Acute GVHD: Grading and Endpoints

Paul Martin, MD

Fred Hutchinson Cancer Research Center, Seattle, WA

Disclosures

- Contract support for clinical trials
  - Osiris: Mesenchymal stem cells for secondary treatment of acute GVHD
  - DOR BioPharma: Beclomethasone diproprionate for prevention of acute GVHD
  - Roche Laboratories: Mycophenolate mofetil for initial treatment of chronic GVHD
- Consultation without compensation
  - Seattle Genetics
  - Lycera
- Consultation with compensation
  - Xymogenetics
- Grant Support
  - National Cancer Institute: Treatment of acute and chronic GVHD

Questions to be Addressed

- What are the problems associated with grading?
- Is survival a good anchor for validation of grading?
- Is grading of GVHD reproducible?
- What endpoints should be used for prevention trials?
  - Symptom burden across time
  - Survival
  - GVHD grade
  - Decision to use steroid treatment

Questions to be Addressed

- What are the problems associated with grading?
- Is survival a good anchor for validation of grading?
- Is grading of GVHD reproducible?
- What endpoints should be used for prevention trials?
  - Symptom burden across time
  - Survival
  - GVHD grade
  - Decision to use systemic treatment

Organ Stages of Acute GVHD

Stage	Skin Percent BSA	Liver Bilirubin	Gut Stool Volume
0	0	< 2.0	≤500
1	<25	2.0 – 2.9	>500*
2	25 – 50	3.0 – 5.9	>1000
3	>50	6.0 – 14.9	>1500
4	Bullae	≥15.0	>2000†

*or persistent anorexia, nausea and vomiting
†or severe abdominal pain with or without ileus

Glucksberg et al. Transplantation 1974;18:295-304
Thomas et al. New Engl J Med 1975;292:895-902
Przepiorka et al. Bone Marrow Transplant 1995;15:825-828

Problems in Staging Skin GVHD

- Confluent vs. non-confluent rash in rule of 9's
- Qualitative character of rash not considered

Problems in Staging Liver GVHD

- Confounding by other factors
  - Toxicity of the pretransplant conditioning regimen
  - Use of ursodiol
  - Hemolysis
  - Toxicity of calcineurin inhibitors
  - Sepsis
- Biopsy seldom obtained

Problems in Staging Gut GVHD

- Stool volume is not an ideal measure
  - Uncertain amount of mixing with urine
  - Incomplete measurement
  - High day-to-day variation due to extraneous factors
    - Oral intake
    - Medication effects (anti-diarrheals, narcotics, oral Mg^{++})
  - Cannot be measured in outpatients
- Confounding by other factors
  - Pretransplant conditioning regimen
  - Infection
- Threshold of 500 mL/day is not appropriate
  - Upper limit of normal in adults is 200 mL/day

“Engraftment Syndrome”

	Author		
Criterion	Spitzer	Gorak	Schmid
Fever	Major	Major	Major
Erythematous rash	Major	Major	Major
Pulmonary capillary leak	Major	Major	Major
Hepatic dysfunction	Minor		
Renal insufficiency	Minor		
Weight gain	Minor	Major	Major
Encephalopathy	Minor		
Timing with neutrophil engraftment	4 days	4 days	7 days
Minimum number of major criteria	2	2	2

Spitzer. Bone Marrow Transplant 2001;27:893-898
Gorak et al. Biol Blood Marrow Transplant 2005;11:542-550
Schmid et al. Biol Blood Marrow Transplant 2008;14:438-444

“Engraftment Syndrome” vs. GVHD

- Prominence of capillary leak features
- Strict association with initial neutrophil engraftment
- Rapid response to high-dose corticosteroid treatment
- Feasibility of rapid steroid withdrawal (?)

