Supplementary material

Population mixing and the risk of childhood leukaemia in Switzerland:

A census-based cohort study

Judith E. Lupatsch[1], Claudia E. Kuehni[1], Felix Niggli[2], Roland A. Ammann[3], Matthias Egger[1,4], Ben D. Spycher[1] for the Swiss Paediatric Oncology Group and the Swiss National Cohort Study Group

1 Institute of Social and Preventive Medicine, University of Bern, Finkenhubelweg 11, 3012 Bern, Switzerland

2 University Children's Hospital Zurich, Steinwiesstrasse 75, 8032 Zurich, Switzerland

3 Department of Paediatrics, Freiburgstrasse 4, University of Bern, 3010, Bern, Switzerland

4 Centre for Infectious Disease Epidemiology and Research, University of Cape Town, Rondebosch, Cape Town, 7700, South Africa

Corresponding author

Ben D. Spycher

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

Email: ben.spycher@ispm.unibe.ch

Phone: +41 31 631 56 97

Fax: +41 31 631 35 20

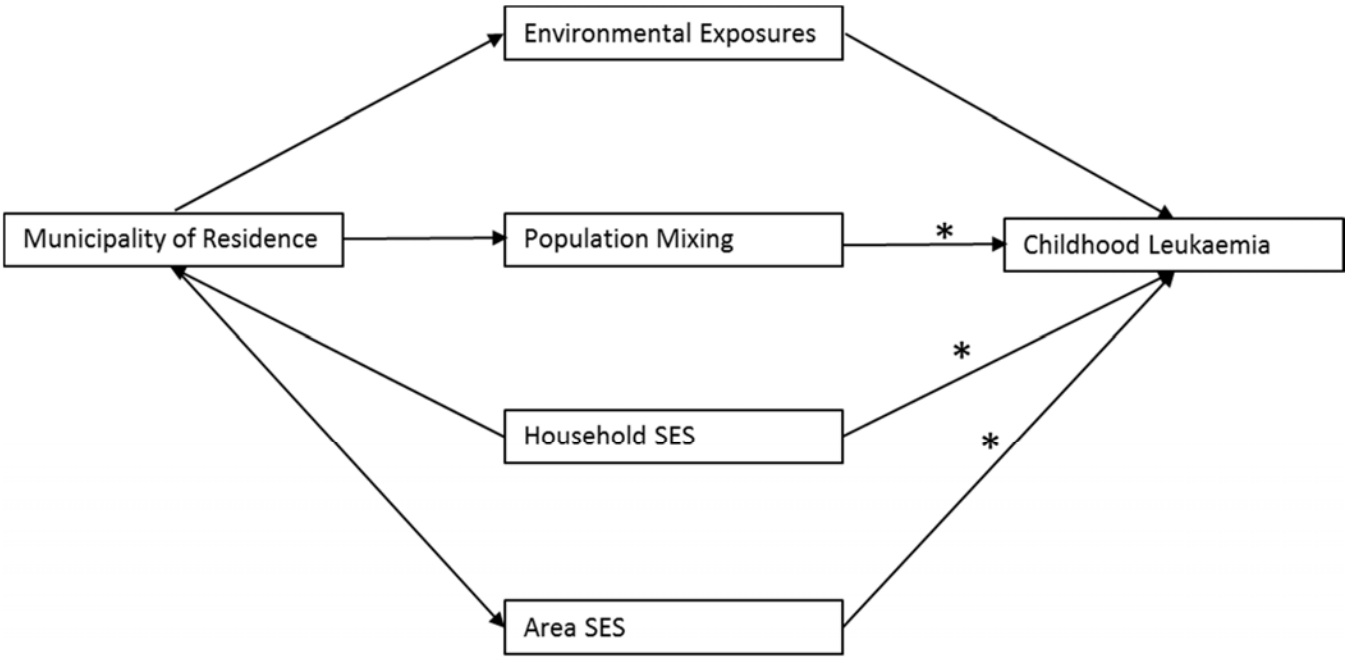

Fig. S1: Directed acyclic graph of relationships between potential confounders, population mixing and childhood leukaemia. * Effects mediated through poorly understood mechanisms possibly involving infections and immune development and responses

Table S1: Characteristics of eligible childhood leukaemia cases included and excluded from analyses

