INTERNAL MEDICINE JOURNAL

doi:10.1111/imj.14942

Scoring system to facilitate diagnosis of Gaucher disease

Atul Mehta,[1] Oliver Rivero-Arias,[2] Magy Abdelwahab,[3] Samantha Campbell,[4] Annabel McMillan,[5] Mark J. Rolfe,[6] Jeremy R. Bright[6] and David J. Kuter[7,8]

[1]Department of Haematology, University College London, and Departments of [4]Hepatology and Gastroenterology, Royal Free Hospital, and [5]Haematology, Royal Free Hospital, University College London School of Medicine, London, and [2]Nuffield Department of Population Health, National Perinatal Epidemiology Unit, University of Oxford, and [6]Oxford PharmaGenesis, Oxford, UK, [3]Department of Paediatric Haematology, Cairo University Paediatric Hospital, Cairo, Egypt, and [7]Department of Medicine, Harvard Medical School, and [8]Center for Hematology, Massachusetts General Hospital, Boston, Massachusetts, USA

Key words
consensus, delayed diagnosis, early diagnosis, receiver-operating characteristic curve, sensitivity and specificity.

Correspondence
Atul Mehta, Department of Haematology, University College London, London, WC1E 6BT, UK.
Email: atul.mehta1@nhs.net

Received 10 October 2019; accepted 10 June 2020.

Abstract

Background: Gaucher disease (GD) manifests heterogeneously and other conditions are often misdiagnosed in its place, leading to diagnostic delays. The Gaucher Earlier Diagnosis Consensus (GED-C) initiative proposed a point-scoring system (PSS) based on the signs and covariables that are most indicative of GD to help clinicians identify which individuals to test for GD.
Aims: To validate the PSS retrospectively in a test population including patients with GD and other conditions with overlapping manifestations.
Methods: Four cohorts of adults with GD, liver disease, haematological malignancy or immune thrombocytopenia were identified from hospital records. Clinical data were audited for GED-C factors identified as potentially indicative of GD and aggregate scores calculated (sum of scores/number of factors) based on published PSS weightings. Threshold discriminatory PSS scores, sensitivity and specificity were determined by receiver-operating characteristic analysis.
Results: Among 100 patients (GD, $n = 25$; non-GD, $n = 75$), analyses based on 11 possible factors estimated group mean (standard deviation) PSS scores of: GD ($n = 14$), 1.08 (0.25); non-GD ($n = 38$), 0.58 (0.31). Mean between-group difference (95% confidence interval) was −0.49 (−0.68, −0.31) and area under the receiver-operating characteristic analysis curve (95% confidence interval) was 0.88 (0.78, 0.97). A threshold PSS score of 0.82 identified all 14 patients with GD in the analysis set (100% sensitivity) and 27 of 38 patients in the non-GD group (71% specificity). Patients with liver disease and haematological malignancy were most likely to have manifestations overlapping GD.
Conclusions: Preliminary validation of the GED-C PSS discriminated effectively between patients with GD and those with overlapping signs.

Introduction

Gaucher disease (GD) is an autosomal recessive condition arising from mutations in the glucocerebrosidase gene that encodes lysosomal glucocerebrosidase.[1] Enzyme deficiency causes the substrate of glucocerebrosidase, glucosylceramide, to accumulate in the bone marrow, liver, lungs, spleen and brain, leading to the clinical

Funding: Analyses were funded by unrestricted independent medical education grants (CME-GBR-15473 and IMEGBR-14397) from Shire International GmbH (now part of Takeda), Zug, Switzerland.
Conflict of interest: The authors declare the following potential conflicts of interest: A. Mehta: advisory boards: Amicus, BioMarin, Protalix, Sanofi and Shire; research educational assignments: Amicus, BioMarin, Protalix, Sanofi and Shire. O. Rivero-Arias acted as data analyst consultant for Oxford PharmaGenesis. M. Abdelwahab: advisory boards: Sanofi; honoraria and travel expenses: Sanofi and Shire. S. Campbell and A. McMillan have nothing to disclose. M. J. Rolfe is an employee of Oxford PharmaGenesis Ltd. J. R. Bright was an employee of Oxford PharmaGenesis Ltd. at the time of study design and initiation. D. J. Kuter: advisory boards: Alnylam, Amgen, Argenx, Bioveratif, Dova, Fujifilm, Genzyme, GlaxoSmithKline, Novartis, Pfizer, Principia, Rigel, Shire and Syntimmune; consulting fees: Alnylam, Amgen, Argenx, Bioveratif, Dova, Fujifilm, Genzyme, GlaxoSmithKline, Kyowa-Kirin, Novartis, Pfizer, Rigel, Shionogi, Shire, Principia and Syntimmune.

