Extent of Dependence of Medical Decision Making on use of p-value: Setting of Treatment for Multiple Myeloma

Keren Osman, MD

Associate Professor of Medicine

Multiple Myeloma/BMT Program

Division of Hematology Oncology

Madhu Mazumdar, PhD

Professor of Biostatistics, Pop Health Sci and Policy

Director, Biostatistics Core, Tisch Cancer Institute

Director, Institute of Healthcare Delivery Science

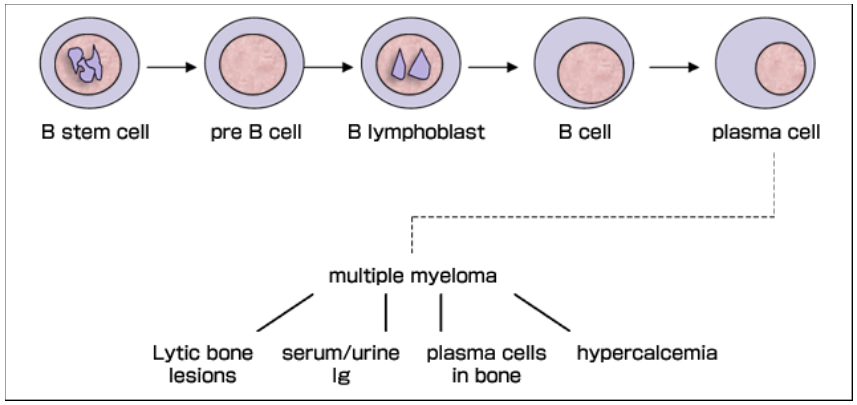
Icahn School of Medicine at Mount Sinai

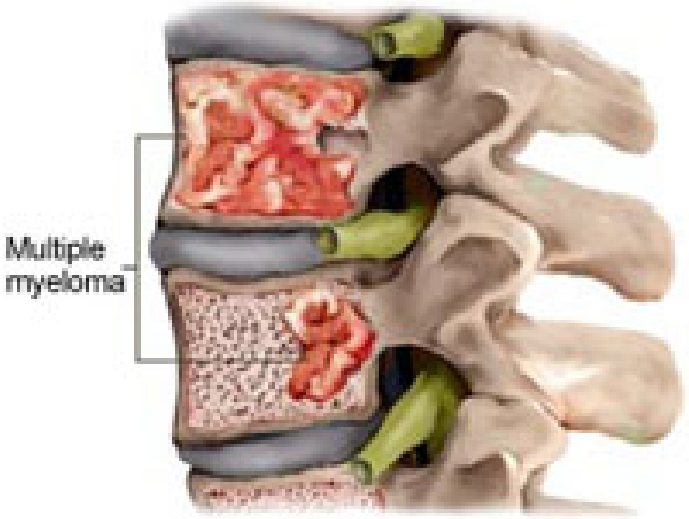
Outline of Our Talk

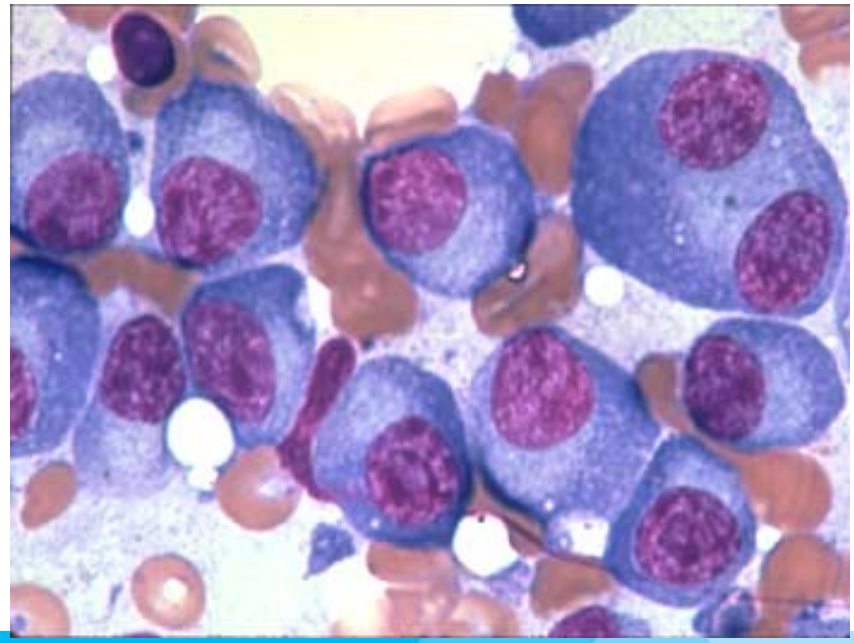
- Intro to Multiple Myeloma (MM) with focus on decision making of Single versus Double Auto Transplant for untreated MM patients
- Comment on the extent of our dependence p-value < 0.05 as the deciding factor; Other unintended misuse of statistics in this field
- Self Correction: New study in MM with robust design and Analysis plans; Increasing use of Bayesian design and analysis in this field
- Strengthening training for additional summary statistics beyond p-value (confidence interval, prior probability, Bayes factor, Posterior probability, False positive rate; Possibility of conversion) and Bayesian design and analysis;
  - Training needed for statisticians and physicians (those collaborating with statisticians and those working on their own)

Multiple Myeloma

- Malignant tumor of plasma cells destroy normal bone tissue causing pain and compromising normal bone marrow function







Epidemiology

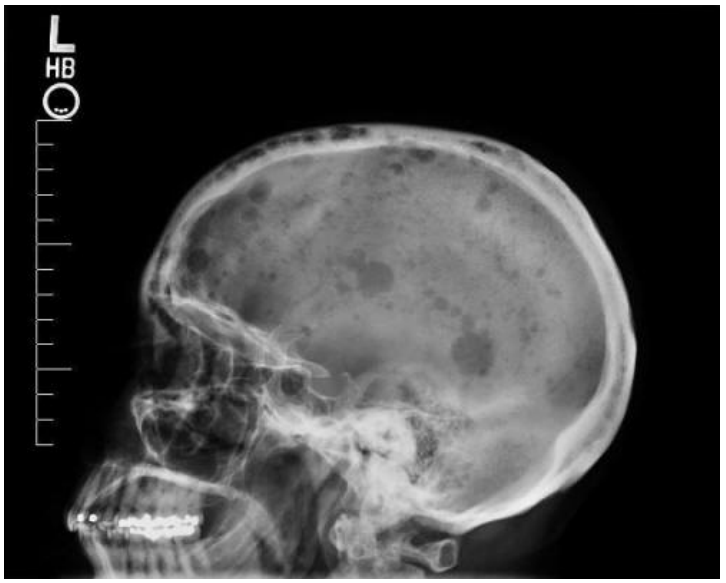
- 45,000 people in the US currently alive with Multiple Myeloma
- 14,600 new cases each year in the US
- Most prevalent hematologic malignancy after NHL.
- 10% of all hematologic malignancies
- 1% of all cancers
- 2% of all cancer deaths

