Spatial clustering of childhood leukaemia in Switzerland: A nationwide study

Authors: Garyfallos Konstantinoudis[1], Christian Kreis[1], Roland A Ammann[2], Felix Niggli[3], Claudia E Kuehni[1], Ben D Spycher[1] for the Swiss Paediatric Oncology Group and the Swiss National Cohort Study Group

Affiliations:

1 Institute of Social and Preventive Medicine (ISPM), University of Bern, Bern, Switzerland.

2 Department of Paediatrics Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland.

3 University Children's Hospital Zurich, Zurich, Switzerland.

Corresponding author:

Ben D Spycher

Institute of Social and Preventive Medicine (ISPM), University of Bern, Finkenhubelweg 11, 3012 Bern, Switzerland.

E-mail: ben.spycher@ispm.unibe.ch

Tel: +41 31 631 33 46

Fax: +41 31 631 35 20

Keywords

cancer registry; childhood cancer; acute lymphoid leukaemia; point patterns; spatial clusters

Abbreviations

ALL – acute lymphoid leukaemia

AML – acute myeloid leukaemia

SCCR – Swiss Childhood Cancer Registry

ICCC3 – International Classification of Childhood Cancers, Third Edition

NN – nearest neighbours

Novelty and impact

This nationwide investigation of spatial clustering of childhood leukaemia is unique in that it used precise geocodes of residence and carefully adjusted for multiple tests. Overall, no evidence of spatial clustering was found. Although individual tests did indicate clustering and a small cluster of acute lymphoblastic leukaemia was identified, Monte Carlo simulations show that such results could easily arise by chance. This highlights the importance of appropriately accounting for multiple testing in clustering studies.

Abstract

The aetiology of childhood leukaemia remains largely unknown. Several hypotheses involve environmental exposures that could implicate spatial clustering of cases. The evidence from previous clustering studies is inconclusive. Most of them used areal data and thus had limited spatial resolution. We investigated whether childhood leukaemia tends to cluster in space using exact geocodes of place of residence both at the time of birth or diagnosis. We included 1871 leukaemia cases diagnosed between 1985 and 2015 at age 0-15 years from the Swiss Childhood Cancer Registry. For each case, we randomly sampled 10 age and sex matched controls from national censuses closest in time. We used the difference of k-functions, Cuzick-Edwards' test and Tango's index for point data to assess spatial clustering and Kulldorff's circular scan to detect clusters. We separately investigated acute lymphoid leukaemia (ALL), acute myeloid leukaemia (AML), different age groups at diagnosis (0-4, 5-15 years) and adjusted for multiple testing. After adjusting for multiple testing, we found no evidence of spatial clustering of childhood leukaemia neither around time of birth ($P = 0.52$) nor diagnosis ($P = 0.51$). Individual tests indicated spatial clustering for leukaemia diagnosed at age 5-15 years, P k-functions = 0.05 and P Cuzick-Edwards' = 0.04 and a cluster of ALL cases diagnosed at age 0-4 years in a small rural area ($P = 0.05$). This study provides little evidence of spatial clustering of childhood leukaemia in Switzerland and highlights the importance of accounting for multiple testing in clustering studies.

Introduction

Childhood leukaemia is the most common childhood cancer, accounting for about a third of incident cases in many high income countries.[1] Its aetiology remains largely unknown, the only established environmental risk factor being ionizing radiation at medium to high doses.[2, 3] Suspected environmental risk factors include low-dose ionising radiation (e.g. background radiation), electromagnetic fields, traffic-related air pollution, pesticides and infections.[2, 4, 5] Geographical variation of risk factors might lead to spatial variation in risk of disease and possibly spatial clustering of incident cases. Spatial clustering might result from emissions originating from point sources or linear structures such as busy roads. Increased risks of childhood leukaemia have for instance been reported in the vicinity of highways,[6] nuclear power plants[7] and petrol stations.[8] Thus investigating spatial clustering of childhood leukaemia might help pinpoint relevant pollution sources and provide new clues about possible environmental risk factors.[9] Furthermore, observing spatial clustering at specific time points, e.g. around diagnosis or birth, may indicate age windows of increased susceptibility.

Several studies have assessed spatial clustering of childhood leukaemia.[10, 11] Most studies aimed to determine whether leukaemia cases tend to occur closer to each other than expected by chance (global clustering tests) while some attempted to identify clusters (cluster detection tests).[12] The majority of studies focused on place of residence at time of diagnosis. Studies in the UK with partly overlapping data mostly reported evidence of clustering or clusters of childhood leukaemia,[9, 13-15] whereas the evidence from other countries was mixed. Evidence of clustering or clusters around time of diagnosis was found in Greece,[16] Hong Kong,[17] South Hungary,[18] Florida[19] and Argentina[20] but not in Sweden,[21] Ohio,[22] Germany,[23] France[24] and Spain.[25] Evidence of clustering around time of birth was found in New Zealand[26] but not in Ohio.[27] Both time points were assessed in a study in Denmark and clustering around time of diagnosis but not around time of birth was reported.[28]

These inconsistent findings might reflect real differences between countries or regions, but could also have resulted from differences in analytical approaches, data aggregation and statistical power. In most previous studies, only areal data was available, i.e. data aggregated to administrative areas, and this

may have reduced the statistical power to detect clustering.[29-32] To our knowledge only three studies have used point pattern data based on precise geocodes.[25, 27, 28] Statistical power is also affected by sample size as well as the shape and frequency of putative clusters (clustering scenario). Multiple testing is another common problem, as multiple tests are typically performed to cover different clustering scenarios, diagnostic subgroups and age groups. To our knowledge only two studies have corrected for multiple testing.[22, 26] Lastly, though many studies have investigated acute lymphoid leukaemia (ALL) only few studies have examined the spatial distribution of acute myeloid leukaemia (AML).[23, 24, 33]

In this nationwide study, we investigated whether there is evidence of spatial clustering of childhood leukaemia in Switzerland using geocoded residential locations of cases and representative controls at birth and at diagnosis. We assessed spatial clustering for different age groups (0-4, 5-15 years) and the two main diagnostic subgroups, ALL and AML. We used three different global clustering tests and a test for cluster detection, and paid particular attention to corrections for multiple testing.

