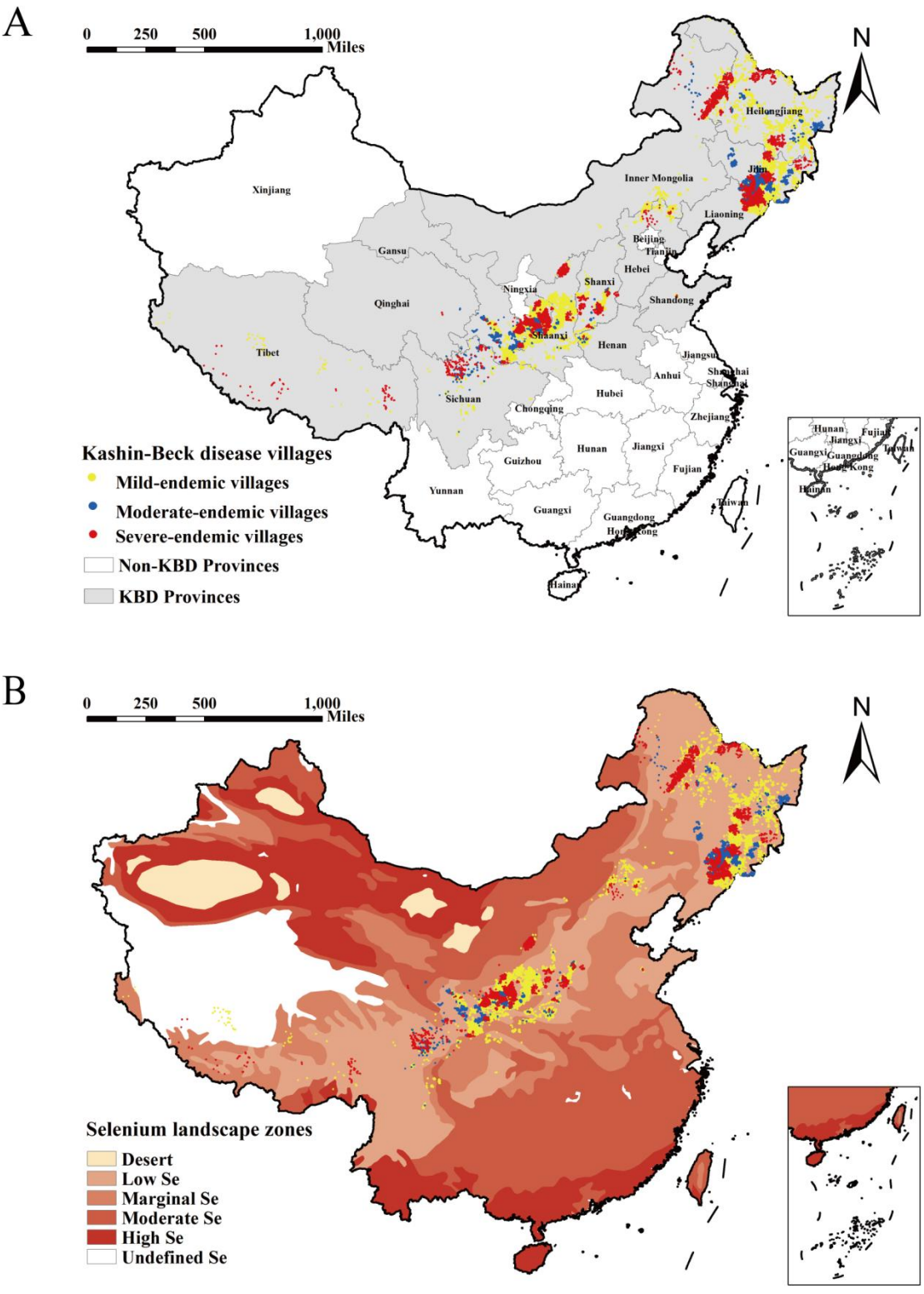


Supplementary Figure S1. Distribution of Kashin-Beck disease endemic villages in China. (A) Distribution of endemic villages within the affected provinces. (B) Overlay of endemic villages with selenium landscape partitions (review No. GS(2024)0650).

Supplementary Table S1. Criteria for designating an area as a Kashin-Beck disease (KBD) endemic area[1].

Criteria	Description
Clinical prevalence	A local resident clinical prevalence of Grade I or higher exceeding 5%.
Radiographic evidence	The presence of cases in children aged 7–12 years with multiple, symmetrical skeletal changes on hand radiographs.

Note. An area was classified as endemic if it met both of the above conditions.

Supplementary Table S2. Classification criteria for the severity of KBD endemic areas[1].

