eAppendix 1: Advanced analyses

Methods

Clusterability of the EU-NN database

Intrinsic dataset clusterability can be measured by comparing the coefficient of determination of the actual dataset with randomly generated, uniformly distributed datasets. If a dataset is composed of uniformly distributed subjects, clustering can still be performed, but will result in a low coefficient of determination as clusters are not distinct. To test the intrinsic clusterability of the EU-NN database, we repeated the clustering with 20 randomly generated datasets and compared the coefficients of determination. These random datasets had the same layout as the EU-NN database with 1078 inclusions and the same number of variables and outcome options and missing values on the variables. We replaced all original values with random numbers from uniform distributions over the outcome options, respecting the categorical or continuous nature of variables.

Cluster distinctness

Silhouette coefficients were calculated for individuals and displayed per cluster to help identify which clusters were most distinctly grouped. The silhouette coefficient represents the ratio of the mean distance between single individuals with others in the same cluster and the individuals in the nearest neighboring cluster. If a particular cluster has large, consistently positive silhouette values, this means the cluster is distinctly grouped.

Cluster reproducibility

Jackknife resampling was implemented as a cross-validation method to test whether the EU-NN database had sufficient entries to ensure that sample size variations are unlikely to change the clustering results. We repeated the clustering algorithm 100 times with random 80% selections of individuals in the EU-NN database patients. We calculated the chance of grouping two individuals together in resampling iterations that were in different clusters in the original clustering. This highlights the clusters between which there is more frequent mixing. A higher mixing probability between two clusters, however, does not necessarily mean that these clusters are not robust, as it can also mean that these clusters are frequently merged into one larger cluster in the resampling iterations for the chosen number of clusters, and are still separate with a larger number of clusters. For more information, see eAppendix 2.

Results

Clusterability of the EU-NN database

Our clustering model explained 40% of the variance in the EU-NN database at seven clusters (Fig. 1A). This is substantially higher than the corresponding coefficients of determination of the randomly generated datasets (mean of 16% and upper limit of the 95% confidence interval at 18%), suggesting intrinsic clusterability of the EU-NN database.

Cluster distinctness

The almost unanimously positive silhouette coefficients of clusters 5 and 6 indicate cluster distinctness compared to the other clusters and to each other (Fig. 1B). The lower silhouette coefficients in other clusters indicate that these clusters are more similar to each other.

Clustering reproducibility

Visual inspection of the resampling repetitions showed similar distinguishing factors for individuals without cataplexy as the original clustering (i.e., sleep drunkenness and weekend-week sleep length difference) or merger into one larger cluster in 77% of repetitions.

In Fig. 1C, it can be seen that there is relatively frequent mixing among individuals in clusters 1-4 and 7. This means that different subdivisions or merging of clusters of people with cataplexy was regularly seen. Clusters 5 and 6 were relatively robustly subdivided from other clusters, further highlighting the ability of the clustering algorithm to distinguish people with cataplexy from those without. Visual inspection of the resampling iterations showed that roughly half of the 42% mixing between clusters 5 and 6 was because these two clusters were already merged together (similar to the six clusters result when clustering the entire database). The remaining mixing between people in clusters 5 and 6 was caused by subdivision based on other differentiating variables, mainly the presence of hypnagogic hallucinations and sleep paralysis (16%).

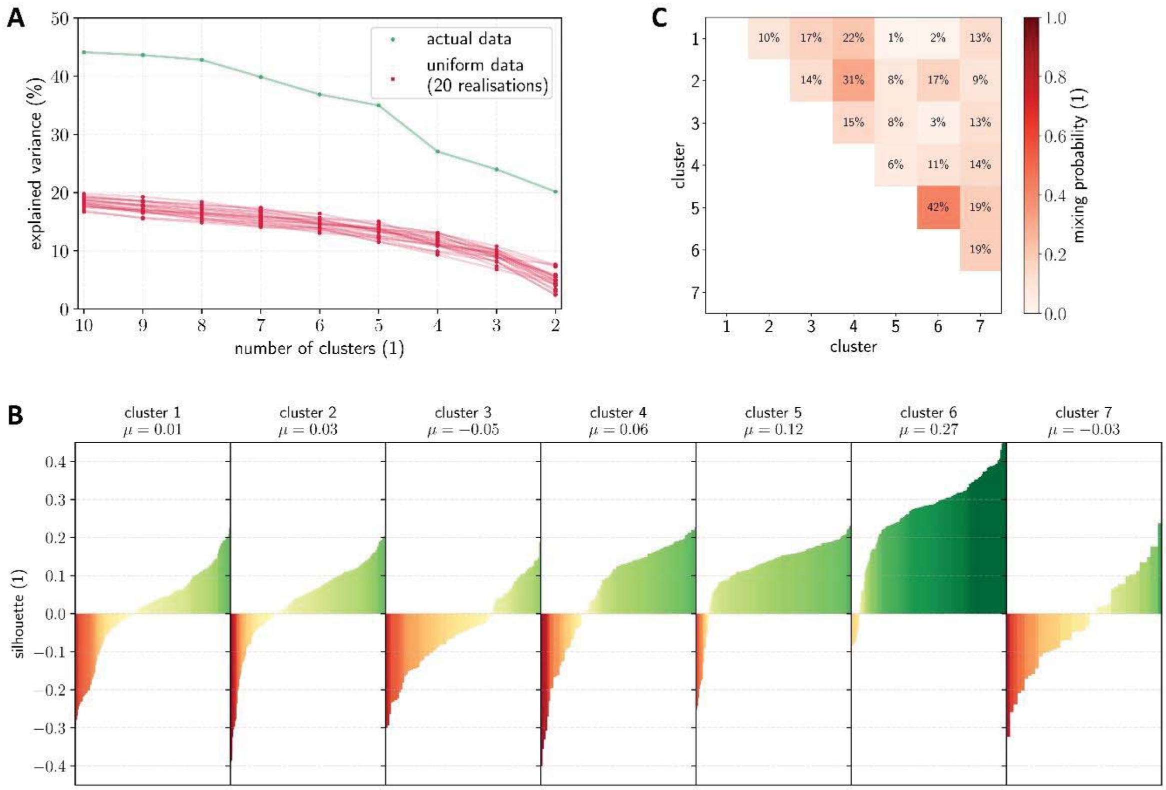


Figure 1. Advanced analyses. (A) Coefficient of determination (R^2) for different numbers of clusters, calculated for the original clustering in green and 20 randomly generated datasets in red. The greater coefficient of determination of the EU-NN database indicates intrinsic clusterability of the database. (**B)** The silhouette coefficient was calculated for each individual and grouped per cluster. Positive values (green) represent individuals distinctly grouped within their cluster, whereas negative values (red) indicate these individuals are closer to the nearest neighboring cluster than to others in their own cluster. Larger silhouette coefficients of clusters 5 and 6 reveal greater cluster distinctness. (**C**) Highlights between which clusters there is more frequent mixing in the jackknife resampling. A higher mixing probability between two clusters however does not necessarily mean that these clusters are not robust, as it can also mean that these clusters are frequently merged into one larger cluster in the resampling iterations for the chosen number of clusters cluster.

