Towards personalised medicine for cancer

From initial therapy to follow-up

Molenaar, R.J.

Creative Commons License (see https://creativecommons.org/use-remix/cc-licenses):
Other

Citation for published version (APA):
Molenaar, R. J. (2017). Towards personalised medicine for cancer: From initial therapy to follow-up.

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (http://dare.uva.nl)
Chapter 17: Defining AML and MDS second cancer risk dynamics after diagnoses of first cancers treated or not with irradiation


Abstract
Risks of acute myeloid leukaemia (AML) and/or myelodysplastic syndromes (MDS) are known to increase after cancer treatments. Their rise-and-fall dynamics and their associations with radiation therapy have, however, not been fully characterised. To improve risk definition, we developed SEERaBomb R software for Surveillance, Epidemiology and End Results (SEER) 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 leukaemia in persons receiving radiation therapy. Knowing the type of first cancer, if it was treated with irradiation, 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. The first two are background mechanisms and the latter two are related to treatment. 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 RR, defined as the ratio of the number of observed cases over the number of 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 by 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.

SEER data comprises three databases of nine, four and five registries. Although the most recent database contains only five registries, it currently accrues more person-years (PY) at risk than the other two databases combined (Figure 1). The first database, with nine 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 or they compared risks with those of the general population. Our study extends the latter of these studies by using all three 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 AML versus MDS, for different first cancers, for males versus females, and for treatment with versus without irradiation, 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 with A-Bomb survivor analyses are made where possible, with the caveat that in the absence of chemotherapy information, SEER irradiation associations do not imply causality and could be indirect, partly or completely, via correlations with chemotherapy.
Materials and Methods

Cancer definitions. Cancer types were defined by International Classification of Disease codes (ICD-9 for non-haematologic cancers and ICD-O-3 for haematologic cancers) as described in eTable 1. Numbers of each cancer type by year show SEER use of ICD-O-3 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-O-3 9920). As MDS and AML 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 versus 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 generalised additive model13 to cases observed using Poisson regression: cases ~ s(age) + s(year) + t(age,year) + offset(log(PY)). Here s() is a one-dimensional spline and t() is a tensor interaction term, which is included to control for possible interactions between age and year, to more accurately compute expected cases for 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 generalised Poisson models is an exponential link, that is, 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 (that is, maximum likelihood) model divided by PY are shown as expected incidence surfaces in Figures 1D-E. Such surfaces provide smoothing (local averaging) of observed age–year incidence rates (that is, 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 (where O is the observed number of cases), and in ratios thereof, RR/RR=O/E × E/E (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 years; this is particularly important for MDS, as 21% of second cancer MDS cases are in the age group 85+ years.

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 prioritised 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.560

Time courses of relative risks after a first cancer. Sex-specific background incidences (surface values in Figures 1D-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 \( O \) is the number of second cancer cases observed for that time interval. RR 95% confidence intervals (CI) were found by assuming that \( O \) is Poisson distributed, as \( qchiq(0.025, 2 \times O) / (2 \times E) \) and \( qchiq(0.975, 2 \times O+2) / (2 \times E) \) in R. For ratios of RR, CI were 2.5 and 97.5% quantiles of 5000 simulations of \( (O/O) \times (E/E) \) with \( O \) and \( O \) Poisson distributed with means equal to observed values with and without irradiation \( (i) \). For cancers with small numbers of observed cases (for example, APL), such ratios were unstable due to too many divisions by zero (owing to \( O \) being small) and are thus not provided (longer time intervals minimise this but result in unacceptable time resolution losses); such RR ratios can, however, still be conceptualised in terms of numerator and denominator RR time courses. RR were plotted at PY-weighted interval midpoints defined by \( interval \ start \ times + PY/cases/2 \), that is, 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 with 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). SEERaBomb adds 0.5 months to survival times to compensate for flooring (for example, times <1 month are reported as 0), it subtracts 0.5 from months of diagnosis (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 Berrington de Gonzalez et al.\(^{554}\) are similar to SEERaBomb estimates. SEERaBomb’s use of all three SEER databases (Figure 1) enables higher resolution second cancer risk estimates than the competing software SEER\(^*\)Stat MP-SIR,\(^{488}\) as the latter does not provide access to SEER registries starting in 2000 (that is, 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 recent 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 nonhaematological first cancers treated with irradiation (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

