CHL 5225 H
Advanced Statistical Methods for Clinical Trials

Two sources for course material

1. Electronic blackboard
   required readings

2. www.andywillan.com/CHL5225H
   code of conduct
   course outline
   schedule
   class notes
   assignment
   grading and evaluation policies
   contact information

CHL 5225 H The Language of Clinical Trials

USEFUL TERMINOLOGY

Treatment Arm (Group)
- the treatment on which the patient is being evaluated

Intervention Arm (Treatment)
- the new or experimental treatment

Control Arm (Standard)
- the standard or current treatment, sometimes no treatment

Placebo
- sham new or experimental treatment

CHL 5225 H The Language of Clinical Trials

USEFUL TERMINOLOGY

Multi-centre trial

- patients recruited from more than one clinical site

Multinational trial

- patients recruited from more than one country

Masking (Blinding)

- single-blind: patients are masked to treatment arm
- double-blind: patients and person evaluating the patients are masked
- triple-blind: patients, evaluators and clinicians masked
- often evaluators are the clinicians, in which case double-blind = triple-blind

CHL 5225 H The Language of Clinical Trials

PHASES

Phase I – Dose finding studies

Usually done on healthy individuals

Often contracted out to Contract Research Organizations (CROs)

O'Quiqley et al. *Stat in Med* 1991 **10**:1647-1664.
Biometrics 1990 **46**: 33-48.
Biometrics 1996 **52**: 673-684.

Storer. *Biometrics* 1989 **45**: 925-937.

Piantadosi and Liu. *Stat in Med* 1996 **15**:1605-1618.

CHL 5225 H The Language of Clinical Trials

PHASES

Phase II – Safety and efficacy studies

- Usually done as a single arm
- Common in oncology

Herson (1984). Chapter 15 in Buyse, Staquet, Sylvester. Cancer Clinical Trials. Oxford: Oxford University Press.

Simon. *Controlled Clinical Trials* 1989 **10**: 1-10.

Thall and Simon. *Controlled Clinical Trials* 1994 **15**: 463-481.

CHL 5225 H The Language of Clinical Trials

PHASES

Phase III – Comparative studies (Randomized Controlled Trials (RCTs)
Patients randomized between two or more treatment arms

Early Phase III – Efficacy (Explanatory) Trials

- Smaller sample size
- Often single country
- Smaller number of clinical sites
- Often placebo controlled (*i.e*. intervention compared to nothing)
- Selected patients (those most likely to benefit from new intervention)
- Sub-clinical outcomes (those most likely to be sensitive to change)
- Tight protocol control (compliance, follow-up)
- Short time horizon
- Positive results lead to late phase III trial
- Answers question “Can the new intervention work?” (biological question)

CHL 5225 H The Language of Clinical Trials

PHASES

Phase III – Comparative studies (Randomized Controlled Trials (RCTs)
Patients randomized to two or more arms

Late Phase III – Effectiveness (Pragmatic, Management) Trials

- Larger sample size
- Often multinational
- Large number of clinical sites
- New intervention usually compared to best alternative (standard care)
- Unselected patients
- Clinical outcomes (those most important to patients and clinicians)
- Loose protocol control
- Longer time horizon
- Positive results lead to adoption of new intervention
- Answers question “Will the new intervention work?” (Clinical question)

CHL 5225 H The Language of Clinical Trials

PHASES

Example: new drug from reducing high blood pressure

Early Phase III – Efficacy (Explanatory) Trials

- 200 or 300 patients, 1 or 2 clinical sites, one country
- Placebo controlled
- Otherwise health patients, not previously treated
- Demonstrate compliance in a “run-in”
- Measure blood pressure as primary outcome
- Monthly visits to clinical site
- Include in the analysis only those who took 80% of drug
- One year time horizon

CHL 5225 H The Language of Clinical Trials

PHASES

Example: new drug for controlling high blood pressure

Late Phase III – Effectiveness (Pragmatic, Management) Trials

- Large multinational trial with numerous clinical sites
- Intervention compared to best active comparator
- All patients with high blood pressure
- Use heart attacks, strokes and death as primary outcomes
- Might include quality of life (QofL) and economic (cost) outcomes
- Patients visit clinical site every 6 months
- Include all randomized patients in analysis
- Five to ten year time horizon

CHL 5225 H The Language of Clinical Trials

PHASES

Phase IV – Post-marketing (RCTs)

- Done by drug companies after drug has been approved
- Large sample size
- Sometimes multinational
- Usually large number of clinical sites
- Intervention compared to best available alternative
- Unselected patients
- Clinical and safety outcomes
- QoL and economic outcomes
- Loose protocol control
- Longer time horizon

CHL 5225 H The Language of Clinical Trials

DESIGNS OF PHASE III TRIALS

Parallel Groups

- Most common
- Patients individually randomized between two or more arms
- Simple or controlled randomization

