Using Unsupervised Learning to Determine Risk Level for Left Ventricular Diastolic Dysfunction

Kaidi Ma[1], Marco Canepa, MD[2,3], James B. Strait, MD, PhD[2,3], Hagit Shatkay, PhD[1]
[1]*Dept. of Computer and Information Sciences, University of Delaware, Newark, DE, USA*
[2]*Longitudinal Studies Section, Clinical Research Branch, National Institute on Aging, NIH, Baltimore, MD, USA*
[3]*Laboratory of Cardiovascular Sciences, Human Cardiovascular Studies Unit, National Institute on Aging, NIH, Baltimore, USA*
kdma@udel.edu, marco.canepa@nih.gov, straitj@mail.nih.gov, shatkay@udel.edu

***Abstract*—Left Ventricular Diastolic Dysfunction (LVDD) is a decompensatory change in the relaxation properties of the heart, the risk for which increases with age. Currently, physicians use a decision-tree-like algorithm to distinguish between discrete LVDD levels. This approach, based on cut-off thresholds, can potentially lead to information loss and possibly to misdiagnosis.**

This paper aims to explore an alternative diagnostic method to determine LVDD risk level, taking into account a wide variety of attributes available in patient records, without pre-setting cut-off thresholds. Using a large dataset derived from the Baltimore Longitude Study of Aging (BLSA), and adjusting the data for age and gender, we employ the Chi Square test and the information gain criterion to identify attributes that correlate well with the physician-assigned grades; such attributes are referred to as *distinguishing attributes*. We then apply the expectation maximization (EM) algorithm, as well as the K-Means, in order to cluster records that are represented using distinguishing attributes.

While clusters resulting from the K-Means are not stable, three stable and tightly-formed clusters, which are obtained from the EM algorithm, roughly correspond to the physician-assigned categories. Based on the results from the EM algorithm, we can compute a patient's probability to have low, high or no risk for LVDD, and use this probability as a basis for defining a risk score to determine the patient's LVDD severity.

Keywords—unsupervised learning; clustering; left ventricle diastolic dysfunction; EM algorithm

I. INTRODUCTION

Left Ventricular Diastolic Dysfunction (LVDD) is a cardiac condition caused by decompensatory changes in the relaxation of the heart, the risk for which increases with age[1]. It is characterized by elevated filling pressures in the left ventricle despite normal or sub-normal diastolic volume. LVDD may result in insufficient pumping of blood to the rest of the body. Diagnosis and treatment is often difficult[2], because similar blood flow pattern may occur for both high and low LVDD severity levels. As diastolic dysfunction is a major cause of heart failure[4], it is important to make a correct early diagnosis so that patients can get appropriate treatment and maximal benefit.

Currently, cardiologists assign four *discrete* grades of LVDD[3], namely: *normal condition* (*grade 0*), *impaired relaxation pattern* (*grade 1*), *pseudonormal filling dynamics* (*grade 2*), and *restrictive filling dynamics* (*grade 3*). This assignment is based on pre-defined cut-off thresholds applied to three attributes (See [3] for the algorithm): the ratio of the early (E) to late (A) ventricular filling velocities (E/A ratio), indexed Left Atrial Volume (LAVi), and the ratio of E to early diastolic mitral annular velocity (denoted Em) (E/Em ratio).

The E/A ratio describes the ratio between passive and active blood flow across the mitral valve[7]. In a healthy heart, the E velocity is higher than the A velocity[10]. With aging, the left ventricular wall may become stiff, leading to decreased passive filling capacity of the left ventricle and an increasing need for active filling[4]. The fact that both people with normal diastolic function and patients with severe LVDD can have an E/A ratio higher than 1, makes early diagnosis of LVDD difficult. Thus additional measures are recommended[18]: *i)* The E/Em ratio is used for estimating left ventricular (LV) filling pressure[9], derived from the movement speed of the ventricular tissue[3]. As elevated filling pressures are the main physiological consequence of diastolic dysfunction (DD), identifying those pressures can help detect DD[8]. Typically an E/Em ratio above *12* is associated with increased LV filling pressures[3]. *ii)* The indexed Left Atrial Volume (LAVi) assessment[11], calculated by indexing the volume of the left atrium to body surface area, reflects the cumulative effects of filling pressures over time[5], unlike E/A or E/Em ratios, which reflect filling pressures at the time of measurement. LAVi is used to differentiate between LVDD grades 0 vs. 2 and 0 vs. 3, when the E/Em ratio is in the intermediate range of 8-12.

