medRxiv preprint doi: https://doi.org/10.1101/2023.02.12.23285595; this version posted July 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Seasonal variation exists in B-Cell Precursor Childhood Acute Lymphoblastic Leukemia diagnosis, but not in Acute Myeloid Leukemia, Brain Tumors or Solid Tumors – a Swedish population-based study

Gleb Bychkov[1Y], Benedicte Bang[1Y], Niklas Engsner[1,2], Mats Marshall Heyman[3], Anna Skarin Nordenvall[1,4], Giorgio Tettamanti[1,5], Nikolas Herold[6,7], Fulya Taylan[1], Emeli Pontén[1], Jan Albert [8,9], Rebecka Jörnsten[10], Claes Strannegård[1,11‡], Ann Nordgren [1,12,13,14‡*]

[1]Department of Molecular Medicine and Surgery, Center for Molecular Medicine, Karolinska Institutet, Stockholm, Sweden
[2]Computer Science and Engineering, Chalmers University of Technology, Gothenburg, Sweden
[3]Department of Women's and Children's Health, Karolinska Institutet, Stockholm, Sweden
[4]Department of Radiology, Karolinska University Hospital, Stockholm, Sweden
[5]Unit of Epidemiology, Institute of Environmental Medicine, Karolinska University Hospital, Stockholm, Sweden
[6]Division of Pediatric Oncology and Hematology, Department of Women's and Children's Health, Karolinska Institutet, Stockholm, Sweden
[7]Pediatric Oncology, Theme of Children's and Women's Health, Karolinska University Hospital Solna, Stockholm, Sweden
[8]Department of Microbiology, Tumor and Cell Biology, Karolinska Institutet, Stockholm, Sweden
[9]Department of Clinical Microbiology, Karolinska University Hospital, Stockholm, Sweden
[10]Mathematical Sciences, University of Gothenburg and Chalmers University of Technology, Gothenburg, Sweden
[11]Applied Information Technology, University of Gothenburg, Gothenburg, Sweden
[12]Clinical Genetics, Karolinska University Hospital, Stockholm, Sweden
[13]Department of Clinical Genetics and Genomics, Sahlgrenska University Hospital, Gothenburg, Sweden
[14]Institute of Biomedicine, Department of Laboratory Medicine, University of Gothenburg, Gothenburg, Sweden

[Y] G.B and B.B are joint first authors and contributed equally to this study.
[‡] C.S and A.N are joint senior authors and contributed equally to this study.
[*] Corresponding author: ann.nordgren@ki.se and Gleb Bychkov gleb.bychkov@ki.se

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2023.02.12.23285595; this version posted July 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Abstract

Background B-cell precursor acute lymphoblastic leukemia (BCP-ALL) is the most common malignancy in children and adolescents. A combination of genetic predisposition, exposures to diverse microbiota, infections, and an immature immune system have been associated with BCP-ALL development. Genetic aberrations causing the progression of preleukemic cells to overt BCP-ALL have been identified, but drivers behind these aberrations remain largely unknown.

Methods We analyzed seasonal variation in 1,380 BCP-ALLs, 385 acute myeloid leukemias (AML), 3,052 solid tumors and 1,945 brain tumors retrieved from the population-based Swedish Childhood Cancer Registry (SCCR), aged 0-18 years at diagnosis and diagnosed between 1995-2017. Cases were first aggregated into three types of quarters (3-month periods) based on the time of BCP-ALL diagnosis. Then, data was analyzed using a Bayesian Generalized Auto Regressive Integrated Moving Average with external variables (GARIMAX) model, adapted for count data via a negative binomial distribution.

Results An informative seasonal variation in BCP-ALL with peak quarters in Jul-Sep and Jun-Aug was identified. Manual inspection revealed that the largest number of BCP-ALL cases (138 (10%)) was observed in August. No seasonal variation was detected in the comparison groups of childhood AML, brain tumors, or solid tumors.

Conclusions Diagnosis of childhood BCP-ALL in Sweden displays seasonal variation with a peak during the summer months, in contrast to other tumor types. We present putative explanation models for this incidence peak that build on the hypothesis of infectious exposure/-s triggering the final progression to BCP-ALL diagnosis in at-risk individuals. Further studies using GARIMAX in larger populations with genetically confirmed BCP-ALL subtypes are warranted.

Keywords: acute lymphoblastic leukemia, ALL, seasonal variation, GARIMAX

medRxiv preprint doi: https://doi.org/10.1101/2023.02.12.23285595; this version posted July 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Introduction

Acute lymphoblastic leukemia (ALL) is the most frequent cancer (25%) in children and adolescents below the age of 18 years. ALL is most commonly (85%) of B-cell precursor origin, termed BCP-ALL, a genetically heterogenous disease with subtypes defined by recurrent genetic aberrations, displaying highly variable prognosis, and used to guide modern treatment strategies (1). The most common subtypes, high hyperdiploidy (HeH)(51-67 chromosomes) and *ETV6*::*RUNX1* fusion, compose around 30% and 25% of cases, respectively, and have the most favorable prognosis. These recurrent genetic aberrations are considered to initiate BCP-ALL, establishing a preleukemic clone, often already in utero (2).