Methods

Study population

We included leukaemia cases recorded in the Swiss Childhood Cancer Registry (SCCR) diagnosed in Switzerland between 1985 and 2015 at age 0-15 years. The SCCR is a nationwide population-based registry with an estimated completeness of 91% during this period and age group, and of approximately 95% since the mid-1990s.[34] The SCCR tracks residential address histories of cases from diagnosis back to birth. Addresses were geocoded to the Swiss grid coordinate system using the geo-referenced building addresses from the Swiss postal system (GeoPost) or manually using the geoportal maintained by the Federal Office of Topography (http://map.geo.admin.ch). We obtained precise geocodes to within 100m for 94% of the available cases (Supporting Information Text S1). We investigated pairs of cases living <50m from each other for possible sibling relationships based on family names (child and parents) recorded in the SCCR and included only one child for each sibling pair identified.

We obtained individual data, including geocoded place of residence, on the Swiss resident population from the national censuses in 1990, 2000 and 2010 through the Swiss National Cohort study.[35] From this population, we randomly sampled 10 controls per case matched for age, gender and time of birth or diagnosis. This case-control ratio represents a compromise between including as many controls as possible (see Diggle's discussion of the Cuzick-Edwards' test[36]) and computational burden. We used a two-stage approach to sample controls as in a previous study of space-time clustering of childhood cancers.[37] Briefly, to select a control, we first selected a municipality – the smallest administrative unit in Switzerland – by weighted random sampling with weights proportional to the estimated population counts in the year of a case's diagnosis (or year of birth for clustering analyses around time of birth). We then selected one of the two censuses nearest to the case's year of diagnosis (or birth) using weighted random sampling with linearly interpolated weights. For instance, for a child diagnosed in 1996, the 1990 census was selected with probability (2000-1996)/10 = 0.4 and the 2000 census with probability (1996-1990)/10 = 0.6. Lastly, we randomly sampled a control without replacement from among the children residing in the same municipality at the selected census and belonging to the same sex and age group (0-4, 5-9, 10-15 years) as the corresponding case.

Outcomes

The SCCR classifies diagnoses according to the International Classification of Childhood Cancers Third Edition (ICCC3).[38] We separately examined leukaemia (ICCC3 main group I), ALL (Ia) and AML (Ib), and age groups 0-4 and 5-15 years for leukaemia and 0-4 years for ALL. All sub-groups were selected *a priori* considering the possibility of differing aetiologies,[2] sample size and consistency with previous studies.

Statistical Analysis

We applied spatial clustering tests appropriate for point pattern data: three global clustering tests, the difference of k-functions,[39] Cuzick-Edwards' test[36] and Tango's index for point data;[40] and the most widespread cluster detection method, Kulldorff's circular scan.[41] We chose these tests for comparability with previous studies[22, 25, 28] and because different tests may be sensitive to different clustering scenarios.[42] Each of the three global tests requires an input parameter defining "closeness" between cases. For Cuzick-Edwards' test the input parameter is k and a case is considered close to

another if it belongs to the k nearest neighbours (k-NN) of the case. For the other tests, the input parameters relate to Euclidean distance. We used a range of different input values for each test, assuring that closeness between two cases was on average comparable across the different tests (Supporting Information Table S1). For k-functions we additionally calculated 95% simulation envelopes (2.5% and 97.5% quantiles of the difference in k-functions in Monte Carlo samples) to highlight distances for which evidence of clustering was strongest. Kulldorff's circular scan draws concentric circles with varying radii around all locations of cases and controls. For each circle, a likelihood ratio is calculated for the number of cases within and outside the circle under the binomial distribution. The circle with the highest likelihood ratio represents the cluster least likely to have occurred by chance and is designated the most likely cluster. A detailed description of the four tests and our implementation is given in the Supporting Information Texts S2, S3 and S4.

We paid particular attention to correct for multiple testing. We did this in two steps using a Monte Carlo procedure. We obtained 999 Monte Carlo samples by randomly permuting case control labels keeping the locations fixed (random labelling). First, we calculated *p*-values adjusted for the multiple input parameters used in each statistical test (The details for this step are reported in Supporting Information Texts S2.1-S2.4). For example, for Cuzick-Edwards' test, this step consisted in first calculating individual *p*-values for each input value of k-NN based on a chi-square approximation of the distribution of the test statistic, minimising over these *p*-values to obtain a profile *p*-value, and ranking the profile *p*-value of the empirical dataset among the profile *p*-values of Monte Carlo samples to obtain the *adjusted p*-value.[40] In a second step, we corrected for the multiple testing due to different statistical tests, diagnostic groups and age groups (24 tests in total) (The details for this step are reported in Supporting Information Text S5). Briefly, we treated each of the Monte Carlo samples as if it were the empirical sample and the empirical sample as if it were a Monte Carlo sample and calculated *adjusted p*-values for each test as in the first step. For each sample, we then selected the smallest *adjusted p*-value over the different tests (minimum *p*-value). Finally, we ranked the minimum *p*-value of the true empirical sample among the minimum *p*-values of the Monte Carlo samples to obtain an *overall p*-value adjusting for all tests performed. This procedure accounts for correlations

between tests making the adjustment less conservative than a Bonferroni correction (Supporting Information Text S5).

In additional analyses, we investigated spatial clustering in the three different sub-periods 1985-1994, 1995-2004 and 2005-2015 and for time of diagnosis conducted a sensitivity analysis excluding cases with less precise geocodes (margin of error > 100m).