Crossover

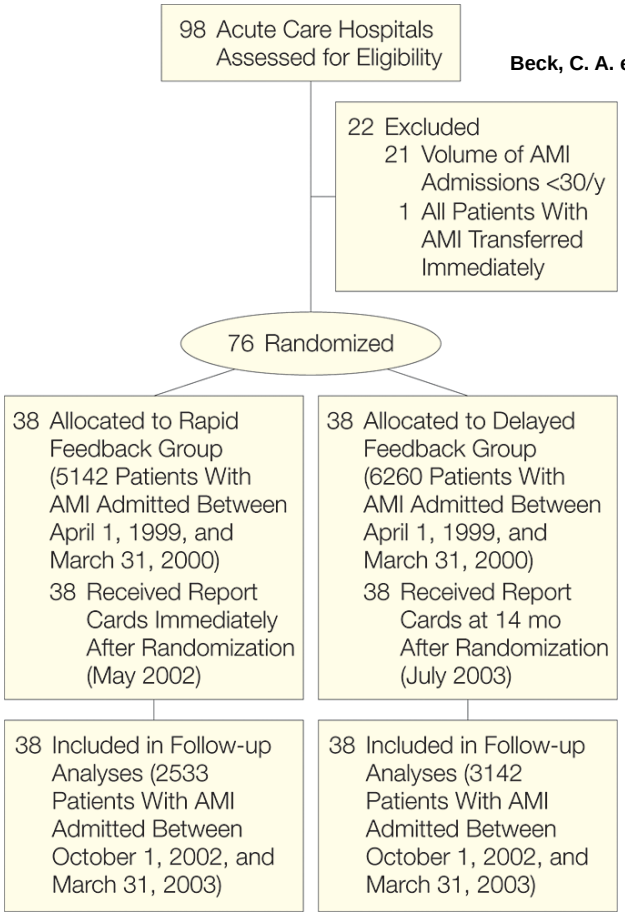
- Patients individually randomized to two or more sequences of treatments
- i.e. A then B or B then A
- Patients act as their own control
- Best for chronic diseases and “short” acting interventions
- Used mostly for early phase III

CHL 5225 H The Language of Clinical Trials

DESIGNS OF PHASE III TRIALS

Cluster Randomized

- Groups of patients are randomized between two or more arms
- Groups can be defined by (i) family practice, (2) town, (3) hospital
- Simple randomization
- Paired randomization
- Stratified randomization
- Usually late phase III



Beck, C. A. et al. JAMA 2005;294:309-317.

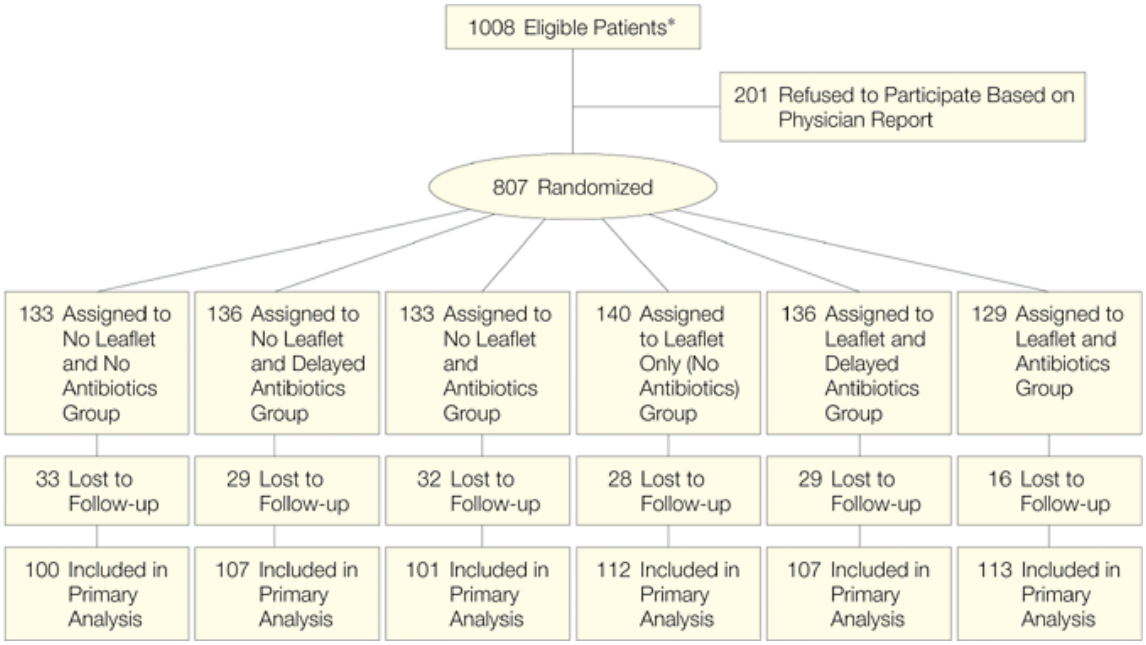
JAMA

CHL 5225 H The Language of Clinical Trials

DESIGNS OF PHASE III TRIALS

Factorial

- Patients randomized "twice": A vs $Placebo_A$ and B vs $Placebo_B$
- Four groups of patients A/B, A/$Placebo_B$, $Placebo_A$/B and $Placebo_A$/$Placebo_B$
- Two trials for the price of one
- Allows for the examination of the interaction between A and B
- Usually late phase III



Little, P. et al. JAMA 2005;293:3029-3035.

JAMA

CHL 5225 H The Language of Clinical Trials

METHODS OF RANDOMIZATION

Parallel Groups (patients enrolled sequentially)

- Simple randomization
  - each patient has same probability of being allocated to a particular arm
  - advantages: simplest analysis; concealment guaranteed, even for unmasked trials
  - disadvantages: clinician don't sleep well; imbalance with respect to important prognostic factors possible

CHL 5225 H The Language of Clinical Trials

METHODS OF RANDOMIZATION

Parallel Groups (patients enrolled sequentially)

- Blocked randomization (block size = b)
  - after k*b patients have been allocated there will be an equal number of patient allocated to each arm, k = 1, 2, . . .
  - b = (no. of arms)*(blocking factor), blocking factor = 1, 2, . . .
  - example: 2 arms, blocking factor = 2, block size = 4
  - possible blocks: (ABAB),(BABA),(AABB),(BBAA),(ABBA),(BAAB)
  - allocation list created by randomly sampling blocks with replacement
  - advantages: ensures the same number of patients on each arm
  - disadvantages: for unmasked trial concealment not guaranteed; clinician don't sleep well; imbalance with respect to important prognostic factors possible