Several studies have suggested that different prenatal environmental and lifestyle factors such as parental age, ethnicity, maternal diet, folate intake, exposure to irradiation, pesticides, and perinatal infections, may have a role in the formation of somatic genetic aberrations, that occur in utero as the first hit that eventually leads to childhood leukemia (3). Previous studies have shown that the *ETV6*::*RUNX1* fusion is detected in cord blood in up to 1-5% of healthy newborns and that the preleukemic clones only give rise to clinically overt leukemia in 0.2%. Progression from pre-leukemic state to overt leukemia, which takes place in only a fraction of at-risk cases, requires secondary somatic genetic aberrations (4-6).

Today, there is limited knowledge about the causes and mechanisms behind the progression from a preleukemic clone to overt childhood BCP-ALL. There is, however, increasing evidence supporting the contribution of genetic predisposition, the microbiome, infections, and the training of the child's immune system to this equation. Several epidemiological and molecular studies suggest that infections, both passed from mother to fetus in utero and occurring in the child after birth, are potential triggers of BCP-ALL. Meanwhile, early

medRxiv preprint doi: https://doi.org/10.1101/2023.02.12.23285595; this version posted July 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

exposure to a diverse microbiota is protective, as indicated by lower frequencies of BCP-ALL in children who have been naturally born, breastfed, exposed to social contacts, and exposed to animals during the first year. Previous studies have also indicated heterogeneity in environmental exposures associated to different molecular subtypes of BCP-ALL (6-11). In addition, recent experimental studies have demonstrated that predisposed transgenic mouse models with *ETV6::RUNX1* fusion or *Pax5*$^{+/-}$ heterozygosity only developed B-ALL when they were exposed to common infections, although with incomplete penetrance, highlighting the role of infections in leukemia development (12, 13).

The seasonal patterns seen in the spread of many common infections, have motivated numerous studies of seasonal variation in ALL diagnosis. Seasonal variation has been detected in some studies but not supported by all, and reported peaks have been scattered throughout the year (Supplementary **Table S5**).

In the current study, we used the population-based Swedish childhood cancer quality register to identify all children diagnosed with BCP-ALL, AML, brain tumors and solid tumors in Sweden between 1995-2018 and applied GARIMAX to study seasonal variation of cancer diagnosis.

Material and methods

Data sources

Sweden has a renowned system of records for citizens in which demographic and healthcare data are collected continuously. All permanent residents are given personal identity numbers that enable linkage between the registers. The Swedish Childhood Cancer Registry (SCCR) and The Total Population Register (RTB) are the two sources of data used in this paper.

medRxiv preprint doi: https://doi.org/10.1101/2023.02.12.23285595; this version posted July 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

The Swedish Childhood Cancer Registry (SCCR) (https://sbcr.se/) is a National Quality Register containing diagnosis-based registries with information about children and adolescents (0-18 years of age) diagnosed with CNS-tumors, solid tumors and hematological malignancies. Registration of ALL and AML started in the 1970s, while CNS- and solid tumors in the 1980s. The registry has an overall coverage of 90%, where the lack of coverage is mostly due to variable criteria for registration of benign and other tumors that have not required oncologic treatment, i.e., coverage for malignant tumors can be considered close to100%. Data about ALL-cases includes information about date of diagnosis, clinical characteristics, treatment, outcome, immunophenotype, genetic subtype, and other clinically important genetic aberrations. The most abundant genetic subtypes of BCP-ALL, HeH and *ETV6::RUNX1*-fusion have been registered since 1992 and 2000, respectively, when robust cytogenetic methods to detect these aberrations were introduced in clinical diagnostic routines. The Total Population Registry (RTB) (14) holds information about date of birth, death, and emigration for all residents of Sweden who reside in the country for ≥12 months.

Patient cohort

We identified a cohort of 1,380 children and adolescents from the Swedish Childhood Cancer Registry (SCCR) diagnosed with BCP-ALL at the age of 0-18 years between January 1st,1995 and December 31st, 2017. 444 had HeH subtype and 272 had *ETV6*::*RUNX1* subtype. In the comparison group, , we identified 385 children with AML, 3,052 children with solid tumors, and 1,945 children with brain tumors. The distribution of BCP-ALL cases by genetic subtype, year of diagnosis, age group, and sex is presented in **Table 1**. The distribution of AML, solid tumor, and brain tumor cases by year of diagnosis, age group, and sex is presented in **Table 2**. Only individuals born and diagnosed in Sweden were included.

medRxiv preprint doi: https://doi.org/10.1101/2023.02.12.23285595; this version posted July 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Methods

ARIMAX (autoregressive integrated moving average with external variables) models assume that the mean at time t depends on previous values (AR) and that additive errors are correlated over time (MA). Integration (I) refers to the process being an integrated one, and external variables (X) are independent variables introduced into the process (15). ARIMA is widely used for detecting seasonal variations across various fields (16-21). The Negative Binomial distribution generalizes ARIMAX (GARIMAX) for low case count data, accounting for excess variability better than the Poisson distribution in limited information settings (22). A Bayesian formulation makes the model less data-hungry and suitable for small sample inference (23).

We implemented the Bayesian Generalized Autoregressive Integrated Moving Average model with exogenous variables (GARIMAX) for the identification of seasonal variation in BCP-ALL diagnosis. The key elements of this model are (i) generalization of the ARIMAX process to a count distribution via a negative binomial distribution, and (ii) a Bayesian formulation for model setup. The generalization allows searching for seasonality in sparse (low value count) data, whereas the Bayesian formulation is beneficial for the application of complex models in small sample settings. We used BIC (Bayesian information criterion) for lag selection.

