Defining AML and MDS second cancer risk dynamics after diagnoses of first cancers treated or not with radiation

T Radivoyevitch, R K Sachs, R P Gale, R J Molenaar, D J Brenner, B T Hill, M E Kalaycio, H E Carraway, S Mukherjee, M A Sekeres, J P Maciejewski

Cite this article as: T Radivoyevitch, R K Sachs, R P Gale, R J Molenaar, D J Brenner, B T Hill, M E Kalaycio, H E Carraway, S Mukherjee, M A Sekeres, J P Maciejewski, Defining AML and MDS second cancer risk dynamics after diagnoses of first cancers treated or not with radiation, *Leukemia* accepted article preview 22 September 2015; doi: 10.1038/leu.2015.258.

Received 10 August 2015; accepted 14 August 2015; Accepted article preview online 22 September 2015

Defining AML and MDS Second Cancer Risk Dynamics after Diagnoses of First Cancers Treated or not with Radiation

Tomas Radivoyevitch[1], Rainer K. Sachs[2], Robert Peter Gale[3], Remco J. Molenaar[4,5], David J. Brenner[6], Brian T. Hill[7], Matt E. Kalaycio[7], Hetty E. Carraway[7], Sudipto Mukherjee[7], Mikkael A. Sekeres[7] and Jaroslaw P. Maciejewski[5,7]

[1]Department of Quantitative Health Sciences, Cleveland Clinic Lerner Research Institute, Cleveland, OH; [2]Department of Mathematics, University of California, Berkeley, CA; [3]Section of Hematology, Department of Medicine, Imperial College London, UK; [4]Department of Cell Biology & Histology, Academic Medical Center, University of Amsterdam, The Netherlands; [5]Department of Translational Hematology and Oncology Research, Cleveland Clinic Taussig Cancer Institute, Cleveland, OH; [6]Center for Radiological Research, Columbia University, New York, NY; [7]Department of Hematology and Medical Oncology, Cleveland Clinic Taussig Cancer Institute, Cleveland, OH.

Running Title: Myeloid 2nd Cancer Risk Dynamics after 1st Cancers

Conflicts of Interest: The authors have nothing to disclose.

Corresponding Author: Tomas Radivoyevitch, Department of Quantitative Health Sciences, Lerner Research Institute, Cleveland Clinic, Desk JJN3-01, 9500 Euclid Ave., Cleveland, OH 44195. Email: radivot@ccf.org, Tel: 216-442-5649, Fax: 216-445-7659.

Word Counts: title 111/150; running title 50/50; abstract 200/200; paper 3981/4000; figures 4, tables 4

Key Words: acute myeloid leukemia; myelodysplastic syndromes; second cancer risks; SEER.

Abstract

Risks of acute myeloid leukemia (AML) and/or myelodysplastic syndromes (MDS) are known to increase after cancer treatments. Their rise-and-fall dynamics and their associations with radiation have, however, not been fully characterized. To improve risk definition we developed SEERaBomb R software for Surveillance, Epidemiology, and End Results second cancer analyses. Resulting high-resolution relative risk (RR) time courses were compared, where possible, to results of A-bomb survivor analyses. We found: 1) persons with prostate cancer receiving radiation-therapy have increased RR of AML and MDS that peak in 1.5-2.5 years; 2) persons with non-Hodgkin lymphoma (NHL), lung and breast first cancers have the highest RR for AML and MDS over the next 1-12 years. These increased RR are radiation-specific for lung and breast cancer but not for NHL; 3) AML latencies were brief compared to those of A-bomb survivors; and 4) there was a marked excess risk of acute promyelocytic leukemia in persons receiving radiation-therapy. Knowing the type of first cancer, if it was treated with radiation, the interval from first cancer diagnosis to developing AML or MDS, and the type of AML, can improve estimates of whether AML or MDS cases developing in this setting are due to background versus other processes.

Introduction

Second cancers can occur after a first cancer by several mechanisms including: 1) coincidence; 2) genetic predispositions *independent* of therapy; 3) prior environmental exposures to carcinogens *independent* of therapy; 4) carcinogenicity of the therapy; and 5) therapy *dependent* genetic predispositions[1]. The first two are background mechanisms and the latter two are treatment-related. It is not possible to state with certainty that a specific second cancer was caused by therapy of a first cancer. Using information at the population level, however, the odds of this can be approximated by relative risks (RR), defined as observed/expected cases. Knowledge of the first cancer type and the time interval between the cancers refines RR estimates. Knowing when RR rise and fall and how much could guide cancer survivor follow-up decisions and could help genetic studies focus on second cancer cases that are less likely to be coincidental.

Surveillance, Epidemiology, and End Results (SEER) data[2] comprises 3 databases of 9, 4 and 5 registries. Although the most recent database contains only 5 registries, it currently accrues more person-years (PY) at risk than the other two databases combined (Figure 1). The first database, with 9 registries, has the longest follow up and has been the focus of studies of second cancer risks. These studies made risk comparisons between treatment types[3-5] or they compared risks to those of the general population[6, 7]. Our study extends the latter of these studies by using all 3 SEER databases to provide RR time course dynamics over a greater number of time-since-diagnosis intervals. New second cancer analysis functions in our R package SEERaBomb enabled this. Using these functions, comparing SEER RR time courses for acute myeloid leukemia (AML) vs. myelodysplastic syndromes (MDS), for different first cancers, for males vs. females, and for treatment with vs without radiation, we provide estimates of therapy-independent initial RR peaks in the interval of 0 to 0.25 years after first cancer diagnoses, of subsequent therapy-associated RR peak heights and timings after 1.5 years, and of subsequent steady state RR values. Comparisons to A-Bomb survivor analyses[8-10] are made where possible, with the caveat that in the absence of chemotherapy information, SEER radiation-associations do not imply causality and could be indirect, partly or completely, via correlations with chemotherapy.

Methods

Cancer Definitions

Cancer types were defined by International Classification of Disease codes (ICD-9 for non-hematologic cancers and ICD-O3 for hematologic cancers) as described in Supplementary Table S1. Numbers of each cancer type by year (Supplementary File SEERdataOverView.xlsx) show SEER use of ICD-O3 code 9987 for therapy-related MDS (tMDS) fell in 2010 and halted in 2012 (Supplementary Figure S1). Indeed, SEER currently enforces tMDS coding as therapy-related AML (tAML; ICD-O3 9920). As MDS[11] and AML[12] have different mutation spectra, from a carcinogenesis perspective (rather than a subsequent therapy perspective), pooling MDS and AML cases is problematic. To compensate for this to the extent possible, straight lines fitted to male and female tAML cases vs age over 2001-2009 were used to predict tAML cases in 2010-2012. Observed minus predicted tAML cases, for each sex and year, were then reassigned as MDS cases (using random draws of tAML cases without replacement and a fixed random number generator seed for reproducibility, see Supplementary Section S1).

