

UC Davis

UC Davis Previously Published Works

Title

Positive predictive value and sensitivity of ICD-9-CM codes for identifying pediatric leukemia

Permalink

<https://escholarship.org/uc/item/1zh3b829>

Journal

Pediatric Blood & Cancer, 69(2)

ISSN

1545-5009

Authors

Weinmann, Sheila
Francisco, Melanie C
Kwan, Marilyn L
et al.

Publication Date

2022-02-01

DOI

10.1002/pbc.29383

Peer reviewed



Published in final edited form as:

Pediatr Blood Cancer. 2022 February ; 69(2): e29383. doi:10.1002/pbc.29383.

Positive predictive value and sensitivity of ICD-9-CM codes for identifying pediatric leukemia

Sheila Weinmann, PhD MPH^{1,2}, Melanie C. Francisco, PhD¹, Marilyn L. Kwan, PhD³, Erin J. A. Bowles, MPH⁴, Alanna Kulchak Rahm, PhD MS, CGC⁵, Robert T. Greenlee, PhD MPH⁶, Natasha K. Stout, PhD⁷, Jason D. Pole, PhD MSc^{8,9,10}, Lawrence H. Kushi, ScD³, Rebecca Smith-Bindman, MD^{11,12}, Diana L. Miglioretti, PhD¹³

¹Center for Health Research, Kaiser Permanente Northwest Portland, OR

²Center for Integrated Health Research, Kaiser Permanente Hawaii Honolulu, Hawaii

³Division of Research, Kaiser Permanente Northern California Oakland, CA

⁴Kaiser Permanente Washington Health Research Institute, Kaiser Permanente Washington Seattle WA

⁵Center for Health Research, Genomic Medicine Institute, Geisinger Health System Danville, PA 17822

⁶Marshfield Clinic Research Institute, Marshfield Clinic Health System Marshfield, WI

⁷Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute Boston, MA

⁸The Hospital for Sick Children, Toronto, Ontario, Canada

⁹ICES, Toronto, Ontario, Canada

¹⁰Centre for Health Services Research, The University of Queensland Brisbane, Australia

¹¹Department of Radiology and Biomedical Imaging, Epidemiology and Biostatistics, University of California San Francisco

¹²The Philip R. Lee Institute for Health Policy, University of California San Francisco

¹³Department of Public Health Sciences, University of California, Davis Davis, CA

Abstract

Background: To facilitate community-based epidemiologic studies of pediatric leukemia, we validated use of ICD-9-CM diagnosis codes to identify pediatric leukemia cases in electronic medical records of six U.S. integrated health plans from 1996–2015 and evaluated the additional contributions of procedure codes for diagnosis/treatment.

Corresponding author: Sheila Weinmann, PhD MPH, 3800 N. Interstate Ave. Portland, OR 97227, Phone: (503)335-2400, FAX: (503)335-6311, Sheila.Weinmann@kpchr.org.

CONFLICT OF INTEREST STATEMENT

The authors report no conflict of interest.

Procedures: Subjects (N=408) were children and adolescents born in the health systems and enrolled for at least 120 days after the date of the first leukemia ICD-9-CM code or tumor registry diagnosis. The gold standard was the health system tumor registry and/or medical record review. We calculated positive predictive value (PPV) and sensitivity by number of ICD-9-CM codes received in the 120-day period following and including the first code. We evaluated whether adding chemotherapy and/or bone marrow biopsy/aspiration procedure codes improved PPV and/or sensitivity.

Results: Requiring receipt of one or more codes resulted in 99% sensitivity (95%CI: 98–100%) but poor PPV (70%; 95%CI: 66–75%). Receipt of two or more codes improved PPV to 90% (95% CI: 86–93%) with 96% sensitivity (95% CI: 93–98%). Requiring at least four codes maximized PPV (95%; 95% CI: 92–98%) without sacrificing sensitivity (93%; 95%CI: 89–95%). Across health plans, PPV for four codes ranged from 84–100% and sensitivity ranged from 83–95%. Including at least one code for a bone marrow procedure or chemotherapy treatment had minimal impact on PPV or sensitivity.

Conclusions: The use of diagnosis codes from the electronic health record has high PPV and sensitivity for identifying leukemia in children and adolescents if more than one code is required.

Keywords

adolescents; children; diagnosis codes; leukemia; positive predictive value; sensitivity

INTRODUCTION

Leukemia is the most common type of childhood and adolescent cancer, accounting for almost one-third of U.S. pediatric cancer cases ages 0–19 years, with an age-adjusted incidence rate of 4.8/100,000 person-years in 2015.¹ It can be difficult to conduct community-based epidemiologic studies of pediatric leukemia in populations without access to a comprehensive tumor registry to identify cases. Although there are statewide tumor registries in most states in the U.S., it can be expensive and complicated to obtain state tumor registry data and match it correctly to health plan members. Also, state registries may not be complete or may not cover the entire geographic area required by the study. Often studies must devote a lot of valuable research resources for records review for case finding,

If electronic databases are available, International Classification of Diseases (ICD) diagnosis codes may be used to identify leukemia cases. However, not all ICD codes for leukemia will flag true cases. Also, non-specific ICD codes for blood-related disorders are sometimes used for coding leukemia (e.g. neoplasm of uncertain behavior of other lymphatic and hematopoietic tissues) but may identify many non-leukemia diagnoses. The few publications examining sensitivity and PPV of International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) codes to identify pediatric leukemia cases in electronic databases reported good results but were restricted to tertiary medical center settings^{2–4} where patient referral and visit patterns may be different from community health care centers and where medical records are often restricted to the hospital setting. Epidemiologic studies of some risk factors for childhood leukemia are best done at the primary care level because that is where clinical data from the child's regular medical care can be found in

the EMR. These types of information are not likely to be included in records of the child's hospitalization at a tertiary medical center. The fact that these studies are population-based and thus representative of a defined population is also an advantage.