μ: mean silhouette coefficient.

eAppendix 2: Included variables & ICSD-3 population characteristics

Variable name	Definition	NT1 (n=752)		NT2 (n=200)		IH (n=126)		Normalized outcome options
		n	Median (IQR) / Percentage	n	Median (IQR) / Percentage	n	Median (IQR) / Percentage	
Presence – Disturbed NS	Presence of disturbed nocturnal sleep	752	63.2%	200	36.0%	126	16.7%	0 = No, 1 = Yes
PSG – Total sleep length	Total sleep length during polysomnography	577	408.0 (358.0-450.0)	163	433.0 (380.0-470.0)	85	443.0 (407.3-483.0)	0-1 = 172-562 minutes
PSG – Sleep efficiency	Sleep efficiency during nocturnal sleep of the polysomnography	596	86.0 (79.0-92.0)	168	93.0 (87.0-96.0)	99	92.2 (89.6-94.7)	0-1 = 52.3-98.4 %
PSG – Sleep latency	Sleep latency during nocturnal sleep of the polysomnography	530	4.5 (2.0-8.8)	150	5.7 (3.0-10.0)	86	7.8 (3.2-13.0)	0-1 = 0-46 minutes
MSLT – Mean sleep latency	Mean sleep latency during multiple sleep latency testing (4 or 5 naps in total)	507	3.0 (1.6-4.8)	156	4.6 (2.3-6.6)	93	7.0 (4.7-9.0)	0-1 = 0.4-15 minutes
MSLT – Mean NREM1 latency	Mean NREM1 latency during multiple sleep latency testing (4 or 5 naps in total)	512	3.0 (1.6-4.8)	149	4.4 (2.3-6.9)	74	6.6 (4.0-8.1)	0-1 = 0.3-15 minutes
MSLT – Mean NREM2 latency	Mean NREM2 latency during multiple sleep latency testing (4 or 5 naps in total)	142	8.3 (4.5-11.7)	42	8.5 (4.7-10.5)	49	9.4 (7.6-11.8)	0-1 = 1.5-17 minutes
Subjective sleep latency	Subjective sleep latency	751	5.0 (2.0-7.5)	200	5.0 (3.0-10.0)	124	5.0 (3.0-10.0)	0-1 = 0-60 minutes
Subject NS length	Subjective nocturnal sleep length change	142	1.0 (1.0-4.0)	31	1.8 (0.5-5.2)	24	1.6 (0.5-4.6)	0-1 = 0.5-10 hours
Subjective daily sleep length	Subjective sleep length per 24 hours	754	7.3 (6.4-8.0)	200	7.8 (7.1-8.5)	126	7.8 (7.1-8.8)	0-1 = 5.9-10.6 hours
Change sleep length since start EDS	Subjective change in sleep length since start of excessive daytime sleepiness	361	Shorter: 26.6% Unchanged: 58.7% Longer: 14.7%	87	Shorter: 4.6% Unchanged: 70.1% Longer: 25.3%	53	Shorter: 7.5% Unchanged: 49.1% Longer: 43.4%	0 = Shorter, 0.5 = Unchanged, 1 = Longer
Time in bed during NS	Subjective time in bed during nocturnal sleep	378	8.0 (7.0-9.0)	102	8.0 (7.0-9.0)	94	8.0 (7.0-9.0)	0-1 = 4-13 hours
PSG – REM latency	REM latency during nocturnal sleep of the polysomnography	572	37.3 (4.0-87.9)	160	59.0 (9.4-82.0)	81	73.0 (61.0-93.0)	0-1 = 0-322 minutes
MSLT – Number of SOREMPs	Number of SOREMPs during multiple sleep latency testing (4 or 5 naps in total)	542	4.0 (3.0-4.0)	162	2.0 (2.0-4.0)	80	0.0 (0.0-0.0)	0-1 = 0-5 SOREMPs
MSLT – Mean REM latency	Mean REM latency during multiple sleep latency testing (4 or 5 naps in total)	338	4.8 (3.0-7.5)	100	8.0 (4.6-12.4)	13	15.0 (10.1-16.4)	0-1 = 0-5 minutes
Presence – Sleep attacks	Presence of sleep attacks	740	84.2%	195	69.2%	125	60.8%	0 = No, 1 = Yes
Subjective daytime sleep length	Subjective daytime sleep length	693	1.0 (1.0-2.0)	170	1.0 (0.5-2.0)	115	1.0 (0.0-2.0)	0-1 = 0-9 hours
Number of naps per day	Subjective number of naps per day	614	3.0 (2.0-4.0)	133	3.0 (2.0-4.0)	70	2.5 (1.0-3.0)	0-1 = 0-10 naps

Frequency of scheduled naps per day	Subjective number of scheduled naps per day	740	Never: 30.4% 1/day: 52.8% >1/day: 16.8%	195	Never: 54.4% 1/day: 36.9% >1/day: 8.7%	125	Never: 57.6% 1/day: 36.8% >1/day:5.6 %	0 = Never, 0.5 = 1/day, 1 = > 1/day
Length of scheduled naps	Subjective length of scheduled naps	514	30.0 (20.0-60.0)	89	30.0 (20.0-60.0)	52	60.0 (30.0-97.8)	0-1 = 10-240 minutes
Feeling refreshed after sleep	Subjective presence of feeling refreshed after sleep	617	No: 12.5% Not always: 24.6% Yes: 62.9%	131	No: 21.4% Not always: 35.1% Yes: 43.5%	75	No: 49.3% Not always: 25.3% Yes: 25.3%	0 = No, 0.5 = Not always, 1 = Yes
Presence – Sleep drunkenness	Presence of sleep drunkenness	752	25.3%	200	42.5%	126	62.7%	0 = No, 1 = Yes
Difficulty waking daytime sleep	Subjective difficulty in waking after daytime sleep	497	On average easy to wake up: 74.4% On average difficult to wake up: 22.7% On average nearly impossible to wake up: 2.8%	115	On average easy to wake up: 53.0% On average difficult to wake up: 41.7% On average nearly impossible to wake up: 5.2%	76	On average easy to wake up: 46.1% On average difficult to wake up: 46.1% On average nearly impossible to wake up: 7.9%	0 = On average easy to wake up, 0.5 = On average difficult to wake up, 1 = On average nearly impossible to wake up
Presence – Cataplexy	Presence and certainty of cataplexy	751	No: 5.5% Possible cataplexy: 4.8% Probable cataplexy: 6.8% Definite cataplexy: 83.0%	200	No: 97.5% Possible cataplexy: 2.0% Probable cataplexy: 0.5% Definite cataplexy: 0.0%	126	No: 96.8% Possible cataplexy: 3.2% Probable cataplexy: 0.0% Definite cataplexy: 0.0%	0 = No, 0.33 = Possible cataplexy, 0.67 = Probable cataplexy, 1 = Definite cataplexy
Cataplexy frequency	Frequency of cataplexy attacks	752	Never: 5.5% < 1/year: 4.3% 1/year – 1/month: 14.6% 1/month – 1/week: 24.6% 1/week – 1/day: 25.0% > 1/day: 26.1%	200	Never: 97.5% < 1/year: 1.0% 1/year – 1/month: 1.0% 1/month – 1/week: 0.5% 1/week – 1/day: 0.0% > 1/day: 0.0%	126	Never: 96.8% < 1/year: 0.0% 1/year – 1/month: 0.0% 1/month – 1/week: 0.8% 1/week – 1/day: 0.8% > 1/day: 1.6%	0 = Never, 0.2 = < 1/year, 0.4 = 1/year – 1/month, 0.6 = 1/month – 1/week, 0.8 = 1/week – 1/day, 1 = > 1/day
Length of single attack	Mean length of a cataplexy attack	752	Never: 5.5% < 10 seconds: 39.1% 10 seconds – 2 minutes: 48.9% 2 minutes – 10 minutes: 5.6% > 10 minutes: 0.9%	200	Never: 97.5% < 10 seconds: 2.0% 10 seconds – 2 minutes: 0.0% 2 minutes – 10 minutes: 0.0% > 10 minutes: 0.5%	126	Never: 96.8% < 10 seconds: 1.6% 10 seconds – 2 minutes: 0.8% 2 minutes – 10 minutes: 0.0% > 10 minutes: 0.8%	0 = Never, 0.25 = < 10 seconds, 0.5 = 10 seconds – 2 minutes, 0.75 = 2 minutes – 10 minutes, 1 = > 10 minutes
Presence – Slurred speech	Presence of slurred speech during a cataplexy attack	752	48.8%	200	2.5%	126	0.8%	0 = No, 1 = Yes
Presence – Muscle twitches	Presence of muscle twitches during a cataplexy attack	752	40.7%	200	2.0%	126	0.0%	0 = No, 1 = Yes
Weakness – Whole body	Presence of weakness in the whole body during a cataplexy attack	752	52.0%	200	2.5%	126	1.6%	0 = No, 1 = Yes

