PRACTICAL APPLICATION

Segmenting Patients and Physicians Using Preferences from Discrete Choice Experiments

Ken Deal

Abstract People often form groups or segments that have similar interests and needs and seek similar benefits from health providers. Health organizations need to understand whether the same health treatments, prevention programs, services, and products should be applied to everyone in the relevant population or whether different treatments need to be provided to each of several segments that are relatively homogeneous internally but heterogeneous among segments. Our objective was to explain the purposes, benefits, and methods of segmentation for health organizations, and to illustrate the process of segmenting health populations based on preference coefficients from a discrete choice conjoint experiment (DCE) using an example study of prevention of cyberbullying among university students. We followed a two-level procedure for investigating segmentation incorporating several methods for forming segments in Level 1 using DCE preference coefficients and testing their quality, reproducibility, and usability by health decision makers. Covariates (demographic, behavioral, lifestyle, and health state variables) were included in Level 2 to further evaluate quality and to support the scoring of large databases and developing typing tools for assigning those in the relevant population, but not in the sample, to the segments. Several segmentation solution candidates were found during the Level 1 analysis, and the relationship of the preference coefficients to the segments was investigated using predictive methods. Those segmentations were tested for their quality and reproducibility and three were found to be very close in quality. While one seemed better than others in the Level 1 analysis, another was very similar in quality and proved ultimately better in predicting segment membership using covariates in Level 2. The two segments in the final solution were profiled for attributes that would support the development and acceptance of cyberbullying prevention programs among university students. Those segments were very different— where one wanted substantial penalties against cyberbullies and were willing to devote time to a prevention program, while the other felt no need to be involved in prevention and wanted only minor penalties. Segmentation recognizes key differences in why patients and physicians prefer different health programs and treatments. A viable segmentation solution may lead to adapting prevention programs and treatments for each targeted segment and/or to educating and communicating to better inform those in each segment of the program/treatment benefits. Segment members' revealed preferences showing behavioral changes provide the ultimate basis for evaluating the segmentation benefits to the health organization.

K. Deal (✉)
McMaster University, 1280 Main St. West, Hamilton,
ON L8S 4M4, Canada
e-mail: deal@mcmaster.ca

Key Points for Decision Makers

- Segmenting patient and physician populations can lead to better understanding of the diversity of needs and desires among segments, and to providing better treatments, prevention programs, health services, and products.
- Segmentation leads to differentiated offerings encouraging greater acceptance and utilization of prevention programs and health treatments.
- Cyberbullying prevention program development should be sensitive to differences in preferences among segments of university students.

1 Introduction

Physicians care for each individual patient and 'all medicine, by definition, should be patient centered' [1]. However, health treatments, care procedures, prevention programs, and pharmaceuticals are developed for groups of patients having needs and desires thought to be similar.

Survey-based methods for investigating patient and physician preferences for treatments look for statistical similarities and differences. Translating patient-centered medicine into subject-by-subject investigation of survey data would introduce a level of detail that would not support usable understanding of health problems and possible solutions nor inferences to relevant populations. At the other extreme, aggregate analysis of survey samples provides overall insights of average needs but may miss key differences among subjects that can lead to better health care. Statistical methods exist to identify differences among subjects and to characterize subjects as being members of groups having meaningful similarities. Consequently, accounting for heterogeneity above the subject level is necessary and productive in survey research of health problems, prevention programs, and treatments. That need is satisfied by segmentation.

The investigation of heterogeneity to understand differences among attitudes of humans and their likely behavior is typically attributed to marketing [2–4]. The American Marketing Association provides a comprehensive broad definition of marketing—"Marketing is the activity, set of institutions, and processes for creating, communicating, delivering, and exchanging offerings that have value for customers, clients, partners, and society at large" [5]. That definition is as applicable to health care as it is to any other sector in society. We will use the breadth of that definition to include the identification and development of health treatments, communication of the characteristics and benefits of those treatments, distribution of treatments to those in need of care, and pricing of treatments, including pharmaceuticals. While marketers refer to customers or consumers, health marketers refer to populations, patients, physicians, or target audiences [6].

One of the most powerful statements in marketing is "If you're not thinking segments, you're not thinking" [7], signifying the critical importance of understanding differences among groups of people and similarities within each of those groups. The concept of segmenting markets has existed since 1956 when W. Smith wrote in the *Journal of Marketing* "Market segmentation, on the other hand, consists of viewing a heterogeneous market (one characterized by divergent demand) as a number of smaller homogeneous markets in response to differing product preferences among important market segments" [8].

Segmentation has evolved dramatically since Smith coined the phrase. However, the essential strategy focuses on improving organizational performance through recognizing that the population is comprised of segments of individuals with different needs, identifying those segments, and differentially providing those services and products that enhance the lives of those in each of the targeted segments as well as improving the well-being of the organization. Segmentation is far more than methods for analyzing data; it forms the essence of more personalized service and improved organizational performance based on user evidence.

Segmentation strategy includes three key activities: segmenting, targeting, and positioning. Segmenting involves methods of identifying those characteristics that most effectively divide the total population into smaller groups, i.e. segments, each of which is relatively homogeneous internally but heterogeneous from other groups. Targeting is the function of evaluating each of the segments to determine whether it should be included in the strategic plans of the organization and whether organizational actions may enhance the lives of those in the segment. Those that are strategically aligned with the organization's strategy will likely be 'targeted' with others not being targeted.

After segmenting and targeting the audience, the organization needs to position its services and products through better product development, marketing, and communication so that those in the appropriate targeted segments are served with treatments and prevention programs that are better suited to their needs and are seen by them as being best. That combination of the three stages of segmenting, targeting, and positioning is often abbreviated as STP.

Following a segmentation strategy by a health organization involves identifying segments that not only can be measured but also make sense to health decision makers, are large enough to reward targeted efforts, and have distinct communication linkages so that the benefits of the differentiated treatments and prevention programs can be offered to those in need. While the costs of following segmentation strategies are often higher, the rewards to patients/physicians and organizations should be high enough to more than offset the costs.

Segmentation has been conducted within health disciplines to identify physician segments for pharmaceutical marketing [9, 10], patients for understanding how to better develop and apply hospital care [11], redesigning an undergraduate medical curriculum [12], students for developing bullying prevention programs [13], for categorizing joint problems [14], heart veins [15], and carcinoid heart disease [16]. In addition, of course, the area of bioinformatics has significantly increased its research into the segmentation of genes [17].

2 Objectives

The objective of this research was to translate knowledge and experience of segmentation in marketing to segmentation of choice information obtained from discrete choice conjoint experiments (DCEs) in health, and to illustrate the tasks that advance segmentation's analytical and implementation processes. This is a practical application article and will focus on methods that can be used effectively on commercial as well as academic studies without becoming immersed in theoretical detail.

To illustrate the segmentation procedure, we selected a data set that was the result of DCE research into ways for developing prevention programs to reduce the incidence of cyberbullying, a significant and growing threat to the physical and mental health of young people (Cunningham et al., personal communication, 2013). We segmented the choice data produced by the survey to identify the preferred cyberbullying reduction programs for each segment.

3 Segmentation Methods

For segmentation to be valuable to the objectives of a health organization, whether a hospital, clinic, or pharmaceutical firm, it must provide information that enhances key decisions and the effectiveness of that organization. Those decision areas typically pertain to how the well-being of different segments of people can be enhanced by understanding differences and developing and providing health treatments, services, and products that are better suited to the segments individually than were treatments previously presented to the undiversified population.

There are three key tasks in segmentation: (1) finding the most useful number of segments; (2) identifying segments; and (3) ensuring that the segmentation supports the objectives of the research team and the health organization. Answers to point 3 are based on whether those segments make sense to the decision makers, whether better organizational decisions can be made because of the segmentation, and if the rewards from segmented relationships with patients and physicians are greater than for an undifferentiated approach of treating the population as homogeneous. The rewards from segmentation need to be commensurately larger than the additional costs of the segmentation analysis plus the increased costs of developing, executing, and maintaining effective health segmentation strategies, which can be substantial. Segmentation strategies should result in greater effectiveness and utilization since the services, treatments, and programs will be developed and targeted to specific needs of each segment.

Segmentation methods depend on identifying response variables, e.g. treatment-specific attitudes that have the potential to group respondents into cohesive segments. Those variables should have firm conceptual, theoretical, and practical reasons for being included in the segmentation analysis—this should not be a fishing trip [18–20].

