Annals of
SURGICAL ONCOLOGY
OFFICIAL JOURNAL OF THE SOCIETY OF SURGICAL ONCOLOGY

ORIGINAL ARTICLE – GLOBAL HEALTH SERVICES RESEARCH

Evaluation of the Impact of the COVID-19 Pandemic on Reliability of Cancer Surveillance Data in the National Cancer Database

Leticia M. Nogueira, PhD, MPH[1], Bryan Palis, MA[2], Daniel Boffa, MD, MBA[3], Sharon Lum, MD, MBA[4], K. Robin Yabroff, PhD[1], and Heidi Nelson, MD, FACS[2]

[1]Surveillance and Health Equity Sciences, American Cancer Society, Kennesaw, GA; [2]Commission on Cancer, American College of Surgeons, Chicago, IL; [3]Division of Thoracic Surgery, Yale School of Medicine, New Haven, CT; [4]Department of Surgery, Loma Linda University School of Medicine, Loma Linda, CA

ABSTRACT

Purpose. This study evaluated the reliability of cancer cases reported to the National Cancer Database (NCDB) during 2020, the first year of the COVID-19 pandemic.

Methods. Total number of cancer cases reported to the NCDB between January 2018 and December 2020 were calculated for all cancers and 21 selected cancer sites. The additive outlier method was used to identify structural breaks in trends compared with previous years. The difference between expected (estimated using the vector autoregressive method) and observed number of cases diagnosed in 2020 was estimated using generalized estimating equation under assumptions of the Poisson distribution for count data. Interrupted time series analysis was used to compare changes in the number of records processed by registrars each month of 2020. All models accounted for seasonality, regional variation, and random error.

Results. There was a statistically significant decrease (structural break) in the number of cases diagnosed in April 2020, with no recovery in number of cases during subsequent months, leading to a 12.4% deficit in the number of cases diagnosed during the first year of the pandemic. While the number of cancer records initiated by cancer registrars also decreased, the number of records marked completed increased during the first months of the pandemic.

Conclusion. There was a significant deficit in the number of cancer diagnoses in 2020 that was not due to cancer registrars' inability to extract data during the pandemic. Future studies can use NCDB data to evaluate the impact of the pandemic on cancer care and outcomes.

In March 2020, soon after the World Health Organization declared the coronavirus disease 2019 (COVID-19) outbreak a pandemic, US states and territories implemented policies and strategies limiting person-to-person interactions to reduce transmission in the community and to accommodate the surge of emergent healthcare needs. These measures led to a reduction in the number of cancer screenings, diagnoses, and treatments measured in healthcare encounters,[1] Medicare claims,[2] and electronic pathology reports.[3]

Although healthcare encounters, medical claims, and pathology reports can be used as proxies for the number of individuals screened, newly diagnosed, or treated for cancer, cancer data collected by certified tumor registrars following registry standards are the gold standard for accurately evaluating the impact of the pandemic on cancer care and outcomes.[4]

Importantly, the COVID-19 pandemic and associated community mitigation strategies might have impacted both the number of individuals diagnosed with cancer and the ability of cancer registrars to collect and report cancer data.

