medRxiv preprint doi: https://doi.org/10.1101/2023.02.12.23285595; this version posted February 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.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 .

Evidence of seasonal variation of childhood acute lymphoblastic leukemia in Sweden

Gleb Bychkov[1◑], Benedicte Bang[1◑], Niklas Engsner[1,3], Mats Heyman[4], Anna Skarin Nordenvall [1,5] Giorgio Tettamanti[1,6], Emeli Ponten[1], Jan Albert [7,8] Rebecka Jörnsten[9], Claes Strannegård[1,2‡], Ann Nordgren [1,10,11,12‡*]

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

◑These authors contributed equally to this work.
‡These authors also contributed equally to this work.
* ann.nordgren@ki.se

Abstract

Background Recent molecular studies of B-cell precursor acute lymphoblastic leukemia (BCP-ALL) have started to delineate the nature and timing of genetic variants and those responsible for subsequent progression to overt leukemia. However, the etiology behind both initiation and progression remains largely unknown. Nonetheless, theories, but also epidemiological evidence, of how exposure to common infections and other microbes in our environment modulates the risk of developing childhood BCP-ALL, have emerged. In light of these theories and the well-known phenomena of seasonality in infectious disease spread, childhood ALL has been analyzed for signs of seasonal variation, with differing results.

Methods In this study we applied the Bayesian Generalized Auto Regressive Integrated Moving Average with external variables (GARIMAX) model, adapted for count data via a negative binomial distribution, to study seasonal variation of incidence in a Swedish population-based cohort of 1601 BCP-ALL cases. The studied cases were

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 February 12, 2023. 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 .

aged 0-18 years at diagnosis and identified from the Swedish Childhood Cancer Registry (SCCR). Also, two subgroups of BCP-ALL represented by the most abundant genetic subtypes, *ETV6/RUNX1* and HeH respectively, were analyzed accordingly. All analyses were performed in two stages. The first stage identified the presence of the repeatable pattern using harmonic functions, and the second stage consisted of the identification of the peak months in the series.

Results An informative seasonal variation in BCP-ALL incidence numbers, displaying a peak in August and September, was detected in the total cohort of 1601 individuals. No seasonality was detected analyzing the subtype groups HeH and *ETV6/RUNX1* positive BCP-ALL, respectively.

Conclusion The manifested seasonality in BCP-ALL with a peak in August-September may suggest that the prolonged period of minimal viral spread during Swedish summer vacation causes a temporary halt in the last step of progression to overt disease and consequently an accumulated number of cases presenting in August-September.

Keywords: acute lymphoblastic leukemia, ALL, seasonal variation

Introduction

Acute lymphoblastic leukemia (ALL) is the most frequent cancer (25%) in children below the age of 15 years [1]. ALL is most commonly (85%) of B-cell precursor origin, termed BCP ALL [2]. To date, 11 distinct molecular subtypes based on recurrent genetic aberrations have been defined in BCP ALL, displaying highly variable prognosis and used to guide modern treatment strategies [3]. The most abundant subtypes, *ETV6/RUNX1* fusion and high hyperdiploidy (HeH), also have the most favorable prognosis [4].

The annual Swedish ALL incidence rate of 4.2 per 100 000 in children 0-15 years of age is comparable to numbers in other affluent societies [5]. A significative peak in incidence is seen at 2-5 years of age [6, 7] mainly comprised by *ETV6/RUNX1* positive and HeH cases [6, 8, 9]. However, markedly lower incidence has been reported especially from sub-Sharan countries [10] where the early incidence peak is less pronounced or even non-existing [11].

The cause of childhood BCP-ALL is yet largely unknown and likely multifactorial, comprised of both environmental and genetic factors. Mapping the temporal aspects of disease initiation and progression has been central to our understanding of its etiology. There is today compelling evidence that BCP-ALL is initiated prenatally. Pre-leukemic clones have been detected in children who later develop BCP-ALL but also in healthy controls [12–18]. As many as 1-5% of healthy neonates have been found to harbor a pre-leukemic clone of ETV6/RUNX1 positive BCP-ALL [19, 20]. However, secondary genetic aberrations are required for the progression of pre-leukemic clones [21, 22], explaining why only a fraction of pre-leukemic cell carriers develop overt BCP-ALL.

Although the nature of both initiating and secondary genetic variants in BCP-ALL have been delineated to some extent, including leukemia predisposing germline variants in an approximated 4,4% [23], little is known of the environmental drivers behind these genetic aberrations. To date, ionizing radiation is the only environmental factor convincingly proven to increase the risk of ALL [24, 25]. However, the role of common pathogens such as viruses and bacteria has been given great interest in both epidemiological and molecular studies in recent years.

The initiation of pre-leukemic clones by infectious agents has to date not been molecularly proven. However, recent studies have indeed confirmed that progression to overt leukemia from a pre-leukemic state may be promoted by a specific pathogen [26], but also that it requires exposure to common infections [27, 28]. On the more general

medRxiv preprint doi: https://doi.org/10.1101/2023.02.12.23285595; this version posted February 12, 2023. 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 .

level, models for infectious exposures selective pressure on leukemia progression have been suggested. Originally, in 1988 Kinlen formulated the hypothesis of "population mixing" after observing an increased incidence of childhood BCP-ALL in immunologically naïve previously isolated populations after exposure to a common mild infectious agents transmitted by individuals from urbanized areas of residence [29–31]. Simultaneously, Greaves suggested a model where delayed exposure to common infections during early childhood causes strong adverse reactions of the immune system once infections are encountered, in turn promoting progression of pre-leukemic cells [32]. As reviewed by Hauer et al [33], more recent data building on these original models suggests early training of the innate immune system has a protective effect against BCP-ALL progression. This is based on observations of reduced BCP-ALL risk following early animal contact, early daycare attendance, vaginal delivery, breastfeeding, having older siblings, early BCG vaccination etc. In summary, both what infections we encounter and when during fetal life and early childhood we encounter them appear to influence the destiny of pre-leukemic cells initiated as a consequence of in some cases genetic predisposition and other yet unknown environmental factors.

One epidemiological aspect studied based on the above hypothesis is seasonal variation in incidence of BCP-ALL and likewise in BCP-ALL patients' season of birth. As the spread of common infections (such as influenza, adeno, corona, rhino and rs-viruses among others) vary strongly with season in temperate areas of the world, the hypothesis has been that if such infections impact disease progression a seasonal pattern in incidence of BCP-ALL, and possible also patients season of birth, may also be seen. The first to report seasonal variation of acute leukemia incidence were Lampin and Gerard [34] in 1934. In a literature search we identified 41 papers published from 1961 an onward, investigating seasonal variation in ALL (predominantly in childhood) by a variety of methods and with inconsistent results (Table 1). 20 publications reported significant seasonal variation of ALL incidence [35, 36, 38, 39, 41, 44, 45, 47, 49, 50, 52, 53, 55, 56, 58, 63, 69–72], while 21 papers reported no evidence of seasonality [37, 40, 42, 43, 46, 48, 49, 51, 54, 56, 57, 59–62, 64–68, 73].

In the current study we applied Generalized Autoregressive Integrated Moving Average model with External seasonal covariates (GARIMAX) to a Swedish population-based cohort of 1601 childhood BCP-ALL cases to investigate presence of seasonal variation in incidence of diagnosis. Also, two subgroups of BCP-ALL represented by the most abundant genetic subtypes, ETV6/RUNX1 and HeH respectively, were analyzed for seasonal variation of incidence.

medRxiv preprint doi: https://doi.org/10.1101/2023.02.12.23285595; this version posted February 12, 2023. 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.

Table 1. Studies of seasonal variation at the date of diagnosis in ALL.

Authors	Year*	Country**	N***	Years****	Age	Methods	Seasonality*****	Other comments
Hassan J. et al [35]	2021	Pakistan	513	2006-2015	All ages	Single-factor analysis of variance and counts, Chi-square test	Yes (June to September)	Monsoon period, BCP- and T-cell ALL analyzed together
Rahimi Pordanjani S. al [36]	2021	Iran	3769	2006-2014	0-14	Joint point regression (regression over aggregated monthly counts)	Yes (June to September)	BCP- and T-cell ALL analyzed together
Bamouni S. et al [37]	2021	France	9493	1990-2014	0-14	Poisson regression with harmonic functions	No	BCP- and T-cell ALL analyzed together
Rahimi Pordanjani S. et al [38]	2020	Iran	3769	2006-2014	0-14	Single factor analysis, temporal trend	Yes (June to September)	BCP- and T-cell ALL analyzed together
Bagirov I.A. [39]	2019	Azerbaijan	991	1998-2014	<29	Single-factor analysis of variance and counts	Yes (Summer)	BCP- and T-cell ALL analyzed together
Nurullah R. et al [40]	2018	Canada	364	1995-2015	0-20	Poisson regression with harmonic functions	No	BCP- and T-cell ALL analyzed together
Shim K.S. et al [41]	2017	South Korea	Appr. 1150	2009-2013	<21	ARIMA	Yes (Dec-Feb)	BCP- and T-cell ALL analyzed together, peak incidence in dec-feb with decreasing trend until sep. strongest correlation to HPIV,
Li S.Y. [42]	2015	China	705	Jan 2001–Dec 2012	All ages	Single-factor analysis of variance and counts, Chi-square test	No	BCP- and T-cell ALL analyzed together
Santoyo-Sánchez A. et al [43]	2014	Mexico	833	Jan 2006-Apr 2012	All ages	Edward's test	No	BCP- and T-cell ALL analyzed together
Kulkarni K.P. et al [44]	2013	North India	446	1990-2006, 2009	4.3 ± 2.4 mean age	Single-factor analysis of seasonal counts, Chi-square test	Yes (Aug-Nov)	BCP- and T-cell ALL analyzed together

medRxiv preprint doi: https://doi.org/10.1101/2023.02.12.23285595; this version posted February 12, 2023. 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.

Authors	Year*	Country**	N***	Years****	Age	Methods	Seasonality*****	Other comments
Goujon-Bellec S. et al [45]	2013	France	6686	1990–2007	0-14	Poisson regression with harmonic functions, negative binomial regression to account for overdispersion	Yes (April, Aug, Dec)	the study showed an increase in childhood ALL risk, which tended to be stronger for 7–14-year-old Bcp-ALL, particularly in girls (peak ?), Seasonal variations in the month of diagnosis were also evidenced for 1–6-year-old boys, with a 10% increase in the risk for all ALL and Bcp-ALL in April, August and December
Mutlu M. et al [46]	2012	Turkey	137	1990-2004	8 months-16 years	Single-factor analysis of monthly counts, Chi-square test	No	BCP- and T-cell ALL analyzed together
Zeng H.M. et al [47]	2011	China	631	Apr 2004-Apr 2010	0-16	Single-factor analysis of monthly counts	Yes (Jan)	Winter, especially January was the peak time for both diagnosis and birth, BCP- and T-cell ALL analyzed together
Basta N.O. et al [48]	2010	Northern England	743	1968-2005	0-6	Poisson regression with harmonic functions fitted to 12 month count period, Chi-square test	No	There was significant sinusoidal variation based on month of birth for acute lymphoblastic leukaemia (ALL) aged 1-6 years (P = 0.04; peak in March), BCP- and T-cell ALL analyzed together
Gao F. et al [49]	2005	Singapore	684	1968-1999	0-19	von Mises distribution and the Mardia test	No	BCP- and T-cell ALL analyzed together
Gao F. et al [49]	2005	The United States	6181	1973-1999	0-19	von Mises distribution and the Mardia test	No	BCP- and T-cell ALL analyzed together
Gao F. et al [49]	2005	Sweden	63	1977-1994	0-19	von Mises distribution and the Mardia test	Yes (Jan)	BCP- and T-cell ALL analyzed together
Karimi M. et al [50]	2003	Iran	221	Apr 1996-Mar 2000	0-14	Chi-square test and normal approximation to Poisson for analyzing	Yes (Oct, Nov)	BCP- and T-cell ALL analyzed together

medRxiv preprint doi: https://doi.org/10.1101/2023.02.12.23285595; this version posted February 12, 2023. 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.

Authors	Year*	Country**	N***	Years****	Age	Methods	Seasonality*****	Other comments
Higgins C.D. et al [51]	2001	The UK	15 835	1972-1986	0-15	Edward's test	No	BCP- and T-cell ALL analyzed together
Sørensen H.T. et al [52]	2001	Denmark	458	1950–1994	0-4	Cosinor analysis	Yes (Oct)	Date of birth, peak month was April, BCP- and T-cell ALL analyzed together
Ross J.A. et al [53]	1999	USA	5532	Jan 1989-31 Dec 1991	0-19	Rodger's test	Yes (summer)	BCP- and T-cell ALL analyzed together
Douglas S. et [54]	1999	England	789	1984-1993	0-14	Cosinor analysis and Normal Approximation to the Poisson Distribution	No	BCP- and T-cell ALL analyzed together
Gilman E.A. et al [55]	1998	Great Britain	805	1971-1994	0-14	Single-factor analysis of variance and counts by season	Yes (summer)	data showed a 16% excess of cases diagnosed in the summer months in children, BCP- and T-cell ALL analyzed together
Westerbeek R.M. et al [56]	1998	NW England	1070	Jan 1954-Dec 1996	0-14	Edward's test	No	BCP- and T-cell ALL analyzed together
Westerbeek R.M. et al [56]	1998	East Anglia, UK	271	1971-1994	0-14	Edward's test	Yes (summer)	BCP- and T-cell ALL analyzed together
Thorne R. et al [57]	1998	South-west of England	420	1976-1995	0-14	Single-factor analysis	No	BCP- and T-cell ALL analyzed together
Badrinath P. et al [58]	1997	East Anglia, UK	271	1971-1994	0-14	Single-factor analysis of variance and counts by season	Yes (summer)	The seasonality was found in the whole ALL group, but there is no suggestion of similar seasonality for any other cell types of leukaemia
Meltzer A.A. et al [59]	1996	Atlanta, Connecticut, Detroit, Hawaii, Iowa, New Mexico, Puerto, Rico, San Francisco, Seattle, and Utah	1487	1973-1986	0-15	Cosinor analysis	No	no evidence of seasonality at date of birth found, BCP- and T-cell ALL analyzed together
Cohen P. [60]	1987	Israel	205	1976-1981	5.9 ± 3.94 mean age	Single-factor analysis of variance and counts	No	No seasonal onset of disease was found, either in the whole group or in subgroups based on cell type

medRxiv preprint doi: https://doi.org/10.1101/2023.02.12.23285595; this version posted February 12, 2023. 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.

