Evidence-Based Decision Making In Healthcare

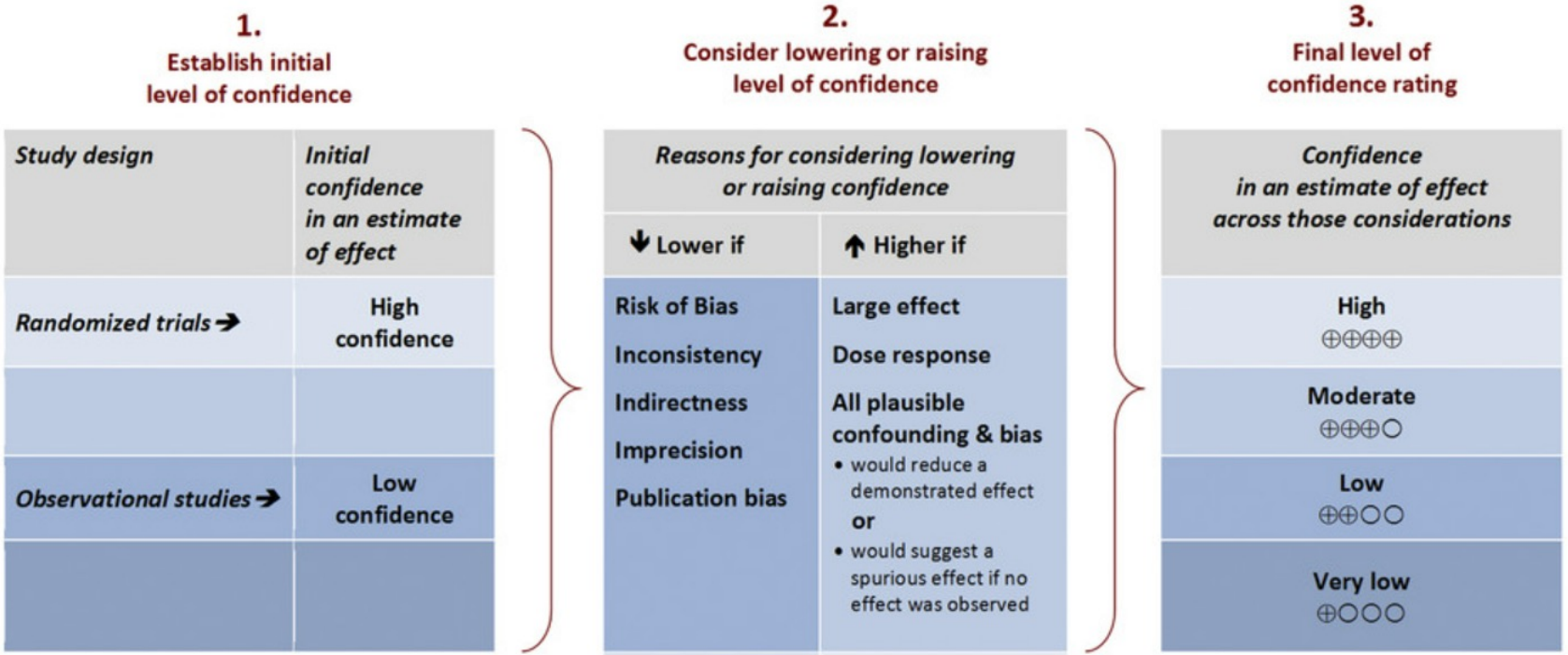
Imprecision, Inconsistency, and Indirectness

