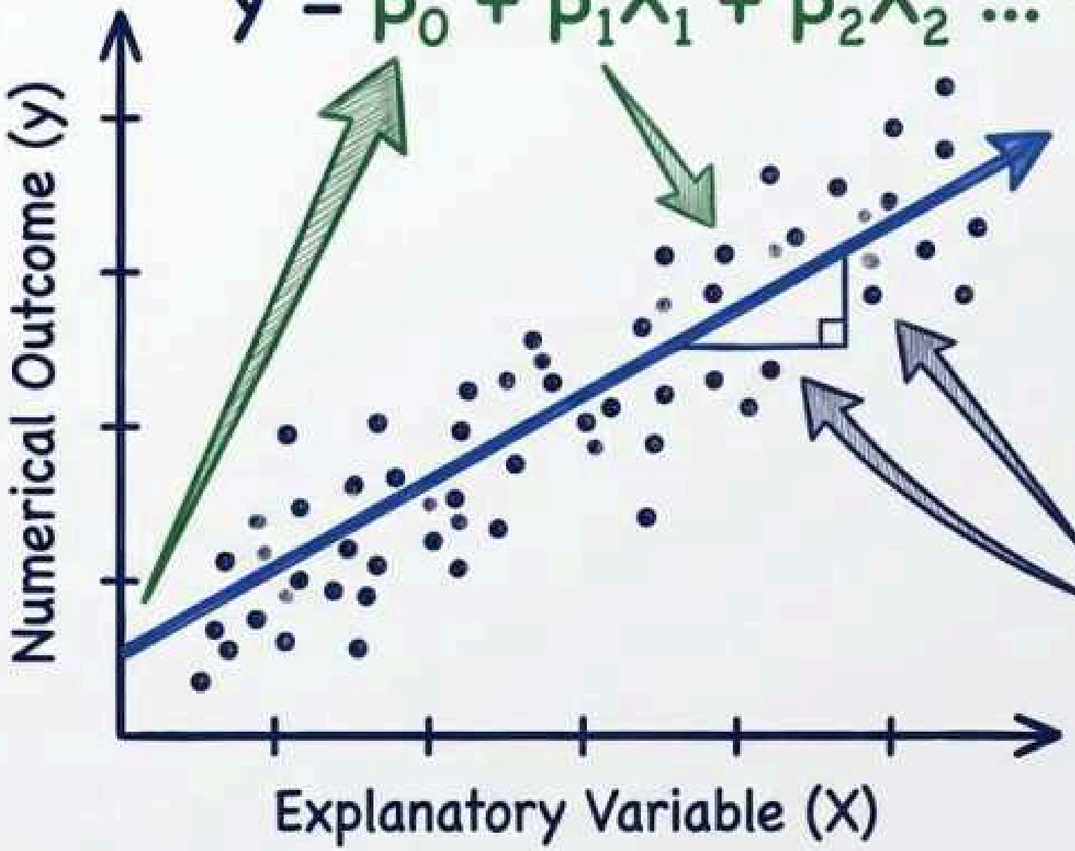
Linear Regression

¶ 6.1

Core Concept: Models how a numerical outcome depends on one or more explanatory variables.

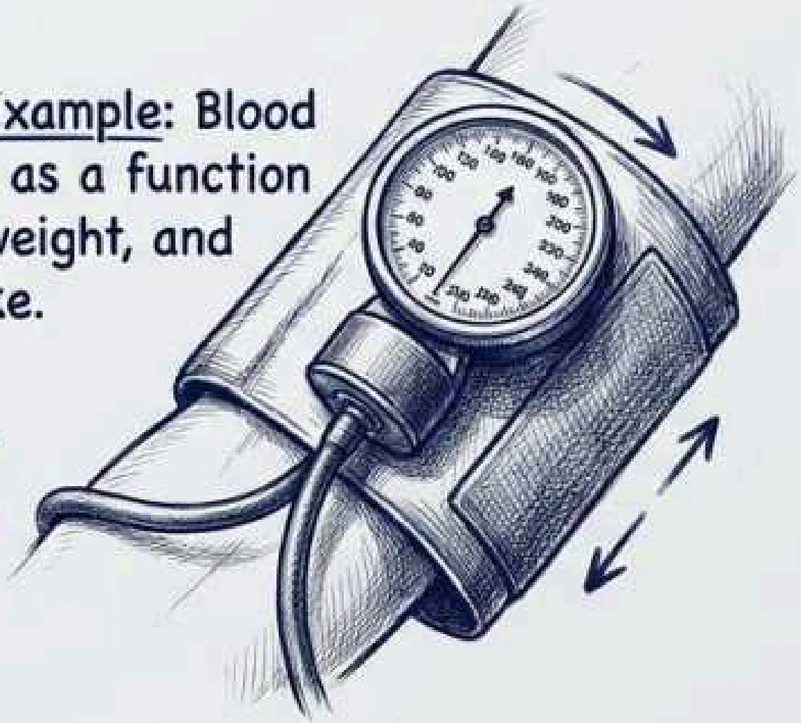
$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \ldots$$

