Teach the Evidence Chain

Each lesson moves from analogy to biology to a bounded clinical decision.

Standard 45-minute lesson flow

Minutes	Teacher and student move
0-5	Carry forward the prior take-home and state today's question.
5-12	Show only the analogy panel. Ask the three observation questions.
12-18	Students write the three rules. Name where the analogy stops.
18-27	Reveal the biomedical panel. Map each rule onto the case.
27-38	Students read labeled evidence and make the concrete decision.
38-45	Reveal the strongest answer, write CER, and state the next unlock.

TEACHER PLAN 1 OF 20 | PRINCIPAL INVESTIGATOR (PI)

From an Observation to a Researchable Question

Students will convert an observation about Mateo into a focused, answerable research question using the PICO framework, and separate researchable from non-researchable questions.

CARRY FORWARD

A complete anatomical story can define structures, functions, repairs, and uncertainties while leaving treatment comparisons open to experiment.

10-YEAR TAKEAWAY

A researchable question names a population, a factor or intervention, a comparison, and a measurable outcome.

Socratic launch

Show the analogy panel only: A camera turns a wide scene into one answerable frame. Do not name the biology yet.

- What is outside the frame?
- Which action and comparison remain inside it?
- How will the camera know when the moment has happened?

Rules students should earn

- Name who is being studied.
- Name the factor or intervention and comparison.
- Name an outcome that can be measured at a set time.

Biology reveal and evidence

ID	Finding supplied in the case
EXP01-E1	The open anatomical question is whether repair timing changes later speech function.
EXP01-E2	A supplied candidate population is medically fit infants with isolated cleft palate.
EXP01-E3	VPI can be scored at age 5 by trained assessors.

Decision and answer guidance

Situation: The team must choose one question before enrolling any participants.

Strongest option: Among eligible infants, does repair at 6 rather than 12 months reduce VPI at age 5?

Evidence required: EXP01-E1 + EXP01-E2

Claim ceiling: You may define the study question. You may not predict which timing is better before data are collected.

Next handoff: We turned an observation into a sharp PICO question. But a question is still not a prediction. How do we turn a PICO question…

TEACHER PLAN 2 OF 20 | PRINCIPAL INVESTIGATOR (PI)

From a Question to a Testable Hypothesis

Students will write a testable hypothesis and its null hypothesis, and identify the independent variable, dependent variable, controlled variables, and control group for a study of Mateo's question.

CARRY FORWARD

A researchable question names a population, a factor or intervention, a comparison, and a measurable outcome.

10-YEAR TAKEAWAY

A falsifiable hypothesis links a defined change to a defined outcome and states what would count against it.

Socratic launch

Show the analogy panel only: A crash-test prediction names the change, the sensor, and the failure result. Do not name the biology yet.

- What is deliberately changed?
- Which sensor records the result?
- Which result would make the prediction fail?

Rules students should earn

- Change one defined independent variable.
- Measure one defined dependent outcome.
- Write a result that could count against the prediction.

Biology reveal and evidence

ID	Finding supplied in the case
EXP02-E1	The proposed study compares repair at 6 months with repair at 12 months.
EXP02-E2	The primary outcome is VPI status at age 5.
EXP02-E3	A valid hypothesis can be challenged by equal or higher VPI in the 6-month group.

Decision and answer guidance

Situation: A draft says, 'Earlier surgery will help somehow,' and the team must decide whether it can guide a study.

Strongest option: Rewrite it with timing, VPI at age 5, and a result that would oppose the prediction.

Evidence required: EXP02-E1 + EXP02-E2

Claim ceiling: You may state a testable prediction. You may not claim the hypothesis is true before the comparison.

Next handoff: We can design a single study and predict what would prove us wrong. But two honest studies on the same question can reach opposite…

TEACHER PLAN 3 OF 20 | PRINCIPAL INVESTIGATOR (PI)

Why We Trust Some Studies More Than Others

Students will rank common study designs in the evidence hierarchy and justify the ranking by how much bias each design controls, for both treatment questions and cause questions.

CARRY FORWARD

A falsifiable hypothesis links a defined change to a defined outcome and states what would count against it.

10-YEAR TAKEAWAY

The strongest design is the one that fits the question and controls its main threats.

Socratic launch

Show the analogy panel only: A tool wall has no single best tool for every job. Do not name the biology yet.

