2024 Summer Institute In Statistics for Clinical & Epidemiological Research

Module 3:

Design, Conduct, and Analysis of Randomized Clinical Trials with Time to Event Primary Endpoints

Lecture 26:

Noninferiority

Scott S. Emerson, M.D., Ph.D.
Professor Emeritus of Biostatistics
University of Washington

1

Outline

- 2004 Case Study: Anti-folate therapy in NSCLC
  - Prior History
  - Design of Trial
  - Trial Results
- Regulatory Setting
  - Scientific Issues
  - Statistical Issues
- My Conclusions
  - Scientific
  - Statistical
  - Logistical

2

2

Case Study

Prior History

3

3

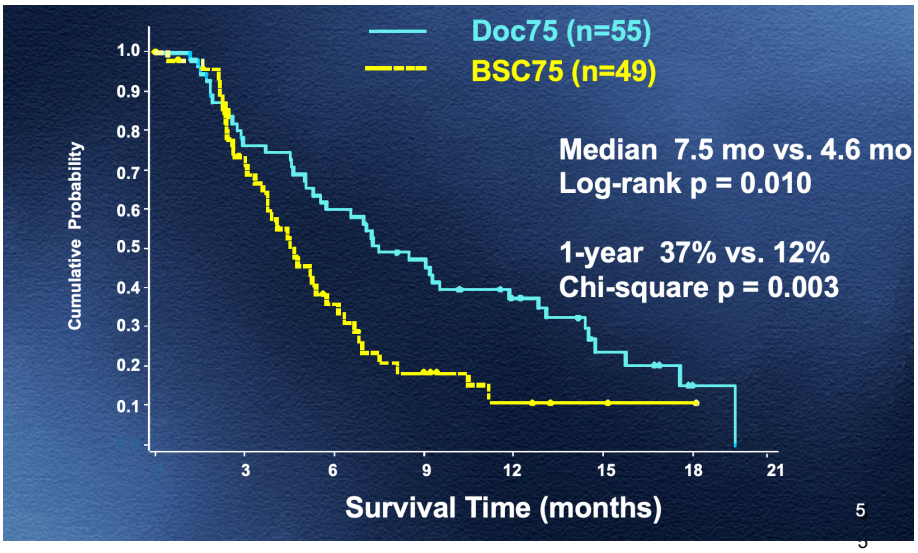
Second Line Therapy in NSCLC

- Non-Small Cell Lung Cancer
- 1997 Standard of treatment
  - ASCO: No proof of effective second line treatment
  - Best supportive care (BSC)
    - Median survival time: 4 – 5 months
    - One-year survival probability: 12 – 18%
- Clinical trials: TAX 317 and TAX 320
  - Docetaxel 100 mg/m2 (later 75 mg/m2) vs BSC
  - Docetaxel 75 and 100 mg/m2 vs Vinorelbine or Ifosfamide

4

4

Overall Survival - TAX 317B



5

5

Efficacy Results – TAX 317

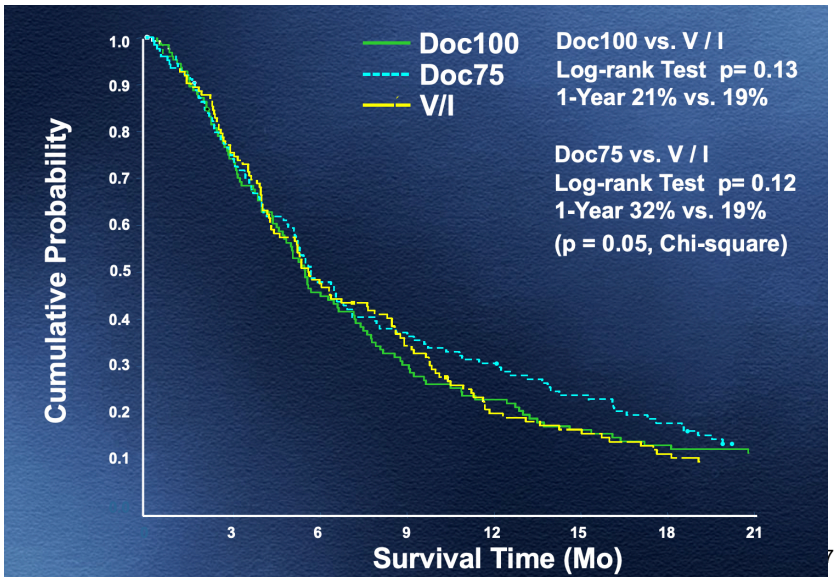
	Doc100 (n=49)	Doc75 (n=55)	BSC (n=49)
Partial Response	6%	6%	—
TTPD	—	12.3 wk	7.0 wk
Median Survival	5.9 mo	7.5 mo	4.6 mo
Log-rank p-value	0.780	0.010*	—
One-Year Survival	19%	37%	12%