The Slope (β): Estimates the exact unit change in the outcome for every single 1-unit change in a specific factor.



Explanator Numerical Variable

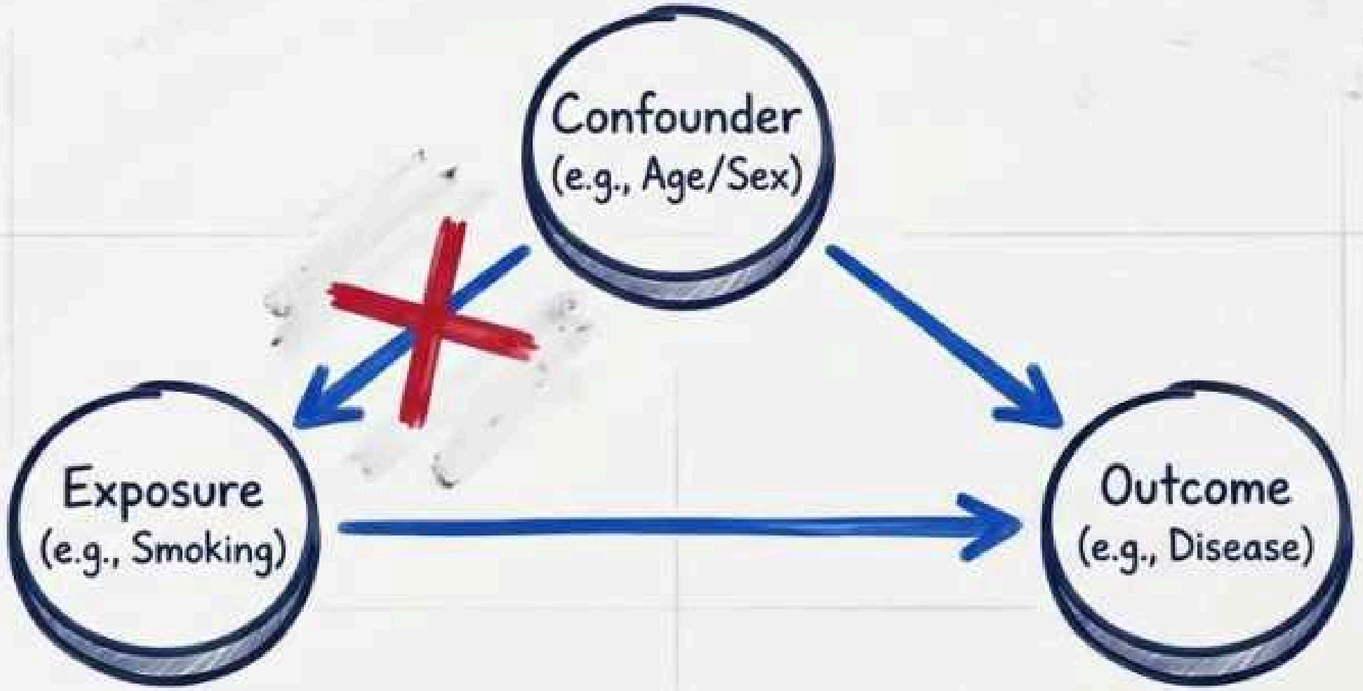
Clinical Example: Blood pressure as a function of age, weight, and salt intake.