Background incidence rates

To estimate AML and MDS cases expected if risks are at background rates (right branch of Figure 1C), we fitted the following generalized additive model[13] to cases observed using Poisson regression:

$$\text{cases} \sim s(\text{age}) + s(\text{year}) + ti(\text{age}, \text{year}) + \text{offset}(\log(\text{PY})) \qquad (1).$$

Here s() is a 1-D spline and ti() is a tensor interaction term, which is included to control for possible interactions between age and year, to more accurately compute expected cases for person-years (PY) in a particular age-year bin; this term is critical for some second cancers (Supplementary Section S2), and for simplicity, our software fits the same model to all cancers. An implicit default in generalized Poisson models (Eq. 1) is an exponential link, i.e. the right-hand side raised to an exponential is the expected number of cases, which, as exp(log(PY)) = PY, is proportional to PY. Expected AML/MDS cases of the best fitting (i.e. maximum likelihood) model divided by PY are shown as expected

incidence surfaces in Figures 1D and 1E. Such surfaces provide smoothing (local averaging) of observed age-year incidence rates (i.e. cases/PY points in these plots). In these fits cases were summed over first, second, and later cancers. First cancers dominate such sums, so the incidences approximately equal risks in individuals never exposed to cancer therapy. This approximation is exact under the null hypothesis that cancer risks are independent of prior cancer therapies, which is implicit in "expected" numbers of second cancer cases (E) of relative risks, RR = O/E (O is the observed number of cases), and in ratios thereof, $RR_i/RR = O_i/O \times E/E_i$ (i denotes treatment with ionizing radiation). Before fitting models, population PY in the age group 85+ years were redistributed to ages 85.5 to 99.5 using male and female US National Vital Statistic Report mortality rates of 2001 (volume 52, issue 14); this is particularly important for MDS, as 21% of second cancer MDS cases are in the age group 85+.

PY at risk after a first cancer

When a SEER subject is diagnosed with a first cancer, the patient's PY at-risk for a second cancer becomes a strip of time that is diagonally directed across ages and calendar years in a single-year resolution PY matrix that has years as columns and ages as rows. The orientation of the strip is diagonal because each increase of a year of age implies an increase of a year of calendar time. Iteratively for each SEER cancer patient, PY strips add values between 0 and 1 to matrix elements under the strip. For example, a person aged 64.3 years when diagnosed with a first cancer in 2003 and aged 66.8 years when diagnosed with a second cancer, contributes 0.7 PY to the age-year bin (64, 2003), 1 PY to the age-year bin (65, 2004), and 0.8 PY to the age-year bin (66, 2005). Resolution of ages was prioritized over calendar years as incidence typically depends on age more than year. Such PY matrices were generated separately for males and females and for each selected time interval after diagnosis of a first cancer. PY strip start and end ages were calculated as first cancer age-at-diagnoses plus starting and ending times of the time-since-diagnosis interval of interest, clipped by age-at-

diagnoses of second cancers and survival times, whichever came first. For computational efficiency PY strips were summed using C++ (via the R package Rcpp); all other codes were written in R[14].

Time courses of relative risks after a first cancer

Sex-specific background incidences (surface values in Figures 1 D-E and in Supplementary Figure S2) multiplied point-wise into PY matrices of specific time-since-diagnosis intervals were summed over product matrix elements to form expected numbers of second cancer cases (E). This yielded RR = O/E where O = second cancer cases observed for that time interval. RR 95% confidence intervals (CI) were found assuming O is Poisson distributed[15], as qchisq(.025, 2*O)/(2*E) and qchisq(.975, 2*O+2)/(2*E) in R. For ratios of RR, CI were 2.5% and 97.5% quantiles of 5,000 simulations of $(O_i/O)x(E/E_i)$ with O_i and O Poisson distributed with means equal to observed values with and without irradiation (i). For cancers with small numbers of observed cases (e.g. APL), such ratios were unstable due to too many divisions by zero (due to O being small) and are thus not provided (longer time intervals minimize this but result in unacceptable time resolution losses); such RR ratios can, however, still be conceptualized in terms of numerator and denominator RR time courses. RR were plotted at PY-weighted interval midpoints defined by interval start times + PY/cases/2, i.e. into the interval by half the average PY strip length. The average age at risk in an interval was taken as the average of interval start ages + PY/2; such expected ages are compared to the average age of observed second cancers in that interval.

SEERaBomb

R codes applicable to any second cancer were placed in new second cancer analysis R functions in our R package SEERaBomb (Supplementary Section S3 and Supplementary Figure S3). Codes that use these functions to produce the AML/MDS specific results of this report are provided as R scripts (Supplementary Appendix). SEERaBomb adds 0.5 months to survival times to compensate for flooring (e.g. times <1 month are reported as 0), it subtracts 0.5 from months of diagnoses (January is coded as

1), adds 0.5 years to ages in years (these are naturally also floored), and scores first and second cancers arising in the same month as being 0.33 months apart. SEERaBomb was validated using simulated data (Supplementary Figure S4) and by showing in Table 1 that RR in Ref.[6] are similar to SEERaBomb estimates (full tables of emulations of RR in Ref.[6] are provided in Supplementary Files validationMales.xlsx and validationFemales.xlsx and further extensions thereof to more time intervals and use of all 3 SEER databases are given in Supplementary Files males(Flipped).xlsx and females(Flipped).xlsx wherein first cancers label sheets and second cancers rows (or vice-versa).

SEERaBomb's use of all 3 SEER databases (Figure 1) enables higher resolution second cancer risk estimates than the competing software SEER*Stat MP-SIR[2], as the latter does not provide access to SEER registries starting in 2000 (i.e. the most recent SEER database), so it covers <50% of SEER cases since 2000; MDS entry into SEER began in 2001, so accessing the most SEER database is particularly important for MDS RR estimates (Supplementary Figure S5).

Results

Risks in US cancer survivors

AML and MDS RR time courses after diagnoses of non-hematological first cancers treated with radiation (with or without chemotherapy) increase in 9 to 12 months, peak in 1.5-2.5 years, and resolve in 10-15 years (Figures 2A-B). Over 1-12 years (Table 2), RR were higher for AML than MDS, higher for females than males, and higher after first cancers treated with radiation than not; for males not treated with radiation, AML risks were marginally elevated and MDS risks marginally decreased. At steady state (times >12 years) all RR CI included 1 or bordered on it (Table 2), indicating that AML/MDS risks after anti-cancer therapy had resolved to risk levels of the general population. To explore observed sex differences we focused on the two most prevalent sex-specific cancers, breast and prostate cancer. AML/MDS RR time courses after breast (Figure 2C) and prostate (Figure 2D) first cancers were similar to those of females and males after any non-hematological first cancer (Figure 2A-

B). RR time course pairs (with vs. without radiation) in Figure 2C-D were plotted as ratios of RR in Figures 2E-F. Figure 2F shows two peaks of MDS risk associations with radiation after prostate cancer, suggesting there are separate early and late MDS-inducing effects of radiation after prostate cancer. Time courses of average ages of observed and expected cases (Figure 2G) do not support conjectures of age differences between the two peaks but they do support MDS depending on age-driven additional hits more than AML. AML/MDS RRs higher in females than in males may reflect non-hematological first cancers arriving at earlier ages in females (Figure 2H-I), and thus lower background rates, if radiation-associated AML/MDS risks are additive more than relative (Figure 2J). Negative correlations of AML/MDS onset ages and RR across first cancer types (Table 3) support risks not being fully relative, i.e. being at least partly additive.

Risks in Japanese A-bomb survivors

A-bomb survivor AML RRs expected after a total body dose of 1 Sv were estimated for 13 time-since-exposure intervals by fitting Eq. (S1) in Supplementary Section S4 to 1950-2001 A-bomb survivor data[8]. The RR estimates peaked at 13.8 years after the A-bomb (Supplementary Figure S6), i.e. considerably later than the peak at 1.5-2.5 years for AML in cancer survivors treated with radiation (Figure 2A). Another difference is that the A-bomb survivor AML RR steady state of 2-4 (combining sexes) beyond 15 years in Supplementary Figure S6 is higher than cancer survivor AML RR steady states of ~1 beyond 12 years in Table 2. The sex-averaged A-bomb survivor AML excess RR (ERR, i.e. RR - 1) rises to a peak of ~9 after 1 Sv and the cancer survivor ERR rises to a peak of ~2.5 (averaging ~1 for males and ~4 for females). This implies that first cancers treated with radiation have, on average, the cancer risks of a whole body A-bomb dose of $(2.5/9)^{0.5}$ = ~0.5 Sv, assuming the relationship between AML risk and radiation is quadratic in dose[8]. Applying 0.5 Sv to the linear dose response fit of steady-state MDS RR among A-bomb survivors[9] predicts a steady-state MDS RR of ~3, which is also

considerably higher than our estimates of ~1. Thus, AML and MDS risk differences exist between US cancer survivors exposed to radiation therapy and Japanese survivors of A-bomb irradiation.

