Outcome Kontinu

a. Independent t-test: Pertanyaan tentang Rata-Rata

Uji t-test dua sampel independen (Student's t-test) digunakan ketika kita ingin mengestimasi perbedaan rata-rata (mean difference) antara dua kelompok independen berskala kontinu. Salah satu miskonsepsi yang paling awet di bidang kedokteran adalah anggapan bahwa uji t-test haram digunakan jika data mentah tidak berdistribusi Normal sempurna.

Secara metodologis, uji t-test sangat tangguh (robust) terhadap penyimpangan normalitas pada ukuran sampel yang cukup besar (biasanya $n \geq 30$ per kelompok) berkat Teorema Limit Pusat (Central Limit Theorem). Masalah yang sebenarnya mengacaukan uji t-test bukanlah normalitas, melainkan adanya pencilan (outlier) yang sangat ekstrem, ketidakseimbangan ukuran kelompok yang berat, serta perbedaan varians antar-kelompok yang sangat mencolok.

b. Student's t-test vs. Welch's t-test

Student's t-test klasik mengasumsikan bahwa varians kedua populasi adalah sama (equal variance). Welch's t-test (unequal variance t-test) adalah bentuk modifikasi yang tidak membutuhkan asumsi kesamaan varians tersebut. Jika varians kedua kelompok berbeda atau ukuran sampel

tidak seimbang, Welch's t-test memberikan tingkat kesalahan Tipe I yang jauh lebih terkontrol dan stabil. Oleh karena itu, para ahli biostatistika sangat menganjurkan penggunaan Welch's t-test secara default tanpa perlu melakukan uji formal kesamaan varians (seperti uji Levene) terlebih dahulu.

c. Contoh Klinis: Independent Mean Difference

Perhatikan skenario berikut: Sebuah uji klinis membandingkan efektivitas dua obat antihipertensi (Obat A vs. Obat B, n = 60 per kelompok) terhadap penurunan SBP setelah 8 minggu.

Tabel 33. Data hipotetik perbandingan penurunan SBP antara dua kelompok terapi

Kelompok	Ukuran Sampel (n)	Mean Perubahan SBP	Simpangan Baku (SD)
Obat A (Intervensi Baru)	60	-12.0 mmHg	± 8.0
Obat B (Standar Care)	60	-7.0 mmHg	± 9.0

Perbedaan rata-rata (mean difference) antara kedua kelompok adalah -5,0 mmHg (Obat A menurunkan SBP 5 mmHg lebih banyak dibanding Obat B). Untuk membuat keputusan klinis, dokter tidak boleh hanya bertanya 'apakah $p < 0,05$?'. Pertanyaan yang sesungguhnya adalah:

- Seberapa presisi estimasi selisih 5 mmHg ini? (Dilihat dari rentang 95% Confidence Interval-nya).
- Apakah penurunan ekstra sebesar 5 mmHg ini memiliki dampak protektif yang signifikan bagi pasien stroke atau kardiovaskular secara klinis? (Clinical Importance).

8. Mann-Whitney: Bukan Sekadar "Uji Median"

Uji Mann-Whitney U (juga dikenal sebagai Wilcoxon rank-sum test) adalah metode berbasis peringkat (rank-based) yang sering dianggap sebagai alternatif otomatis uji t-test jika asumsi normalitas dilanggar. Namun, mengajarkan bahwa Mann-Whitney membandingkan median secara universal adalah tidak tepat.

Mann-Whitney membandingkan seluruh distribusi data menggunakan informasi peringkat. Menyebutnya sebagai uji perbandingan median hanya benar jika kita berasumsi bahwa bentuk distribusi kedua kelompok adalah identik (hanya berbeda lokasinya). Jika bentuk distribusinya berbeda, Mann-Whitney sebenarnya menguji probabilitas

bahwa nilai subjek acak dari Kelompok A lebih besar daripada nilai subjek acak dari Kelompok B (probability of superiority).

Kapan uji berbasis peringkat ini sangat berguna?

1) Outcome yang diukur bersifat ordinal (seperti derajat keparahan edema, stadium kanker, atau skor visual nyeri 0–10).
2) Data kontinu sangat miring (highly skewed) dan dipengaruhi oleh pencilan (outlier) ekstrem yang tidak dapat dieliminasi secara biologis.
3) Ukuran sampel sangat kecil (misal n < 10) di mana estimasi rata-rata dan simpangan baku menjadi tidak stabil.

9. Dua Pengukuran Berpasangan

Ketika data diperoleh dari subjek yang sama (berpasangan), kita harus menggunakan metode analisis yang sesuai untuk menjaga integritas korelasi intra-subjek:

Tabel 34. Metode Kandidat untuk data berpasangan (paired)

Kategori Target Analisis	Metode Statistik	Filosofi Utama Uji
Rata-rata perubahan kontinu	Paired t-test	Menganalisis rata-rata dari selisih (difference) within-subject. Asumsi normalitas berlaku pada selisihnya, bukan data mentah.
Perubahan peringkat / ordinal	Wilcoxon signed-rank test	Membandingkan besaran dan arah peringkat dari selisih berpasangan.
Perubahan kategori biner	McNemar test	Berfokus pada pasangan tidak selaras (discordant pairs) sebelum dan sesudah intervensi.

PERINGATAN KLINIS

Kesalahan fatal yang sering dilakukan PPDS: Menggunakan independent t-test untuk membandingkan nilai sebelum dan sesudah intervensi pada pasien yang sama. Ini melanggar asumsi independensi dan dapat mengaburkan efek intervensi yang sebenarnya.

10. Outcome Kategorik: Chi-Square dan Fisher Exact

a. Filosofi Chi-Square

Uji chi-square (uji kuadrat-chi) digunakan untuk mengevaluasi asosiasi antara dua variabel kategorik pada tabel kontingensi. Filosofinya adalah membandingkan frekuensi kejadian yang diamati (observed counts) di setiap sel dengan frekuensi kejadian yang diharapkan (expected counts) seandainya kedua variabel tersebut sepenuhnya tidak berhubungan (independen).

Penting untuk diingat bahwa chi-square harus dihitung menggunakan jumlah absolut pasien (actual counts), bukan menggunakan persentase atau rata-rata proporsi. Memasukkan persentase langsung ke dalam rumus chi-square akan merusak kalkulasi matematisnya.

b. Fisher's Exact Test

Uji chi-square mengandalkan perkiraan distribusi asimtotik yang memerlukan jumlah sampel yang cukup besar. Aturan praktisnya, jika terdapat tabel kontingensi 2x2 di mana salah satu sel memiliki frekuensi harapan (expected count) kurang dari 5, uji chi-square menjadi tidak valid. Pada kondisi ini, kita harus menggunakan Fisher's exact test yang menghitung probabilitas eksak dari kombinasi tabel kontingensi tersebut secara langsung.

c. Uji Signifikansi Tidak Menggantikan Ukuran Efek

Bagi seorang dokter, p-value dari uji chi-square (misal p = 0,02) hanya menjawab pertanyaan 'apakah ada bukti asosiasi?'. Keputusan klinis memerlukan informasi ukuran efek (effect measure) biner untuk memahami besarnya dampak terapi. Kita memerlukan: Risk Difference (RD) untuk melihat manfaat absolut, Risk Ratio (RR) atau Odds Ratio (OR) untuk melihat manfaat relatif, serta Number Needed to Treat (NNT) untuk melihat berapa pasien yang harus diterapi untuk mencegah satu kejadian buruk.

d. Contoh Kategorik: Kejadian Hipoglikemia

Skenario: Sebuah studi membandingkan efek dua regimen insulin (Regimen A vs. Regimen B, n = 100 per kelompok) terhadap efek samping hipoglikemia berat selama 6 bulan follow-up.

Tabel 35. Tabel kontingensi kejadian hipoglikemia pada dua regimen insulin

Regimen Insulin	Hipoglikemia (+)	Hipoglikemia (-)	Total Pasien
Regimen A	12	88	100
Regimen B	22	78	100
Total	34	166	200

Dari data ini, kita dapat menghitung:

- Risiko Hipoglikemia Regimen A = 12% dan Regimen B = 22%.
- Risk Difference (RD) = 12% - 22% = -10% (Regimen A menghemat kejadian hipoglikemia sebesar 10% secara absolut).
- Risk Ratio (RR) = 12% / 22% = 0,55 (Regimen A menurunkan risiko relatif hipoglikemia sebesar 45%).
- Number Needed to Harm (NNH) = 1 / 0,10 = 10 (Setiap mengobati 10 pasien dengan Regimen B alih-alih Regimen A, kita memperkirakan akan ada tambahan 1 pasien yang mengalami efek samping hipoglikemia).

11. Tiga Atau Lebih Kelompok: Mengapa Bukan Banyak T-Test?

Ketika kita membandingkan tiga kelompok atau lebih (misalnya kelompok Diet Rendah Karbohidrat, Diet Mediterania, dan Diet Standar), kita mungkin tergoda untuk menjalankan uji t-test berulang-ulang untuk setiap pasang kelompok (A vs B, A vs C, B vs C).

Pendekatan ini sangat dilarang karena memicu masalah multiplisitas (multiplicity). Jika setiap uji menggunakan batas signifikansi $\alpha = 0,05$, maka probabilitas kumulatif untuk membuat setidaknya satu kesalahan Tipe I (menemukan perbedaan yang signifikan padahal sebenarnya tidak ada) akan meningkat secara eksponensial (dari 5% naik menjadi hampir 14% untuk 3 perbandingan).

a. Filosofi ANOVA

Analysis of Variance (ANOVA) mengatasi masalah ini dengan menyediakan uji global tunggal (omnibus test). ANOVA membandingkan variabilitas antarkelompok (variation between groups) dengan variabilitas di dalam kelompok (variation within groups). Jika variasi antarkelompok jauh lebih besar daripada variasi biologis/acak di dalam kelompok, ANOVA akan menghasilkan $p < 0,05$.

Hasil ANOVA yang signifikan hanya memberi tahu kita bahwa 'setidaknya ada satu pasang kelompok yang rata-ratanya berbeda secara nyata'. ANOVA tidak menunjukkan secara spesifik pasangan kelompok mana yang berbeda.

b. Post-hoc dan Multiple Comparisons

Untuk melacak pasangan kelompok mana yang berbeda setelah ANOVA menunjukkan hasil signifikan, kita harus menjalankan analisis pasca-uji (post-hoc comparison) seperti uji Tukey HSD, Bonferroni, atau Scheffe. Pemilihan metode post-hoc harus didasarkan pada pertanyaan riset awal: Apakah kita ingin membandingkan seluruh pasangan secara eksploratif (gunakan Tukey), atau hanya membandingkan kelompok intervensi dengan satu kontrol standar (gunakan Bonferroni/Dunnett)?

c. Welch ANOVA dan Kruskal-Wallis

Sama halnya dengan t-test, jika asumsi kesamaan varians dilanggar pada perbandingan ≥3 kelompok, kita harus menggunakan Welch ANOVA. Jika data berupa skala peringkat/ordinal atau data kontinu sangat miring, kita menggunakan Kruskal-Wallis sebagai alternatif berbasis peringkat.

Tabel 36. Welch ANOVA dan Kruskal-Wallis sebagai alternatif untuk perbandingan lebih dari dua kelompok

Metode Analisis	Tipe Target Pertanyaan	Kondisi Penggunaan	Landasan Filosofis
Welch ANOVA	Perbedaan rata-rata (mean difference) untuk ≥3 kelompok	Varians antar-kelompok heterogen atau ukuran sampel tidak seimbang	Fokus tetap pada perbandingan rata-rata populasi, menggunakan koreksi derajat bebas.
Kruskal-Wallis test	Perbedaan distribusi peringkat untuk ≥3 kelompok	Data skala ordinal atau kontinu yang sangat miring dengan pencilan	Generalisasi dari Mann-Whitney untuk ≥3 kelompok; tidak mengasumsikan distribusi tertentu.

12. Kasus Integratif Dan Studi Kasus Gizi

a. Kasus Integratif: Variabel Sama, Struktur Berbeda

Sebagai penegasan kembali, nama variabel klinis (seperti kadar HbA1c) sama sekali tidak mengunci uji statistik Anda. Metode analisis sepenuhnya ditentukan oleh struktur data dan target pertanyaan Anda.

Tabel 37. Contoh bagaimana variabel kadar hba1c memerlukan uji berbeda sesuai pertanyaan

Pertanyaan Penelitian Klinis	Struktur Data	Target Estimasi	Metode Statistik Kandidat
Berapa perbedaan kadar HbA1c akhir antara kelompok obat baru vs plasebo?	2 kelompok independen	Mean Difference (Selisih Rata-rata)	Welch's t-test (atau Student's t-test)
Apakah kadar HbA1c pasien berubah setelah menjalani program diet 12 minggu?	Kelompok berpasangan (pre-post)	Mean Change (Rata-rata perubahan)	Paired t-test / Wilcoxon signed-rank
Apakah rata-rata kadar HbA1c berbeda di antara pasien dengan 3 jenis diet berbeda?	3 kelompok independen	Rata-rata global kelompok	ANOVA / Welch ANOVA / Kruskal-Wallis
Bagaimana pola perubahan HbA1c pasien yang diukur pada bulan 0, 3, 6, dan 12?	Repeated measures longitudinal	Trajectory perubahan seiring waktu	Repeated Measures ANOVA / Mixed-Effects Model

b. Kasus Gizi: Tiga Pola Diet

Sebuah studi klinis fiktif mengevaluasi efek tiga jenis pola diet (Diet Keto, Diet Mediterania, dan Diet Rendah Lemak) terhadap penurunan berat badan setelah 12 minggu pada 150 pasien obesitas (n = 50 per kelompok). Sebelum melangkah ke analisis ANOVA, klinisi yang bijak akan mengajukan pertanyaan penalaran berikut:

- Apakah subjek di ketiga kelompok benar-benar independen dan tidak saling berinteraksi?
- Apakah penurunan berat badan (kg) berdistribusi cukup simetris tanpa adanya pencilan ekstrem?
- Apakah kita memiliki rencana kontras spesifik sebelum penelitian berjalan, atau kita ingin membandingkan seluruh pasangan pasca-uji?

13. "Normal Atau Tidak Normal?" Reframing

Kita harus menghentikan ritual mekanis yang terlalu menyederhanakan biostatistika menjadi 'gerbang tunggal uji normalitas

formal (seperti Shapiro-Wilk)'. Uji normalitas formal memiliki kelemahan inheren:

- Pada ukuran sampel kecil, uji normalitas sering gagal mendeteksi penyimpangan yang signifikan karena kekuatan uji (power) yang sangat rendah.
- Pada ukuran sampel besar, uji normalitas sangat sensitif terhadap penyimpangan minor yang sebenarnya tidak memengaruhi performa uji parametrik (teorema limit pusat membuat t-test tetap valid).

Calon dokter harus mengintegrasikan visualisasi grafis (seperti histogram, boxplot, dan Q-Q plot), pemahaman biologis variabel, ukuran sampel, serta sifat ketahanan (robustness) metode parametrik dalam mengambil keputusan analitis.

14. Effect Size, Confidence Interval, Dan Clinical Importance

Setiap perbandingan kelompok harus selalu bermuara pada estimasi besaran efek (effect size) beserta interval kepercayaannya (confidence interval) untuk memandu interpretasi klinis.

Tabel 38. Jenis ukuran efek yang relevan berdasarkan tipe outcome

Tipe Outcome Klinis	Ukuran Besaran Efek (Effect Size) yang Relevan
Kontinu Independen	Mean Difference (Selisih Rata-rata); Standardized Mean Difference (SMD)
Kategorik Biner (Dichotomous)	Risk Difference (RD), Risk Ratio (RR), Odds Ratio (OR), NNT/NNH
Ordinal / Peringkat (Rank-based)	Probability of Superiority, Rank correlation contrast
Kontinu Berpasangan (Paired)	Mean paired change (Rata-rata selisih berpasangan)

Selalu bedakan antara signifikansi statistik dan kepentingan klinis. Perbedaan rata-rata yang sangat kecil (misal penurunan SBP 0,5 mmHg) bisa menghasilkan $p < 0,001$ jika sampelnya mencapai puluhan ribu pasien, namun efek sekecil ini sama sekali tidak penting secara klinis bagi pasien Anda. Sebaliknya, perbedaan bermakna (misal penurunan SBP 8 mmHg) mungkin menghasilkan $p = 0,08$ pada sampel kecil, namun data tersebut tetap kompatibel dengan efek klinis yang menguntungkan.

15. Mitos dan Klarifikasi

Mari kita luruskan beberapa mitos biostatistika yang masih sering menyesatkan (Tabel 39).

Tabel 39. Mitos umum dalam perbandingan kelompok dan klarifikasinya

Mitos Statistik yang Keliru	Klarifikasi Ilmiah yang Benar
'Jika uji normalitas Shapiro-Wilk menghasilkan p < 0,05, kita wajib menggunakan uji non-parametrik.'	Salah. t-test tetap robust pada sampel besar berkat Teorema Limit Pusat. Kita harus melihat visualisasi data dan menilai apakah rata-rata tetap merupakan estimand yang bermakna.
'Uji Mann-Whitney adalah uji perbandingan median.'	Tidak secara umum. Mann-Whitney membandingkan peringkat keseluruhan distribusi. Hanya diinterpretasikan sebagai uji median jika bentuk distribusi kedua kelompok identik.
'Hasil ANOVA yang signifikan (p < 0,05) membuktikan bahwa seluruh kelompok berbeda.'	Salah. ANOVA hanya membuktikan bahwa setidaknya ada satu pasang kelompok yang berbeda. Kita memerlukan uji post-hoc untuk melacak pasangan mana yang berbeda.
'Uji t-test berpasangan adalah uji t-test biasa pada dua kolom data.'	Salah. t-test berpasangan menganalisis satu kolom baru yang berisi nilai selisih (difference) within-subject, bukan membandingkan dua kelompok independen.
'Uji chi-square yang signifikan berarti efek klinisnya pasti besar.'	Salah. Chi-square hanya menguji kekuatan bukti terhadap asosiasi. Besaran efek klinis harus dinilai terpisah menggunakan RR, OR, atau RD.

16. Rangkuman

a. Perbandingan kelompok harus selalu dimulai dari perumusan pertanyaan klinis dan kontras yang ingin diestimasi, bukan langsung melompat ke pemilihan uji statistik.
b. Struktur data independen dan berpasangan (paired) memiliki filosofi dan metode analisis yang berbeda secara fundamental.
c. Uji t-test berfokus pada perbandingan rata-rata, sedangkan uji Mann-Whitney membandingkan distribusi berbasis peringkat.
d. Welch's t-test lebih disukai secara default dibanding Student's t-test klasik karena tidak memerlukan asumsi kesamaan varians.
e. Uji chi-square menguji asosiasi pada counts (angka mutlak), dan harus selalu dilengkapi dengan ukuran efek klinis seperti RR, OR, atau RD.
f. Membandingkan tiga kelompok atau lebih dengan uji t-test berulang-ulang akan meningkatkan kesalahan Tipe I akibat masalah multiplisitas.
g. Uji normalitas formal bukanlah gerbang mutlak penentu uji statistik; keputusan harus menggabungkan visualisasi data dan ukuran sampel.
h. Signifikansi statistik (p-value) tidak sama dengan kepentingan klinis (clinical importance).

i. Interval kepercayaan (confidence interval) memberikan informasi krusial mengenai presisi dan rentang efek yang kompatibel dengan data.
j. Inferensi yang valid memerlukan desain penelitian yang kokoh untuk meminimalkan bias dan perancu, sebelum statistik dijalankan.

17. Latihan dan Evaluasi

Latihan Formatif Pilihan Ganda

1) Seorang peneliti mengukur tekanan darah sistolik pasien hipertensi sebelum dan sesudah intervensi diet rendah garam selama 12 minggu pada pasien yang sama. Struktur data utama ini adalah:
   a. Independen
   b. Berpasangan (paired)
   c. Clustered antar-rumah sakit
   d. Time-to-event

Kunci Jawaban: b

Pembahasan: *Pengukuran dilakukan pada subjek yang sama sebelum dan sesudah terapi, sehingga setiap pasien menjadi kontrol bagi dirinya sendiri. Ini adalah struktur data berpasangan (paired).*

2) Manakah dari pernyataan berikut yang paling tepat mengenai uji non-parametrik Mann-Whitney?
   a. Ia selalu menguji perbedaan median antara dua kelompok secara mutlak.
   b. Ia hanya valid digunakan jika data sampel berukuran besar.
   c. Ia membandingkan peringkat keseluruhan distribusi dan mengukur probability of superiority jika bentuk distribusi berbeda.
   d. Ia memiliki kekuatan statistik yang persis sama dengan paired t-test.

Kunci Jawaban: C

Pembahasan: *Mann-Whitney adalah uji berbasis peringkat yang membandingkan distribusi keseluruhan, bukan sekadar membandingkan median secara kaku.*

3) Dalam uji ANOVA satu arah (one-way ANOVA), hasil overall p-value < 0,05 berarti:
   a. Seluruh kelompok memiliki rata-rata yang berbeda satu sama lain.
   b. Setidaknya terdapat bukti bahwa tidak semua rata-rata populasi kelompok adalah sama.
   c. Kelompok dengan ukuran sampel terbesar pasti berbeda dengan kontrol.

d. Besaran efek klinis di antara seluruh kelompok pasti berkategori besar.

Kunci Jawaban: b

Pembahasan: ANOVA adalah uji global (omnibus). Hasil signifikan hanya membuktikan adanya perbedaan pada minimal satu pasang kelompok, yang harus dilacak lebih lanjut menggunakan uji post-hoc

4) Pada analisis outcome biner menggunakan uji chi-square, mengapa p-value saja tidak cukup untuk keputusan klinis?
   a. Karena chi-square tidak bisa digunakan pada sampel di atas 100 pasien.
   b. Karena chi-square hanya mendeteksi normalitas data.
   c. Karena p-value tidak mengukur besarnya dampak atau efek klinis nyata (seperti perbedaan risiko absolut atau relatif) bagi pasien.
   d. Karena chi-square hanya valid untuk data kontinu.

Kunci Jawaban: c

Pembahasan: P-value hanya mengukur kekuatan bukti statistik terhadap hipotesis nol. Penilaian klinis membutuhkan ukuran efek absolut (RD) dan relatif (RR/OR) beserta confidence interval-nya.

5) Apakah alasan utama memilih paired t-test dibandingkan independent t-test pada pre-post design?
   a. Karena ukuran sampel penelitian sangat kecil.
   b. Karena data berasal dari subjek yang sama sehingga korelasi intra-subjek harus dipertahankan untuk mengeliminasi variabilitas antarindividu.
   c. Karena paired t-test tidak memerlukan asumsi normalitas sama sekali.
   d. Karena outcome yang diukur berupa variabel biner.

Kunci Jawaban: b

Pembahasan: Desain berpasangan memanfaatkan subjek sebagai kontrol diri sendiri. Paired t-test menguji rata-rata selisih berpasangan secara valid sesuai struktur desainnya.

18. Checklist: Before You Compare Groups

Gunakan delapan pertanyaan kritis ini sebelum Anda membandingkan kelompok:

Tabel 40. Checklist delapan pertanyaan sebelum membandingkan kelompok

Langkah Evaluasi	Pertanyaan Kritis yang Harus Dijawab
Outcome?	Apakah variabel outcome berskala kontinu, biner, ordinal, count, atau time-to-event?
Estimand?	Apakah kita ingin mengestimasi perbedaan rata-rata, risiko absolut/relatif, ranking, atau perubahan dalam-pasien?
Groups?	Berapa jumlah kelompok yang dibandingkan (dua kelompok atau lebih dari dua kelompok)?
Independence?	Apakah observasi bersifat independen, berpasangan (paired), repeated measures, atau berkelompok (clustered)?
Assumptions?	Apakah ada pencilan ekstrem, ketidakseimbangan varians, atau expected counts yang terlalu rendah?
Effect Size?	Berapa nilai estimasi besaran efek klinisnya? Jangan hanya melaporkan p-value.
Uncertainty?	Seberapa presisi estimasi efek tersebut berdasarkan rentang Confidence Interval-nya?
Clinical Meaning?	Apakah perbedaan yang terukur tersebut benar-benar penting bagi kehidupan dan keselamatan pasien?

Jembatan ke Bab Selanjutnya

Kini kita telah menguasai filosofi dasar dalam membandingkan kelompok secara valid dan klinis. Pertanyaan berikutnya yang sangat menantang adalah: ketika kita menemukan bahwa dua variabel bergerak bersama (berasosiasi), apakah hubungan tersebut mencerminkan hubungan sebab-akibat (kausalitas), dipengaruhi oleh faktor perancu (confounding), ataukah sekadar kebetulan sampling? Kita akan membedahnya secara tuntas pada Bab 5: Association ≠ Causation.