Results

Study Population

We identified 1871 eligible cases of childhood leukaemia in the SCCR. For analyses around time of birth, after excluding those born before 1985, born outside of Switzerland or with missing geocodes and one case from a sibling pair, we included 1297 leukaemia cases (of whom 850 had identical place of birth and diagnosis) of which 1042 had been diagnosed with ALL and 180 with AML (Figure 1). For the analysis at time of diagnosis, we included 1865 leukaemia cases, of which 1485 were ALL and 272 AML. Table 1 shows the distribution of cases stratified by sex and calendar period. ALL accounted for almost 80% of the leukaemia cases and boys outnumber girls as expected.

Spatial Clustering

After adjusting for multiple testing, we found no evidence of spatial clustering of childhood leukaemia cases, neither around time of birth (*overall P* = 0.52) nor diagnosis (*overall P* = 0.51) (Table 2). When we disregarded the multiple testing due to different diagnostic and age groups, individual tests did show evidence of spatial clustering.

Based on k-functions, the strongest evidence for spatial clustering was for leukaemia at age 5-15 years using residence at diagnosis (*adjusted P* = 0.05) (Table 2). Figure 2 shows that the observed difference in k-functions between cases and controls for that subgroup stays within the 95% simulation envelopes for distances from 500m to 5km, providing no evidence of clustering. By contrast, for distances of less than 500m we did observed evidence of clustering, with the empirical difference in k-functions exceeding the envelope at 100m (Figure 2). The graphs for the other sub-groups did not show any departure (data not shown).

Similarly, Cuzick-Edwards' test also showed the strongest evidence of clustering for leukaemia cases aged 5-15 years around the time of diagnosis (*adjusted P* = 0.04) (Table 2). Varying the number of k-NN, evidence was strongest for 1-NN (unadjusted *P* = 0.01, Supporting Information Table S2) reflecting distances of 270m on average.

The strongest evidence from Tango's test was also observed for leukaemia diagnosed at 5-15 years for time of diagnosis (*adjusted P* = 0.05) (Table 2). Varying the input parameter θ, evidence was strongest for $\theta = 354$ (unadjusted *P* = 0.01, Supporting Information Table S3) reflecting distances of 250m.

Kulldorff's circular scan showed the strongest evidence of a cluster for ALL at age 0-4 years with similar results at the time of birth (*adjusted P* = 0.05) and diagnosis (*adjusted P* = 0.05) (Table 2). In both analyses, the most likely cluster consisted of 5 cases (4 males and 1 female) living within a circle of approximately 500m radius, born during 1994-2009 and diagnosed during 1999-2012. In that circle, the number of expected cases (based on Monte Carlo sampling) was 0.39 suggesting a relative risk of 12.8. The same group of cases also accounted for the low *p*-values observed for leukaemia at age 0-4 years at both birth and diagnosis. This is unsurprising given that 75% of cases diagnosed at this age in our study did not relocate between birth and diagnosis. There was no evidence of additional clusters of ALL at 0-4 years around time of birth (*P* of the second most likely cluster = 0.36). Similarly there was no evidence of clusters in the other diagnostic or age groups (Table 2).

For cases of leukaemia aged 0-15 years, we also tested for spatial clustering within the sub-periods 1985-1994, 1995-2004 and 2005-2015 (without *overall* adjustment for multiple diagnostic and age groups). The sub-period analysis showed no evidence of spatial clustering around the time of birth (Table 3). In contrast, there was evidence of clustering around the place of diagnosis during 1995-2004, with *adjusted P* = 0.04 for k-functions (strongest evidence at 1km), *adjusted P* = 0.05 for Cuzick-Edwards' (1-NN corresponding to 270m on average) and *adjusted P* = 0.04 for Tango's test ($\theta = 2121$ corresponding to 1.5km distance). Evidence was weaker for the other sub-periods. Kulldorff's circular scan was indicative of a cluster during 2005-2015 (*adjusted P* = 0.05) pinpointing the same rural area as for the entire period.

In a sensitivity analysis for place of diagnosis excluding the 104 cases with less precise geocodes (margin of error >100m) the *overall P* did not provide evidence of clustering (*overall P* = 0.15). However, compared to the main analysis, the *adjusted P* dropped markedly (Supporting Information Table S4). K-functions, Cuzick-Edwards' and Tango's test also indicated spatial clustering of leukaemia at age 0-15 years. Results of Kulldorff's scan remained similar, again pinpointing the same set of five cases as the most likely cluster.

Post-hoc analysis

In a post-hoc analysis we investigated the risk of ALL in the municipality where the most likely cluster lies. We identified 10 children diagnosed with leukaemia (8 with ALL) during the study period 1985-2015 with place of residence in that municipality. The incidence risk ratio for ALL adjusted for age, sex and year of diagnosis comparing children living within that municipality to those living outside was 4.28 (95% CI 2.13 - 8.57). When we restricted the analysis to children aged 0-4 years, the adjusted risk ratio increased to 7.37 (95% CI 3.50 - 15.5) (for details see Supporting Information Text S6).

Discussion

Main findings

This nationwide study of childhood leukaemia in Switzerland covering the period 1985-2015 found no *overall* evidence of spatial clustering of leukaemia cases, neither around time of birth nor diagnosis. When we disregarded multiple testing, the strongest evidence from individual tests was for leukaemia at age 5-15 years at place of diagnosis during the entire period and for leukaemia at age 0-15 years at diagnosis during the sub-period 1995-2004. Evidence of spatial clustering tended to be stronger at small spatial scales representing distances of <500m. The strongest evidence for a cluster based on the Kulldorff's circular scan was for a group of 5 ALL cases aged 0-4 years at diagnosis living in a rural area of Switzerland at birth within a circle of radius 500m. We found no evidence of spatial clustering of AML cases.