* **44% reduction in risk of death compared to BSC**

6

6

Overall Survival - TAX 320



7

Efficacy Results – TAX 320

	Doc100 (n=124)	**Doc75 (n=124)**	V/I (n=122)
Partial Response	11%	**7%**	1%
TTPD	8.4 wk	**8.5 wk**	7.9 wk
Median Survival	5.5 mo	**5.7 mo***	5.6 mo
One-Year Survival	21%	**32%***	19%

*** Log-rank p=0.13, Chi square p=0.05**

8

8

Docetaxel Toxicities (75 mg/m^2)

Toxicity	Any	Grade 3/4
Neutropenia	84.1	65.3
Infection	33.5	10.2
Diarrhea	22.7	2.8
Febrile Neutropenia	-	6.3 *
Neurosensory	23.3	1.7
Alopecia	56.3	-

*** Grade 4 neutropenia with fever >38°C with iv antibiotics or hospitalization**

9

9

Second Line Therapy in NSCLC

- Non-Small Cell Lung Cancer
- 2004 standard of treatment
  - ASCO: Docetaxel for second line treatment
  - (Increasing use of docetaxel as first line)

10

10

Case Study

Pivotal Trial Design

11

11

Intervention

- Pemetrexed
  - Anti-folate
  - Administered with folic acid supplementation

12

12

Pilot Studies

- Single arm studies of pemetrexed with or without other chemotherapy
- First line NSCLC
- Second line NSCLC
- (Also studied in mesothelioma; FDA approval Feb 04)

13

13

2000: On To Phase III

- Ideal
  - Randomized, double blind RCT of new treatment against placebo to show efficacy
- Real world
  - Clinical trials that simultaneously ensure
    - Scientific / statistical credibility
    - Individual ethics of patients on trial
    - Group ethics of patient population

14

14

Design of Pivotal Study

- Randomized with docetaxel active control
- BSC considered not feasible in US
- Combination chemotherapy not feasible due to toxicity
- Docetaxel only approved agent for second line NSCLC

15

15

Potential Hypotheses

- Survival as a primary endpoint
  - Secondary ORR, PFS, TTP, DoR
- Hazard ratio compares survival distributions
- Superiority not necessary for efficacy
  - Efficacy = superior to placebo
- Equivalence study
  - Establish HR approximately 1 with high precision
- Noninferiority study
  - Establish HR not too much greater than 1

16

16

Prespecified Hypotheses

- Superiority or Noninferiority
- Superior if CI excludes HR $\geq$ 1
- Noninferior if CI excludes HR $\geq$??
  - European agency suggested 1.1
    - Power calculation assumed true HR 0.83
    - 400 events (520 subjects) to have 80% power
    - Observe estimated HR < 0.90 to rule out true HR > 1.1
  - Consistent with retaining > 50% docetaxel effect

17

17

Percent Retention Method

- Rothmann, et al. (Jan, 2003)
- Historical trial (TAX 317)
  - Docetaxel : BSC HR 0.56 (95% CI 0.35, 0.88)
- Hypothetically "significant results" from planned trial
  - Pemetrexed : Docetaxel HR 0.90 (95% CI 0.74, 1.10)
- "Induced" Pemetrexed : BSC comparison
  - Use estimates AND standard errors to estimate
    - Pemetrexed : BSC HR and 95% CI
    - Express that ratio as a proportion of Docetaxel : BSC HR
      - Also provide 95% CI for the "percent retention"

