ULTIMATE BIOSTATS GUIDE FOR USMLE

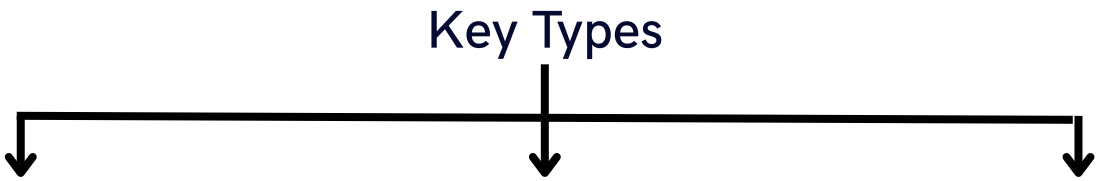
Build a Powerful Foundation in Just 1 Day.

Welcome to the Ultimate Biostats Guide for USMLE— your high-yield, 1-day crash course to unlocking biostatistics for Step 1, Step 2 CK and Step 3.

Contents

COMPREHENSIVE BIOSTATISTICS STUDY GUIDE

1. Observational Studies



CASE SERIES

Collection of individual cases with similar characteristics (no control group)

CROSS-SECTIONAL

Measures exposure and outcome at the same time (snapshot)

ECOLOGICAL

Analyzes data at the population level (not individual)

Read FA for the above 3

Case-Control vs. Cohort
USMLE Smash Edition

Case-Control = "Rare Disease Detective"

- Disease is ALREADY KNOWN (cases vs. controls) → You start with sick people and ask: "What caused this?"
- Perfect for rare diseases (e.g., mesothelioma, HIV, childhood cancers) because you don't need to wait for cases to appear—you already have them!
- USMLE Trick: If the question mentions a rare disease (or gives you people with vs. without disease), think CASE-CONTROL!

Example Q:
"*Researchers compare 100 pancreatic cancer patients (cases) to 100 healthy controls. They find that 80% of cases smoked, vs. 20% of controls. What study is this?*"
→ Answer: Case-control (started with disease, looked back for risk factors).

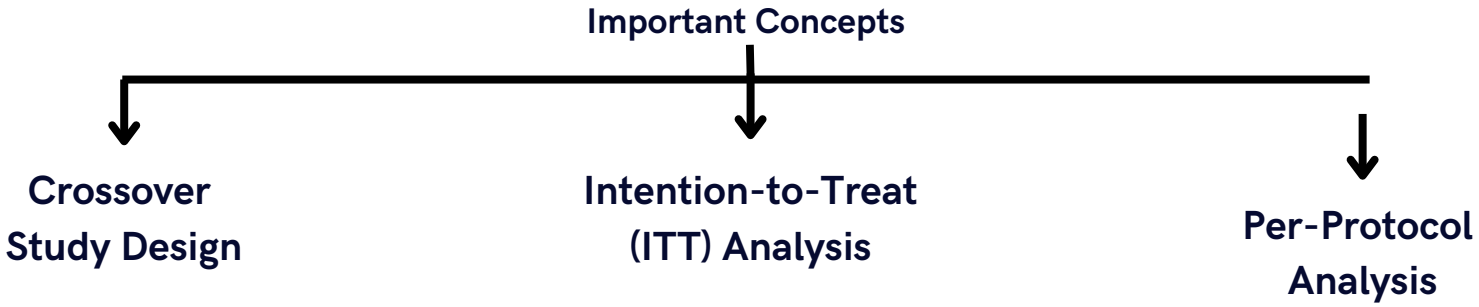
Cohort = "Exposure → Future Outcome"

- Exposure is ALREADY KNOWN (smokers vs. non-smokers) → You track them forward to see "Who gets sick?"
- Best for common diseases (e.g., heart disease, diabetes) because you need enough people to develop the outcome.
- USMLE Trick: If the question gives you exposed vs. unexposed groups and asks "What happens later?", think COHORT!