	Cases in analyses	Cases not in analyses	*p*
Total	536 (100%)	106 (100%)	
Tumor ICCC3[a] code			
(Ia) Acute lymphoblastic leukaemia	420 (78.4%)	72 (67.9%)	0.053[b]
(Ib) Acute myeloid leukaemia	82(15.3%)	28(26.4%)	
(Ic) Chronic myeloproliferative disease	9(1.7%)	2(1.9%)	
(Id) Myelodysplastic syndrome	19(3.5%)	2(1.9%)	
(Ie) Unspecified and other specified leukaemia	6(1.1%)	2(1.9%)	
Median age at diagnosis in years (IQR)	8.1 (4.7-12.2)	9.5(5.5-12.8)	0.137[c]
Female	217(40.5%)	40(37.7%)	0 .598[b]
Male	319(59.5%)	66(62.3%)	

[a] *ICCC3 International Classification of Childhood Cancer – third edition [1]*

[b] *p-value Fisher's exact test,* [c] *p-value for non-parametric equality of median test*

Text S1: Comparison of different estimators of Shannon's entropy: a Monte Carlo experiment

In a Monte Carlo experiment, we estimated the bias, variance and mean square error of various proposed estimators of Shannon's entropy. For a given vector of multinomial probabilities $p = [p_1, \dots, p_M]^T$ the Shannon entropy [2] is defined as

$$H = -\sum_{i=1}^{M} p_i \ln p_i \,. \tag{1}$$

We can use observed frequencies n_i $(i = 1, \dots, M)$ to estimate $H(p)$. Simply replacing probabilities by the observed relative frequencies $\hat{p} = [\hat{p}_1, \dots, \hat{p}_M]^T$, where $\hat{p}_i = n_i/N$ and $N = \sum_{i=1}^{M} n_i$, results in the naive estimator

$$\widehat{H}_n = -\sum_{i=1}^{M} \hat{p}_i \ln \hat{p}_i \tag{2}$$

which is known to be biased. The expected value of $\widehat{H}_n$ can be approximated as [3]

$$E\left[\widehat{H}_n\right] = H - \frac{M-1}{2N} + \frac{1}{12N^2}\left(1 - \sum_{i=1}^{M} p_i^{-1}\right) + \frac{1}{12N^3}\sum_{i=1}^{M}\left(p_i^{-1} - p_i^{-2}\right) + \dots$$

Based on this approximation, Law et al. [4] used the estimator

$$\widehat{H}_L = \widehat{H}_n + \frac{M-1}{2N} - \frac{1}{12N^2}\left(1 - \sum_{i=1}^{M} \hat{p}_i^{-1}\right) - \frac{1}{12N^3}\sum_{i=1}^{M}\left(\hat{p}_i^{-1} - \hat{p}_i^{-2}\right). \tag{3}$$

However, this is inappropriate as the probabilities are again simply replaced by relative frequencies [5]. As alternatives to $\widehat{H}_N$, we evaluated a set of estimators proposed in [6] and [5], namely

$$\widehat{H}_\psi = \sum_{i=1}^{M} \hat{p}_i \left(\ln N - \psi(n_i) - \frac{(-1)^{n_i}}{n_i(n_i+1)}\right), \tag{4}$$

where $\psi(n)$ is the digamma function, and

$$\widehat{H}_S(\xi) = \sum_{i=1}^{M} \hat{p}_i \left(\psi(N) - \psi(n_i) - (-1)^{n_i} \int_0^{1/\xi - 1} \frac{t^{n_i - 1}}{1+t}\, dt\right), \tag{5}$$

with ξ taking on values 0.5, exp(-0.5), and 1.

The sampling distribution for the Monte Carlo experiment was based on 5-year in-migration data from the census 2000. For each Swiss municipality, we calculated a vector of relative frequencies with length M equal to the number of districts in Switzerland. Values represented the relative frequency of migrants coming from a particular district as a proportion of total in-migration into the municipality and were ordered according to size such that first elements represented the most important districts of origin for that municipality. We fitted a Dirichlet distribution to these frequency vectors using the 'dirmult' command of the R package 'dirmult'. We then randomly selected 100 municipalities and, for each of these, randomly sampled a vector of multinomial probabilities from the fitted Dirichlet distribution p_j $(j = 1, \dots, 100)$. For each selected municipality j, we then drew 100 random frequency vectors n_{jk} $(k = 1, \dots, 100)$ from a multinomial distribution with parameters N_j and p_j where N_j is the total number of in-migrants into the municipality. We calculated entropy estimators (equations 2-5) for each of the 100 frequency vectors and compared them with actual entropy (equation 1) based on the true probability vector p_j.

