How Clinical Trials Work

29th Annual SoCal Alzheimer's Disease Research Conference

Presented September 21, 2018

Daniel L. Gillen, PhD

Professor and Chair of Statistics,

Professor of Public Health & Epidemiology

Goals of Medical Research

1. Identify risk factors for disease
2. Identify treatments for disease
3. Identify strategies for prevention of disease
4. Basic science

Chronology of Scientific Investigation

- Anecdotal observations
  - Case report(s) leading to hypothesis
  - “That's not an experiment you have there, that's an experience.” - Sir Ronald A. Fisher (1890 - 1962)
  - “The plural of anecdote is not data.” - Roger Brinner
- Preclinical experiments
  - Laboratory, animal studies of mechanisms, toxicology

Chronology of Scientific Investigation

- Designed observational study
  - Case - control study
    - Sample diseased and non-diseased; examine exposures
    - Efficient for rare diseases; multiple risk factors
  - Cohort study (prospective or retrospective)
    - Sample exposed and nonexposed; examine disease
    - Efficient for common diseases; multiple diseases

Chronology of Scientific Investigation

- Designed interventional study
  - Clinical trial
    - Assign subjects to treatments
    - Prospectively examine outcomes
    - Can look at multiple responses (adverse effects)
    - Can infer cause and effect (when well designed)

Goals of Clinical Trials

- Experimentation in human volunteers
  - Investigate new treatment or preventive agent
    - Safety: Do adverse effects outweigh any benefit?
    - Efficacy: Does treatment beneficially alter disease?
    - Effectiveness: Would adoption benefit population?
  - Investigation of existing treatments
    - Relative benefits: Is one treatment clearly superior to another?
    - Harm: Should a therapy currently in use be removed from the market?

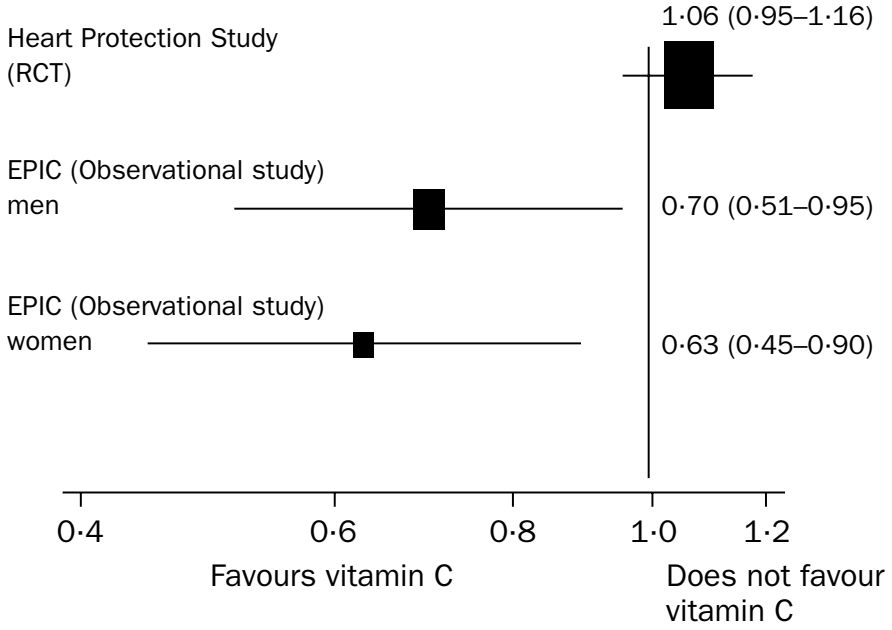
Why the need for randomized CTs?

Case study 1: Vitamin C and CHD

- Multiple observational studies have reported a protective effect of antioxidant vitamins on cardiovascular disease
- For example, Khaw et al (Lancet, 2001) estimated a 30% and 37% decrease in cardiovascular disease-specific mortality at 5 years in males and females, respectively. (The EPIC Study)
- The Heart Protection Study of antioxidant vitamin supplementation (Lancet, 2002) was a placebo-controlled randomized study to test this hypothesis

Why the need for randomized CTs?

Comparison of results between the Heart Protection Study and EPIC



Estimates of the effects of an increase of 15·7 μmol/L plasma vitamin C on coronary heart disease 5-year mortality

Estimates from men and women from a large observational epidemiological (EPIC) study[3] and from a large randomised controlled trial (Heart Protection Study).[7] In the EPIC study, results were adjusted for age, systolic blood pressure, cholesterol, body-mass index, smoking, diabetes, and vitamin supplement use.

Lawlor et al., The Lancet, 2004 (363) 1724-27

Why the need for randomized CTs?