Segmentation is different from sample stratification where manifest, i.e. observable, variables are used to separate the sample into groups. Some individuals naturally expect that patients' needs and actions will be different based on sex, income, education, and other demographic dimensions. And, of course, they are different. However, are those differences important for understanding how to provide better treatments for those people? Maybe, but usually not. In most cases, people do not subscribe to treatments or buy products and services based on their demographics. *"If there is a 'most useful' segmentation variable, it would be benefits sought from a product, because the selection of benefits can determine a total business strategy"* [21].

3.1 The Segmentation Process

Most segmentation literature and methods are based on understanding differences in attitudes and, in this article, those attitudes will be choices or preferences obtained from the DCE. After the preference-based segments are formed, other methods strive to explain people who occupy those segments through their demographics, socioeconomics, and mainly manifest, or observable, dimensions, but may also include attitudes not included in the DCE.

While some use a one-step segmentation procedure that includes both response variables and covariates simultaneously, or does not include covariates at all, we follow a two-level process in this research where response variables are used in Level 1 to partition the sample and assign respondents to the segments and then covariates are used to better explain the members of the segments in Level 2 (Table 1). While Level 1 segmentation provides the basis for forming differentially beneficial treatments for each segment, Level 2 segmentation describes the path for supplying those treatments. Vermunt [22] described a three-step process for latent class analysis (LCA) that is somewhat similar to this approach.

Groups are formed in Level 1 segmentation to express the latent structure of the sample using DCE choices or estimated preference coefficients. The segment assignment variable is a latent variable, i.e. a variable that is not the result of directly-asked questions but is formed through the analysis of the DCE preference coefficients. No 'segmentation question' is asked in the survey.

Segmentation using clustering/partitioning methods is often called unsupervised learning, which refers to attempts to determine latent structures in data where explicit segment identification is absent [23, 24]. As explained by

Table 1 A two-level segmentation process

Level 1	Step 1	Discover the latent structure of the sample through segmentation
	Step 2	Develop an algorithm to predict segment membership based on preference (choice) information
Level 2		Develop an algorithm to predict segment membership based on the covariates (scoring the database and/or typing tool)

Dolnicar and Lazarevski [18] and Dolnicar and Leisch [23], health decision makers often believe that there is a natural segmentation structure among patients or physicians, as shown in pane (a) of Fig. 1. Those tidy circular clusters are rather rare in segmentation studies of real people completing surveys. In most cases, pane (b) in Fig. 1 represents the more likely reality of sample-based segmentation. Visually, it appears that there is one natural segment in the stretched oval in pane (b), and from a segmentation perspective it could be called unstructured. However, the data might contain a pseudo cluster structure [23] that may support two or more latent segments that can be reproduced well by segmentation methods, thereby leading to constructed segments. A pseudo cluster structure is one where visually differentiated groups do not appear when viewing the data, as in panel (b) of Fig. 1. Research methods may be able to construct stable and reproducible segments from data having pseudo cluster structure that have good internal homogeneity and good external heterogeneity and provide valuable insights for health analysts and decision makers.

After identifying segmentation solution candidates in step 1 of Level 1, the analysis moves into the realm of supervised learning (step 2, Level 1) where the goal is developing and testing one or several algorithms for accurately predicting segment membership based on the preference coefficients. The analysis is called supervised because of the need to conform to the segmentation structure discovered in step 1 of Level 1. The purpose is to estimate how well the algorithm would likely perform on a new group of patients from the same population but not from the sample used to build the algorithm. This sample-splitting practice becomes problematic when sample sizes are small.

The segmentation process often cycles between the challenges of determining the correct number of segments and the best grouping of individuals into segments. That cycling occurs because of the need to find a segmentation that is highly predictable by the preference coefficients and then by predicting using only a set of variables not included in the segmentation, the covariates.

There are three natural endpoints for a segmentation: (1) gaining a better understanding of the population through the segments based on preference coefficients only; (2) scoring a data set to identify segment membership by using predictive covariates that are in the population data base and in the sample, but were not used to form the segments; and (3) developing a typing or prototyping tool [4] using very few questions built on covariates to provide fast assignment of patients/physicians who were not in the sample into the segments. Endpoints 2 and 3 rely on the development of an algorithm that uses the covariates to predict segment membership.

Scoring and typing tools have the same end purpose—to assign people not in the sample to one of the segments identified based on the sample analysis and often using variables that were not included in the segmentation analysis. Scoring a large database of patients into each segment is often referred to as data mining. Typing tools are used frequently by pharmaceutical firms to quickly identify the best segment to type a patient into in order to more effectively communicate about a drug to best meet the

Fig. 1 Clusters of different types

Pane a)   Pane b)
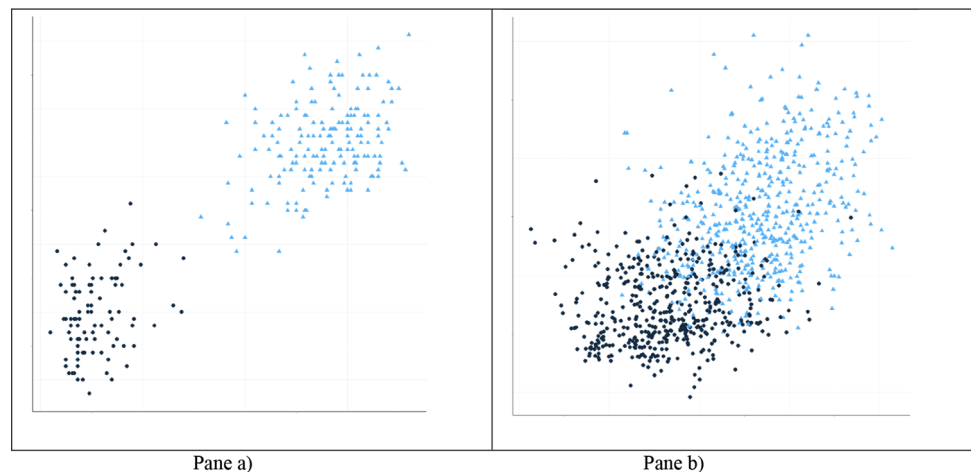
needs of patients in each of the targeted segments [4]. For example, a pharmaceutical firm identified several target segments based on biologic predictor variables and had the resulting typing tool programmed into its website for psoriasis so that they can ask visiting patients five communication preferences and demographic questions, for example, assign them to the appropriate segments, and determine the best way to identify and communicate the drug characteristics best suited to those in each targeted segment. The key, of course, is to keep the tool succinct and to ensure that it is fairly accurate.

After the predictive algorithms have been developed, they are translated into a mark-up language such as Predictive Model Markup Language (PMML) to score the data bases [25]. For smaller data bases and situations where speed is not required, those algorithms can be developed in statistical package syntax such as SPSS.

3.2 Sample Size Effects

The size of the sample directly affects the number of segments that may be extracted. It is often recommended that surveys should be designed to draw samples that will allow for 200 respondents per segment [26]. Large samples can be attained when the respondents are patients or individuals from the general population. However, it is often either physically impossible to attain those size samples among some medical specialists or prohibitively expensive due to substantial incentives needed to engage physicians. Smaller sample sizes present significant challenges for identifying segments that reflect actual differences among groups in the parent population.

3.3 Finding the Most Useful Number of Segments

Ideally, one or more Level 1 segmentations are found using some unsupervised learning method in step 1 as described below. Then, in step 2 of Level 1, the segment identifiers for each segmentation solution and preference coefficients are taken into a supervised learning method, such as random forests (RF) [27], where a predictive algorithm is developed.

As in most statistical analyses, it is wise to split the sample into training and holdout, or testing, subsamples prior to beginning the analysis. In this way, the algorithm that is developed on the training sample and predicts segment membership can be tested in an unbiased manner on the holdout sample. RF was developed based on the principle of ensemble analysis by recognizing that one decision tree may not contain the best solution so hundreds of trees (the default is often 500) are grown to increase the chance of finding better solutions. The RF method provides for internal unbiased testing by splitting the training sample

into a so-called bagged sample that is used for investigating each of the many trees, and an out-of-bag (OOB) sample that is used for testing the quality of the solution that is based on the bagged sample.