BAB 5
ASSOCIATION ≠ CAUSATION

Memisahkan Hubungan Asosiatif dari Hubungan Sebab-Akibat dalam Pengambilan Keputusan Klinis

1. Pendahuluan

Dalam praktik klinis dan penelitian kedokteran, dokter sangat sering menemukan dua variabel yang bergerak bersama secara beriringan: kadar biomarker tertentu yang terukur lebih tinggi pada pasien dengan kondisi yang lebih berat, pasien yang menerima suatu intervensi khusus menunjukkan angka mortalitas yang berbeda, atau pola asupan gizi tertentu yang secara konsisten berkaitan dengan suatu kesudahan klinis (*outcome*). Hubungan semacam ini di dalam statistik disebut sebagai **asosiasi** atau korelasi. Namun, salah satu pelajaran terpenting dalam biostatistika yang harus dipegang teguh oleh seorang klinisi adalah: asosiasi yang teramati tidak dengan sendirinya menentukan atau membuktikan adanya hubungan sebab-akibat (kausalitas).

Bab ini secara khusus disusun untuk membangun kompetensi dan disiplin berpikir mahasiswa kedokteran terutama Peserta Program Pendidikan Dokter Spesialis (PPDS) dalam memisahkan pertanyaan deskriptif atau asosiatif dari pertanyaan kausal. Kita akan mempelajari konsep korelasi, memahami perancu (*confounding*), mengenali hubungan sebab-akibat terbalik (*reverse causation*), serta memahami peran variabel perantara (*mediator*) dan variabel benturan (*collider*) melalui bantuan diagram kausal sederhana yang disebut **Directed Acyclic Graph (DAG)**. Fokus utama pembahasan di sini bukanlah aspek matematis formal yang rumit dari causal inference, melainkan pembentukan kerangka penalaran kritis klinis yang proporsional saat menginterpretasikan hasil-hasil penelitian di dunia nyata.

> **PRINSIP UTAMA: Asosiasi vs Kausalitas.** *Association adalah pola keterikatan atau pergerakan bersama yang dapat kita amati secara empiris langsung di dalam data. Sebaliknya, Causal Effect (efek kausal) adalah pertanyaan teoretis yang membutuhkan definisi intervensi, kontras hipotesis, urutan waktu (temporal ordering), kesesuaian desain, dan asumsi tambahan yang eksplisit. Analisis statistik secanggih apa pun tidak akan pernah bisa menggantikan penalaran kausal (causal reasoning) yang berbasis pada patofisiologi dan metodologi desain.*

Mengapa pemahaman ini sangat krusial bagi keselamatan pasien? Karena kesalahan dalam mengira korelasi sebagai kausalitas (*correlation implies causation*) dapat berakibat fatal secara klinis. Dokter dapat menyarankan terapi atau tindakan yang sebenarnya tidak berkhasiat, atau bahkan membahayakan, hanya karena melihat adanya hubungan statistik yang kuat dalam studi observasional. Sebaliknya, klinisi juga tidak boleh bersikap ekstrem dengan menganggap semua studi observasional sama sekali tidak berguna untuk menjawab pertanyaan kausal. Literatur metodologi modern menunjukkan bahwa pertanyaan kausal sangat mungkin dipelajari dari data observasional, asalkan target efek kausalnya didefinisikan dengan jernih dan asumsi-asumsi pengontrolan bias diuji secara eksplisit (Hernán dan Robins, 2020; Greenland et al., 1999).

2. Capaian Pembelajaran

Setelah mempelajari bab ini dengan saksama, mahasiswa diharapkan mampu mencapai kompetensi klinis berikut:

- Membedakan secara konseptual antara istilah asosiasi, korelasi, prediksi, dan efek kausal dalam konteks data klinis.
- Menjelaskan urgensi temporal ordering (urutan waktu) dan desain penelitian dalam menarik kesimpulan sebab-akibat.
- Mengidentifikasi serta membedakan peran confounder, mediator, dan collider pada sebuah diagram kausal (DAG) sederhana.
- Menjelaskan secara metodologis mengapa strategi "memasukkan/menyesuaikan semua variabel" (adjust semua variabel) ke dalam model statistik dapat menghasilkan bias yang merugikan.
- Mengenali dan mendeteksi fenomena reverse causation (sebab-akibat terbalik) dan confounding by indication (perancu berdasarkan indikasi) pada kasus-kasus klinis sehari-hari.
- Menggunakan bahasa ilmiah yang hati-hati dan proporsional ketika mengomunikasikan hasil penelitian yang desainnya hanya mendukung tingkat klaim asosiatif.
- Menilai secara kritis apakah suatu analisis multivariat benar-benar menjawab pertanyaan kausal atau sekadar menghasilkan ukuran asosiasi yang disesuaikan (adjusted association).

Kompetensi yang Belum Ditargetkan pada Bab Ini:

- Menurunkan persamaan matematika formal untuk potential outcomes atau do-calculus.
- Mengukur atau memprogram estimasi statistik tingkat lanjut seperti Inverse Probability Weighting (IPW) atau G-computation secara mandiri.
- Membuat dan menggambar DAG kompleks yang melibatkan puluhan variabel menggunakan software khusus.
- Menganggap bahwa sebuah gambar DAG yang rapi secara otomatis membuktikan kebenaran hubungan kausal di lapangan.

3. Pertanyaan Besar dan Peta Konsep

Mari kita mulai dengan sebuah skenario klinis nyata yang sering dihadapi oleh para dokter spesialis di unit perawatan intensif (ICU):

> **CLINICAL TRIGGER: Kasus Pemantik: Transfusi Darah dan Mortalitas.** *Seorang residen PPDS membaca sebuah laporan studi kohort observasional yang menunjukkan bahwa pasien ICU yang menerima transfusi sel darah merah (PRC) dalam jumlah lebih banyak secara signifikan memiliki angka mortalitas yang lebih tinggi (28% vs 12%; p = 0,002). Residen tersebut langsung menyimpulkan: 'Transfusi darah merusak sistem imun pasien ICU, meningkatkan risiko kematian, dan kita harus menghentikan atau membatasi transfusi seketat mungkin.' Apakah kesimpulan kausal langsung seperti ini sah secara ilmiah?*

Jawaban kritisnya adalah: **Sama sekali belum sah**. Sebelum kita melompat pada kesimpulan kausal bahwa 'transfusi meningkatkan mortalitas', kita harus memahami alasan mendasar mengapa transfusi tersebut diberikan (indikasi transfusi), tingkat keparahan penyakit pasien *(severity of illness)*, kondisi komorbiditas, urutan waktu kejadian, serta bagaimana data dikumpulkan. Pada situasi ini, pasien dengan kondisi klinis yang lebih berat (lebih kritis, mengalami perdarahan aktif hebat, atau syok) secara fisiologis pasti memiliki kemungkinan yang jauh lebih tinggi untuk menerima transfusi darah, sekaligus memiliki risiko yang memang sudah tinggi untuk meninggal sejak awal. Fenomena inilah yang secara metodologis disebut sebagai ***confounding by indication*** (perancu berdasarkan indikasi medis).

Gambar di bawah ini merangkum peta jalan berpikir kritis klinis yang bergerak dari sekadar melihat pola keterikatan data kasar (crude association) hingga menilai pemenuhan elemen-elemen kausalitas secara berjenjang.



Gambar 1. Peta alur berpikir dari asosiasi menuju kesimpulan kausal

4. Association, Correlation, Prediction, dan Causation

Untuk meluruskan pemahaman, kita harus mendefinisikan empat istilah yang sangat sering digunakan secara tumpang tindih dalam diskusi medis, padahal masing-masing memiliki batasan klaim dan konsekuensi analitik yang berbeda:

a. Association (Asosiasi): Kondisi ketika nilai dari dua variabel bergerak bersama dalam pola tertentu secara empiris. Jika kita mengetahui nilai variabel X, kita dapat memperkirakan sebaran nilai variabel Y, tanpa mengasumsikan adanya mekanisme biologis atau arah pengaruh sebab-akibat.

b. Correlation (Korelasi): Bentuk kuantifikasi linear atau monoton dari asosiasi antara dua variabel numerik atau ordinal (misalnya menggunakan koefisien Pearson atau Spearman). korelasi murni mengukur kekuatan pola pergerakan linear, bukan mekanisme sebab-akibat.

c. Prediction (Prediksi): Penggunaan satu atau kumpulan variabel prediktor (X) untuk memperkirakan atau menebak nilai outcome (Y) pada tingkat individu atau populasi masa depan dengan akurasi setinggi mungkin. Prediktor yang sangat baik tidak harus menjadi penyebab terjadinya outcome.

d. Causal Effect (Efek Kausal): Perbedaan kesembuhan atau outcome (Y) yang akan benar-benar terjadi jika kita secara aktif mengintervensi atau mengubah nilai paparan (X) pada suatu subjek, dibandingkan jika subjek yang sama menerima kondisi pembanding lainnya.

ANALISIS KLINIS: Predictive Importance vs Causal Importance. *Sebuah penanda biologis (biomarker) seperti C-Reactive Protein (CRP) atau kadar laktat darah dapat menjadi prediktor mortalitas yang sangat akurat di ICU. Namun, hal ini tidak berarti bahwa CRP atau laktat adalah penyebab langsung kematian. Menurunkan nilai CRP secara paksa dengan obat anti-inflamasi tanpa mengobati infeksi bakteri di bawahnya tidak akan menurunkan risiko mortalitas pasien. Di sini, CRP memiliki arti prediktif tinggi, tetapi arti kausalnya nol. Sebaliknya, suatu faktor bisa saja memiliki efek kausal yang nyata tetapi kemampuan prediksi individualnya tampak lemah karena outcome klinis tersebut dipengaruhi oleh lusinan faktor fisiologis lainnya.*

5. Korelasi: Apa yang Diberitahukan dan Tidak Diberitahukan

Dalam pelaporan data klinis, koefisien korelasi Pearson (r) atau Spearman (rho) sangat sering ditampilkan untuk menunjukkan hubungan dua variabel kontinu. Pearson mengukur kekuatan hubungan linear, sedangkan Spearman bekerja menggunakan peringkat untuk mendeteksi hubungan monotonik.

PITFALL METODOLOGIS: Miskonsepsi Koefisien Korelasi. *Nilai korelasi r = 0,70 antara asupan garam dan tekanan darah sistolik sama sekali tidak berarti '70% dari kenaikan tekanan darah pasien disebabkan oleh konsumsi garam'. Nilai r juga tidak menjamin bahwa menurunkan asupan garam sebesar x gram akan menurunkan tekanan darah sebesar y mmHg pada pasien Anda. Korelasi murni mendeskripsikan kerapatan titik-titik data pada garis lurus, bukan kemiringan efek intervensi klinis.*

Klinisi juga harus waspada bahwa nilai korelasi yang mendekati nol tidak otomatis membuktikan 'tidak adanya hubungan apa pun' di antara kedua variabel tersebut. Hubungan yang bersifat non-linear (misalnya kurva berbentuk U pada hubungan antara tekanan darah sistolik dan mortalitas) dapat menghasilkan koefisien korelasi linear yang sangat kecil atau nol. Selain itu, kesalahan pengukuran (measurement error), pembatasan rentang data (restriction of range), atau penggabungan subkelompok pasien yang heterogen secara fisiologis dapat menyamarkan nilai korelasi riil. Oleh sebab itu, visualisasi data mentah melalui scatter plot selalu jauh lebih informatif daripada sekadar melaporkan satu angka koefisien korelasi statistik.

6. Temporal Ordering: Sebab Harus Mendahului Akibat

Secara logika keilmuan, syarat mutlak sebuah hubungan dapat diklaim sebagai sebab-akibat adalah **temporal ordering**, yaitu paparan atau intervensi (sebab) harus secara konseptual terjadi dan terukur mendahului terjadinya outcome (akibat). Desain studi potong-lintang (cross-sectional) mengukur paparan dan outcome pada satu titik waktu yang sama secara simultan, sehingga sangat sulit untuk menegaskan arah waktu hubungan ini.

> **CONTOH DALAM GIZI KLINIS: Asupan Energi dan Keparahan Penyakit.** *Sebuah studi cross-sectional di bangsal geriatri menemukan bahwa pasien lanjut usia dengan tingkat keparahan penyakit yang tinggi memiliki asupan energi dan protein yang jauh lebih rendah. Apakah asupan energi yang rendah tersebut yang menyebabkan penyakit mereka bertambah parah? Ataukah sebaliknya, kondisi penyakit yang memburuk secara sistemik menurunkan nafsu makan pasien sehingga asupan energinya menurun? Kedua hipotesis ini sama-sama konsisten dengan data statistik kasar yang diperoleh.*

Masalah pelik ini disebut sebagai ***reverse causation*** (sebab-akibat terbalik), di mana proses perjalanan penyakit atau kesudahan klinis sebenarnya memengaruhi paparan yang diukur, namun secara statistik tampak seolah-olah paparan tersebutlah yang menjadi faktor risiko timbulnya penyakit.

7. Confounding: *"Third Variable"* yang Bukan Sekadar Variabel Ketiga

Secara konseptual, perancu (confounding) terjadi ketika perbandingan kelompok yang terpapar dan tidak terpapar tidak memberikan kontras yang adil (comparability) untuk menilai efek kausal sesungguhnya, akibat adanya faktor luar yang memengaruhi baik pemilihan paparan maupun timbulnya outcome (common cause).

Mari kita kembalikan pada contoh kasus transfusi darah di ICU. Variabel keparahan penyakit awal (severity of illness) bertindak sebagai perancu utama karena memenuhi kriteria struktur kausal dasar:

- Severity memengaruhi kemungkinan pasien menerima transfusi darah (pasien yang lebih sakit parah/syok lebih sering ditransfusi).

- Severity secara independen memengaruhi risiko mortalitas (pasien yang lebih sakit parah memiliki peluang meninggal yang lebih besar, terlepas dari apakah mereka menerima transfusi darah atau tidak).
- Severity bukan merupakan akibat atau jalur perantara (mediator) dari tindakan transfusi darah itu sendiri.

> **MUTIARA KLINIS: Confounding by Indication.** *Perancu berdasarkan indikasi medis adalah tantangan terbesar dalam studi observasional yang menilai efektivitas terapi. Keputusan dokter untuk memberikan suatu pengobatan, dosis obat, atau tindakan operasi selalu didasarkan pada kondisi klinis spesifik pasien saat itu. Alasan klinis (indikasi) pemberian terapi ini hampir selalu berkaitan erat dengan prognosis pasien ke depan, sehingga mengacaukan efek terapi murni yang ingin diukur.*

8. Crude VS Adjusted Association

Untuk mengatasi confounding pada studi observasional, peneliti biasanya menggunakan analisis regresi multivariat untuk menghitung nilai estimasi yang telah disesuaikan (adjusted association). Mari kita perhatikan contoh visualisasi hipotetik di bawah ini:

Tabel 41. Contoh hipotetik perubahan estimasi odds ratio (or) sebelum dan sesudah adjustment

Model Analisis	Estimasi (OR)	95% Confidence Interval (CI)	Interpretasi Pedagogik Klinis
Model 1: Crude (Kasar)	**2,8**	1,9 - 4,1	Asosiasi kuat teramati sebelum disesuaikan dengan variabel lain. Risiko bias confounding sangat tinggi.
Model 2: Adjusted (Disesuaikan)	**1,6**	1,1 - 2,5	Asosiasi melemah setelah mengontrol kovariat terpilih (misal usia, severity). Confounding sebagian teratasi.

Perubahan nilai estimasi dari OR 2,8 menjadi adjusted OR 1,6 menunjukkan adanya pengaruh confounding yang cukup besar. Namun, klinisi harus menyadari bahwa perubahan angka ini tidak menjamin bahwa seluruh perancu telah dibersihkan sepenuhnya dari analisis. Masih terdapat

risiko adanya perancu sisa (residual confounding) akibat pengukuran variabel yang kurang presisi, atau perancu yang tidak terukur (unmeasured confounding) karena variabel perancu penting tersebut memang tidak dikumpulkan datanya oleh peneliti sejak awal.

> **DISIPLIN BAHASA: Bahasa Pelaporan yang Aman.** *Hindari mengklaim bahwa 'confounding sudah teratasi sepenuhnya dengan regresi'. Bahasa ilmiah yang jauh lebih aman dan jujur adalah: 'Setelah melakukan penyesuaian (adjustment) terhadap variabel usia, jenis kelamin, dan derajat keparahan penyakit awal, hubungan asosiasi yang teramati mengalami pelemahan, namun tetap bernilai positif signifikan.'*

9. Mengapa "Adjust Semua Variabel" Dapat Salah

Salah satu mitos paling berbahaya yang berkembang dalam analisis statistik kedokteran adalah anggapan bahwa model statistik akan semakin 'valid', 'sempurna', atau 'bebas bias' jika kita memasukkan sebanyak mungkin variabel kontrol ke dalam model regresi. Praktik mengontrol semua variabel baseline tanpa dasar teoretis yang jelas atau sekadar memasukkan variabel yang memiliki nilai $p < 0,25$ pada analisis univariat adalah kesalahan konseptual yang fatal.

Keputusan untuk menyesuaikan (adjust) suatu variabel tidak boleh didasarkan pada signifikansi statistik semata, melainkan harus didasarkan pada struktur kausal klinis yang masuk akal. Memasukkan variabel yang salah justru dapat menghilangkan efek kausal murni yang ingin kita ukur (over-adjustment bias), atau bahkan menciptakan hubungan asosiasi palsu yang baru (collider bias). Untuk memahami bahaya ini, kita harus membedakan dengan jelas tiga peran variabel pendamping: confounder, mediator, dan collider.

10. Mediator, Collider, dan Dag

a. Mediator (Variabel Perantara)

Mediator berada secara langsung di dalam jalur kausal yang menghubungkan paparan (X) menuju outcome (Y). Struktur kausalnya adalah: X → Mediator → Y. Sebagai contoh, jika kita meneliti efek dari program intervensi diet rendah kalori (X) terhadap penurunan tekanan darah sistolik (Y), maka penurunan berat badan pasien (M) bertindak sebagai mediator.

Jika tujuan klinis kita adalah mengetahui total effect (manfaat keseluruhan) dari program diet tersebut terhadap penurunan tekanan darah sistolik, maka kita tidak boleh mengontrol atau melakukan

adjustment terhadap penurunan berat badan (M). Jika kita memasukkan penurunan berat badan ke dalam model regresi, kita akan memblokir dan menghilangkan bagian efek utama yang justru ingin kita ukur, sehingga kesimpulan kita akan salah bahwa program diet 'tidak efektif'.

b. Collider (Variabel Benturan)

Collider adalah sebuah variabel yang merupakan akibat bersama dari dua variabel lainnya. Struktur kausalnya adalah: X → Collider ← Other Cause. Keunikan collider adalah: pada kondisi awal, X dan Other Cause sama sekali tidak memiliki hubungan asosiasi statistik apa pun. Namun, jika kita secara sengaja melakukan conditioning—baik dengan memasukkan collider ke model regresi atau menyeleksi sampel hanya pada kelompok subjek yang memenuhi kriteria collider tersebut—kita secara tidak sadar akan membuka jalur non-kausal (backdoor path) yang menciptakan hubungan asosiasi palsu yang menyesatkan antara X dan Other Cause.

CONTOH NYATA COLLIDER: Bias Seleksi Masuk ICU (Collider Bias). *Bayangkan jika masuk perawatan ICU (Collider) dipengaruhi oleh dua kondisi medis independen: penyakit paru obstruktif kronis / PPOK (X) dan kecelakaan lalu lintas parah (Other Cause). PPOK dan kecelakaan mobil jelas tidak memiliki hubungan sebab-akibat apa pun di populasi umum. Namun, jika seorang peneliti melakukan studi yang hanya merekrut pasien yang dirawat di ICU, maka peneliti akan menemukan korelasi negatif yang kuat seolah-olah 'pasien PPOK terlindung dari risiko kecelakaan lalu lintas'. Mengondisikan sampel pada collider ICU menciptakan bias asosiasi palsu.*

c. Directed Acyclic Graph (DAG) sebagai Alat Bantu Berpikir

Directed Acyclic Graph (DAG) adalah representasi grafis visual yang digunakan untuk menggambarkan asumsi terstruktur mengenai arah hubungan sebab-akibat antar-variabel (Greenland et al., 1999). DAG membantu peneliti klinis menetapkan variabel mana yang wajib dikendalikan (untuk menutup jalur perancu) dan variabel mana yang haram untuk disentuh (untuk menghindari over-adjustment dan bias collider).

INGAT: DAG Bukan Mesin Pembukti Otomatis. *Sebuah gambar DAG yang digambar dengan rapi menggunakan garis dan panah tidak dengan sendirinya membuktikan bahwa hubungan kausal tersebut benar-benar ada di dunia nyata. Arah panah harus murni berasal dari pemahaman patofisiologi biologis, urutan waktu logis, dan literatur klinis yang valid. DAG membantu mengeksplisitkan asumsi teoretis kita agar dapat*

diperdebatkan secara transparan, bukan membuktikan kausalitas secara empiris.

11. Confounder ≠ Mediator ≠ Collider

Tabel di bawah ini merangkum perbedaan mendasar ketiga peran variabel tersebut sebagai pedoman cepat bagi klinisi sebelum menyusun model statistik multivariat:

Tabel 42. Perbedaan konseptual dan konsekuensi analitik confounder, mediator, dan collider

Peran Variabel	Struktur Hubungan Kausal	Konsekuensi Tindakan Adjustment dalam Regresi
Confounder (Perancu)	X ← C → Y (C memengaruhi paparan X dan outcome Y)	WAJIB di-adjust / dikontrol untuk mendapatkan taksiran efek paparan yang adil dan bebas bias.
Mediator (Perantara)	X → M → Y (M berada pada jalur efek X ke Y)	HARAM di-adjust jika ingin mengukur total effect keseluruhan. Adjustment hanya dilakukan jika ingin mengukur direct effect.
Collider (Benturan)	X → Collider ← Other Cause (Collider dipengaruhi akibat bersama)	HARAM di-adjust atau diseleksi secara sepihak. Adjustment akan membuka bias asosiasi palsu yang baru.

12. RCT dan Studi Observasional Dalam Causal Inference

Uji klinis acak terkontrol (Randomized Controlled Trial / RCT) dipandang sebagai standar tertinggi dalam pembuktian efek kausal karena proses randomisasi secara probabilistik memastikan kelompok intervensi dan kontrol memiliki karakteristik baseline yang setara (exchangeability), baik untuk perancu yang disadari maupun yang tidak terukur. Namun, RCT di dunia nyata tidak pernah kebal dari keterbatasan implementasi klinis.

Masalah ketidakpatuhan pasien (non-adherence), kehilangan tindak lanjut (loss to follow-up), data yang hilang (missing outcomes), perpindahan kelompok terapi (crossover), dan bias pengukuran tetap dapat merusak validitas kausalitas dari sebuah RCT. Oleh sebab itu, pedoman pelaporan internasional trial modern (seperti CONSORT 2025) menekankan pentingnya pelaporan estimasi efek absolut, kepatuhan analisis rencana awal (intention-to-treat analysis), pelaporan efek merugikan (harms), serta batasan-batasan generalisasi secara jujur (Hopewell et al., 2025).

Sebaliknya, pada studi observasional (seperti kohort atau kasus-kontrol), karena penentuan paparan tidak dikontrol oleh peneliti melainkan berjalan secara natural, maka penarikan kesimpulan kausal sangat bergantung pada kekuatan desain dalam mengumpulkan data perancu yang lengkap serta pemenuhan asumsi-asumsi metodologi (von Elm et al., 2007).

13. Bahasa Kausal Yang Proporsional

Seorang dokter yang ilmiah harus menggunakan bahasa penulisan laporan penelitian yang proporsional dan tidak berlebihan (overclaim). Kekuatan kata kerja yang dipilih harus mutlak mencerminkan kekuatan desain penelitian di belakangnya:

Bahasa Asosiatif (Aman untuk Studi Observasional): 'Paparan X dikaitkan dengan risiko outcome Y yang lebih tinggi', 'Kadar biomarker X berasosiasi secara positif dengan...', 'Variabel X memprediksi kejadian Y.'

Bahasa Kausal (Hanya Sah untuk RCT dengan Implementasi Baik atau Studi Eksperimental): 'Tindakan intervensi X menurunkan angka mortalitas secara signifikan', 'Pemberian obat X menghambat terjadinya komplikasi Y.'

DISIPLIN PENULISAN:
Cegah Overclaim Kausal.
Kepatuhan penulisan istilah asing harus ditaati. Jangan pernah menggunakan kata kerja sebab-akibat langsung seperti 'meningkatkan', 'menyebabkan', atau 'menurunkan' saat Anda hanya memiliki data dari studi observasional cross-sectional atau kohort sederhana tanpa pembuktian teoretis kausal yang kokoh.