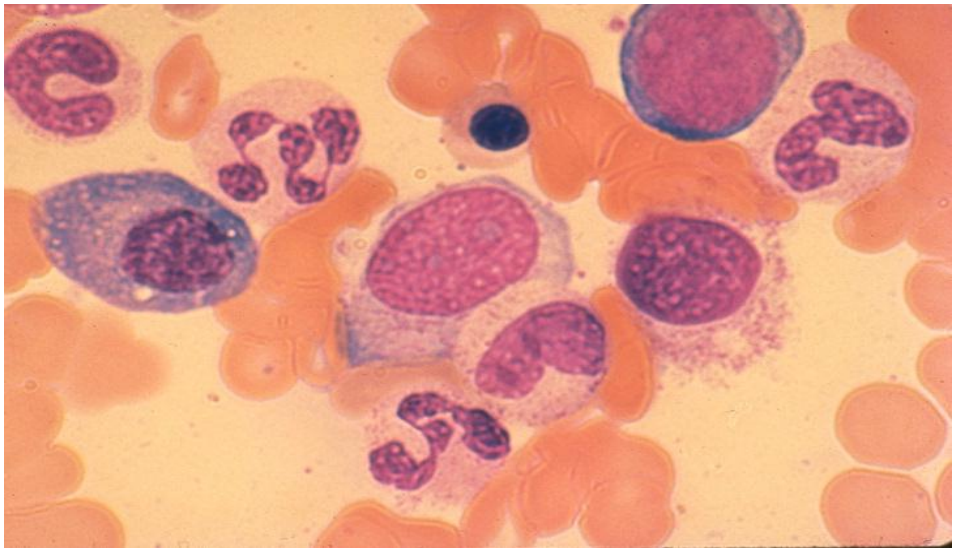
Epidemiology

- Blacks > Whites
- Slight male predominance
- Median age 65 years
- Etiology unknown
- Incurable

Clinical Presentation

- Monoclonal (M) Protein (93%)
- Increased plasma cells in the bone marrow (96%)
- Anemia (73%)
- Lytic Bone Lesions (67%)
- Hypercalcemia >11 (13%)
- Renal failure (19%)





Clinical Presentation

- Bone pain because of lytic lesions
- Bone fractures from myeloma related osteoporosis
- Extreme fatigue from anemia
- Increased risk of infections such as pneumonia and shingles from dysregulated immunity
- Kidney dysfunction or complete kidney failure

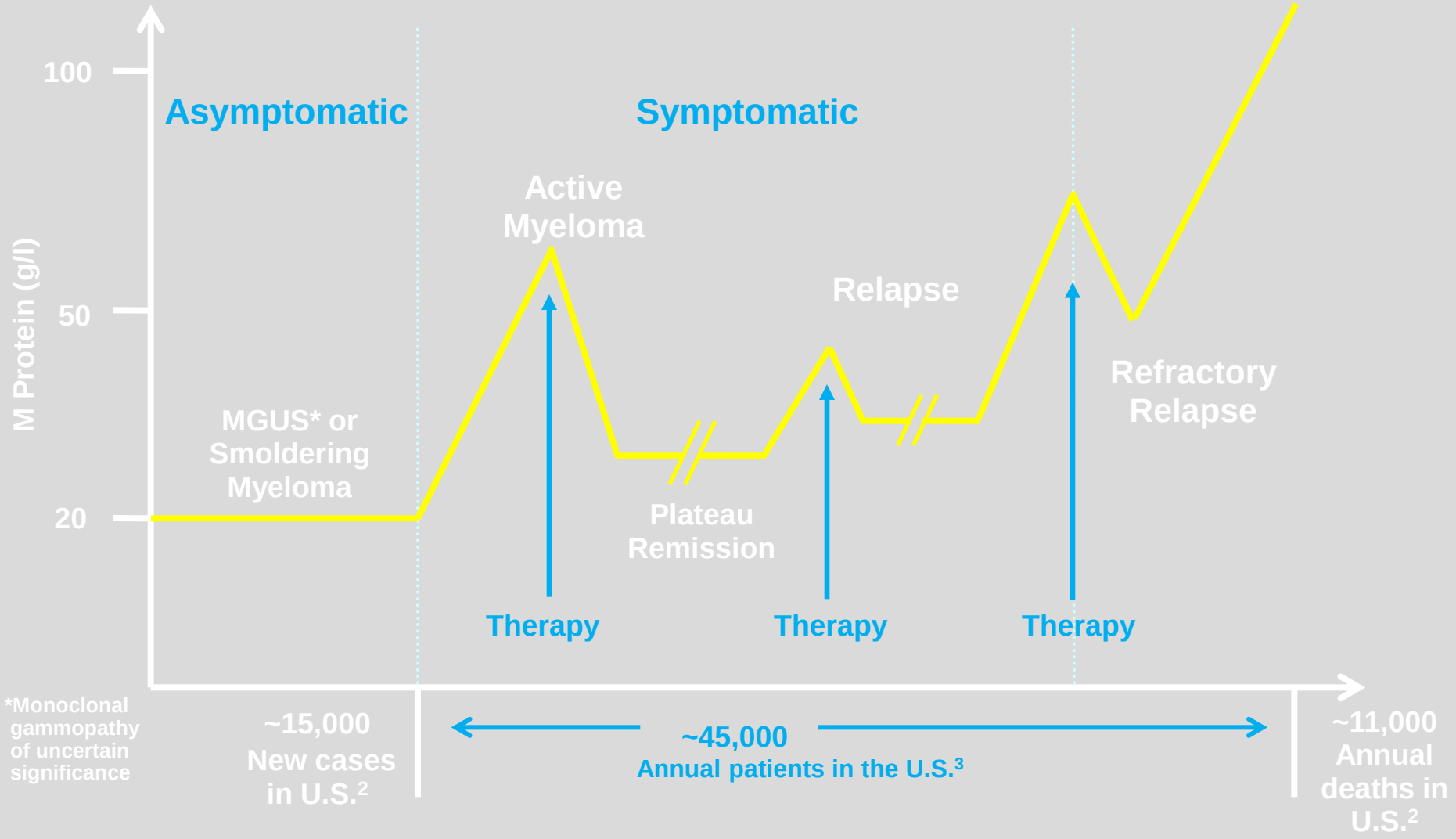
Kyle et al. NEJM 2004

How do we approach therapy?

