Disclosing progress in cancer survival with less delay

Lina Jansen 1, Chloé C. Schröder 1, Katharina Emrich 2, Bernd Holleczech 1, Ron Pritzkuleit 4 and Hermann Brenner 1,5,6

1Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany
2Cancer Registry of Rhineland-Palatinate, Institute for Medical Biostatistics, Epidemiology and Informatics, University Medical Center, Johannes Gutenberg University Mainz, Mainz, Germany
3Saarland Cancer Registry, Saarbrücken, Germany
4Cancer Registry of Schleswig-Holstein, Lübeck, Germany
5Division of Preventive Oncology, German Cancer Research Center (DKFZ), and National Center for Tumor Diseases (NCT), Heidelberg, Germany
6German Cancer Consortium (DKTK), German Cancer Research Center, Heidelberg, Germany

Cancer registration plays a key role in monitoring the burden of cancer. However, cancer registry (CR) data are usually made available with substantial delay to ensure best possible completeness of case ascertainment. Here, we investigate empirically with routinely available data whether such a delay is mandatory for survival analyses or whether data can be used earlier to provide more up-to-date survival estimates. We compared distributions of prognostic factors and period relative survival estimates for three population-based CRs in Germany (Schleswig-Holstein (SH), Rhineland-Palatinate (RP), Saarland (SA)) computed on datasets extracted one (DY+1) to 5 years after the year of diagnosis (DY+5; reference). Analyses were conducted for seven cancer sites and various survival analyses scenarios. The proportion of patients registered in the datasets at a given time varied strongly across registries with 57% (SH), 2% (RP) and 26% (SA) registered in DY+1 and >93% in all registries in DY+3. Five-year survival estimates for the most recent three-year period were comparable to estimates from the reference dataset already in DY+1 (mean absolute deviations = 0.2–0.6% units). Deviations >1% units were only observed for pancreatic and lung cancer in RP and leukemia in SA (all ≤1.5% units). For estimates of 1-year survival based on the most recent 1-year period only, slightly longer delays were required, but reasonable estimates were still obtained after 1–2 years, depending on the CR and cancer site. Thus, progress in cancer survival could be disclosed in a more timely manner than commonly practiced despite delays in completeness of registration.

Introduction
High-quality population-based cancer registry (CR) data play a key role in monitoring the burden of cancer in a population over time. The estimation of population-based survival estimates from this data is an important tool to assess the overall effectiveness of health services in the management of patients with cancer. Due to its implications for health policy and health care planning, as well as for clinicians and patients as prognostic information, such monitoring should be as up-to-date as possible. However, there is often a substantial delay in data registration and release. For example, in the United States (US), the dataset of the Surveillance, Epidemiology and End Results (SEER) Program includes only patients with cancer diagnosed at least 3 years prior to publication of the dataset.1 This delay in the use and publication of data is introduced to guarantee sufficient completeness in the ascertainment of cancer cases. While completeness of registration is a prerequisite for deriving valid estimates of cancer incidence and mortality,
valid survival estimates might be obtained even before an almost complete ascertainment of cancer patients has been reached (albeit with increased standard errors due to the smaller sample size) as long as completeness of registration is unrelated to the prognosis of patients. Brenner and Hakulinen have investigated the impact of selective under-reporting of cancer patients on period and cohort survival estimates in a simulation study. While these simulations allow estimating the bias expected for survival estimates for specific scenarios of under-reporting, the extent to which delayed reporting is in fact selective and requires delay in the use of CR data for derivation of valid survival estimates in real-life settings of cancer registration has not been investigated.

Earlier use of data for survival analyses is particularly important for period analyses, as this method does not require a long follow-up to estimate long-term survival. With period analyses, reliable estimates of short- as well as long-term survival can be obtained 1 year after the year of diagnosis, as only 1-year follow-up information is needed for these patients. It has been extensively shown that the period estimate for a specific year of diagnosis closely predicts later observed survival of patients diagnosed during this year. Therefore, using period analyses, early use of data does not only allow estimation of short-term survival but can also allow more up-to-date estimates of long-term survival.

Here, we provide an approach to determine the required delay using routinely available CR data. We used annual CR database backups to investigate the impact of the length of delay on changes in patient characteristics and in period survival estimates. This approach was used to determine the required length of delay for various survival analyses scenarios for three population-based German cancer registries.