18

18

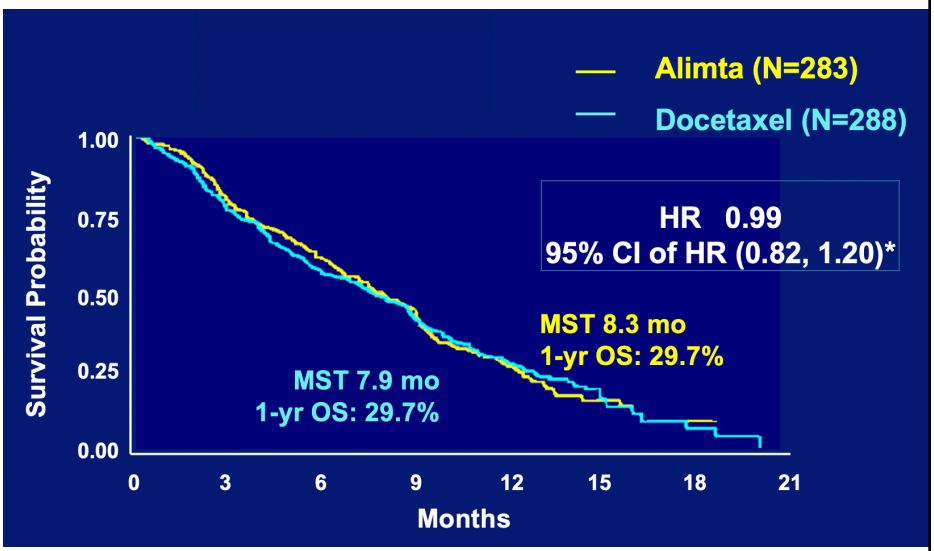
Case Study

2003 Trial Results

19

19

Survival (ITT)

