06-01

CCSS Concept Proposal

1. TITLE: Determining the best comparison group for a cancer survivor study

2. WORKING GROUP/INVESTIGATORS: This proposed publication will be within the Epidemiology/Biostatistics Working Group. The proposed investigators will include:

Anne Jurek	jurek@epi.umn.edu	612-624-1639
George Maldonado	GMPhD@umn.edu	612-626-2104
Ann Mertens	mertens@epi.umn.edu	612-626-2187
Kiri Ness	ness@epi.umn.edu	612-626-2160
Wendy Leisenring	wleisenr@fhcrc.org	206-667-4374
Les Robison	les.robison@stjude.org	901-495-5817

3. BACKGROUND AND RATIONALE: A crucial task when designing an epidemiologic study is selecting an appropriate control group (i.e., comparison (usually unexposed) group). Cancer survivor studies are no exception. Specifically, for the Childhood Cancer Survivor Study (CCSS), siblings were chosen as the comparison group. The choice of using siblings over the United States general population for the comparison group has led reviewers to wonder how different the siblings and general population may be[1-4]. This leads to the question, "Are siblings the best choice for a comparison group for all health outcomes?"

   To answer this question, counterfactual reasoning will be employed[5]. The target population is five-year survivors of childhood cancer who returned the long-term follow-up baseline survey in 1997. The factual events are the health outcomes experienced by the target population, the five-year survivors of childhood cancer.

   The ideal comparison group is comprised of the health outcomes that would have been experienced by the five-year survivors of childhood cancer if instead they had not had childhood cancer. These health outcomes are counterfactual given the factual; therefore, a substitute is needed. The use of a substitute will introduce confounding if the substitute is not what would have been experienced by the target population if the target population had not had childhood cancer. Hence, confounding is a result of using an imperfect substitute. By using an imperfect substitute, uncontrollable factors arise which are called risk factors (i.e., "confounders" associated with both the health outcome and risk factor of interest (e.g., radiation)).

4. SPECIFIC AIMS/OBJECTIVES:
   - Determine the best comparison group for a long-term, follow-up study of cancer survivors.

5. ANALYSIS FRAMEWORK:
   The factual data will be 1997 CCSS data from the childhood cancer survivors. To that data, 1997 data from three sources will be compared to determine which source would be the best substitute for the counterfactual and, consequently, most likely introduce the least amount of confounding. The three comparison groups (i.e., possible substitutes) will be: the siblings of cancer survivors (CCSS data); the Behavioral Risk Factor Surveillance System (BRFSS); and the National Health Interview Survey (NHIS) Sample Adult File. The latter two sources are two United States population data sources. The analysis will be percent-joint distribution analyses of risk factors. Our initial approach for determining which comparison group will be chosen is to calculate the percent-joint distributions of risk factors for each comparison group. Then we will sum the absolute value of differences between the strata of the percent-joint distributions for all strata. Next we will divide that by the number of possible strata. This will give a strata-percent

difference score for each data source. We will then select the comparison group with the smallest strata-percent difference score since it most likely will introduce the least amount of confounding.

The following common risk factors will be analyzed: age, gender, obesity, race, and smoking status. These risk factors have comparable data from all three data sources. The health outcomes that we are interested in are: lung fibrosis (gender and smoking status) [6], cardiovascular disease (BMI, smoking status, gender, race, and age), physical function (age and sex) [7], and obesity (race and gender) [8]. From CCSS baseline questionnaires for both cases and siblings, we will need the following variables to compute the analysis: ccssid, qstkind, parent, d_birth, dcompq, age, sex, height, weight, race, evsmoke, and smokenow.

For example, lung fibrosis is a reported pulmonary condition in the long-term follow-up study[6]. Risk factors for lung fibrosis are gender and smoking status[6]. The first step will be to create a table of percent distributions for the joint distribution of gender and smoking status (See Table 1).

Table 1. Percent-Joint Distribution of Gender and Smoking Status by Group

	Childhood Cancer Survivors				Surveyed Siblings				BRFSS data				NHIS Sample Adult File data			
	Males		Females		Males		Females		Males		Females		Males		Females	
	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N
Never																
Former																
Current																
Total	100	550	100	417	100	620	100	680	100	37,274	100	48,849	100	10,414	100	13,092

Next we will calculate the absolute value of the differences in the percent-joint distributions for each gender-smoking status stratum. Then we will sum these differences for all levels of the risk factors. Finally, we will divide by 6 (3 smoking categories * 2 gender categories) to calculate the strata-percent difference score. The comparison group will the smallest strata-percent difference score will be selected as the best comparison group and should be used for analyzing lung fibrosis.

6. SPECIAL CONSIDERATIONS: This concept proposal focuses on applying counterfactual theory in practice. Dr. Maldonado[5] is "the leading proponent of formal thinking about counterfactuals in epidemiology." Even though counterfactual reasoning is not new, its application in epidemiologic studies is not yet commonly employed. Nonetheless, the goal of this work is to put into practice methodological concepts which can help epidemiologists design better health studies.

References

1. Cuttini et al. Survivors of Childhood Cancer: Using Siblings as a Control Group. Letter. *Pediatrics*. 2003; 112(6 Pt 1):1454-5.

2. Zebrack et al. Reply: Survivors of Childhood Cancer: Using Siblings as a Control Group. *Pediatrics*. 2003; 112(6 Pt 1):1455.

3. Greenberg, ER. Random Digit Dialing for Control Selection. A review and a caution on its use in

studies of childhood cancer. *American Journal of Epidemiology*. 1990; 131(1):1-5.

4. Goldstein, AM, Hodge, SE, and Haile, RW. Selection Bias in Case-Control Studies Using Relatives as the Controls. *International Journal of Epidemiology*. 1989; 18(4): 985-9.

5. Maldonado and Greenland. Estimating Causal Effects. *International Journal of Epidemiology*. 2002; 31:422-9.

6. Mertens et al. Pulmonary Complications in Survivors of Childhood and Adolescent Cancer. *Cancer*. 2002; 95(11):2431-41.

7. Ness et al. Limitations on Physical Performance and Daily Activities among Long-Term Survivors of Childhood Cancer. *Annals of Internal Medicine*. 2005; 143:639-47.

8. Oeffinger et al. Obesity in Adult Survivors of Childhood Acute Lymphoblastic Leukemia: A Report from the Childhood Cancer Survivor Study. *Journal of Clinical Oncology*. 2003; 21(7):1359-65.