Weakness – Face/jaw	Presence of weakness in the face or jaw during a cataplexy attack	752	68.1%	200	2.5%	126	2.4%	0 = No, 1 = Yes
Weakness – Neck/drop head	Presence of weakness in the neck or dropping the head during a cataplexy attack	752	57.6%	200	2.0%	126	2.4%	0 = No, 1 = Yes
Weakness – Arms/hands	Presence of weakness in the arms or hands during a cataplexy attack	752	48.5%	200	2.5%	126	2.4%	0 = No, 1 = Yes
Weakness – Knees	Presence of weakness in the knees during a cataplexy attack	752	75.4%	200	2.5%	126	3.2%	0 = No, 1 = Yes
Trigger – Emotions	Are emotions a trigger for cataplexy attacks?	690	Never: 7.5% Most of the time: 25.2% Always: 67.2%	199	Never: 97.5% Most of the time: 1.0% Always: 1.5%	126	Never: 96.8% Most of the time: 0.8% Always: 2.4%	0 = Never, 0.5 = Most of the time, 1 = Always
Trigger – Laughing	Is laughing a trigger for cataplexy attacks?	679	87.6%	199	2.5%	126	2.4%	0 = No, 1 = Yes
Trigger – Telling a joke	Is telling a joke a trigger for cataplexy attacks?	679	72.9%	199	2.0%	126	0.8%	0 = No, 1 = Yes
Trigger – Being startled	Is being startled a trigger for cataplexy attacks?	679	40.4%	199	0.5%	126	2.4%	0 = No, 1 = Yes
Trigger – Meeting acquaintance	Is meeting an acquaintance a trigger for cataplexy attacks?	679	37.1%	199	0.5%	126	2.4%	0 = No, 1 = Yes
Trigger – Being angry	Is being angry a trigger for cataplexy attacks?	679	47.7%	199	1.5%	126	3.2%	0 = No, 1 = Yes
Length muscle tone loss	Length of muscle tone loss during a cataplexy attack	711	< 10 seconds: 62.7% 10 seconds – 2 minutes: 34.6% > 2 minutes: 2.7%	5	< 10 seconds: 80.0% 10 seconds – 2 minutes: 20.0% > 2 minutes: 0.0%	4	< 10 seconds: 50.0% 10 seconds – 2 minutes: 25.0% > 2 minutes: 25.0%	0 = < 10 seconds, 0.5 = 10 seconds – 2 minutes, 1 = > 2 minutes
Presence – Hypnagogic hallucinations	Presence of hypnagogic hallucinations	752	No: 46.0% Possible hypnagogic hallucinations: 2.1% Probable hypnagogic hallucinations: 6.4% Definite hypnagogic hallucinations: 45.5%	200	No: 70.0% Possible hypnagogic hallucinations: 1.0% Probable hypnagogic hallucinations: 6.0% Definite hypnagogic hallucinations: 23.0%	126	No: 71.4% Possible hypnagogic hallucinations: 1.6% Probable hypnagogic hallucinations: 5.6% Definite hypnagogic hallucinations: 21.4%	0 = No, 0.33 = Possible hypnagogic hallucinations, 0.67 = Probable hypnagogic hallucinations, 1 = Definite hypnagogic hallucinations
Hypnagogic hallucinations frequency	Frequency of hypnagogic hallucinations	752	Never: 73.4% < 1/year: 2.8% 1/year – 1/month: 4.8%	200	Never: 87.0% < 1/year: 1.0% 1/year – 1/month: 3.5%	125	Never: 88.8% < 1/year: 0.0% 1/year – 1/month: 3.2%	0 = Never, 0.2 = < 1/year, 0.4 = 1/year – 1/month, 0.6 = 1/month – 1/week, 0.8 = 1/week – 1/day, 1 = > 1/day

			1/month – 1/week: 4.1% 1/week – 1/day: 10.0% > 1/day: 4.9%		1/month – 1/week: 3.0% 1/week – 1/day: 5.5% > 1/day: 0.0%		1/month – 1/week: 4.0% 1/week – 1/day: 3.2% > 1/day: 0.8%	
Appears waking up	Do the hallucinations occur when waking?	542	16.6%	166	9.0%	105	7.6%	0 = No, 1 = Yes
Appears falling asleep	Do the hallucinations occur when falling asleep?	542	30.6%	166	10.8%	105	9.5%	0 = No, 1 = Yes
Appears with sleep paralysis	Do the hallucinations co-occur with sleep paralysis?	406	25.1%	60	15.0%	36	11.1%	0 = No, 1 = Yes
Presence – Sleep paralysis	Presence of sleep paralysis	752	No: 52.3% Possible sleep paralysis: 1.2% Probable sleep paralysis: 5.5% Definite sleep paralysis: 41.1%	200	No: 77.0% Possible sleep paralysis: 1.5% Probable sleep paralysis: 3.5% Definite sleep paralysis: 18.0%	126	No: 76.2% Possible sleep paralysis: 1.6% Probable sleep paralysis: 7.1% Definite sleep paralysis: 15.1%	0 = No, 0.33 = Possible sleep paralysis, 0.67 = Probable sleep paralysis, 1 = Definite sleep paralysis
Sleep paralysis frequency	Frequency of sleep paralysis	752	Never: 76.7% < 1/year: 5.3% 1/year – 1/month: 4.4% 1/month – 1/week: 5.2% 1/week – 1/day: 5.9% > 1/day: 2.5%	200	Never: 88.0% < 1/year: 0.5% 1/year – 1/month: 5.5% 1/month – 1/week: 2.0% 1/week – 1/day: 3.5% > 1/day: 0.5%	125	Never: 89.6% < 1/year: 0.0% 1/year – 1/month: 4.8% 1/month – 1/week: 3.2% 1/week – 1/day: 1.6% > 1/day: 0.8%	0 = < 1/year, 0.25 = 1/year – 1/month, 0.5 = 1/month – 1/week, 0.75 = 1/week – 1/day, 1 = > 1/day
Appears waking up	Does the sleep paralysis occur when waking?	534	31.1%	169	10.7%	104	9.6%	0 = No, 1 = Yes
Appears falling asleep	Does the sleep paralysis occur when falling asleep?	534	31.1%	169	10.7%	104	9.6%	0 = No, 1 = Yes
Epworth sleepiness scale	Epworth sleepiness scale score	752	17.0 (15.0-20.0)	200	16.0 (13.0-18.0)	126	16.0 (13.0-18.3)	0-1 = 4-24
Presence – EIDS warning signs	Presence of warning signs before episodes of irresistible daytime sleep	294	69.4%	61	68.9%	46	91.3%	0 = No, never, 0.5 = Sometimes, 1 = Yes, always
EIDS length per day	Subjective length of episodes of irresistible daytime sleep per day	623	>60 minutes: 3.5% 15-60 minutes: 29.4% 1-15 minutes: 60.7% <1 minute: 6.4%	134	>60 minutes: 9.7% 15-60 minutes: 29.9% 1-15 minutes: 53.0% <1 minute: 7.5%	77	>60 minutes: 13.0% 15-60 minutes: 33.8% 1-15 minutes: 40.3% <1 minute: 13.0%	0 = >60 minutes, 0.33 = 15-60 minutes, 0.67 = 1-15 minutes, 1 = <1 minute
PVT mean reaction time	Mean reaction time during the psychomotor vigilance task. If multiple tests were performed, the outcomes were averaged.	42	76.5 (58.25-266.0)	20	250.5 (234.8-268.8)	0	-	0-1 = 46-285 milliseconds
Presence – Automatic behavior	Presence of automatic behavior	252	72.6%	51	43.1%	39	28.2%	0 = No, 1 = Yes