- Which tool measures length?
- Which tool reveals a hidden structure?
- Why would ranking every tool from weak to strong be misleading?

Rules students should earn

- Name the question type first.
- Choose the design that can answer it ethically.
- Identify the main bias or alternative explanation the design must control.

Biology reveal and evidence

ID	Finding supplied in the case
EXP03-E1	Repair timing is an assignable clinical choice when both options are acceptable.
EXP03-E2	A suspected harmful pregnancy exposure cannot be assigned.
EXP03-E3	A gene mechanism requires perturbation and molecular or tissue outcomes.

Decision and answer guidance

Situation: A teammate declares that an RCT is always the best design for every cleft question.

Strongest option: Match each design to the treatment, harm, prognosis, or mechanism question.

Evidence required: EXP03-E1 + EXP03-E2

Claim ceiling: You may select a design family. You may not call any design free of bias or automatically decisive.

Next handoff: A randomized trial sits near the top because we can assign who gets the treatment. But we could never assign a baby to a…

Studying a Cause You Cannot Assign

Students will explain how a case-control study works backward from outcome to exposure, compute and interpret a simple odds ratio with its 95 percent confidence interval, and name…

CARRY FORWARD

The strongest design is the one that fits the question and controls its main threats.

10-YEAR TAKEAWAY

Case-control studies efficiently study rare outcomes, but their odds ratios remain vulnerable to selection, recall, and confounding.

Socratic launch

Show the analogy panel only: A detective starts with a rare event and looks backward for clues. Do not name the biology yet.

- Which group is selected because the event happened?
- Where does the search move in time?
- Which clue could be remembered differently by the two groups?

Rules students should earn

- Select cases by outcome and comparable controls without it.
- Measure prior exposure the same way.
- Read an odds ratio with its interval and bias risks.

Biology reveal and evidence

ID	Finding supplied in the case
EXP04-E1	The table has 30 exposed and 70 unexposed cases, plus 15 exposed and 85 unexposed controls.
EXP04-E2	The reported 95 percent confidence interval is 0.9 to 3.4.
EXP04-E3	Parents of affected infants may recall an exposure differently while searching for an explanation.

Decision and answer guidance

Situation: The team must decide whether the supplied result proves that the exposure caused clefts.

Strongest option: Report an uncertain association and plan better exposure measurement plus confounder control.

Evidence required: EXP04-E1 + EXP04-E2

Claim ceiling: You may report the observed association and uncertainty. You may not claim individual blame or causation.

Next handoff: The case-control design lets us chase causes from the outside world without assigning harm. But Mateo's sparse family history keeps nagging us. How would…

TEACHER PLAN 5 OF 20 | EPIDEMIOLOGIST ON THE CLEFT RESEARCH TEAM

Measuring Inheritance Over Time and in Twins

Students will explain how a cohort study follows exposure forward to outcome, how a twin study compares identical and fraternal concordance to estimate heritability, and what high heritability…

CARRY FORWARD

Case-control studies efficiently study rare outcomes, but their odds ratios remain vulnerable to selection, recall, and confounding.

10-YEAR TAKEAWAY

Cohorts establish time order, while twin comparisons estimate population patterns under assumptions rather than one person's cause.

Socratic launch

Show the analogy panel only: A movie follows events forward while a matched split screen compares two paths. Do not name the biology yet.

- Which panel shows the exposure before the outcome?
- What is more similar within one pair?
- Which shared condition could make a pair look genetically alike?

Rules students should earn

- Define the starting population before outcomes occur.
- Measure exposure and follow-up consistently.
- State the assumptions behind twin comparisons.

Biology reveal and evidence

ID	Finding supplied in the case
EXP05-E1	A cohort records an exposure before birth outcomes are known.
EXP05-E2	Monozygotic twins share more DNA than dizygotic twins on average.
EXP05-E3	Twin estimates rely on assumptions, and heritability changes across populations and settings.

Decision and answer guidance

Situation: The proposal claims a twin study can cleanly separate genes from environment for one infant.

Strongest option: Use cohort and twin evidence with explicit assumptions and population-level wording.

Evidence required: EXP05-E1 + EXP05-E2

Claim ceiling: You may discuss time order and population variance. You may not assign Mateo's cleft a heritability percentage or single cause.

