Evidence-based detection of pulmonary arterial hypertension in systemic sclerosis: The DETECT study

Web Only file

APPENDICES

LIST OF INVESTIGATORS

We are indebted to all the local investigators and staff members for their collaboration and commitment, in particular to the following:

Austria: D Bonderman, H Olschewski, G Klein. **Bosnia**: S Sokolovic. **Canada**: JE Pope, P Lee, J-R Goulet, M Baron, E Sutton, J Markland, J Dunne, N Hirani, D Robinson, D Lien, D Langleben. **People's Republic of China**: M Li. **Czech Republic**: R Becvar. **Germany**: E Grünig, U Müller-Ladner, M Aringer, F Behrens, J Distler, P Willeke, C Baerwald, S Rosenkranz, J Thoden, G Riemekasten, D Skowasch, R Ewert, SM Weiner. **Hungary**: I Czuriga/G Szucs, R Faludi/L Czirjak. **Netherlands**: MC Vonk. **Norway**: JT Gran, E Rødevand, L-T Bertelsen. **Poland**: S Sierakowski. **Romania**: S Rednic. **Russia**: E Nasonov. **Slovakia**: J Lukac. **Spain**: M Egurbide, I Castellvi. **Switzerland**: O Distler, D Weilenmann, C Chizzolini. **Turkey**: M Inanc, I Ertenli, Y Kabasakal, S Ozbek. **United Kingdom**: CP Denton. **United States**: K Phillips, N Sandorfi, B Kahaleh, R Simms, A Goldberg, L Shapiro, N Rothfield, J Condemi, S Volkov, RW Martin, G Karpouzas, D Khanna, H Kenney, A Gerbino, T Osborn, J Molitor, ME Csuka, K Highland, M Cadeiras, M Harris, V Steen, L Hummers, G Mitri, G Rios, F Wigley.

METHODOLOGY

Appendix 1: Analysis sets and patient groups

Analysis sets and patient groups were defined as shown in Table S1. Patients were classified as non-pulmonary hypertension (PH), or World Health Organization (WHO) group 1 PH (PAH), WHO group 2 PH (PH due to left heart disease) or WHO group 3 PH (PH due to lung disease/hypoxia), according to current guidelines.[1,2] The WHO group 3 definition was based on Study Scientific Committee consensus.

Table S1. Analysis sets and patient groups

Analysis sets	**Definition**
All enrolled set (N=488)	All patients who had signed the consent form, met eligibility criteria and were enrolled in the study
RHC analysis set (N=466)	All patients with RHC test results who were in the PH group or in the non-PH group, as defined below
PAH analysis set (N=408)	All patients with RHC test results who were in the PAH group or in the non-PH group, as defined below
Patient groups	**Definition**
PH	mPAP ≥25 mm Hg at rest
WHO group 1 PH **(PAH)**	mPAP ≥25 mm Hg at rest and PCWP ≤15 mm Hg
WHO group 2 PH **(PH due to left heart disease)**	mPAP ≥25 mm Hg at rest and PCWP >15 mm Hg
WHO group 3 PH **(PH due to lung disease/hypoxia)**	mPAP ≥25 mm Hg at rest and PCWP ≤15 mm Hg and (FVC <60% or [FVC 60–70% and 'moderate-severe' parenchymal lung disease on HRCT or HRCT not available])
Non-PH	mPAP <25 mm Hg at rest

Abbreviations: FVC, forced vital capacity; HRCT, high resolution computed tomography; mPAP, mean pulmonary arterial pressure; PAH, pulmonary arterial hypertension; PH, pulmonary hypertension; PCWP, pulmonary capillary wedge pressure; RHC, right heart catheterisation; WHO, World Health Organization.

The analyses presented herein pertain to the PAH analysis set using the PAH and the non-PH patient groups only.

Appendix 2: Complete list of variables (number of variables)

Demographic and clinical parameters (68)

Demographics (3)

Gender; Age; Race.

Physical exam (8)

Systolic blood pressure; Diastolic blood pressure; Presence of peripheral oedema; Presence of crackles, rales or wheezing; Body mass index; WHO functional class (I/II vs. III/IV), 6-minute walk distance (6MWD), Borg dyspnoea index.

Systemic sclerosis (SSc) clinical characteristics (12)

SSc disease duration; SSc subtype (diffuse cutaneous vs. limited cutaneous); SSc subtype (overlap/mixed vs. limited cutaneous); Presence of digital ulcers in past 12 months; Modified Rodnan skin score; Physician global assessment scale; Physician skin disease activity assessment last year; Physician skin disease activity assessment last month; Physician skin disease activity assessment current overall; Overall scleroderma compared to 1 month ago; Overall scleroderma compared to 1 year ago; Raynaud's condition score.

Other SSc clinical characteristics (current/past) and medical history (33)

Raynaud's phenomenon; Calcinosis; Telangiectasias; Anginal pain; Syncope/near syncope; Palpitations; Pericarditis; Myocarditis; Dyspnoea; Cough; Proteinuria; Gastro-oesophageal reflux; Dysphagia; Aspiration; Diarrhoea; Biliary cirrhosis; Dyspareunia; Erectile dysfunction;

Arthralgia; Myalgia; Muscle weakness; Tendon friction rub; Loss in joint range of motion; Depression; Sjogren/Sicca syndrome; Fatigue/malaise; Systemic lupus erythematosus; Rheumatoid arthritis; Polymyositis; Dermatomyositis; Mixed connective tissue disease; History of any connective tissue disease; Current/previous smoker.

