# BMJ Open

# Impact of loss-to-follow-up on cancer survival estimates for small populations: a simulation study using Hospital-Based Cancer Registries in Japan

Ayako Okuyama [ORCID],[1] Matthew Barclay,[2] Cong Chen,[3,4] Takahiro Higashi[1]

[1]Center for Cancer Control and Information Services, National Cancer Center, Tokyo, Japan
[2]Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK
[3]National Cancer Registration and Analysis Service, Public Health England, London, UK
[4]Health Data Insight, Cambridge, UK

**Correspondence to**
Dr Ayako Okuyama;
aokuyama@ncc.go.jp

## ABSTRACT

**Objectives** The accuracy of the ascertainment of vital status impacts the validity of cancer survival. This study assesses the potential impact of loss-to-follow-up on survival in Japan, both nationally and in the samples seen at individual hospitals.

**Design** Simulation study

**Setting and participants** Data of patients diagnosed in 2007, provided by the Hospital-Based Cancer Registries of 177 hospitals throughout Japan.

**Primary and secondary outcome measures** We performed simulations for each cancer site, for sample sizes of 100, 1000 and 8000 patients, and for loss-to-follow-up ranging from 1% to 5%. We estimated the average bias and the variation in bias in survival due to loss-to-follow-up.

**Results** The expected bias was not associated with the sample size (with 5% loss-to-follow-up, about 2.1% for the cohort including all cancers), but a smaller sample size led to more variable bias. Sample sizes of around 100 patients, as may be seen at individual hospitals, had very variable bias: with 5% loss-to-follow-up for all cancers, 25% of samples had a bias of <1.02% and 25% of samples had a bias of > 3.06%.

**Conclusion** Survival should be interpreted with caution when loss-to-follow-up is a concern, especially for poor-prognosis cancers and for small-area estimates.

## Strengths and limitations of this study

► The ascertainment of vital status impacts the validity of cancer survival.
► This study assessed impact of loss-to-follow-up for small populations using Hospital-Based Cancer Registries in Japan.
► The expected bias was not associated with the sample size, but a smaller sample size led to more variable bias.
► The probable range of the loss-to-follow-up rate was simulated without the tendency of incompleteness in the real data.
► Survival estimates must be interpreted with caution even with small rates of loss-to-follow-up, particularly for smaller samples.

## BACKGROUND

Survival statistics are the most-used measures to estimate cancer patients prognosisand the likely course of their disease, and are of great interest to patients, clinicians, researchers and policy makers.[1] Cancer survival estimates are employed by policymakers to compare cancer outcomes between different populations and time periods.[1] In Japan, the equalisation of cancer medical services is a key issue for cancer control strategy.[2] Hospital-level survival estimates would be useful for assessing progress toward equalisation of services, but interpreting these estimates requires an understanding of limitations, biases, underlying stochastic processes, and clinical and biological insight.[3]

In Japanese survival statistics of the Hospital-Based Cancer Registry (HBCR), data of hospitals that have over 90% of follow-up for their cancer patients were used for survival estimates (the high-completeness hospitals) in order to avoid overestimates of survival, as we do not have an official system to link death certificates with HBCR data for patients diagnosed before 2015. Loss-to-follow-up is still a potential limitation of Japanese cancer survival estimates: around 5.7% of patients diagnosed in 2007 were lost-to-follow-up even in the high-completeness hospitals included in the national reporting.[4] The National Cancer Center collects data from the HBCR of every designated cancer-care hospital, starting with cancer cases diagnosed in 2007.[4] In the HBCRs, patients' vital status during follow-up is obtained by matching hospital databases with resident cards held by provincial governments using patients' names and postal addresses and by recording deaths occurring in hospital. Loss-to-follow-up may occur when patients move around the country or when mis-recording of patients' names or postal addresses leads to linkage failure. For these patients, survival was censored at their

most recent visit to the hospital, an approach giving accurate results only if the reasons they were lost-to-follow-up were not related to survival.

Use of cancer survival estimates to assess equalisation of cancer medical services naturally involves comparing estimates based on the relatively small number of patients treated by individual hospitals. Previous studies exploring the impact of loss-to-follow-up on survival estimates in a large sample (national survival statistics) focused on the expected impact[5–12] and demonstrated that even modest levels of under-registration of deaths might lead to severe overestimation of long-term survival. To understand the possible impact of loss-to-follow-up on results for small areas and individual hospitals, it is important to describe the variance of the bias as well as the expected bias. This study uses Monte Carlo simulations to assess the impact of loss-to-follow-up on Japanese cancer survival estimates for different sample sizes to represent the variation resulting when reporting at hospital and other scales.