**Methods**

**Data sources**

In Germany, cancer registration has been implemented on a federal state level. While the "Federal Cancer Registry Data Act" (Bundeskrebsregisterdatengesetz) enacted in 1995 regularized the implementation, it lacked standards or specifications for certain structures. Consequently, each state developed its own registration system and CR law. To improve the comparability across German cancer registries, the Association of Population-based Cancer Registries in Germany was founded to develop methodological standards. However, there are key differences across registries. Here, we use data from three population-based cancer registries in Germany (Schleswig-Holstein, Rhineland-Palatinate and Saarland), which were chosen based on data availability. The cancer registries started registration in 1967 (Saarland), 1997 (Rhineland-Palatinate) and 1998 (Schleswig-Holstein) and together covered a population of 8.0 million people (9.6% of the total German population) in 2017. A key difference across these registries is that the CR of Schleswig-Holstein is the only one that has used an electronic notification system since its implementation, while Rhineland-Palatinate received approximately 20% of notifications and pathology reports electronically and other notifications by paper. The registries of Saarland and Rhineland-Palatinate only recently fully implemented electronic notification. With respect to follow-up, a linkage with all deceased persons in the federal state has been conducted in Saarland and Schleswig-Holstein. In Rhineland-Palatinate, it was implemented in autumn 2009 retrospectively for all years (of death) since 1998. Furthermore, in Rhineland-Palatinate linkage could only be made with death certificates with an indication of cancer until the 2014 dataset. While traceback of death certificate notified cases is routinely conducted in Saarland, in the other two cancer registries it has only been done for specific years, based on staff capacity.

For our analyses, each CR provided copies of their annually stored data backup. The data included all cancer patients registered since 1997 (Saarland) and 1998 (Rhineland-Palatinate, Schleswig-Holstein). Datasets were extracted for 2009–2016 in Schleswig-Holstein, 2008–2016 in Rhineland-Palatinate and 2007–2016 in Saarland. In addition, follow-up data were provided until the end of 2016 in Schleswig-Holstein and until end of 2013 in Rhineland-Palatinate and Saarland and were linked to each previous dataset using a persistent patient identifier.

Patients aged 15 or older with a diagnosis of breast (10th revision of the International Classification of Disease (ICD-10) C50), colorectal (C18–21), lung (C33–34), prostate (C61), pancreatic (C25) cancer, and hematological cancers (lymphomas: Hodgkin [C81] and non-Hodgkin lymphoma [C82–85], myeloma [C90.0] and leukemia [C91–96]) were included. These cancer sites were selected to include the most common tumors, tumors with short and long survival, and solid tumors.
as well as hematological cancers. According to common practice in population-based cancer survival analyses, cases notified by death certificate only were excluded.

The characteristics age at diagnosis, stage (for solid tumors) and subsite (for hematological cancers) were explored. Stage was grouped according to the recommendations of the European Network of Cancer Registry in localized/local spread, regional spread and advanced. In Saarland, a cancer diagnosis was originally coded according to the 2nd edition of the International Classification of Disease Oncology (ICD-O-2) and codes were subsequently converted to ICD-10 codes.

**Statistical analyses**

Analyses were conducted separately for each CR and each cancer site. For hematological cancers, survival scenarios were investigated separately for lymphoma and leukemia.

For each registry, five datasets were created, one for each length of delay between diagnosis and dataset extraction (1 year [DY+1] to 5 years [DY+5]). For example, the dataset DY+5 included all patients diagnosed in 2002, who were included in the CR in 2007 at the latest, while the dataset DY+1 included these patients only if they were already included in the CR in 2003. Depending on the availability of cancer patient data one to 5 years after diagnosis, the following years of diagnosis were included in the analyses: 2008–2011 in Schleswig-Holstein, 2007–2011 in Rhineland-Palatinate and 2006–2011 in Saarland. The dataset DY+5 was used as the reference assuming that sufficient completeness was achieved 5 years after the year of diagnosis. Using these datasets, changes in the number of patients and the distribution of age at diagnosis, stage (for solid tumors) and subtype (for hematological cancers) between the datasets DY+1 to DY+5 were explored to investigate when distributions of these prognostic factors start to be comparable to the reference distribution (DY+5).

In addition, 5-year age-standardized relative cohort survival was compared across dataset versions to assess whether prognosis is related to the timing of registration. For these analyses, the datasets had to be restricted to patients with linked follow-up information and to years of diagnosis with at least 5-year follow-up (Schleswig-Holstein: 2008–2011, Rhineland-Palatinate: 2007–2008, Saarland: 2006–2008).

Relative survival was computed as the ratio of the observed survival in the group of cancer patients divided by the expected survival of a comparable group in the general population. Expected survival was estimated by the Ederer II method using life tables stratified by age, sex and calendar period as obtained from the German Federal Statistical Office. Age standardization was conducted according to the International Cancer Survival Table 1. Proportions of death certificate only (DCO) notified cases for the following years after the year of diagnosis in percent