Diagnostic Utility of Biopsy

Organ (n)	Percent Positive	Positive Predictive Value	Negative Predictive Value	Sensitivity	Specificity
Skin (263)	70	0.98	0.40	0.77	0.90
Liver (15)	33	1.0	0.30	0.42	1.0
Gut (214)	47	1.0	0.28	0.55	1.0

Weisdorf et al. BBMT 2003;9:512-518

Seattle Grading of Acute GVHD

Grade	Skin Stage	Liver Stage	Gut Stage	Decrease in Performance
0	0	0	0	None
1	1 – 2	0	0	None
2	3	1	1	Mild
3		2 – 3	2 – 4	Marked
4	4	4		Extreme

Thomas et al. New Engl J Med 1975;292:895-902

IBMTR Grading of Acute GVHD

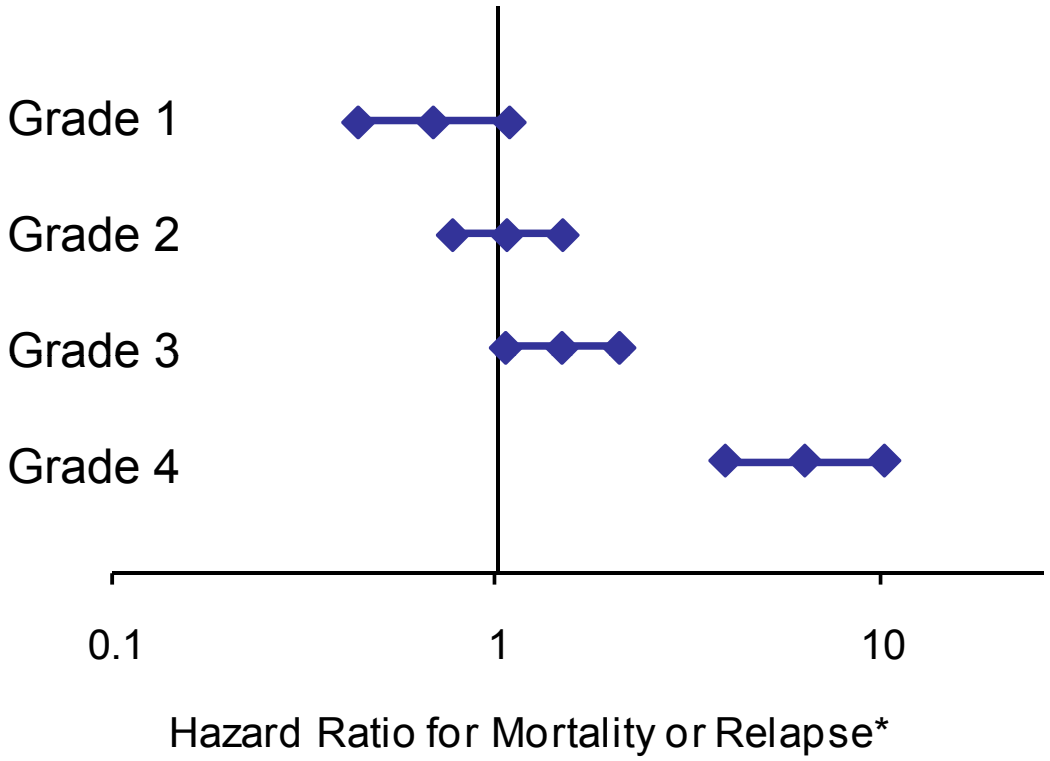
Grade	Skin Stage	Liver Stage	Gut Stage
0	0	0	0
1	1	0	0
2	2	1 – 2	1 – 2
3	3	3	3
4	4	4	4

Rowlings et al. Brit J Haematol 1997;97:855-864

Questions to be Addressed

- What are the problems associated with grading?
- Is survival a good anchor for validation of grading?
- Is grading of GVHD reproducible?
- What endpoints should be used for prevention trials?
  - Symptom burden across time
  - Survival
  - GVHD grade
  - Decision to use systemic treatment