Endemic area type	Classification criteria
Mild	Local resident clinical prevalence of Grade I or higher, or radiographic detection rate among children aged 7–12 years ≤ 10%.
Moderate	Local resident clinical prevalence of Grade I or higher or radiographic detection rate among children aged 7–12 years > 10% and ≤ 20%.
Severe	Local resident clinical prevalence of Grade I or higher or radiographic detection rate among children aged 7–12 years > 20%.

Supplementary Table S3. Description and sources of predictor variables used in the Bayesian multilevel ordinal logistic regression model

Variables	Description	Type	Assignment / Units	Data source	Resolution	Temporal coverage
Outcome variables						
KBD villages	KBD endemic severity level of the village	Ordinal Categorical	0 = Control, 1 = Mild endemic village, 2 = Moderate endemic village, 3 = Severe endemic village	National KBD Surveillance System	-	-
Predictor variables						
GDP	Gross domestic oroduct density	Continuous	10,000 yuan/km²	Resource and Environmental Science Data Platform (https://www.resdc.cn/)	1 km	1995-2000
Rail	Euclidean distance to the nearest railway	Continuous	km	National Basic Geographic Information Website	1:1,000,000	1980-2000
Altitude	Elevation above sea level	Continuous	m	Resource and Environmental Science Data Platform (https://www.resdc.cn/)	250 m	2003
Slope	Terrain slope	Continuous	%	ArcGIS 10.8.1	250 m	2003

Bio1	Annual mean temperature	Continuous	℃	WorldClim(https://worldclim.org/)	1 km	1970-2000
Bio12	Annual precipitation	Continuous	mm	WorldClim(https://worldclim.org/)	1 km	1970-2000
RH	Relative humidity	Continuous	%	High-resolution Gridded Dataset (https://crudata.uea.ac.uk/cru/data/hrg/)	1 km	1980-2000
Bio3	Isothermality	Continuous	%	WorldClim(https://worldclim.org/)	1 km	1970-2000
T_GRAVEL	Gravel content	Continuous	%vol	National Cryosphere Desert Data Center (http://www.ncdc.ac.cn)	1:1,000,000	2009
T_SAND	Sand content	Continuous	%wt	National Cryosphere Desert Data Center (http://www.ncdc.ac.cn)	1:1,000,000	2009
T_SILT	Silt content	Continuous	%wt	National Cryosphere Desert Data Center (http://www.ncdc.ac.cn)	1:1,000,000	2009
Porosity	Soil porosity	Continuous	%	National Tibetan Plateau/Third Pole Environment Data Center (http://data.tpdc.ac.cn)	1:1,000,000	2013
T_ESP	Exchangeable sodium percentage	Continuous	%	National Cryosphere Desert Data Center (http://www.ncdc.ac.cn)	1:1,000,000	2013

T_ECE	Soil electrical conductivity	Continuous	dS/m	National Cryosphere Desert Data Center (http://www.ncdc.ac.cn)	1:1,000,000	2013
PH	Soil pH	Continuous	-log(H+)	National Tibetan Plateau/Third Pole Environment Data Center (http://data.tpdc.ac.cn)	1:1,000,000	2013
AN	Soil alkali-hydrolysable nitrogen	Continuous	mg/kg	National Tibetan Plateau/Third Pole Environment Data Center (http://data.tpdc.ac.cn)	1:1,000,000	2013
Selenium	Selenium landscape zone	categorical	1 = Low Se, 2 = Marginal Se, 3 = Moderately-high Se	National Geographic Resource Science Subcenter, National Earth System Science Data Center, National Science & Technology Infrastructure of China (http://gre.geodata.cn)	1:1,000,000	1989

Supplementary Table S4. Thresholds for classifying selenium content levels in geo-ecosystems (μg/g)

Level	Total Se in topsoil	Water-soluble Se in topsoil	Se in cereal grains	Se in children's hair
Deficient	< 0.125	< 0.003	< 0.025	< 0.200
Marginal	0.125–0.175	0.003–0.006	0.025–0.040	0.200–0.250
Moderate	0.175–0.400	0.006–0.008	0.040 - 0.070	0.250–0.500
High	0.400–3.00	0.008–0.020	0.070–1.0	0.500–3.0
Excessive	≥ 3.00	≥ 0.020	≥ 1.0	≥ 3.0

Note. The selenium zoning data used in this study were sourced from a vector map provided by the National Earth System Science Data Center (developed by the Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences), with the original data extracted from the *Endemic Disease and Environment Atlas of the People's Republic of China*. This classification system is based on a comprehensive survey covering 27 provinces (autonomous regions and municipalities directly under the Central Government) of China. By measuring selenium concentrations in soil profiles, grain, and human hair samples, the environmental thresholds for low-selenium areas were defined in accordance with geographical zoning theories, ultimately forming an eco-landscape map reflecting the spatial distribution of selenium in topsoil, grain, and human hair across China[2-6]. The categories adopted in this study (non-low selenium, low selenium, marginal selenium, and moderately high selenium zones) are established based on this framework, and the selenium zoning attributes of each village were obtained by spatially matching village locations with the selenium zoning vector map.

