STUDENT CASE FILE | EXPERIMENTAL DESIGN DOMAIN | LESSONS 1 TO 20

The Baby Mateo Case

One observed structural difference. Twenty evidence updates. One bounded clinical synthesis.

Composite patient snapshot

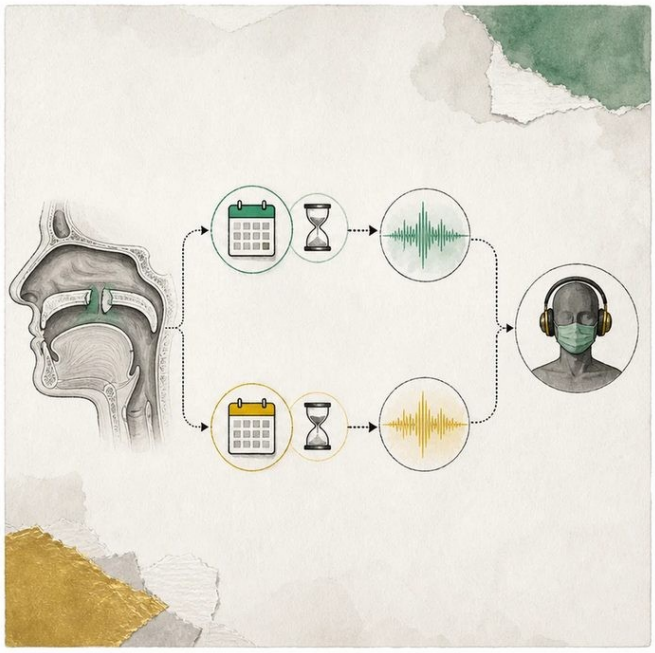
Patient	Mateo, a composite teaching case. No real patient data.
Birth	Term newborn. Breathing and tone are stable.
Observed structure	Complete unilateral left cleft lip and palate.
Initial whole-body exam	No additional congenital anomalies identified on the first structured examination.
Inquiry target	Is the cleft isolated or part of a syndrome? What evidence supports the cause model and care plan?
Evidence rule	Use only the labeled findings supplied in the lesson. Do not invent history, tests, symptoms, or family details.

EVIDENCE SUPPLIED

Every required finding has a stable ID.

EVIDENCE NOT SUPPLIED

Do not invent missing symptoms, tests, exposures, or family history.

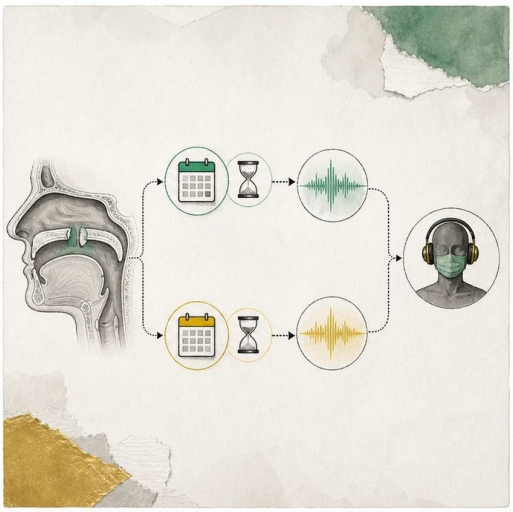


Illustration, not a clinical photograph. Composite craniofacial teaching model.

EVIDENCE UPDATE 1 OF 20 | PRINCIPAL INVESTIGATOR (PI)

From an Observation to a Researchable Question

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



Biomedical teaching illustration: Frame the palate-timing question with PICO

ID	Finding supplied in the case	Why it matters
EXP01-E1	The open anatomical question is whether repair timing changes later speech function.	The question concerns a treatment choice, not the cause of Mateo's cleft.
EXP01-E2	A supplied candidate population is medically fit infants with isolated cleft palate.	The population is specific enough to set eligibility rules.
EXP01-E3	VPI can be scored at age 5 by trained assessors.	The outcome and time point can be defined before the study begins.

Decision in this update: Choose the researchable question and label its population, intervention, comparison, outcome, and time.

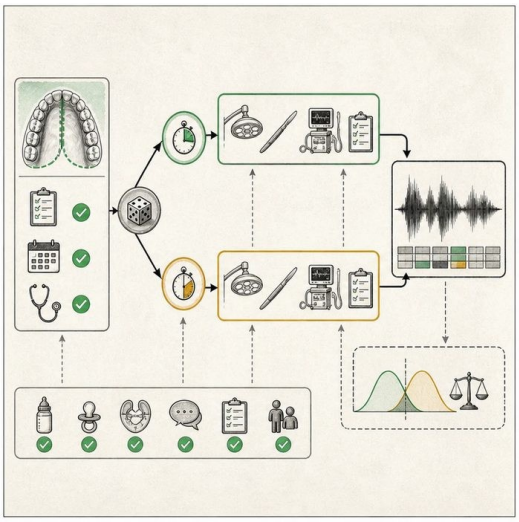
Claim ceiling: You may define the study question. You may not predict which timing is better before data are collected.