The observed bias, variance, and mean square error for estimators 2, 4, and 5 are plotted against in-migration levels N_j in *Figure S1*. Results for the estimator 3 are not shown as it performed systematically worse than the other estimators. For all estimators bias and variance were larger for small levels of in-migration with a steep increase of mean square error for communities with <100 in-migrants. Both bias and variance tended to be smaller for the naïve estimator than for the alternative estimators. Average mean square error over the 100 replications was 1.432, 3.050, 3.953, 3.047, 3.382 for the estimators $\widehat{H}_n$, $\widehat{H}_\psi$, $\widehat{H}_S(0.5)$, $\widehat{H}_S(\exp(-0.5))$, and $\widehat{H}_S(1)$ respectively. For comparison, actual entropies ranged from 2.345 to 2.586 over the 100 replications with a mean of 2.467.

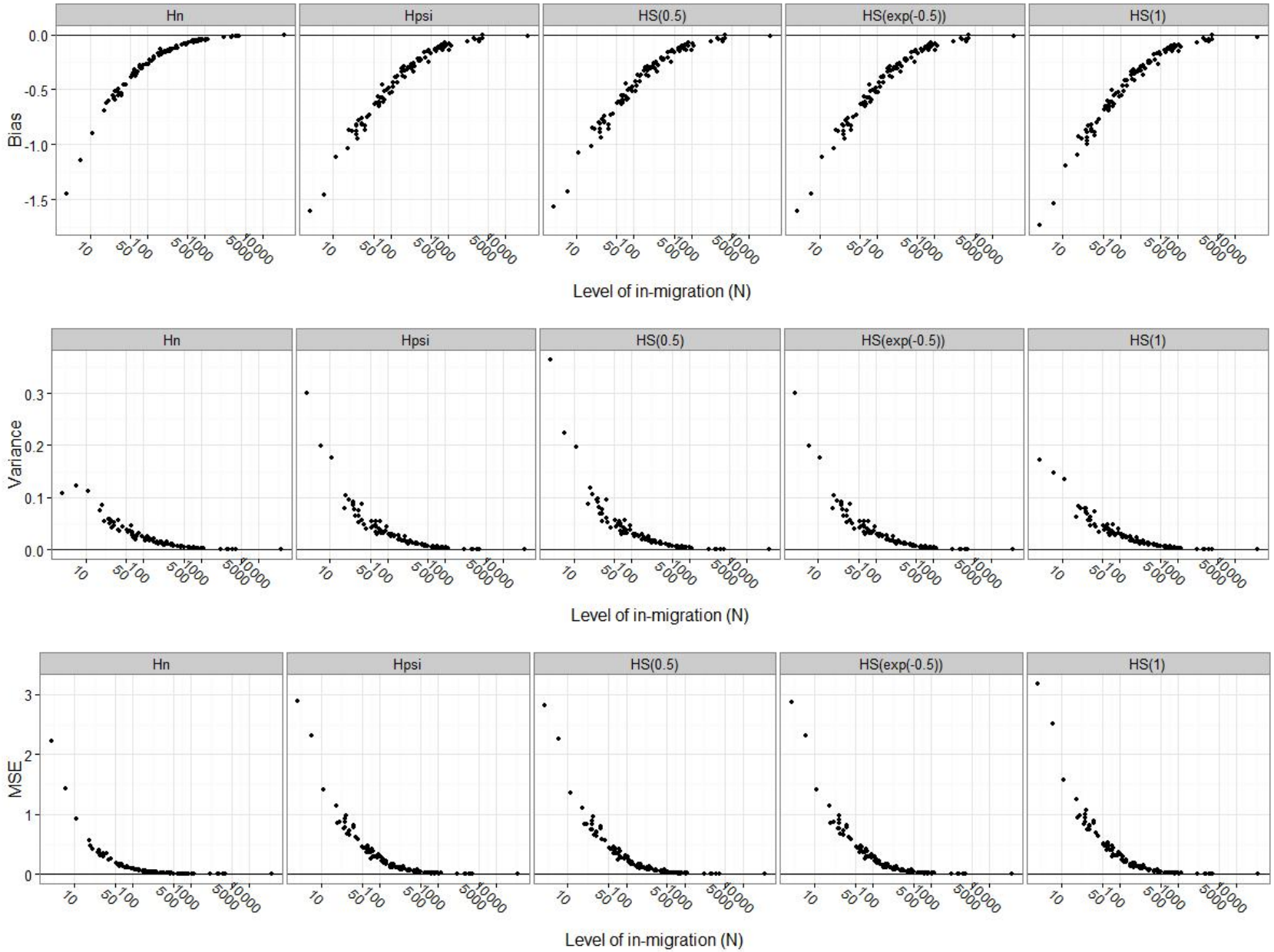

Fig. S2: Bias, variance, and mean square error (MSE) of estimated diversity of migrant origin using different estimators of Shannon's entropy: Results from a Monte Carlo for Swiss municipalities. Formulas of estimators are given in equations 2, 4, and 5

