PHC6711: Measurement in Epidemiology & Outcomes Research
Assignment #1
Due: Jan 16, 2015

Please read the article, titled" Epidemiologic study of residential proximity to transmission lines and childhood cancer in California: description of design, epidemiologic methods and study population" (*SO: Journal of Exposure Science and Environmental Epidemiology (2015) 25, 45–52*) and answer the following questions.

1. What is the research objective of this study?

2. What is the true primary exposure in this study?

3. What is (are) the exposure variable(s) used to measure the true primary exposure?

4. What information did the authors use to develop the variable(s)?

5. Please discuss the limitations and strengthens of the selected surrogate measures of exposure regarding dose (duration, frequency and intensity) and time of exposure?

Journal of Exposure Science and Environmental Epidemiology (2015) 25, 45–52

www.nature.com/jes

ORIGINAL ARTICLE

Epidemiologic study of residential proximity to transmission lines and childhood cancer in California: description of design, epidemiologic methods and study population

Leeka Kheifets[1], Catherine M. Crespi[2], Chris Hooper[3], Sona Oksuzyan[4], Myles Cockburn[5], Thomas Ly[6] and Gabor Mezei[7]

We conducted a large epidemiologic case-control study in California to examine the association between childhood cancer risk and distance from the home address at birth to the nearest high-voltage overhead transmission line as a replication of the study of Draper et al. in the United Kingdom. We present a detailed description of the study design, methods of case ascertainment, control selection, exposure assessment and data analysis plan. A total of 5788 childhood leukemia cases and 3308 childhood central nervous system cancer cases (included for comparison) and matched controls were available for analysis. Birth and diagnosis addresses of cases and birth addresses of controls were geocoded. Distance from the home to nearby overhead transmission lines was ascertained on the basis of the electric power companies' geographic information system (GIS) databases, additional Google Earth aerial evaluation and site visits to selected residences. We evaluated distances to power lines up to 2000 m and included consideration of lower voltages (60–69 kV). Distance measures based on GIS and Google Earth evaluation showed close agreement (Pearson correlation >0.99). Our three-tiered approach to exposure assessment allowed us to achieve high specificity, which is crucial for studies of rare diseases with low exposure prevalence.

Journal of Exposure Science and Environmental Epidemiology (2015) **25,** 45–52; doi:10.1038/jes.2013.48; published online 18 September 2013

Keywords: childhood leukemia; electromagnetic fields; geographic information system; childhood cancer; power lines; case-control study

INTRODUCTION

Leukemia is the most common malignant disease among children.[1,2] Childhood leukemia has been extensively researched over the past several decades, leading to a significantly increased understanding of the disease. Although this has resulted in tremendous therapeutic improvements and radically improved overall and disease-free survival for childhood leukemia cases, disease etiology remains mostly unexplained.[3,4]

The association between exposure to extremely low frequency (ELF) magnetic fields (MF) and childhood leukemia has been studied since 1979 when Wertheimer and Leeper[5] reported an association between wire codes and childhood cancer, including leukemia. Since then, more than 30 epidemiologic studies with various study designs, exposure measurements and methods have been conducted to investigate the association between residential exposure to ELF MF and childhood leukemia.[6] Several pooled analyses have found an approximate doubling of childhood leukemia risk for relatively high estimated residential MF exposure.[7–9]

Draper et al.[10] in 2005 reported an association between childhood leukemia and distance from the home address at birth to the nearest high-voltage (mostly 275–400 kV) overhead transmission line in the United Kingdom. Compared with children who lived more than 600 m from a high-voltage line at birth, children who lived within 200 m had a relative risk of 1.69 (95% CI: 1.13–2.53); those who lived between 200 and 600 m had a relative risk of 1.23 (95% CI: 1.02–1.49).[10] As the MF are very small at distances beyond 50 m,[11] the apparent risk extended to a distance greater than would be expected if fields from the high-voltage lines were a causal agent. Therefore, the study results did not appear to be fully compatible with the extensive literature on ELF MF and childhood leukemia. Among suggested explanations of note was the dependence of the association on the selected control subjects rather than on the leukemia cases themselves, suggesting the potential for control selection bias in the study.[12] The study by Draper et al. 2005[10] found no association between residential distance to power lines and childhood cancers of the central nervous system (CNS).

The current paper presents a detailed description of the study design, method of case ascertainment, control selection, exposure assessment and data analysis plan of a study that attempts to replicate and extend the study of Draper et al.[10] in a different geographical location. As in the study by Draper, and unlike many other studies focused on the address at diagnosis, we focus on birth addresses.

[1]Department of Epidemiology, UCLA Fielding School of Public Health, Los Angeles, California, USA; [2]Department of Biostatistics, UCLA Fielding School of Public Health, Los Angeles, California, USA; [3]Enertech Consultants, Campbell, California, USA; [4]Department of Epidemiology, UCLA Fielding School of Public Health, Los Angeles, California, USA; [5]Department of Preventive Medicine, University of Southern California, Los Angeles, California, USA; [6]Department of Preventive Medicine, University of Southern California, Los Angeles, California, USA and [7]Environment Department, Electric Power Research Institute, Palo Alto, California, USA. Correspondence to: Dr. Leeka Kheifets, Department of Epidemiology, UCLA Fielding School of Public Health, 71-279 CHS 650 Charles E. Young Drive South, Los Angeles, CA 90095-1772, USA.
Tel.: +1 310 825 6950; +1 310 2796568. Fax: +1 310 206 6039.
E-mail: kheifets@ucla.edu
Received 30 December 2012; revised 1 May 2013; accepted 12 June 2013; published online 18 September 2013