<table>
<thead>
<tr>
<th>First cancer</th>
<th>First cancer treated without radiation</th>
<th>First cancer treated with radiation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bladder</td>
<td>Lung</td>
</tr>
<tr>
<td>Breast(^a)</td>
<td>1.07 (0.98,1.16)</td>
<td>0.89 (0.86,0.93)</td>
</tr>
<tr>
<td></td>
<td>0=515</td>
<td>0=2381</td>
</tr>
<tr>
<td></td>
<td>1.14 (1.01,1.28)</td>
<td>1.15 (1.1,1.21)</td>
</tr>
<tr>
<td></td>
<td>0=290</td>
<td>0=1701</td>
</tr>
<tr>
<td>Breast(^b)</td>
<td>1.06, 0=435</td>
<td>0.86, 0=1903</td>
</tr>
<tr>
<td></td>
<td>1.06, 0=192</td>
<td>1.10, 0=1136</td>
</tr>
<tr>
<td>Prostate(^a)</td>
<td>0.82 (0.78,0.86)</td>
<td>0.72 (0.69,0.75)</td>
</tr>
<tr>
<td></td>
<td>0=1593</td>
<td>0=2903</td>
</tr>
<tr>
<td></td>
<td>1.42 (1.35,1.49)</td>
<td>0.93 (0.89,0.97)</td>
</tr>
<tr>
<td></td>
<td>0=1577</td>
<td>0=2067</td>
</tr>
<tr>
<td>Prostate(^b)</td>
<td>0.77, 0=1072</td>
<td>NA, 0=2115</td>
</tr>
<tr>
<td></td>
<td>1.32, 0=1026</td>
<td>0.87, 0=1450</td>
</tr>
</tbody>
</table>

*RR are for second cancers over periods of time >5 years after first cancer diagnoses at ages of 20 to 84 years. \(^a\)Data used with SEERaBomb was through 2012; RR 95% CI are in parentheses; \( O \) = observed second cancer numbers. \(^b\)From the Online Supplement of Lancet Oncology 2011; 12: 353-60;\(^{554}\) data was through 2002; 95% CI were not provided.
AML and MDS risks after therapy for a first cancer

Figure 1. SEER cancer cases and person-years (PY) used in this chapter.

(A) The SEER database 73 that began in 1973 contains more person-years (PY) at risk (light gray area) than the databases 92 (gray 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 gray) or SEER-13 (92 (gray) plus the portion of 73 (light gray) directly above it); it does not allow access to 00 data (black). In contrast, SEERaBomb second cancer analyses use cases and PY in all three 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), owing to hurricane Katrina, Louisiana PY and cases in the second 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 (that is, 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 tumours identified by SEER sequence codes in the range of 60–88 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 log10 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.

prostate (Figure 2D) first cancers were similar to those of females and males after any nonhaematological first cancer (Figures 2A–B). RR time course pairs (with versus without radiation) in Figures 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 nonhaematological first cancers arriving at earlier ages in females (Figures 2H-I), and thus lower background rates, if irradiation-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 Supplementary Equation S1 in Supplementary Section S4 to 1950-2001 A-bomb survivor data.556 The RR estimates peaked at 13.8 years after the A-bomb (Supplementary Figure S6), that is, 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, that is, 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.556 Applying 0.5 Sv to the linear dose response fit of steady-state MDS RR among A-bomb survivors557 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 leukaemia (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 leukaemia (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.561 Thus, counter-intuitively, A-bomb data do not support radiation induction of APL. As background APL incidence remains low with increasing ages (Figure 3A)561 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), Table 2. AML/MDS RR after any non-haematological first cancer.