Pulmonary function tests and haemoglobin (12)

Forced vital capacity (FVC); FVC % predicted; Forced expiratory volume in 1 second (FEV1); FEV1 % predicted; Pulmonary diffusing capacity for carbon monoxide (DLCO); DLCO % predicted; DLCO/alveolar volume (DLCO/VA); DLCO/VA % predicted; Total lung capacity % predicted; Residual volume % predicted; FVC % predicted/DLCO % predicted; Haemoglobin.

Serum laboratory (13)

N-terminal pro-brain natriuretic peptide (NTproBNP); Endothelin-1; von Willebrand factor; C-reactive protein; Creatinine; Estimated glomerular filtration rate; Serum urate; Erythrocyte sedimentation rate; Anti-centromere antibody (ACA); Anti-topoisomerase-I (Scl-70) antibody; Anti-U3-RNP (fibrillarin) antibody; Anti-U1-RNP antibody; Anti-RNA polymerase antibody.

Electrocardiography (3)

Right ventricular strain; Right axis deviation (RAD); Right bundle branch block.

Echocardiography (28)

Aortic root; Left atrium; Inferior vena cava; Interventricular septum; Tricuspid annular plane systolic excursion (TAPSE); Posterior wall; Right atrium (RA) area; Right ventricle (RV) area; RV diameter; Left ventricle (LV) end-diastolic dimension; LV end-systolic dimension; Tissue Doppler imaging (TDI) tricuspid annulus S; TDI tricuspid annulus E'; TDI tricuspid annulus A'; TDI mitral annulus S; TDI mitral annulus E'; TDI mitral annulus A'; Pulsed-wave Doppler

mitral inflow E'; Pulsed-wave Doppler mitral inflow A'; Tricuspid regurgitant jet (TR) velocity; Pulmonary regurgitant velocity; Aortic valve (normal/abnormal); Mitral valve (normal/abnormal); Tricuspid valve (normal/abnormal); Pulmonary valve (normal/abnormal); Pericardial effusion (yes/no); Qualitative assessment of RV pump function; Qualitative assessment of LV pump function.

The total number of variables was 112.

Appendix 3: Analytical methods

Univariable logistic regression (ULR) and multivariable logistic regression (MLR) modelling were the main analytical methods, including linear and non-linear functional relations, where the binary outcome variable was PAH versus non-PH.

Variable selection in MLR was performed in different stages by means of stepwise forward procedure and clinical judgement (see Appendix 4 and Appendix 5).

Discriminatory performance to distinguish between PAH and non-PH patients was examined by receiver operating characteristic (ROC) curve analysis. A ROC curve shows the relationship between the true-positive rate (sensitivity on y-axis) and false-positive rate (1−specificity on x-axis). The ROC area under the curve (AUC), also called the concordance statistic (C-statistic), formed the criterion for assessing the discriminatory ability of a model. A risk prediction model with perfect discrimination (AUC=100%) has a ROC curve that passes through the upper left corner (100% sensitivity, 100% specificity) and pure chance discrimination (AUC=50%) has a ROC curve that is a diagonal line. The ROC AUC and its 95% confidence intervals were calculated for each ULR and MLR model.

A risk cut-off can be calculated to classify subjects as having PAH or non-PH for a fitted model. Among others, this can be done by pre-specifying either sensitivity or specificity

levels. A two-by-two classification table (Table S2) was created for each assessed cut-off showing frequencies, together with the discriminatory performance statistics: sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV), and their 95% confidence intervals, when relevant.

Table S2. Two-by-two classification of diagnostic test characteristics

	PAH (+ve) (by RHC)	Non-PH (-ve) (by RHC)	Total	
Algorithm (+ve) ≥x score cut-off	a (true positive)	b (false positive)	a+b	a/(a+b) **Positive predictive value (PPV)**
Algorithm (–ve) <x score cut-off	c (false negative)	d (true negative)	c+d	d/(c+d) **Negative predictive value (NPV)**
Total	a+c	b+d	a+b+c+d	
	a/(a+c) **Sensitivity**	d/(b+d) **Specificity**		

Note: (a + b)/(a + b + c + d) = RHC referral rate; c/(a + c) = missed PAH diagnoses rate.

Calibration of MLR models was assessed by the Hosmer-Lemeshow χ^2 goodness-of-fit test (HL-test), where a significant p-value implies lack of model fit.