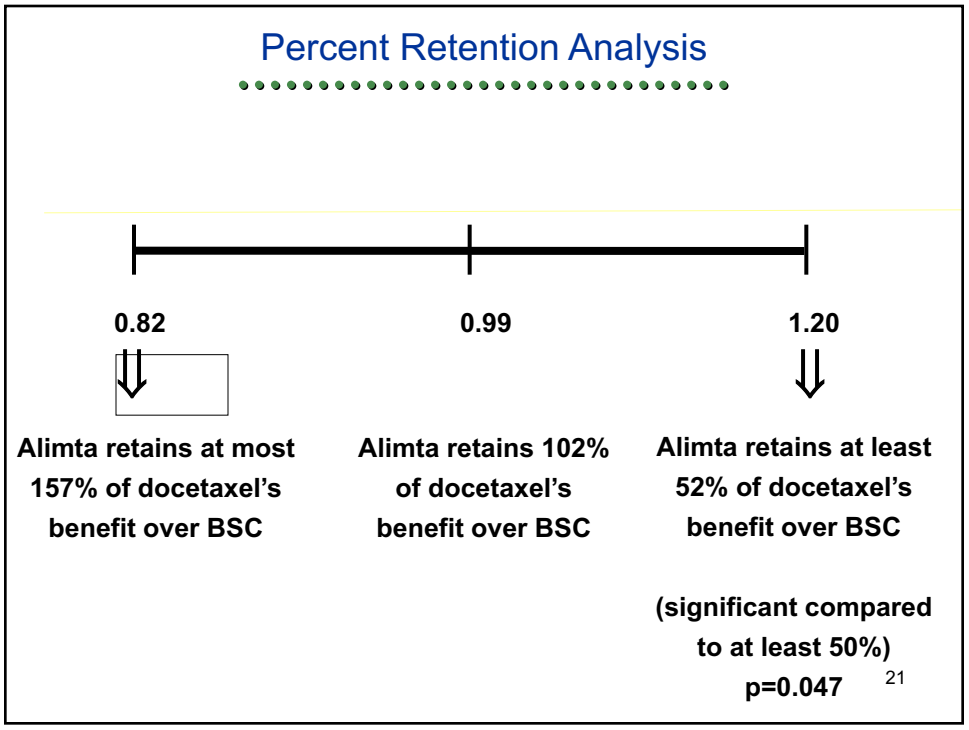


ITT = intent to treat
MST = median survival time

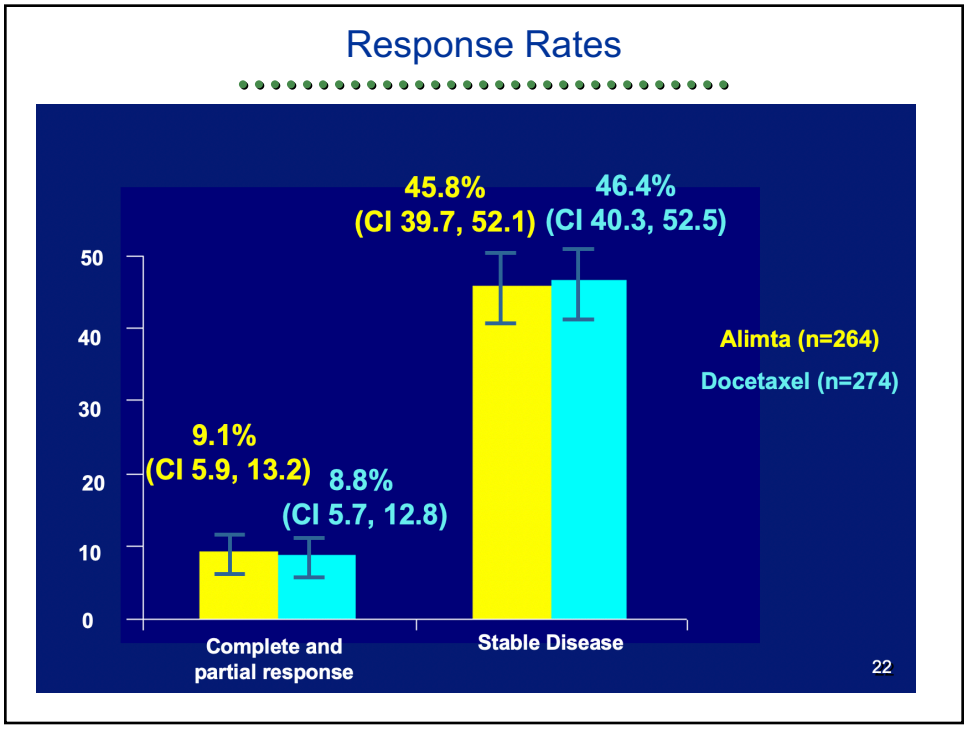
** Does not rule out a margin greater than 1.11*

20

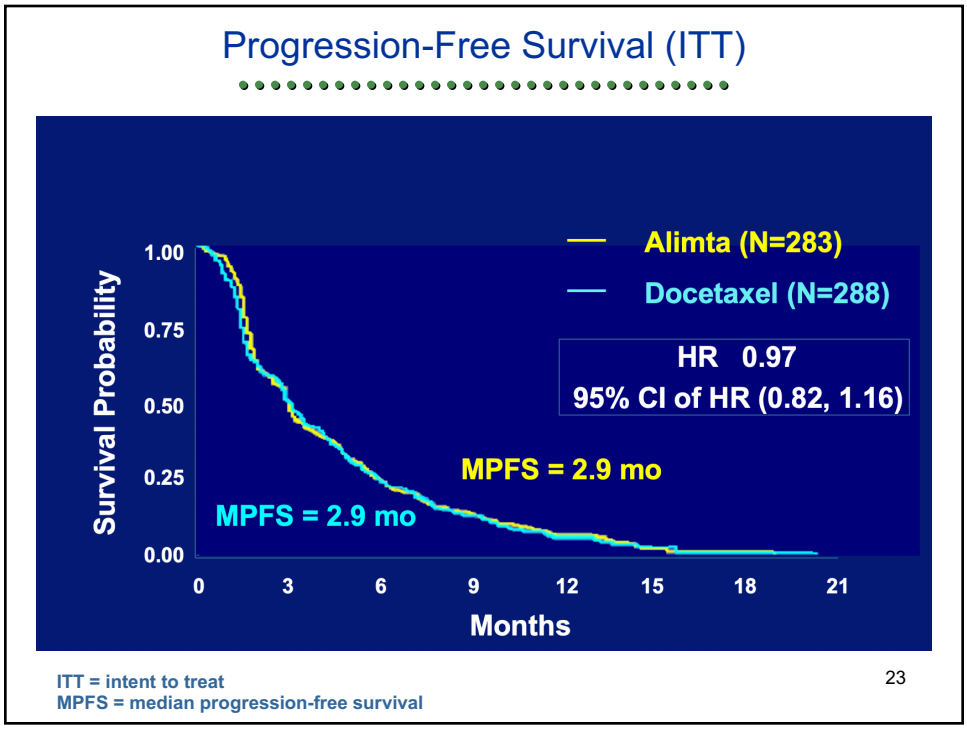
20


Percent Retention Analysis
0.82
0.99
1.20
Alimta retains at most 157% of docetaxel's benefit over BSC
Alimta retains 102% of docetaxel's benefit over BSC
Alimta retains at least 52% of docetaxel's benefit over BSC
(significant compared to at least 50%)
p=0.047
21