<table>
<thead>
<tr>
<th>Radiation</th>
<th>2nd cancer</th>
<th>Sex</th>
<th>O*</th>
<th>E*</th>
<th>RR* (95% CI)</th>
<th>Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>yes AML</td>
<td>F</td>
<td>848</td>
<td>306.93</td>
<td>2.76 (2.58, 2.96)</td>
<td>1-12</td>
<td></td>
</tr>
<tr>
<td>yes AML</td>
<td>M</td>
<td>801</td>
<td>546.25</td>
<td>1.47 (1.37, 1.57)</td>
<td>1-12</td>
<td></td>
</tr>
<tr>
<td>yes MDS</td>
<td>F</td>
<td>557</td>
<td>324.74</td>
<td>1.72 (1.58, 1.86)</td>
<td>1-12</td>
<td></td>
</tr>
<tr>
<td>yes MDS</td>
<td>M</td>
<td>850</td>
<td>687.24</td>
<td>1.24 (1.16, 1.32)</td>
<td>1-12</td>
<td></td>
</tr>
<tr>
<td>no AML</td>
<td>F</td>
<td>1176</td>
<td>795.38</td>
<td>1.48 (1.40, 1.57)</td>
<td>1-12</td>
<td></td>
</tr>
<tr>
<td>no AML</td>
<td>M</td>
<td>1429</td>
<td>1372.06</td>
<td>1.04 (0.99, 1.10)</td>
<td>1-12</td>
<td></td>
</tr>
<tr>
<td>no MDS</td>
<td>F</td>
<td>853</td>
<td>741.93</td>
<td>1.15 (1.07, 1.23)</td>
<td>1-12</td>
<td></td>
</tr>
<tr>
<td>no MDS</td>
<td>M</td>
<td>1463</td>
<td>1533.65</td>
<td>0.95 (0.91, 1.00)</td>
<td>1-12</td>
<td></td>
</tr>
<tr>
<td>yes AML</td>
<td>F</td>
<td>90</td>
<td>79.16</td>
<td>1.14 (0.91, 1.40)</td>
<td>&gt;12</td>
<td></td>
</tr>
<tr>
<td>yes AML</td>
<td>M</td>
<td>67</td>
<td>75.98</td>
<td>0.88 (0.68, 1.12)</td>
<td>&gt;12</td>
<td></td>
</tr>
<tr>
<td>yes MDS</td>
<td>F</td>
<td>118</td>
<td>97.56</td>
<td>1.21 (1.00, 1.45)</td>
<td>&gt;12</td>
<td></td>
</tr>
<tr>
<td>yes MDS</td>
<td>M</td>
<td>132</td>
<td>123.24</td>
<td>1.07 (0.90, 1.27)</td>
<td>&gt;12</td>
<td></td>
</tr>
<tr>
<td>no AML</td>
<td>F</td>
<td>283</td>
<td>319.87</td>
<td>0.88 (0.78, 0.99)</td>
<td>&gt;12</td>
<td></td>
</tr>
<tr>
<td>no AML</td>
<td>M</td>
<td>245</td>
<td>286.51</td>
<td>0.86 (0.75, 0.97)</td>
<td>&gt;12</td>
<td></td>
</tr>
<tr>
<td>no MDS</td>
<td>F</td>
<td>329</td>
<td>355.81</td>
<td>0.92 (0.83, 1.03)</td>
<td>&gt;12</td>
<td></td>
</tr>
<tr>
<td>no MDS</td>
<td>M</td>
<td>411</td>
<td>429.4</td>
<td>0.96 (0.87, 1.05)</td>
<td>&gt;12</td>
<td></td>
</tr>
</tbody>
</table>

*O, E, and RR are observed and expected cases and relative risk, respectively. Interval is the time after 1st cancer diagnosis in years.
AML and MDS risks after therapy for a first cancer

Figure 2. AML/MDS RR time courses after diagnoses of nonhaematological 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 haematopoietic stem cell (HSC) mutant clones, haematological 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 years 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 (that is, 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 nonhaematological 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 haematological 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 versus 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.
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 AMLt, RR peaked higher than other AML subtypes combined (Figure 3B), and that for AML more than for AMLt, 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 RR. Radiation is highly associated with AML/MDS after diagnoses of lung first cancers (Figures 4A-B). In contrast, after NHL (Figures 4C-D), 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 with an MDS RR of ~2 after prostate cancer that may be due to early MDS detection, NHL and chronic lymphocytic leukaemia (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, that is, 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 (Figures 4E-F) were similar to those of NHL in being slow-moving and independent of
AML and MDS risks after therapy for a first cancer

Figure 3. Acute promyelocytic leukaemia (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 nonhaematological first cancers treated with radiation are higher than RR of other AMLs and MDS. (C) After nonhaematological 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.