Supplementary Table S5. Baseline characteristics and univariate analysis

Variables	Total	Non-endemic village	KBD endemic village			Statistic	*P*
			Mild	Moderate	Severe		
Altitude (m)	882.55±758.89	814.51 ± 790.82	929.19±616.89	972.09±867.40	980.72±811.99	*F*=94.95	<0.01
Slope (%)	3.54 ± 4.60	2.59 ± 4.01	4.78 ± 4.98	4.17 ± 4.97	4.14 ± 4.82	*F*=524.42	<0.01
Bio1 (℃)	7.61 ± 3.93	8.10 ± 4.05	7.88 ± 3.89	6.12 ± 3.22	6.17 ± 3.35	*F*=448.98	<0.01
Bio12 (mm)	579.21± 172.46	539.60±201.14	611.41± 129.58	659.01 ± 95.76	613.03±128.80	*F*=669.80	<0.01
RH(%)	63.72 ± 7.22	63.31 ± 7.59	63.46 ± 7.13	64.92 ± 7.49	64.99 ± 5.61	*F*=94.55	<0.01
Bio3(%)	28.06 ± 4.11	27.95 ± 4.25	28.26 ± 3.46	27.88 ± 4.77	28.15 ± 4.37	*F*=13.34	<0.01
T_GRAVEL (%vol)	6.61 ± 4.06	6.93 ± 4.22	6.57 ± 4.01	6.15 ± 3.44	5.81 ± 3.72	*F*=108.55	<0.01
T_SAND(%wt)	41.22 ± 14.43	41.45 ± 16.18	41.49 ± 12.92	38.81 ± 10.25	41.10 ± 12.39	*F*=25.08	<0.01
T_SILT(%wt)	38.04 ± 10.63	37.72 ± 11.67	38.04 ± 9.84	40.06 ± 8.09	38.12 ± 9.28	*F*=34.18	<0.01
Porosity(%)	1.24 ± 0.10	1.25 ± 0.09	1.23 ± 0.10	1.20 ± 0.10	1.21 ± 0.10	*F*=364.36	<0.01
T_ESP(%)	1.82 ± 2.90	2.17 ± 3.76	1.53 ± 1.51	1.23 ± 1.77	1.48 ± 1.51	*F*=170.92	<0.01
T_ECE(dS/m)	0.35 ± 1.43	0.45 ± 1.94	0.26 ± 0.37	0.23 ± 0.83	0.24 ± 0.59	*F*=55.90	<0.01
PH(-log(H+))	7.49 ± 0.92	7.63 ± 0.85	7.46 ± 0.92	7.00 ± 1.02	7.30 ± 0.97	*F*=446.66	<0.01
AN (mg/kg)	110.13 ± 91.81	91.54 ± 65.87	119.51± 106.51	161.90±114.15	129.83±107.02	*F*=619.97	<0.01

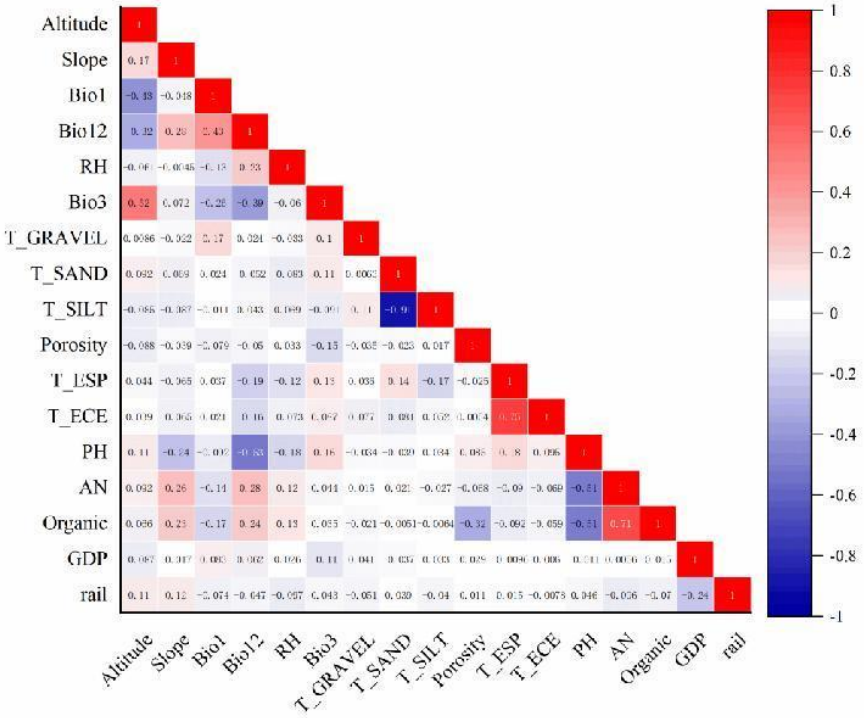
Organic (g/100g)	2.60 ± 2.52	2.16 ± 1.91	2.71 ± 2.77	3.89 ± 3.24	3.26 ± 3.01	F=516.85	<0.01
GDP (10,000 yuan/km²)	1038.91± 3440.61	1191.82 ± 3759.51	1028.79 ± 3056.47	460.76 ± 1802.92	824.48 ± 3570.23	F=39.94	<0.01
Rail (km)	56.62 ± 99.00	49.06 ± 93.86	59.65 ± 78.06	60.14 ± 72.19	74.79 ± 148.18	F=91.28	<0.01
Se, n (%)						χ^2=5882.28	<0.01
Low	25,156 (76.18)	9,654 (58.47)	8,627 (93.26)	2,356 (98.08)	4,519 (92.98)		
Marginal	5,437 (16.46)	4,608 (27.91)	542 (5.86)	43 9 (1.79)	244 (5.02)		
Mod-high	2,387 (7.23)	2,238 (13.56)	53 (0.57)	3 (0.12)	93 (1.91)		
Undefined	42 (0.13)	10 (0.06)	28 (0.30)	0 (0.00)	4 (0.08)		