Next handoff: The twin study tells us genes carry most of the risk, but heritability is just a number; it does not name a single gene.…

Finding a Risk Gene Among Millions of Bases

Students will explain how GWAS scans the whole genome and how the transmission disequilibrium test (TDT) uses parents as built-in controls to flag a real risk allele.

CARRY FORWARD

Cohorts establish time order, while twin comparisons estimate population patterns under assumptions rather than one person's cause.

10-YEAR TAKEAWAY

Gene discovery combines broad scans, family structure, quality control, and independent replication.

Socratic launch

Show the analogy panel only: A citywide signal scan needs family checkpoints and a second city. Do not name the biology yet.

- Where does the broad scan find candidate signals?
- How does the family gate check transmission?
- Why must another city repeat the result?

Rules students should earn

- Scan broadly without choosing only favorite genes.
- Use family or population structure to test the signal.
- Require quality control and independent replication.

Biology reveal and evidence

ID	Finding supplied in the case
EXP06-E1	A GWAS compares millions of common variants across many people.
EXP06-E2	A case-parent trio can test whether one allele is transmitted to affected children more often than expected.
EXP06-E3	Genotyping quality, ancestry structure, sample size, and independent replication affect credibility.

Decision and answer guidance

Situation: One sample shows a strong signal near a developmental gene, but no second cohort has tested it.

Strongest option: Call it a candidate locus and require quality checks plus independent replication.

Evidence required: EXP06-E1 + EXP06-E2

Claim ceiling: You may nominate a replicated risk locus. You may not name a causal gene or patient diagnosis from proximity alone.

Next handoff: A scan and a trio test both flagged rs642961 near IRF6. But a flag is not proof. With millions of SNPs and hundreds of…

TEACHER PLAN 7 OF 20 | BIOSTATISTICIAN ON THE CLEFT RESEARCH TEAM

Is the Association Real, or Just Chance?

Students will interpret an odds ratio, a p-value, and a 95 percent confidence interval, and decide whether an association is statistically significant.

CARRY FORWARD

Gene discovery combines broad scans, family structure, quality control, and independent replication.

10-YEAR TAKEAWAY

Effect size and confidence interval show magnitude and precision; a p-value measures model compatibility, not truth.

Socratic launch

Show the analogy panel only: A forecast needs an estimate and range, not a yes-or-no stamp. Do not name the biology yet.

- Which mark gives the best estimate?
- What does the width of the band show?
- Why can a tiny gauge reading not prove the forecast is true?

Rules students should earn

- Read the effect size first.
- Use the interval to judge precision and compatible values.
- Interpret the p-value with design, bias, and prior evidence.

Biology reveal and evidence

ID	Finding supplied in the case
EXP07-E1	Study A reports a risk ratio of 0.60 with a 95 percent interval from 0.37 to 0.98.
EXP07-E2	Study B reports a risk ratio of 0.55 with an interval from 0.18 to 1.67.
EXP07-E3	A p-value is calculated under a statistical model and null hypothesis.

Decision and answer guidance

Situation: A teammate wants to label Study A true and Study B false based only on whether p is below 0.05.

Strongest option: Compare effect sizes, intervals, designs, and bias before stating the claim.

Evidence required: EXP07-E1 + EXP07-E2

Claim ceiling: You may describe magnitude, precision, and statistical compatibility. You may not convert a threshold into scientific truth.

Next handoff: One p-value below 0.05 feels convincing, but a GWAS runs about a million tests at once. If 0.05 lets through a 1-in-20 fluke, how…

TEACHER PLAN 8 OF 20 | BIOSTATISTICIAN ON THE CLEFT RESEARCH TEAM

Why Gene Studies Use Such Tiny P-Values

Students will explain why testing many hypotheses at once inflates false positives, and apply a Bonferroni-style correction to reach the genome-wide significance threshold p < 5 x 10^-8.

CARRY FORWARD

Effect size and confidence interval show magnitude and precision; a p-value measures model compatibility, not truth.

10-YEAR TAKEAWAY

Many tests inflate false positives, so thresholds and independent replication must be planned before results are seen.

Socratic launch

Show the analogy panel only: Thousands of airport alarms require a stricter screen and a second check. Do not name the biology yet.

- What happens when thousands of sensors each can false-alarm?
- Why set the alert rule before scanning?
- What does the second scanner add?

Rules students should earn

- Count how many tests are being run.
- Predefine a correction or threshold suited to the analysis.
- Replicate the signal in independent data.