Handoff to the next update: We turned an observation into a sharp PICO question. But a question is still not a prediction. How do we turn a PICO question into a hypothesis we can actually prove wrong, with variables we control and a clear way to fail?

EVIDENCE UPDATE 2 OF 20 | PRINCIPAL INVESTIGATOR (PI)

From a Question to a Testable Hypothesis

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



Biomedical teaching illustration: Convert the PICO question into a testable timing hypothesis

ID	Finding supplied in the case	Why it matters
EXP02-E1	The proposed study compares repair at 6 months with repair at 12 months.	Repair timing is the factor that differs between groups.
EXP02-E2	The primary outcome is VPI status at age 5.	The outcome must be measured the same way in both groups.
EXP02-E3	A valid hypothesis can be challenged by equal or higher VPI in the 6-month group.	The prediction is falsifiable rather than protected from failure.

Decision in this update: Choose the revision and identify the independent variable, dependent variable, and falsifying result.

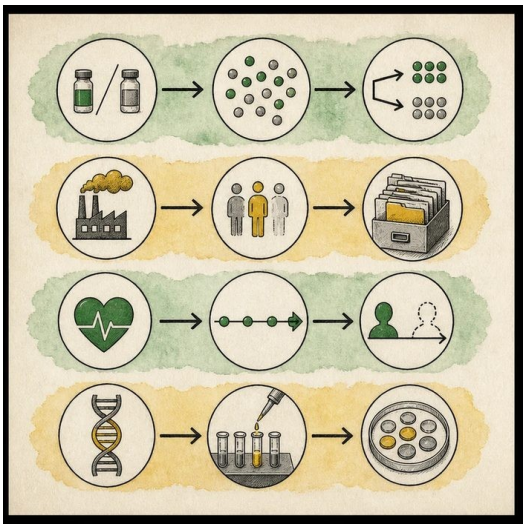
Claim ceiling: You may state a testable prediction. You may not claim the hypothesis is true before the comparison.

Handoff to the next update: We can design a single study and predict what would prove us wrong. But two honest studies on the same question can reach opposite answers. Why do scientists trust some study designs more than others?

EVIDENCE UPDATE 3 OF 20 | PRINCIPAL INVESTIGATOR (PI)

Why We Trust Some Studies More Than Others

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



Biomedical teaching illustration: Match four cleft questions to four study designs

ID	Finding supplied in the case	Why it matters
EXP03-E1	Repair timing is an assignable clinical choice when both options are acceptable.	A randomized trial can compare the treatment options under equipoise.
EXP03-E2	A suspected harmful pregnancy exposure cannot be assigned.	An observational design is required for that cause question.
EXP03-E3	A gene mechanism requires perturbation and molecular or tissue outcomes.	A bench experiment answers a different question than a treatment trial.

Decision in this update: Choose the design rule and match the three supplied questions to appropriate designs.

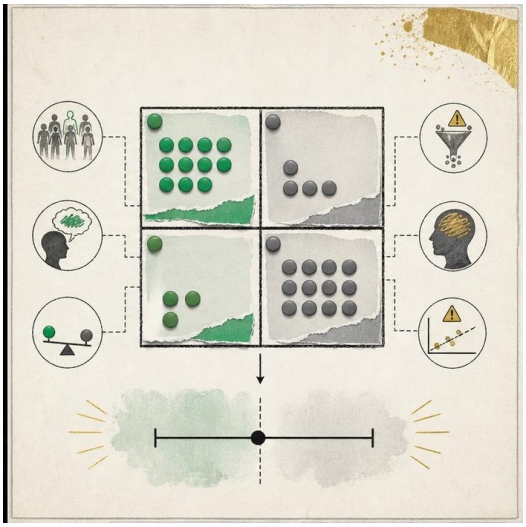
Claim ceiling: You may select a design family. You may not call any design free of bias or automatically decisive.

Handoff to the next update: A randomized trial sits near the top because we can assign who gets the treatment. But we could never assign a baby to a harmful chemical in the womb. So how does science study a cause it is forbidden to hand out?

EVIDENCE UPDATE 4 OF 20 | EPIDEMIOLOGIST ON THE CLEFT RESEARCH TEAM

Studying a Cause You Cannot Assign

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



Biomedical teaching illustration: Test a suspected exposure with a supplied 2 by 2 table

ID	Finding supplied in the case	Why it matters
EXP04-E1	The table has 30 exposed and 70 unexposed cases, plus 15 exposed and 85 unexposed controls.	The odds ratio is about 2.4.
EXP04-E2	The reported 95 percent confidence interval is 0.9 to 3.4.	The interval includes 1.0, so the data remain compatible with no odds difference.
EXP04-E3	Parents of affected infants may recall an exposure differently while searching for an explanation.	Recall bias could inflate or reduce the observed association.