- Why the difference? Most likely, unadjusted confounding by socioeconomic and lifestyle factors (not adjusted for in EPIC)

	Vitamin C quartile (range μmol/L)				Odds ratio (95% CI)*	p
	1 (0·00–20·46)	2 (20·47–39·27)	3 (39·28–59·64)	4 (59·65–190·47)		
Socioeconomic indicators						
Childhood						
Manual social class (%)	91·6 (89·3–93·4)	88·5 (86·0–90·7)	83·9 (81·0–86·5)	82·1 (79·2–84·7)	0·88 (0·84–0·92)	<0·0001
No bathroom in house (%)	45·1 (41·7–48·4)	40·7 (37·5–44·0)	33·7 (30·6–36·9)	33·4 (30·3–36·6)	0·92 (0·90–0·95)	<0·0001
No hot water in house (%)	39·5 (36·2–42·8)	37·5 (34·3–40·8)	31·7 (28·6–34·8)	29·9 (26·9–33·0)	0·93 (0·91–0·97)	<0·0001
Shared bedroom (%)	57·6 (54·2–60·9)	56·0 (52·6–59·3)	49·9 (46·6–53·2)	47·9 (44·6–51·3)	0·93 (0·90–0·97)	<0·0001
No car access (%)	86·5 (84·0–88·7)	84·7 (82·1–87·0)	79·9 (77·1–82·5)	78·9 (76·0–81·6)	0·91 (0·87–0·94)	<0·0001
Completed full-time education by age 18 years (%)	94·7 (93·0–96·0)	90·2 (88·0–92·1)	84·1 (81·5–86·4)	83·4 (80·8–85·7)	0·82 (0·78–0·86)	<0·0001
Adult						
Manual social class (%)	61·7 (58·2–65·0)	56·1 (52·6–59·4)	46·0 (42·6–49·3)	44·7 (41·4–48·0)	0·88 (0·85–0·91)	<0·0001
Local authority housing (%)	19·1 (16·7–21·8)	11·2 (9·3–13·4)	10·3 (8·5–12·5)	8·5 (6·9–10·5)	0·85 (0·81–0·88)	<0·0001
No car access (%)	35·8 (32·5–39·2)	27·3 (24·4–30·5)	22·6 (19·9–25·5)	22·6 (19·9–25·5)	0·90 (0·87–0·93)	<0·0001
State pension only (%)	36·7 (33·5–40·1)	28·7 (25·7–31·9)	22·3 (19·6–25·3)	24·4 (21·6–27·4)	0·89 (0·86–0·92)	<0·0001

Lawlor et al., The Lancet, 2004 (363) 1724-27

Why the need for randomized CTs?

Case study 2: Hormone replacement therapy

- Observational research on postmenopausal hormone therapy suggests a 40–50% reduction in coronary heart disease incidence among women using these preparations.
- Women's Health Initiative (WHI) performed a randomized, placebo controlled trial of daily use of 0.625 mg of conjugated equine estrogen (CEE) and 2.5 mg of medroxyprogesterone acetate (MPA) (E+P)
- WHI randomized trial stopped early by the DSMB partly due to increased risk of coronary heart disease and stroke in E+P arm

Why the need for randomized CTs?

Comparison of E+P results between the WHI RCT and the WHI observational study

TABLE 1. Cardiovascular disease incidence rates in the Women's Health Initiative clinical trial (1994–2002) and observational study (1994–2003) of estrogen plus progestin

	Clinical trial			Observational study		
	Placebo	Estrogen plus progestin	Ratio	Control	Estrogen plus progestin	Ratio
No. of women	8,102	8,506		35,551	17,503	
Coronary heart disease						
No. of events	147	188		615	158	
Annualized incidence (%)	0.33	0.39	1.18	0.32	0.16	0.50
Age-adjusted* annualized incidence (%)	0.33	0.40	1.21	0.28	0.20	0.71
Stroke						
No. of events	107	151		490	123	
Annualized incidence (%)	0.24	0.31	1.29	0.25	0.13	0.52
Age-adjusted* annualized incidence (%)	0.24	0.32	1.33	0.22	0.17	0.77
Venous thromboembolism						
No. of events	76	167		336	153	
Annualized incidence (%)	0.17	0.35	2.10	0.17	0.16	0.94
Age-adjusted* annualized incidence (%)	0.17	0.35	2.10	0.16	0.17	1.06

* Age adjusted to the 5-year age distribution in the clinical trial cohort.

Prentice et al., American Journal of Epidemiology, 2005 (462) 405-14.

Why the need for randomized CTs?

Why the difference? Many possible explanations have been postulated:

1. Unmeasured confounding – E+P users in observational study had somewhat better cardiovascular risk profiles compared to comparison group
2. Outcome ascertainment – WHI RCT ascertained silent MIs
3. Formulation and dose of E+P in RCT
4. Survival bias – Time-varying effect of RCT

Why the need for randomized CTs?

Survival bias would result if early deleterious effects of E+P "removed" patients from the study sample

TABLE 3. Age-adjusted* lifetime hormone therapy use history, expressed as percentage at baseline, in the Women's Health Initiative clinical trial (1994–2002) and observational study (1994–2003)

	Clinical trial		Observational study	
	Placebo	Estrogen plus progestin	Control	Estrogen plus progestin
Unopposed estrogen duration, years				
Never	89.4	89.4	91.4	86.6
<2	5.1	5.1	4.3	3.4
2–5	3.0	2.8	2.2	3.4
>5	2.5	2.7	2.1	6.7
Recency of estrogen-alone use				
Never user	89.4	89.4	91.4	86.6
Current user	0.6	0.6	0.0	0.0
Within past 1–4 years	1.8	1.8	1.9	3.9
Past use, 5–10 years ago	1.9	1.7	1.6	4.1
Past use, >10 years ago	6.3	6.6	5.1	5.4
Estrogen-plus-progestin duration, years				
Never	82.7	82.2	86.0	0.0
<2	7.4	7.2	6.6	14.7
2–5	5.0	5.1	3.4	23.2
>5	5.0	5.5	4.0	62.1

Prentice et al., American Journal of Epidemiology, 2005 (462) 405-14.