The aim of this study was to determine whether changes in the number of cancer cases reported to the National Cancer Database (NCDB) took place during 2020, the first
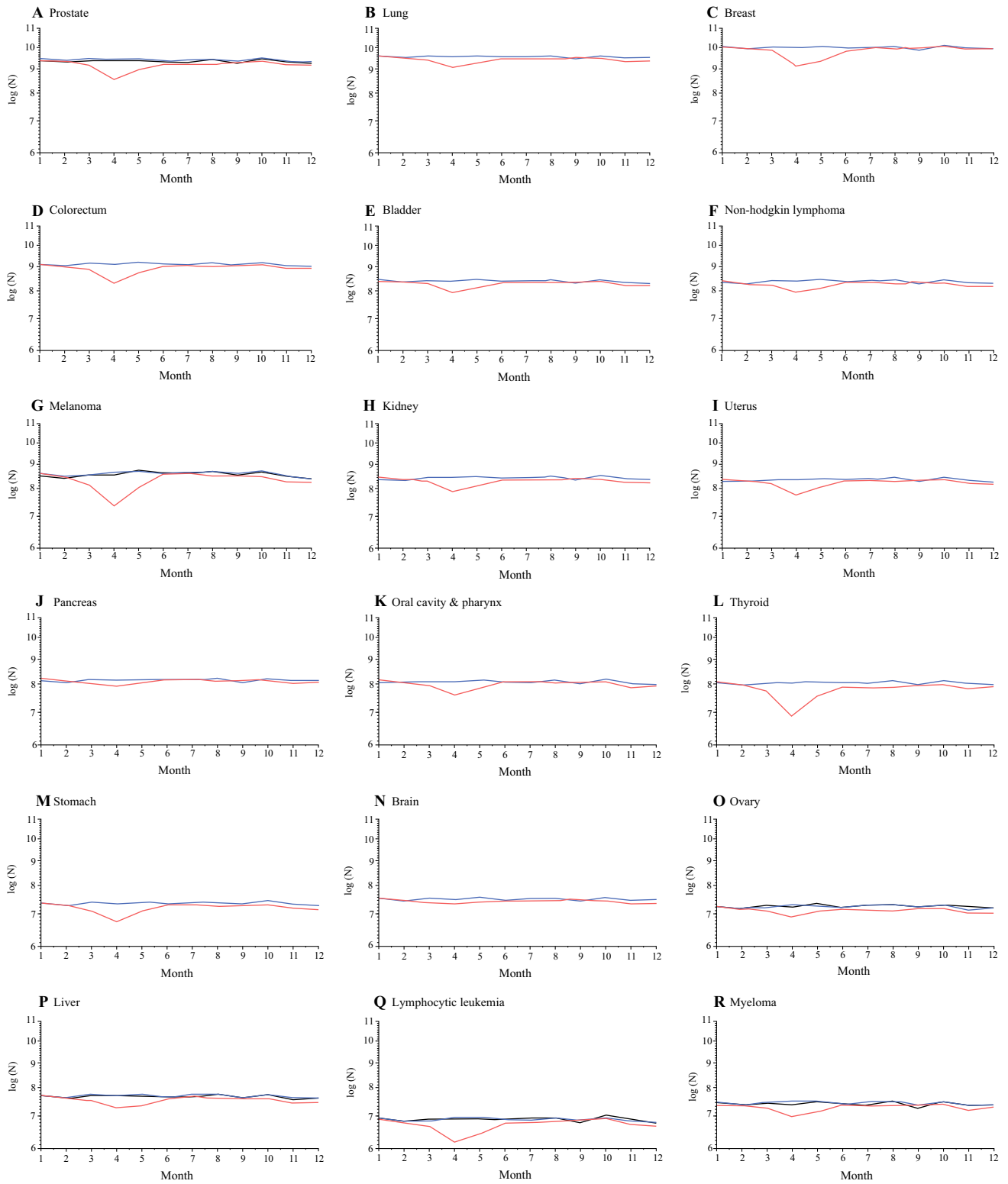
FIG. 1 Decrease in expected number of cancer cases diagnosed in 2020
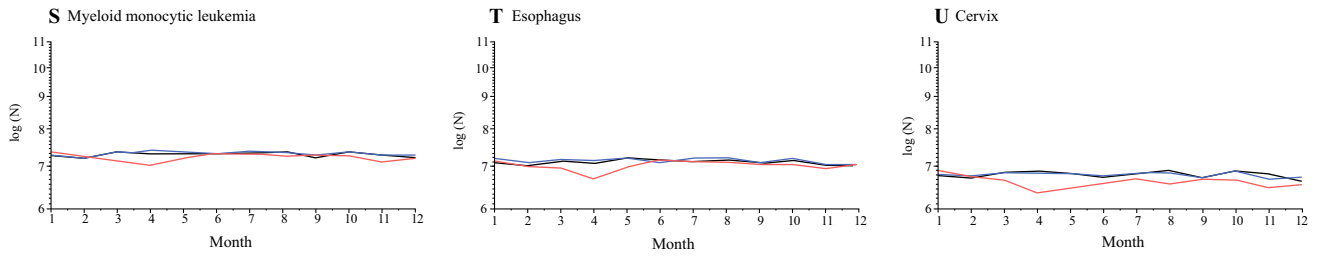
FIG. 1 continued

year of the COVID-19 pandemic, and if identified, whether changes were significant, reflect alterations in cancer diagnoses, or limitations in in registrars' ability to report data. Answering these questions is critical prior to incorporating 2020 data from the NCDB into research studies of people in the USA diagnosed with cancer.