Dr Jay K. Varma
https://drjayvarma.com

GRADE

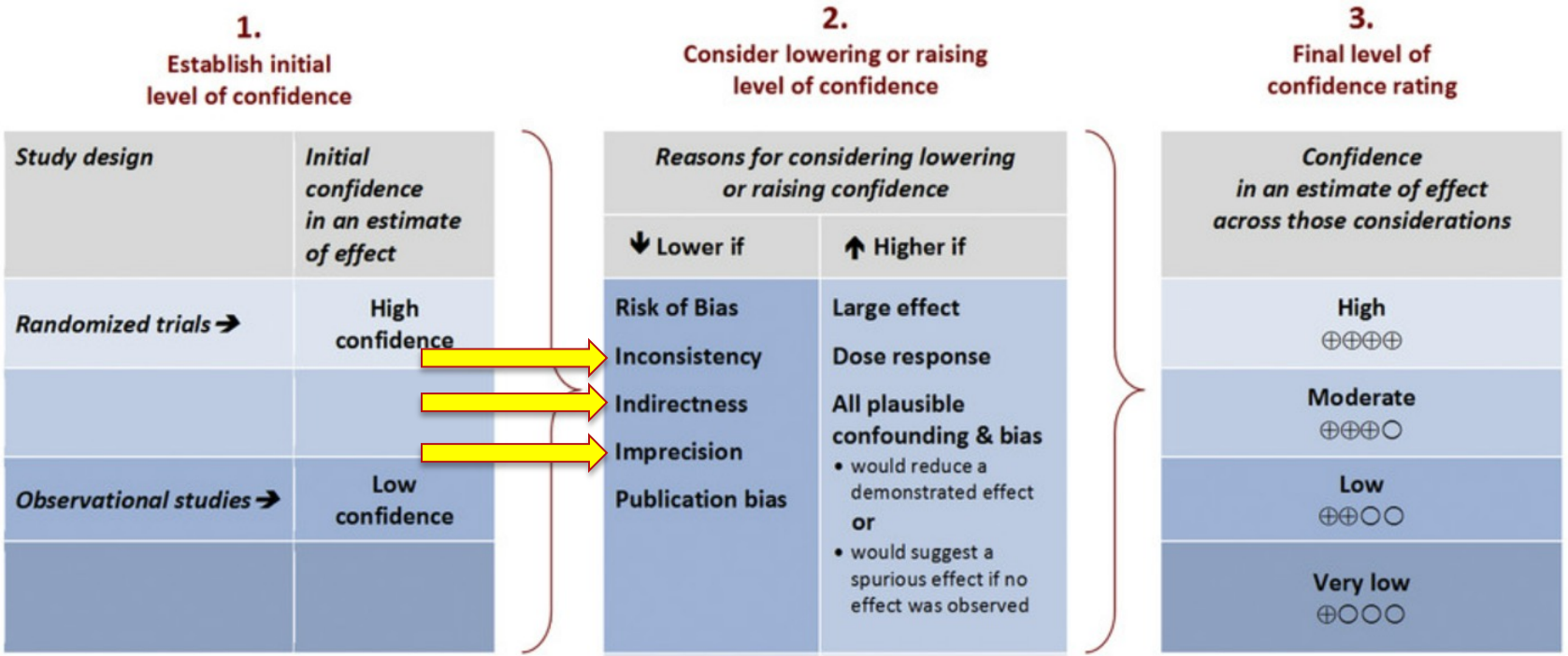
- Summary of evidence and systematic approach to make recommendations
- Reviews quality of evidence with the study design
  - 5 reasons to rate down, 3 reasons to rate up

Rating Quality of Evidence



Source: Environment International 2016

Rating Quality of Evidence



Source: Environment International 2016

Grading Quality of Evidence

Grade	Definition
High 	We are **very confident** that the true effect lies close to that of the estimate of the effect
Moderate 	We are **moderately confident** in the effect estimate: the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low 	Our **confidence** in the effect estimate **is limited**: the true effect may be substantially different from the estimate of the effect
Very Low 	We have **very little confidence** in the effect estimate: the true effect is likely to be substantially different from the estimate of effect