Biology reveal and evidence

ID	Finding supplied in the case
EXP08-E1	Testing one million variants at p below 0.05 would produce many chance hits under the null.
EXP08-E2	For many common-variant GWAS analyses, p below 5 x 10^-8 is a conventional threshold.
EXP08-E3	The candidate peak repeats in an independent cohort with the same direction of effect.

Decision and answer guidance

Situation: A peak has p equal to 2 x 10^-6 in the discovery sample and no replication result.

Strongest option: Label it suggestive and require the planned threshold, quality checks, and replication.

Evidence required: EXP08-E1 + EXP08-E2

Claim ceiling: You may apply the supplied common-variant GWAS convention. You may not treat that number as universal or proof of mechanism.

Next handoff: Gene studies use a tiny p-value because a million tests breed a million chances to be fooled, and replication seals the deal. But even…

Taking a Gene to the Bench

Students will describe how a knockout, an in situ hybridization, and an immunohistochemistry experiment each test a different part of a gene's job, and match each method to…

CARRY FORWARD

Many tests inflate false positives, so thresholds and independent replication must be planned before results are seen.

10-YEAR TAKEAWAY

Orthogonal assays answer different mechanism questions: where RNA or protein is and what changes after perturbation.

Socratic launch

Show the analogy panel only: Three flashlights reveal message, material, and system failure. Do not name the biology yet.

- Which light shows where the message is?
- Which shows where the working part is?
- Which reveals what changes when the part is missing?

Rules students should earn

- Match each assay to one question.
- Include negative and positive controls.
- Combine location evidence with perturbation evidence before claiming function.

Biology reveal and evidence

ID	Finding supplied in the case
EXP09-E1	RNA signal appears in developing palatal shelves at the relevant stage.
EXP09-E2	Protein staining appears in palatal epithelium and is absent in the no-primary-antibody control.
EXP09-E3	Perturbed embryos show altered shelf elevation more often than matched controls.

Decision and answer guidance

Situation: The team has funding for three assays and must decide whether one expression image alone proves function.

Strongest option: Combine RNA location, protein location, perturbation, and controls.

Evidence required: EXP09-E1 + EXP09-E2

Claim ceiling: You may connect expression and perturbation evidence. You may not treat location alone as causal proof.

Next handoff: A knockout removes a gene and a stain shows where it sits, but both are slow and blunt; building a conditional-knockout mouse takes many…

CRISPR as an Experimental Tool

Students will explain how CRISPR-Cas9 edits a chosen DNA site using a guide RNA, and identify the two checks (sequence-verify the edit, search for off-target cuts) that make…

CARRY FORWARD

Orthogonal assays answer different mechanism questions: where RNA or protein is and what changes after perturbation.

10-YEAR TAKEAWAY

CRISPR is a controlled perturbation only when the edit, mosaicism, phenotype, and off-target controls are verified.

Socratic launch

Show the analogy panel only: A precision editor needs verification stations after every change. Do not name the biology yet.

- How is the intended change confirmed?
- Where could an unintended edit hide?
- What if only some copies were changed?

Rules students should earn

- Sequence the intended target after editing.
- Assess mosaicism and plausible off-target sites.
- Compare multiple guides, controls, and repeated phenotypes.

Biology reveal and evidence

ID	Finding supplied in the case
EXP10-E1	Sequencing confirms a frameshift in 78 percent of reads from the target tissue.
EXP10-E2	A second guide produces a similar phenotype while a non-targeting guide does not.
EXP10-E3	The top predicted off-target sites show no detectable edits in the supplied assay.

Decision and answer guidance

Situation: One embryo has a palate defect after injection, but the target was not sequenced and no control guide was used.

Strongest option: Withhold the gene-function claim and require edit plus specificity checks.

Evidence required: EXP10-E1 + EXP10-E2

Claim ceiling: You may interpret the supplied edit and control checks. You may not claim all off-target effects are absent or a human treatment is ready.

Next handoff: We can now edit a gene precisely in a mouse, but Mateo is not a mouse. When is a mouse actually a good stand-in…

TEACHER PLAN 11 OF 20 | PRINCIPAL INVESTIGATOR (WITH THE LAB'S ANIMAL-WELFARE LIAISON)

When Is a Mouse a Good Stand-In for Mateo?

