Supplementary methods.

Data Sources: The Brecon Group Register is a longitudinal cohort of children and young people identified with diabetes and resident within Wales.[1] The Register includes patient name, date of birth, date and location of diagnosis, family history of diabetes, and home address. This Register, collected since 1995, has been linked to routine administrative data within the Secure Anonymised Information Linkage (SAIL) Databank, held by the Health Information Research Unit at Swansea University, UK. Unaffected population controls were selected from individuals who lived in the surrounding area and were also recorded in the SAIL databank.[1,2]

Inclusion Criteria: Patients with T1D and matched controls were required to be resident (defined by the Administrative Register) [3] in Wales at the time of the subject's diagnosis, and individuals with evidence of having diabetes (consisting of a hospital admission for any type of diabetes) were excluded from the control population.

Admissions Definition: Admissions were counted from 30 days after the diagnosis of T1D, in order to exclude admissions related to the events surrounding diagnosis, until the end of May 2012 and compared to the amount of time resident in Wales according to NHS records, to calculate an admission rate per person year of follow up in children and young adults resident in Wales. Hospital admissions, for all causes, were defined using the Patient Episode Database for Wales (PEDW), which records all episodes of inpatient and daycase activity in NHS Wales hospitals, including planned and emergency admissions, minor and major operations and hospital stays for giving birth.[4] Hospital activity for Welsh residents treated in hospitals in England is also included. Data are collected and coded at each hospital to optimise coverage and then transferred to the NHS Wales Informatics Service for validation and merging into the main database. Multiple admissions on a single day were counted as one admission for the purposes of this study.

Multi-level Poisson Modelling approach: The matched cohort design of the study was incorporated into the analysis by the introduction of a variance parameter. Violation of the Poisson model assumptions, i.e. the mean of admission rates is not equal to the variance, was tested by addition of a hyper variance parameter.[5] The robustness of the model estimates using approximate likelihood methods was compared to estimates from Markov Chain Monte Carlo (MCMC) algorithm.[6,7] Effective sample size (ESS) is also reported to indicate the amount of chain mixing, and results are reported as incident rate ratios (IRR) comparing cases to controls. 95% credible intervals and p-values are also reported. An IRR of 1.2 would represent a 20% difference in rate of admissions between cases and controls.

Supplementary Table 1. Variable Provenance

Variable	Source	Notes
No of Admissions 30 days after diagnosis	Patient Episode Database for Wales	Multiple admissions for an individual on the same day are only counted as a single admission. All types of inpatient admissions, for any reason, are included. Welsh residents treated in English border hospitals are also included.
Date of diagnosis with T1D	Brecon group register	Cases Only.
Follow-up time	NHS Administrative Register	The total number of days from 30 days after diagnosis until 31-05-2012 for which the individual was listed as resident in Wales.
Week of birth	NHS Administrative Register	Perfect agreement with Brecon Group register.
Gender	NHS Administrative Register	Perfect agreement with Brecon Group register.
County of residence at diagnosis	NHS Administrative Register	
Follow-up time in county	NHS Administrative Register	Total number of days from 30 days after diagnosis until 2012-05-31 for which the individual was resident in the same county as at the time of diagnosis.
Socio economic status	Welsh Index of Multiple Deprivation 2008	SES defined by location of residence at the time of the subject's diagnosis with T1D. Residences were defined into groups using the Office for National Statistics lower super output area definition. Residence is obtained from the NHS Administrative Register.
Urban rural status	The Office of National Statistics	Urban defined as a population greater than 10,000 individuals per super output area.
Treatment centre size	Derived	Cardiff, Newport, and Swansea were defined as large centres, each diagnosing over 10% of individuals in the Brecon group register. Remaining centres were defined as small.
Continuous residency	Derived	Continuous residency in the study period. Based on NHS Administrative Register.
Border county	Derived	Monmouthshire, Powys, Wrexham, and Flintshire border England.
Age at diagnosis with T1D	Derived	Derived from week of birth and date of diagnosis.

Supplementary Table 2. Incidence rate ratio of hospital admissions between individuals with type I diabetes compared to normal population controls using different model specification and estimation.