Why the need for randomized CTs?

But is there an early deleterious effect of E+P on cardiovascular risk?

TABLE 5. Estrogen-plus-progestin hazard ratios as a function of years from initiation of the current episode of estrogen-plus-progestin use in the Women's Health Initiative clinical trial (1994–2002) and observational study (1994–2003)

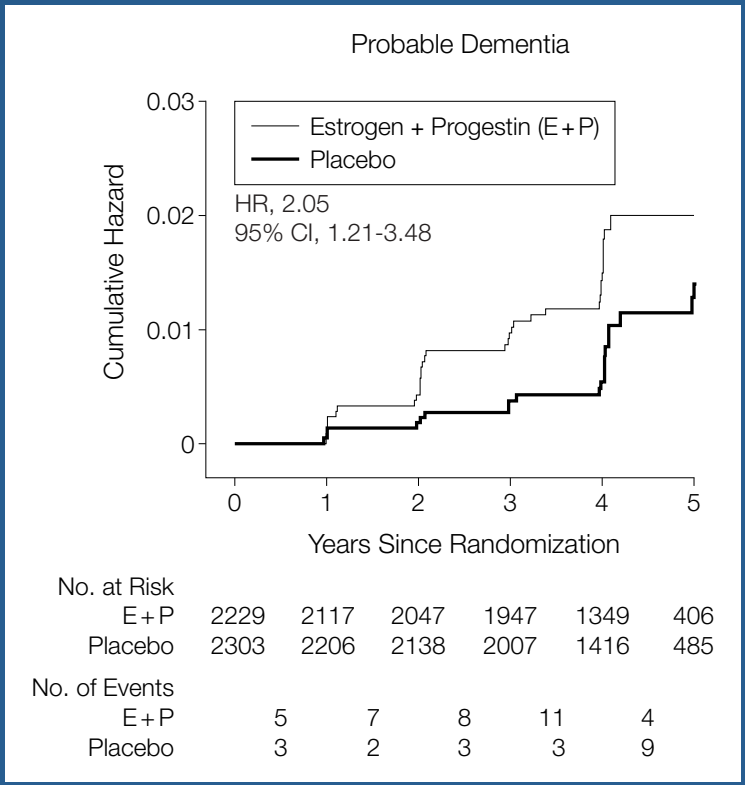
Time from estrogen-plus-progestin initiation* (years)	Clinical trial			Observational study		
	Cases in estrogen-plus-progestin group (no.)	Hazard ratio	95% confidence interval	Cases in estrogen-plus-progestin group (no.)	Hazard ratio	95% confidence interval
Coronary heart disease						
<2	80	1.68	1.15, 2.45	5	1.12	0.46, 2.74
2–5	80	1.25	0.87, 1.79	27	1.05	0.70, 1.58
>5	28	0.66	0.36, 1.21	126	0.83	0.67, 1.01
Stroke						
<2	43	1.15	0.71, 1.87	7	2.10	0.96, 4.56
2–5	79	1.49	1.02, 2.17	12	0.48	0.24, 0.93
>5	29	0.74	0.39, 1.39	104	0.89	0.71, 1.18
Venous thromboembolism						
<2	73	3.10	1.85, 5.19	7	2.37	1.08, 5.19
2–5	72	1.89	1.24, 2.88	27	1.52	1.01, 2.29
>5	22	1.31	0.64, 2.67	119	1.24	0.99, 1.55

Prentice et al., American Journal of Epidemiology, 2005 (462) 405-14.

Why the need for randomized CTs?

More appropriate for today's meeting, similar results were always observed in the WHI Dementia substudy

- Observed 2-fold increase in the risk of probable dementia (95% CI: 1.21-3.48)
- Previous observational studies reported 30-35% reductions



Shumaker et al., JAMA, 2003 (289) No. 20, 2651-62.

Take-Home Message

- Association versus Causation
  - Truly determining causation requires a suitable interventional study (experiment)
    - Comparisons tell us about associations
    - Associations in the presence of an appropriate experimental design allows us to infer causation
    - But even then, we need to be circumspect in identifying the true mechanistic cause
      - E.g., a treatment that causes headaches, and therefore aspirin use, may result in lower heart attack rates due entirely to the use of aspirin

Levels of Evidence

- U.S. Preventive Services Task Force
  - Level I: At least one properly designed RCT
  - Level II
    - II-1: Well-designed, nonrandomized CT
    - II-2: Well-designed, multicenter cohort/case-control
    - II-3: Multiple time series with/without intervention; Dramatic results from uncontrolled trial
  - Level III: Opinions of respected authorities

    = *Eminence based* (not their wording!)