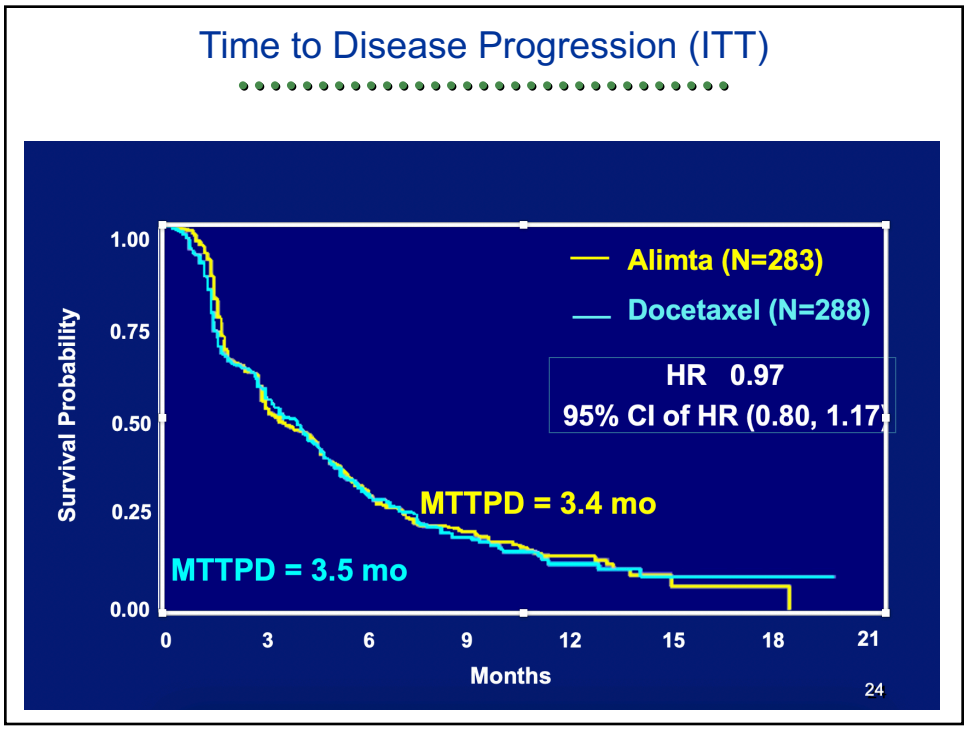
21


Response Rates
45.8%
(CI 39.7, 52.1)
46.4%
(CI 40.3, 52.5)
50
40
30
20
10
0
9.1%
(CI 5.9, 13.2)
8.8%
(CI 5.7, 12.8)
Alimta (n=264)
Docetaxel (n=274)
Complete and partial response
Stable Disease
22

22


Progression-Free Survival (ITT)
Survival Probability
1.00
0.75
0.50
0.25
0.00
0
3
6
9
12
15
18
21
Months
Alimta (N=283)
Docetaxel (N=288)
HR 0.97
95% CI of HR (0.82, 1.16)
MPFS = 2.9 mo
MPFS = 2.9 mo
ITT = intent to treat
MPFS = median progression-free survival
23

23


Time to Disease Progression (ITT)
Survival Probability
1.00
0.75
0.50
0.25
0.00
0
3
6
9
12
15
18
21
Months
Alimta (N=283)
Docetaxel (N=288)
HR 0.97
95% CI of HR (0.80, 1.17)
MTTPD = 3.4 mo
MTTPD = 3.5 mo
24

24

Toxicities

	Percent of Patients		
Toxicity	Alimta (N=265)	Docetaxel (N=276)	p-value
Infect w Gr 3/4 Neutropenia	**0**	**5.8**	**<0.001**
Fatigue	15.8	16.7	0.817
Nausea	3.8	2.5	0.466
Vomiting	1.5	1.4	1.0
Stomatitis	1.1	1.1	1.0
Diarrhea	**0.4**	**4.0**	**0.006**
Pulmonary Toxicity	6.8	9.8	0.217
Neurosensory (Gr 2-4)	**0.8**	**4.3**	**0.012**
Alopecia (all grades)	**6.7**	**37.7**	**<0.001**