| Cancer registry      | Cancer site          | Registered in |         |
|----------------------|----------------------|---------------|---------|
|                      |                      | DY+1 | DY+2 | DY+3 | DY+4 | DY+5 |
| Schleswig-Holstein   | Colorectal cancer    | 2.2  | 5.1  | 10.7 | 10.5 | 10.1 |
|                      | Pancreatic cancer    | 19.0 | 29.2 | 44.1 | 43.1 | 41.0 |
|                      | Lung cancer          | 6.2  | 12.8 | 23.4 | 22.9 | 21.6 |
|                      | Breast cancer        | 2.1  | 4.1  | 9.1  | 8.9  | 8.7  |
|                      | Prostate cancer      | 2.4  | 4.4  | 9.8  | 9.7  | 9.4  |
|                      | Blood                | 8.3  | 12.0 | 21.1 | 19.9 | 18.9 |
|                      | Mean over cancer sites| 6.7  | 11.3 | 19.7 | 19.2 | 18.3 |
| Rhineland-Palatinate | Colorectal cancer    | 0.0  | 4.9  | 10.0 | 9.8  | 9.2  |
|                      | Pancreatic cancer    | 0.0  | 27.8 | 39.0 | 34.9 | 31.2 |
|                      | Lung cancer          | 0.0  | 15.8 | 24.3 | 21.3 | 18.0 |
|                      | Breast cancer        | 0.0  | 3.8  | 7.7  | 7.8  | 7.5  |
|                      | Prostate cancer      | 0.0  | 4.1  | 9.3  | 9.1  | 8.7  |
|                      | Blood                | 0.0  | 13.7 | 24.0 | 22.1 | 20.7 |
|                      | Mean over cancer sites| 0.0  | 11.7 | 19.1 | 17.5 | 15.9 |
| Saarland             | Colorectal cancer    | 0.0  | 2.5  | 3.5  | 3.3  | 3.3  |
|                      | Pancreatic cancer    | 1.6  | 19.3 | 21.7 | 19.7 | 19.3 |
|                      | Lung cancer          | 0.4  | 9.9  | 12.0 | 10.8 | 10.6 |
|                      | Breast cancer        | 0.0  | 1.1  | 1.9  | 1.9  | 1.9  |
|                      | Prostate cancer      | 0.0  | 1.2  | 2.1  | 2.1  | 2.1  |
|                      | Blood                | 0.0  | 7.1  | 9.3  | 8.8  | 8.6  |
|                      | Mean over cancer sites| 0.3  | 6.9  | 8.4  | 7.8  | 7.6  |

Abbreviations: DY, year of diagnosis.

---

Int. J. Cancer: 147, 838–846 (2020) © 2019 The Authors. *International Journal of Cancer* published by John Wiley & Sons Ltd on behalf of UICC
Standard using the weight for cancer sites with a steeply increasing age distribution (15–44 years: 7%, 45–54 years: 12%, 55–64 years: 23%, 65–74 years: 23%, 75+ years: 29%).

As earlier use of data is particularly interesting when using period analyses, we investigated various period survival analyses scenarios. Cancer survival statistics are typically reported for several recent calendar years and 5-year relative survival is the most frequently used survival estimate. Therefore, we estimated age-standardized 5-year period survival for 3-year calendar periods on datasets extracted one to 5-years after the period. For all cancer registries, person-times at risk, observed deaths, and expected deaths in the absence of cancer were computed for 2006–2008 and 2009–2011 separately on the datasets extracted 1–5 years after the last year of the 3-year period (Supporting Information Table S1). Then, estimates were pooled over the two periods to obtain one estimate for each CR, cancer site and delay. The estimates based on the datasets extracted 5 years after the year of diagnosis were used as the reference.

Analyses were additionally conducted for 1-year calendar periods, which are less commonly used, but might be relevant to investigate recent trends in cancer survival. Here, we investigated 1-year and 5-year period survival, pooled over all estimable 1-year period for each CR (Schleswig-Holstein: 2008–2011, Rhineland-Palatinate: 2007–2011, Saarland: 2006–2011). The approach is exemplified for five-year survival estimated for 1-year periods in 2008–2011 in Supporting Information Table S2.

All calculations were carried out with SAS software (version 9.3). Cohort and period survival estimates were computed using macros developed for registry analyses.

Data availability
The data that support the findings of our study are available from the cancer registries Schleswig-Holstein, Rhineland-Palatinate, and Saarland. Restrictions apply to the availability of these data, which were used under license for our study. Data are available from the authors with the permission of the cancer registries.

Results
In Schleswig-Holstein, 41,105 patients were registered and included in the data 5 years after the year of diagnosis (DY+5). As shown in Table 1, the proportion of DCO cases increased from DY+1 to DY+3 (mean over all cancer sites: 7 and 20%) and was stable or decreased afterward. Figure 1 shows the proportion of patients registered in DY+1 to DY+4 compared to the reference dataset (DY+5) after the exclusion of DCO cases. The overall proportion increases strongly from 57% in DY+1 to 89% in DY+2 and then slightly to 95% (DY+3) and 98% (DY+4). Notably, patients with hematological cancers were registered later with only 44% registered in DY+1. A similar tendency was observed for cancers with worse prognosis (pancreatic cancer: 51% and lung cancer: 54%).