Random samples of variables are selected for splitting each node into branches. Thus, two levels of randomness are used in the RF method—one for respondents and the other for variables. An OOB prediction error rate is calculated for the bagged sample estimate and a prediction error rate is determined when using that algorithm on the holdout sample.

It sometimes happens that a good candidate segmentation at Level 1 is not very predictable by the covariates and/or might not make sense to the decision maker. That necessitates investigating the second-best Level 1 segmentation for Level 2 quality. This process may cycle through several Level 1 segmentation solutions before finding one that provides the best combination of Level 1 and Level 2 quality measures and makes sense to the decision maker. This possible cycling is shown in the upper part of Fig. 2. "Another important criterion is the interpretability of the cluster structure, that is, its face validity. To be managerially relevant, the number of clusters must be small enough to allow strategy development. At the same time, each segment must be large enough to warrant strategic attention" [28]. In health organizations, the decision makers need two or more segments that are each rich in their idiosyncratic characteristics but not so numerous that segment differences are not distinct enough to support truly differentiated offerings of health treatments, prevention programs, services, and products.

3.4 Measures of the Quality of Segmentation Solutions

The quality of segments depends on the similarity or internal compactness of the members of each segment and the distance separating each segment from the others; and directly related is the ability to predict segment membership based on the preference coefficients and also on covariates. Statistical quality should relate to meaningfulness of the segmentation to the decision maker and that requires thorough understanding of the objectives of the health organization by the research team.

As mentioned earlier, most survey-based segmentations do not look like pane (a) of Fig. 1 but more like panel (b) where there do not appear to be visual barriers between the segments. Most segmentation quality measures relate to the distinctiveness of segments and the ability to predict segment membership based on the response variables and the covariates. While there are many indices of quality, the following represent those that are used most often. These are called internal indices because they pertain to the particular segmentation itself and not to relationships among segmentation solutions.

Fig. 2 Sequence of analysis for segmentation. *CCEA* Convergent Cluster & Ensemble Analysis, *CHAID* Chi Square Automatic Interaction Detection, *CART* Classification and Regression Trees
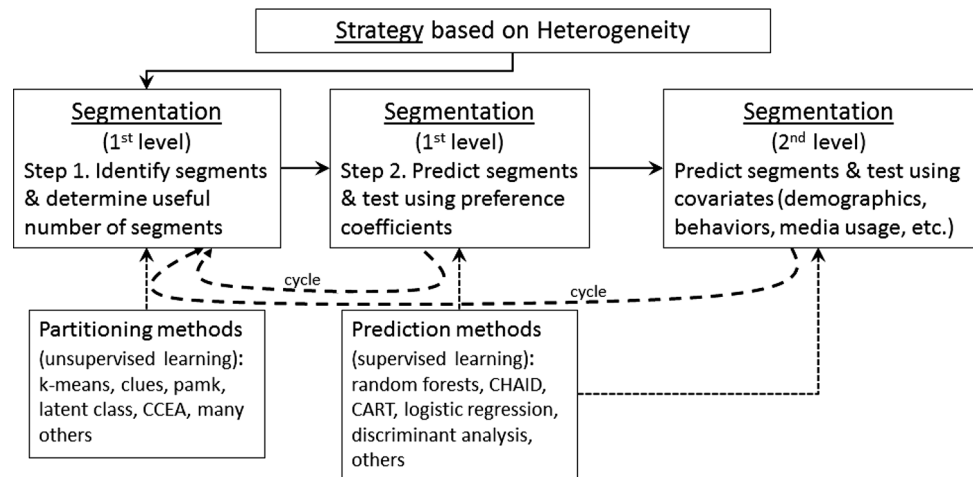
a) The silhouette index [29] assesses the quality of segments by representing the average dissimilarity between each data value in a segment to all of the data values in the closest neighboring segment. Larger positive silhouette values (the range is −1 to +1) indicate that a data value does not belong to a neighboring segment and is properly assigned; an index is produced for each person, each segment, as well as an overall average index. The graphical shapes in the silhouette plot provide insights into those cases that seem to belong in the segment to which they were assigned more than they seem to belong to the closest neighboring segment.

b) The Calinski and Harabasz (CH) fit index is an unbounded and positive measure of the ratio of between-segment variation to within-segment variation [30]. The largest CH index should indicate the optimal number of segments.

c) The Bayesian Information Criterion (BIC) [31] is used for selecting among models when the method involves log-likelihood calculations in probability methods such as LCA. Also, the Akaike Information Criterion [32, 33] and Consistent Akaike Information Criterion [34], as well as log-likelihood, entropy, and average weight of evidence [35], are used with LCA and other methods that contain the information needed by those methods. The minimum BIC is supposed to indicate the best segmentation solution.

Logic strives for consistency and scientific investigations require reproducibility of findings. External indices measure the similarity between different segmentation solutions.

The adjusted Rand index (ARI) [36], Hubert and Arabie's adjusted for chance Rand index [37], Morey and Agresti's ARI [38], Fowlkes and Mallows index [39], and Jaccard index [40] are used to compare the similarity of pairs of segments whether from the same method or different methods [41].

If the true segment membership were known, ARI would provide a measure of the quality of the segmentation by comparing each derived segment to the true segment. Perfect matching of two segments produces an ARI of 1. When the true segments are not known, the ARI provides a measurement of the reproducibility or reliability of the estimation process, which depends on the random nature of the sample and the stochastic nature of the algorithms [23]. When similarity is obtained, it may be assumed that natural clusters exist in the data. Internal and external indices are used in step 1 of Level 1. Since prediction methods are used in step 2 of Level 1 and in Level 2, measures of prediction accuracy are needed. Hit ratios, or confusion matrices, indicate the likely accuracy of predicting segment membership when using either the DCE information (Level 1) or the covariates (Level 2) as predictors. While hit ratios can be calculated on the training sample, it is best to test the algorithm on a holdout sample and calculate a corresponding hit ratio to provide a better idea of the unbiased usefulness of the segment assignment algorithm on patients and physicians not in the sample.

Graphical presentation of the accuracy of prediction is provided by the Receiver Operating Characteristic (ROC) curve [17, 42] where the vertical axis shows the increase in accuracy over not having a predictive model. It plots the percentage of true positives out of the positives (y axis) versus the percentage of false positives (true negatives but classed as positives) out of the negatives (x axis) [43].

3.5 Partitioning Methods

Some procedures attempt to identify the best number of segments and the best assignment of individuals to segments in one step while others must be manually directed to produce a specific number of segments. No one method

has been identified as best in all situations. Some methods are probabilistic models, such as LCA [44, 45], and others use distance-minimizing mathematical procedures, such as k-means clustering.

The DCE data can be used for segmentation analysis in its original choice format in Sawtooth Software's Latent Class [46, 47] and Statistical Innovation's Latent GOLD (LG) [48]. Other methods require respondent-level preference coefficients derived from the choice data, usually by hierarchical Bayes (HB) [49, 50], mixed logit methods [51, 52], or from latent class.

Segmentation methods range from those having no natural statistical tests (k-means) to others that have many measures such as LCA. K-means cluster analysis is the oldest and best-known of the frequently used mathematical methods for segmenting samples [53]. It generates a starting solution, calculates distances between individuals and the segment centers, and then iterates towards better solutions until the difference between two candidate solutions change by less than some minimum. While accessible, fast, and commonly used, the k-means class of methods do not have known statistical properties that support formal inference [54]. Many alterations have been made to the basic k-means algorithm to infuse it with greater power.

We use the following two methods to illustrate useful and easily applied, but highly evolved, procedures based on the logic of k-means, and then discuss other methods that have shown great promise of enriching the tool box of segmentation methods. While many other methods are available, we have found the following procedures to be particularly helpful. Fig. 2 provides a flow diagram of the analytical process for segmentation.

clues (CLUstEring based on local Shrinking) [55, 56] is a non-parametric algorithm that uses a shrinking process determined by the k-nearest-neighbor approach and searches for the number of segments that maximize either the CH fit index or the silhouette index. The algorithm estimates the best number of segments and the corresponding partition.

pamk (Partitioning Around Medoids) [57] estimates the number of segments by optimizing the average silhouette width [29] or CH index [30] based on a data matrix or dissimilarity matrix. pamk is a k-means type of method that uses distances to medoids, i.e. representative cases, rather than means. Both clues and pamk are packages in the R statistical environment [58].