Approach to Therapy

- **1950-69**
  - Melphalan (Alkeran) first synthesized 1953
  - Glucocorticoids
- **1970-89**
  - Combination chemotherapy
    - vincristine, doxorubicin, dexamethasone (VAD)
  - Barlogie et al. published the VAD regimen in the *New England Journal of Medicine* -- 1984
  - High-dose chemotherapy--1986
  - Stem cell transplantation (SCT)
- **1990-2017**
  - Thalidomide
  - Bisphosphonates
  - VELCADE® (bortezomib) for Injection
  - Revlimid (lenalidomide)
  - Investigational combinations

Multiple Myeloma Disease Paradigm

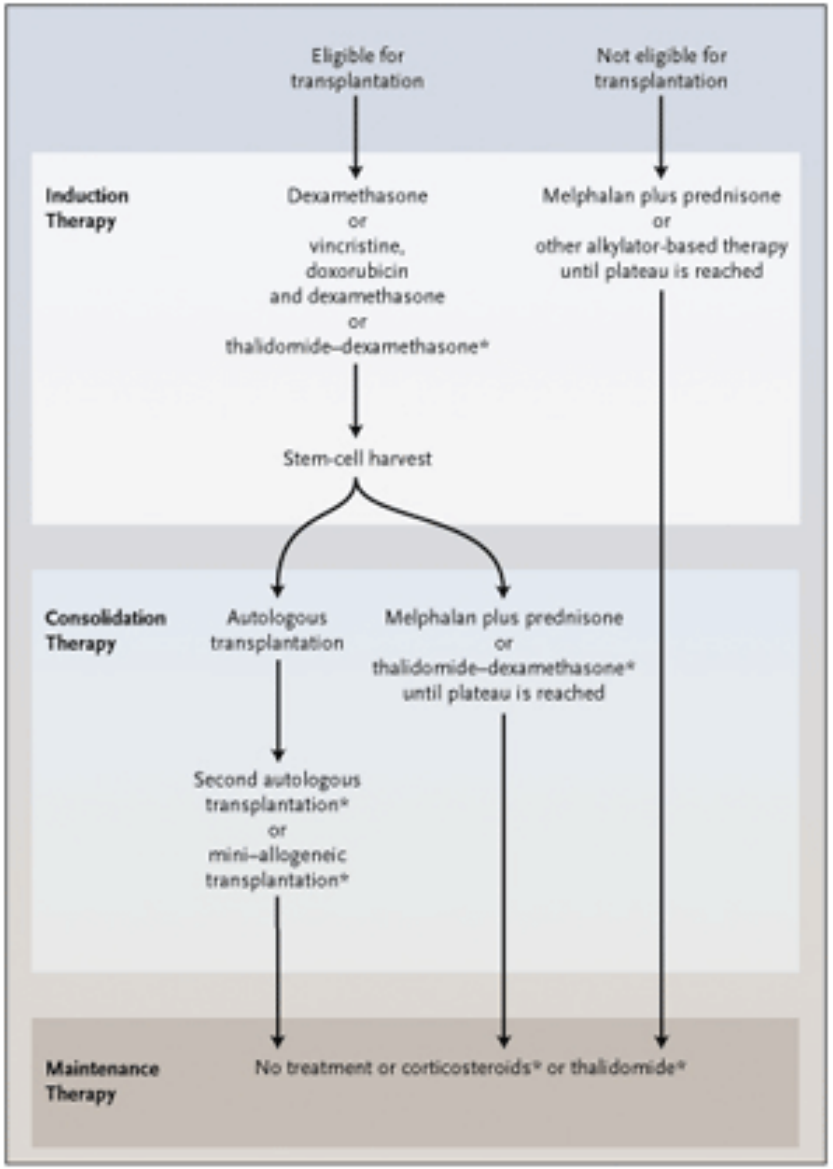


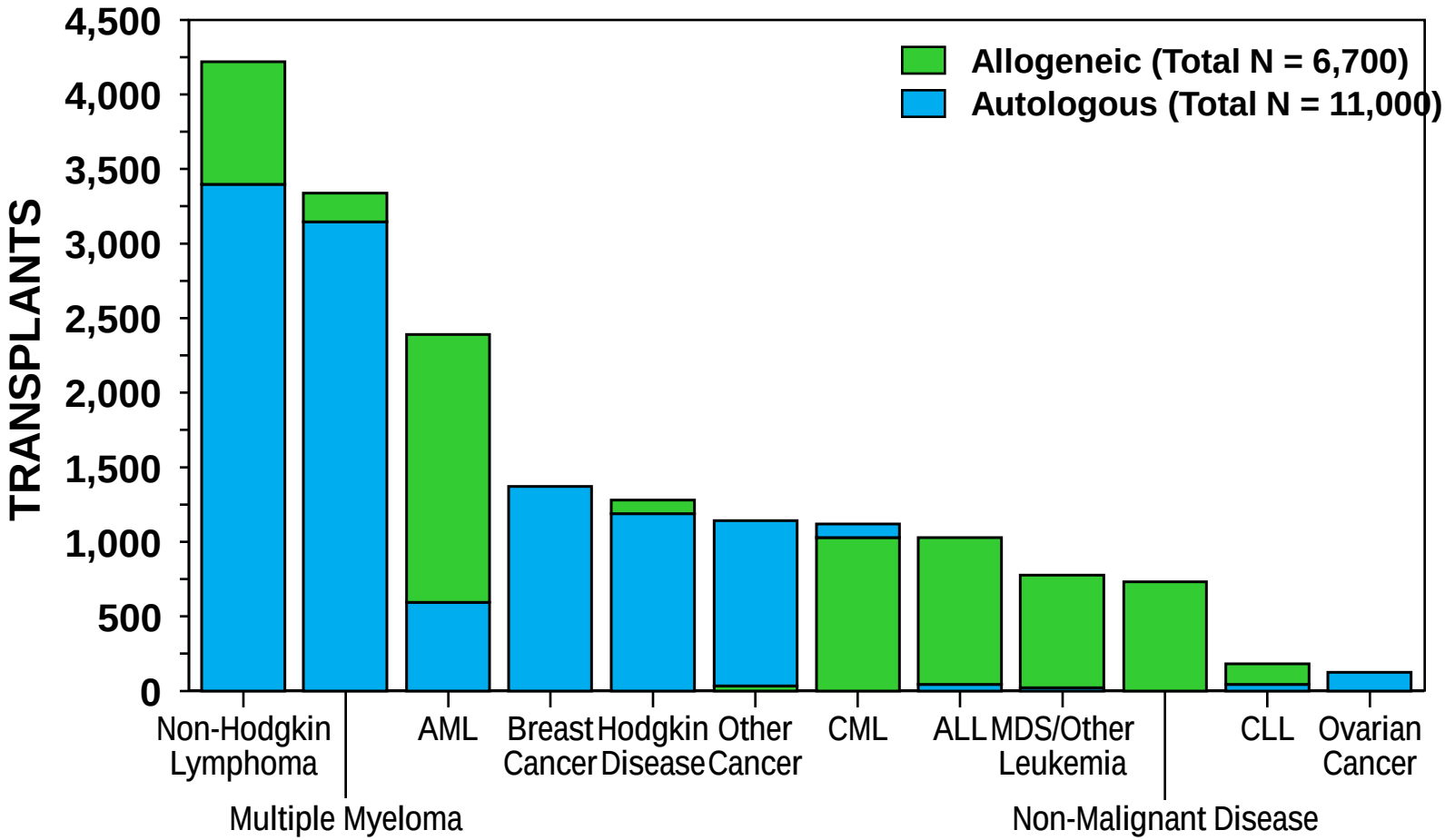
1. Adapted from International Myeloma Foundation; 2001. Reprinted with permission
2. American Cancer Society. Cancer Facts & Figures; 2003

Standard Treatment Paradigm

- Induction
- Consolidation
  - Stem Cell Transplant
- Maintenance
- Outcomes measured /compared:
  - Response rates
  - Overall survival
  - Event-free Survival (EFS)
  - Progression-free Surv (PFS)
  - P-value