radiation therapy. We therefore pooled these first cancers with NHL and other lymphoid cancers including CLL and hairy cell leukaemia (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, that is, also supporting treatment-independent lymphoid cancer linkage to MDS more than to AML; for nonhaematological first cancers this ratio was ~1 (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 nonhaematological first cancers treated with radiation. In comparison, in Japanese A-bomb survivors, risks of all leukaemia types combined over 1945–1959 peaked 4–7 years after exposure562 and AML risks in 1950–2001 A-bomb survivors peaked in 10–15 years.556 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 four decades after the peak but returned to one in 10–15 years in cancer survivors. These differences between A-bomb and cancer survivors may reflect genetics, whole-body versus partial-body radiation, the impact of chemotherapy, or environmental factors.

In persons with a first cancer treated with radiation, MDS differs from AML in that, based on A-bomb survivor data, its irradiation dose–response is expected to be linear557 rather than quadratic.8 This suggests that radiation-induced AML risks are driven by two-track events (that is, 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 leukaemia,563-566 but lack of a linear component in the AML dose–response556 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.493 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
Figure 4. Risks of AML/MDS after diagnoses of various first cancers.

(A) Risks after lung first cancers are higher with versus 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 versus without radiation and remain elevated longer than RR after all nonhaematological cancers combined (Figures 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 nonirradiated cases across NHL, HL, MM, hairy cell leukaemia (HCL), small lymphocytic lymphoma (SLL) and chronic lymphocytic leukaemia (CLL) yields a broad AML/MDS RR time course that peaks at ~5 years.
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, that is, an AML/MDS case detected early by first cancer tests would likely present otherwise in 6–9 months. Based on nonirradiated nonhaematological 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. Higher values of ~6 (AML) and 12–16 (MDS) after NHL and ~17 (MDS) after CLL (Table 4) may reflect pre-existing first hit multipotent haematopoietic stem cell 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 therapy568 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 haematopoietic stem cell clones, a second possible cause of lung cancer versus 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 stabilised or diminished in size by therapy. A study of risks after androgen deprivation therapy569 did not report decreases in AML/MDS, but was not powered to detect one. Estrogens promote haematopoietic stem cell growth,570 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.