## METHODS

### Data source

A subset of incident cases diagnosed in 2007 in the HBCR of 177 designated cancer-care hospitals throughout Japan was used for the analysis.

The inclusion criteria were:

1. Patients aged 15 years or older at diagnosis.
2. Patients received the first course of treatment at the hospital.
   a. This prevents duplicate records of patients treated at multiple hospitals.
3. Patients had 5 years of complete follow-up after diagnosis. Therefore, we only used data of known follow-up cases for simulation (complete cases).

In total, 163 084 cancer cases were included: 25 890 cases for stomach, 20 603 cases for colon and rectum, 8694 cases for liver, 20 623 cases for lung, 14 764 cases for female breast and 72 510 for other cancers. The mean age of cancer cases was 68.4 years (SD 10.9) for stomach, 68.0 years (SD 11.5) for colorectal, 69.0 years (SD 9.6) for liver, 69.2 years (SD 10.1) for lung and 58.2 years (SD 13.0) for female breast.

### Data generating mechanism

Sampling with replacement was used to generate 10 000 datasets for each of the 108 permutations of loss-to-follow-up rates, cancer site and sample size discussed below.

1. Loss-to-follow-up rate: Loss-to-follow-up rates of 1%, 2%, 3%, 4% and 5% were considered, as well as an empirical estimate based on real loss-to-follow-up rates in the high-completeness hospital data of 2007 (6.3% for those aged 15–39; 3.8% for 40–49; 2.7% for 50–59; 2.4% for 60–69; 3.0% for 70–79; and 5.0% for those aged 80 or older). When we simulated the loss-to-follow-up rate of 1% up to 5%, we randomly selected loss-to-follow-up cases across the whole cancer-specific

cohorts. In age-specific simulations, we selected these cases randomly after stratifying by age group.
2. Cancer site: Survival estimates were produced for each major cancer site (stomach, colorectal, liver, lung and female breast) individually, as well as for all cancers combined.
3. Sample size: Samples of 100, 1000 and 8000 patients were considered, representing potential incidence at different levels of geography (eg, an individual hospital at 100 patients and Japan at 8000 or more).

Loss-to-follow-up was applied by marking patients identified as 'lost to follow-up' as alive regardless of their true vital status. The probability of loss-to-follow-up in this simulation study was not related to the survival time of the patient.

### Performance measure

We assessed performance using bias in the 5-year survival estimate due to loss-to-follow-up; it was calculated as the difference between the Kaplan-Meier estimators based on the accurate survival data and on the data with loss-to-follow-up in each of the 10 000 datasets for each of the 108 permutations.

The average bias across the 10 000 datasets, the empirical SD and the IQR of the bias were assessed.

All analyses were performed using Stata V.14.0 (Stata Corporation).

## RESULTS

### Survival for complete cases

The 5-year crude survival rate of all cancers for complete cases was 56.6% (95% CI, 56.4% to 56.8%). The 5-year crude survival rate for complete cases was 61.5% (95% CI, 60.9% to 62.1%) for stomach, 62.9% (95%CI, 62.2% to 63.5%) for colorectal, 31.0% (95%CI, 30.1% to 32.0%) for liver, 34.0% (95%CI, 33.4% to 34.6%) for lung and 88.2% (95%CI, 87.6% to 88.7%) for female breast. Figure 1 shows the result of simulation of loss-to-follow-up rates of 1%, 2%, 3%, 4% and 5%. Figure 2 shows the result of simulation of empirical estimates based on real loss-to- follow-up rates in the high-completeness hospital data of 2007.

### Sample size

The expected bias was not associated with the sample size (figure 1), but a smaller sample size led to a more variable bias. Sample sizes of around 100 patients, as may be seen at individual hospitals, had a very variable bias: with 5% loss-to-follow-up for all cancers, 25% of samples had a bias of <1.02% and 25% of samples had a bias of >3.06%. This appears to be exacerbated for sites with poor prognosis (figure 1; table 1). In liver cancer, 4.76% of estimates have >5%-points of error in samples of 100 patients. With larger samples, particularly with 8000 or more cases, the variation in bias becomes inconsequential.