Probability model methods for segmentation are largely represented by LCA (finite mixture models) [45–47]. LCA, as applied in LG, uses either the choice data directly or estimated preference coefficients. LG can treat the covariates as active during the Level 1 segmentation or inactive in Level 1 and active during Level 2 by continuing the analysis using SI-CHAID (Statistical Innovations' Chi Square Automatic Interaction Detection) [59], CHAID, RF, CART (Classification and Regression Trees), logistic regression, discriminant analysis and others. Latent Class is an application from Sawtooth Software [46] that can be very easily used on the choice data from the DCE when using Sawtooth Software for DCE design, fielding, and analysis.

Almost any segmentation data set can support a large number of segmentation solutions, and finding the best of those segmentations is challenging. Cluster ensemble analysis (CEA) [60, 61] recognizes the fact that rarely is there only one best segmentation solution and typically there are many viable candidates. CEA uses a large number of segmentation contenders as the raw material on which it develops potentially still better solutions.

The end result of a segmentation analysis is a two-column matrix where the first column contains the respondent identification numbers and the second column contains integers that assign each person to one of the segments. The input to CEA is many columns of integers, each of which represents one segmentation solution, i.e. its particular assignment of respondents to segments. For example, column 1 would be the respondent identification numbers and column 2 would be a set of integers from 1 to 2 that assign respondents to one of the two segments. Column 3 would be a set of integers that assign respondents to one of the two segments from another segmentation solution and column 4 would be a set of integers from 1 to 3 that assign respondents to one of the three segments from yet another segmentation, and so on. The input matrix to CEA is often comprised of scores of segmentation solutions, each of which is represented by just one column of integers.

One CEA software package, Convergent Cluster & Ensemble Analysis (CCEA) by Sawtooth Software [62], uses an internally-generated ensemble of multiple segments using three methods, and that ensemble can be enhanced further by merging solutions produced by other applications, such as those mentioned earlier. Starting solutions are acknowledged [62] to be important to the resulting segment identification, and CCEA includes sophisticated methods for beginning the solution process. The end result is a consensus segmentation supposed to be better than all input segmentations for each number of segments requested. A consensus segmentation solution is produced for each of the requested number of segments. Those consensus solutions are then analyzed further using other methods before deciding on the best overall segmentation.

4 An Example of Applying Segmentation to an increasing Health Threat: Cyberbullying

Bullying is a significant threat to both the physical and mental health of young people in many countries [63–65]. While it is not a new phenomenon, discussions of its

incidence and social and health impacts have intensified in recent years. Widespread use of the Internet by young people has facilitated the spread of cyberbullying.

Cunningham et al. designed an adaptive choice-based conjoint (ACBC) study to investigate the preferences and aversion that university students may have towards a set of six attributes of cyberbullying prevention programs (Cunningham et al., personal communication, 2013), including the content of anti-cyberbullying advertising, the possible sources of that advertising, reporting policies, reporting processes, consequences for cyberbullying, and the amount of time students would contribute to the program. Each attribute was described by four levels that were exhaustive of relevant possibilities within attributes and mutually exclusive across attributes (Fig. 3).

Using the data from the cyberbullying study of Cunningham et al., personal communication, 2013, we directed our investigation to determine whether all students have the same preferences for cyberbullying prevention program features or if there are important differences among students that might cause some to prefer a program while others might react against the program. While that Cunningham study did not segment the sample, it investigated differences among cyberbullying roles.

5 Methods: The Discrete Choice Conjoint Experiment

This DCE was designed as an ACBC experiment using Sawtooth Software SSI Web v7 [66]. There were six

Fig. 3 Cyberbullying attributes and attribute levels (all categorical)

Reporting Policy	
A1L1	None
A1L2	Encouraged to report cyberbullying
A1L3	Required to report cyberbullying
A1L4	Required to report cyberbullying. Consequences for not reporting.
Anti-bullying Advertising Features	
A2L1	None
A2L2	University President
A2L3	Popular students
A2L4	Famous actors, rock stars, & athletes
Your Commitment to Prevention	
A3L1	None
A3L2	2 hours per week
A3L3	4 hours per week
A3L4	6 hours per week
Advertising Focuses on	
A4L1	None
A4L2	The definition of cyberbullying
A4L3	The effects of cyberbullying on victims
A4L4	Ways students can prevent cyberbullying
Cyberbullying Reporting Process	
A5L1	None
A5L2	Complainants file signed written report
A5L3	Anonymous Crime Stoppers tip line
A5L4	Anonymous on-line reporting
Consequences for Cyberbullies	
A6L1	None
A6L2	Loss of university internet access
A6L3	Suspension for one semester
A6L4	Reported to police

attributes, each with four levels (Fig. 3). The DCE with 20 choice sets and three alternatives per choice set comprised the choice tournament, which is the last stage of the ACBC process and the stage that is most similar to more traditional DCEs. The online survey was conducted among 1,004 undergraduate students at a southern Ontario university who received course credit for their participation. The McMaster University Research Ethics Board approved the survey design and implementation plan. The research team closely followed the recommendations of the ISPOR Good Research Practices for Conjoint Analysis Task Force [67].

5.1 Results: Level 1 Segmentation, Finding Segmentation Candidates

The DCE was designed to discern the preferences for attributes that could comprise cyberbullying prevention programs. Students' preferences for cyberbullying programs were investigated for the types defined by victim, victim/bully, and bully (Cunningham et al., personal communication, 2013).

Respondent-level preference coefficients, or part-worth utilities (PWU), were obtained from HB[1] analysis and used in several segmentation methods.

LG, pamk, and clues were used to find good segmentation candidates for the Level 1 segmentation. CEA was conducted using CCEA to help improve on those solutions, as well as including many other segmentations that were generated internally.

LG was run to investigate from two through nine segment solutions. The lowest BIC (225,819.993) indicated that eight segments was the best solution. The CAIC value (226,199.3) was also lowest for the eight-segment solution. However, the LG model-based classification error was lower for the two-segment solution, while that solution had a higher BIC of 229,159.592. As commonly happens, not all of the statistical indicators lined up to support the eight-segment solution.

___

[1] HB is a two-level procedure that shares information among respondents at the higher level using a multivariate normal distribution while estimating at the lower level using multinomial logit regression. While coefficients at the individual level could have been obtained from LCA, HB assumes a continuous distribution of heterogeneity which appears to be closer to this data than the discrete assumption of LCA [75]. Those individual estimates can be obtained from LCA by multiplying the PWUs for segments by respondents' posterior probabilities of membership in each segment. The coefficients from LCA may be less accurate than HB at estimating respondent preferences [47]. Also, HB may be somewhat more effective in alleviating the independence from irrelevant alternatives (IIA) problems [76]. For sparse data sets, HB seems to capture more of the heterogeneity, while LCA may produce slightly less biased estimates [77].

Both clues and pamk strive to identify the optimal number of segments and both are packages in the R statistical environment. clues and pamk each determined that it was optimal to divide the sample of 1,004 respondents into two segments, although each method suggested somewhat different assignments of respondents to segments. Since clues and pamk do not directly produce likelihood measures, BIC was not used. The average silhouette width and the CH index were highest for the two segment solutions from each method. Two different two-segment solutions were obtained from each method (see Table 2).

CCEA version 3.0.12 was used to estimate a consensus segmentation based on an ensemble of 70 groupings determined by k-means and hierarchical methods within CCEA plus the seven segmentation solutions produced external to CCEA by LG, clues, and pamk. CCEA was executed to determine if still better segmentation solutions might exist.

Since there is no universally accepted indicator of overall superiority of one segmentation compared to another, we used the average silhouette width and the CH index in step 1 of Level 1. Similarity between segment solutions was measured by Hubert and Arabie's ARI (see Table 3).

The results presented in Table 2 reflect comparisons among the consensus segmentation solutions for two and three segments from CCEA, and the optimal two-segment solutions produced by clues and pamk and the eight-segment solution and two two-segment solutions generated by LG.