Discussion in the context of previous studies

Few studies have investigated spatial clustering of childhood leukaemia using exact geocodes.[25, 27, 28] A case-control study in San Francisco investigating place of residence at birth found no evidence of spatial clustering.[27] A study in Denmark investigating both residence at diagnosis and at birth reported evidence of spatial clustering only for ALL cases aged 2-6 years at place of diagnosis.[28] A study in Spain reported evidence of leukaemia clustering for only one of 5 investigated regions. This study also searched for clusters using Kulldorff's circular scan and found weak evidence of a small leukaemia cluster in Barcelona.[25] Two other studies have also analysed point pattern data but with less precise geocodes. A study in Ohio using street level precision and residence at diagnosis reported no evidence of global clustering or local clusters.[22] A study in New Zealand using centroids of meshblocks (a small geographic unit) of children's residence at birth found evidence of ALL clustering for children aged 10-14 years.[26] That study did not search for clusters. Though all of the above studies either used different statistical tests and/or multiple diagnostic or age subgroups, none of them adjusted for multiple testing.

Our results are less comparable with studies that have used areal data. Several of these studies have reported evidence of global clustering[9, 13, 14, 16, 18] or clusters[19, 20] whereas others found weak or no evidence.[23, 24, 43-45] The heterogeneity between the results might partly be attributable to the modifiable areal unit problem whereby results of spatial analyses may be greatly affected by the size and shape of area boundaries.[46]

There have been reports of clusters of childhood leukaemia in Sellafield, Cumbria in the UK,[47] Krümmel, Elbmarsch in Germany[48] and Fallon, Nevada in the US.[49] However, these clusters were discovered circumstantially and are not the result of a systematic scan over a large pre-specified area. No causative factor has been established for any of these clusters.

In a recent study in Switzerland, we reported evidence of space-time clustering of leukaemia cases at birth.[37, 50] That analysis was only sensitive to temporary, localized risk increases, a pattern consistent with epidemics of infections, which might point to an infectious aetiology. In contrast, the present analysis is most sensitive to localised risk increases that are stable over prolonged periods of time and

are thus more indicative of a pollution source. These two patterns are not mutually exclusive and the presence of both could indicate that childhood leukaemia can have different aetiologies.

Strengths and Limitations

The main strength of our study was the availability of precise geocodes, not only for cases but for the entire population, allowing us to sample a large number of representative controls. We were able to investigate both residence at diagnosis and birth. We ascertained cases from a national cancer registry with high coverage. Furthermore, we considered four different tests in order to cover a range of possible clustering scenarios. While the difference of k-functions, Cuzick-Edwards' and Tango's test are sensitive to the general tendency of cases to be closer to each other than would be expected by chance, Kuldorff's scan method is sensitive to a single pronounced cluster. Furthermore, while k-functions and Tango's test use Euclidian distance, Cuzick-Edwards' test uses nearest neighbours and thus weights distances differently in scarcely and densely populated areas. Finally, we carefully adjusted for multiple testing. Our correction was less conservative than the commonly used Bonferroni method, which, contrary to our method, does not account for correlations between tests. In doing so, we noticed that the test statistics for the difference of k-functions and Tango's test are highly correlated and future studies may opt for only one of them in order to mitigate the multiple testing problem (Supporting Information Text S5 and S7).

Though the SCCR has a high coverage, a small proportion of cases were not registered and missed by our study. This may have reduced the statistical power to detect spatial clustering. We could not adjust for full residential histories as these data were not available for controls. Also, as all other studies, we could not account for time spent away from homes. However, place of residence is a reasonable surrogate of young children's local environment as they spend much of their time in or near their homes.[51] Lastly, we cannot guarantee that control locations were fully representative of the population at risk in non-census years. Although our control sampling scheme accounted for changes in the geographic distribution of the population between census years at the municipal level, we could not control for uneven population shifts within municipalities.

Interpretation of findings

Overall we found no evidence of spatial clustering of childhood leukaemia. Although result tables each contain two to five *p*-values around the standard 5%-level, careful correction for multiple testing shows that such tables are not exceptional ($P > 0.5$) and can easily arise by chance alone.

The strongest evidence for clustering was around diagnosis for the age group 5-15 years at small distances (100 m for k-functions and the first NN for the Cuzick-Edwards' test). This suggests that if indeed there was spatial clustering which our global tests had insufficient statistical power to confirm, it is likely to have involved highly localised risk increases. Such clustering would be compatible with point sources of pollution, which give rise to increased exposure levels in their immediate vicinity.

The most likely cluster occurred in a rural area comprising 5 cases of ALL aged 0-4 years. The cluster represents a more than 10-fold risk increase within a circle with radius of 500m. Had such a cluster come into focus incidentally, it would indeed seem extraordinary. However, considering that it resulted from a systematic, country-wide scan, involving the evaluation of a vast number of circles of varying radii and centre points, and that multiple diagnostic and age groups were investigated, the cluster is by no means extraordinary. In fact, 266 of the 999 Monte Carlo samples produced clusters of cases by chance alone that attained 5%-significance level for at least one of the six diagnostic and age groups. This highlights the importance of adjusting for multiple testing.

Conclusion

Overall, this study provided no evidence of spatial clustering in Switzerland during the period 1985-2015. However, we cannot exclude the presence of weak spatial clustering, the effects of which may have been too small for our analyses to detect. If indeed there was clustering, it is likely to have occurred at a small geographic scale, a scenario compatible with the presence of numerous highly localised pollution sources. Our study also highlights the importance of adjusting for multiple testing and demonstrates that localised excesses of childhood leukaemia cases, which may appear extraordinary when brought to one's attention, may in fact occur quite often by chance.

Acknowledgements

The authors thank Prof Dr Lutz Dümbgen for his valuable advice on the multiple testing problem and his comments on the Supporting Information Text S5.

This work was supported by Swiss Cancer Research (4012-08-2016, 3515-08-2014, 3049-08-2012), the Swiss Federal Office of Public Health (08.001616, 10.002946, 12.008357) and the Swiss Cancer League (02224-03-2008). BD Spycher was supported by a Swiss National Science Foundation fellowship (PZ00P3_147987).

