

CHILDHOOD CANCER SURVIVOR STUDY Analysis Concept Proposal

1. Title: Evaluation of cancer treatment data quality: a report from the Childhood Cancer Survivor Study

2. Working Group and Investigators: This study will be carried out within the Statistics and Epidemiology Working Group. Proposed investigators include:

John Whitton	jwhitton@fhcrc.org	(206) 667-6895
Yutaka Yasui	ywasui@fhcrc.org	(206) 667-4459
Les Robison	robison@epi.umn.edu	(612) 626-2902
Ann Mertens	mertens@epi.umn.edu	(612) 626-2187
Pauline Mitby	mitby@epi.umn.edu	(612) 624-9139
John Potter	jpotter@fhcrc.org	(206) 667-4683
Marilyn Stovall	mstovall@mail.mdanderson.org	(713) 792-3240

Possibly:

Joe Neglia	neglia@tc.umn.edu	(612) 626-2778
Dan Green	Daniel.Green@RoswellPark.org	(716) 845-2334

3. Background and Rationale: A key component of the CCSS is the treatment data for each patient. These data are abstracted from each patient's medical records by data managers in the participating institutions. Because of the key role of these data, a system was devised for monitoring the quality of the abstraction process.

During the peak period of treatment data collection, a random sample of five medical record abstraction forms (MRAFs) was re-abstracted per quarter in each institution, for up to five quarters. The re-abstraction forms were sent to the institution's CCSS Principal Investigator, with instructions to have someone other than the original abstractor be responsible for the re-abstraction, if possible.

Data on chemotherapy were reabstracted using the same format as the original MRAF. In the surgery section of the MRAF, a copy of the Operation Notes was sent instead of a re-abstraction of this section, to check on ICD-9 codings. The number of radiation-therapy treatments was reabstracted, but the detailed radiation data were not re-sent to MD Anderson Cancer Center. The radiation data abstracted at MD Anderson Cancer Center were subject to a separate program of reabstraction and quality control, which will be described in this paper.

The reabstracted MRAF data were scanned and edited in the same way as the original data, including all error-checking procedures. The results of the original abstraction and the subsequent reabstraction are therefore directly comparable.

Analysis of the differences between data recorded on the original MRAFs and the reabstracted data will enable us to estimate the overall precision of the CCSS treatment data. The analysis will have wider implications, indicating the general accuracy of cancer treatment data abstracted from medical records.

4. Specific Hypotheses:

1. At least 95% of the original and reabstracted medical-records data are identical.
2. The original data do not systematically under- or over-estimate the amount of treatment (number of surgeries, number of drugs, drug doses).
3. Reporting of general treatment modalities (whether received chemo, surgery or radiation) is reported more accurately than more specific data. Data on whether the patient received particular chemotherapy drugs are more accurate than detailed dose and route information.
4. Accuracy of chemotherapy dose reporting reflects the data quality estimates (Total dose abstracted; total dose, some estimated; incomplete or partial dose) recorded on the MRAF. Data initially reported to be of high quality will have fewer discrepancies than lower-quality estimates.
5. Data quality depends on the number of drugs and surgeries given, but does not depend on cancer diagnosis.
6. The quality of data abstraction improves over time: data collected early in the data-abstraction period will be less accurate than later abstractions.

5. Analysis Framework:

Outcomes of interest:

Differences between original and reabstracted MRAF data:
whether received chemotherapy, radiation, surgery (Yes/No);
number of surgeries;
number of drugs;
specific chemotherapy drugs and routes of administration;
drug doses;
drug dates.

Subject population:

CCSS cases whose medical records were reabstracted

Explanatory variables:

Institution
Diagnosis
Chemotherapy dose quality score (total dose abstracted; total dose, some estimated; incomplete or partial dose; no dose data)
Number of chemotherapy drugs and surgeries
Dates of original abstraction and reabstraction.

Binary (Yes/No) variables will be created for each of the six outcomes of interest. Logistic regression models will be used to examine the relationship between the outcomes and the explanatory variables.

Examples of specific tables:

		Discrepancies in whether received:							
		Chemo-therapy		Surgery		Radiation		Chemo, surg or RT	
Diagnosis	# of MRAFs	N	%	N	%	N	%	N	%
Leukemia									
CNS									
HD									
NHL									
Kidney (Wilms)									
Neuroblastoma									
Soft tissue sarcoma									
Bone cancer									

		Discrepancies in:							
		Drug name		Drug route		Drug dose (any difference)		Drug dose (>20% difference)	
Diagnosis	MRAFs	N	%	N	%	N	%	N	%
Leukemia									
CNS									
HD									
NHL									
Kidney (Wilms)									
Neuroblastoma									
Soft tissue sarcoma									
Bone cancer									

(similar tables for institution and date of abstraction)

			Discrepancies in:			
			Drug dose (any difference)		Drug dose (>20% difference)	
Dose quality	# of MRAFs	# of drugs	N	%	N	%
Total dose abstracted						
Total dose, some estimated						
Incomplete or partial dose						