Brief History of Legal Requirements

- Wiley Act (1906) – Labeling
- Food, Drug, and Cosmetics Act of 1938 – Safety
- Kefauver – Harris Amendment (1962) – Efficacy / effectiveness
  - " [If] there is a lack of substantial evidence that the drug will have the effect ... shall issue an order refusing to approve the application. "
  - "...The term 'substantial evidence' means evidence consisting of adequate and well-controlled investigations, including clinical investigations, by experts qualified by scientific training"
- FDA Amendments Act (2007)

Phases of Clinical Trials

- A series of studies used to support adoption of a new standard of treatment
  - Phase I: Initial safety / dose finding
  - Phase II: Preliminary efficacy / further safety
  - Phase III:
    - Therapeutics: Establish effectiveness
    - Prevention: Establish efficacy
  - Phase IV:
    - Therapeutics: Post-marketing surveillance
    - Prevention: Effectiveness

Phase I Clinical Trials

- Initial safety / dose finding in humans
  - Goals:
    - Pharmacokinetics / pharmacodynamics
    - Incidence of major adverse effects
    - Decide whether it is ethical to continue testing in humans
  - Methods
    - Relatively small number of participants
      - Participants often not true target population
    - Sometimes dose escalation
    - Sometimes no comparison group

Phase II Clinical Trials

- Preliminary evidence of efficacy
  - Goals:
    - Screening trial to look for any evidence of treatment efficacy
    - Incidence of major adverse effects
    - Decide if worth studying in larger samples
  - Methods
    - Relatively small number of participants
      - Participants closer to true target population
    - Outcome often a surrogate
    - Sometimes no comparison group

Phase II Clinical Trials

- ANAVEX®2-73 Phase IIa Study:
  - 57-week Phase 2a Alzheimer's disease study
  - N=32 participants with mild to moderate AD
  - Single arm
  - Primary endpoint: No. of participants with treatment related adverse events
  - Secondary endpoints: MMSE, ADCS-ADL, HAM-D
  - Subgroup Findings:
    - Improvements in MMSE and ADCS-ADL after exclusion of SIGMAR1 and COMT variants

Phase II Clinical Trials

- BAN2401 Study 201 Phase II Study
  - Study inclusion:
    - Early AD: MCI due to AD or mild AD
    - Amyloid pathology
    - MMSE range: 22-30
    - CDR global range: 0.5 (MCI); 0.5-1.0 (mAD)
  - Six arms (Placebo, 5 dose/regimen arms)
  - N=856 patients randomized
  - Endpoints assessed over 18 months (Amyloid beta, ADCOMS, ADAS-Cog, CDR-SB)

Phase III Clinical Trials

- Establishment of efficacy / effectiveness
  - Goals:
    - Obtain measure of treatment's efficacy on disease process
    - Incidence of major adverse effects
    - Therapeutic index
    - Modify clinical practice (obtain regulatory approval)
  - Methods
    - Relatively large number of participants from true target population (almost)
    - Clinically relevant outcome

Phase IV Clinical Trials

- Therapeutic: Post-marketing surveillance
  - Goals:
    - Monitor for rare serious events
    - (Some "Phase IV" trials are of more interest for marketing than for science)
- Prevention: Effectiveness

Ethics and Oversight

- Mechanisms for ensuring ethical treatment of study subjects
  - Before starting the study:
    - Institutional review board (IRB)
    - Protocol Review Committee (PRC)
  - During conduct of the study:
    - Data safety monitoring board (DSMB) or Independent Data Monitoring Committee (IDMC)
  - After studies completed:
    - Regulatory agencies (e.g., FDA)

Essential Elements of Trial Design

- A clinical trial must answer a meaningful question and be able to discriminate between viable hypotheses.
- At minimum, this requires critical thought regarding specification of the:
    - Study population
    - Outcome of interest
    - Treatment regimen
    - Treatment allocation
    - Outcome ascertainment
    - Methods for comparing arms

Inclusion/Exclusion Criteria

- Precise definition of target patient population is crucial
  - Scientific:
    - Materials and methods of scientific experiment
    - Increased homogeneity to decrease variation in response
    - Patients most likely to yield largest response
  - Clinical:
    - Generalization of safety outcomes
    - Generalization of efficacy outcomes

Inclusion/Exclusion Criteria

- Ideal setting
  - The study sample should look like a random sample from the subpopulation of all diseased patients who would ultimately be judged suitable for the intervention.
    - Negligible impact of restrictions due to clinical trial procedures
    - Negligible impact of restrictions due to locale of clinical trial
    - High participation rate by eligible patients

Inclusion/Exclusion Criteria

- Inclusion criteria: definition of ultimate target population
  - Objective criteria of disease
  - Disease severity
  - Contraindications of treatment

- Exclusion criteria: exceptions required for clinical trial setting
  - Contraindications of treatment (or control)
  - Requirements for evaluation of outcome
  - Requirements for compliance (generalizability?)
  - Inability/unwillingness to provide informed consent

Choice of Outcome

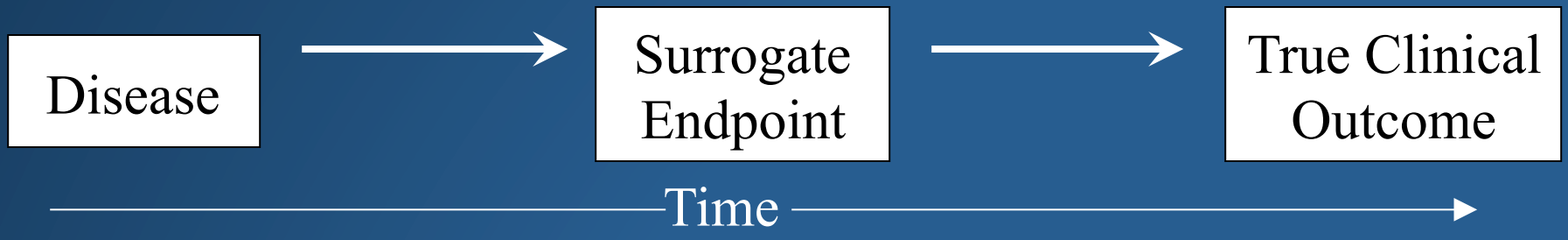
- Generally, subjects participating in a clinical trial are hoping that they will benefit in some way from the trial
- Clinical endpoints are therefore of more interest than purely biological endpoints (multiple?)
- Consider (in order)
  - The most relevant clinical endpoint (survival, quality of life, functioning)
  - The endpoint the treatment is most likely to affect
  - The endpoint that can be assessed most accurately and precisely