Judge a proposed mouse study by face validity and construct validity, and apply the 3Rs (Replacement, Reduction, Refinement) and ARRIVE reporting to decide whether the model can honestly…

CARRY FORWARD

CRISPR is a controlled perturbation only when the edit, mosaicism, phenotype, and off-target controls are verified.

10-YEAR TAKEAWAY

A model earns relevance by matching the mechanism and outcome needed for the question while following Replace, Reduce, and Refine.

Socratic launch

Show the analogy panel only: A scale model is useful only for the feature being tested. Do not name the biology yet.

- Which model can test load?
- Which model only matches appearance?
- What question could be answered without a whole animal?

Rules students should earn

- Define the exact mechanism and outcome the model must reproduce.
- Use the least sentient suitable system.
- Replace, Reduce, and Refine animal use while preserving a valid answer.

Biology reveal and evidence

ID	Finding supplied in the case
EXP11-E1	The question asks whether a pathway changes mammalian palatal shelf elevation and fusion.
EXP11-E2	A cell culture can test signaling but cannot reproduce whole-palate elevation.
EXP11-E3	A mouse study uses prior data for sample size, shared controls, analgesia, humane endpoints, and blinded scoring.

Decision and answer guidance

Situation: The team wants to use the largest animal available because it seems more human-like.

Strongest option: Choose the least burdensome model that can measure the required mechanism and outcome.

Evidence required: EXP11-E1 + EXP11-E2

Claim ceiling: You may judge fitness for the stated question. You may not call any model a complete stand-in for Mateo.

Next handoff: We decided a mouse can be a fair stand-in when its construct and face validity hold up. But knocking out a gene and seeing…

TEACHER PLAN 12 OF 20 | LAB SCIENTIST (CAUSAL-INFERENCE BENCH SEAT)

Knock It Out, Then Put It Back: Proving a Gene Causes a Defect

Explain why a knockout alone cannot prove causation, and how adding a rescue (and a redundancy control) completes the causal argument.

CARRY FORWARD

A model earns relevance by matching the mechanism and outcome needed for the question while following Replace, Reduce, and Refine.

10-YEAR TAKEAWAY

Loss plus targeted rescue strengthens causation when the rescue is specific, verified, and compared with controls.

Socratic launch

Show the analogy panel only: Remove one circuit part, then restore that same part. Do not name the biology yet.

- What changes after the part is removed?
- Which replacement restores function?
- Why test a wrong or empty replacement?

Rules students should earn

- Show a phenotype after specific loss of function.
- Restore the targeted function in the relevant place and time.
- Compare rescue with mock and unrelated controls.

Biology reveal and evidence

ID	Finding supplied in the case
EXP12-E1	Wild-type cultures fuse in 18 of 20 samples, while knockout cultures fuse in 4 of 20.
EXP12-E2	Targeted epithelial rescue restores fusion in 15 of 20 knockout samples.
EXP12-E3	Empty-vector rescue produces fusion in 5 of 20 samples.

Decision and answer guidance

Situation: The team must decide whether knockout alone is enough for its strongest causal claim.

Strongest option: Use knockout plus targeted rescue and controls to strengthen the model-specific causal claim.

Evidence required: EXP12-E1 + EXP12-E2

Claim ceiling: You may claim that targeted restoration strengthens causation in this model. You may not claim strict sufficiency or a patient-specific cause.

Next handoff: In a mouse we control everything, so we can prove cause by cutting a gene and putting it back. But we cannot assign a…

The Fairest Test: How to Compare Two Treatments in Real Children

Explain how randomization, blinding, and intention-to-treat make a two-treatment comparison fair, and identify each, plus equipoise, in a study description.

CARRY FORWARD

Loss plus targeted rescue strengthens causation when the rescue is specific, verified, and compared with controls.

10-YEAR TAKEAWAY

Fair treatment trials require equipoise, concealed randomization, unbiased outcome assessment, and analysis by assigned group.

Socratic launch

Show the analogy panel only: A fair race uses concealed lanes, blinded judges, and every assigned runner. Do not name the biology yet.

- Who knows the lane before assignment?
- What is hidden from judges?
- Why keep runners in their assigned lane during analysis?

Rules students should earn

- Randomize only between acceptable options under genuine uncertainty.
- Conceal assignment and blind outcome assessment when feasible.
- Use intention-to-treat to preserve the assigned comparison.