In the Radiation-Induced Cancer (RIC) study,^{5,6} we validated ICD-9-CM diagnosis codes associated with leukemia in six U.S. integrated health systems using their electronic medical record (EMR) databases. Study aims were (1) to calculate the positive predictive value (PPV) and sensitivity of ICD-9-CM codes for childhood and adolescent leukemia; (2) to identify the accuracy associated with using different numbers of codes for case identification; and (3) to determine whether adding procedure codes for chemotherapy and/or bone marrow biopsy/aspiration improves identification of true leukemia cases.

METHODS

Population

Six integrated health systems participating in the Radiation-Induced Cancer Study contributed data to this analysis, including Kaiser Permanente (KP) Hawaii, KP Northern California, KP Northwest (Oregon/Southwest Washington), KP Washington, Geisinger (Pennsylvania), and Marshfield Clinic Health System (Wisconsin). The Institutional Review Boards of all collaborating institutions and the University of California, Davis Statistical Coordinating Center approved the study and granted a waiver of individual consent. The time period of the study was 1/1/1996 through 9/30/2015, ending on the last date when ICD-9-CM codes were used in these health systems.

Our study subjects were born in one of the six health systems between 1996–2015 (an eligibility requirement of the parent study from which our dataset was obtained); children not born in the health system were excluded. All potential leukemia cases were also required to have either a tumor registry leukemia diagnosis (ICD-O-3 histology/morphology codes 9820–9992, Supplemental Table S1) and/or one or more ICD-9-CM codes indicating leukemia or other less-specific codes for blood-related neoplasms or disorders (Supplemental Table S2) prior to 6/1/2015. This date was chosen to allow at least 120 days of follow-up time before the transition to ICD-10-CM administrative codes. Subjects must also have been continuously enrolled for at least 120 days after the tumor registry diagnosis date or the date of the first leukemia ICD-9-CM code, whichever was earlier (reference date). This observation period was chosen in order to allow sufficient time to capture codes for chemotherapy treatment and multiple ICD-9-CM codes. An exception to the 120-day continuous enrollment requirement was inclusion of those who died within the 120-day period.

Data collection

Data were obtained from electronic medical record (EMR) research databases available at each health system that were standardized according to the Health Care Systems Research Network (HCSRN) Virtual Data Warehouse (VDW)⁷ common data model. These included data from the health systems and local tumor registries. We identified cancer diagnoses from January 1, 1996 through December 31, 2017 primarily from the VDW tumor table at

each site, which originates from the individual site's linkages with the NCI Surveillance, Epidemiology, and End Results (SEER)-affiliated or local accredited health system cancer registries, with additional linkages to state cancer registries. All International Classification of Diseases for Oncology, 3rd Edition (ICD-O-3) histology and topography codes were extracted.

Diagnosis and procedure codes were obtained from outpatient, inpatient, and emergency department patient encounters as well as from billing codes. We counted only one code per ICD-9-CM code type (any encounter type) per day. We obtained dates of health system eligibility from membership files and ICD-9-CM diagnoses and dates of diagnosis from diagnosis files. We classified the ICD-9-CM codes into codes specific for leukemia (leukemia codes), and codes for non-specific blood-related disorders (non-specific codes). We enumerated the number of leukemia and non-specific blood-related ICD-9-CM codes per subject within the 120-day observation period following their first diagnosis code (including the first code in the code count), overall and by code category; this first code was the index date. From the tumor registry datasets, we collected date of diagnosis, age at diagnosis, and tumor characteristics. From inpatient, emergency department, ambulatory visit, radiology, laboratory, and other clinical encounter data, we identified chemotherapy administration and bone marrow biopsy and aspiration procedures using Current Procedural Terminology (CPT[®]) and ICD-9-CM procedure codes (Supplemental Table S3). Chemotherapy administration and bone marrow biopsy and aspiration procedures were identified within 7 days prior to until 120 days after the date of the first diagnosis code.

Validation methods

The gold standard for this analysis was a leukemia case included in a tumor registry or identified using leukemia-specific ICD-9-CM codes and confirmed through medical record review. Select tumor registry diagnoses for lymphoblastic leukemia/lymphoma and myeloid leukemia/neoplasms were reviewed by a pediatric oncologist to determine whether a diagnosis of leukemia or another blood-related disorder, such as lymphoma or myelodysplastic syndrome, was most appropriate; the histology codes that were adjudicated are in Supplemental Table S1. If the tumor registry contained a non-leukemia cancer diagnosis and no leukemia diagnosis within \pm 90 days of the index date, we considered the subject to have a cancer other than leukemia but no leukemia. Trained medical record abstractors conducted individual medical record review for subjects with no cancer diagnoses noted in the tumor registry within 90 days of the index date.

Analysis

Data analysis for this paper was performed using SAS/STAT[®] software.⁸ We calculated the total number of possible leukemia related ICD-9-CM codes (including leukemia codes and non-specific codes) within the observation period per subject, overall, and stratified by code categories (leukemia vs non-specific blood-related disorders), and by health system. We calculated PPV and sensitivity of each category of codes. Exact Clopper-Pearson 95% confidence limits were computed for all binomial proportions.

We calculated the PPV and sensitivity varying the number of codes found in the 120-day observation period (e.g. at least 1 code, at least 2 codes, etc.), and by code categories, health system, and child age at the time of the first ICD-9-CM diagnosis date (<1 year, 1–5 years, >5 years old). We also calculated PPV and sensitivity requiring at least one treatment code within the observation period (chemotherapy or bone marrow biopsy/aspiration procedure) in addition to the minimum number of ICD-9-CM codes. In a separate analysis, we calculated PPV and sensitivity by the number of ICD-9-CM codes received within different observation periods (1 month, 2 months, 3 months, and 4 months).

RESULTS

During the study period, 857,760 age-eligible health system members at the six health systems were identified, including 1,208 children with at least one leukemia or non-specific blood-related disorder code or tumor registry leukemia diagnosis (Figure 1). Overall, 1,085 (90%) of these children were enrolled in the health system for at least 120 days, or died prior to 120 days, and met study inclusion criteria.