METHODS

Case Ascertainment and Control Selection

The California Cancer Registry (CCR), a statewide population-based cancer registry, was used to obtain information on all childhood leukemia and CNS cancer cases diagnosed between 1988 and 2008 among children who had been born in California and were younger than 15 years of age and residing in California at the time of diagnosis. Cancer incidence reporting is mandatory in the state of California, and the CCR is the agency responsible for registering incident cases of cancer. The CCR meets or exceeds all standards of the Surveillance, Epidemiology and End Results (SEER) Program and North American Association of Central Cancer Registries (NAACCR) for the timely and complete reporting of incident cancer cases in the state, exceeding 99% coverage of incident cancers. Variables recorded by the CCR include date of diagnosis, date of birth, race/ethnicity, sex and residential address at the time of diagnosis as well as information on cancer subtypes and characteristics.[13]

A total of 6645 childhood leukemia cases and 3858 childhood CNS cancer cases meeting our inclusion criteria were identified from the CCR. Cases were linked to the California Birth Registry (CBR; California Department of Public Health, Vital Statistics Branch). Linkage was successful in 87.1% ($n = 5788$) of leukemia cases and in 85.7% ($n = 3308$) of CNS cancer cases. Three groups of control subjects, with birth dates between 1986 and 2007, were selected for the included cases from the CBR. A primary control (control no. 1) for each case was randomly selected from the CBR and matched to the corresponding case (1 to 1) on date of birth (± 6 months) and sex. For each case born during years for which the residential address of the mother was electronically available in the birth certificate records (1997 and later), two additional controls were selected from the CBR using a two-stage cluster sampling design. One of these additional controls (control no. 2) was matched to the case (1 to 1) on date of birth (± 6 months), sex and birth address zip code. The second of the additional controls (control no. 3) was matched to the primary control (1 to 1) on date of birth (± 6 months), sex and birth address zip code. Controls were eligible for inclusion in the study only if they had not been diagnosed with any type of cancer in California before the time of the diagnosis of the corresponding case. Variables obtained from the CBR, for both cases and controls, included mother's residential address at the time of birth, date of birth, sex, race of parents, ethnicity of parents (Hispanic/non-Hispanic), birth weight, maternal and paternal ages, maternal and paternal education, source of payment for delivery, history of pregnancy terminations, number of live births living and deceased, and maternal complications during pregnancy. The specific variables collected by the CBR changed over the time period of the study such that certain variables (e.g., maternal education) were collected only in certain years.[14]

Of the 5788 successfully linked leukemia cases, 4721 (82%) were acute lymphocytic leukemia cases. Table 1 provides overall numbers of cases and controls. Table 2 shows demographic characteristics of cases and primary controls. There were more male than female cases for both leukemia and CNS cancer. The median age at diagnosis was 3 years for leukemia and 4 years for CNS cancer. About half of the cases and primary controls were of Hispanic origin, and about 80% were White (Hispanic or non-Hispanic).[14,15]

There was substantial missingness on some variables (Table 3). This missingness was due largely to differences in the information collected on birth certificates from year to year rather than non-response. No differences in patterns of missingness were detected between cases and controls.

Residential Address Information

From 1998 onwards, the birth records contained the complete residential address of the mother at the time of birth in an electronic format (street number and name, town/city, and zip code). Before 1997, only the zip code of the home address at the time of birth was available electronically. Year 1997 was a transition year in which only a portion of addresses were computerized. For all cases and controls before 1997, we obtained a paper copy of the birth certificate section containing full address information, and manually entered this information into our database, using double entry to minimize data entry errors. We entered

Table 2. Socio-demographic characteristics of childhood leukemia and CNS/brain tumor cases and primary controls, California birth registry, 1986–2007.

	Leukemia cases, N *(%)*	*Leukemia primary controls,* N *(%)*	*CNS cancer cases,* N *(%)*	*CNS cancer controls,* N *(%)*
All	5788	5788	3308	3308
Child's sex				
Male	3227 (54)	3227 (54)	1769 (53)	1769 (53)
Child's age, years				
<1	415 (7)	407 (7)	347 (10)	332 (10)
1–5	3696 (64)	3681 (64)	1613 (49)	1622 (49)
6–9	1003 (17)	1001 (17)	762 (23)	760 (23)
10–14	674 (12)	687 (12)	586 (18)	591 (18)
Hispanic origin of child				
Yes	3114 (55)	2807 (50)	1385 (43)	1636 (51)
Missing	105	154	77	86
Child's race				
White	4550 (82)	4339 (79)	2564 (81)	2436 (78)
Black	290 (5)	490 (9)	257 (8)	298 (9)
Asian	631 (11)	583 (11)	295 (9)	345 (11)
Other	103 (2)	106 (2)	48 (2)	53 (2)
Missing	214	270	144	176

Abbreviation: CNS, central nervous system.

Table 1. Counts of birth addresses of childhood leukemia and CNS cancer cases and controls.

Diagnosis	*Total,* N	*Residence within 2000 m of transmission line(s) of four largest companies*		*Google Earth validation*	
		Yes	*No*	*Four largest companies*	*Other California utilities*
Leukemia					
Cases	5788	3193	2595	1528	420
Primary controls	5788	3150	2638	1521	402
Additional controls	4434	2406	2028		
CNS/brain					
Cases	3308	1814	1494	683	237
Primary controls	3308	1790	1518	700	241
Additional controls	2420	1314	1106		
Total	25 046	13 667	11 379		

Abbreviation: CNS, central nervous system.

Table 3. Summary of missingness on potential control variables, cases and primary controls.