Adjusting for Confounders

¶ 6.2

The Power of Multiple Factors: Regression includes several variables at once to estimate the effect of one factor while holding others mathematically constant.



Reality Check: In observational studies, adjusting removes some confounding, but can never eliminate it entirely.

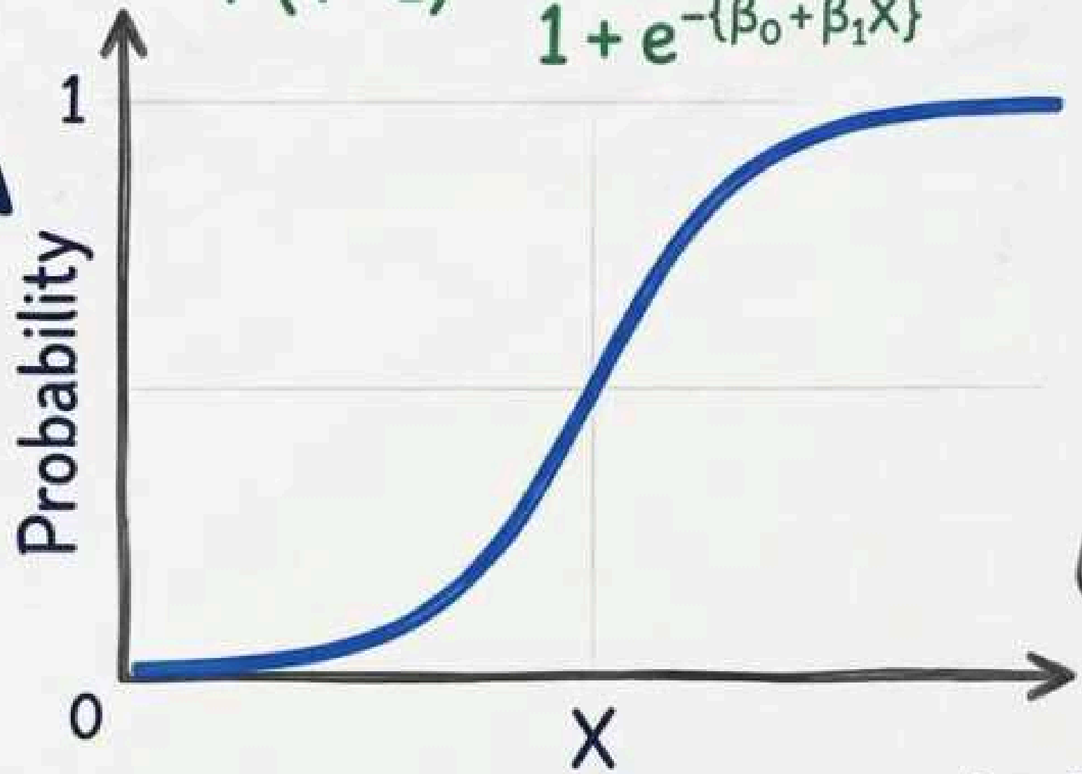
Gemini Notebook

Logistic Regression & Odds

¶ 6.3

$$P(Y=1) = \frac{1}{1 + e^{-\{\beta_0 + \beta_1 X\}}}$$

The Outcome: **Binary** (Yes/No)—for example, does the patient develop diabetes?





The Output: Produces **Adjusted Odds Ratios** for each factor rather than linear slopes.

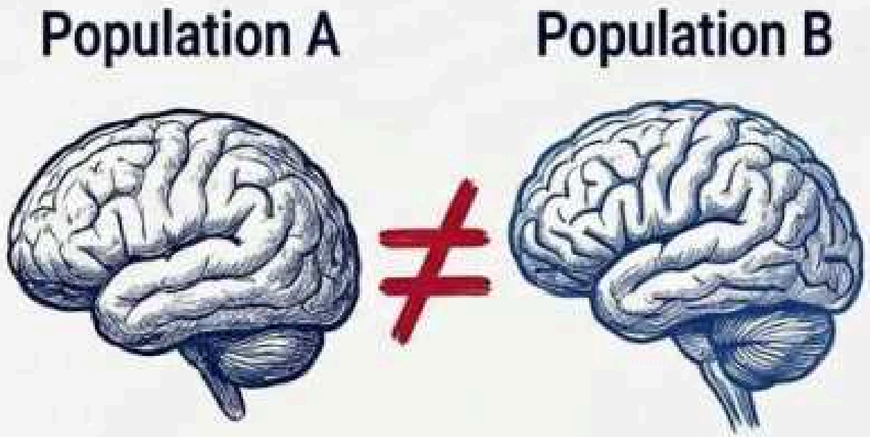
Application: This mathematical engine powers modern **clinical risk scores**, estimating a patient's exact chance of a future event.