Example Q:
"Researchers follow 1,000 smokers and 1,000 non-smokers for 20 years. Smokers develop lung cancer at 10x the rate of non-smokers. What study is this?
Answer: Cohort (started with exposure, followed for disease).

USMLE BIOSTAT <1minute: Click Here

2. Clinical Trials



a). Crossover Study Design

What? Patients switch treatment groups (e.g., Drug A → Drug B).

USMLE Keywords:

- "Each participant receives both treatments"
- "Washout period" (to avoid carryover effects)
- "Same group compared to itself"

MCQ Clues:

- If question mentions "patients switching treatments" → Crossover!

2. Intention-to-Treat (ITT) Analysis

What? Analyze ALL randomized patients (even if they dropped out or didn't follow the protocol).

USMLE Keywords:

- "Analyzed by original assigned groups"
- "Includes non-compliant patients"
- "Preserves randomization"

MCQ Clues:

- If question says "all participants analyzed regardless of adherence" → ITT!
- More conservative (real-world bias included).
- Reduces Type I error (false positives).

3. Per-Protocol Analysis

What? Only analyze patients who completed the study as planned.

USMLE Keywords:

- "Only compliant participants included"
- "Excludes dropouts/crossovers"
- "Ideal conditions"

MCQ Clues:

- If question says "only those who finished treatment analyzed" → Per-Protocol!
- Overestimates treatment effect (removes non-responders).
- Higher risk of bias (less real-world).

USMLE MCQ Decision Tree

1. "Patients switched treatments?" → Crossover.
2. "Analyzed as originally assigned (even dropouts)?" → ITT.
3. "Only perfect completers analyzed?" → Per-Protocol.

💡 **HY Mnemonic:**
"ITT = Include The Total" (everyone stays, warts and all).
"Per-Protocol = Perfect Patients" (only the obedient ones count).

- ITT includes all randomized participants.
- Per-Protocol includes only those who followed the study protocol.

Phases of Clinical Trials (FA + Neil Randy Shorts)
Link to neil randy clinical trial videos:
- Phase I: Safety, small group.
- Phase II: Efficacy, side effects.
- Phase III: Large-scale comparison to standard treatment.
- Phase IV: Post-marketing surveillance.

🔗 First Aid Read Properly
And NEIL RANDY BIOSTATS SHORTS Q 36-40

Your NBME-style MCQs and explanations are strong—clear, accurate, and high-yield. Just a few small refinements for precision, formatting, and flow

---

NBME-Style MCQ 1: Clinical Trial Phase Identification

Question:
A new drug for metastatic melanoma is tested in a small group of 30 patients to determine the maximum tolerated dose (MTD) and evaluate dose-limiting toxicities. No control group is used, and the primary endpoint is safety. Which phase of clinical trial is this?

A) Phase I
B) Phase II
C) Phase III
D) Phase IV

NBME-Style MCQ 2: Clinical Trial Phase Identification

Question:
Researchers are conducting a randomized, double-blind study comparing a novel antihypertensive drug to the current standard therapy in 2,000 patients across multiple centers. The primary outcome is the reduction in systolic blood pressure at 6 months. Which phase of clinical trial is this?

A) Phase I
B) Phase II
C) Phase III
D) Phase IV

Answers & Explanations:

Answer to MCQ 1:

A) Phase I

Why?

Involves a small sample (20–80 participants).

Primary goal is safety, pharmacokinetics, and dose-limiting toxicity.

Determines maximum tolerated dose (MTD).

Typically non-randomized and uncontrolled.

Answer to MCQ 2:

C) Phase III

Why?

Involves large patient population (several hundred to thousands).

Randomized, double-blind, multicenter—hallmarks of Phase III.

Compares new treatment vs. standard of care to assess efficacy and safety before approval.