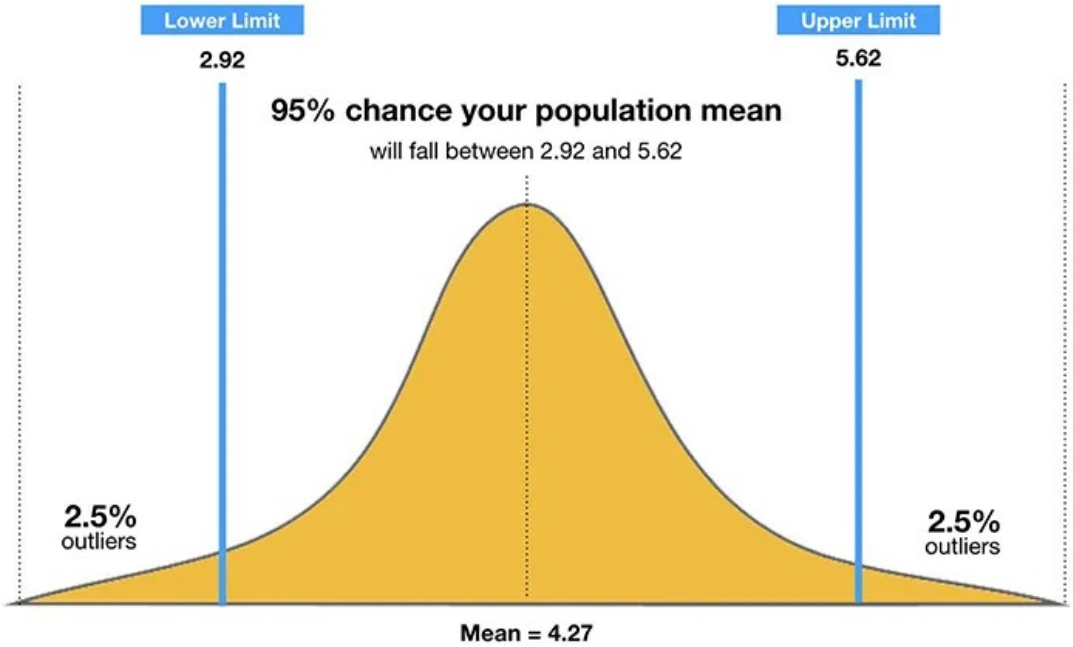
IMPRECISION

Confidence Intervals

Primary criteria are width of CI intervals

- Dependent on sample size and number of events
- Wide confidence intervals (CI) around the estimate of treatment effect occur when studies have relatively **few patients** and **few events**
- If CI is not sufficiently narrow, then **downgrade** quality of evidence

Confidence Interval



- The **95% CI** contains the true **mean** of the population
- As the **sample size increases**, the range of interval values will narrow (*more precise*), and the mean is **more accurate**

Optimal Information Size

Secondary Criteria: Optimal Information Size (OIS)

- If total number of patients included in **systematic review** is ***less*** than number generated by conventional sample size calculation for single adequately powered trial, then **downgrade** for quality of evidence

Power & Sample Size

Factors to consider for Power & Sample Size calculations:

Delta (Effect Size):	With a larger effect size, we either achieve a higher power with a fixed sample size or require a smaller sample for achieving a fixed desired power.
Standard deviation:	It represents the variability in the distribution of hospital stay. With a higher standard deviation, we need a larger sample to achieve the desired power.
Alpha:	Threshold for type I error. It's usually set to be 0.05.
Beta:	Threshold for type II error. (Note that 1-beta is the "power" and it's usually set to be 0.2)

Type I and Type II Error

α, β, and Δ

α = Probability of Type I error
β = Probability of type II error
Power = 1 - β
Δ = The difference (CI)

Type I error = Rejecting the H_0, when H_0 is true
Type II error = Fail to rejecting the H_0, when H_0 is false