There were 408 children who had at least a single specific leukemia code or tumor registry diagnosis of leukemia (Figure 1). These children received an average of 18 ICD-9-CM leukemia diagnosis codes in the 120-day period following and including their first diagnosis code (standard deviation 17.5, range 0–122 codes). Most of these children (N=271, 66%) had a tumor registry diagnosis within 90 days of the index date (date of first diagnosis code); 96% of these were leukemia diagnoses, 4% were cancers of other sites. When the medical records of the remaining 137 children without a tumor registry diagnosis were reviewed, 28 (20%) of those were confirmed as leukemia. Overall, 288 (71%) children with at least one specific leukemia code or tumor registry diagnosis had confirmed leukemia diagnoses.

The median difference between first diagnosis code and confirmed diagnosis date was 0 days, with an interquartile range of 0–1 days. The tumor registries of our integrated health plans follow the North American Association of Central Cancer registries (NAACCR) rules for determining date of diagnosis (NAACCR item #390), which is the ‘Date of initial diagnosis by a recognized medical practitioner for the tumor being reported whether clinically or microscopically confirmed.’⁹

Among the 677 patients with only non-specific blood-related disorder codes and no leukemia diagnosis in the registry, we reviewed the medical records for a random sample of 94 subjects (14%) (Figure 1). No leukemia cases were identified in these children. Therefore, the 677 subjects with only non-specific blood-related disorder codes and no leukemia diagnosis in the registry were excluded from further analysis.

Demographic characteristics of 408 study subjects with at least one leukemia code or a leukemia diagnosis in the tumor registry are presented in Table 1. Subjects were White (48%), Asian (17%), Black (5%), or other/unknown race (5%), and 24% were of Hispanic ethnicity. On average, subjects were 3.7 years of age at the time of the first ICD-9-CM diagnosis code recorded (median: 2.8, range: 0–17 years). More than a quarter of subjects were younger than 18 months old at the time of the first diagnosis code. In confirmed

leukemia cases, half of the cases were 36 months or younger. There were 219 children with acute lymphoblastic leukemia (ALL), 55 with acute myeloblastic leukemia (AML), and 14 with other leukemia types.

All but one subject had at least one leukemia or leukemia-related ICD-9-CM code recorded in the EMR during the 120-day observation period; the subject with no ICD-9-CM codes was identified solely through the tumor registry (Table 2). The most common ICD-9-CM codes first recorded were 204.00: *Acute lymphoid leukemia, without mention of having achieved remission* (42%) and 208.90: *Unspecified leukemia, without mention of having achieved remission* (22%).

PPV ranged from 70% (95% confidence interval (CI): 66,77%) when requiring only one ICD-9-CM code to 99% (95% CI: 97,100%) when requiring at least eight ICD-9-CM codes (Table 3 and Figure 2). Sensitivity decreased from 99% (95% CI: 98,100%) to 89% (95% CI: 85,93%), respectively. There was little improvement in PPV after four codes. For four leukemia-specific codes, PPV varied little across age groups (range 92,97%), while sensitivity was lower (81%) in the youngest age group, compared to the older age groups (95%) (Table 3).

Requiring at least one bone marrow procedure code and/or chemotherapy treatment had little impact on the PPV and sensitivity if more than one ICD-9-CM code was used (Table 3 and Figure 2). However, when requiring only one ICD-9-CM code, the bone marrow procedure/chemotherapy treatment variable was influential in increasing the PPV from 70% (95% CI: 66,75%) to 79% (95% CI: 74,83%).

PPV for four codes ranged from 84% to 100% across sites (Supplemental Table S4). Sensitivity across sites ranged from 83% to 95%. We also examined PPV and sensitivity for shorter observation periods after the reference date (Supplemental Table S4). When requiring at least four ICD-9-CM diagnosis codes, PPV increased slightly from 95% for a 120-day observation period to 97% (95% CI: 83,98%) for a 30-day observation period, but sensitivity dropped from 93% (95% CI: 90,96%) to 77% (95% CI: 71,81%).

DISCUSSION

We found that in integrated health system EMR databases, most childhood and adolescent leukemia cases could be identified using multiple leukemia-specific ICD-9-CM codes in a defined period. Requiring at least two diagnosis codes within the 120-day observation period improved PPV compared to requiring only one code, especially for children younger than 1 year old. Presumably one code may reflect a suspected rather than confirmed diagnosis (rule-out code). Requiring at least four codes in the 120-day observation period maximized PPV without sacrificing sensitivity.

Including at least one code for a bone marrow procedure or chemotherapy treatment had minimal impact on PPV or sensitivity. For most confirmed leukemia cases without an identified treatment code within the observation period, treatment occurred more than 120 days after the first ICD-9-CM code. For other cases these may have occurred outside the health system.

We found variability in PPV and sensitivity across sites. Even when requiring four diagnosis codes, the PPV was below 90% at two of the six health care systems. This likely reflects differences in coding practices across the health systems. Since the number of potential cases in most of the health system was generally small, a few misclassifications would have a large negative effect on PPV and sensitivity. In new populations, validation of a small sample of potential cases could be valuable.

We identified two papers evaluating use of ICD-9-CM diagnosis codes to identify leukemia; both from the same U.S. children's hospital group - one validated AML among 1,686 children³ and the other validated ALL among 8,733 children.² Their algorithms required one hospital discharge diagnosis ICD-9-CM code specific to ALL, AML, or unspecified leukemia plus manual chart review verifying induction chemotherapy treatment in absence of any ICD-9-CM code suggesting another malignancy.^{2,3} PPV for the ALL algorithm was 93% with sensitivity of 88%. The AML algorithm reported PPV of 100% with a sensitivity of 96%. Another children's hospital group developed a computable phenotype for leukemia and lymphoma to be used in EMR datasets. This algorithm employs Systemized Nomenclature of Medicine (SNOMED) codes, chemotherapy data, and provider specialty information; the authors report PPV of 96% and sensitivity of 100% in two test institutions.⁴

These tertiary care populations are very different from our community population, though the PPVs and sensitivity are similar to what we found. Because these studies' hospitalized cases included ICD-9-CM codes for discharge diagnoses, cases could be identified using only one diagnosis code. Enumerating the number of codes was important in our community-based population with outpatient and inpatient ICD-9-CM codes. Because the current published literature is based on tertiary care settings where patient referral and visit patterns may be different from community health care centers or population-based integrated health plans, this study provides additional value.