CHL 5225 H The Language of Clinical Trials

METHODS OF RANDOMIZATION

Parallel Groups (patients enrolled sequentially)

- Stratified blocked randomization (block size = b)
  - patients stratified by important prognostic factor and a blocked allocation list is created for each stratum
  - example: suppose age is expected to affect outcome
  - create 3 strata: 18-25; 26-45; 46-65
  - each age stratum has their own allocation list
  - most multi-centre RCTs stratify randomization by clinical site
  - often use more than one block size randomly
  - advantages: ensures each each arm will have the same age distribution
  - disadvantages: for unmasked trial concealment not guaranteed; with several factors the number strata become too large; analysis is more complicated, *i.e*. should include stratification factor as covariate

CHL 5225 H The Language of Clinical Trials

METHODS OF RANDOMIZATION

Parallel Groups (patients enrolled sequentially)

- Adaptive randomization
  - next patient is allocated to the arm that minimizes the imbalance with respect to prognostic factors according to some loss function

Loss function : $L = w_{age}(\mu_A - \mu_B) + w_{sex}(\pi_A - \pi_B)$

where μ_i = average age on treatment arm i

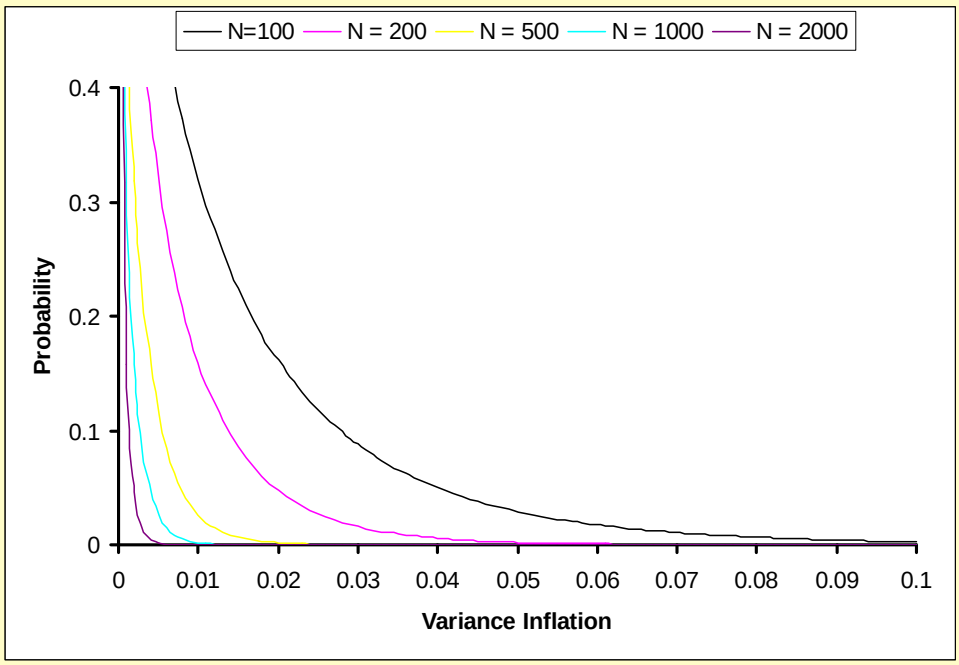
and π_i = proportion of males on treatment arm i

$L_A < L_B \Rightarrow$ allocate patient to A

$L_A > L_B \Rightarrow$ allocate patient to B

- advantages: handles large number of prognostic factors
- disadvantages: complicated; appropriate analysis not known

CHL 5225 H The Language of Clinical Trials



CHL 5225 H The Language of Clinical Trials

INTENT TO TREAT PRINCIPLE

Patients are analyzed in the arm to which they were randomized, regardless of what treatment they received and whether or not they were eligible to be randomized

This is the analysis that accurately assess the impact of adopting the intervention

Often used for “Effectiveness” trials

Meant to minimize “selection” bias and improve credibility

CHL 5225 H The Language of Clinical Trials

DRUG COMPANY FUDGE ON THE INTENT TO TREAT PRINCIPLE

Intent to Treat Population (Sample)
- All eligible patients

Safety Population (Sample)
- All eligible patients who had a least one dose of drug

Efficacy Population (Sample)
- All eligible patients who had 80% of the drug they were allocated to