🔥 **USMLE High-Yield Takeaways:**

Phase	Key Features
Phase I	Small group (20–80), safety, dosage, pharmacokinetics. No control group.
Phase II	Larger group (100–300), efficacy & side effects. May include placebo.
Phase III	Large (1,000+), randomized & blinded, compares to standard of care.
Phase IV	Post-marketing surveillance. Long-term effects, rare adverse events.

💡 **Remember: Phase I = Is it safe? → Phase II = Does it work? → Phase III = Is it better? → Phase IV = What happens long-term?**

HY USMLE Q#1200- Biostat CLICK HERE

---

3. Bradford Hill Criteria & Terminology

- Just read once (for understanding -that is it!)

4. Measures of Association

a). Relative Risk (RR)

Definition: Tells you how much more (or less) likely an outcome is in the **exposed** group compared to the **unexposed** group.
Used in: Cohort studies (prospective or retrospective).

Formula:

RR=Risk in ExposedRisk in Unexposed

- **a / (a + b) =** risk of disease in exposed group
- **c / (c + d) =** risk of disease in unexposed group

Interpretation:

- **RR = 1 →** No association
- **RR > 1 →** Exposure increases risk
- **RR < 1 →** Exposure is protective

🧠 **Mnemonic: "RR = Risk Ratio" →** used when you're comparing **actual risk** (cohort).

b). Odds Ratio (OR)

Definition: Compares the odds of exposure among **cases vs controls.**
Used in: **Case-control studies** (because you already know disease outcome).

Formula: OR=ad/bc

- **a/c =** odds of exposure in diseased
- **b/d =** odds of exposure in non-diseased

Interpretation:

- **OR = 1 →** No association
- **OR > 1 →** Exposure associated with disease
- **OR < 1 →** Exposure protective

🧠 **Mnemonic: "OR = Odds Ratio" →** used when you're comparing odds (case-control).

🪄 **HY Tip: When disease is rare, OR ≈ RR — USMLE loves this.**

c). Absolute Risk Reduction (ARR)

Definition: The actual decrease in risk from a treatment or intervention.
Used in: Clinical trials.

Formula:
ARR=Risk in Control−Risk in Treatment

Interpretation:
Tells you the real-world benefit of a treatment.

🧠 **Mnemonic:**
ARR = Actual Risk Removed by the treatment.

Questions:
A study investigates a new anticoagulant in preventing DVT after orthopedic surgery.
In the treatment group (n = 500), 15 patients developed DVT.
In the placebo group (n = 500), 40 patients developed DVT.

1) - What is the Absolute Risk Reduction (ARR)?
A. 0.025
B. 0.05
C. 0.08
D. 0.13
E. 0.20

2). – Relative Risk Reduction (RRR)
Using the data above, what is the Relative Risk Reduction (RRR)?
A. 0.25
B. 0.33
C. 0.50
D. 0.60
E. 0.75

3) – Number Needed to Treat (NNT)
Based on the ARR from the same study, what is the Number Needed to Treat (NNT) to prevent 1 DVT?
A. 8
B. 13
C. 20
D. 25
E. 40

Answers and Explanations

Question 1 – ARR
Risk in treatment = 15/500 = 0.03
Risk in placebo = 40/500 = 0.08
ARR = 0.08 – 0.03 = 0.05
Answer: B. 0.05

Question 2 – RRR
RRR = ARR / Risk in placebo = 0.05 / 0.08 = 0.625
Closest option:
Answer: D. 0.60

Question 3 – NNT
NNT = 1 / ARR = 1 / 0.05 = 20
Answer: C. 20

First Aid recommended:

USMLE STEP 1: BIOSTAT QUESTIONS RR CLICK HERE

USMLE STEP 1: BIOSTAT QUESTIONS CLICK HERE (Thank You DEVIKA)

---

5. Rates & Definitions

a) Case Fatality Ratio (CFR)

What? % of people with a disease who die from it.