The threshold-based grading suffers several shortcomings, including the use of only a few available measurements in the diagnosis, and the potential loss of information through the use of pre-set thresholds. We propose an alternative framework to determine the LVDD risk level, without using cut-off thresholds. Our scoring framework is based on unsupervised learning, and does not assign discrete grades, but rather indicates the probability of a condition to be severe, given attributes that are tested to be informative about LVDD severity levels.

By assigning probabilities, this framework avoids the drawback of using explicit cut-off threshold, and has the potential to better utilize patient data. Moreover, using probability rather than a discrete grade has the advantage of providing more information about a patient's potential risk level as opposed to a discrete binary decision of the severity group to which a person does (or does not) belong. A clear high probability value, e.g. 0.9, may help physicians to confidently assign a high LVDD severity level, while a flatter probability distribution (e.g. 0.55 vs. 0.45) can assist in recognizing borderline cases.

The rest of the paper is organized as follows: Section II introduces the data and the terminology. Section III explains the clustering algorithms that are used. Section IV presents the results while Section V concludes the paper, discusses outcomes limitations, and outlines directions for the future work.

978-1-4799-5669-2/14/$31.00 ©2014 IEEE

II. DATA AND DISTINGUISHING ATTRIBUTES

The dataset throughout the work consists of records from a subset of patients who participated in the Baltimore Longitude Study of Aging (BLSA) [7]. A total of 6,642 records were gathered from 1,888 participants during multiple visits over a period of 34 years (1979-2012). Four discrete grades were assigned to the patients using the decision-tree-like algorithm[3].

We focus on records with assigned LVDD grades of 0-2, leaving out records with grade 3, because it is straightforward for a physician to identify the most severe cases and there is less room for improving this aspect. In contrast, lower severity levels are relatively harder for physicians to differentiate. As multiple visits are not aligned in time across patients, we use only the last visit record of each patient as our base dataset, for a total of 890 records (430 are male and 460 are female; 533 people are grade 0, 258 are graded 1 and 99 are graded 2).

Each record is described by 182 attributes, excluding identification fields (*id* and *visit time*). Two types of attributes are assessed: *continuous* and *discrete*. Most of the continuous attributes are Doppler ultrasound measurements[9], describing either heart structure or blood flow. Most discrete attributes consist of *Yes/No* answers to clinical inquiries. Not all attributes are informative for distinguishing among LVDD grades. We aim to identify attributes that correlate well with – and are potentially predictive of – LVDD severity level, and refer such attributes as *distinguishing attributes*. To find out these attributes, we use physician-assigned LVDD severity levels as ground-truth class labels, and employ the Pearson's Chi Square test[12] to identify attributes that correlate strongly with LVDD grades.

The magnitude of association between an attribute and the LVDD severity level may be exaggerated due to confounding factors[6], e.g. age and gender. To identify attributes that do not reflect age or gender, but rather truly reflect the propensity of a patient (of any age) to demonstrate a certain LVDD severity level, we separated the data into *female* and *male* groups, as well as into different age groups, and created 10 female sets and 10 male sets. The association of each attribute with the physician-assigned grade is tested separately, within each of the 20 datasets, as well as within the original dataset as a whole. Table I shows the 20 datasets with their age range and the number of records within each set.

To ensure each dataset provides statistically significant measures, we allow overlapping age interval across age-groups. An advantage of having such an overlap is that we can test an attribute on a variety of age scopes to ensure the attribute is independent of age. Attributes that consistently distinguish between the same pair of LVDD grades regardless of the variability in age boundaries, are more likely to be associated with LVDD grade regardless of the specific age. We also removed attributes that have a high ($\geq$ 40%) rate of missing values, because: *1)* It is impractical to impute data when the missing rate is so high; and *2)* Ignoring missing values in an attribute with high missing rate introduces bias and skews to the estimates of parameters like mean and variance.

Having removed high missing rate attributes, we employ the Chi Square test to check if there is a significant association between a remaining attribute and the LVDD severity level. The continuous attributes are discretized by using *equal-width binning*[14] before the test. We examine one attribute and one pair of LVDD grades, S_i and S_j, at a time, to check whether the distributions of the attribute's value are the same or different between records with S_i compared to those with S_j, Three pairs of LVDD severity grades are compared: 0 vs. 1, 1 vs. 2, and 0 vs. 2. If the p-value generated from the Chi Square test is lower than 0.05, it suggests that the distributions of the attribute values are statistically significantly different, between one LVDD grade vs. another. If all p-values associated with the attribute are low across all three pairs of grades, we conclude that the attribute is informative and can distinguish well among all LVDD severity levels.