manifestations of GD.[1] The disease has been classified as types 1, 2 and 3 or non-neuronopathic, acute and chronic neuronopathic GD respectively.[2,3] In the general population, the prevalence of GD typically ranges from 1/40 000 to 1/60 000, but rises to 1/800 among Ashkenazi Jews.[1] Type 1 GD is the most prevalent form in the Western hemisphere, with a global prevalence of 1/50 000–100 000.[4] The rare nature and heterogeneous presentation of GD may lead to multiple referrals across different medical specialties, often delaying diagnosis and treatment.[5] Moreover, delays may be compounded by poor awareness of GD among clinicians, and alternative diagnostic possibilities associated with greater mortality than GD take precedence at differential diagnosis.[2,5,6]

Diagnostic pathways tailored to GD specialists have been published,[4,7] but information designed to guide non-specialists is lacking. The Gaucher Earlier Diagnosis Consensus (GED-C) initiative is an ongoing global project designed to facilitate diagnosis of GD in clinical practice and across medical disciplines, ultimately aiming to reduce diagnostic delays. The GED-C initiative recruited 22 experts in GD, who used Delphi consensus methodology[8] to identify factors most likely to be indicative of type 1 GD.[9] Nine major factors were agreed (splenomegaly, thrombocytopenia, bone-related manifestations, anaemia, hyperferritinaemia, hepatomegaly, gammopathy, family history of GD, Ashkenazi Jewish ancestry), as well as 10 minor factors.[9]

The subsequent aim of the GED-C initiative was to use these factors to create a simple scoring-based algorithm for use by non-specialist physicians or patients as an aid for the diagnosis of GD. Scoring systems are an established aid in diagnosing several diseases, such as systemic lupus erythematosus.[10] The factors and a prototype point scoring system (PSS) generated by the GED-C panel are shown in Table 1.[9] Combining the weightings allocated to each factor in the prototype PSS allows an aggregate score to be calculated for each patient. The recommendation to test for GD would depend on whether this score exceeded a predetermined threshold value.[9] It is common to use diagnostic models to estimate whether signs and symptoms predict the presence of a condition.[11] Based on this approach, odds ratios can be used to weight each predictive factor significantly associated with that condition.[12,13] Irrespective of how weightings are derived, their collective discriminatory power should be tested retrospectively and prospectively to determine the sensitivity and specificity with which an outcome can be predicted based on a given aggregate score.

Here, we retrospectively investigated patients with GD and patients with overlapping clinical signs but no GD diagnosis. We first compared the signs and patient covariables at the time of diagnosis in a cohort of adults with GD with those in three groups of patients who also had liver and spleen enlargement (chronic liver disease (LD) with portal hypertension; haematological malignancy (HM)) and thrombocytopenia or bleeding (immune thrombocytopenia (ITP)); then, we externally validated the prototype PSS in this patient population to determine its power to discriminate between those with and without GD. We also report two prospective case studies of patients tested with the PSS and demonstrate the challenges of estimating a diagnostic model within a sample population of patients with signs and covariables identified as important in the GED-C initiative.

Table 1 Prototype point-scoring system for diagnostic testing in Gaucher disease (GD), based on factors identified as potentially indicative of type 1 GD (adapted from Mehta *et al.*)[9]

	Weighting	Clinical sign or covariable
Major signs and covariables	3 points	Splenomegaly (≥3 × normal)
	2 points	Thrombocytopenia, mild or moderate (platelet count, 50–140 × 10^9/L)
		Bone issues, including pain, crises, avascular necrosis and fractures
		Family history of GD
		Anaemia, mild or moderate (haemoglobin, F ≥90–130 g/dL; M ≥90–140 g/dL)
		Hyperferritinaemia, mild or moderate (serum ferritin, 300–1000 μg/L)
		Jewish ancestry
		Hepatomegaly, mild or moderate (≤3 × normal)
		Gammopathy, monoclonal or polyclonal
	1 point	Anaemia, severe (haemoglobin, <90 g/dL)
		Hyperferritinaemia, severe (serum ferritin, >1000 μg/L)
		Hepatomegaly, severe (>3 × normal)
		Thrombocytopenia, severe (platelet count, <50 × 10^9/L)
Minor signs and covariables	0.5 points†	Bleeding, bruising or coagulopathy
		Leukopenia

†Other minor signs potentially indicative of type 1 GD were identified in the Gaucher Earlier Diagnosis Consensus (elevated serum angiotensin-converting enzyme levels; growth retardation, including low bodyweight; low bone mineral density; fatigue; asthenia; gallstones; dyslipidaemia; family history of Parkinson disease),[9] but were excluded here because data were unavailable in the hospital records. F, female; M, male.