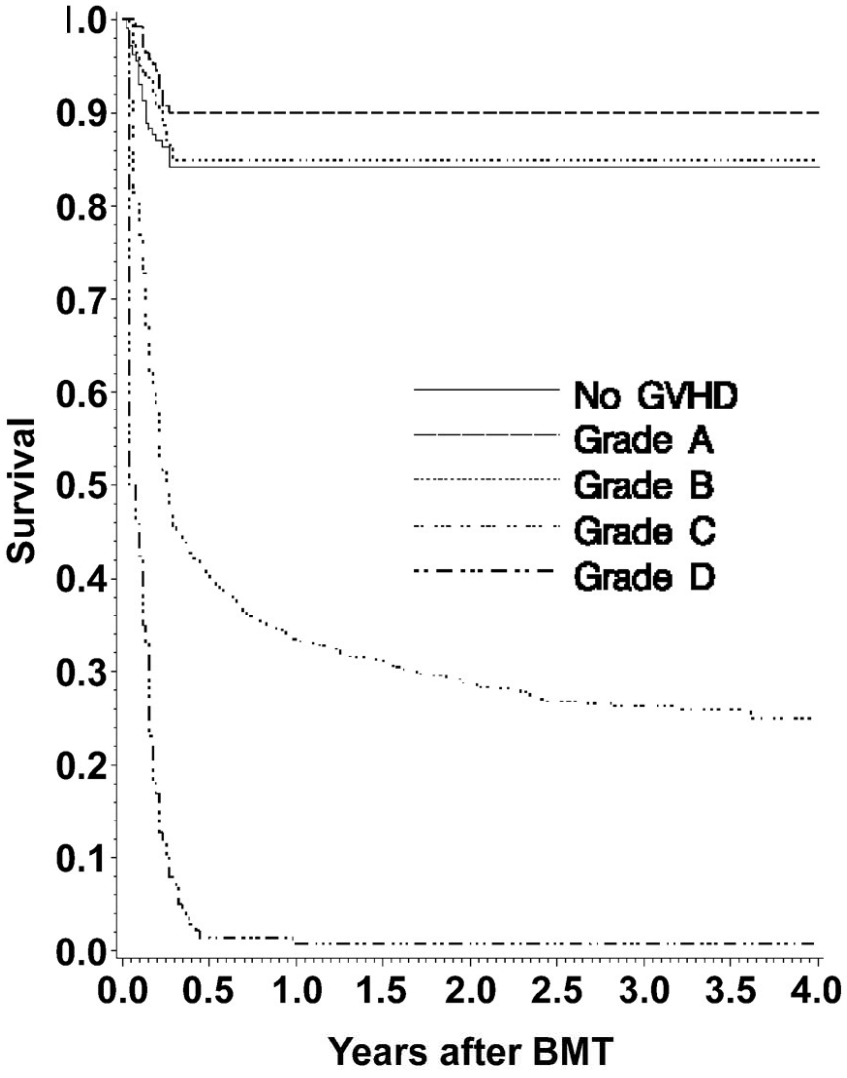
Outcome Associated with Maximum GVHD Grade