Radiation therapy associations with translocation-mediated AML RR

Acute promyelocytic leukemia (APL) is a subtype of AML that is associated with a chromosome translocation and is thus favored to occur after radiation exposure. Using chronic myeloid leukemia (CML) as a radiation-induced translocation-mediated positive control, at a time when 5 of 18 CML cases in A-bomb survivors occurred in those exposed to >1 Sv, 13/13 APL cases occurred in those exposed to <1 Sv[16]. Thus, counter-intuitively, A-bomb data does not support radiation induction of APL. As background APL incidence remains low with increasing ages (Figure 3A)[17] wherein PY-at-risk for a second cancer are high, we reasoned that a high signal-to-noise ratio may exist in SEER for therapy-related APL. Similar arguments can also be made for other AMLs associated with translocations and inversions (AMLti), defined here as t(6,9), inv(3), inv(16), t(8,21), t(9,11), and t(1,22) combined. We found that APL and AMLti RR peaked higher than other AML subtypes combined (Figure 3B), and that for APL more than for AMLti, the increase is specific to first cancer patients treated with radiation (Figure 3C).

First cancers with high AML/MDS RR

Table 3 shows AML/MDS RR over 1-12 years for different common first cancers. Non-Hodgkin lymphoma (NHL) and lung first cancers are followed by high AML/MDS RRs. Radiation is highly associated with AML/MDS after diagnoses of lung first cancers (Figure 4A & 4B). In contrast, after NHL (Figure 4C & 4D), slower RR time courses have no (females) or even a negative (males) association with radiation. Table 4 shows that NHL is followed by elevated risks both immediately (<0.25 years) and at steady state (>12 years). Regarding the interval of 0-0.25 years, compared to an MDS RR of ~2 after

prostate cancer that may be due to early MDS detection, NHL and chronic lymphocytic leukemia (CLL) yield much higher RR of 12-17. These RR are also considerably higher than the highest AML RR of ~7 after NHL in males, i.e. treatment-independently, NHL and CLL are linked more to MDS than to AML.

AML and MDS associations with all lymphoid first cancers combined

Hodgkin lymphoma (HL) and multiple myeloma (MM) also yielded high AML/MDS RR over 1-12 years (Table 3). Their RR time courses (Figure 4E-F) were similar to those of NHL in being slow-moving and independent of radiation therapy. We therefore pooled these first cancers with NHL and other lymphoid cancers including CLL and hairy cell leukemia (HCL). Furthermore, because radiation effects appear to be dwarfed by other effects, we also pooled first cancers treated with and without radiation. The resulting RR time courses (Figure 4G) peaked at ~5 years, resolved in >15 years, and had an MDS:AML immediate peak ratio of ~3, i.e. also supported treatment-independent lymphoid cancer linkage to MDS more than to AML; for non-hematological first cancers this ratio was ~1, see Figure 2B.

Discussion

We obtained high-resolution time-courses of the risks of developing AML/MDS after a first cancer. Risks peaked 1.5-2.5 years after diagnoses of non-hematological first cancers treated with radiation. In comparison, in Japanese A-bomb survivors, risks of all leukemia types combined over 1945-1959 peaked 4-7 years after exposure[18] and AML risks in 1950-2001 A-bomb survivors peaked in 10-15 years[8]. In addition to these latency differences there is also a difference in post-peak steady states: AML RRs remained elevated in A-bomb survivors for 4 decades after the peak but returned to 1 in 10-15 years in cancer survivors. These differences between A-bomb and cancer survivors may reflect genetics, whole- vs. partial-body radiation, the impact of chemotherapy, or environmental factors.

In persons with a 1st cancer treated with radiation, MDS differs from AML in that, based on A-bomb survivor data, its radiation dose-response is expected to be linear[9] rather than quadratic[8]. This suggests that radiation-induced AML risks are driven by two-track events (*i.e.* events caused by two independent particles), which suggests large deletions (on a scale of up to chromosomes) and/or translocations drive this process; one-track events can also cause translocations, if the target loci are tethered as for *BCR/ABL* formation in chronic myeloid leukamia[19-22], but lack of a linear component in the AML dose-response[8] speaks against such mechanisms being common for AML. In contrast, A-bomb survivor MDS risks linear in dose suggest that MDS results from one-track events. Such events include mutations that are recurrently found in AML and MDS and also found in clones in over 10% of individuals over 70 years of age[23]. Whole-body doses are more likely than partial body radiation therapy to create such clones, due to less cell killing and more cells exposed, so it is possible that higher steady-state risks in A-bomb survivors resulted from greater numbers of radiation-induced clones.

In theory, treatment-induced second cancer excess risks are expected to start at zero with a slope that is initially zero, to rise to a peak, and to fall back to steady state. First cancer blood tests revealing latent co-occurring cancers that are recorded as second cancers, will, however, confound RR estimates at early times. This RR component has an initial risk spike that falls within ~3 months into an equal magnitude trough lasting perhaps 6 months for AML/MDS, i.e. an AML/MDS case detected early by first cancer tests would likely present otherwise in 6-9 months. Based on non-irradiated non-hematological first cancers (Figure 2B) and prostate cancer (Table 4), initial RR over 0 to 0.25 years are expected to be on the order of ~2. Higher values of 2.6-3.9 after lung cancer (Table 4) could result from smoking causing both cancers[24]. Higher values of ~6 (AML) and 12-16 (MDS) after NHL and ~17 (MDS) after CLL (Table 4) may reflect pre-existing 1st hit multipotent hematopoietic stem cell (HSC) expanded clones predisposing to both AML/MDS and NHL, or MDS and CLL.

A large meta-analysis of long-term survivors of NHL showed no additional risk with radiation therapy[25] consistent with our finding of similarly elevated AML/MDS RR of 5-6 after NHL treated with or

without radiation. In contrast, AML/MDS RR below 2 after lung first cancers not treated with radiation increased substantially if radiation was used. This difference may reflect adjuvant chemotherapy for locally advanced lung cancer commonly including drugs not considered carcinogenic. In contrast, NHL is often treated with carcinogenic drugs such as alkylators. Thus, beyond mutated HSC clones, a second possible cause of lung cancer *vs*. NHL differences in AML/MDS risks may be chemotherapy. Consequently, the contribution of radiation therapy may be more important in lung cancer than in NHL.

AML/MDS risks below background rates after prostate cancers not treated with radiation (Figure 2D) suggests undetectable neoplastic clones were stabilized or diminished in size by therapy. A study of risks after androgen deprivation therapy[26] did not report decreases in AML/MDS, but was not powered to detect one. Estrogens promote HSC growth[27, 28], so a role of hormones in sex differences in AML/MDS background rates (after the age of ~55 years in Figure 3A) should perhaps be explored.