Methods

Patients and data collection

Clinical and laboratory data from a cohort of 25 adult patients with GD attending the Royal Free Hospital (London, UK) were audited. Comparative data for

cohorts of 25 adults with LD, HM and ITP attending the same hospital were extracted sequentially. Patients with LD were all candidates for liver transplantation and had cirrhosis with portal hypertension, patients with HM all had lymphoma and patients with ITP were all severely affected with refractory ITP requiring treatment with thrombopoietin receptor agonists. Baseline characteristics were those recorded at diagnosis in the GD cohort and at presentation for specialist assessment in the comparator cohorts. Data were collected relating to factors identified by the GED-C panel as potentially indicative of GD. These GD factors were categorised by: serum ferritin levels (normal (<300 μg/L) or high (≥300 μg/L)); haematological parameters (haemoglobin (sex-adjusted), normal (women >130 g/L; men >140 g/L) or low (women ≤130 g/L; men ≤140 g/L); platelet count, normal (≥140 × 10^9/L) or low (<140 × 10^9/L); white blood cell count, normal (>4.5–11 × 10^9/L) or low (≤4.5 × 10^9/L); presence or absence of gammopathy; presence or absence of bleeding/bruising/coagulopathy); organ-related characteristics (hepatomegaly, splenomegaly); and family history and other characteristics (Jewish ancestry, family history of GD, bone pain). Data were also collected for two factors not identified in the GED-C consensus (splenectomy, presence or absence of lymph-node enlargement); these two factors were excluded from the PSS validation but included in the diagnostic modelling. The regional ethics committee confirmed that no approval was needed for the analyses.

Analyses

Demographic characteristics and factors potentially indicative of GD were summarised descriptively at baseline or at presentation for specialist assessment for each condition (GD, LD, HM and ITP) and for non-GD (LD, HM and ITP combined). Standard statistical inference compared whether the frequency of factors was different between the GD and non-GD groups. Mean differences between groups were compared by parametric *t* test; frequencies between groups were compared using Pearson χ^2 test. STATA (version 15; StataCorp LLC, College Station, TX, USA) was used for all analyses.

PSS validation and ROC analysis

A score was estimated for each participant using the prototype PSS (Table 1). Scores for each factor were stratified (3 points, 2 points, 1 point and 0.5 points) based on the GED-C panel's consensus of the likelihood of their association with GD.[9] The highest score (3 points) was awarded to factors deemed the most important potential indicators of GD.[9] Participant scores were calculated as the sum of weights associated with each factor present in that participant, divided by the number of factors included. The parametric *t* test was used to compare mean scores for GD versus non-GD. Receiver-operating characteristic (ROC) analysis was used to determine the performance accuracy of the potential score in terms of sensitivity (detecting the condition when it is present) and specificity (not detecting the condition when it is absent).

A scenario analysis was conducted to understand the impact of including different numbers and combinations of factors in the prototype PSS, and to determine the threshold value needed to recommend a GD test in each scenario. The base case assumed that information would be available about splenomegaly and about five further factors (anaemia, thrombocytopenia, leukopenia, bleeding/bruising, bone pain) based on data from routine blood tests and patient histories. The PSS was also tested prospectively in two patients, who presented with relevant signs to the Center for Haematology, Massachusetts General Hospital (Boston, MA, USA).

Diagnostic modelling

Maximum-likelihood, logistic regression models were tested to determine odds ratios for potential GD risk factors. Categorical definitions for each factor are defined above. Optimism-corrected c-index estimates were calculated using internal bootstrap validation based on 500 simulations to assess the performance of the model.

Results

Patient population

Baseline characteristics

Adults with GD, LD, HM or ITP (n = 25 per group; N = 100) were included in the analysis (Table 2). In both the GD and non-GD groups just over half of individuals were men and most were white. The GD group included more patients with Jewish ancestry and more with a family history of GD than did other groups. Comparing clinical characteristics between the GD and other groups (Table 2), mean serum ferritin levels were generally higher in the GD than in the non-GD groups, but the difference only reached significance versus LD. Only patients with GD had undergone splenectomy; hepatomegaly and splenomegaly were significantly more frequent with GD than with other conditions, the exception being LD, in which splenomegaly was as prevalent. Mean haemoglobin levels suggested all groups were moderately anaemic; the ITP group was severely thrombocytopenic, the GD and LD groups were on average moderately thrombocytopenic; mean platelet counts in the HM group were in the normal range. Gammopathy