14. Studi Kasus dan Mini-Case

a. Dataset yang Sama dengan Pertanyaan yang Berbeda

Mari kita perhatikan bagaimana variabel biomarker inflamasi C-Reactive Protein (CRP) yang sama dapat dianalisis secara berbeda total bergantung pada pertanyaan penelitian klinis yang diajukan:

Tabel 43. Satu variabel (CRP) dengan beragam pertanyaan dan analisis konseptual

Pertanyaan Klinis Dokter	Target Analisis Statistik	Keluarga Analisis dan Kesimpulan Sah
Apakah nilai CRP yang tinggi berhubungan dengan mortalitas di ICU?	Crude / Adjusted Association	Membuktikan adanya pola asosiasi statistik kasat mata dalam data. Berguna untuk deskripsi awal.
Dapatkah nilai CRP membantu	Predictive Model dan Validation	CRP digunakan sebagai variabel prediktor klinis dalam

memperkirakan probabilitas kematian pasien?		model prediksi multivariat (misal skor risiko).
Apakah tindakan medis menurunkan kadar CRP akan menurunkan mortalitas?	Causal Inference / RCT	Membutuhkan desain uji klinis acak (RCT) intervensi penurun CRP untuk membuktikan kebenaran efek sebab-akibat.
Apakah efektivitas terapi X berbeda pada pasien dengan CRP tinggi vs rendah?	Effect Modification / Interaction	Menganalisis koefisien interaksi statistik (Treatment CRP) untuk menilai variasi manfaat klinis.

b. Latihan Mandiri: Selesaikan Mini-Case di Bawah Ini

Untuk melatih ketajaman berpikir klinis, bacalah keempat kasus hipotesis di bawah ini dan diskusikan jawaban teoretisnya:

1) **Mini-Case 1: Sepsis Berat dan Vasopresor.** Sebuah studi observasional kohort melaporkan bahwa pasien sepsis berat yang menerima vasopresor dosis tinggi memiliki risiko kematian yang jauh lebih besar. Penulis menyimpulkan vasopresor menyebabkan kematian dan harus dihindari. Apakah simpulan ini tepat?

   Jawaban Pembahasan: Tidak tepat. Kasus ini merupakan contoh klasik dari Confounding by Indication. Pasien dengan derajat syok sepsis yang jauh lebih berat secara klinis mutlak membutuhkan vasopresor dosis tinggi untuk bertahan hidup, sekaligus memiliki peluang kematian biologis yang memang sangat tinggi sejak awal. Vasopresor adalah indikasi keparahan, bukan penyebab kematian.

2) **Mini-Case 2: Nyeri Lutut dan Aktivitas Fisik.** Studi potong-lintang (cross-sectional) menemukan hubungan signifikan bahwa individu dengan aktivitas fisik rendah memiliki tingkat nyeri lutut yang lebih parah. Apakah aktivitas fisik rendah menyebabkan nyeri lutut?

   Jawaban Pembahasan: Tidak selalu. Desain studi yang cross-sectional mengukur kedua variabel pada satu waktu bersamaan, memicu bias Reverse Causation (sebab-akibat terbalik). Sangat besar kemungkinan bahwa nyeri lutut yang hebatlah yang membuat pasien malas bergerak dan membatasi aktivitas fisiknya.

3) **Mini-Case 3: Program Gaya Hidup dan Berat Badan.** Program olahraga intensif terbukti menurunkan tekanan darah sistolik. Peneliti memasukkan variabel penurunan berat badan ke dalam model regresi,

lalu menyimpulkan 'olahraga tidak berefek terhadap tekanan darah setelah dikontrol berat badan'. Apa kekeliruan analisis ini?

Jawaban Pembahasan: Peneliti melakukan Over-adjustment bias. Penurunan berat badan bertindak sebagai Mediator (jalur antara) bagaimana olahraga bekerja secara fisiologis untuk menurunkan tekanan darah. Mengontrol mediator berarti menghancurkan efek total intervensi utama yang sedang dievaluasi.

4) **Mini-Case 4: Rekrutmen Sampel Hanya Pasien ICU.** Peneliti ingin melihat hubungan merokok dan kanker paru, tetapi seluruh sampel penelitian hanya diambil dari pasien yang dirawat di ICU. Mengapa desain ini bermasalah?

Jawaban Pembahasan: Ini adalah bentuk bias seleksi akibat pengondisian terhadap Collider (rawat inap ICU). admission ICU dipengaruhi secara independen oleh banyak kondisi berat yang berkaitan dengan merokok maupun kanker, sehingga korelasi di dalam ICU tidak akan pernah mencerminkan hubungan biologis yang sebenarnya di populasi umum.

15. Mitos dan Klarifikasi

Berikut adalah beberapa mitos konseptual yang sangat sering beredar di kalangan akademisi kedokteran beserta klarifikasi ilmiahnya:

Tabel 44. Mitos yang umum beredar mengenai kausalitas dan klarifikasinya

Mitos / Pendapat yang Keliru	Klarifikasi Ilmiah yang Benar
"Jika analisis regresi multivariat saya sudah menunjukkan p < 0,05, maka hubungan tersebut pasti bersifat kausal."	Salah. Regresi multivariat hanya menghasilkan adjusted association (asosiasi yang disesuaikan). Kesimpulan kausal murni membutuhkan kebenaran struktur teoretis klinis dan desain studi yang kuat.
"Semua variabel demografi baseline yang berbeda nyata pada uji bivariat wajib dimasukkan sebagai perancu (adjust) di regresi."	Tidak selalu. Pengendalian variabel harus dituntun oleh struktur kausal (DAG). Memasukkan mediator atau collider ke model justru akan merusak estimasi dan melahirkan bias baru.
"Variabel yang memiliki nilai p > 0,05 pada uji statistik kasar (crude) terbukti bukan merupakan perancu (non-confounder)."	Salah besar. Status perancu ditentukan oleh peran konseptualnya sebagai common cause (sebab bersama), bukan oleh status nilai signifikansi statistik (p-value) di dalam dataset sampel.
"Penelitian dengan jumlah sampel yang sangat besar (n = 100.000) otomatis terbebas	Sama sekali tidak. Ukuran sampel yang besar hanya akan mempertinggi presisi statistik (mempersempit CI dan SE), tetapi tidak akan

Mitos / Pendapat yang Keliru	Klarifikasi Ilmiah yang Benar
dari kesalahan bias perancu."	pernah mampu memperbaiki atau menghilangkan bias desain.
"Uji korelasi tidak memiliki manfaat klinis apa pun karena tidak membuktikan adanya hubungan sebab-akibat."	Korelasi tetap sangat berguna untuk mendeskripsikan keterikatan awal data (description) dan membantu memformulasikan hipotesis eksperimental baru di masa mendatang.

16. Checklist 8 Pertanyaan Sebelum Menggunakan Bahasa Kausal

Sebelum Anda menuliskan kalimat kesimpulan akhir atau menggunakan kata kerja bernuansa sebab-akibat, periksalah draf manuskrip Anda menggunakan delapan daftar pertanyaan kritis di bawah ini:

a. Question (Pertanyaan Inti): Apakah saya sedang menanyakan hubungan asosiasi, akurasi prediksi, atau benar-benar menginginkan efek kausal dari intervensi?
b. Time (Urutan Waktu): Apakah paparan (exposure) yang diteliti terbukti benar-benar diukur mendahului terjadinya outcome klinis?
c. Comparability (Keterbandingan): Mengapa sebagian pasien menerima paparan tersebut dan sebagian lagi tidak? Apakah alasan tersebut memengaruhi prognosis?
d. Confounding (Perancu): Apakah ada sebab bersama (common cause) masuk akal lainnya yang dapat memengaruhi baik paparan maupun outcome secara simultan?
e. Adjustment (Penyesuaian): Apakah variabel-variabel yang dimasukkan ke dalam model regresi sudah tepat sebagai perancu (bukan mediator atau collider)?
f. Selection (Seleksi Sampel): Apakah kriteria masuk sampel penelitian berisiko menyeleksi subjek berdasarkan akibat bersama (collider bias)?
g. Residual Bias (Bias Sisa): Apakah masih ada kemungkinan bias sisa atau perancu yang tidak terukur yang dapat menerangkan hubungan asosiasi teramati?
h. Language (Pemilihan Kata): Apakah kata kerja yang saya gunakan sudah aman dan proporsional sesuai dengan desain dan batas kekuatan studi?

17. Rangkuman

a. Asosiasi statistik (association) bukanlah bukti adanya hubungan sebab-akibat (causation).

b. Akurasi prediksi individu (prediction) dapat berjalan sangat baik tanpa membutuhkan adanya mekanisme kausal langsung pada variabel prediktornya.
c. Temporal ordering (urutan waktu) adalah fondasi konseptual wajib yang menyatakan bahwa sebab harus mendahului akibat.
d. Perancu (confounding) berkaitan erat dengan masalah keterbandingan kelompok (comparability) akibat adanya sebab bersama (common cause).
e. Nilai estimasi yang disesuaikan (adjusted estimate) tidak pernah secara otomatis setara dengan nilai efek kausal (causal estimate).
f. Mediator berada di jalur efek langsung; mengontrol mediator (over-adjustment) akan memblokir dan mendistorsi nilai manfaat total intervensi.
g. Collider adalah akibat bersama; mengondisikan collider (selection bias) akan melahirkan hubungan asosiasi palsu yang baru.
h. Directed Acyclic Graph (DAG) adalah peta jalan visual untuk memperjelas asumsi kausalitas teoretis, bukan pembuktian empiris otomatis.
i. Ukuran sampel yang sangat besar hanya akan mempersempit ketidakpastian (precision), tetapi tidak akan pernah mampu meniadakan bias confounding.
j. Bila data tidak mencukupi untuk penarikan kesimpulan kausal, jawaban yang paling ilmiah dan bertanggung jawab adalah mengakui adanya keterbatasan tersebut.

18. Latihan dan Evaluasi

Latihan Formatif (Pilihan Ganda)

1) Variabel X terbukti secara statistik sangat kuat memprediksi risiko kejadian mortalitas di ICU pada analisis regresi. Kesimpulan klinis yang paling tepat dan aman adalah:
   a. Variabel X terbukti merupakan penyebab langsung terjadinya mortalitas di ICU.
   b. Menurunkan nilai variabel X secara agresif dipastikan akan menurunkan risiko kematian pasien.
   c. Variabel X memiliki nilai kegunaan prediktif klinis, tanpa harus bertindak sebagai penyebab timbulnya mortalitas.
   d. Variabel X merupakan collider murni yang tidak boleh digunakan dalam analisis klinis.

Kunci Jawaban: C

Pembahasan: Kemampuan prediksi (predictive value) berfokus pada seberapa baik informasi X membantu menebak Y. Suatu biomarker dapat menjadi prediktor yang sangat baik (seperti CRP) karena berkaitan dengan proses inflamasi sistemik, tanpa harus menjadi penyebab patofisiologis langsung dari kematian.

2) Variabel C secara biologis terbukti menyebabkan timbulnya paparan X dan secara independen memicu terjadinya outcome Y. Dalam diagram struktur kausal sederhana X ← C → Y, maka variabel C bertindak sebagai:
   a. Mediator
   b. Confounder (Perancu)
   c. Collider
   d. Kesudahan klinis (Outcome)

Kunci Jawaban: b

Pembahasan: Variabel C bertindak sebagai perancu (confounder) karena merupakan sebab bersama (common cause) dari paparan X dan outcome Y. Penyesuaian terhadap variabel C wajib dilakukan dalam model regresi.

3) Jika variabel M terbukti berada pada jalur kausal antara paparan dan outcome (X → M → Y), dan target klinis kita adalah mengetahui total effect (manfaat keseluruhan) dari X terhadap Y, maka tindakan kita terhadap variabel M adalah:
   a. M wajib dimasukkan ke dalam model regresi sebagai variabel penyesuai.
   b. Penyesuaian terhadap M tidak boleh dilakukan karena akan menghilangkan sebagian efek utama paparan yang ingin diukur.
   c. M harus dihapus dari seluruh dataset karena merusak validitas data.
   d. Menghitung korelasi bivariat antara M dan Y terlebih dahulu.

Kunci Jawaban: b

Pembahasan: Variabel M bertindak sebagai mediator (perantara). Jika tujuan kita adalah menilai total effect, melakukan adjustment terhadap mediator (over-adjustment) akan memblokir sebagian jalur transmisi efek fisiologis, sehingga menipu analisis seolah-olah intervensi utama X tidak berkhasiat.

4) Sebuah penelitian potong-lintang (cross-sectional) melaporkan adanya hubungan statistik yang kuat di mana pasien dengan tingkat asupan

energi rendah memiliki keparahan penyakit kronis yang lebih parah. Isu metodologis utama yang harus diwaspadai di sini adalah:

a. Reverse Causation (Sebab-Akibat Terbalik)
b. Confounding by Indication
c. Kehilangan tindak lanjut (Attrition bias)
d. Kesalahan klasifikasi paparan gizi

Kunci Jawaban: a

Pembahasan: Karena kedua variabel diukur secara simultan, sangat sulit menentukan mana sebab dan akibat. Penyakit yang bertambah parah dapat menyebabkan hilangnya nafsu makan dan asupan gizi pasien menurun (sebab-akibat terbalik).

5) Manakah dari pernyataan di bawah ini yang paling mencerminkan prinsip pelaporan biostatistika kedokteran yang jujur dan bertanggung jawab?

a. Nilai adjusted OR 2,1 membuktikan secara pasti adanya efek kausal obat terhadap kesembuhan.'
b. Penelitian kohort observasional ini bebas dari bias karena ukuran sampel yang sangat besar.'
c. Interpretasi kausal selalu membutuhkan kesesuaian desain, urutan waktu, dan kebenaran asumsi struktur kausal klinis, bukan sekadar pelaporan adjusted OR regresi.'
d. Variabel pengganggu baseline yang tidak memiliki nilai $p < 0,05$ tidak akan pernah memicu bias.'

Kunci Jawaban: c

Pembahasan: Nilai statistik hanyalah model di bawah asumsi tertentu. Pembuktian kausalitas membutuhkan evaluasi kualitas metodologi desain secara menyeluruh, bukan sekadar melihat p-value koefisien regresi.

BAB 6
REGRESI, CONFOUNDING, DAN INTERACTION

Sistematisasi Analisis Multivariat dan Interpretasi Skala Efek

1. Pendahuluan

Bab ini merupakan kelanjutan alami dari prinsip *causal thinking* yang telah diletakkan pada Bab 5 (*Association ≠ Causation*). Di dunia nyata medis, variabel klinis jarang berdiri sendiri; penyakit dipengaruhi oleh konstelasi faktor biologis, perilaku, dan lingkungan yang saling berinteraksi. Model regresi hadir sebagai bahasa formal untuk mendeskripsikan dan mengestimasi hubungan-hubungan tersebut secara simultan di bawah asumsi dan struktur data tertentu.

Dalam penelitian kedokteran, analisis regresi multivariat (*multivariable analysis*) sering digunakan untuk mengendalikan faktor perancu (*confounding adjustment*) yang didasarkan pada pertimbangan klinis substantif. Namun, sangat penting untuk dipahami bahwa hasil penyesuaian matematika (*adjusted estimate*) tidak serta-merta mengubah hubungan asosiatif menjadi kausalitas murni. Panduan pelaporan STROBE menegaskan pentingnya menjelaskan secara eksplisit alasan di balik pemilihan variabel pengganggu, serta melaporkan estimasi efek tanpa klaim kausalitas yang melampaui batas desain riset.

> **PRINSIP UTAMA BAB INI:**
> *Model regresi bukanlah sekadar "alat pengolah data" untuk memasukkan sebanyak mungkin variabel ke dalam komputer demi mendapatkan p-value yang signifikan. Regresi adalah model struktural untuk mengestimasi ukuran efek (effect size) pada skala tertentu. Setiap adjustment harus berlandaskan alasan klinis yang matang, dan interaksi (interaction) harus diinterpretasikan secara hati-hati pada skala efek yang sesuai.*

2. Capaian Pembelajaran

Setelah menyelesaikan bab ini, mahasiswa kedokteran diharapkan mampu:

- Menjelaskan filosofi regresi sebagai model estimasi hubungan klinis, bukan sekadar prosedur uji signifikansi nilai p.

- Membedakan konsep crude estimate (estimasi kasar) dan adjusted estimate (estimasi yang disesuaikan) serta menguraikan alasan perubahan nilainya secara klinis.
- Memilih keluarga model regresi yang sesuai berdasarkan jenis outcome data (Linear, Logistik, Count/Poisson, Cox Survival, dan Longitudinal/Clustered Data).
- Menafsirkan koefisien regresi secara proporsional sesuai dengan skala model (koefisien beta, Odds Ratio, Hazard Ratio, dan Rate Ratio) dan coding variabelnya.
- Menjelaskan mengapa adjusted estimate tidak otomatis merepresentasikan efek kausal yang bebas bias.
- Membedakan secara konseptual peran variabel pengganggu (confounding) dari modifikasi efek (interaction / effect modification).
- Menjelaskan keterbatasan pemodelan seperti nonlinearity, collinearity, data sparsity, dan overfitting secara konseptual.
- Menafsirkan model data longitudinal (Mixed-effects model dan GEE) serta perbedaan interpretasi subject-specific vs. population-average.
- Menuliskan hasil laporan regresi multivariat secara jujur, objektif, dan relevan bagi penatalaksanaan pasien.

3. Peta Konsep dan Clinical Trigger: Mengapa or Berubah?

Bayangkan sebuah studi kohort observasional yang meneliti hubungan antara **obesitas** (sebagai exposure) dan kejadian **hipertensi** (sebagai outcome) pada 5.000 subjek dewasa. Peneliti membangun tiga model analisis secara bertahap, dan memperoleh hasil estimasi kekuatan hubungan (Odds Ratio / OR) yang berbeda-beda seperti disajikan pada tabel di bawah ini:

Tabel 45. Evolusi Perubahan Estimasi Kekuatan Hubungan (Odds Ratio) Hipertensi

Model Analisis	Estimasi OR	95% Confidence Interval	Keterangan Klinis
Model 1: Crude (Kasar)	2,80	1,90 – 4,10	Hubungan bivariat mentah tanpa penyesuaian.
Model 2: Adjusted Usia + Seks	2,10	1,40 – 3,20	OR melemah setelah mengontrol faktor demografi dasar.
Model 3: Adjusted Usia + Seks + Aktivitas + DM	1,65	1,05 – 2,60	OR semakin mendekat ke nilai null (1,0) setelah variabel perilaku dikontrol.

Pertanyaan Kritis untuk Penalaran Klinis:

a. Apakah penurunan nilai OR dari 2,80 menjadi 1,65 menunjukkan bahwa Model 3 'paling benar' secara mutlak?
b. Apakah semua variabel yang ditambahkan dalam Model 3 pasti bertindak sebagai faktor perancu (confounder)?
c. Apakah adjusted OR 1,65 tersebut secara otomatis mewakili efek kausal dari obesitas terhadap hipertensi?

Jawaban atas pertanyaan di atas menuntut pemahaman mendalam. Penurunan nilai estimasi (pergeseran koefisien) menunjukkan adanya perubahan informasi akibat pengkondisian variabel dalam model. Namun, tanpa pengetahuan substantif tentang desain, waktu pengukuran (temporal ordering), kualitas data, dan struktur hubungan biologis riil (apakah variabel pengganggu tersebut merupakan confounder, mediator, atau collider), penambahan variabel secara mekanis bisa berisiko melahirkan kesimpulan yang salah arah.

4. Filosofi Regresi

Regresi secara mendasar bukanlah sebuah ritual statistik untuk menguji hipotesis nihil (H0) lewat p-value. Regresi adalah sebuah kerangka kerja matematis (*estimation framework*) untuk menggambarkan bagaimana distribusi dari variabel hasil (*outcome/dependent variable*) berubah atau bergeser secara sistematis sebagai fungsi dari satu atau beberapa variabel prediktor (*exposure/independent variable*).

Bentuk fungsional model regresi ini sangat bervariasi tergantung pada karakteristik skala pengukuran outcome, fungsi penghubung (*link function*), distribusi peluang sisaan (residual), dan struktur korelasi antar-observasi. Oleh karena itu, istilah 'regresi' merujuk pada rumpun keluarga model statistik yang luas, bukan sekadar uji regresi linear sederhana (Harrell, 2015; Vittinghoff et al., 2012).

5. Outcome Menentukan Keluarga Model

Pemilihan model statistik tidak boleh didasarkan pada coba-coba (*trial and error*) melainkan dituntun secara logis oleh tipe dari variabel outcome utama riset Anda. Tabel berikut menyajikan peta konseptual untuk menjodohkan tipe outcome dengan keluarga model regresi yang tepat:

Tabel 46. Panduan pemilihan keluarga model regresi berdasarkan tipe outcome

Tipe OutcomeTipe Outcome	Model RegresiModel Regresi	Ukuran Efek LazimUkuran Efek Lazim	Pertanyaan Klinis IntiPertanyaan Klinis Inti
Kontinu	Linear Regression	Koefisien β (Mean Difference)	Berapa perbedaan rata-rata perubahan SBP untuk setiap perbedaan prediktor?
Biner (Ya/Tidak)	Logistic Regression	Odds Ratio (OR)	Bagaimana perbandingan odds kejadian penyakit stroke menurut status obesitas?
Hitungan (Count)	Poisson / Negative Binomial	Rate Ratio (IRR)	Berapa laju kejadian eksaserbasi asma per tahun jika pasien menggunakan inhaler baru?
Time-to-Event	Cox Proportional Hazards	Hazard Ratio (HR)	Bagaimana rasio laju mortalitas sesaat (hazard) pasien kanker antar regimen terapi?
Berulang (Longitudinal)	Mixed-Effects / GEE	Subject-Specific / Population-Average	Bagaimana pola trajektori penurunan kadar HbA1c pasien selama 12 bulan follow-up?

6. Linear Regression: Apa Arti B?

Pada model regresi linear multivariat, koefisien beta (**β**) mengukur estimasi perbedaan rata-rata outcome klinis untuk setiap perbedaan satu satuan unit variabel prediktor, dengan asumsi seluruh variabel kovariat lain di dalam model dipertahankan konstan secara matematis (*all other things held constant*).

Tabel 47. Contoh interpretasi koefisien β pada model linear

Struktur Model	Nilai β Estimasi	Interpretasi Klinis yang Tepat dan Hati-hati
SBP ~ Usia (tahun)	+0,55 mmHg / tahun	Setiap perbedaan usia 1 tahun berkaitan dengan rata-rata SBP sekitar 0,55 mmHg lebih tinggi pada populasi studi ini.
HbA1c akhir ~	−0,60%	Dengan mengondisikan kadar

Struktur Model	Nilai β Estimasi	Interpretasi Klinis yang Tepat dan Hati-hati
Intervensi + Baseline HbA1c		baseline HbA1c yang setara, kelompok intervensi memiliki kadar HbA1c rata-rata 0,60% lebih rendah dibanding kontrol.

7. Logistic Regression: Odds Bukan Risiko

Ketika outcome penelitian bertipe biner (misalnya Hidup vs. Meninggal, atau Sembuh vs. Tidak), regresi logistik memodelkan logaritma natural dari odds (*log-odds*). Setelah dieksponensialkan, koefisien model dilaporkan sebagai Odds Ratio (OR).

Salah satu miskonsepsi paling mendarah daging di kalangan klinisi adalah menyamakan Odds Ratio dengan Risk Ratio (RR). Odds membandingkan probabilitas terjadinya event dengan probabilitas tidak terjadinya event ($p / (1 - p)$), sedangkan risiko membandingkan event dengan keseluruhan total subjek (p). Jika prevalensi penyakit dalam populasi cukup tinggi (>10%), OR akan overestimate—nilainya tampak jauh lebih ekstrem menjauhi nilai null (1,0) dibanding RR yang sebenarnya. Oleh karena itu, jangan menerjemahkan OR secara serampangan sebagai 'peningkatan risiko sekian kali lipat'!

Tabel 48. Panduan interpretasi output regresi logistik

Keluaran Model (Output)	Interpretasi Klinis yang Benar
Adjusted OR 1,80 (95% CI: 1,20 – 2,70)	Pada model yang dispesifikasikan, setelah menyesuaikan kovariat lain, odds kejadian stroke diperkirakan 1,8 kali lebih tinggi pada kelompok obesitas dibanding kontrol.
Adjusted OR 0,65 (95% CI: 0,48 – 0,89)	Faktor paparan dikaitkan dengan odds outcome yang lebih rendah (efek protektif). Makna klinisnya harus ditimbang bersama risiko dasar (baseline risk) dan desain studi.

8. Count Outcomes: Poisson dan Negative Binomial

Data hitungan (count) diskrit non-negatif seperti jumlah kunjungan rawat inap, frekuensi episode hipoglikemia, atau jumlah eksaserbasi dalam periode setahun tidak cocok dianalisis dengan model linear biasa (karena tidak mungkin bernilai negatif dan cenderung miring/skewed). Regresi Poisson memodelkan laju kejadian (*incident rate*) dengan memasukkan

unsur waktu paparan (*exposure time/offset*) dari masing-masing pasien. Jika variasi data melebihi rata-ratanya (*overdispersion*), model Negative Binomial harus digunakan agar estimasi ketidakpastian (confidence interval) tidak menjadi terlalu sempit secara semu.

