



# THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

### Equine grass sickness in Scotland

**Citation for published version:**

Wylie, C, Shaw, D, Fordyce, F, Lilly, A, Pirie, R & McGorum, B 2016, 'Equine grass sickness in Scotland: A case-control study of environmental geochemical risk factors', *Equine Veterinary Journal*, vol. 48, no. 6, pp. 779-785. <https://doi.org/10.1111/evj.12490>

**Digital Object Identifier (DOI):**

[10.1111/evj.12490](https://doi.org/10.1111/evj.12490)

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Peer reviewed version

**Published In:**

Equine Veterinary Journal

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



1 **Equine grass sickness in Scotland: A case-control study of environmental geochemical**  
2 **risk factors**

3 Wylie C.E. (1)\*, Shaw D.J. (1), Fordyce F.M. (2), Lilly, A. (3), Pirie, R.S. (1) & McGorum  
4 B.C. (1)

5 (1) Royal (Dick) School of Veterinary Studies & The Roslin Institute, Easter Bush Veterinary  
6 Centre, The University of Edinburgh, Roslin, EH25 9RG, United Kingdom

7 (2) British Geological Survey, West Mains Road, Edinburgh, EH9 3LA, United Kingdom

8 (3) James Hutton Institute, Craigiebuckler, Aberdeen, AB15 8QH, United Kingdom

9

10 \* Current address; Rosssdales Equine Hospital, Cotton End Road, Exning, Newmarket, Suffolk,  
11 CB8 7NN

12 Contact: [claire.wylie@rossdales.com](mailto:claire.wylie@rossdales.com)

13

14 **Key Words:** equine grass sickness, EGS, equine dysautonomia, epidemiology, soil  
15 geochemistry, trace elements

16 **Word Count: 4978**

17

18 **Summary:**

19 **Reasons for performing study:** We hypothesised that the apparent geographical distribution  
20 of equine grass sickness (EGS) is partly attributable to sub-optimal levels of soil macro- and  
21 trace elements in fields where EGS occurs. If proven, altering levels of particular elements  
22 could be used to reduce the risk of EGS.

23 **Objectives:** To determine whether the geographical distribution of EGS cases in Eastern  
24 Scotland is associated with the presence or absence of particular environmental chemical  
25 elements.

26 **Study design:** Retrospective time-matched case-control study.

27 **Methods:** This study utilised data for 455 geo-referenced EGS cases and 910 time-matched  
28 controls in Eastern Scotland, and geo-referenced environmental geochemical data from the  
29 British Geological Survey Geochemical Baseline Survey of the Environment stream sediment  
30 (G-BASE) and the James Hutton Institute, National Soil Inventory of Scotland (NSIS) datasets.

31 **Results:** Multivariable statistical analyses identified clusters of three main elements associated  
32 with cases from (i) the G-BASE dataset – higher environmental *Ti* and lower *Zn*, and (ii) the  
33 NSIS dataset – higher environmental *Ti* and lower *Cr*. There was also some evidence from  
34 univariable analyses for lower *Al*, *Cd*, *Cu*, *Ni* and *Pb* and higher *Ca*, *K*, *Mo*, *Na* and *Se*  
35 environmental concentrations being associated with a case. Results were complicated by a  
36 high degree of correlation between most geochemical elements.

37 **Conclusions:** The work presented here would appear to reflect soil- not horse-level risk factors  
38 for EGS, but due to the complexity of the correlations between elements, further work is  
39 required to determine whether these associations reflect causality, and consequently whether  
40 interventions to alter concentrations of particular elements in soil, or in grazing horses, could

41 potentially reduce the risk of EGS. The effect of chemical elements on the growth of those soil  
42 micro-organisms implicated in EGS aetiology also warrants further study.

43

#### 44 **Acknowledgements**

45 Published with permission of the Executive Director of the British Geological Survey (BGS).

46 We are grateful to the BGS University Funding Initiative (BUFI) for part-funding, and give

47 particular thanks to Dr Chris Johnson (BGS) for data extraction. The BGS work is funded by

48 the Natural Environment Research Council (NERC). The work by AL is funded by the Scottish

49 Government's Rural and Environment Science and Analytical Services (RESAS) Division.

50 These organisations accept no responsibility for any inaccuracies or omissions in data, nor for

51 any loss or damage directly or indirectly caused to any person or body by reason of, or arising

52 out of, any use of this data. The authors also acknowledge support from the Equine Grass

53 Sickness Fund, World Horse Welfare, The Horse Trust, RDSVS clinicians, referring vets and

54 owners.

55

## 56 **Introduction**

57 Equine grass sickness (EGS) has a restricted geographical distribution, with Great Britain, and  
58 Scotland in particular, having the highest prevalence worldwide [1; 2]. While the aetiology  
59 remains unknown, increasing evidence suggests a toxico-infectious form of botulism with  
60 additional risk factors [3-7]. The spatial distribution may be attributable to the dispersal of  
61 botulinum spores or to geographically-restricted environmental risk factors. EGS has been  
62 associated with acidic soils and soils with high concentrations of nitrogen (*N*) [8-13]. In  
63 comparison with clay soils; loam or sandy soils are associated with a higher risk of EGS  
64 recurrence, while chalk and other texture soils are associated with lower risk of recurrence [14].  
65 Previous research into macro- and trace element intake in relation to EGS is limited and  
66 inconclusive. A small case-control study (*n*=23 cases, *n*=11 controls) quantified Aluminium  
67 (*Al*), Arsenic (*As*), Barium (*B*), Calcium (*Ca*), Cadmium (*Cd*), Chromium (*Cr*), Copper (*Cu*),  
68 Iron (*Fe*), Mercury (*Hg*), Potassium (*K*), Magnesium (*Mg*), Manganese (*Mn*), Molybdenum  
69 (*Mo*), *N*, Nickel (*Ni*), Phosphorus (*P*), Lead (*Pb*) and Sulphur (*S*) in soil and herbage from EGS  
70 and control fields [15]. Herbage from EGS fields had significantly higher levels of *Cr*, *Fe* and  
71 *Pb*, and significantly lower levels of *Hg*, while concentrations of all elements in soils were not  
72 significantly different. Greig [16] and Doxey *et al.* [17] found no differences in systemic  
73 concentrations of *Cu*, *Mg* and glutathione peroxidase between EGS and healthy co-grazing  
74 horses. McGorum *et al.* [13] identified differences in systemic levels of *Cu*, *Mg*, and Selenium  
75 (*Se*) between EGS and control horses, but it was unclear whether these represented risk factors  
76 for EGS, metabolic consequences of EGS, or were unrelated to EGS.

77 The aim of this study was to determine whether the geographical distributions of geo-  
78 referenced EGS and control horses in Eastern Scotland were associated with the presence or  
79 absence of particular environmental chemical elements.

80

81 **Materials and Methods**

82 The analysis was undertaken as a retrospective time-matched case-control study, employing  
83 much larger populations than previous studies.

84 *Case and control data*

85 Horse data used are detailed in Wylie *et al.* [18]. Briefly, the postcode of 455 EGS cases were  
86 identified from clinical case records at the Royal (Dick) School of Veterinary Studies (RDSVS)  
87 between 01/01/1990 and 01/06/2006. Each case had two time-matched controls ( $n=910$ ),  
88 which were grazing equids referred to RDSVS for reasons other than EGS, immediately prior  
89 and after the EGS case. Equids originating outwith Scotland were excluded.

90

91 *Geochemical data*

92 It was not possible to obtain and analyse soil samples from locations where cases and controls  
93 were grazing; therefore, geochemical data from two Scotland-wide derived datasets, namely  
94 the British Geological Survey (BGS) Geochemical Baseline Survey of the Environment (G-  
95 BASE) stream sediment dataset and the James Hutton Institute (JHI) National Soil Inventory  
96 of Scotland (NSIS) were used as broad-scale proxies. Data from both the G-BASE stream  
97 sediment and NSIS soil datasets were considered as part of the study as they had different  
98 merits. Whilst the G-BASE stream sediment dataset may be further removed from the soil-  
99 horse exposure route (with some enhancement and/or depletion of certain elements); it  
100 uniquely provides detailed information on the surface geochemistry of Scotland (1 sample per  
101 1.5 km<sup>2</sup>) thereby permitting more accurate geographically located sediment data for each horse.  
102 These samples closely relate to the local bedrock of the stream catchment, and are considered

103 to be a sufficient proxy for the source of chemical elements in soil in a given area [19-21]. The  
104 NSIS dataset contains less spatial detail (1 sample per 100 km<sup>2</sup>), but has been extrapolated to  
105 provide information on the chemistry of the main soil types across Scotland. These are more  
106 strongly influenced by climate, topography, land-use and biological processes than stream  
107 sediments, albeit only for more broad scale differences [22]. The G-BASE dataset had the  
108 advantage that is it spatially detailed, but is not a measure of actual soil chemistry; just a proxy  
109 for it [20; 21]. The NSIS dataset had the advantage that it is a measure of soil chemistry, but is  
110 not as spatially detailed. As such, both datasets provide information on chemical element  
111 distributions across Scotland, but with some variations due to the different sample types and  
112 methods used to generate the data.