Formula:
CFR=Deaths from diseaseTotal cases of disease×100
CFR=
Total cases of disease
Deaths from disease
×100

Example:

- 200 COVID-19 cases → 10 deaths → CFR = 5%.

USMLE Clue:

- "Of those infected, how many die?"
- Used for **acute diseases** (e.g., Ebola, COVID-19).

b) Mortality Rate (Death Rate)

What? % of the TOTAL population that dies from a disease.

Formula:
Mortality Rate=Deaths from diseaseTotal population×100,000
Mortality Rate=
Total population
Deaths from disease
×100,000

Example:

- 10 heart attack deaths in a city of 100,000 → Mortality rate = 10/100,000.

USMLE Clue:

- "How common is death from this disease in the whole population?"
- Includes both sick and healthy people.

c) Attack Rate (Risk)

What? % of exposed people who get the disease.

Formula:
Attack Rate=New cases during outbreakPopulation at risk×100
Attack Rate=
Population at risk
New cases during outbreak
×100

Example:

- 50 people eat contaminated food → 20 get food poisoning → Attack rate = 40%.

USMLE Clue:

- "How contagious/harmful is the exposure?"
- Used for outbreaks (e.g., food poisoning, measles).

- **CFR:** "Case = Sick, Fatality = Death" → **How many sick people die?**
- **Mortality Rate:** "Total population deaths" → **Heart disease in a city.**
- **Attack Rate:** "Attack = Outbreak" → **Food poisoning at a party.**

---

6. Demographic Transition

- Read from First Aid(stages of population changes).

---

7. Likelihood Ratio

- Positive LR = Sensitivity / (1 - Specificity).
- Negative LR = (1 - Sensitivity) / Specificity.

CLICK HERE for HY PRECISE Explanation

8. Survival Analysis (Kaplan-Meier Curve)

- Estimates survival probability over time.
- Used in cohort studies and clinical trials.

USMLE STEP 1 KAPLAN MEIER CURVE w/ Questions
Click Here

9. Diagnostic Tests

- Sensitivity, Specificity, PPV, NPV

CLICK HERE

USMLE Step 1 Biostatistics

USMLE Step 1 Questions Biostats

ROC Curve (First Aid is sufficient)

10. Precision vs. Accuracy

- Precision= Consistency (low variability).
- Accuracy = Closeness to true value.

Click Here for the HY note

---

11. Incidence vs. Prevalence

- Incidence = New cases over time.
- Prevalence = Total cases at a given time.
- First Aid has an excellent explanation.

---

12. Lead Time vs. Length Time Bias

- Lead Time Bias: Early detection falsely increases survival time.
- Length Time Bias: Slow-growing diseases are detected more often in screening

CLICK HERE

---

13. Types of Bias

- Availability Bias: Overestimating based on recent examples.
- Premature Closure: Diagnosing without full evaluation.
- Confirmation Bias: Favoring information that supports pre-existing beliefs.
- First Aid covers most biases; use ChatGPT for tricky distinctions.

HY DIVINE VIDEO FOR BIASES

NBME-Style MCQ

Question 1:
A 30-year-old man presents with fever, headache, and neck stiffness. The physician immediately diagnoses bacterial meningitis because they recently managed a similar case last week. Workup later reveals viral meningitis.
Which cognitive bias is most likely at play?

A) Availability bias
B) Anchoring bias
C) Premature closure
D) Confirmation bias

Question 2:
A 50-year-old woman presents with chest pain. Because she has a history of GERD, the physician diagnoses acid reflux and does not reconsider the diagnosis even after an ECG shows ST elevations.
Which cognitive bias does this reflect?

A) Confirmation bias
B) Availability bias
C) Anchoring bias
D) Premature closure

Question 3:
A patient presents with fatigue and cold intolerance. The physician suspects hypothyroidism and continues ordering thyroid function tests despite normal results, ignoring signs of anemia and depression.
Which cognitive bias is this?