25

25

Regulatory Setting

Scientific Issues

26

26

Noninferiority Trials

- ICH E-10 Guidelines
  - Active control treatment must truly be active in the study population
- Possible differences from historical trial
  - Patient population
  - (Ancillary) treatments
  - Clinical endpoints
    - Measure summarizing distribution

27

27

Clinically Important Effects

- Therapeutic index
  - New treatment is better than placebo
  - New treatment is safe
- But: Need to be able to mount ethical clinical trial
  - New treatment cannot be too much worse than existing treatment in serious disease

28

28

Clinically Important Effects: ODAC

- Pemetrexed has a more favorable toxicity profile than docetaxel?
  - Unanimous yes
- Supporting efficacy data on tumor response and progression-free survival outweigh the uncertainly about loss of docetaxel survival effect by using post-study pemetrexed?
  - Unanimous yes
- Given the potential confounding effect of crossover and the problem of estimating control effect, is there sufficient evidence to warrant regular approval?
  - 8 No; 5 Yes
- Recommended accelerated but not full approval

29

29

Clinically Important Effects

- Richard Pazdur, MD (FDA, Oncology Drugs):

"The active control in a clinical trial, in this case docetaxel, should have a pronounced and measurable effect, and we should have multiple trials so we could perform meta-analysis. In this case, there is neither. In addition, the primary objective—survival—was not achieved, and the significant crossover from pemetrexed to docetaxel obscures the differences between the two drugs."

30

30

Regulatory Setting

Statistical Issues

31

31

General Issues

- Multiple comparisons
  - Can we simultaneously consider superiority and noninferiority?
    - Science vs Statistics vs Game theory
- Pivotal results
  - Usual Phase III standards:
    - Two independent level 0.025 trials
  - Pivotal study:
    - Level .000625?
    - Level .004?

32

32

Noninferiority Trials

- Issues in setting the "margin"
  - What measure compares distributions?
  - Is the treatment effect random?
  - How much of a decrease in effect is acceptable?
  - Need to avoid "cherry picking" worst historical results
  - How to account for variability in the estimate(s) from historical trials?

33

33

Precedence

- Is the treatment effect random?
  - Ideally use meta-analysis of multiple trials
- How much of a decrease in effect is acceptable?
  - 10%, 20%, retain 50% of active control effect?
- How to account for variability in the estimate(s) from historical trials?
  - Use worst case from historical 95% CI?
    - 95-95 rule
  - Explicitly account for variability in historical trial

34

34

My Conclusions

Scientific Issues

35

35

Active Control Trials

- Rationale for decisions about efficacy when using active controls
- Control treatment is truly active in study population
  - Superiority
  - Noninferiority if a margin can be established
- Control treatment is standard of care
  - Superiority
  - Noninferiority if superior on secondary endpoints

36

36

Noninferiority Trials

- Assessing possible differences from historical trial
- Patient population
  - Baseline risk factors
- (Ancillary) treatments
  - Post randomization factors
- Clinical endpoints
  - Similarity of response under active treatment for new and historical trials
  - Measure summarizing distribution
    - Transitivity

37

37

Baseline Comparison

	TAX 317 Data (N=204)	JMEI (N=571)
Median Age	61	58
Female	33%	28%
Performance status (2)	15%	12%
Stage IV	79%	75%
Number of prior chemo (2)	25%	6%
Prior platinum	100%	91%
Prior taxane	0%	27%
Best response to prior chemo (other than CR/PR)	65%	64%

38

38

Post Study Chemotherapy

Type	No. of Patients in Each Arm (%)	
	Alimta (N=265)	Docetaxel (N=276)
≥ 1 Chemotherapy	126 (47.5)	107 (38.8)
Platinum	9 (3.4)	15 (5.4)
Docetaxel	85 (32.1)	11 (4.0)
Paclitaxel	4 (1.5)	3 (1.1)
Vinorelbine	6 (2.3)	25 (9.1)
Gemcitabine	17 (6.4)	32 (11.6)
Other chemo	22 (8.4)	34 (12.3)
Gefitinib	5 (1.9)	21 (7.6)

39

39

Survival by Post Study Chemotherapy

Patient Population	Alimta (N=265)	MS	Docetaxel (N=276)	MS
No post-study chemo	139	6.2 mo	169	5.0 mo
Any post-study chemo	126	9.8 mo	107	10.8 mo
Post-study docetaxel therapy	85	9.6 mo	11	10.1 mo
Other chemotherapy	41	10.6 mo	96	11.2 mo