						Variances			
	Model	IRR (SE)	(95% CI)	P<=	ESS	Level	Variance(SE)	(95% CI)	ESS
Unmatched	1	4.738 (1.029)	(4.569, 4.569)	0.0001					
Matched	2	4.786 (1.030)	(4.610, 4.610)	0.0001		1	1.045 (0.046)	(0.956, 1.135)	
Over dispersed	3	5.397 (1.325)	(3.890, 3.890)	0.0001		1	30.628 (0.547)	(29.555, 31.700)	
						2	0.117 (0.229)	(-0.331, 0.565)	
MCMC Over dispersed	4	5.742 (1.078)	(5.282, 6.254)	0.0001	7255	1	1.217 (0.049)	(1.125, 1.317)	6572
DIC=19246.96						2	0.108 (0.025)	(0.060, 0.158)	1499

Model 1 is a single level Poisson model comparing the rates of hospital admission between individuals with type I diabetes compared to normal population controls estimated using quasi likelihood approach. Model 2 is multi-level Poisson model which accounts for the matched design using a 2 level variance component model, estimated using quasi likelihood approach. Model 3 is a multi-level Poisson model which accounts for the matched design and over-dispersion using a 3 level variance component model, estimated using quasi likelihood. Results from models 1 to 3 are reported using maximum quasi-likelihood Incident Rate Ratio, asymptotic standard errors (SE), 95 % confidence intervals and two-sided p-values. Model 4 is the same as model 3, except estimated using MCMC. Results are reported using the mean of the posterior distribution to indicate Incident Rate Ratio, the standard deviation of the posterior chain is used to indicate the parameter standard error, 95% posterior probability intervals (95%CI) represent the 2.5^{th} and 97.5^{th} percentiles of the posterior distribution, and directional posterior probabilities (P<=). Effective Sample Size (ESS) indicates the effectiveness of MCMC chain mixing. Bayesian Deviance Information Criterion (DIC) is used to indicate model fit. Data are restricted to individuals with continuous residency, and not living in a county bordering England.

Supplementary Table 3. Incidence rate ratio of hospital admissions between individuals with type I diabetes compared to normal population controls at different levels of exposures of interest.

Model			Parameter Estimates					Variances					
(DIC)	Description	Category	IRR	(SE)	(95% CI)	P<=	ESS	Level	Variance	(SE)	(95% CI)		ESS
1		Male (Ref)	5.618	(1.126)	(4.969, 6.375)	0.0001	2091	1	~	~	~	~	~
(19344)	Sex	SEX	1.116	(1.011)	(0.933, 1.338)	0.1158	2196	2	1.411	(0.060)	(1.308, 1.520)		3504
								3	0.104	(0.070)	(0.055, 0.159)		1570
2		SES 1 (Ref)	5.464	(1.213)	(4.351, 6.792)	0.0001	813	1	~	~	~	~	~
(19321)		SES 2	1.01	(1.002)	(0.744, 1.378)	0.4739	1396	2	1.424	(0.055)	(1.320, 1.537)		6016
	Social Economic Status	SES 3	1.198	(1.028)	(0.895, 1.619)	0.1147	1249	3	0.093	(0.027)	(0.042, 0.147)		935
		SES 4	1.018	(1.003)	(0.764, 1.363)	0.4519	1223						
		SES 5	1.196	(1.026)	(0.901, 1.585)	0.107	1094						
3		Small (Ref)	6.43	(1.119)	(5.716, 7.249)	0.0001	3135	1	~	~	~	~	~
(19337)	Centre Size	Large	0.831	(0.983)	(0.691, 0.996)	0.0227	3278	2	1.417	(0.054)	(1.315, 1.526)		8269
								3	0.122	(0.027)	(0.072, 0.178)		1730
4		<5 (Ref)	8.045	(1.224)	(6.725, 9.668)	0.0001	1181	1	~	~	~	~	~
(19328)	Age at Diagnosis (yrs)	5 to 10	0.78	(0.967)	(0.619, 0.984)	0.0184	1274	2	1.412	(0.131)	(1.309, 1.525)		731
		10 +	0.572	(0.931)	(0.455, 0.721)	0.001	1340	3	0.115	(2.489)	(0.065, 0.174)		476
5		Rural (Ref)	5.652	(3.638)	(4.765, 6.628)	0.0067	197	1	~	~	~	~	~
(19420)	Urban Rural	Urban	1.075	(1.080)	(0.888, 1.322)	0.2275	186	2	1.419	(5.003)	(1.315, 1.537)		273
								3	0.122	(7.162)	(0.069, 0.183)		215
6		Case	5.949	(1.086)	(5.438, 6.518)	0.0001	6451	1	~	~	~	~	~
(19330)	Re: County							2	1.421	(0.054)	(1.318, 1.531)		7494
								3	0.112	(0.027)	(0.062, 0.169)		1432
								4	0.012	(0.232)	(0.002, 0.043)		229
7		<90 (Ref)	4.477	(1.174)	(3.628, 5.520)	0.0001	904	1	~	~	~	~	~
(19303)		90 to 94	1.131	(1.017)	(0.868, 1.463)	0.1782	1173	2	1.419	(0.054)	(1.316, 1.529)		6741
	Date of Birth	95 to 99	1.595	(1.064)	(1.232, 2.062)	0.0002	1221	3	0.103	(0.026)	(0.053, 0.156)		847
		00 to 04	1.785	(1.098)	(1.299, 2.445)	0.0001	1853						
		>05	1.682	(1.151)	(0.993, 2.863)	0.0265	5026						