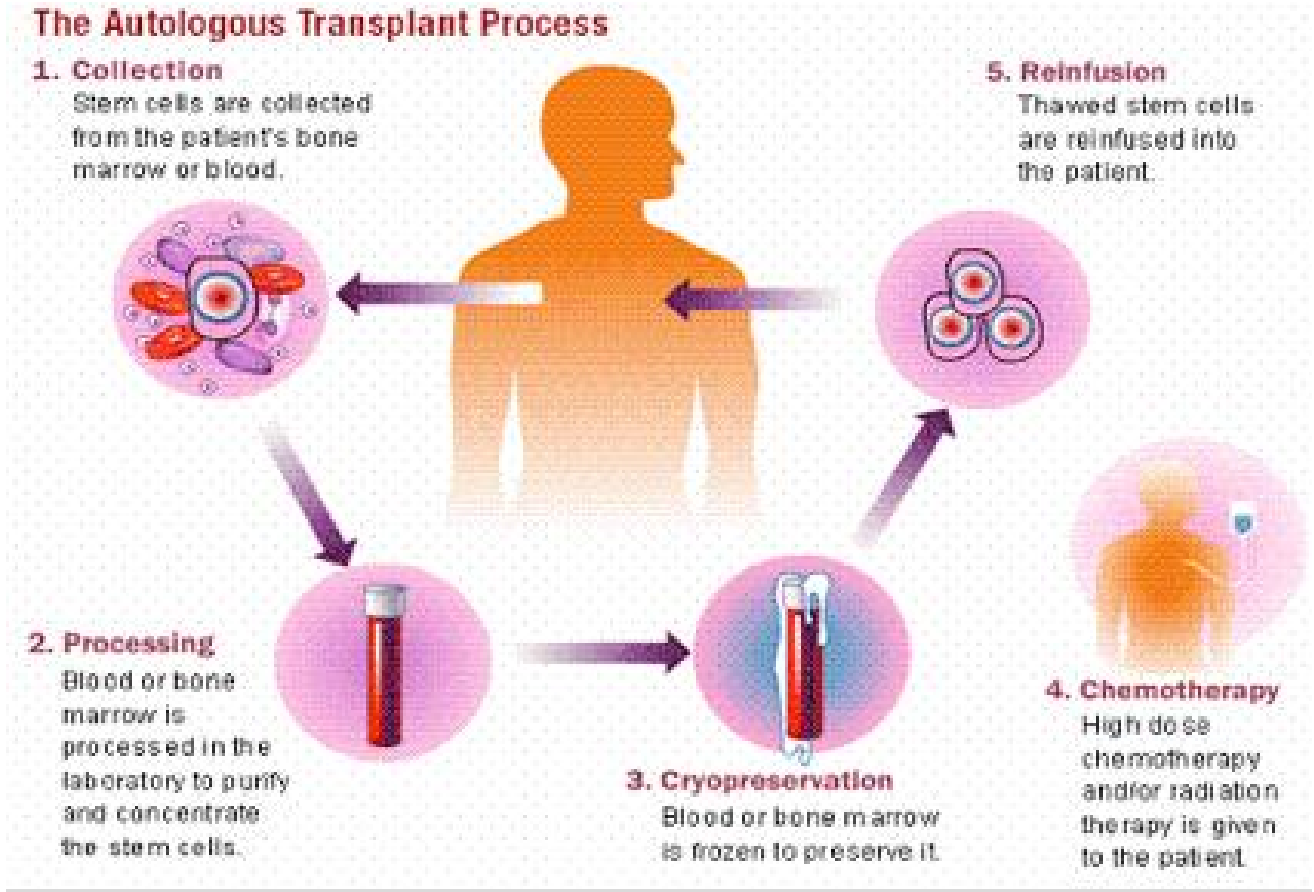
Kyle et al. NEJM 2004




TRANSPLANTS
4,500
4,000
3,500
3,000
2,500
2,000
1,500
1,000
500
0
Allogeneic (Total N = 6,700)
Autologous (Total N = 11,000)
Non-Hodgkin Lymphoma
Multiple Myeloma
AML
Breast Cancer
Hodgkin Disease
Other Cancer
CML
ALL
MDS/Other Leukemia
Non-Malignant Disease
CLL
Ovarian Cancer

Role of Transplantation

- High Dose Therapy with Autologous Stem Cell Support (ASCT)
  - Dose intensity 200mg/m2 melphalan
  - Dose Density
    - “some is good, more is better”
  - Rescue patient with autologous stem cells


The Autologous Transplant Process
1. Collection
Stem cells are collected from the patient's bone marrow or blood.
2. Processing
Blood or bone marrow is processed in the laboratory to purify and concentrate the stem cells.
3. Cryopreservation
Blood or bone marrow is frozen to preserve it.
4. Chemotherapy
High dose chemotherapy and/or radiation therapy is given to the patient.
5. Reinfusion
Thawed stem cells are reinfused into the patient.

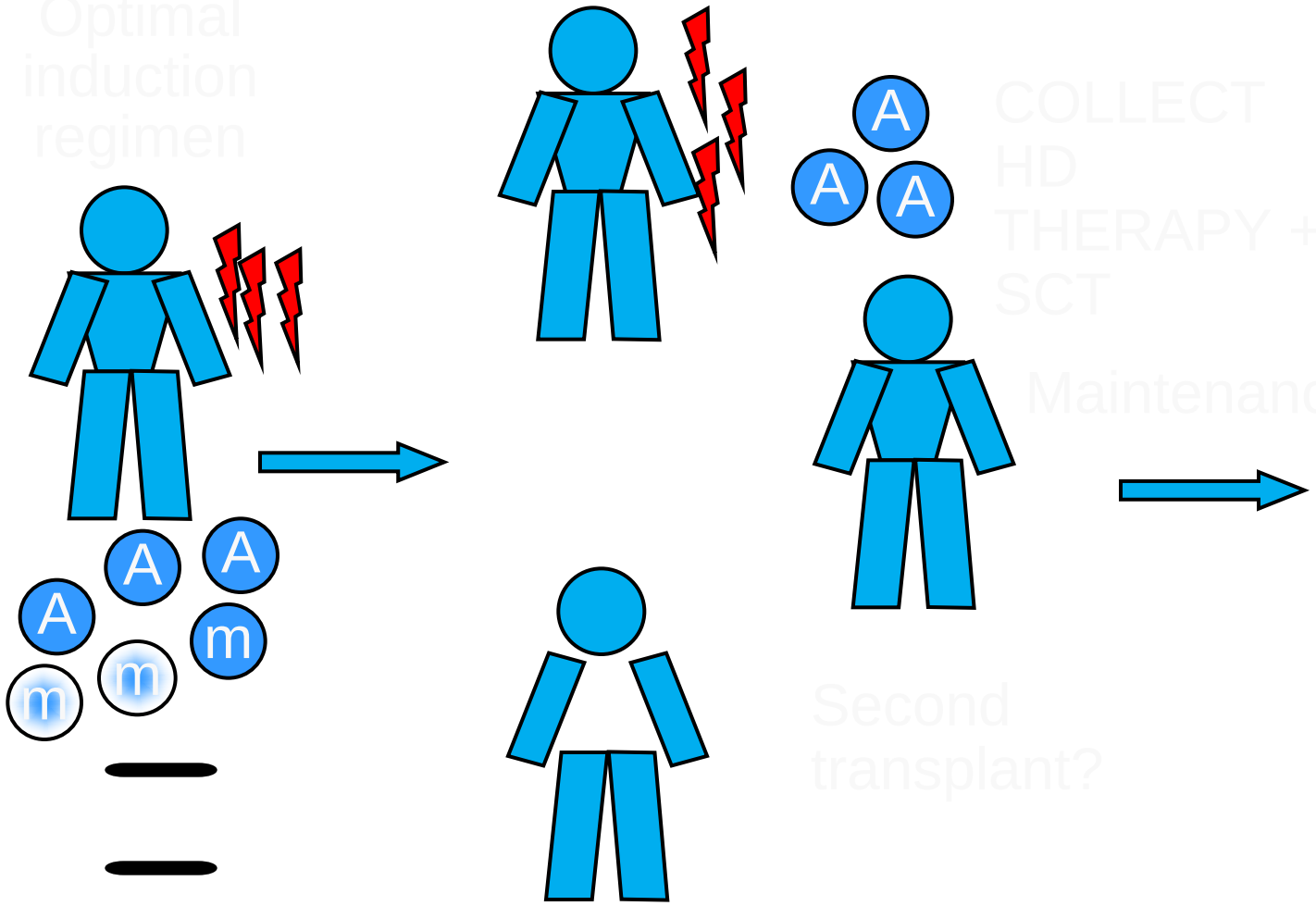
Autologous Transplantation vs Conventional Chemotherapy for Newly Diagnosed Myeloma