Authors	Year*	Country**	N***	Years****	Age	Methods	Seasonality*****	Other comments
van Steensel-Moll H.A. et al [61]	1983	the Netherlands	293	1973-80	0-14	Edward's test	No	BCP- and T-cell ALL analyzed together
Walker A.M., van Noord P.A. [62]	1982	The USA	1783	1969–1977	All ages	Single factor analysis	No	No strong evidence was found for seasonality in the diagnosis of acute leukemias as a whole or for subgroups based on cell type
Zannos-Mariolea L. et al [63]	1975	Greece	151	-	0-14	Single-factor analysis, Chi-squared test	Yes (winter)	BCP- and T-cell ALL analyzed together
Hems G., Stuart A. [64]	1972	Scotland	978	1939–1968	0-15	Single-factor analysis, Chi-square test	No	BCP- and T-cell ALL analyzed together
Gunz F.W., Spears G.F. [65]	1968	New Zealand	288	1953-1964	All age groups	Single-factor analysis, Chi-squared test	No	Significant seasonal variations in the onset were found in adults, BCP- and T-cell ALL analyzed together
Till M.M. et al [66]	1967	Greater London, England	374	1952-1961	0-9	Single factor analysis	No	BCP- and T-cell ALL analyzed together
Mainwaring D. [67]	1966	Liverpool	74	1955-64	0-14	Single-factor analysis	No	Younger age group more common in summer, BCP- and T-cell ALL analyzed together
Meighan S.P. et al [68]	1965	Oregon, the USA	214	1950-1961	0-14	Single factor analysis, Chi-squared test	No	BCP- and T-cell ALL analyzed together
Knox G. [69]	1964	Northumberland, Durham	185	1951-1960	0-14	Single-factor analysis, Chi-Squared test	Yes (summer)	BCP- and T-cell ALL analyzed together
Lanzkowsky P. [70]	1964	South Africa	27	Only states data collected "in the past six years".	0-12	Single-factor analysis	Yes (summer)	BCP- and T-cell ALL analyzed together
Fraumeni J.F. [71]	1963	Washington DC, the US	237	1958-1961	0-15	Single-factor analysis and variance, Chi-square test	Yes (spring)	BCP- and T-cell ALL analyzed together

medRxiv preprint doi: https://doi.org/10.1101/2023.02.12.23285595; this version posted February 12, 2023. 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 .

Authors	Year*	Country**	N***	Years****	Age	Methods	Seasonality*****	Other comments
Lee J.A.M. [72]	1963	England and Wales	548	1946-1960	0-18	Single-factor analysis	Yes (summer)	BCP- and T-cell ALL analyzed together
Hayes D.M. [73]	1961	North Carolina	184	1943-1950	All ages	Statistical comparisons between the seasonal curve and a hypothetical random distribution	No	BCP- and T-cell ALL analyzed together

* Year of publication
** Number of cases in analyzed cohort
*** Country or region of data collection
**** Observed period
***** Seasonality for date of diagnosis reported in the study Yes/No. Particular month/season in brackets.

medRxiv preprint doi: https://doi.org/10.1101/2023.02.12.23285595; this version posted February 12, 2023. 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 .

Materials 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) is a National Quality Registry containing information about children diagnosed with tumors and hematological malignancies between 0 and 18 years of age stretching back to the 1970s for ALL and 1980s for all other malignancies. The registry has an overall coverage of 89% for all diagnoses, however, coverage for ALL specifically is estimated to be as good as 100% at present. It includes information about clinical characteristics, treatment, outcome, immunophenotype, genetic subtype, and other clinically important genetic aberrations. The most abundant BCP-ALL genetic subtypes, 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) holds information on the date of birth, death, and emigration for all Swedish citizens.

The main data sources of this paper the Swedish Childhood Cancer Registry (SCCR), and The Total Population Register (RTB).

The study population

From the Swedish Childhood Cancer Registry (SCCR) we identified a cohort of 1601 children and adolescents diagnosed with BCP-ALL at age of <18 years between January 1, 1995 and December 31, 2017. Only individuals born and diagnosed in Sweden were included in the study. For the purpose of this study, anonymized data of date of diagnosis, date of birth, and genetic subtype were collected. Patient characteristics are summarized in Table 2. The overall mean age at diagnosis was 6.8 years, with *ETV6/RUNX1* carriers having a lower mean age than HeH and the non-HeH/ *ETV6/RUNX1* BCP-ALL group) (5.0 vs. 5.3 and 8.1 years). In the cohort, 448 cases were of HeH and 272 cases of *ETV6/RUNX1*-fusion BCP-ALL subtype. The mean age at diagnosis was 6.8 years for all cases, 5.0 for *ETV6/RUNX1*-positive cases, 5.3 for HeH cases, and 8.1 for other and undefined genetic subtypes.

To study seasonality by date of diagnosis, we included all 1601 BCP-ALL cases described above. Individual cases of BCP-ALL were summed up in a total of 92 quarters to facilitate time series of quarterly counts. Three types of quarters within a year were defined as follows: Jan-Mar, Apr-Jun, Jul-Sep, Oct-Dec (first quarter type), Feb-Apr, May-Jul, Aug-Oct, Nov-Jan (second quarter type), and Mar-May, Jun-Aug, Sep-Nov, Dec-Feb (third quarter type).

Statistical methods

We implemented the Bayesian Generalized Auto Regressive Integrated Moving Average model with exogenous variables (GARIMAX) [74, 75] for identification of seasonal variation in BCP-ALL. The key elements of the model are (i) generalization of the ARIMAX process to a count distribution via negative binomial distribution, and (ii) a Bayesian formulation for model setup. The generalization allowed us to search for seasonality in the sparse (counts of low values) data [76], whereas the Bayesian formulation is beneficial in the applications of the complex models in small sample settings [77]. We used BIC (Bayesian information criterion) for the model selection. Chi-square tests were performed for exploratory purposes.

medRxiv preprint doi: https://doi.org/10.1101/2023.02.12.23285595; this version posted February 12, 2023. 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 .

Table 2. Frequencies and proportions of baseline clinical characteristics among Swedish patients diagnosed with BCP-ALL

	BCP-ALL overall	HeH	*ETV6/RUNX1*	Non HeH/ *ETV6/RUNX1*, other types
Year of diagnosis				
1995-2005	795 (0.4966)	218 (0.4866)	93 (0.3419)	484 (0.5494)
2005-2017	806 (0.5034)	230 (0.5134)	179 (0.6581)	397 (0.4506)
Age at diagnosis				
0-5	837 (0.5228)	295 (0.6585)	169 (0.6213)	373 (0.4234)
5-10	424 (0.2648)	105 (0.2344)	87 (0.3199)	232 (0.2633)
10-18	340 (0.2124)	48 (0.1071)	16 (0.0588)	276 (0.3133)
Sex				
Male	906 (0.5659)	236 (0.5268)	162 (0.5956)	508 (0.5766)
Female	695 (0.4341)	212 (0.4732)	110 (0.4044)	373 (0.4234)

ARIMAX model

ARIMAX stands for autoregressive integrated moving average with external variables and was proposed by Box and Jenkins [78] in 1970. The assumption in the AR (autoregressive) process is that the mean at time t (expected value) depends on the previous realization of the process and the MA (moving average) part is that the additive error is correlated across time. The I (integration) part refers to the fact that the process is an integration of a process. The X part (external variable) introduces independent variables to the process.

The ARIMAX is denoted by parameters (p, d, q), where p is the order of autoregression (indicates how many previous observations to use in the AR), d is the differencing order (indicates how many times to differentiate the response variable), and q is the order of moving average (indicates how many previous errors to use in the MA).

The ARIMA model is widely used for modeling seasonal patterns and trends in economics in the analysis of Gross Domestic Product, inflation, demand [79–81], and finance in predicting stock prices [82]. It has also been widely used in medical research. For example, the ARIMA model was implemented to forecast COVID-19 cases using Johns Hopkins data [83], and to analyze malaria cases in Sri Lanka [75]. ARIMAX is a powerful statistical method in the analysis and forecasting of time series data.

The ARIMAX(p, d, q) for y_t in the general form can be written:

$$\Delta^d \Phi_p(L)[y_t - x_t^T \beta] = \Theta_q(L) u_t, \tag{1}$$

where y_t is an observation of the series at time t, p is the order of autoregression, q is the order of moving average. Moreover, $\Phi_p(L) = (1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p)$, and $\Theta_q(L) = (1 + \theta_1 L + \theta_2 L^2 + ... + \theta_q L^q)$, where $\phi_1, ..., \phi_p$ are the coefficients for the autoregressive part of the process, $\theta_1, ..., \theta_q$ are the coefficients of the moving average part of the model. L is a backshift operator with $L^i y_t = y_{t-i}$. The lag operator can be multiplied such that $L^i L^j y_t = y_{t-i-j}$. u_t is a white noise, or uncorrelated, error process at time t, Δ^d is a differencing operator of order d. The differentiation method Δ is chosen to be log differentiation, so $\Delta^d y_t$ denotes log differentiation of y_t taken d times. For example $\Delta y_t = log(y_t) - log(y_{t-1})$. β is a vector of coefficients for covariates x_t^T.

medRxiv preprint doi: https://doi.org/10.1101/2023.02.12.23285595; this version posted February 12, 2023. 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 .

Generalization of ARIMAX

While the classical ARIMAX model assumes normally distributed data, BCP-ALL case data can only take integer values from minimum value 0 to maximum value 16, which breaks the assumption of the classical model. Generalization of ARIMAX allows modeling the series of non-normal distributions [76].

The go-to generalization to the discrete data is performed via Poisson distribution. We can see that several papers in our literature search used Poisson regression as a method for obtaining seasonality in leukemia case data [37, 40, 45]. One of the assumptions of Poisson distribution is that the mean and variance of the process are the same. The observed excess variability compared with the Poisson count model motivates use of a negative binomial formulation within the ARIMAX framework. Conditional variance of the negative binomial distribution exceeds the conditional mean. The source of the overdispersion is the unobserved heterogeneity caused by hidden variables as the harmonic and seasonal covariates are just proxies of infectious agents. NB distribution compensates for the lack of fit by introducing an extra parameter [84, 85]. In the literature survey generalization via NB distribution was used by Goujon-Bellec et al [45]. The parameters of NB distribution are pr and r, where pr denotes the probability of success, and r is the number of successes before trials stop. Poisson distribution is the limiting form of NB distribution when $r \to \infty$.

The parameter of interest, pr, is assumed to depend on the season.

The score of the ARIMAX is generalized to the parameter of NB distribution pr_t by link function $pr_t = g(\lambda_t)$ [76], where $g(.)$ is a link function, and λ_t is the score of ARIMAX process (predicted y_t in equation 1), the score λ_t becomes:

$$\lambda_t = y_t - \Delta^d y_t + \Delta^d x_t^T \beta + \sum_{i=1}^{p} [\Delta^d y_{t-i} - \Delta^d x_t^T \beta] + \sum_{j=0}^{q} u_t, \tag{2}$$

The link function used in the paper is $g(x) = \frac{r}{r+x}$. So, given pr_t and estimated r $y_t \sim NB(pr_t, r)$. The generalized ARIMAX model called GARIMAX model [75].

Following Zeger and Qaqish [86] we implement "ZQ1" transformation of the ARIMAX by adding a constant c. Addition of the constant allows avoiding the problem of computing the logarithm of observations with the zero value in the log difference integration transformation of the model. "ZQ1" transformation suggests $y'_t = y_t + c$, where $0 < c \leq 1$.

GARIMAX in the Bayesian setup

The formulation of the model in the Bayesian setup gives several advantages over the classical maximum likelihood estimation. The first one is that the analysis is no longer performed on a single estimate, but rather on the distributions of the underlying parameters. The Bayesian models with the correctly specified priors and estimation using MCMC allow for the parameters' interval estimates to be appropriate in small samples [77]. This advantage of the Bayesian framework allows for unbiased inference even in small samples.

We assume a stationary model for the GARIMAX process, so AR and MA coefficients $\phi_1, ...\phi_p$ and $\theta_1, ..., \theta_q$ should be constrained such that the resulting process is invertible and stationary. We followed Jones [87] in assigning prior distributions for AR and MA coefficients. The algorithm for generation of the sample of $\phi_1, ..., \phi_p$ can be summarized as:

Algorithm:

medRxiv preprint doi: https://doi.org/10.1101/2023.02.12.23285595; this version posted February 12, 2023. 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 .

- Generate value $k_1, ..., k_p$ following the $k_p \sim \text{Beta}([\frac{1}{2}(j+1)], [\frac{1}{2}j]+1)$, where $p = 1, ..., P$, where P is the number of lags of autoregression, square brackets denote the integer part of the value in them (round to the closest integer);
- perform transformation $r_p = 2k_p - 1$ for all p;
- assign $y_p^{(p)} = r_p$ for all p;
- and then for $j = 2, ..., p$ and for $i = 1 : (j-1)$ iteratively compute $y_i^{(j)} = y_i^{(j-1)} - r_j y_{j-i}^{(j-1)}$;
- y_i is the sample for ϕ_i, where $i \in 1, .., p$

The same procedure is performed for the θ coefficients, but instead of the lags for the autoregression (p) the lags for the moving average (q) are used.