113 Data regarding a range of elements were collated from the two geochemical datasets (G-BASE,  
114 NSIS) (Table 1). Unfortunately neither dataset contained data regarding environmental *S* [13],  
115 while limited *Se* data (45 samples) were available for the NSIS dataset only.

116

#### 117 *BGS G-BASE stream sediment dataset*

118 BGS G-BASE data are based upon fine-fraction (<150 µm) first and second order stream  
119 sediment total element concentrations determined by Direct Reading Optical Emission  
120 Spectrometry (DC-OES) and Atomic Adsorption Spectrometry (AAS) collected at a sampling  
121 density of 1 per 1.5 km<sup>2</sup> ( c.49,000 sample points) across Scotland [23]. Data were provided  
122 as geographic information system (GIS) interpolated raster maps (based on a 250 m grid)  
123 (ArcMap10.1; Environmental Systems Research Institute, ESRI<sup>®</sup>), showing concentrations of  
124 each element at any given location (Figure 1a and Table 1). For each horse location (Ordnance  
125 Survey [OS] easting and northing), geochemical values at that site were extracted by GIS  
126 overlay. Data for *Ca*, *K*, *Mg* and Titanium (*Ti*) were lacking for northern Sutherland/Caithness

127 and Orkney, resulting in two (0.2%) controls and eight (1.8%) cases being excluded from  
128 statistical analysis.

129

### 130 *JHI NSIS Soil Dataset*

131 Information on topsoil geochemistry from JHI comprised concentrations derived from a 10km  
132 grid survey across Scotland subsequently grouped by major Soil Associations (defined by  
133 parent material types) and median values derived to allow spatial extrapolation between the  
134 10km sample points (Figure 1b). Data comprised element concentrations for the uppermost  
135 soil horizon (layer) from 719 samples ( $n=292$  mineral topsoils,  $n=427$  organic surface layers)  
136 determined by Inductively Coupled Plasma Atomic Emission Spectrometry or Graphite  
137 Furnace Atomic Absorption Spectrometry after Aqua regia extraction (Table 2) [19]. More  
138 geochemical elements were available compared to the G-BASE dataset but the same values  
139 were ascribed to many horses as many owners' addresses were located over the same Soil  
140 Association found in different parts of the sampled area. In total, 1115 sites (384 cases, 731  
141 controls) were allocated a median concentration for a range of elements. Data were unavailable  
142 for a further 250 sites (71 cases, 179 controls) which were located within unmapped, often  
143 urban, areas.

144

### 145 *Data Management and Statistical Analyses*

146 Prior to statistical analysis, G-BASE data for all elements except *Mo* were  $\log_{10}$  transformed  
147 (Supplementary Figure 1a), and geometric means (GM) and 95% confidence intervals (CI)  
148 calculated for cases and controls. *Mo* data could not be normalised with any transformations  
149 as 23% of values=0 mg kg<sup>-1</sup>, therefore the percentage of values >0 mg kg<sup>-1</sup> were calculated.



150 For NSIS-derived data for all elements except *Se*, overall medians and interquartile ranges  
151 (IQR) were determined for cases and controls due to the distribution of data (Supplementary  
152 Figure 1b). However, for the risk factor analysis, NSIS-derived values were  $\log_{10}$  transformed  
153 to try and generate meaningful estimates of risk [24]. For *Se*, data were converted into  
154 percentage of values  $>0.22 \text{ mg kg}^{-1}$ . Correlations between elements within the G-BASE and  
155 NSIS-derived datasets were considered separately, using Spearman rank correlations.

156 For the risk factor analysis a three-stage approach was adopted. First, univariable conditional  
157 logistic regression analyses (hereafter '*univariable conditional*') were conducted to examine  
158 individual element relationships between cases *versus* controls and variables. G-BASE and  
159 NSIS-derived element datasets were analysed separately. Each case and its two matched  
160 controls were entered into the *univariable conditional* as a set. Data from the two geochemical  
161 databases were not combined due to differences in sample type, methods of data collection,  
162 sample density and method of spatial interpolation.

163 The next two stages of the analysis were both involved in generating final G-BASE and NSIS-  
164 derived multivariable conditional logistic regression models (hereafter '*multivariable*  
165 *conditional*') based on the *univariable conditional* results. The complication was the  
166 statistically significant correlations found between most of the elements (see Results and  
167 Supplementary Table 1). The approach adopted here was to consider one variable per cluster  
168 of correlated variables (where  $r_s > |0.50|$ ) – the variable with the most statistically significant  
169 univariate result and enter them into an initial *multivariable conditional* model and then employ  
170 step-wise model deletion until each final *multivariable conditional* model was obtained.

171 Finally, due to potential concerns that results were being driven by the exact variables selected  
172 to go into the *multivariable conditional* model, a set of sensitivity type analyses were carried  
173 out. A series of *multivariable conditional* models were run where individual elements and pairs

174 of elements were initially excluded in order to determine whether common elements remained  
175 in the final models and whether some elements only remained in a final model if particular  
176 element(s) were excluded.

177 *Univariable conditional* logistic regression results for both G-BASE and NSIS-derived are  
178 expressed as coefficients from the models, Wald chi-square *P*-values and for all variables  
179 (except G-BASE *Mo* and NSIS-derived *Se*), odds ratios of log<sub>10</sub> transformed data (OR) and  
180 corresponding CI. For G-BASE *Mo* and NSIS-derived *Se* odds ratios of the percentage of  
181 horses >0 mg kg<sup>-1</sup> and >0.22 mg kg<sup>-1</sup> respectively (OR [ $\pm$ 95%CI]) were calculated. For all the  
182 *multivariable conditional* models log likelihoods, degrees of freedom and overall *P*-values are  
183 also given. To maintain the case-control structure for the NSIS-derived analysis only 246  
184 complete sets remained in areas where the soil had been mapped.

185 A one percent level (*P*<0.01) was taken to indicate statistical significance throughout. All  
186 analyses were done in R Statistical Package (version 3.0.0 © 2013 The R Foundation for  
187 Statistical Computing), using ‘*epicalc*’ (v 2.15.1) for risk factor analysis.

188

## 189 **Results**

190

### 191 *Conditional logistic regression analysis with G-BASE stream sediment data*

192 *Univariable conditional* analyses revealed a greater probability of being a case for lower  
193 concentrations of sediment *Cr*, *Pb* and *Zn* (OR<0.5, Table 1; Supplementary Figure 1a) and  
194 higher concentrations of *Ca*, *K* and *Ti* (OR>4.8), and cases associated with *Mo*>0 mg kg<sup>-1</sup> (87%  
195 vs. 76%, OR=1.98), with no differences for the other elements. However, these results are  
196 complicated by 14 of the possible 66 correlations (21%) between the 12 numerical element

197 values having  $r_s > |0.50|$ , and 53 (80%) showing any statistical association (Supplementary Table  
198 1a).

199 However of the variables that were statistically significant in the *univariable conditional*  
200 analyses only 2 pairs of variables had  $r_s > |0.50|$  - the positive correlations between *Ca* and *Ti*,  
201 and between *Pb* and *Zn*. Of these *Ti* and *Zn* had the most significant coefficients in the  
202 *univariable conditional* analyses and they were therefore entered into the initial stepwise  
203 deletion *multivariable conditional* model. For the other 3 significant variables from the  
204 *univariable conditional* analysis – *Cr*, *Ga* and *K* – which were all correlated with each other to  
205 a lesser degree  $r_s > |0.15|$ , *K* had the most significant coefficient in the *univariable conditional*  
206 analyses and was also entered into the initial stepwise deletion *multivariable conditional*  
207 model. Finally, the binary measure of *Mo* > 0 mg kg<sup>-1</sup> was added to the initial *multivariable*  
208 *conditional* model.