Note. KBD, Kashin-Beck disease; Bio1, annual mean temperature; Bio12, annual precipitation; RH, relative humidity; Bio3, isothermality; T_GRAVEL, gravel volume percentage; T_SAND, sand content; T_SILT, silt content; T_ESP, exchangeable sodium percentage; T_ECE, soil electrical conductivity; AN, alkali-hydrolysable nitrogen; GDP, gross domestic product; Rail, distance to nearest railway; Se, selenium. Note: Continuous variables are presented as mean ± standard deviation, and categorical variables are presented as n (%). *P* for continuous variables were calculated using one-way analysis of variance (ANOVA), and *P* for categorical variables were calculated using the chi-square test. $P < 0.01$ indicates statistical significance.

Supplementary Table 6S. Univariate analysis after excluding samples from undefined selenium landscape areas

Variables	Total	Non-endemic village	KBD endemic village			Statistic	*P*
			Mild	Moderate	Severe		
Se, n(%)						χ^2=5849.90<0.001	
Low Se	25156	9654 (58.51)	8627	2356	4519		
Marginal Se	5437	4608 (27.93)	542 (5.88)	43 (1.79)	244 (5.02)		
Moderately-high	2387 (7.24)	2238 (13.56)	53 (0.57)	3 (0.12)	93 (1.92)		

Note. Se, selenium. Note: Values are presented as n (%). χ^2, Chi-square test.



Supplementary Figure S2. Correlation matrix among environmental variables.

Supplementary Table S7. VIF values of environmental variables

All variables		**Selected variables**	
Variable	VIF	Variable	VIF
Altitude	1.69	Altitude	1.66
Slope	1.29	Slope	1.26
Bio1	1.96	Bio1	1.78
Bio12	2.83	Bio12	2.34
RH	1.19	RH	1.17
Bio3	1.64	Bio3	1.61
T_GRAVEL	1.16	T_GRAVEL	1.07

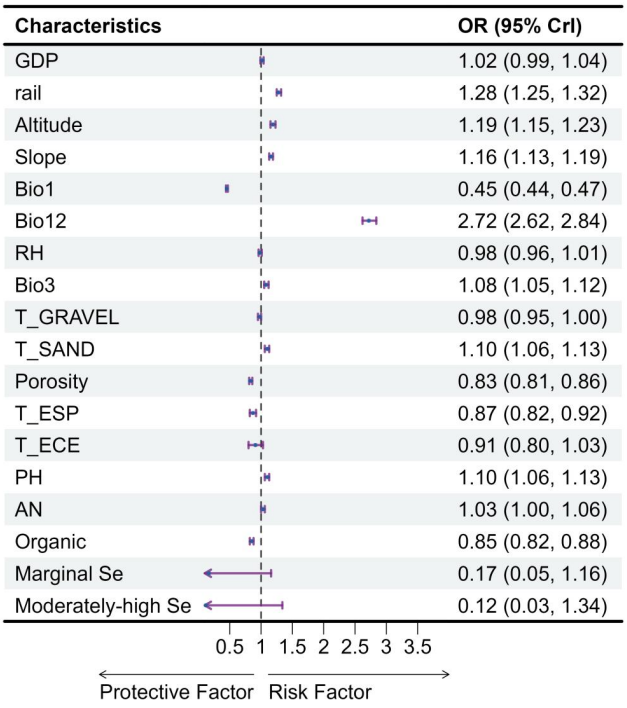
T_SAND	6.19	T_SAND	1.16
T_SILT	6.28	—	—
Porosity	1.28	Porosity	1.28
T_ESP	2.72	T_ESP	2.67
T_ECE	2.59	T_ECE	2.58
PH	1.98	PH	1.92
AN	2.37	AN	2.36
Organic	2.64	Organic	2.63
GDP	1.08	GDP	1.08
rail	1.14	rail	1.11

Note. VIF, variance inflation factor. “All variables” represents VIF values before collinearity screening; “Selected variables” represents VIF values after removing collinear variables.

