

# Clean Sweep for Data



Tanya Pedersen  
February 5, 2007



---

---

---

---

---

---

---

---

## Outline

- ◆ Study Outcomes
  - What do we look for when analyzing data?
- ◆ Data Queries
  - How is data flagged?
  - Types of data queries
  - When does it become my problem?
  - Common data queries



---

---

---

---

---

---

---

---

## Critical Data – Study Outcomes

- ◆ In order to analyze patient outcomes, an indicator and date is needed
  - Indicator: Did the event occur? (Y/N)
  - Date: What is the interval from Tx to event?



---

---

---

---

---

---

---

---

### Critical Data – Study Outcomes

- ◆ Engraftment
  - ANC
  - Platelets
- ◆ GVHD
  - Acute
  - Chronic
- ◆ Survival
  - Overall
  - Disease-free/Leukemia-free



---

---

---

---

---

---

---

---

### Critical Data – Study Outcomes

- ◆ Cause of Death
- ◆ Disease Progression
  - Lymphoma
- ◆ Relapse
  - Leukemia
- ◆ Treatment-related Mortality
  - Death due to transplant, not relapse



---

---

---

---

---

---

---

---

### Other Critical Data

- ◆ Date of diagnosis
- ◆ Date of birth
- ◆ Karnofsky/Lansky
- ◆ Disease stage at transplant
- ◆ Conditioning regimen
- ◆ GVHD prophylaxis



---

---

---

---

---

---

---

---

## How a Data Query is Born

- ◆ Study Data Sets
  - Defined by inclusion/exclusion criteria
  - Data cleaned by statistician
  - Inconsistent/out of range/missing data flagged
  - Database/form checked
  - Sent to CRC for validation/resolution by team



---

---

---

---

---

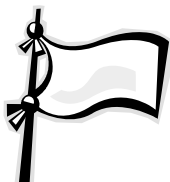
---

---

---

## "Flaggable" Data

- ◆ Missing
- ◆ Out of range
- ◆ Inconsistent



---

---

---

---

---

---

---

---

## Missing Data

- ◆ Question not answered on form
  - Not marked "unknown"
- ◆ Data not collected on previous version form
  - Supplemental forms



---

---

---

---

---

---

---

---

## Out of Range Data

- ◆ Reported value outside validation range in CIBMTR/NMDP system
  - Ex. TBI dose: 100 – 1600 cGy
  - Not marked “verified”
- ◆ Clinical outliers
  - Ex. ANC recovery > 42 days
- ◆ Statistical outliers
  - Too high or too low in relation to the rest of data



---

---

---

---

---

---

---

---

## Inconsistent Data

- ◆ Data doesn't jibe with other data
  - Post-Tx relapse date before Date of Tx
  - Cause of death = CGVHD (or relapse, etc.) but not reported on form
  - AGVHD max grade 4 but organ systems all stage 1



---

---

---

---

---

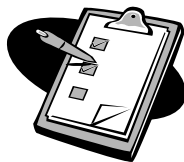
---

---

---

## When Does a Query Become My Problem?

- ◆ Check against other data
- ◆ Check database
- ◆ Check form
- ◆ Check audit records
- ◆ Query to the team



---

---

---

---

---

---

---

---

### Common Data Queries

- ◆ Acute GVHD
  - Onset < 7 or > 180 days post-Tx with no DLI or subsequent Tx
- ◆ Chronic GVHD
  - Onset < 60 days post-Tx
- ◆ Platelet Engraftment
  - > 20K: > 6 months post-Tx
  - > 50K: > 1 year post-Tx



---

---

---

---

---

---

---

---

### Common Data Queries

- ◆ ANC Engraftment
  - > 500: > 60 days if no subsequent Tx
- ◆ GVHD occurrence and no engraftment
- ◆ Others?



---

---

---

---

---

---

---

---

### Why Query Old Transplants?

- ◆ Validation ranges
  - New and improved
- ◆ First time data has been studied
  - Disease-specific
- ◆ Data not previously collected
  - Supplemental form



---

---

---

---

---

---

---

---

## Why Short Timelines?

- ◆ Regulatory reasons
- ◆ Data sent for publication



---

---

---

---

---

---

---

---

## Questions?



---

---

---

---

---

---

---

---