209 From this initial *multivariable conditional* model, two variables remained in the final model -  
210 higher *Ti* concentrations (OR=17.24) and lower *Zn* concentrations (OR=0.25) was associated  
211 with being a case (Table 2). This final *multivariable conditional* model was also observed in  
212 with a number of the sensitivity models run, and at least one of these variables was present in  
213 all the sensitivity models (Supplementary text and Supplementary Table 2a).

#### 214 *Conditional logistic regression analysis with NSIS-derived soil data*

215 As with the G-BASE data, *univariable conditional* analyses revealed a greater probability of  
216 being a case with lower soil *Cr*, *Pb*, and *Zn* concentrations (OR<0.2) and higher concentrations  
217 of *Ti* (OR=77.2, Table 1, Supplementary Figure 1b). However, unlike the G-BASE data,  
218 relationships were also observed with lower concentrations of *Co* and *Cu* (OR<0.23) and no  
219 relationships were observed for *Ca* or *K*. There were additional elements in the soil sample  
220 data, and from these, lower concentrations of *Al*, *Cd* and *Ni* (OR<0.23) were associated with

221 being a case as were higher concentrations of *Na* (OR=8.9). Finally, the proportion of cases  
222 (75%) that had *Se* concentrations  $>0.22 \text{ mg kg}^{-1}$  were greater than those of controls (60.6%,  
223 OR=2.52).

224 Similar to the G-BASE dataset, most of the NSIS-derived elements were significantly  
225 correlated with each other, but the degree of correlation was even greater with 44% (N=68) -  
226 of these  $r_s > |0.50|$  (Supplementary Table 1b), and 49 of these 68 (72%) involving pairs of  
227 variables statistically significant in the *univariable conditional* analyses (Table 1b,  
228 Supplementary Table 1b). Taking the 2 most significant coefficients from the *univariable*  
229 *conditional* analyses – *Ti* and *Cr*, then of the other significant coefficients *Ti* was positively  
230 correlated with *Na*, and negatively correlated with *Cd*, *Cu*, and *Pb*; and *Cr* was positively  
231 correlated with *Al*, *Cd*, *Co*, *Cu*, *Fe*, *Pb*, *Ni*, *P* and *Zn*. Therefore, *Ti* & *Cr* were added to the  
232 initial *multivariable conditional* model.

233 From this initial *multivariable conditional* model, both variables remained in the final model -  
234 higher *Ti* concentrations (OR=38) and lower *Cr* concentrations (OR=0.09) was associated with  
235 being a case (Table 2b). This final *multivariable conditional* model was also observed in with  
236 a number of the sensitivity models run, and at least one of these variables was present in all but  
237 one of the sensitivity models (Supplementary text and Supplementary Table 2b).

238

## 239 **Discussion**

240 This is the largest case-control study regarding the role of geochemical elements as potential  
241 risk factors for EGS. As we were unable to obtain and analyse soil samples from where the  
242 cases and controls were grazed, we adopted a broad scale approach, utilising for the first time  
243 in EGS research two large scale geochemical element databases that exist in Scotland. These

244 two databases were employed as they conferred complementary advantages: whilst the G-  
245 BASE stream sediment dataset provides more geographically accurate values for each horse,  
246 the NSIS-derived data were for actual soil, albeit at a lower resolution. Neither dataset was  
247 detailed enough to identify particular grazing pastures at risk of EGS. Nevertheless in the  
248 absence of such data, we investigated whether general patterns in disease incidence were  
249 associated with the distribution of soil chemical elements.

250 A major complication in the analysis was considerable confounding between different elements  
251 within both databases. This is expected as geochemical relationships, which are largely  
252 controlled by the parent material, often result in certain elements occurring together in higher  
253 concentrations at the same location [25]. For example, soils developed over mafic igneous  
254 rocks tend to have higher *Fe*, *Mg*, *Cr*, *Ni* and *Cu* concentrations than soils developed over  
255 granite rock types due to differences in the mineral composition of these rock parent materials.  
256 This made interpretation of the results complex. However, the *multivariable conditional*  
257 models identified a number of elements that were associated with an EGS case. In the G-BASE  
258 data, cases were associated with two elements -  $\uparrow Ti$  and  $\downarrow Zn$  (Table 2). Furthermore, the  
259 sensitivity analyses always containing one of these 2 variables in four predominating clusters  
260 of four elements -  $\uparrow Ca\uparrow Ti\downarrow Zn$ ,  $\uparrow Ca\uparrow Ti\downarrow Pb$ ,  $\uparrow Ti\downarrow Zn$ , and  $\uparrow Ca\downarrow Zn$ , depending on which  
261 elements were included in the initial models (Supplementary Table 2a).

262 A similar situation was observed with the NSIS-derived data – cases were associated with two  
263 elements -  $\uparrow Ti$  and  $\downarrow Cr$  (Table 2). These two variables again dominated in the sensitivity  
264 analyses, with sensitivity analyses models containing one of these 2 variables in three  
265 predominating clusters with four elements –  $\downarrow Cr\uparrow Na$ ,  $\downarrow Cr\uparrow Ti$ , and  $\downarrow Ni\uparrow Ti$ , again depending  
266 on which elements were in the initial models (Supplementary Table 2b). Other elements from  
267 both databases were either present in one or two particular combinations of elements or only

268 appeared in the *univariable conditional* analyses, with EGS potentially associated with lower  
269 environmental concentrations of *Al*, *Cd*, *Co*, and *Cu* (NSIS), and higher environmental  
270 concentrations of *Ca*, *K*, *Mo* (G-BASE), and *Se* (NSIS) (Tables 1&2). The significance of  
271 these other elements in the pathophysiology of the disease is difficult to ascertain given the  
272 confounding.

273 Further investigation is required to determine whether these relationships represent risk factors  
274 acting at horse- or soil-level; surrogate indicators of other risk factors; or simply spatial  
275 coincidence unrelated to disease causality. Clarification of these associations is a pre-requisite  
276 to determine whether interventions to alter concentrations of particular elements in soil, or in  
277 grazing horses, could potentially reduce the risk of EGS. It is unlikely that the aforementioned  
278 associations represent sufficient causal complex factors for EGS because environmental levels  
279 of all the elements were largely within acceptable ranges [26; 27], and were not sufficiently  
280 high or low to cause clinical toxicity or deficiency. More likely, sub-optimal macro- and trace  
281 element status of horses grazing EGS fields represents a contributory horse-level risk factor,  
282 by exacerbating the detrimental effects of other causal agents such as microbial neurotoxins.

283 In addition, for most of the elements in the two sets of clusters there is no apparent known  
284 biological impact on horses. It appears biologically implausible that the association between  
285 EGS and higher levels of *Ti* reflect horse-level risks. The association with EGS and low  
286 environmental levels of *Cr* is difficult to interpret since, despite its important biological  
287 functions, essentiality of *Cr* in the horse has not been determined [28]. Of the elements  
288 predominating in lower-environmental-concentration clusters, *Zn* is perhaps the most likely to  
289 reflect horse-level risk factors, since *Zn* is an essential requirement for horses and for neuronal  
290 homeostasis [29]. There was no difference in plasma *Zn* concentrations of EGS, healthy co-  
291 grazing and control horses in a previous study [13]; although the small study population may  
292 have limited statistical power.

293 More likely, these associations between EGS and geochemical parameters could be due to  
294 interactions in the soil, for example, lower environmental concentrations of *Cd*, *Cu*, *Ni* and *Zn*  
295 may reflect the low levels of these elements in sandy soils [30; 31], which are associated with  
296 increased rate of recurrence of EGS compared with clay, chalk and other textures of soil [14].  
297 Similarly, the association between EGS and high *Ti* may reflect the generally higher *Ti* levels  
298 in clay soils [21]. As high levels of metals inhibit the growth of some soil microbes [32; 33],  
299 including potentially causal soil micro-organisms such as *C. botulinum* and *Fusarium* spp., the  
300 effect of high or low concentrations of elements on those soil micro-organisms that have been  
301 implicated in EGS aetiology warrants further study.