Background cancer incidences generally increase with age. Use of relative as opposed to additive metrics of risks are warranted if some of this age dependency is also present in exposure-induced risks. This is plausible if the exposure modulates a step in the multi-stage process of carcinogenesis and depends on age to drive the others. Traditionally, radiation-induced solid neoplasms have been modeled using relative risks[29] and radiation-induced leukemias have been modeled using additive risks[10]. From a statistical perspective, models are more parsimonious if their induced-risk components borrow estimates of age dependence parameters from background incidence data. The most recent analysis of Japanese A-bomb survivor data now proposes a relative risk model of radiation-induced AML[8]. The tentative nature of quantifying exposure related AML risks using relative risks is seen in Figure 2I: if the incidence in the first age group exposed to high doses is high by chance a relative risk model is preferred, but if the second age group has a low incidence by chance, and the incidence at high doses is thus roughly independent of age, an absolute/additive risk model is preferred. Lack of a clear answer in Figure 2I is consonant with choices based on statistical criteria changing between Preston *et al* in 1994[10] and Hsu *et al* in 2013[8]. If cancer-therapy-induced AML risks are additive and not relative, first cancers diagnosed at young ages will yield higher AML RR not

because the therapies are more carcinogenic but because the background AML rates are lower. Evidence that this is partly true for AML/MDS is provided in Table 3 wherein RR decreases correlate with increasing ages of observed AML/MDS second cancer cases. Thus, if our primary objective was to compare strengths of associations of different first cancer diagnoses with AML/MDS, a metric better than RR might have been the absolute risk AR = (O-E)/PY, *i.e*. the increase in incidence above the background incidence, measured on the same absolute scale as the background incidence. Our primary objective, however, was to rank tissue-banked second cancer AML/MDS cases by the odds that they are not background cases, to prioritize DNA deep sequencing efforts to find radiation susceptibility genes. For this objective, RR is the appropriate metric. For an analysis using both metrics see Morton et al.[7] Regarding ratios of RR with *vs*. without radiation as metrics of radiation association with second cancers, that breast and prostate cancer yield similar ratio magnitudes (Figures 2E & 2F) suggests that, perhaps *via* numerator and denominator age dependence cancellation, such RR ratios control for age better than RR alone.

A disadvantage of methods that compare treatment types[30, 31] is seen in the AML/MDS RR time courses after NHL (Figure 4C-D): if both treatments yield similar risk time courses, differences may not be detected and absolute dynamics may be overlooked. SEERaBomb and SEER*Stat MP-SIR avoid this by providing second cancer RR estimates relative to general population risks. The resulting RR time course plots reveal risk diversity and shed light on the plausibility of assumptions of model-based methods[30, 31]. For example, Figure 2C suggests that for AML after breast first cancers, constant proportional hazards may not be unreasonable for comparing those treated with vs. without radiation.

A logical next goal is risk decomposition. Risk component parameterizations, such as representing treatment-induced ERR as $At^n\exp(-kt)$ where t is time since the first cancer diagnosis, must be defined and tested for their ability to be fitted as a sum to epidemiologically measured risks. Improving the resolution of measured risks is a first step toward this goal. We accomplished this step here by developing new, enabling, broadly applicable functions within our R package SEERaBomb.

Our results are as follows: (1) AML/MDS RR peak ~2 years after non-hematological first cancers treated with radiation. This is sooner than expected based on A-bomb survivor data; (2) after prostate cancer not treated with radiation AML/MDS risks decrease. After radiation-therapy risks increase, uni-modally for AML and bi-modally for MDS; (3) radiation therapy has a stronger association with APL than with other AML types; and 4) strengths of radiation association with AML/MDS by first cancer type rank as lung cancer > breast cancer > lymphoid cancers.

Funding. This work was supported by: the Cleveland Clinic, the National Aeronautics and Space Administration (NNJ13ZSA001N) (TR), the National Institute for Health Research Biomedical Research Centre funding scheme (RPG), and the Academic Medical Center of Amsterdam (RJM).

Author Contributions. JPM initiated and motivated the study. TR developed the software and wrote the first draft. RKS, DJB and RPG helped with the A-bomb data analyses. RJM helped with SEER*Stat comparisons. RPG, JPM, MAS, BTH, RJM, HEC and SM contributed feedback/guidance and writing.

Acknowledgements. We thank Drs Li Li and Cornelis Van Noorden for suggestions and the R community for software. The content of this report is solely the responsibility of the authors and does not necessarily represent the official views of the National Cancer Institute or the National Institutes of Health. We used data obtained from the Radiation Effects Research Foundation (RERF), Hiroshima and Nagasaki, Japan. RERF is a public interest foundation funded by the Japanese Ministry of Health, Labour and Welfare (MHLW) and the U.S. Department of Energy (DOE), the latter in part through DOE award DE-HS0000031 to the National Academy of Sciences. Conclusions in this report are those of the authors and do not necessarily reflect the scientific judgment of RERF or its funding agencies.

Supplementary information is available at Leukemia's website

References

1. Gale RP, Bennett JM, Hoffman FO. Therapy-related AML: a slip of the lip can sink a ship. *Leukemia research* 2014 Mar; **38**(3)**:** 418-420.

2. SEER. Surveillance, Epidemiology, and End Results (SEER) Program Research Data (1973-2012), National Cancer Institute, DCCPS, Surveillance Research Program, Surveillance Systems Branch (http://www.seer.cancer.gov), released April 2015, based on the November 2014 submission. 2015.

3. Mukherjee S, Reddy CA, Ciezki JP, Abdel-Wahab M, Tiu RV, Copelan E*, et al.* Risk for developing myelodysplastic syndromes in prostate cancer patients definitively treated with radiation. *Journal of the National Cancer Institute* 2014 Mar; **106**(3)**:** djt462.

4. Armitage JO, Carbone PP, Connors JM, Levine A, Bennett JM, Kroll S. Treatment-related myelodysplasia and acute leukemia in non-Hodgkin's lymphoma patients. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2003 Mar 1; **21**(5)**:** 897-906.

5. Kaplan HG, Malmgren JA, Atwood MK. Increased incidence of myelodysplastic syndrome and acute myeloid leukemia following breast cancer treatment with radiation alone or combined with chemotherapy: a registry cohort analysis 1990-2005. *BMC cancer* 2011; **11:** 260.

6. Berrington de Gonzalez A, Curtis RE, Kry SF, Gilbert E, Lamart S, Berg CD*, et al.* Proportion of second cancers attributable to radiotherapy treatment in adults: a cohort study in the US SEER cancer registries. *The Lancet Oncology* 2011 Apr; **12**(4)**:** 353-360.

7. Morton LM, Dores GM, Tucker MA, Kim CJ, Onel K, Gilbert ES*, et al.* Evolving risk of therapy-related acute myeloid leukemia following cancer chemotherapy among adults in the United States, 1975-2008. *Blood* 2013 Apr 11; **121**(15)**:** 2996-3004.

8. Hsu WL, Preston DL, Soda M, Sugiyama H, Funamoto S, Kodama K*, et al.* The incidence of leukemia, lymphoma and multiple myeloma among atomic bomb survivors: 1950-2001. *Radiation research* 2013 Mar; **179**(3)**:** 361-382.

9. Iwanaga M, Hsu WL, Soda M, Takasaki Y, Tawara M, Joh T*, et al.* Risk of myelodysplastic syndromes in people exposed to ionizing radiation: a retrospective cohort study of Nagasaki atomic bomb survivors. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2011 Feb 1; **29**(4)**:** 428-434.

10. Preston DL, Kusumi S, Tomonaga M, Izumi S, Ron E, Kuramoto A*, et al.* Cancer incidence in atomic bomb survivors. Part III. Leukemia, lymphoma and multiple myeloma, 1950-1987. *Radiation research* 1994 2/1994; **137**(2 Suppl)**:** S68-S97.

11. Haferlach T, Nagata Y, Grossmann V, Okuno Y, Bacher U, Nagae G, *et al.* Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. *Leukemia* 2014 Feb; **28**(2)**:** 241-247.

12. Cancer Genome Atlas Research N. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *The New England journal of medicine* 2013 May 30; **368**(22)**:** 2059-2074.

13. Hastie T, Tibshirani R. Generalized additive models for medical research. *Statistical methods in medical research* 1995 Sep; **4**(3)**:** 187-196.

14. Ihaka R, Gentleman R. R:a language for data analysis and graphics. *Journal of Computational and Graphical Statistics* 1996; **5:** 299-314.