Many children in children's hospitals are referred from outside primary care health settings where they are diagnosed. The children sent to children's hospitals may also be more likely to be sicker patients and require more intensive treatment. The patients in the children's hospital cohorts cited as references in our study all had standard chemotherapy regimens as an eligibility requirement, whereas that was not an eligibility criterion in our study, though we did look at whether requiring chemotherapy improved identification of leukemia cases. Identifying patients in the general pediatric population will facilitate cancer incidence studies specific to outpatient care both before and after hospital discharge, something that would be more difficult when using hospital-based cohorts. It will also allow for identification of patients who died before treatment could begin and those with non-standard therapy.

Since only one leukemia case was identified from the population receiving only non-specific ICD-9-CM codes in the absence of specific leukemia codes, future studies can restrict pediatric leukemia case-finding to the list of specific leukemia codes for nearly complete case ascertainment. Including non-specific codes will likely add little to case identification but may be worthwhile if resources are available to conduct chart review of these potential cases.

This study has several strengths. It is a large population-based study over a 20-year time period and included six different integrated health systems across the U.S, thus providing a respectable sample size for this study of a rare cancer. We had comprehensive EMR data with access to medical records for chart verification of leukemia cases and demographic variables. We also had tumor registry data to serve as the gold standard for validating ICD-9-CM codes.

One study limitation was that it was not possible to manually review medical records for all children who were not included in the tumor registry. We therefore may have missed true leukemia diagnoses for patients with no, or other, diagnosis codes, leading to an over-estimate in sensitivity. Given our requirement of continuous health plan enrollment and tumor registry surveillance, any such over-estimate is likely to be minor. Because our parent study protocol did not include manually reviewing medical records for children with no leukemia codes or tumor registry records, we were unable to calculate specificity; however because leukemia is a rare cancer, specificity is likely to be very high. Another limitation is that this work is focused on validating ICD-9-CM codes, which are no longer in use and have been superseded by ICD-10-CM codes. It is possible that the findings here may differ when examined for ICD-10-CM codes. The 22-year time period covered by the parent study (children born into the health systems from 1996–2016, with diagnosis data collected through 10/31/2016) includes only about one year in which ICD-10 codes were used for leukemia diagnosis. Only 12 subjects with specific ICD-10 leukemia codes were identified in our study population during this time. Though the positive predictive value in this small group was 100% and the sensitivity was 87%, these small numbers do not allow for a proper validation of the ICD-10-CM codes. However, ICD-9-CM codes will continue to be of relevance for research purposes, as it is likely that many investigations of childhood leukemia will continue to include cases diagnosed in the past given the rarity of the disease. There was also variability in PPV and sensitivity across sites. We do not have specific information on the reasons for differential PPV and sensitivity across our six health systems, though specific health plan coding practices may play a role. The eligibility requirement that subjects be born in the health system and have continuous enrollment through the index date, needed for the parent study on cumulative lifetime radiation dose, meant that the study population was not a cross-sectional representation of pediatric health plan members; older children and adolescents were less well-represented in the population than infants and young children. However, we were able to stratify the results by age group with a reasonable number in the group 6 to 17 years.

In conclusion, our study found that the use of diagnostic codes from the EMR has high PPV and sensitivity for identifying pediatric leukemia if receipt of more than one code is required, though we did find some variability across our six health systems. These analyses will help researchers assess the reliability of ICD-9-CM codes in identifying most cases of childhood and adolescent leukemia. For studies where tumor registries are not available and to supplement tumor registries that may not cover the entire service area, we recommend requiring a minimum of 3–4 codes depending on how important it is to balance PPV with the possibility of missing cases. This study adds to the evidence that ICD-9-CM codes can enhance cancer research potential in health care settings, in particular where a study will need to go back several years to assemble a sufficient number of cases of this rare disease.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

This study was supported by the National Cancer Institute (R01CA185687 and R50CA211115).

We gratefully acknowledge the support of Yolanda Prado, Charisma Jenkins, Yannica-Theda Martinez, Joanne M Mor, Donna Gleason, Arthur Truong, Kay Theis, Casey Luce, Deborah Seger, Mary Lyons, Glen Buth, Diane Kohnhorst, Matthew Lakoma, Mallory Snyder, Dustin Hartzel, Catarina Maney, Deanna Jarrett, Kamala Deosaransingh, Lisa Moy, Julie Munneke, and Aleyda V Solorzano Pinto for data collection and administrative support.

DATA AVAILABILITY STATEMENT

The data that support these findings are not publicly available because they contain information that could compromise research participant privacy and confidentiality. The authors will make the data available upon reasonable request and with appropriate human subjects approval and data use agreements.

Abbreviations:

ALL	Acute lymphoblastic leukemia
AML	acute myeloblastic leukemia
CI	confidence interval
CPT	Current Procedural Terminology
ICD	International Classification of Diseases
ICD9-CM	International Classification of Diseases, 9 th Revision, Clinical Modification
KP	Kaiser Permanente
PPV	positive predictive value

REFERENCES

- Howlader N, Noone AM, Krapcho M, Miller D, Brest A, Yu M, Ruhl J, Tatalovich Z, Mariotto A, Lewis DR, Chen HS, Feuer EJ, Cronin KA. SEER Cancer Statistics Review 1975–2016. In: Institute NC, ed. Bethesda, MD. 2019.
- Fisher BT, Harris T, Torp K, Seif AE, Shah A, Huang Y-SV, Bailey LC, Kersun LS, Reilly AF, Rheingold SR, Walker D, Li Y, Aplenc R. Establishment of an 11-year cohort of 8733 pediatric patients hospitalized at United States free-standing children's hospitals with de novo acute lymphoblastic leukemia from health care administrative data. *Medical Care*. 2014;52(1):e1–e6. [PubMed: 22410405]
- Kavcic M, Fisher BT, Torp K, Li Y, Huang Y-S, Seif AE, Vujkovic M, Aplenc R. Assembly of a cohort of children treated for acute myeloid leukemia at free-standing children's hospitals in the United States using an administrative database. *Pediatr Blood Cancer*. 2013;60(3):508–511. [PubMed: 23192853]