TABLE I. DATASETS CREATED FOR AGE AND GENDER ADJUSTMENT. *AGE SCOPE* SHOWS THE RANGE OF AGE FOR EACH DATASET. *COUNT* SHOWS THE SIZE OF EACH DATASET. '*F*' IN THE DATASET NAME, INDICATES ONLY FEMALE RECORDS ARE INCLUDED, WHILE 'M' INDICATES MALE.

Age Scope	Female		Male	
	Dataset Name	Count	Dataset Name	Count
50-64	F5064	139	M5064	99
50-70	F5070	204	M5070	159
≥50	F50up	375	M50up	407
60-74	F6074	175	M6074	159
65-80	F6580	163	M6580	196
≥65	F65up	236	M65up	308
75-90	F7590	108	M7590	170
≥75	F75up	120	M75up	185
≥81	F81up	73	M81up	112
Full range	Female	430	Male	460

The Chi Square test results indicated that only two attributes statistically significantly differentiate among *all three pairs* of LVDD severity levels, namely E/Em ratio and Mitral Valve Early Filling Point (MV E Point). However, quite a few attributes were found to differentiate well between two out of the three pairs of LVDD grades. While such attributes are likely to be useful for distinguishing between certain pairs of severity-levels, using them for representing patients may diminish the contribution of other informative attributes tested to be correctly distinguishing between the severity-level-pairs for which these attributes were found to be non-informative. To minimize this type of effect, we set *0.4* as an upper bound on the p-value for selecting distinguishing continuous attributes, while we employ the information gain[13] (IG), which is used for measuring how well an attribute discriminates between target classes (LVDD grades), as an additional measure to ensure that the selected discrete attributes are informative. We found that if a discrete attribute contributes an IG below *0.018*, the attribute is unlikely to differentiate well between a pair of grades. Therefore discrete attributes whose IG is below 0.018 are removed. For the remaining attributes, if all three p-values obtained from the Chi Square tests are below 0.4 and two of them are below 0.05, these attributes are selected as *distinguishing attributes*; each of them can statistically significantly distinguish at least two pairs of LVDD grades without impeding the differentiation with regard to the third pair. Table II shows the distinguishing attributes and the pairs of LVDD grades that are distinguished by them. Each distinguishing attribute can differentiate between at least two pairs of LVDD grades.

III. CLUSTERING METHODS

Having identified the distinguishing attributes, we employ two well-known clustering methods, K-means[19] and the Expectation Maximization[15] (EM), to cluster patients based on the distinguishing attribute values that are used to represent each records. We use our own implementation rather than open-source tools such as Weka[16], because the latter do not

TABLE II. DISTINGUISHING ATTRIBUTES AND THE PAIRS OF LVDD GRADES THAT ARE STATISTICALLY SIGNIFICANTLY DIFFERENTIATED.

	Attribute Name	Pair of LVDD grades that are statistically differentiated		
		0 vs. 1	*0 vs. 2*	*1 vs. 2*
Continuous Attribute	E/A Ratio	Y		Y
	Doppler Heart Rate (Echo HR)	Y		Y
	End Diastolic Volume Teich Algorithm (EDV Teich)	Y	Y	
	E/Em Ratio	Y	Y	Y
	Left Atrium (LA) Dimension	Y	Y	
	Indexed LA Volume (LAVi)		Y	Y
	Lateral Em	Y	Y	
	MDRD Formula for Creatinine Clearance Rate (MDRDcr_cl)	Y	Y	
	Mitral Valve Early Filling Point (MV E Point)	Y	Y	Y
	Relative Wall Thickness (RWT)	Y	Y	
	Septal Am		Y	Y
Discrete Attribute	β-Blockers	Y	Y	
	Angiotensin Type (ANG Medint)	Y	Y	
	Renin-Angiotensin-Aldosterone System Drug (RAAS Drugs)	Y	Y	
	Simplified Renal Disease Formula (MDRDrenfail)	Y	Y	
	Heart Failure Type (HF Medint)		Y	Y
	Atherosclerosis (Athero)	Y	Y	
	Coronary Artery Disease (CAD)	Y	Y	

support customized initialization of the cluster's parameters. Both clustering algorithms were applied to three datasets: female, male, and the dataset with both genders. We do not run experiments over each of the 20 datasets described in Section II, as those were only devised for selecting distinguishing attributes. Moreover, many sets are small, and do not reflect the whole patient population. As such, we use three large datasets of female, male and gender-mixed set in the experiments.