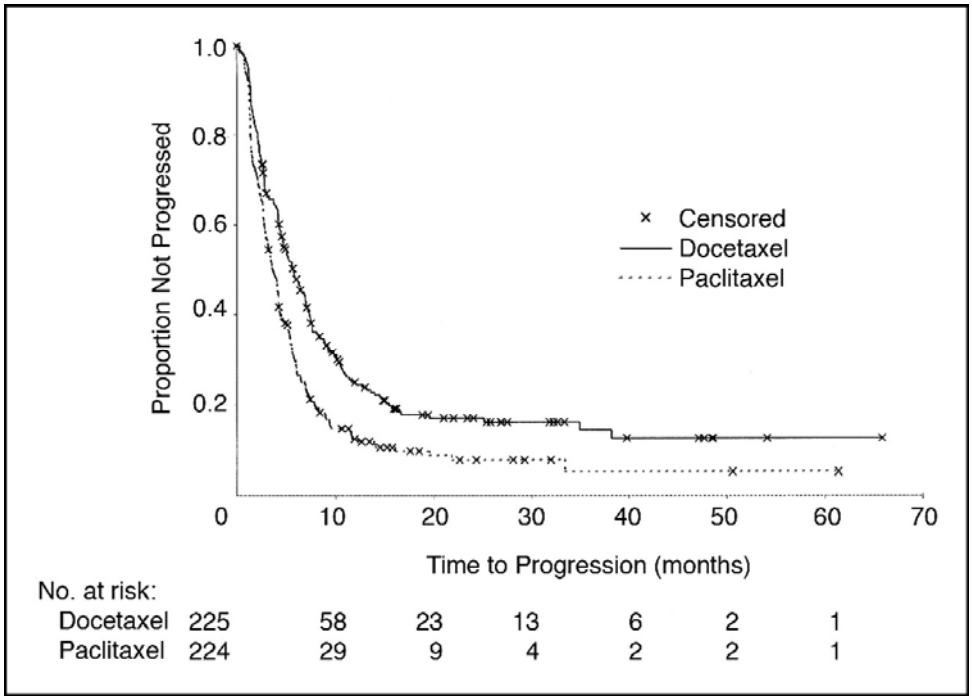
CHL 5225 H The Language of Clinical Trials

TYPES OF OUTCOME VARIABLES

Time to Event

- Very common
- Time to death, relapse, stroke, symptom relief, discharge from hospital
- Usually characterized by censoring
  - lost to follow-up
  - administrative (analysis done before everyone has an event)
- Survival curves to describe data
- Log-rank test to compare treatment arms
- Cox proportional hazards for covariate (prognostic factor) adjustment

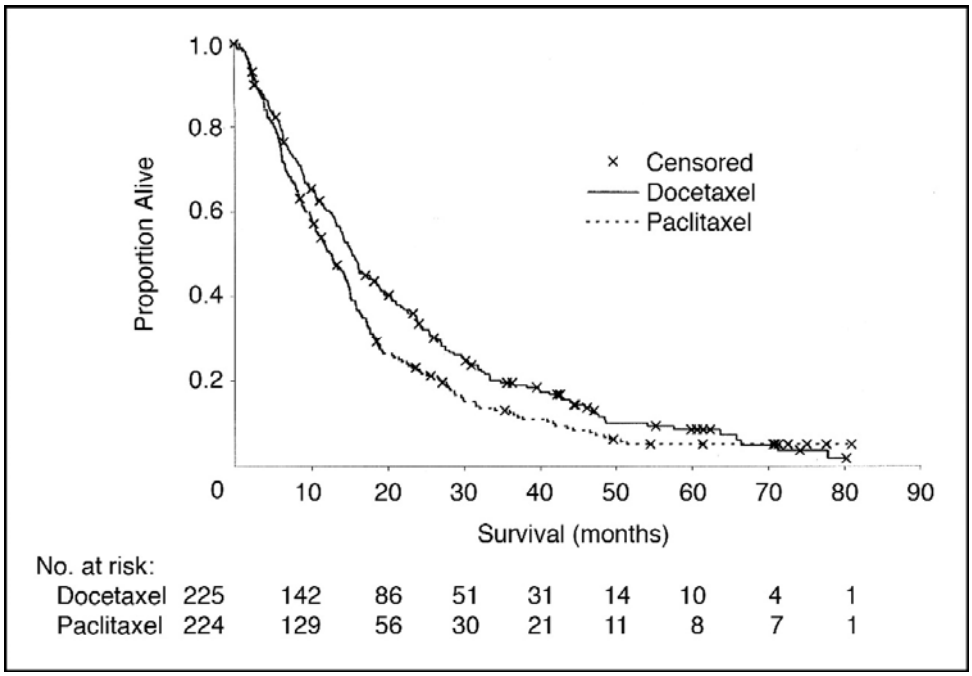
Fig 1. Kaplan-Meier plot of time to progression in the intent-to-treat population in each treatment group



Jones, S.E. et al. J Clin Oncol; 23:5542-5551 2005

JOURNAL OF CLINICAL ONCOLOGY

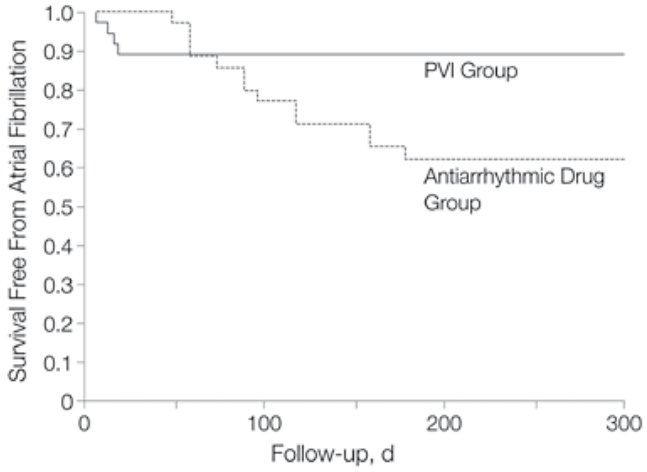
Fig 2. Kaplan-Meier plot of survival in the intent-to-treat population in each treatment group