9. Cox Regression: Hazard Ratio Bukan Risk Ratio

Pada analisis survival, outcome dinilai berdasarkan aspek 'apakah' event terjadi dan 'kapan' event terjadi, lengkap dengan adanya data tersensor (*censored data*). Model Cox Proportional Hazards menghasilkan Hazard Ratio (HR) yang membandingkan laju kejadian sesaat (hazard) pada waktu t antar dua kelompok.

Penting untuk diingat bahwa HR tidak sama dengan Risk Ratio (RR) akumulatif. HR mengasumsikan rasio hazard konstan sepanjang waktu pengamatan (*proportional hazards assumption*). Jika kurva survival saling berpotongan (*crossing survival curves*), hal ini menunjukkan adanya pelanggaran asumsi proportional hazards (misalnya akibat efek obat yang lambat atau toksisitas awal), sehingga interpretasi satu nilai HR tunggal bisa menyesatkan klinisi.

10. Crude Vs Adjusted: Apa yang Berubah?

Perubahan nilai estimasi efek dari crude (tidak disesuaikan) menjadi adjusted (disesuaikan) setelah memasukkan variabel lain ke dalam model multivariat memiliki beberapa skenario penjelasan klinis:

Tabel 49. Alasan di balik perubahan estimasi efek setelah adjustment

Skenario Perubahan	Mekanisme Metodologis	Konsekuensi Klinis dan Tindakan
Confounding Berkurang	Variabel yang dimasukkan adalah confounder sejati.	Estimasi adjusted lebih mendekati efek kausal yang sebenarnya.
Over-adjustment Bias	Variabel yang dimasukkan adalah mediator (jalur antara).	Menghapus sebagian efek intervensi yang ingin kita ukur. Jangan di-adjust.
Collider Bias (Selection Bias)	Variabel yang dimasukkan adalah collider (akibat dari dua hal).	Menciptakan hubungan semu yang menyesatkan. Jangan dimasukkan ke model.
Data Sparsity / Missing Data	Adanya missing data mengubah sampel subjek yang dianalisis.	Perubahan nilai bisa jadi akibat bergesernya populasi subjek, bukan karena penyesuaian perancu.

11. Confounder, Mediator, dan Collider dalam Model

Untuk menyistematiskan cara membangun model regresi, klinisi harus mengenal peran masing-masing variabel berdasarkan arah waktu (temporal) dan logika biologis:

Tabel 50. Ringkasan peran dan tindakan kovariat dalam model regresi

Jenis Variabel	Struktur Hubungan	Konsekuensi Bila Di-adjust	Keputusan Klinis Ideal
Confounder	C → X dan C → Y	Mengurangi bias perancu, mendekatkan estimasi ke nilai riil.	Wajib dimasukkan ke model (adjustment).
Mediator	X → M → Y	Menghilangkan total effect, menyisakan direct effect saja.	Jangan di-adjust jika tujuannya mengukur total efek intervensi.
Collider	X → C ← Y	Membuka jalur bias non-kausal (selection bias).	Wajib dikeluarkan dari model (jangan di-adjust).

12. Interaction dan Effect Modification: Ketika Efek Tidak Sama

Interaksi atau modifikasi efek (*effect modification*) terjadi ketika kekuatan atau arah hubungan antara exposure utama dan outcome berbeda secara signifikan pada strata level variabel lain (modifier). Namun, besarnya interaksi ini sangat bergantung pada skala yang dipilih: skala aditif (selisih risiko) atau skala multiplikatif (rasio risiko) (Knol dan VanderWeele, 2012).

Tabel 51. Contoh Kasus: Rasio Risiko Sama, tetapi manfaat absolut berbeda

Subgrup Pasien	Risiko Kontrol	Risiko Terapi	Risk Difference (Absolut)	Risk Ratio (Relatif)
Risiko Dasar Rendah	5%	4%	−1% (ARR = 1%)	0,80 (Penurunan 20%)
Risiko Dasar Tinggi	30%	24%	−6% (ARR = 6%)	0,80 (Penurunan 20%)

Meskipun secara relatif (multiplikatif) efek terapi sama (RR 0,80), secara absolut (aditif) manfaat klinisnya jauh lebih besar pada pasien berisiko dasar tinggi (ARR 6% vs. 1%). Ini menunjukkan mengapa

penulisan laporan interaksi sebaiknya tidak hanya mengandalkan nilai p interaksi multiplikatif dari software regresi logistik, melainkan harus menyajikan estimasi per strata secara transparan.

Jebakan Analisis Subgrup: Kesalahan fatal yang sering ditemui adalah menyatakan interaksi terjadi karena 'efek signifikan pada laki-laki (p=0,02) tetapi tidak signifikan pada perempuan (p=0,20)'. Perbedaan antara dua nilai p-value bukanlah bukti formal adanya interaksi. Uji interaksi harus dilakukan dengan membandingkan parameter antar-subgrup secara langsung menggunakan variabel interaksi (*treatment × subgroup*) di dalam model regresi.

13. Nonlinearity, Collinearity, Sparsity, dan Overfitting

Sebelum mempublikasikan model regresi multivariat Anda, klinisi harus mewaspadai empat keterbatasan kestabilan data:

- Nonlinearity: Mengasumsikan hubungan usia dan tekanan darah berbentuk garis lurus linear bisa jadi keliru. Perubahan risiko usia 30 ke 31 tahun tentu tidak setara dengan usia 80 ke 81 tahun. Gunakan fungsi spline jika hubungan kontinunya melengkung.
- Collinearity: Keadaan di mana dua prediktor saling berkorelasi sangat kuat (misalnya berat badan dan BMI). Ini menyebabkan estimasi koefisien menjadi tidak stabil dan confidence interval melebar secara ekstrem.
- Data Sparsity: Terjadi bila kombinasi antara exposure dan outcome biner di dalam sel tabel kontingensi sangat sedikit. Ini memicu estimasi OR yang bias dan bernilai tidak masuk akal (separation).
- Overfitting: Keadaan di mana model terlalu kompleks (memasukkan terlalu banyak variabel) dibanding jumlah informasi/event yang tersedia. Sebagai aturan umum praktis, model regresi logistik atau Cox membutuhkan minimal 10 hingga 15 kejadian outcome (events) untuk setiap 1 variabel prediktor bebas yang dimasukkan (events per variable).

14. Longitudinal Dan Clustered Data

Jika satu subjek pasien diukur berulang kali sepanjang waktu (longitudinal) atau pasien-pasien berkelompok dalam satu kluster rumah sakit (clustered data), observasi di dalam subjek tersebut cenderung saling berkorelasi (*intra-class correlation*). Menganalisis data ini menggunakan regresi logistik atau linear biasa sangat dilarang karena melanggar asumsi independensi dan akan menghasilkan standar error yang terlalu sempit secara semu. Dua model kandidat utama yang digunakan:

- **Mixed-Effects Models (Random Effects):** Memodelkan heterogenitas antar-subjek. Menghasilkan interpretasi *subject-specific (conditional)* (misal: bagaimana peluang kesembuhan seorang pasien individu berubah sepanjang waktu).
- **Generalized Estimating Equations (GEE):** Memodelkan rata-rata populasi dengan matriks korelasi kerja. Menghasilkan interpretasi *population-average (marginal)* (misal: bagaimana proporsi kesembuhan rata-rata dari populasi pasien berubah sepanjang waktu).

15. Decision Framework dan Checklist Membaca Tabel Regresi

Tabel 52. Checklist 10 pertanyaan kritis sebelum membaca tabel regresi

Pertanyaan Kritis	Aspek Evaluasi	Panduan untuk Pembaca Kedokteran
Outcome	Apa outcome studi?	Apakah tipe data kontinu, biner, count, survival, atau longitudinal?
Model	Apa keluarga model?	Apakah model regresi linear, logistik, Poisson, Cox, atau mixed-effects?
Skala	Apa ukuran efeknya?	Apakah dilaporkan sebagai β, Odds Ratio (OR), Hazard Ratio (HR), atau IRR?
Kovariat	Mengapa di-adjust?	Apakah penyesuaian didasarkan teori klinis (DAG) atau sekadar signifikansi bivariat?
Struktur	Bagaimana struktur data?	Apakah observasi independen atau memiliki korelasi repeated/clustered?
Interaksi	Adakah interaksi?	Apakah efek paparan berbeda menurut kelompok lain? Pada skala apa?
Presisi	Bagaimana ketidakpastian?	Apakah interval kepercayaan 95% (CI) sempit (presisi tinggi) atau lebar?
Kestabilan	Adakah overfitting?	Apakah jumlah event memadai untuk jumlah variabel yang dianalisis?
Kausalitas	Apakah klaim kausal?	Apakah desain studi benar-benar mendukung klaim sebab-akibat (causation)?
Relevansi	Apa maknanya bagi pasien?	Apakah besarnya magnitude perbedaan/perubahan ini bermakna secara klinis?

16. Rangkuman

a. Regresi adalah model estimasi efek klinis secara sistematis, bukan mesin penghasil *p-value*.
b. Variabel outcome utama penelitian secara mutlak menentukan keluarga model regresi yang akan dibangun.
c. Crude estimate dan adjusted estimate menjawab pertanyaan statistik yang berbeda; bergesernya nilai tidak otomatis membuktikan keberadaan perancu sejati.
d. Model yang memasukkan lebih banyak variabel kovariat belum tentu lebih valid *(overfitting)*. Penyelarasan variabel pengganggu *(confounding adjustment)* membutuhkan dasar clinical thinking yang matang.
e. Hindari menyamakan Odds Ratio (OR) hasil regresi logistik atau Hazard Ratio (HR) hasil regresi Cox sebagai Risiko Relatif (Risk Ratio / RR) langsung.
f. Uji interaksi formal harus dilakukan dengan membandingkan parameter antar strata secara simultan, bukan dengan membandingkan p-value terpisah di tiap subgrup.
g. Penyimpangan data seperti nonlinearity, collinearity, data sparsity, dan overfitting harus dinilai secara diagnostik sebelum kesimpulan ditarik.
h. Data longitudinal atau clustered membutuhkan pemodelan khusus (Mixed-effects model atau GEE) yang memodelkan korelasi intra-subjek secara tepat.

17. Latihan dan Evaluasi

1) Koefisien treatment (β) pada model regresi linear multivariat paling tepat didefinisikan sebagai:
   a. Probabilitas kesembuhan pasien setelah diobati.
   b. Estimasi perbedaan rata-rata outcome antar kelompok dengan mengondisikan variabel kovariat lain konstan.
   c. Rasio risiko relatif kejadian efek samping terapi.
   d. Bukti kausal mutlak efektivitas obat.

Kunci Jawaban: b

Pembahasan Rasional: Pada regresi linear, koefisien β mengukur perbedaan rata-rata (mean difference) outcome kontinu yang dikondisikan (conditional) terhadap variabel lain di dalam model regresi.

2) Sebuah studi kohort sepsis di ICU melaporkan adjusted OR mortalitas 30-hari untuk status renta (frailty) adalah 2,20 (95% CI: 1,35 – 3,59). Interpretasi yang paling akurat adalah:

a. Pasien renta memiliki risiko kematian 2,2 kali lipat lebih tinggi dibanding pasien tidak renta.
b. Setelah mengontrol kovariat, odds mortalitas 30 hari pasien renta diperkirakan 2,2 kali odds pasien tidak renta.
c. Probabilitas kematian pasien renta adalah sebesar 220%.
d. Model ini pasti bias karena tidak memiliki nilai p-value.

Kunci Jawaban: b

Pembahasan Rasional: Odds ratio membandingkan odds kejadian biner (p/(1-p)), bukan risiko (probability) langsung. Menyatakan OR sebagai 'risiko naik sekian kali lipat' adalah salah kaprah statistik, terutama pada kasus ICU dengan prevalensi kematian yang tinggi.

3) Mengapa memasukkan variabel mediator (jalur antara) ke dalam model adjustment regresi dianggap sebagai kesalahan metodologis jika tujuannya mengukur total efek?
a. Karena mediator akan memicu multikolinearitas murni.
b. Karena penyesuaian terhadap mediator akan menghapus sebagian efek intervensi yang ingin kita ukur (over-adjustment bias).
c. Karena mediator selalu mengubah tipe data kontinu menjadi biner.
d. Karena mediator tidak memiliki p-value

Kunci Jawaban: b

Pembahasan Rasional: Mediator (M) berada di tengah jalur kausal (X → M → Y). Menyesuaikan (adjusting) mediator akan memotong transmisi efek dari X ke Y, sehingga total efek intervensi tidak dapat terukur secara utuh.

4) Manakah pernyataan yang paling tepat mengenai analisis subgrup dan uji interaksi?
a. Interaksi nyata terjadi jika efek signifikan pada subgrup A ($p<0,05$) tetapi tidak signifikan pada subgrup B ($p>0,05$).
b. Perbedaan antara dua p-value subgrup terpisah tidak membuktikan adanya interaksi formal. Perlu dilakukan uji interaksi langsung.
c. Analisis subgrup boleh dilakukan tanpa batasan jumlah variabel karena tidak meningkatkan risiko false positive.
d. Interaksi aditif dan multiplikatif selalu bernilai identik.

Kunci Jawaban: b

Pembahasan Rasional: Perbedaan signifikansi (p-value) antar subgrup secara matematis bukanlah uji statistik formal untuk menguji interaksi. Modifikasi efek harus diuji dengan menyertakan term interaksi langsung (treatment × subgroup) di dalam model regresi.

5) Anda mengukur berat badan pasien hemodialisis setiap minggu selama 3 bulan. Mengapa regresi linear biasa (ordinary least squares) dilarang digunakan?
   a. Karena berat badan tidak memiliki nilai rata-rata.
   b. Karena data berulang pada subjek yang sama cenderung berkorelasi, sehingga melanggar asumsi independensi data regresi biasa.
   c. Karena regresi linear hanya boleh diuji pada sampel berukuran kurang dari 30.
   d. Karena data hemolisis selalu bertipe nominal.

   Kunci Jawaban: b

   Pembahasan Rasional: Pengukuran berulang *(repeated measures)* pada subjek yang sama menghasilkan data dependen yang berkorelasi. Regresi biasa mengasumsikan independensi sisaan; pelanggaran ini akan mengecilkan standar error secara semu, meningkatkan risiko kesalahan tipe I *(false positive)*.

BAB 7
KEDOKTERAN DIAGNOSTIK (DIAGNOSTIC MEDICINE)

Dari Hasil Tes Menuju Probabilitas Penyakit dan Keputusan Klinis (Bayesian Thinking)

1. Pendahuluan

Praktik klinis kedokteran pada hakikatnya adalah serangkaian proses pengambilan keputusan di bawah kondisi ketidakpastian. Setiap hari, dokter berhadapan dengan pertanyaan mendasar: 'Apakah pasien ini menderita penyakit X?', 'Apakah tes penunjang Y diperlukan?', atau 'Bagaimana saya menginterpretasikan hasil laboratorium yang positif ini pada pasien tanpa gejala?'. Biostatistika dalam kedokteran diagnostik menyediakan kerangka logis-matematis untuk memformalkan pembaruan keyakinan klinis berdasarkan bukti objektif.

Fokus utama bab ini bukan pada menghafal rumus matematika kering, melainkan membangun penalaran klinis (clinical biostatistical thinking). Kita akan belajar bagaimana memahami bahwa hasil tes diagnostik tidak pernah memberikan kepastian absolut. Sebaliknya, tes diagnostik berfungsi sebagai alat pengubah probabilitas (probability modifier) yang menggeser tingkat keyakinan klinis kita dari sebelum tes dilakukan (pre-test probability) menjadi setelah tes diperoleh (post-test probability).

> **★ PRINSIP UTAMA KEDOKTERAN DIAGNOSTIK**
> *Akurasi Diagnostik (Diagnostic Accuracy) bukan merupakan karakteristik tetap suatu tes.*
> *Kinerja klinis tes di dunia nyata sangat bergantung pada spektrum populasi yang diuji (spectrum effect), desain pemilihan subjek, serta keandalan standar acuan yang digunakan.*

2. Capaian Pembelajaran

Setelah menyelesaikan bab ini, mahasiswa diharapkan mampu memiliki kompetensi klinis berikut:

a. Mengidentifikasi target condition, index test, dan reference standard secara tepat dari skenario klinis.
b. Menyusun dan menganalisis tabel kontingensi 2×2 sebagai dasar penentuan akurasi diagnostik.

c. Menjelaskan perbedaan konseptual antara Sensitivitas/Spesifisitas (disease-oriented) dengan Nilai Prediktif (patient-oriented).
d. Menganalisis bagaimana prevalensi populasi memengaruhi Nilai Prediktif Positif (PPV) dan Negatif (NPV).
e. Menggunakan Likelihood Ratio (LR+ dan LR-) secara konseptual untuk memperbarui probabilitas pra-tes menjadi pasca-tes.
f. Menjelaskan trade-off pemilihan ambang batas (threshold), serta menginterpretasikan kurva ROC dan nilai Area Under the Curve (AUC).
g. Membedakan akurasi diagnostik (diagnostic accuracy/discrimination) dari kemanfaatan klinis nyata (clinical usefulness).
h. Mengenali berbagai bias metodologis dalam studi diagnostik (seperti bias verifikasi, bias seleksi, dan bias pembacaan) serta menerapkan panduan pelaporan STARD 2015 dan instrumen QUADAS-2.

3. Peta Konsep Diagnostik dan Clinical Trigger

Clinical Trigger: Seorang pasien pria berusia 45 tahun tanpa riwayat keluhan kardiovaskular menjalani pemeriksaan skrining kesehatan mandiri. Hasil pemeriksaan Elektrokardiografi (EKG) menunjukkan adanya depresi segmen ST ringan (positif untuk iskemia miokard). Pasien sangat panik dan bertanya kepada Anda: 'Dokter, EKG saya positif. Apakah saya pasti menderita penyakit jantung koroner? Apakah saya akan segera terkena serangan jantung?'

Untuk menjawab pertanyaan kritis ini secara bertanggung jawab, seorang dokter tidak boleh hanya mengandalkan ingatan hafalan bahwa EKG memiliki 'sensitivitas sekian persen'. Dokter harus menyadari bahwa untuk menerjemahkan hasil tes positif menjadi probabilitas sakit riil pada pasien tersebut, diperlukan integrasi beberapa parameter penting: karakteristik klinis awal pasien (pre-test probability), spesifisitas tes EKG, dan konsekuensi dari tindakan lanjut berikutnya.

Tabel 53. Lima komponen dasar penilaian kedokteran diagnostik

Komponen	Pertanyaan Klinis Kritis	Contoh Penerapan Hipertensi / Kardiovaskular
Target condition	Kondisi patologis apa yang sebenarnya ingin diidentifikasi?	Sepsis, anemia defisiensi besi, infark miokard akut, emboli paru.
Index test	Tes penunjang apa yang sedang dievaluasi akurasinya?	Pemeriksaan biomarker baru (misal: procalcitonin), EKG, skor klinis.

Reference standard	Metode pembuktian terbaik apa yang dianggap sebagai penentu kebenaran?	Biopsi patologi anatomi, kultur darah, angiografi koroner (gold standard).
Threshold (Cut-off)	Kapan hasil tes kuantitatif dinyatakan positif vs negatif?	Kadar Troponin I >= 0.04 ng/mL, diameter lingkaran lengan, skor risiko klinis.
Consequence	Apa dampak klinis dari hasil salah positif (FP) atau salah negatif (FN)?	Overtreatment dan kecemasan pasien (FP) vs Kematian akibat terlambat tatalaksana (FN).

4. Tabel 2×2: Fondasi Berpikir Diagnostik

Tabel kontingensi 2×2 *(confusion matrix)* adalah struktur dasar yang digunakan untuk membandingkan hasil uji dari index test dengan status penyakit yang sebenarnya berdasarkan reference standard. Melalui tabel ini, seluruh indikator akurasi diagnostik dapat diturunkan.

Tabel 54. Struktur tabel kontingensi 2×2 akurasi diagnostik

Hasil Index Test	Sakit (Target +)	Tidak Sakit (Target -)	Total Baris
Hasil Tes Positif (Test +)	True Positive (TP) Tes positif pada orang sakit	False Positive (FP) Tes positif pada orang sehat	Total Positif (TP + FP) Menentukan PPV
Hasil Tes Negatif (Test -)	False Negative (FN) Tes negatif pada orang sakit	True Negative (TN) Tes negatif pada orang sehat	Total Negatif (FN + TN) Menentukan NPV
Total Kolom	Total Sakit (TP + FN) Menentukan Sensitivitas	Total Sehat (FP + TN) Menentukan Spesifisitas	Total Sampel (N) Accuracy = (TP+TN)/N

★ JANGAN MENGHAFAL TANPA MEMAHAMI DENOMINATOR!

Sensitivitas (Se) = TP / (TP + FN) --> Seberapa baik tes mendeteksi subjek yang sakit.

Spesifisitas (Sp) = TN / (TN + FP) --> Seberapa baik tes menyingkirkan subjek yang sehat.

Positive Predictive Value (PPV) = TP / (TP + FP) --> Probabilitas sakit jika hasil tes positif.

> *Negative Predictive Value (NPV) = TN / (TN + FN) --> Probabilitas sehat jika hasil tes negatif.*
> *Accuracy (ACC) = (TP + TN) / (TP + TN + FP + FN) --> Proporsi hasil tes yang benar secara keseluruhan.*

5. Sensitivitas dan Spesifisitas

Sensitivitas dan Spesifisitas adalah ukuran kinerja intrinsik dari suatu alat diagnosis. Artinya, kedua nilai ini dihitung dengan menempatkan status penyakit (reference standard) sebagai denominator awal.

a. Sensitivitas (Kepekaan): Bertolak dari kelompok orang yang benar-benar sakit (Target +). Jika suatu tes memiliki sensitivitas 90%, ini berarti di antara 100 pasien yang benar-benar sakit, tes tersebut akan mendeteksi 90 pasien sebagai positif (True Positive), sementara 10 pasien lainnya akan terlewatkan secara keliru (False Negative atau missed case).
b. Spesifisitas (Keandalan): Bertolak dari kelompok orang yang benar-benar sehat (Target -). Jika suatu tes memiliki spesifisitas 80%, ini berarti di antara 100 orang yang benar-benar sehat, tes akan memberikan hasil negatif pada 80 orang (True Negative), sedangkan 20 orang sehat lainnya akan mendapatkan hasil positif palsu (False Positive).

Mitos 'SnNout' dan 'SpPIn'

Dalam pendidikan kedokteran klasik, sering diajarkan mnemonik praktis:

- SnNout: Tes dengan Sensitivitas tinggi (Sn) jika memberikan hasil Negatif (N), sangat berguna untuk menyingkirkan diagnosis (out). Logikanya, jika tes sangat peka dan hasilnya ternyata negatif, sangat tidak mungkin pasien tersebut sakit karena persentase False Negative sangat kecil.
- SpPIn: Tes dengan Spesifisitas tinggi (Sp) jika memberikan hasil Positif (P), sangat berguna untuk mengonfirmasi/memasukkan diagnosis (in). Logikanya, jika tes sangat spesifik dan hasilnya positif, kemungkinan besar pasien tersebut benar-benar sakit karena persentase False Positive sangat kecil.

Meskipun mnemonik ini sangat populer, penggunaan mekanis tanpa menyadari keterbatasan operasionalnya dapat membahayakan keselamatan pasien. Sebab, kinerja riil di lapangan tetap ditentukan oleh seberapa besar pergeseran probabilitas (Likelihood Ratio) dan prevalensi penyakit di populasi.

6. PPV dan NPV: Membaca dari Arah Hasil Tes

Jika Sensitivitas dan Spesifisitas menjawab pertanyaan berorientasi alat ('Seberapa baik alat ini mengklasifikasikan status penyakit?'), maka Nilai Prediktif menjawab pertanyaan klinis langsung dari pasien setelah hasil tes keluar ('Dokter, hasil tes saya positif. Berapa persen kemungkinan saya benar-benar sakit?').

Positive Predictive Value (PPV) adalah probabilitas bahwa subjek benar-benar sakit jika hasil tesnya positif. Sebaliknya, Negative Predictive Value (NPV) adalah probabilitas bahwa subjek benar-benar sehat jika hasil tesnya negatif.

a. Pengaruh Prevalensi Terhadap Nilai Prediktif

Satu aspek konseptual paling krusial yang wajib dipahami oleh mahasiswa kedokteran adalah bahwa Nilai Prediktif bukanlah nilai statis. Berbeda dengan Sensitivitas dan Spesifisitas yang cenderung stabil, PPV dan NPV sangat dipengaruhi oleh prevalensi penyakit pada populasi tempat tes tersebut diterapkan.

