

resided or was treated and checked for coding errors that might explain discrepancies in date of birth or SSN.

Statistical analysis

Vital status on January 1, 2002 was updated from NWTs follow-up records in April, 2005. The updated cohort of subjects known to have died between 1979 and 2001 was used to determine which of the following factors were associated with an NDI match: sex, ethnicity (non-Hispanic Caucasian, African American, Asian, Hispanic, unknown/other), Social Security Number (SSN) on file, year of death, country of birth (United States (US), foreign or unknown), age at death, geographic location of treatment institution (northern border, southern border, non-border state), institutional compliance (the fraction of cases from the registering institution in active follow-up) and NWTs recorded cause of death (tumor, toxicity or infection, toxicity with viable tumor present, other/unknown). Each factor was examined separately and then all were examined together using multiple logistic regression. Records submitted to the NDI due to unknown survival status as of June 2003, but which when updated in 2005 showed survival beyond that date, were used to assess the specificity of the NDI matches.

Survival estimates based on information acquired through standard NWTs follow-up procedures were compared with those based on NWTs follow-up augmented with results from the NDI search. All analyses used a closing date of January 1, 2002. Deaths prior to 1979 were included in the analyses so as to be consistent with other reports from the NWTs and provide accurate overall survival estimates. All survival analyses were stratified by NWTs study number into the "early" era (NWTs-1,2) and "modern" era (NWTs-3,4) to account for changes in survival which occurred mainly due to substantial therapeutic advances made over the course of the clinical trials. For the standard analysis, which used only the updated 2005 NWTs information, subjects lost to follow-up before the closing date had their records censored. For the augmented analysis, subjects last known alive prior to closing who had no NDI match had their date last seen updated to January 1, 2002, whereas those who had a match were considered to have died on the date supplied by the NDI.

Because we did not submit the entire NWTs cohort to the NDI we could not directly compare the results of a mortality study using an NDI search alone with mortality based on active follow-up through the NWTs. As a representation of what might have happened had the NWTs halted follow-up and relied solely on the NDI results, a third analysis was preformed in which subjects known to the NWTs to have died after 1978 but who were not matched through the NDI were considered to have been alive on January 1, 2002. For all analyses, overall survival

percentages and 95 percent Confidence Intervals (CIs) were estimated at 15 years from diagnosis by actuarial methods [7,28].

Actuarial methods were also used to calculate 10 year percentages and 95% CIs of loss to follow-up by the NWTs among subjects enrolled prior to January 1, 1990. The policy of the NWTs is to start tracking procedures once a subject has been out of contact for two years. The selection of a cut-off date twelve years prior to the study closing date allows each subject to have had the potential for at least ten years of follow-up.

To determine how well passive follow-up through the NDI could substitute for active follow-up, the number of deaths ascertained through the NDI search was compared with an expected number based on NWTs mortality rates. The entire cohort of 6,217 subjects was used to determine numbers of deaths and person-years of follow-up by age, time since diagnosis, gender, study number (NWTs-1-4), stage of disease (I-V), histology (favorable or anaplastic Wilms tumor, CCSK, RTK or other) and the availability of an SSN. The Lexis package [29] developed for the R statistical [30] was used to form the seven dimensional tables. Mortality rates were calculated by dividing numbers of deaths by person-years of follow-up in each cell of the table. For subjects who had their follow-up augmented by the NDI, i.e. those who in June 2005 were still last known to be alive prior to closing and for whom no NDI match was found, a similar table was constructed of augmented person-years at risk from the later of date last seen or January 1, 1979 to closing. Multiplying the mortality rates for the entire cohort by the augmented person-years gave the expected number of deaths in each cell. Observed and expected mortality rates for this period of augmented follow-up were compared by availability of a SSN and other factors. Poisson based p-values for testing whether the observed number of deaths could be adequately explained by expected mortality rates were calculated using Byar's approximation as described by Breslow and Day [31].

Results

A total of 706 NDI matches were established among the 786 subjects known as of June, 2003 to have died between 1979 and 2001. No NDI matches were identified in the group of known decedents with deaths prior to 1979. A further 21 matches were identified from the cohort of 3,406 NWTs subjects with vital status unknown on January 1, 2002. Three of these deaths were also ascertained in April 2005 when NWTs follow-up was used to update records of the 3,406 subjects whose vital status was unknown at the time of the NDI submission. Thus the NDI record search matched 709 of the 789 deaths (89.8 percent) known by the NWTs to have occurred between