The reference dataset of Rhineland-Palatinate included 63,874 patients. DCO cases were first included in DY+2 (12%), increased to 20% in DY+3 and decreased afterward, especially for the cancer sites with poor prognosis (lung and pancreatic cancer; Table 1). Only 2% of the patients were already registered in DY+1 and even in DY+2, only 58% were
Table 2. Period estimates of 5-year age-standardized relative survival for 3-year periods for the most complete dataset (DY+5) and differences to these estimates for the following years after the year of diagnosis for various cancer sites and registries

| Registry            | Cancer Site | All in DY+5 (SE) | Difference to DY+5 |
|---------------------|-------------|------------------|--------------------|
|                     |             |                  | DY+1  | DY+2  | DY+3  | DY+4  |
| Schleswig-Holstein  | Colorectum  | 65.9 (0.5)       | +0.6  | +0.5  | +0.1  | −0.1  |
|                     | Pancreas    | 10.1 (0.7)       | −0.2  | +0.3  | −0.2  | −0.3  |
|                     | Lung        | 17.2 (0.4)       | +0.1  | +0.6  | +0.1  | −0.2  |
|                     | Breast      | 86.5 (0.4)       | +0.8  | +0.6  | +0.2  | 0.0   |
|                     | Prostate    | 93.4 (0.5)       | +0.4  | +0.5  | +0.1  | −0.1  |
|                     | Lymphoma    | 69.8 (0.8)       | −0.2  | 0.0   | −0.1  | −0.2  |
|                     | Leukemia    | 58.8 (1.1)       | +0.2  | +0.7  | +0.1  | −0.7  |
| Mean absolute difference |      |                  | 0.4   | 0.5   | 0.1   | 0.2   |
| Rhineland-Palatinate| Colorectum  | 65.7 (0.4)       | +0.6  | +0.3  | 0.0   | 0.0   |
|                     | Pancreas    | 8.3 (0.6)        | −1.1  | −0.4  | −0.7  | −0.4  |
|                     | Lung        | 15.9 (0.4)       | −1.5  | −1.3  | −0.6  | −0.3  |
|                     | Breast      | 84.4 (0.3)       | +0.4  | +0.2  | 0.0   | −0.1  |
|                     | Prostate    | 92.1 (0.4)       | +0.3  | +0.2  | −0.1  | −0.1  |
|                     | Lymphoma    | 65.8 (0.8)       | +0.1  | +0.1  | −0.5  | −0.3  |
|                     | Leukemia    | 53.8 (1.0)       | −0.2  | −1.0  | −1.1  | −0.7  |
| Mean absolute difference |      |                  | 0.6   | 0.5   | 0.4   | 0.3   |
| Saarland            | Colorectum  | 64.9 (0.7)       | +0.7  | +0.1  | −0.1  | 0.0   |
|                     | Pancreas    | 9.3 (1.0)        | −0.5  | −0.8  | −0.9  | −0.5  |
|                     | Lung        | 17.1 (0.5)       | −0.1  | −0.5  | −0.5  | −0.2  |
|                     | Breast      | 83.5 (0.7)       | +0.3  | 0.0   | −0.2  | −0.1  |
|                     | Prostate    | 91.7 (1.1)       | +0.7  | +0.1  | −0.1  | −0.1  |
|                     | Lymphoma    | 62.8 (1.3)       | +0.4  | −0.2  | −0.2  | −0.1  |
|                     | Leukemia    | 45.3 (1.7)       | −1.1  | −1.0  | −1.4  | −0.6  |
| Mean absolute difference |      |                  | 0.5   | 0.4   | 0.5   | 0.2   |

Cells are shaded in light and dark gray if the absolute difference to DY+5 is larger than 1% unit and 3% units, respectively. Period estimates were estimated for the pooled 3-year periods 2006–2008 and 2009–2011.

Abbreviations: DY, year of diagnosis; SE, standard error.

Among hematological cancers, in all cancer registries, lymphoma cases were over and leukemia cases underrepresented in earlier datasets. Overall, deviations were mostly stronger or more long-lasting for pancreatic cancer and hematological cancers compared to other cancers. In Schleswig-Holstein, absolute differences for all factors were <3% units in DY+2 compared to DY+5 except for a lower proportion of advanced pancreatic cancer cases (~3.5% units). In Rhineland-Palatinate, in DY+2, absolute differences were larger than 3% units only for pancreatic cancer stage (+4.1% more advanced cancers) and 5-year cohort survival for hematological cancer (+4.7% units). In Saarland, all absolute differences were <3% units in DY+2.

Table 2 shows age-standardized 5-year period survival estimates for 3-year periods computed on datasets extracted 1 (DY+1) to 5 (DY+5) years after the last year included in the period. For all cancer registries and all cancer sites, absolute differences to DY+5 were already smaller than 1.6% units in DY+1, resulting in mean absolute differences of 0.4, 0.6 and
0.5% units in Schleswig-Holstein, Rhineland-Palatinate and Saarland, respectively. Absolute differences >1% units were only observed for pancreatic and lung cancer and leukemia in Rhineland-Palatinate and Saarland. A consistent pattern between the length of the delay and the absolute difference (>1% unit) was only observed for lung cancer in Rhineland-Palatinate with an underestimation of lung cancer survival by 1.5, 1.3, 0.6 and 0.3% units in DY+1 to DY+4.