		Pts (n)	CR (%)	EFS (mos)	OS (mos)
Barlogie et al	Conventional*	116	–	22	48
	HDT	123	40	49	62
Lenhoff et al	Conventional*	274	–		46% @ 48
	HDT	274	34	27	61% @ 48
Attal et al	Conventional	100	5	18	37
	HDT	100	22	27	52% @ 60
Fermand et al	Conventional	96	–	19	50
	HDT	94	–	24	55
Blade et al	Conventional	83	11	34	67
	HDT	81	30	43	67
Child et al	Conventional	200	9	20	42
	HDT	201	44	32	55

* Historical controls

Fermand J. Blood. 1998;92:3131.
Blade J. Blood. 2001;98:815a.
Barlogie B. Blood. 1997;89:789.
Lenhoff S. Blood. 2000;95:7.
Attal M. N Eng J Med. 1996;335:91.

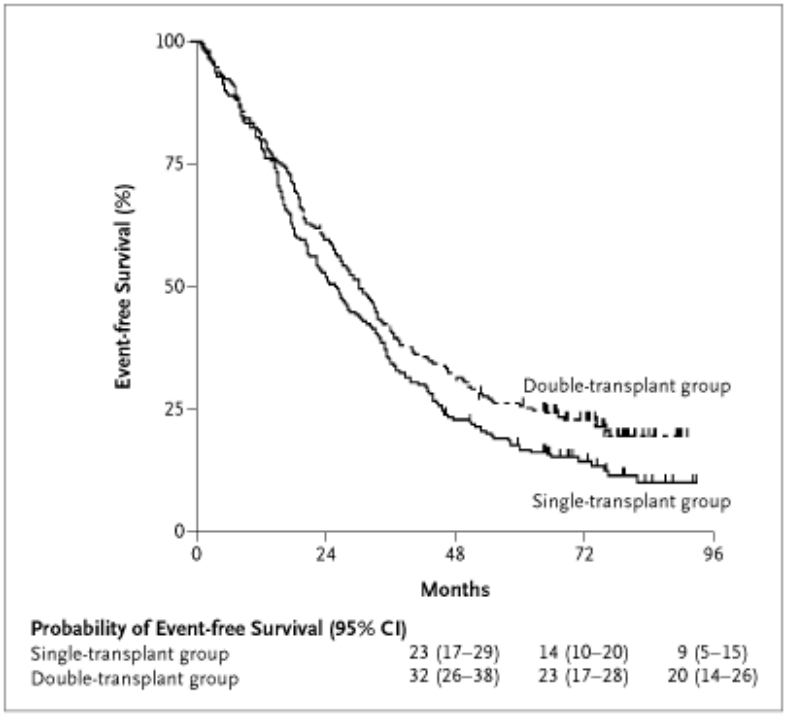
One or Two SCT?

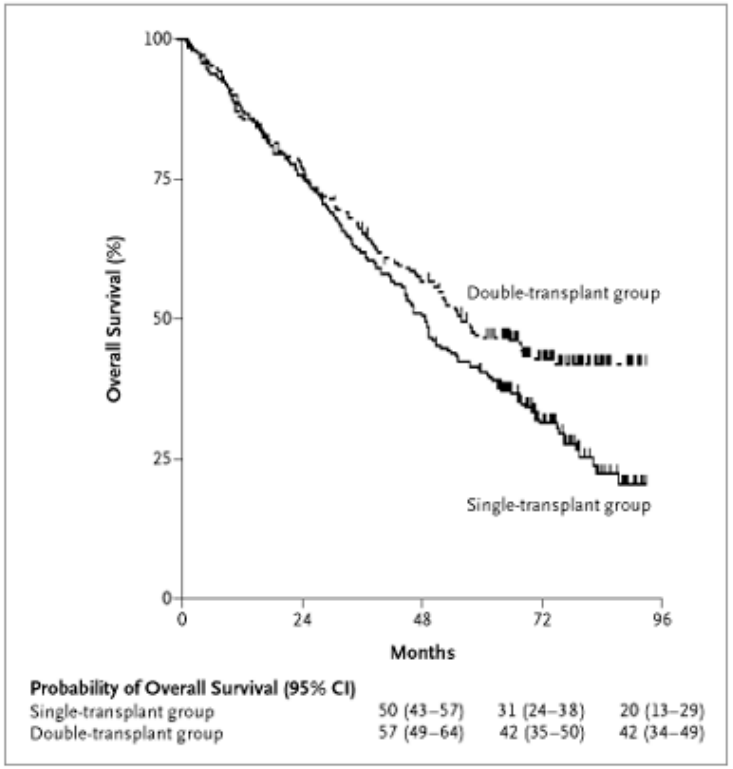


Single or Double Transplant?

IFM 94 Trial Double is superior to single transplant in terms of EFS and OS

Confirmed by the Arkansas Experience : Total therapy II 2006





True only for patients who do not achieve a Cr or VGPR to induction therapy or to the first transplant

Attal et al. 2003 NEJM; 349

Barlogie et al. 2006 Blood 107

ASBMT Guideline and Medical Decision Making

- A dogma was born from this study
- Medical decisions of treating with double transplantation for CR or VGPR patients was established

- Now looking back with Dr. Mazumdar, we see some misuse of statistics were made
- She will expand …..