Variable	*Leukemia cases and primary controls, percent nonmissing*	*CNS/brain tumor cases and primary controls, percent nonmissing*
Child race	96% (11092/11576)	95% (6296/6616)
Source of payment for delivery	85% (9854/11576)	85% (5642/6616)
Father's education	71% (8168/11576)	69% (4551/6616)
Mother's education	42% (4819/11576)	41% (2684/6616)
Birth order	99.89% (11564/11576)	99.89% (6609/6616)
Birth weight	99.98% (11574/11576)	100% (6616/6616)
Maternal age	99.98% (11574/11576)	99.98% (6615/6616)
Paternal age	94% (10928/11576)	94% (6218/6616)

Abbreviation: CNS, central nervous system.

Table 4. Accuracy levels for geocoding of residential addresses at birth, childhood leukemia, and CNS cancer cases and controls.

Geocode matching geography type	N	*(%)*	*Median area of matched geography*
Street/street segment	16 718	71.9	3859.8 m^2
Parcel	3852	16.6	697.6 m^2
Zip code	2313	10.0	313.2 km^2
City	264	1.1	171.9 km^2
Other geography (e.g., county and state)	102	0.4	42 396.7 km^2

Excludes birth addresses for the year 1997, which was a transitional year between paper and electronic birth records at the California Birth Registry and had inconsistent electronic address recording.

the information exactly as it appeared on the birth certificate, including entries such as "unknown" and "99999" (used by some counties as a code for missing zip codes) to ensure that the manually entered data matched the format and completeness of the electronically available records (1998 onwards).

All records for birth and diagnosis addresses were geocoded using the University of Southern California (USC) Geographic Information System (GIS) Laboratory's open-source geocoder, which uses parcel level data for Los Angeles County and street level data for the whole of California. Details on the geocoding practice, reference data, and procedures can be found in previous publications.[16] For the purposes of the current study, we recorded geocoded latitude and longitude, and the geographic feature of the geocoded match, which depended on the completeness of the address. The geographic features matches were tax-assessor parcel (the boundary of an inhabited property or dwelling) centroid, street segment (most often a specific block of a street) centroid, street centroid, and US Postal Service (USPS) ZIP Code Tabulation Area (ZCTA) centroid (USPS ZIP codes correspond to a collection of mail delivery routes; ZCTAs are generalized geographic representations of USPS ZIP code service areas),[17] town/city centroid, county centroid, or state centroid for those with unknown addresses). These geographic objects exhibit spatial heterogeneity from region to region in terms of size, shape, and distribution in a manner that defies simple summarization of their "typical" characteristics. For example, ZCTAs shrink as the number of mail recipients increases, resulting in compact geometries in dense urban areas and large sprawling ones in rural areas. The geocoder utilized exploits these variations to improve geocoding accuracy.[16]

The area of the matching geography in the geocoding process (e.g. tax-assessor parcel, street or street segment, and so on) was calculated for each record. For street segments that are curvilinear, this represents the total area encompassed by the bounding polygon for that segment (as outlined in Goldberg et al.[16]). The median area measure for all records in each geocode matching type is presented in Table 4.

Latitude and longitude coordinates were obtained for the majority of addresses with a single pass of the geocoder. About 8% of addresses were reconciled using an interactive process: the address was passed to a satellite imaging/aerial photography software package that provides a "best guess" of the actual location. The technician then corrected the address as needed to obtain an exact latitude/longitude address from the aerial photograph. This resulted in parcel or street segment level accuracy for 88.5% of the geocoded addresses (Table 4).

Exposure Assessment

All aspects of exposure assessment were conducted blinded to case-control status of the subjects. On the basis of the results of an earlier feasibility study,[18] we adopted a three-tiered approach to ascertain distance from the home address to nearby overhead transmission lines: distance based on the electric power companies' GIS databases for transmission lines within 2000 m of a residence; additional Google Earth aerial imagery evaluation for transmission lines within 200 m of a residence; and site visits for residences of leukemia cases and primary controls that had overhead transmission lines close enough to produce MF above background levels (see Figure 1 for voltage by distance criteria). Underground lines were not included in the study. Figure 1 provides an overview of the distance evaluation protocol for leukemia cases and primary controls. Distance evaluation for CNS cancer cases and primary controls, as well as for the additional controls for leukemia cases was similar except that site visits were not conducted.

The four largest electric power companies provide electric service to over 85% of California residences and own about 85% of total linear kilometers of transmission lines above 69 kV (Figure 2).[19] Another 43 small utilities service the remaining residences. The four major electric power companies have reliable GIS systems that cover all transmission lines and some lower-voltage lines owned and operated by them. For the current study, all transmission lines over 100 kV within 2000 m of a home birth address of any of the included cases and controls were identified within the service territories of the four largest electric power companies. In addition, information on lines with lower voltages (above 60 kV) was also available in the GIS systems of two companies. Initial GIS information from the four largest utilities generated over 30 000 potential intersections of transmission lines within 2000 m of over 13 000 study subject residences (Table 1). Availability of GIS systems for smaller utilities was uncertain and logistically difficult, and thus distances to their lines were determined from aerial imagery (about 7% of cases and controls).