Weekend-week sleep length diff.	Subjective difference in sleep length during weekend and week days	750	0.0 (0.0-1.0)	199	1.0 (0.0-2.0)	125	1.0 (0.0-2.0)	0-1 = -2-6 hours
Presence – Fatigue	Presence of fatigue complaints	245	59.2%	51	64.7%	37	59.5%	0 = No, 1 = Yes
Fatigue severity scale	Fatigue severity scale score	58	5.0 (4.0-6.0)	27	5.0 (4.0-6.0)	7	5.0 (3.0-6.0)	0-1 = 3-6
Sex	Sex	753	Male: 56.5% Female: 43.5%	200	Male: 53.0% Female: 47.0%	126	Male: 46.0% Female: 54.0%	0 = Male, 1 = Female
Sleep disorders 1st degree relatives	Presence of sleep disorders in 1st degree relatives	557	No: 84.6% Other sleep disorder than hypersomnia disorder: 7.4% Hypersomnia disorder other than narcolepsy: 3.4% Narcolepsy: 4.7%	142	No: 82.4% Other sleep disorder than hypersomnia disorder: 8.5% Hypersomnia disorder other than narcolepsy: 4.9% Narcolepsy: 4.2%	92	No: 66.3% Other sleep disorder than hypersomnia disorder: 8.7% Hypersomnia disorder other than narcolepsy: 16.3% Narcolepsy: 8.7%	0 = No, 0.33 = Other sleep disorder than hypersomnia disorder, 0.67 = Hypersomnia disorder other than narcolepsy, 1 = Narcolepsy
Daily caffeine beverages	Daily intake of caffeine beverages	229	2.0 (1.0-4.0)	51	2.0 (0.0-4.0)	39	2.0 (0.0-3.0)	0-1 = 0-5 beverages
BMI	Body mass index	751	27.0 (23.5-30.8)	200	24.6 (21.9-27.5)	125	24.4 (20.1-26.4)	0-1 = 17.7-43.3
Weight gain since start EDS	Subjective increase in weight since onset of excessive daytime sleepiness	243	46.9%	55	21.8%	41	17.1%	0 = No, 1 = Yes
Hypocretin level	Hypocretin-1 level as measured in cerebrospinal fluid	236	12.0 (0.0-55.0)	29	228.0 (143.5-337.0)	46	302.5 (253.0-368.0)	0-1 = 0-382 pg / mL
HLA – DQB1*0602	HLA – DQB1*0602 positivity	552	93.8%	100	53.0%	58	43.1%	0 = No, 1 = Yes
HLA – DQA1*0102	HLA – DQA1*0102 positivity	185	74.6%	34	50.0%	2	0.0%	0 = No, 1 = Yes
Age onset EDS Age onset cataplexy Age onset disturbed NS	Age at onset of excessive daytime sleepiness, cataplexy or disturbed nocturnal sleep	736 709 469	19.8 (14.5-29.5) 23.3 (16.3-33.0) 23.8 (16.1-34.8)	197 5 68	19.3 (14.8-28.3) 23.8 (9.4-37.4) 24.0 (17.0-34.7)	123 4 24	18.3 (14.8-26.5) 20.4 (9.4-30.4) 26.7 (15.1-37.8)	Age EDS: 0-1 = 3.6-56.1 years Age cataplexy: 0-1 = 5.9-60.7 years Age disturbed NS: 0-1 = 6.6-60.9 years
Diff. onset EDS-Cat Diff. onset Disturbed NS-Cat Diff. onset Disturbed NS-EDS	Difference in age at onset between excessive daytime sleepiness, cataplexy and/or disturbed nocturnal sleep	701 455 465	0.2 (0.0-3.0) 0.0 (-0.5-0.5) 0.0 (-3.8-0.0)	5 4 67	0.5 (-1.2-14.6) -7.0 (-20.0-9.8) 0.0 (-0.6-0.0)	4 4 22	0.0 (-0.4-3.2) 0.0 (0.0-2.9) 0.0 (-7.3-0.0)	EDS-Cat: 0-1 = -4.7-28 years Disturbed NS-EDS: 0-1 = -32-12 years Disturbed NS-Cat: 0-1 = -22.8-20 years
Onset cataplexy winter/spring/summer/autumn Onset EDS winter/spring/summer/autumn	Season of onset of excessive daytime sleepiness, cataplexy or disturbed nocturnal sleep. These variables initially were single variables listing the onset month of each symptom. These initial variables and weightings were split into four seasonal variables.	344 344 344 344 371 371 371 371 196 196	44.8% 17.2% 24.1% 14.0% 46.9% 19.9% 19.4% 13.7% 44.4% 18.9%	4 4 4 4 113 113 113 113 33 33	75.0% 0.0% 25.0% 0.0% 54.9% 13.3% 15.9% 15.9% 51.5% 15.2%	3 3 3 3 42 42 42 42 8 8	0.0% 100.0% 0.0% 0.0% 50.0% 23.8% 7.1% 19.0% 75.0% 0.0%	0 = No, 1 = Yes

Onset disturbed NS winter/spring/summer/autumn		196 196	21.9% 14.8%	33 33	12.1% 21.2%	8 8	25.0% 0.0%	
Presence – Violence during sleep	Presence of violent behavior during sleep	524	29.6%	147	11.6%	101	9.9%	0 = No, 1 = Yes
PSG – AHI	Apnea-hypopnea index as measured during the polysomnography	552	2.0 (0.3-6.6)	157	2.0 (0.3-5.0)	99	1.3 (0.4-3.5)	0-1 = 0-53 / hour
PSG – PLMI	Periodic leg movement index as measured during the polysomnography	489	5.0 (0.0-16.2)	132	1.1 (0.0-5.0)	97	1.5 (0.0-8.2)	0-1 = 0-93 / hour
Presence – Breathing stops	Presence of breathing stops during sleep	516	24.0%	144	20.1%	103	8.7%	0 = No, 1 = Yes
Presence – Nightmares	Presence of nightmares during sleep	484	53.5%	106	31.1%	78	15.4%	0 = No, 1 = Yes
Presence – Sleep walking	Presence of sleep walking during sleep	752	8.1%	200	7.0%	126	4.0%	0 = No, 1 = Yes
Presence – Loud snoring	Presence of loud snoring during sleep	752	28.6%	200	16.0%	126	21.4%	0 = No, 1 = Yes
Presence – Nocturnal eating	Presence of nocturnal eating during sleep	254	34.3%	56	17.9%	39	7.7%	0 = No, 1 = Yes
Presence – Memory complaints	Presence of memory complaints	253	62.8%	56	53.6%	40	60.0%	0 = No, 1 = Yes
Presence – Mood complaints	Presence of mood complaints	257	48.6%	55	36.4%	40	47.5%	0 = No, 1 = Yes
Presence – Uncontrollable eating	Presence of uncontrollable eating	243	33.3%	52	15.4%	41	26.8%	0 = No, 1 = Yes
Presence – Smoking	Presence of smoking	256	39.8%	63	25.4%	36	8.3%	0 = No, 1 = Yes
Presence – Other addictive behavior	Presence of other addictive behavior	196	9.2%	47	6.4%	36	0.0%	0 = No, 1 = Yes
Presence – Libido/erection problems	Presence of libido and/or erection problems	145	27.6%	30	6.7%	25	8.0%	0 = No, 1 = Yes

eAppendix 3: Supplementary methods

The supplementary methods include the database preprocessing steps, explanations on the distance metric that was used for the hierarchical clustering algorithm, the combination of clustering evaluation metrics to determine the number of clusters, and the mathematical definitions of the clustering outcome measures.