Biology reveal and evidence

ID	Finding supplied in the case
EXP13-E1	Both timings are accepted options and experts remain uncertain about their balance of outcomes.
EXP13-E2	A central system reveals assignment only after an eligible infant is enrolled.
EXP13-E3	Speech assessors receive coded recordings and analysis keeps participants in assigned groups.

Decision and answer guidance

Situation: A surgeon wants to choose timing after seeing the next assignment and remove crossovers from analysis.

Strongest option: Use concealed randomization, coded assessment, and intention-to-treat.

Evidence required: EXP13-E1 + EXP13-E2

Claim ceiling: You may design the comparison when both options are acceptable. You may not randomize repair versus knowingly harmful no repair.

Next handoff: We can design a fair trial on paper. But for Mateo's care, families face a real fight that lasted decades: at what age should…

TEACHER PLAN 14 OF 20 | PRINCIPAL INVESTIGATOR (WITH THE TEAM BIOSTATISTICIAN)

The TOPS Trial: How Surgeons Tested the Best Time to Repair a Palate

Walk a real RCT from question to result, and read its primary outcome (risk ratio, 95% CI, p-value) to state what it proved and what it did not.

CARRY FORWARD

Fair treatment trials require equipoise, concealed randomization, unbiased outcome assessment, and analysis by assigned group.

10-YEAR TAKEAWAY

TOPS supports a narrow claim: earlier repair lowered VPI in its eligible population, with uncertainty and limits.

Socratic launch

Show the analogy panel only: Two calendar windows lead to one coded five-year review. Do not name the biology yet.

- What differs between the two paths?
- What remains standardized?
- Why are the final recordings coded?

Rules students should earn

- Read eligibility before applying the result.
- Report effect size, interval, and outcome definition.
- Carry safety, setting, and cleft-type limits into the conclusion.

Biology reveal and evidence

ID	Finding supplied in the case
EXP14-E1	TOPS randomized 558 medically fit infants with nonsyndromic isolated cleft palate at 23 centers.
EXP14-E2	VPI at age 5 occurred in 8.9 percent of the 6-month group and 15.0 percent of the 12-month group among analyzable participants.
EXP14-E3	The risk ratio was 0.59 with a 95 percent interval from 0.36 to 0.99.

Decision and answer guidance

Situation: The team asks whether TOPS proves every infant with cleft lip and palate should have repair at 6 months in every center.

Strongest option: Apply the result to similar eligible infants while discussing uncertainty, setting, and other outcomes.

Evidence required: EXP14-E1 + EXP14-E2

Claim ceiling: You may state the primary TOPS finding for its eligible population. You may not prescribe timing for Mateo or generalize to all clefts and settings.

Next handoff: TOPS rested on one judgment for every child: did this 5-year-old's speech sound right, or was it nasal and unclear? But good speech is…

How Do You Measure Something as Fuzzy as Speech?

Turn a fuzzy concept like 'good speech' into a defined, standardized outcome measure, and explain why patient-reported outcomes capture success a clinician's score can miss.

CARRY FORWARD

TOPS supports a narrow claim: earlier repair lowered VPI in its eligible population, with uncertainty and limits.

10-YEAR TAKEAWAY

An operational definition makes speech scoring repeatable, while patient-reported outcomes capture effects clinician scores can miss.

Socratic launch

Show the analogy panel only: A calibrated ruler and several listeners turn a fuzzy judgment into a shared measure. Do not name the biology yet.

- Where are the scoring boundaries?
- Why do listeners use the same examples?
- Which experience is only visible to the speaker?

Rules students should earn

- Define each outcome before data collection.
- Train and test raters with coded samples.
- Pair clinical measures with validated patient-reported outcomes when relevant.

Biology reveal and evidence

ID	Finding supplied in the case
EXP15-E1	The protocol defines VPI using specified ratings from standardized speech samples.
EXP15-E2	Assessors are trained, blinded to timing, and agreement is checked on duplicate coded recordings.
EXP15-E3	A validated child or caregiver measure records speech participation and burden.

Decision and answer guidance

Situation: The draft outcome is simply 'speech sounds better,' with no task, rule, or patient voice.

Strongest option: Define the task, rubric, training, blinding, reliability check, and reported outcome.

Evidence required: EXP15-E1 + EXP15-E2

Claim ceiling: You may define reliable study outcomes. You may not claim one score represents every language, dialect, or lived experience.