Appendix 4: Variable reduction process by univariable and multivariable analysis

a) Methods

In order to manage a large number of variables (112) relative to the number of PAH patients, four groups of clinically related variables were formed for the model building process:

standard demographic and clinical variables; serum tests; electrocardiography (ECG); and echocardiography (ECHO). Selection of a final set of variables for detection purposes was performed in a series of model building stages to obtain a reduced set of good-performing variables. The first variable selection stage of the large set of potential candidate variables started after the ULR analysis, where only the statistically good-performing variables (Wald χ^2 test statistic p-value <0.15 for linear terms and <0.05 for quadratic terms) were selected and carried forward to the MLR analysis stage. The second variable selection stage consisted of MLR within the groups of variables to further reduce the set of potential candidate variables by stepwise forward selection procedure (SELECTION=STEPWISE method in SAS Proc Logistic; with entry criterion slentry =0.15 and retention criterion slstay =0.10). Interaction effect between two variables was tested for its statistical significance (Wald χ^2 test statistic p-value <0.15), whenever it was suspected to exist by the Study Scientific Committee. A review of the results of statistically good-performing variables was performed by the Study Scientific Committee and variables were further selected based on clinical plausibility and/or feasibility with particular regard to resource limitations in standard real-world practice. Clinical feasibility means that for variables related to the same clinical condition, if they had a similar performance, the easiest to measure in clinical practice was selected.

b) Results

Using the described sequence of statistical analyses, combined with clinical judgment, the initial set of 112 variables was reduced to 13 good-performing and clinically well-accepted variables (Table S3 and Table S4). These thirteen variables formed the basis for constructing the final detection algorithm as a two-step decision tree (see Appendix 5).

Table S3. Variable selection process

Variable selection step	Variable selection criteria	Number of variables selected				Total number of variables selected
		Demographic and clinical	Serum laboratory	ECG variables	ECHO variables	
All variables	–	68	13	3	28	112
Univariable logistic regression	$p<0.15$	25	8	2	12	47
Multivariable logistic regression	Entry criterion $p<0.15$, retention criterion $p<0.10$	7	4	2	4	17
Study Scientific Committee consensus	Clinical plausibility and practical feasibility	4	4	1	4	13
2-step decision tree (multivariable logistic regression)	Entry criterion $p<0.15$, retention criterion $p<0.10$	2	3	1	2	8

Table S4. Univariable analysis of the 13 selected variables of which eight were used for the DETECT algorithm

Variables	Summary Statistics		Univariable Logistic Regression	
	Non-PH patients	PAH patients	Wald Chi-square (p-value)	ROC AUC, % (95% CI)
Demographic and clinical characteristics				
Systemic sclerosis characteristics Current/past telangiectasias, n/N (%)	218/321 (67.9)	76/87 (87.4)	11.82 (<0.001)	59.7 (55.4, 64.1)
Physical examination Peripheral oedema present, n/N (%)	39/318 (12.3)	19/87 (21.8)	4.97 (0.026)	54.8 (50.0, 59.5)
WHO functional class, n/N (%) I/II III/IV	256/306 (83.7) 50/306 (16.3)	56/87 (64.4) 31/87 (35.6)	14.66 (<0.001)	59.6 (54.2, 65.1)
Pulmonary function tests, n FVC % predicted/DLCO % predicted mean (SD) median (Q1, Q3)	321 1.8 (0.5) 1.7 (1.4, 2.0)	87 2.2 (0.7) 2.1 (1.7, 2.5)	33.19 (<0.001)	71.5 (65.6, 77.4)
Serum tests				
n NTproBNP, log_{10} mean (SD) median (Q1, Q3) Serum urate, mg/100 mL mean (SD) median (Q1, Q3)	306 2.1 (0.5) 2.1 (1.8, 2.3) 4.7 (1.5) 4.5 (3.7, 5.4)	80 2.4 (0.5) 2.3 (2.0, 2.8) 5.9 (1.5) 5.8 (4.7, 6.8)	25.42 (<0.001) 30.16 (<0.001)	67.5 (60.9, 74.2) 71.9 (65.9, 77.9)

Anti-centromere antibody positive, n/N (%) Scl-70 positive, n/N (%)	77/306 (25.2) 91/299 (30.4)	40/80 (50.0) 12/79 (15.2)	17.64 (<0.001) 6.99 (0.008)	62.4 (56.4, 68.4) 57.6 (52.9, 62.4)
Electrocardiography				
Right axis deviation* present, n/N (%)	10/291 (3.4)	11/83 (13.3%)	10.19 (0.001)	54.9 (51.1, 58.7)
Echocardiography				
TAPSE, mm, n mean (SD) median (Q1, Q3)	284 22.7 (4.8) 23.0 (20.0, 26.0)	76 21.5 (4.3) 21.0 (19.0, 24.0)	3.92 (0.048)	59.0 (51.7, 66.2)
Right atrium area, cm^2, n mean (SD) median (Q1, Q3)	286 13.4 (4.7) 12.6 (10.2, 15.4)	82 17.1 (6.2) 15.7 (13.9, 19.0)	24.45 (<0.001)	71.2 (65.0, 77.3)
Right ventricle area, cm^2, n mean (SD) median (Q1, Q3)	291 15.0 (5.4) 14.7 (12.0, 17.4)	82 19.3 (6.8) 18.4 (13.7, 22.6)	25.81 (<0.001)	69.3 (62.5, 76.0)
TR velocity, m/s, n mean (SD) median (Q1, Q3)	303 2.4 (0.6) 2.4 (2.1, 2.7)	84 3.0 (0.8) 3.0 (2.5, 3.5)	62.28 (<0.001)	79.5 (73.7, 85.3)

*QRS axis ≥90°. Abbreviations: AUC, area under the curve; DLCO, pulmonary diffusing capacity for carbon monoxide; FVC, forced vital capacity; NTproBNP, N-terminal pro-brain natriuretic peptide; PAH, pulmonary arterial hypertension; PH, pulmonary hypertension; ROC; receiver operating characteristic; TAPSE, tricuspid annular plane systolic excursion; TR, tricuspid regurgitant jet; WHO, World Health Organization.