For the seasonal coefficients of the harmonic functions and the coefficients of the seasonal matrix normal distribution was selected, $\beta_1, \ldots, \beta_k \sim \text{Norm}(0, 0.1)$. Distribution is centered around the 0 value, the prior states that there is no evidence of seasonality of BCP-ALL no its subtypes before the data is introduced.

The last parameter to be estimated in the model is the parameters of the Negative Binomial distribution r, which represent the number of failures until the trials are stopped. Prior distribution for r is gamma [88], $r \sim \text{Gamma}(0.01, 0.01)$.

The model is estimated using MCMC (Markov chain Monte Carlo) simulation method. JAGS [89] software is used to implement the simulation. Python is the main programming language used for data preparation, visualization and calls for JAGS software.

Model choice

To identify a number of lags for autoregression and moving average parameters, 12 models were estimated. The maximum 3rd order of the lags was chosen to identify the best GARIMAX model for each analysed time series. We assume a weakly stationary series after taking log difference. Bayesian information criteria (BIC) was chosen as a score for model selection. The model with the lowest BIC was then selected as a basis for the seasonality checks. The BIC allows finding a balance between the complexity of the model and its performance. The BIC for the model m is defined as:

$$BIC_m = -2LL_m + log(N)k, \tag{3}$$

where LL_m is the median log-likelihood computed by the model m, N is the number of observations of the analyzed series, and k is the number of the estimated parameters of the model.

General procedure

We implemented a two-step procedure for identification of seasonal wave for every analysed time series. The preliminary step aims to identify the number of lags for AR and MA coefficients using the BIC score without any seasonal covariates.

The first step uses harmonic functions as a covariate for the GARIMAX model. The harmonic covariate is $x_t^T = [sin(2\pi t \div 4), cos(2\pi t \div 4)]$.

During the second step, we run the same specification of the GARIMAX model with a quarterly seasonal matrix as a covariate. The third step identifies the particular quarter when the wave has its peak. The covariates in this case become a seasonal matrix, in which each column value corresponds to a specific quarter of the year, except

medRxiv preprint doi: https://doi.org/10.1101/2023.02.12.23285595; this version posted February 12, 2023. 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 .

for the so-called "base quarter", which is taken as the quarter with the lowest number of cases.

The main underlying process that models the dynamics of the observations at time t is ARIMAX. The score is then generalized via Negative Binomial distribution, using of the discrete NB distribution. The generalization addresses the assumption of the count nature of date of diagnosis quarterly case time series data.

Exploratory analysis

As a supplementary and exploratory analysis, the most popular test in the literature survey (Chi-square test) was performed on date of diagnosis and date of birth data of BCP-ALL. Chi-square test is a statistical test for categorical data, and it is used to determine whether the data is significantly different from the expected value. The test statistics can be formally written as:

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

, where χ^2 is the test statistic, O is observed frequency, E is the expected frequency. P-values were adjusted using Benjamini-Hochberg procedure [90] and all tests were performed using the Python programming language.

Also as an exploratory measure, we generated descriptive data on distribution of age at diagnosis and genetic subtypes, for the whole cohort as well as for cases diagnosed in each respective month of the year.

Results

The best models for all three types of quarters are presented in the table 3.

Table 3. The best GARIMAX (p, d, q) specifications for each quarterly series

	BCP-ALL	HeH	*ETV6/RUNX1*
1st type quarter	GARIMAX(3, 1, 0)	GARIMAX(3, 1, 0)	GARIMAX(4, 1, 0)
2nd type quarter	GARIMAX(4, 1, 0)	GARIMAX(4, 1, 0)	GARIMAX(3, 1, 0)
3rd type quarter	GARIMAX(2, 1, 0)	GARIMAX(3, 1, 0)	GARIMAX(2, 1, 0)

All BIC scores for the models are presented in S1 Appendix

After identification of the order of autoregression and moving average, we estimated the model with harmonic functions as covariates to identify presence of seasonal wave in the series.

Tables 4 and 5 report summary statistics for posterior distributions of the coefficients of seasonal harmonic functions. The first column of the table is the name of the analyzed series, the second column is the specification of the GARIMAX model, the column "Waves" specifies the type of the harmonic function, the forth column reports the median value of the distribution, the last two column report 95% credibility interval. If the credibility interval fully consists of positive or negative values, it means that 95% of the posterior does not contain 0 and it is unlikely that that the covariate has no effect on the response variable. The covatiate in this case is said to be informative in the Bayesian setup, the corresponding term in the classical statistics is significant. If the credibility interval contains the 0 value, it means that the big mass of the posterior is centered around 0 value and the covariate is uninformative.

Supplementary Figures S1 Fig, S2 Fig, S3 Fig show graphical distributions of posterior densities of the coefficients of seasonal harmonic functions (the visual

medRxiv preprint doi: https://doi.org/10.1101/2023.02.12.23285595; this version posted February 12, 2023. 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 .

Table 4. Summary of posterior distributions of the coefficients of seasonal harmonic functions, BCP-ALL

Series	(p,d,q)	BIC scores	Waves	Median	2.5%	97.5%
BCP-ALL 1st	(3, 1, 0)	526.02	Sin wave (β_1)	-0.107	-0.1863	-0.0288
type quarter			Cos wave (β_2)	-0.0314	-0.114	0.0486
BCP-ALL 2nd	(3, 1, 0)	517.08	Sin wave (β_1)	-0.1042	-0.2092	-0.0054
type quarter			Cos wave (β_2)	-0.0366	-0.1525	0.0699
BCP-ALL 3rd	(4, 1, 0)	535.54	Sin wave (β_1)	-0.0322	-0.1159	0.0497
type quarter			Cos wave (β_2)	-0.1212	-0.2017	-0.0419

representation of the table 4 and 5). The red line on the figure represents the zero value of the coefficients.

The credibility intervals of posterior densities of the harmonic functions do not include 0 values for sine harmonic waves of BCP-ALL of the 1st and 2nd specified quarter type, and cosine harmonic wave of BCP-ALL of the 3rd quarter type.

Table 5. Summary of the posterior distributions of the coefficients of the seasonal harmonic functions, BCP-ALL subtypes

Series	(p,d,q)	BIC scores	Waves	Median	2.5%	97.5%
HeH 1st	(4, 1, 0)	403.86	Sin wave (β_1)	-0.0303	-0.2153	0.1812
type quarter			Cos wave (β_2)	-0.0521	-0.2377	0.1491
HeH 2nd	(4, 1, 0)	405.01	Sin wave (β_1)	-0.0462	-0.2462	0.1496
type quarter			Cos wave (β_2)	0.018	-0.1832	0.2066
HeH 3rd	(3, 1, 0)	410.39	Sin wave (β_1)	0.0653	-0.1312	0.2645
type quarter			Cos wave (β_2)	-0.0985	-0.2946	0.0963
ETV6/RUNX1 1st	(2, 1, 0)	354.84	Sin wave (β_1)	0.1052	-0.0572	0.2626
type quarter			Cos wave (β_2)	0.0562	-0.1057	0.2184
ETV6/RUNX1 2nd	(3, 1, 0)	320.42	Sin wave (β_1)	-0.0999	-0.3933	0.1864
type quarter			Cos wave (β_2)	-0.0366	-0.3352	0.2707
ETV6/RUNX1 3rd	(2, 1, 0)	343.50	Sin wave (β_1)	-0.0911	-0.2583	0.0779
type quarter			Cos wave (β_2)	-0.0272	-0.1865	0.1311

The credibility intervals of posterior distribution of the coefficients of seasonal harmonic functions do include the value 0, which means that harmonic functions do not provide much information in explaining the dynamics of the case counts of BCP-ALL subtypes HeH and *ETV6/RUNX1*. The informative seasonal waves are: sine wave of BCP-ALL series of the 1st type quarter, sine wave of BCP-ALL series of the 2nd type quarter, the cosine wave of BCP-ALL series of the 3rd type quarter. No informative seasonal waves for the BCP-ALL subtypes were found.

Tables 6 and 7 report the summary for the posterior distributions of the coefficients of the seasonal matrix. The first column is the name of the series. Base quarter for each series was chosen such that all other quarters show positive median values. The last three columns of the table are median and 95% credibility intervals.

Figures S4 Fig, S5 Fig, S6 Fig show the visual representation of the posterior

medRxiv preprint doi: https://doi.org/10.1101/2023.02.12.23285595; this version posted February 12, 2023. 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 .

Table 6. Summary of the posterior distributions of the coefficients of the seasonal matrix, BCP-ALL

Series	Base quarter	Estimated quarters	Median	2.5%	97.5%
BCP-ALL 1st type quarter	Jan-Mar	Apr-Jun	0.0952	-0.4695	0.308
		Jul-Sep	0.2095	0.0463	0.3796
		Oct-Dec	0.0294	-0.5489	0.2582
BCP-ALL 2nd type quarter	Nov-Jan	Feb-Apr	0.0031	-0.2901	0.2919
		May-Jul	0.0813	-0.1445	0.3041
		Aug-Oct	0.2088	0.0730	0.4986
BCP-ALL 3rd type quarter	Dec-Feb	Mar-May	0.1354	-0.0422	0.3124
		Jun-Aug	0.2487	-0.0788	0.4233
		Sep-Nov	0.1901	-0.0105	0.4421

distributions of the coefficients of the seasonal matrix. The base quarter presented as a blank picture.

Table 7. Summary of the posterior distributions of the coefficients of the seasonal matrix, BCP-ALL subtypes

Series	Base quarter	Estimated quarters	Median	2.5%	97.5%
HeH 1st type quarter	Oct-Dec	Jan-Mar	-0.0041	-0.2813	0.2755
		Apr-Jun	0.1593	-0.1335	0.4614
		Jul-Sep	0.228	-0.0374	0.4962
HeH 2nd type quarter	May-Jul	Feb-Apr	0.0288	-0.4287	0.4466
		Aug-Oct	0.1227	-0.3106	0.5008
		Nov-Jan	0.0274	-0.3739	0.4129
HeH 3rd type quarter	Mar-May	Jun-Aug	0.2912	-0.0625	0.6493
		Sep-Nov	0.0941	-0.3191	0.4861
		Dec-Feb	0.0677	-0.3165	0.4321
ETV6/RUNX1 1st type quarter	Apr-Jun	Jan-Mar	0.3394	0.0308	0.6457
		Jul-Sep	0.1986	-0.1151	0.5093
		Oct-Dec	0.3135	-0.0090	0.6271
ETV6/RUNX1 2nd type quarter	May-Jul	Feb-Apr	0.2626	-0.1695	0.6922
		Aug-Oct	0.4715	0.0572	0.9090
		Nov-Jan	0.2790	-0.1615	0.7104
ETV6/RUNX1 3rd type quarter	Mar-May	Jun-Aug	0.0262	-0.2791	0.3425
		Sep-Nov	0.2180	-0.1384	0.5679
		Dec-Feb	0.0202	-0.2842	0.3186

Jul-Sep and Aug-Oct are the informative peak quarters for BCP-ALL series for the first and the second quarter types accordingly. Peak seasons contain August and September, the months with the highest number of diagnosed BCP-ALL cases in our studied Swedish cohort. It is important to observe the significance on both stages of the analysis, as the seasonal matrix does not consider repeatability of seasonal variation, while harmonic seasonal waves do not provide information concerning the peak months.

The high hyperdiploid genetic subtype of BCP-ALL does not show any informative result for any quarter types not in the first nor the second steps of analysis. Diagnosis of *ETV6/RUNX1* positive BCP-ALL shows an increase in Jan-March and Aug-Oct quarters, which might suggest a two-peaked seasonal wave, which however did not yet form the strong periodicity tested by strictly periodic harmonic functions. On the other hand, the second step of analysis does not test for seasonality, it tests for the increase of

medRxiv preprint doi: https://doi.org/10.1101/2023.02.12.23285595; this version posted February 12, 2023. 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 estimated quarters from the base quarter. Hence, the only result that consistently holds is seasonality in the whole group of BCP-ALL cases. In this group, the informative harmonic function was found in all quarter types and in the first and second specified quarters with an increase containing August and September.

Table 8. Chi-square tests at the date of diagnosis

		BCP-ALL			
Jan-Mar	386	Feb-Apr	398	Mar-May	406
Apr-Jun	406	May-Jul	391	Jun-Aug	427
Jul-Sep	433	Aug-Oct	446	Sep-Nov	412
Oct-Dec	376	Nov-Jan	366	Dec-Feb	356
p-value	0.1919		0.0387		0.0686
Adj p-value	0.3454		0.3483		0.3087
		HeH			
Jan-Mar	102	Feb-Apr	112	Mar-May	100
Apr-Jun	120	May-Jul	108	Jun-Aug	132
Jul-Sep	124	Aug-Oct	119	Sep-Nov	111
Oct-Dec	102	Nov-Jan	109	Dec-Feb	105
p-value	0.3027		0.8824		0.1509
Adj p-value	0.4541		0.8824		0.4527
		ETV6/RUNX1			
Jan-Mar	73	Feb-Apr	69	Mar-May	66
Apr-Jun	57	May-Jul	54	Jun-Aug	65
Jul-Sep	67	Aug-Oct	80	Sep-Nov	76
Oct-Dec	75	Nov-Jan	69	Dec-Feb	65
p-value	0.4101		0.1697		0.7375
Adj p-value	0.5273		0.3818		0.8297

Tables 8 and 9 report the results of exploratory Chi-square tests of BCP-ALL (including subtypes) seasonal variation. The lowest p-value was obtained for seasonality of BCP-ALL diagnosis using the second type of quarters, which showed a peak in August - October. No significant p-values were found for seasonality in diagnosis of BCP-ALL subtypes (HeH and *ETV6/RUNX1*). Neither did we find any significant p-values for seasonality of date of birth in BCP-ALL cases, the lowest p-value being 0.07 (BCP-ALL third quarter type with peak in March-May).

As all the tests were performed on the same BCP-ALL data, HeH and *ETV6/RUNX1* being the subtypes of BCP-ALL, the p-values of Chi-square test were adjusted using the Benjamini-Hochberg procedure. After this adjustment of p-values, there are no significant results of Chi-square test to be reported.