Jika prevalensi penyakit sangat rendah (misalnya skrining populasi umum tanpa gejala), maka PPV akan anjlok secara dramatis, sekalipun tes memiliki spesifisitas yang tinggi. Sebaliknya, jika prevalensi penyakit tinggi (misalnya di klinik rujukan sub-spesialis), PPV akan meningkat tajam.

Tabel 55. Perbandingan PPV pada dua klinik dengan prevalensi berbeda (Tes yang Sama: Se=90%, Sp=90%)

Karakteristik Klinik	Prevalensi Penyakit	Positive Predictive Value (PPV)	Negative Predictive Value (NPV)
Klinik A (Skrining Populasi Umum)	1% (100 dari 10.000)	8.3% (Hanya 1 dari 12 orang positif yang benar-benar sakit)	99.9% (Hampir semua hasil negatif adalah benar-benar sehat)
Klinik B (Rujukan Sub-spesialis)	20% (2.000 dari 10.000)	69.2% (Sebagian besar hasil positif adalah benar-benar sakit)	97.3% (Hasil negatif tetap sangat andal untuk rule-out)
Catatan Metodologis	**Prevalensi Rendah**	**PPV Menurun Tajam (Banyak False Positive)**	**NPV Meningkat Sangat Tinggi**
Catatan Metodologis 2	**Prevalensi Tinggi**	**PPV Meningkat Tajam (Sedikit False Positive)**	**NPV Menurun Sedikit**

7. Pre-Test Probability dan Likelihood Ratio

Ketergantungan PPV/NPV pada prevalensi populasi memicu tantangan klinis: bagaimana menerapkan hasil studi kelompok pada satu pasien individual yang sedang duduk di hadapan kita? Solusi ilmiah yang paling elegan adalah menggunakan pendekatan Bayesian Klinis dengan Likelihood Ratio (Rasio Kemungkinan).

a. Pre-test Probability (Probabilitas Pra-tes): Tingkat keyakinan awal dokter bahwa pasien menderita penyakit tertentu sebelum tes penunjang dilakukan. Nilai ini tidak lahir dari ruang hampa, melainkan dibangun dari anamnesis komprehensif, pemeriksaan fisik terarah, dan data epidemiologi lokal.
b. Likelihood Ratio (LR): Mengukur seberapa besar perubahan probabilitas yang dihasilkan oleh suatu hasil tes tertentu. Kelebihan utama LR adalah ia menggabungkan Sensitivitas dan Spesifisitas menjadi indikator tunggal yang tidak dipengaruhi oleh prevalensi penyakit.
c. Likelihood Ratio Positif (LR+): Rasio antara probabilitas hasil tes positif pada orang sakit dibandingkan orang sehat. Rumus: LR+ = Sensitivitas / (1 - Spesifisitas). Nilai LR+ > 10 menunjukkan kemampuan yang sangat kuat untuk mengonfirmasi diagnosis ('rule in').
d. Likelihood Ratio Negatif (LR-): Rasio antara probabilitas hasil tes negatif pada orang sakit dibandingkan orang sehat. Rumus: LR- = (1 - Sensitivitas) / Spesifisitas. Nilai LR- < 0.1 menunjukkan kemampuan yang sangat kuat untuk menyingkirkan diagnosis ('rule out').

Alur Kerja Bayesian di Tempat Praktik

Untuk menghitung probabilitas pasca-tes secara formal, dokter menggunakan tiga langkah berurutan:

- Ubah Pre-test Probability (P) menjadi Pre-test Odds: Odds = P / (1 - P)
- Hitung Post-test Odds: Post-test Odds = Pre-test Odds × Likelihood Ratio (LR+ atau LR-)
- Ubah kembali Post-test Odds menjadi Post-test Probability: P_post = Odds_post / (1 + Odds_post).

Tabel 56. Interpretasi praktis nilai likelihood ratio (LR) untuk keputusan klinis

Nilai LR+	Nilai LR-	Kekuatan Pergeseran Keyakinan dan Tindakan Klinis
> 10	< 0,1	Sangat Kuat (Sangat andal untuk konfirmasi / menyingkirkan penyakit secara mutlak).
5 - 10	0,1 – 0,2	Sedang / Moderat (Cukup bermakna untuk menggeser keputusan medis dalam banyak kasus).
2 - 5	0,2 – 0,5	Lemah / Kecil (Hanya memberikan sedikit pergeseran, sering membutuhkan tes tambahan).
1 - 2	0,5 – 1,0	Sangat Lemah / Minimal (Hampir tidak mengubah keyakinan awal dokter).
~ 1	~ 1	Tidak Berguna (Hasil tes tidak memberikan informasi tambahan apa pun).

8. Threshold, Kurva ROC, dan AUC

Banyak parameter pemeriksaan klinis tidak dilaporkan sebagai biner sederhana (ya/tidak), melainkan berupa nilai kontinu (seperti kadar Troponin I, Hemoglobin, atau tekanan darah sistolik). Untuk menentukan status diagnosis, kita harus memotong distribusi kontinu tersebut pada titik tertentu yang disebut Ambang Batas *(Threshold atau Cut-off)*.

Trade-off Threshold: Pemindahan batas cut-off akan selalu mengubah nilai Sensitivitas dan Spesifisitas secara berlawanan. Jika kita menurunkan batas cut-off agar lebih mudah dinyatakan positif, kita akan menangkap lebih banyak pasien yang sakit (Sensitivitas naik), tetapi konsekuensinya kita juga akan menjaring lebih banyak orang sehat sebagai positif palsu (Spesifisitas turun). Sebaliknya, menaikkan batas cut-off akan meningkatkan Spesifisitas tetapi mengorbankan Sensitivitas.

Kurva ROC *(Receiver Operating Characteristic)*

Kurva ROC adalah visualisasi grafis dari trade-off antara Sensitivitas (pada sumbu Y) melawan 1 - Spesifisitas (pada sumbu X) pada seluruh rentang batas ambang yang dimungkinkan. Semakin melengkung kurva tersebut mendekati sudut kiri atas grafik, semakin baik pula kinerja diskriminasi keseluruhan dari tes tersebut.

Area Under the Curve (AUC): Merupakan kuantifikasi numerik dari kurva ROC. Nilai AUC berkisar dari 0.5 (uji tidak lebih baik daripada tebakan koin acak) hingga 1.0 (uji klasifikasi sempurna tanpa ada kesalahan). Secara probabilistik, AUC 0.85 menunjukkan bahwa jika kita mengambil satu pasien sakit dan satu orang sehat secara acak, terdapat

probabilitas 85% bahwa skor tes pasien sakit akan lebih ekstrem/tinggi daripada orang sehat.

> **★ KETERBATASAN YOUDEN INDEX**
> *Youden Index (J) dihitung dengan formula: J = Sensitivitas + Spesifisitas - 1. Meskipun Youden Index sering digunakan oleh perangkat lunak statistik untuk merekomendasikan 'Optimal Cut-off' matematis, dokter tidak boleh menggunakannya secara membabi buta. Sebab, Youden Index mengasumsikan bahwa konsekuensi kerugian antara salah positif (FP) dan salah negatif (FN) adalah sama, padahal dalam kenyataan klinis nyata, konsekuensinya sering kali sangat timpang.*

9. Diagnostic Accuracy ≠ Clinical Usefulness

Kesalahan umum dalam literatur medis adalah mengasumsikan bahwa jika suatu tes diagnostik baru memiliki nilai AUC yang sangat tinggi, maka tes tersebut harus segera diadopsi secara luas di rumah sakit. Dalam kerangka berpikir biostatistika klinis, akurasi diagnostik tinggi hanyalah prasyarat awal, bukan jaminan kemanfaatan medis nyata (clinical usefulness).

Zweig dan Campbell (1993) menguraikan empat tingkatan evaluasi tes diagnostik yang terstruktur:

- Analytical Validity: Apakah alat ukur tersebut andal dan konsisten secara teknis laboratorium?
- Diagnostic Accuracy: Seberapa baik alat tersebut memisahkan kelompok sakit dari sehat dibandingkan reference standard?
- Clinical Usefulness: Apakah informasi tambahan dari tes tersebut mengubah keputusan pengobatan dokter ke arah yang lebih tepat?
- Patient Outcome: Apakah perubahan keputusan medis tersebut benar-benar menghasilkan penurunan morbiditas, mortalitas, atau peningkatan kualitas hidup pasien?

Contoh Gizi Klinis: Penggunaan skor skrining risiko malnutrisi baru di ICU. Tes tersebut memiliki akurasi sangat tinggi untuk mendeteksi deplesi nutrisi. Namun, jika tim klinis tidak memiliki kapasitas logistik untuk memberikan terapi nutrisi enteral dini secara intensif, maka deteksi dini tersebut tidak akan mengubah keputusan medis maupun memperbaiki status klinis pasien. Tes tersebut akurat, tetapi tidak memiliki kemanfaatan klinis (utility) pada setting tersebut.

10. Bias pada Studi Diagnostic Accuracy

Studi diagnostik observasional sangat rentan terhadap berbagai penyimpangan sistematis (bias) yang dapat mendistorsi nilai estimasi akurasi menjadi terlalu optimis (overestimate) atau justru salah arah.

a. Bias Seleksi Spektrum (Spectrum Bias): Terjadi jika studi hanya merekrut kelompok pasien dengan derajat penyakit yang sangat berat dan membandingkannya dengan kelompok kontrol yang sangat sehat (murni). Pada dunia nyata, tes akan dihadapkan pada spektrum klinis yang jauh lebih abu-abu, sehingga akurasinya akan tampak jauh lebih rendah.
b. Bias Verifikasi Parsial (Partial Verification Bias): Terjadi jika keputusan untuk merujuk pasien menjalani pemeriksaan reference standard (gold standard) dipengaruhi oleh hasil index test. Sebagai contoh, hanya pasien dengan hasil index test positif yang dikirim untuk menjalani angiografi invasif. Hal ini akan mendistorsi Sensitivitas menjadi tampak sangat tinggi dan Spesifisitas tampak sangat rendah.
c. Bias Pembacaan (Review Bias): Terjadi jika pembaca index test sudah mengetahui hasil dari reference standard (atau sebaliknya). Penilaian harus dilakukan secara independen dan tersamar (blinded) untuk mencegah sugesti subyektif pembaca.

> **★ STANDAR PELAPORAN DAN PENILAIAN BIAS**
> *STARD 2015 (Standards for Reporting Diagnostic Accuracy) adalah panduan pelaporan transparan studi diagnostik (30 item penting).*
> *QUADAS-2 (Quality Assessment of Diagnostic Accuracy Studies) digunakan untuk menilai risiko bias dan aplikabilitas pada empat domain: patient selection, index test, reference standard, serta flow dan timing.*

11. Studi Kasus dan Mini-Case Diagnostik

1) **Mini-Case:** Dugaan Emboli Paru di IGD

   Kasus: Seorang wanita berusia 28 tahun datang ke IGD dengan keluhan nyeri dada pleuritik mendadak pasca penerbangan jarak jauh 14 jam. Secara klinis, pre-test probability untuk Emboli Paru (EP) dinilai rendah-sedang (skor Wells = 2, probabilitas sekitar 10%). Residen IGD mengusulkan pemeriksaan D-dimer. Hasil EKG normal.

   Diskusi dan Pembahasan: D-dimer dikenal sebagai tes dengan Sensitivitas sangat tinggi (96%) tetapi Spesifisitas rendah (50%) untuk

EP. Jika hasil tes D-dimer ternyata NEGATIF, maka nilai LR- = (1 - 0.96) / 0.50 = 0.08. Menggunakan kalkulasi Bayes: Pre-test Odds = 0.10 / 0.90 = 0.11. Post-test Odds = 0.11 × 0.08 = 0.009. Post-test Probability = 0.009 / (1 + 0.009) = 0.009 (Kurang dari 1%). Hasil negatif ini sangat andal untuk secara aman menyingkirkan diagnosis EP (rule out) tanpa perlu CT Angiografi yang mahal dan beradiasi.

2) Mini-Case: Biomarker Sepsis Baru di ICU

Kasus: Sebuah studi di jurnal kedokteran ternama melaporkan biomarker X memiliki AUC 0.92 untuk mendeteksi sepsis dini pada pasien ICU bedah tersier. Seorang klinisi ingin mengimplementasikan tes ini sebagai alat skrining rutin di IGD Puskesmas kecamatan.

Diskusi dan Pembahasan: Rencana ini mengabaikan prinsip applicability dan spectrum effect. Pasien di ICU bedah tersier memiliki karakteristik keparahan yang sangat tinggi (spectrum penyakit berat). Di IGD Puskesmas umum, sebagian besar pasien demam bukan disebabkan oleh sepsis berat melainkan infeksi virus ringan. Akurasi tes (Sensitivitas, Spesifisitas, dan PPV) dipastikan akan anjlok drastis jika diterapkan pada populasi dengan prevalensi dan derajat keparahan yang jauh lebih rendah daripada populasi asli tempat model tersebut divalidasi.

12. Rangkuman

a. Tes diagnostik tidak memberikan kepastian absolut; fungsi utamanya adalah menggeser probabilitas penyakit.
b. Sensitivitas dan Spesifisitas dinilai berdasarkan status penyakit yang sebenarnya (reference standard sebagai denominator).
c. Nilai Prediktif (PPV/NPV) berorientasi pada hasil tes pasien, dan nilainya sangat bergantung pada prevalensi penyakit di populasi.
d. Likelihood Ratio (LR) adalah indikator pergeseran probabilitas terbaik karena tidak dipengaruhi prevalensi populasi.
e. Ambang batas (threshold) menentukan keseimbangan sensitivitas dan spesifisitas; tidak ada 'optimal cut-off' yang murni matematis tanpa menimbang risiko klinis.
f. Akurasi diskriminasi (AUC) tidak otomatis mencerminkan kegunaan klinis (clinical usefulness) dalam manajemen pasien nyata.
g. Studi diagnostik wajib meminimalkan bias seleksi spektrum, bias verifikasi, dan bias review untuk menghasilkan estimasi akurasi yang valid.

13. Latihan dan Evaluasi Formatif

1) Sebuah tes skrining baru memiliki sensitivitas 95% dan spesifisitas 90%. Jika tes ini diterapkan pada populasi dengan prevalensi penyakit 1%, pernyataan manakah yang paling tepat?
   a. Sebagian besar hasil tes positif adalah benar-benar sakit.
   b. Positive Predictive Value (PPV) dari tes ini akan bernilai sangat rendah.
   c. Negative Predictive Value (NPV) dari tes ini akan bernilai sangat rendah.
   d. Tes ini tidak memiliki nilai akurasi karena prevalensinya kecil.

Kunci Jawaban: b

Pembahasan: PPV sangat dipengaruhi prevalensi. Pada prevalensi 1%, di antara 10.000 orang, terdapat 100 orang sakit dan 9.900 orang sehat. Tes positif akan menjaring 95 orang sakit (true positive) dan 990 orang sehat (false positive). PPV = 95 / (95 + 990) = 8.7%. Artinya, lebih dari 91% hasil positif adalah palsu (false positive). Maka PPV bernilai sangat rendah.

2) Mnemonik 'SnNout' secara konseptual menyatakan bahwa jika sebuah tes dengan sensitivitas tinggi memberikan hasil negatif, maka tes tersebut sangat andal untuk menyingkirkan diagnosis. Indikator statistik apa yang paling kuat mendukung fungsi 'rule-out' ini?
   a. Nilai Likelihood Ratio positif (LR+) yang sangat tinggi (>10).
   b. Nilai Likelihood Ratio negatif (LR-) yang sangat rendah (<0.1).
   c. Nilai Area Under the Curve (AUC) tepat di angka 0.50.
   d. Nilai akurasi keseluruhan yang mencapai 100%.

Kunci Jawaban: b

Pembahasan: Fungsi menyingkirkan penyakit secara andal (rule out) dicerminkan oleh Likelihood Ratio negatif (LR-) yang sangat kecil (di bawah 0.1 atau lebih rendah). Ini memastikan odds pasca-tes turun mendekati nol.

3) Seorang peneliti ingin menentukan cut-off optimal untuk pemeriksaan kadar protein urin kuantitatif guna mendeteksi preeklampsia berat. Ia memilih cut-off yang memberikan nilai Youden Index (J) maksimum. Apa keterbatasan klinis utama dari pendekatan ini?
   a. Youden Index hanya bisa dihitung jika data terdistribusi Normal.
   b. Youden Index mengasumsikan kerugian akibat salah positif dan salah negatif memiliki bobot klinis yang setara.
   c. Nilai Youden Index tidak dipengaruhi oleh perubahan sensitivitas.
   d. Pendekatan ini selalu menghasilkan spesifisitas yang bernilai nol.

Kunci Jawaban: b

Pembahasan: Secara matematis, Youden Index memberikan bobot seimbang (1:1) antara Sensitivitas dan Spesifisitas. Namun dalam praktik medis, melewatkan kasus preeklampsia berat (salah negatif) jauh lebih berbahaya daripada salah positif (pemeriksaan konfirmasi lanjutan). Maka penentuan threshold harus mempertimbangkan konsekuensi klinis ini.

4) Bias verifikasi parsial (partial verification bias) dalam studi diagnostik terjadi ketika:
   a. Peneliti menggunakan dua reference standard yang berbeda untuk kelompok yang berbeda secara subyektif.
   b. Keputusan untuk melakukan pemeriksaan reference standard dipengaruhi oleh hasil index tes yang diperoleh sebelumnya.
   c. Pemeriksa reference standard mengetahui hasil index test sebelum melakukan penilaian.
   d. Subjek penelitian dipilih dari kelompok sehat murni saja.

Kunci Jawaban: b

Pembahasan: Bias verifikasi parsial (atau work-up bias) terjadi jika rujukan menjalani pemeriksaan standar emas dipengaruhi hasil index test. Hal ini mendistorsi sensitivitas menjadi bias ke atas dan spesifisitas menjadi bias ke bawah.

5) Manakah dari dimensi evaluasi tes diagnostik berikut yang menilai apakah penggunaan alat penunjang baru tersebut benar-benar menghasilkan perbaikan status kesehatan atau kesembuhan pasien secara nyata di dunia medis?
   a. Analytical validity
   b. Diagnostic accuracy
   c. Patient outcome / Clinical utility
   d. Receiver Operating Characteristic (ROC)

Kunci Jawaban: c

Pembahasan: Kemanfaatan klinis akhir suatu tes harus dibuktikan melalui peningkatan kualitas hidup atau kesembuhan pasien (Patient outcome / clinical utility), bukan hanya dari tingginya akurasi teoretis.

BAB 8
SURVIVAL dan PROGNOSIS

Analisis Waktu Kejadian (Time-to-Event) dan Evaluasi Kemampuan Prediksi Model

1. Pengantar: Mengapa Analisis Survival Diperlukan?

Dalam kedokteran klinis, prognosis adalah salah satu pilar utama pelayanan pasien. Ketika seorang pasien terdiagnosis penyakit kronis seperti gagal ginjal kronis atau kanker paru, pertanyaan mereka jarang sekali hanya berupa "Apakah saya akan bertahan hidup?". Sebaliknya, pertanyaan kritis mereka hampir selalu berfokus pada dimensi waktu: *"Berapa lama waktu yang saya miliki sebelum penyakit ini memburuk?"* atau *"Berapa persen peluang saya untuk tetap bertahan hidup bebas dari komplikasi dalam lima tahun ke depan?"*

Mengapa kita tidak menggunakan uji statistik standar seperti regresi logistik untuk menjawab pertanyaan ini? Setidaknya ada dua hambatan metodologis utama yang membuat metode standar seperti regresi logistik atau uji *t-test* tidak valid digunakan:

a. **Waktu Tindak Lanjut *(Follow-up)* yang Tidak Seragam:** Dalam studi prospektif, pasien tidak masuk ke dalam penelitian pada hari yang sama dan tidak selalu diikuti dalam durasi waktu yang identik. Ada pasien yang diamati selama 5 tahun, namun ada yang baru masuk di tahun ke-3 dan hanya diikuti selama 2 tahun.
b. **Kehadiran Data Tersensor *(Censored Data)*:** Pada akhir masa penelitian, banyak pasien yang belum mengalami kejadian klinis yang dinilai (misalnya belum meninggal atau belum mengalami kekambuhan), hilang dari pengamatan (lost-to-follow-up), atau memutuskan keluar dari studi. Kita tidak boleh menganggap mereka 'bebas komplikasi selamanya' ataupun 'pasti mati' status mereka adalah tidak diketahui secara pasti setelah titik waktu terakhir pengamatan.

> **Hambatan Analisis Konvensional:** *Regresi logistik memperlakukan outcome sebagai variabel biner murni (ya/tidak) tanpa memedulikan kapan kejadian terjadi. Jika Anda mengamati 100 pasien stroke selama 5 tahun, pasien yang meninggal di hari pertama dan pasien yang meninggal di hari terakhir tahun ke-5 akan dicatat sama-sama sebagai 'Meninggal = Ya'. Analisis survival mempertahankan informasi dimensi waktu ini.*

2. Konsep Utama: Data Tersensor *(Censoring)*

Penyensoran *(censoring)* adalah karakteristik unik dari analisis survival. Sensor terjadi ketika kita memiliki informasi parsial tentang waktu kejadian seorang pasien, namun kita tidak tahu secara pasti kapan kejadian riilnya terjadi.

Bentuk sensor yang paling umum dalam penelitian medis adalah **Right-Censoring (Penyensoran Kanan)**, yang berarti kejadian klinis yang diteliti hanya diketahui terjadi di sebelah kanan (setelah) titik waktu terakhir pasien tersebut diobservasi secara aktif. Terdapat tiga penyebab umum terjadinya right-censoring:

a. **Studi Berakhir sebelum Pasien Mengalami Kejadian:** Hingga tanggal akhir pengumpulan data yang dispesifikasikan dalam protokol studi, pasien masih dalam kondisi hidup atau bebas dari kejadian.
b. **Hilang Kontak (Lost-to-Follow-up):** Pasien pindah alamat, tidak dapat dihubungi, atau tidak melakukan kontrol rutin sehingga status klinis terakhirnya tidak terlacak.
c. **Kejadian Kompetitif (Competing Event):** Pasien mengalami kejadian lain yang menyebabkan pengamatan terhadap kejadian utama tidak lagi relevan (misalnya pasien meninggal karena kecelakaan lalu lintas ketika studi sedang menilai kematian akibat kanker paru).

Tabel 57. Meringkas perbedaan mendasar antara data kejadian (event) dan data tersensor (censoring) untuk memberikan landasan pemahaman sebelum melangkah ke kurva kelangsungan hidup

Aspek Evaluasi	Data Kejadian (Event) vs. Data Tersensor (Censoring)
Definisi Klinis	Kejadian klinis utama (kematian, kekambuhan, infark miokard) benar-benar terjadi pada waktu tertentu yang tercatat akurat.
Waktu Tersensor	Pasien tidak mengalami kejadian selama periode pengamatan aktif. Waktu sensor adalah waktu kontak klinis terakhir yang diketahui.

Poin Penting dari Tabel 8.1: Data tersensor tidak boleh dibuang dari analisis! Data tersebut tetap memberikan informasi berharga bahwa pasien tersebut berhasil bertahan hidup (survival) sekurang-kurangnya hingga waktu penyensoran terjadi.

3. Estimasi Non-Parametrik: Kurva Kaplan-Meier

Kurva Kaplan-Meier (KM) merupakan metode standar non-parametrik untuk mengestimasi fungsi survival S(t). Fungsi survival S(t) didefinisikan sebagai probabilitas seorang individu bertahan hidup bebas dari kejadian melampaui waktu t.

Prinsip matematika dasar kurva KM adalah menghitung probabilitas survival kondisional pada setiap waktu kejadian nyata terjadi. Kurva KM digambarkan sebagai grafik tangga menurun. Penurunan anak tangga hanya terjadi pada titik waktu ketika ada subjek yang mengalami kejadian (event). Tanda centang atau titik vertikal kecil pada kurva menandakan adanya subjek yang tersensor pada titik waktu tersebut.

> **Cara Membaca Kurva KM:** *Ketika membaca kurva Kaplan-Meier, perhatikan tabel 'Number at Risk' di bawah sumbu horizontal. Jumlah subjek berisiko akan berkurang seiring waktu karena dua hal: subjek mengalami event, atau subjek tersensor. Jika jumlah subjek di akhir masa tindak lanjut sangat sedikit (misalnya < 10% dari jumlah awal), estimasi kurva di bagian ujung tersebut menjadi sangat tidak presisi dan memiliki tingkat ketidakpastian yang sangat tinggi.*

Tabel 58. Memberikan contoh perhitungan ilustratif bagaimana probabilitas kelangsungan hidup kumulatif dihitung secara interval

Waktu (t)	Subjek Berisiko (n)	Kejadian (d)	Tersensor (c)	Survival Kumulatif S(t)
0	50	0	0	1,00 (100%)
12 bulan	50	2	3	48/50 = 0,96 (96%)
24 bulan	45	4	1	0,96 (41/45) = 0,875 (87,5%)

4. Membandingkan Kurva: Uji Log-Rank

Uji Log-rank adalah uji statistik non-parametrik yang paling sering digunakan untuk membandingkan kurva kelangsungan hidup antara dua atau lebih kelompok independen (misalnya kelompok obat baru vs. plasebo).