Decision in this update: Choose the interpretation and cite the odds ratio, interval, and one bias risk.

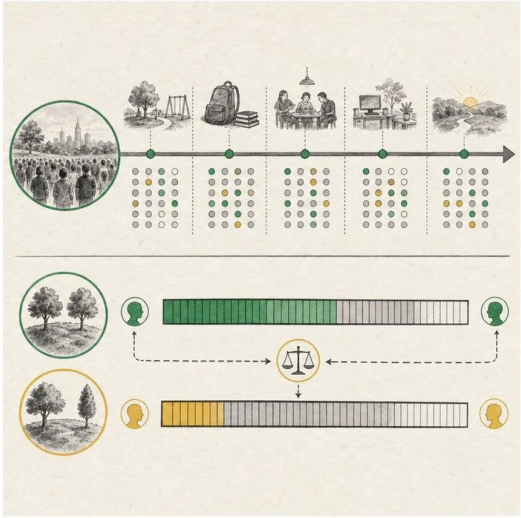
Claim ceiling: You may report the observed association and uncertainty. You may not claim individual blame or causation.

Handoff to the next update: 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 we measure how much of cleft risk is inherited rather than environmental, when you cannot give a memory questionnaire about DNA?

EVIDENCE UPDATE 5 OF 20 | EPIDEMIOLOGIST ON THE CLEFT RESEARCH TEAM

Measuring Inheritance Over Time and in Twins

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



Biomedical teaching illustration: Compare a prospective cohort with a twin concordance study

ID	Finding supplied in the case	Why it matters
EXP05-E1	A cohort records an exposure before birth outcomes are known.	This supports time order and reduces some recall problems.
EXP05-E2	Monozygotic twins share more DNA than dizygotic twins on average.	Higher concordance can support a genetic contribution at the population level.
EXP05-E3	Twin estimates rely on assumptions, and heritability changes across populations and settings.	The estimate does not assign a percentage cause to Mateo.

Decision in this update: Choose the defensible claim and name one strength plus one assumption for each design.

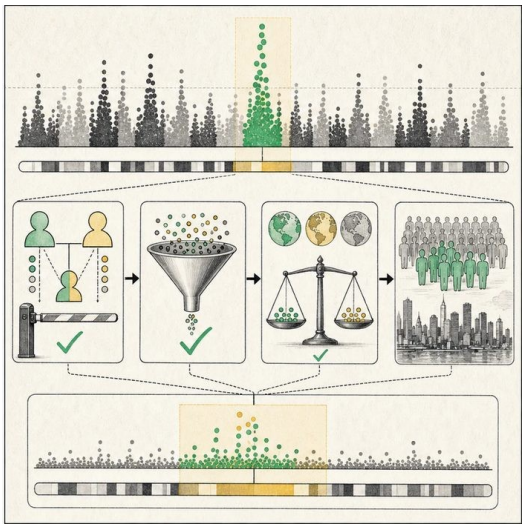
Claim ceiling: You may discuss time order and population variance. You may not assign Mateo's cleft a heritability percentage or single cause.

Handoff to the next update: The twin study tells us genes carry most of the risk, but heritability is just a number; it does not name a single gene. Out of roughly twenty thousand genes, how do scientists track down the specific risk gene hiding among millions of bases?

EVIDENCE UPDATE 6 OF 20 | DATA SCIENTIST ON THE CLEFT RESEARCH TEAM

Finding a Risk Gene Among Millions of Bases

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



Biomedical teaching illustration: Move from a GWAS signal to a replicated cleft-risk locus

ID	Finding supplied in the case	Why it matters
EXP06-E1	A GWAS compares millions of common variants across many people.	It can reveal risk-associated regions without selecting one candidate in advance.
EXP06-E2	A case-parent trio can test whether one allele is transmitted to affected children more often than expected.	Family structure can reduce some population-stratification problems.
EXP06-E3	Genotyping quality, ancestry structure, sample size, and independent replication affect credibility.	A single peak is not yet a proven causal gene.

Decision in this update: Choose the next step and distinguish a locus, an associated variant, and a causal mechanism.