Next handoff: We now know how to measure a fuzzy outcome like speech carefully and the same way for every child. But even a perfectly measured…

TEACHER PLAN 16 OF 20 | BIOSTATISTICIAN

What Could Fool Us Into the Wrong Conclusion?

Identify bias, confounding, and the surgeon-as-confounder problem in a real cleft study, and name the design defenses (matching, randomization, blinding).

CARRY FORWARD

An operational definition makes speech scoring repeatable, while patient-reported outcomes capture effects clinician scores can miss.

10-YEAR TAKEAWAY

Bias is directional design or measurement error; confounding mixes effects; each threat needs a matched safeguard.

Socratic launch

Show the analogy panel only: A crooked mirror and a hidden magnet distort results in different ways. Do not name the biology yet.

- Which problem changes the measurement itself?
- Which adds another cause linked to both sides?
- Would a larger sample fix either problem?

Rules students should earn

- Name the direction and stage of each bias.
- Draw the confounder linked to exposure and outcome.
- Choose a safeguard that fits the threat.

Biology reveal and evidence

ID	Finding supplied in the case
EXP16-E1	Earlier-repair participants are more likely to receive care at high-volume centers with strong speech services.
EXP16-E2	Assessors know each child's group and expect earlier repair to work.
EXP16-E3	Follow-up is 92 percent in one group and 71 percent in the other, with access-related reasons.

Decision and answer guidance

Situation: The team says a sample of 5,000 makes bias and confounding disappear.

Strongest option: Match each supplied threat with a design, measurement, or analysis safeguard.

Evidence required: EXP16-E1 + EXP16-E2

Claim ceiling: You may reduce named threats. You may not claim adjustment removes all unmeasured bias or confounding.

Next handoff: We answered today's question: bias, confounding, and weak blinding can all fool us. But even a careful single study can be a fluke or…

How Do We Combine Many Studies Into One Answer?

Explain how a systematic review pools studies using PRISMA, read a PRISMA flow diagram, and state when studies are too different to combine (heterogeneity).

CARRY FORWARD

Bias is directional design or measurement error; confounding mixes effects; each threat needs a matched safeguard.

10-YEAR TAKEAWAY

Systematic reviews reduce cherry-picking, while meta-analysis is credible only when studies and pooling assumptions support it.

Socratic launch

Show the analogy panel only: A recipe search needs a sieve before ingredients enter one bowl. Do not name the biology yet.

- Which records are found before the sieve?
- Who sets the inclusion holes and when?
- Why might some ingredients remain outside the bowl?

Rules students should earn

- Predefine eligibility, search, outcomes, and analysis.
- Assess risk of bias before pooling.
- Interpret heterogeneity and pool only when the combined estimate has clear meaning.

Biology reveal and evidence

ID	Finding supplied in the case
EXP17-E1	The protocol was registered before screening and lists databases, eligibility, outcomes, and analysis.
EXP17-E2	Two studies use blinded age-5 VPI ratings, while one uses an unvalidated parent question at age 2.
EXP17-E3	The forest plot shows effects in different directions and major clinical differences across centers.

Decision and answer guidance

Situation: The authors want one dramatic number even though one study measures a different outcome and timing.

Strongest option: Present the review and pool only a justified compatible subset.

Evidence required: EXP17-E1 + EXP17-E2

Claim ceiling: You may synthesize and pool compatible studies when justified. You may not claim a pooled estimate repairs weak or incompatible evidence.

Next handoff: Combining studies only works if the studies were done on real children fairly. But every study we pool was run on real babies and…

What Makes Research on Children Ethical?

Explain why a pediatric study needs IRB approval, parental informed consent, child assent, and genuine equipoise, and apply these rules to decide whether a proposed cleft study could…

CARRY FORWARD

Systematic reviews reduce cherry-picking, while meta-analysis is credible only when studies and pooling assumptions support it.

10-YEAR TAKEAWAY

Child research requires favorable risk and benefit, IRB review, parental permission, and affirmative assent when the child is capable.

Socratic launch

Show the analogy panel only: A field trip needs safety review, adult permission, and child agreement. Do not name the biology yet.

- Who reviews the plan before anyone joins?
- Whose permission is needed?
- How can a child show an affirmative choice?

Rules students should earn

- Use added protections for children and minimize risk.
- Obtain parental permission unless an approved exception applies.
- Seek affirmative, age-appropriate assent when the child is capable.