Database preprocessing

In total, 97 variables were input into the hierarchical clustering algorithm after removing the follow-up and near zero-variance variables (i.e., the values of the variables were almost the same in all people) and the variables with more than 95% missing values. The number of included variables was substantially higher for some aspects of hypersomnolence (e.g., cataplexy) and, as we wanted to give each aspect a potentially similar influence on the clustering, we used expert opinion to group variables in 15 overarching categories as presented in Fig. 1 of the main manuscript. We equally divided the clustering weighting over the 15 categories, before evenly splitting it over the variables within that category. For example, the category cataplexy had 1/15th of the total weight, so that each of the 17 cataplexy-related variables had 1/17th of 1/15th of the total weight.

All variables were normalized to the range [0, 1] to make them comparable. For categorical variables this was done by evenly distributing the outcome options over the normalized range. To ensure sufficient spread in continuous variables and meanwhile to reduce the influence of outliers, the outermost 10 values of continuous variables (about 2% of raw continuous data) were changed to the corresponding 11th highest or lowest value, before normalization. This data pre-processing step refined the dataset while it preserved the distributions of the subjects.

Distance metrics

Different distance metrics can be used to determine the distances between individuals and different linkage types to determine which clusters will be merged next by the clustering algorithm. The Gower's distance was most suitable for the EU-NN database as it is the simplest distance metric where input may contain mixed categorical, continuous or missing data.[1] In case of a missing value, this variable was not included in the distance calculations for this individual. Different linkage methods (single, average and complete) were tested.[2] Single and average linkage consistently assigned individual outliers to separate clusters. Complete linkage was therefore chosen, because it tended to group individuals with outlying values most inclusively.

Clustering evaluation metrics

Standard clustering evaluation metrics were used to determine how well the clustering algorithm performed with different numbers of clusters. Metrics included the coefficient of determination (R^2), mean silhouette, inter- and intra-cluster distance and its ratio, and the Dunn's index.[2, 3] The larger the values of Dunn's index, mean silhouette, and inter-/intra-cluster-mean distance ratio, the better the model fits the dataset. As a rule of thumb, models with smaller numbers of clusters and R^2 higher than 0.3 are generally preferred.[4]

Mathematical definitions of distances, clustering evaluation metrics & mixing probability

In the definitions below, we consider M individuals (or, more technically, their parameter vectors) $\{\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_M\}$ distributed over N different clusters $\{C_1, \ldots, C_N\}$. The outcome measures can be calculated for any step in the clustering process; therefore, N can be any number between M and 2 (inclusive). Let $|C_i|$ denote the size of the cluster i, and let C_{ij} be its j-th individual.

For example, if cluster 2 contains individual with IDs 2, 42 and 100, then $C_2 = (\mathbf{p}_2, \mathbf{p}_{42}, \mathbf{p}_{100})$ and $|C_2| = 3$. Moreover, $C_{2,1} = \mathbf{p}_2$, $C_{2,2} = \mathbf{p}_{42}$ *and* $C_{2,3} = \mathbf{p}_{100}$.

Distance between individuals: Gower's distance

In this work, we define distances between individualsvia Gower's metric. Gower's metric is a weighted Manhattan metric that accounts for missing data. If somebody has a missing value on one of the dimensions, then that dimension will be ignored for all distance calculations involving thatindividual. The dimension will however be used when it is known for both individuals in a pair. Let there be L dimensions, with weights $\{w_1, w_2, \ldots, w_L\}$. Let O_{ij} be a Boolean variable denoting whether the j-th dimension has been observed for the i-th individual (1: observed, 0: missing).

The Gower's distance d_P between individuals $\mathbf{p}_i$ and $\mathbf{p}_j$ is then

$$d_P\left(\mathbf{p}_i, \mathbf{p}_j\right) = \frac{\sum_{k=1}^{L} \mathcal{O}_{ik}\mathcal{O}_{jk}w_k \left| \left(\mathbf{p}_i\right)_k - \left(\mathbf{p}_j\right)_k \right|}{\sum_{k=1}^{L} \mathcal{O}_{ik}\mathcal{O}_{jk}w_k}$$

Distance between clusters: complete linkage

We define the distance between two clusters by means of an additional rule, called *complete linkage,* applied to the set of all distances between two individuals (taking one from each cluster).

Under complete linkage, the inter-cluster distance d_C is chosen to be the largest distance of this set:

$$d_C\left(C_i, C_j\right) = \max_{k\in\{1,\ldots,|C_i|\}, l\in\{1,\ldots,|C_j|\}} d_P\left(C_{ik}, C_{jl}\right)$$

Mean inter-cluster distance

For each cluster pair, we calculate the inter-cluster distance $d_C(C_i, C_j)$, and define the *mean inter-cluster distance* $< d_C >$ as the arithmetic mean of these distances.

Note that the choice of linkage type governs the meaning of $d_C(C_i, C_j)$ and, in turn, of $< d_C >$.

$$\langle d_C \rangle = \frac{2}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} d_C(C_i, C_j)$$

Mean intra-cluster distance

For each cluster, we consider all pairs *{$\boldsymbol{p}_i$, $\boldsymbol{p}_j$}* of parameter vectors within it. For each such pair, we calculate the distance *$d_P(\boldsymbol{p}_i,\boldsymbol{p}_j)$*. We define the *mean intra-cluster distance* d_{intra} as the arithmetic mean of these distances.

Note that the choice of linkage type does not affect this outcome measure, and that it is not defined when $N = M$ (i.e., when all individuals are in separate clusters, so that no intra-cluster distances exist).

$$d_{\mathrm{intra}} = \frac{\sum_{i=1}^{N} \sum_{j=1}^{|C_i|-1} \sum_{k=j+1}^{|C_i|} d_P(C_{ij}, C_{ik})}{\sum_{i=1}^{N} \frac{1}{2}|C_i|(|C_i|-1)}$$

Silhouette

The silhouette is defined on a per-individual basis. Intuitively, this measure is high if an individual is similar to other individuals of its cluster and distinct from the individuals of other clusters.

More specifically, let the *i*-th individual, represented by $\boldsymbol{p}_i$, be the *j*-th cluster's *k*-th individual: $\boldsymbol{p}_i = C_{jk}$. Let a($\boldsymbol{p}_i$) be the mean Gower's distance of $\boldsymbol{p}_i$ to other members of its cluster. Besides, consider all clusters of which $\boldsymbol{p}_i$ is not a member, and for each, consider the mean distance between $\boldsymbol{p}_i$ and its members. Let b($\boldsymbol{p}_i$) be the minimum of these mean distances.