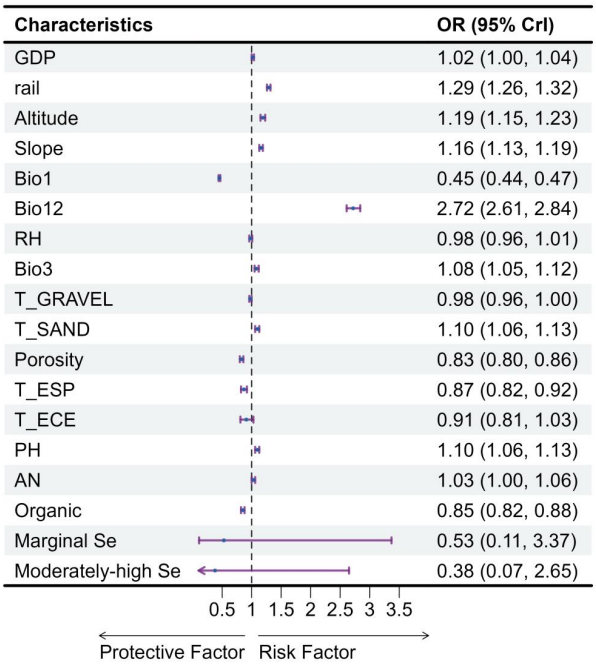
Supplementary Table S8. Results of Bayesian multilevel ordinal logistic null model

Type	**Parameter**	***Estimate***	**Uncertainty**		**Diagnosti**
		β/σ/ICC	*SE*	*95% CrI*	*Rhat*
Fixed effects	Intercept 1 (control vs mild+)	0.24	0.86	(-1.68, 1.69)	1.003
	Intercept 2 (mild vs moderate+)	1.67	0.86	(-0.26, 3.11)	1.003
	Intercept 3 (moderate vs severe)	2.18	0.86	(0.25, 3.63)	1.003
Random effects	Between-Se-zone standard deviation (σ)	1.87	0.77	(0.84, 3.81)	1.001
	Intraclass correlation (*ICC*)	0.47	-	(0.18, 0.82)	-

Note. *Estimate*, posterior mean of the fixed-effect regression coefficient; *β*, regression coefficient; *σ*, standard deviation of random effect; *ICC*, intraclass correlation coefficient; *SE*, standard error; *CrI*, credible interval; *Rhat*, Gelman-Rubin statistic. Note: Intercept 1 indicates the threshold between control and mild-or-higher severity, Intercept 2 indicates the threshold between mild and moderate-or-higher severity, and Intercept 3 indicates the threshold between moderate and severe severity. Random effects reflect between-selenium-zone variability.



Supplementary Figure S3. Forest plot of the Bayesian two-level ordinal logistic regression under a diffuse prior specification (Normal(0, 2)). Squares indicate posterior odds ratio (*OR*), and horizontal lines indicate 95% credible intervals (*CrIs*) for each predictor.

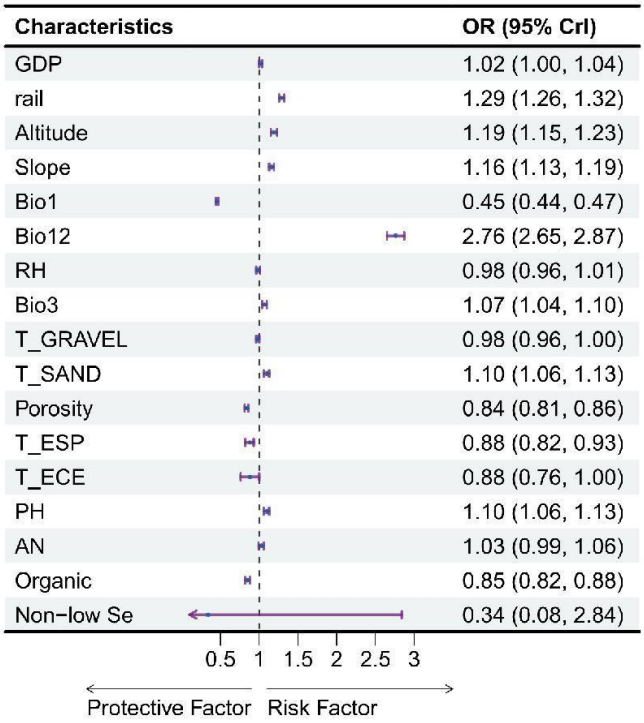


Supplementary Figure S4. Forest plot of the Bayesian two-level ordinal logistic regression under a concentrated prior specification (Normal(0, 1)). Squares indicate posterior odds ratio (*OR*), and horizontal lines indicate 95% credible intervals (*CrIs*) for each predictor.