The work of the Swiss Childhood Cancer Registry is supported by the Swiss Paediatric Oncology Group (www.spog.ch), Schweizerische Konferenz der kantonalen Gesundheitsdirektorinnen und –direktoren (www.gdk-cds.ch), Swiss Cancer Research (www.krebsforschung.ch), Kinderkrebshilfe Schweiz (www.kinderkrebshilfe.ch), Ernst-Göhner Stiftung, Stiftung Domarena and National Institute of Cancer Epidemiology and Registration (www.nicer.ch).

We thank the Swiss Federal Statistical Office for providing mortality and census data and for the support which made the Swiss National Cohort and this study possible. This work was supported by the Swiss National Science Foundation (grant nos. 3347CO-108806, 33CS30_134273 and 33CS30_148415).

Member of study groups

The members of the Swiss Pediatric Oncology Group Scientific Committee:

R A Ammann (Bern), R Angst (Aarau), M Ansari (Geneva), M Beck Popovic(Lausanne), E Bergstraesser (Zurich), P Brazzola (Bellinzona), J Greiner (St. Gallen), M Grotzer (Zurich), H Hengartner (St Gallen), T Kuehne (Basel), K Leibundgut (Bern), F Niggli (Zurich), J Rischewski (Lucerne), N von der Weid (Basel)

The members of the Swiss National Cohort Study Group:

Matthias Egger (Chairman of the Executive Board), Adrian Spoerri and Marcel Zwahlen (all Bern), Milo Puhan (Chairman of the Scientific Board), Matthias Bopp (both Zurich), Nino Künzli (Basel), Fred Paccaud (Lausanne) and Michel Oris (Geneva).

Conflict of interest

The authors declare that they have no conflict of interest.

Ethical standard

Ethics approval was granted through the Ethics Committee of the Canton of Bern to the SCCR on the 22th of July 2014 (KEK-BE: 166/2014).

References

1. Stiller CA. Aetiology and Epidemiology. In: Plowman PN, Pinkerton CR, Pieters R. *Paediatric Oncology*ed. London: Arnold, 2004: 3-24.
2. Wiemels J. Perspectives on the causes of childhood leukemia. *Chemico-biological interactions* 2012;**196**: 59-67.
3. Wakeford R. The cancer epidemiology of radiation. *Oncogene* 2004;**23**: 6404-28.
4. Eden T. Aetiology of childhood leukaemia. *Cancer Treat Rev* 2010;**36**: 286-97.
5. Whitehead TP, Metayer C, Wiemels JL, Singer AW, Miller MD. Childhood Leukemia and Primary Prevention. *Curr Prob Pediatr Ad* 2016;**46**: 317-52.
6. Spycher BD, Feller M, Roosli M, Ammann RA, Diezi M, Egger M, Kuehni CE. Childhood cancer and residential exposure to highways: a nationwide cohort study. *European journal of epidemiology* 2015;**30**: 1263-75.
7. Kaatsch P, Spix C, Schulze-Rath R, Schmiedel S, Blettner M. Leukaemia in young children living in the vicinity of German nuclear power plants. *International journal of cancer* 2008;**122**: 721-6.
8. Brosselin P, Rudant J, Orsi L, Leverger G, Baruchel A, Bertrand Y, Nelken B, Robert A, Michel G, Margueritte G, Perel Y, Mechinaud F, et al. Acute childhood leukaemia and residence next to petrol stations and automotive repair garages: the ESCALE study (SFCE). *Occupational and environmental medicine* 2009;**66**: 598-606.
9. McNally RJ, Alexander FE, Vincent TJ, Murphy MF. Spatial clustering of childhood cancer in Great Britain during the period 1969-1993. *International journal of cancer* 2009;**124**: 932-6.
10. Little J. *Epidemiology of Childhood Cancer.* ed. Lyon, France: IARC Scientific Publications No. 149, 1999. 385.
11. McNally RJ, Eden TO. An infectious aetiology for childhood acute leukaemia: a review of the evidence. *British journal of haematology* 2004;**127**: 243-63.
12. Kulldorff M. Tests of spatial randomness adjusted for an inhomogeneity: A general framework. *J Am Stat Assoc* 2006;**101**: 1289-305.
13. Alexander F, Cartwright R, McKinney PA, Ricketts TJ. Investigation of spacial clustering of rare diseases: childhood malignancies in North Humberside. *Journal of Epidemiology & Community Health* 1990;**44**: 39-46.

14. Alexander F. Investigations of localised spatial clustering and extra-Poisson variation. In: Draper G. *The geographical epidemiology of childhood leukaemia and non-Hodgkin lymphomas in Great Britain, 1966-83* ed., vol. 83. London: HMSO, 1991: 69-76.
15. Committee on Medical Aspects of Radiation in the Environment (COMARE), Eleventh Report. The distribution of childhood leukaemia and other childnood cancers in Great Britain 1969-1993. Health Protection Agency, 2006.
16. Petridou E, Alexander FE, Trichopoulos D, Revinthi K, Dessypris N, Wray N, Haidas S, Koliouskas D, Kosmidis H, Piperopoulou F, Tzortzatou F. Aggregation of childhood leukemia in geographic areas of Greece. *Cancer causes & control : CCC* 1997;**8**: 239-45.
17. Alexander FE, Chan LC, Lam TH, Yuen P, Leung NK, Ha SY, Yuen HL, Li CK, Li CK, Lau YL, Greaves MF. Clustering of childhood leukaemia in Hong Kong: association with the childhood peak and common acute lymphoblastic leukaemia and with population mixing. *British journal of cancer* 1997;**75**: 457-63.
18. Nyari TA, Ottoffy G, Bartyik K, Thurzo L, Solymosi N, Cserni G, Parker L, McNally RJQ. Spatial clustering of childhood acute lymphoblastic leukaemia in Hungary. *Pathology and Oncology Research* 2013;**19**: 297-302.
19. Amin R, Hendryx M, Shull M, Bohnert A. A cluster analysis of pediatric cancer incidence rates in Florida: 2000–2010. *Statistics and Public Policy* 2014;**1**: 69-77.
20. Agost L. Analysis of spatial-temporal clusters of childhood cancer incidence in the province of Cordoba, Argentina (2004-2013). *Archivos argentinos de pediatria* 2016;**114**: 534-43.
21. Hjalmars U, Kulldorff M, Gustafsson G, Nagarwalla N. Childhood leukaemia in Sweden: using GIS and a spatial scan statistic for cluster detection. *Statistics in medicine* 1996;**15**: 707-15.
22. Wheeler DC. A comparison of spatial clustering and cluster detection techniques for childhood leukemia incidence in Ohio, 1996-2003. *International journal of health geographics* 2007;**6**: 13.
23. Schmiedel S, Blettner M, Kaatsch P, Schuz J. Spatial clustering and space-time clusters of leukemia among children in Germany, 1987-2007. *European journal of epidemiology* 2010;**25**: 627-33.
24. Demoury C, Goujon-Bellec S, Guyot-Goubin A, Hemon D, Clavel J. Spatial variations of childhood acute leukaemia in France, 1990-2006: global spatial heterogeneity and cluster detection at 'living-zone' level. *Eur J Cancer Prev* 2012;**21**: 367-74.
25. Ramis R, Gomez-Barroso D, Tamayo I, Garcia-Perez J, Morales A, Pardo Romaguera E, Lopez-Abente G. Spatial analysis of childhood cancer: a case/control study. *PLoS One* 2015;**10**: e0127273.