40

40

Major Issue

- Did apparent similarity of survival arise solely from the cross-over to the current Standard of Care (docetaxel)?
  - Were we in effect just testing immediate vs delayed docetaxel?
- What to make of
  - More Alimta patients receiving post study chemotx: 48% vs 39%
  - 32% of Alimta patients receiving post docetaxel (vs 4%)
  - Longer median survival estimate for
    - Chemotx following docetaxel (10.8 mo) vs Alimta (9.8 mo)
    - Docetaxel following docetaxel (10.1 mo) vs Alimta (9.6 mo)
    - Other chemotx following docetaxel (11.2 mo) vs Alimta (10.6 mo)

41

41

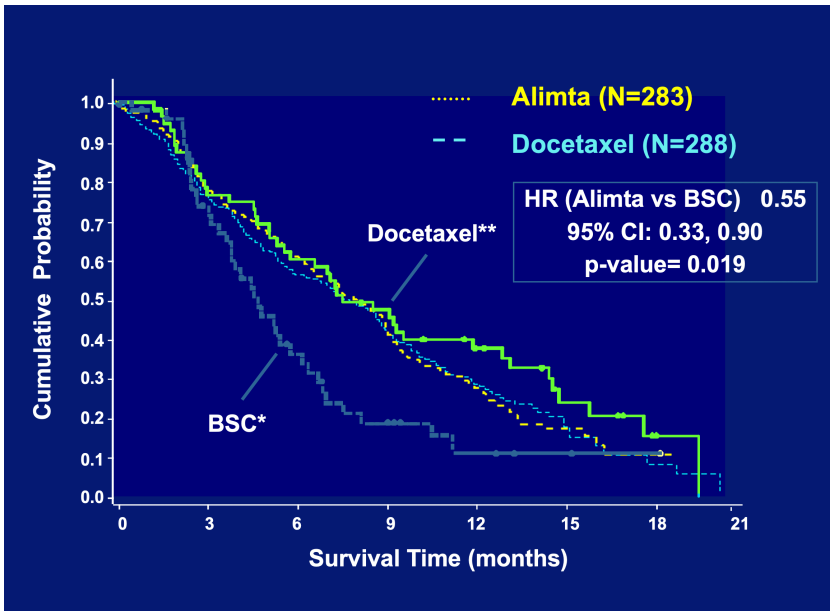
Time Varying Covariate Analysis

- Adjust for post-study chemo as time varying covariate
- Better (but still not perfect) approach
- Prior to post-study chemo
  - Alimta : Docetaxel HR 0.84 (95%CI 0.65, 1.08)
- Post-study chemo to No post-study chemo
  - Docetaxel arm: HR 1.12 (95%CI 0.81, 1.53)
  - Alimta arm: HR 1.58 (95%CI 1.17, 2.12)
- Clearly no strong benefit of docetaxel after Alimta
  - Did use of Alimta make docetaxel ineffective?
  - Does higher use of docetaxel explain worse survival?

42

42

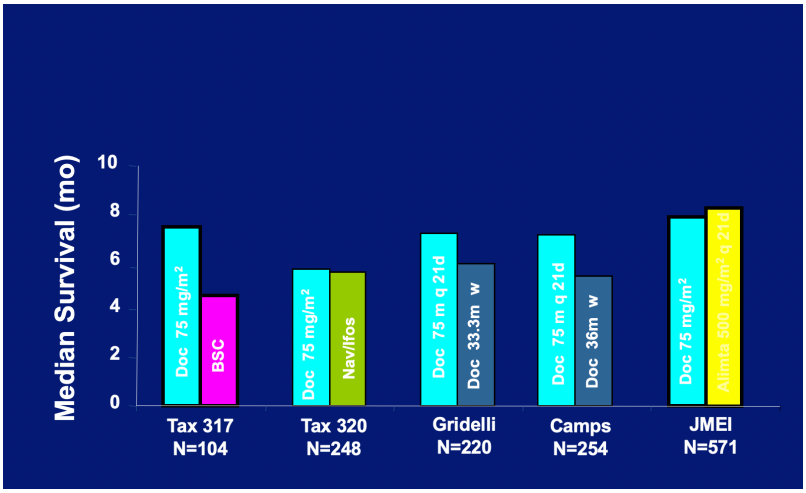
Consistency Across Studies



43

43

Consistency Across Studies



44

44

Scientific Comparability

- Some issues with comparability are inevitable, however
  - Baseline variables largely similar
    - And adjustment for disparities preserves effect
  - Post randomization treatments differ
    - But no real evidence that conferred advantage
  - Similar response to Docetaxel across studies
    - Suggests no large random treatment effect
  - (Sensitivity analysis)