Appendix 5: Construction of PAH detection algorithm as a two-step decision tree

a) Methods

The 13 variables selected as good-performers during the above stages of model building were divided into nine non-ECHO variables and four ECHO variables to align the detection algorithm with real-world practice where the rheumatologist managing the patient accesses non-ECHO data prior to referral to a cardiologist for echocardiography. These two groups of variables formed the basis for developing a two-step detection algorithm as a decision tree, as shown in Figure 2 of the manuscript.

The development of the decision tree entailed building two models by MLR; the nine non-ECHO variables were candidate variables in the first step of the decision tree and the four ECHO variables were candidate variables in the second step of the decision tree. In order to carry forward non-ECHO PAH risk information, the step 1 linear risk prediction score was included at step 2 (Figure S1). The decision tree allows patients with a low risk of PAH to be filtered out in the first step, which saves further assessments by the cardiologist. A stepwise forward selection procedure was applied within each step where the variable selection criteria were: entry criterion 0.15 and retention criterion 0.10.

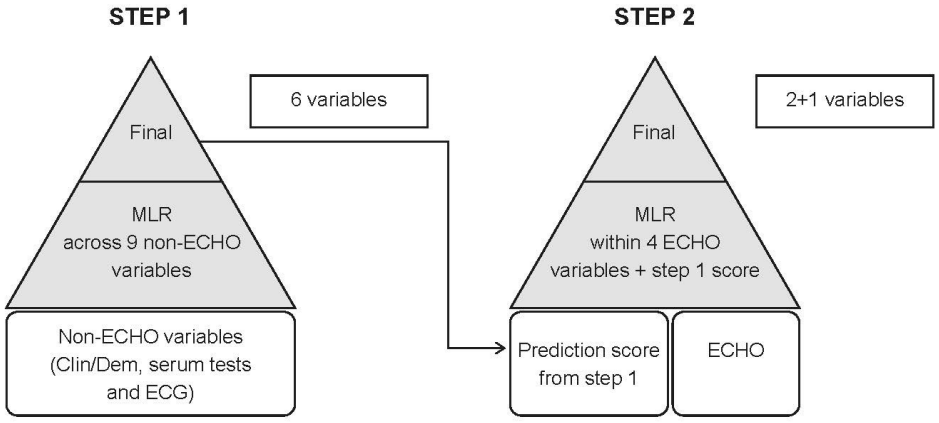
No significant lack of fit was observed for the final MLR model at step 1 (HL p-value =0.23), nor for the final MLR model at step 2 (HL p-value =0.74).

Quadratic terms for the continuous variables were included in the variable selection process in a hierarchical manner, that is, selection of a quadratic term implies selection of the corresponding linear term as well. For those variables where the quadratic term was significant, different approaches were applied to further examine non-linear relationships using: flexible non-parametric methods (Generalized Additive Models [GAM] with penalised

splines,[3] Restricted Cubic Splines,[4] or variable transformations [e.g., $\log_{10}$ or quintiles categorisation]).

Based on the prediction scores of the two MLR models of the decision tree, cut-offs were selected to classify patients for referral to echocardiography at step 1, and for referral to RHC at step 2. As there is a trade-off between sensitivity and specificity, different levels of sensitivity and specificity can be selected to identify cut-offs to achieve acceptable global performance characteristics of the detection algorithm. The objective of step 1 was to achieve a low rate of missed PAH diagnoses (i.e. low false-negative rate) by selecting a high sensitivity, which was fixed at 97%. In step 2, the level of specificity was selected in order to obtain a high PPV (high rate of positive PAH diagnoses). Therefore, a range of specificities (35% to 85%) was used to examine the impact on the rate of missed diagnoses.

Figure S1. Construction of the two-step decision tree algorithm



Abbreviations: Clin/Dem, demographic and clinical parameters; ECG, electrocardiography; ECHO, echocardiography; MLR, multivariable logistic regression.

b) Results

The forward stepwise selection procedure reduced the number of variables from nine to six in step 1 and from four to two in step 2. None of the statistically selected variables was replaced by another non-ECHO variable by the Study Scientific Committee at step 1, whereas one ECHO variable (RV area) was replaced by another ECHO variable (RA area) as the latter is regarded as easier to measure and likely to be more reproducible. This replacement had a minimal effect on the performance of the step 2 model (AUC of 89% and 88% using RV area and RA area, respectively).