### Table 4. AML/MDS initial and Steady State RR after non-myeloid first cancers.

<table>
<thead>
<tr>
<th>Radiation</th>
<th>First Cancer</th>
<th>2nd Cancer</th>
<th>Sex</th>
<th>O*</th>
<th>E*</th>
<th>RR*</th>
<th>Int**</th>
</tr>
</thead>
<tbody>
<tr>
<td>no</td>
<td>NHL</td>
<td>AML</td>
<td>M</td>
<td>20</td>
<td>3.05</td>
<td>6.56 (4.01, 10.13)</td>
<td>&lt;0.25</td>
</tr>
<tr>
<td>no</td>
<td>lung</td>
<td>AML</td>
<td>M</td>
<td>24</td>
<td>9.18</td>
<td>2.61 (1.67, 3.89)</td>
<td>&lt;0.25</td>
</tr>
<tr>
<td>no</td>
<td>CLL</td>
<td>MDS</td>
<td>M</td>
<td>26</td>
<td>1.54</td>
<td>16.92 (11.05, 24.79)</td>
<td>&lt;0.25</td>
</tr>
<tr>
<td>no</td>
<td>NHL</td>
<td>MDS</td>
<td>F</td>
<td>27</td>
<td>2.01</td>
<td>13.42 (8.84, 19.53)</td>
<td>&lt;0.25</td>
</tr>
<tr>
<td>no</td>
<td>NHL</td>
<td>MDS</td>
<td>M</td>
<td>42</td>
<td>3.28</td>
<td>12.82 (9.24, 17.33)</td>
<td>&lt;0.25</td>
</tr>
<tr>
<td>no</td>
<td>lung</td>
<td>MDS</td>
<td>M</td>
<td>27</td>
<td>9.07</td>
<td>2.98 (1.96, 4.33)</td>
<td>&lt;0.25</td>
</tr>
<tr>
<td>no</td>
<td>prostate</td>
<td>MDS</td>
<td>M</td>
<td>39</td>
<td>23</td>
<td>1.70 (1.21, 2.32)</td>
<td>&lt;0.25</td>
</tr>
<tr>
<td>yes</td>
<td>NHL</td>
<td>MDS</td>
<td>M</td>
<td>10</td>
<td>0.64</td>
<td>15.65 (7.50, 28.78)</td>
<td>&lt;0.25</td>
</tr>
<tr>
<td>yes</td>
<td>lung</td>
<td>MDS</td>
<td>M</td>
<td>20</td>
<td>5.19</td>
<td>3.85 (2.35, 5.95)</td>
<td>&lt;0.25</td>
</tr>
<tr>
<td>no</td>
<td>NHL</td>
<td>AML</td>
<td>M</td>
<td>20</td>
<td>6.18</td>
<td>3.24 (1.98, 5.00)</td>
<td>&gt;12</td>
</tr>
<tr>
<td>no</td>
<td>NHL</td>
<td>AML</td>
<td>F</td>
<td>14</td>
<td>4.56</td>
<td>3.07 (1.68, 5.15)</td>
<td>&gt;12</td>
</tr>
<tr>
<td>no</td>
<td>colon</td>
<td>AML</td>
<td>M</td>
<td>19</td>
<td>34.84</td>
<td>0.55 (0.33, 0.85)</td>
<td>&gt;12</td>
</tr>
<tr>
<td>no</td>
<td>NHL</td>
<td>MDS</td>
<td>M</td>
<td>28</td>
<td>8.17</td>
<td>3.43 (2.28, 4.95)</td>
<td>&gt;12</td>
</tr>
<tr>
<td>no</td>
<td>NHL</td>
<td>MDS</td>
<td>F</td>
<td>15</td>
<td>5.8</td>
<td>2.59 (1.45, 4.27)</td>
<td>&gt;12</td>
</tr>
<tr>
<td>no</td>
<td>GI CIS</td>
<td>MDS</td>
<td>M</td>
<td>20</td>
<td>9.63</td>
<td>2.08 (1.27, 3.21)</td>
<td>&gt;12</td>
</tr>
<tr>
<td>no</td>
<td>breast CIS</td>
<td>MDS</td>
<td>F</td>
<td>13</td>
<td>22.32</td>
<td>0.58 (0.31, 1.00)</td>
<td>&gt;12</td>
</tr>
<tr>
<td>no</td>
<td>melanoma</td>
<td>MDS</td>
<td>M</td>
<td>14</td>
<td>26.31</td>
<td>0.53 (0.29, 0.89)</td>
<td>&gt;12</td>
</tr>
<tr>
<td>yes</td>
<td>NHL</td>
<td>MDS</td>
<td>F</td>
<td>8</td>
<td>2.39</td>
<td>3.35 (1.45, 6.60)</td>
<td>&gt;12</td>
</tr>
<tr>
<td>yes</td>
<td>NHL</td>
<td>MDS</td>
<td>M</td>
<td>10</td>
<td>3.15</td>
<td>2.90 (1.39, 5.33)</td>
<td>&gt;12</td>
</tr>
<tr>
<td>yes</td>
<td>thyroid</td>
<td>MDS</td>
<td>F</td>
<td>8</td>
<td>3.16</td>
<td>2.53 (1.09, 4.99)</td>
<td>&gt;12</td>
</tr>
<tr>
<td>yes</td>
<td>breast</td>
<td>MDS</td>
<td>F</td>
<td>71</td>
<td>53.28</td>
<td>1.33 (1.04, 1.68)</td>
<td>&gt;12</td>
</tr>
</tbody>
</table>

*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 >5 (with radiation) and >10 (without radiation) for the steady state interval of >12 y. *CIS = carcinoma in situ. Int= interval
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 multistage process of carcinogenesis and depends on age to drive the others. Traditionally, radiation-induced solid neoplasms have been modeled using relative risks and radiation-induced leukaemias have been modeled using additive risks. 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. 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. and Hsu et al. If cancer-therapy-induced AML risks are additive and not relative, then 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 $AR=(O-E)/PY$, that is, 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 prioritise 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. Regarding ratios of RR with versus without radiation as metrics of radiation association with second cancers, that breast and prostate cancer yield similar ratio magnitudes (Figures 2E-F) 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 is seen in the AML/MDS RR time courses after NHL (Figures 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. For example, Figure 2C suggests that for AML after breast first cancers, constant proportional hazards may not be unreasonable for comparing those treated with versus without radiation.