A) Confirmation bias
B) Availability bias
C) Anchoring bias
D) Premature closure

Question 4:

A 24-year-old woman presents with abdominal pain. The clinician diagnoses irritable bowel syndrome based on a quick history and exam. Despite red flags like weight loss and anemia, no further evaluation is done.

What type of cognitive bias best explains this?

A) Availability bias
B) Confirmation bias
C) Anchoring bias
D) Premature closure

Answers & Explanations

MCQ 1: A) Availability bias

The physician's judgment is overly influenced by a recent case of bacterial meningitis.
Availability bias = relying on memorable or recent examples, even if statistically less likely.

MCQ 2: C) Anchoring bias

The doctor fixated early on GERD and ignored new info (ECG changes).
Anchoring = locking onto an initial diagnosis and failing to adjust.

MCQ 3: A) Confirmation bias

Physician seeks evidence to support their idea (hypothyroidism) and ignores contradictory data.
Confirmation bias = selectively interpreting findings to match your hunch.

MCQ 4: D) Premature closure

Physician stopped thinking too early once IBS was assumed—despite red flags.
Premature closure = ending diagnostic process before it's complete.

14. Types of Bias

- **Confounding**: Distorts association between exposure and outcome.
- **Effect Modification**: When the effect of exposure differs across subgroups.

"Effect modifier changes HOW exposure affects outcome (only impacts outcome). Confounder affects BOTH exposure AND outcome."

Effect Modifier (USMLE HY):
✅ ONLY modifies outcome's response to exposure
✅ Example:
- Drug works better in men (sex modifies drug→outcome effect)
- Not linked to exposure itself (sex doesn't cause you to take the drug)

Confounder (Different!):
❌ Linked to BOTH exposure AND outcome
❌ Example:
- Smoking causes coffee drinking (exposure) AND lung cancer (outcome)

How USMLE Tests This:

MCQ Example 1:
"A study finds aspirin prevents heart attacks in diabetics but not non-diabetics. Diabetes in this context is a:"
Answer: Effect modifier (changes aspirin→outcome effect)

MCQ Example 2:
"Smokers are more likely to drink coffee and develop lung cancer. In studying coffee→cancer, smoking is a:"
Answer: Confounder (affects both coffee drinking AND cancer)

🔥 3-Second USMLE Hack:

1. "Does it change HOW exposure affects outcome?" → Effect modifier
2. "Does it affect BOTH exposure AND outcome?" → Confounder

b). Randomization (Best for RCTs!)

What? **Randomly assigning patients to groups to eliminate confounding.**

Example:
In a drug trial, random assignment ensures smokers are equally distributed in treatment & placebo groups.

USMLE Clues:
"Randomized controlled trial (RCT)" → gold standard!
If question says "randomly assigned" → confounding is minimized!

MCQ Example:
"Researchers randomly assign 500 patients to Drug A or placebo to test efficacy. What does randomization primarily prevent?"
Answer: Confounding (balances known/unknown confounders).

c). Stratification (Splitting Data to Control Confounding/Effect Mod)

What? **Dividing data into subgroups (strata) to adjust for variables.**

Example:
Study smoking & lung cancer → stratify by age to remove age's effect.

USMLE Clues:
If question says "analyzed separately by gender/age" → stratification!
Used to:

- Control confounding (e.g., adjust for smoking).
- Check for effect modification (e.g., does drug work differently in men vs. women?).

MCQ Example:
"A study on obesity and heart disease analyzes results separately for smokers and non-smokers. What is this called?"
Answer: Stratification (by smoking status).

15. Descriptive Statistics (Mean, Median, Mode, Skewness)

- Left Skew: Mean < Median < Mode.
- Right Skew: Mode < Median < Mean.
- First Aid diagrams are very helpful.

---

16. P-Value & Hypothesis Testing

- **P-value**: Probability of observing results if null is true.
- **Type I Error (α)**: False positive (rejecting true null).
- **Type II Error (β)**: False negative (failing to reject false null).
- **Power (1-β):** Probability of correctly rejecting null.

a). What is a P-value?