4. Phillips CA, Razzaghi H, Aglio T, McNeil MJ, Salvesen-Quinn M, Sopfe J, Wilkes JJ, Forrest CB, Bailey LC. Development and evaluation of a computable phenotype to identify pediatric patients with leukemia and lymphoma treated with chemotherapy using electronic health record data. *Pediatr Blood Cancer*. 2019;66(9):e27876.
5. Smith-Bindman R, Kwan M, Marlow E, Theis M, Bolch W, Cheng S, Bowles E, Duncan J, Greenlee R, Kushi L, Pole J, Rahm A, Stout N, Weinmann S, Miglioretti D. Trends in Use of Medical Imaging in US Health Care Systems and in Ontario, Canada, 2000–2016. *JAMA*. 2019;322:843. [PubMed: 31479136]
6. Kwan ML, Miglioretti DL, Marlow EC, Aiello Bowles EJ, Weinmann S, Cheng SY, Deosaransingh KA, Chavan P, Moy LM, Bolch WE, Duncan JR, Greenlee RT, Kushi LH, Pole JD, Rahm AK, Stout NK, Smith-Bindman R. Trends in Medical Imaging During Pregnancy in the United States and Ontario, Canada, 1996 to 2016. *JAMA Netw Open*. 2019;2(7):e197249.
7. Ross TR, Ng D, Brown JS, Pardee R, Hornbrook MC, Hart G, Steiner JF. The HMO Research Network Virtual Data Warehouse: A Public Data Model to Support Collaboration. *EGEMS (Washington, DC)*. 2014;2(1):1049–1049. [PubMed: 25848584]
8. SAS System for Windows [computer program]. Version 9.4. Cary, NC, USA: SAS Institute Inc; 2016.
9. Thornton ML, ed Standards for Cancer Registries Volume II: Data Standards and Data Dictionary, Version 21, 22nd ed. Springfield, Ill.: North American Association of Central Cancer Registries, August 2020, revised Sept. 2020, Oct. 2020, Nov. 2020, May 2021.

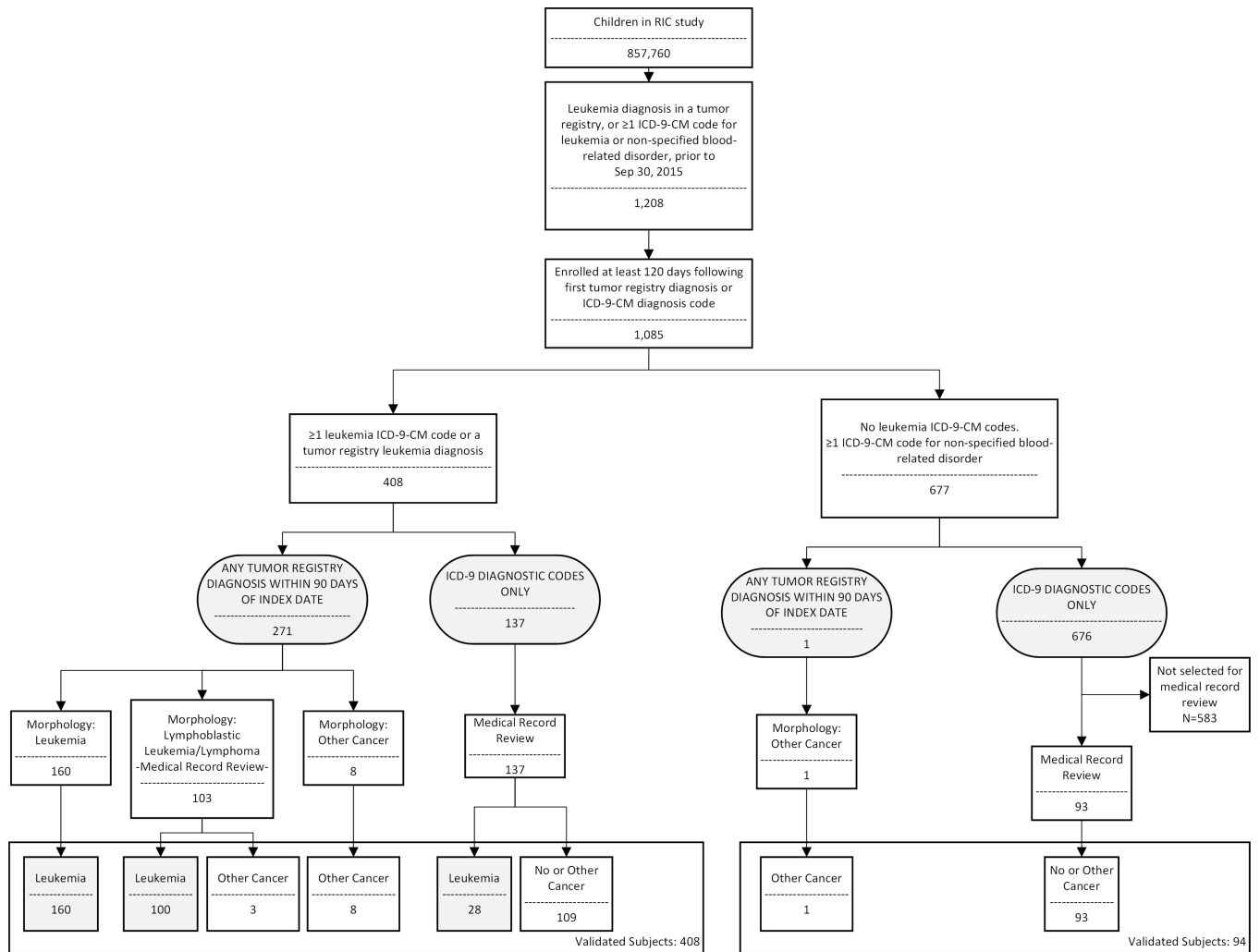
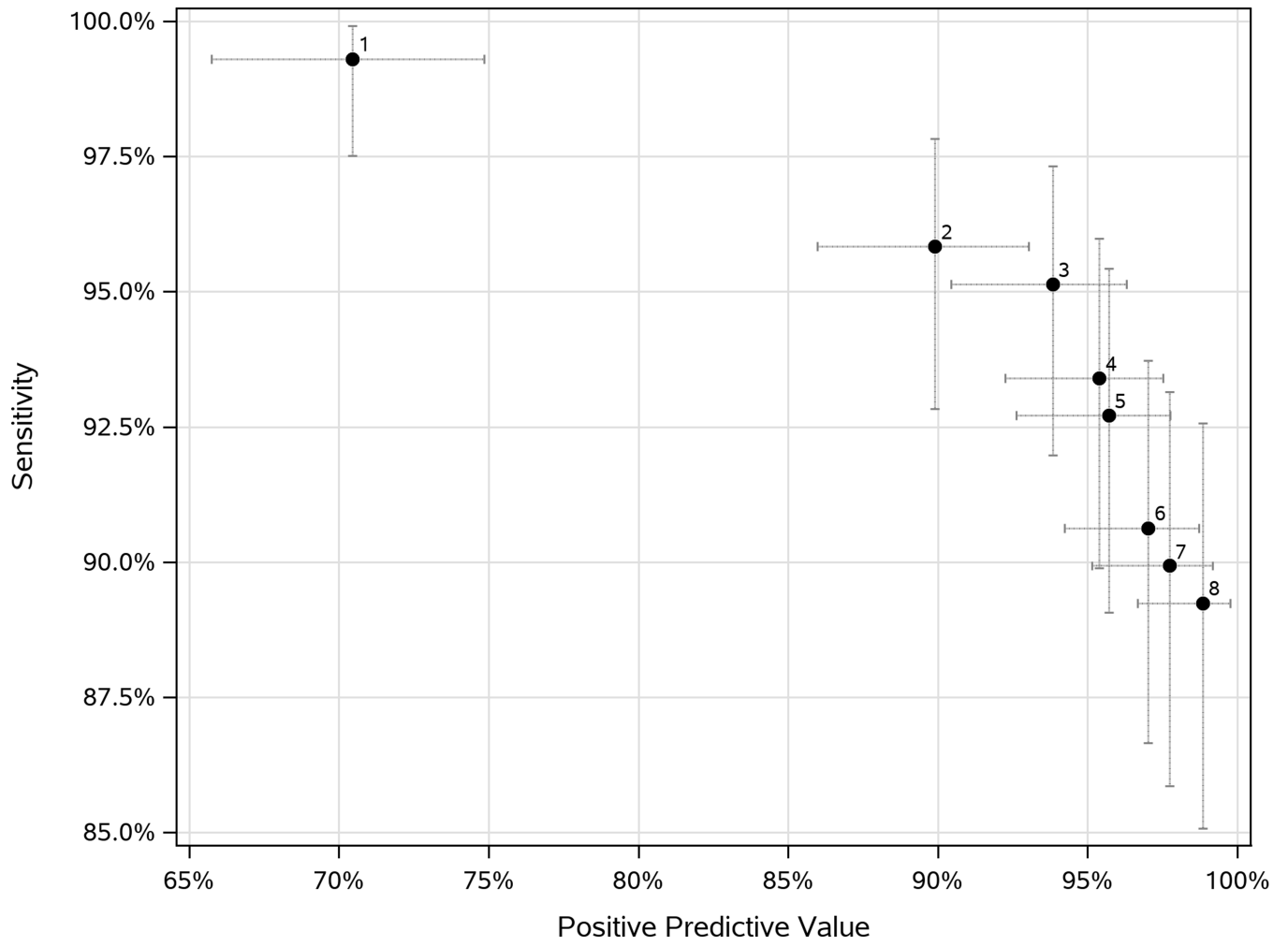