302 Of the elements out-with the main clusters, lower concentrations of *Cu* could be acting in a  
303 similar way as for *Zn*, as again, *Cu* is an essential requirement for horses and for neuronal  
304 homeostasis [29; 34]. Consistent with potential involvement of low *Cu* status in EGS, healthy  
305 horses co-grazing with EGS horses had lower plasma *Cu* levels than control horses grazing  
306 fields in Eire, where EGS is rare [13]. While high dietary *Mo* levels reduce the bioavailability  
307 of *Cu* in ruminants, this does not appear to occur in horses [35], suggesting that the association  
308 between higher environmental *Mo* levels and EGS is not attributable to a detrimental effect of  
309 *Mo* on *Cu* bioavailability. The complexity in association between EGS, environmental levels  
310 of *Co* and other elements in the NSIS analysis makes the *Co* finding difficult to interpret.  
311 Associations between EGS and low environmental concentrations of *Al*, and *Cd* are, as with *Ni*  
312 and *Pb*, unlikely to reflect horse-level risks because again these elements are not essential for  
313 horses. As with *Ca* it appears biologically implausible that the associations between EGS and  
314 higher levels of *K* reflect horse-level risks. Finally, higher environmental *Se* levels also appears  
315 unlikely to be a horse-level risk factor for EGS because many horses grazing in Scotland have  
316 low systemic selenium status [12; 17], and clinical signs of *Se* toxicity [36; 37] are inconsistent  
317 with those of EGS. As with the main elements, some of the other minor elements' associations

318 between EGS and geochemical parameters, such as associations between EGS and lower  
319 environmental levels of *Cd* and *Cu* could act at the soil-level, mirroring the low *Ni/Zn* levels  
320 observed, potentially reflective of sandy soils [30; 31].

321 Unfortunately because many factors influence the uptake and availability of elements from  
322 soils into plants and animals including pH, redox, organic matter, soil wetness and plant and  
323 animal species [38], the associations between EGS and environmental levels of particular  
324 elements do not necessarily imply parallel associations between EGS and *in vivo*  
325 concentrations of these elements. Consequently, there is a requirement for further evaluation  
326 of the macro- and trace element status of horses grazing EGS and control fields, using larger  
327 numbers of horses than previous studies.

328 Limitations regarding the horse dataset were discussed in Wylie *et al.* [18]. Data analysis used  
329 spatial referencing on the postcode of the horses' premises. In some cases matching postcodes  
330 may have reflected matching premises, and this will have had the effect of reducing the power  
331 of the study to detect association between cases and geochemical processes. Given that the  
332 geochemical data were derived from relatively sparse sampling frames and spatially  
333 extrapolated, there is a lack of precision in these data; therefore, having the exact location  
334 would improve future work. Differences between the two datasets may result from possible  
335 geological association between different clusters, different spatial resolution of the data sets,  
336 GBASE acting just as a proxy measure for actual soil chemistry and/or the lack of data for the  
337 250 cases/controls in the NSIS dataset. The G-BASE data were based on interpolated point  
338 source information, whereas NSIS data were used to provide median element concentrations  
339 over Soil Associations. The averaging process and differing sample types, collection and  
340 analytical methods account for the generally higher levels of elements in sediments than soils  
341 (Table 1). The G-BASE stream sediment data are a broad proxy for soils only, but have the  
342 advantage that values could be assigned more accurately to the geographic location of each



343 horse. A limitation in the NSIS-derived dataset was that the same Soil Association median  
344 values were distributed across different parts of the country complicating the analysis and  
345 degree of precision with regards to the results.

346

347 Nevertheless, these are the two most comprehensive soil geochemical databases available and  
348 in the absence of data from the fields where the horses were grazing prior to arrival at the  
349 RDSVS, they allow an initial evaluation of the possible relationship between soil geochemistry  
350 and the factors that influence a disease where the aetiology remains unexplained.

351 The approach taken in this study has been to utilise relative simple analyses to see what broad  
352 level results were found. Clearly further complex evaluation of environmental geochemical  
353 risk factors for EGS, perhaps utilising data with a higher resolution and more comprehensive  
354 derivation could be carried out, but this would require the use of more advanced  
355 epidemiological techniques such as additive Bayesian graphical networks (*e.g.* [39]),  
356 geostatistical model-based interpolation (*e.g.* [40]) or generalised linear or point process  
357 spatial models (*e.g.* [41]). Given the scope of the study and the lack of precision of some of  
358 the geochemical variables and exact locations of horses with regard to premises these  
359 approaches were not adopted for the current study.

360

361

## 362 **Conclusions**

363 Although the two geochemical datasets used in this study are the most comprehensive available  
364 for Scotland, their resolution was not sufficiently detailed to facilitate identification of specific  
365 fields that are at risk of EGS. However, across the East of Scotland in general, statistically  
366 significant associations were identified between the occurrence of EGS and clusters of

367 combinations of elements ( $\uparrow Ti \downarrow Zn$  and  $\uparrow Ti \downarrow Cr$ ) in stream sediment and soil. The relationships  
368 were complicated by the correlations between most variables with some elements associated  
369 with a case if other elements excluded from analyses. Further work is required to determine  
370 whether the associations reflect disease causality, and consequently whether interventions to  
371 alter concentrations of particular elements in soil, or in grazing horses, could potentially reduce  
372 the risk of EGS. Given that the associations identified appear more likely to reflect soil- than  
373 horse-related risk factors, further investigation into the effect of soil element concentrations on  
374 the growth of those soil micro-organisms implicated in EGS aetiology is warranted.

375

376 **References**

377

378 [1] Proudman, C.J. (2008) Geographical distribution of equine grass sickness. In: *Equine Grass*  
379 *Sickness Fund 20th Anniversary Seminar*, Ed: E.G.S. Fund, Equine Grass Sickness Fund, Royal  
380 (Dick) School of Veterinary Studies.

381

382 [2] Wylie, C.E. and Proudman, C.J. (2009) Equine grass sickness: epidemiology, diagnosis, and  
383 global distribution. *Vet. Clin. North Am. Equine Pract.* **25**, 381-399.

384

385 [3] Nunn, F.G., Pirie, R.S., McGorum, B., Wernery, U. and Poxton, I.R. (2007) Preliminary study of  
386 mucosal IgA in the equine small intestine: specific IgA in cases of acute grass sickness and  
387 controls. *Equine Vet. J.* **39**, 457-460.

388

389 [4] Hunter, L.C., Miller, J.K. and Poxton, I.R. (1999) The association of Clostridium botulinum type  
390 C with equine grass sickness: a toxicoinfection? *Equine Vet. J.* **31**, 492-499.

391

392 [5] Hunter, L.C. and Poxton, I.R. (2001) Systemic antibodies to Clostridium botulinum type C: do  
393 they protect horses from grass sickness (dysautonomia)? *Equine Vet. J.* **33**, 547-553.

394

395 [6] McCarthy, H., French, N., Edwards, G., Poxton, I., Kelly, D., Payne-Johnson, C., Miller, K. and  
396 Proudman, C. (2004) Equine grass sickness is associated with low antibody levels to  
397 Clostridium botulinum: a matched case-control study. *Equine Vet. J.* **36**, 123 - 129.

398

399 [7] Newton, J.R., Wylie, C.E., Proudman, C.J., McGorum, B.C. and Poxton, I.R. (2010) Equine grass  
400 sickness: are we any nearer to answers on cause and prevention after a century of research?  
401 *Equine Vet. J.* **42**, 477-481.

402

403 [8] Gilmour, J.S. and Jolly, G.M. (1974) Some aspects of the epidemiology of equine grass sickness.  
404 *Vet. Rec.* **95**, 77-81.

405

406 [9] Doxey, D.L., Gilmour, J.S. and Milne, E.M. (1991) A comparative study of normal equine  
407 populations and those with grass sickness (dysautonomia) in eastern Scotland. *Equine Vet. J.*  
408 **23**, 365-369.

409

410 [10] Wood, J.L., Milne, E.M. and Doxey, D.L. (1998) A case-control study of grass sickness (equine  
411 dysautonomia) in the United Kingdom. *Vet. J.* **156**, 7-14.

412

413 [11] McCarthy, H.E., Proudman, C.J. and French, N.P. (2001) The epidemiology of equine grass  
414 sickness - a literature review (1990-1999). *Vet. Rec.* **149**, 293 - 300.