Descriptive data displayed a similar distribution of age at diagnosis comparing the entire cohort and cases diagnosed each respective month Table 10, including the peak incidence months August and September. Also the distribution of different subtypes of BCP-ALL was similar when comparing the peak months August and September to all other 10 months respectively and to the whole year (entire cohort). Thus, cases diagnosed during peak months August and September did not stand out in any apparent way neither regarding age of diagnosis nor subtype.

Discussion

Exposure to infectious disease is today a well-established suspect in the search for environmental involved in both initiation and progression of BCP-ALL. We know that pre-leukemic clones of many BCP-ALL subtypes are initiated during fetal life in a

medRxiv preprint doi: https://doi.org/10.1101/2023.02.12.23285595; this version posted February 12, 2023. 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 .

Table 9. Chi-square tests at the date of birth

		BCP-ALL			
Jan-Mar	421	Feb-Apr	416	Mar-May	446
Apr-Jun	420	May-Jul	405	Jun-Aug	385
Jul-Sep	385	Aug-Oct	398	Sep-Nov	390
Oct-Dec	375	Nov-Dec	382	Dec-Feb	380
p-value	0.2382		0.6774		0.0689
Adj p-value	0.7146		0.8709		0.6201
		HeH			
Jan-Mar	123	Feb-Apr	112	Mar-May	116
Apr-Jun	103	May-Jul	111	Jun-Aug	111
Jul-Sep	117	Aug-Oct	119	Sep-Nov	110
Oct-Dec	105	Nov-Dec	106	Dec-Feb	111
p-value	0.4818		0.8571		0.9782
Adj p-value	0.8672		0.9642		0.9782
		ETV6/RUNX1			
Jan-Mar	64	Feb-Apr	62	Mar-May	76
Apr-Jun	79	May-Jul	84	Jun-Aug	68
Jul-Sep	62	Aug-Oct	58	Sep-Nov	62
Oct-Dec	67	Nov-Dec	68	Dec-Feb	66
p-value	0.4648		0.1236		0.6755
Adj p-value	1		0.5562		1

Table 10. Number of cases and proportions of children diagnosed during specified months and periods

	Aug	Sep	Aug-Sep	Oct-Jul	All months
ALL	162 (0.1012)	144 (0.0899)	306 (0.1911)	1295 (0.8089)	1601
HeH	48 (0.1071)	40 (0.0893)	88 (0.1964)	360 (0.8036)	448
ETV6/RUNX1	27 (0.0993)	22 (0.0809)	49 (0.1802)	223 (0.8198)	272
other sub-types	87 (0.0987)	82 (0.0931)	169 (0.1918)	712 (0.8082)	881

substantial portion of cases [91–94], and that the chromosomal rearrangement representing each subtype is considered the initiating genetic event or "first hit". Epidemiological data has pointed to some specific viral infections increasing the risk of BCP-ALL following maternal infection during fetal life, although not all studies support these findings [95–97]. Thus, the cause of initiating genetic events stands unresolved. We performed an exploratory analysis applying Chi-square test to data of quarterly aggregated at time of birth of BCP-ALL cases, but did not obtain any significant difference in incidences between months. Four previous studies have indeed reported seasonality in time of birth for ALL cases (B- and T-cell ALL analyzed together), two of which specifically in the 1-6 year age group, with peaks ranging from jan-april [47, 48, 52]. Yet, other reports have not detected date of birth seasonality [99]. Further studies aiming at associating temporal waves in time of birth of BCP-ALL cases to preceding dittos of infectious disease hold the potential to guide molecular studies of leukemia initiating infections.

For preleukemic clones to progress into overt leukemia additional "hits" are required, as first suggested in the "two-hit hypothesis" by Greaves [32]. That a fraction of healthy

medRxiv preprint doi: https://doi.org/10.1101/2023.02.12.23285595; this version posted February 12, 2023. 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 .

neonates harbor small populations of pre-leukemic clones at birth without ever developing BCP-ALL emphasizes the importance of such subsequent events [19], but also offers hope for development of preventive measures [33]. Epidemiological studies have identified environmental factors such as early (<1 years of age) day care attendance [100–102], birth mode [103–106], and early contact with livestock and pets [107], all proxies for exposure to a variety of microbes, to be modulating the risk of BCP-ALL progression. These findings are in agreement with the theory that delayed exposure to infections results in an aberrant immunological response fueling progression of preleukemic cells if present. This delayed exposure to infections and minimized contact with microbes is a clear consequence of lifestyle in affluent societies, which indeed also experience higher incidence rates of BCP-ALL [10].

While early (within first year of life) exposure to microbes appears to protect against BCP-ALL development, infectious disease has also been suggested to promote BCP-ALL progression possibly through inflicting the second hit or hits leading to overt diesease. In 1917, Ward was the first to acknowledge the peak of BCP-ALL incidence in 1-6 year olds, suggesting infections could be a trigger of disease given that common infections are more abundant in this age group than others [108]. This was followed by Kinlen's idea of BCP-ALL being a rare response to a common infectious agent which becomes apparent by an increase in incidence upon in-mixing of migrants to an isolated immunologically naïve population [29–31]. Biological examples suggested to support this notion is the space-time clustering of increased BCP-ALL incidence with infectious epidemies [109–113]. In addition, a study reporting rapidly decreased incidence of childhood BCP-ALL soon after the implementation of infection control measures in response to the 2003 SARS outbreak in Hong Kong is also thought to lend supports of viral etiology. A two month long closing of schools was followed by rigorous hygiene routines after re-opening and other social distancing measure lasting for a total of 6 months. Substantial decline in communicable common infectious disease following these measures was proposed to protect at risk individuals, i.e. carriers of pre-leukemic clones, from acquiring 2nd hit/-s needed for progression [114]. This is further supported by two studies showing that exposure to infectious agents is required for developing BCP-ALL in mice housing pre-leukemic ETV6/RUNX1 positive pre-leukemic clones or susceptibility conveyed by Pax5 heterozygozity [27, 28].

Further, effects of severe acute respiratory syndrome coronavirus 2 (SARSCoV-2) pandemic on BCP-ALL incidence have been debated, speculating that both increasing incidence in response to a novel widespread pathogen (population-mixing hypothesis) and declining incidence as a consequence of decreased exposure to infectious disease (second-hit hypothesis) to be possible short term (weeks to months) outcomes [115, 116]. To date, in line with findings in Hong Kong, a report from Norway found distinctly decreased incidence numbers during the first months of the SARSCoV-2 pandemic [117] and similar indications have been reported from a region in Italy [118]. In contrast, recent data based on the German Childhood Cancer Registry indicate a significant increase in age standardized incidence rates (ASIR) of childhood (0-14y) lymphoid leukemia was seen during the Covid-19 pandemic in 2020 and 2021 [119, 120]. Like in Hong Kong [114], extensive infection control measures were instated in Germany during this time. Although the reason for this increased incidence is unknown it may lend further support to infectious exposures importance to BCP-ALL progression. It is however crucial to point out that increased ASIR was reported also for other childhood cancer types who's etiology is not considered susceptible to infectious exposure, which implies there may be other explanations to the reported ASIR of childhood lymphoid leukemia. Meanwhile, two reports from USA and Canada did not detect any change in incidence of pediatric leukemia and lymphoma as a group [121, 122]. Due to limitations in sample size and observation time, the currently available reports on this topic should

medRxiv preprint doi: https://doi.org/10.1101/2023.02.12.23285595; this version posted February 12, 2023. 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 .

be interpreted with caution. To draw any conclusions about the effects of SARSCoV-2 infection and the restrictions to prevent it's spread, BCP-ALL incidence will need to be closely monitored, and will undoubtedly be so henceforth. Not least will observing long-term effects (years) of decreased infectious exposure on future incidence rates allow for scrutinizing of the delayed infectious exposure hypothesis [123, 124].

In light of the above, seasonal variation of ALL (most studies do not discriminate between B- and T-cell origin) as a proxy for infectious exposure has been extensively studied over the last decades, applying an array of different methods as discussed in more detail below. We can conclude from our review of previous studies (Table 1) that there is no consistent proof of a seasonal wave in time of ALL diagnosis. And, in cases where seasonality is detected, time of peak incidence is scattered throughout the year. However, when studying seasonal variation, it is important to consider that patterns of seasonality may differ from country to country depending on factors such as climate zones, affluence, way of life et.c. affecting infectious panorama and seasonality of communicable infectious disease.

In the current study we report an informative seasonality at date of diagnosis specifically for BCP-ALL, with peak incidence in August-September. Descriptive data on distribution of age at diagnosis and genetic subtype did not differ for cases diagnosed in August-September compared to other months nor the entire cohort. In Sweden almost all children attend pre-school from age 1 and a long summer holiday is customary, usually beginning in late June or early July and extending into early August. One possible explanation for the timing of our observed incidence peak is that decreased spreading of infectious disease during summer, a consequence of the prolonged summer holiday in Sweden, causes a temporary halt in the final steps of disease progression for some individuals that would otherwise have presented clinically. This would be in accordance with observations in Hong Kong during SARS-epidemic [114]. However, in contrast to the Hong Kong example cases are then instead accumulated in August-September when at risk-individuals are again exposed to infections in schools and pre-schools. This rationale builds on the assumption that an infection quite rapidly pushes disease progression the last step to giving non-ignorable clinical symptoms of BCP-ALL. It is known that symptoms of ALL do indeed evolve in only days to weeks but the length of latency from second genetic hit to clinical presentation remains unknown and may very well be variable. Thus, if such a "last" infection before diagnosis causes the second genetic hit or just puts selective pressure on already "ready-to-go" leukemic cells remains to be understood.

An alternative explanation for our observed August-September BCP-ALL incidence peak would be that some specific common pathogens have a slightly sharper ability than others to cause genetic second hits, and thus progression to overt disease, in pre-leukemic cells. Entero and varicella viruses for example have seasonal peaks during summer [125]. Again, the challenge of associating a peak in BCP-ALL incidence to a peak in spread of certain pathogens is the undetermined and likely variable latency from second genetic hit to clinical presentation.

Analysis of seasonal variation in series of aggregated counts may face several challenges. As for epidemiological studies in general, small sample size is a common problem resulting in low statistical power, which increases probability of reporting false negative results as statistical tests performed on small samples are only capable of detecting large effects [126, 127]. In our literature review, merely 9 out of 42 previous publications on seasonality in ALL (as summarized in Table 1) had more than 1000 ALL cases in their studied cohort.

The largest cohort studied to date is that of 15 835 cases of childhood leukmeia (73 % lymphatic) born and diagnosed between 1953-1995 in UK, published by Higgins, C. D. et al in 2001. [51] For the 1282 cases born and diagnosed before 1962, a suggestive

medRxiv preprint doi: https://doi.org/10.1101/2023.02.12.23285595; this version posted February 12, 2023. 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 .

but after not statistically significant incidence peak in August-September was identified. The authors however call for caution when interpreting this indication; since cases form this time period were extracted from death-records, retrieving date of diagnosis retrospectively, a "complication to death" could have introduced an apparent seasonality. Further, no significant seasonality in childhood leukemia incidence was found when analysing the entire cohort. A possible explanation was suggested to be the fact that seasonality was examined by date of diagnosis rather than clinical onset, between which there may be a discrepancy in time masking seasonality. Based on this possible discrepancy, one could argue for increasing the aggregation period from month to quarter as was done in the present study. Quarterly transformation allows the date of onset and date of diagnosis to be in the same period of time series analysis. The main drawback of quarterly transformation is that it requires performing the analysis on three different subsets of quarters, which increases the probability of finding false positive results using classical statistical methods. We believe that our cohort of 1601 cases diagnosed during a time-span of 22 years (1995-2017) as well as formulation of the GARIMAX model in the Bayesian setup allows to provide the unbiased results in this paper.

Our literature review revealed three previously applied types of data transformation, an obligate step before performing analysis with any method. The first transformation aggregates tabular data from registries to counts by seasons or months, summing up the data for all observed years in the sample, resulting in four (seasons) or 12 (months) bin histograms. The second transformation counts number of cases per month or season in the sample, without summation between different years, resulting in time series of monthly or seasonal counts. The third transformation uses individual case-by-case statistical methods. To our knowledge, this third type of transformation was only previously used by Gao et al [49] and was also applied in the present study. Type of data transformation does to some extent depend on the chosen method for analysis, but different combinations may be applied. Therefore, choice of data transformation type is a variable that may affect output.

The most common method for detection of the seasonal variation is the Chi-square test, which was used in 13 previous studies included in our review of previous publications, out of which seven reported seasonal variation in ALL/childhood leukmeia [35, 39, 42, 44, 46, 48, 50, 63–65, 68, 69, 71]. The test is implemented to histogram data transformation and answers the question whether there is likely higher relative frequency in one group than the other, the other group usually being the mean of all data.

Exploratory analysis without statistical tests was used in 11 publications (stated as "single-factor analysis" in Table1) 1) [38, 47, 55, 57, 58, 60, 62, 66, 67, 70, 72] with seven papers reporting positive results for seasonal variation in ALL/childhood leukmeia. The most recent paper with descriptive statistics was published in 2011, marking a shift towards use of more strict hypothesis-driven methods.

Edward's test [129] was implemented in four reviewed publications [43, 51, 56, 61]. The model detects sinusoidal curve within a 12-month period histogram. J A Ross et al [53] implement Rodger's test [130], a modification of Edward's test, which evaluates the significance for cyclic trends based on the efficient score vector calculated for each seasonal peak of aggregated cases. Five studies that applied these similar tests reported significant seasonal variation of ALL/childhood leukemia diagnosis in East Anglia (UK) [56] and USA [53], but neither in Mexico [43], UK (all regions) [51], NW England [56] nor the Netherlands [61].