The GARIMAX analysis consisted of two steps. First, harmonic functions were applied as a covariate to detect periodicity, i.e., sinus or cosine waves, in all three quarter-types. In the second step, a quarterly seasonal matrix was applied as a covariate to detect the particular quarter when the wave has its peak. The seasonal matrix was constructed by factoring the quarters: the quarters of the year are treated as distinct categorical variables in the model.

medRxiv preprint doi: https://doi.org/10.1101/2023.02.12.23285595; this version posted July 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

This process involves creating indicator (dummy) variables for each quarter, allowing the model to account for seasonal effects by incorporating the specific influence of each quarter as separate factors. A "base quarter," the one with the lowest number of cases, was identified for each quarter type. Using the seasonal matrix, each base quarter was compared to the three remaining quarters to evaluate increases in quarterly case numbers. Statistical significance, termed 'informative' in the Bayesian framework, was assessed using the posterior distribution. If the credibility interval contains only positive or only negative values, then 95% of the posterior does not include 0, indicating it is unlikely that the covariate has no effect on the response variable. To confirm the presence of seasonality in the time series, both informative periodicity in the quarter type and an informative peak quarter were required. A more detailed method section is available in Supplementary information.

Based on month of diagnosis, the analysis was first performed for the entire cohort, including BCP-ALL (1,380 cases), AML (385 cases), solid tumors (3,052 cases), and brain tumors (1,945 cases). The analysis was subsequently performed for each of the two large BCP-ALL subgroups: the HeH subtype (444 cases) and the *ETV6::RUNX1* subtype (272 cases). Three quarter types, each consisting of four quarters, were defined in a total of 12 different quarters distributed over a year. 1st quarter type: Jan-Mar, Apr-Jun, Jul-Sep, Oct-Dec. 2nd quarter type Feb-Apr, May-Jul, Aug-Oct, Nov-Jan. 3rd quarter type: Mar-May, Jun-Aug, Sep-Nov, Dec-Feb (**Figure 1**). This generated a total of 92 individual quarters from January 1st, 1995 and December 31st, 2017.

Results

In total, 1,380 BCP-ALL cases, 385 AML cases, 3,052 solid tumor cases, and 1,945 brain tumor cases that fulfilled inclusion criteria were identified in SCCR and included in the study

medRxiv preprint doi: https://doi.org/10.1101/2023.02.12.23285595; this version posted July 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

(**Figure 2**). For each of the two large BCP-ALL subgroups 444 cases were identified with the HeH subtype, and 272 cases with the *ETV6::RUNX1* subtype. The distribution of age at diagnosis in the entire BCP-ALL cohort (**Figure 3**) displayed a well-known pattern with peak incidence at ages 2-6 years (6).

Initially, the order of autoregression was identified for all groups (BCP-ALL, AML, Solid tumors, and Brain tumors) using Bayesian Information Criterion (BIC) score. BIC scores of all tested models of GARIMA are presented in supplementary **Table S1**. The best GARIMAX (p, d, q) specifications for each quarterly series are summarized in supplementary **Table S2**. By applying harmonic functions as a covariate to the GARIMAX model, informative waves were identified in the 1st and 3rd quarter types in the entire BCP-ALL cohort. The 95% credibility interval of the posterior distribution for the coefficients of the seasonal harmonic functions did not include 0, indicating informative (non-random) periodicity in these two quarter types. The 2nd quarter type did not show any informative periodicity in the BCP-ALL cohort (**Table 3,** Supplementary **Figure S1**).

Applying a quarterly seasonal matrix as a covariate to GARIMAX, two informative peak quarters were detected in the BCP-ALL cohort: Jul-Sep (1st quarter type) and Jun-Aug (3rd quarter type). The 2nd quarter type had no informative peak quarter. (**Table 4** and Supplementary **Figure S2**). Both of the above steps of GARIMAX analysis are required to identify seasonal variation in the cohort and define an informative peak quarter (season). In our analysis, the data show an informative seasonal wave in the diagnosis of BCP-ALL in the 1st and 3rd quarter type, with informative peak quarters in Jul-Sep and Jun-Aug, respectively.

medRxiv preprint doi: https://doi.org/10.1101/2023.02.12.23285595; this version posted July 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

The four months encompassed by these two peak quarters were June, July, August, and September. July and August were included in both peak quarters. When manually examining the absolute numbers of BCP-ALL cases diagnosed in these respective months, August indeed had the highest number of cases (138) compared to all other 11 months. The number of cases in July (110) was below average (115) and the 5th lowest of all months. June had the 2nd largest number of cases (126) and September the 3rd largest (125) together with April (125). The total range of case numbers per month was 91 (Dec) to 138 (Aug). (**Table 5**)

Subgroup analysis using GARIMAX was performed on two subgroups of the BCP-ALL cohort; cases with HeH and *ETV6*::*RUNX1*, respectively. We did not detect any informative periodicity (first test) in any of the three-quarter types analyzed in either subgroup. (**Table 3**, Supplementary **Figures S3**, **S5**) Informative peak quarters (second test) were identified in both HeH (Jul-Sep, 1st quarter type, Jun-Aug, 3rd quarter type) and *ETV6*::*RUNX1* (Oct-Dec, 1st quarter type and Aug-Oct, 2nd quarter type) subgroups. (**Table 4**, Supplementary **Figures S4**, **S6**) However, since informative periodicity was lacking, we conclude that no informative seasonal variation was detected in either subgroup.