METHODS

The NCDB is a nationwide hospital-based cancer registry, which includes approximately 70% of all newly diagnosed cancers in the USA.[5] NCDB data elements are abstracted by certified tumor registrars who undergo formal training through the National Cancer Registrars

Association. Variable definitions are standardized with other cancer registries and participating sites undergo periodic data audits to ensure data reliability. Although the current data submission to the NCDB is real time (registrars continuously submit cases), registrars update submissions and add new information (e.g., treatment, date of last contact) to cases diagnosed in prior months. The last date of data input to the NCDB that was considered for the 2020 diagnosis year was 15 March 2022.

To evaluate changes in the number of cancer cases diagnosed in 2020, we first calculated total number of cancer diagnoses reported to the NCDB between January 2018 and December 2020. We then used the additive outlier method to identify changes (structural breaks) in

TABLE 1 Identified outliers (structural breaks) from number of cases expected to be diagnosed each month in 2020, NCDB

Cancer site	Structural break month year	Structural break estimate (SE)	p-value
All	April 2020	−0.31 (0.09)	0.0002
Prostate	April 2020	−0.52 (0.11)	< 0.0001
Lung	April 2020	−0.22 (0.07)	0.002
Breast	April 2020	−0.45 (0.13)	0.0003
Colon and rectum	April 2020	−0.46 (0.10)	< 0.0001
Bladder	April 2020	−0.29 (0.08)	0.0003
Non-Hodgkin lymphoma	April 2020	−0.27 (0.08)	0.0008
Melanoma	April 2020	−0.69 (0.16)	< 0.0001
Kidney	April 2020	−0.30 (0.08)	0.0002
Uterus	April 2020	−0.49 (0.11)	< 0.0001
Pancreas	April 2020	−0.17 (0.06)	0.007
Oral cavity and pharynx	April 2020	−0.42 (0.10)	< 0.0001
Thyroid	April 2020	−0.79 (0.16)	< 0.0001
Stomach	April 2020	−0.25 (0.07)	0.0008
Ovary	April 2020	−0.26 (0.07)	0.0003
Lymphocytic leukemia	April 2020	−0.41 (0.11)	0.0001
Myeloma	April 2020	−0.26 (0.08)	0.002
Myeloid and monocytic leukemia	April 2020	−0.22 (0.08)	0.003
Esophagus	April 2020	−0.34 (0.08)	< 0.0001
Cervix	April 2020	−0.34 (0.10)	0.0007

The additive outlier method was used to identify structural breaks in natural logarithm of number of cancer cases diagnosed each month and year between January 2017 and December 2020 accounting for seasonality, regional variation, and random error

TABLE 2 Absolute decrease in number of cancer diagnoses and deficit in cancer cases reported to the NCDB compared with expected in 2020