### Loss-to-follow-up rate

Higher loss-to-follow-up rates gave more bias in the survival estimates (figure 1; online supplementary appendix table
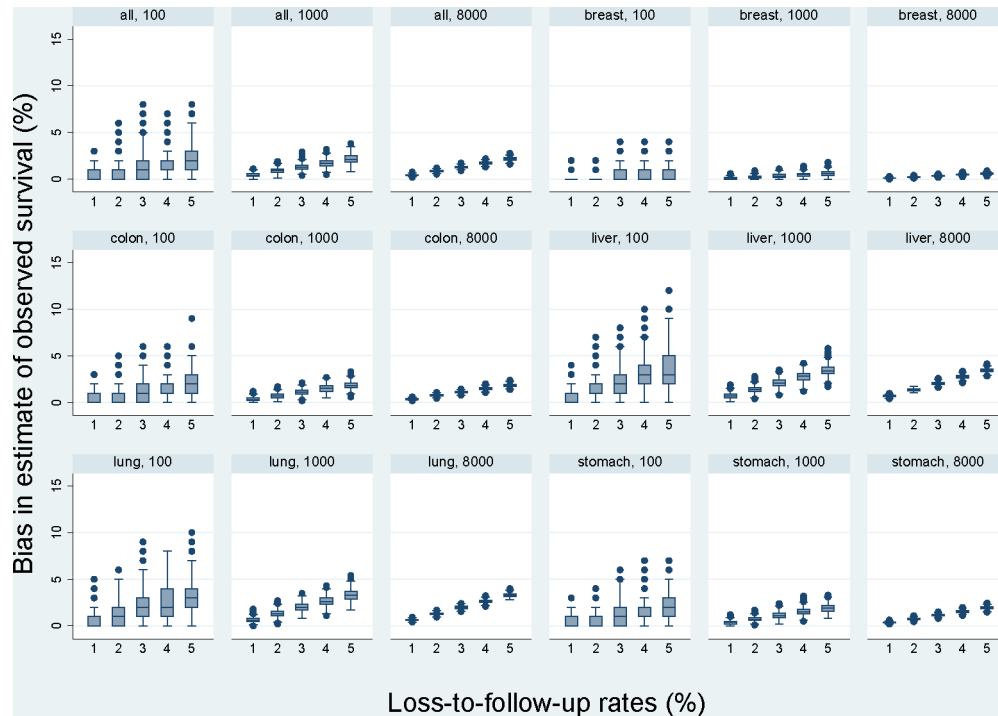
**Figure 1** Distribution of bias resulting from 1% up to 5% loss-to-follow-up by different sample sizes.

1), but had less impact on the spread of the bias. If we look at liver cancer, the expected bias with 1% loss-to-follow-up in a sample of 1000 was 0.71% with an SD of <0.27%, but with 5% loss-to-follow-up it was 3.49% with an SD of 0.59%.

### Cancer site

Loss-to-follow-up introduced more bias and more variable bias for cancer sites with poor prognosis (figures 1 and 2; table 1). The expected bias with loss-to-follow-up, similar to that observed empirically for lung cancer, was around



**Figure 2** Distribution of bias resulting from observed rates of loss-to-follow-up by different sample size. Empirical estimate based on real loss-to-follow-up rates in the high-completeness hospital data of 2007 (6.3% for those aged 15–39; 3.8% for 40–49; 2.7% for 50–59; 2.4% for 60–69; 3.0% for 70–79; and 5.0% for those aged 80 or older).
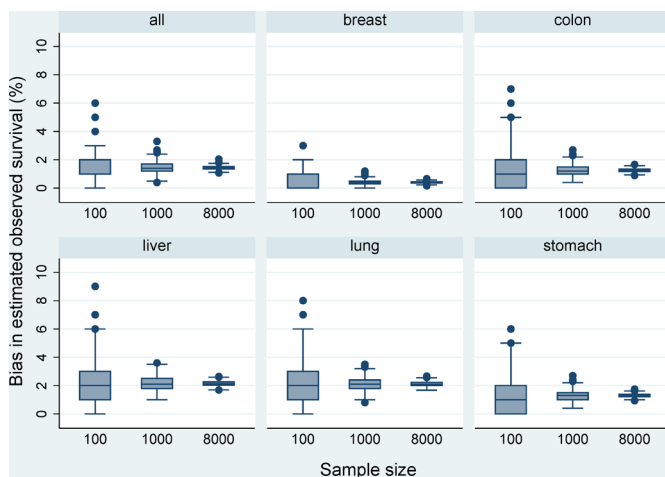
2.12% (SD 0.16%–1.43%), while for breast cancer it was only 0.40% (SD 0.07%–0.63%).