The average silhouette width and the CH index were highest for clues2, 0.211 and 314.32, respectively ($n1 = 500$, $n2 = 504$), and second highest for pamk1, 0.208 and 308.54 ($n1 = 475$, $n2 = 529$), providing two good candidates (Table 2). Segmentation of patients, physicians, and consumers in general depend on finding structure in data that may not have true natural density structure, as shown in Fig. 1, called reproducible clustering/segmentation [23]. Considering reproducibility criteria, the pamk1, clues2, CCEA1 and LG2 solutions have the highest ARIs (Table 3). The highest ARI is between pamk1 and clues2 (0.814), and the second highest is between clues2 and CCEA1 (0.779). In this context, reproducibility of a segment solution is an additional indicator of the fundamental properties of quality [62]. It is obvious that the LG3 eight-segment solution is inferior to all other solutions, as indicated by the very low silhouette index and CH, and that was dropped from further consideration.

A helpful procedure for assessing practical quality is to develop an assignment algorithm using a fraction of the total sample (the training sample) and then test how accurately that algorithm assigns respondents in the

Table 2 Indicators of the quality of alternative segmentation solutions

Segmentation alternatives	Segment sizes			Silhouettes, average widths Calinski–Harabasz Index	Random forest[a]	
	Segment 1	Segment 2	Segment 3		OOB est. of error rate on training sample (70 %)	Error rate on test sample (30 %)
CCEA1 (two segments)	442	562		0.2055293[b] 303.2595[c]	6.41 %	4.97 %
CCEA2 (three segments)	405	330	269	0.1198374 216.8515	8.40 %	8.28 %
clues1 (two segments)	656	348		0.2052618 287.3386	3.28 %	4.64 %
clues2 (two segments)	500	504		0.2111669 314.3168	3.70 %	5.63 %
pamk1 (two segments)	475	529		0.2081598 308.540	6.84 %	4.97 %
pamk2 (two segments)	510	494		0.191653 273.1479	7.12 %	7.95 %
LG1 (two segments)	631	373		0.1844168 264.9917	5.13 %	5.63 %
LG2 (two segments)	568	436		0.1972908 288.4449	5.70 %	3.64 %
LG3, (8 segments)	S1 219	S2 156 S3 143 S4 119 S5 106 S6 105 S7 89 S8 67		0.048472 91.05573	14.96 %	15.56 %

CCEA Convergent Cluster & Ensemble Analysis, *clues* CLUstEring based on local Shrinking, *pamk* Partitioning Around Medoids, *LG* Latent Gold, *OOB* out-of-bag

[a] Rattle was used to execute the random forest method ($n = 1,004$; training $= 702$, testing $= 302$); lower error rates are better

[b] Higher values are better; 0 is minimum, 1 is maximum

[c] Higher values are better; no maximum

Table 3 Measures of similarity between the segment solutions, Pairwise Hubert and Arabie Adjusted Rand Index

HA, ARI	CCEA1 (two segments)	CCEA2 (three segments)	LG1 (two segments)	LG2 (two segments)	LG3 (eight segments)	clues1 (two segments)	clues2 (two segments)	pamk1 (two segments)
CCEA (three segments)	0.604[a]							
LG1 (two segments)	0.616	0.473						
LG2 (two segments)	0.696	0.433						
LG3 (eight segments)	0.170	0.258	0.176					
clues1 (two segments)	0.323	0.559	0.309	0.309	0.150			
clues2 (two segments)	0.779	0.552	0.528	0.647	0.182	0.475		
pamk1 (two segments)	0.677	0.539	0.606	0.606	0.180	0.408	0.814	
pamk2 (two segments)	0.594	0.387	0.450	0.616	0.186	0.401	0.647	0.600

HA ARI Hubert and Arabie Adjusted Rand Index, *CCEA* Convergent Cluster Ensemble Analysis, *LG* Latent Gold, *clues* CLUstEring based on local Shrinking, *pamk* Partitioning Around Medoids

[a] Higher values indicate closer similarity between segmentation solutions

remaining part of the sample (the holdout sample). One very effective method for developing the algorithm and executing the analysis is RF [27], a tree-splitting method that generates and tests the algorithm on hundreds of trees and automatically tests the effectiveness on part of the sample that is not used at each node of the tree, called OOB analysis [25, 27, 68, 69]. RandomForest is an R package [70] and Rattle [25] is an R package wrapped in an easy-to-use Graphical User Interface that facilitates the use of randomForest.

We used the randomForest package (in R and in Rattle) to further evaluate the segmentation in step 2 of Level 1. Each of the segment classifications other than LG3 were taken into Rattle and each segment assignment was used as the criterion variable with the preference coefficients used as predictors. The randomForest method in Rattle was run with a random sample split of 70 % (702) for training and 30 % (302) for the holdout.

The clues1 segmentation solution had the lowest OOB estimated error rate of 3.28 % for the training sample and an error rate of 4.64 % for the holdout sample (Table 2)—the average silhouette width of 0.205 was slightly lower that the estimate for pamk1 and clues2 and higher, i.e. better, than most others. The CH index was high for clues1, although not the highest among these final solutions. LG2 had a somewhat higher OOB error, lower holdout sample error, slightly lower silhouette index, and slightly higher CH index. The pamk1 solution had a very good holdout sample error rate of 4.97 %, second-highest silhouette and second-highest CH index. All these solutions will be investigated further. While the clues1 OOB error rate was better than others (3.28 %), its error rate on the holdout sample was higher than the LG2 solution.

The ROC plot in Fig. 4 for clues2 shows the substantial strength of that solution to provide correct positives versus false positive predictions [43]. The optimal ROC curve would be horizontal at the 1.0 level and have an area under the curve (AUC) of 1.0; the clues2 curve is very close to that ideal, with an AUC of 0.994. The diagonal is what would be expected if no improvement over chance were found.
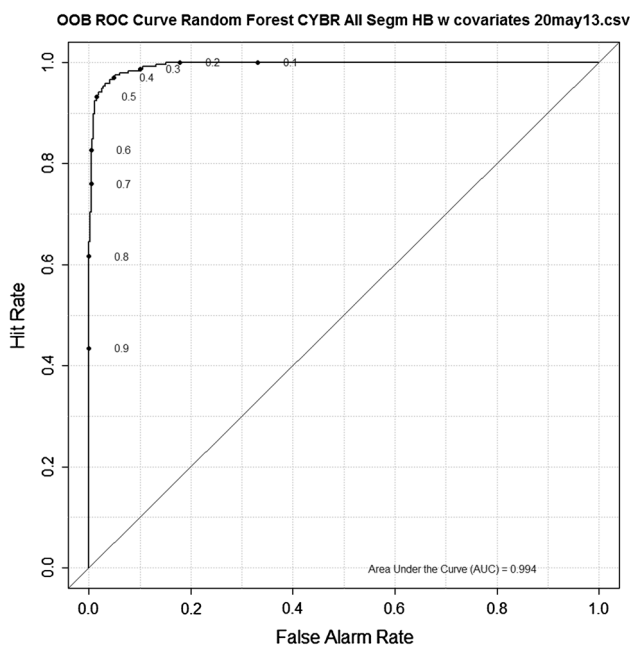

Fig. 4 The Receiver Operating Characteristic curve for the clues2 segmentation

5.2 Results: Level 2 Segmentation, Predicting Segment Membership Using Covariates

Our objective for the Level 2 segmentation was to identify a set of variables (covariates) that was not used in the segmentation, i.e. not from the DCE, that could be used to predict segment membership. An extensive set of questions was asked regarding the frequencies (never, to once per day—5-point scale) that respondents were bullied or bullied others throughout their education from elementary school grade 1 through university, and the frequencies of witnessing, being a victim or being the perpetrator of cyberbullying in university. In addition, scaled agree/disagree attitude questions were asked regarding barriers to preventing cyberbullying, intentions regarding cyberbullying prevention, cyberbullying norms, and benefits of cyberbullying prevention. Questions were asked regarding self-perception, frequencies of engaging in computing activities, computing proficiencies, sex, age, first language, and whether an immigrant.

In practical applications, the range of covariates is typically constrained by the variables in the customer information system or focus on a small number of questions that may be helpful in predicting segment membership and can be easily asked of a physician or patient. There was no such information system for the student respondents in this study. Since, obviously, none of the covariates were used to develop the segmentations, we did not expect to find the predictive ability to be as high as when using the preference coefficients.

Level 2 investigation was executed using RF in Rattle. The segment identifiers were loaded into Rattle along with covariates. Seventy percent of the sample was randomly assigned to the training sample and the remaining 30 % to the testing, or holdout, sample.