Table 2 Patient and disease characteristics at baseline

Parameter	GD (*n* = 25)	LD (*n* = 25)		HM (*n* = 25)		ITP (*n* = 25)	
			P-value		*P*-value		*P*-value
Demographic characteristics							
Male	14 (56)	16 (64)	0.56	14 (56)	1.00	13 (52)	0.78
Age, mean (SD)† (years)	31 (13.2)	48 (13.1)	–	72 (11.7)	–	59 (21.0)	–
Ethnicity			0.24		0.097		0.048
White	22 (88)	20 (80)		18 (72)		17 (68)	
Asian/Asian British	1 (4)	4 (16)		1 (4)		7 (28)	
Black/African/Caribbean/black British	0	1 (4)		2 (8)		0	
Other	0	0		4 (16)		1 (4)	
Missing	*2 (8)*						
Serum ferritin level, mean (SD) (μg/L)	857 (1340)	213 (242)	0.029	546 (1088)	0.42	224 (340)	0.060
Missing	*6 (24)*	*2 (8)*		*4 (16)*		*7 (28)*	
Haematology							
Haemoglobin, mean (SD) (g/L)	114 (21.3)	102 (18.8)	0.54	109 (24.0)	0.48	105 (26.0)	0.22
Missing	*3 (12)*						
Platelet count, mean (SD) ($\times 10^9$/L)	120 (86.8)	94 (70.9)	0.26	181 (77.9)	0.017	23 (26.5)	<0.001
Missing	*4 (16)*						
White blood cell count, mean (SD) ($\times 10^9$/L)	5 (2.9)	4 (1.8)	0.046	10 (4.7)	<0.001	5 (1.8)	0.47
Missing	*6 (24)*						
Gammopathy	9 (36)	19 (86)	<0.001	5 (26)	0.49	1 (8)	0.076
Missing	*0*	*3 (12)*		*6 (24)*		*13 (52)*	
Lymph-node enlargement	0	7 (32)	0.003	15 (60)	<0.001	4 (16)	0.041
Missing	*1 (4)*	*3 (12)*					
Organ-related							
Hepatomegaly	14 (67)	3 (12)	<0.001	1 (4)	<0.001	0	<0.001
Missing	*4 (16)*						
Splenectomy	5 (23)	0	0.012	0	0.012	0	0.012
Missing	*3 (12)*						
Splenomegaly	17 (77)	17 (68)	0.48	4 (16)	<0.001	0	<0.001
Missing	*3 (12)*						
Family history and other							
Jewish ancestry	9 (39)	0	0.006	5 (20)	0.15	4 (16)	0.072
Missing	*2 (8)*	*10 (40)*					
Family history of GD	8 (36)	0	0.003	0	<0.001	0	<0.001
Missing	*3 (12)*	*5 (20)*					
Bone pain	17 (81)	0	<0.001	2 (8)	<0.001	0	<0.001
Missing	*4 (16)*						

Data are *n* (%) unless stated otherwise. *P*-values are for comparison with the GD group. †GD, age at diagnosis; non-GD groups, age at last record used. GD, Gaucher disease; HM, haematological malignancy; ITP, immune thrombocytopenia; LD, liver disease; SD, standard deviation.

was more common with LD than with GD or HM, and only one case of gammopathy was recorded among patients with ITP, although gammopathy data were missing for about one-third of non-GD patients.

Prototype PSS validation and ROC analysis

Data for 11 factors identified in the audit were available from 52 patients. The prototype PSS seemed to discriminate robustly between patients with GD and other patients in the population (mean PSS difference (95% confidence interval (CI)) for non-GD score minus GD score, −0.49 (−0.68, −0.31)). Moreover, the diagnostic accuracy (i.e. sensitivity and specificity) of the prototype PSS using data from all 11 factors was high (ROC area under the curve (95% CI), 0.88 (0.78, 0.97); Figure 1). Scores equal to or greater than a threshold value of 0.82 identified all patients with GD in the analysis set. The ability of the prototype PSS to discriminate between different conditions in the test population is illustrated in Figure 2.

Scenario analysis of the PSS

Implementation of the PSS in clinical practice may be limited by data availability for all of the 11 factors

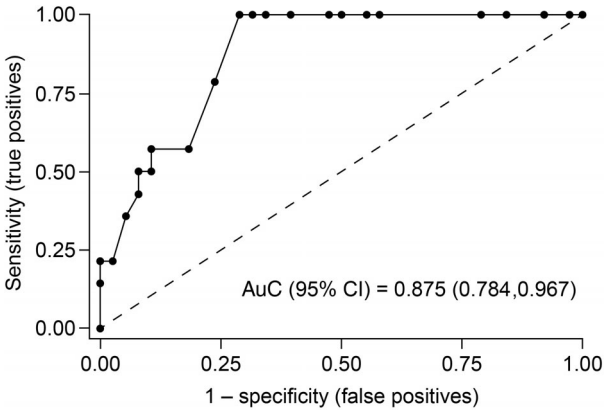

Figure 1 The receiver-operating characteristic (ROC) curve illustrating the accuracy† of the prototype point-scoring system (11 factors) for identifying Gaucher disease in the sample population (*n*/*N* = 52/100). †The closer the area under the ROC curve is to 1, the more accurate the diagnostic tool. *n*, number of patient records included in the analysis; *N*, sample population. AuC, area under the curve; CI, confidence interval.

assessed here. It was assumed that only patients with unexplained splenomegaly would be assessed with the PSS, thus different scenarios were examined starting with a base case of splenomegaly plus five other factors (anaemia, thrombocytopenia, leukopenia, bleeding/bruising and bone pain) for which information would typically be available (92 out of 100 patients in this analysis). Area under the ROC curve, the cut-off score necessary to achieve 100% sensitivity and the associated level of specificity attained in each of nine scenarios (including the optimum scenario of 11 factors) are summarised in Table 3; group mean data and between-group means for each scenario are summarised in Figure 3.