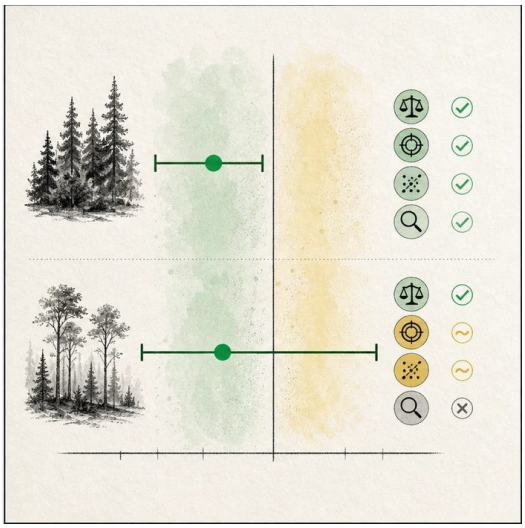
Claim ceiling: You may nominate a replicated risk locus. You may not name a causal gene or patient diagnosis from proximity alone.

Handoff to the next update: A scan and a trio test both flagged rs642961 near IRF6. But a flag is not proof. With millions of SNPs and hundreds of trios, some hits happen by pure luck. How do we tell a real association from a lucky roll of the dice?

EVIDENCE UPDATE 7 OF 20 | BIOSTATISTICIAN ON THE CLEFT RESEARCH TEAM

Is the Association Real, or Just Chance?

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



Biomedical teaching illustration: Interpret two treatment estimates without a significance shortcut

ID	Finding supplied in the case	Why it matters
EXP07-E1	Study A reports a risk ratio of 0.60 with a 95 percent interval from 0.37 to 0.98.	The estimate suggests lower risk, with values near no difference still close to the interval edge.
EXP07-E2	Study B reports a risk ratio of 0.55 with an interval from 0.18 to 1.67.	The result is much less precise and includes possible benefit, no difference, or harm.
EXP07-E3	A p-value is calculated under a statistical model and null hypothesis.	It is not the probability that the scientific claim is true or false.

Decision in this update: Choose the interpretation and explain why the two intervals support different precision claims.

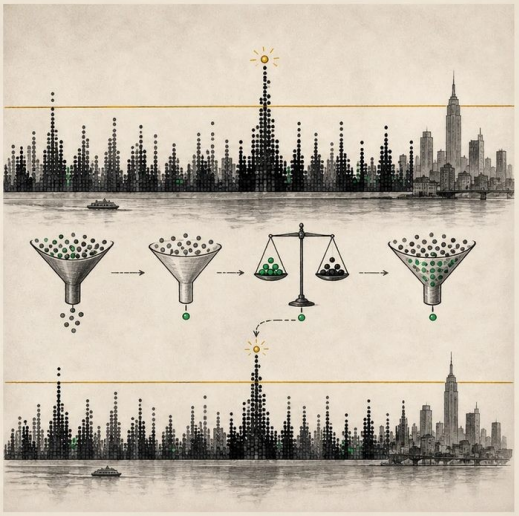
Claim ceiling: You may describe magnitude, precision, and statistical compatibility. You may not convert a threshold into scientific truth.

Handoff to the next update: 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 many flukes slip through a million tests, and why do gene studies demand a far tinier p-value?

EVIDENCE UPDATE 8 OF 20 | BIOSTATISTICIAN ON THE CLEFT RESEARCH TEAM

Why Gene Studies Use Such Tiny P-Values

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



Biomedical teaching illustration: Control false positives in a common-variant GWAS

ID	Finding supplied in the case	Why it matters
EXP08-E1	Testing one million variants at p below 0.05 would produce many chance hits under the null.	A usual single-test threshold is not adequate for a genome-wide scan.
EXP08-E2	For many common-variant GWAS analyses, p below 5 x 10^-8 is a conventional threshold.	The convention is not universal for every genetic analysis.
EXP08-E3	The candidate peak repeats in an independent cohort with the same direction of effect.	Replication lowers concern that the first signal was sample-specific chance or error.

Decision in this update: Choose the report language and explain how multiplicity and replication change the claim.

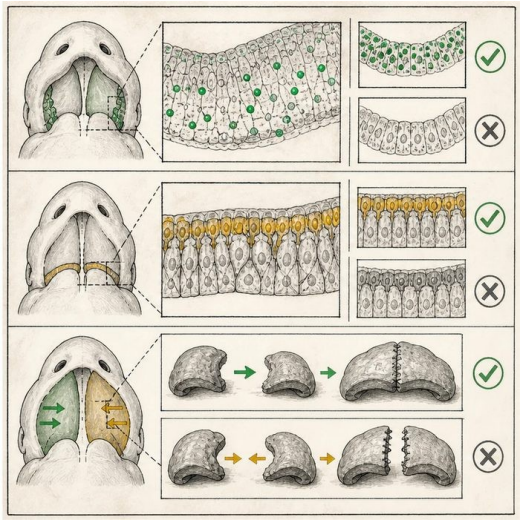
Claim ceiling: You may apply the supplied common-variant GWAS convention. You may not treat that number as universal or proof of mechanism.