Gemini Notebook

The Limits of Regression

¶ 6.4







1. **Hidden Variables:** Models only adjust for what is measured. Unmeasured confounders remain.

2. **Population Mismatch:** A model built on one population may not predict well in another.

3. **Overfitting:** Adding too many factors to a small data set fits the noise rather than the signal.

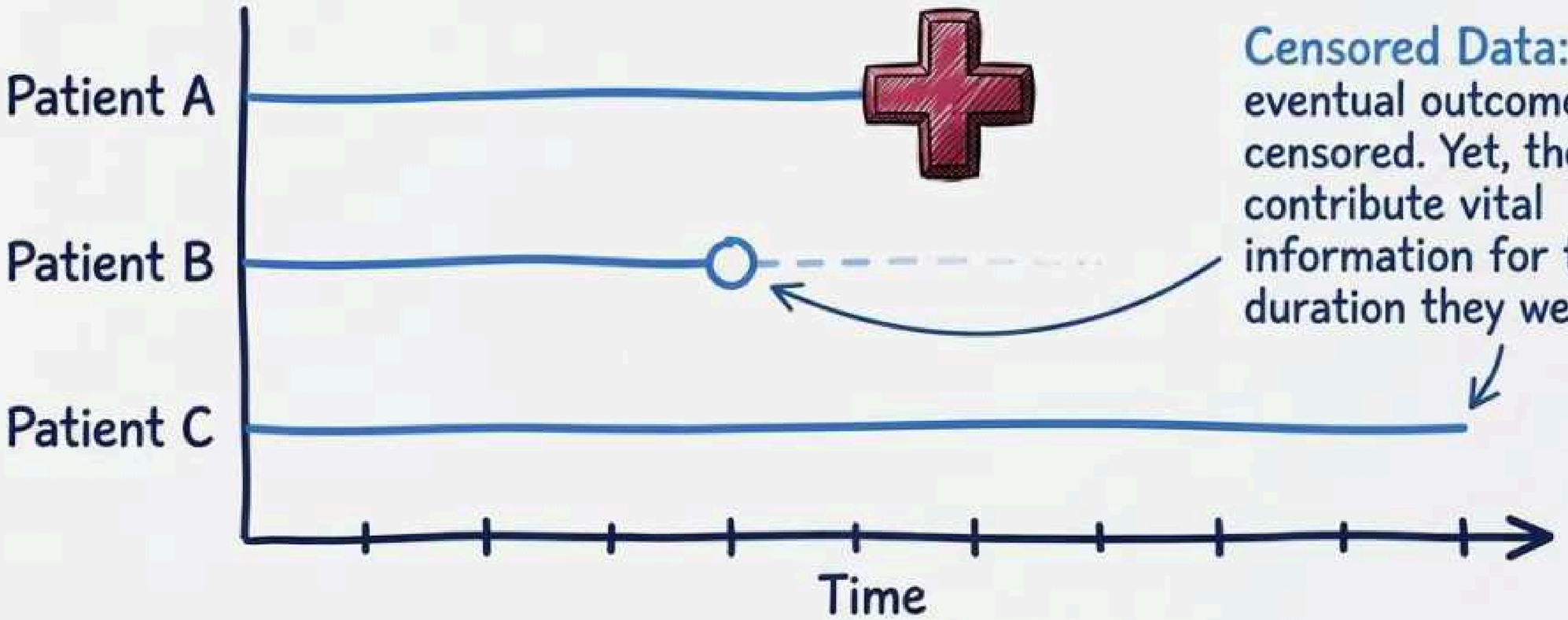
Gemini Notebook

Survival Analysis & Censoring

¶ 6.5



The Challenge: Tracking events over time (death, relapse, recovery) when patients leave the study early or stay well until the end.