	H_0	
	True	False
Reject H_0	Type I error	✓
Fail to reject H_0	✓	Type II error

Example: H_0 = John's car is safe to drive

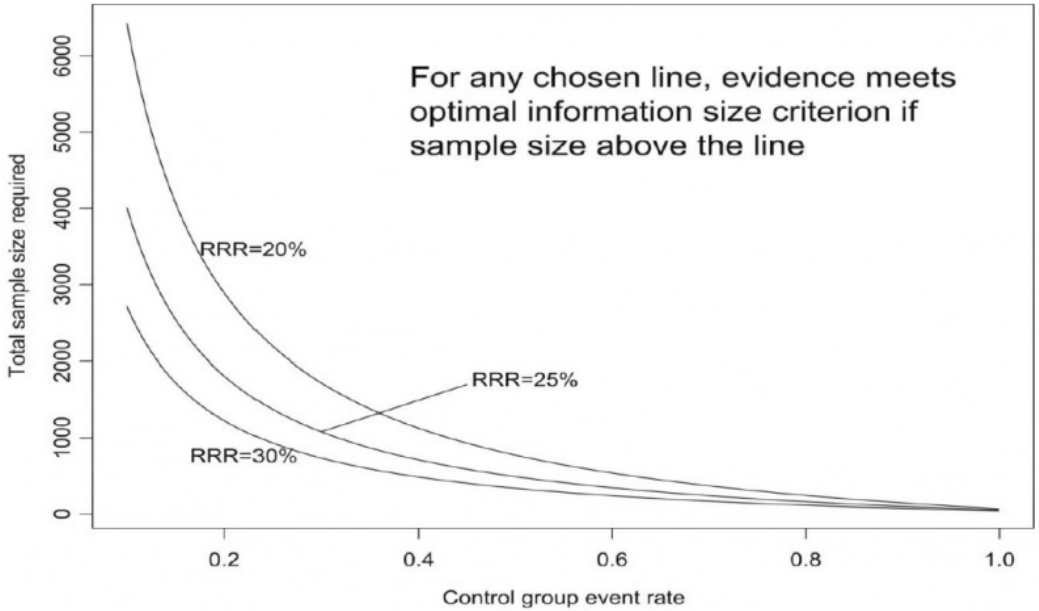
Scenario (1): John thinks his car may be safe to drive, in fact, it is not safe.
Scenario (2): John thinks his car may be safe to drive, in fact it is safe.
Scenario (3): John thinks his car may NOT be safe to drive, in fact, it is not safe.
Scenario (4): John thinks his car may NOT be safe to drive, in fact, it is safe.

Optimal Information Size

Simulation: Generating sample size

- Systematic review of fluoroquinolone prophylaxis for patients with neutropenia and infection-related mortality (95% CI: 0.21, 0.69).
- Total number of events only 69, total patients 1022
- Considering control group risk=6.9%, α=0.05, β=0.2, RRR of 25% → OIS=6400 patients
- Fails to meet OIS, so rate down for imprecision

Optimal Information Size



Required sample size (α=0.05, β=0.2) for RRR of 20%, 25%, 30% across varying control group risks

If best estimate of control group risk is 0.2 and want to measure RRR of 25%, OIS ~ 2000 patients

Optimal Information Size

Total Number of Events	Relative Risk Reduction	Implications for meeting OIS threshold
100 or less	≤ 30%	Will almost never meet threshold whatever control event rate
200	30%	Will meet threshold for control event rates for ~25% or greater
200	25%	Will meet threshold only control event rates for ~50% or greater
200	20%	Will meet threshold only for control event rates for ~80% or greater
300	≥ 30%	Will meet threshold
300	25%	Will meet threshold for control event rates ~25% or greater
300	20%	Will meet threshold for control event rates ~60% or greater
400 or more	≥ 25%	Will meet threshold for any control event rate
400 or more	20%	Will meet threshold for control event rates of ~40% or greater

Probiotics in Crohn's Disease

- Point estimate of risk ratio (0.96) suggests no difference
- CI includes both reduction in likelihood of remission ***and*** an increase in the likelihood (95% CI: 0.56, 1.69)

	Remission	No remission	Total
Treatment	4 (0.8)	1	5
Control	5 (0.83)	1	6
Total	9	2	11
Risk Ratio (RR)	0.96 (0.8 / 0.83) ➔ No difference		
CI (95%)	0.5 (0.56 and 1.69)		

Rate down quality of evidence, because:

Few events (OIS)

CI includes both appreciable benefit (RR>1.25) and harm (RR<0.75)

INCONSISTENCY

What is Inconsistency?