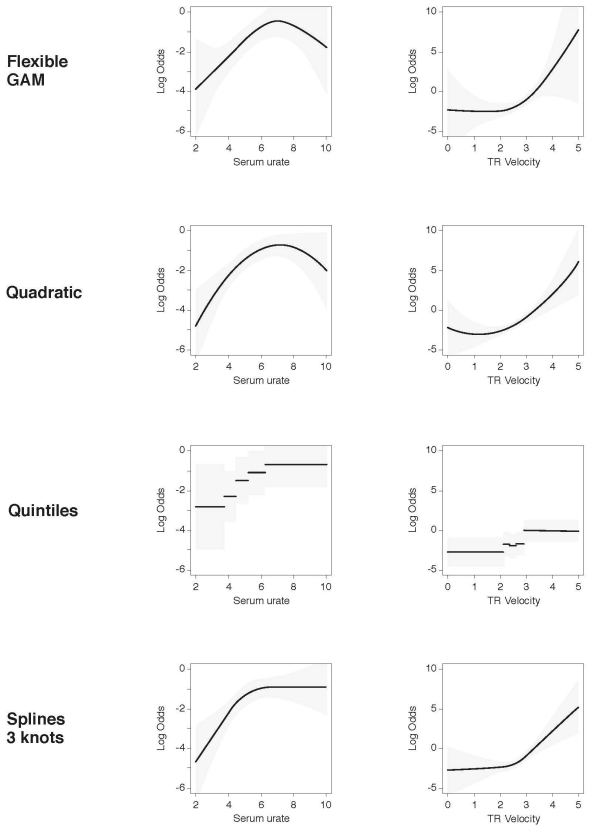
For serum urate at step 1 and TR velocity at step 2, it was necessary to fit a non-linear relationship. Different approaches were applied to adjust for non-linearity: flexible GAM, quadratic function, categorisation with quintiles, and flexible spline functions (Figure S2). Among all these options, no relevant differences were observed in terms of discrimination (same AUC=84% for all approaches). Splines with three knots were finally used in order to avoid the inversion of PAH risk at the extremes of the curves, as recommended by the Study Scientific Committee.

MLR model coefficient estimates and their statistical significance (Wald test) are given in Table 2 of the main manuscript. A positive relationship was observed between all final variables and PAH risk, as indicated by the positive value of the regression coefficients of the main terms.

The global performance of the detection algorithm at a pre-defined sensitivity of 97% at step 1 and different specificity levels at step 2 is given in Table S5. This table shows that as the specificity at step 2 increases, the global specificity and global PPV also increase. The table also shows the natural trade-off between sensitivity and specificity, i.e., as one increases the other decreases.

Characteristics of patients that were misdiagnosed as non-PAH by the decision algorithm are presented in Table S6. At step 1 (97% sensitivity) two out of 77 PAH patients who had complete data for step 1 were misdiagnosed. For a range of specificity levels between 35% and 85% at step 2, the higher the specificity level the greater the number of missed diagnoses. Characteristics of the 24 patients with missed PAH diagnoses based on current current ESC/ERS guidelines are presented in Table S7.

Figure S2. Different approaches to assess non-linear functional relationships of serum urate (step 1) and TR velocity (step 2) of the decision tree



Partial fitting for serum urate at step 1 (left column) and TR velocity at step 2 (right column) with different non-linear function. The Y-axis represents the log-odds of having PAH in the MLR model. Abbreviations: GAM, Generalized Additive Models; TR, tricuspid regurgitant jet.

Table S5. Global performance of the final two-step decision tree algorithm at 97% sensitivity at step 1 and different levels of specificity at step 2

Specificity step 2 (%)	Overall sensitivity (%)	Overall specificity (%)	Overall PPV (%)	Overall NPV (%)	Overall missed PAH diagnoses (false negatives, %)
35	96	48	35	98	4
40	89	52	35	94	11
45	89	56	37	94	11
50	88	60	39	94	12
55	88	64	41	95	12
60	88	68	44	95	12
65	85	72	47	94	15
70	83	78	50	94	17
75	80	80	54	93	20
80	80	83	60	94	20
85	74	88	64	92	26

Abbreviations: NPV, negative predictive value; PAH, pulmonary arterial hypertension; PPV, positive predictive value.

Table S6. Details of the patients with missed PAH diagnoses using a sensitivity of 97% at step 1 and a range of specificities at step 2

			Step 1 variables						Step 2 variables	
Step	**Criteria**	**Cumulative number of missed PAH diagnoses**	**RAD (presence)**	**Telang. (presence)**	**ACA (presence)**	**NT-proBNP**	**FVC % pred./DLCO % pred.**	**Serum urate**	**RA area**	**TR velocity**
1	Sens 97%	2	No	Yes	No	204	1.70	2.3	16.4	2.90
	Sens 97%		No	No	No	47	1.74	3.7	15.0	2.52
2	Spec 35%	+1=3	No	Yes	No	71	2.21	3.8	8.5	2.59
	Spec 45%	+5=8	No	Yes	No	25	1.70	5.2	13.0	2.30
			No	Yes	No	47	1.32	4.6	17.7	2.52
			No	Yes	No	81	1.32	3.6	18.0	2.90
			No	No	Yes	210	1.45	6.1	10.0	2.14
			No	Yes	No	102	1.38	4.3	13.0	2.71
	Spec 60%	+1=9	No	Yes	No	44	2.37	4.9	8.8	2.25
	Spec 65%	+2=11	No	Yes	Yes	71	1.46	3.5	7.0	3.20
			No	Yes	Yes	46	2.07	4.3	16.7	2.00
	Spec 70%	+1=12	No	Yes	No	187	2.29	7.3	11.5	1.06
	Spec 80%	+2=14	No	Yes	No	41	1.85	6.3	14.0	2.78
			No	Yes	No	155	2.04	5.4	15.8	2.40
	Spec 85%	+5=19	No	Yes	Yes	95	2.29	4.5	10.0	2.75
			No	Yes	Yes	135	1.86	5.5	16.0	2.25
			No	Yes	Yes	215	1.92	5.2	10.3	2.67
			No	Yes	No	287	1.96	7.6	16.9	2.53
			No	Yes	Yes	166	1.78	6.3	11.4	2.65