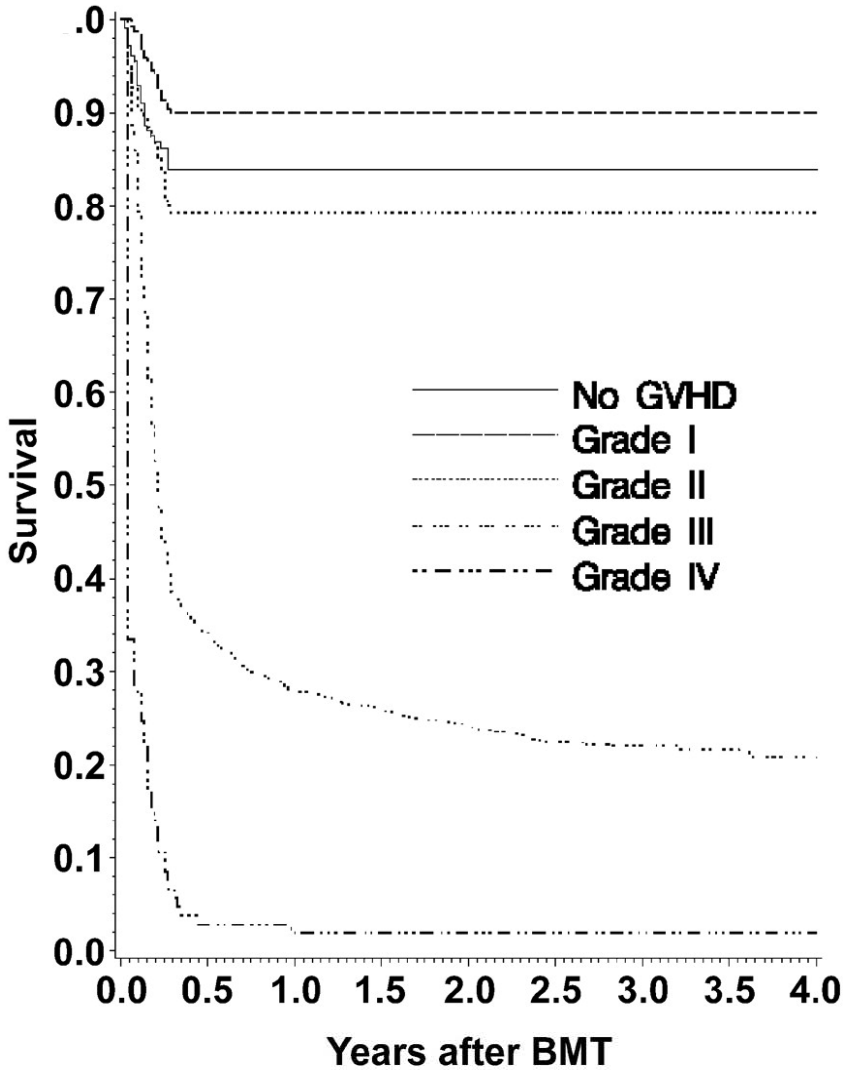


*time-dependent model adjusted for patient and donor age, pretransplant diagnosis and disease stage, method of GVHD prevention, HLA match, recipient CMV antibody

Weisdorf et al. BBMT 2003;9:512-518

Peak Grades Correlated with Survival





Cahn et al. Blood 2005;106:1495-1500

Odds Ratios of Mortality*

Maximum Grade	100-Days		1-Year	
	IBMTR	Seattle	IBMTR	Seattle
None	1.0	1.0	1.0	1.0
A,B or I,II	0.27	0.29	0.80	0.80
C or III	0.87	1.04	1.42	2.20
D or IV	3.60	4.30	12.3	13.1