- Inconsistency is an unexplained heterogeneity or variability of study outcomes between studies
- Reduces the confidence in the generalizability of findings
- Lower quality of evidence

Why Assess Inconsistency?

- Determine if it is appropriate to combine study results to estimate the right outcome
- Apples to apples *vs.* apples to oranges

Questions to Ask

- Are there different patient subgroups with different effects?
- Are there different interventions?
- Are there different outcomes defined in the studies?
- Are there differences in the quality of the studies?

Rating Inconsistency

- Magnitude of inconsistency in the results
- Ambiguous or conflicting conclusions
- Rating for inconsistency
  - Not serious (no change in the grading of quality)
  - Serious (downgrade quality by 1 level)
  - Very Serious (downgrade quality by 2 levels)

Differences in direction, but minimal variability

- Differences in direction small
- Differences in point estimate small

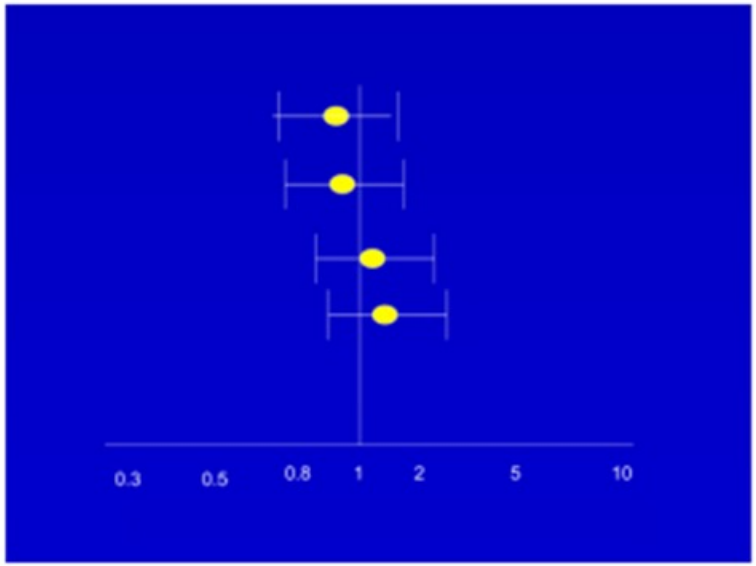


Fig. 1. Differences in direction, but minimal heterogeneity.

Inconsistency Large, but Differences between Small and Large Benefit

- Large variability
- Differences are between small and large treatment effects

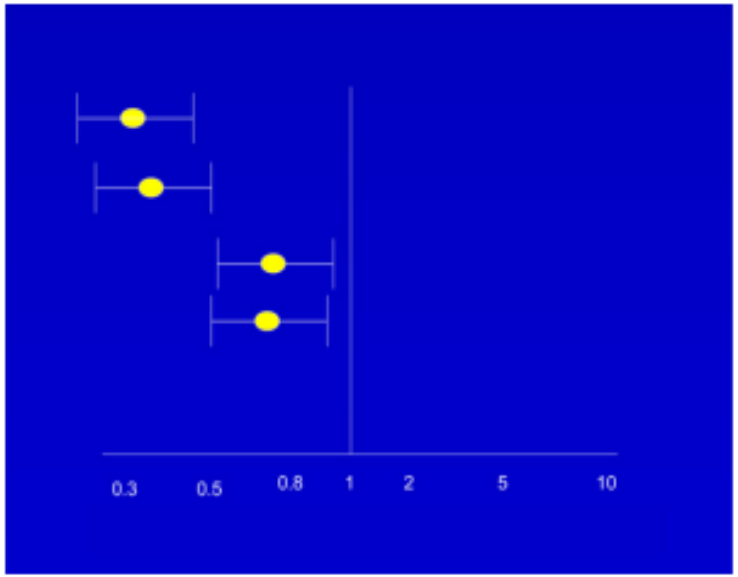


Fig. 2. Substantial heterogeneity, but of questionable importance.