A logical next goal is risk decomposition. Risk component parameterizations, such as representing treatment-induced ERR as $Ae^{-kt}$ 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.

In summary, our results are as follows: (1) AML/MDS RR peak approximately two years after nonhaematological 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, unimodally for AML and bimodally 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, then breast cancer, then lymphoid cancers.
Supplementary Methods

S1. Cancer Definitions
Definitions of cancer types used are provided in Table S1. Haematological first cancers excluded from Figure 2A & 2B and Table 2 are the myeloid cancers AML, MDS, MDS/MPN, CMML, CML, and MPN (PV+ET+PMF), which are also excluded from the Tables 3 & 4, and the lymphoid cancers ALL, CLL, HCL (hairy cell leukaemia), HL (Hodgkin lymphoma), MM (multiple myeloma), NHL, and other leukaemias (OL, i.e. mixed- and no lineage leukaemias).

S2. Background incidences
SEER background incidence rates can involve age-year interactions (Figure S2). Interactions are included in our AML and MDS incidence models not to prove that interactions exist, but to control for them if they do.

S3. SEERaBomb
Installation
The most recent version of SEERaBomb can be installed from GitHub by typing

```r
library(devtools); install_github("radivot/SEERaBomb/SEERaBomb")
```

at the R prompt. This builds it from source, so Rtools is needed if using Windows.

Setup
Figure S3 depicts SEERaBomb manipulations used to merge ASCII files of 3 SEER databases into R binaries. The function mkSEER (Figure S3A) merges both cancer files (Figure S3B) and the following population files: populations/white_black_other/yr1973_2012.seer9/singleages.txt; populations/expanded.race.by.hispanic/yr1973_2012.seer9/singleages.txt; and populations/expanded.race.by.hispanic/yr1992_2012.seer9/singleages.txt. A subtlety here was that whilst SEER databases starting in 1973 and 2000 have population folders that match their cancer folders, the 1992 database population folder includes populations already in the 1973 folder and thus does not match its cancer folder; these population person years (PY) had to be removed from the 1992 population file before merging it with the 1973 and 2000 population files. To extend background incidence estimates to ages 85-99, the function mkSEER() redistributes PY in the 85+ age group of SEER population files to PY in 85-99 single-year bins. This was accomplished using National Vital Statistics Report US mortality rates of 2001 (Vol. 52, Issue 14) for males and females (with races pooled) to determine population proportions at each age; 2010 rates were not used because this introduced abrupt slope increases in incidence in Figure 2H at the age of 85 years, consistent with current mortality rates increasing with age at a lower rate than the average for the SEER population, which ranges over the years 1973-2012 (rates older than 2001 were not tested because they are not readily available as Excel files from...

Supplementary Figure S1. tAML (9920) and tMDS (9987) SEER coding.
A jump in tAML coding of males in 2010 is apparent. As tMDS coding fell to 0 in 2012, tAML coding of females increased too. To compensate for this misclassification of MDS as AML, line fits of tAML versus age over 2001-2009 (not shown) were used to predict 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 (with a fixed random number generator seed for reproducibility).

Supplementary Figure S2. Age-year interactions are apparent for NHL (left) and liver cancer (right).
SEERaBomb Software Validation Using Simulated SEER Data

Hypothetical cohorts of 1,000,000 individuals born each year between 1873 and 2012, who all live to ages of 100 years, and who have only two cancer possibilities A and B, were simulated. Cancer B was assumed to have a relative risk (RR) of 5 between 1 and 5 years after A. It was also assumed to be detectable 0.5 years earlier than otherwise if cancer A was detected first. Averages of 20 simulations are shown in Figure S4. In this figure the initial peak in B risk is due to early detection as a result of cancer A. Excess cases in this peak equal the number of missing cases in the subsequent trough, which lasts 0.5 years after diagnosis of A. Cancer inter-arrival times were assumed to be exponentially distributed with rates of 0.001/PY and 0.0002/PY for A and B, respectively, independent of age. Early detections of B were assumed to occur in, on average, 0.1 years of detection of A, also with an exponential distribution. Software was debugged until RR CI included 1 when RR = 1 was expected, as the number of simulations increased. No detectable biases remained.