When the period of analysis was shortened to 1 year, deviations generally increased (Table 3). Mean absolute differences to DY+5 started to be lower than 3% unit in DY+1 in Schleswig-Holstein and Saarland and in DY+2 in Rhineland-Palatinate. They were <1% unit 1 year later in each registry. Overall, differences were largest for pancreatic and lung cancer and leukemia. In Schleswig-Holstein, survival was underestimated for these cancer sites in DY+1 (−2.9, −1.5 and 0.27% units). These deviations started to be <1% unit in DY+2 (lung), DY+3 (pancreatic) and DY+4 (leukemia), respectively. In Rhineland-Palatinate, strong differences to DY+5 were observed in DY+1. In DY+2, absolute differences were <3% units for all cancer sites except leukemia (4.0% units) and further decreased in DY+3 and DY+4. In Saarland, absolute differences >1% units were only observed for colorectal cancer (+2.4% units), prostate cancer (+1.7% units) and leukemia (+3.7% units). These differences resolved in DY+2, but in DY+3 survival was slightly underestimated for pancreatic cancer (−1.2% units) and leukemia (−1.3% units).

Deviations overall increased when estimating 1-year instead of 5-year survival on 1-year calendar periods (Table 4). However, mean absolute differences to DY+5 were already <3% units in DY+1 and <1.0% units in DY+2 in Schleswig-Holstein (1.9 and 0.3% units) and Saarland (2.7 and 0.6% units). Absolute differences >3% units in DY+1 in Schleswig-Holstein and Saarland were observed for pancreatic cancer (−6.6 and −5.6% units), lung cancer (−3.2 and 3.1% units) and leukemia (+5.3% units). In Rhineland-Palatinate, survival estimates in DY+1 differed strongly from DY+5 with a mean absolute difference of 6.2% units. In DY+2 and DY+3, the mean absolute difference reduced to 2.2 and 1.0% units, respectively. Again, strongest differences were generally observed for pancreatic cancer, lung cancer and leukemia.

Table 3. Period estimates and standard error of 5-year age-standardized relative survival for 1-year period for the most complete dataset (DY+5) and differences to these estimates for the following years after year of diagnosis for various cancer sites and registries

| Registry         | Cancer Site | All in DY+5 (SE) | Difference to DY+5 |
|------------------|-------------|-----------------|--------------------|
|                  |             |                 | DY+1   | DY+2   | DY+3   | DY+4   |
| Schleswig-Holstein | Colorectum  | 67.2 (0.7)      | +1.0   | +0.1   | −0.1   | −0.1   |
|                  | Pancreas    | 12.4 (1.3)      | −2.9   | −1.0   | −0.5   | −0.2   |
|                  | Lung        | 18.4 (0.7)      | −1.5   | 0.0    | −0.2   | 0.0    |
|                  | Breast      | 87.2 (0.7)      | +0.3   | +0.2   | +0.1   | 0.0    |
|                  | Prostate    | 94.5 (0.6)      | +0.2   | 0.0    | 0.0    | −0.1   |
|                  | Lymphoma    | 71.6 (1.3)      | 0.0    | −0.5   | −0.4   | 0.0    |
|                  | Leukemia    | 60.4 (1.8)      | −2.7   | −1.5   | −1.6   | −0.9   |
| Mean absolute difference | Colorectum | 65.0 (0.5) | +6.9 | +1.1 | +0.2 | +0.3 |
|                  | Pancreas    | 8.2 (0.7)       | −0.2   | −0.7   | −0.9   | −0.2   |
|                  | Lung        | 15.8 (0.4)      | +2.3   | −1.1   | −1.2   | −0.2   |
|                  | Breast      | 84.9 (0.5)      | +2.3   | +0.7   | +0.2   | 0.0    |
|                  | Prostate    | 92.0 (0.5)      | +2.2   | +1.0   | +0.2   | +0.1   |
|                  | Lymphoma    | 65.8 (1.0)      | +7.4   | +1.6   | −0.3   | −0.1   |
|                  | Leukemia    | 54.2 (1.3)      | +16.1  | +4.0   | −1.2   | −0.7   |
| Mean absolute difference | Colorectum | 64.2 (0.8) | +2.4 | +0.8 | −0.1 | −0.1 |
|                  | Pancreas    | 9.3 (1.2)       | −0.7   | +0.1   | −1.2   | −0.4   |
|                  | Lung        | 17.4 (0.6)      | −0.9   | +0.1   | −0.8   | −0.3   |
|                  | Breast      | 83.7 (0.8)      | +1.0   | +0.4   | −0.1   | −0.1   |
|                  | Prostate    | 90.7 (1.4)      | +1.7   | +0.6   | −0.1   | −0.1   |
|                  | Lymphoma    | 62.4 (1.5)      | +0.9   | +0.4   | −0.3   | −0.3   |
|                  | Leukemia    | 43.9 (1.9)      | +3.7   | +0.5   | −1.3   | −0.4   |
| Mean absolute difference | Colorectum | 62.4 (0.8) | +1.5 | +0.4 | +0.6 | +0.2 |

Cells are shaded in light and dark gray if the absolute difference to DY+5 is larger than 1% unit and 3% units, respectively. Period estimates were estimates for 1-year period in 2008–2011 for Schleswig-Holstein, 2007–2011 for Rhineland-Palatinate and 2006–2011 for Saarland.