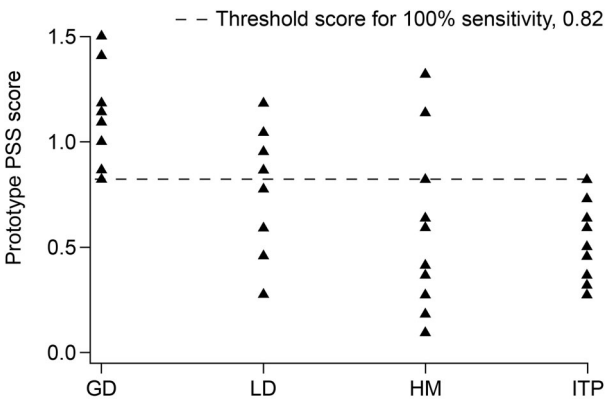

Figure 2 Scatter plot of prototype point-scoring system (PSS) scores (11 factors) by condition. GD, Gaucher disease; HM, haematological malignancy; ITP, immune thrombocytopenia; LD, liver disease.

Diagnostic accuracy was generally good across the different scenarios despite including fewer than the optimum number of 11 factors: the area under the ROC curve ranged from 0.75 to 0.90, the highest value being associated with the base case plus hepatomegaly. Indeed, adding hepatomegaly to the base case increased the level of specificity associated with 100% sensitivity more than any other factor (from 37% to 53%); only the specificity of the optimum scenario was greater (71%). Information about gammopathy and hyperferritinaemia improved specificity but did not increase that seen with hepatomegaly even when all three factors were included. Information about family history of GD or Jewish ancestry made little difference to the specificity of the base case in this analysis population.

Prospective case studies

Two patients presenting with factors included in the PSS were scored and tested prospectively for GD. Threshold discriminatory values were determined for these particular sets of factors using the same approach as applied in the scenario analysis. Both patients had PSS scores considerably greater than the threshold discriminatory values determined using test population data, and both tested positive for GD.

Patient 1

This was an asymptomatic 67-year-old Jewish patient (Jewish ancestry; 2 points) who had platelet counts in the range of 90–120 $\times 10^9$/L (mild to moderate thrombocytopenia; 2 points) and haemoglobin concentration in the range of 100–105 g/L (mild to moderate anaemia; 2 points) for decades. The patient had also had a slightly enlarged spleen for 3 years (14 cm on ultrasound with a 41 mm density (splenomegaly <3 × normal); 0 points) and a serum ferritin level of 401 μg/L (mild hyperferritinaemia; 2 points). Following serial computerised tomography scans of the spleen and subsequent laparoscopic surgical evaluation because of a differential diagnosis of lymphoma, the spleen was eventually removed, leading to a diagnosis of GD. Based on five factors possibly indicative of GD, the patient's PSS score was 1.60 (2 + 2 + 2 + 0 + 2 = 8; 8/5 = 1.60), well in excess of the theoretical threshold of 0.60.

Patient 2

This was a 19-year-old patient who had experienced 'growing pains' for 3–4 years (bone issues; 2 points), and had an enlarged spleen (28 cm long on ultrasound (splenomegaly ≥3 × normal); 3 points), a platelet count of 57 $\times 10^9$/L (mild to moderate thrombocytopenia; 2 points), easy bruising (bleeding/bruising/coagulopathy; 0.5 points) and a haemoglobin concentration of 128 g/L

Table 3 Prototype point-scoring system scenario analysis of different combinations of factors potentially indicative of Gaucher disease (GD)

Scenario	Major factors									Minor factors		*n*	AuC (95% CI)	Cut-off score for 100% sensitivity	Specificity at 100% sensitivity
	Splenomegaly	Hepatomegaly	Thrombocytopenia	Anaemia	Gammopathy	Hyperferritinaemia	Bone issues†	Jewish ancestry	Family history of GD	Bleeding‡	Leukopenia				
1. All factors	X	X	X	X	X	X	X	X	X	X	X	52	0.875 (0.784, 0.967)	0.818	71%
2. Base case§	X		X	X			X			X	X	92	0.819 (0.715, 0.923)	0.666	37%
3. Base case + hepatomegaly	X	X	X	X			X			X	X	92	0.900 (0.830, 0.971)	0.714	53%
4. Base case + hyperferritinaemia	X		X	X		X	X			X	X	76	0.808 (0.702, 0.915)	0.714	48%
5. Base case + gammopathy	X		X	X	X		X			X	X	70	0.745 (0.629, 0.862)	0.714	45%
6. Base case + all clinical information	X	X	X	X	X	X	X			X	X	61	0.826 (0.718, 0.934)	0.777	53%
7. Base case + Jewish ancestry	X		X	X			X	X		X	X	82	0.863 (0.767, 0.959)	0.571	39%
8. Base case + family history of GD	X		X	X			X		X	X	X	87	0.861 (0.773, 0.949)	0.571	40%
9. Base case + all family information	X		X	X			X	X	X	X	X	79	0.891 (0.811, 0.971)	0.500	40%

†Including pain, crises, avascular necrosis and fractures. ‡Including bruising or coagulopathy. §Base case was defined as splenomegaly plus thrombocytopenia, anaemia, bone pain, bleeding/bruising and leukopenia. AuC, area under curve; CI, confidence interval.