Database Merging Improves Resolution

The impact of merging SEER databases on RR 95% confidence intervals (CI) is apparent in Figure S5. The competing software SEER*Stat MP-SIR (multiple primary standardised incidence ratios) currently provides access to either SEER-9 data or SEER-13 data (which clips off SEER-9 data prior to 1992 but uses data in 4 additional registries). Importantly, SEER*Stat MP-SIR does not provide access to 5 SEER registries that began in 2000 (see Figure 1 of the main text to appreciate the loss of cases and PY at risk). Because registries that began in 2000 contain more MDS cases than the other registries combined, statistical power that SEERaBomb provides by including them is particularly noticeable for MDS risks at short times since first cancer diagnoses (Figure S5). Other advantages of SEERaBomb include: synergy with R packages such as mgcv for splines, XLConnect for tables, and ggplot2 and rgl for graphics; availability across platforms; and open source on GitHub that enables code change tracking, collaboration, and public scrutiny.
The SEERaBomb workflow

SEERaBomb uses the following workflow: 1) use mkSEER to merge SEER ASCII data into R data binaries (Figure S3); 2) if needed, modify the cancer field, e.g. redefine AML to include APL; 3) if needed, modify the treatment field, e.g. if a treatment variable is wanted that differs from the default of having factor levels of no radiation (SEER codes 0 and 7), radiation (SEER codes 1-6), or unknown (SEER codes 8 and 9); 4) bundle cancer and PY data into a seerSet object using the function seerSet; 5) use mk2D to add background incidence fits to this object and use plot2D to view them; 6) use tsd (times since diagnosis) to compute RR at various user defined intervals after the first cancer; and 7) produce Excel files from the seerSet object outputted by tsd using the R function mkExcel, or extract results from the seerSet object using mkDF to produce RR time course plots. User interfaces of SEERaBomb’s R functions are described in SEERaBomb help pages.

S4. Japanese A-bomb survivor AML risk time course estimation

A-bomb survivor data analyses of Hsu et al. 556 yielded AML RR estimates of 4.6 (1.7, 11.2), 10.0 (4.5, 20.2), 4.1 (2.1, 7.7), 2.9 (1.4, 5.7), 2.8 (1.6, 5.0) and 2.5 (1.4, 4.4) for the time-since-exposure intervals (5,10], (10,15], (15,25], (25,35], (35,45], and (45,55] years. They thus predict a peak RR of ~10 in 10-15 years and a steady-state RR in the range of 2.5 to 4 after 15 years. We sought to both confirm this result and to increase its resolution on the time axis by considering 13 intervals instead of 6 as in Hsu et al. To do this, Poisson regression was used to fit the following model of expected AML cases to the A-bomb data:

\[ E = (BK(\text{age}) \cdot (1 + D^2 F_t)) \cdot \text{PY} \]  

where PY is the number of person-years; a quadratic dose-response in dose $D$ in Sv was chosen based on Hsu et al. 556; BK is a background incidence spline fitted by Poisson regression to data with doses <0.01 Sv using 10-year age interval bins with both sexes combined; and the 13 $F_t$ parameters correspond to the excess relative risks (ERR = RR - 1) predicted to exist at various time intervals after an exposure to 1 Sv. The parameters $F_t$ were fitted to 1950-2001 data (http://www.rerf.jp/library/dl_e/lsshempy.html) using mle2 of the R package bbmle, which was also used to form profile likelihood 95% CI. We then formed RR estimates by adding 1 to ERR estimates. Our results (Figure S6) confirm results in Hsu et al. Comments. The A-bomb haematological malignancy dataset has 14 time-since-exposure intervals, but the most recent interval has too few PY so it was merged with the penultimate interval to yield 13 $F_t$ parameters. Sexes were pooled because likelihood optimizations did not converge with separate sets of such parameters for each sex.

Supplementary Figure S5. Statistical power gained by using all SEER registries (SEER-18) relative to only the original 9 SEER registries (SEER-9).

Supplementary Figure S6. A-bomb survivor AML relative risk time course predicted for a 1 Sv exposure.

Plotted are maximum likelihood estimates and 95% profile likelihood limits of the 13 $F_t$ parameters of Eq. (S1) fitted to both cities (Nagasaki and Hiroshima) and sexes combined. The horizontal line is RR = 1.