Analysis using the GARIMAX model was performed on three non-ALL cohorts: cases with AML, solid tumors, and brain tumors. We did not detect any informative periodicity (first test) in any of the three-quarter types analyzed in any of these cohorts (Supplementary **Figures S7, S9, and S11**, Supplementary **Table S3**). Informative peak quarters (second test) were identified in solid tumors cohorts (Apr-Jun, 1st quarter type, and Mar-May, Sep-Nov for the 3rd quarter type), but not in AML, nor in brain tumors cohorts (Supplementary **Figures S8, S10**, and **S12** Supplementary **Tables S4**). However, since informative periodicity was

medRxiv preprint doi: https://doi.org/10.1101/2023.02.12.23285595; this version posted July 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

lacking, we conclude that no informative seasonal variation was detected in any of these cohorts.

Discussion

In this study, we used GARIMAX to analyze a Swedish population-based cohort of 1,380 BCP-ALL cases, aged between 0 and 18 years at diagnosis and an aged-matched cohort with AML, brain tumors, and solid tumors and found evidence for seasonal variation of BCP-ALL diagnoses with peaks in quarters Jun-Aug and Jul-Sep with July and August included in both peak quarters. In the comparison groups, we did not detect any seasonal variation.

Our literature review (**Table S5**) revealed that seasonal variation in ALL diagnosis has been identified inconsistently across studies. Peaks, when detected, occur at various times throughout the year. Only 11 of 42 studies on ALL seasonality included cohorts of more than 1000 cases. Differences in methods, cohort sizes, age at diagnosis, ethnicity, population susceptibility, and disease types (T- and B- or only B-cell ALL) may explain these varying results. Other factors may also contribute, including regional or climate-specific differences in seasonal patterns, daylight hours, cultural behaviors, socioeconomic conditions, microbiota exposure, antibiotic use, and infectious disease prevalence.

Since several genetic ALL subgroups, such as *ETV6::RUNX1* gene fusion, are found in neonatal blood spots, they are considered the first genetic hits that cause the expansion of preleukemic B-cell clones in utero (4-7), we hypothesized that the second hit that occurs after birth that eventually leads to leukemia progression, may be caused by common infectious agents, such as different viral infections that cause accumulated hematopoietic stress in predisposed and immunologically immature children. To further explore the meaning of our

medRxiv preprint doi: https://doi.org/10.1101/2023.02.12.23285595; this version posted July 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

finding of a seasonal peak in BCP-ALL diagnosis from an etiologic perspective, we hypothesize about four theoretic etiological models assuming different induction time-frames, i.e., time from exposure to clinically overt disease.

Peak diagnosis indicated by GARIMAX Jun-Sep with highest number of cases in August may be associated with:

1. A peak in the general load of common infections during winter months in Sweden (assuming induction time around 6-9 months)
2. One or a few common infections peaking during winter months in Sweden that are more potent as drivers of disease progression (assuming induction time around 6-9 months)
3. One or a few common infections peaking during summer months in Sweden that are more potent as drivers of disease progression (assuming induction time around 1-2 months)
4. A short but marked decrease in spread of common infections during Swedish summer holidays, postponing a portion of cases in July to be accumulated during August when at risk individuals again are exposed to common infections (assuming induction time around 2-4 weeks).

We consider all of the explanatory models above as possible, but today, the main limitation lies in the lack of available data on viral agents, making it difficult to deem one more likely than the others. Also, other environmental factors, such as lifestyle factors, chemical agents or the large seasonal differences in daylight in Sweden cannot be excluded (24).

medRxiv preprint doi: https://doi.org/10.1101/2023.02.12.23285595; this version posted July 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Sweden is well known for its prolonged summer vacation when a vast majority of children and their parents are on summer vacation for 4-8 weeks in mid-June to mid-August. Therefore, a lock-down effect with reduced numbers of BCP-ALLs during longer vacations, and higher numbers, after returning to daycare/school may well reflect differences in infectious exposures. The possibility of a doctor´s delay in diagnosing the child or a parent's delay in seeking care for their children during summer vacation, as an explanation for our findings, was considered since lower detection rates and higher mortality rates were reported in adult patients with cancer, diagnosed during the holiday season in Sweden (25). Although we cannot confidently exclude these explanations, we do consider them less likely as symptoms of BCP-ALL at diagnosis are acute and progress quickly. Also, we did not find any seasonal variation in the diagnosis of childhood AML, brain tumors, or solid tumors, which supports the presence of a true seasonal variation in BCP-ALL rather than a doctor's delay.

Previous studies show environmental exposure heterogeneity in BCP-ALL subtypes, with *ETV6::RUNX1* linked to space-time clustering (11). This led to our subgroup analyses of HeH and *ETV6::RUNX1*, but the GARIMAX model yielded inconclusive results for these subtypes. Notably, the HeH subtype showed peak quarters (Jul-Sep and Jun-Aug) similar to the entire BCP-ALL cohort, suggesting HeH may drive observed seasonal variations. However, the limited number of HeH cases (444) likely affected the harmonic function's periodicity detection. Future studies pooling Nordic datasets and examining genetic subgroups separately and together could provide clearer insights into BCP-ALL etiologies.

Limitations

medRxiv preprint doi: https://doi.org/10.1101/2023.02.12.23285595; this version posted July 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

As described in the methods-section, our cases were aggregated into quarters based on their month of diagnosis. Although our method, through both generalization and formulation in the Bayesian setup, enables analysis of (in the context) small sample sizes, a limitation of our study is the sample size, since limited case numbers in each analyzed time frame, would, in itself have prevented informative results. This was, for example, the case in analyses of seasonal variation in cytogenetic subgroups represented by the two most common subtypes of BCP-ALL, HeH and *ETV6*::*RUNX1*. Another limitation that prevented studies of viral and other infectious agents is that the dataset does not contain detailed information about common infections.