Substantial Heterogeneity, of Unequivocal Importance

- Magnitude of difference in treatment effects identical to previous example
- But previous example showed same direction
- Here you have mix of benefit & harm→ grade ⬇

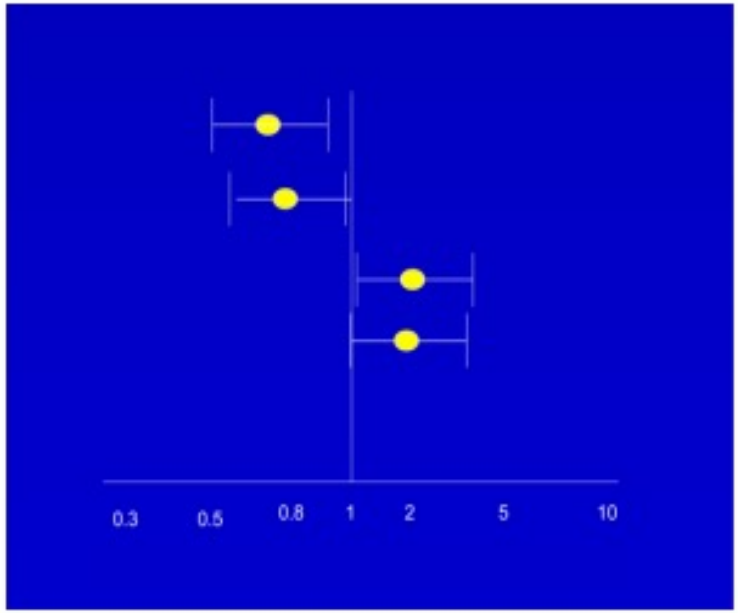


Fig. 3. Substantial heterogeneity, of unequivocal importance.

Inconsistency Leads to Decreased Quality of Evidence

- Rate down for inconsistency, not up for consistency
- Search for alternative explanation
  - If observed inconsistency cannot be explained, assess seriousness of inconsistency and why the authors may have made the conclusion
- Determine if subgroup analysis is appropriate

Potential Causes of Inconsistency

- Population Subgroups
  - Larger relative effects in sicker populations
  - Treatment on different diagnosis: cancer vs CHF
  - Age
- Interventions
  - Larger effects with higher drug doses
- Outcomes
  - Duration of follow-up
- Study Methods
  - Intervention vs observation
  - Randomized clinical trials with higher and lower risk of bias

Rate Down Quality of Evidence

- Point estimates vary widely across studies
- Confidence intervals show minimal/no overlap
- Statistical test for heterogeneity
  - Large I^2 –variation in point estimates due to true variation among study differences
  - Low p-value

Assessing Heterogeneity

- Visually inspect
  - Tables describing heterogeneity
  - Differing baseline characteristics
  - Do the estimates line up, do they overlap?
- Review the statistical test for heterogeneity (I^2)

INDIRECTNESS

Why Do We Care about Indirectness?

- Evidence can be direct or indirect
- If indirect evidence, quality decreases when differences in…
  - Population
  - Intervention
  - Outcome measures, including surrogate outcomes
  - Comparison (A vs. placebo, B vs. placebo, not A vs. B)
  - Mechanism (A and B different, but same class of drugs)

Outcome: ALS drug

- Specify every important outcome of interest
- Use of substitute or surrogate endpoints
- Has the surrogate been validated, and/or repeatedly established in RCT?

HEALTH

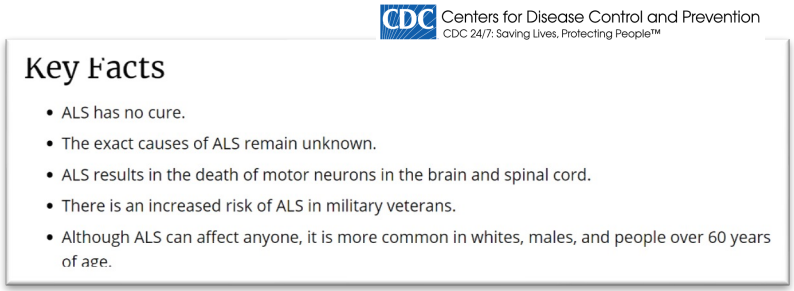
2 College Students Dreamed Up an A.L.S. Treatment. The Results Are In.