All models use a multi-level Poisson model. Results are reported using the mean of the posterior distribution to indicate Incident Rate Ratio, the standard deviation of the posterior chain is used to indicate the parameter standard error, 95% posterior probability intervals (95%CI) represent the 2.5th and 97.5th percentiles of the posterior distribution, and directional posterior probabilities (P<=). Effective Sample Size (ESS) indicates the effectiveness of MCMC chain mixing. Bayesian Deviance Information Criterion (DIC) is used to indicate model fit. The reference category comparing the IRR between cases and controls is indicated, and the IRR of the exposure of interest is represented by the interaction between the exposure of interest and case control status. Level 2 variance indicates the over-dispersion parameter, level 3 variance indicates the matching criteria, and level 4 (Model 6) variance indicates the county level variance. Data are restricted to individuals with continuous residency, and not living in a county bordering England.

Supplementary Table 4. Incidence rate ratio of hospital admissions between individuals with type I diabetes compared to normal population controls at different levels of exposures of interest whilst adjusting for significant variables explored in Table 3.

Model (DIC)		Category	Parameter Estimates IRR	(SE)	(95% CI)	P<=	ESS	Variances Level	Variance(SE)	(95% CI)	ESS
	Baseline (SEP 1, Small, <1990, <5)		7.260	(1.538)	(4.857, 11.133)	0.0001	302	1	~ ~	~ ~	~
	Social Economic Status	SES 2	0.951	(0.992)	(0.691, 1.317)	0.3814	1348	2	1.429 (0.055)	(1.323, 1.538)	2605
		SES 3	1.140	(1.021)	(0.836, 1.560)	0.2030	1184	3	0.063 (0.025)	(0.014, 0.114)	325
		SES 4	0.970	(0.995)	(0.725, 1.316)	0.4193	1086	4	0.009 (11.488)	(0.001, 41.303)	147
		SES 5	1.176	(1.024)	(0.883, 1.583)	0.1332	1096				
(19276)	Centre Size	Large	0.840	(0.984)	(0.698, 1.010)	0.0319	2945				
	Date of Birth	90 to 94	1.011	(1.002)	(0.771, 1.321)	0.4686	747				
		95 to 99	1.225	(1.032)	(0.903, 1.666)	0.1001	520				
		00 to 04	1.179	(1.034)	(0.795, 1.758)	0.2080	498				
		>05	1.097	(1.029)	(0.604, 1.992)	0.3827	1392				
	Age at Diagnosis (yrs)	Age	0.786	(0.982)	(0.682, 0.910)	0.0007	560				

Results are from a multi-variable multi-level models estimated using MCMC. Results are reported using the mean of the posterior distribution to indicate Incident Rate Ratio, the standard deviation of the posterior chain is used to indicate the parameter standard error, 95% posterior probability intervals (95%CI) represent the 2.5th and 97.5th percentiles of the posterior distribution, and directional posterior probabilities (P<=). Effective Sample Size (ESS) indicates the effectiveness of MCMC chain mixing. Bayesian Deviance Information Criterion (DIC) is used to indicate model fit. The reference category comparing the IRR between cases and controls in the highest social class, small centres, born prior to 1990 and with an age at diagnosis less than 5 years is indicated, and the IRR of the exposure of interest is represented by an interaction with case control status. The level 2 variance indicates the over dispersion parameter, the level 3 variance indicates the matched set, and the level 4 variance indicates the county at diagnosis. Age is modelled as a linear change per 5 year increase in the age at diagnosis. Data are restricted to individuals with continuous residency, and not living in a county bordering England.

References

1. Ford DV, Jones KH, Verplancke JP, et al. The SAIL Databank: building a national architecture for e-health research and evaluation. BMC health services research. 2009;9:157.
2. Lyons RA, Jones KH, John G, et al. The SAIL databank: linking multiple health and social care datasets. BMC medical informatics and decision making. 2009;9:3.
3. John G. NHS Administrative Register. Cardiff: NHS Wales Informatics Service (UK). http://www.wales.nhs.uk/sitesplus/documents/888/Gareth%20John.pdf Accessed 10 July 2013.
4. SAIL Databank (UK) SU. Patient Episode Database for Wales (PEDW). http://www.saildatabank.com/data-dictionary/sail-datasets/patient-episode-database-for-wales-(pedw).aspx. Accessed 10 July 2013.
5. Gelman A, Hill J. Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge: Cambridge University Press; 2009:335.
6. Browne WJ. MCMC Estimation in MLwiN v2.1. Centre for Multilevel Modelling, University of Bristol. 2009.
7. Browne WJ, Draper D. A comparison of Bayesian and likelihood-based methods for fitting multilevel models. Bayesian Anal. 2006;1(3):473-513.