Abbreviations: ACA, anti-centromere antibody; DLCO, pulmonary diffusing capacity for carbon monoxide; FVC, forced vital capacity; NTproBNP, N-terminal pro-brain natriuretic peptide; PAH, pulmonary arterial hypertension; RA, right atrium; RAD, right axis deviation; Telang; telangiectasias; TR, tricuspid regurgitant jet.

Table S7. Characteristics of the 24 patients with missed PAH diagnoses using the ESC/ERS guidelines[1]

mPAP (mm Hg)	PCWP (mm Hg)	PVR (dyn·sec/cm^5)	TR velocity (m/s)	RA area (cm^2)	NTproBNP (pg/mL)	RAD (presence)	ACA (presence)
36	12	400.00	3.30	15.5	123	No	Yes
25	12	208.00	2.50	15.8	55	-	No
25	11	169.70	*	14.8	92	Yes	No
26	10	261.22	3.04	14.4	-	No	-
31	5	520.00	3.35	12.0	175	Yes	Yes
26	14	147.69	2.14	10.0	210	No	Yes
27	14	231.11	1.06	11.5	187	No	No
26	10	266.67	*	12.5	160	No	Yes
26	10	196.92	2.78	14.0	41	No	No
30	9	236.62	2.47	13.4	-	-	-
27	13	200.00	*	14.8	191	No	Yes
29	13	272.34	2.62	15.2	133	No	Yes
28	14	169.70	2.65	11.4	166	No	Yes
25	9	237.04	2.75	10.0	95	No	Yes
26	13	226.09	3.00	13.9	383	No	No
30	12	257.14	2.59	8.5	71	No	No
28	13	260.87	2.40	15.8	155	No	No
25	13	160.00	2.90	18.0	81	No	No
25	5	210.53	3.15	15.0	258	No	No
29	8	254.55	2.30	13.0	25	No	No
30	14	312.20	2.71	13.0	102	No	No
25	11	238.30	2.52	15.0	47	No	No
26	15	275.00	2.90	-	-	No	-
37	15	320.00	*	16.0	135	No	Yes

Abbreviations: ACA, anti-centromere antibody; mPAP, mean pulmonary arterial pressure; NTproBNP, N-terminal pro-brain natriuretic peptide; PCWP, pulmonary capillary wedge pressure; PVR, pulmonary vascular resistance; RA, right atrium; RAD, right axis deviation; TR, tricuspid regurgitant jet.

[1] As defined in Appendix 10. * TR not detectable.

Appendix 6: Construction of prediction nomograms

Prediction nomograms were developed for the algorithm in order to provide a tool that can be used in clinical practice (Figure 3 in manuscript).

a) Methods

Model functions of the two-step decision tree were graphically represented as two prediction nomograms. In conventional nomograms, the risk score is left-aligned starting from a patient with the smallest possible value for each covariable. In such a nomogram, highly specific covariables (primarily identifying non-PH) can contribute many points for clinically normal findings, compared with sensitive covariables that primarily identify PAH cases. Although this is methodologically correct, this feature of the conventional nomogram makes it difficult for physicians to interpret the clinical influence of each covariable on the risk of PAH. To solve this problem, we refined the nomogram for step 1 of the algorithm by centring each covariable at the mean value of non-PH patients (except for current/past telangiectasias where the majority of patients had ‘presence’ but we chose to centre around ‘absence’ as this is closer to a healthy patient).

b) Results

See Figure 3 in manuscript.

Appendix 7: Bootstrap validation of models

a) Methods

Internal validation[5,6] was used to assess over-optimism of the fitted models that can occur when a large set of potential candidate predictor variables are fitted with a relatively small

set of cases. Bootstrap re-sampling method[7] was used for internal validation of the model building procedure starting with 12 variables. Bootstrapping was conducted on 12 variables rather than 13 because right ventricle area was eliminated in the final model for reasons of feasibility in clinical practice. Therefore, right ventricle area was not included in the bootstrap validation process.

The bootstrap process was conducted by repeating the following steps 2000 times:

i. Generation of a training sample by random selection with replacement from the original data. A validation sample was generated with all the patients not included in the training sample, approximately one-third of the patients

ii. Variable selection and model fitting was performed as in Appendix 3 using the training sample, where sensitivity was fixed 97% at step 1 and specificity at 35% for step 2

iii. For validation purposes, diagnostic performance statistics (ROC AUC, sensitivity, specificity, PPV and NPV) were calculated for each step of the decision tree and overall using the validation sample.

Mean and 95% confidence intervals for ROC AUC, sensitivity, specificity, PPV and NPV of the 2000 validation samples were computed.