Biology reveal and evidence

ID	Finding supplied in the case
EXP18-E1	The study involves children, coded recordings, and no change from accepted care.
EXP18-E2	The protocol gives caregivers plain-language information and time for questions.
EXP18-E3	Children who can understand receive an age-appropriate explanation and are asked for affirmative agreement.

Decision and answer guidance

Situation: A protocol says parental signature alone is enough and children should not be told because that is faster.

Strongest option: Require IRB-approved protections, permission, and capability-based assent.

Evidence required: EXP18-E1 + EXP18-E2

Claim ceiling: You may apply the supplied educational ethics framework. You may not replace local IRB review or legal guidance.

Next handoff: We answered today's question: an ethical study needs IRB approval, real consent and assent, and genuine equipoise. But clearing ethics only means a study…

How Does the World Check That a Study Is Trustworthy?

Explain how reproducibility, reporting standards (CONSORT for trials, STROBE for observational studies), and peer review let outsiders check a study, and use a reporting checklist to spot what…

CARRY FORWARD

Child research requires favorable risk and benefit, IRB review, parental permission, and affirmative assent when the child is capable.

10-YEAR TAKEAWAY

Transparent reporting lets others appraise and reproduce methods, but it cannot repair a weak design.

Socratic launch

Show the analogy panel only: A lab notebook must let an outside team repeat the route. Do not name the biology yet.

- Which detail lets the next team repeat the method?
- Where are missing participants shown?
- What if the original design was unfair but perfectly reported?

Rules students should earn

- Report planned and actual methods, participants, outcomes, and analyses.
- Use the guideline matched to the study type.
- Separate reporting from design quality, replication, and peer review.

Biology reveal and evidence

ID	Finding supplied in the case
EXP19-E1	The trial report includes eligibility, allocation, participant flow, outcomes, harms, and analysis.
EXP19-E2	The study type determines the matched guideline: CONSORT, STROBE, PRISMA, or ARRIVE.
EXP19-E3	A fully reported study still has unconcealed allocation and severe missing data.

Decision and answer guidance

Situation: An author says completing a reporting checklist proves the study is trustworthy.

Strongest option: Use the matched guideline, then judge design, conduct, analysis, and replication separately.

Evidence required: EXP19-E1 + EXP19-E2

Claim ceiling: You may judge reporting completeness and exposed strengths or weaknesses. You may not call a study reproducible or valid from a checklist alone.

Next handoff: We answered today's question: reporting standards, reproducibility, and peer review let the world check a study instead of trusting it on faith. Across this…

TEACHER PLAN 20 OF 20 | PRINCIPAL INVESTIGATOR (THE WHOLE TEAM)

What New Question About Mateo Would You Investigate, and How?

Design an original, ethical, well-controlled study to answer a new question about Mateo's cleft using the full experimental-design checklist, and state what converging evidence taught us about how…

CARRY FORWARD

Transparent reporting lets others appraise and reproduce methods, but it cannot repair a weak design.

10-YEAR TAKEAWAY

A defensible study plan aligns question, design, variables, sampling, analysis, ethics, reporting, and a claim ceiling.

Socratic launch

Show the analogy panel only: A blueprint board checks every connection before construction. Do not name the biology yet.

- Which card determines the design?
- Where could one broken connection invalidate the plan?
- Which boundary stops the claim from outrunning evidence?

Rules students should earn

- Make every method trace back to the question.
- Supply data, safeguards, and decision rules before results are known.
- Write the strongest claim the design could support and no stronger.

Biology reveal and evidence

ID	Finding supplied in the case
EXP20-E1	The question asks whether a structured speech follow-up program is associated with improved age-8 communication after palate repair.
EXP20-E2	The supplied synthetic CSV contains 24 coded records with program group, center, baseline score, age-8 score, and missing-data flags.
EXP20-E3	The analysis compares adjusted estimates, reports intervals and missingness, protects coded data, and follows STROBE.

Decision and answer guidance

Situation: The board must select one complete plan from three proposals using only the supplied case file and dataset.

Strongest option: Approve the aligned coded-records cohort with its observational claim ceiling.

Evidence required: EXP20-E1 + EXP20-E2

Claim ceiling: You may defend an adjusted observational association using the supplied dataset. You may not claim causation or require outside data.

Next handoff: We answered the whole domain's question with this one: you can take a fresh question about Mateo's cleft and design a sound, ethical study…