Censored Data: Unknown eventual outcomes are censored. Yet, they still contribute vital information for the exact duration they were followed.

Gemini Notebook

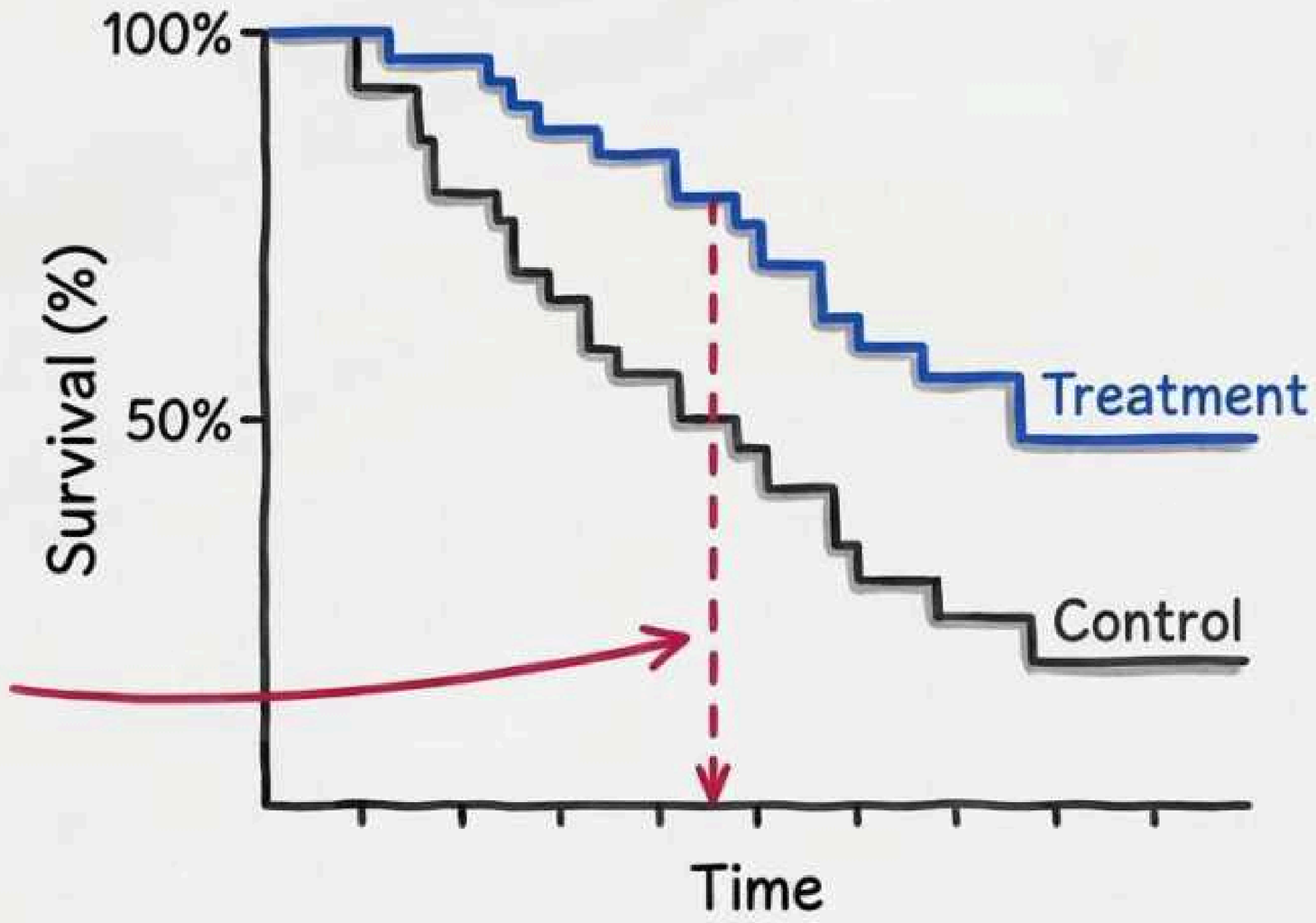
The Kaplan-Meier Method (1958)

¶ 6.6

Visualizing Survival: Shows the proportion of patients still free of the event at each point in time.