Supplementary Table S9. Multilevel ordinal logistic regression (two-category Se)

Type	Variables	*Estimate*	*Est.Error*	*95% CrI(β)*	*Rhat*
Fixed effects	Intercept				
	Intercept1(control vs mild+)	-0.58	0.66	(-2.09, 0.58)	1.001
	Intercept2(mild vs moderate+)	1.14	0.66	(-0.38, 2.30)	1.001
	Intercept3(moderate vs severe)	1.71	0.66	(0.19, 2.88)	1.001
	Level 1(village)				
	GDP	0.02	0.01	(-0.00, 0.04)	1.000
	rail	0.25	0.01	(0.23, 0.28)	1.000
	Altitude	0.18	0.02	(0.14, 0.21)	1.001
	Slope	0.15	0.01	(0.13, 0.17)	1.002
	Bio1	-0.79	0.02	(-0.82, -0.75)	1.001
	Bio12	1.01	0.02	(0.97, 1.05)	1.000
	RH	-0.02	0.01	(-0.04, 0.01)	1.001
	Bio3	0.07	0.02	(0.04, 0.10)	1.000
	T_GRAVEL	-0.02	0.01	(-0.05, 0.00)	1.000
	T_SAND	0.09	0.02	(0.06, 0.12)	1.000
	Porosity	-0.18	0.02	(-0.22, -0.15)	1.000
	T_ESP	-0.13	0.03	(-0.20, -0.07)	1.000
	T_ECE	-0.13	0.07	(-0.27, -0.00)	1.000
	PH	0.09	0.02	(0.06, 0.13)	1.001
	AN	0.03	0.02	(-0.01, 0.06)	1.000
	Organic	-0.16	0.02	(-0.20, -0.13)	1.000
	Level 2(Se)(Reference: low Se)				
	Non-low Se	-1.07	0.97	(-2.52, 1.04)	1.000
Random effects					
	Level 2(between-Se-zone)	1.17	-	(0.04, 3.36)	1.000

Note. *Estimate*, posterior mean of the fixed-effect regression coefficient; *β*, regression coefficient; *Est.Error*, standard error of the estimate; *CrI*, credible interval; *Rhat*, Gelman–Rubin convergence diagnostic; Se, selenium.



Supplementary Figure S5. Forest plot of Bayesian two-level ordinal logistic regression with selenium categorized into two groups (low Se vs. non-low Se). Note: Squares indicate posterior odds ratio (*OR*), and horizontal lines indicate 95% credible intervals (*CrIs*) for each predictor.

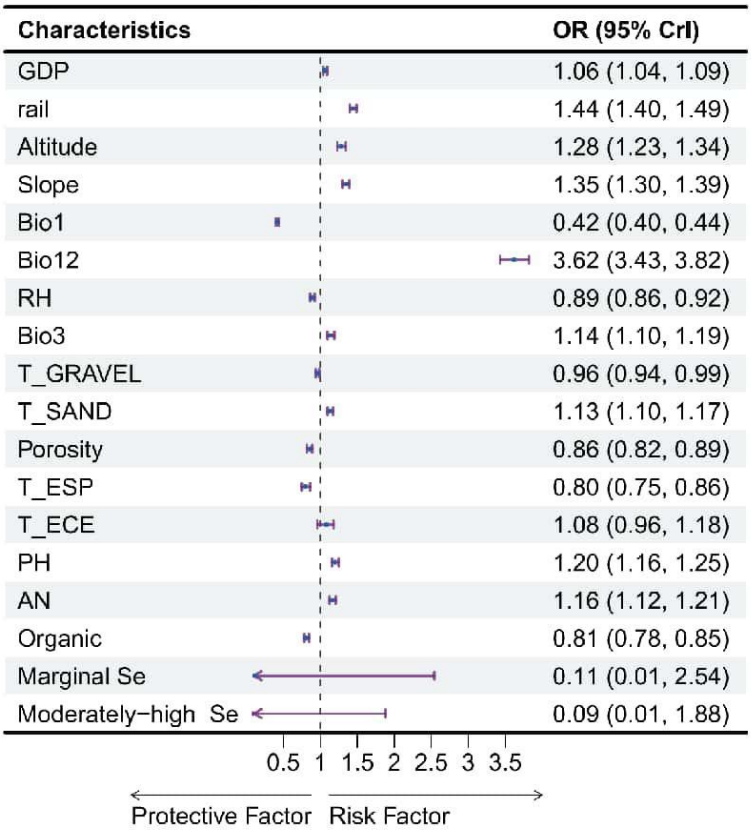
Supplementary Table S10. Results of Bayesian multilevel binary logistic null model