The assignment algorithm developed in RF in the training sample ($n = 702$) was then used to predict segment membership within the holdout sample ($n = 302$). The seven segmentation solutions in Table 4 seemed to be strong candidates based on the Level 1 segmentation. The average error of prediction in the holdout sample ranges from 27.48 to 33.44 %. While this improves on the 50 % success rate of guessing segment assignment, without prior information, this level of prediction is often surpassed.

The prediction errors in the holdout samples for segment 1 and for segment 2 can be radically different, while it is preferable that these errors be more equally balanced. The clues1 solution that did so well in the Level 1 segmentation encountered severe problems in Level 2 where the prediction error for segment 1 was an impressive 4.19 % but the prediction error for segment 2 was a very disappointing 68.47 %, indicating that the majority of segment 2 members were predicted to be in segment 1—that solution is not

Table 4 Error rates and balance in error rates among solution contenders

	Level 1 segmentation		Level 2 segmentation					Root-mean-square deviation
	Training	Holdout	Training sample		Holdout sample			
	Error	Error		Error	Segment 1	Segment 2	Error	
pamk1								
Segment 1	6.85[a]	7.19	Segment 1	38.99	89	50	35.97	
Segment 2	6.83	3.07	Segment 2	29.78	46	117	28.22	
Mean	6.84	4.97	Mean	34.19	Mean		31.79	3.89
clues2								
Segment 1	3.38	6.90	Segment 1	33.24	104	41	28.28	
Segment 2	4.03	4.46	Segment 2	37.18	59	98	37.58	
Mean	3.70	5.63	Mean	35.19	Mean		33.11	4.66
pamk2								
Segment 1	7.59	9.22	Segment 1	27.10	107	34	24.11	
Segment 2	6.61	6.83	Segment 2	40.54	67	94	41.61	
Mean	7.12	7.95	Mean	33.48	Mean		33.44	8.77
LG2								
Segment 1	2.75	0.52	Segment 1	25.45	134	45	25.14	
Segment 2	9.06	14.81	Segment 2	45.37	53	70	43.09	
Mean	5.13	5.63	Mean	34.33	Mean		32.45	9.13
CCEA1								
Segment 1	5.40	0.58	Segment 1	28.02	134	38	22.09	
Segment 2	7.67	10.77	Segment 2	47.28	60	70	46.15	
Mean	6.41	4.97	Mean	36.61	Mean		32.45	12.15
LG1								
Segment 1	3.86	1.12	Segment 1	15.33	166	28	14.43	
Segment 2	7.99	7.32	Segment 2	55.85	55	53	50.93	
Mean	5.70	3.64	Mean	30.63	Mean		27.48	18.97
clues1								
Segment 1	1.51	3.14	Segment 1	8.60	183	8	4.19	
Segment 2	6.75	7.21	Segment 2	63.29	76	35	68.47	
Mean	3.28	4.64	Mean	27.07	Mean		27.81	21.54

pamk Partitioning Around Medoids, *clues* CLUstEring based on local Shrinking, *LG* Latent Gold, *CCEA* Convergent Cluster & Ensemble Analysis

[a] Lower error rates are better

satisfactory. A similar situation occurs for the LG1 solution. While the majority of segment 2 members are predicted to belong to segment 2 for the CCEA1 and LG1 solutions, there is still much discrepancy in predictive accuracy between the segments. The pamk1, pamk2, clues2 and LG2 solutions should all be contenders for the best segmentation.

Unless it is important to more accurately report membership in one segment than another, it is desirable for a solution to have prediction errors that are fairly well-balanced over the segments. We measured prediction balance by calculating the root-mean-square deviation (RMSD) of the errors for each segment from the overall prediction error. The RMSDs for the pamk1 solution and the clues2 solution are very close at 3.89 and 4.66, respectively. The other solutions all have one segment prediction error above 40 % with larger RMSDs.

Based on the error for the holdout sample in the Level 2 segmentation and for the Level 1 segmentation, the pamk1 solution appears to have the best record, but clues2 is a very close contender. The pamk1 and the clues2 solutions have the highest ARIs between solutions in Table 3 of 0.814, suggesting reasonable reproducibility between the two solutions. Which solution wins may relate to what might be considered a logical or useful solution by the decision maker in the health organization.

For the pamk1 solution, the RF analysis identified a set of covariates that best predicted segment membership,

including barriers (I don't have time to participate in a cyberbullying prevention program), intent (I would participate in a cyberbullying prevention program), norms (it is important to report cyberbullying), intent (enrol in a university-sponsored cyberbullying education program), norms (you should participate in cyberbullying prevention programs; there should be consequences for cyberbullying; Universities should have cyberbullying prevention programs; everyone should participate in a cyberbullying prevention program), intent (report a student who cyberbullies), that predicted segment membership with an overall error rate in the holdout sample of 31.79 %. The prediction error of membership in segment 1 was 35.97 % and 28.22 % for segment 2.

The graph of the values of the PWUs for each of the two pamk1 segments appears in Fig. 5, with the attributes and attribute levels in Fig. 3. There is relatively large separation between the PWUs for each attribute level except for A2L3, having popular students as spokespeople in communications, A3L2, the commitment of 2 h per week of personal time to cyberbullying prevention, A4L2, the definition of cyberbullying, A4L3, the effects of cyberbullying on victims, and A5L4, anonymous online reporting.

Segment 2 prefers a no-reporting policy (A1L1) substantially more so than does segment 1, and segment 2 has a much lower utility for required reporting with a penalty for not reporting (A1L4). Segment 2 is highly negatively affected by the time contribution (A3L4) to an anti-bullying program, while segment 1 has a higher preference for contributing 2 h per week (A3L2) than no time commitment (A3L1).

Those in segment 2 prefer no reporting process (A5L1) more than do those in segment 2, although both most prefer anonymous online reporting (A5L4). Segment 1 is very

opposed to there being no consequences for cyberbullying (A6L1) while those in segment 2 greatly prefer that there be no consequences or just the loss of university Internet privileges (A6L2). Segment 2 members have substantially lower preferences for the consequences of suspension for one semester (A6L3) and having cyberbullying reported to police (A6L4) than do segment 1 members. The two segments differ most on the consequences of cyberbullying, and that attribute 6 is most important in influencing subjects' choices among bullying prevention programs.

There are significantly higher percentages of bullies and those who stated they were bullies and victims in segment 2. This finding is consistent with the preference tendencies mentioned in the previous paragraph. Segment 2 contains 34.2 % males, which is significantly higher than the 23.3 % males in segment 1.

6 Discussion

Segmentation enables health organizations to develop differential strategies to serve those in each targeted segment better than could be done with general undifferentiated strategies. A segmentation strategy should increase the value of the health organization to its stakeholders, whether patients, physicians, decision makers, benefactors, government, investors, or others. The segmentation process includes (1) the analysis of data, usually from samples, to identify the segments, and (2) the development of the segmentation strategy by the decision makers for the health organization.

The segmentation procedure described in this article moves between two levels and includes two sets of variables: (1) the basis variables on which the segmentation is

Fig. 5 Part-worth utilities for the pamk 1 solution (Please refer to Fig. 3 for attribute descriptions)
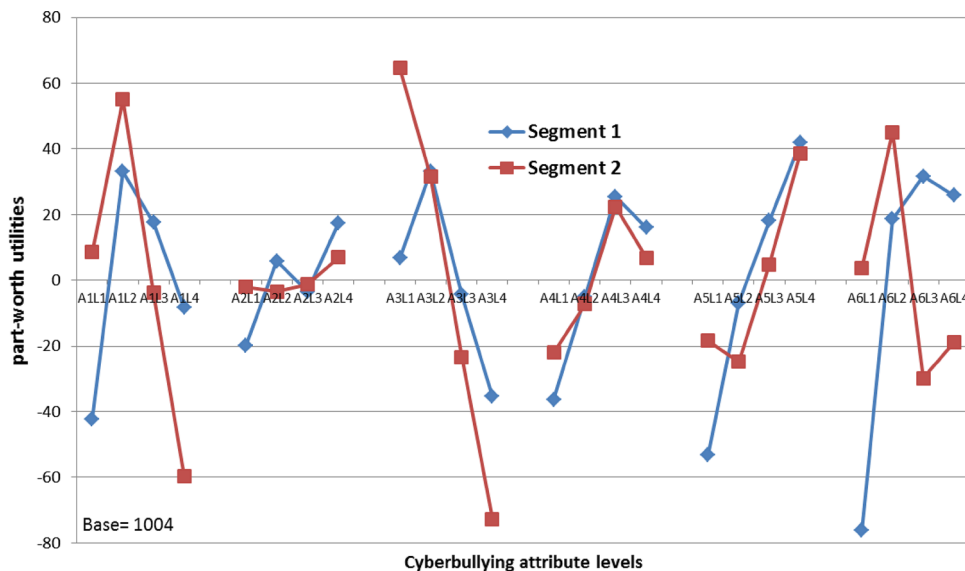
initially defined (Level 1); and (2) the predictor variables (covariates) that help to implement the segmentation strategy among the relevant population (Level 2). The most effective segmentations are built on attitudinal basis variables that help explain why patients react to treatment options, prevention programs, services, and products. The strongest attitudinal variables for segmentation are often preference coefficients from DCEs, but attitudinal variables from batteries of scaled questions can be used. Demographic, health state, behavioral, and lifestyle variables often comprise the set of covariates that are used to predict likely membership in the segments among members of the population outside of the sample.