Google Earth aerial photographs were examined to further validate the distances obtained based on the GIS databases of the electric power companies (numbers of subjects provided in Table 1) and also to identify

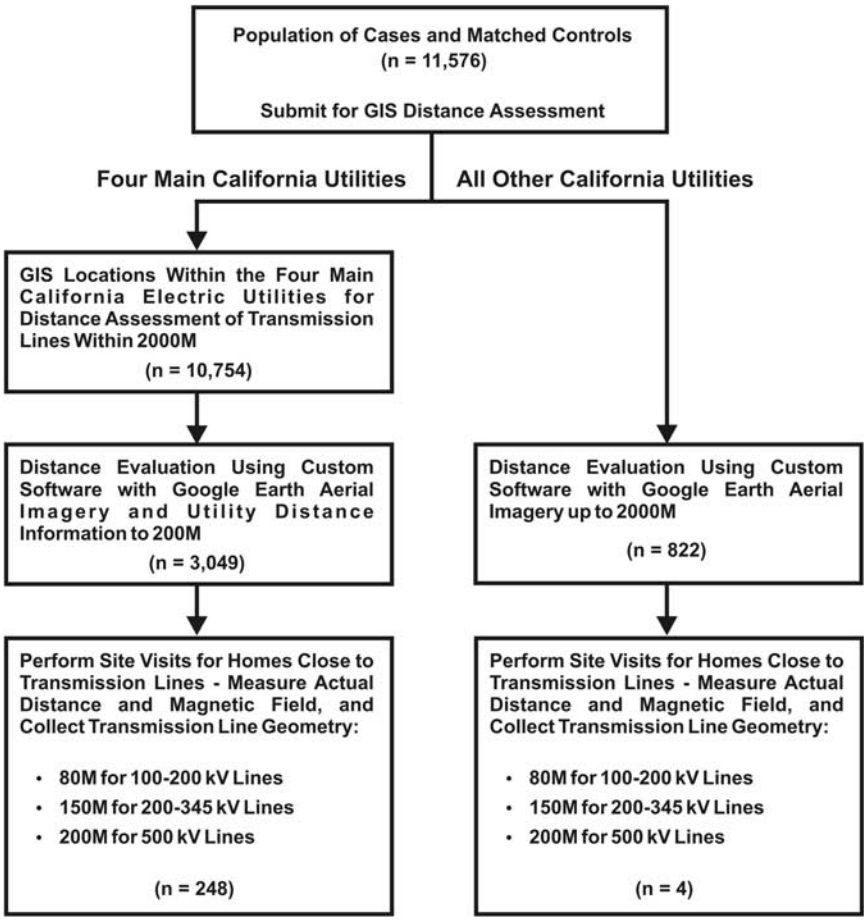

Figure 1. Tiered distance evaluation.

any home addresses outside of the service territories of the four companies that had transmission lines within 2000 m. On the basis of the evaluation of utility data using Google Earth, apparent anomalies in the routing data were observed for one utility; these transmission line routes were reviewed and manual changes to the utility routing were performed when needed.

For all home addresses in the four largest companies' service territories that were within 2000 m of a transmission line and had a geocode accuracy level of parcel, street segment, or street centroid, we determined the distances to company transmission lines using a web-based visualization tool with Google Earth imagery and linear measuring tools. A technician reviewed the aerial photography in the 2000 m buffer around the geocoded residence location, and attempted to identify the intersecting lines. For each identified line, the distance from the geocoded residential address to the nearest point on the line was measured and recorded. If no lines were visible, it was noted, as were any additional lines not identified through the respective electric power company's GIS data.

Higher-voltage transmission lines generally have distinctive right-of-ways (particularly in urban and suburban areas) that allows easy identification of power lines. In rural locations, transmission structures are very visible in open areas, and right-of-ways are again visible in forested or vegetated areas. Lower-voltage lines (for example, 60–138 kV), however, can be more problematic in the urban and suburban areas as they can be routed along main streets and roads which are harder to identify. In rural locations, lower-voltage lines can utilize wooden poles and therefore may be confused with distribution lines. For our study, the majority of residential locations occur within urban and suburban areas, which made identification of transmission lines using Google Earth images very reliable.

For identification of transmission lines in locations where no prior utility information was available, care was taken to ensure that lines were not overlooked. In these cases, lower-voltage lines were often identified and then were later excluded based on Google Street View review as being either of low voltage or beyond distance criteria. In particular, we visually assessed conductor arrangement and relative spacing, insulator length (which can be an indicator of voltage), and ground clearance for several different spans of a power line to make a determination for inclusion or exclusion within our criteria.

In addition, a few residential locations were tested to evaluate our process. We selected a few residences within service areas where line information from the utility was provided for power lines down to as low as 60–69 kV for cross-check. For these residences, site visits were performed to determine whether lower-voltage lines would have been overlooked using Google Earth had they not been identified by the utility. We found transmission lines whenever we expected to find them, and did not find any additional transmission lines that were not previously identified.

Early work using Google Earth to validate all distances to lines for all lines within 2000 m (according to utility GIS information) revealed that when utility information indicated that a line was more than 200 m away from a residence, there was a 96.0% chance (88/2190) that Google Earth validation would confirm the residence as more than 200 m away and a 99.95% chance (2189/2190) that validation would locate the residence as more than 100 m away. Thus, later Google Earth validation work was confined to validating utility distances within 200 m (with confirmation that utility distances of 200–2000 m were beyond the site visit distance criteria). More than 8000 residence-to-transmission line distances, including both the four largest utilities and other services areas and a sample of addresses at diagnosis, were evaluated using aerial photo imagery.

Figure 3 provides a scatterplot of Google Earth distance *versus* utility distance for 3795 residence-to-line observations for which both were obtained (the vast majority of observations were beyond 200 m and precise distance for them was not validated). The two distance measures showed close agreement; the Pearson correlation coefficient was 0.998