Given that the diagnostic performance statistics were calculated using patients (validation samples) independent from those used for model fitting (training samples), these values can be considered as surrogates for the diagnostic performance of the two-step decision tree applied to new patients outside of the DETECT study.[6]

The proportion of times a potential candidate variable was selected in the 2000 training decision tree models is informative about its robustness as a predictor in the decision tree

(more than 30% indicates a weak, more than 50% a moderately strong, and more than 70% a strong predictor[8,9]).

b) Results

The results of the bootstrap validation process are presented in Table S8 and Table S9. High ROC AUC estimates (~80%), were obtained for both steps of the decision tree algorithm using the validation samples. The overall sensitivity was 87% (95% CI: 71%, 100%) and the overall specificity 53% (95% CI: 35%, 72%; Table S8). All selected variables in the final two-step decision tree algorithm were selected more than 60% of the times in the bootstrap process, indicating high model robustness. The exception was RA area, which was selected 48% of the times (Table S9). The reason for this is that RA area replaced RV area in the final model as a result of a clinical decision by the Study Scientific Committee.

The mean ROC AUC in the bootstrap validation samples (Table S8) at step 1 and step 2 of the decision tree (79% and 83%, respectively) was slightly lower than that in the final models (84% and 88%, respectively). The sensitivity at step 1 was also lower in the validation samples versus the final model (94% vs. 97%). The specificity at step 2 was slightly lower in the validation sample than in the final model (33% vs. 35%). All these measures are slightly lower in the bootstrap validation samples than in the final models, as expected.

Table S8. Summary of diagnostic performance statistics obtained in the bootstrap process for the two-step decision tree using the validation samples

	ROC AUC, % Mean (95% CI)	Sensitivity, % Mean (95% CI)	Specificity, % Mean (95% CI)	PPV, % Mean (95% CI)	NPV, % Mean (95% CI)
Step 1 validation	79 (71, 87)	94 (79, 100)	27 (8, 54)	27 (18, 36)	95 (88, 100)
Step 2 validation	83 (73, 91)	94 (83, 100)	33 (19, 49)	35 (24, 47)	93 (81, 100)
Overall validation	Not available	87 (71, 100)	53 (35, 72)	35 (24, 47)	94 (88, 100)

Abbreviations: AUC, area under the curve; CI, confidence interval; NPV, negative predictive value; PPV, positive predictive value.

The design features of the decision tree imply that the overall sensitivity is smaller than or equal to the intermediate sensitivities of steps 1 and 2, that the overall specificity is larger than or equal to the intermediate specificities of steps 1 and 2, and that the overall PPV equals the intermediate PPV of step 2.

Table S9. Number of times that each variable was selected to be included in the final decision tree algorithm by means of bootstrap using 2000 training samples

	% of times that a variable was selected in the bootstrap process	
Step 1 variables	**Linear term**	**Quadratic term**
FVC % pred./DLCO % pred*	99	52
Serum urate*	97	75
Telangiectasias*	91	–
NTproBNP (log_{10})*	86	25
Right axis deviation*	81	–
Serum ACA*	62	–
Peripheral oedema	28	–
WHO functional class	26	–
Serum Scl-70	16	–
Step 2 variables		
TR velocity*	100	69
Right atrium area*	48	5
TAPSE	24	13

*Variable in the final model. Abbreviations: ACA, anti-centromere antibody; DLCO, pulmonary diffusing capacity for carbon monoxide; FVC, forced vital capacity; NTproBNP, N-terminal pro-brain natriuretic peptide; TR, tricuspid regurgitant jet; TAPSE, tricuspid annular plane systolic excursion; WHO, World Health Organization.

Appendix 8: Sensitivity analyses – handling of missing and sparse data

a) Methods

Analyses were performed on available data (see below TR velocity) and each model fit used all patients who had available data for each of the candidate variables. Because the reporting of some data was not mandatory, some variables (e.g., 6MWD) had a large amount of missing data and were excluded from the multivariable models, thus optimising the use of the maximum number of patients for the clinically important variables in model building.

Categorical variables which had sparse data in some categories (e.g., WHO functional class and SSc sub-type) were re-categorised into clinically meaningful categories to improve model estimation.

For patients in whom TR velocity was reported to be absent and no specific value was provided, the value was imputed as the mean of all available values ≤2.8 m/s within the PAH and within the non-PH groups. This method of imputation was applied in consensus with the Study Scientific Committee. The imputed TR velocity values were used in the main regression analyses: sensitivity analyses without imputation were performed to confirm the suitability of imputed values. The reason for using imputed values was to minimise the loss of patients available for constructing the detection algorithm since TR velocity is a strong predictor.

The effect of the imputation of TR velocity was evaluated comparing the performance of the decision tree algorithm with and without imputation.

b) Results

Since TR velocity is used in step 2 of the decision tree, the effect of using imputation or not was only observed in this step and in the overall performance. The global performance results with and without imputation (97% sensitivity in step 1 and 35% specificity at step 2) are shown in Table S10 below.