Deeper look at Attal 2003 NEJM Study

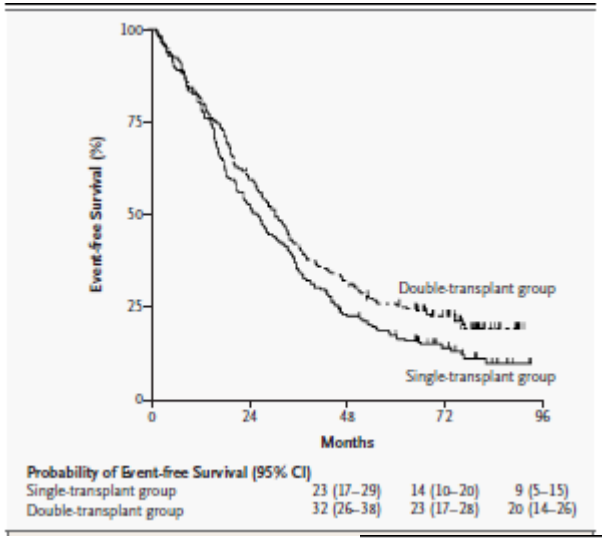
Design: N=399 previously untreated patients <60 years were randomized to single versus double ASCT

Primary Endpoint: Rate of complete response (CR)

Power Calculation: 95% power to determine a 20% difference in CR rate (25% versus 45%)

Intent-to-treat Analysis: Observed diff in overall response rate (ORR): CR or very good partial response (90% reduction of paraprotin); 42% versus 50% (p=0.1)

Conclusion: DT improves OS, specially among those who did not have ORR

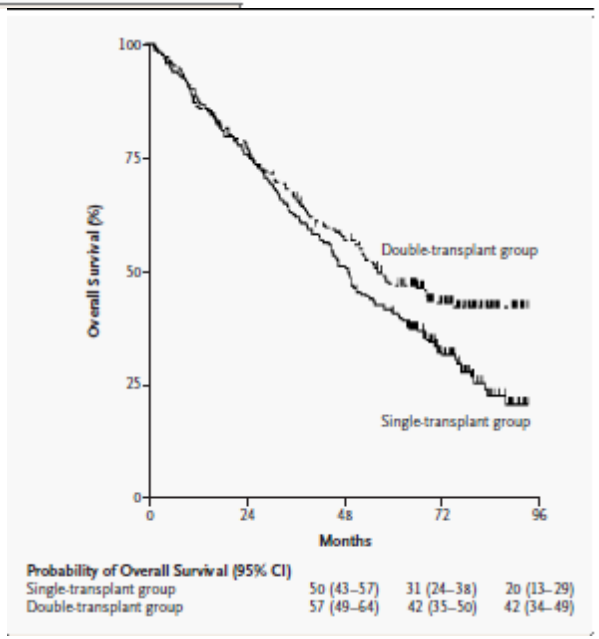


EFS: Time to progression, relapse, or death

7-yr EFS: 10% vs 20% (p=0.03)

OS: Time to death

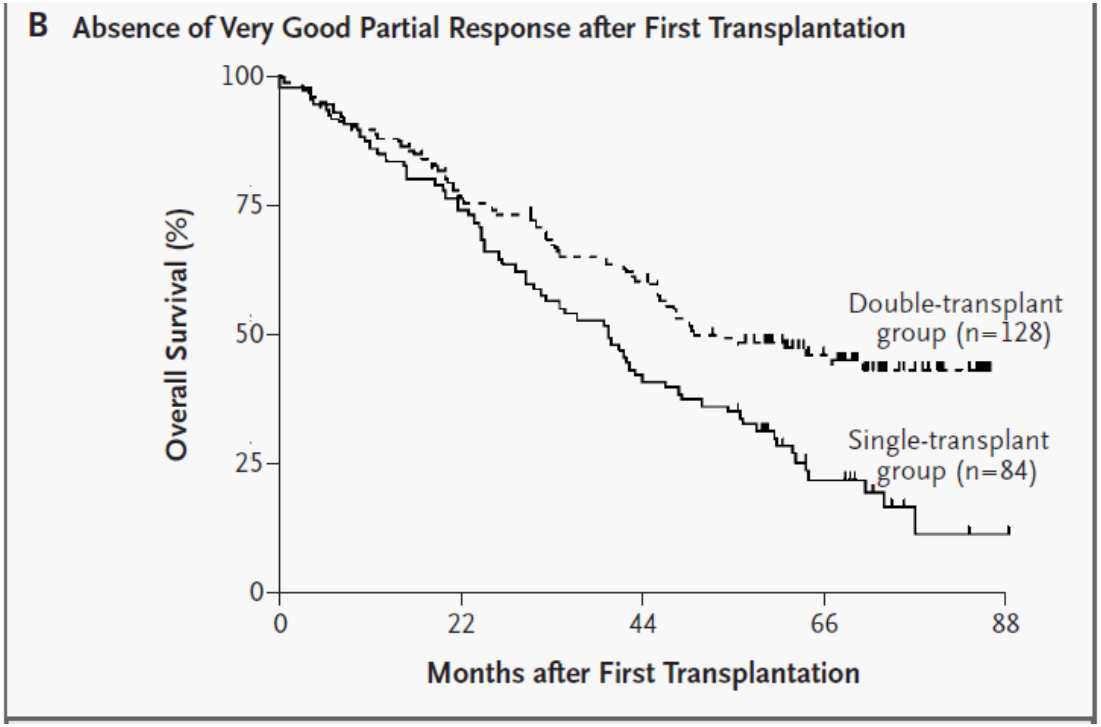
7-yr OS: 21% vs 42% (p=0.01)

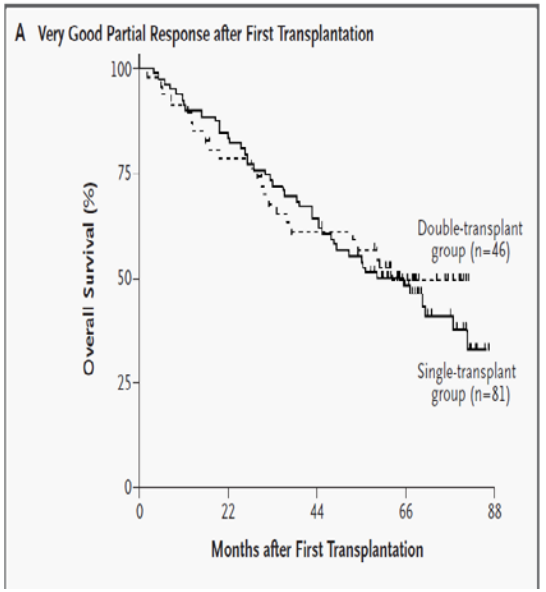


Subgroup analysis: Overall Survival According to Whether Patients had at least a very good partial response (90% decrease in serum paraprotein level)

Results indicate that DT could benefit patients who donot have a good partial response within 3 months after undergoing a single transplantation;

7-year OS rate for Fig B: 11% vs 43% (p< 0.001)





7-yr EFS for fig A:

(p=0.7)

Challenges Associated with conducting and Reporting Subgroup Analysis (Wang, Lagakos, Ware, Hunter, Drazen;NEJM 2007)

- Subgroup analysis comparison was not based on primary endpoint
- Analysis of subgroups not based on interaction test
- Not randomized trial anymore; therefore adjustment of covariates needed
- Survival stratified by treatment Response (Landmark Analysis)