Log-Rank Test: The statistical test used to mathematically compare the difference between the curves.

Median Survival: The exact time point by which half of the patients have experienced the event.

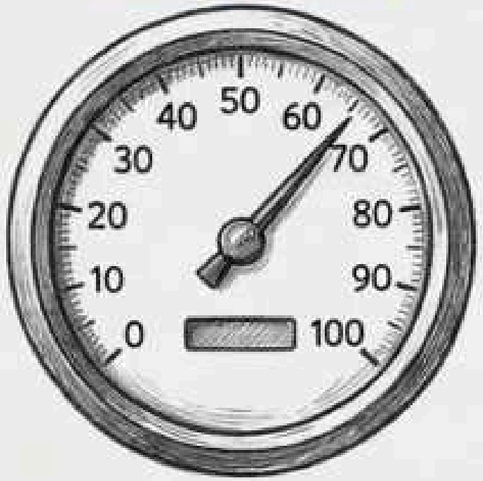


Gemini Notebook

Cox Proportional Hazards (1972)

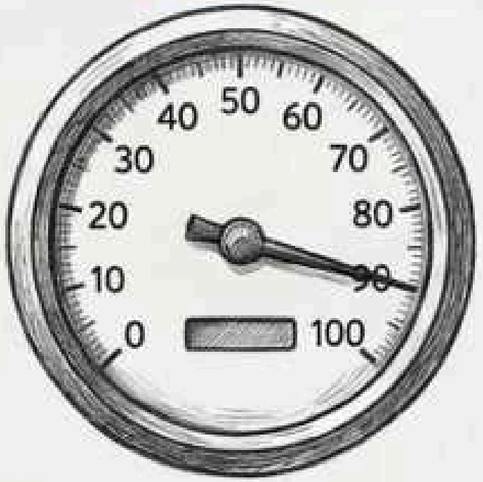
¶ 6.7

Treatment



$$h(t) = h_0(t)\, e^{(\beta_1 X_1 + \ldots)}$$

Control



The Regression of Survival

Produces Hazard Ratios (HR) to compare the rate of events between groups at any given moment.

Reading the Ratio

An HR = 0.7 means the treated group experiences events at about 70% of the control group's rate, safely adjusted for other factors in the model.

Gemini Notebook

Evaluating the Whole Evidence

¶ 6.8

Systematic Review: An exhaustive search for every relevant study on a question using a rigidly defined method.

Meta-Analysis: Statistically combining these gathered results to yield a single, highly precise estimate.

The Goal: Single studies can mislead; the aggregated whole reveals the truth (e.g., Cochrane Collaboration reviews).



Gemini Notebook

The Hazards of Meta-Analyses

¶ 6.9



Publication Bias: Studies with positive results are more likely to be published, artificially inflating how effective treatments look while negative results are buried.



Quality Control: Garbage In = Garbage Out. Combining poor-quality individual studies inevitably produces a poor-quality answer. Readers must rigorously check the foundational ingredients.