45

45

My Conclusions

Statistical Issues

46

46

Noninferiority Trials

- Issues with induced hazard ratio estimates
- Transitivity of hazard ratio estimate
  - Proportional hazards or same survival and censoring distribution
- Setting the margin
  - Science versus statistics
  - Game theory
- Multiple comparisons
  - Only a single estimate and CI is used
- Sensitivity analysis

47

47

Transitivity

- The weighting scheme used in the weighted logrank statistics also introduces intransitivity to studies
- The weights are stochastically determined from
  - Each group's survivor function
  - The censoring distribution

$$T = \sum_t \frac{(O_t - E_t)^2}{E_t}$$

$$W = \sqrt{\frac{N_0 + N_1}{N_0 N_1}} \sum_t \frac{n_0 n_1}{n_0 + n_1} \left(\hat{\lambda}_1(t) - \hat{\lambda}_0(t) \right)$$

- Hence we can obtain A > B > C > A
  - Very distressing to regulatory agencies, if not all scientists

48

48

Setting the Margin

- Scientific issues should govern how we would react to particular hazard ratios if we knew the truth
- Statistical issues should govern whether we have discriminated between hypotheses that would cause us to act differently

49

49

Sensitivity Analysis

- We can consider the possibility that the new trial only has P% of the study participants similar to the historical trial
- Assume that neither the active control nor the new treatment are effective in the (100-P)%
- How large can P be and have the new study suggest a statistically significant beneficial effect of Alimta over BSC?
  - Turns out to be related to the percent retention analysis:
    - So long as P = 48%

50

50

My Conclusions

Logistical Issues

51

51

Pivotal Trial

- Clearly there are several unresolved issues
- Comparability of patient population
- Effect of ancillary treatments
- Consistency of treatment effects across studies
- Strength of evidence
  - Induced Alimta : BSC HR 0.55 (95% CI 0.33-0.90)
  - Does not exclude 1 with p-value < .004

52

52

Nevertheless

- There are also mitigating factors
- Alimta did show evidence of anti-cancer effect in mesothelioma
- There is a promising estimate of therapeutic index in NSCLC
  - Best estimate: equal efficacy, better toxicity
- This trial is as large as any reported trial in 2nd line NSCLC
  - Unlikely to do a replication of this trial exactly
  - Could a second trial be done in other settings?
    - First line
    - Combinations?

53

53

Time Marched On: 2008

- First line NSCLC trial Alimta + cisplatin vs gemcitabine + cisplatin
  - HR: 0.94 (95% CI 0.84 – 1. 05)
  - Prespecified subgroups
    - Non squamous cell HR: 0.84 (95% CI 0.74 – 0.96)
    - Squamous cell HR: 1.23 (95% CI 1.00 – 1. 51)
- Retrospective (*post hoc)* analysis of 2nd line Alimta vs docetaxel
  - Adjusted for ECOG PS, time since prior tx, stage, sex
  - HR: 0.93 (95% CI 0.76 – 1. 13) (2004 ODAC, no adj for sex)
  - Subgroups motivated by 1st line results
    - Non squamous cell HR: 0.78 (95% CI 0.61 – 1.00)
    - Squamous cell HR: 1.56 (95% CI 1.08 – 2.26)
- FDA approval for non squamous NSCLC
  - 1st line: Alimta (+ folate) + cisplatin
  - 2nd line: monotherapy (+ folate)

54

54

What About Censoring at Change of Therapy

- This case provides data about possible impact of the approach based on censoring subjects when they change therapies
  - Overall survival was primary endpoint pre specified in protocol
- Censoring at change of therapy suggested clear noninferiority
  - Prior to post-study chemo
    - Alimta : Docetaxel HR 0.84 (95%CI 0.65, 1.08)
- But on the full analysis, there were suggestions of differential behavior of chemotherapy in general and docetaxel in particular
- The per randomization analysis can be trusted, even if still leaving a dilemma re standards for noninferiority
  - Alimta : Docetaxel HR 0.99 (95%CI 0.82, 1.20)

55

55