Choice of Outcome

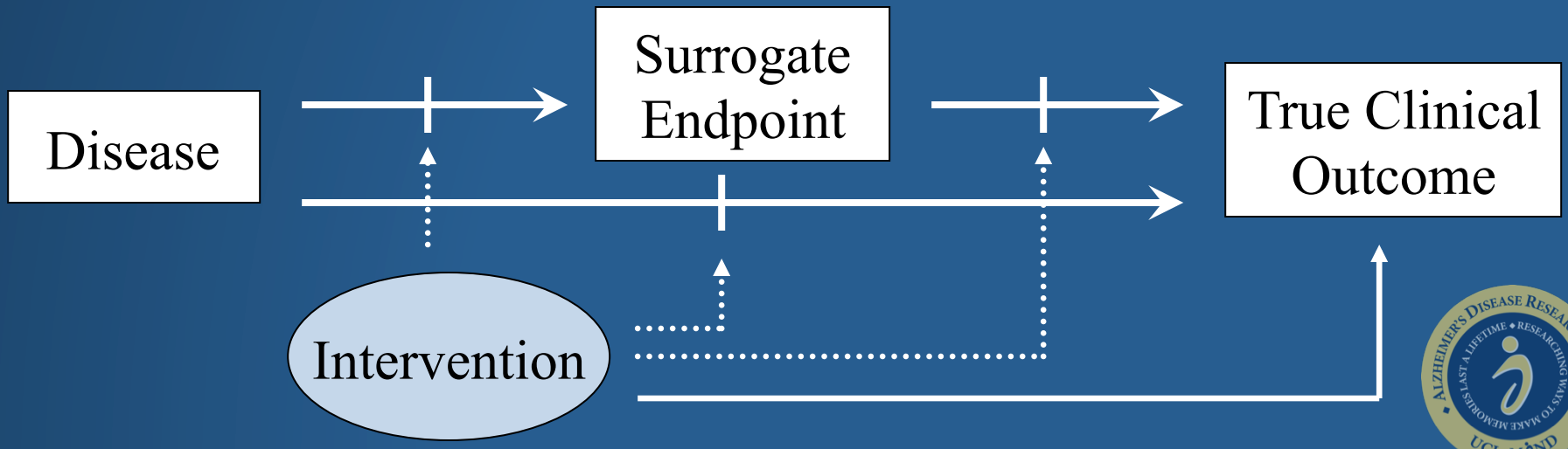
- Problem:
  - Relevant clinical outcomes are often relatively rare events that occur after a significant delay
    - Believe that earlier interventions have greater chance of benefit
  - Difficulty in measuring clinical outcome
    - Quality of life needs to be assessed over a sufficiently long period of time
  - Motivates the use of ”surrogate” biological endpoints that can be measured in a shorter timeframe

Choice of Outcome

- Ideal surrogate



- Need to worry about alternate pathways



Comparison Group (Common Choices)

1. No comparison group (single arm study)
   - Appropriate when an absolute criterion for treatment effect exists (rare)
2. Historical controls
   - Absolute criterion derived from historical trials
   - Sample from historical clinical trial (better)
   - Potential for bias due to confounding
3. Internal controls
   - Cross-over design (patients serve as his/her control)
   - Not always logistically possible

Comparison Group (Common Choices)

4. Concurrent control groups
   - Two or more treatment arms
   - Placebo or standard therapy
   - Active treatments
   - Multiple levels of same treatment

Most common to have concurrent control groups due to the difficulty in ensuring valid conclusions when using no comparison group or historical controls

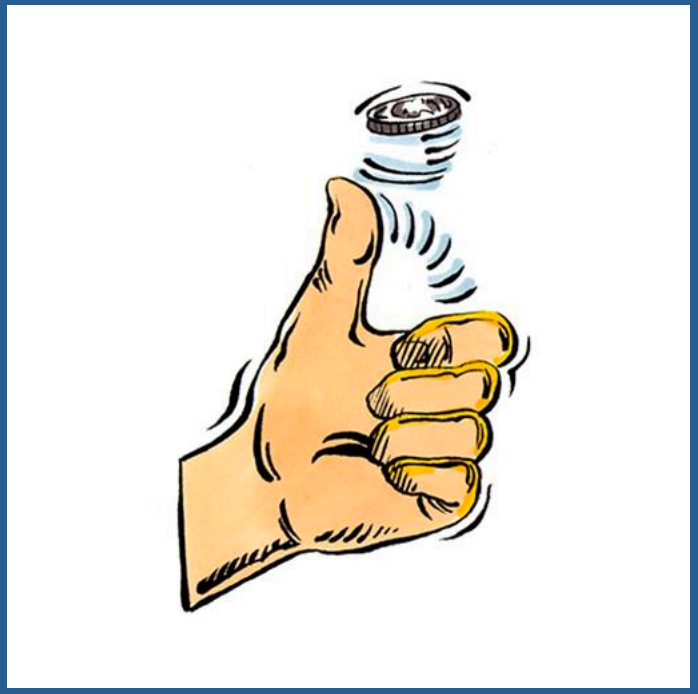
How to Assign Treatments?