*adjusted for patient age, disease stage and donor type

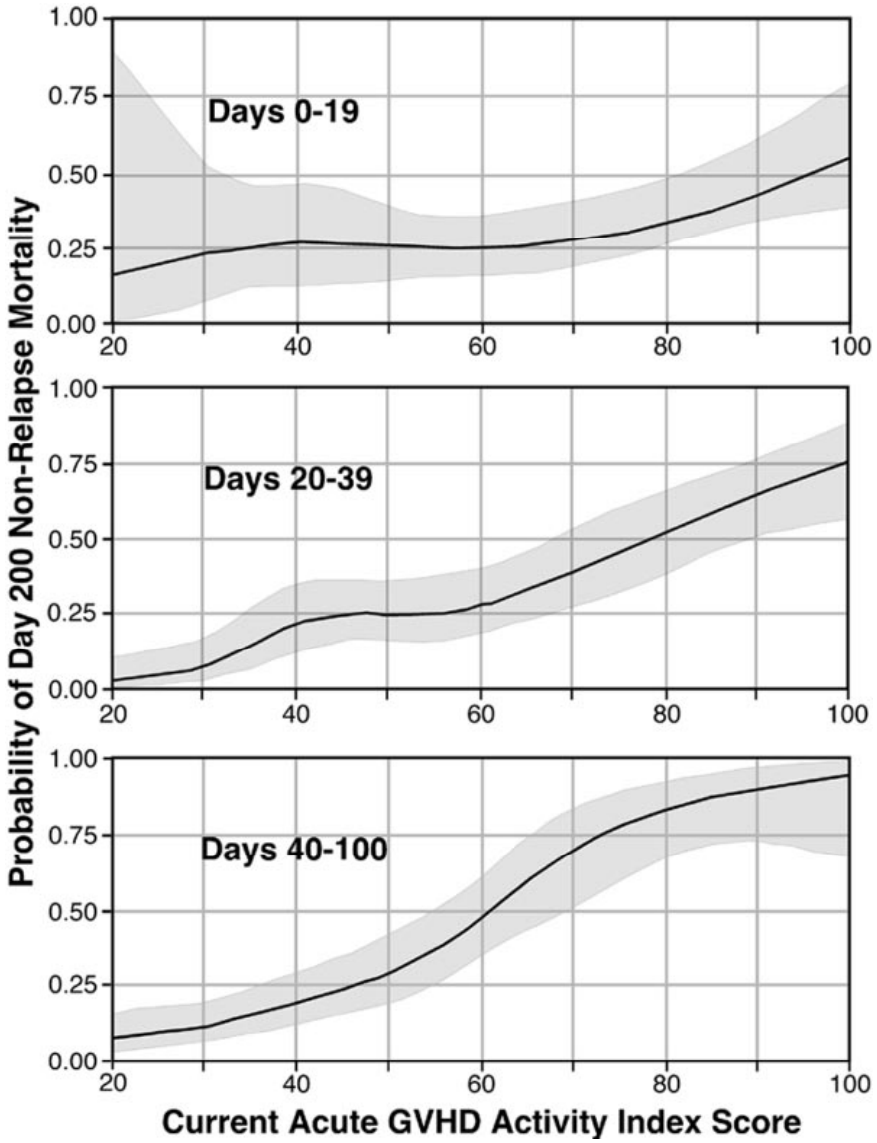
Cahn et al. Blood 2005;106:1495-1500

GVHD Activity Index

Factor	Weight
Liver	
Total serum bilirubin 2.0 – 4.9	16
Total serum bilirubin ≥5.0	26
Oral caloric intake <40% of requirements	20
Any systemic treatment for GVHD	17
Performance	
ECOG 1 – 2	20
ECOG 3 – 4	37

Leisenring et al. Blood 2006;108:749-755

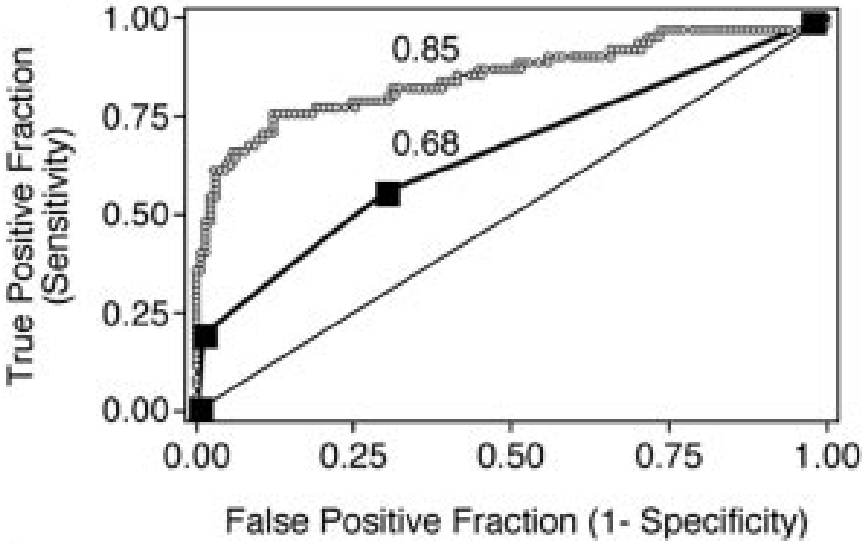
Correlation of Index with Day 200 Mortality



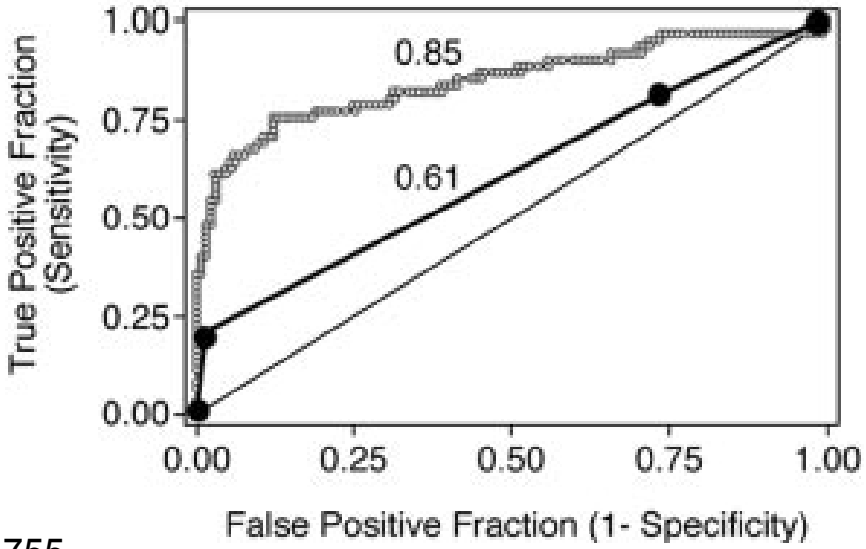
Leisenring et al. Blood 2006;108:749-755

