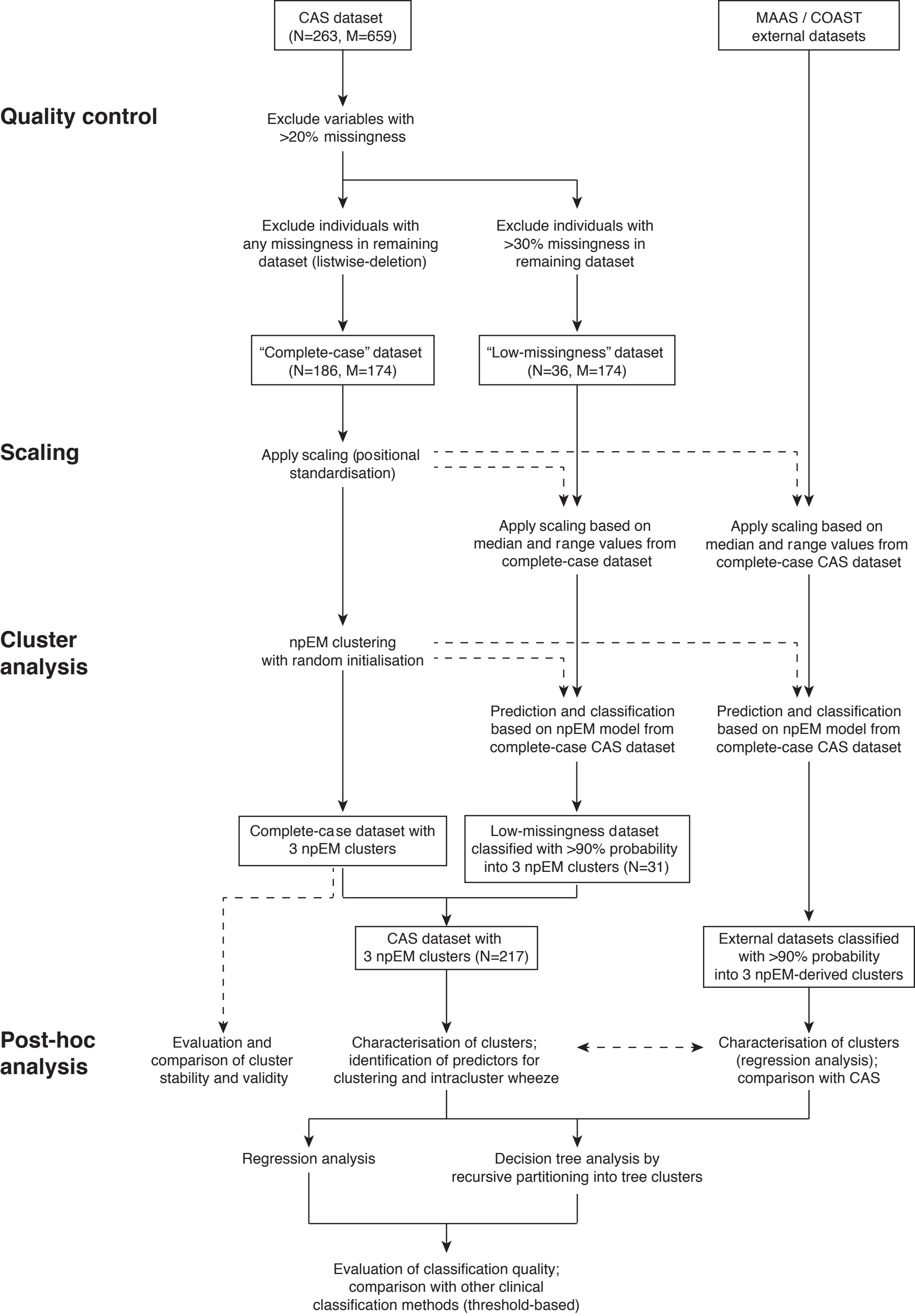
CAS dataset
(N=263, M=659)
MAAS / COAST
external datasets
Quality control
Exclude variables with
>20% missingness
Exclude individuals with
any missingness in remaining
dataset (listwise-deletion)
Exclude individuals with
>30% missingness in
remaining dataset
"Complete-case" dataset
(N=186, M=174)
"Low-missingness" dataset
(N=36, M=174)
Scaling
Apply scaling (positional
standardisation)
Apply scaling based on
median and range values from
complete-case dataset
Apply scaling based on
median and range values from
complete-case CAS dataset
Cluster
analysis
npEM clustering
with random initialisation
Prediction and classification
based on npEM model from
complete-case CAS dataset
Prediction and classification
based on npEM model from
complete-case CAS dataset
Complete-case dataset with
3 npEM clusters
Low-missingness dataset
classified with >90% probability
into 3 npEM clusters (N=31)
CAS dataset with
3 npEM clusters (N=217)
External datasets classified
with >90% probability
into 3 npEM-derived clusters
Post-hoc
analysis
Evaluation and
comparison of cluster
stability and validity
Characterisation of clusters;
identification of predictors for
clustering and intracluster wheeze
Characterisation of clusters
(regression analysis);
comparison with CAS
Regression analysis
Decision tree analysis by
recursive partitioning into tree clusters
Evaluation of classification quality;
comparison with other clinical
classification methods (threshold-based)