The initialization of cluster parameters is non-random, and is based on the physician assigned LVDD grades, as these assignments reflect some prior knowledge about LVDD categories. This information would be lost if random initialization is used. As there are three LVDD grades that are of interest in the data, we use three clusters in the experiments, although future work will examine a larger number of clusters, to possibly find concrete subclasses of patients within the data. Notably, the clustering algorithms as a whole *do proceed in an unsupervised fashion*, even when the initialization is based on labels that were pre-assigned to the data.

Each algorithm requires initialization of several parameters. The maximum iteration number for both clustering methods is set to 100. For K-Means, we initialize the centroid of each cluster by random seeding from different LVDD grades. A random *grade 0* record is used as the initial centroid for cluster 0, likewise, a random *grade 1* record is used for cluster 1, and a random *grade 2* record is used for cluster 2. This seeding reflects a priori bias toward having three clusters that roughly correspond to three major severity levels. To test the stability of the K-means, we run 20 experiments on each of the datasets; each experiment initialized using a different seed assignment. The parameters for the EM include: for each cluster, the initial Gaussian *mean* and *variance* of each continuous distinguishing attribute, and an initial value for seeding the mode of each discrete distinguishing attribute for each cluster. To initialize these parameters, as well as to utilize the prior know-ledge of physician assigned categories, we sample a small set according to a fixed (small) *sampling fraction*[17]. For a sampling fraction of *a%*, we draw at random *a%* of the records in each LVDD category to initialize each potentially respective cluster. From the sample, we compute the Gaussian mean and variance of each distinguishing attribute A_i for each cluster. For a discrete attribute D_k, we find the mode of D_k in the sample to initialize the value of D_k in the cluster. We keep the sampling fractions low so as to not over-bias the EM, and let the algorithm converge in an unsupervised manner based on all the data it receives. To test stability of the EM, we chose four low fractions: 10%, 20%, 30%, and 40% to create different sample sets. We run 20 experiments for each fraction and compare the clusters obtained from the EM with the physician-assigned categories.

IV. RESULTS AND DISCUSSION

Clustering is usually applied to datasets where no prior class information is available. However, to validate the connection between the obtained clusters and the physician-assigned LVDD grades, we use Table III to analyze the resulting clusters, showing the distribution of records in each cluster (and each LVDD category) for each dataset and demonstrating how stable in size the resulting cluster is across different runs.

Clusters obtained from the K-means showed high variability and instability depending on the initial centroids; the distribution of instances in clusters varies greatly across experiments, including results with empty cluster, or uniformly distributed clusters. As such, we do not further discuss these results here. In contrast, the clusters produced by the EM show stability across all runs, and the respective cluster sizes remain similar across all experiments under different initializations. Approximately 98% of the records in each dataset were always assigned into the same cluster across different runs. Based on the data shown in Table III, the ratio between the size of the assigned cluster 0 and the size of dataset remains at around 37%; the size of cluster 1 roughly remains at around 43%; the size of cluster 2 roughly remains at around 18%.

According to the resulting clusters from the EM, almost all records in cluster 0 are LVDD grade 0, while most records that were LVDD grade 1 are placed in cluster 1, and most records that were assigned grade 2 are placed in cluster 2. In addition, the means of the E/A ratio in each cluster are very similar to the means in the corresponding pre-assigned LVDD categories. The standard deviation for most attributes in each cluster is smaller compared to the standard deviation measured within the pre-assigned categories. This indicates that tighter clusters are formed by the EM algorithm. Since the resulting clusters are stable and tight, cluster 0 can be viewed as corresponding to *no LVDD risk*, cluster 1 can be viewed as to *low LVDD risk* and cluster 2 can be regarded as relatively *high LVDD risk*.

Our clusters are based on patient-representation through 18 distinguishing attributes, as opposed to the three used by physicians in the decision-tree-like process. Having more distinguishing attributes as part of the clusters characteristics, can provide additional information to physicians regarding recurring patterns across different severity levels of LVDD. More importantly, the availability of the clusters and their respective mean and variance, allows us to calculate, for each person, a probability to be at a low, high or no risk for LVDD. Therefore, these clusters can serve as a way to assign continuous risk scores to a patient, without employing pre-set thresholds.