Accuracy in Predicting Day 200 Mortality

Seattle



IBMTR



Leisenring et al. Blood 2006;108:749-755

Questions to be Addressed

- What are the problems associated with grading?
- Is survival a good anchor for validation of grading?
- Is grading of GVHD reproducible?
- What endpoints should be used for prevention trials?
  - Symptom burden across time
  - Survival
  - GVHD grade
  - Decision to use systemic treatment

Comparison of Scoring Methods

Maximum Grade — %	Transplant Center	Computer Algorithm	Panel Review
0	41	38	42
I	13	12	13
II	28	23	23
III	11	20	16
IV	7	7	6

Weisdorf et al. BBMT 2003;9:512-518

Changes in Maximum Scoring

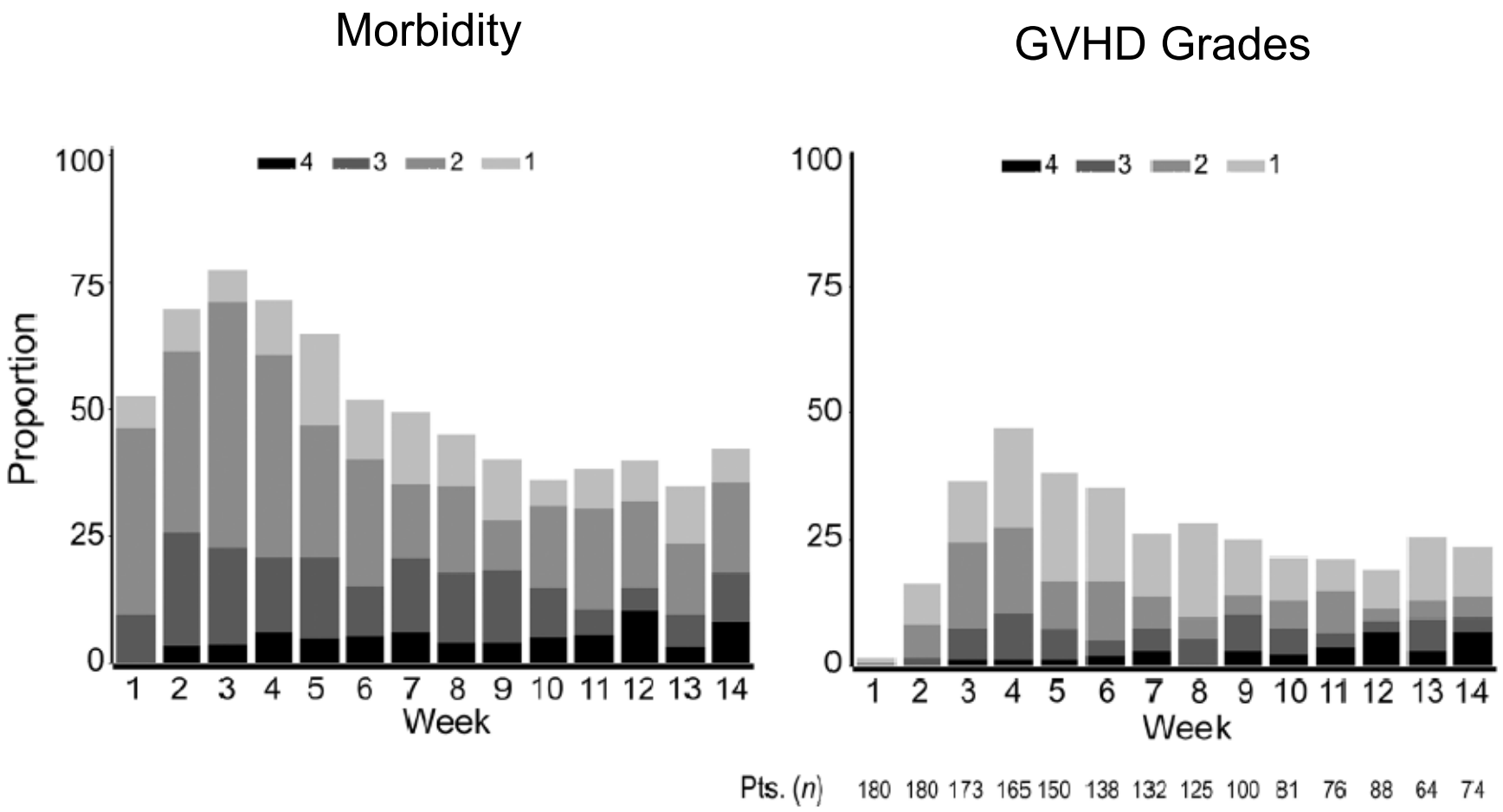
Change — n (%)	TC → ALG	ALG → PR
0 – II → III – IV	42 (10)	1 (0.2)
III – IV → 0 – II	9 (2)	17 (4)
Grade increase	73 (18)	3 (0.7)
Grade decrease	42 (10)	28 (7)