If a($\boldsymbol{p}_i$) < b($\boldsymbol{p}_i$), the silhouette s($\boldsymbol{p}_i$) = 1 - a($\boldsymbol{p}_i$) / b($\boldsymbol{p}_i$) > 0. (This is the preferred case.)

If a($\boldsymbol{p}_i$) = b($\boldsymbol{p}_i$), the silhouette s($\boldsymbol{p}_i$) = 0.

If a($\boldsymbol{p}_i$) > b($\boldsymbol{p}_i$), the silhouette s($\boldsymbol{p}_i$) = b($\boldsymbol{p}_i$) / a($\boldsymbol{p}_i$) – 1 < 0.

Although the choice of linkage type affects how clusters form, it does not affect the definition of this outcome measure.

$$a(\mathbf{p}_i) = a(C_{jk}) = \frac{1}{|C_j|-1} \sum_{l=1, l\neq k}^{|C_j|} d_P(C_{jk}, C_{jl})$$

$$b(\mathbf{p}_i) = b(C_{jk}) = \min_{l=1, l\neq j}^{N} \frac{1}{|C_l|} \sum_{m=1}^{|C_l|} d_P(C_{jk}, C_{lm})$$

$$s(\mathbf{p}_i) = \frac{b(\mathbf{p}_i) - a(\mathbf{p}_i)}{\max(a(\mathbf{p}_i), b(\mathbf{p}_i))}$$

Dunn index

Like the silhouette, the family of Dunn-like indices attempts to capture the distinctness of the clusters. However, it is not defined on a per-individual basis but for the dataset as a whole. We define the Dunn index (DI) as the ratio between the smallest inter-cluster distance and the largest intra-cluster distance. As such, it is high only if all clusters are compact relative to even the smallest separation between clusters.

Note that, due to the inclusion of the inter-cluster distance in the definition, the Dunn index depends on the choice of linkage type. It is defined only when at least one cluster has at least two distinct individuals. Other Dunn-like indices can be defined by quantifying the notion of cluster compactness differently.

$$\mathrm{DI} = \frac{\min_{i=1,j=i+1}^{N-1,N} d_C\left(C_i, C_j\right)}{\max_{i=1,j=1,k=j+1}^{N,|C_i|-1,|C_i|} d_P\left(C_{ij}, C_{ik}\right)}$$

Coefficient of determination

Interpreting the center of a cluster as the prototypical individual for that cluster, the coefficient of determination R^2 (between 0 and 100%) quantifies the degree to which the set of N typical individuals summarizes the dataset as a whole. The measure is high if most individuals are similar to the typical individual of the cluster they are assigned to, and dissimilar from the center of the whole dataset.

Specifically, R^2 measures the goodness of fit of the clustering, when viewed as a model with the N typical patient vectors as its inferred parameters. The distance between individual $\mathbf{p}_i = C_{jk}$ and the center $< C_j >$ of corresponding cluster C_j is a fitting residual, and R^2 is conventionally defined as one minus the ratio between the sum of squares of these residuals, and the sum of squares of the residuals of the simplest model. The simplest model is the 1-cluster model (i.e., when the dataset is not partitioned).

Note that, as the number of clusters increases, R^2 is bound to increase. However, an increased model complexity (and thus an increased goodness of fit) does not imply increased reproducibility or predictive power.

Although the choice of linkage type affects how clusters form, it does not affect this definition.

$$\langle C_i \rangle = \frac{1}{|C_i|} \sum_{j=1}^{|C_i|} C_{ij}, \ \ \langle \mathbf{p} \rangle = \frac{1}{M} \sum_{i=1}^{M} \mathbf{p}_i$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N} \sum_{j=1}^{|C_i|} d_P\left(C_{ij}, \langle C_i \rangle\right)^2}{\sum_{i=1}^{M} d_P\left(\mathbf{p}_i, \langle \mathbf{p} \rangle\right)^2}$$

Jackknife resampling – mixing probability

If the clustering is robust against addition or removal of data, individuals that were in different clusters in the original run should remain in different clusters in jackknife runs. If members of the *i*-th original cluster are indeed seldom grouped with members of the *j*-th original cluster during jackknife runs, the mixing probability M_{ij} is low; if such groupings occur often, M_{ij} is high.

After running a clustering run yielding clusters $\{C_1, \ldots, C_N\}$, we consider K jackknife runs in which we incorporate only a fraction f of all individuals, yielding a total of fM included individuals per jackknife run. Let J_{kij} be a Boolean variable denoting whether the *j*-th individual of the *i*-th cluster from the original run occurs in the *k*-th jackknife run. Finally, denote the *i*-th cluster of the *k*-th jackknife run by C_i^k.

The mixing probability M_{ij} is the fraction of pairs of individuals, one from original cluster C_i and one from original cluster C_j, who are grouped together in jackknife iterations:

$$M_{ij} = \frac{\sum_{k=1}^{K} \sum_{l=1}^{|C_i|} \sum_{m=1}^{|C_j|} \sum_{n=1}^{N} \mathbb{I}_{C_{il} \in C_n^k \wedge C_{jm} \in C_n^k}}{\sum_{k=1}^{K} \sum_{l=1}^{|C_i|} \sum_{m=1}^{|C_j|} J_{kil} J_{kjm}}$$

References

1. Pavoine S, Vallet J, Dufour AB, Gachet S, Daniel H. On the challenge of treating various types of variables: application for improving the measurement of functional diversity. *Oikos*. 2009;118(3):391-402; doi:10.1111/j.1600-0706.2008.16668.x
2. Halkidi M, Batistakis Y, Vazirgiannis M. On Clustering Validation Techniques. *J Intell Inf Syst*. 2001;17:107-145; doi:10.1023/A:1012801612483
3. Rousseeuw PJ. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Math*. 1987;20:53-65; doi:10.1016/0377-0427(87)90125-7
4. Moore DS, Notz WI, Flinger MA. *The Basic Practice of Statistics*. 6 ed. W.H. Freeman and Company; 2013.
5. Khatami R, Luca G, Baumann CR, et al. The European Narcolepsy Network (EU-NN) database. *J Sleep Res*. 2016;25(3):356-364; doi:10.1111/jsr.12374

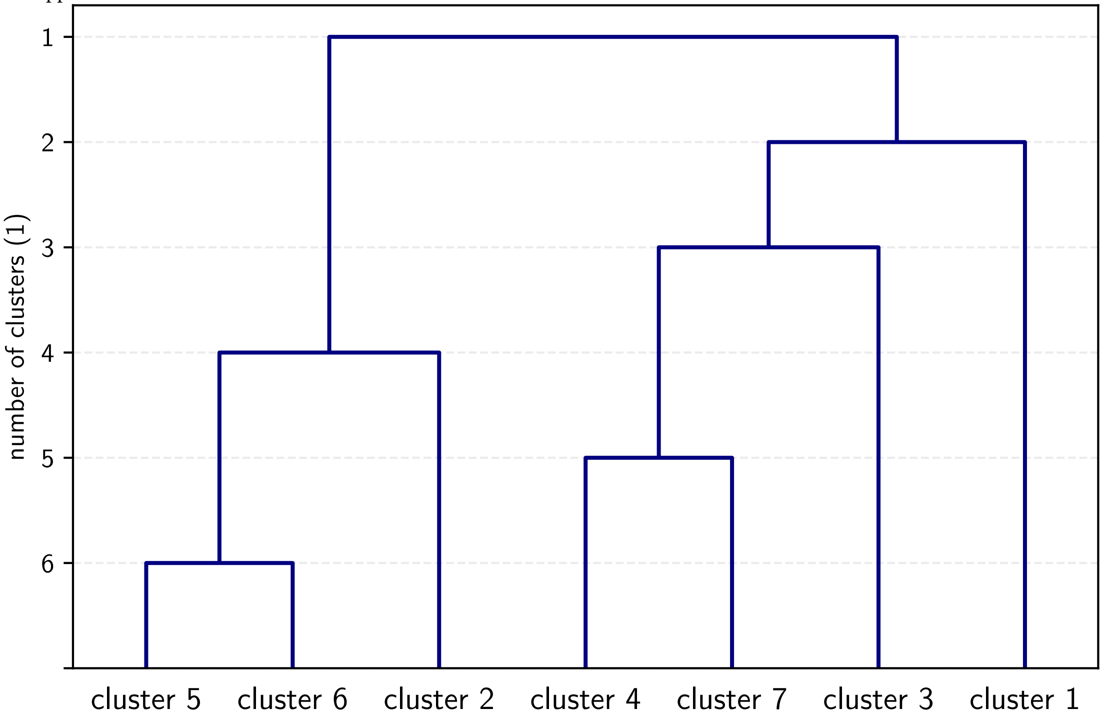
eAppendix 4: Ethical approval and centers of inclusion