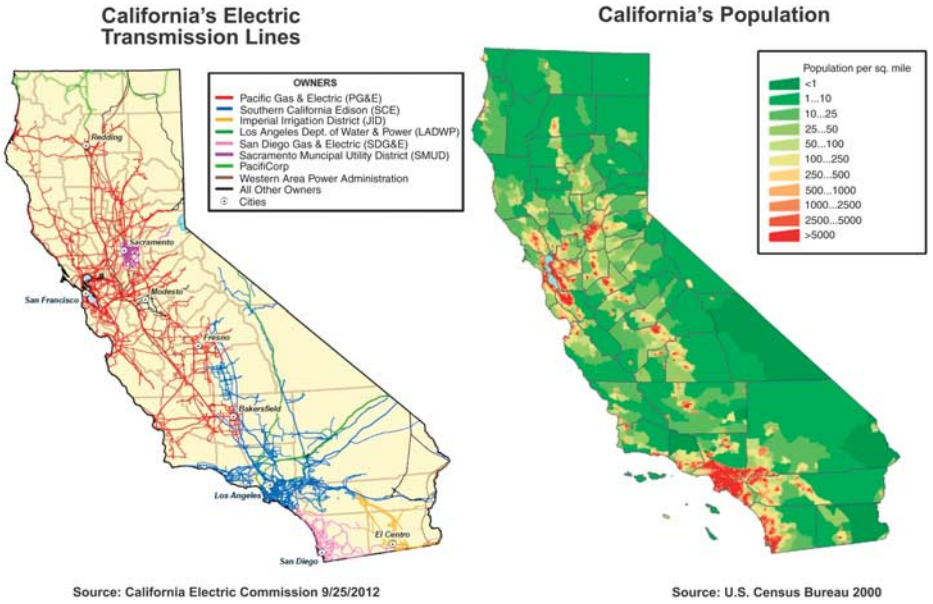

Figure 2. California electric grid and population map.

overall and 0.899 for lines within 200 m (according to utility information). The mean difference (validated minus utility) was -0.6 m, suggesting no systematic upward or downward bias of utility distance information. The two measurements were within ± 3 m for 50% of observations and within ± 35 m for 90% of observations. Further evaluation of outliers showed that many of them were not due to a real disagreement but rather by the presence of foreign lines, that is, lines belonging to a different utility.

We also evaluated the extent to which Google Earth validation of utility-provided distances improved the classification of observations into Draper distance categories (Table 5). Of 1605 utility residence-to-line distance observations within 200 m, 83% (1333/1605) were found to be correctly classified into the Draper distance categories based on Google Earth validation; 6% (103/1605) of utility distances were corrected to be in a closer distance category, and 11% (169/1605) were corrected to be in a farther distance category.

Site visits. Criteria for selection of residences for site visits were formulated based on a pilot study in which we obtained both GIS distance and actual measured distance from a site visit for 25 different residences within each of the four major electric utility service areas.[18] These criteria, based on a combination of distance and line voltage, were designed to ensure that site visits were conducted on all homes for which nearby transmission lines could influence exposure levels; these criteria were: within 80 m of 100–200 kV lines; or within 150 m of 200–345 kV lines; or within 200 m of 500 kV lines. In addition, as site visits were resource intensive, they were limited to leukemia case and primary control birth addresses for which reliable MF calculations would be possible. Thus, site visits were limited to residences with geocode accuracy levels of parcel, street segment, or street centroids that were within the service territories of one of the four major electric power companies.

For each residence, distances from the center of the home, as well as from the closest and farthest points of the home, to the closest transmission line were recorded. For residences located within apartment complexes with no access, exact location of the residence could not be determined and only minimum and maximum distances of the complex to the transmission line were measured. GPS information for each measurement location was recorded.

Magnetic field readings were taken in front of the residence at ground level and at heights of one and two meters. When possible, field readings were also taken underneath nearby transmission lines. The date and time of day were recorded and photographs of the residence, transmission lines, and distribution lines were taken. A sketch of the power line structures with horizontal conductor spacing, vertical conductor spacing, mid-span sag (if available), and circuit direction towards residence were collected for use in magnetic field calculations. These magnetic field measurements will be used for the validation of models for calculated fields.

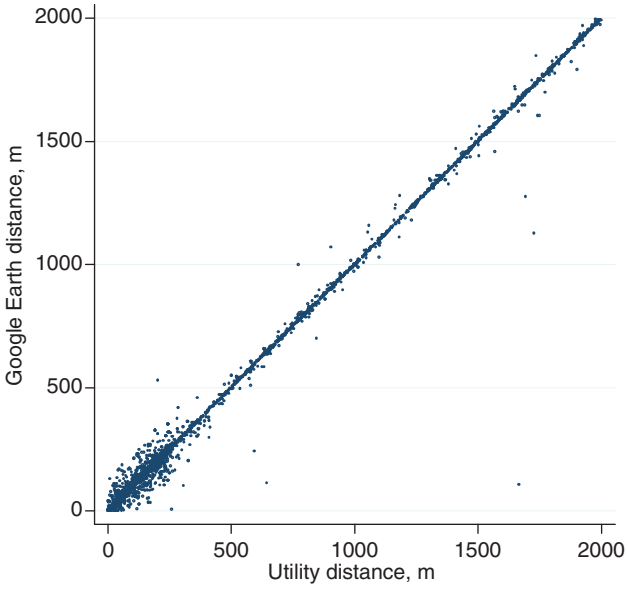

Figure 3. Scatterplot of Google Earth distance *versus* utility distance, $N = 3795$ residence-to-line distance observations.

A total of 297 homes corresponding to birth addresses of leukemia cases and primary controls within the service areas of the four largest utilities were site visited. By design, we wanted to be more inclusive, so that we capture all addresses that might fall within our site visit criteria. As a result, upon in-person site visit, there were 37 instances in which the measured distance of the address to overhead lines was actually farther than the distance criteria for site visits; this occurred mostly because of discrepancies between the location of the actual physical address and location of the geocoded point upon which distances were initially based, as in large apartment complexes or mobile home parks, and/or between the actual locations of power lines and the locations indicated in utility records. For an additional 11 residences, the physical address could not be located and one additional address was not a residence. This left 248 leukemia cases and primary controls within the service areas of the four largest utilities for whom detailed site visit data were collected. In addition, 95 birth addresses (of brain cancer cases and controls as well leukemia cases and controls with addresses near underground lines and/or

Table 5. Original and validated classifications of utility distances into Draper distance categories, $N = 1605$ utility residence-to-line distance observations within 200 m.