Other issues:

- Multiplicity of testing not adjusted (~ 15 p-values computed in the paper; not clear what was pre-specified)

- Statistical test might have been inappropriate (difficult to say from sparse reporting); binomial proportion test for comparing 7-yr EFS and OS rate or survival analysis based comparison

Single versus Double Auto Transplant for untreated multiple myeloma : Evidence behind ASBMT Guidelines (2015)

Table 7

Prospective Studies Examining Single versus Tandem Auto-HCT

Author	Conditioning Regimen	TRM/ORR	EFS	OS	Level of Evidence
Attal, 2003 [46]	TBI 8 Gy and Mel 140 mg/m^2 versus Mel 140 mg/m^2 followed by TBI 8 Gy and Mel 140 mg/m^2; IFN maintenance offered to all pts	TRM 4% versus 6% ORR 84% versus 88%	Favoring tandem arm; 25 mo versus 36 mo (P = .03)	Favoring tandem arm 48 versus 58 mo (P = .01)	1++
Cavo, 2007 [47]	Mel 200 mg/m^2 d-2 versus Mel 200 mg/m^2 followed by Mel 140 mg/m^2 d-2 and Bu 1 mg/kg PO × 12 d-5-to -3; maintenance IFN offered to all pts	TRM 3% versus 4% ORR NS CR + nCR 33% versus 47% (P = .01)	Favoring tandem arm; 23 versus 35 mo (P = .001)	65 mo versus 71 mo (P = .9)	1++
Sonnevold, 2007 [48]	Mel 70 mg/m^2 i.v. × 2 versus Mel 70 mg/m^2 i.v. × 2; Cy 120 mg/kg i.v. and TBI 9 Gy; maintenance IFN offered to all pts	TRM not stated; ORR 88% for entire group CR 13% versus 32%	Favoring tandem arm; 21 mo versus 22 mo (P = .013)	55 mo versus 50 mo (P = .51)	1++

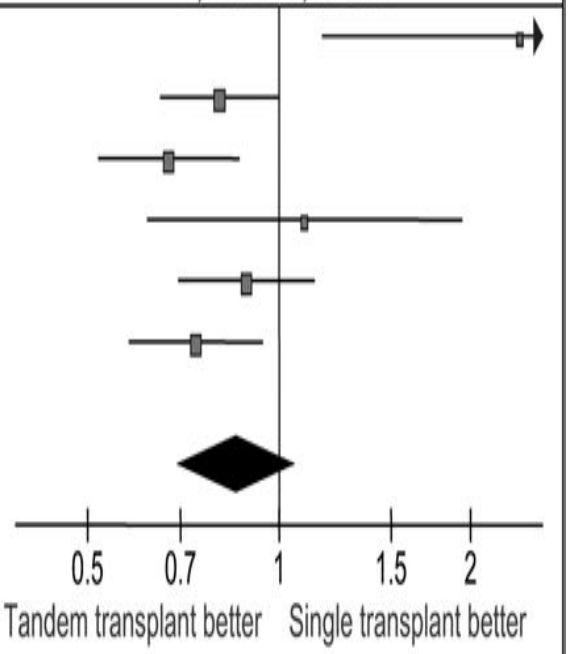
ORR indicates overall response rate; NS, not significant; PO, orally.
TRM, treatment-related mortality.

Forest plot of EFS with tandem vs single transplant for myeloma: Meta-Analyses

Study	Tandem transplant Events	Tandem transplant Total	Single transplant Events	Single transplant Total	Hazard Ratio IV, Random, 95% CI
Abdelkefi, 2008*	22	97	10	98	2.39 [1.17, 4.88]
Attal, 2003	132	200	149	199	0.81 [0.65, 1.00]
Cavo, 2007*	102	158	124	163	0.67 [0.52, 0.86]
Fermand, 2003*	79	114	78	113	1.10 [0.62, 1.94]
Goldschmidt, 2007*	129	180	131	178	0.89 [0.70, 1.13]
Sonneveld, 2007	134	155	139	148	0.74 [0.58, 0.94]
Total (95% CI)		**904**		**899**	**0.86 [0.70, 1.05]**
Total events	597		631		

Heterogeneity: $Tau^2 = 0.04$; $Chi^2 = 13.16$, df = 5 (P = 0.02); $I^2 = 62\%$

Test for overall effect: Z = 1.49 (P = 0.14)



Single versus Double Auto Transplant for untreated multiple myeloma : Evidence of RCT with OS and Meta-Analysis

1. Based on the conflicting data from the prospective randomized trials and the above meta-analyses, there is insufficient evidence to support tandem auto-HCT as the standard of care for myeloma patients.

2. However, there are cases when this may be considered, based in the IFM data (Attal 2003), in patients with less than a very good partial response after a first auto-HCT (grade D) or as part of a clinical trial.

- Grade D Evidence: **Extrapolated result** from a 2+ study (well conducted study with low risk of confounding and bias) or Nonanalytic Stdies; e.g., case reports or case series or expert opinion

Single versus Double Auto Transplant for untreated multiple myeloma : Evidence of RCT with OS and Meta-Analysis

3. It is important to note that in the current era of novel therapies (immunomodulatory drugs (IMiDs) and proteasome inhibitors), the role of up-front tandem transplantation has not yet been decided.

4. A Blood and Marrow Transplant Clinical Trials Network (BMT CTN) 0702 trial, in which one third of 750 patients have been randomized to a tandem transplantation along with novel therapies, may ultimately alter future transplantation algorithms.

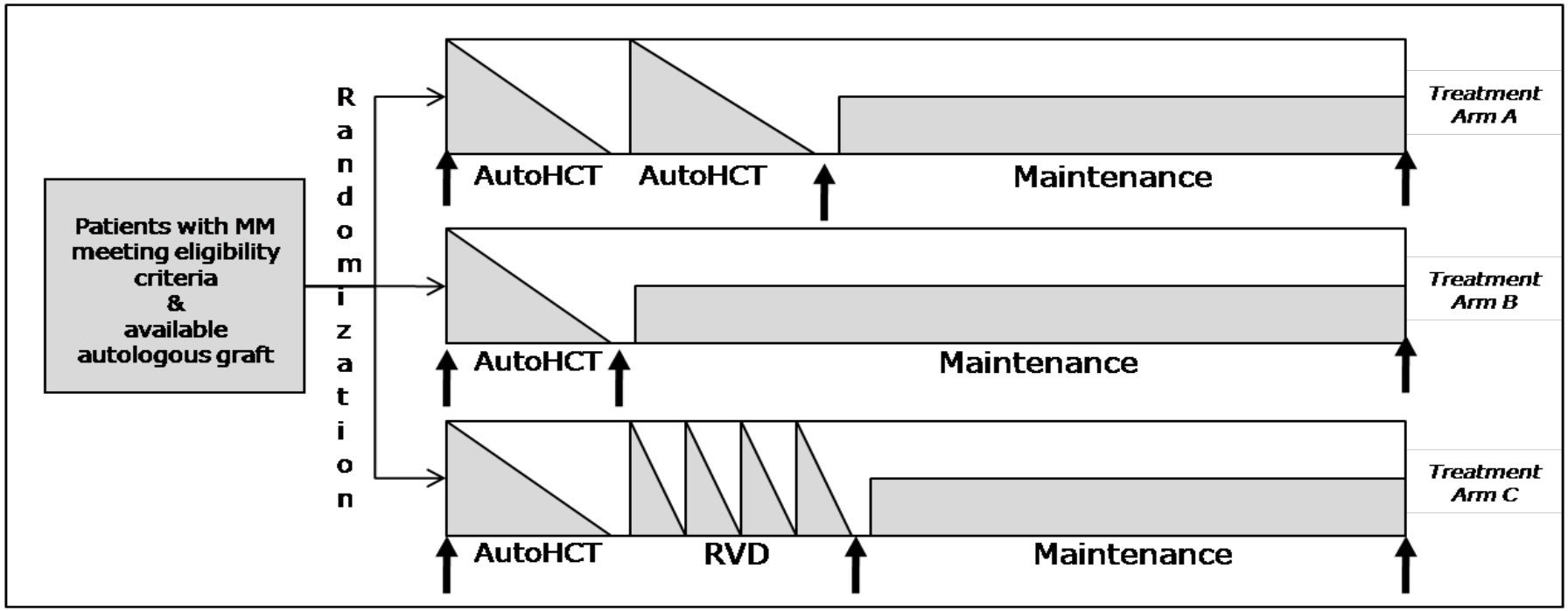
- Seemed like a fair interpretation of p-value in light of study limitations and meta-analysis
- Despite grade D evidence, tandem transplantation for CR and VGPR (#2) continues to be used frequently