415

- 416 [12] Macpherson, R. (1978) Selenium deficiency. *Vet. Rec.* **103**, 60.
- 417
- 418 [13] McGorum, B.C., Wilson, R., Pirie, R.S., Mayhew, I.G., Kaur, H. and Aruoma, O.I. (2003) Systemic  
419 concentrations of antioxidants and biomarkers of macromolecular oxidative damage in horses  
420 with grass sickness. *Equine Vet. J.* **35**, 121-126.
- 421
- 422 [14] Newton, J.R., Hedderson, E.J., Adams, V.J., McGorum, B.C., Proudman, C.J. and Wood, J.L.  
423 (2004) An epidemiological study of risk factors associated with the recurrence of equine grass  
424 sickness (dysautonomia) on previously affected premises. *Equine Vet. J.* **36**, 105-112.
- 425
- 426 [15] Edwards, S.E., Martz, K.E., Rogge, A. and Heinrich, M. (2010) Edaphic and Phytochemical  
427 Factors as predictors of Equine Grass Sickness Cases in the UK. *Frontiers in Pharmacology* **1**,  
428 12.
- 429
- 430 [16] Greig, J.R. (1942) Grass sickness in horses. A review of the present knowledge of the disease,  
431 with particular reference to the nature of the casual agent. *Transactions of the Highland and*  
432 *Royal Agricultural Society of Scotland* **54**, 1-27.
- 433
- 434 [17] Doxey, D.L., Milne, E.M., Gilmour, J.S. and Pogson, D.M. (1991) Clinical and biochemical  
435 features of grass sickness (equine dysautonomia). *Equine Vet. J.* **23**, 360-364.
- 436
- 437 [18] Wylie, C.E., Shaw, D.J., Fordyce, F.M., Lilly, A. and McGorum, B.C. (2014) Equine grass sickness  
438 in Scotland: a case-control study of signalment- and meteorology-related risk factors. *Equine*  
439 *Vet. J.* **46**, 64-71.
- 440
- 441 [19] Paterson, E., Campbell, C.D., Coull, M.C. and Shand, C.A. (2011) *Geochemical Atlas for Scottish*  
442 *Topsoils*, Ed: T.M. Institute, Aberdeen.
- 443
- 444 [20] BGS (1991) *Regional Geochemistry of the East Grampians Area*, British Geological Survey,  
445 Keyworth.
- 446
- 447 [21] BGS (1993) *Regional Geochemistry of Southern Scotland and Part of Northern England*, British  
448 Geological Survey, Keyworth.
- 449
- 450 [22] Berrow, M.L. and Ure, A.M. (1986) Trace element distribution and mobilization in Scottish soils  
451 with particular reference to cobalt, copper and molybdenum. *Environ Geochem Health* **8**, 19-  
452 24.
- 453
- 454 [23] Johnson, C., Breward, N., Ander, E.L. and Ault, L. (2005) G-BASE: baseline geochemical  
455 mapping of Great Britain and Northern Ireland. *Geochemistry: Exploration, Environment and*  
456 *Analysis* **5**, 347-357.
- 457

- 458 [24] Hosmer, D.W. and Lemeshow, S. (2000) Applied Logistic Regression: Second Edition, John  
459 Wiley & Sons. Inc, New York.
- 460  
461 [25] De Vos, W. and Tarvainen, T. (2006) *FOREGS Geochemical Atlas of Europe*, Geological Survey  
462 of Finland, Espoo.
- 463  
464 [26] Reimann, C. and de Caritat, P. (1998) *Chemical Elements in the Environment*, Springer, Berlin.
- 465  
466 [27] Anon (2005) *Guidance for Developing Ecological Soil Screening Levels*, Ed: U.S.E.P. Agency,  
467 Washington, U.S.A.
- 468  
469 [28] Galyean, M.L., Perino, L.J. and Duff, G.C. (1999) Interaction of cattle health/immunity and  
470 nutrition. *J. Anim. Sci.* **77**, 1120-1134.
- 471  
472 [29] de Auer, D.E., Ng, J.C. and Seawright, A.A. (1988) Assessment of copper and zinc status of farm  
473 horses and training thoroughbreds in south-east Queensland. *Aust. Vet. J.* **65**, 317-320.
- 474  
475 [30] Fergusson, J.E. (1990) *The Heavy Elements: Chemistry, Environmental Impact and Health*  
476 *Effects*, Pergamon Press, Oxford.
- 477  
478 [31] Kabata-Pendias, A. (2001) *Trace Elements in Soil and Plants*, CRC Press Inc, Boca Raton.
- 479  
480 [32] Leita, L., de Nobili, M., Muhlbachova, G., Mondini, C., Marchil, L. and Zerbi, G. (1995)  
481 Bioavailability and effects of heavy metals on soil microbial biomass survival during laboratory  
482 incubation. *Biol Fertil Soils* **19**, 103-108.
- 483  
484 [33] Brookes, P.C. and McGrath, S.P. (1984) Effects of metal toxicity on the size of the soil microbial  
485 biomass. *J Soil Sci* **35**, 341-347.
- 486  
487 [34] Suttle, N., Small, J. and Jones, D. (1995) Overestimation of copper deficiency in horses. *Vet.*  
488 *Rec.* **136**, 131.
- 489  
490 [35] Pearce, S.G., Firth, E.C., Grace, N.D. and Fennessy, P.F. (1999) The effect of high pasture  
491 molybdenum concentrations on the copper status of grazing horses in New Zealand. *New*  
492 *Zealand Journal of Agricultural Research* **42**, 93-99.
- 493  
494 [36] Stowe, H.D. and Herdt, T.H. (1992) Clinical assessment of selenium status of livestock. *J. Anim.*  
495 *Sci.* **70**, 3928-3933.
- 496  
497 [37] Horvath, C.J., Stowe, H.D. and Miller, E.R. (1992) Effects of monensin on selenium status and  
498 related factors in genetically hypo- and hyperselenemic growing swine. *Am. J. Vet. Res.* **53**,  
499 2109-2118.

- 500  
501 [38] Mertz, W. (1987) *Trace Elements in Human and Animal Nutrition*, Academic Press, London.
- 502  
503 [39] Greiner, M., Smid, J., Havelaar, A.H. and Muller-Graf, C. (2013) Graphical models and Bayesian  
504 domains in risk modelling: application in microbiological risk assessment. *Prev. Vet. Med.* **110**,  
505 4-11.
- 506  
507 [40] Firestone, S.M., Cogger, N., Ward, M.P., Toribio, J.A., Moloney, B.J. and Dhand, N.K. (2012)  
508 The influence of meteorology on the spread of influenza: survival analysis of an equine  
509 influenza (A/H3N8) outbreak. *PLoS ONE* **7**, e35284.
- 510  
511 [41] Sanchez, J., Stryhn, H., Flensburg, M., Ersboll, A.K. and Dohoo, I. (2005) Temporal and spatial  
512 analysis of the 1999 outbreak of acute clinical infectious bursal disease in broiler flocks in  
513 Denmark. *Prev. Vet. Med.* **71**, 209-223.
- 514  
515

516 Table 1a. Summary of the G-BASE (455 cases and 910 controls) and NSIS-derived (246 cases and 492 controls) geochemical parameters. G-  
517 BASE data are geometric means ( $\pm 95\%$  CI), except for *Molybdenum* - %  $>0$  mg kg<sup>-1</sup> ( $\pm 95\%$  CI). NSIS-derived data are medians ( $\pm$ IQR), except  
518 for *Selenium* – %  $>0.22$  mg kg<sup>-1</sup> ( $\pm 95\%$  CI).