- ICH guidelines (www.ich.org) part E9 Statistical Principles
  - "The most important design techniques for avoiding bias in clinical trials are *blinding* and *randomisation*, and these should be normal features of most controlled clinical trials intended to be included in a marketing application."
  - Here, bias is a tendency of a statistical estimate to deviate in one direction from a "true value"
- Randomization should be concealed and should occur just prior to initiation of treatment (whenever possible)

Forms of Randomization

Completely Randomized Design

- Assign each new patient to an arm with a constant probability
- Eg. 1:1 randomization would be a simple coin flip
- Easy to implement
- Can lead to imbalance in important factors by chance



Forms of Randomization

- Consider differences in baseline stoke severity in a multi-center randomized clinical trial comparing tissue plasminogen activator (tPA) for the treatment of acute ischemic stroke (Marler et al., *Neurology*, 2000)

Baseline NIHSS score	*tPA-treated patients, % (n = 153)*	*Patients given placebo, % (n = 167)*
0-5	19.0	4.2
6-10	24.2	27.5
11-15	17.0	21.0
16-20	21.6	19.8
>20	18.3	27.5

tPA = tissue plasminogen activator

- "The marked imbalance in baseline stroke severity in the 91 to 180-minute groups of the NINDS trial suggests that the NINDS trial lacks internal validity." -Mann, West J. of Med (2002)

Forms of Randomization

Stratified Randomization

- A priori define strata that are important predictors of the outcome
  - Disease severity, age, region of the world
- Within each strata, assign each new patient to an arm with a constant probability
  - Assures balance in important prognostic variables

Blocked Randomization

- Ensures that after ever K patients are enrolled, the number of patients allocated to each arm is fixed
- Important in smaller trials and when there is a "learning curve" associated with treatment/control

Forms of Randomization

Adaptive Randomization

- Can be *baseline adaptive* or *response adaptive*
- Baseline adaptive
  - Modify chance of being randomized to each arm to ensure balance across groups with respect to a summary of prognostic factors measured prior to randomization (eg. average age plus average disease severity score)
- Response adaptive
  - Modify chance of being randomized to each arm depending upon outcomes observed in each arm at the time of randomization
  - Can lead to severely imbalance arms and bias in some cases

Forms of Randomization

- BAN2401 Study 201 Phase II Study
  - Six arms (Placebo, 5 dose/regimen arms)
  - Response adaptive randomization
    - Periodically analyze data and randomize to most promising dose
  - In end, most patients randomized to second most efficacious dose? -> EMA restriction on APOE4 in highest dose group
    - Broken randomization?
    - Differential APOE4 prevalence in highest dose groups

Blinding

- Types of blinding
  - Single blind experiments : Participant is unaware of treatment assignment
  - Double blind experiments : Neither the participant nor treatment provider know treatment assignment
  - Triple blind experiments : Monitoring committee also blinded (can also refer to blinded adjudication)

"Blinding seeks to prevent ascertainment bias, protects the sequence after allocation, and cannot always be implemented."

(Schultz, JAMA; 1995; 274(18)1456:1458)

Blinding

- Methylene blue (LMTM) phase III anti-tau trial
  - Double-blind, "placebo"-controlled trial
  - N=891 individuals with probable AD
  - Three arms
    - LMTM at 150 mg daily
    - 250 mg daily,
    - placebo consisting of 8 mg LMTM daily
  - LMTM known to result in blue urine so the placebo incorporated low dose LMTM to maintain the blind

Does Blinding Really Matter?

- Noseworthy (1994). Neurology 1994;44:16-20.
  - All patients examined and response judged by both a blinded and unblinded neurologist.