It is conceptually and statistically challenging to identify the 'best' segmentation and there are often several candidates for that rank. While identifying and understanding heterogeneity in the sample based on the basis variables is stimulating enough on its own; continuing the analysis through to identifying a good usable segmentation at Level 1 that can then be reasonably predicted by the predictor covariates in Level 2 is equally interesting and provides the information needed to fulfil segmentation strategies for health care.

The purpose of Level 1 segmentation is to identify segments that may help the health organization decision makers to better understand their patients or physicians. The methods used in step 1 of Level 1 are often called partitioning or clustering and can be found in health disciplines as well as many other sectors. The term 'unsupervised learning' is used for that type of analysis since the segments can be formed freely and not conform to any predetermined structure.

After identifying viable segmentation candidates, it is important to determine the quality of each. Several internal quality indices were used in this article, including silhouette plots, CH index, and the BIC—and others are available from the literature. Those indices are called internal because they pertain to each individual segmentation solution and not to any others.

External indices attempt to measure the segmentation quality across two or more segmentation solutions. For example, the ARI measures the similarity or reproducibility of two segmentation solutions. Reproducibility is an important trait of segmentation since health strategies based on the segmentation seek stability and dependability throughout the parent population and over a reasonably long time period.

Step 2 of Level 1 and Level 2 focus on the predictive ability of a segmentation solution, and evaluate the quality of the segmentation solution using a measure of predictive quality. This is supervised learning since the process must conform to the segmentation structure found in step 1 of Level 1. Step 2 of Level 1 continues to work with the basis variables, e.g. the DCE preference coefficients. RF, decision trees such as CART and CHAID, discriminant analysis, logistic regression, machine learning algorithms, and other methods are used to determine the ability of those basis variables to predict segment membership and to identify those variables that are most strongly related to each segment. Segment names often evolve from the variables that are most attached to each segment.

That best segmentation from Level 1, i.e. the assignment of subjects to segments, is combined with the demographic and other covariates, and becomes the concentration of Level 2. The most important considerations in Level 2 are (1) the ability of the predictor variables to accurately identify segment memberships identified in Level 1, and (2) the real-world usability of the segmentation. The analytical methods used in Level 2 include those used in step 2 of Level 1, e.g. RF, etc. A key objective in Level 2 is to find the smallest set of predictor variables that assign subjects to segments with the highest accuracy. Attention must be paid to the accuracy of predicting each segment class, not just the overall average predictive ability.

The demographic variables of age, sex, origin, computer usage, computer skills, and self-perception were not very good at predicting segment membership—the error rate was between 44 and 48 %, indicating success was close to chance. The ineffective prediction of segment membership based on core demographic variables is not unusual [71]. The most effective segmentations are built on attitudinal data closely related to the relevant health treatment, prevention program, service, or product [3, 4, 9, 72–74]. Segmentation is not a linear process and may cycle between Level 1 and Level 2. The predictive analysis continues until the segmentation solution with the greatest predictive ability and the most useful structure for health decision makers is determined. As illustrated with the cyberbullying study, what might appear to be the best segmentation in Level 1 may be less than optimal in Level 2. Because the possibility of cycling between Level 1 and Level 2 is so high, it is normally best to produce several good candidates at Level 1 so that they can be investigated more easily if the presumed best from Level 1 fails the Level 2 testing. The natural end result of the segmentation analyses is an algorithm for accurately assigning patients who are in the relevant population but were not part of the sample to the segments.

The cyberbullying study was used to illustrate the segmentation process for health sector studies. Other data sets might be easier or more difficult to analyze. Prediction of segment membership results can vary widely depending on the data and the methods—it is critically important to develop an excellent survey instrument that is designed to be analyzed for segmentation purposes. The analytical methods used on the cyberbullying segmentation in this

article all have excellent reputations. While some seemed to produce better results for this study, none should be considered to be generally superior or inferior to others.

Whether the solution makes sense to the health decision makers and is perceived as being real and not just the result of the statistical method might be the most important quality consideration [19]. Although unscientific, following an organizational strategy of segmentation requires confidence from the health decision makers. Only when they understand the results and appreciate the basics of the segmentation process will the segmentation be accepted and form the basis for better treatment, prevention programs, care, and organizational strategies.

6.1 Limitations

Segmentation may be adversely effected by the randomness of the sample and the randomness of the segmentation methods [23]. Not all segmentation methods were used and others might have found still better solutions. While the sample was large, it was drawn from one institution and attitudes might be more diverse across universities.

7 Conclusions

When health organizations ignore patient and physician heterogeneity, they face the prospect of offering treatments that appeal to the average but might actually fall between two or several very distinctly different foci of preferences. Highly effective measures of heterogeneity exist to help with this important task. We found two segments of university students who support quite different types of programs for lessening cyberbullying. While it may be tempting to some to develop a cyberbullying prevention program that would appeal most to those who want mandatory reporting with penalties for not reporting and more severe controls, such a program may backfire if the segment wanting lesser penalties were to work against it. It may be preferable to institute a prevention program that appeals mainly to the segment that holds higher social values but does not cause a sizable minority to sabotage that program. This may help universities and other institutions to develop programs that appeal to targeted segments and do not trigger actions against those programs from other segments.

References

1. Bensing J. Bridging the gap, the separate worlds of evidence-based medicine and patient-centered medicine. Patient Educ Couns. 2000;39:17–15.
2. Lerer L. Pharmaceutical marketing segmentation in the age of the internet. Int J Med Mark 2. 2002;2(2):159–66.
3. Bassi F. Latent class factor models for market segmentation: an application to pharmaceuticals. Stat Methods Appl. 2007;16:270–87.
4. Vaughn S, Sarianne S. Examining physician segments. Pharm Represent. 2009;39(4):12–5.
5. American Marketing Association. The American Marketing Association releases new definition for marketing. http://www.marketingpower.com/AboutAMA/Documents/American%20Marketing%20Association%20Releases%20New%20Definition%20for%20Marketing.pdf.. Accessed 13 Nov 2013.
6. Andreasen AR. Redesigning the marketing universe. Keynote address, World Marketing Summit, Dhaka, 2 Mar 2012.
7. Levitt T. The marketing imagination. New York: The Free Press; 1983.
8. Smith W. Product differentiation and market segmentation as alternative marketing strategies. J Mark. 1956;21:3–8.
9. Greengrove K. Needs-based segmentation: principles and practice. Int J Mark Res. 2002;44(4):405–21.
10. Ferrandiz J. The impact of generic goods in the pharmaceutical industry. Health Econ. 1999;8(7):599–612.
11. Cunningham C, Deal K, Chen Y. Adaptive choice-based conjoint analysis: a new patient-centered approach to the assessment of health service preferences. Patient. 2010;3(4):257–73.
12. Cunningham C, Deal K, Neville A, Miller H, Lohfeld L. Modeling the problem-based learning preferences of McMaster University undergraduate medical students using a discrete choice conjoint experiment. Adv Health Sci Educ Theory Pract. 2006;3(2):245–66.
13. Cunningham C, Vaillancourt T, Rimas H, Deal K, Cunningham L, Short K, Chen Y. Modeling the bullying prevention program preferences of educators: a discrete choice conjoint experiment. J Abnorm Child Psychol. 2009;37(7):929–43. doi:10.1007/s10802-009-9324-2.
14. Yin Y, Zhang X, Williams R, et al. LOGISMOS—Layered optimal graph image segmentation of multiple objects and surfaces: cartilage segmentation in the knee joint. IEEE Trans Med Imag. 2010;29(12):2023–37.
15. Schaap M, van Walsum T, Neefjes L, et al. Robust shape regression for supervised vessel segmentation and its application to coronary segmentation in CTA. IEEE Trans Med Imag. 2010;30(11):1974–86.
16. Van Gerven MA, Jurgelenaite R, Taal BG, et al. Predicting carcinoid heart disease with noisy-threshold classifier. Artif Intell Med. 2007;40(1):45–55.
17. Giuly RJ, Martone M, Ellisman M. Method: automatic segmentation of mitochondria utilizing patch classification, contour pair classification, and automatically seeded level sets. BMC Bioinform. 2012;13:29.
18. Dolnicar S, Lazarevski K. Methodological reasons for the theory/practice divide in market segmentation. J Mark Manag. 2009;25(3–4):357–73.
19. Aldenderfer M, Blashfield R. Cluster analysis. Newbury Park: Sage Publications; 1984.