Abbreviations: DY, year of diagnosis; SE, standard error.
Our study shows that up-to-date period survival estimates can be obtained much earlier than they are using current standard practices. Standard survival estimates, such as 5-year relative period survival for the most recent 3-year periods, can be reliably estimated 1 year after the most recent year of diagnosis instead of 3 years of delay in the investigated cancer registries. For estimates of short-term (1-year) survival based on the most recent 1-year period only, slightly longer delay was required, but reasonable estimates were still obtained after 1–2 years, depending on the CR and cancer site.

Timeliness of CR data varies across countries and the required delay is usually estimated using completeness estimates. For example, within the Californian CR the median time from the date of diagnosis until availability for research was 382 days. In the Norwegian CR, incidence gets reported 1 year after the closing of a certain year of diagnosis. An underreporting of 2.2% of cases was estimated at this time point, which was considered to be acceptable for analyses to allow up-to-date use of the data. In Northern Ireland, first cancer statistics are published approximately 15 months after the latest incident year. Within the SEER program, member registries should report their data within 22–24 months after the diagnosis year. In Germany, CR data is usually made available 3 years after the year of diagnosis. Our study shows that standard period survival estimates can already be obtained with preliminary, incomplete CR data. For example, in Schleswig-Holstein, although only 57% of cases were registered in the first year after the year of diagnosis, the mean absolute difference between the 5-year period survival estimates (for the most recent 3-year calendar period) obtained from the dataset with 1 year compared to 5 years delay was 0.4% units with a maximum deviation of 0.6% units. Therefore, the choice of the required delay need not be interlinked with the completeness estimate but should be based on other measures to avoid unnecessary delays in data usage.

To the best of our knowledge, only one previous study has examined timing of registration and its impact on cohort survival. In our study from the Northern Ireland CR, late ascertainment was associated with worse prognosis for all cancers combined and with older age, late or unknown stage of disease and specific cancer sites such as small intestine, bone, leukemia and

| Registry          | Cancer Site | All in DY+5 (SE) | DY+1 | DY+2 | DY+3 | DY+4 |
|-------------------|-------------|------------------|------|------|------|------|
| Schleswig-Holstein| Colorectum  | 85.0 (0.4)       | +1.3 | +0.2 | -0.1 | 0.0  |
|                   | Pancreas    | 39.0 (1.4)       | -6.6 | 0.0  | -0.6 | -0.1 |
|                   | Lung        | 47.8 (0.7)       | -3.2 | +0.7 | -0.1 | +0.2 |
|                   | Breast      | 97.3 (0.3)       | +0.2 | +0.2 | +0.0 | 0.0  |
|                   | Prostate    | 98.8 (0.2)       | +0.2 | 0.0  | +0.0 | -0.1 |
|                   | Lymphoma    | 84.9 (0.8)       | -0.3 | -0.7 | -0.6 | -0.1 |
|                   | Leukemia    | 79.3 (1.2)       | -1.6 | -0.4 | -0.6 | -0.4 |

Mean absolute difference

| Registry          | Cancer Site | All in DY+5 (SE) | DY+1 | DY+2 | DY+3 | DY+4 |
|-------------------|-------------|------------------|------|------|------|------|
| Rhineland-Palatinate| Colorectum  | 84.7 (0.3)       | +5.7 | +0.7 | +0.2 | +0.2 |
|                   | Pancreas    | 35.5 (1.1)       | -4.1 | -5.7 | -2.9 | -0.1 |
|                   | Lung        | 44.2 (0.6)       | +5.2 | -3.5 | -2.0 | -0.2 |
|                   | Breast      | 96.0 (0.2)       | +2.4 | +0.3 | +0.3 | +0.1 |
|                   | Prostate    | 98.2 (0.3)       | +1.5 | +0.2 | +0.1 | 0.0  |
|                   | Lymphoma    | 80.9 (0.7)       | +8.1 | +1.8 | -0.1 | -0.1 |
|                   | Leukemia    | 73.2 (1.0)       | +16.6| +3.2 | +1.1 | -0.5 |

Mean absolute difference

| Registry          | Cancer Site | All in DY+5 (SE) | DY+1 | DY+2 | DY+3 | DY+4 |
|-------------------|-------------|------------------|------|------|------|------|
| Saarland          | Colorectum  | 83.0 (0.5)       | +2.7 | +0.7 | -0.1 | -0.1 |
|                   | Pancreas    | 30.8 (1.6)       | -5.6 | -1.0 | -2.9 | -0.7 |
|                   | Lung        | 42.5 (0.8)       | -3.1 | -0.5 | -1.4 | -0.5 |
|                   | Breast      | 95.8 (0.4)       | +0.5 | +0.2 | 0.0  | 0.0  |
|                   | Prostate    | 96.9 (0.4)       | +1.2 | +0.4 | 0.0  | 0.0  |
|                   | Lymphoma    | 78.3 (1.2)       | +0.7 | 0.0  | -0.5 | -0.2 |
|                   | Leukemia    | 65.0 (1.7)       | +5.3 | +1.1 | +1.1 | -0.3 |

Mean absolute difference

Discussion

Our study shows that up-to-date period survival estimates can be obtained much earlier than they are using current standard practices. Standard survival estimates, such as 5-year relative period survival for the most recent 3-year periods, can be reliably estimated 1 year after the most recent year of diagnosis instead of 3 years of delay in the investigated cancer registries. For estimates of short-term (1-year) survival based on the most recent 1-year period only, slightly longer delay was required, but reasonable estimates were still obtained after 1–2 years, depending on the CR and cancer site.