Standard protocol approvals, registrations, and patient consents

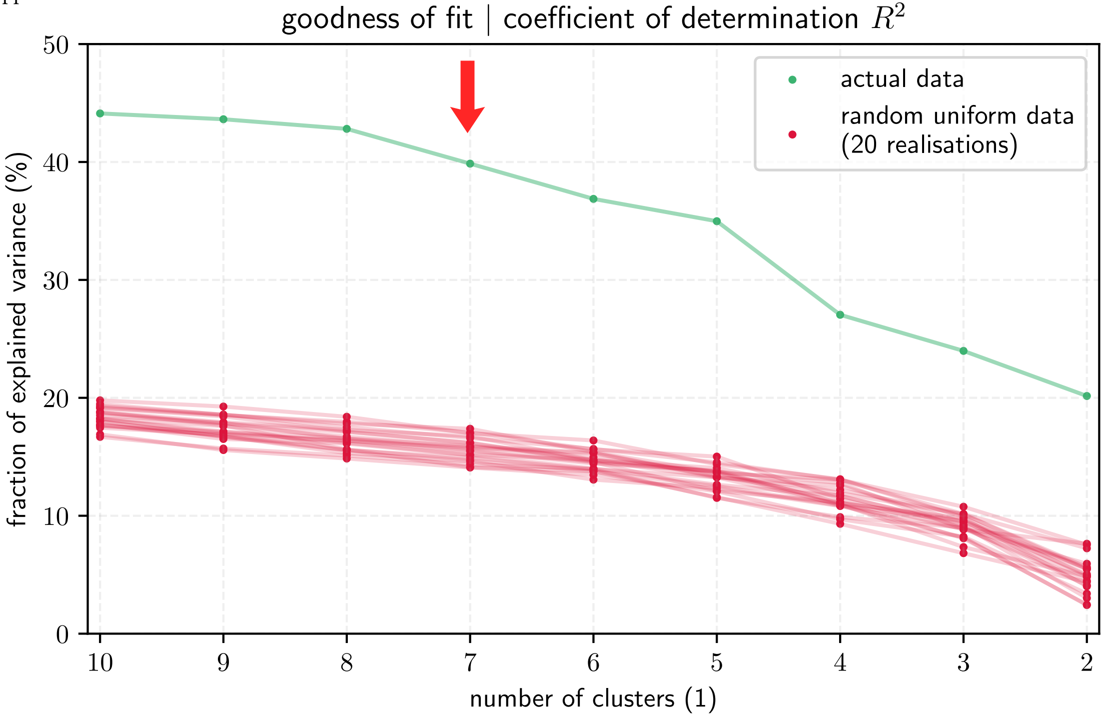
Each center of EU-NN obtained ethical approval by a national Institutional Review Board before entering individuals in the database.[5] The scientific committee of EU-NN approved the study protocol. All individuals provided their informed consent to be included into the EU-NN database.

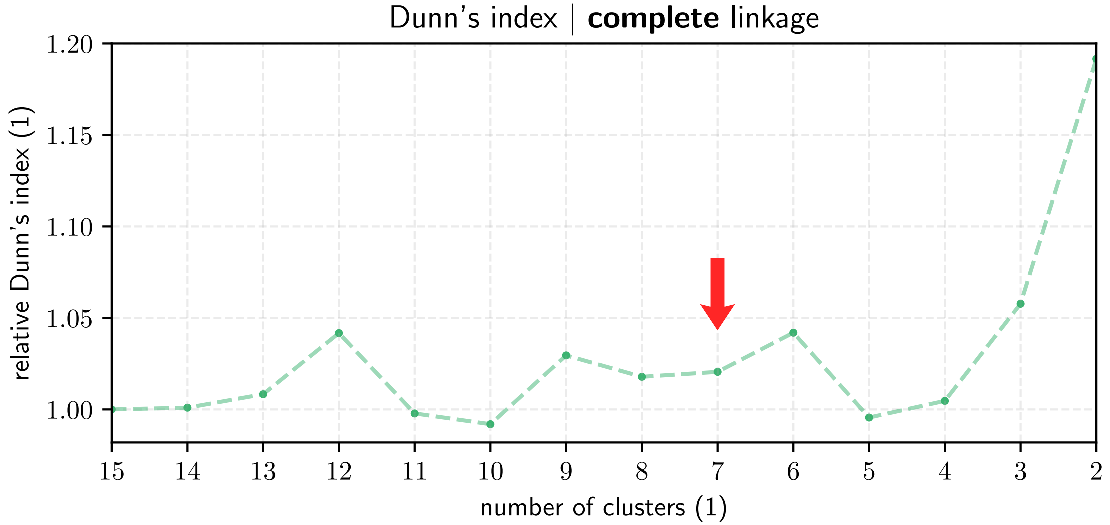
Center abbreviation	Number of inclusions	Main contact person
BAR (Barmelweid)	58	Ramin Khatami
BCN (Barcelona)	56	Joan Santamaria
BER (Bern)	3	Claudio Bassetti
BOL (Bologna)	125	Giuseppe Plazzi
HEE (Heeze)	48	Sebastiaan Overeem
HEL (Helsinki)	50	Markku Partinen
HGUGM (Madrid)	41	Rosa Peraita-Adrados
INN (Innsbrück)	31	Birgit Högl
KOS (Košice)	10	Eva Feketeova
LAU (Lausanne)	34	Raphael Heinzer
LEI (Leiden)	44	Gert Jan Lammers
LUG (Lugano)	8	Mauro Manconi
MAD (Madrid)	64	Rafael del Rio-Villegas
MON (Montpellier)	250	Yves Dauvilliers
MUE (Münster)	11	Anna Heidbreder
NIMH (Klecany)	7	Jitka Bušková
OPO (Porto)	33	Antonio Martins da Silva
PAL (Palma de M.)	5	Francesca Canellas
PRA (Prague)	45	Karel Šonka
SCH (Schwalmstadt)	142	Geert Mayer
WAR (Warschau)	17	Aleksandra Wierzbicka