Jones, S.E. et al. J Clin Oncol; 23:5542-5551 2005

JOURNAL OF CLINICAL ONCOLOGY

Kaplan-Meier Curve of Survival Free From Atrial Fibrillation



No. at Risk

PVI Group	32	28	28	28	28	28	28
Antiarrhythmic Drug Group	35	34	23	19	13	13	13

JAMA

CHL 5225 H The Language of Clinical Trials

TYPES OF OUTCOME VARIABLES

Time to Event

Treatment contrasts of interest

- ratio of median survival (from survival curves)
- hazard ratio (from Cox proportional hazards)
- difference in mean survival is more clinically relevant
  - must assume a parametric form for the survival distribution
  - not estimable from Cox proportional hazards
  - survival curves only provide restricted means, unadjusted for covariates
  - Inverse probability weighting provides restricted means while adjusting for covariates

CHL 5225 H The Language of Clinical Trials

TYPES OF OUTCOME VARIABLES

Time to Event—Difference in mean survival

1 / (difference in mean survival) = Number Needed to Treat (NNT)

NNT is the number of patients you need to treat on the new treatment (T), rather than the standard (S), to expect to realize one extra unit of effectiveness (in this case one year)

Suppose mean survival on S = 5 years

and mean survival on T = 5.25 years

NNT = 1 / 0.25 = 4

Total expected survival time for 4 patients on T = 4*5.25 = 21

Total expected survival time for 4 patients on S = 4*5 = 20

CHL 5225 H The Language of Clinical Trials

TYPES OF OUTCOME VARIABLES

Binary

- Very common
- Patient died, relapsed, had stroke, became free of symptoms (within a specific time period)
- Often a composite outcome (e.g. a patient would be considered a failure if he or she died OR had a stroke OR had a major bleeding event
- Often no censoring
- Results presented in 2 by 2 contingency table
- Fisher exact test or Chi-squared test to compare treatment arms
- Logistic regression for covariate (prognostic factor) adjustment
- Life table methods (survival tables) for censored data

One-Year Follow-up Results by Treatment Group

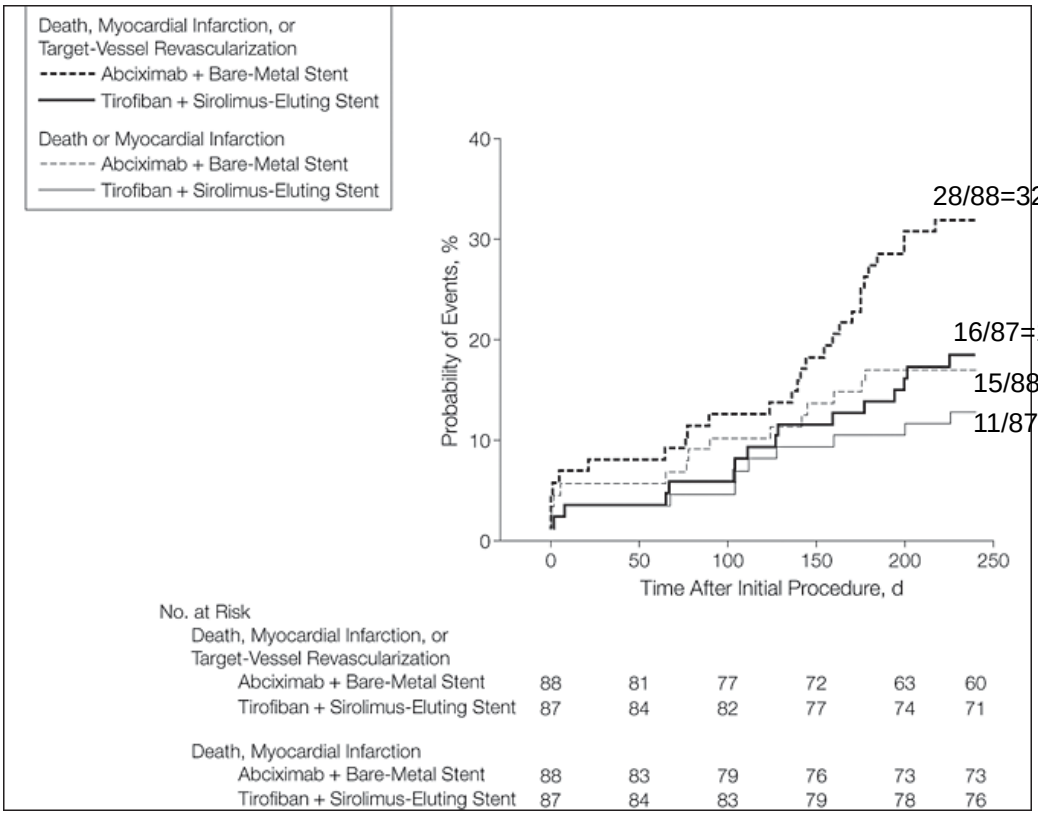
Table 2. One-Year Follow-up Results by Treatment Group

	No. (%) of Patients		
	Pulmonary Vein Isolation Group (n = 32)	Antiarrhythmic Drug Group (n = 35)	*P* Value
Symptomatic atrial fibrillation recurrence	4 (13)	22 (63)	<.001
Hospitalization	3 (9)	19 (54)	<.001
Thromboembolic events*	0	0	NA
Bleeding	2 (6.3)	1 (2.9)	.60
Bradycardia	0	3 (8.6)	.20
Pulmonary vein stenosis†			
Mild	1 (3)	0	.50
Moderate	1 (3)	0	.50
Severe	0	0	NA

Abbreviation: NA, not applicable.
*Defined as transient ischemic events, stroke, deep vein thrombosis, or pulmonary embolism.
†Mild pulmonary vein stenosis is defined as less than 50%; moderate, 50% to 70%; and severe, more than 70%.

Wazni, O. M. et al. JAMA 2005;293:2634-2640.

JAMA



Valgimigli, M. et al. JAMA 2005;293:2109-2117.