Timeliness of CR data varies across countries and the required delay is usually estimated using completeness estimates. For example, within the Californian CR the median time from the date of diagnosis until availability for research was 382 days. In the Norwegian CR, incidence gets reported 1 year after the closing of a certain year of diagnosis. An underreporting of 2.2% of cases was estimated at this time point, which was considered to be acceptable for analyses to allow up-to-date use of the data. In Northern Ireland, first cancer statistics are published approximately 15 months after the latest incident year. Within the SEER program, member registries should report their data within 22–24 months after the diagnosis year. In Germany, CR data is usually made available 3 years after the year of diagnosis. Our study shows that standard period survival estimates can already be obtained with preliminary, incomplete CR data. For example, in Schleswig-Holstein, although only 57% of cases were registered in the first year after the year of diagnosis, the mean absolute difference between the 5-year period survival estimates (for the most recent 3-year calendar period) obtained from the dataset with 1 year compared to 5 years delay was 0.4% units with a maximum deviation of 0.6% units. Therefore, the choice of the required delay need not be interlinked with the completeness estimate but should be based on other measures to avoid unnecessary delays in data usage.

To the best of our knowledge, only one previous study has examined timing of registration and its impact on cohort survival. In our study from the Northern Ireland CR, late ascertainment was associated with worse prognosis for all cancers combined and with older age, late or unknown stage of disease and specific cancer sites such as small intestine, bone, leukemia and
multiple myeloma. In our study, we also observed later registration for hematological cancer sites in all three registries but we did not find a consistent pattern between the length of delay and survival. It can be expected that results differ across countries and even across registries within the same country, as the registration processes differ. The Northern Ireland CR primarily uses electronic methods for case ascertainment, while among the included German CRs an electronic notification system was only used fully in Schleswig-Holstein during the time of the study. Furthermore, the Northern Ireland CR routinely conducts a traceback of death certificate notified cases, which was only routinely conducted in the Saarland CR. With a DCO proportion of 1% in 1993–2010, the case ascertainment is overall more complete in the Northern Ireland CR than in the investigated German CRs.16 These differences illustrate that results from our study should not be generalized to other CRs and it is important to investigate reporting patterns and, consequently, the required delay to obtain accurate survival estimates for each CR separately. However, our study exemplifies an easy way to estimate the required delay with routinely available data and could be easily conducted in cancer registries to avoid unnecessary delays in reporting of survival statistics.

Even within Germany, differences in registration practice across federal states exist that might cause different registration timing and, thus, might explain our observed intra-registry differences in the analyses. One main difference is that an electronic notification system was used fully only in Schleswig-Holstein during the time of the study. Such a system helps to accelerate the processes of registration in general, as paper forms do not have to be typed in a database and several linkages would be possibly easier and faster than before. Another CR in Germany that used an electronic notification system is the CR of North Rhine-Westphalia, which is able to report important epidemiological information for the state within 2 years.19,20 Our result support that a modernization to electronic systems might lead to an improvement of timeliness, as cases were registered earlier and deviations with shorter delays were generally smaller in Schleswig-Holstein compared to the other two cancer registries.

Despite differences in registration practice, some patterns were consistently found in all three registries. The proportion of DCO cases increased in all registries with the length of the delay during the first 3 years and decreased afterward. As information on date of diagnosis is usually not available for DCO cases and substituted by the year of death, the number of DCO cases cannot increase with the length of the delay as long as death certificates are immediately available and linked to the CR data. Thus, the pattern we observed is solely based on delayed availability of death certificates and/or delayed linkage with this information. Regarding survival, estimates computed on earlier datasets showed stronger bias or a more long-lasting bias for pancreatic, lung and hematological cancers than for the other cancer sites. Due to the consistency across registries, it can be hypothesized that these differences do not arise due to processes within registries but due to delayed notification by reporting sources. Pancreatic and lung cancer have the worst prognosis and, thus, a lower probability of being registered during the patient’s lifetime.21,22 Delayed reporting for patients with leukemia has previously been observed in the US23 and was explained by the dependence on the reporting of licensed practitioners, as these patients are often treated by practitioners and are not in clinics for surgery. However, it is not directly evident why the reporting patterns might differ for these sites, wherefore further insights into the data, especially with respect to the notification sources, are needed.