Think of it like a **"luck meter"**:

P < 0.05 → "Whoa, only a <5% chance this happened by dumb luck! Probably real!"
P ≥ 0.05 → "Meh, could've just been luck."

Example:

You test a **magic pill for headaches**:
P = 0.03 → Only 3% chance the results are fake. "Statistically significant!"
P = 0.50 → 50% chance it's fake. "Probably garbage."

b). Statistical vs. Clinical Significance

Real-life example:

Study finds: New drug lowers blood pressure by 0.5 mmHg with P = 0.04.
Statistically significant? (P < 0.05)
Clinically important? (0.5 mmHg won't save lives!)

USMLE HY Point:
P-value = Math truth.
Clinical importance = Real-world impact.

USMLE MCQ Attack Plan:
"P < 0.05?" → "Statistically significant!"
"Tiny effect size?" → "Not clinically significant!"

Example Q:
"A weight-loss drug shows P=0.01 but only reduces weight by 0.1kg. What's true?"
A) Statistically significant, but may lack clinical importance.

Mnemonic:
"P < 0.05 = Not fake!"
"Big P = Probably crap!"

---

17. Statistical vs. Clinical Significance

- Statistical: P < 0.05 (mathematically significant).
- Clinical: Real-world importance of the finding.

---

18. Confidence Intervals & Meta-Analysis

- CI: Range where true effect likely lies.
- Meta-Analysis: Combines multiple studies for stronger evidence.
- Meta Analysis increases the Power of the study and external Validity
- Internal Validity: Study design accuracy.
- External Validity: Generalizability to real-world settings.

CLICK HERE for Further HY points

19. Statistical Tests

Test	Use Case	Data Type
T-test	Compare 2 means	Quantitative
ANOVA	Compare >2 means	Quantitative
Chi-Square Test	independence/categorical	**Qualitative**

a) Chi-Square Test

When? Comparing categorical data (yes/no, dead/alive, smoker/non-smoker).

Example:
Does smoking (yes/no) link to lung cancer (yes/no)?

USMLE Clue Words:

- "Proportions," "percentages," "categories"
- "Counts of people/things in groups"

MCQ Example:
"A study compares vaccine effectiveness (got flu: yes/no) between two groups. Which test?"
➔ **Answer**: Chi-square (categorical outcome!

b) T-Test

When? Comparing 2 groups' averages (blood pressure, lab values).

Example:
Does Drug A lower BP more than Placebo?

USMLE Clue Words:

- "Mean difference," "two groups," "continuous data"
- "Before vs. after" (if paired, use paired t-test)

MCQ Example:
"Researchers compare mean cholesterol levels in statin vs. placebo groups. Which test?"
➔ **Answer**: T-test (comparing 2 means!).

c) ANOVA

When? Comparing ≥3 groups' averages (e.g., Drug A vs. B vs. Placebo).

Example:
Which of 3 diets works best for weight loss?

USMLE Clue Words:

- "Multiple groups," "more than two means"
- "One outcome, multiple treatments"

MCQ Example:
"A trial tests blood glucose levels in patients on Diet X, Y, or Z. Which test?"
➔ **Answer**: ANOVA (3+ groups!).

20. Correlation Coefficient (r)

- Measures strength & direction of linear relationship (-1 to +1).
- First Aid plus these 2 videos Below after reading FA

Video 1

Video 2

Drug Ads Superb Advice & Practice Questions

DIVINE PODCAST INTERVENTION

WATCH THIS PLAYLIST AFTER COMPLETING THIS DOCUMENT

CLICK HERE

Great Way to quickly Revise Biostats- Do 1 daily

Biostat Summary Step 1 Extra Stuff

Biostat Summary Step 1 The Basics

Biostat Summary Step 1 Exam Review

Biostat Summary Step 1 Exam Review