### DISCUSSION

Loss-to-follow-up rates consistent with those seen in Japan in 2007 incidence data lead to highly variable bias in survival estimates for small populations. Hospital-level survival comparisons should be avoided, as some loss-to-follow-up is likely to introduce a large overestimation of survival rates for some hospitals but not for others, even if the loss-to-follow-up rate is the same for all hospitals. Previous studies have demonstrated that higher loss-to-follow-up leads to survival estimates with more bias, especially for cancers with poor prognosis;[9] this study also found similar trends in both large and small samples.

Our empirical data shows that a higher proportion of patients aged 15–39 years and over 80 years were lost to follow-up than in other groups. Follow-up currently relies on linking patients to their resident card to ascertain their vital status. Loss-to-follow-up among younger people may be due to patients marrying and changing their names or due to moving to other places. Lower follow-up for elderly patients may be due to their moving more frequently after discharge from the hospital because of difficulties living alone, for example, into a nursing home. The number of older patients has recently been increasing in the designated cancer-care hospitals.[13] This will make it difficult to estimate unbiased survival for cancer patients without any official system to follow-up for cancer patients.

Our study has several limitations. First, we considered a probable range for the loss-to-follow-up rate for the

**Table 1** Summary results across 10 000 simulations for each cancer site, with loss-to-follow-up probabilities by age group based on those seen in real data

| Cancer site | 'True' 5-year survival rate (%) | Loss-to-follow-up | Sample size | Expected bias (%) | SD of the bias (%) | IQR of bias (%) | Percentage with bias of 5%-points or more (%) |
|---|---|---|---|---|---|---|---|
| All | 56.6 | Real | 100 | 1.45 | 1.19 | 1.00–2.00 | 1.30 |
| All | 56.6 | Real | 1000 | 1.45 | 0.38 | 1.20–1.70 | <0.01 |
| All | 56.6 | Real | 8000 | 1.44 | 0.13 | 1.35–1.53 | <0.01 |
| Breast | 88.2 | Real | 100 | 0.40 | 0.63 | 0.00–1.00 | 0.01 |
| Breast | 88.2 | Real | 1000 | 0.40 | 0.20 | 0.30–0.50 | <0.01 |
| Breast | 88.2 | Real | 8000 | 0.40 | 0.07 | 0.35–0.45 | <0.01 |
| Colon | 62.9 | Real | 100 | 1.28 | 1.11 | 0.00–2.00 | 0.82 |
| Colon | 62.9 | Real | 1000 | 1.26 | 0.35 | 1.00–1.50 | <0.01 |
| Colon | 62.9 | Real | 8000 | 1.26 | 0.12 | 1.18–1.35 | <0.01 |
| Liver | 31.0 | Real | 100 | 2.16 | 1.44 | 1.00–3.00 | 4.76 |
| Liver | 31.0 | Real | 1000 | 2.15 | 0.46 | 1.80–2.50 | <0.01 |
| Liver | 31.0 | Real | 8000 | 2.15 | 0.16 | 2.04–2.26 | <0.01 |
| Lung | 34.0 | Real | 100 | 2.11 | 1.43 | 1.00–3.00 | 4.18 |
| Lung | 34.0 | Real | 1000 | 2.12 | 0.46 | 1.80–2.40 | <0.01 |
| Lung | 34.0 | Real | 8000 | 2.12 | 0.16 | 2.01–2.22 | <0.01 |
| Stomach | 61.5 | Real | 100 | 1.29 | 1.12 | 0.00–2.00 | 0.88 |
| Stomach | 61.5 | Real | 1000 | 1.29 | 0.36 | 1.00–1.50 | <0.01 |
| Stomach | 61.5 | Real | 8000 | 1.29 | 0.13 | 1.20–1.38 | <0.01 |