TABLE III. ***A***. THE DISTRIBUTION OF RECORDS IN EACH CATEGORY ASSIGNED BY A PHYSICIAN, ACROSS THE THREE MAIN DATASETS; ***B.*** THE DISTRIBUTION OF RECORDS AMONG THE CLUSTERS OBTAINED FROM THE EM ALGORITHM, ACROSS THE THREE MAIN DATASETS, UNDER THE FOUR DIFFERENT INITIALIZATION SETTINGS. (SAMPLING FRACTIONS).

A.	**Physician-Assigned LVDD Category**			
	Group 0	***Group 1***	***Group2***	***Total***
Male dataset	260 (60.5%)	109 (25.3%)	61 (14.2%)	430 (100%)
Female dataset	273 (59.3%)	149 (32.4%)	38 (8.3%)	460 (100%)
Base dataset	533 (59.9%)	258 (29.0%)	99 (11.1%)	890 (100%)
B.	**Clusters Resulting from the EM Using Male Dataset**			
Sampling Fraction	***Cluster 0***	***Cluster 1***	***Cluster 2***	***Total***
10%	159 (37.0%)	191 (44.4%)	80 (18.6%)	430 (100%)
20%	161 (37.4%)	181 (42.1%)	88 (20.5%)	430 (100%)
30%	160 (37.2%)	182 (42.3%)	88 (20.5%)	430 (100%)
40%	160 (37.2%)	182 (42.3%)	88 (20.5%)	430 (100%)
Clusters Resulting from the EM Using Female Dataset				
10%	167 (36.3%)	216 (47.0%)	77 (16.7%)	460 (100%)
20%	171 (37.2%)	212 (46.1%)	77 (16.7%)	460 (100%)
30%	169 (36.7%)	213 (46.3%)	78 (17%)	460 (100%)
40%	171 (37.2%)	211 (45.9%)	78 (17%)	460 (100%)
Clusters Resulting from the EM Using Base Dataset				
10%	339(38.1%)	399(44.8%)	152 (17.1%)	890 (100%)
20%	334(37.5%)	390(43.8%)	166 (18.7%)	890 (100%)
30%	334(37.5%)	390(43.8%)	166 (18.7%)	890 (100%)
40%	334(37.5%)	390(43.8%)	166 (18.7%)	890 (100%)

As mentioned in Section I, these probabilities can represent a level-of-confidence, by showing how likely a patient is to be in each LVDD risk level, thus having a specifically important advantage in borderline cases, as opposed to a discrete decision of the severity group to which a person does (or not) belong. For instance, when using the decision procedure, it is difficult to decide on the correct LVDD grade for a patient, whose E/A ratio is 1.24, E/Em ratio is 10.4 and LAVi is 28.01. While the decision procedure assigns an LVDD grade of *2*, this assignment is rather arbitrary; a small measurement error (e.g. 0.011) could change the LAVi to just below the 28 ml/m^2 threshold, leading to an LVDD grade of *0*. In contrast, our probabilistic model assigns the same patient with a probability 0.927 to be at *no risk* of LVDD, a probability of 0.001 to be at *low risk*, and 0.072 to be at *high risk*. Such a probabilistic assignment is clearer, reflecting both the most likely case and the alternatives, while also likely to be more reliable as it incorporates information from 18 significant measures as opposed to only three.

V. CONCLUSION AND FUTURE WORK

In this study, we proposed an alternative way to assign LVDD risk level by clustering data using the EM algorithm, and viewing the probability of a patient to belong to each cluster as a basis for a continuous risk score. We first employed the Chi Square test to select informative distinguishing attributes. We next employed unsupervised learning to cluster patient records, where the latter were represented as vectors of distinguishing attributes. While the clusters obtained from the K-means were neither meaningful nor stable, the clusters obtained from the EM consistently correspond to physician-assigned LVDD categories. Having the stable clusters, we can compute the probability of a person to have low, high or no LVDD risk. The probabilities can serve as a basis for a continuous risk score to determine the LVDD risk level of a patient.