Validated distances7	*Utility distances*			
	0–49 m N *(%)*	*50–99 m* N *(%)*	*100–199 m* N *(%)*	*Total* N *(%)*
0–49 m	263 (83)	37 (10)	7 (1)	307 (19)
50–99 m	44 (14)	305 (79)	59 (7)	408 (25)
100–199 m	11 (4)	45 (12)	765 (85)	821 (51)
200–299 m	0 (0)	0 (0)	69 (8)	69 (4)
Total	318 (100)	387 (100)	900 (100)	1605 (100)

lower-voltage lines) were site visited for quality control purposes. An additional six site visits were conducted (four for leukemia cases and primary controls) in services areas of smaller utilities that met site visit distance criteria.

Calculation of magnetic field levels. MF were calculated for all childhood leukemia and CNS cancer cases and primary controls. Residences that did not have a transmission line within 200 m were assigned a magnetic field level of 0 μT. For leukemia cases and primary controls with site visits (all of which were within service areas of the four largest electric power companies), we utilized utility load data (historical loading, phasing, and direction of current flow) and line geometry data based on site visit observations to calculate historic magnetic field levels. For CNS cases and primary controls within service areas of the four largest electric power companies (for which site visits were not conducted, other than a small number for quality control), and for leukemia and CNS cancer cases and primary controls outside of the four largest service areas (again, except few visited for quality control), historic magnetic field levels were estimated based on distance and assumed voltage rating. A look-up table was created for these residences based upon magnetic field results generated from calculation results for residences in the largest four utilities service areas.

Computer modeling was used to calculate historical magnetic field at each residence for the birth date of cases and controls and the date of diagnosis (cases only). Two different types of computer modeling software were used to calculate magnetic field levels, depending upon the complexity of the nearby transmission line circuits. For situations in which there was only one nearby transmission line, or where there were multiple lines but they were parallel to one another, a two-dimensional computer model was used.[20] For situations in which there were multiple non-parallel transmission lines present, or where there were significant changes in the elevation of the terrain, a more complex three-dimensional computer model was required.[21]

In general, more complete information on loading was available for more recent years (2001 through 2007), and estimation was needed for many earlier cases and controls (1986 through 2000). For the majority of cases there is a gap of 1–22 years between birth year and closest year with loading data. When historical load data were not available, meetings were held with utility system planners to estimate earlier loading data based on load data trends. Each transmission line was evaluated based on geographic location with respect to energy generation facilities and customer load centers, changes to historical energy generation, and year of construction for the transmission line in question as well as other significant transmission line construction which may alter load flow patterns. If the electric utility had historical information available on transmission line changes and/or upgrades, this was incorporated into the computer model.

Number of years of extrapolation, historical changes in transmission line geometry, particularly the phasing arrangement on multiple circuits, and load flow could all affect the quality of the calculated historical magnetic field. Availability and quality of such data were coded for use in sensitivity analyses that will consider data quality.

Statistical Analytical Plan

As with any study of this size and complexity, many analytical choices are available. The guiding principles of our analysis plan were to specify primary and secondary analyses *a priori* to focus on analyses that will yield evidence to support or refute an exposure-response relationship or otherwise advance science in this area and reduce the chance of spurious findings.

Given the results of the Draper study, as well as of the pooled analyses, our primary analysis will focus on leukemia. The CNS cancer results will be used mostly to evaluate bias. For primary analyses, we will focus on the data of highest quality: subjects with geocode matches of street; street segments or parcels; best available distance measure; and best calculated field estimate. In addition, the focus of the analysis will be on replicating the Draper findings; the primary distance metric will be the shortest distance to any transmission line 220 kV or above. Secondarily, we will evaluate extensions to the methodology used by the Draper study, such as increased distance to 2000 m, inclusion of lower voltages, consideration of data quality, examination of leukemia subtypes, particularly acute lymphoblastic leukemia (ALL), and the use of multiple controls and other features of our design. Sensitivity analyses will be conducted on the basis of the quality of information for magnetic field calculations.

To compare our findings with the pooled analyses, our primary analysis of calculated MF will be a categorical analysis, with cutpoints similar to the ones used previously. The numbers of subjects with high calculated field exposure levels will likely be low. The choice for the highest cutpoint will depend on the number of cases and controls observed at exposure levels of 0.4 μT and higher. For reliable estimates, we will use the cutpoint of 0.4 μT for the highest exposure category if there are at least five cases and five controls above this level. Otherwise, we will use a cutpoint of 0.3 μT. The reference group will be exposure <0.1 μT. As secondary analysis, we will visualize the exposure-response relationship using regression spline modeling using a continuous exposure variable.

All analyses will be conducted with and without subjects with Down's syndrome, a known strong risk factor for childhood leukemia.[22] We will control for race/ethnicity and socioeconomic status (SES) in all analyses. Proxies for SES available include father's education and payment source for delivery from the birth records as well as census tract SES. Other potential confounders include parity, birth weight, and maternal and paternal ages. In secondary analyses, we will control for a more complete set of potential confounders. Missing data will be multiply imputed under a missing at random assumption,[23,24] a reasonable assumption given that most missingness is due to differences in the variables collected on birth certificates from year to year.

DISCUSSION

Because of continuing health concerns related to transmission lines and power-frequency MF, it is important to validate the Draper study in a different location and population. We have conducted a large, statewide, record-based case-control study of childhood leukemia and CNS cancer in California. The study design was based on extensive pilot work and aimed at minimizing biases and improving exposure assessment. To enhance scientific integrity, we developed an analysis plan that specifies primary and secondary analyses *a priori*.