Cancer site	Expected[a]		Observed	Difference[b]		
	N	SE	N	N	%	p-value
Total	1,405,412	4304.0	1,231,119	174,293	12.4	< 0.0001
Prostate	145,382	716.6	117,033	28,349	19.5	0.00
Lung	169,864	524.0	145,618	24,246	14.3	0.02
Breast	266,381	953.4	231,970	34,411	12.9	0.08
Colon and rectum	108,152	277.1	92,212	15,940	14.7	0.03
Bladder	52,977	195.9	47,292	5685	10.7	0.08
Non-Hodgkin lymphoma	52,181	177.2	45,917	6264	12.0	0.03
Melanoma	65,463	266.1	51,957	13,506	20.6	0.04
Kidney	55,406	196.5	47,405	8001	14.4	0.02
Uterus	51,191	215.5	45,214	5977	11.7	0.08
Pancreas	42,408	166.4	39,216	3192	7.5	0.04
Oral cavity and pharynx	39,357	148.0	35,281	4076	10.4	0.07
Thyroid	37,376	129.2	29,920	7456	19.9	0.02
Stomach	18,829	66.1	16,165	2664	14.1	0.03
Brain	21,805	55.3	20,017	1788	8.2	0.01
Ovary	16,702	62.5	14,747	1955	11.7	0.003
Liver	26,598	100.0	22,776	3822	14.4	0.02
Lymphocytic leukemia	11,774	61.2	10,057	1717	14.6	0.02
Myeloma	20,526	103.5	17,724	2802	13.7	0.02
Myeloid and monocytic leukemia	18,229	83.3	16,520	1709	9.4	0.06
Esophagus	15,100	67.3	13,337	1763	11.7	0.02
Cervix	10,682	42.8	8941	1741	16.3	0.02

N: number of cases, *SE*: standard error

[a]The expected number of cancer diagnoses in 2020 was estimated using the vector autoregressive method accounting for seasonality and random error.

[b]The difference between expected and observed number of cases diagnosed in 2020 was estimated with logistic regressions using the Poisson distribution.
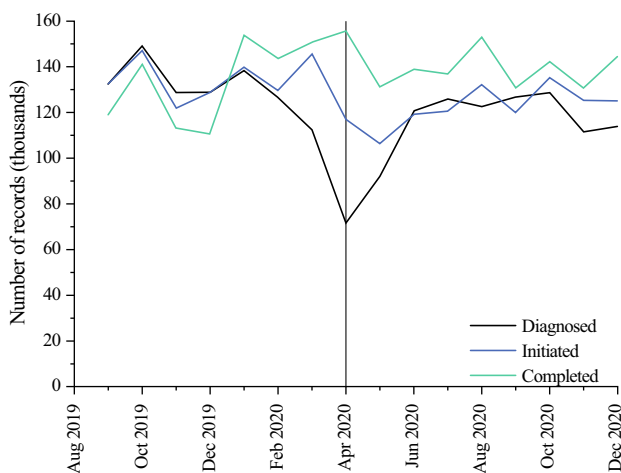

FIG. 2 Number of records diagnosed, initiated, and completed by cancer registrars during the COVID-19 pandemic, NCDB 2019–2020

natural logarithm of number of cancer diagnoses reported to the NCDB each month between January 2018 and December 2020 accounting for seasonality, regional variation, and random error.[6]

Next, we quantified the change in number of observed cancer cases diagnosed in 2020 compared with the expected number of cancer cases on the basis of historic patterns. We estimated the expected number of cancer cases diagnosed between January 2020 and December 2020 using monthly cancer counts of cancer diagnoses from January 2018 to December 2019 and the vector autoregressive method to account for seasonality and random error.[7]

We then estimated the difference between expected and observed number of cases diagnosed in 2020 using generalized estimating equation under assumptions of the Poisson distribution for count data. This analysis was conducted for all cancers combined (over 50 cancer sites, including one category for "other" cancer sites) and also separately for each of the 21 selected cancer sites: prostate, lung, breast, colorectum, bladder, non-Hodgkin lymphoma,

TABLE 3 Comparison between number of cancer cases diagnosed and number of cancer records processed by cancer registrars in NCDB-participating facilities during the initial months of the COVID-19 pandemic

Month	Change in number of records diagnosed		Change in number of records initiated		Difference		Change in number of records completed		Difference	
	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value
March	−0.37	0.01	−0.08	0.32	0.29	0.08	0.21	0.01	0.58	0.0004
April	−0.40	0.01	−0.18	< 0.0001	0.22	0.14	0.16	0.06	0.56	0.001
May	−0.14	0.27	−0.17	0.0008	−0.04	0.78	0.07	0.30	0.21	0.14
June	0.05	0.46	−0.08	0.0478	−0.13	0.10	0.09	0.16	0.04	0.64

p values < 0.05 are given in bold

Interrupted time series analysis was used to identify immediate changes in the number of records processed each month after the COVID-19 pandemic was declared (March–June 2020) and to test for differences in natural logarithm of number of records by types (i.e., diagnosed versus initiated and completed) using linear generalized estimating equation models.

melanoma, kidney, uterus, pancreas, oral cavity and pharynx, thyroid, stomach, brain, ovary, liver, lymphocytic leukemia, myeloma, myeloid and monocytic leukemia, esophagus, and cervix.