*Real loss-to-follow-up means that empirical estimate based on real loss-to-follow-up rates in the high-completeness hospital data of 2007 (6.3% for those aged 15–39; 3.8% for 40–49; 2.7% for 50–59; 2.4% for 60–69; 3.0% for 70–79 and 5.0% for those aged 80 or older).

simulation work. We do not know the incompleteness of death ascertainment in the real data, which may affect the true survival rate used as a baseline, and so we cannot estimate true survival from this simulation. Second, we used data for patients diagnosed in 2007. This data is the data collected in first year in Japan by HBCR. Cancer registries may have lacked experience gathering data items needed to estimate survival, and some data, such as follow-up-days or last-contact date, may have been miss-recorded. Third, when patients are known to be lost-to-follow-up in real data, their survival is censored at the date they were last known to have been alive. Such loss-to-follow-up is likely to be informative censoring, and our results could be viewed as an approximate upper bound on the bias. Finally, this simulation only considers absolute survival. Previous studies report that, in general, relative survival is much more affected by incomplete registration of deaths than absolute survival, and potential problems are much larger for relative survival estimates in older patients compared with younger.[7] Even more care will be needed when interpreting relative survival estimates.

## CONCLUSION
Loss-to-follow-up of the level currently observed in HBCR in designated cancer-care hospitals in Japan introduces substantial and variable bias into cancer survival estimates for small samples. Until we achieve >99%

completeness in ascertainment of survival status, we must interpret survival estimates with caution, particularly for smaller samples (eg, individual hospitals). An official follow-up system is required to support unbiased estimation of survival.

## Patient and public involvement
This study is based on historical hospital-level data collected for the purpose of national disease registration. The research question and outcome measures are aimed at assessing standard epidemiological outputs, and patients were not involved in this study.

**ORCID iD**
Ayako Okuyama http://orcid.org/0000-0003-4732-3272

## REFERENCES

1  Mariotto AB, Noone A-M, Howlader N, *et al*. Cancer survival: an overview of measures, uses, and interpretation. *J Natl Cancer Inst Monogr* 2014;2014:145–86.
2  Sobue T. Evaluation of equalization of cancer medical services in Japan using public database in Japan. *J Epidemiol* 2018;28.
3  Dickman PW, Adami HO. Interpreting trends in cancer patient survival. *J Intern Med* 2006;260:103–17.
4  Nishimoto H, Shibata A, Okuyama A. The hospital-based cancer registries in Japan: cancer survival at the designated cancer care hospitals in 2007, 2015. Available: https://ganjoho.jp/reg_stat/statistics/brochure/hosp_c_reg_surv.html [Accessed 7th Oct 2019].
5  Sriamporn S, Swaminathan R, Parkin DM, *et al*. Loss-adjusted survival of cervix cancer in Khon Kaen, Northeast Thailand. *Br J Cancer* 2004;91:106–10.
6  Swaminathan R, Rama R, Shanta V. Lack of active follow-up of cancer patients in Chennai, India: implications for population-based survival estimates. *Bull World Health Organ* 2008;86:509–15.
7  Brenner H, Hakulinen T. Implications of incomplete registration of deaths on long-term survival estimates from population-based cancer registries. *Int J Cancer* 2009;125:432–7.
8  Johnson CJ, Weir HK, Yin D, *et al*. The impact of patient follow-up on population-based survival rates. *J Registry Manag* 2010;37:86–103.
9  Ganesh B, Swaminathan R, Mathew A, *et al*. Loss-adjusted hospital and population-based survival of cancer patients. *IARC Sci Publ* 2011;162:15–21.
10  Rutherford MJ, Møller H, Lambert PC. A comprehensive assessment of the impact of errors in the cancer registration process on 1- and 5-year relative survival estimates. *Br J Cancer* 2013;108:691–8.
11  Weir HK, Johnson CJ, Mariotto AB, *et al*. Evaluation of North American association of central cancer registries' (NAACCR) data for use in population-based cancer survival studies. *J Natl Cancer Inst Monogr* 2014;2014:198–209.
12  Pinheiro PS, Morris CR, Liu L, *et al*. The impact of follow-up type and missed deaths on population-based cancer survival studies for Hispanics and Asians. *J Natl Cancer Inst Monogr* 2014;2014:210–7.
13  Okuyama A, Higashi T. Patterns of cancer treatment in different age groups in Japan: an analysis of hospital-based cancer registry data, 2012-2015. *Jpn J Clin Oncol* 2018;48:417–25.