Strengths

The GARIMAX model tests yearly seasonal variations, unlike chi-square which tests significant differences in stacked observations. GARIMAX also accounts for the discrete nature of BCP-ALL data through Negative-Binomial generalization, unlike ARIMA which assumes continuous data (26). This makes GARIMAX a key strength of our study. Additionally, using the population-based SCCR dataset enabled comparisons with other childhood cancers (AML, solid tumors, brain tumors). Aggregating cases by quarter also helped account for year-to-year variations in environmental factors affecting BCP-ALL progression, like viral infection peaks.

Altogether, we report seasonal variation in BCP-ALL diagnosis, with peak quarters in June-August and July-September. Our study improves upon previous reports by applying a powerful statistical model and examining a theoretically etiologically distinct cohort of BCP-ALL cases (in terms of immunophenotype and age at diagnosis) while also including comparisons with other childhood cancer types (AML, solid tumors, and brain tumors).

medRxiv preprint doi: https://doi.org/10.1101/2023.02.12.23285595; this version posted July 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

However, expanding the cohort size would have strengthened the study and enabled more detailed subgroup analyses of molecular subtypes.

Ethics Declaration

The study was approved by the Regional Ethical Review Board in Stockholm, Sweden (ethics permit numbers 2018/1849-32) in accordance with the Declaration of Helsinki.

Acknowledgments

The authors wish to thank Gunilla Karlsson Hedestam for valuable discussions, Henrik Passmark, Case Officer at The Swedish National Board of Health and Welfare and Jonas Janegren at Statistics Sweden for their great support in collecting the dataset. This study was supported by grants from the Swedish Childhood Cancer Fund, the Swedish Cancer Society, the Cancer Society of Stockholm, the Swedish Research Council, the Stockholm Regional Council, Mary Béves Stiftelse för Barncancerforskning, the Swedish Rare Diseases Research foundation (Sällsyntafonden), Berth von Kantzow's foundation, Karolinska Institutet Research Foundation, Hållsten Research foundation, Stiftelsen Frimurare Barnhuset in Stockholm and the Stockholm County Council. One of the authors of this publication is a member of the European Reference Network on Rare Congenital Malformations and Rare Intellectual Disability ERN-ITHACA [EU Framework Partnership Agreement ID: 3HP-HP-FPA ERN-01-2016/739516]. Funders had no role in study design, data collection and analysis, the decision to publish, or the preparation of the manuscript.

Authorship

Contribution: A.N., B.B., C.S., G.B., N.E., and R.J. designed the study. G.B., N.E., R.J., A.N., and B.B. performed analysis. G.B., R.J. and N.E. performed the GARIMAX analysis.

medRxiv preprint doi: https://doi.org/10.1101/2023.02.12.23285595; this version posted July 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

N.E., G.B., F.T., G.T., A.N., M.H., and A.N.S. prepared the dataset. G.B., N.E., R.J., C.S., G.T., and A.S. contributed with knowledge in statistics and epidemiology. A.N., M.H., J.A., N.H., E.P., and B.B. contributed with medical knowledge and generation of hypothesis. G.B. and B.B. prepared figures and tables. A.N., B.B., and G.B. wrote the manuscript. All authors contributed to data interpretation, revised the manuscript and approved the final version.

Conflict-of-interest disclosure

The authors declare no competing financial interest.

References

1. Pfister SM, Reyes-Múgica M, Chan JK, Hasle H, Lazar AJ, Rossi S, Ferrari A, Jarzembowski JA, Pritchard-Jones K, Hill DA, Jacques TS. A summary of the inaugural WHO classification of pediatric tumors: transitioning from the optical into the molecular era. Cancer discovery. 2022 Feb 2;12(2):331-55.
2. Moorman AV. The clinical relevance of chromosomal and genomic abnormalities in B-cell precursor acute lymphoblastic leukaemia. Blood reviews. 2012 May 1;26(3):123-35.
3. Benítez L, Castro-Barquero S, Crispi F, Youssef L, Crovetto F, Fischer U, Kameri E, Bueno C, Camos M, Menéndez P, Heinäniemi M. Maternal lifestyle and prenatal risk factors for childhood leukemia: a review of the existing evidence. Fetal Diagnosis and Therapy. 2024 May 6.
4. Mori H, Colman SM, Xiao Z, et al. Chromosome translocations and covert leukemic clones are generated during normal fetal development. Proc Natl Acad Sci USA. 2002;99(12):8242-8247.
5. Schäfer D, Olsen M, Lähnemann D, Stanulla M, Slany R, Schmiegelow K, Borkhardt A, Fischer U. Five percent of healthy newborns have an ETV6-RUNX1 fusion as revealed by DNA-based GIPFEL screening. Blood. 2018 Feb 15;131(7):821-826. doi: 10.1182/blood-2017-09-808402. Epub 2018 Jan 8. PMID: 29311095; PMCID: PMC5909885.
6. Hauer J, Fischer U, Borkhardt A. Toward prevention of childhood ALL by early-life immune training. Blood, The Journal of the American Society of Hematology. 2021 Oct 21;138(16):1412-28.
7. Greaves M. A causal mechanism for childhood acute lymphoblastic leukaemia. Nature Reviews Cancer. 2018 Aug;18(8):471-84.
8. Zhang J, Walsh MF, Wu G, Edmonson MN, Gruber TA, Easton J, Hedges D, Ma X, Zhou X, Yergeau DA, Wilkinson MR. Germline mutations in predisposition genes in pediatric cancer. New England Journal of Medicine. 2015 Dec 10;373(24):2336-46.
9. Vicente-Dueñas C, Janssen S, Oldenburg M, Auer F, González-Herrero I, Casado-García A, Isidro-Hernández M, Raboso-Gallego J, Westhoff P, Pandyra AA, Hein D. An intact gut microbiome protects genetically predisposed mice against leukemia.

medRxiv preprint doi: https://doi.org/10.1101/2023.02.12.23285595; this version posted July 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Blood, The Journal of the American Society of Hematology. 2020 Oct 29;136(18):2003-17.