To evaluate the impact of the pandemic on the ability of registrars to collect cancer data in 2020, we first calculated the number of records with date of diagnosis (month and year) between September 2019 and December 2020. Then, we calculated the number of records with date initiated [i.e., date (month and year) the electronic abstract was created in the database] between September 2019 and December 2020. Finally, we calculated the number of records with date completed [i.e., date (month and year) when specified data elements are completed and pass relevant data quality checks] between September 2019 and December 2020 (Supplementary Fig. 1). This analysis included all records reported to the NCDB, including those for the same cancer diagnosis processed at different times and/or different facilities (duplicates) in an effort to evaluate registrars' workload.

We hypothesized that if the pandemic impacted the ability of registrars to process cancer data, there would be a decline in the number of all three types of records (diagnosed, initiated, and completed) at the start of the pandemic. To test this hypothesis, we used interrupted time series analysis to identify immediate changes in the number of records processed each month after the COVID-19 pandemic was declared (March–June 2020) and tested for differences between natural logarithm of number of record by types (i.e., initiated and completed) using linear generalized estimating equation models and natural logarithm of number of records diagnosed as the reference group.[8] All analyses were performed using SAS 9.4. Statistical significance was set at two-sided $\alpha = 0.05$.

RESULTS

Deviations from Trends in Monthly Number of Cancer Cases

In April 2020, there was a statistically significant structural break (decrease) in the number of cases diagnosed overall and for each cancer site compared with what was expected given previous years' trend and seasonality (Table 1). This was the only statistically significant structural break from previous trends detected in 2020. In the months that followed, there was no evidence of recovery of absent cases as the subsequent number of cases diagnosed each month through December 2020 did not exceed projections (Fig. 1).

Deficit in Number of Cases Diagnosed in 2020

The absolute decrease in number of cancer cases reported to NCDB in 2020 was 174,293, leading to a 12.4% deficit compared with the expected number of cases (Table 2). Breast, lung, and prostate cancers had the largest absolute decrease in the number of cases (34,411; 24,246; and 28,349 fewer cases than expected in 2020, respectively). The greatest deficit in cancer diagnoses reported to the NCDB was observed among thyroid, melanoma, and prostate cancers (19.9%, 20.6%, and 19.5% lower compared with expected diagnoses in 2020, respectively).

Assessment of Changes in Cancer Registrar Reporting

The deficit in cancer diagnoses reported to the NCDB in 2020 was not due to registrars' inability to process records during the pandemic (Fig. 2). There was a statistically significant decline in the number of records with diagnosis date between March and April 2020. Similarly, there was a

statistically significant decline in the number of records initiated by registrars (new cancer cases) in April 2020, with no difference between number of reported diagnoses and number of records initiated during these months (Table 3). In contrast, the number of records marked completed by registrars increased significantly between March and April 2020, demonstrating continuous ability to process data (Table 3). There were no statistically significant changes in number of cases processed with diagnosis, initiation, or completion date after May 2020.

DISCUSSION

There was a significant decrease in the number of cancer cases reported to the NCDB in 2020, and this deficit was not due to cancer registrars' inability to process cancer data during the first year of the COVID-19 pandemic. Therefore, NCDB data can be reliably used to evaluate the impact of the pandemic on cancer care and outcomes.

Patterns in decline were similar to those reported elsewhere, with greatest deficit observed in April 2020.[1,3,9] Number of cancer diagnoses did not exceed expected number of cancer cases through the end of 2020. This is consistent with the finding that the deficit in new diagnoses in April 2020 was not made up by increased diagnoses in the months that followed (in which case more than expected numbers would have been seen in May–December 2020).