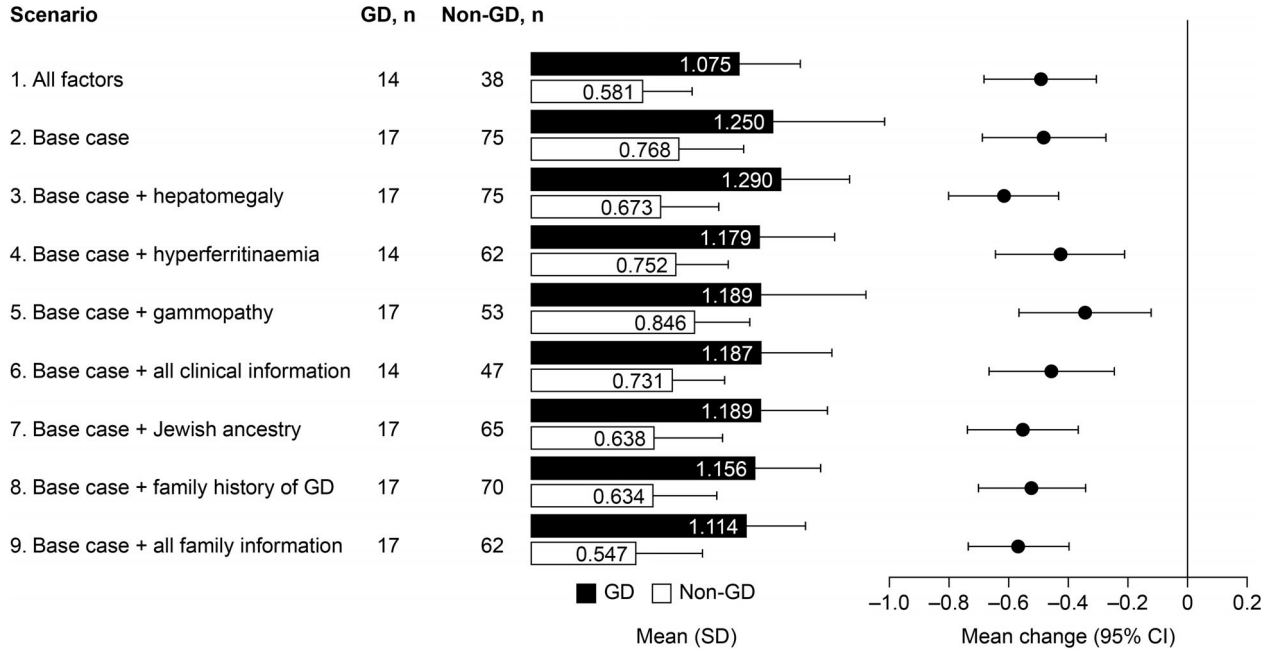

Figure 3 Group mean point-scoring system scores and difference between means in the scenario analysis. Base case – six factors were included: splenomegaly, anaemia, thrombocytopenia, leukopenia, bruising/bleeding and bone pain. Combinations of the base case and the other five factors are shown. CI, confidence interval; GD, Gaucher disease; SD, standard deviation.

(mild to moderate anaemia; 2 points). The patient had no Jewish ancestry (0 points). Initial clinical opinion was leukaemia but GD was diagnosed following bone marrow biopsy. Based on six factors possibly indicative of GD, the patient's PSS score was 1.58 (2 + 3 + 2 + 0.5 + 2 + 0 = 9.5; 9.5/6 = 1.58), well in excess of the theoretical threshold of 0.33.

Comparison of GD risk factors in GD and non-GD groups

Between-group comparison of factors potentially indicative of GD that were included in regression models are summarised in Table 4. Abnormalities of many of these factors were significantly more prevalent in the GD than in the non-GD group. Non-significant between-group differences were seen in haemoglobin concentration, platelet count, white blood cell count and the presence of gammopathy; values for comparisons of all factors are presented in Supporting Information Table S1.

Multivariate logistic regression analysis

Data for regression modelling were available from 49 patients with complete information for all GD risk factors. No factors were significant predictors of GD in the multivariate regression. Odds ratios and CI for each factor are shown in Table S2. The c-index (95% CI) estimate was 0.99 (0.98,1.00).

Discussion

The prototype PSS reported previously[9] has been adapted to create a simple scoring algorithm that predicted GD with a high level of sensitivity and specificity in a test population that included cohorts of patients with GD and of patients with other diagnoses but with signs overlapping those in GD. Several factors were significantly more common among patients with than without GD in this study and these probably contribute most to the PSS algorithm in terms of discriminatory power. It was striking that the prevalence of factors such as anaemia, thrombocytopenia, leukopenia and gammopathy, which are common in GD and contribute to the aggregate PSS score, were no more prevalent in GD than in the non-GD group. This perhaps illustrates why such factors in isolation or in combination offer clinicians little help in reaching a differential diagnosis of GD.

