10.1148/radiol.2019181695

Appendix E1

Results of the Validation of the Classification Criteria

The prevalence of definite DISH, according to the classification criteria, at the last CT scan of cohort II was 9.5% (201/2121) for definite DISH and 18.5% (393/2121) for early DISH, as the result of the exclusion of participants with severe disc height loss and/or facet joint ankylosis. After assessment of the first CT scans using the nested case-control design, a mismatch was seen in both the size of the cohort (2121 versus 2047) and in the size of the early DISH (393 versus 361) and no DISH groups (1527 versus 1496), because all cases with loss of the intervertebral disc height and/or ankylosis of the facet joints at the first CT scan were also excluded and because of the use of a sampling fraction. If all cases were excluded based on the presence of intervertebral disc height loss and/or ankylosis of the facet joints at both the first and last time point, the prevalence of definite DISH at the last CT scan was 9.3% (190/2047) and the prevalence of early DISH was 17.6% (361/2047). The prevalence of definite DISH at the first CT scan was 4.9% (100/2047) after correction for the full study population with the sampling fraction of 0.454 (Table E3). The classification criteria had a sensitivity of 96% (88/92; 95%CI 89%–99%), specificity of 86% (1593/1855; 95%CI 84%–87%), PPV of 25% (88/350; 95%CI 21%–30%), NPV of 100% (1593/1597; 95%CI 99%–100%) and a risk ratio of 100 (95%CI 37.1–271.6). The intrarater reliability was 0.77 [95% CI 0.67–0.88] and the interrater reliability ranged from 0.59 [95% CI 0.47–0.72]–0.83 [95% CI 0.74–0.93] between each of the observers, corresponding to ‘substantial’ to ‘almost perfect’ agreement (Table E4). The Fleiss’ kappa between all seven observers was 0.67 [95% CI 0.66–0.68].

Table E1: Results of cohort II using the diagnostic criteria for the clinical setting without assessment of the height of the intervertebral disc and the presence of ankylosis of the facet joints

Diagnostic criteria First CT scan	Last CT scan	Definite DISH	Early DISH	No DISH
Definite DISH		128	1 (2.2)	0 (0)
Early DISH		99	129 (284.2)	25 (55.1)
			154 (339.3)	
No DISH		4	64 (141.0)	705 (1553.4)
			769 (1694.5)	
Total		231	194 (427.5)	730 (1608.5)
Total original data (no sample fraction)		231	447	1589

The number of participants is presented with the calculated number of expected participants using the sample fraction in parenthesis. The minor mismatch between the original data (as a result of the scoring of the presence of (early-)DISH in all participants at the last CT scan) and the calculated data (data based on 1155 (231+(4*231)) cases) is the consequence of the estimate as a result of using a nested case-control design. Abbreviations: CT = computed tomography, DISH = diffuse idiopathic skeletal hyperostosis.

Table E2: The results of the interrater reliability of the criteria for (early-)DISH between each of the observers

	Diagnostic criteria for DISH					
Observer	1	2	3	4	5	6
Observer						
2	**0.89 [0.81–0.96]**					
3	0.87 [0.80–0.95]	0.84 [0.76–0.92]				
4	0.82 [0.73–0.91]	0.84 [0.74–0.93]	0.78 [0.67–0.88]			
5	0.80 [0.70–0.89]	0.81 [0.72–0.91]	0.80 [0.70–0.90]	0.83 [0.74–0.91]		
6	0.79 [0.70–0.89]	0.84 [0.75–0.92]	**0.75 [0.65–0.86]**	0.85 [0.76–0.94]	0.80 [0.70–0.90]	
7	0.85 [0.76–0.93]	0.86 [0.78–0.94]	0.83 [0.73–0.92]	0.80 [0.70–0.90]	0.85 [0.76–0.94]	0.85 [0.77–0.93]