Element	G-BASE		NSIS	
	Control	EGS	Control	EGS
<i>Aluminium (Al) [mg kg<sup>-1</sup>]</i>	-	-	31160 (26140-31300)	26140 (26140-31160)
<i>Barium (Ba) [mg kg<sup>-1</sup>]</i>	-	-	119.8 (119.8-125.8)	120 (118.8-125.8)
<i>Cadmium (Cd) [mg kg<sup>-1</sup>]</i>	-	-	1.3 (0.7-1.4)	0.7 (0.7-1.4)
<i>Calcium (Ca) [mg kg<sup>-1</sup>]</i>	7794 (7579-8080)	9224 (8795-9653)	2772 (2464-2772)	2466 (2464-2772)
<i>Chromium (Cr) [mg kg<sup>-1</sup>]</i>	144.9 (140.7-149.2)	132.4 (126.1-139.1)	41.3 (38.5-43.6)	38.5 (38.5-41.3)
<i>Cobalt (Co) [mg kg<sup>-1</sup>]</i>	31.1 (30.5-31.8)	30.5 (29.5-31.4)	7.5 (6.2-8.6)	6.2 (6.2-8.6)
<i>Copper (Cu) [mg kg<sup>-1</sup>]</i>	29.1 (28.2-30.1)	28.1 (26.76-29.56)	8.3 (7.1-12.15)	7.1 (7.1-12.2)
<i>Gallium (Ga) [mg kg<sup>-1</sup>]</i>	11.4 (11.3-11.6)	11.1 (10.8-11.4)	-	-
<i>Iron (Fe) [mg kg<sup>-1</sup>]</i>	44310 (43540-45080)	44170 (42980-45430)	25650 (21830-26330)	21830 (21830-25650)
<i>Lead (Pb) [mg kg<sup>-1</sup>]</i>	48.8 (46.7-51.1)	42.3 (39.69-45.2)	30.65 (21.7-30.65)	21.7 (21.7-30.65)
<i>Magnesium (Mg) [mg kg<sup>-1</sup>]</i>	8985 (8744-9286)	9347 (8924-9708)	4272 (3320-5115)	4272 (3320-4272)
<i>Manganese (Mn) [mg kg<sup>-1</sup>]</i>	1705 (1628-1705)	1705 (1628-1860)	423.2 (380.9-446.3)	423.2 (380.9-438.5)
<i>Nickel (Ni) [mg kg<sup>-1</sup>]</i>	-	-	22.7 (16.8-22.7)	16.8 (16.8-22.7)
<i>Phosphorous (P) [mg kg<sup>-1</sup>]</i>	-	-	922 (770-922)	781.5 (770-922)
<i>Potassium (K) [mg kg<sup>-1</sup>]</i>	18177 (17845-18509)	18924 (18509-19339)	3740 (2774-4894)	4448 (2774-4894)
<i>Sodium (Na) [mg kg<sup>-1</sup>]</i>	-	-	620 (456.5-717)	717 (456.5-717)
<i>Strontium (Sr) [mg kg<sup>-1</sup>]</i>	-	-	26.9 (24-33.7)	26.9 (24-33.7)
<i>Titanium (Ti) [mg kg<sup>-1</sup>]</i>	6360 (6240-6480)	6960 (6780-7140)	199 (178-221)	220.5 (178-249.5)
<i>Zinc (Zn) [mg kg<sup>-1</sup>]</i>	134.1 (128.9-139.4)	111.5 (106.6-116.7)	60.8 (46.75-61.3)	46.8 (46.8-61.3)
<i>Molybdenum (Mo) [%<math>&gt;0</math> mg kg<sup>-1</sup>]</i>	76.40% (73.4-79.1)	86.60% (83.1-89.6)	-	-
<i>Selenium (Se) [%<math>&gt;0.22</math> mg kg<sup>-1</sup>]</i>	-	-	60.60% (53.7-67.1)	75.00% (66.1-82.6)

519

520 Table 1b. Summary of the G-BASE (455 cases and 910 controls) and NSIS-derived (246 cases and 492 controls) geochemical parameters.  
 521 *Univariable conditional* coefficients, (Coef ( $\pm$  standard error)), odds ratios (OR ( $\pm$ 95% CI)) of log<sub>10</sub> transformed data - except for G-BASE *Mo*  
 522 and NSIS-derived *Se*, where it is percentage of horses. Associated Wald P-values are also given.

523

Element	G-BASE			NSIS		
	Coef	OR	P-value	Coef	OR	P-value
<i>Aluminium (Al) [mg kg<sup>-1</sup>]</i>	-	-	-	-0.318 (0.854)	0.04 (0.01-0.22)	<0.001
<i>Barium (Ba) [mg kg<sup>-1</sup>]</i>	-	-	-	-2.220 (0.896)	0.11 (0.02-0.63)	0.011
<i>Cadmium (Cd) [mg kg<sup>-1</sup>]</i>	-	-	-	-1.870 (0.429)	0.15 (0.07-0.36)	<0.001
<i>Calcium (Ca) [mg kg<sup>-1</sup>]</i>	1.580 (0.283)	4.85 (2.78-8.44)	<0.001	0.287 (0.688)	1.33 (0.35-5.13)	0.675
<i>Chromium (Cr) [mg kg<sup>-1</sup>]</i>	-0.892 (0.276)	0.41 (0.24-0.70)	0.001	-3.170 (0.787)	0.04 (0.01-0.20)	<0.001
<i>Cobalt (Co) [mg kg<sup>-1</sup>]</i>	-0.590 (0.454)	0.55 (0.23-1.35)	0.194	-1.600 (0.556)	0.20 (0.07-0.60)	0.003
<i>Copper (Cu) [mg kg<sup>-1</sup>]</i>	-0.306 (0.260)	0.74 (0.44-1.23)	0.239	-1.530 (0.491)	0.22 (0.08-0.57)	0.002
<i>Gallium (Ga) [mg kg<sup>-1</sup>]</i>	-0.809 (0.485)	0.28 (0.10-0.81)	0.020	-	-	-
<i>Iron (Fe) [mg kg<sup>-1</sup>]</i>	-0.090 (0.495)	0.91 (0.35-2.41)	0.856	-2.020 (0.795)	0.13 (0.03-0.63)	0.011
<i>Lead (Pb) [mg kg<sup>-1</sup>]</i>	-0.686 (0.190)	0.48 (0.33-0.71)	<0.001	-3.140 (0.726)	0.04 (0.01-0.18)	<0.001
<i>Magnesium (Mg) [mg kg<sup>-1</sup>]</i>	0.423 (0.304)	1.53 (0.84-2.77)	0.164	-0.643 (0.561)	0.53 (0.18-1.58)	0.252
<i>Manganese (Mn) [mg kg<sup>-1</sup>]</i>	0.269 (0.279)	1.31 (0.76-2.26)	0.336	-1.010 (0.770)	0.37 (0.08-1.65)	0.191
<i>Nickel (Ni) [mg kg<sup>-1</sup>]</i>	-	-	-	-2.010 (0.453)	0.13 (0.06-0.33)	<0.001
<i>Phosphorous (P) [mg kg<sup>-1</sup>]</i>	-	-	-	-2.130 (0.929)	0.12 (0.02-0.73)	0.022
<i>Potassium (K) [mg kg<sup>-1</sup>]</i>	1.730 (0.556)	5.63 (1.89-16.74)	0.002	-0.420 (0.497)	0.66 (0.25-1.74)	0.398
<i>Sodium (Na) [mg kg<sup>-1</sup>]</i>	-	-	-	2.180 (0.763)	8.85 (1.99-39.46)	0.004
<i>Strontium (Sr) [mg kg<sup>-1</sup>]</i>	-	-	-	-1.980 (1.020)	0.14 (0.02-1.02)	0.053
<i>Titanium (Ti) [mg kg<sup>-1</sup>]</i>	2.940 (0.518)	18.87 (6.84-52.08)	<0.001	4.350 (0.988)	77.15 (10.91-545.84)	<0.001
<i>Zinc (Zn) [mg kg<sup>-1</sup>]</i>	-1.430 (0.274)	0.24 (0.14-0.41)	<0.001	-2.350 (0.718)	0.10 (0.02-0.39)	0.001
<i>Molybdenum (Mo) [%&gt;0 mg kg<sup>-1</sup>]</i>	0.683 (0.159)	1.98 (1.45-2.70)	<0.001	-	-	-
<i>Selenium (Se) [%&gt;0.22 mg kg<sup>-1</sup>]</i>	-	-	-	0.925 (0.343)	2.52 (1.29-4.94)	0.007



524 Table 2. Summary of the *multivariable conditional* model analyses carried out for the (a) G-BASE and (b) NSIS-derived elements. Here  
 525 variables were inserted into the initial step-wise deletion final model by considering the variables that were statistically significant in Table 1  
 526 (except *Se* and *Co* for the NSIS data) and taking one variable per cluster of correlated variables outlined in Supplementary Table 1. OR = odds  
 527 ratio ( $\pm 95\%$  CI), P-value – Wald’s P-value for individual terms.