Throughout this study, we have used the last visit record to represent each patient. We shall extend it to include the complete set of records per patient, while accounting for the temporal aspect of the extended data. We have also focused on clustering data into three clusters, while utilizing physician-assigned LVDD labels to initialize the parameters. As the three clusters resulting from the EM are stable, in future studies we shall increase the number of clusters used, in order to possibly uncover subclasses within the data, which can be indicative of subgroups of patients with distinct LVDD forms.

ACKNOWLEDGMENT

This research by Dr. Marco Canepa, and Dr. James B. Strait was supported in part by the Intramural Research Program of the NIH, National Institute on Aging. Dr. Hagit Shatkay was supported in part by NIH Grant U54GM104941.

REFERENCES

[1] C. Satpathy, T.K. Mishra, et al, "Diagnosis and Management of Diastolic Dysfunction and Heart Failure," Am Fam Physician, Mar 2006, 73(5): 841-46.

[2] C. Gutierrez, D. G. Blanchard, "Diastolic Heart Failure: Challenges of Diagnosis and Treatment," Am Fam Physician, Jun 2004, 69(11): 2609-16.

[3] M. Canepa, J.B. Strait, et al, "Contribution of Central Adiposity to Left Ventricular Diastolic Function (From the Baltimore Longitudinal Study of Aging)," Am J Cardiol, Apr 2012, 109(8): 1171-78.

[4] J.B. Strait, E.G. Lakatta, "Aging-Associated Cardiovascular Changes and Their Relationship to Heart Failure," Heart Fail Clin, Jan 2012, 8(1): 143-64.

[5] A.M. Pritchett, D.W. Mahoney, et al "Diastolic Dysfunction and Left Atrial Volume: a population-based study," J Am Coll Cardiol, Jan 2005, 45(1): 87-92.

[6] G. Wunsch, "Confounding and Control," Demographic Research, Feb 2007, 16(4): 97-120.

[7] R.M. Lang, M. Bierig, et al, "Recommendations for Chamber Quantification," Eur J Echocardiog, Mar 2006, 7(2): 79-108.

[8] S.F. Nagueh, C.P. Appleton, et al, "Recommendations for The Evaluation of Left Ventricular Diastolic Function by Echocardiography," Eur J Echocardiogr, Mar 2009, 10(2): 165-193.

[9] S.R. Ommen, R.A. Nishimura, et al, "Clinical Utility of Doppler Echocardiography and Tissue Doppler Imaging in the Estimation of Left Ventricular Filling Pressures, A Comparative Simultaneous Doppler-Catheterization Study," Circulation, Oct 2000, 102(15): 1788-94.

[10] L. Mandinov, F.R. Eberli, et al, "Review: Diastolic Heart Failure," Cardiovasc Research, Mar 2000, 45(4): 813-25.

[11] T.S. Tsang, M.E. Barnes, et al, "Left Atrial Volume as A Morphophysiologic Expression of Left Ventricular Diastolic Dysfunction and Relation to Cardiovascular Risk Burden," Am J Cardiol, Dec 2002, 90(12): 1284-89.

[12] P.E. Greenwood, M.S. Nikulin, "A Guide to Chi-Squared Testing," Wiley, New York, 1996, pp. 10-17.

[13] T.M. Mitchell, "Machine Learning," McGraw-Hill, Inc, New York, NY, USA, 1997, pp. 55-58.

[14] I. Witten, E. Frank, "Data Mining: Practical Machine Learning Tools and Techniques," Second Edition, Morgan Kaufmann, 2005, p. 139.

[15] A. Dempster, N. Laird, D. Rubin, "Maximum Likelihood from Incomplete Data via The EM Algorithm," J. of the Royal Statistical Society, Series B, 1977, 39(1): 1-38.

[16] M. Hall, E. Frank, et al, "The WEKA Data Mining Software: An Update," SIGKDD Explorations, 2009, 11(1): 10-18.

[17] S.L. Lohr, "Sampling: Design and Analysis," Second Edition, Duxbury, Pacific grove, CA, 2010, pp. 36-38.

[18] W.J. Paulus, C. Tschope, et al, "How to Diagnose Diastolic Heart Failure: A Consensus Statement on the Diagnosis of Heart Failure with Normal Left Ventricular Ejection Fraction by the Heart Failure and Echocardiography," Eur Heart J, Apr 2007, 28(20): 2539-50.

[19] J. B. MacQueen, "Some Methods for classification and Analysis of Multivariate Observations," Berkeley Symp on Math Statist and Prob, Berkeley, University of California Press, 1967, 1:281-297