Type	**Parameter**	***Estimat***	**Uncertainty**		**Diagnostics**
		β/σ/ICC	*SE*	*95% CrI*	*Rhat*
Fixed effects	Intercept	-1.05	1.67	(-2.70, 0.98)	1.012
Random effects	Between-Se-zone standard deviation(σ)	1.67	0.64	(0.81, 3.24)	1.011
	Intraclass Correlation(*ICC*)	0.42	-	(0.17, 0.76)	-

Note. *β*, regression coefficient; σ, standard deviation of random effect; ICC, intraclass correlation coefficient; SE, standard error; *CrI*, credible interval; Rhat, Gelman-Rubin statistic.

Supplementary Table S11. Results of Bayesian multilevel binary logistic full model

Type	Variables	*Estimate*	*Est.Error*	*95% CrI(β)*	*Rhat*
Fixed effects	Intercept	0.41	0.47	(-1.53, 1.95)	1.000
	Level 1(village)				
	GDP	0.06	0.01	(0.04, 0.09)	1.000
	rail	0.37	0.02	(0.34, 0.40)	1.001
	Altitude	-2.18	0.64	(-4.34, 0.93)	1.001
	Slope	-2.43	0.66	(-4.82, 0.63)	1.000
	Bio1	0.25	0.02	(0.21, 0.29)	1.000
	Bio12	0.30	0.02	(0.27, 0.33)	1.000
	RH	-0.87	0.02	(-0.91, -0.83)	1.000
	Bio3	1.29	0.03	(1.23, 1.34)	1.000
	T_GRAVEL	-0.12	0.02	(-0.15, -0.09)	1.001
	T_SAND	0.13	0.02	(0.10, 0.17)	1.000
	Porosity	-0.04	0.01	(-0.07, -0.01)	1.001
	T_ESP	0.13	0.02	(0.09, 0.16)	1.000
	T_ECE	-0.15	0.02	(-0.20, -0.12)	1.000
	PH	-0.22	0.03	(-0.29, -0.15)	1.000
	AN	0.08	0.05	(-0.04, 0.17)	1.001
	Organic	0.18	0.02	(0.15, 0.22)	1.000
	Level 2(Se)(reference: low Se)				
	Marginal Se	-2.18	0.64	(-4.34, 0.93)	1.001
	Moderately-high Se	-2.43	0.66	(-4.82, 0.63)	1.000
Random					
	Level 2(between-Se-zone)	0.58	-	(0.02, 2.68)	1.000