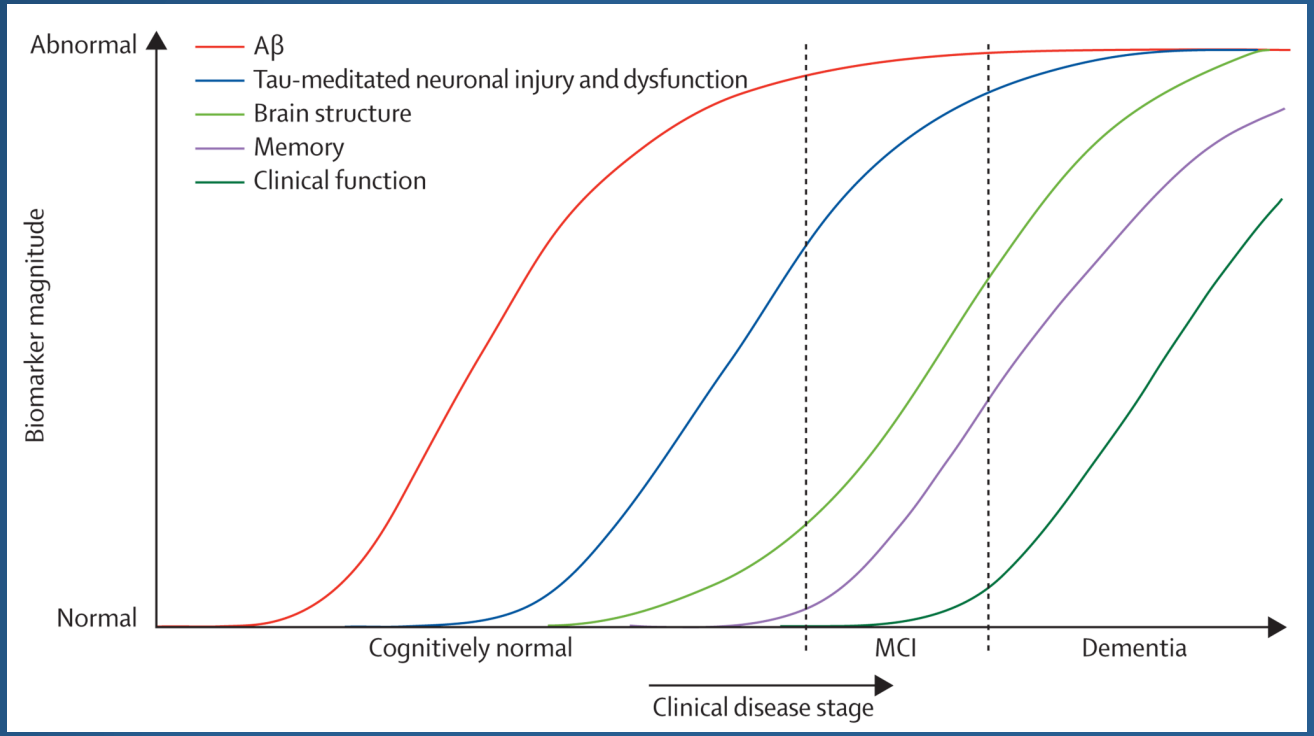
Table 1. *p* Value* of between-treatment comparison of proportion of subjects improved, stable, or worse

Assessment (no. pts)	IV cyclo versus placebo		PLEX versus placebo	
	Blinded	Unblinded	Blinded	Unblinded
6 Months (165)	0.159	0.069	0.246	0.047
12 Months (144)	0.295	0.084	0.086	0.004
18 Months (108)	0.418	0.255	0.106	0.072
24 Months (91)	0.088†	0.152	0.201	0.031
Final (mean, 30.4 months; 165)	0.290†	0.490	0.990	0.590

* Derived from chi-square test of the 2 (treatment) × 3 (improved, stable, worse) frequency table at each assessment point.
† Trend favoring placebo; all other comparisons favor active therapy.
IV cyclo Intravenous cyclophosphamide group (group I).
PLEX Plasma exchange group (group II).

Special Challenges in Early AD Trials

- Target population - Inclusion/exclusion criteria
- Outcome definition
- Outcome assessment
- Trial duration



Jack, Jr et al, Lancet Neurol. 2010 9(1): 119-139

FDA Draft Guidance for Early AD Trials

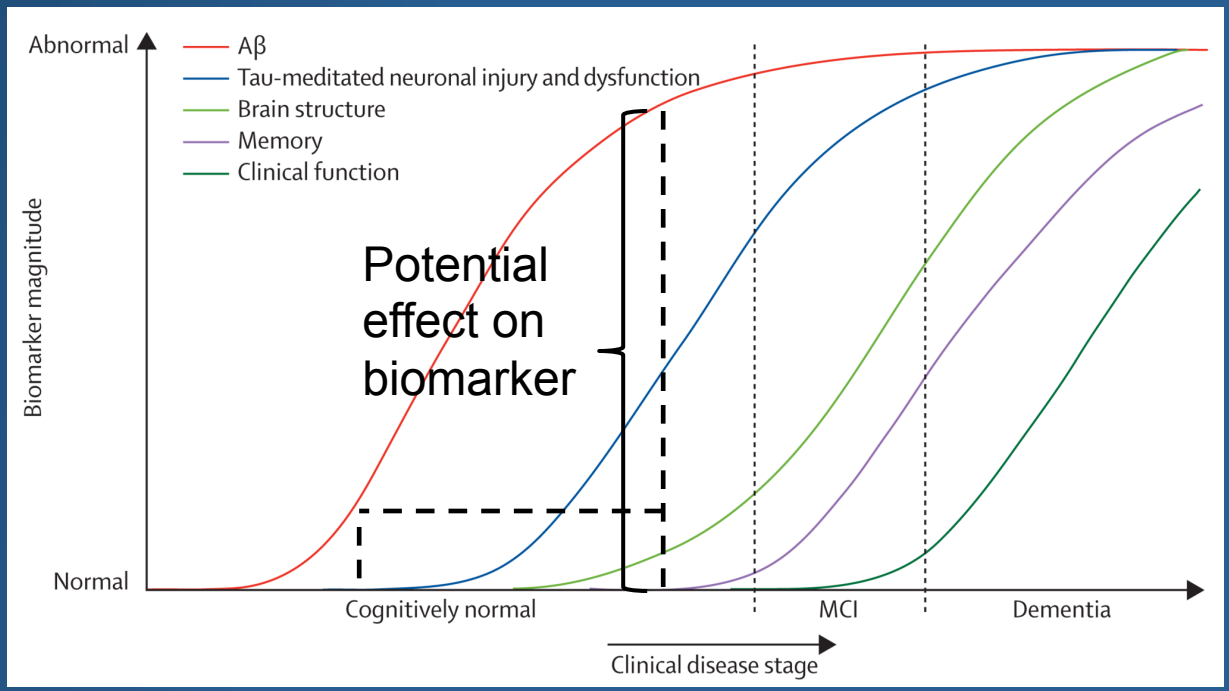
Basic approach: Start with the target population then consider relevant outcomes