10. Kinlen L. Evidence for an infective cause of childhood leukaemia: comparison of a Scottish new town with nuclear reprocessing sites in Britain. The Lancet. 1988 Dec 10;332(8624):1323-7.
11. Kreis C, Lupatsch JE, Niggli F, Egger M, Kuehni CE, Spycher BD, Swiss Paediatric Oncology Group and the Swiss National Cohort Study Group. Space-time clustering of childhood leukemia: evidence of an association with ETV6-RUNX1 (TEL-AML1) fusion. PLoS One. 2017 Jan 27;12(1):e0170020.
12. Rodríguez-Hernández G, Hauer J, Martín-Lorenzo A, Schäfer D, Bartenhagen C, García-Ramírez I, Auer F, González-Herrero I, Ruiz-Roca L, Gombert M, Okpanyi V, Fischer U, Chen C, Dugas M, Bhatia S, Linka RM, Garcia-Suquia M, Rascón-Trincado MV, Garcia-Sanchez A, Blanco O, García-Cenador MB, García-Criado FJ, Cobaleda C, Alonso-López D, De Las Rivas J, Müschen M, Vicente-Dueñas C, Sánchez-García I, Borkhardt A. Infection Exposure Promotes *ETV6-RUNX1* Precursor B-cell Leukemia via Impaired H3K4 Demethylases. Cancer Res. 2017 Aug 15;77(16):4365-4377. doi: 10.1158/0008-5472.CAN-17-0701. Epub 2017 Jun 19. PMID: 28630052.
13. Martin-Lorenzo□A, Hauer□J, Vicente-Duenas□C, Auer□F, Gonzalez-Herrero□I, Garcia-Ramirez□I, et al□Infection exposure is a causal factor in B-precursor acute lymphoblastic leukemia as a result of Pax5 inherited susceptibility. *Cancer Discov*□2015;5:1328–43
14. Ludvigsson JF, Almqvist C, Bonamy AK, Ljung R, Michaëlsson K, Neovius M, et al. Registers of the Swedish total population and their use in medical research. Eur J Epidemiol. 2016;31:125–36.
15. Box GE, Jenkins GM, Reinsel GC, Ljung GM. Time series analysis: forecasting and control. John Wiley & Sons; 2015 May 29
16. Petrevska B. Predicting tourism demand by ARIMA models. Economic research-Ekonomska istraživanja. 2017 Dec 1;30(1):939-50.
17. Ahmar AS, Gs AD, Listyorini T, Sugianto CA, Yuniningsih Y, Rahim R, Kurniasih N. Implementation of the ARIMA (p, d, q) method to forecasting CPI Data using forecast package in R Software. InJournal of Physics: Conference Series 2018 Jun 1 (Vol. 1028, p. 012189). IOP Publishing.
18. Yang B, Li C, Li M, Pan K, Wang D. Application of ARIMA model in the prediction of the gross domestic product. In2016 6th international conference on mechatronics, computer and education informationization (MCEI 2016) 2016 Dec (pp. 1258-1262). Atlantis Press.
19. Mondal P, Shit L, Goswami S. Study of effectiveness of time series modeling (ARIMA) in forecasting stock prices. International Journal of Computer Science, Engineering and Applications. 2014 Apr 1;4(2):13.
20. Benvenuto D, Giovanetti M, Vassallo L, Angeletti S, Ciccozzi M. Application of the ARIMA model on the COVID-2019 epidemic dataset. Data in brief. 2020 Apr 1;29:105340.
21. Briët OJ, Amerasinghe PH, Vounatsou P. Generalized seasonal autoregressive integrated moving average models for count data with application to malaria time series with low case numbers. PloS one. 2013 Jun 13;8(6):e65761.
22. Benjamin MA, Rigby RA, Stasinopoulos DM. Generalized autoregressive moving average models. Journal of the American Statistical association. 2003 Mar 1;98(461):214-23.

medRxiv preprint doi: https://doi.org/10.1101/2023.02.12.23285595; this version posted July 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