To the extent possible, we designed a study that maximizes our ability to test the findings of Draper et al.[10] Replication of the study in a different location and inclusion of non-leukemia cases would contribute to the evaluation of possible bias in the Draper study.[12] At the same time, when appropriate, we introduced methodological improvements that should allow us to advance science. For example, for years for which electronic birth records were available, we have selected two additional controls per case, which will allow us to examine an effect of matching on geography. In addition, we have substantially extended the evaluated distances to power lines out to 2000 m and included consideration of lower voltages. In many cases, multiple transmission lines can be present and/or are colocated within common right-of-ways. Our consideration of multiple line configurations in the measurement of distance is another refinement, and could aid in the interpretation of the results. We also plan to examine changes in risk estimates with the degree of accuracy in exposure assessment, which would depend on accuracy of geocoding and availability of power line and historical changes. Finally, the size

of our study may be sufficient to conduct a subgroup analysis for ALL, for which we would expect a stronger association, should one exist.

Case-control studies are prone to bias, and in the area of MF, participation bias has been proposed as the most likely alternative explanation for the observed relationship between MF and childhood leukemia.[25] Our study, as other studies that are entirely based on population-based registries with complete registration of births and cancers, eliminates participation and differential information bias (recall bias). Misclassification of outcome status is also unlikely in our study owing to the completeness and high accuracy of the CCR. One of the strengths of registry-based non-contact studies is the absence of participation bias. Other sources of bias, however, such as possible confounding and differential mobility, could be present and will need to be evaluated.

The role of confounding as an explanation for an observed association between MF and childhood leukemia has been examined extensively, with no known confounder identified.[26,27] We plan to examine and adjust for, as appropriate, a number of potential confounders that are available in registries or in GIS databases, such as ethnicity, perinatal factors, parental education and age, pesticides, and various proxies for SES. Information on other potential confounders, such as viral infections, parental occupation, and diet is unavailable. Thus, residual confounding, although unlikely as these factors are unlikely to be strongly correlated with MF or distance to power lines, nevertheless remains a possibility.

There are some differences between our case and control source populations in terms of residential mobility. Cases must have been born in California and diagnosed in California. Controls must have been born in California and not diagnosed with cancer in California. We do not know whether a control resided in California at the time of diagnosis of the corresponding case. Thus, cases may be more residentially stable than controls and likely to have lived more of their lives in the state than controls. The hypothesis has been advanced that people who live close to power lines might be more mobile, that is, they may be people who move frequently.[28] Mobility might be related to increased exposure to viruses or other infections which could be associated with higher leukemia risk.[29] Thus, mobility is a potential confounder. To address this, we will investigate the association between residential stability and magnetic field exposure in our sample through case-only analysis and by investigating distances over which cases moved between birth and diagnosis. Nevertheless, our exposure assessment is limited to the address at birth and does not capture exposure at other addresses where a child lived before diagnosis.

A potential limitation of the study is missing data on potential confounders. However, as information is missing mainly due to differences in the information collected on birth certificates from year to year rather than non-response, the potential for biases is probably small, and the impact is mainly on the precision of the estimates.

The accuracy of the utility GIS distance information in terms of distance of residences from transmission lines was generally good. Most discrepancies between utility distances and Google Earth validated distances were minor and due to such factors as the spreading out of the locations of lines on maps in order to accommodate labeling of the lines.

When the prevalence of exposure and incidence of disease are both low, as is the case in this study, high specificity (probability that unexposed individuals are correctly classified as "unexposed") is critical. If even a few unexposed individuals are classified as "exposed," this can bias the risk estimate toward the null considerably. Low sensitivity (probability that exposed individuals are correctly classified as "exposed") does not affect the risk estimate nearly as much, although it does affect statistical power by lowering the size of the "exposed" group. Thus, the absence of information on other sources of EMF exposure, such as distribution lines, domestic appliances, and exposures outside of home should not materially bias our results, although it may affect power. Our three-tiered exposure assessment protocol is designed to achieve high specificity, by focusing increasing attention and resources on subjects who appear to be close to lines to ensure that subjects are not classified as close to lines unless this has been carefully confirmed. Our results confirm our ability to achieve high specificity. Inclusion of lower transmission voltages, when available, is important as it will increase statistical power.

Studies on childhood cancer and proximity to high-voltage power lines are underway in Denmark and France as well.[30,31] In addition, the original study by Draper et al.[10] in the United Kingdom is being expanded to include lower voltages. These studies can be pooled with our effort to evaluate for consistency and to aid in the investigation of distance *versus* MF hypothesis.

CONCLUSIONS

This paper describes the methods used in a large case-control epidemiologic study of residential proximity to high-voltage transmission lines and childhood cancers in California. Particular attention was paid to the improvement in exposure assessment and to minimizing, or at least systematically examining, biases that are inherent in any complex case-control study. On the basis of the results of pilot work conducted before the initiation of the study, informed decisions could be made with respect to various design options and methodological choices for the main study. This large record-based study provides for a powerful test on a hypothesized association between distance to power lines and childhood leukemia.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This project was supported by a research contract from the Electric Power Research Institute (EPRI) to the UCLA and by the UCLA Faculty Grants Program. Crespi was also supported by National Institutes of Health grant CA16042. The study was approved by University of California, Los Angeles (UCLA) Office for the Protection of Research Subjects, University of Southern California IRB, and California Committee for the Protection of Human Subjects (CPHS). We thank Kevin Hooper for GE validation and site visit work. We are grateful to participating utilities for time and effort in providing crucial data for this study.