Uji ini bekerja dengan membandingkan jumlah kejadian yang diamati (Observed) dengan jumlah kejadian yang diharapkan (Expected) secara teoretis pada setiap titik waktu kejadian jika hipotesis nol (bahwa tidak ada perbedaan kelangsungan hidup antar kelompok) adalah benar.

Batasan Uji Log-Rank: *Uji Log-rank mengasumsikan bahwa rasio bahaya (hazard ratio) antara kedua kelompok konstan sepanjang waktu. Jika kurva Kaplan-Meier saling menyilang (misalnya Terapi Bedah memiliki kematian tinggi di awal namun unggul jangka panjang dibandingkan Terapi Obat), uji Log-rank standar menjadi tidak valid karena asumsi bahaya proporsional dilanggar. Dalam skenario ini, uji statistik alternatif seperti Generalized Wilcoxon atau Weighted Log-rank harus dipertimbangkan.*

5. Regresi Survival: Model Cox Proportional Hazards (PH)

Meskipun kurva Kaplan-Meier dan uji Log-rank sangat berguna, keduanya memiliki keterbatasan besar: tidak dapat menyesuaikan efek dari beberapa variabel kovariat (konfounder) secara bersamaan. Untuk analisis survival multivariat, standar baku dalam kedokteran adalah menggunakan **Model Regresi Cox Proportional Hazards (Cox PH)**.

Output utama dari model Cox PH adalah **Hazard Ratio (HR)**. Hazard didefinisikan sebagai laju instan terjadinya suatu kejadian pada titik waktu t, dengan syarat subjek tersebut telah bertahan hidup hingga waktu t. Dengan demikian, Hazard Ratio membandingkan laju kejadian instan antara dua kelompok pengamatan:

HR = 1.0: Laju kejadian di kedua kelompok adalah sama (tidak ada perbedaan efektivitas klinis).

HR > 1.0: Laju kejadian pada kelompok intervensi/pajanan lebih tinggi dibandingkan kelompok kontrol (faktor risiko).

HR < 1.0: Laju kejadian pada kelompok intervensi/pajanan lebih rendah dibandingkan kelompok kontrol (faktor protektif).

Perbedaan Kritis HR vs RR: *Hazard Ratio (HR) tidak sama dengan Relative Risk (RR). RR mengukur akumulasi kejadian di akhir studi (misalnya total kematian setelah 5 tahun). Sebaliknya, HR mengukur laju kejadian dari waktu ke waktu selama periode studi berjalan. Jika suatu pengobatan memperlambat kematian tanpa mengubah jumlah total kematian di akhir studi 5 tahun, HR akan menunjukkan nilai signifikan < 1.0, sedangkan RR di akhir tahun ke-5 mungkin mendekati 1.0.*

6. Penerapan Klinis: Model Prediksi Prognosis di Sisi Tempat Tidur

Dalam praktik kedokteran berbasis bukti, dokter menggunakan skor prognosis (seperti Skor Framingham untuk penyakit kardiovaskular, atau

skor CHA2DS2-VASc untuk fibrilasi atrium) untuk memandu keputusan klinis. Skor-skor ini dibangun berdasarkan koefisien regresi dari model Cox PH. Kemampuan model prognosis dievaluasi menggunakan dua dimensi kinerja statistik:

a. **Diskriminasi (Discrimination / Harrell's C-index):** Mengukur seberapa baik skor dapat membedakan antara pasien yang akan mengalami kejadian lebih cepat dengan pasien yang bertahan hidup lebih lama. C-index berkisar dari 0.5 (tebakan acak) hingga 1.0 (diskriminasi sempurna).
b. **Kalibrasi (Calibration):** Mengukur kesesuaian antara probabilitas kejadian yang diprediksi oleh model dengan persentase kejadian riil yang teramati pada sekelompok pasien nyata.

7. Evaluasi Formatif dan Latihan Mandiri

1) Seorang peneliti melakukan uji klinis acak (RCT) fase III untuk membandingkan obat kemoterapi baru (Regimen X) dengan terapi standar pada pasien kanker pankreas stadium lanjut selama 3 tahun. Pada akhir masa studi, 20% pasien pada Regimen X masih hidup, 10% hilang kontak karena pindah rumah, dan sisanya meninggal dunia. Bagaimana status pasien yang hilang kontak tersebut diklasifikasikan dalam analisis survival?
   a. Dikeluarkan dari seluruh analisis karena datanya tidak lengkap.
   b. Diklasifikasikan sebagai data tersensor (censored) pada waktu kontak klinis terakhir.
   c. Dianggap mengalami kejadian (event) kematian pada hari ke-1 setelah kontak terakhir.
   d. Diimputasi menggunakan metode regresi linier untuk menebak tanggal kematian riil.

Kunci Jawaban: b

Pembahasan Rasional: Pasien yang hilang kontak (lost-to-follow-up) tidak boleh dibuang karena akan memicu bias seleksi, dan tidak boleh diasumsikan meninggal karena kita tidak tahu pasti. Status mereka diklasifikasikan sebagai 'tersensor kanan' (right-censored) pada waktu kontak terakhir yang diketahui secara aktif. Data kelangsungan hidup mereka hingga tanggal tersebut tetap dihitung berkontribusi dalam analisis.

2) Saat menganalisis hasil kurva Kaplan-Meier untuk studi pencegahan sekunder infark miokard, Anda melihat bahwa kurva untuk kelompok intervensi diet mediterania dan kelompok diet rendah lemak saling menyilang pada bulan ke-6. Apa implikasi metodologis utama dari fenomena kurva menyilang ini?

a. Kekuatan uji *(power of study)* meningkat secara drastis setelah bulan ke-6.
b. Asumsi proportional hazards dilanggar, sehingga hasil uji Log-rank standar menjadi tidak valid.
c. Kedua kurva dipastikan memiliki nilai $p < 0.05$ pada seluruh titik waktu.
d. Analisis harus dihentikan karena tidak ada perbedaan klinis sama sekali.

Kunci Jawaban: b

Pembahasan Rasional: Uji Log-rank mengasumsikan bahwa hazard ratio (rasio bahaya instan) antar kelompok konstan sepanjang masa pengamatan (Proportional Hazards Assumption). Jika kurva Kaplan-Meier saling menyilang, ini menandakan efek intervensi berbalik arah atau berubah seiring waktu, yang secara otomatis melanggar asumsi tersebut. Hasil uji Log-rank konvensional dalam situasi ini menjadi tidak valid atau kehilangan kekuatan statistiknya.

3) Sebuah studi kohort mengevaluasi hubungan antara kebiasaan merokok dengan kejadian penyakit jantung koroner (PJK) selama 10 tahun tindak lanjut. Model regresi Cox multivariat menghasilkan Hazard Ratio (HR) untuk kelompok perokok sebesar 2.40 (95% CI: 1.80 s.d. 3.20) setelah disesuaikan dengan usia, diabetes, dan hipertensi. Bagaimana interpretasi klinis yang tepat dari nilai HR tersebut?
   a. Perokok memiliki risiko kumulatif menderita PJK 2.40 kali lipat lebih tinggi di akhir tahun ke-10.
   b. Laju instan terjadinya PJK pada perokok adalah 2.40 kali lipat laju pada non-perokok pada setiap titik waktu pengamatan.
   c. Kebiasaan merokok meningkatkan durasi kelangsungan hidup bebas PJK rata-rata sebesar 2.40 tahun.
   d. Terdapat probabilitas sebesar 2.40% bahwa perokok akan mengalami PJK mendadak.

Kunci Jawaban: b

Pembahasan Rasional: Hazard Ratio (HR) mencerminkan perbandingan laju instan (instantaneous rate) terjadinya kejadian antara kelompok terpapar dibandingkan kelompok kontrol sepanjang periode observasi, bukan risiko kumulatif di akhir masa tindak lanjut (yang merupakan definisi Relative Risk). Dengan demikian, laju terjadinya PJK pada perokok adalah 2.40 kali lipat lebih cepat pada setiap titik waktu dibandingkan kelompok non-perokok setelah menyelaraskan konfounder.

4) Seorang kardiolog ingin menguji kegunaan praktis dari skor risiko prognosis baru untuk memprediksi gagal jantung akut dalam 1 tahun. Hasil evaluasi model pada sampel independen menunjukkan nilai C-index sebesar 0.52. Bagaimana keputusan klinis terbaik mengenai penggunaan skor risiko baru tersebut?
   a. Skor tersebut sangat siap diaplikasikan karena memiliki akurasi di atas 50%.
   b. Skor tersebut harus segera dipatenkan karena memberikan hasil diskriminasi yang superior.
   c. Skor tersebut tidak layak digunakan secara klinis karena performa diskriminasinya setara tebakan acak (koin seimbang).
   d. Skor tersebut hanya boleh digunakan jika prevalensi gagal jantung di populasi bernilai sangat rendah.

Kunci Jawaban: c

Pembahasan Rasional: *Harrell's C-index berkisar antara 0.5 hingga 1.0. Nilai C-index sebesar 0.50 menandakan model sama sekali tidak memiliki kemampuan diskriminasi (setara dengan melempar koin acak untuk memprediksi prognosis pasien). Skor dengan C-index 0.52 menunjukkan kinerja yang sangat buruk dan tidak layak digunakan untuk memandu keputusan medis.*

5) Manakah dari skenario klinis berikut yang menggambarkan bias penyensoran informatif (*informative censoring*) yang dapat mendistorsi keabsahan kurva Kaplan-Meier?
   a. Pasien tersensor karena periode penelitian berakhir sesuai protokol uji klinis.
   b. Pasien kanker stadium lanjut mengundurkan diri dari penelitian karena efek samping obat uji yang terlalu beracun dan memperburuk kondisi fisik mereka.
   c. Pasien dipindahkan ke pusat perawatan lain karena alasan perpindahan domisili keluarga.
   d. Pasien meninggal dunia karena kecelakaan pesawat saat perjalanan bisnis.

Kunci Jawaban: b

Pembahasan Rasional: *Analisis survival mengasumsikan bahwa proses penyensoran bersifat non-informatif (penyensoran terjadi secara acak dan tidak terkait dengan prognosis klinis pasien). Jika pasien kanker keluar dari penelitian karena kondisi fisik mereka memburuk akibat toksisitas obat (informative censoring), menyensor mereka akan menyebabkan kurva kelangsungan hidup tampak 'lebih baik' secara semu (overestimasi survival)*

karena kelompok pasien yang tersisa adalah mereka yang sehat saja. Ini melanggar asumsi dasar analisis survival.

BAB 9
CRITICAL APPRAISAL (TELAAH KRITIS)

Menilai Validitas, Kepentingan, dan Penerapan Bukti Klinis pada Tempat Tidur

1. Pendahuluan

Bagi seorang dokter di era modern, tantangan terbesar bukanlah menemukan informasi klinis, melainkan menyaring banjir informasi tersebut. Setiap hari, ribuan artikel medis baru diterbitkan di seluruh dunia. Namun, tidak semua bukti ilmiah yang dipublikasikan memiliki kualitas metodologi yang layak untuk dijadikan landasan terapi bagi pasien.

Critical appraisal (telaah kritis) adalah proses evaluasi secara sistematis terhadap literatur ilmiah untuk menilai validitas metodologi, signifikansi klinis hasil, dan aplikabilitas praktisnya pada pasien. Bab ini dirancang untuk membekali Anda dengan lensa kritis struktural, bukan sekadar mencentang checklist akreditasi, melainkan membangun penalaran klinis berbasis bukti (Evidence-Based Medicine/EBM).

> **Pesan Kunci Bab Ini**
> *Telaah kritis bukanlah aktivitas birokrasi akademis atau sekadar mencentang daftar dokumen pelaporan (seperti CONSORT atau STROBE). Ini adalah latihan pertahanan klinis (clinical defense mechanism) untuk melindungi pasien Anda dari kesimpulan salah yang disebabkan oleh desain penelitian yang buruk, analisis yang bias, atau klaim komersial yang berlebihan.*

2. Struktur Dasar Telaah Kritis: Tiga Pilar EBM

Dalam melakukan telaah kritis terhadap artikel jurnal kedokteran, seorang klinisi harus selalu berpijak pada tiga pilar utama yang diusulkan oleh kelompok pemikir Evidence-Based Medicine (EBM) di McMaster University. Ketiga pilar ini dirumuskan sebagai tiga pertanyaan kritis:

***Are the results of the study valid*? (Apakah hasil penelitian ini valid?)** Pilar ini menguji validitas internal (internal validity) untuk memastikan bahwa kesimpulan studi merupakan refleksi kebenaran biologis yang mendekati kenyataan, bukan produk dari bias sistematis, perancu (confounding), atau variasi acak semata.

***What are the results*? (Apa hasil penelitian ini dan apakah penting?)** Pilar ini mengukur signifikansi klinis dan statistik. Fokusnya bergeser dari sekadar melihat nilai p (p-value) ke arah penilaian ukuran efek (effect size) seperti Risk Ratio, Hazard Ratio, Mean Difference, beserta ketidakpastian estimasinya yang digambarkan oleh selang kepercayaan (Confidence Interval/CI).

***Will the results help me in caring for my patient*? (Apakah hasil ini dapat diterapkan pada pasien saya?)** Pilar ini menguji validitas eksternal (external validity/applicability). Kita mengevaluasi kesesuaian karakteristik pasien dalam studi dengan pasien kita di tempat tidur, serta memperhitungkan keseimbangan antara manfaat, risiko efek samping, biaya, dan preferensi pasien.

Tabel 59. Kerangka kerja tiga pilar telaah kritis jurnal kedokteran

Pilar Telaah Kritis	Pertanyaan Evaluasi Kunci	Metrik dan Instrumen Utama
Validitas Internal (Internal Validity)	• Apakah randomisasi berjalan baik? • Apakah ada blinding ganda? • Apakah follow-up pasien lengkap (attrition)? • Apakah analisis menggunakan Intention-to-Treat (ITT)?	• Cochrane Risk of Bias (RoB 2) untuk RCT • Newcastle-Ottawa Scale (NOS) untuk studi observasional • Diagram Alir CONSORT
Kepentingan Klinis (Clinical Importance)	• Seberapa besar efek terapi? • Apakah interval kepercayaan (CI) sempit (presisi)? • Apakah batas bawah CI masih bermakna klinis? • Berapa nilai NNT (Number Needed to Treat)?	• Ukuran Efek Absolut (Risk Difference) • Ukuran Efek Relatif (RR, OR, HR) • Nilai NNT dan NNH • Selang Kepercayaan (95% CI)
Aplikabilitas (External Validity)	• Apakah pasien saya berbeda secara biologis/demografis dengan subjek studi? • Apakah intervensi layak dilakukan di faskes saya? • Apakah rasio manfaat-mudarat menguntungkan?	• Perbandingan kriteria inklusi/eksklusi • Analisis utilitas biaya (Cost-Utility) • Diskusi nilai dan preferensi pasien

3. Menilai Validitas Internal: Anatomi Bias dan Perancu

Menilai validitas internal berarti mendeteksi apakah hasil studi menyimpang dari kenyataan akibat adanya bias sistematis. Ada lima domain bias utama yang harus diwaspadai klinisi saat membaca uji klinis teracak samar (Randomized Controlled Trial/RCT):

a. **Bias yang timbul dari proses randomisasi (*Bias arising from the randomization process*):** Apakah alokasi pasien benar-benar acak dan tidak dapat diprediksi (allocation concealment)? Jika alokasi dapat ditebak (misalnya berdasarkan hari kedatangan genap-ganjil), dokter dapat secara sadar atau tidak mengarahkan pasien yang lebih sakit ke kelompok terapi aktif, merusak keseimbangan kelompok sejak awal.

b. **Bias akibat penyimpangan dari intervensi yang dimaksud (*Bias due to deviations from intended interventions*):** Apakah pasien dan personel medis mengetahui intervensi yang diberikan (blinding)? Jika tidak disamarkan, pasien dalam kelompok plasebo mungkin mencari terapi tambahan di luar protokol, atau dokter mungkin memberikan perawatan suportif yang lebih intensif pada kelompok intervensi aktif.

c. **Bias akibat hilangnya data hasil (*Bias due to missing outcome data/Attrition Bias*):** Apakah banyak pasien yang keluar atau hilang dari pengamatan (lost-to-follow-up)? Aturan praktisnya adalah bila kehilangan >20% data, validitas internal sangat terancam. Analisis harus menggunakan prinsip Intention-to-Treat (ITT)—menganalisis pasien sesuai kelompok acak asalnya, meskipun mereka berhenti minum obat—bukan analisis Per-Protocol yang cenderung melebih-lebihkan manfaat terapi.

d. **Bias dalam pengukuran hasil (*Bias in measurement of the outcome/Detection Bias*):** Apakah penilai hasil (outcome assessor) disamarkan? Blinding bagi penilai sangat krusial, terutama jika penilaian bersifat subjektif (misalnya skala nyeri, perbaikan radiologis, atau derajat depresi). Penilai yang tahu kelompok terapi pasien cenderung memberikan skor yang bias secara tidak sadar.

e. **Bias dalam pemilihan hasil yang dilaporkan (*Bias in selection of the reported result/Reporting Bias*):** Apakah peneliti hanya melaporkan hasil yang secara statistik signifikan? Ini dideteksi dengan membandingkan protokol asli uji klinis terdaftar (misalnya di ClinicalTrials.gov) dengan manuskrip yang dipublikasikan untuk memastikan tidak ada perubahan hasil primer (outcome switching) di tengah jalan.

> **Mitos Checklist yang Menyesatkan**
> *Pemberian skor numerik pada validitas studi (misalnya 'Studi ini memiliki skor Jadad 5/5, maka sangat valid') telah ditinggalkan oleh komunitas ilmiah global. Kualitas metodologi tidak bisa dirangkum dalam satu angka. Sebagai gantinya, gunakan pendekatan berbasis domain (seperti Cochrane RoB 2) untuk menilai tingkat risiko bias (Low Risk, Some Concerns, High Risk) pada tiap aspek spesifik demi pemahaman yang kontekstual.*

4. Menilai Kepentingan Klinis: Melampaui Nilai P

Sering kali penulis jurnal medis menekankan signifikansi statistik ($p < 0.05$) untuk menarik perhatian pembaca. Namun, bagi seorang praktisi medis yang menghadapi pasien nyata, signifikansi statistik tidak selalu berarti signifikansi klinis (clinical importance). Untuk membedakannya, kita harus mengevaluasi ukuran efek absolut dan presisi estimasi:

a. **Ukuran Efek Absolut (*Risk Difference* / RD):** Menunjukkan selisih absolut risiko kejadian antara kelompok terapi dan kontrol. RD adalah basis untuk menghitung Number Needed to Treat (NNT = 1 / RD). Nilai NNT memberi tahu kita berapa banyak pasien yang perlu diobati dengan regimen baru selama periode tertentu untuk mencegah satu hasil klinis buruk tambahan.
b. **Ukuran Efek Relatif (RR, OR, HR):** Menggambarkan proporsi pengurangan risiko pada kelompok intervensi dibanding kontrol. Efek relatif ini sering kali disalahgunakan dalam materi pemasaran farmasi untuk membuat efek obat tampak jauh lebih besar daripada yang sebenarnya. Pengurangan risiko relatif (RRR) sebesar 50% bisa berarti pencegahan kejadian dari 2% menjadi 1% (RD hanya 1%, NNT = 100), atau dari 50% menjadi 25% (RD = 25%, NNT = 4).
c. **Lebar Selang Kepercayaan (95% Confidence Interval):** Selang kepercayaan menggambarkan rentang nilai yang masuk akal bagi parameter populasi yang sebenarnya berdasarkan data sampel kita. Jika batas bawah selang kepercayaan 95% sangat dekat dengan nilai ekuivalen (RR = 1 atau RD = 0), maka manfaat terapi tersebut sangat marjinal, meskipun nilai p-nya sangat kecil karena ukuran sampel yang sangat besar.

Tabel 60. Contoh kasus perbandingan signifikansi statistik vs. kepentingan klinis

Skenario Studi	Ukuran Sampel (n)	Penurunan Tekanan Darah Sistolik (Rerata)	95% Confidence Interval (CI)	Keputusan Klinis dan Signifikansi
Penelitian A (Sampel Sangat Besar)	n = 10,000	1.2 mmHg (p = 0.002)	[0.8 to 1.6] mmHg	Signifikan secara statistik (p < 0.05) namun TIDAK penting secara klinis. Penurunan TDS sebesar 1.2 mmHg tidak mengubah stratifikasi risiko stroke pasien secara bermakna.
Penelitian B (Sampel Proporsional)	n = 250	12.5 mmHg (p = 0.001)	[8.4 to 16.6] mmHg	Sangat signifikan secara statistik DAN sangat penting secara klinis. Rentang batas bawah CI (8.4 mmHg) pun tetap memiliki manfaat proteksi kardiovaskular yang kuat.

5. Menilai Aplikabilitas: Menghubungkan Bukti dengan Pasien Nyata

Pilar ketiga dari telaah kritis adalah aplikabilitas (generalizability/external validity). Bukti klinis terbaik yang terbukti valid dan penting di dalam studi RCT terkontrol ketat akan sia-sia jika tidak dapat diterapkan pada pasien nyata yang Anda hadapi di tempat praktik sehari-hari. Evaluasi aplikabilitas memerlukan integrasi keahlian klinis Anda melalui empat langkah praktis:

a. **Langkah 1: Menilai Kesesuaian Subjek (Patient Match):** Apakah karakteristik biologis, patofisiologis, dan demografis pasien Anda berada dalam cakupan kriteria inklusi dan eksklusi populasi studi? Misalnya,

jika pasien Anda adalah wanita berusia 82 tahun dengan multipel komorbiditas, berhati-hatilah menerapkan hasil RCT yang hanya merekrut subjek berusia 40-65 tahun tanpa penyakit penyerta.

b. **Langkah 2: Kelayakan Infrastruktur (Feasibility of Intervention):** Apakah teknologi, obat-obatan, keterampilan operator, dan pemantauan efek samping yang digunakan dalam studi tersebut tersedia dan dapat diakses dengan aman di rumah sakit atau klinik Anda? Terapi mutakhir dengan tuntutan laboratorium canggih tidak dapat diterapkan di puskesmas daerah terpencil.
c. **Langkah 3: Perbandingan Risiko-Manfaat Individual (Individualized Benefit-Harm Balance):** Apakah probabilitas manfaat klinis bersih (net clinical benefit) pada pasien Anda melebihi potensi bahaya efek samping? Seorang pasien dengan risiko perdarahan gastrointestinal yang tinggi mungkin tidak mendapatkan rasio manfaat-mudarat yang menguntungkan dari pemberian terapi antikoagulan ganda baru, meskipun studi menunjukkan penurunan kejadian stroke iskemik secara populasi.
d. **Langkah 4: Nilai dan Preferensi Pasien (Patient Values dan Preferences):** Apakah pasien menyetujui konsekuensi logistik, finansial, dan efek samping dari intervensi tersebut? Keputusan klinis berbasis EBM sejati selalu melibatkan proses diskusi bersama (shared decision-making) antara dokter dan pasien, menghormati hak otonomi pasien atas tubuhnya sendiri.

6. Evaluasi Formatif: Studi Kasus Latihan Mandiri

1) Dalam telaah kritis terhadap uji klinis teracak samar (RCT) yang membandingkan obat antihipertensi baru dengan terapi standar, Anda menemukan bahwa 25% pasien di kelompok intervensi keluar dari studi karena efek samping batuk ringan, sementara kelompok kontrol hanya 5% yang keluar. Di akhir studi, peneliti melakukan analisis Per-Protocol (hanya pada pasien yang menyelesaikan terapi penuh) dan melaporkan bahwa obat baru secara signifikan menurunkan tekanan darah sistolik rata-rata sebesar 15 mmHg dibandingkan kontrol ($p < 0.01$). Apa penilaian kritis Anda terhadap hasil studi ini?
    a. Hasil sangat valid karena analisis Per-Protocol memastikan kepatuhan penuh subjek terhadap intervensi biologis.
    b. Studi ini memiliki risiko attrition bias yang tinggi; analisis Intention-to-Treat (ITT) seharusnya digunakan untuk mencegah bias ini.

c. Nilai $p < 0.01$ membuktikan bahwa randomisasi bekerja sempurna dan mengompensasi hilangnya subjek tersebut.
d. Tingkat dropout 25% adalah hal yang wajar dalam uji klinis skala besar dan tidak memengaruhi validitas internal.