15. Garwood F. Fiducial limits for the Poisson distribution. *Biometrika* 1936; **28:** 437-442.

16. Matsuo T, Tomonaga M, Bennett JM, Kuriyama K, Imanaka F, Kuramoto A, *et al.* Reclassification of leukemia among A-bomb survivors in Nagasaki using French-American-British (FAB) classification for acute leukemia. *Japanese journal of clinical oncology* 1988 Jun; **18**(2)**:** 91-96.

17. Radivoyevitch T, Ramsey MJ, Tucker JD. Estimation of the target stem-cell population size in chronic myeloid leukemogenesis. *Radiation and environmental biophysics* 1999 9/1999; **38**(3)**:** 201-206.

18. Tomonaga M. Leukaemia in Nagasaki atomic bomb survivors from 1945 through 1959. *Bulletin of the World Health Organization* 1962; **26:** 619-631.

19. Radivoyevitch T, Kozubek S, Sachs RK. Biologically based risk estimation for radiation-induced CML. Inferences from BCR and ABL geometric distributions. *Radiation and environmental biophysics* 2001 3/2001; **40**(1)**:** 1-9.

20. Radivoyevitch T, Hlatky L, Landaw J, Sachs RK. Quantitative modeling of chronic myeloid leukemia: insights from radiobiology. *Blood* 2012 May 10; **119**(19)**:** 4363-4371.

21. Radivoyevitch T, Jankovic GM, Tiu RV, Saunthararajah Y, Jackson RC, Hlatky LR, *et al.* Sex differences in the incidence of chronic myeloid leukemia. *Radiation and environmental biophysics* 2014 Mar; **53**(1)**:** 55-63.

22. Gale RP, Hlatky L, Sachs RK, Radivoyevitch T. Why is there so much therapy-related AML and MDS and so little therapy-related CML? *Leukemia research* 2014 Oct; **38**(10)**:** 1162-1164.

23. Jaiswal S, Fontanillas P, Flannick J, Manning A, Grauman PV, Mar BG, *et al.* Age-Related Clonal Hematopoiesis Associated with Adverse Outcomes. *The New England journal of medicine* 2014 Nov 26.

24. Varadarajan R, Ford L, Sait SN, Block AW, Barcos M, Wallace PK*, et al.* Metachronous and synchronous presentation of acute myeloid leukemia and lung cancer. *Leukemia research* 2009 Sep; **33**(9)**:** 1208-1211.

25. Pirani M, Marcheselli R, Marcheselli L, Bari A, Federico M, Sacchi S. Risk for second malignancies in non-Hodgkin's lymphoma survivors: a meta-analysis. *Ann Oncol* 2011 Aug; **22**(8)**:** 1845-1858.

26. Wallner LP, Wang R, Jacobsen SJ, Haque R. Androgen deprivation therapy for treatment of localized prostate cancer and risk of second primary malignancies. *Cancer Epidemiol Biomarkers Prev* 2013 Feb; **22**(2)**:** 313-316.

27. Carreras E, Kovats S. Girl power: estrogen promotes HSC self-renewal. *Cell Stem Cell* 2014 Feb 6; **14**(2)**:** 137-138.

28. Nakada D, Oguro H, Levi BP, Ryan N, Kitano A, Saitoh Y*, et al.* Oestrogen increases haematopoietic stem-cell self-renewal in females and during pregnancy. *Nature* 2014 Jan 23; **505**(7484)**:** 555-558.

29. Thompson DE, Mabuchi K, Ron E, Soda M, Tokunaga M, Ochikubo S*, et al.* Cancer incidence in atomic bomb survivors. Part II: Solid tumors, 1958-1987. *Radiation research* 1994 Feb; **137**(2 Suppl)**:** S17-67.

30. Cox DR. Regression models and life tables. *Journal of the Royal Statistical Society, SerB* 1972 1972; **34:** 187-220.

31. P. FJ, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *J Am Stat Assoc* 1999; **94:** 496-509.

Figure Legends

Figure 1. SEER cancer cases and person-years (PY) used in this study. A) The SEER database 73 that began in 1973 contains more person years (PY) at risk (light grey area) than the databases 92 (grey area) and 00 (black area) that began in 1992 and 2000. Cancer numbers are proportional to PY, so database 73 also has the greatest number of cancer cases. SEER*Stat MP-SIR allows access to only either SEER-9 (73, light grey) or SEER-13 (92 [grey] plus the portion of 73 [light grey] directly above it); it does not allow access to 00 data (black). In contrast, SEERaBomb second cancer analyses use cases and PY in all 3 SEER databases. **B)** MDS entry into SEER began in 2001 and as a result, there are more MDS cases in the 00 (black) SEER database than in the other two databases combined. In A) and B) in 2005 in 00 (black), due to hurricane Katrina, Louisiana PY and cases in the 2nd half of 2005 exist in a separate database not included here. **C)** First cancer cases are used to compute PY at risk of a second cancer (left branch) and all (i.e. first, second, and higher) AML and MDS cases and corresponding SEER population PY since 1973 and 2001, respectively, are used to compute AML and MDS background incidences (right branch). The branches merge to compute expected cases (E) under a null hypothesis that prior cancers are irrelevant. Observed (O) cases then yield RR = O/E. First cancer patients treated with ionizing radiation (IR) have SEER cancer treatment radiation codes 1-6, those without IR have codes 0 or 7; those with codes 8-9 have unknown IR status and were thus excluded from this study. Benign tumors identified by SEER sequence codes in the range of 60 to 88 were also excluded. **D-E)** 2D-spline fits to female AML **(D)** and MDS **(E)** incidence *versus* year and age (plots for males were similar). Incidence units are $\log_{10}$ of cases per 100,000 PY. Points on the bottom plane correspond to age-years with zero cases; for AML, fewer ages with zero cases with increasing years result from increases in PY at risk, particularly as the number of SEER registries increased from 9 to 13 in 1992 and from 13 to 18 in 2000.

Figure 2. AML/MDS RR time courses after diagnoses of non-hematological first cancers. A-B) Peaks are higher in females than in males and higher with radiation (A) than without (B), being essentially undetectable in males not treated with radiation. To avoid correlations possibly attributable to pre-existing hematopoietic stem cell (HSC) mutant clones, hematological first cancers (defined in Supplementary Section S1) were excluded. **C-D)** Risk time courses after breast (C) and prostate (D) first cancers are similar to those of all female and male first cancers. After breast first cancers, AML RRs with and without radiation therapy have similar time course shapes. After prostate first cancers treated with radiation, MDS RR show two modes, an early mode between 0.75 and 2.75 y and a late mode over 3-15 y. **E-F)** As in C and D but with risks after radiation therapy relative to risks after treatment without radiation (instead of relative to general population risks). These ratios of RR for treatment with radiation relative to treatment without radiation can be viewed as metrics of association of radiation with AML and MDS. **E)** For breast first cancers, peak ratios are smaller than RR themselves and comparable in magnitude to those after prostate first cancers. **F)** With prostate first cancers not treated with radiation as the baseline the existence of two MDS risk modes is revealed more than in D. **G)** Average ages of observed cases are higher than expected average ages (i.e. of PY at risk in that interval), more so for MDS than for AML. **A-F)** RR were estimated over the following time intervals in years: (0,0.25], (0.25,0.5], (0.5,0.75], (0.75,1], (1,1.5], (1.5,2], (2,2.5], (2.5,3], (3,4], (4,5], (5,6], (6,8], (8,10], (10,12], and (12,∞). RR were plotted as points at PY-weighted times; MDS times in A and B were shifted by +0.05 years to increase CI visibility. **H)** At 55-60 years of age, the incidence of non-hematological cancers transitions from being higher in females to being higher in males. This implies that female ages at times of second cancer diagnoses are likely to be younger. **I)** Sex differences in hematological cancers are relatively steady across ages. **H-I)** incidence increases due to screening visible at the ages of 40, 50 and 65 years (vertical gray lines) confirm SEER data signal availability. **J)** In Japanese A-bomb survivors, with both sexes pooled, AML incidence vs age in the high dose group can be interpreted as being either independent of age or proportional to background, depending on belief in the first high dose data point versus the second. The former favors absolute risk

models, the latter relative risk models; parallel curves on a log-scale correspond to multiplicative risks. Dose group definitions are: low, <0.01 Sv; medium, 0.01-0.4 Sv; and high, ≥0.4 Sv. 95% CI assume Poisson distributed cases.