Handoff to the next update: 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 a rock-solid statistical link never proves IRF6 does anything to tissue. How would we leave the spreadsheet, go to the bench, and test what IRF6 actually does to a developing palate?

EVIDENCE UPDATE 9 OF 20 | LAB SCIENTIST ON THE CLEFT RESEARCH TEAM

Taking a Gene to the Bench

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



Biomedical teaching illustration: Test a candidate developmental gene with three assays

ID	Finding supplied in the case	Why it matters
EXP09-E1	RNA signal appears in developing palatal shelves at the relevant stage.	The gene is transcribed in a useful place and time.
EXP09-E2	Protein staining appears in palatal epithelium and is absent in the no-primary-antibody control.	The protein location is supported by a specificity control.
EXP09-E3	Perturbed embryos show altered shelf elevation more often than matched controls.	Changing the gene is associated with a phenotype, but other effects still need checks.

Decision in this update: Choose the assay set and state the specific question each assay answers.

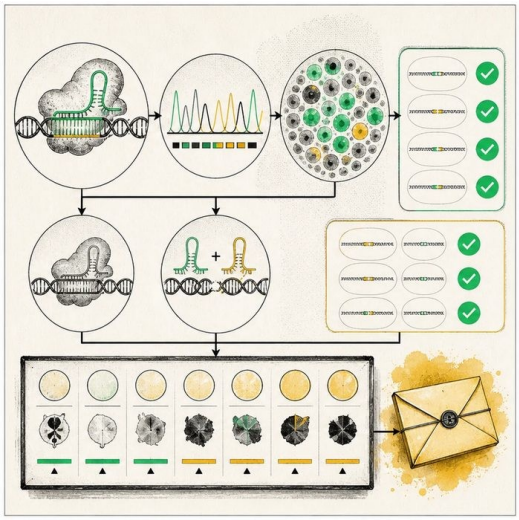
Claim ceiling: You may connect expression and perturbation evidence. You may not treat location alone as causal proof.

Handoff to the next update: 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 crosses and many months. Is there a faster, more precise way to edit a gene on purpose and read out exactly what it does?

EVIDENCE UPDATE 10 OF 20 | LAB SCIENTIST ON THE CLEFT RESEARCH TEAM

CRISPR as an Experimental Tool

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



Biomedical teaching illustration: Verify a CRISPR perturbation before reading the palate phenotype

ID	Finding supplied in the case	Why it matters
EXP10-E1	Sequencing confirms a frameshift in 78 percent of reads from the target tissue.	The edit occurred, but the sample is mosaic.
EXP10-E2	A second guide produces a similar phenotype while a non-targeting guide does not.	Independent perturbation and a negative control strengthen specificity.
EXP10-E3	The top predicted off-target sites show no detectable edits in the supplied assay.	The tested concern is reduced, not eliminated everywhere in the genome.

Decision in this update: Choose the claim status and cite the minimum verification steps needed next.

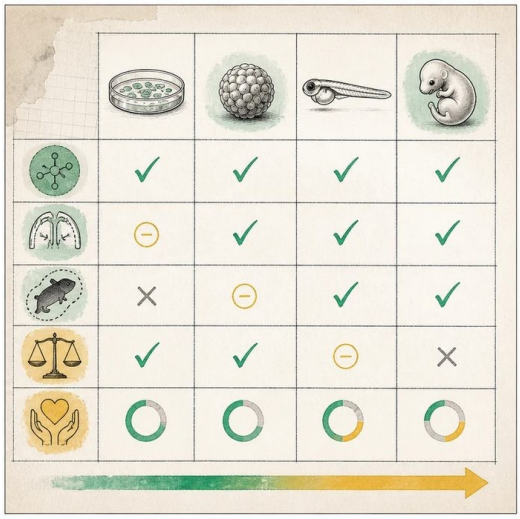
Claim ceiling: You may interpret the supplied edit and control checks. You may not claim all off-target effects are absent or a human treatment is ready.

Handoff to the next update: 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 for him, and when does the species gap make the answer untrustworthy?

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

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

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



Biomedical teaching illustration: Select a model for palatal shelf elevation

ID	Finding supplied in the case	Why it matters
EXP11-E1	The question asks whether a pathway changes mammalian palatal shelf elevation and fusion.	A model must represent the relevant tissue movement and timing.
EXP11-E2	A cell culture can test signaling but cannot reproduce whole-palate elevation.	It can replace animals for some early questions, not the full outcome.
EXP11-E3	A mouse study uses prior data for sample size, shared controls, analgesia, humane endpoints, and blinded scoring.	The design applies Reduce and Refine while protecting validity.

Decision in this update: Choose the model sequence and justify it with mechanism fit plus the 3Rs.