Weisdorf et al. BBMT 2003;9:512-518

Questions to be Addressed

- What are the problems associated with grading?
- Is survival a good anchor for validation of grading?
- Is grading of GVHD reproducible?
- What endpoints should be used for prevention trials?
  - Symptom burden across time
  - Survival
  - GVHD grade
  - Decision to use systemic treatment

Comparison of Morbidity vs. GVHD Grades

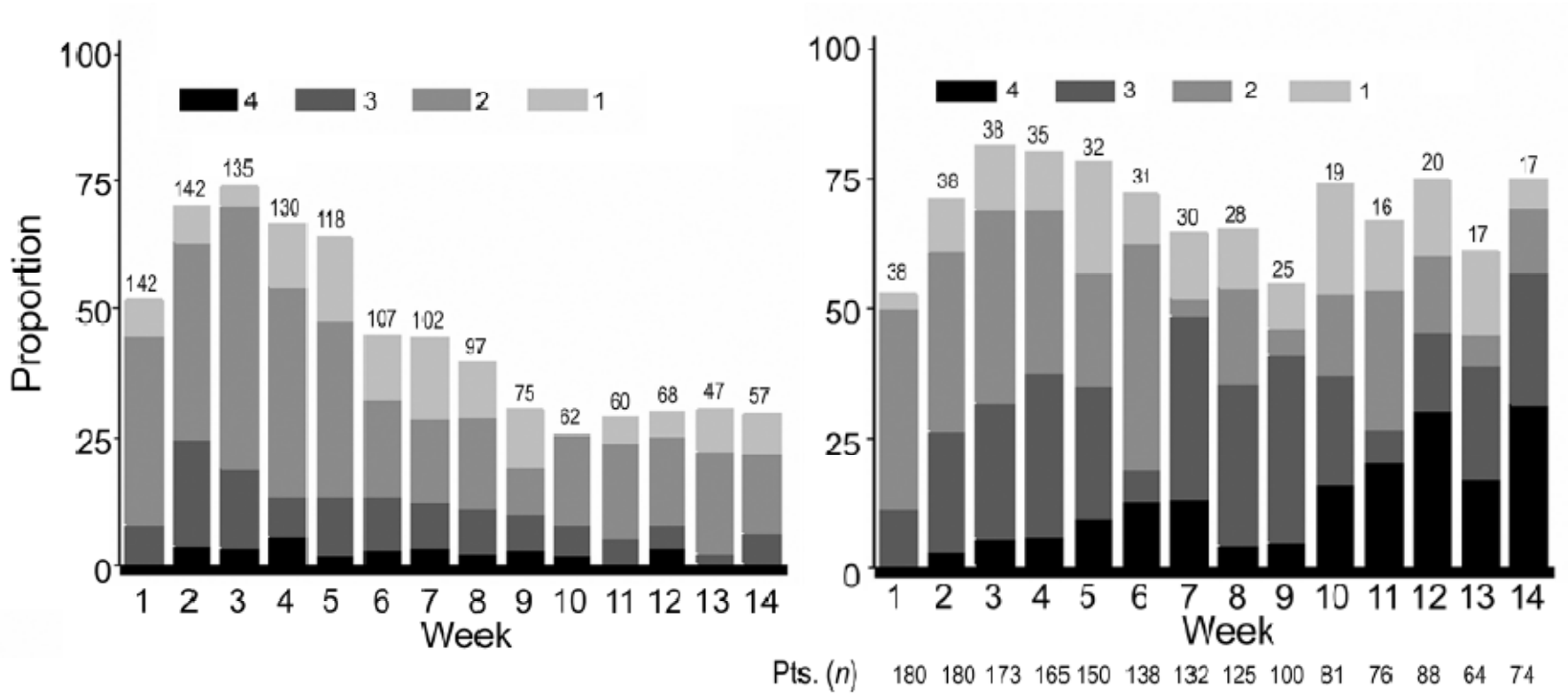


Carnevale-Schianca et al. Biol Blood Marrow Transplant 2009, in press

Morbidity in Grades 0 – II vs. III – IV GVHD

Grades 0 – II

Grades III – IV



Carnevale-Schianca et al. Biol Blood Marrow Transplant 2009, in press

Questions to be Addressed

- What are the problems associated with grading?
- Is survival a good anchor for validation of grading?
- Is grading of GVHD reproducible?
- What endpoints should be used for prevention trials?
  - Symptom burden across time
  - Survival
  - GVHD grade
  - Decision to use systemic treatment

Survival as an Endpoint for Prevention Studies

- This benchmark of success is rarely reached.
- Attainable only if
  - Risk of GVHD in study population is high, and
  - Risk of fatal non-GVHD complications is low, and
  - Risks of fatal non-GVHD complications are balanced between arms, and
  - Intervention has large beneficial effect, and
  - Intervention does not have fatal adverse effects

Percent Uncertainty in Survival Explained by Maximum GVHD Grade before Day 100

Endpoint	IBMTR	Seattle
100 days	1.36	2.04
6 months	13.79	8.01
1 year	6.48	6.44

Cahn et al. Blood 2005;106:1495-1500

Effect of GVHD on 1-year Mortality

GVHD Grade	Observed Frequency*	1-year Mortality*	Attributable Mortality
0 – II	0.75	0.16	0.12
III	0.18	0.70	0.13
IV	0.07	0.98	0.07
Total Observed Mortality			0.32

*Data from Cahn et al. Blood 2005;106:1495

Effect of GVHD on 1-year Mortality

GVHD Grade	Observed Frequency*	1-year Mortality*	Attributable Mortality
0 – II	0.75	0.16	0.12
III	0.18	0.70	0.13
IV	0.07	0.98	0.07
Total Observed Mortality			0.32

GVHD Grade	Expected Frequency	1-year Mortality	Attributable Mortality