Figure 3. Acute promyelocytic leukemia (APL) association with radiation therapy. **A)** Background incidence rates of APL and AMLti (translocation and inversion mediated) are flat versus age relative to other AML types, so at older ages where second cancer PY coverage is greatest, these endpoints may have high signal-to-noise ratios. **B)** APL and AMLti RR after non-hematological first cancers treated with radiation are higher than RR of other AMLs and MDS. **C)** After non-hematological first cancers not treated with radiation, over 1-2 years in females and 2-3 years in males, AMLti lower CI limits do not include AML/MDS RR means while APL CI do. **B-C)** Ratios of RR could not be computed for these plots due to too few observed cases causing too many simulated divisions by 0.

Figure 4. Risks of AML/MDS after diagnoses of various first cancers. A) Risks after lung first cancers are higher with vs. without radiation; without radiation small initial risks stay flat or trend downward. **B)** Radiation risks relative to no therapy are similar to those relative to general population risks because RR after first lung cancers not treated with radiation are ~1, i.e. radiation is strongly associated with AML/MDS risks after lung cancer. **C)** AML/MDS RR after NHL are similar with vs. without radiation and remain elevated longer than RR after all non-hematological cancers combined (Figure 2A-B). **D)** Risks after radiation therapy relative to risks after treatment without radiation reveal some radiation prophylaxis of AML/MDS in male NHL cases; in female NHL cases, radiation is not associated with AML/MDS risks. **E-F)** AML/MDS RR peak sooner after Hodgkin lymphomas (HL) than after multiple myelomas (MM), and are approximately the same whether treated with or without radiation. **G)** Pooling irradiated and non-irradiated cases across NHL, HL, MM, hairy cell leukemia (HCL), small lymphocytic lymphoma (SLL) and chronic lymphocytic leukemia (CLL) yields a broad AML/MDS RR time course that peaks at ~5 years.

Table 1. Software validation based on RR* of second bladder- and lung cancers in SEER-9 data.

	First cancer treated without radiation		First cancer treated with radiation	
First cancer	➔Bladder	➔Lung	➔Bladder	➔Lung
Breast[#]	1.07 (0.98,1.16) O=515	0.89 (0.86,0.93) O=2381	1.14 (1.01,1.28) O=290	1.15 (1.1,1.21) O=1701
Breast[^]	1.06, O=435	0.86, O=1903	1.06, O=192	1.11, O=1136
Prostate[#]	0.82 (0.78,0.86) O=1593	0.72 (0.69,0.75) O=2903	1.42 (1.35,1.49) O=1577	0.93 (0.89,0.97) O=2067
Prostate[^]	0.77, O=1072	NA, O=2115	1.32, O=1026	0.87, O=1450

*RR are for second cancers over periods of time >5 years after first cancer diagnoses at ages of 20 to 84 years.
[#]Data used with SEERaBomb was through 2012; RR 95% CI are in parentheses; O = observed second cancer numbers.
[^]From the Online Supplement of Lancet Oncology 2011; 12: 353-60; data was through 2002; 95% CI were not provided.
Cancers shown were selected based on high numbers of cases and proximity of lungs to breasts and bladders to prostates.

Table 2. AML/MDS RR after any non-hematological 1st cancer

Radiation	Second Cancer	Sex	O*	E*	RR* (95% CI)	Time interval after 1st cancer diagnosis in y
yes	AML	Female	870	308.43	2.82 (2.64, 3.01)	1-12
yes	AML	Male	830	548.47	1.51 (1.41, 1.62)	1-12
yes	MDS	Female	563	325.52	1.73 (1.59, 1.88)	1-12
yes	MDS	Male	865	688.53	1.26 (1.17, 1.34)	1-12
no	AML	Female	1209	797.22	1.52 (1.43, 1.60)	1-12
no	AML	Male	1486	1375.21	1.08 (1.03, 1.14)	1-12
no	MDS	Female	861	743.33	1.16 (1.08, 1.24)	1-12
no	MDS	Male	1483	1536.1	0.97 (0.92, 1.02)	1-12
yes	AML	Female	95	80.08	1.19 (0.96, 1.45)	>12
yes	AML	Male	69	77.43	0.89 (0.69, 1.13)	>12
yes	MDS	Female	119	98.16	1.21 (1.00, 1.45)	>12
yes	MDS	Male	137	124.27	1.10 (0.93, 1.30)	>12
no	AML	Female	284	320.52	0.89 (0.79, 1.00)	>12
no	AML	Male	252	287.67	0.88 (0.77, 0.99)	>12
no	MDS	Female	332	356.35	0.93 (0.83, 1.04)	>12
no	MDS	Male	417	430.44	0.97 (0.88, 1.07)	>12

*O, E, and RR are observed and expected cases and relative risk, respectively.

Table 3. AML/MDS RR 1-12 years after non-myeloid first cancers

First Cancer	Second Cancer	Sex	Age	O*	E*	RR* (95% CI)	Radiation Therapy
NHL	AML	Female	65.12	55	9.03	6.09 (4.59, 7.93)	Yes
lung	AML	Female	66.67	52	10.48	4.96 (3.71, 6.51)	Yes
lung	AML	Male	68.69	79	19.91	3.97 (3.14, 4.95)	Yes
NHL	AML	Male	61.12	47	12.75	3.69 (2.71, 4.90)	Yes
breast	AML	Female	63.44	539	178.69	3.02 (2.77, 3.28)	Yes
uterus	AML	Female	71.22	68	31.16	2.18 (1.69, 2.77)	Yes
oral	AML	Male	65.88	39	21.73	1.79 (1.28, 2.45)	Yes
prostate	AML	Male	75.82	497	414.82	1.20 (1.10, 1.31)	Yes
NHL	MDS	Female	68.82	38	9.48	4.01 (2.84, 5.50)	Yes
NHL	MDS	Male	71.01	43	13.76	3.12 (2.26, 4.21)	Yes
lung	MDS	Female	72.47	32	11.38	2.81 (1.92, 3.97)	Yes
lung	MDS	Male	73.21	48	19.07	2.52 (1.86, 3.34)	Yes
oral	MDS	Male	70.26	34	19.83	1.71 (1.19, 2.40)	Yes
breast	MDS	Female	72.61	343	205.02	1.67 (1.50, 1.86)	Yes
HL	AML	Male	49.71	57	3.24	17.59 (13.32, 22.79)	No
ovary	AML	Female	64.97	145	25.28	5.74 (4.84, 6.75)	No
NHL	AML	Female	66.62	154	28.23	5.46 (4.63, 6.39)	No
NHL	AML	Male	64.63	209	40.92	5.11 (4.44, 5.85)	No
MM	AML	Male	69.31	58	13.07	4.44 (3.37, 5.74)	No
lung	AML	Female	70.58	60	31.5	1.90 (1.45, 2.45)	No
breast	AML	Female	66.13	410	247.49	1.66 (1.50, 1.83)	No
bladder	AML	Male	74.79	180	153.46	1.17 (1.01, 1.36)	No
prostate	AML	Male	76.92	574	658.08	0.87 (0.80, 0.95)	No
NHL	MDS	Female	71.24	156	33.74	4.62 (3.93, 5.41)	No
NHL	MDS	Male	66.51	211	49.79	4.24 (3.69, 4.85)	No
MM	MDS	Male	70.39	55	15.63	3.52 (2.65, 4.58)	No
ovary	MDS	Female	67.95	67	22.42	2.99 (2.32, 3.80)	No
CLL	MDS	Male	71.98	61	29.17	2.09 (1.60, 2.69)	No
lung	MDS	Female	74.95	58	36.46	1.59 (1.21, 2.06)	No
prostate	MDS	Male	79.49	639	755.83	0.85 (0.78, 0.91)	No
melanoma	MDS	Male	76.97	55	86.57	0.64 (0.48, 0.83)	No
NHL	AML	Female	65.12	55	9.03	6.09 (4.59, 7.93)	Yes

*O, E, and RR are observed and expected cases and relative risk, respectively.