Our study has specific strengths and limitations. We provide an easy-to-implement approach to determine the required delay for up-to-date registry-based monitoring of cancer survival and illustrate its application for cancer registries with various practices of cancer registration and various delays in completeness of registration. Although this delay was found to be consistently shorter than the commonly practiced delay in using data for survival analyses across registries with strongly varying practices of cancer registration, these patterns should be confirmed by similar analyses for other countries or even other registries within Germany, as registration practices differ across cancer registries. Furthermore, these analyses should be repeated when registration processes change, as it is currently the case in Germany due to the implementation of the nationwide clinical cancer registration.4

A further strength of our study is the investigation of various years of diagnosis and many cancer sites including solid as well as hematological tumors and tumors with high and low survival rates. However, we have not investigated rare tumors, which might show a different notification pattern and we only investigated the timeliness of the reporting of the prognostic factors age and stage. Especially, for the evaluation of clinical CR data, an assessment of further tumor-related factors and treatment information will be required.

In conclusion, our study shows that CR data could be used much earlier for standard survival analyses as it is currently done: for standard period survival analyses in Germany, the delay could be shortened from 3 to 1 year, with some caution for pancreatic cancer, lung cancer and leukemia when estimating short-term survival or using short calendar periods. Routine implementation of the presented method for estimating the validity of the most recent survival estimates rather than exclusively relying on estimates of delays in completeness of cancer registration could help to avoid unnecessary delays in the monitoring of cancer survival.

Acknowledgements
We thank Dianne Pulte for her assistance in the English Language editing of the manuscript. The study was partly funded by the German Cancer Aid (grant numbers: 108257, 110446).

Conflict of interest
The authors declare that they have no conflict of interest.
References

1. Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) Research Data (1973–2015). National Cancer Institute, DCCPS, Surveillance Research Program, released April 2018, based on the November 2017 submission.

2. Brenner H, Hakulinen T. Population-based monitoring of cancer patient survival in situations with imperfect completeness of cancer registration. Br J Cancer 2005;92:576–9.

3. Brenner H, Hakulinen T. Up-to-date long-term survival curves of patients with cancer by period analysis. J Clin Oncol 2002;20:826–32.

4. Holleczek B, Katalinic A. Toward a comprehensive cancer registration in Germany. Eur J Cancer Prev 2017;26:S132–8.

5. Nennecke A, Barnes B, Brenner H, et al. Data quality or differences in oncological care? Standards of reporting for cancer survival analyses based on registry data. Gesundheitswesen 2013;75:94–8.

6. Nennecke A, Brenner H, Eberle A, et al. Cancer survival analysis in Germany—heading towards representative and comparable findings. Gesundheitswesen 2010;72:692–9.

7. World Health Organization. ICD-10: international statistical classification of diseases and related health problems, Tenth revision, 2nd edn. Geneva: World Health Organization, 2004.

8. Berrino F, Brown C, Möller T, et al. ENCR recommendations—condensed TNM for coding the extent of disease. Lyon, France: European Network of Cancer Registries, 2002.

9. Percy C, van Holten V, Muir CS, et al. International classification of diseases for oncology, 2nd edn. Geneva: World Health Organization, 1990.

10. Ederer F, Heise H. Instructions to IBM 650 programmers in processing survival computations. Bethesda, MD: National Cancer Institute, 1959.

11. Federal Statistical Office Germany. GENESIS-Online database. Life Tables (12621-0001). Wiesbaden: Federal Statistical Office Germany; 2019.

12. Corazziari I, Quinn M, Capocaccia R. Standard cancer patient population for age standardising survival ratios. Eur J Cancer 2004;40:2307–16.

13. Brenner H, Gefeller O, Hakulinen T. A computer program for period analysis of cancer patient survival. Eur J Cancer 2002;38:690–5.

14. Smith-Gagen J, Cress RD, Drake CM, et al. Factors associated with time to availability for cases reported to population-based cancer registries*. Cancer Causes Control 2005;16:449–54.

15. Larsen IK, Småstuen M, Johannesen TB, et al. Data quality at the cancer registry of Norway: an overview of comparability, completeness, validity and timeliness. Eur J Cancer 2009;45:1218–31.

16. Donnelly C, Cairnduff V, Chen JJ, et al. The completeness and timeliness of cancer registration and the implications for measuring cancer burden. Cancer Epidemiol 2017;49:101–7.

17. Bray F, Parkin DM. Evaluation of data quality in the cancer registry: principles and methods. Part I: comparability, validity and timeliness. Eur J Cancer 2009;45:747–55.

18. Bray F, Znaor A, Cueva P, et al. Planning and developing population-based cancer registration in low- and middle-income settings. IARC Technical Publication No. 43. Lyon: International Agency of Research on Cancer, 2014.

19. Jahresbericht 2016, Krebsgeschehen in Nordrhein-Westfalen 2014. Münster: Landeskrebsregister NRW; 2016.

20. Report 2012 mit Datenbericht 2010. Münster: Epidemiologisches Krebsregister NRW gGmbH; 2012.

21. Brenner H, Holleczek B. Deriving valid population-based cancer survival estimates in the presence of nonnegligible proportions of cancers notified by death certificates only. Cancer Epidemiol Biomarkers Prev 2011;20:2480–6.

22. Brenner H, Jansen L. Determinants and interpretation of death certificate only proportions in the initial years of newly established cancer registries. Eur J Cancer 2013;49:931–7.

23. Boscoe FP. Trends in reporting delay in United States central cancer registries. J Registry Manag 2014;41:135–8.