Predictive scoring systems are commonly generated using multivariate regression analyses to identify which factors are significant independent predictors of a given outcome.[12,13] Regression analysis in this test population, using the same factors as used in the PSS, yielded good-quality models that were not overfitted, but did not

Table 4 Between-group comparison of categorical factors potentially indicative of Gaucher disease (GD)

Factor, *n* (%)	Non-GD (*n* = 75)	GD (*n* = 25)	*P*-value
Serum ferritin levels			<0.001
Normal (<300 μg/L)	45 (60)	5 (20)	
Abnormal (≥300 μg/L)	17 (23)	14 (56)	
Missing	*13 (17)*	*6 (24)*	
Haematology			
Haemoglobin			0.42
Normal (F >130 g/L; M >140 g/L)	6 (8)	3 (12)	
Low (F ≤130 g/L;M ≤140 g/L)	69 (92)	19 (76)	
Missing	*0*	*3 (12)*	
Platelet count			0.63
Normal (≥140 × 10^9/L)	21 (28)	7 (28)	
Low (<140 × 10^9/L)	54 (72)	14 (56)	
Missing	*0*	*4 (16)*	
White blood cell count			0.53
Normal (>4.5–11 × 10^9/L)	41 (55)	10 (40)	
Leukopenia (≤4.5 × 10^9/L)	34 (45)	9 (36)	
Missing	*0*	*6 (24)*	
Gammopathy	25 (47)	9 (36)	0.35
Missing	*22 (29)*	*0*	
Lymph-node enlargement	26 (36)	0	<0.001
Missing	*3 (4)*	*1 (4)*	
Organ-related			<0.001
Hepatomegaly	4 (5)	14 (67)	
Missing	*0*	*4 (16)*	
Splenectomy	0	5 (23)	<0.001
Missing	*0*	*3 (12)*	
Splenomegaly	21 (28)	17 (77)	<0.001
Missing	*0*	*3 (12)*	
Family history and other			0.001
Jewish ancestry	9 (14)	9 (39)	
Missing	*10 (13)*	*2 (8)*	
Family history of GD	0	8 (36)	<0.001
Missing	*5 (7)*	*3 (12)*	
Bone pain	2 (3)	17 (81)	<0.001
Missing	*0*	*4 (16)*	

Data are presented as *n* (%) for categorical measures. Proportion differences between groups compared using Pearson χ^2 test. F, female; M, male.

identify any significant predictors among the factors tested. The CI associated with many of the factors included in the regression analyses were large, indicating that the models were underpowered.

One limitation of this study is that the PSS algorithm was developed using data from patients with established GD, which could potentially affect the ability of the PSS to detect early disease. However, perhaps the key limitation of our analysis is its generalisability, which cannot be ascertained without prospective validation in a larger and more diverse patient group. Testing a scoring system prospectively is an important part of validation but is inherently challenging with the GED-C PSS owing to the very low prevalence of GD in the general population. Prospective validation would have to be undertaken in populations in which GD is overrepresented, such as the Ashkenazi Jewish community. Even so, thousands of patients would have to be screened to power a prospective study adequately. Limited prospective testing to consolidate how the algorithm is applied and how well it performs may be possible by studying a population-level data set such as the Danish National Patient Registry.[14] Other options might involve specialists at GD centres, surgical groups scheduling splenectomy, and haematologists or hepatologists encountering splenomegaly, but these routes would probably yield too few patients for robust prospective validation. Given this, it seems unfeasible to pursue multivariate regression analysis prospectively in GD.

Rather than relying on clinical routes of validation, an alternative might be to focus on patients with no GD diagnosis but with a major sign of the disease such as splenomegaly, as was assumed in the base case described here. Reaching a diagnosis is possibly the most intractable problem in GD type 1, owing to its rarity and to clinicians' lack of awareness of the disease.[5] A patient-led route may complement existing referral mechanisms and may be more effective than trying to raise the profile of GD among general clinicians. Development of an online calculator based on the PSS for use by patients in partnership with their primary care provider may facilitate referral for GD testing. Linking an on-line version of the algorithm to search engine results relating to organomegaly or bone conditions may be feasible and could reach patients with relevant unexplained signs. An online calculator, or a checklist of major and minor signs, may also be useful on referral to specialist secondary care, perhaps saving patients from unnecessary bone marrow biopsy or even splenectomy.