- However, the field is self correcting

Ongoing Trial with Novel Therapy is Testing a Third Option and has Planned Subgroup Analysis with Interaction Test

- New Clinical trial BMT CTN 0702 pre-specifies use of interaction test for subgroup analysis

Outline of PRIMeR Ancillary Study



Use of Bayesian Design and Analysis Increasing in this field

- **Phase I/II trial** combining high-dose melphalan and autologous transplant with bortezomib for MM: a dose- and schedule-finding study. Lonial S et al .***Clin Cancer Res* 2010**

- **Durable remission with salvage second autotransplants** in patients with multiple myeloma. Shah N et al. *Cancer* 2012

- **Application of a Bayesian approach to treatment selection** in a rare disease sub-population (case of MM included). Hinsley S et al. from Leeds Inst of Clin Trials Research (abstract presented to SCT, May 2016)

Novel Statistical Methods Developed Using reanalysis of data from MM clinical research:

- **Using Joint Utilities of the Times to Response and Toxicity to Adaptively Optimize Schedule-Dose Regimes**. Thall PF, Nguyen HQ, Braun TM, Qazilbash M *Biometrics 2013*

- **The Bayesian basket design for genomic variant-driven phase II trial**. Simon R, Geyer S, Subramanian J, Roychowdhury S; *Seminars in Oncology, 2017*

- **Use of Bayesian Decision Analysis to Minimize Harm in Patient-Centered Randomized Clinical Trials in Oncology**. Montazerhodjat V, Chaudhuri SE, Sargent DJ, Lo AW; *JAMA Oncol 2017*

Mandatory Training at Institutional level to begin:

- Mount-Sinai School of Medicine is creating **mandated training for all faculty in statistical concepts** through recommendation from 'Task Force on Enhancing Translational Discovery in Biomedical Research'

- Collecting materials and examples to create modules in CITI Program as addendum to recently released Biostatistics course (https://about.citiprogram.org/en/series/fundamentals-of-biostatistics/)

- Webinar on NIH Clearing House on **'Improving Experimental Rigor and Enhancing Data Reproducibility in Neuroscience'**
  - Presented by a laboratory scientist and a quantitative scientist
  - Discusses issues of underpowered study, multiple testing, p-hacking

Recommendation for Statistical Training:

- https://about.citiprogram.org/en/series/fundamentals-of-biostatistics/ **(Instructor: Dr. Seth Schwartz, Professor, Miami Miller SOM)**

- 13 week, 2 classes per week, college level introductory biostatistics course; 45 minutes time commitment for teaching, interactive quizzes and exams

- Course Content: includes topics below and more …
  - Population and Sample - Confidence Intervals
  - Sensitivity and Specificity - Analysis of Variance
  - Probability and Odds - Multiple Regression

- Format amenable to adding modules on other topics easily: prior probability/odds, Bayes factor, Posterior probability, False positive rate; Possibility of conversion with examples from many fields that can be searched); Bayesian design and analysis.

- Might be more effective since this is familiar format for physicians

Supporting Materials for Webinar on underpowered study, multiple testing, p-hacking ….

- **Pre-webinar Information**
  - Before the webinar gets started, read these articles chosen by the moderator for useful background knowledge:

    - Rigor or Mortis: Best Practices for Preclinical Research in Neuroscience (**Steward** and Balice-Gordon, *Neuron,* 2014)
    - Power Failure: Why Small Sample Size Undermines the Reliability of Neuroscience(**Button** et al. *Nature Neuroscience,* 2013)
    - Six Red Flags for Suspect Work (Begley, *Nature,* 2013)
    - Director's Blog: P-Hacking (Insel, NIMH Director's Blog, 2014)

- **Post-webinar Information allows** downloading discussion questions prompting you to think about these issues in other fields and links to an **online calculator for performing a power analysis** using the provided means (mu1 and mu 2) and standard deviation (sigma) for two independent samples.

Speakers:

- **Oswald Steward, PhD** is the senior associate dean for research, professor of neurobiology and behavior, and the director of the Reeve-Irvine Research Center at the School of Medicine at the University of California at Irvine. He has authored several publications that identify problems with replication and reproducibility and outline best practices for pre-clinical neuroscience research.

- **Katherine Button, PhD** is a lecturer in the department of psychology at the University of Bath. She advocates for improving the transparency and rigor of psychological and neuroscience research and has authored high-profile publications focused on reliability of neuroscience research.

In Conclusion ….

- Proper use and Interpretation of p-value is not possible; We tried very hard and failed
- Moving the medical decision enterprise to Bayesian design and analysis needs to be tried with gusto
- Training is needed for physicians (those collaborating with statisticians and those working on their own)
- Further training is needed for statisticians
- Very small funding for statisticians on grants was a problem for getting fully involved in collaborations; making time for bringing this change will be challenging
- However, our institution is showing promise to support moving to right direction

- Please email Madhu.mazumdar@mountsinai.org if you are interested in creating content for CITI Program training

Reference:

- **Single versus double autologous stem-cell transplantation for multiple myeloma.** N Engl J Med. 2003; Attal M et al. InterGroupe Francophone du Myélome.

- **Hematopoietic Stem Cell Transplantation for Multiple Myeloma: Guidelines from the American Society for Blood and Marrow Transplantation.** Biol Blood Marrow Transplant, 2015 Shah, N., Callander, N., Ganguly, S., et al.

- **The challenge of subgroup analyses--reporting without distorting.** N Engl J Med. 2006 Lagakos SW.

- **Statistics in medicine--reporting of subgroup analyses in clinical trials.** Wang R, Lagakos SW, Ware JH, Hunter DJ, Drazen JM. **N Engl J Med**. 2007