Initial model	Final model	Likelihood ratio (LR) test	OR	P-value
<b>(a) G-BASE</b>				
<i>K + Ti + Zn + Mo</i>	$\log_{10} Ti$	LR=64.4, df=2, P<0.001	17.28 (6.39-46.71)	<0.001
	+ $\log_{10} Zn$		0.24 (0.14-0.41)	<0.001
<b>(a) NSIS-derived</b>				
<i>Cr + Ti</i>	$\log_{10} Cr$	LR=31.2, df=2, P<0.001	0.09 (0.02-0.38)	<0.001
	+ $\log_{10} Ti$		37.99 (5.36-269.43)	<0.001

528

529

530 Figure 1. Distribution of (a) cases of EGS and (b) controls across Scotland. (c) an interpolated  
531 map of Cu concentrations from the BGS G-BASE stream sediment dataset; (d) Major Soil  
532 Associations. Map source for (a) and (b): 2001 Census Output Area Boundaries. Crown  
533 copyright 2003. Crown copyright material is reproduced with the permission of the Controller  
534 of HMSO.

535

536

537

538

539

540 Supplementary Figure 1. Distribution plots of the (a) G-BASE and (b) NSIS-derived  
541 geochemical elements sub-divided into whether location is associated with a control or a case  
542 horse. Width of each distribution is proportional to the most common value for that element  
543 within controls or cases. Horizontal bars black (controls) and white (cases) are the median  
544 value for that element.

545 Supplementary Table 1 – Spearman rank correlation coefficients of association ( $r_s$ ) between (a) G-BASE variables for the 455 case-control sets;  
 546 (b) NSIS-derived variables for the 246 case-control sets where complete NSIS-derived data were available. Numbers in **bold italic** have  $P < 0.01$  &  
 547 those in **red**  $r_s > |0.5|$ .

(a)	Cr	Co	Cu	Ga	Fe	Pb	Mg	Mn	K	Ti	Zn
Ca	<b>-0.302</b>	0.047	<b>0.195</b>	<b>-0.123</b>	<b>0.205</b>	<b>0.070</b>	<b>0.323</b>	<b>0.092</b>	0.062	<b>0.582</b>	<b>-0.073</b>
Cr		0.053	0.066	-0.047	0.062	<b>0.088</b>	<b>0.192</b>	-0.021	<b>-0.259</b>	<b>-0.137</b>	<b>0.131</b>
Co			<b>0.548</b>	<b>0.499</b>	<b>0.728</b>	<b>0.438</b>	0.054	<b>0.614</b>	<b>-0.362</b>	<b>0.161</b>	<b>0.584</b>
Cu				<b>0.245</b>	<b>0.508</b>	<b>0.653</b>	0.047	<b>0.155</b>	<b>-0.311</b>	<b>0.269</b>	<b>0.581</b>
Ga					<b>0.443</b>	<b>0.214</b>	<b>0.278</b>	<b>0.282</b>	<b>0.152</b>	<b>0.084</b>	<b>0.327</b>
Fe						<b>0.385</b>	<b>0.220</b>	<b>0.539</b>	<b>-0.284</b>	<b>0.373</b>	<b>0.488</b>
Pb							<b>-0.229</b>	<b>0.212</b>	<b>-0.357</b>	<b>0.087</b>	<b>0.755</b>
Mg								0.007	<b>0.373</b>	<b>0.360</b>	<b>-0.218</b>
Mn									<b>-0.181</b>	<b>0.160</b>	<b>0.331</b>
K										0.008	<b>-0.488</b>
Ti											-0.045

548

(b)	Ba	Cd	Ca	Cr	Co	Cu	Fe	Pb	Mg	Mn	Ni	P	K	Na	Sr	Ti	Zn
Al	<b>0.124</b>	<b>0.713</b>	0.074	<b>0.947</b>	<b>0.783</b>	<b>0.657</b>	<b>0.934</b>	<b>0.786</b>	<b>0.438</b>	<b>0.448</b>	<b>0.865</b>	<b>0.768</b>	-0.038	<b>-0.198</b>	<b>0.209</b>	<b>-0.319</b>	<b>0.718</b>
Ba		<b>-0.236</b>	<b>-0.654</b>	<b>0.141</b>	<b>-0.185</b>	-0.015	0.044	<b>0.365</b>	<b>0.376</b>	<b>-0.138</b>	0.015	<b>-0.379</b>	<b>0.743</b>	0.047	<b>-0.165</b>	<b>-0.121</b>	0.001
Cd			<b>0.456</b>	<b>0.737</b>	<b>0.930</b>	<b>0.870</b>	<b>0.749</b>	<b>0.617</b>	-0.008	<b>0.514</b>	<b>0.838</b>	<b>0.719</b>	<b>-0.599</b>	<b>-0.445</b>	<b>0.607</b>	<b>-0.523</b>	<b>0.890</b>
Ca				<b>0.109</b>	<b>0.506</b>	<b>0.352</b>	<b>0.230</b>	<b>-0.110</b>	-0.048	<b>0.200</b>	<b>0.243</b>	<b>0.287</b>	<b>-0.671</b>	<b>0.128</b>	<b>0.400</b>	<b>0.125</b>	<b>0.342</b>
Cr					<b>0.763</b>	<b>0.645</b>	<b>0.861</b>	<b>0.717</b>	<b>0.461</b>	<b>0.491</b>	<b>0.929</b>	<b>0.751</b>	-0.059	<b>-0.237</b>	<b>0.170</b>	<b>-0.377</b>	<b>0.706</b>
Co						<b>0.853</b>	<b>0.831</b>	<b>0.658</b>	<b>0.168</b>	<b>0.486</b>	<b>0.861</b>	<b>0.670</b>	<b>-0.484</b>	<b>-0.280</b>	<b>0.629</b>	<b>-0.387</b>	<b>0.892</b>
Cu							<b>0.700</b>	<b>0.681</b>	-0.087	<b>0.412</b>	<b>0.689</b>	<b>0.510</b>	<b>-0.566</b>	<b>-0.466</b>	<b>0.605</b>	<b>-0.509</b>	<b>0.961</b>
Fe								<b>0.721</b>	<b>0.458</b>	<b>0.609</b>	<b>0.838</b>	<b>0.759</b>	<b>-0.108</b>	-0.041	<b>0.235</b>	<b>-0.136</b>	<b>0.715</b>
Pb									<b>0.097</b>	<b>0.319</b>	<b>0.691</b>	<b>0.470</b>	0.018	<b>-0.455</b>	<b>0.400</b>	<b>-0.559</b>	<b>0.774</b>
Mg										<b>0.210</b>	<b>0.416</b>	<b>0.196</b>	<b>0.571</b>	<b>0.652</b>	<b>-0.337</b>	<b>0.421</b>	-0.024
Mn											<b>0.593</b>	<b>0.563</b>	<b>-0.184</b>	0.016	<b>-0.123</b>	<b>0.112</b>	<b>0.387</b>
Ni												<b>0.780</b>	<b>-0.194</b>	<b>-0.245</b>	<b>0.319</b>	<b>-0.391</b>	<b>0.776</b>
P													<b>-0.356</b>	<b>-0.266</b>	<b>0.102</b>	<b>-0.221</b>	<b>0.548</b>
K														<b>0.438</b>	<b>-0.529</b>	<b>0.290</b>	<b>-0.505</b>
Na															<b>-0.486</b>	<b>0.863</b>	<b>-0.469</b>
Sr																<b>-0.631</b>	<b>0.640</b>
Ti																	<b>-0.569</b>

549

550 Supplementary text: Sensitivity analysis: G-BASE

551

552 As part of a sensitivity analysis of what variables were being entered into the initial model a  
553 series of additional analyses were carried out. First data from the seven elements significant in  
554 univariable analyses were entered into a *multivariable conditional* model with a final model  
555 obtained. Then additional models were run excluding either specific single or pairs of variables  
556 and final models obtained.

557 From the initial model all the seven elements higher *Ca* and *Ti* concentrations (OR>2.42) and  
558 lower *Zn* concentrations (OR=0.25) associated with being a case (Supplementary Table 2a).  
559 Excluding any one of the other four elements from the initial *multivariable conditional* models  
560 did not result in changes to the final model, and excluding *Ca* or *Ti* alone resulted in a model  
561 with the two other elements remaining (*i.e.*  $\uparrow Ti \downarrow Zn / \uparrow Ca \downarrow Zn$  – Supplementary Table 2a). The  
562 exception was where *Zn* was excluded, where a lower concentration of *Pb* was now associated  
563 with being a case (OR=0.47, Supplementary Table 2a). *Pb* and *Zn* were associated with the  
564 highest correlation coefficient ( $r_s=0.755$ , Supplementary Table 1a).