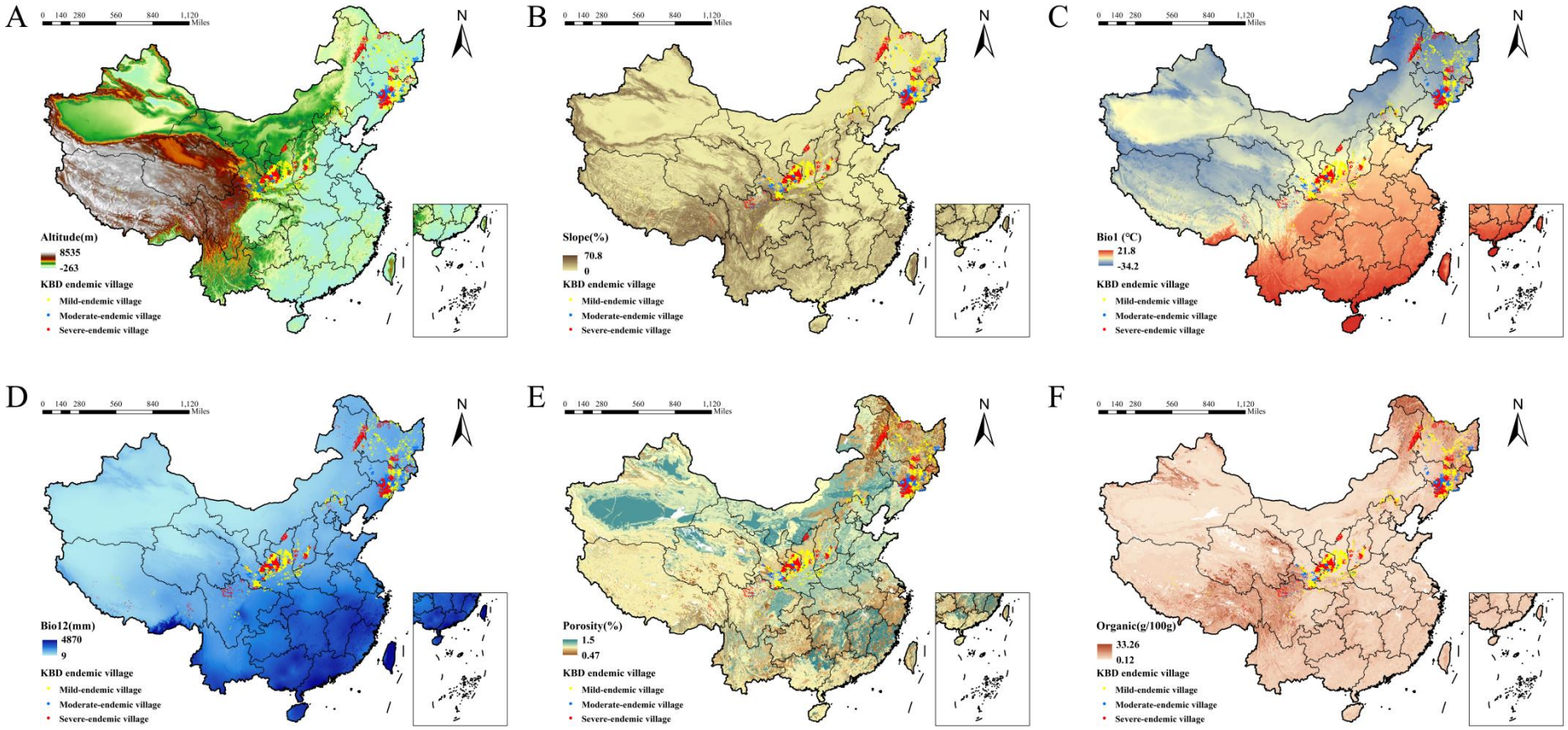
Supplementary Figure S6. Forest plot of Bayesian multilevel binary logistic regression. Fixed effects represent estimated odds ratios (*ORs*) and 95% credible intervals (*CrIs*) for the risk of KBD-endemic villages (relative to non-endemic villages).

Supplementary Table S12. Interaction effects among environmental factors

Parameter	*Estimate*	*Est.Error*	*Rhat*	*95% CrI*
T_ESP*				
Marginal Se	0.11	0.06	1.005	(-0.01, 0.24)
Mod-high Se	0.45	0.07	1.004	(0.32, 0.57)
T_ECE*				
Marginal Se	-0.08	0.20	1.026	(-0.49, 0.29)
Mod-high Se	0.30	0.09	1.011	(0.10, 0.47)
PH*				
Marginal Se	0.05	0.03	1.002	(-0.01, 0.12)
Mod-high Se	-0.28	0.10	1.001	(-0.49, -0.07)
AN*				
Marginal Se	0.29	0.04	1.002	(0.21, 0.36)
Mod-high Se	-0.90	0.16	1.004	(-1.23, -0.60)

Parameter	*Estimate*	*Est.Error*	*Rhat*	*95% CrI*
Organic*				
Marginal Se	0.04	0.04	1.006	(-0.03, 0.12)
Mod-high Se	-0.44	0.08	1.003	(-0.59, -0.29)

Note.Se, selenium; *Est.Error*, standard error of the estimate; *CrI*, credible interval; *Rhat*, Gelman–Rubin statistic. Interaction effects between soil factors (T_ESP, T_ECE, PH, AN, and Organic) and selenium background levels (reference: low Se).



Supplementary Figure S7. Overlay maps of KBD endemic villages and key environmental factors.

1. National Health and Family Planning Commission of People’s Republic of China. Criteria for delimitation and classification of Kashin–Beck disease endemic area (GB 16395–2011). 2011. (In Chinese).
2. Li HR, Yang LS, Tan JA, et al. Progress on Selenium Deficiency inGeographical Environment and its Health Impacts in China .Current Biotechnology, 2017; 7, 381-6. (In Chinese).
3. Compilation Committee of the Endemic Disease and Environment Atlas of the People's Republic of China. Endemic Disease and Environment Atlas of the People's Republic of China. Beijing: Science Press, 1989. (In Chinese).
4. Tan JA. Environmental Selenium and Health. Beijing: People's Medical Publishing House, 1989. (In Chinese).
5. Tan JA. Environmental Life Elements and Keshan Disease: A Study on Eco-Chemical Geography. Beijing: China Medical Science and Technology Press, 1996. (In Chinese).
6. Research Group of Environment and Endemic Diseases, Institute of Geography, Chinese Academy of Sciences. A Study on China's Low-Selenium Belt and the Etiology of Keshan Disease and Kashin-Beck Disease. Bulletin of Chinese Academy of Sciences, 1988, 1:54-60. (In Chinese)