23. McNeish D. On using Bayesian methods to address small sample problems. Structural Equation Modeling: A Multidisciplinary Journal. 2016 Sep 2;23(5):750-73.
24. Zhong C, Wang R, Morimoto LM, Longcore T, Franklin M, Rogne T, Metayer C, Wiemels JL, Ma X. Outdoor artificial light at night, air pollution, and risk of childhood acute lymphoblastic leukemia in the California Linkage Study of Early-Onset Cancers. Scientific reports. 2023 Jan 11;13(1):583.
25. Wikén I, Andersson TM, Radkiewicz C. Seasonal effects on cancer incidence and prognosis. Acta Oncologica. 2023 Feb 1;62(2):103-9.
26. Shim KS, Kim MH, Shim CN, Han M, Lim IS, Chae SA, Yun SW, Lee NM, Yi DY, Kim H. Seasonal trends of diagnosis of childhood malignant diseases and viral prevalence in South Korea. Cancer Epidemiol. 2017 Dec;51:118-124. doi: 10.1016/j.canep.2017.11.003. Epub 2017 Nov 8. PMID: 29127946; PMCID: PMC7102810.
27. Liboschik T, Fokianos K, Fried R. tscount: An R package for analysis of count time series following generalized linear models. Journal of Statistical Software. 2017 Nov 30;82:1-51.
28. Bamouni S, Hémon D, Faure L, Clavel J, Goujon S. Seasonal variations in childhood leukaemia incidence in France, 1990–2014. Cancer Causes & Control. 2021 Jul;32:693-704.
29. Nurullah R, Kuhle S, Maguire B, Kulkarni K. Seasonality in pediatric cancer. The Indian Journal of Pediatrics. 2018 Sep;85:785-7.
30. Goujon-Bellec S, Mollié A, Rudant J, Guyot-Goubin A, Clavel J. Time trends and seasonal variations in the diagnosis of childhood acute lymphoblastic leukaemia in France. Cancer epidemiology. 2013 Jun 1;37(3):255-61.
31. Ross GJ, Preece DA. The negative binomial distribution. Journal of the Royal Statistical Society: Series D (The Statistician). 1985 Sep;34(3):323-35.
32. Hilbe JM. Negative binomial regression. Cambridge University Press; 2011 Mar 17.
33. Zeger SL, Qaqish B. Markov regression models for time series: a quasi-likelihood approach. Biometrics. 1988 Dec 1:1019-31.
34. Jones MC. Randomly choosing parameters from the stationarity and invertibility region of autoregressive–moving average models. Journal of the Royal Statistical Society: Series C (Applied Statistics). 1987 Jun;36(2):134-8.
35. Lukacs E. A characterization of the gamma distribution. The Annals of Mathematical Statistics. 1955 Jun 1;26(2):319-24.
36. Plummer M. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. InProceedings of the 3rd international workshop on distributed statistical computing 2003 Mar 20 (Vol. 124, No. 125.10, pp. 1-10).
37. Biau DJ, Kernéis S, Porcher R. Statistics in brief: the importance of sample size in the planning and interpretation of medical research. Clinical Orthopaedics and Related Research®. 2008 Sep 1;466(9):2282-8.
38. Chan AW, Altman DG. Epidemiology and reporting of randomised trials published in PubMed journals. The Lancet. 2005 Mar 26;365(9465):1159-62.
39. Higgins CD, dos-Santos-Silva I, Stiller CA, Swerdlow AJ. Season of birth and diagnosis of children with leukaemia: an analysis of over 15 000 UK cases occurring from 1953-95. British journal of cancer. 2001 Feb;84(3):406-12.
40. Gao F, Nordin P, Krantz I, Chia KS, Machin D. Variation in the seasonal diagnosis of acute lymphoblastic leukemia: evidence from Singapore, the United States, and Sweden. American journal of epidemiology. 2005 Oct 15;162(8):753-63.

medRxiv preprint doi: https://doi.org/10.1101/2023.02.12.23285595; this version posted July 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

41. Hassan J, Adil SO, Haider Z, Zaheer S, Anwar N, Nadeem M, Ansari SH, Shamsi T. Seasonal variations in hematological disorders: A 10□year single□center experience. International Journal of Laboratory Hematology. 2021 Feb;43(1):93-8.
42. Bagirov IA. The seasonal dynamics of morbidity of acute lymphoblastic leukemia in Azerbaijan. Problemy Sotsial'noi Gigieny, Zdravookhraneniia i Istorii Meditsiny. 2019 Sep 1;27(5):911-4.
43. Li SY, Ye JY, Meng FY, Li CF, Yang MO. Clinical characteristics of acute lymphoblastic leukemia in male and female patients: A retrospective analysis of 705 patients. Oncology letters. 2015 Jul 1;10(1):453-8.
44. Kulkarni KP, Marwaha RK. Seasonality in diagnosis of childhood acute lymphoblastic leukemia: impact on disease presentation, survival outcome and resources. Journal of Pediatric Hematology/Oncology. 2013 Jan 1;35(1):81-2.
45. Mutlu M, Erduran E. The relationship between seasonal variation in the diagnosis of acute lymphoblastic leukemia and its prognosis in children. Turkish Journal of Hematology. 2012 Jun;29(2):188.
46. Basta NO, James PW, Craft AW, McNally RJ. Season of birth and diagnosis for childhood cancer in Northern England, 1968–2005. Paediatric and perinatal epidemiology. 2010 May;24(3):309-18.
47. Karimi M, Yarmohammadi H. Seasonal variations in the onset of childhood leukemia/lymphoma: April 1996 to March 2000, Shiraz, Iran. Hematological oncology. 2003 Jun;21(2):51-5.
48. Zannos-Mariolea L, Haidas S, Tzortzatou F, Dentaki-Svolaki K, Kiossoglou K. Epidemiology of acute leukemia of childhood in Greece (author's transl). Nouvelle Revue Francaise D'hematologie. 1975 Nov 1;15(6):649-55.
49. Hems G, Stuart A. Childhood leukaemia in Scotland, 1939–68. Scottish Medical Journal. 1972 Jan;17(1):13-7.
50. Gunz FW, Spears GF. Distribution of acute leukaemia in time and space. Studies in New Zealand. British Medical Journal. 1968 Dec 1;4(5631):604-8.
51. Meigham SS, Knox G. Leukemia In Childhood: Epidemiology In Oregon. Cancer. 1965 Jul 1;18:811-4.
52. Knox G. Epidemiology of childhood leukaemia in Northumberland and Durham. British journal of preventive & social medicine. 1964 Jan;18(1):17.
53. Fraumeni JF. Seasonal variation in leukaemia incidence. British Medical Journal. 1963 Nov 11;2(5369):1408.
54. Pordanjani SR, Kavousi A, Mirbagheri B, Shahsavani A, Etemad K. Temporal trend and spatial distribution of acute lymphoblastic leukemia in Iranian children during 2006-2014: a mixed ecological study. Epidemiology and Health. 2020;42.
55. Zeng HM, Guo Y, Yi XL, Zhou JF, An WB, Zhu XF. Large sample clinical analysis of patients with children acute leukemia in single center. Zhongguo shi yan xue ye xue za zhi. 2011 Jun 1;19(3):692-5.
56. Gilman EA, Sorahan T, Lancashire RJ, Lawrence GM, Cheng KK. Seasonality in the presentation of acute lymphoid leukaemia. British journal of cancer. 1998 Feb;77(4):677.
57. Thorne R, Hunt LP, Mott MG. Seasonality in the diagnosis of childhood acute lymphoblastic leukaemia. British Journal of Cancer. 1998 Feb;77(4):678.
58. Badrinath P, Day NE, Stockton D. Seasonality in the diagnosis of acute lymphocytic leukaemia. British journal of cancer. 1997 Jun;75(11):1711-3.
59. Walker AM, van Noord PA. No seasonality in the diagnosis of acute leukemia in the United States. Journal of the National Cancer Institute. 1982 Dec 1;69(6):1283-8.