Figure 1.
Consort diagram



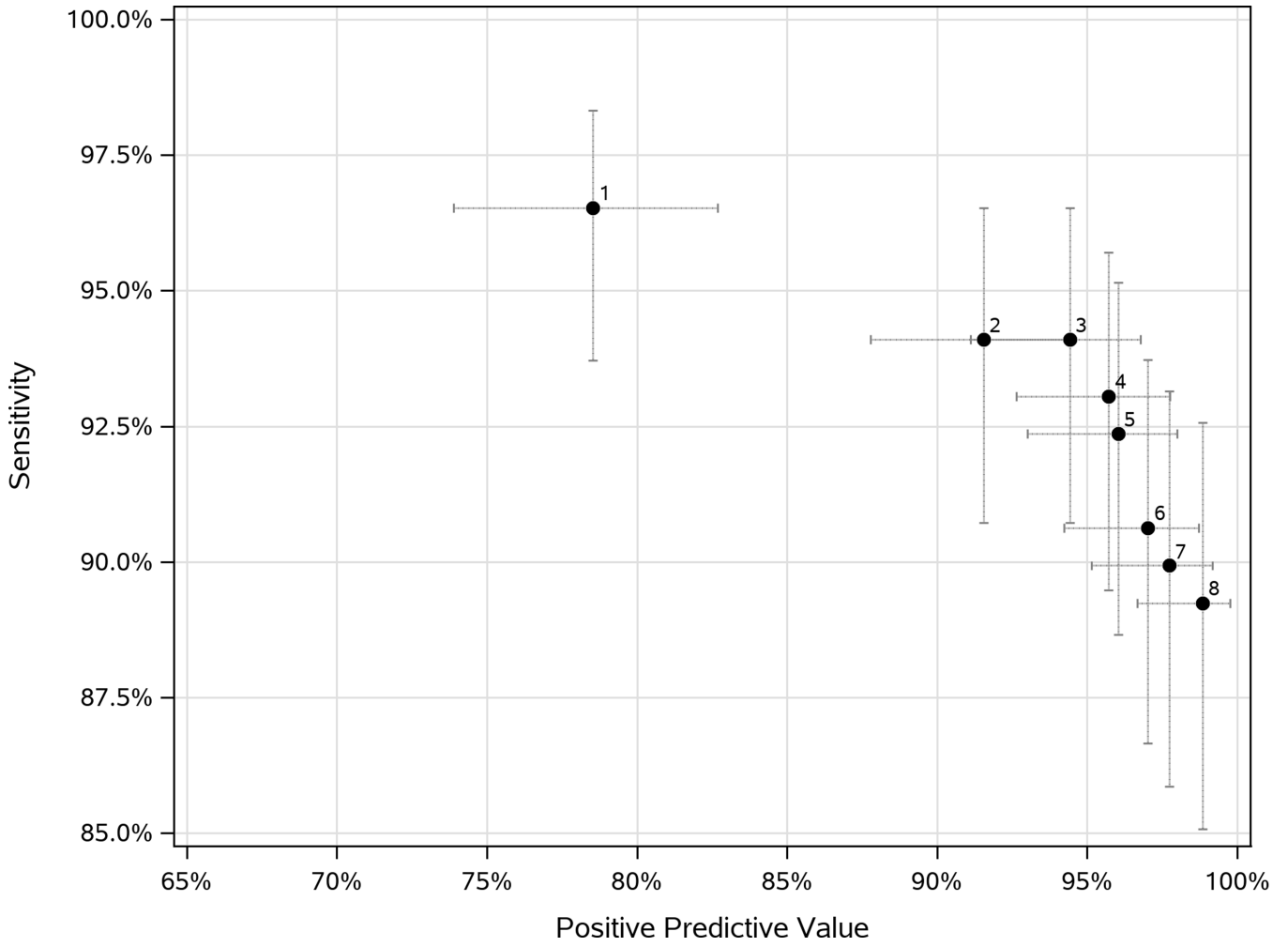


Figure 2. Sensitivity versus PPV by number of specific leukemia ICD-9 diagnosis codes (A) and number of specific leukemia ICD-9 diagnosis codes in combination with at least one chemotherapy and/or bone marrow procedure code (Px) (B) within the 120-day observation period. Data point labels indicate number of ICD-9 diagnosis codes. Horizontal and vertical bars at each data point indicate 95% confidence intervals for PPV and sensitivity, respectively.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1.

Population characteristics among children with a leukemia diagnosis code validated by medical record review or a leukemia diagnosis in the tumor registry.

	All (N=408)	No Leukemia (N=120)	Leukemia (N=288)	Yield of Leukemia Cases* (70.6)
	N (%)			row %
Number of subjects by health system				
<i>A</i>	32 (7.8)	14 (11.7)	18 (6.3)	56.3
<i>B</i>	286 (70.1)	81 (67.5)	205 (71.2)	71.7
<i>C</i>	49 (12.0)	17 (14.2)	32 (11.1)	65.3
<i>D</i>	13 (3.2)	1 (.8)	12 (4.2)	92.3
<i>E</i>	18 (4.4)	5 (4.2)	13 (4.5)	72.2
<i>F</i>	10 (2.5)	2 (1.7)	8 (2.8)	80.0
Confirmed cancer diagnosis				
<i>No leukemia</i>	120 (29.4)	120 (100.0)		
<i>Acute lymphoblastic leukemia</i>	219 (53.7)		219 (76.0)	
<i>Acute myeloid leukemia</i>	55 (13.5)		55 (19.1)	
<i>Chronic leukemia</i>	8 (2.0)		8 (2.8)	
<i>Other leukemia</i>	6 (1.5)		6 (2.1)	
ICD-9 Diagnosis Code Validation Method				
<i>Tumor Registry</i>	271 (66.4)	11 (9.2) [†]	260 (90.3)	95.9
<i>Chart Review (no tumor registry record)</i>	137 (33.6)	109 (90.8)	28 (9.7)	20.4
Chemotherapy and/or Bone Marrow Procedure				
<i>No treatment or procedure codes</i>	53 (13.0)	44 (36.7)	9 (3.1)	17.0
<i>Bone/Bone marrow procedure codes only</i>	64 (15.7)	36 (30.0)	28 (9.7)	43.8
<i>Chemotherapy treatment codes only</i>	24 (5.9)	16 (13.3)	8 (2.8)	33.3
<i>Chemotherapy and bone marrow procedure codes</i>	267 (65.4)	24 (20.0)	243 (84.4)	91.0
Diagnosis Year				
<i>1996–2000</i>	44 (10.8)	19 (15.8)	25 (8.7)	56.8
<i>2001–2005</i>	101 (24.8)	39 (32.5)	62 (21.5)	61.4
<i>2006–2010</i>	127 (31.1)	33 (27.5)	94 (32.6)	74.0
<i>2011–2015</i>	136 (33.3)	29 (24.2)	107 (37.2)	78.7
Diagnosis Age (years)				
<i>mean (SD)</i>	3.74 (3.47)	3.19 (3.28)	3.97 (3.52)	
<i>min-max</i>	0–17.1	0–14.1	0–17.1	
Diagnosis Age (years)				
<i><1 year old</i>	68 (16.7)	31 (25.8)	37 (12.8)	54.4
<i>1–5 years old</i>	250 (61.3)	61 (50.8)	189 (65.6)	75.6
<i>6–17 years old</i>	90 (22.1)	28 (23.3)	62 (21.5)	68.9
Gender				
<i>Male</i>	229 (56.1)	70 (58.3)	159 (55.2)	69.4

	All (N=408)	No Leukemia (N=120)	Leukemia (N=288)	Yield of Leukemia Cases* (70.6)
	N (%)			row %
<i>Female</i>	179 (43.9)	50 (41.7)	129 (44.8)	72.1
Race/Ethnicity				
<i>Hispanic</i>	97 (23.8)	21 (17.5)	76 (26.4)	78.4
<i>White</i>	198 (48.5)	56 (46.7)	142 (49.3)	71.7
<i>Black</i>	21 (5.1)	10 (8.3)	11 (3.8)	52.4
<i>Asian</i>	70 (17.2)	22 (18.3)	48 (16.7)	68.6
<i>Other</i>	6 (1.5)	1 (.8)	5 (1.7)	83.3
<i>Unknown or Not Reported</i>	16 (3.9)	10 (8.3)	6 (2.1)	37.5

[†] A tumor registry diagnosis of a non-leukemia cancer within 90 days of the reference date was used to validate no leukemia.