26. Dockerty JD, Sharples KJ, Borman B. An assessment of spatial clustering of leukaemias and lymphomas among young people in New Zealand. *Journal of epidemiology and community health* 1999;**53**: 154-8.

27. Selvin S, Ragland KE, Chien EY-L, Buffler PA. Spatial analysis of childhood leukemia in a case/control study. *International Journal of Hygiene & Environmental Health* 2004;**207**: 555-62.

28. Schmiedel S, Jacquez GM, Blettner M, Schuz J. Spatial clustering of leukemia and type 1 diabetes in children in Denmark. *Cancer causes & control : CCC* 2011;**22**: 849-57.

29. Ozonoff A, Jeffery C, Manjourides J, White LF, Pagano M. Effect of spatial resolution on cluster detection: a simulation study. *International journal of health geographics* 2007;**6**: 52.

30. Meliker JR, Jacquez GM, Goovaerts P, Copeland G, Yassine M. Spatial cluster analysis of early stage breast cancer: a method for public health practice using cancer registry data. *Cancer Causes & Control* 2009;**20**: 1061-9.

31. Olson KL, Grannis SJ, Mandl KD. Privacy protection versus cluster detection in spatial epidemiology. *Am J Public Health* 2006;**96**: 2002-8.

32. Schmiedel S, Blettner M, Schuz J. Statistical power of disease cluster and clustering tests for rare diseases: a simulation study of point sources. *Spatial and spatio-temporal epidemiology* 2012;**3**: 235-42.

33. Alexander FE, Ricketts TJ, Williams J, Cartwright RA. Methods of Mapping and Identifying Small Clusters of Rare Diseases with Applications to Geographical Epidemiology. *Geogr Anal* 1991;**23**: 158-73.

34. Schindler M, Mitter V, Bergstraesser E, Gumy-Pause F, Michel G, Kuehni CE, Swiss Paediatric Oncology G. Death certificate notifications in the Swiss Childhood Cancer Registry: assessing completeness and registration procedures. *Swiss Med Wkly* 2015;**145**: w14225.

35. Bopp M, Spoerri A, Zwahlen M, Gutzwiller F, Paccaud F, Braun-Fahrlander C, Rougemont A, Egger M. Cohort Profile: the Swiss National Cohort--a longitudinal study of 6.8 million people. *International journal of epidemiology* 2009;**38**: 379-84.

36. Cuzick J, Edwards R. Spatial Clustering for Inhomogeneous Populations. *J Roy Stat Soc B Met* 1990;**52**: 73-104.

37. Kreis C, Grotzer M, Hengartner H, Spycher BD, Swiss Paediatric Oncology G, the Swiss National Cohort Study G. Space-time clustering of childhood cancers in Switzerland: A nationwide study. *International journal of cancer* 2016;**138**: 2127-35.

38. Steliarova-Foucher E, Stiller C, Lacour B, Kaatsch P. International Classification of Childhood Cancer, third edition. *Cancer* 2005;**103**: 1457-67.

39. Diggle PJ, Chetwynd AG. Second-order analysis of spatial clustering for inhomogeneous populations. *Biometrics* 1991;**47**: 1155-63.

40. Tango T. A class of multiplicity adjusted tests for spatial clustering based on case-control point data. *Biometrics* 2007;**63**: 119-27.

41. Kulldorff M. A spatial scan statistic. *Communications in Statistics - Theory and Methods* 1997;**26**: 1481-96.

42. Waller LA, Hill EG, Rudd RA. The geography of power: Statistical performance of tests of clusters and clustering in heterogeneous populations. *Statistics in medicine* 2006;**25**: 853-65.

43. Thorpe N, Shirmohammadi A. Herbicides and nitrates in groundwater of Maryland and childhood cancers: A geographic information systems approach. *Journal of Environmental Science and Health - Part C Environmental Carcinogenesis and Ecotoxicology Reviews* 2005;**23**: 261-78.

44. Mosavi-Jarrahi A, Moini M, Mohagheghi MA, Alebouyeh M, Yazdizadeh B, Shahabian A, Nahvijo A, Alizadeh R. Clustering of childhood cancer in the inner city of Tehran metropolitan area: a GIS-based analysis. *International journal of hygiene and environmental health* 2007;**210**: 113-9.

45. Bailony MR, Hararah MK, Salhab AR, Ghannam I, Abdeen Z, Ghannam J. Cancer registration and healthcare access in West Bank, Palestine: A GIS analysis of childhood cancer, 1998–2007. *International Journal of Cancer* 2011;**129**: 1180-9.