medRxiv preprint doi: https://doi.org/10.1101/2023.02.12.23285595; this version posted July 16, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

60. Till MM, Hardisty RM, Pike MC, Doll R. Childhood leukaemia in greater London: a search for evidence of clustering. British Medical Journal. 1967 Sep 9;3(5568):755.
61. Mainwaring D. Epidemiology of acute leukaemia of childhood in the Liverpool area. British Journal of Preventive & Social Medicine. 1966 Oct;20(4):189.
62. Lanzkowsky P. Variation in leukaemia incidence. British Medical Journal. 1964 Apr 4;1(5387):910.
63. Lee JA. Seasonal variation in leukaemia incidence. British Medical Journal. 1963 Sep 9;2(5357):623.
64. Cohen P. The influence on survival of season of onset of childhood acute lymphoblastic leukemia (ALL). Chronobiology international. 1987 Jan 1;4(2):291-7.
65. Edwards JH. The recognition and estimation of cyclic trends. Annals of human genetics. 1961 Sep;25(1):83-7.
66. Santoyo-Sánchez A, Ramos-Peñafiel C, Palmeros-Morgado G, Mendoza-García E, Olarte-Carrillo I, Martínez-Tovar A, Collazo-Jaloma J. Leucemias agudas Características clínicas y patrón estacional. Revista Médica del Instituto Mexicano del Seguro Social. 2014;52(2):176-81.
67. Westerbeek RM, Blair V, Eden OB, Kelsey AM, Stevens RF, Will AM, Taylor GM, Birch JM. Seasonal variations in the onset of childhood leukaemia and lymphoma. British journal of cancer. 1998 Jul;78(1):119-24.
68. van Steensel-Moll HA, Valkenburg HA, Vandenbroucke JP, Van Zanen GE. Time space distribution of childhood leukaemia in the Netherlands. Journal of Epidemiology & Community Health. 1983 Jun 1;37(2):145-8.
69. Ross JA, Severson RK, Swensen AR, Pollock BH, Gurney JG, Robison LL. Seasonal variations in the diagnosis of childhood cancer in the United States. British journal of cancer. 1999 Oct;81(3):549-53.
70. Roger JH. A significance test for cyclic trends in incidence data. Biometrika. 1977 Apr 1;64(1):152-5.
71. Sørensen HT, Pedersen L, Olsen JH, Rothman KJ. Seasonal variation in month of birth and diagnosis of early childhood acute lymphoblastic leukemia. Jama. 2001 Jan 10;285(2):168-9.
72. Douglas S, Cortina-Borja M, Cartwright R. A Quest for Seasonally in Presentation of Leukaemia and Non-Hodgkin's Lymphoma. Leukemia & lymphoma. 1999 Jan 1;32(5-6):523-32.
73. Meltzer AA, Annegers JF, Spitz MR. Month-of-birth and incidence of acute lymphoblastic leukemia in children. Leukemia & lymphoma. 1996 Jan 1;23(1-2):85-92.
74. Mardia KV. Statistics of directional data. Journal of the Royal Statistical Society Series B: Statistical Methodology. 1975 Jul;37(3):349-71.
75. Pordanjani SR, Kavousi A, Mirbagheri B, Shahsavani A, Etemad K. Geographical pathology of acute lymphoblastic leukemia in Iran with evaluation of incidence trends of this disease using joinpoint regression analysis. Archives of Iranian Medicine. 2021 Mar 1;24(3):224-32.
76. Hayes DM. The seasonal incidence of acute leukemia. A contribution to the epidemiology of the disease. Cancer. 1961 Nov;14(6):1301-5.