Cosinor Analysis [52, 54, 59] is performed by fitting the sinusoidal curves (harmonic functions) to 12-month histograms. Poisson regression with harmonic functions, an extension of cosinor analysis assuming not-normal distribution of errors, was applied in four previous studies [37, 40, 45, 48]. Moreover, Gao et al [49] investigated the seasonal

medRxiv preprint doi: https://doi.org/10.1101/2023.02.12.23285595; this version posted February 12, 2023. 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 .

variation of ALL in USA, Singapore, and a western region of Sweden, using von Mises distribution in combination with Mardia test statistics [131], reporting a seasonal peak in January in Sweden but only with 63 observations in the sample.

Joint point regression rendered the report of an increased number of ALL-diagnoses in spring and summer in Iran, studying cases diagnosed between 2006 and 2014. This method was applied to monthly histograms and allowed identification of changes in trends of studied data. [36].

Finally, Kyu Seok Shim et al [41] implemented ARIMA (Autoregressive Integrated moving average model) to analyse seasonal variation in the time series of ALL-diagnosis in South Korea, reporting positive results for the presence of seasonality with a peak in winter. The authors also showed that the seasonal wave of ALL-diagnosis correlated with that of HPIV (Human para-influenza viruses) with an assumed 1-2 month period of latency. To our knowledge, this is the only study so far correlating seasonal wave of childhood ALL to a specific viral agent, thus addressing the question of potential specific viral agents promoting progression to overt leukemia. However, the correlation to HPIV was not only made for ALL but also several other pediatric cancers which are not suspected to have a viral etiology.

All described methods above exhibit positive and negative sides. The methods that employ harmonic functions (combination of sine and cosine functions) assume perfect repeatability of the infectious agents from year to year. Since seasonal variation of infections might not be regular, this assumption may pose a problem when looking for seasonal variation in childhood ALL-diagnosis as as a proxy for infectious exposure contributing to disease progression. It is also important to keep in mind that the errors of the regressions with harmonic functions might not follow the normal distribution in count data scenarios. On the other hand, methods that are performed on aggregated histograms of months or seasons are harshly affected by outliers. Thus, a random increase in number of cases in a monthly bin without true periodicity may nevertheless show significant seasonal variation, rendering false conclusions.

There are several advantages of implementing the GARIMAX model to search for seasonal variation in BCP-ALL case count data. ARIMA is a classical model that is widely used in different fields of studies to detect seasonal variation. Generalization of the model via Negative Binomial distribution adapts the model to low case count data allowing accounting for excess variability as compared to the generalization via Poisson distribution in the limited information setting. Formulation in the bayesian setup makes the model less data hungry, and allows for inference in small sample settings. The main disadvantage of the GARIMAX model is that it requires transformation to case count data, thus blocking the opportunity to infer individual cases' contribution to seasonal variation. Also, the model does not allow for inference at date of birth as it requires stationary or semi-stationary time series. Aggregation by date of birth is impossible as the series exhibit informational delay, i.e. many children who were already born are not yet diagnosed in the data. One possible solution would be to cut time series by 18 years (childhood ALL is not diagnosed beyond the age of 18 years), but this approach would shift the analyzed period such that the sample becomes too small for the model even in the Bayesian setup. Chi-square test has previously been widely applied for analysis of seasonal variation in date of birth of childhood ALL patients, but the test requires p-value adjustment which is impacted by outliers as discussed above.

Conclusion

In the present study we identified seasonal variation of childhood BCP-ALL incidence numbers with a peak in August-September. Results were obtained applying modern statistical methods, Generalized Autoregressive Integrated Moving Average model with

medRxiv preprint doi: https://doi.org/10.1101/2023.02.12.23285595; this version posted February 12, 2023. 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 .

External seasonal covariates (GARIMAX), to a Swedish population-based cohort of 1601 childhood BCP-ALL cases from the Swedish childhood cancer registry. The cause of our findings is yet to be determined. Nonetheless, based on the current knowledge and hypotheses regarding childhood BCP-ALL etiology, we suggest two possible explanatory models. If a final unspecific viral infection pushes leukemic clones to progress and expand in the bone marrow and cause non-ignorable clinical symtoms, then the August-September incidence peak could be explained by a rebound effect following a halt in such progression during Swedish summer (attributable to deceased spread of viral infections due to closure of schools and pre-schools).Alternatively, peak incidence in August-September could be attributable to a viral agent with seasonal spread during summer potent in causing a 2nd-hit in pre-leukmeic clones or. Which model is more likely depends on how long latency from 2nd hit to overt leukmeia is, a factor yet unknown in vivo in humans. Further studies to understand these etiological factors are required. Also, further studies of larger cohorts using modern statistical tools are needed to examine the possible temporal pattern of BCP-ALL at both time of diagnosis and birth.

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 Henrik Passmark, Case Officer at The Swedish National Board of Health and Welfare and Jonas Janegren at Statistics Sweden for their great support to collect 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, decision to publish, or preparation of the manuscript.

Authorship

Contribution: A.N., B.B., C.S., G.B., N.E. and R.J. designed the study. G.B., performed analysis; G.B., R.J. and N.E. performed the GARIMAX analysis.; N.E., G.B., G.T. and A.N.S. prepared the dataset; A.N., M.H., J.A., E.P. and B.B. contributed with medical knowledge and generation of hypothesis. G.B. prepared the figures; A.N., B.B. and G.B. wrote the manuscript; All authors contributed to data interpretation; and all authors revised the manuscript and approved the final version.

Conflict-of-interest disclosure

The authors declare no competing financial interests.

medRxiv preprint doi: https://doi.org/10.1101/2023.02.12.23285595; this version posted February 12, 2023. 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 .

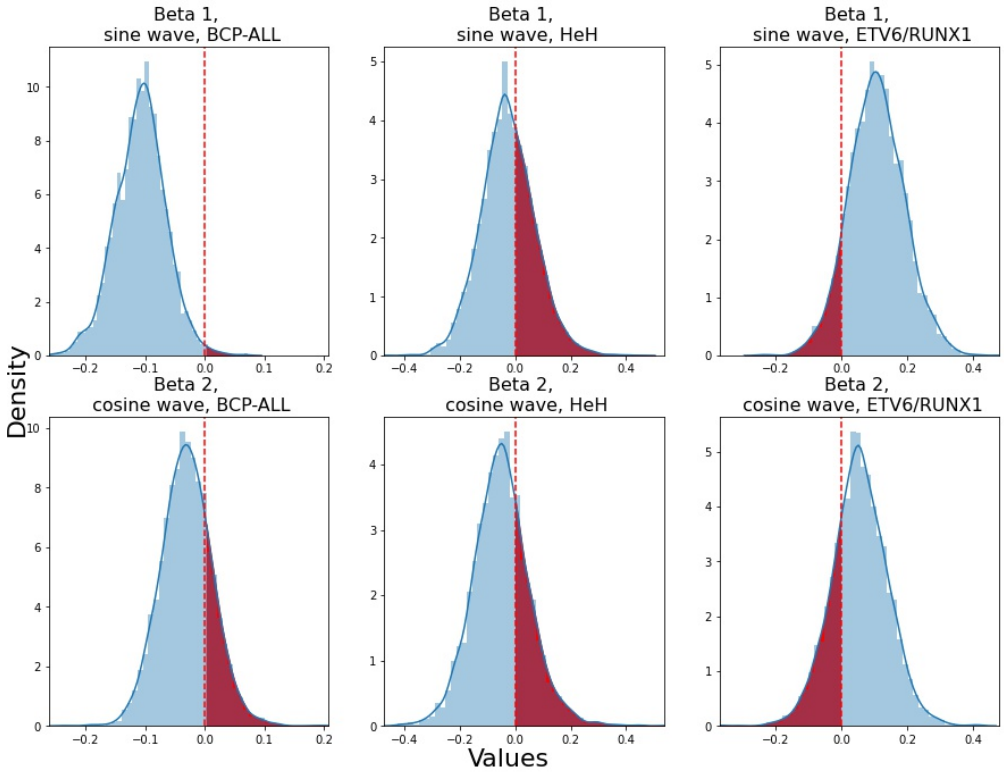
Supporting information

S1 Appendix. BIC scores of the tested models, GARIMAX

model	BIC (BCP-ALL, 1st type quarter)	BIC (BCP-ALL, 2nd type quarter)	BIC (BCP-ALL, 3rd type quarter)	BIC (HeH, 1st type quarter)	BIC (HeH, 2nd type quarter)	BIC (HeH, 3rd type quarter)	BIC (*ETV6/RUNX1*, 1st type quarter)	BIC (*ETV6/RUNX1*, 2nd type quarter)	BIC (*ETV6/RUNX1*, 3rd type quarter)
ARIMA(1, 1, 0)	536.50	531.53	549.50	428.18	421.06	420.63	368.89	332.03	364.38
ARIMA(2, 1, 0)	532.38	529.57	546.49	405.83	418.37	417.19	**354.84**	321.83	**343.50**
ARIMA(3, 1, 0)	**526.02**	**517.08**	538.76	405.57	405.08	**410.39**	356.29	**320.42**	346.14
ARIMA(4, 1, 0)	526.34	521.73	**535.54**	**403.86**	**405.01**	411.03	356.60	324.28	344.56
ARIMA(1, 1, 1)	541.49	536.55	554.51	433.17	426.02	425.75	373.79	336.66	369.25
ARIMA(1, 1, 2)	546.16	541.56	556.74	425.91	427.56	423.54	358.53	329.55	362.57
ARIMA(1, 1, 3)	548.19	541.70	551.69	422.69	431.26	424.99	364.81	333.18	358.54
ARIMA(1, 1, 4)	543.40	531.30	543.18	415.95	420.50	419.07	364.82	335.27	362.48
ARIMA(2, 1, 1)	537.23	534.24	551.31	410.50	422.83	421.80	357.72	325.60	346.75
ARIMA(2, 1, 2)	541.95	538.99	555.21	414.42	426.74	423.00	357.06	328.88	348.63
ARIMA(2, 1, 3)	546.09	543.84	556.42	415.80	419.59	427.08	359.55	329.33	352.33
ARIMA(2, 1, 4)	530.28	528.66	547.67	413.12	412.82	421.63	362.01	331.42	355.33
ARIMA(3, 1, 1)	530.86	521.79	543.33	409.99	409.42	414.76	357.84	322.92	348.26
ARIMA(3, 1, 2)	535.95	526.73	547.68	410.96	414.36	419.11	356.05	326.12	348.52
ARIMA(3, 1, 3)	538.11	531.66	547.70	414.79	413.72	421.85	360.64	327.57	351.21
ARIMA(3, 1, 4)	535.02	532.17	551.20	417.83	416.58	425.72	363.41	331.81	351.90
ARIMA(4, 1, 1)	531.05	526.54	540.27	408.08	409.16	415.11	356.12	326.30	345.40
ARIMA(4, 1, 2)	536.19	531.67	544.77	412.81	414.22	418.75	358.30	329.44	348.42
ARIMA(4, 1, 3)	538.63	536.71	546.90	416.42	416.33	420.47	359.65	330.17	350.05
ARIMA(4, 1, 4)	536.79	536.76	548.43	419.37	418.14	424.98	361.90	333.16	351.84

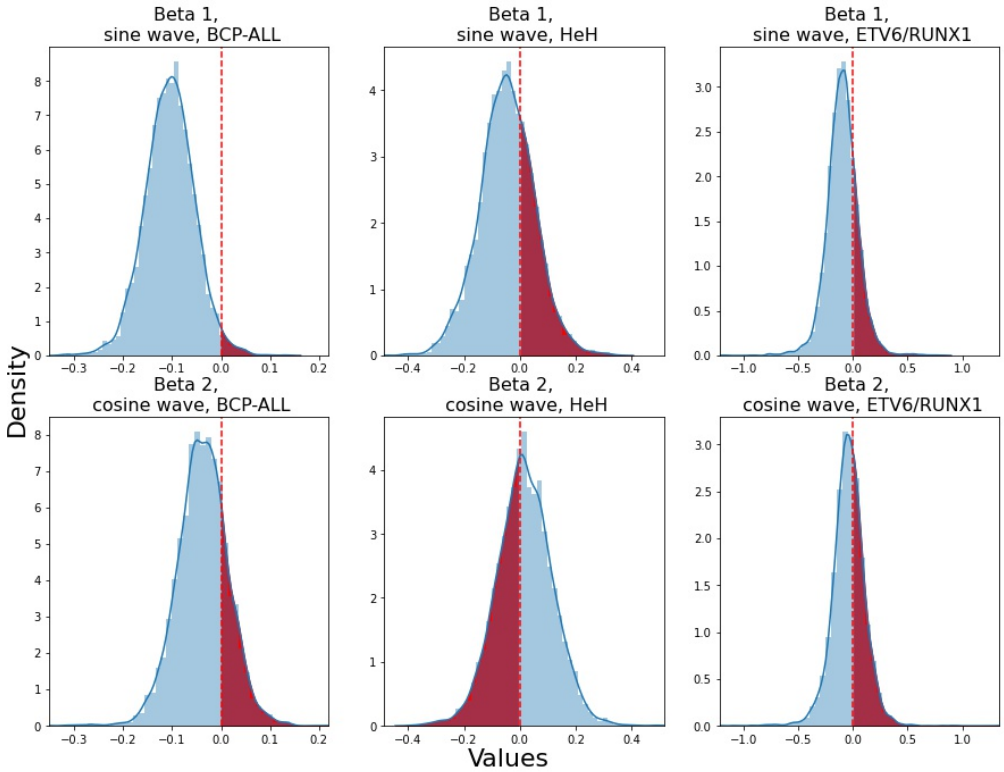
medRxiv preprint doi: https://doi.org/10.1101/2023.02.12.23285595; this version posted February 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

S1 Fig. The posterior distributions of the coefficients of the harmonic functions, date of diagnosis, GARIMAX, first type quarter