Thresholds for inclusion in this table were RR CI not including 1 and observed cases ≥30 (with radiation) or ≥50 (without radiation). Spearman correlation between age (mean of observed cases) and RR: r_s = -0.73; P = 5.5e-6.

Table 4. AML/MDS initial and steady state RR after non-myeloid first cancers

First Cancer	Second Cancer	Sex	O*	E*	RR*	Radiation	Time interval in y
NHL	AML	Male	20	3.05	6.56 (4.01, 10.13)	No	<0.25
lung	AML	Male	24	9.18	2.61 (1.68, 3.89)	No	<0.25
CLL	MDS	Male	26	1.54	16.88 (11.03, 24.74)	No	<0.25
NHL	MDS	Female	27	2.01	13.43 (8.85, 19.54)	No	<0.25
NHL	MDS	Male	42	3.28	12.80 (9.23, 17.31)	No	<0.25
lung	MDS	Male	27	9.07	2.98 (1.96, 4.33)	No	<0.25
prostate	MDS	Male	39	23	1.70 (1.21, 2.32)	No	<0.25
NHL	MDS	Male	10	0.64	15.62 (7.49, 28.73)	Yes	<0.25
lung	MDS	Male	20	5.19	3.85 (2.35, 5.95)	Yes	<0.25
NHL	AML	Male	20	6.18	3.24 (1.98, 5.00)	No	>12
NHL	AML	Female	14	4.56	3.07 (1.68, 5.15)	No	>12
colon	AML	Male	19	34.84	0.55 (0.33, 0.85)	No	>12
NHL	MDS	Male	28	8.17	3.43 (2.28, 4.95)	No	>12
NHL	MDS	Female	15	5.8	2.59 (1.45, 4.27)	No	>12
giCIS	MDS	Male	20	9.63	2.08 (1.27, 3.21)	No	>12
melanoma	MDS	Male	14	26.31	0.53 (0.29, 0.89)	No	>12
HL	AML	Female	5	0.95	5.26 (1.71, 12.28)	Yes	>12
HL	MDS	Male	5	0.98	5.10 (1.66, 11.91)	Yes	>12
NHL	MDS	Female	8	2.39	3.35 (1.45, 6.60)	Yes	>12
NHL	MDS	Male	10	3.45	2.90 (1.39, 5.33)	Yes	>12
thyroid	MDS	Female	8	3.16	2.53 (1.09, 4.99)	Yes	>12
breast	MDS	Female	71	53.28	1.33 (1.04, 1.68)	Yes	>12

*O, E, and RR are observed and expected cases and relative risk, respectively. Shown are significant RR with observed cases of ≥10 (with radiation) or ≥20 (without radiation) for the interval of <0.25 y and observed cases ≥5 (with radiation) and ≥10 (without radiation) for the steady state interval of >12 y. #CIS = *carcinoma in situ*