In a primary care setting, availability of information for the various factors considered here may be limiting. Among patients with unexplained splenomegaly, primary care physicians should have ready access to data regarding anaemia, thrombocytopenia, leukopenia, bruising/bleeding and bone pain, and may have knowledge of Jewish ancestry or family history of GD. In our analyses, information about liver enlargement would seem to be the most useful addition diagnostically and providing such guidance as part of an online algorithm might help to expedite differential diagnosis. Our findings also suggest that in secondary and tertiary care, awareness of GD among hepatologists should be raised; whereas previous publications[4,15] have suggested diagnosis of GD is most likely delayed because of its similarity to HM, surprisingly just as many patients with LD share factors potentially indicative of GD. Based on ROC

analysis, all of the scenarios examined here had reasonable diagnostic accuracy but in order to capture all patients with GD, the level of specificity decreased notably when fewer than 11 factors were included. Despite this, results of two prospective case studies were encouraging.

Conclusions

Subject to further validation, the PSS algorithm reported here may prove to be a powerful discriminatory tool in predicting whether splenomegaly is attributed to GD. Reducing unnecessary investigations, time to diagnosis, and therefore to appropriate management, would alleviate some of the anxiety experienced by patients and their families and is critical to ensuring the best patient outcomes.

Acknowledgements

Medical writing and editorial support were provided by Oxford PharmaGenesis, Oxford, UK, and funded by the same unrestricted medical education grants. Shire International GmbH was not involved in the planning, design or delivery of the initiative. No reimbursement for involvement was offered to any of the authors or to their institutions.

References

1 Stirnemann J, Belmatoug N, Camou F, Serratrice C, Froissart R, Caillaud C *et al*. A review of Gaucher disease pathophysiology, clinical presentation and treatments. *Int J Mol Sci* 2017; **18**: 441.
2 Cassinerio E, Graziadei G, Poggiali E. Gaucher disease: a diagnostic challenge for internists. *Eur J Intern Med* 2014; **25**: 117–24.
3 Schiffmann R, Sevigny J, Rolfs A, Davies EH, Goker-Alpan O, Abdelwahab M *et al*. The definition of neuronopathic Gaucher disease. *J Inherit Metab Dis* 2020; **130**: 164–9.
4 Mistry PK, Cappellini MD, Lukina E, Özsan H, Mach Pascual S, Rosenbaum H *et al*. A reappraisal of Gaucher disease – diagnosis and disease management algorithms. *Am J Hematol* 2011; **86**: 110–5.
5 Mistry PK, Sadan S, Yang R, Yee J, Yang M. Consequences of diagnostic delays in type 1 Gaucher disease: the need for greater awareness among hematologists-oncologists and an opportunity for early diagnosis and intervention. *Am J Hematol* 2007; **82**: 697–701.
6 Thomas AS, Mehta AB, Hughes DA. Diagnosing Gaucher disease: an ongoing need for increased awareness amongst haematologists. *Blood Cells Mol Dis* 2013; **50**: 212–7.
7 Di Rocco M, Andria G, Deodato F, Giona F, Micalizzi C, Pession A. Early diagnosis of Gaucher disease in pediatric patients: proposal for a diagnostic algorithm. *Pediatr Blood Cancer* 2014; **61**: 1905–9.
8 Hsu C, Sandford B. The Delphi technique: making sense of consensus. *Pract Assess Res Eval* 2007; **12**: 1–8.
9 Mehta A, Kuter DJ, Salek SS, Belmatoug N, Bembi B, Bright J *et al*. Presenting signs and patient co-variables in Gaucher disease: outcome of the Gaucher earlier diagnosis consensus (GED-C) Delphi initiative. *Intern Med J* 2019; **49**: 578–91.
10 Petri M, Orbai AM, Alarcón GS, Gordon C, Merrill JT, Fortin PR *et al*. Derivation and validation of the Systemic Lupus International Collaborating Clinics classification criteria for systemic lupus erythematosus. *Arthritis Rheum* 2012; **64**: 2677–86.
11 Moons KG, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW *et al*. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med* 2015; **162**: W1–73.
12 Tomita K, Sano H, Chiba Y, Sato R, Sano A, Nishiyama O *et al*. A scoring algorithm for predicting the presence of adult asthma: a prospective derivation study. *Prim Care Respir J* 2013; **22**: 51–8.
13 Lykiardopoulos B, Hagström H, Fredrikson M, Ignatova S, Stål P, Hultcrantz R *et al*. Development of serum marker models to increase diagnostic accuracy of advanced fibrosis in nonalcoholic fatty liver disease: the new LINKI algorithm compared with established algorithms. *PLoS One* 2016; **11**: e0167776.
14 Schmidt M, Schmidt SA, Sandegaard JL, Ehrenstein V, Pedersen L, Sorensen HT. The Danish National Patient Registry: a review of content, data quality, and research potential. *Clin Epidemiol* 2015; **7**: 449–90.
15 Mehta A, Belmatoug N, Bembi B, Deegan P, Elstein D, Göker-Alpan Ö *et al*. Exploring the patient journey to diagnosis of Gaucher disease from the perspective of 212 patients with Gaucher disease and 16 Gaucher expert physicians. *Mol Genet Metab* 2017; **122**: 122–9.

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's web-site:

Table S1. Between-group comparison of baseline characteristics and GD risk factors.
Table S2. Odds ratios for GD risk factors.