eAppendix: 5A

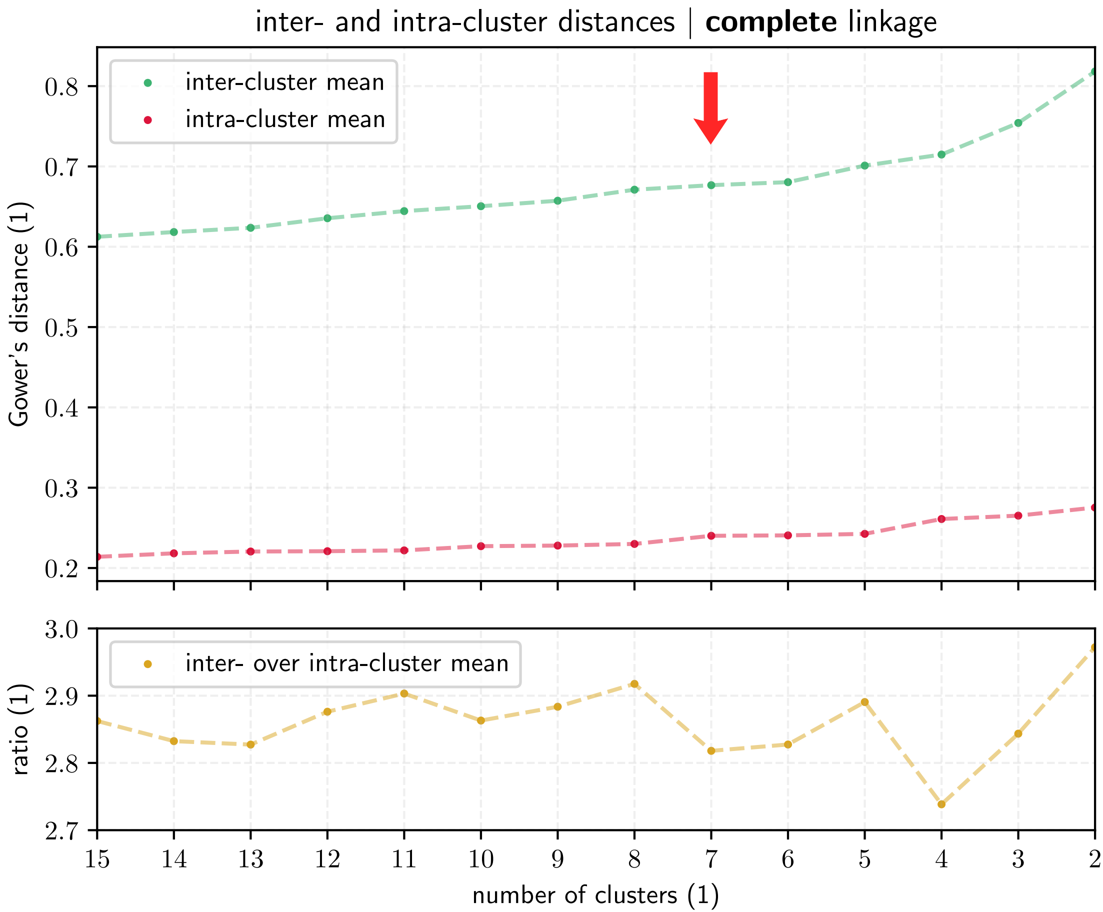


eAppendix: 5B

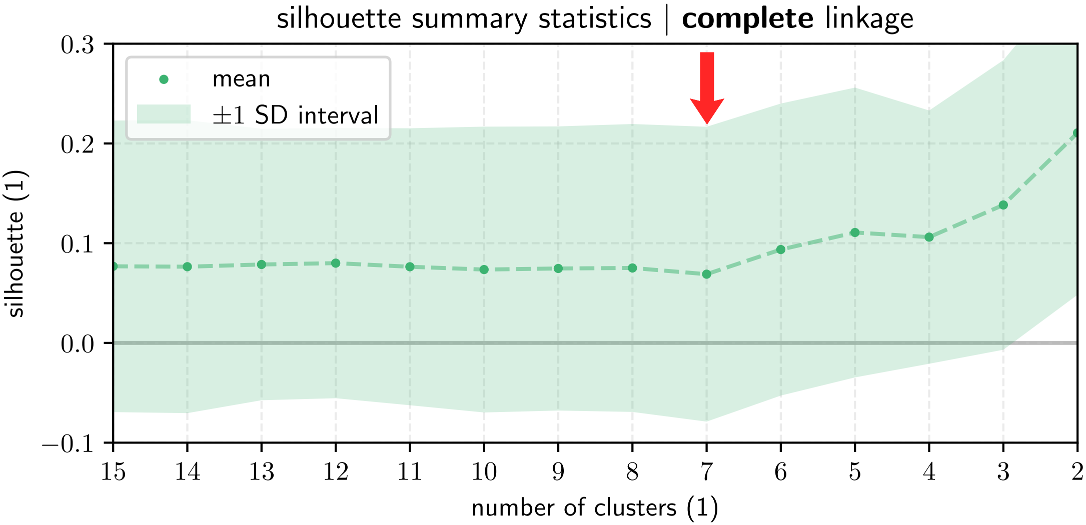




Dunn's index | **complete** linkage
relative Dunn's index (1)
1.20
1.15
1.10
1.05
1.00
15
14
13
12
11
10
9
8
7
6
5
4
3
2
number of clusters (1)



inter- and intra-cluster distances | **complete** linkage
inter-cluster mean
intra-cluster mean
Gower's distance (1)
0.8
0.7
0.6
0.5
0.4
0.3
0.2
inter- over intra-cluster mean
ratio (1)
3.0
2.9
2.8
2.7
15
14
13
12
11
10
9
8
7
6
5
4
3
2
number of clusters (1)



silhouette summary statistics | **complete** linkage
mean
±1 SD interval
silhouette (1)
0.3
0.2
0.1
0.0
−0.1
15
14
13
12
11
10
9
8
7
6
5
4
3
2
number of clusters (1)

eAppendix 5: legends

eAppendix 5. A: In this work, we have analysed agglomerative hierarchical clustering of the EU-NN database at seven clusters. The dendrogram displays the step-wise merger process from seven clusters (bottom) to one (top), showing the continuation of the process beyond our choice of termination. The cluster id on the x-axis correspond to the cluster id as described in the results of the manuscript.

eAppendix 5. B: The coefficient of determination (R^2) is a measure of goodness of fit, and a standard clustering evaluation metric. We show R^2 as a function of the number of clusters, for both the original clustering (green), and for 20 simulated random datasets with no intrinsic clustering (red). The greater coefficient of determination of the EU-NN database indicates intrinsic clusterability.

eAppendix 5. C: Displays the Dunn index for different numbers of clusters, as a standard clustering evaluation metric. The Dunn index is the ratio between the smallest inter-cluster distance and the largest intra-cluster distance. As such, it is high only if all clusters are compact relative to even the smallest separation between clusters.

eAppendix 5. D: Displays the mean inter-cluster and intra-cluster distance and the ratio between the two, as standard clustering evaluation metrics. Higher inter-cluster distance, lower intra-cluster distance and a higher ratio indicate more distinct clusters and hereby better clustering.

eAppendix 5. E: Displays the mean and standard deviation of the sillhouette for different numbers of clusters, as a standard clustering evaluation metric. High sillhouette values are preferable and indicate that an individual is more similar to other individuals of its cluster and distinct from the individuals of other clusters.

eAppendix 6: Overview of raw values per variable per cluster

eAppendix 6: Overview of raw values per variable per cluster

eAppendix 6: Overview of raw values per variable per cluster

eAppendix 6: Overview of raw values per variable per cluster

eAppendix 6: Overview of raw values per variable per cluster

eAppendix 6: legend

eAppendix 6: The histograms display the distribution of each variable for each cluster. Just above each distribution, the number of observations (N) and the sample mean (μ) and standard deviation (σ) are indicated. For details on individual variables, see Supplementary material 1.