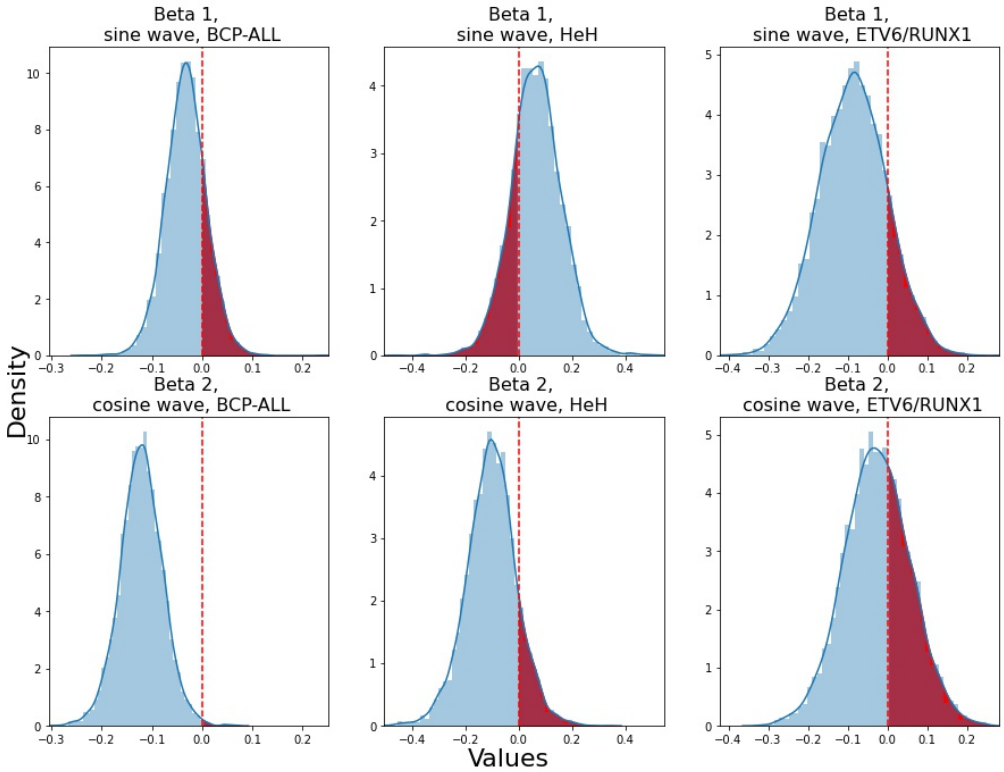
medRxiv preprint doi: https://doi.org/10.1101/2023.02.12.23285595; this version posted February 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

S2 Fig. The posterior distributions of the coefficients of the harmonic functions, date of diagnosis, GARIMAX, second type quarter



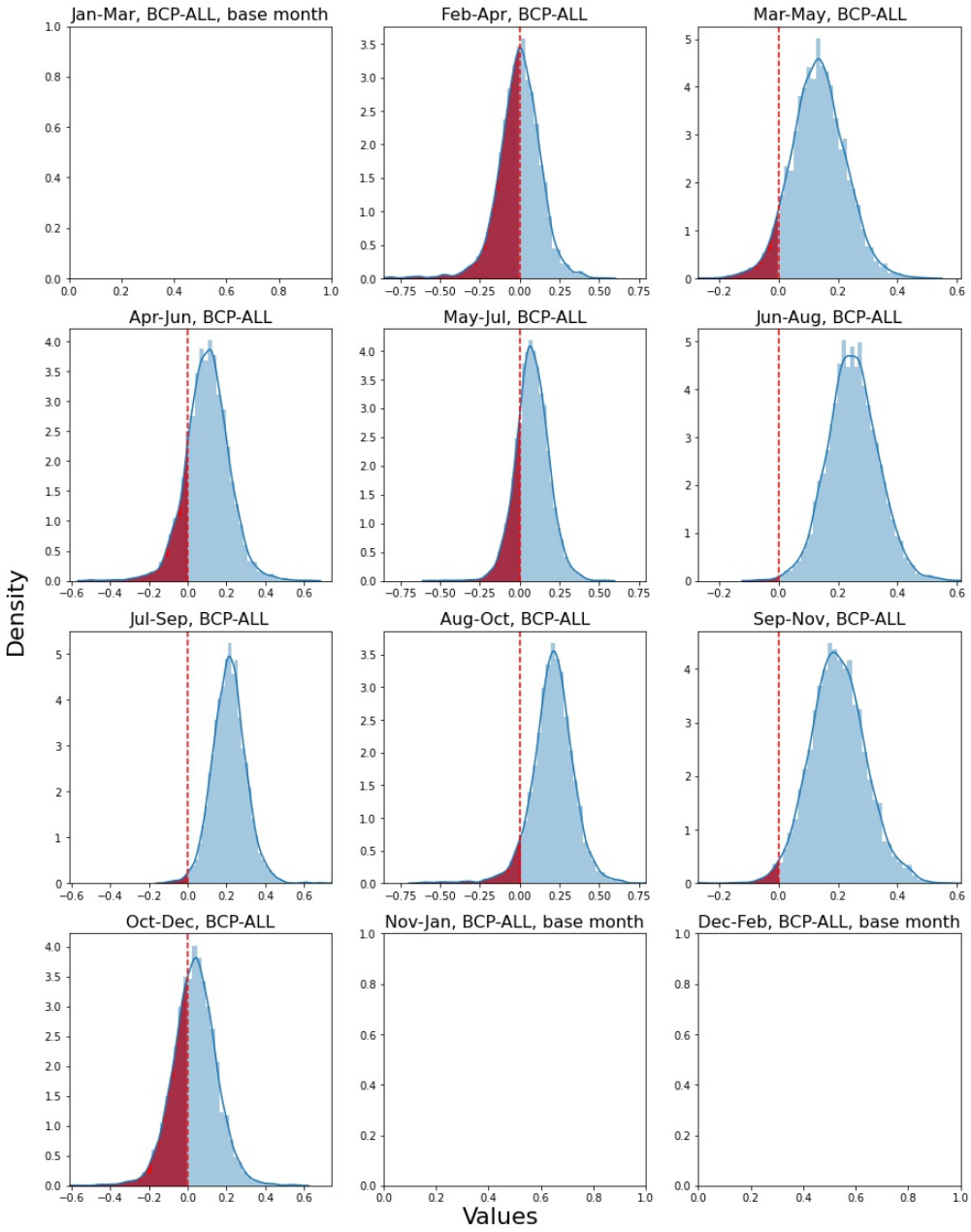
medRxiv preprint doi: https://doi.org/10.1101/2023.02.12.23285595; this version posted February 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

S3 Fig. The posterior distributions of the coefficients of the harmonic functions, date of diagnosis, GARIMAX, third type quarter



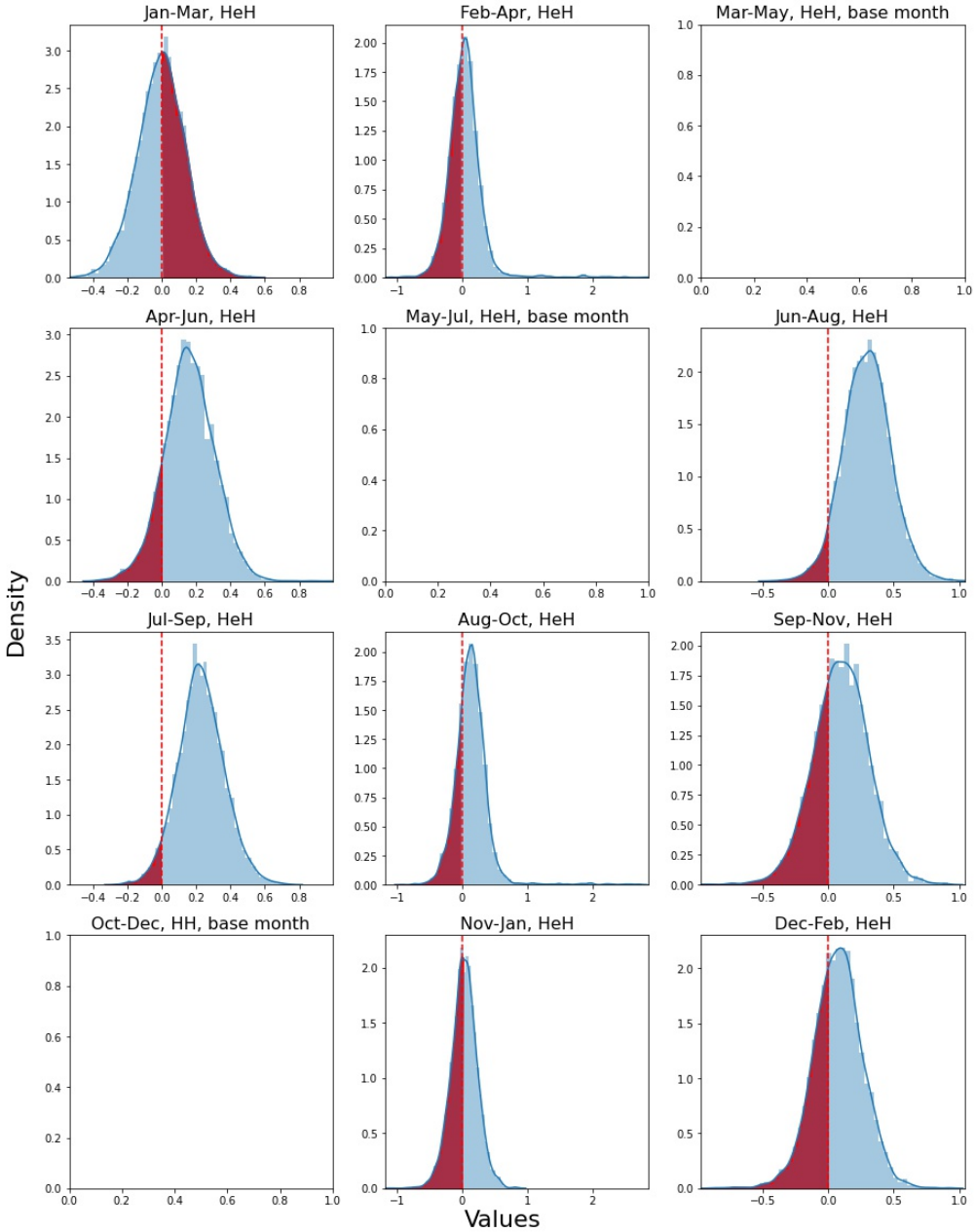
medRxiv preprint doi: https://doi.org/10.1101/2023.02.12.23285595; this version posted February 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

S4 Fig. The posterior distributions of the coefficients of the seasonal matrix, BCP-ALL



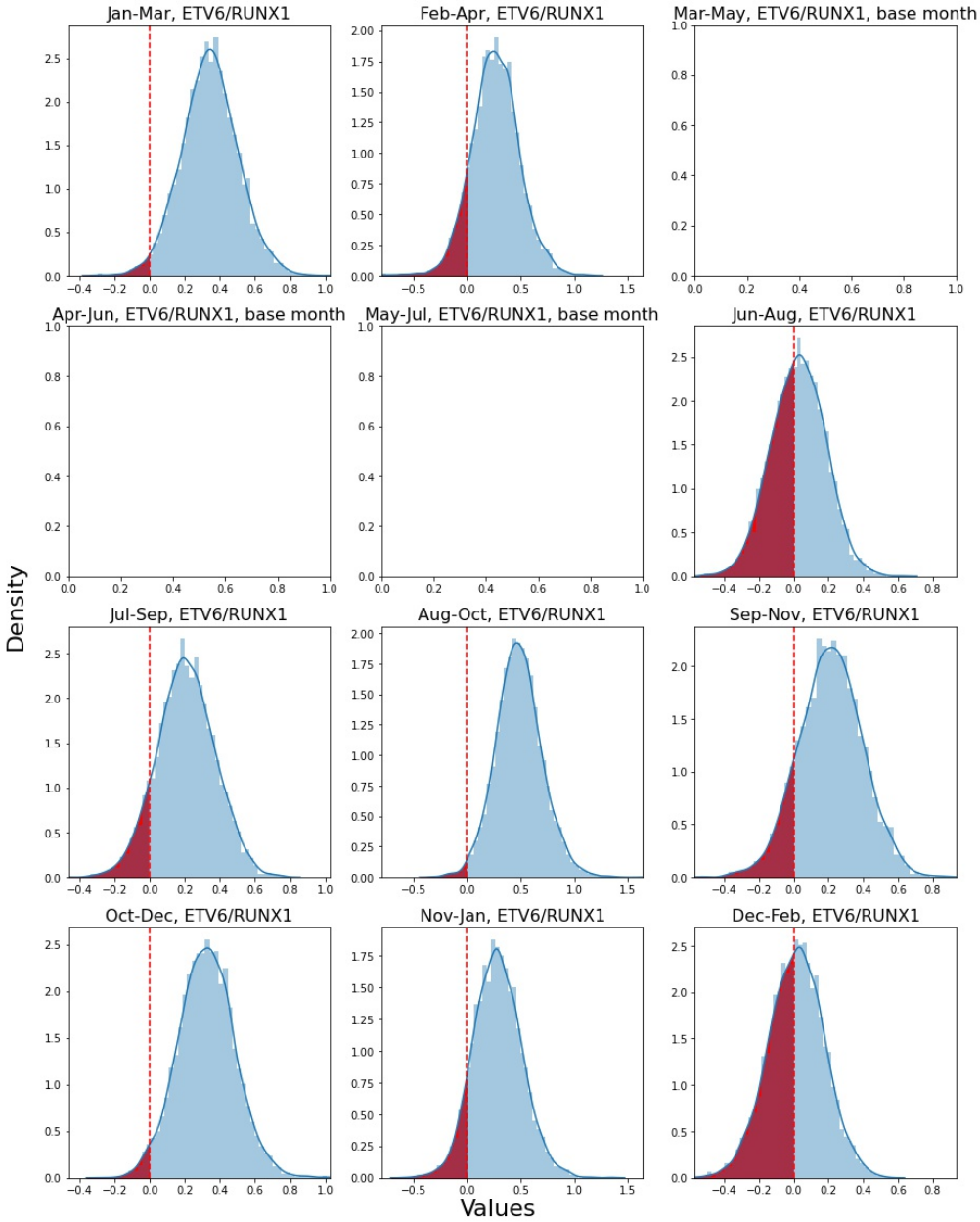
medRxiv preprint doi: https://doi.org/10.1101/2023.02.12.23285595; this version posted February 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

S5 Fig. The posterior distributions of the coefficients of the seasonal matrix, HeH



medRxiv preprint doi: https://doi.org/10.1101/2023.02.12.23285595; this version posted February 12, 2023. 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 .