Gemini Notebook

¶ 6.10



Appraising the Literature: Six Critical Questions

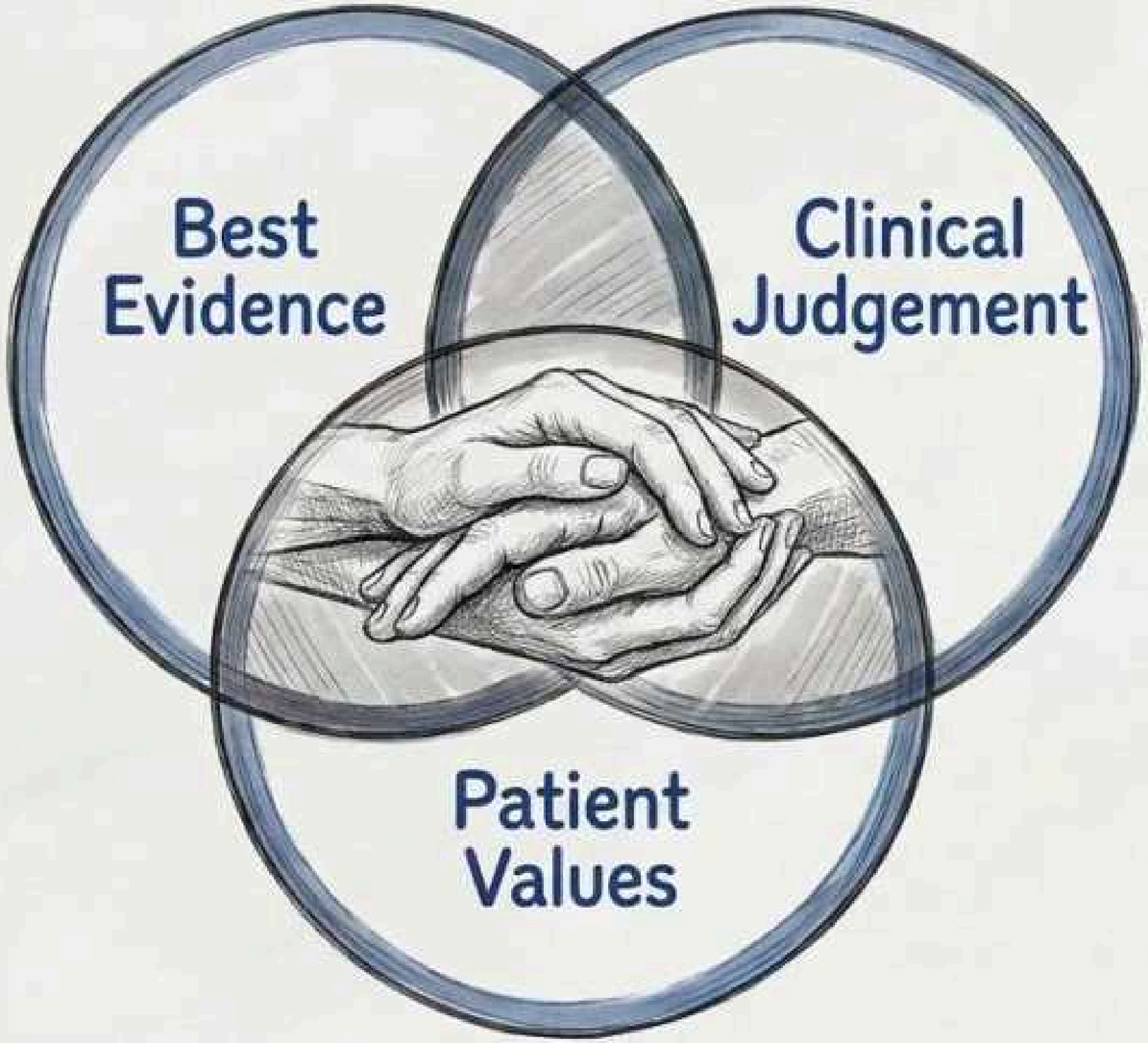
- ☑ **Population:** What was the exact question, and who were the participants?
- ☑ **Randomization:** Were the comparison groups formed randomly?
- ☑ **Outcomes:** Was the main outcome chosen in advance, and does it actually matter to patients?
- ☑ **Impact:** How large is the effect in absolute terms, and what is the Confidence Interval (CI)?
- ☑ **Bias:** Who funded this study? Could anything other than the treatment explain this result?

Gemini Notebook

Statistics Informs, Humans Decide

¶ 6.11

Context is King:
A treatment with a real statistical benefit might be right for one patient but entirely wrong for another.



Good Medicine:
Requires fusing the best rigorous statistical evidence with nuanced clinical judgement, tailored to the specific risk-tolerance and preferences of the patient in front of you.

Gemini Notebook

The Golden Rules of Data

¶ 6.12

1. Describe data honestly.
2. Express uncertainty with Confidence Intervals.
3. Read p-values with extreme care.
4. Insist on absolute risks.
5. Association ≠ Causation.
6. Judge every study by its design.



Questions or feedback? Email: garyanderson@mail.grokbot.com X: @garyanderson

Gemini Notebook