0 – II	0.95	0.16	0.15
III	0.05	0.70	0.04
IV	0	0.98	0
Total Expected Mortality			0.19

*Data from Cahn et al. Blood 2005;106:1495

Questions to be Addressed

- What are the problems associated with grading?
- Is survival a good anchor for validation of grading?
- Is grading of GVHD reproducible?
- What endpoints should be used for prevention trials?
  - Symptom burden across time
  - Survival
  - GVHD grade
  - Decision to use systemic treatment

Prevention Endpoints Proposed for Discussion

- Grades III – IV or C – D GVHD
  - Advantages
    - Correlation with decreased survival
  - Disadvantages
    - Difficult to define with precision
    - Susceptible to underestimation by investigators
    - Incidence influenced by efficacy of treatment

- Grades II – IV or B – D GVHD
  - Advantages
    - Correlation with symptoms attributable to GVHD
  - Disadvantages
    - Difficult to define with precision
    - Variation among centers in diagnosis of upper GI GVHD

Questions to be Addressed

- What are the problems associated with grading?
- Is survival a good anchor for validation of grading?
- Is grading of GVHD reproducible?
- What endpoints should be used for prevention trials?
  - GVHD grade
  - Survival
  - Symptom burden across time
  - Decision to use systemic treatment

Factors that Influence Decision

- Anticipated risk of GVHD
  - Unrelated donors
  - Ability to give planned immunosuppressive treatment
- Severity of manifestations
- Rate of change in severity over time
- Anticipated toxicity of systemic treatment

Prevention Endpoints Proposed for Discussion

- Decision to give systemic treatment
  - Advantages
    - Objective fact
    - Not susceptible to underestimation
  - Disadvantages
    - Subject to bias (blinding would be mandatory)
    - Susceptible to overestimation
    - Relationship to clinical benefit not defined

Need for “No Harm” Endpoints*

- Adverse events
- Risk of infections
- Non-relapse mortality
- Recurrent malignancy
- Survival

*Trials typically not adequately powered to rule out small differences