JAMA

CHL 5225 H The Language of Clinical Trials

TYPES OF OUTCOME VARIABLES

Binary—Parameters of Interest

Let p_T and p_S be the probability of the event occurring in a patient that is randomized to Treatment and Standard, respectively

Relative Risk (RR) = p_T / p_S

If RR = 0.8, then patients on T have 20% lower chance of experiencing event

Odds Ratio (OR) $= \dfrac{p_T/(1-p_T)}{p_S/(1-p_S)} = RR\dfrac{1-p_S}{1-p_T}$

If event is “rare”, say p_T and p_S <= 0.1, then $\dfrac{1-p_S}{1-p_T}$ is close to 1 and OR is close to RR

OR has better statistical properties

Logistic regression, used for covariate adjustment, yields ORs

CHL 5225 H The Language of Clinical Trials

TYPES OF OUTCOME VARIABLES

Binary—Parameters of Interest

Risk Difference (RD) = $p_S - p_T$

Number Needed to Treat (NNT) = 1 / RD

NNT = the number of patients that need to be treated with T rather than S to expect to prevent one event

If $p_S = 0.1$ and $p_T = 0.05$, then NNT = 1/0.05 = 20

Expected number of event from treating 20 patients with S = 0.1*20 = 2

Expected number of event from treating 20 patients with T = 0.05*20 = 1

CHL 5225 H The Language of Clinical Trials

TYPES OF OUTCOME VARIABLES

Continuous

- Less common
- blood pressure, symptom scores, quality of life (QoL)
- Usually characterized by repeated measurements
- Often characterized by missing values
- Repeated measures ANOVA to compare treatment arms
- GLM for for covariate (prognostic factor) adjustment
- Area under the curve (AUC) methods for QoL yields
  quality-adjusted survival time

Table 3. Posttreatment Scores and Group Effect From End-point Analyses With Pretreatment as a Covariate

Shear, K. et al. JAMA 2005;293:2601-2608.

	No. (%)				
	Complicated Grief Treatment	Interpersonal Psychotherapy	F	*df*	*P* Value
Modified Intention to Treat					
Inventory of Complicated Grief	n = 49	n = 46			
Pretreatment	45.8 (8.0)	44.2 (9.9)			
Posttreatment	28.6 (16.2)	31.4 (12.9)			
Difference	17.2 (15.3)	12.8 (11.9)	1.86	92	.18
Beck Anxiety Inventory	n = 47	n = 45			
Pretreatment	17.5 (12.0)	15.4 (10.2)			
Posttreatment	9.3 (10.7)	9.4 (8.7)			
Difference	8.2 (8.7)	6.0 (9.3)	0.69	89	.41
Beck Depression Inventory	n = 47	n = 45			
Pretreatment	23.9 (10.3)	22.4 (9.8)			
Posttreatment	13.4 (10.0)	15.2 (10.8)			
Difference	10.4 (9.6)	7.2 (7.2)	2.70	89	.10
Work and Social Adjustment Scale	n = 49	n = 46			
Pretreatment	20.3 (10.1)	20.5 (9.6)			
Posttreatment	12.5 (10.5)	16.2 (11.0)			
Difference	7.8 (11.3)	4.2 (9.5)	3.59	92	.06
Treatment Completers					
Inventory of Complicated Grief	n = 35	n = 34			
Pretreatment	46.4 (8.4)	43.4 (9.8)			
Posttreatment	25.8 (15.7)	30.6 (13.8)			
Difference	20.6 (15.0)	12.8 (10.7)	5.18	66	.03

JAMA

CHL 5225 H The Language of Clinical Trials

Continuous Outcome

ANOCOVA

$$y_{post,i} = \beta_0 + \beta_1 y_{pre,i} + \beta_T T_i + e_i$$

where T_i is dummy indicator for treatment
i.e. $T_i = I$(patient i on new treatment arm)

T-test on difference

$$y_{post,i} - y_{pre,i} = \beta_0 + \beta_T T_i + e_i$$

T-test on post-randomization observation

$$y_{post,i} = \beta_0 + \beta_T T_i + e_i$$

$$V\left(y_{post,i} - y_{pre,i}\right) = 2(1-\rho)V\left(y_{post,i}\right)$$

CHL 5225 H The Language of Clinical Trials

Patient follow-up–Compliance

"Drop-outs"

- referring to patients who do not complete allocated treatment
- often used only for intervention arm (drug company)
- poor use of terminology
- "non-completers" is a more accurate term

"Drop-ins"

- referring to patients allocated to control who receive intervention
- poor use of terminology
- "Crossovers" is a more accurate term

Much better to characterize the treatments patients receive using proper statistics

CHL 5225 H The Language of Clinical Trials

Patient follow-up–Compliance

Consider a trial with a 8 doses

Doses	3	4	5	6	7	8
Patients	200	180	164	150	140	100
Proportion	1	0.9	0.82	0.75	0.7	0.5