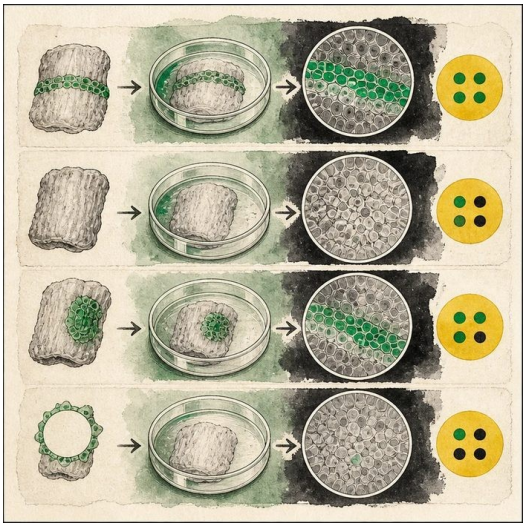
Claim ceiling: You may judge fitness for the stated question. You may not call any model a complete stand-in for Mateo.

Handoff to the next update: 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 a cleft is not yet proof the gene caused it; something else changed by the same edit could be to blame. So next: how do we prove the gene itself, and nothing else, caused the cleft? We chase that next time.

EVIDENCE UPDATE 12 OF 20 | LAB SCIENTIST (CAUSAL-INFERENCE BENCH SEAT)

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

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



Biomedical teaching illustration: Test knockout and epithelial rescue of a palate phenotype

ID	Finding supplied in the case	Why it matters
EXP12-E1	Wild-type cultures fuse in 18 of 20 samples, while knockout cultures fuse in 4 of 20.	Loss of function is associated with a large defect in this model.
EXP12-E2	Targeted epithelial rescue restores fusion in 15 of 20 knockout samples.	Restoring function in the relevant tissue recovers much of the outcome.
EXP12-E3	Empty-vector rescue produces fusion in 5 of 20 samples.	The delivery procedure alone does not explain the targeted rescue.

Decision in this update: Choose the claim and compare all four supplied groups.

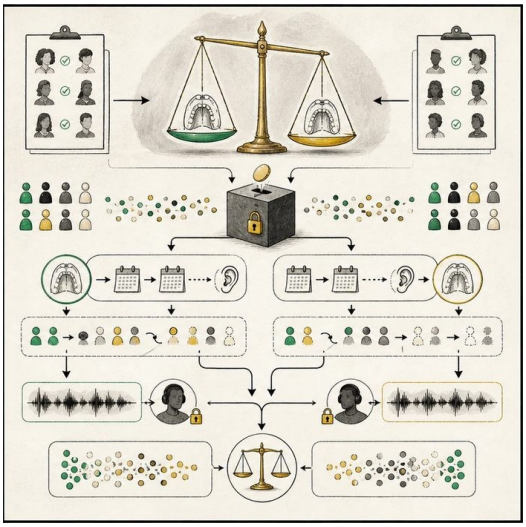
Claim ceiling: You may claim that targeted restoration strengthens causation in this model. You may not claim strict sufficiency or a patient-specific cause.

Handoff to the next update: In a mouse we control everything, so we can prove cause by cutting a gene and putting it back. But we cannot assign a gene, or a surgery, to a baby like Mateo on purpose. When the subjects are children and the question is which treatment is better, how do we make a fair comparison without controlling their biology directly? We chase that next time.

EVIDENCE UPDATE 13 OF 20 | PRINCIPAL INVESTIGATOR (CLINICAL-TRIAL DESIGN SEAT)

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

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



Biomedical teaching illustration: Design a fair comparison of two accepted palate-repair timings

ID	Finding supplied in the case	Why it matters
EXP13-E1	Both timings are accepted options and experts remain uncertain about their balance of outcomes.	Clinical equipoise can support randomization.
EXP13-E2	A central system reveals assignment only after an eligible infant is enrolled.	Allocation concealment reduces selection manipulation.
EXP13-E3	Speech assessors receive coded recordings and analysis keeps participants in assigned groups.	Blinding and intention-to-treat protect the comparison.

Decision in this update: Choose the fair-trial plan and explain the threat controlled by each safeguard.

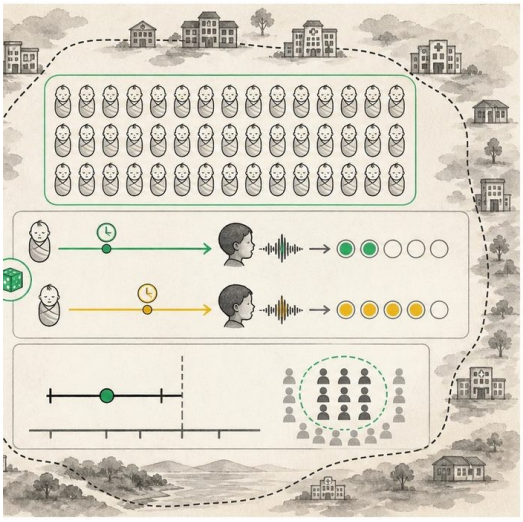
Claim ceiling: You may design the comparison when both options are acceptable. You may not randomize repair versus knowingly harmful no repair.