A study of their therapy and clinical trials of other experimental treatments are offering glimmers of hope that paralysis from the disorder can be slowed.

By Pam Belluck

The New York Times

PRINT EDITION Awash in Optimism | September 8, 2020, Page D1



CDC Centers for Disease Control and Prevention
CDC 24/7: Saving Lives, Protecting People™

Key Facts

- ALS has no cure.
- The exact causes of ALS remain unknown.
- ALS results in the death of motor neurons in the brain and spinal cord.
- There is an increased risk of ALS in military veterans.
- Although ALS can affect anyone, it is more common in whites, males, and people over 60 years of age.

ClinicalTrials.gov Identifier: NCT03127514

AMX0035 in ALS - Primary Outcomes

- Functional rating scale-revised (ALSFRS-R) slope change
- Number with adverse events
- Number in each group able to remain on study drug until planned discontinuation

[Time Frame: 24 weeks]

ClinicalTrials.gov Identifier: NCT03127514

AMX0035 in ALS – Secondary Outcomes

- Accurate testing of limb isometric strength (ATLIS) total score change
- Change in level of phosphorylated axonal neurofilament H subunit (pNF-H)
- Rate of decline in slow vital capacity (SVC)
- Death, tracheostomy, and hospitalization

[Time Frame: 24 weeks]

FDA Specialty Panel Concerns

The FDA's Position:

The primary endpoint in the open-label extension study (AMX0035-OLE) was safety. Open-label efficacy analyses are often difficult to interpret.

Additionally, we note that while the assessment of AMX0035 on a composite endpoint based on survival, hospitalization, and tracheostomies was listed in the hierarchy of efficacy endpoints in the OLE study protocol, death alone was not included in this list of endpoints. Analyses of the three components of the composite survival endpoint were planned, but the death analysis was not given priority over the other two components of the composite (or the composite itself).

As noted above, the Division does not agree with the inclusion of tracheostomy or hospitalizations in the definition of survival, as there is considerable variation in clinical practice as to when to hospitalize a patient or perform a tracheostomy due to differences in standard of care by treating physicians and patient preference; tracheostomies may also be placed for the management of secretions.

FDA also notes the Applicant's comment above that the OLE study completed on March 1, 2021. It was unclear why the March 1 date was chosen for study completion, as there were two patients still receiving treatment at that time that were terminated from the study by the Sponsor. All other patients had either died, discontinued from the study, or completed the prespecified 132 weeks of treatment (see Section 4.5).

Combined FDA and Applicant Briefing Document

NDA# 216660
Drug Name: AMX0035/ sodium phenylbutyrate (PB) and taurursodiol (TURSO)
Applicant: Amylyx Pharmaceuticals, Inc.

Peripheral and Central Nervous System Drugs Advisory Committee (PCNS) Meeting
March 30, 2022
Division of Neurology 1/Office of Neuroscience
Center for Drug Evaluation and Research

The New York Times

New A.L.S. Treatment Lacks Evidence of Benefit, F.D.A. Panel Finds

With a 6-4 vote, the group of independent advisers to the agency narrowly concluded that results from another clinical trial are needed to assess whether the therapy, called AMX0035, can help patients.

FDA Reviews Again – Approves

- “residual uncertainty about the evidence of effectiveness”
- “without any significant safety signals of concern”
- “given the serious and life-threatening nature of A.L.S. and the substantial unmet need, this level of uncertainty is acceptable in this instance”

https://www.accessdata.fda.gov/drugsatfda_docs/nda/2022/216660Orig1s000SumR.pdf

What GRADE Says

- *“In general, evidence based on surrogate outcomes should usually trigger rating down, whereas the other types of indirectness will require a more considered judgment.”*