46. Openshaw S. *The Modifiable Areal Unit Problem*ed.: Geo Books, 1984.

47. Bunch KJ, Vincent TJ, Black RJ, Pearce MS, McNally RJ, McKinney PA, Parker L, Craft AW, Murphy MF. Updated investigations of cancer excesses in individuals born or resident in the vicinity of Sellafield and Dounreay. *British journal of cancer* 2014.

48. Hoffmann W, Terschueren C, Richardson DB. Childhood leukemia in the vicinity of the Geesthacht nuclear establishments near Hamburg, Germany. *Environmental health perspectives* 2007;**115**: 947-52.

49. 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;**196**: 102-9.

50. Kreis C, Lupatsch JE, Niggli F, Egger M, Kuehni CE, Spycher BD, Swiss Paediatric Oncology G, the Swiss National Cohort Study G. Space-Time Clustering of Childhood Leukemia: Evidence of an Association with ETV6-RUNX1 (TEL-AML1) Fusion. *PLoS One* 2017;**12**: e0170020.

51. Jacquez GM. Current practices in the spatial analysis of cancer: flies in the ointment. *International journal of health geographics* 2004;**3**: 22.

Tables

Table 1: Leukaemia cases available for analysis around time of birth and diagnosis according to gender and calendar period.

	Total	Female	1985-1994	1995-2004	2005-2015
		N (%)	N (%)	N (%)	N (%)
Time of birth					
Leukaemia 0-15	1297	539 (41.6)	496 (38.2)	506 (39.0)	295 (22.7)
Leukaemia 0-4	768	333 (43.4)	251 (32.7)	278 (36.2)	239 (31.2)
Leukaemia 5-15	529	206 (38.9)	245 (46.3)	228 (43.1)	56 (10.6)
ALL 0-15	1042	439 (42.1)	399 (38.3)	402 (38.6)	241 (23.1)
ALL 0-4	627	278 (44.3)	207 (33.0)	224 (35.7)	196 (31.3)
AML 0-15	180	72 (40.0)	75 (41.7)	71 (39.4)	34 (18.9)
Time of diagnosis					
Leukaemia 0-15	1865	774 (41.5)	551 (29.5)	590 (31.6)	724 (38.8)
Leukaemia 0-4	944	405 (42.9)	293 (31.0)	287 (30.4)	364 (38.6)
Leukaemia 5-15	921	369 (40.1)	258 (28.0)	303 (32.9)	360 (39.1)
ALL 0-15	1485	621 (41.8)	440 (29.6)	465 (31.3)	580 (39.1)
ALL 0-4	782	342 (43.7)	246 (31.5)	233 (29.8)	303 (38.7)
AML 0-15	272	109 (40.1)	87 (32.0)	91 (33.5)	94 (34.6)

Abbreviations: ALL Acute Lymphoid Leukaemia, AML Acute Myeloid Leukaemia

Table 2: Results of different tests for spatial clustering of childhood leukaemia using residence at birth and at diagnosis.

Diagnostic group	Age	Difference of k-functions P^a (distance in m)	Cuzick-Edwards' test P^a (no. of NN)	Tango's Index P^a (parameter θ^b)	Kulldorff's scan statistic P^a (Radius in m)
Time of birth					
Leukaemia	0-15	0.22 (600)	0.40 (6)	0.24 (1414)	0.30 (796)
	0-4	0.75 (600)	0.50 (6)	0.72 (1414)	0.06 (493)
	5-15	0.48 (1000)	0.80 (1)	0.59 (636)	0.97 (1762)
ALL	0-15	0.11 (600)	0.06 (6)	0.11 (849)	0.75 (4330)
	0-4	0.33 (100)	0.46 (6)	0.67 (1414)	0.05 (493)
AML	0-15	0.45 (1000)	0.14 (76)	0.63 (2121)	0.29 (20,987)
					overall $P^c = 0.52$
Time of diagnosis					
Leukaemia	0-15	0.11 (250)	0.11 (1)	0.13 (636)	0.13 (438)
	0-4	0.42 (4000)	0.33 (11)	0.38 (7071)	0.07 (517)
	5-15	0.05 (100)	0.04 (1)	0.05 (354)	0.22 (3089)
ALL	0-15	0.42 (5000)	0.33 (1)	0.38 (7071)	0.10 (411)
	0-4	0.49 (5000)	0.46 (6)	0.42 (7071)	0.05 (517)
AML	0-15	0.61 (5000)	0.58 (52)	0.70 (7071)	0.88 (2698)
					overall $P^c = 0.51$

Abbreviations: ALL Acute Lymphoid Leukaemia, AML Acute Myeloid Leukaemia, NN nearest neighbours

[a] Data represent *p*-values adjusting for the multiple input parameters used (Supporting Information Table S1). We refer to these *p*-values as *adjusted P* in the manuscript. The number in parenthesis represents the input parameter for which evidence of clustering was strongest.

[b] The parameter θ determines the weights associated with a given distance between two points (see Supporting Information Text S4).

[c] *p* -value adjusted for the multiple diagnostic groups and tests performed.

Table 3: Results of tests for spatial clustering of childhood leukaemia for different calendar periods.

	Difference of k-functions P^a (distance in m)	Cuzick-Edwards' test P^a (no. of NN)	Tango's Index P^a (parameter θ^b)	Kulldorff's scan statistic P^a (radius in m)
Time of birth				
1985-1994	0.17 (1000)	0.58 (52)	0.23 (2121)	0.32 (27,119)
1995-2004	0.62 (600)	0.23 (3)	0.44 (849)	0.97 (273)
2005-2015	0.81 (450)	0.36 (1)	0.86 (1414)	0.90 (5986)
Time of diagnosis				
1985-1994	0.17 (5000)	0.29 (32)	0.12 (7071)	0.09 (3497)
1995-2004	0.04 (1000)	0.05 (1)	0.04 (2121)	0.99 (640)
2005-2015	0.50 (100)	0.06 (2)	0.35 (354)	0.05 (595)

Abbreviations: NN nearest neighbours

[a] Data represent *p*-values adjusting for the multiple input parameters used (Supporting Information Table S1). We refer to these *p*-values as *adjusted P* in the manuscript. The number in parenthesis represents the input parameter for which evidence of clustering was strongest.