Table S10. Global discrimination performance with and without TR velocity imputation (97% sensitivity at step 1 and 35% specificity at step 2)

	Overall sensitivity (%)	Overall specificity (%)	Overall PPV (%)	Overall NPV (%)	Overall missed PAH diagnoses (false negatives, %)
With TR velocity imputation (n=356)	96	48	35	98	4
Without TR velocity imputation (n=336)	95	48	35	97	5

Abbreviations: NPV, negative predictive value; PAH, pulmonary arterial hypertension; PPV, positive predictive value; TR, tricuspid regurgitant velocity.

Appendix 9: Sensitivity analyses – influence of omitting a variable in the decision tree

a) Methods

The performance of the global decision tree algorithm was evaluated leaving out one variable at a time in order to assess the effect of a missing value in real practice. The omitted variable was imputed using the mean value of non-PH patients.

b) Results

The results of using the decision tree algorithm with one missing variable at a time are shown in Table S11. The effect of imputing a missing variable on the overall performance is related to the discriminatory ability of the variable, showing a worse performance when the missing variable has higher discriminatory ability (e.g., TR velocity). Overall the effect was small.

Table S11. Global discrimination performance when leaving out one variable at a time (97% sensitivity at step 1 and 35% specificity at step 2)

	Overall sensitivity (%)	Overall specificity (%)	Overall PPV (%)	Overall NPV (%)	RHC referral rate (%) (positive screening tests/ all patients)	Overall missed PAH diagnoses (false negatives, %)
No missing variables	96	48	35	98	62	4
Excluding one variable*						
Right axis deviation	96	50	36	98	61	4
Current/past telangiectasias	96	38	31	97	70	4
Serum ACA	93	52	36	96	58	7
Serum NTproBNP	94	47	34	97	62	6
FVC % pred./DLCO % pred.	93	56	38	96	55	7
Serum urate	94	39	31	96	69	6
Right atrium area	95	44	32	97	64	5
TR velocity	92	46	33	95	62	8

*The omitted variable was imputed using the mean value of non-PH patients. Abbreviations: ACA, anti-centromere antibody; DLCO, pulmonary diffusing capacity for carbon monoxide; FVC, forced vital capacity; NPV, negative predictive value; NTproBNP, N-terminal pro-brain natriuretic peptide; PAH, pulmonary arterial hypertension; PPV, positive predictive value; RHC, right heart catheterisation; TR, tricuspid regurgitant jet.

Appendix 10: Interpretation of ESC/ERS guidelines

For performance comparisons of the DETECT model and application of the ESC/ERS guidelines to the DETECT population, the ESC/ERS guidelines were interpreted as RHC being recommended if any of the following situations was present:

1. TR velocity >3.4 m/s
2. TR velocity >2.8–≤3.4 m/s **AND** symptomatic (defined as at least one of the following DETECT parameters: current anginal pain, current syncope/near syncope, current dyspnoea, presence of peripheral oedema)
3. TR velocity ≤2.8 m/s **AND** symptomatic (defined as above) **AND** presence of additional echocardiography variables suggestive of PH (right atrium area >16 cm^2 and/or ratio of right ventricular diameter/left ventricular end diastolic diameter >0.8).

Appendix 11: Statistical software

Statistical analyses were performed using SAS® (versions 9.2 and 9.3) for descriptive analyses and variables reduction process. The R-software (version 2.13.2) was used mainly for the decision tree modelling. Logistic regression models were fitted using 'Function lrm' from the R package 'Design'. Nomograms were produced using a modified version of 'Function nomogram' from the Design package, in order to obtain the 'centred' nomogram (Appendix 6).

REFERENCES

1. Badesch DB, Champion HC, Sanchez MA, et al. Diagnosis and assessment of pulmonary arterial hypertension. *J Am Coll Cardiol* 2009; **54:** S55–S66.

2. Simonneau G, Robbins IM, Beghetti M, et al. Updated clinical classification of pulmonary hypertension. *J Am Coll Cardiol* 2009; **54:** S43–S54.

3. Wood SN. Generalized Additive Model. An Introduction with R. Texts in Statistical Science. Boca Raton, Fl: Chapman & Hall/CRC; 2006. ISBN 1-58488-474-6.

4. Harrell FE Jr. Regression modeling strategies. Springer series in statistics. Springer, 2001.

5. Harrell FE Jr, Margolis PA, Gove S, et al. Development of a clinical prediction model for an ordinal outcome: the World Health Organization Multicentre Study of Clinical Signs and Etiological agents of Pneumonia, Sepsis and Meningitis in Young Infants. WHO/ARI Young Infant Multicentre Study Group. *Stat Med* 1998; **17:** 909–44.

6. König IR, Malley JD, Weimar C, Diener HC, Ziegler A; German Stroke Study Collaboration. Practical experiences on the necessity of external validation. *Stat Med* 2007; **26:** 5499–511.

7. Efron B and Tibshirani R. An introduction to the bootstrap. Monographs on statistics and applied probability. Boca Raton, Fl: Chapman & Hall/CRC, 1993.

8. Chen C-H, George SL. The bootstrap and identification of prognostic factors via Cox's proportional hazards regression model. *Stat Med* 1985; **4:** 39–46.

9. Sauerbrei W, Schumacher M. A bootstrap resampling procedure for model building: application to the Cox regression model. *Stat Med* 1992; **11:** 2093–109.