Kunci Jawaban: b

Pembahasan: Kehilangan subjek penelitian (attrition) di atas 20% merupakan ancaman besar bagi validitas internal RCT. Analisis Per-Protocol yang mengecualikan pasien drop-out akan menghancurkan efek randomisasi asli, karena kelompok yang dianalisis tidak lagi setara. Pasien yang bertahan mungkin adalah kelompok dengan daya tahan tubuh lebih baik atau respons yang optimal secara biologis. Prinsip Intention-to-Treat (ITT)—di mana setiap pasien dianalisis berdasarkan alokasi kelompok acak asalnya, terlepas dari kepatuhannya—harus digunakan untuk menjaga integritas randomisasi dan memberikan estimasi efek klinis di dunia nyata yang lebih jujur.

2) Anda membaca sebuah artikel uji klinis obat antikoagulan baru untuk mencegah stroke sekunder pada pasien fibrilasi atrium. Artikel melaporkan bahwa obat baru menurunkan risiko stroke sebesar 50% dibandingkan obat standar (Relative Risk Reduction / RRR = 50%, $p = 0.01$). Setelah Anda menelaah data mentah, kejadian stroke adalah 2% pada kelompok kontrol dan 1% pada kelompok intervensi. Berapakah nilai Number Needed to Treat (NNT) untuk obat baru tersebut?
   a. NNT = 2 (Kita perlu mengobati 2 pasien untuk mencegah 1 stroke).
   b. NNT = 50 (Kita perlu mengobati 50 pasien untuk mencegah 1 stroke).
   c. NNT = 100 (Kita perlu mengobati 100 pasien untuk mencegah 1 stroke).
   d. NNT = 1 (Obat ini bersifat preventif mutlak).

   Kunci Jawaban: c

   Pembahasan: Untuk menghitung NNT, kita harus terlebih dahulu mencari selisih risiko absolut (Risk Difference / RD atau Absolute Risk Reduction / ARR). Kejadian stroke pada kelompok kontrol adalah 2% (0.02) dan pada kelompok intervensi adalah 1% (0.01). Maka, RD = 2% - 1% = 1% (0.01). Nilai NNT dihitung dengan rumus 1 / RD, yaitu 1 / 0.01 = 100. Hal ini berarti kita perlu mengobati 100 pasien fibrilasi atrium dengan obat baru tersebut untuk mencegah terjadinya 1 stroke tambahan. Kasus ini menunjukkan bagaimana pelaporan efek relatif (RRR = 50%) terkesan sangat bombastis secara pemasaran, padahal dampak absolutnya di tingkat klinis lebih sederhana (NNT = 100).

3) Mengapa penyamaran alokasi (allocation concealment) sebelum rekrutmen pasien dianggap sebagai salah satu pelindung validitas internal terkuat dalam uji klinis teracak samar (RCT)?
   a. Karena menyamarkan hasil laboratorium pasien dari dokter yang melakukan analisis data akhir.
   b. Karena mencegah peneliti/dokter secara sadar maupun tidak mengarahkan pasien tertentu (misalnya yang prognosisnya lebih buruk) ke kelompok terapi tertentu.
   c. Karena menjamin bahwa pasien tidak akan mengalami dropout di tengah jalan selama penelitian.
   d. Karena memastikan bahwa ukuran sampel penelitian akan selalu mencapai kekuatan statistika (power) maksimum.

Kunci Jawaban: b

Pembahasan: Penyamaran alokasi (allocation concealment) menjaga agar tidak ada pihak (baik pasien maupun dokter) yang mengetahui kelompok terapi mana yang akan dimasuki pasien sebelum rekrutmen secara resmi disepakati. Jika penyamaran ini rusak (misalnya amplp tidak tembus pandang dibuka sebelum waktunya), dokter dapat secara subjektif memilah pasien: pasien yang sangat sakit diarahkan ke kelompok intervensi baru karena kasihan, sedangkan pasien yang relatif sehat ditaruh di kelompok kontrol. Ini merusak kesetaraan dasar kelompok hasil randomisasi dan memicu bias seleksi yang fatal.

4) Sebuah studi kohort observasional melaporkan adanya hubungan antara kebiasaan konsumsi suplemen vitamin D dengan penurunan risiko kanker kolorektal (Adjusted Hazard Ratio = 0.82, $p < 0.05$). Namun, setelah Anda memeriksa bagian metodologi, studi tersebut tidak menyesuaikan (adjust) variabel status sosioekonomi, aktivitas fisik, dan indeks massa tubuh (BMI). Apa komentar kritis Anda?
   a. Hasil tetap valid sepenuhnya karena Adjusted Hazard Ratio dihitung menggunakan model regresi COX multivariat.
   b. Studi ini mengalami potensi residual confounding yang besar; orang yang secara teratur mengonsumsi vitamin D mungkin juga memiliki gaya hidup yang lebih sehat (olahraga, makan serat) yang menjadi perancu sejati penurunan risiko kanker.
   c. Uji klinis teracak samar (RCT) dan studi observasional memiliki tingkat kerentanan bias perancu yang sama.
   d. Hasil penelitian ini membuktikan bahwa vitamin D secara kausalitas langsung menghambat pertumbuhan sel kanker di usus besar.

Kunci Jawaban: b

Pembahasan: Studi observasional sangat rentan terhadap perancu (confounding) karena subjek tidak dialokasikan secara acak (randomisasi). Jika variabel penting seperti gaya hidup aktif, indeks massa tubuh, dan status sosial ekonomi tidak disesuaikan dalam model statistik, maka akan timbul bias residual confounding. Sangat mungkin hubungan yang diamati bukan karena khasiat suplemen vitamin D, melainkan karena perilaku hidup sehat lainnya yang terasosiasi erat dengan konsumsi suplemen.

5) Seorang dokter ingin menerapkan hasil penelitian terapi imunologi mutakhir untuk kanker paru stadium IV yang dipublikasikan di jurnal internasional ternama pada pasiennya di puskesmas kecamatan. Terapi tersebut membutuhkan pemantauan laboratorium khusus setiap minggu dan penanganan efek samping badai sitokin di unit ICU tingkat lanjut. Menurut pilar 'aplikabilitas' telaah kritis, apa langkah keputusan terbaik dokter tersebut?
   a. Tetap memberikan terapi tersebut karena dipublikasikan di jurnal internasional bereputasi tinggi dengan hasil valid.
   b. Menolak menerapkannya karena infrastruktur puskesmas tidak memenuhi kelayakan operasional (feasibility) untuk memantau efek samping secara aman.
   c. Meminta pasien membeli obat tersebut sendiri di apotek luar tanpa perlu melakukan pemantauan laboratorium.
   d. Mengabaikan aspek aplikabilitas karena signifikansi statistik ($p < 0.05$) adalah indikator kepastian mutlak di atas tempat tidur.

Kunci Jawaban: b

Pembahasan: Pilar ketiga telaah kritis adalah aplikabilitas klinis yang mencakup penilaian kelayakan fasilitas setempat (feasibility). Meskipun hasil studi valid secara metodologi dan memiliki kepentingan klinis tinggi di pusat medis tersier, terapi tersebut sangat berbahaya jika dilakukan di tingkat puskesmas primer tanpa fasilitas ICU dan pemantauan laboratorium mingguan. Mengorbankan keselamatan pasien dengan memaksakan intervensi tanpa sistem proteksi efek samping melanggar prinsip etik kedokteran non-maleficence (first, do no harm).

BAB 10
STATISTIKA KOMPUTER DAN INTERPRETASI OUTPUT

Menjembatani Teori ke Praktik Klinis: Membaca Hasil Analisis Software Tanpa Kehilangan Logika Medis

1. Prinsip Utama: "Garbage In, Garbage Out" (GIGO)

Ketersediaan software statistik modern (seperti SPSS, Jamovi, Stata, R, atau GraphPad Prism) memudahkan mahasiswa kedokteran untuk melakukan analisis statistik hanya dengan beberapa klik. Namun, kemudahan operasional ini menyimpan risiko besar jika pengguna tidak memahami logika klinis di balik data. Prinsip utama statistika komputer adalah **"Garbage In, Garbage Out" (GIGO)** jika data yang dimasukkan memiliki kualitas buruk, tidak terstandardisasi, atau tidak memiliki landasan teoretis klinis yang jelas, maka keluaran (output) dari software tersebut tidak akan memiliki nilai klinis sama sekali.

Sebuah software statistik tidak memiliki penalaran klinis (clinical reasoning). Software akan patuh menjalankan perintah regresi atau uji t pada variabel apa pun yang Anda masukkan, bahkan jika secara biologis hubungan tersebut tidak masuk akal. Oleh karena itu, peran utama seorang dokter peneliti bukanlah sekadar menjadi operator komputer, melainkan menjadi penilai kritis sebelum data diinput, serta penafsir yang kritis ketika output komputer telah diterbitkan.

> **Pesan Kunci Pedagogis (Bayesian Thinking)**
> Komputer adalah kalkulator raksasa, bukan dokter. Software statistik mampu mendeteksi korelasi numerik yang sangat kuat antara ***jumlah stetoskop yang dimiliki seorang dokter dengan kesembuhan pasien***, tetapi ia tidak tahu bahwa hubungan tersebut bersifat non-kausal (spurious). Logika klinis dan struktur pemodelan yang matang adalah satu-satunya pembatas agar data tidak menyesatkan keputusan klinis Anda.

2. Struktur dan Penyiapan Data: Tidy Data Format

Sebelum data dapat diproses oleh software statistik, struktur lembar sebar (spreadsheet) harus memenuhi kaidah **Tidy Data**. Banyak peneliti pemula melakukan kesalahan fatal dengan membuat format tabel yang

ramah dibaca manusia secara visual tetapi tidak dapat dipahami oleh mesin komputer. Tiga aturan emas dalam penyiapan data statistik komputer:

a. Setiap baris merupakan satu observasi tunggal (misalnya, satu pasien unik).
b. Setiap kolom merupakan satu variabel unik (misalnya, Usia, Jenis_Kelamin, Tekanan_Darah_Sistolik).
c. Setiap sel hanya berisi satu nilai tunggal (hindari menggabungkan teks seperti '120/80 mmHg' ke dalam satu kolom numerik).

Tabel 61. Perbandingan format penyusunan data Untidy vs. Tidy

No Pasien	TD Sistolik (mmHg)	Kelompok Intervensi	Status Klinis
001	120	A (Standard)	Sembuh
002	145	B (Baru)	Meninggal
003	130	A (Standard)	Sembuh
004	160	B (Baru)	Sembuh

3. Panduan Integratif Membaca Output Multivariat

Saat melakukan pemodelan multivariat, software statistik akan menerbitkan tabel koefisien yang cukup padat. Kemampuan membaca dan menafsirkan tabel ini secara akurat sangat krusial dalam menyajikan bukti ilmiah. Berikut adalah tiga jenis output utama yang paling sering digunakan dalam studi kedokteran.

a. Output Regresi Linear Multivariat (Outcome Kontinu)

Digunakan saat variabel dependen (outcome) bersifat kontinu (misalnya, tekanan darah dalam mmHg, kadar HbA1c, atau lama rawat inap dalam hari). Parameter kunci yang harus dipahami adalah koefisien beta (unstandardized coefficient B), nilai p, dan R-squared (R^2).

Tabel 62. Contoh output regresi linear multivariat (Outcome: Kadar HbA1c %)

Variabel Independen	Koefisien (B)	Standard Error	Nilai p	95% Confidence Interval
Intersep (Konstanta)	8.24	0.35	<0.001	7.55 s.d. 8.93
Terapi Baru (vs. Standard)	-0.75	0.15	<0.001	-1.04 s.d. -0.46
Usia Pasien (per Tahun)	0.02	0.01	0.045	0.001 s.d. 0.039

Catatan: Nilai R-squared (R^2) model ini adalah 0.38, menunjukkan bahwa kombinasi variabel terapi baru dan usia pasien mampu menjelaskan 38% variabilitas kadar HbA1c pasien.

b. Output Regresi Logistik Multivariat (Outcome Biner)

Digunakan ketika outcome bersifat dikotomis/biner (misalnya, kesembuhan pasien [ya/tidak], mortalitas [ya/tidak], atau relaps [ya/tidak]). Output yang diterbitkan oleh software harus ditranslasikan dari koefisien log-odds (B) menjadi Odds Ratio (OR) dengan melakukan eksponensial (e^B).

Tabel 63. Contoh output regresi logistik multivariat (outcome: mortalitas 30 hari)

Variabel Independen	Odds Ratio (OR)	Standard Error	Nilai p	95% Confidence Interval
Skor SOFA (per unit kenaikan)	1.25	0.08	<0.001	1.12 s.d. 1.39
Usia > 65 Tahun (vs. <= 65)	2.10	0.22	0.003	1.36 s.d. 3.24
Vasinasi Sepsis (vs. Tidak)	0.45	0.18	0.012	0.24 s.d. 0.84

4. Miskonsepsi Umum dan Cara Menghindarinya

Mahasiswa kedokteran sering kali tergelincir pada perangkap interpretasi otomatis saat membaca hasil software. Dua miskonsepsi yang paling sering terjadi adalah:

a. **Menganggap Nilai p < 0.05 Sebagai Bukti Efek Klinis yang Besar:** Nilai p hanya mengukur kekuatan bukti terhadap hipotesis nol. Pada sampel berukuran sangat besar, efek klinis yang sangat kecil dan tidak relevan bagi pasien (misalnya, penurunan tekanan darah sebesar 0.5 mmHg) bisa menghasilkan nilai p < 0.001. Selalu rujuk ukuran efek (b, OR, atau HR) dan rentang klinis 95% CI.

b. **Multikolinearitas yang Diabaikan:** Memasukkan beberapa variabel yang mengukur aspek biologis serupa secara bersamaan (misalnya, kadar urea darah dan serum kreatinin) dapat merusak model regresi multivariat Anda, menyebabkan standar error melonjak dan arah koefisien berubah secara tidak logis.

5. Latihan Formatif: Membaca Output dan Studi Kasus

1) Seorang dokter melakukan analisis regresi linear multivariat untuk memprediksi kadar glukosa darah puasa (mg/dL) pasien diabetes

melitus tipe 2. Model menghasilkan koefisien B untuk variabel indeks massa tubuh (IMT) sebesar 3.4 (p = 0.012, 95% CI: 0.8 s.d. 6.0). Apa arti koefisien tersebut secara klinis?

a. Setiap kenaikan 1 unit IMT diasosiasikan dengan kenaikan rerata kadar glukosa darah puasa sebesar 3.4 mg/dL setelah mengontrol variabel lain dalam model.
b. Sebanyak 3.4% variasi kadar glukosa darah puasa dapat dijelaskan oleh variasi IMT pasien.
c. IMT pasien diabetes memiliki probabilitas sebesar 3.4% untuk memicu komplikasi metabolik fatal.
d. Kadar glukosa darah puasa pasien akan meningkat secara eksponensial sebesar e^3.4 untuk setiap unit kenaikan IMT.

Kunci Jawaban: a

Pembahasan Rasional: Koefisien B unstandardized pada regresi linear merepresentasikan perubahan rerata pada variabel outcome (glukosa darah) untuk setiap unit kenaikan pada variabel prediktor (IMT) setelah disesuaikan dengan variabel lain di dalam model.

2) Pada Tabel 10.3 (regresi logistik mortalitas 30 hari), variabel 'Vaksinasi Sepsis' memiliki Odds Ratio (OR) sebesar 0.45 (95% CI: 0.24 s.d. 0.84, p = 0.012). Bagaimana interpretasi klinis yang tepat terhadap temuan ini?

a. Pemberian vaksinasi sepsis meningkatkan risiko kematian sebesar 45%.
b. Pasien yang menerima vaksinasi memiliki odds kematian 30 hari sebesar 55% lebih rendah dibandingkan kelompok non-vaksinasi setelah disesuaikan dengan skor SOFA dan usia.
c. Efek perlindungan vaksinasi sepsis tidak signifikan karena batas bawah CI di bawah 0.50.
d. Komputer salah menghitung karena tidak mungkin Odds Ratio bernilai kurang dari 1.

Kunci Jawaban: b

Pembahasan Rasional: Nilai OR < 1 menunjukkan efek protektif. Nilai OR 0.45 berarti odds mortalitas pada kelompok intervensi hanya 0.45 kali (atau reduksi odds sebesar 55%) dibandingkan kelompok pembanding. Karena rentang 95% CI (0.24-0.84) tidak mencakup nilai nol atau 1, efek protektif ini signifikan secara statistik.

3) Mengapa seorang klinisi tidak boleh hanya melaporkan nilai p-value tanpa menyertakan ukuran efek (effect size) dan confidence interval (CI)?

a. Karena nilai p-value sangat dipengaruhi oleh ukuran sampel (sample size); sampel yang sangat besar dapat memicu temuan p-value signifikan secara statistik meski efek klinisnya sangat kecil.
b. Karena p-value tidak mencerminkan arah pengaruh asosiasi variabel.
c. Karena program komputer SPSS atau Jamovi tidak dapat menyelesaikan perhitungan statistik jika CI tidak diaktifkan secara manual.
d. Jawaban A dan B benar.

Kunci Jawaban: d

Pembahasan Rasional: Nilai p-value murni fungsional terhadap ukuran sampel dan tidak menunjukkan arah atau kekuatan dampak klinis riil bagi pasien. Tanpa arah (positif/negatif) dan interval keyakinan, temuan penelitian kehilangan nilai kepraktisan klinisnya.

4) Apa konsekuensi memasukkan dua variabel independen yang sangat berkorelasi erat (seperti LDL dan Kolesterol Total) ke dalam satu model regresi linear multivariat?
   a. Meningkatkan kekuatan prediksi (R-squared) secara eksponensial.
   b. Menyebabkan multikolinearitas, merusak kestabilan estimasi parameter, memperlebar standar error (SE), dan menghasilkan nilai p yang menyesatkan.
   c. Mengubah model regresi linear secara otomatis menjadi model regresi logistik.
   d. Menyebabkan data berdistribusi normal secara artifisial.

Kunci Jawaban: b

Pembahasan Rasional: Multikolinearitas terjadi ketika dua atau lebih variabel prediktor saling berkorelasi sangat kuat. Ini melemahkan daya pisang model dan dapat menyebabkan estimasi koefisien menjadi sangat tidak stabil.

5) Manakah dari format berikut yang memenuhi kaidah Tidy Data untuk software statistik komputer?
   a. Kolom 'Tekanan Darah' berisi teks gabungan '130/90' dan '140/95'.
   b. Kolom 'Mortalitas' berisi status numerik biner (0 untuk hidup, 1 untuk meninggal), di mana setiap baris mewakili satu pasien unik.
   c. Setiap baris diisi dengan nama dokter pemeriksa, dan data masing-masing pasien digabung dalam sel yang sama.
   d. Menyatukan kolom intervensi obat dan kelompok kontrol di dalam satu baris pasien yang sama secara berulang.

Kunci Jawaban: b

Pembahasan Rasional: Aturan emas tidy data mensyaratkan setiap baris mewakili satu observasi tunggal (satu pasien), setiap kolom adalah satu variabel (mortalitas dikodekan numerik biner 0/1), dan sel berisi nilai tunggal.

DAFTAR PUSTAKA

Akobeng, A.K. (2007). Understanding Diagnostic Tests 3: Receiver Operating Characteristic Curves. Acta Paediatrica, 96(5), 644–647. Https://Doi.Org/10.1111/J.1651-2227.2006.00178.X

Altman, D.G., Bland, J.M. (1994). Diagnostic Tests 1: Sensitivity and Specificity. BMJ, 308, 1552. Https://Doi.Org/10.1136/Bmj.308.6943.1552

Altman, D.G., Bland, J.M. (1994b). Diagnostic Tests 2: Predictive Values. BMJ, 309, 102. Https://Doi.Org/10.1136/Bmj.309.6947.102

Altman, D.G., Bland, J.M. (1995a). Absence of Evidence is Not Evidence of Absence. BMJ, 311(7003), 485. Https://Doi.Org/10.1136/Bmj.311.7003.485

Altman, D.G., Bland, J.M. (1995). The Normal Distribution. BMJ, 310(6975), 298. Https://Doi.Org/10.1136/Bmj.310.6975.298

Altman, D.G., Bland, J.M. (1996). Comparing Several Groups using Analysis of Variance. BMJ, 312(7044), 1472. Https://Doi.Org/10.1136/Bmj.312.7044.1472

Altman, D.G., Bland, J.M. (1996). Detecting Skewness from Summary Information. BMJ, 313(7066), 1200. Https://Doi.Org/10.1136/Bmj.313.7066.1200

Altman, D.G., Bland, J.M. (2005). Standard Deviations and Standard Errors. BMJ, 331(7521), 903. Https://Doi.Org/10.1136/Bmj.331.7521.903

Altman, D.G., Bland, J.M. (2009). Parametric V Non-Parametric Methods for Data Analysis. BMJ, 338, A3167. Https://Doi.Org/10.1136/Bmj.A3167

Bossuyt, P.M., Reitsma, J.B., Bruns, D.E., Gatsonis, C.A., Glas, A.S., Irwig, L., Lijmer, J.G., Moher, D., Rennie, D., De Vet, H.C.W., Kressel, H.Y., Rifai, N., Golub, R.M., Altman, D.G., Hooft, L., Korevaar, D.A., Cohen, J.F. (2015). STARD 2015: An Updated List of Essential Items for Reporting Diagnostic Accuracy Studies. BMJ, 351, H5527. Https://Doi.Org/10.1136/Bmj.H5527

Cohen, J.F., Korevaar, D.A., Altman, D.G., Bruns, D.E., Gatsonis, C.A., Hooft, L., Irwig, L., Levine, D., Reitsma, J.B., De Vet, H.C.W., Bossuyt, P.M.M. (2016). STARD 2015 Guidelines for Reporting Diagnostic Accuracy Studies: Explanation and Elaboration. BMJ Open, 6(11), E012799. Https://Doi.Org/10.1136/Bmjopen-2016-012799

Collins, G.S., Moons, K.G.M., Dhiman, P., Riley, R.D., Beam, A.L., Van Calster, B., Ghassemi, M., Liu, X., Reitsma, J.B., Van Smeden, M., Boulesteix, A.L., Camaradou, J.C., Celi, L.A., Denaxas, S., Denniston, A.K., Glocker, B.,

Golub, R.M., Harvey, H., Heinze, G., Logullo, P. (2024). TRIPOD+AI Statement: Updated Guidance for Reporting Clinical Prediction Models That Use Regression or Machine Learning Methods. BMJ, 385, E078378. Https://Doi.Org/10.1136/Bmj-2023-078378

Collins, G.S., Reitsma, J.B., Altman, D.G., Moons, K.G.M. (2015). Transparent Reporting of A Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD): The TRIPOD Statement. BMJ, 350, G7594. Https://Doi.Org/10.1136/Bmj.G7594

Cox, D.R. (1972). Regression Models and Life-Tables. Journal of The Royal Statistical Society: Series B (Methodological), 34(2), 187–202. Https://Doi.Org/10.1111/J.2517-6161.1972.Tb00899.X

Dans, A.L., Dans, L.F., Guyatt, G.H., Richardson, S. (1998). Users' Guides to The Medical Literature: XIV. How to Decide on the Applicability of Clinical Trial Results to Your Patient. JAMA, 279(7), 545–549. Https://Doi.Org/10.1001/Jama.279.7.545

Deeks, J.J., Altman, D.G. (2004). Diagnostic Tests 4: Likelihood Ratios. BMJ, 329, 168–169. Https://Doi.Org/10.1136/Bmj.329.7458.168

Fine, J.P., Gray, R.J. (1999). A Proportional Hazards Model for The Subdistribution of A Competing Risk. Journal of The American Statistical Association, 94(446), 496–509. Https://Doi.Org/10.1080/01621459.1999.10474144

Gardner, M.J., Altman, D.G. (1986). Confidence Intervals Rather Than P Values: Estimation Rather Than Hypothesis Testing. British Medical Journal, 292(6522), 746–750. Https://Doi.Org/10.1136/Bmj.292.6522.746

Greenland, S., Pearl, J., Robins, J.M. (1999). Causal Diagrams for Epidemiologic Research. Epidemiology, 10(1), 37–48. Https://Doi.Org/10.1097/00001648-199901000-00008

Greenland, S., Senn, S.J., Rothman, K.J., Carlin, J.B., Poole, C., Goodman, S.N., Altman, D.G. (2016). Statistical Tests, P Values, Confidence Intervals, and Power: A Guide to Misinterpretations. European Journal of Epidemiology, 31, 337–350. Https://Doi.Org/10.1007/S10654-016-0149-3

Harrell, F.E. (2015). Regression Modeling Strategies: with Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis. Springer. Https://Doi.Org/10.1007/978-3-319-19425-7

Hart, A. (2001). Mann-Whitney Test is Not Just A Testof Medians: Differences in Spread Can Be Important. BMJ, 323(7309), 391–393. Https://Doi.Org/10.1136/Bmj.323.7309.391

Hernán, M.A., Robins, J.M. (2020). Causal Inference: What If. Chapman dan Hall/CRC. Https://Miguelhernan.Org/Whatifbook