[b] The parameter θ determines the weights associated with a given distance between two points (see Supporting Information Text S4).

Figure legends

Figure 1. Flow chart of the study population.

Figure 2. The difference in k-functions between cases and controls for leukaemia at age 5-15 years at time of diagnosis. The shaded area is defined by the 95% simulation envelopes based on 999 Monte Carlo samples generated by randomly permuting the case-control status conditional on the entire set of case-control locations (random labelling): At any given distance, 95% of the calculated k-function differences from the Monte Carlo samples are within the grey zone. Values of the difference within the shaded area indicate the absence of spatial clustering. The solid line represents the empirical difference and the black dashed line the mean difference in the Monte Carlo samples. For small distances, the plot is enlarged in a separate window.

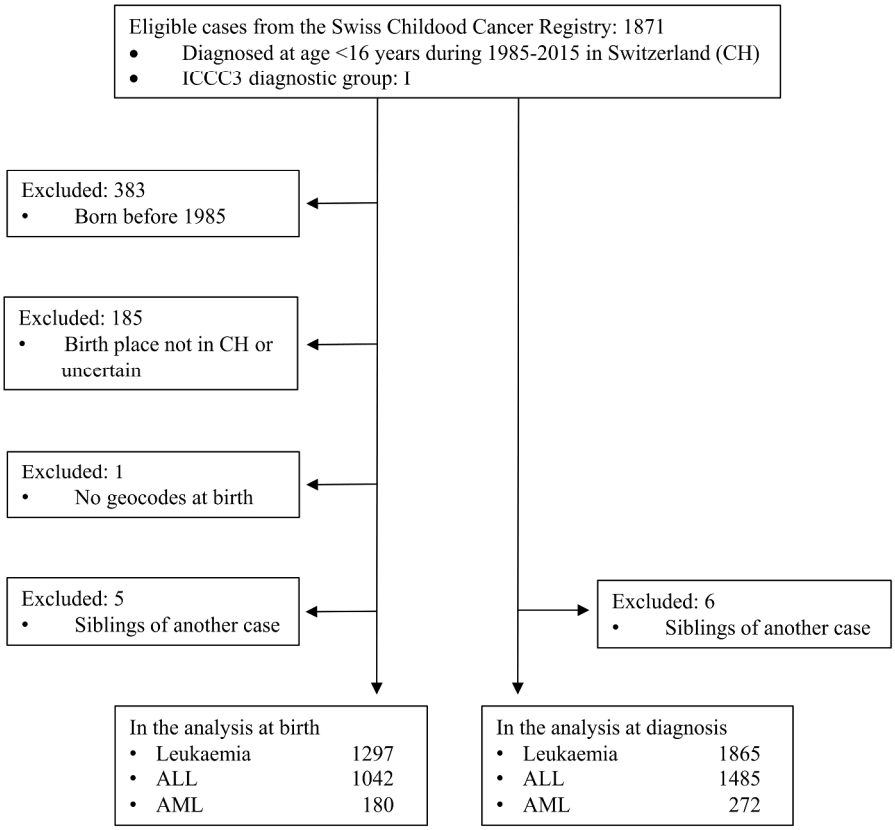

Figure 1. Flow chart of the study population.

159x142mm (600 x 600 DPI)

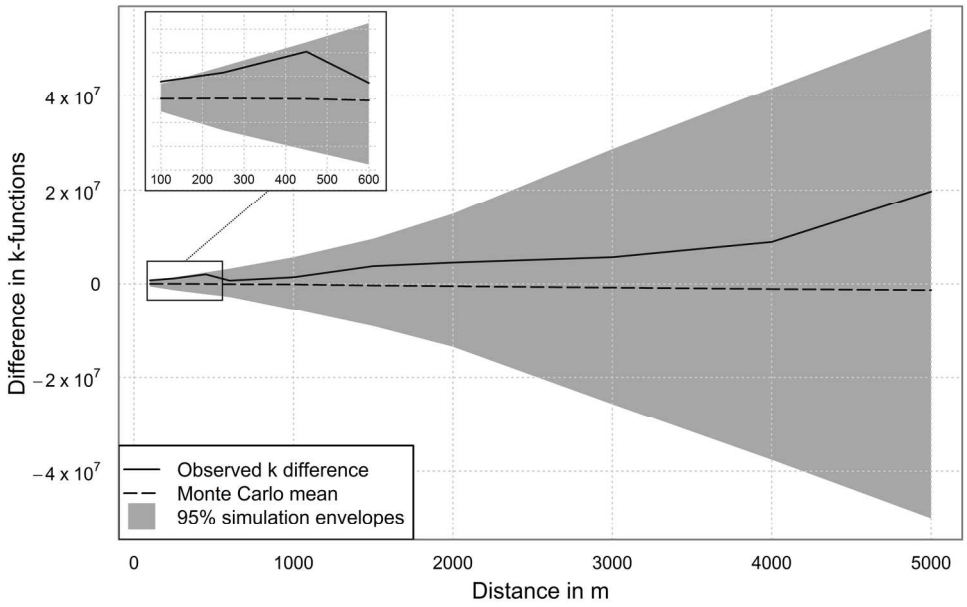

Figure 2. The difference in k-functions between cases and controls for leukaemia at age 5-15 years at time of diagnosis. The shaded area is defined by the 95% simulation envelopes based on 999 Monte Carlo samples generated by randomly permuting the case-control status conditional on the entire set of case-control locations (random labelling): At any given distance, 95% of the calculated k-function differences from the Monte Carlo samples are within the grey zone. Values of the difference within the shaded area indicate the absence of spatial clustering. The solid line represents the empirical difference and the black dashed line the mean difference in the Monte Carlo samples. For small distances, the plot is enlarged in a separate window.

90x57mm (600 x 600 DPI)