Handoff to the next update: 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 a cleft palate be repaired? Earlier might help speech, but might harm facial growth. How did scientists run a real trial to settle that exact question, and what did they decide? We chase that next time.

EVIDENCE UPDATE 14 OF 20 | PRINCIPAL INVESTIGATOR (WITH THE TEAM BIOSTATISTICIAN)

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

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



Biomedical teaching illustration: Read the TOPS trial from eligibility to claim ceiling

ID	Finding supplied in the case	Why it matters
EXP14-E1	TOPS randomized 558 medically fit infants with nonsyndromic isolated cleft palate at 23 centers.	The study population did not include every cleft type or setting.
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.	The earlier group had fewer primary-outcome events.
EXP14-E3	The risk ratio was 0.59 with a 95 percent interval from 0.36 to 0.99.	The estimate favors earlier repair, but other outcomes and uncertainty matter.

Decision in this update: Choose the briefing and cite population, event rates, risk ratio, and one limit.

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.

Handoff to the next update: 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 not a number a machine prints out. How do you actually measure something as fuzzy as speech, fairly and the same way for hundreds of children? We chase that next time.

EVIDENCE UPDATE 15 OF 20 | STUDY COORDINATOR (OUTCOME-MEASUREMENT SEAT)

How Do You Measure Something as Fuzzy as Speech?

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

Biomedical teaching illustration: Operationalize VPI and speech participation

ID	Finding supplied in the case	Why it matters
EXP15-E1	The protocol defines VPI using specified ratings from standardized speech samples.	Every group is measured with the same rule.
EXP15-E2	Assessors are trained, blinded to timing, and agreement is checked on duplicate coded recordings.	The process tests whether scoring is reliable.
EXP15-E3	A validated child or caregiver measure records speech participation and burden.	The study includes an outcome that clinical resonance scores may miss.

Decision in this update: Choose the measurement plan and explain what each measure adds.

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

Handoff to the next update: We now know how to measure a fuzzy outcome like speech carefully and the same way for every child. But even a perfectly measured outcome can still point us to the wrong cause. A real cleft trial once found its result was driven not by the treatment but by something hiding in the background. What could fool us, even when our measurements are flawless? We chase that next time.

EVIDENCE UPDATE 16 OF 20 | BIOSTATISTICIAN

What Could Fool Us Into the Wrong Conclusion?

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

Biomedical teaching illustration: Red-team a speech study for selection, measurement, and confounding

ID	Finding supplied in the case	Why it matters
EXP16-E1	Earlier-repair participants are more likely to receive care at high-volume centers with strong speech services.	Center resources could confound an observational comparison.
EXP16-E2	Assessors know each child's group and expect earlier repair to work.	Expectation could create directional measurement bias.
EXP16-E3	Follow-up is 92 percent in one group and 71 percent in the other, with access-related reasons.	Differential attrition could change the comparison.

Decision in this update: Choose the red-team response and propose one matched safeguard per evidence card.

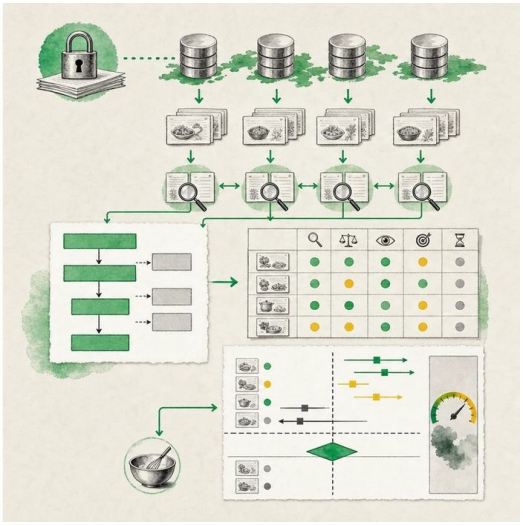
Claim ceiling: You may reduce named threats. You may not claim adjustment removes all unmeasured bias or confounding.

Handoff to the next update: 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 can disagree with the next study. How do we gather every study on a question and combine them into one trustworthy answer? We chase that next time.

EVIDENCE UPDATE 17 OF 20 | DATA SCIENTIST

How Do We Combine Many Studies Into One Answer?

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



Biomedical teaching illustration: Move from a PRISMA flow diagram to a defensible forest plot

ID	Finding supplied in the case	Why it matters
EXP17-E1	The protocol was registered before screening and lists databases, eligibility, outcomes, and analysis.	The process is harder to change after seeing results.
EXP17-E2	Two studies use blinded age-5 VPI ratings, while one uses an unvalidated parent question at age 2.	Outcome differences may limit a meaningful pooled estimate.
EXP17-E3	The forest plot shows effects in different directions and major clinical differences across centers.	Heterogeneity must be explained rather than hidden by one average.