- Stage 1: Characteristic pathophysiologic changes but no evidence of clinical impact
- Stage 2: Characteristic pathophysiologic changes **AND** subtle detectable abnormalities on sensitive neuropsychological measures; **NO** functional impairment
- Stage 3: Characteristic pathophysiologic changes **AND** subtle (or more) detectable abnormalities on sensitive neuropsychological measures **AND** mild but detectable functional impairment
- Stage 4: Overt dementia

FDA Draft Guidance for Early AD Trials

Basic approach: Guidance for outcomes guided by logistical constraints on trial duration

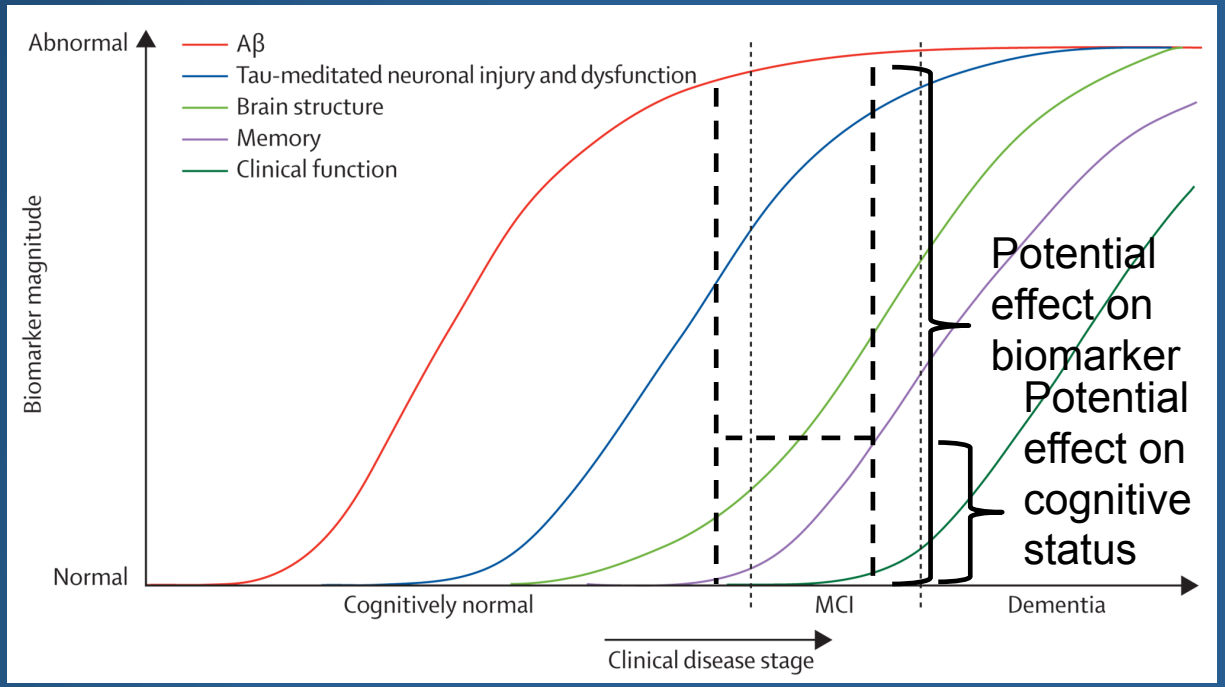
Stage 1: Effect on various biomarkers, analyzed as a primary efficacy measure, may, in principle, serve as the basis for an accelerated approval



Base figure: Jack, Jr et al, Lancet Neurol. 2010 9(1): 119-139

FDA Draft Guidance for Early AD Trials

Stage 2: FDA will consider strongly justified arguments that a persuasive effect on sensitive measures of neuropsychological performance may provide adequate support for a marketing approval (no effect on functional changes)



Base figure: Jack, Jr et al, Lancet Neurol. 2010 9(1): 119-139

FDA Draft Guidance for Early AD Trials

Basic approach: Guidance for outcomes guided by logistical constraints on trial duration

Stage 1: Effect on various biomarkers, analyzed as a primary efficacy measure, may, in principle, serve as the basis for an accelerated approval

Stage 2: FDA will consider strongly justified arguments that a persuasive effect on sensitive measures of neuropsychological performance may provide adequate support for a marketing approval (no effect on functional changes)

Stage 3: An integrated scale that adequately and meaningfully assesses both daily function and cognitive effects in early AD OR independent assessment of daily function and cognitive effects.

Summary

- Clinical trials represent the gold standard for establishing cause and effect in interventional therapies
- Need to focus on answering a well-defined clinically relevant question
  - Target population
  - Outcome of interest
  - Comparator
- Design elements can reduce (eliminate) bias
  - Randomization
  - Blinding
- Recent guidelines in AD recognize the need for tailored responses depending upon the target population
  - But we still need to be cautious of surrogate endpoints…