S6 Fig. The posterior distributions of the coefficients of the seasonal matrix, *ETV6/RUNX1*



medRxiv preprint doi: https://doi.org/10.1101/2023.02.12.23285595; this version posted February 12, 2023. 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 .

References

1. Lähteenmäki P. Årsrapport 2019 Svenska Barncancerregistret.
2. Bhojwani D, Yang JJ, Pui CH. Biology of childhood acute lymphoblastic leukemia. Pediatric Clinics. 2015 Feb 1;62(1):47-60.
3. Pfister SM, Reyes-Múgica M, Chan JKC, Hasle H, Lazar AJ, Rossi S, Ferrari A, Jarzembowski JA, Pritchard-Jones K, Hill DA, Jacques TS, Wesseling P, López Terrada DH, von Deimling A, Kratz CP, Cree IA, Alaggio R A Summary of the Inaugural WHO Classification of Pediatric Tumors: Transitioning from the Optical into the Molecular Era. Cancer Discov. 2022 Feb;12(2):331-355. doi: 10.1158/2159-8290.CD-21-1094. Epub 2021 Dec 17. PMID: 34921008.
4. Moorman AV. The clinical relevance of chromosomal and genomic abnormalities in B-cell precursor acute lymphoblastic leukaemia. Blood Rev. 2012 May;26(3):123-35.
5. Gustafsson, G., Kogner, P. & Heyman, M. Childhood Cancer Incidence and Survival in Sweden 1984-2010 - Report 2013. Karolinska Institutet, Stockholm, Sweden, 2014.
6. Marcotte EL, Spector LG, Mendes-de-Almeida DP, Nelson HH. The Prenatal Origin of Childhood Leukemia: Potential Applications for Epidemiology and Newborn Screening. Front Pediatr. 2021 Apr 23;9:639479. doi: 10.3389/fped.2021.639479. PMID: 33968846; PMCID: PMC8102903.
7. Greaves MF, Colman SM, Beard ME, Bradstock K, Cabrera ME, Chen PM, Jacobs P, Lam-Po-Tang PR, MacDougall LG, Williams CK. Geographical distribution of acute lymphoblastic leukaemia subtypes: second report of the collaborative group study. Leukemia. 1993 Jan 1;7(1):27-34.
8. Inaba H, Pui CH. Advances in the Diagnosis and Treatment of Pediatric Acute Lymphoblastic Leukemia. J Clin Med. 2021 Apr 29;10(9):1926. doi: 10.3390/jcm10091926. PMID: 33946897; PMCID: PMC8124693.
9. Iacobucci I, Mullighan CG. Genetic Basis of Acute Lymphoblastic Leukemia. J Clin Oncol. 2017 Mar 20;35(9):975-983.
10. Steliarova-Foucher E, Colombet M, Ries LA, Moreno F, Dolya A, Bray F, Hesseling P, Shin HY, Stiller CA, Bouzbid S, Hamdi-Cherif M. International incidence of childhood cancer, 2001–10: a population-based registry study. The Lancet Oncology. 2017 Jun 1;18(6):719-31.
11. Greaves MF, Alexander FE. An infectious etiology for common acute lymphoblastic leukemia in childhood?. Leukemia. 1993 Mar 1;7(3):349-60.
12. Ford AM, Bennett CA, Price CM, Bruin MC, Van Wering ER, Greaves M. Fetal origins of the TEL-AML1 fusion gene in identical twins with leukemia. Proceedings of the National Academy of Sciences. 1998 Apr 14;95(8):4584-8.
13. Wiemels JL, Cazzaniga G, Daniotti M, Eden OB, Addison GM, Masera G, Saha V, Biondi A, Greaves MF. Prenatal origin of acute lymphoblastic leukaemia in children. The Lancet. 1999 Oct 30;354(9189):1499-503.

medRxiv preprint doi: https://doi.org/10.1101/2023.02.12.23285595; this version posted February 12, 2023. 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 .

14. Mori H, Colman SM, Xiao Z, Ford AM, Healy LE, Donaldson C, Hows JM, Navarrete C, Greaves M. Chromosome translocations and covert leukemic clones are generated during normal fetal development. Proceedings of the National Academy of Sciences. 2002 Jun 11;99(12):8242-7.

15. Taub JW, Konrad MA, Ge Y, Naber JM, Scott JS, Matherly LH, Ravindranath Y. High frequency of leukemic clones in newborn screening blood samples of children with B-precursor acute lymphoblastic leukemia. Blood, The Journal of the American Society of Hematology. 2002 Apr 15;99(8):2992-6.

16. Maia AT, Tussiwand R, Cazzaniga G, Rebulla P, Colman S, Biondi A, Greaves M. Identification of preleukemic precursors of hyperdiploid acute lymphoblastic leukemia in cord blood. Genes, Chromosomes and Cancer. 2004 May;40(1):38-43.

17. Bateman CM, Alpar D, Ford AM, Colman SM, Wren D, Morgan M, Kearney L, Greaves M. Evolutionary trajectories of hyperdiploid ALL in monozygotic twins. Leukemia. 2015 Jan;29(1):58-65.

18. Hein D, Borkhardt A, Fischer U. Insights into the prenatal origin of childhood acute lymphoblastic leukemia. Cancer and Metastasis Reviews. 2020 Mar;39(1):161-71.

19. 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.

20. Škorvaga M, Nikitina E, Kubeš M, Košík P, Gajdošechová B, Leitnerová M, Copáková L, Belyaev I. Incidence of common preleukemic gene fusions in umbilical cord blood in Slovak population. PLoS One. 2014 Mar 12;9(3):e91116.

21. Mullighan CG, Goorha S, Radtke I, Miller CB, Coustan-Smith E, Dalton JD, Girtman K, Mathew S, Ma J, Pounds SB, Su X. Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. Nature. 2007 Apr;446(7137):758-64.

22. Greaves M. A causal mechanism for childhood acute lymphoblastic leukaemia. Nature Reviews Cancer. 2018 Aug;18(8):471-84.

23. 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.

24. Greaves, M. Aetiology of acute leukaemia. The Lancet 349, 344–349 (1997).

25. Little MP, Tawn EJ, Tzoulaki I, Wakeford R, Hildebrandt G, Paris F, Tapio S, Elliott P. A systematic review of epidemiological associations between low and moderate doses of ionizing radiation and late cardiovascular effects, and their possible mechanisms. Radiation research. 2008 Jan;169(1):99-109.

26. He JR, Ramakrishnan R, Hirst JE, Bonaventure A, Francis SS, Paltiel O, Håberg SE, Lemeshow S, Olsen S, Tikellis G, Magnus P. Maternal infection in pregnancy and childhood leukemia: a systematic review and meta-analysis. The Journal of pediatrics. 2020 Feb 1;217:98-109.

medRxiv preprint doi: https://doi.org/10.1101/2023.02.12.23285595; this version posted February 12, 2023. 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 .

27. Martín-Lorenzo A, Hauer J, Vicente-Dueñas C, Auer F, González-Herrero I, García-Ramírez I, Ginzel S, Thiele R, Constantinescu SN, Bartenhagen C, Dugas M. Infection Exposure Is a Causal Factor in B-cell Precursor Acute Lymphoblastic Leukemia as a Result of Pax5-Inherited SusceptibilityCausal Role of Infection in pB-ALL Development. Cancer discovery. 2015 Dec 1;5(12):1328-43.

28. Rodríguez-Hernández G, Hauer J, Martín-Lorenzo A, Schäfer D, Bartenhagen C, García-Ramírez I, Auer F, Gonzalez-Herrero I, Ruiz-Roca L, Gombert M, Okpanyi V. Infection Exposure Promotes ETV6-RUNX1 Precursor B-cell Leukemia via Impaired H3K4 DemethylasesKDM Family Drives ETV6-RUNX1 pB-ALL Progression. Cancer Research. 2017 Aug 15;77(16):4365-77.

29. Kinlen L. Evidence for an infective cause of childhood leukaemia: comparison of a Scottish new town with nuclear reprocessing sites in Britain. Lancet. 1988;2(8624):1323-1327. doi:10.1016/s0140-6736(88)90867-7

30. Kinlen LJ, Clarke K, Hudson C. Evidence from population mixing in British New Towns 1946-85 of an infective basis for childhood leukaemia. The Lancet. 1990 Sep 8;336(8715):577-82.

31. Kinlen LJ, Balkwill A. Infective cause of childhood leukaemia and wartime population mixing in Orkney and Shetland, UK. The Lancet. 2001 Mar 17;357(9259):858.

32. Greaves MF. Speculations on the cause of childhood acute lymphoblastic leukemia. Leukemia. 1988;2(2):120-5.

33. Hauer J, Fischer U, Borkhardt A. Toward prevention of childhood ALL by early-life immune training. Blood. 2021 Oct 21;138(16):1412-28.

34. Lambin P, Gerard MJ. Variations de fréquence saisonnières de la leucémie aiguë. Sang 1934;8:730–2

35. 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. Int J Lab Hematol. 2021 Feb;43(1):93-98.

36. Rahimi Pordanjani S, 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. Arch Iran Med. 2021 Mar 1;24(3):224-232.

37. 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(7):693-704.

38. Rahimi Pordanjani S, 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. Epidemiol Health. 2020;42:e2020057.

39. Bagirov IA. The seasonal dynamics of morbidity of acute lymphoblastic leukemia in Azerbaijan. Probl Sotsialnoi Gig Zdravookhranenniiai Istor Med. 2019;27(5):911-914.

40. Nurullah R, Kuhle S, Maguire B, Kulkarni K. Seasonality in Pediatric Cancer. Indian J Pediatr. 2018 Sep;85(9):785-787. doi: 10.1007/s12098-017-2561-4. Epub 2017 Dec 14. PMID: 29238940.

medRxiv preprint doi: https://doi.org/10.1101/2023.02.12.23285595; this version posted February 12, 2023. 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. 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.

42. 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. Oncol Lett. 2015 Jul;10(1):453-458.

43. 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 [Clinical features of acute leukemia and its relationship to the season of the year]. Rev Med Inst Mex Seguro Soc. 2014 Mar-Apr;52(2):176-81. Spanish

44. Kulkarni KP, Marwaha RK. Seasonality in diagnosis of childhood acute lymphoblastic leukemia: impact on disease presentation, survival outcome and resources. J Pediatr Hematol Oncol. 2013 Jan;35(1):81-2.

45. 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 Epidemiol. 2013 Jun;37(3):255-61.

46. Mutlu M, Erduran E. The relationship between seasonal variation in the diagnosis of acute lymphoblastic leukemia and its prognosis in children. Turk J Haematol. 2012 Jun;29(2):188-90.

47. 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;19(3):692-5. Chinese.

48. Basta NO, James PW, Craft AW, McNally RJ. Season of birth and diagnosis for childhood cancer in Northern England, 1968-2005. Paediatr Perinat Epidemiol. 2010 May;24(3):309-18.

49. 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. Am J Epidemiol. 2005 Oct 15;162(8):753-63.

50. Karimi M, Yarmohammadi H. Seasonal variations in the onset of childhood leukemia/lymphoma: April 1996 to March 2000, Shiraz, Iran. Hematol Oncol. 2003 Jun;21(2):51-5

51. Higgins, C. D., et al. 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, 84.3: 406-412

52. Sørensen HT, Pedersen L, Olsen J, Rothman K. Seasonal variation in month of birth and diagnosis of early childhood acute lymphoblastic leukemia. JAMA. 2001 Jan 10;285(2):168-9.

53. Ross JA, Severson RK, Swensen AR, Pollock BH, Gurney JG, Robison LL. Seasonal variations in the diagnosis of childhood cancer in the United States. Br J Cancer. 1999 Oct;81(3):549-53.

54. Douglas S, Cortina-Borja M, Cartwright R. A quest for seasonality in presentation of leukaemia and non-Hodgkin's lymphoma. Leuk Lymphoma. 1999 Feb;32(5-6):523-32.

medRxiv preprint doi: https://doi.org/10.1101/2023.02.12.23285595; this version posted February 12, 2023. 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 .

55. Gilman EA, Sorahan T, Lancashire RJ, Lawrence GM, Cheng KK. Seasonality in the presentation of acute lymphoid leukaemia. Br J Cancer. 1998 Feb;77(4):677-8.

56. 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. Br J Cancer. 1998 Jul;78(1):119-24.

57. Thorne R, Hunt LP, Mott MG. Seasonality in the diagnosis of childhood acute lymphoblastic leukaemia. Br J Cancer. 1998 Feb;77(4):678.

58. Badrinath P, Day NE, Stockton D. Seasonality in the diagnosis of acute lymphocytic leukaemia. Br J Cancer. 1997;75(11):1711-3.

59. Meltzer AA, Annegers JF, Spitz MR. Month-of-birth and incidence of acute lymphoblastic leukemia in children. Leuk Lymphoma. 1996 Sep;23(1-2):85-92.

60. Cohen P. The influence on survival of season of onset of childhood acute lymphoblastic leukemia (ALL). Chronobiol Int. 1987;4(2):291-7.

61. van Steensel-Moll HA, Valkenburg HA, Vandenbroucke JP, van Zanen GE. Time space distribution of childhood leukaemia in the Netherlands. J Epidemiol Community Health. 1983 Jun;37(2):145-8

62. Walker AM, van Noord PA. No seasonality in the diagnosis of acute leukemia in the United States. Journal of the National Cancer Institute, 1982, 69.6: 1283-1288.