Decision in this update: Choose the synthesis plan and cite protocol, outcome compatibility, and heterogeneity evidence.

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

Handoff to the next update: Combining studies only works if the studies were done on real children fairly. But every study we pool was run on real babies and families. What rules protect children in research in the first place, and who decides what is allowed? We chase that next time.

EVIDENCE UPDATE 18 OF 20 | BIOETHICIST / IRB LIAISON

What Makes Research on Children Ethical?

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

Biomedical teaching illustration: Apply Subpart D protections to a pediatric speech study

ID	Finding supplied in the case	Why it matters
EXP18-E1	The study involves children, coded recordings, and no change from accepted care.	The IRB must assess risk, privacy, and added protections.
EXP18-E2	The protocol gives caregivers plain-language information and time for questions.	Permission should be informed and voluntary.
EXP18-E3	Children who can understand receive an age-appropriate explanation and are asked for affirmative agreement.	Assent is more than silence or failure to object.

Decision in this update: Choose the ethics correction and explain the role of the IRB, permission, and assent.

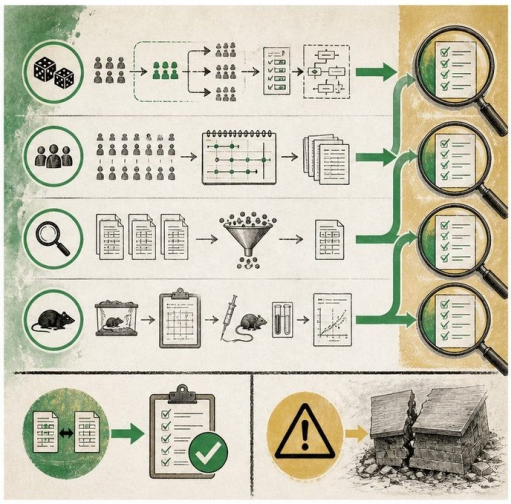
Claim ceiling: You may apply the supplied educational ethics framework. You may not replace local IRB review or legal guidance.

Handoff to the next update: 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 was allowed to run, not that its result is true. Suppose a study cleared every ethics check and was run perfectly. How does the rest of the world confirm it is trustworthy and not just one team's claim? We chase that next time.

EVIDENCE UPDATE 19 OF 20 | SCIENCE WRITER

How Does the World Check That a Study Is Trustworthy?

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



Biomedical teaching illustration: Match CONSORT, STROBE, PRISMA, and ARRIVE to four studies

ID	Finding supplied in the case	Why it matters
EXP19-E1	The trial report includes eligibility, allocation, participant flow, outcomes, harms, and analysis.	Readers can check what happened from enrollment to result.
EXP19-E2	The study type determines the matched guideline: CONSORT, STROBE, PRISMA, or ARRIVE.	One checklist does not fit every design.
EXP19-E3	A fully reported study still has unconcealed allocation and severe missing data.	Transparency exposes the weakness but does not remove it.

Decision in this update: Choose the editorial decision and match each supplied study to its guideline.

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.

Handoff to the next update: 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 whole domain you have learned to ask a researchable question, build a hypothesis, pick a design, control bias, run the stats, clear ethics, and report it all. So here is the last test: what new question about Mateo's cleft would YOU investigate, and how would you design a study to answer it? We chase that next time, and this time the study is yours.

EVIDENCE UPDATE 20 OF 20 | PRINCIPAL INVESTIGATOR (THE WHOLE TEAM)

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

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

Biomedical teaching illustration: Defend one complete Mateo-related study protocol

ID	Finding supplied in the case	Why it matters
EXP20-E1	The question asks whether a structured speech follow-up program is associated with improved age-8 communication after palate repair.	The outcome and population are specific, but the program is not randomly assigned.
EXP20-E2	The supplied synthetic CSV contains 24 coded records with program group, center, baseline score, age-8 score, and missing-data flags.	Students can complete the design without outside recruitment or invented measurements.
EXP20-E3	The analysis compares adjusted estimates, reports intervals and missingness, protects coded data, and follows STROBE.	The strongest result is an adjusted association, not proof of cause.

Decision in this update: Choose the defensible protocol and justify question, design, variables, sample, analysis, ethics, reporting, and claim ceiling.

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

Handoff to the next update: 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, and you discovered his cleft is nonsyndromic and multifactorial by reasoning across the evidence, not by being told. The studies you read this year were designed by scientists years ago. The next generation of evidence about clefts like Mateo's will come from studies that students like you design. So the question that opens onto the rest of your life: what will you ask, and who will you become to answer it?