Figure 1

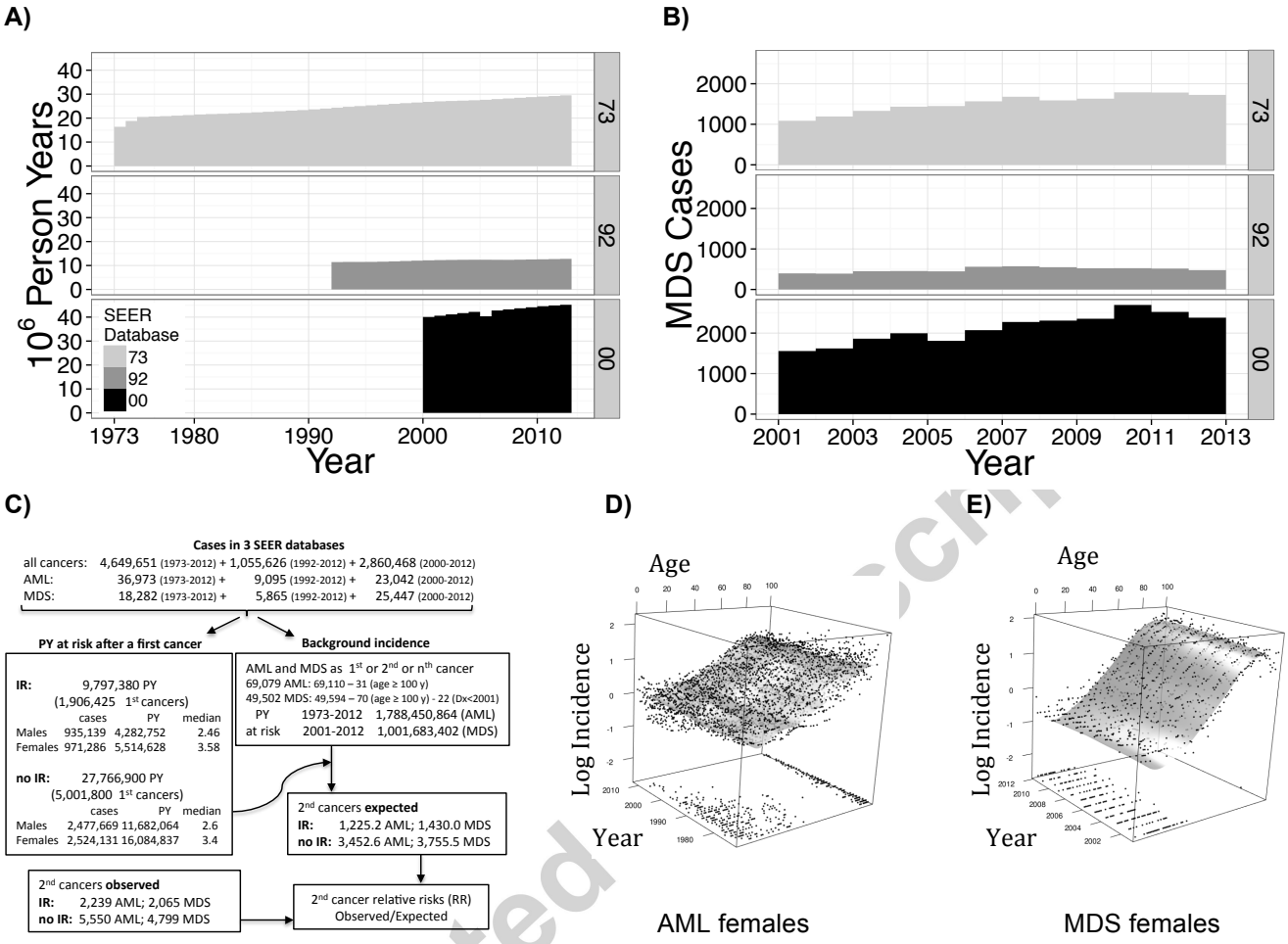

Figure 2

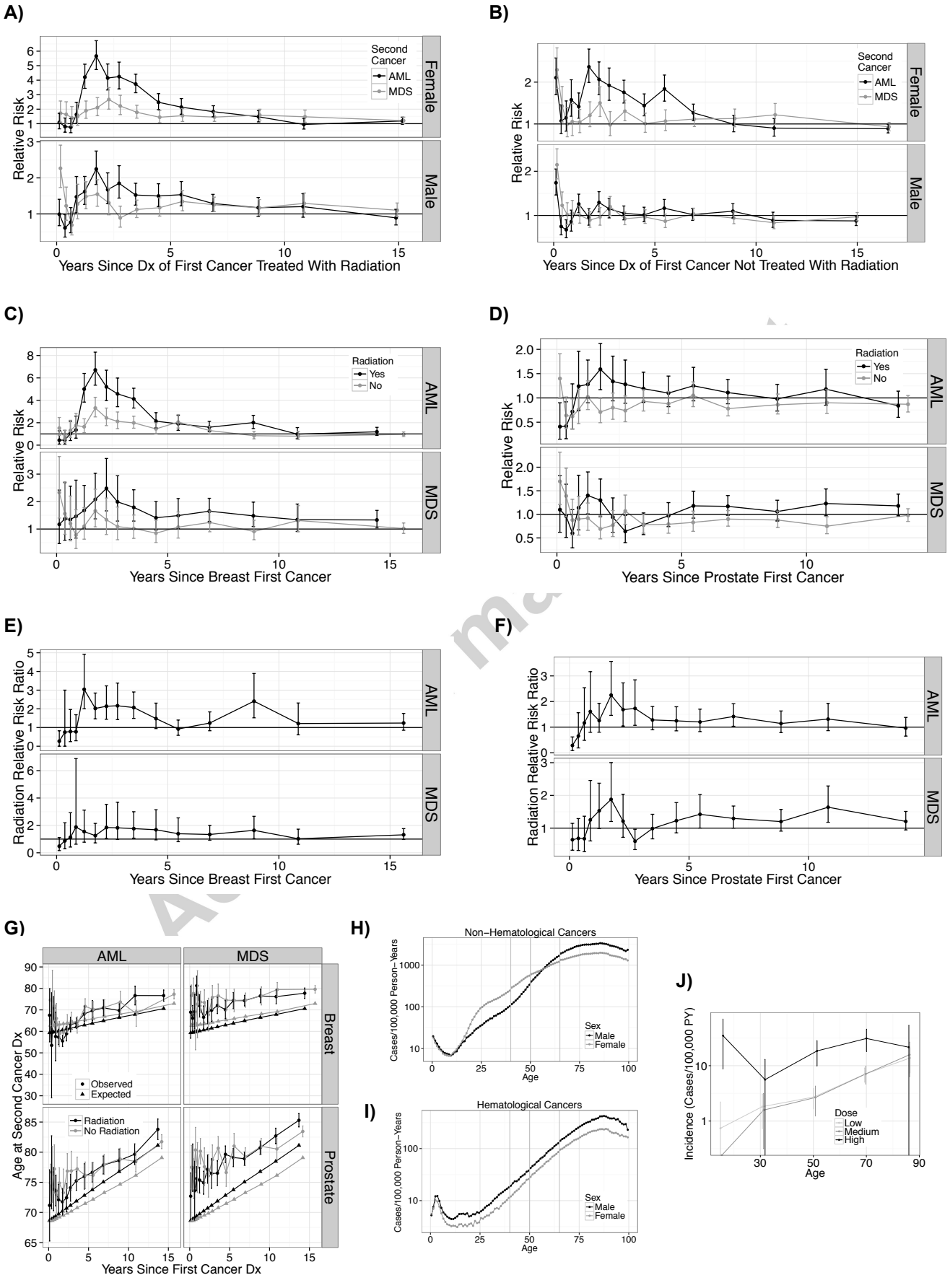

Figure 3

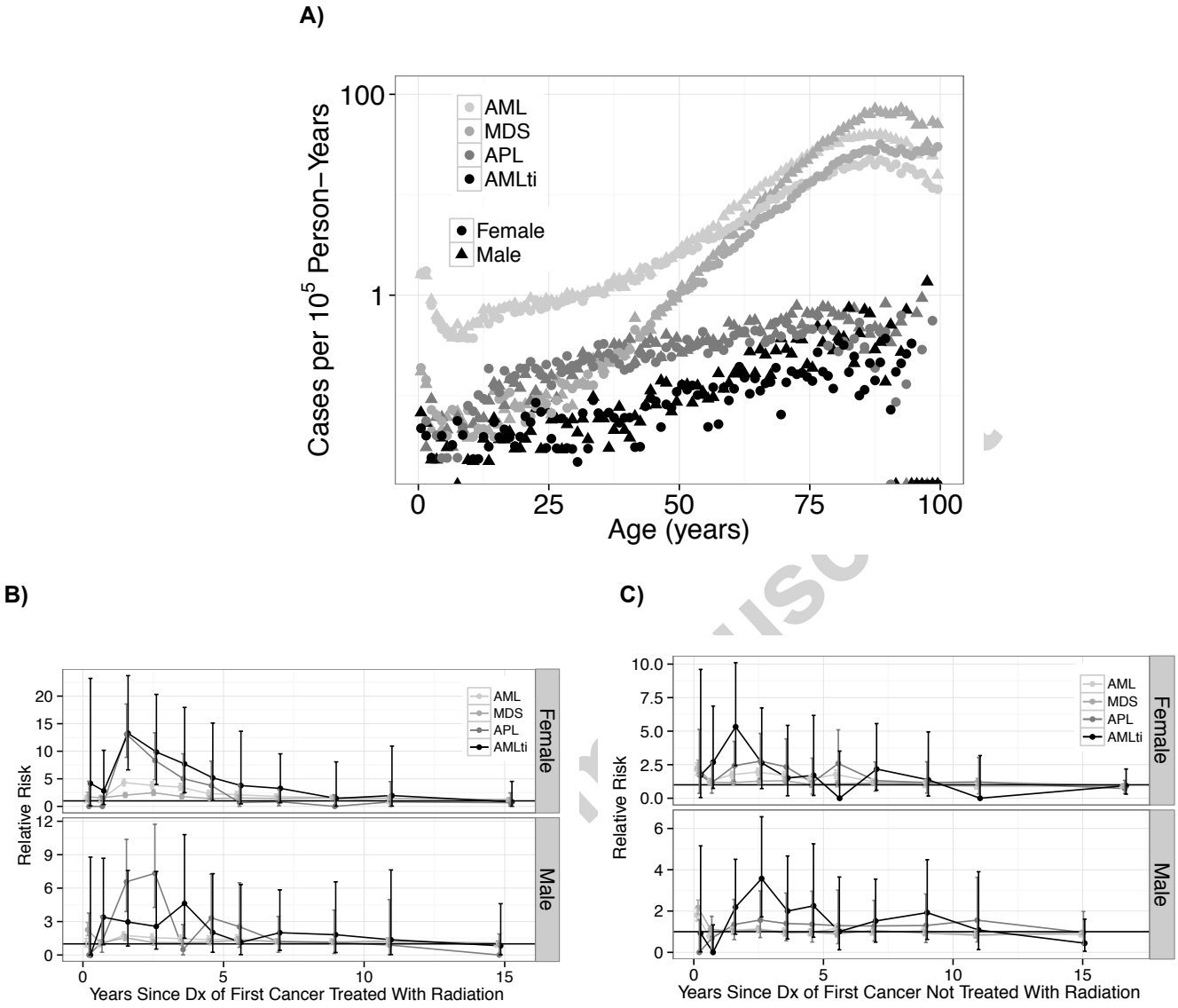

Figure 4