All the observers used the criteria for DISH on all 90 CT scans. The results are linear kappas with the 95% confidence interval in brackets. The observers had variable levels of expertise with radiographic scoring and DISH: observer 1 (JW, nonradiologist, research student, 0 years of experience), observer 2 (WPG, nonradiologist, PhD candidate, 1 year of experience), observer 3 (JSK, nonradiologist, PhD candidate, 3 years of experience), observer 4 (WF, radiologist in training, 2 years of experience), observers 5 (SFO, radiologist, 6 years of experience), observer 6 (FAAMH, radiologist, 6 years of experience) and observer 7 (JJV, nonradiologist, orthopedic surgeon, 12 years of experience). Abbreviations: DISH = diffuse idiopathic skeletal hyperostosis, CT = computed tomography.

Bold numbers represent the range of the results.

Table E3: Results of cohort II using the classification criteria for the research setting with exclusion of the cases with loss of the intervertebral disc height and/or ankylosis of the facet joints

Classification criteria First CT scan	Last CT scan	Definite DISH	Early DISH	No DISH
Definite DISH		98	1 (2.2)	0
Early DISH		88	103 (227.0)	16 (35.3)
			119 (262.2)	
No DISH		4	60 (132.2)	663 (1460.9)
			723 (1593.1)	
Total		190	164 (361.2)	679 (1496.1)
Exclusion based on		41	30 (66.1)	51 (112.4)
-Loss of height IVD				
-Facet ankylosis				
Total with excluded participants		231	194 (427.3)	730 (1608.5)

The number of participants is presented with the calculated number of expected participants using the sample fraction in parenthesis. The prevalence of (early-)DISH according to the classification criteria at the last CT-scan was based on the data of all the participants, without correction using the sampling fraction. A mismatch was seen in both the size of the cohort (2121 versus 2047) and in the size of the early DISH (393 versus 361) and no DISH groups (1527 versus 1496), because all cases with loss of the intervertebral disc height and/or ankylosis of the facet joints at the first CT scan were also excluded and because of the use of a sampling fraction. Abbreviations: CT = computed tomography, DISH = diffuse idiopathic skeletal hyperostosis, IVD = intervertebral disc.

Table E4: The results of the interrater reliability of the classification criteria for DISH

	Classification criteria for DISH					
Observer	1	2	3	4	5	6
Observer						
2	0.72 [0.60–0.84]					
3	0.63 [0.50–0.77]	**0.83 [0.74–0.93]**				
4	0.68 [0.56–0.80]	0.76 [0.64–0.87]	0.80 [0.70–0.90]			
5	0.75 [0.64–0.86]	0.66 [0.54–0.78]	**0.59 [0.47–0.72]**	0.62 [0.49–0.74]		
6	0.63 [0.50–0.76]	0.77 [0.66–0.87]	0.75 [0.65–0.85]	0.75 [0.64–0.87]	0.60 [0.47–0.72]	
7	0.69 [0.56–0.81]	0.75 [0.65–0.86]	0.68 [0.56–0.81]	0.71 [0.59–0.82]	0.70 [0.58–0.82]	0.65 [0.52–0.77]

The results are linear kappas with the 95% confidence interval in brackets. The observers used the criteria for DISH excluding cases with poor intervertebral disc height and/or apophyseal ankylosis on 90 CT scans. The observers had variable levels of expertise with radiographic scoring and DISH: observer 1 (JW, nonradiologist, research student, 0 years of experience), observer 2 (WPG, nonradiologist, PhD candidate, 1 year of experience), observer 3 (JSK, nonradiologist, PhD candidate, 3 years of experience), observer 4 (WF, radiologist in training, 2 years of experience), observers 5 (SFO, radiologist, 6 years of experience), observer 6 (FAAMH, radiologist, 6 years of experience) and observer 7 (JJV, nonradiologist, orthopedic surgeon, 12 years of experience). Abbreviations: DISH = diffuse idiopathic skeletal hyperostosis, CT = computed tomography.

Bold numbers represent the range of the results.