Table S2. Time to event analyses with exposures as continuous variables (any leukaemia <16 years)

Outcome	All municipalities		Rural municipalities		Urban municipalities		Test for interaction	
	Crude HR (95% CI)	Adjusted[a] HR (95% CI)	Crude HR (95% CI)	Adjusted[a] HR (95% CI)	Crude HR (95% CI)	Adjusted[a] HR (95% CI)	Crude	Adjusted[a]
Population Growth	0.44 (0.12 - 1.59)	0.25 (0.07 - 0.97)	1.52 (0.16 - 14.83)	1.45 (0.15 - 14.41)	0.27 (0.05 - 1.33)	0.90 (0.64 - 1.28)	0.234	0.079
Level of in-migration	0.38 (0.09 - 1.66)	0.26 (0.05 - 1.20)	0.20 (0.02 - 2.63)	0.24 (0.01 - 4.19)	0.34 (0.05 - 2.23)	0.19 (0.03 - 1.38)	0.234	0.079
Diversity of origin	1.20 (0.99 - 1.45)	1.18 (0.97 - 1.43)	1.35 (0.89 - 2.04)	1.40 (0.91 - 2.14)	1.21 (0.97 - 1.51)	1.13 (0.90 - 1.41)	0.766	0.877

Hazard ratio (HR) per increase in population mixing by 1 unit, i.e. by 100% increase for population growth and level of in-migration

[a] adjusted for level of education of head of household, flat rent, crowding, the Swiss neighbourhood index of socioeconomic position, distance to highways, ionising background radiation (dose rates from terrestrial gamma and cosmic radiation), electromagnetic fields from broadcast transmitters, distance to high voltage power lines

Table S3. Sensitivity Analysis for diversity of origin excluding municipalities with less than a 100 in-migrants (any leukaemia <16 years)

Outcome	Diversity of origin (quintiles)	All municipalities		Rural municipalities		Urban municipalities		Test for interaction	
		Crude HR (95% CI)	Adjusted[a] HR (95% CI)	Crude HR (95% CI)	Adjusted[a] HR (95% CI)	Crude HR (95% CI)	Adjusted[a] HR (95% CI)	Crude	Adjusted[a]
All leukaemia, <16 y	1	1	1	1	1	1	1	0.948	0.949
	2	0.95 (0.70 - 1.28)	0.92 (0.68 - 1.25)	0.84 (0.47 - 1.50)	0.84 (0.47 - 1.50)	1.01 (0.71 - 1.44)	0.94 (0.66 - 1.35)		
	3	1.01 (0.76 - 1.36)	0.94 (0.70 - 1.27)	0.91 (0.51 - 1.63)	0.86 (0.47 - 1.56)	1.08 (0.77 - 1.52)	0.97 (0.69 - 1.36)		
	4	1.17 (0.88 - 1.55)	1.13 (0.84 - 1.50)	1.26 (0.72 - 2.19)	1.30 (0.74 - 2.30)	1.19 (0.85 - 1.66)	1.08 (0.77 - 1.51)		
	5	1.19 (0.88 - 1.61)	1.13 (0.83 - 1.54)	1.24 (0.65 - 2.34)	1.27 (0.66 - 2.45)	1.21 (0.86 - 1.72)	1.09 (0.76 - 1.55)		

[a] adjusted for level of education of head of household, flat rent, crowding, the Swiss neighbourhood index of socioeconomic position, distance to highways, ionising background radiation (dose rates from terrestrial gamma and cosmic radiation), electromagnetic fields from broadcast transmitters, distance to high voltage power lines

References

1. Steliarova-Foucher E, Stiller C, Lacour B, Kaatsch P. International Classification of Childhood Cancer, third edition. Cancer. 2005;103:1457-67.
2. Shannon CE. A Mathematical Theory of Communication. At&T Tech J. 1948;27:379-423.
3. Peet RK. The Measurement of Species Diversity. Annu Rev Ecol Syst. 1974;5:285-307.
4. Law GR, Parslow RC, Roman E, United Kingdom Childhood Cancer Study I. Childhood cancer and population mixing. Am J Epidemiol. 2003;158:328-36.
5. Schurmann T. Bias analysis in entropy estimation. J Phys a-Math Gen. 2004;37:L295-L301.
6. Grassberger P. Entropy estimates from insufficient samplings. arXiv:physics/0307138v2. 2008.