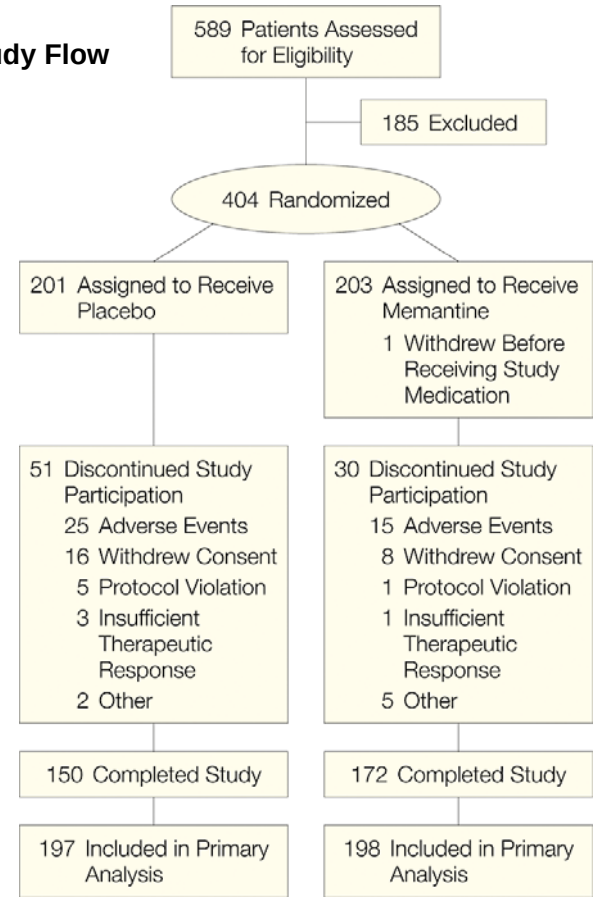
CHL 5225 H The Language of Clinical Trials

Journal Reporting of Randomized Clinical Trials

Example:

Tariot PN et al.
Memantine Treatment in Patients With Moderate to Severe Alzheimer Disease Already Receiving Donepezil
JAMA 2004; **291**:317-314

Study Flow



JAMA

Tariot, P. N. et al. JAMA 2004;291:317-324.

Table 1. Baseline Demographic and Clinical Characteristics*

Characteristics	Placebo (n = 201)	Memantine (n = 202)
Men	67 (33)	74 (37)
Women	134 (67)	128 (63)
Age, mean (SD), y	75.5 (8.73)	75.5 (8.45)
Weight, mean (SD), kg	66.4 (14.12)	70.7 (14.31)†
White race	186 (92.5)	182 (90.1)
MMSE score, mean (SD)	10.2 (2.98)	9.9 (3.13)
Duration of donepezil treatment, mean (SD), wk	129 (70.3)	126 (64.9)
Donepezil dose, mean (SD), mg	9.49 (1.88)	9.25 (1.79)
Any concurrent medical condition	149 (74.1)	149 (73.8)
Any concomitant medication during treatment	197 (98.0)	197 (97.5)
Tocopherol	120 (59.7)	131 (64.9)
Multivitamins	78 (38.8)	80 (39.6)
Acetylsalicylic acid	76 (37.8)	73 (36.1)
Ascorbic acid	35 (17.4)	43 (21.3)
Paracetamol	25 (12.4)	32 (15.8)
Ginkgo biloba	24 (11.9)	31 (15.3)
Calcium	21 (10.4)	25 (12.4)

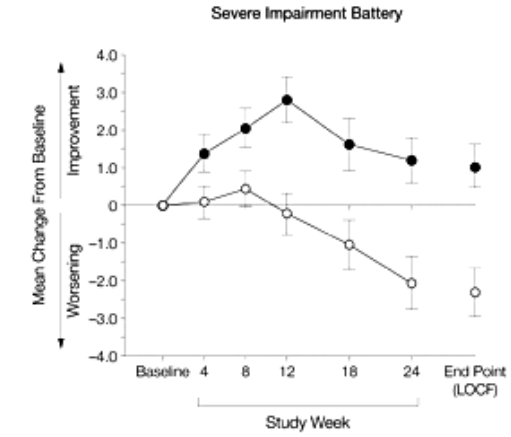
Abbreviation: MMSE, Mini-Mental State Examination.
*Data are No. (%) unless otherwise specified. One randomized patient discontinued the study prior to receiving any treatment and was not included in the analyses.
†$P = .003$.

Table 2. Efficacy Outcomes at Week 24 (Observed Case) and at End Point (LOCF)*

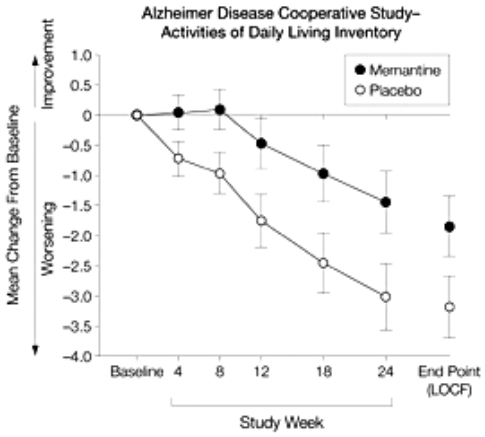
	Least Squares Mean Score (SE)							
			Change From Baseline					
	Baseline		End Point LOCF†			Week 24 Observed Case		
Outcome Measure	Placebo	Memantine	Placebo	Memantine	*P* Value	Placebo	Memantine	*P* Value
SIB	80.0 (1.13)	78.0 (1.11)	−2.5 (0.69)	0.9 (0.67)	<.001	−2.4 (0.74)	1.0 (0.70)	<.001
No. of patients	197	198	196	198		153	171	
ADCS-ADL$_{19}$	35.8 (0.74)	35.5 (0.73)	−3.4 (0.51)	−2.0 (0.50)	.03	−3.3 (0.55)	−1.7 (0.51)	.02
No. of patients	197	198	197	198		152	172	
CIBIC-Plus‡	NA	NA	4.66 (0.075)	4.41 (0.074)	.03	4.64 (0.087)	4.38 (0.081)	.03
No. of patients	197	198	196	198		152	172	
NPI	13.4 (1.08)	13.4 (1.07)	3.7 (0.99)	−0.1 (0.98)	.002	2.9 (1.06)	−0.5 (0.99)	.01
No. of patients	197	198	189	193		152	171	
BGP Care Dependency Subscale	9.8 (0.46)	9.5 (0.45)	2.3 (0.38)	0.8 (0.37)	.001	2.2 (0.40)	0.6 (0.37)	.001
No. of patients	196§	198	179	185		151	172	