REFERENCES

1 Howlader N, Noone AM, Krapcho M, Neyman N, Aminou R, Altekruse SF *et al. SEER Cancer Statistics Review, 1975-2009 (Vintage 2009 Populations),* Bethesda, MD, USA, 2012.
2 Pui C-H. *Childhood Leukemias*. Cambridge University Press: Cambridge, UK, 2006, 960 p.
3 Greaves M. Childhood leukaemia. *Bmj* 2002; **324**: 283–287.
4 Schuz J, Grell K, Kinsey S, Linet MS, Link MP, Mezei G *et al.* Extremely low-frequency magnetic fields and survival from childhood acute lymphoblastic leukemia: an international follow-up study. *Blood Cancer J* 2012; **2**: e98.
5 Wertheimer N, Leeper E. Electrical wiring configurations and childhood cancer. *Am J Epidemiol* 1979; **109**: 273–284.
6 Kheifets L, Shimkhada R. Childhood leukemia and EMF: review of the epidemiologic evidence. *Bioelectromagnetics* 2005; (Suppl 7): S51–S59.
7 Ahlbom A, Day N, Feychting M, Roman E, Skinner J, Dockerty J *et al.* A pooled analysis of magnetic fields and childhood leukaemia. *Br J Cancer* 2000; **83**: 692–698.
8 Greenland S, Sheppard AR, Kaune WT, Poole C, Kelsh MA. A pooled analysis of magnetic fields, wire codes, and childhood leukemia. Childhood Leukemia-EMF Study Group. *Epidemiology* 2000; **11**: 624–634.
9 Kheifets L, Ahlbom A, Crespi CM, Draper G, Hagihara J, Lowenthal RM *et al.* Pooled analysis of recent studies on magnetic fields and childhood leukaemia. *Br J Cancer* 2010; **103**: 1128–1135.

10 Draper G, Vincent T, Kroll ME, Swanson J. Childhood cancer in relation to distance from high voltage power lines in England and Wales: a case-control study. *Bmj* 2005; **330**: 1290.
11 Kroll ME, Swanson J, Vincent TJ, Draper GJ. Childhood cancer and magnetic fields from high-voltage power lines in England and Wales: a case-control study. *Br J Cancer* 2010; **103**: 1122–1127.
12 Kheifets L, Feychting M, Schuz J. Childhood cancer and power lines: results depend on chosen control group. *Bmj* 2005; **331**: 635.
13 The California Cancer Registry. Mission Statement & Purpose Sacramento, CA [2009], Available from http://www.ccrcal.org/aboutthecr.html.
14 Oksuzyan S, Crespi CM, Cockburn M, Mezei G, Kheifets L. Birth weight and other perinatal characteristics and childhood leukemia in California. *Cancer Epidemiol* 2012; **36**: e359–e365.
15 Oksuzyan S, Crespi CM, Cockburn M, Mezei G, Kheifets L. Birth weight and other perinatal factors and childhood CNS tumors: a case-control study in California. *Cancer Epidemiol* 2013; **37**: 402–409.
16 Goldberg D, Cockburn M. Improving geocode accuracy with candidate selection criteria. *Transactions in GIS* 2010; **14**: 149–176.
17 United States Census Bureau. Geography: Zip Code Tabulation Areas 2012 [cited 2012 April 2], Available from http://www.census.gov/geo/reference/zctas.html.
18 Electric Power Research Institute. Pilot Study for the Replication of the Draper Study of Leukemia, Brain Tumors, and Distance to Power Lines in California. Palo Alto, CA: EPRI 2008, Contract No.: 1014939.
19 O'Neill Ean editorCalifornia's Transmission Challenges for Interconnecting Renewables 2011 Mach 29.
20 Bonneville Power Administration (BPA). Description of Equations and Computer Program for Predicting Audible Noise, Radio Interference, Television Interference, and Ozone from A-C Transmission Lines 1977 September, Report No.: Contract No.: ERJ-77-167.
21 Electric Power Research Institute (EPRI). EMF Workstation 2012. 3420 Hillview Avenue, Palo Alto, CA 2012, Contract No.: 1023794.
22 Lange B. The management of neoplastic disorders of haematopoiesis in children with Down's syndrome. *Br J haematol* 2000; **110**: 512–524.
23 Rubin DB, Schenker N. Multiple imputation in health-care databases: an overview and some applications. *Stat Med* 1991; **10**: 585–598.
24 Little R, Rubin D. *Statistical Analysis with Missing Data*, 2nd Edition ed (Wiley: New York, NY, USA, 2002.
25 Mezei G, Kheifets L. Selection bias and its implications for case-control studies: a case study of magnetic field exposure and childhood leukaemia. *Int J Epidemiol* 2006; **35**: 397–406.
26 Greenland S, Kheifets L. Leukemia attributable to residential magnetic fields: results from analyses allowing for study biases. *Risk Anal* 2006; **26**: 471–482.
27 Langholz B. Factors that explain the power line configuration wiring code-childhood leukemia association: what would they look like? *Bioelectromagnetics* 2001; (Suppl 5): S19–S31.
28 Jones TL, Shih CH, Thurston DH, Ware BJ, Cole P. Selection bias from differential residential mobility as an explanation for associations of wire codes with childhood cancer. *J Clin Epidemiol* 1993; **46**: 545–548.
29 Sahl JD. Viral contacts confound studies of childhood leukemia and high-voltage transmission lines. *Cancer Causes Control* 1994; **5**: 279–283.
30 Bessou J, Deschamps F, Figueroa L, Cougnaud D. Methods used to estimate residential exposure to 50 Hz magnetic fields from overhead power lines in an epidemiological study in France. *J Radiol Protection* 2013; **33**: 349–365.
31 Pedersen C. Distance from residence to nearest power line and the risk of childhood leukaemia: a population-based case-control study in Denmark. *Childhood Cancer 2012*. Children With Cancer UK: London, UK, 2012.