Hopewell, S., Chan, A.W., Collins, G.S., Hróbjartsson, A., Moher, D., Schulz, K. F., Tunn, R., Aggarwal, R., Berkwits, M., Berlin, J.A., Boutron, I., Cates, C.J., Cook, J.A., Creinin, M.D., Dickersin, K., Elbourne, D., Fiedler, M., Flanagin, A., Glasziou, P.P., Altman, D.G. (2025). CONSORT 2025 Statement: Updated Guideline For Reporting Randomised Trials. BMJ, 389, E081123. Https://Doi.Org/10.1136/Bmj-2024-081123

Hopewell, S., Chan, A.W., Collins, G.S., Hróbjartsson, A., Moher, D., Schulz, K. F., Tunn, R., Aggarwal, R., Berkwits, M., Berlin, J.A., Bhandari, N., Butcher, N.J., Campbell, M.K., Chidebe, R.C.W., Elbourne, D., Farmer, A., Fergusson, D.A., Golub, R.M., Goodman, S.N., Boutron, I. (2025). CONSORT 2025 Explanation and Elaboration: Updated Guideline for Reporting Randomised Trials. BMJ, 389, E081124. Https://Doi.Org/10.1136/Bmj-2024-081124

Hróbjartsson, A., Boutron, I., Hopewell, S., Moher, D., Schulz, K.F., Collins, G. S., Tunn, R., Aggarwal, R., Berkwits, M., Berlin, J.A., Bhandari, N., Butcher, N.J., Campbell, M.K., Chidebe, R.C.W., Elbourne, D.R., Farmer, A.J., Fergusson, D.A., Golub, R.M., Goodman, S.N., Chan, A.W. (2025). SPIRIT 2025 Explanation and Elaboration: Updated Guideline for Protocols of Randomised Trials. BMJ, 389, E081660. Https://Doi.Org/10.1136/Bmj-2024-081660

International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use. (2019). ICH E9(R1): Addendum on Estimands And Sensitivity Analysis In Clinical Trials To The Guideline on Statistical Principles for Clinical Trials. Https://Database.Ich.Org/Sites/Default/Files/E9-R1_Step4_Guideline_2019_1203.Pdf

Jaeschke, R., Singer, J., Guyatt, G.H. (1989). Measurement of Health Status: Ascertaining The Minimal Clinically Important Difference. Controlled Clinical Trials, 10(4), 407–415. Https://Doi.Org/10.1016/0197-2456(89)90005-6

Kaplan, E.L., Meier, P. (1958). Nonparametric Estimation from Incomplete Observations. Journal of The American Statistical Association, 53(282), 457–481. Https://Doi.Org/10.1080/01621459.1958.10501452

Knol, M.J., Egger, M., Scott, P., Geerlings, M.I., Vandenbroucke, J.P. (2009). When One Depends on The Other: Reporting of Interaction in Case-Control and Cohort Studies. Epidemiology, 20(2), 161–166. Https://Doi.Org/10.1097/EDE.0b013e31818f6651

Knol, M.J., Vanderweele, T.J. (2012). Recommendations for Presenting Analyses of Effect Modification and Interaction. International Journal of Epidemiology, 41(2), 514–520. Https://Doi.Org/10.1093/Ije/Dyr218

Lachat, C., Hawwash, D., Ocké, M.C., Berg, C., Forsum, E., Hörnell, A., Larsson, CL., Sonestedt, E., Wirfält, E., Åkesson, A., Kolsteren, P., Byrnes, G., De Keyzer, W., Van Camp, J. (2016). Strengthening The Reporting of Observational Studies in Epidemiology Nutritional Epidemiology *(STROBE-Nut):* An Extension of The STROBE Statement. PLOS Medicine, 13(6), E1002036. Https://Doi.Org/10.1371/Journal.Pmed.1002036

Lang, T.A., Altman, D.G. (2015). Basic Statistical Reporting for Articles Published in Biomedical Journals: The "Statistical Analyses and Methodsin The Published Literature" or The SAMPL Guidelines. International Journal of Nursing Studies, 52(1), 5–9. Https://Doi.Org/10.1016/J.Ijnurstu.2014.09.006

Liang, K.Y., Zeger, S.L. (1986). Longitudinal Data Analysis Using Generalized Linear Models. Biometrika, 73(1), 13–22. Https://Doi.Org/10.1093/Biomet/73.1.13

Little, R.J.A., Rubin, D.B. (2019). Statistical Analysis with Missing Data (3rd Ed.). Wiley. Https://Doi.Org/10.1002/9781119482260

Lumley, T., Diehr, P., Emerson, S., Chen, L. (2002). The Importance of The Normality Assumption in Large Public Health Data Sets. Annual Review of Public Health, 23, 151–169. Https://Doi.Org/10.1146/Annurev.Publhealth.23.100901.140546

Murad, M.H., Montori, V.M., Ioannidis, J.P.A., Jaeschke, R., Devereaux, P.J., Prasad, K., Neumann, I., Carrasco-Labra, A., Agoritsas, T., Hatala, R., Meade, M.O., Wyer, P., Cook, D.J., Guyatt, G. (2014). How to Read A Systematic Review And Meta-Analysis And Apply The Results To Patient Care: Users' Guides To The Medical Literature. JAMA, 312(2), 171–179. Https://Doi.Org/10.1001/Jama.2014.5559

Page, M.J., Mckenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., Brennan, S.E., Chou, R., Glanville, J., Grimshaw, J.M., Hróbjartsson, A., Lalu, MM., Li, T., Loder, E. W., Mayo-Wilson, E., Mcdonald, S., Moher, D. (2021). The PRISMA 2020 Statement: An Updated Guideline For Reporting Systematic Reviews. BMJ, 372, N71. Https://Doi.Org/10.1136/Bmj.N71

Pearl, J., Mackenzie, D. (2018). The Book of Why: The New Science of Cause and Effect. Basic Books.

Pepe, M.S. (2003). The Statistical Evaluation of Medical Tests for Classification and Prediction. Oxford University Press. Https://Doi.Org/10.1093/Oso/9780198509844.001.0001

Richardson, W.S., Wilson, M.C., Nishikawa, J., Hayward, R.S. (1995). The Well-Built Clinical Question: A Key to Evidence-Based Decisions. ACP Journal Club, 123(3), A12–A13. Https://Doi.Org/10.7326/ACPJC-1995-123-3-A12

Rohrer, J.M. (2018). Thinking Clearly About Correlations and Causation: Graphical Causal Models for Observational Data. Advances in Methods and Practices in Psychological Science, 1(1), 27–42. Https://Doi.Org/10.1177/2515245917745629

Royston, P., Parmar, M.K.B. (2013). Restricted Mean Survival Time: An Alternative to The Hazard Ratio for The Design and Analysis of Randomized Trials with A Time-To-Event Outcome. BMC Medical Research Methodology, 13, 152. Https://Doi.Org/10.1186/1471-2288-13-152

Ruxton, G.D. (2006). The Unequal Variance T-Test is An Underused Alternative to Student's T-Test and The Mann–Whitney U Test. Behavioral Ecology, 17(4), 688–690. Https://Doi.Org/10.1093/Beheco/Ark016

Sterne, J.A.C., Hernán, M.A., Reeves, B.C., Savović, J., Berkman, N.D., Viswanathan, M., Henry, D., Altman, D.G., Ansari, M.T., Boutron, I., Carpenter, J.R., Chan, A.W., Churchill, R., Deeks, J.J., Hróbjartsson, A., Kirkham, J., Jüni, P., Loke, Y.K., Pigott, T.D., ... Higgins, J.P.T. (2016). ROBINS-I: A Tool for Assessing Risk of Bias in Non-Randomised Studies of Interventions. BMJ, 355, I4919. Https://Doi.Org/10.1136/Bmj.I4919

Sterne, J.A.C., Savović, J., Page, M.J., Elbers, R.G., Blencowe, N.S., Boutron, I., Cates, C.J., Cheng, H.Y., Corbett, M.S., Eldridge, S.M., Emberson, J.R., Hernán, M.A., Hopewell, S., Hróbjartsson, A., Junqueira, D.R., Jüni, P., Kirkham, J.J., Lasserson, T., Li, T., ... Higgins, J.P.T. (2019). Rob 2: A Revised Tool for Assessing Risk of Bias in Randomised Trials. BMJ, 366, L4898. Https://Doi.Org/10.1136/Bmj.L4898

Swinscow, T.D.V., Campbell, M.J. (1997). Statistics At Square One (9th Ed.). BMJ Publishing Group. Https://Www.Bmj.Com/About-Bmj/Resources-Readers/Publications/Statistics-Square-One

Vandenbroucke, J.P., Von Elm, E., Altman, D.G., Gøtzsche, P.C., Mulrow, C.D., Pocock, S.J., Poole, C., Schlesselman, J.J., Egger, M. (2007). Strengthening The Reporting of Observational Studies in Epidemiology (STROBE): Explanation and Elaboration. PLOS

Medicine, 4(10), E297. Https://Doi.Org/10.1371/Journal.Pmed.0040297

Vanderweele, T.J. (2019). Principles of Confounder Selection. European Journal of Epidemiology, 34(3), 211–219. Https://Doi.Org/10.1007/S10654-019-00494-6

Vittinghoff, E., Glidden, D.V., Shiboski, SC., Mcculloch, C.E. (2012). Regression Methods in Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models (2nd Ed.). Springer. Https://Doi.Org/10.1007/978-1-4614-1353-0

Von Elm, E., Altman, D.G., Egger, M., Pocock, S.J., Gøtzsche, P.C., Vandenbroucke, J.P. (2007). The Strengthening The Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for Reporting Observational Studies. The Lancet, 370(9596), 1453–1457. Https://Doi.Org/10.1016/S0140-6736(07)61602-X

Wasserstein, R.L., Lazar, N.A. (2016). The ASA's Statement on P-Values: Context, Process, and Purpose. The American Statistician, 70(2), 129–133. Https://Doi.Org/10.1080/00031305.2016.1154108

Whiting, P.F., Rutjes, A.W.S., Westwood, M.E., Mallett, S., Deeks, J.J., Reitsma, J. B., Leeflang, M.M.G., Sterne, J.A.C., Bossuyt, P.M.M. (2011). QUADAS-2: A Revised Tool for The Quality Assessment of Diagnostic Accuracy Studies. Annals of Internal Medicine, 155(8), 529–536. Https://Doi.Org/10.7326/0003-4819-155-8-201110180-00009

Zweig, M.H., Campbell, G. (1993). Receiver-Operating Characteristic (ROC) Plots: A Fundamental Evaluation Tool in Clinical Medicine. Clinical Chemistry, 39(4), 561–577. Https://Doi.Org/10.1093/Clinchem/39.4.561

GLOSARIUM ISTILAH

Istilah / Terminologi (APA Style)	Definisi dan Konteks Penalaran Medis
Absolute Risk Reduction (ARR)	Penurunan Risiko Absolut. Selisih absolut laju insiden kejadian klinis (mortalitas, efek samping) antara kelompok kontrol dengan kelompok intervensi. Metrik ini sangat krusial karena merefleksikan manfaat klinis riil di sisi tempat tidur pasien.
Adjustment (Penyesuaian)	Proses statistik multivariat untuk mengontrol pengaruh variabel perancu (confounder) pada model statistik, sehingga diperoleh estimasi kekuatan hubungan atau efek intervensi yang murni (adjusted estimate).
Association (Asosiasi)	Pola ketergantungan statistik antara dua variabel. Keberadaan asosiasi tidak secara otomatis membuktikan adanya hubungan sebab-akibat (kausalitas) biologis.
Attrition Bias	Bias akibat berkurangnya subjek penelitian selama masa tindak lanjut (lost to follow-up). Attrition bias yang tinggi merusak validitas internal karena karakteristik kelompok intervensi dan kontrol tidak lagi seimbang.
Bayesian Reasoning	Metode penalaran klinis yang memperbarui tingkat kepercayaan terhadap suatu diagnosis (probabilitas pasca-tes) setelah memperoleh informasi atau hasil pemeriksaan penunjang baru.
Bias	Kesalahan sistematis dalam desain, pelaksanaan, atau analisis studi yang menghasilkan estimasi yang secara konsisten menyimpang dari nilai kebenaran populasi yang sesungguhnya.
Biner (Binary Data)	Tipe data kategorikal yang hanya memiliki dua kemungkinan nilai yang saling lepas (misalnya: hidup/mati, sembuh/gagal, positif/negatif).
Calibration (Kalibrasi)	Derajat kesesuaian antara probabilitas kejadian klinis yang diprediksi oleh model statistik dengan frekuensi kejadian klinis riil yang terobservasi pada kelompok pasien.
Causal Effect (Efek Kausal)	Perubahan pada nilai outcome yang terjadi secara eksklusif akibat adanya manipulasi

	atau intervensi pada variabel pajanan, dengan mengasumsikan kondisi lainnya tetap konstan.
Censoring (Penyensoran)	Kondisi khas pada analisis data survival di mana waktu terjadinya peristiwa klinis (event of interest) tidak dapat teramati secara presisi akibat pasien hilang dari pemantauan atau studi berakhir sebelum kejadian terjadi.
C-index (Harrell's Concurrence)	Metrik untuk mengukur kemampuan diskriminasi model prediksi prognosis, berkisar antara 0.5 (tebakan acak) hingga 1.0 (diskriminasi sempurna).
Collider	Variabel yang merupakan efek bersama dari dua variabel lainnya dalam suatu diagram kausal (DAG). Melakukan adjustment terhadap collider secara keliru akan memicu bias seleksi (collider-stratification bias).
Confidence Interval (CI)	Rentang nilai terhitung dari sampel yang menggambarkan ketidakpastian estimasi parameter populasi pada tingkat kepercayaan tertentu (biasanya 95%). CI yang sempit merefleksikan presisi estimasi yang tinggi.
Confounder (Faktor Perancu)	Variabel ketiga yang berhubungan secara independen dengan pajanan (exposure) sekaligus dengan hasil klinis (outcome), tetapi bukan merupakan perantara (mediator) di jalur biologis keduanya.
Correlation (Korelasi)	Metrik statistik bivariat (seperti koefisien Pearson atau Spearman) yang mengukur kekuatan dan arah hubungan linear antara dua variabel kontinu.
Cox Proportional Hazards	Model regresi semiparametrik untuk data survival multivariat yang digunakan untuk mengestimasi Hazard Ratio (HR) dengan asumsi bahwa rasio hazard antarkelompok konstan sepanjang waktu.
Directed Acyclic Graph (DAG)	Diagram grafis terarah tanpa siklus yang memetakan hubungan sebab-akibat teoritis antarkovariat untuk memandu pemilihan variabel perancu yang wajib dikontrol secara objektif.
Discrimination (Diskriminasi)	Kemampuan suatu model prognosis atau tes diagnostik untuk membedakan secara akurat antara pasien yang akan mengalami kejadian klinis dengan yang tidak.

Estimand	Target kuantitatif yang didefinisikan secara klinis secara terperinci sebelum analisis data dimulai, menetapkan populasi, intervensi, outcome, interaksi antarvariabel, dan strategi penanganan data hilang.
Exposure (Pajanan)	Variabel independen berupa faktor risiko, karakteristik pasien, atau intervensi medis yang dievaluasi pengaruhnya terhadap kejadian klinis (outcome).
Garbage In, Garbage Out (GIGO)	Prinsip komputasi statistik bahwa kualitas dan validitas hasil analisis komputer (output) sepenuhnya dibatasi oleh kualitas, akurasi, dan struktur data yang diinput.
Hazard Ratio (HR)	Rasio laju instan terjadinya suatu peristiwa klinis (event) per unit waktu pada kelompok intervensi dibandingkan kelompok kontrol dalam pemodelan survival.
Interaction (Interaksi)	Kondisi di mana efek dari suatu pajanan atau intervensi terhadap outcome klinis bervariasi nilainya tergantung pada keberadaan atau level variabel ketiga (efek modifikasi).
Internal Validity	Derajat kepastian bahwa perbedaan hasil klinis yang teramati antara kelompok studi benar-benar disebabkan oleh intervensi yang diuji, bukan akibat bias atau perancu.
Kaplan-Meier Estimator	Metode non-parametrik yang digunakan untuk mengestimasi dan memplot kurva probabilitas kelangsungan hidup (survival probability) sepanjang waktu pengamatan.
Likelihood Ratio (LR)	Rasio kemungkinan yang menunjukkan seberapa besar hasil pemeriksaan penunjang (positif atau negatif) akan menggeser probabilitas diagnosis klinis pasien.
Log-Rank Test	Uji hipotesis non-parametrik yang digunakan untuk membandingkan keseluruhan profil kelangsungan hidup (survival curves) di antara dua atau lebih kelompok independen.
Mediator	Variabel antara yang berada di dalam jalur biologis/mekanis hubungan sebab-akibat antara pajanan (exposure) dan hasil klinis (outcome).
Missing Data	Kondisi tidak lengkapnya nilai variabel pada subjek penelitian akibat kesalahan entri data, ketidakhadiran pasien, atau kegagalan alat ukur.

Multicollinearity	Adanya korelasi linear yang sangat kuat di antara dua atau lebih variabel prediktor independen dalam model regresi multivariat, yang merusak presisi estimasi koefisien.
Nominal Data	Tipe data kategorikal tanpa memiliki urutan tingkat atau hierarki biologis bawaan (misalnya: golongan darah, jenis kelamin, pilihan spesialisasi).
Number Needed to Treat (NNT)	Jumlah pasien rata-rata yang harus menerima intervensi aktif selama periode waktu tertentu untuk mencegah terjadinya satu kasus hasil klinis buruk tambahan.
Odds Ratio (OR)	Ukuran kekuatan asosiasi berupa rasio odds terjadinya peristiwa klinis pada kelompok terpajan dibandingkan odds kelompok tidak terpajan.
Ordinal Data	Tipe data kategorikal yang memiliki urutan tingkatan atau derajat hierarki bawaan yang jelas, namun tanpa jarak matematis yang seragam (misalnya: derajat keparahan penyakit).
Precision (Presisi)	Ukuran kedekatan hasil estimasi sampel yang diperoleh secara berulang dari proses sampling. Direfleksikan oleh sempitnya rentang Confidence Interval (CI).
Prevalence (Prevalensi)	Proporsi pasien yang menderita penyakit tertentu pada titik waktu spesifik di suatu populasi tertentu.
P-value	Probabilitas memperoleh hasil uji statistik yang sama ekstremnya atau lebih ekstrem dari nilai sampel yang teramati, dengan asumsi bahwa hipotesis nol (tidak ada perbedaan/efek) benar.
QUADAS-2	Instrumen standar internasional yang divalidasi untuk menilai risiko bias dan aplikabilitas studi akurasi tes diagnostik dalam telaah kritis.
Receiver Operating Characteristic (ROC)	Kurva grafis yang memplot Sensitivitas (Sumbu Y) versus 1-Spesifisitas (Sumbu X) untuk seluruh kemungkinan nilai ambang batas (cut-off) suatu tes diagnostik.
Right-Censoring	Bentuk penyensoran survival yang paling sering terjadi di mana masa tindak lanjut pasien terhenti secara prematur sebelum kejadian klinis yang dinilai sempat terjadi.

Risk Ratio (RR)	Rasio probabilitas terjadinya hasil klinis buruk pada kelompok intervensi dibandingkan probabilitas pada kelompok kontrol dalam desain penelitian prospektif.
Sensitivity (Sensitivitas)	Kemampuan suatu uji diagnostik untuk mengidentifikasi secara akurat subjek yang benar-benar menderita penyakit (laju positif asli).
Specificity (Spesifisitas)	Kemampuan suatu uji diagnostik untuk mengidentifikasi secara akurat subjek yang benar-benar sehat/bebas penyakit (laju negatif asli).
Standard Deviation (SD)	Metrik deskriptif yang mengukur besarnya variabilitas atau penyebaran nilai biologis subjek secara individual di sekitar nilai rata-rata kelompok.
Standard Error (SE)	Metrik inferensial yang mengukur variabilitas sampling dari estimasi statistik (misalnya rata-rata) jika diambil sampel secara berulang dari populasi.
STARD 2015	Standar internasional berupa pedoman pelaporan 30 butir checklist wajib untuk meningkatkan transparansi pelaporan studi akurasi diagnostik.
Tidy Data	Format penyimpanan data terstandardisasi di mana setiap variabel menempati satu kolom tersendiri, setiap observasi (pasien) menempati satu baris, dan setiap nilai menempati satu sel.
Time-to-event Data	Tipe data gabungan yang mengukur waktu dari awal pemantauan hingga terjadinya suatu peristiwa klinis spesifik atau titik penyensoran berakhir.
Youden Index (J)	Metrik optimasi matematis (Sensitivitas + Spesifisitas - 1) yang digunakan sebagai salah satu rujukan penentuan nilai batas optimal (cut-off) suatu uji diagnostik.

PROFIL PENULIS

Dr. dr. Omat Rachmat, Sp.OT(K)Spine



Dr. dr. Omat Rachmat Hasbulloh, Sp.OT(K)Spine adalah Dekan Fakultas Kedokteran dan Ilmu Kesehatan (FKIK) Universitas Sultan Ageng Tirtayasa (Untirta). Ia merupakan Dokter Spesialis Ortopedi dan Traumatologi dengan subspesialisasi Konsultan Tulang Belakang (Spine). Sebagai praktisi klinis yang aktif menangani pasien dan membimbing mahasiswa, ia mengintegrasikan pengalaman klinis dan metodologi riset kedokteran ke dalam buku ini. Kontribusinya berfokus pada penerapan biostatistika sebagai instrumen penalaran kritis untuk menghadapi variabilitas data klinis dalam pengambilan keputusan medis yang bertanggung jawab demi keselamatan pasien. Email: omat@untirta.ac.id.

dr. Rukman Abdullah, M.Pd.Ked



dr. Rukman Abdullah, M.Pd.Ked. adalah dokter dan dosen di Departemen Pendidikan Kedokteran FKIK Untirta. Ia menyelesaikan pendidikan profesi Dokter di Universitas Jember dan melanjutkan studi pascasarjana bidang Magister Pendidikan Kedokteran (*Medical Education*) di Universitas Indonesia (UI). Dalam penyusunan buku ini, dr. Rukman menerapkan prinsip-prinsip edukasi kedokteran untuk menyajikan materi biostatistika secara runut dan terstruktur. Fokusnya adalah membantu pembaca, khususnya Peserta Program Pendidikan Dokter Spesialis (PPDS), membangun alur penalaran bertahap dalam memahami kedokteran diagnostik, regresi, konfontasi faktor pengganggu (*confounding*), hingga analisis survival. Email: rukman.abdullah@untirta.ac.id.

Bohari, S.Gz., M.Kes



Bohari, S.Gz., M.Kes. adalah akademisi dan peneliti di Jurusan Gizi FKIK Untirta. Menyelesaikan studi sarjana di bidang ilmu gizi dan menempuh pendidikan Magister Kesehatan, ia memiliki fokus perhatian pada metodologi penelitian dan pengelolaan data riset kuantitatif di bidang kesehatan. Perannya dalam buku ini adalah memastikan aspek teknis biostatistika seperti pemosisian asosiasi versus kausalitas, penentuan besar efek (*effect size*), dan interpretasi hasil uji statistik disajikan secara proporsional, akurat, serta didukung oleh literatur ilmiah yang valid dan dapat ditelusuri. Email: bohari@untirta.ac.id.

BUKU AJAR

BIOSTATISTIKA DAN KOMPUTER STATISTIKA

BAGI MAHASISWA KEDOKTERAN

PANDUAN BERPIKIR KRITIS KLINIS



SINOPSIS

Banyak dokter menganggap biostatistika sebagai kumpulan rumus dan uji yang harus dihafal. Buku ini menawarkan cara pandang berbeda: biostatistika sebagai cara berpikir untuk menghadapi variabilitas, ketidakpastian, dan bukti ilmiah dalam praktik klinis sehari-hari.

Disusun khusus bagi Peserta Program Pendidikan Dokter Spesialis, buku ini membangun penalaran secara bertahap mulai dari mengapa biostatistika penting, merumuskan pertanyaan klinis menjadi pertanyaan statistik, hingga memahami data, membandingkan kelompok, dan membedakan asosiasi dari kausalitas.

Pembahasan berlanjut ke topik yang lebih aplikatif: regresi dan confounding, kedokteran diagnostik, analisis survival dan prognosis, hingga cara membaca besar efek dan kepentingan klinis suatu hasil penelitian secara proporsional.

Dengan bahasa yang komunikatif dan berpijak pada sumber ilmiah yang dapat ditelusuri, buku ini mengajak pembaca untuk tidak sekadar menghitung, melainkan bernalar agar mampu menilai bukti secara kritis dan menerapkannya secara bertanggung jawab demi kepentingan pasien.

- **Berpikir Kritis**
- **Berbasis Bukti**
- **Relevan dengan Praktik Klinis**
- **Bahasa yang Komunikatif**

Pallantikang, Kabupaten Maros
Website:penerbit.salnesia.id/index.php/press

ISBN 978-634-05-5952-1 (PDF)
9 786340 559521