Abbreviations: ADCS-ADL$_{19}$, 19-item Alzheimer Disease Cooperative Study–Activities of Daily Living Inventory; BGP, Behavioral Rating Scale for Geriatric Patients; CIBIC-Plus, Clinician's Interview-Based Impression of Change Plus Caregiver Input; LOCF, last observation carried forward; NA, not applicable; NPI, Neuropsychiatric Inventory; SIB, Severe Impairment Battery.
*SIB range of possible scores, 0 to 100; higher score indicates better function. ADCS-ADL$_{19}$ range of possible scores, 0 to 54; higher score indicates better function. CIBIC-Plus was defined as a change score, therefore baseline values are not applicable; range of possible scores, 1 (marked improvement) to 7 (marked worsening). NPI range of possible scores, 0 to 144; higher scores indicate worse symptoms. BGP range of possible scores, 0 to 70; higher scores indicate worse function.
†For the end point LOCF approach, only postbaseline assessments were carried forward.
‡Arithmetic mean.
§One patient had an incomplete BGP baseline assessment and was not included.

SIB and ADCS-ADL19 by Visit (Observed Case) and at End Point (LOCF)



No. of Patients							
Memantine	198	197	190	185	181	171	198
Placebo	197	194	180	169	164	153	196
LS Mean Difference		−1.2	−1.5	−3.1	−2.7	−3.4	−3.4
P Value		.06	.03	<.001	.006	<.001	<.001



No. of Patients							
Memantine	198	198	190	185	181	172	198
Placebo	197	195	182	170	163	152	197
LS Mean Difference		−0.8	−1.1	−1.3	−1.4	−1.6	−1.4
P Value		.03	.01	.02	.03	.02	.03

JAMA

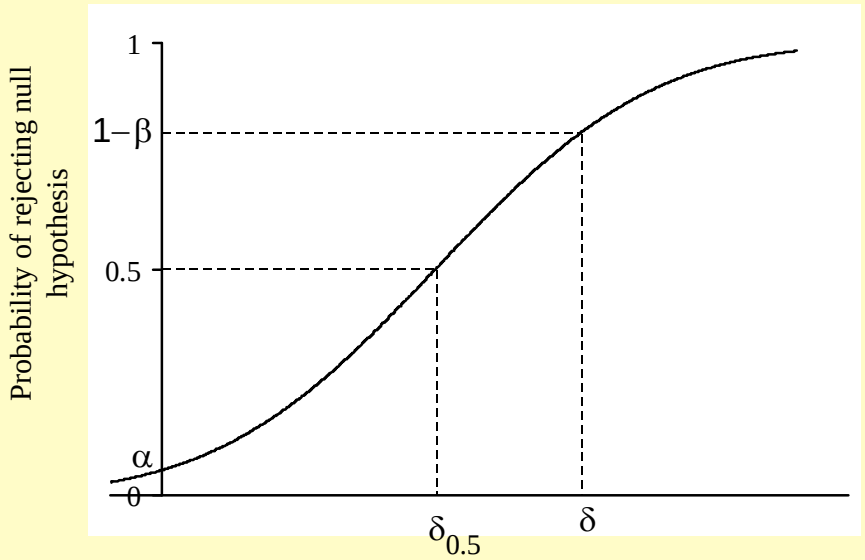
Table 3. Adverse Events Reported in at Least 5% of Patients in Either Treatment Group*

	Adverse Event, No. (%)	
	Placebo (n = 201)	Memantine (n = 202)
Agitation	24 (11.9)	19 (9.4)
Confusion	4 (2.0)	16 (7.9)
Fall	14 (7.0)	15 (7.4)
Influenza-like symptoms	13 (6.5)	15 (7.4)
Dizziness	16 (8.0)	14 (6.9)
Headache	5 (2.5)	13 (6.4)
Urinary tract infection	10 (5.0)	12 (5.9)
Urinary incontinence	6 (3.0)	11 (5.4)
Accidental injury	16 (8.0)	10 (5.0)
Upper respiratory tract infection	13 (6.5)	10 (5.0)
Peripheral edema	8 (4.0)	10 (5.0)
Diarrhea	17 (8.5)	9 (4.5)
Fecal incontinence	10 (5.0)	4 (2.0)

*Patients may have reported more than 1 adverse event.

CHL 5225 H The Language of Clinical Trials

Sample Size and Power Curves



CHL 5225 H The Language of Clinical Trials

Sample Size and Power Curves

Power curves will pass through $(0,\alpha)$ because we're testing the hypothesis

H: $\Delta = 0$ vs. A: $\Delta > 0$ at the level α

(i.e. probability of reject H when it's true is α)

Steepness of the power curve will increase with sample size

CHL 5225 H The Language of Clinical Trials

Multiplicity and Controlling Type I Error (i.e. false positives)

- Multiple arms
- Multiple outcomes (endpoints)
- Multiple subgroups
- Interim analyses

CHL 5225 H The Language of Clinical Trials

Exercise

Search through recent issues of

- Journal of the American Medical Association
- New England Journal of Medicine
- Lancet
- Journal of Clinical Oncology

and located one or two reports of randomized clinical trials

Read the articles and identify some of the issues we discussed today