63. Zannos-Mariolea L, Haidas S, Tzortzatou F, Dentaki-Svolaki K, Kiossoglou K. Epidémiologie de la leucémie aiguë chez l'enfant en Grèce [Epidemiology of acute leukemia of childhood in Greece (author's transl)]. Nouv Rev Fr Hematol. 1975 Nov-Dec;15(6):649-55. French.

64. Hems G, Stuart A. Childhood leukaemia in Scotland, 1939-68. Scott Med J. 1972 Jan;17(1):13-7.

65. Gunz FW, Spears GF. Distribution of acute leukaemia in time and space. Studies in New Zealand. Br Med J. 1968;4(5631):604-608

66. Till MM, Hardisty RM, Pike MC, Doll R. Childhood leukaemia in greater London: a search for evidence of clustering. Br Med J. 1967 Sep 23;3(5568):755-8.

67. Mainwaring D. Epidemiology of acute leukaemia of childhood in the Liverpool area. Br J Prev Soc Med. 1966 Oct;20(4):189-94.

68. Meighan SS, Knox G. Leukemia in childhood. Epidemiology in Oregon. Cancer. 1965 Jul;18(7):811-4.

69. Knox G. Epidemiology of childhood leukaemia in Northumberland and Durham. Br J Prev Soc Med. 1964;18(1):17-24

70. Lanzkowsky P. Variation in Leukaemia Incidence. Br Med J. 1964 Apr 4;1(5387):910.

71. Fraumeni JF. Seasonal variation in leukemia incidence (letter to the editor). Br Med J 1963;2:1408–9

72. Lee JAH. Seasonal variation in leukemia incidence (letter to the editor). Br Med J 1963;2:623

medRxiv preprint doi: https://doi.org/10.1101/2023.02.12.23285595; this version posted February 12, 2023. 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 .

73. Hayes, Donald M. The seasonal incidence of acute leukemia. A contribution to the epidemiology of the disease. Cancer, 1961, 14.6: 1301-1305

74. Benjamin MA, Rigby RA, Stasinopoulos DM. Generalized autoregressive moving average models. Journal of the American Statistical association. 2003 Mar 1;98(461):214-23.

75. 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.

76. 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.

77. McNeish D. On using Bayesian methods to address small sample problems. Structural Equation Modeling. : A Multidisciplinary Journal 2016 Sep 2;23(5):750-73.

78. Box GE, Jenkins GM, Reinsel GC, Ljung GM. Time Series Analysis: Forecasting and Control. John Wiley & Sons; 2015 May 29.

79. Petrevska B. Predicting tourism demand by ARIMA models. Economic research-Ekonomska istraživanja. 2017 Dec 1;30(1):939-50.

80. 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, No. 1, p. 012189). IOP Publishing.

81. Yang B, Li C, Li M, Pan K, Wang D. Application of ARIMA model in the prediction of the gross domestic product. 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.

82. 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.

83. Benvenuto D, Giovanetti M, Vassallo L, Angeletti S, Ciccozzi M. Application of the ARIMA model on the COVID-2019 epidemic dataset. Data Brief. 2020 Feb 26;29:105340. doi: 10.1016/j.dib.2020.105340. PMID: 32181302; PMCID: PMC7063124.

84. Ross GJ, Preece DA. The negative binomial distribution. Journal of the Royal Statistical Society: Series D (The Statistician). 1985 Sep;34(3):323-35.

85. Hilbe JM. Negative binomial regression. Cambridge University Press; 2011 Mar 17.

86. Zeger SL, Qaqish B. Markov regression models for time series: a quasi-likelihood approach. Biometrics. 1988 Dec 1:1019-31.

87. 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.

medRxiv preprint doi: https://doi.org/10.1101/2023.02.12.23285595; this version posted February 12, 2023. 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 .

88. Lukacs E. A characterization of the gamma distribution. The Annals of Mathematical Statistics. 1955 Jun 1;26(2):319-24.

89. 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).

90. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal statistical society: series B (Methodological). 1995 Jan;57(1):289-300.

91. Ford AM, Bennett CA, Price CM, Bruin MC, Van Wering ER, Greaves M. Fetal origins of the TEL-AML1 fusion gene in identical twins with leukemia. Proceedings of the National Academy of Sciences. 1998 Apr 14;95(8):4584-8.

92. Maia AT, Tussiwand R, Cazzaniga G, Rebulla P, Colman S, Biondi A, Greaves M. Identification of preleukemic precursors of hyperdiploid acute lymphoblastic leukemia in cord blood. Genes, Chromosomes and Cancer. 2004 May;40(1):38-43.

93. Gale KB, Ford AM, Repp R, Borkhardt A, Keller C, Eden OB, Greaves MF. Backtracking leukemia to birth: identification of clonotypic gene fusion sequences in neonatal blood spots. Proceedings of the National Academy of Sciences. 1997 Dec 9;94(25):13950-4.

94. Hein D, Dreisig K, Metzler M, Izraeli S, Schmiegelow K, Borkhardt A, Fischer U. The preleukemic TCF3-PBX1 gene fusion can be generated in utero and is present in≈ 0.6% of healthy newborns. Blood, The Journal of the American Society of Hematology. 2019 Oct 17;134(16):1355-8.

95. He JR, Ramakrishnan R, Hirst JE, Bonaventure A, Francis SS, Paltiel O, Håberg SE, Lemeshow S, Olsen S, Tikellis G, Magnus P. Maternal infection in pregnancy and childhood leukemia: a systematic review and meta-analysis. The Journal of pediatrics. 2020 Feb 1;217:98-109.

96. Naumburg E, Bellocco R, Cnattingius S, Jonzon A, Ekbom A. Perinatal exposure to infection and risk of childhood leukemia. Medical and pediatric oncology. 2002 Jun;38(6):391-7.

97. Nyari TA, Dickinson HO, Parker L. Childhood cancer in relation to infections in the community during pregnancy and around the time of birth. International journal of cancer. 2003 May 10;104(6):772-7.

98. Infante-Rivard C, Fortier I, Olson E. Markers of infection, breast-feeding and childhood acute lymphoblastic leukaemia. British journal of cancer. 2000 Dec;83(11):1559-64.

99. Kajtár P, Fazekasné KM, Méhes K. Month of birth in childhood acute lymphoblastic leukemia. Orvosi hetilap. 2003 Sep 1;144(38):1869-71.

100. Ma X, Buffler PA, Selvin S, Matthay KK, Wiencke JK, Wiemels JL, Reynolds P. Daycare attendance and risk of childhood acute lymphoblastic leukaemia. British journal of cancer. 2002 May;86(9):1419-24.

101. Gilham C, Peto J, Simpson J, Roman E, Eden TO, Greaves MF, Alexander FE. Day care in infancy and risk of childhood acute lymphoblastic leukaemia: findings from UK case-control study. Bmj. 2005 Jun 2;330(7503):1294.

medRxiv preprint doi: https://doi.org/10.1101/2023.02.12.23285595; this version posted February 12, 2023. 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 .

102. Kamper-Jørgensen M, Woodward A, Wohlfahrt J, Benn CS, Simonsen J, Hjalgrim H, Schmiegelow K. Childcare in the first 2 years of life reduces the risk of childhood acute lymphoblastic leukemia. Leukemia. 2008 Jan;22(1):189-93

103. Greenbaum S, Sheiner E, Wainstock T, Segal I, Ben-Harush M, Sergienko R, Walfisch A. Cesarean delivery and childhood malignancies: a single-center, population-based cohort study. The Journal of Pediatrics. 2018 Jun 1;197:292-6.

104. Wang R, Wiemels JL, Metayer C, Morimoto L, Francis SS, Kadan-Lottick N, DeWan AT, Zhang Y, Ma X. Cesarean section and risk of childhood acute lymphoblastic leukemia in a population-based, record-linkage study in California. American journal of epidemiology. 2017 Jan 15;185(2):96-105.

105. Marcotte EL, Thomopoulos TP, Infante-Rivard C, Clavel J, Petridou ET, Schüz J, Ezzat S, Dockerty JD, Metayer C, Magnani C, Scheurer ME. Caesarean delivery and risk of childhood leukaemia: a pooled analysis from the Childhood Leukemia International Consortium (CLIC). The Lancet Haematology. 2016 Apr 1;3(4):e176-85.

106. Sevelsted A, Stokholm J, Bønnelykke K, Bisgaard H. Cesarean section and chronic immune disorders. Pediatrics. 2015 Jan;135(1):e92-8.

107. Orsi L, Magnani C, Petridou ET, Dockerty JD, Metayer C, Milne E, Bailey HD, Dessypris N, Kang AY, Wesseling C, Infante-Rivard C. Living on a farm, contact with farm animals and pets, and childhood acute lymphoblastic leukemia: pooled and meta-analyses from the Childhood Leukemia International Consortium. Cancer medicine. 2018 Jun;7(6):2665-81.

108. Ward G. The infective theory of acute leukaemia. Br J Child Dis. 1917;14:10-20.

109. Heath Jr CW, Hasterlik RJ. Leukemia among children in a suburban community. The American Journal of Medicine. 1963 Jun 1;34(6):796-812.

110. Francis SS, Selvin S, Yang W, Buffler PA, Wiemels JL. Unusual space-time patterning of the Fallon, Nevada leukemia cluster: Evidence of an infectious etiology. Chemico-biological interactions. 2012 Apr 5;196(3):102-9.

111. Cazzaniga G, Bisanti L, Randi G, Deandrea S, Bungaro S, Pregliasco F, Perotti D, Spreafico F, Masera G, Valsecchi MG, Biondi A. Possible role of pandemic AH1N1 swine flu virus in a childhood leukemia cluster. Leukemia. 2017 Aug;31(8):1819-21.

112. Kroll ME, Draper GJ, Stiller CA, Murphy MF. Childhood leukemia incidence in Britain, 1974–2000: time trends and possible relation to influenza epidemics. Journal of the National Cancer Institute. 2006 Mar 15;98(6):417-20.

113. 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.

114. Li CK, Zee B, Lee J, Chik KW, Ha SY, Lee V. Impact of SARS on development of childhood acute lymphoblastic leukaemia. : Leukemia. 2007 Jul;21(7):1353-6. doi: 10.1038/sj.leu.2404729. PMID: 17579654; PMCID: PMC7099337.

115. Taub JW, Ge Y, Xavier AC. COVID-19 and childhood acute lymphoblastic leukemia. : Pediatric Blood & Cancer. 2020 Jul;67(7).

medRxiv preprint doi: https://doi.org/10.1101/2023.02.12.23285595; this version posted February 12, 2023. 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 .

116. Greaves M. COVID-19 and childhood acute lymphoblastic leukemia. : Pediatric Blood & Cancer. 2020 Dec;67(12):e28481.

117. Jarvis KB, Lind A, LeBlanc M, Ruud E. Observed reduction in the diagnosis of acute lymphoblastic leukaemia in children during the COVID-19 pandemic. : Acta Paediatrica (Oslo, Norway: 1992). 2021 Feb;110(2):596.

118. Ferrari A, Zecca M, Rizzari C, Porta F, Provenzi M, Marinoni M, Schumacher RF, Luksch R, Terenziani M, Casanova M, Spreafico F. Children with cancer in the time of COVID-19: an 8-week report from the six pediatric onco-hematology centers in Lombardia, Italy. : Pediatric blood & cancer. 2020 Aug;67(8).

119. Schüz J, Borkhardt A, Bouaoun L, Erdmann F. The impact of the COVID-19 pandemic on the future incidence of acute lymphoblastic leukaemia in children: Projections for Germany under a COVID-19 related scenario. : International Journal of Cancer. 2022 Jul 1;151(1):153-5.

120. Erdmann F, Spix C, Schrappe M, Borkhardt A, Schüz J. Temporal changes of the incidence of childhood cancer in Germany during the COVID-19 pandemic: Updated analyses from the German Childhood Cancer Registry. : The Lancet Regional Health–Europe. 2022 Jun 1;17.

121. Silver LJ, Desai P, Shah S, Krystal J, Taylor M, Murphy K. New pediatric leukemia/lymphoma diagnoses during the COVID-19 pandemic: A New York perspective. : Pediatric Blood & Cancer. 2022 Jul 23.

122. Pelland-Marcotte MC, Xie L, Barber R, Elkhalifa S, Frechette M, Kaur J, Onysko J, Bouffet E, Fernandez CV, Mitchell D, Rayar M. Incidence of childhood cancer in Canada during the COVID-19 pandemic. : CMAJ. 2021 Nov 29;193(47):E1798-806.

123. Rechavi Y, Rechavi G, Rechavi E. Origin of childhood leukaemia: COVID-19 pandemic puts the 'delayed infection'hypothesis to the test. : Internal Medicine Journal. 2021 Oct;51(10):1761.

124. Lillie K. Leukaemia and lockdown: the delayed infection model of childhood acute lymphoblastic leukaemia and the COVID-19 pandemic. : Pediatric Blood & Cancer. 2021 Oct;68(10):e29194.

125. Fisman D. Seasonality of viral infections: mechanisms and unknowns. : Clinical Microbiology and Infection. 2012 Oct 1;18(10):946-54.

126. 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;466(9):2282-8

127. Chan AW, Altman DG. Epidemiology and reporting of randomised trials published in PubMed journals. : The Lancet. 2005 Mar 26;365(9465):1159-62.

128. Stewart A, Kneale GW. Role of local infections in the recognition of haemopoietic neoplasms. : Nature. 1969 Aug;223(5207):741-2.

129. Edwards JH. The recognition and estimation of cyclic trends. : Annals of human genetics. 1961 Sep;25(1):83-7

130. Roger JH. A significance test for cyclic trends in incidence data. : Biometrika. 1977 Apr 1;64(1):152-5

medRxiv preprint doi: https://doi.org/10.1101/2023.02.12.23285595; this version posted February 12, 2023. 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 .

131. Mardia KV. Statistics of directional data. : New York, NY: Academic Press, 1972