565 One of these four final *multivariable conditional* models ( $\uparrow Ca \uparrow Ti \downarrow Zn$ ,  $\uparrow Ca \uparrow Ti \downarrow Pb$ ,  $\uparrow Ti \downarrow Zn$ ,  
566 and  $\uparrow Ca \downarrow Zn$ ) were obtained with all but four of the models where pairs of elements were  
567 dropped (Supplementary Table 2a), and even in two of these four other models where *Ca+Zn*  
568 and *Ti+Zn* were dropped, they resulted in models with the other two elements *i.e.*  $\uparrow Ti \downarrow Pb$  and  
569  $\uparrow Ca \downarrow Pb$ , respectively (results not shown). The exception to just these four elements were in  
570 the final two *multivariable conditional* models where a higher percentage of *Mo*>0 mg kg<sup>-1</sup>  
571 (OR>1.59) combined with lower *Zn* (if *Ca* and *Ti* excluded) or higher *Ti* (if *Pb* and *Zn* were  
572 excluded) were associated with being a case (Supplementary Table 2a).

573

574 Sensitivity analysis: NSIS

575  
576 The correlations between NSIS variables resulted in even more complicated sensitivity  
577 *multivariable conditional* analyses for the NSIS dataset. While  $Se > 0.22 \text{ mg kg}^{-1}$  had to be  
578 excluded from the sensitivity due to Cox model singularity problems; initial exploration found  
579 that *Co* was dominating the final models, but its associated OR had changed from being  
580 apparently highly protective in the univariable analyses (OR=0.2) to apparently a very big risk  
581 factor (OR=222.5) in a final *multivariable conditional* model (results not shown). This lack of  
582 confidence in the OR estimate resulted in *Co* also being excluded from the NSIS *multivariable*  
583 *conditional* sensitivity models.

584 If data from the other nine elements significant in the univariable analyses were then entered  
585 into a NSIS *multivariable conditional* sensitivity model a final model of lower *Cr*  
586 concentrations (OR=0.04) and greater *Na* concentrations (OR=10.52) were associated with  
587 being a case (Supplementary Table 2b). Dropping any of *Cd*, *Cu*, *Ni* or *Ti* from the initial  
588 *multivariable conditional* model resulted in no change to the final model (Supplementary Table  
589 2b). However, if any of *Al*, *Pb*, *Na* or *Zn* were excluded then while *Cr* remained in the final  
590 *multivariable conditional* model, *Na* was replaced by *Ti*, with greater concentrations associated  
591 with being a case (OR=38.0, Supplementary Table 2b). In addition, if *Cr* was excluded, then  
592 while *Na* again dropped out of the final model to be replaced by *Ti* (OR=20.2), the model now  
593 included lower concentrations of *Ni* (OR=0.24) being associated with being a case.

594 One of these three final *multivariable conditional* models ( $\downarrow Cr \uparrow Na$ ,  $\downarrow Cr \uparrow Ti$ , and  $\downarrow Ni \uparrow Ti$ ) was  
595 obtained with all but five of the *multivariable conditional* models where pairs of elements were  
596 dropped from the initial model (Supplementary Table 2b). The only exceptions were if *Cr* and  
597 either *Cd*, *Cu* or *Ni* were excluded a final model consisting of lower concentrations of *Al*  
598 (OR=0.11) together with greater concentrations of *Ti* (OR=37.3) being associated with a case,  
599 and if *Ti* and either *Cr* or *Na* were excluded then the final model collapsed to just lower

600 concentrations of *Ni* (OR=0.13) being associated with being a case (Supplementary Table 2b).

601 Supplementary Table 2. Summary of the sensitivity *multivariable conditional* model analyses carried out for the (a) G-BASE and (b) NSIS-  
602 derived elements. Here either initial step-wise deletion final model with all the variables significant in Table 1 (except *Se* and *Co* for the NSIS  
603 data) are shown first, or sub-set *multivariable conditional* models where one or two parameters were initially excluded. OR = odds ratio ( $\pm 95\%$   
604 CI), P-value – Wald’s P-value for individual terms.

605

(a) G-BASE				
Initial model	Final model	Likelihood ratio (LR) test	OR	P-value
<i>Ca + Cr + Pb + K + Ti + Zn + Mo</i>	$\log_{10} Ca$	LR=71.6, df=3, P<0.001	2.43 (1.26-4.68)	0.007
	+ $\log_{10} Ti$		7.06 (2.17-23.01)	0.003
	+ $\log_{10} Zn$		0.25 (0.14-0.43)	<0.001
<i>Cr, K, Pb, Mo</i> excluded (singly or as pairs)	$\log_{10} Ca$	LR=71.6, df=3, P<0.001	2.43 (1.26-4.68)	0.007
	+ $\log_{10} Ti$		7.06 (2.17-23.01)	0.003
	+ $\log_{10} Zn$		0.25 (0.14-0.43)	<0.001
<i>Ca</i> excluded / <i>Ca + Cr</i> or <i>Pb</i> or <i>K</i> or <i>Mo</i> excluded	$\log_{10} Ti$	LR=64.4, df=2, P<0.001	17.28 (6.39-46.71)	<0.001
	+ $\log_{10} Zn$		0.24 (0.14-0.41)	<0.001
<i>Ti</i> excluded / <i>Ti + Cr</i> or <i>Pb</i> or <i>K</i> or <i>Mo</i> excluded	$\log_{10} Ca$	LR=60.8, df=2, P<0.001	4.43 (2.55-7.69)	<0.001
	+ $\log_{10} Zn$		0.26 (0.15-0.44)	<0.001
<i>Zn</i> excluded / <i>Zn + Cr</i> or <i>K</i> or <i>Mo</i> excluded	$\log_{10} Ca$	LR=57.8, df=3, P<0.001	2.77 (1.45-5.31)	0.002
	+ $\log_{10} Ti$		7.02 (2.12-23.25)	0.001
	+ $\log_{10} Pb$		0.47 (0.32-0.70)	<0.001
<i>Ca + Ti</i> excluded	$\log_{10} Zn$	LR=39.4, df=2, P<0.001	0.29 (0.16-0.52)	<0.001
	+ % <i>Mo</i> >0		1.60 (1.15-2.21)	0.005
<i>Pb + Zn</i> excluded	$\log_{10} Ti$	LR=44.1, df=2, P<0.001	11.70 (4.14-33.10)	<0.001
	+ % <i>Mo</i> >0		1.72 (1.24-2.38)	0.001

606

607

(b) NSIS-derived				
Initial model	Final model	Likelihood ratio test	OR	P-value
<i>Al + Cd + Cr + Cu + Pb + Ni + Na + Ti + Zn</i>	$\log_{10} Cr$ + $\log_{10} Na$	LR=27.7, df=2, P<0.001	0.04 (0.01-0.19) 10.52 (2.30-48.06)	<0.001 0.002
<i>Cd or Cu or Ni or Ti excluded</i> / <i>Ti + Al or Cd or Cu or Pb or Ni or Zn</i> / <i>Cd + Cu or Ni</i> / <i>Cu + Ni excluded</i>	$\log_{10} Cr$ + $\log_{10} Na$	LR=27.7, df=2, P<0.001	0.04 (0.01-0.19) 10.52 (2.30-48.06)	<0.001 0.002
<i>Al or Pb or Na or Zn excluded</i> / <i>Al + Cd or Cu or Pb or Ni or Na or Zn</i> / <i>Cd + Pb or Na or Zn</i> / <i>Cu + Pb or Na or Zn</i> / <i>Pb + Ni or Na or Zn</i> / <i>Ni + Ni or Zn</i> / <i>Na + Zn excluded</i>	$\log_{10} Cr$ + $\log_{10} Ti$	LR=31.2, df=2, P<0.001	0.09 (0.02-0.38) 37.99 (5.36-269.43)	<0.001 <0.001
<i>Cr excluded</i> / <i>Cr + Al or Pb or Na or Zn excluded</i>	$\log_{10} Ni$ + $\log_{10} Ti$	LR=28.5, df=2, P<0.001	0.24 (0.09-0.62) 20.2 (2.49-163.47)	0.003 0.005
<i>Cr + Cd or Cu or Ni excluded</i>	$\log_{10} Al$ + $\log_{10} Ti$	LR=28.3, df=2, P<0.001	0.11 (0.02-0.49) 37.31 (5.28-263.56)	0.003 <0.001
<i>Ti + Cr or Na excluded</i>	$\log_{10} Ni$	LR=20.5, df=1, P<0.001	0.13 (0.06-0.33)	<0.001

608