The deficit in number of cancer cases reported to the NCDB in 2020 was significant for all cancer sites combined and for each of the 21 selected cancer sites. The most commonly diagnosed cancer sites in the US had the greatest absolute decrease in number of cases reported to the NCDB in 2020.[10] Cancer sites often diagnosed and/or treated at outpatient settings, such as prostate cancer and melanoma, which have been previously reported to have the lowest case coverage in NCDB,[5] and for which clinical guidelines have recently changed in an attempt to mitigate overdiagnosis (prostate and thyroid),[10,11,12] had the greatest percentage deficit in cases.

The significant decrease in number of cancer diagnoses reported to the NCDB in 2020 was not due to registrars' inability to process records during the pandemic. As expected, both the number of cancer records with diagnosis dates and the number of new cancer records created by registrars (initiation dates) in March and April of 2020 decreased. In contrast, the number of records with completion dates in March and April of 2020 increased, reflecting the ability of registrars to process cancer data during the first two months of the pandemic.

This study has limitations. Although the NCDB captures over 70% of individuals newly diagnosed with cancer in the US and cancer cases captured by NCDB closely resemble those of population-based cancer registries,[5] NCDB is not population-based. Additionally, it is possible that Commission on Cancer (CoC)-accredited facilities reporting to NCDB have increased access to resources and were better able to abstract and report cancer cases during the pandemic than non-accredited facilities. Future studies should evaluate the impact of the pandemic on reliability of data collected by other national cancer registries.

Taken together, our results demonstrate that the decline in the number of cancer records reported to the NCDB in 2020 accurately reflect the impacts of the pandemic on cancer care and were not biased by changes in data collection, as cancer registrars in CoC-accredited institutions were able to process cancer data during the pandemic. Future studies should evaluate how the pandemic, its associated community mitigation measures, and regional variation in care impacted cancer diagnosis, care, and outcomes.

REFERENCES

1. London JW, et al. Effects of the COVID-19 pandemic on cancer-related patient encounters. *JCO Clin Cancer Inform*. 2020;4:657–65.
2. Patt D, et al. Impact of COVID-19 on cancer care: how the pandemic is delaying cancer diagnosis and treatment for American seniors. *JCO Clin Cancer Inform*. 2020;4:1059–71.
3. Yabroff KR, et al. Association of the COVID-19 pandemic with patterns of statewide cancer services. *J Natl Cancer Inst*. 2022;114(6):907–9.
4. National Cancer Registrar Association, N. NCRA'S Council on Certification History. (2022) Available from: https://www.ncra-usa.org/CTR/Certification-About
5. Mallin K, et al. Incident cases captured in the National Cancer Database compared with those in U.S. population based central cancer registries in 2012–2014. *Ann Surg Oncol*. 2019;26(6):1604–12.
6. De Jong P and Penzer J. Diagnosing shocks in time series. *J Am Stat Assoc*. 1998;93(442):796–806.
7. Miller KD, et al. Updated methodology for projecting U.S.- and state-level cancer counts for the current calendar year: part II: evaluation of incidence and mortality projection methods. *Cancer Epidemiol Biomarkers Prev*. 2021;30(11):1993–2000.

8. Kontopantelis E, et al. Regression based quasi-experimental approach when randomisation is not an option: interrupted time series analysis. *BMJ*. 2015;350:h2750.

9. Kaufman HW, et al. Changes in the number of US patients with newly identified cancer before and during the coronavirus disease 2019 (COVID-19) pandemic. *JAMA Netw Open*. 2020;3(8):e2017267.

10. Siegel RL, et al. Cancer statistics, 2022. *CA Cancer J Clin*. 2022;72(1):7–33.

11. Bibbins-Domingo K, et al. Screening for thyroid cancer: US preventive services task force recommendation statement. *JAMA*. 2017;317(18):1882–7.

12. Grossman DC, et al. Screening for prostate cancer: US preventive services task force recommendation statement. *JAMA*. 2018;319(18):1901–13.