* The number of leukemia cases in the row divided by all subjects in that row.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2.

Patterns of diagnosis codes in first 120 days of reference date among children with at least one leukemia diagnostic code and/or a leukemia diagnosis in the tumor registry

	No Leukemia (N=120)	Leukemia (N=288)	All (N=408)	Yield of Leukemia Cases* (70.6%)
	N (%)			Row %
Specific leukemia codes only	113 (94.2)	276 (95.8)	389 (95.3)	71.0
Specific and nonspecific codes	7 (5.8)	10 (3.5)	17 (4.2)	58.8
Nonspecific codes only [‡]	0	1 (0.3)	1 (0.2)	100.0
No diagnostic codes [‡]	0	1 (0.3)	1 (0.2)	100.0

[‡]Leukemia diagnosis identified in tumor registry

*The number of leukemia cases in the row divided by all subjects in that row.

Table 3: Positive predictive value and sensitivity of leukemia codes by number of codes recorded within 120 days.

	Subjects	Positive Predictive Value/Sensitivity, % (95% CI)							
		1	2	3	4	5	6	7	8
<i>Specific leukemia codes</i>									
	408	PPV 70 (66, 75)	90 (86, 93)	94 (90, 96)	95 (92, 98)	96 (93, 98)	97 (94, 99)	98 (95, 99)	99 (97, 100)
		Sensitivity 99 (98, 100)	96 (93, 98)	95 (92, 97)	93 (90, 96)	93 (89, 95)	91 (87, 94)	90 (86, 93)	89 (85, 93)
<i>Specific leukemia + treatment codes</i>									
	408	PPV 79 (74, 83)	92 (88, 94)	94 (91, 97)	96 (93, 98)	96 (93, 98)	97 (94, 99)	98 (95, 99)	99 (97, 100)
		Sensitivity 97 (94, 98)	94 (91, 97)	94 (91, 97)	93 (89, 96)	92 (89, 95)	91 (87, 94)	90 (86, 93)	89 (85, 93)
Specific leukemia codes stratified by health system:									
<i>A</i>									
	68	PPV 56 (38, 74)	84 (60, 97)	84 (60, 97)	84 (60, 97)	84 (60, 97)	84 (70, 100)	93 (68, 100)	93 (68, 100)
		Sensitivity 100 (81, 100)	89 (65, 99)	89 (65, 99)	89 (65, 99)	89 (65, 99)	83 (59, 96)	78 (52, 94)	78 (52, 94)
<i>B</i>									
	68	PPV 72 (66, 77)	91 (86, 94)	95 (91, 98)	97 (94, 99)	97 (94, 99)	98 (95, 99)	98 (95, 100)	99 (96, 100)
		Sensitivity 100 (97, 100)	97 (94, 99)	97 (93, 99)	95 (91, 97)	94 (89, 97)	91 (86, 95)	91 (86, 94)	90 (85, 94)
<i>C</i>									
	68	PPV 65 (50, 78)	83 (67, 94)	86 (70, 95)	88 (73, 97)	88 (73, 97)	91 (76, 98)	94 (79, 99)	100 (88, 100)
		Sensitivity 100 (89, 100)	94 (79, 99)	94 (79, 99)	94 (79, 99)	94 (79, 99)	94 (79, 99)	94 (79, 99)	94 (79, 99)
<i>D</i>									
	68	PPV 92 (64, 100)	100 (72, 100)	100 (72, 100)	100 (69, 100)	100 (69, 100)	100 (69, 100)	100 (69, 100)	100 (69, 100)
		Sensitivity 100 (74, 100)	92 (62, 100)	92 (62, 100)	83 (52, 98)	83 (52, 98)	83 (52, 98)	83 (52, 98)	83 (52, 98)
<i>E</i>									
	68	PPV 72 (47, 90)	93 (66, 100)	100 (74, 100)	100 (74, 100)	100 (74, 100)	100 (74, 100)	100 (74, 100)	100 (74, 100)
		Sensitivity 100 (75, 100)	100 (75, 100)	92 (64, 100)	92 (64, 100)	92 (64, 100)	92 (64, 100)	92 (64, 100)	92 (64, 100)
<i>F</i>									
	68	PPV 78 (40, 97)	88 (47, 100)	100 (59, 100)	100 (59, 100)	100 (59, 100)	100 (59, 100)	100 (59, 100)	100 (59, 100)
		Sensitivity 88 (47, 100)	88 (47, 100)	88 (47, 100)	88 (47, 100)	88 (47, 100)	88 (47, 100)	88 (47, 100)	88 (47, 100)
Specific leukemia codes stratified by age:									
<i><1 year old</i>									
	68	PPV 54 (41, 66)	82 (66, 92)	89 (73, 97)	91 (76, 98)	91 (75, 98)	96 (82, 100)	100 (87, 100)	100 (87, 100)
		Sensitivity 97 (86, 100)	86 (71, 95)	84 (68, 94)	81 (65, 92)	78 (62, 90)	73 (56, 86)	70 (53, 84)	70 (53, 84)
<i>1-5 years old</i>									
	250	PPV 76 (70, 81)	92 (88, 96)	96 (92, 98)	97 (93, 99)	95 (91, 98)	94 (89, 97)	93 (89, 96)	92 (87, 95)
		Sensitivity 100 (98, 100)	97 (93, 99)	97 (93, 99)	94 (85, 98)	95 (87, 99)	97 (88, 100)	97 (88, 100)	98 (91, 100)
<i>6-17 years old</i>									
	90	PPV 69 (58, 78)	87 (77, 94)	91 (81, 97)	95 (87, 99)	95 (87, 99)	92 (82, 97)	92 (82, 97)	92 (82, 97)
		Sensitivity 98 (91, 100)	98 (91, 100)	97 (89, 100)	95 (87, 99)	95 (87, 99)	95 (87, 99)	95 (87, 99)	95 (87, 99)