20. Everitt BS. Unresolved problems in cluster analysis. Biometrics. 1979;35:169–82.

21. Aaker D. Developing business strategies. 5th ed. New York: Wiley; 1998. p. 47.

22. Vermunt J. Latent class modeling with covariates: two improved three-step approaches. Polit Anal. 2010;18:450–69.

23. Dolnicar S, Leisch F. Evaluation of structure and reproducibility of cluster solutions using the bootstrap. Market Lett. 2010;21:83–101.

24. Retzer J, Shan M. Cluster ensemble analysis and graphical depiction of cluster partitions. Proceedings of the 2007 Sawtooth Software Conference, Sequim (WA); 2007.

25. Williams G. Data mining with Rattle and R: the art of excavating data for knowledge discovery. New York: Springer Science+Business Media; 2011.

26. Orme B. Getting started with conjoint analysis: strategies for product design and pricing research. Madison: Research Publishers LLC: p. 65.

27. Breiman L. Random forests. Mach Learn. 2001;45(5–3):2.

28. Arabie P, Hubert L. Cluster analysis in marketing research. In: Bagozzi R, editor. Advanced methods of marketing research, Cambridge: Blackwell; 1994. p. 160–189.

29. Rousseeuw P. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Comput Appl Math. 1986;20:53–65. doi:10.1016/0377-0427(87)90125-7.

30. Calinski T, Habarasz J. A dendrite method for cluster analysis. Commun Stat. 1974;3:1–17.

31. Schwartz G. Estimating the dimension of a model. Ann Stat. 1978;6:461–4.

32. Akaike, H. Information theory as an extension of the maximum likelihood principle. In: Petrov BN, Csaki F, editors. Second international symposium on information theory. Budapest: Akademiai Kiado; 1973. p. 267–8.

33. Bozdogan H. Model selection and Akaike's information criterion (AIC): the general theory and its analytical extensions. Psychometrika. 1987;52:345–70.

34. Sugiura N. Further analysis of the data by Akaike's information criterion and the finite corrections. Commun Stat Theory Methods. 1978;A7:13–26.

35. Banfield JD, Raftery AE. Model-based gaussian and non-gaussian clustering. Biometrics. 1993;49:803–21.

36. Rand WM. Objective criteria for the evaluation of clustering methods. J Am Stat Assoc. 1971;66:846–50.

37. Hubert L, Arabie P. Comparing partitions. J Class. 1985;2:193–218.

38. Morey LC, Agresti A. An adjustment to the rand statistic for chance agreement. Classif Soc Bull. 1981;5:9–10.

39. Fowlkes EB, Mallows CL. A method for comparing two hierarchical clusterings. J Am Stat Assoc. 1983;78(383):553–69.

40. Hultsch L. Untersuchung zur Besiedlung einer Sprengfläche im Pockautal durch die Tiergruppen Heteroptera (Wanzen) und Auchenorrhyncha (Zikaden).

41. Krieger AM, Green PE. A generalized Rand-index method for consensus clustering of separate partitions of the same data base. J Classif. 1999;16:63–89.

42. Zweig M, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clin Chem. 1993;39(4):561–77.

43. Vuk M, Curk T. ROC curve, lift chart and calibration plot. Metodoloski zvezki. 2006;3(1):89–108.

44. Goodman LA. The analysis of systems of qualitative variables when some of the variables are unobservable. Part I: a modified latent structure approach. Am J Sociol. 1974;79:1179–259.

45. Magidson J, Vermunt JK. Latent class factor and cluster models, bi-plots and related graphical displays. Sociol Methodol. 2001;31:223–64.

46. The CBC Latent Class Technical Paper. Version 3. Sawtooth Software Technical Paper Series, 2004.

47. Latent Class v4.5, Sawtooth Software Inc., 26 Sep 2012.

48. Vermunt JK, Magidson J. Latent Gold Choice 4.0 user's guide. Statistical Innovations; 2005.

49. Allenby G, Arora N, Ginter J. On the heterogeneity of demand. J Market Res. 1998;35:384–9.

50. Rossi P, Allenby G, McCullough R. Bayesian statistics and marketing. New York: Wiley; 2005.

51. Revelt D, Train K. Mixed logit with repeated choices: households' choices of appliance efficiency level. Rev Econ Stat. 1998;30(4):647–57.

52. Johnson FR, Mansfield C. Survey design and analytical strategies for better healthcare stated-choice studies. The Patient. 2008;1(4):299–307.

53. MacQueen JB. Some methods for classification and analysis of multivariate observations. In: Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability. Vol. 1. University of California Press; 1967: p. 281–297.

54. Witek E. Comparison of model-based clustering with heuristic clustering methods. Folia Oeconomica. 2011;255:191–7.

55. Wang X, Qiu W, Zamar RH. CLUES: a non-parametric clustering method based on shrinking. Comput Stat Data Anal. 2007;52(1):286–98.

56. Chang F, Qiu W, Zamar RH, Lazarus R, Wang X. Clues: An R package for nonparametric clustering based on local shrinking. J Stat Softw. 2010;33:4.

57. Kaufman L, Rousseeuw PJ. Finding groups in data. New York: Wiley; 2005.

58. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2013. http://www.R-project.org/.

59. Magidson, J. SI-CHAID 4.0 user's guide. Statistical Innovations; 2005.

60. Retzer J, Shan M. Cluster ensemble analysis and graphical depiction of cluster partitions. In: Proceedings of the 2007 Sawtooth Software Conference, Sequim (WA); 2007.

61. Strehl A, Ghosh J. Cluster ensembles: a knowledge reuse framework for combining multiple partitions. J Mach Learn Res. 2002;3:583–617.

62. Orme B, Johnson R. Improving K-means cluster analysis: ensemble analysis instead of highest reproducibility replicates. Sawtooth Software Research Paper Series; 2008.

63. Arseneault L, Walsh E, Trzesniewski K, Newcombe R, Caspi A, Moffitt TE. Bullying victimization uniquely contributes to adjustment problems in young children: a nationally representative cohort study. Pediatrics. 2006;118(1):130–8. doi:10.1542/peds.2005-2388.

64. Arseneault L, Bowes L, Shakoor S. Bullying victimization in youths and mental health problems: 'Much ado about nothing'? Psychol Med. 2010;40:717–29.

65. Kim YS, Leventhal BL, Koh YJ, Hubbard A, Boyce WT. School bullying and youth violence: causes or consequences of psychopathologic behavior? Arch Gen Psychiatry. 2006;63(9):1035–41. doi:10.1001/archpsyc.63.9.1035.

66. Sawtooth Software. http://www.sawtoothsoftware.com/version/ssiweb/ssiweb_history.html.. Accessed 13 Nov 2013.

67. Bridges JFP, Hauber AB, Marshall D, et al. Conjoint analysis applications in health—a checklist: a report of the ISPOR good research practices for conjoint analysis task force. Value Health. 2011;14:403–13.

68. Chen C, Liaw A, Breiman L. Using random forest to learn imbalanced data. UC Berkeley: Department of Statistics; 2004.

69. Svetnik V, Liaw A, Tong C, et al. Random forest: a classification and regression tool for compound classification and QSAR modeling. J Chem Inf Comput Sci. 2003;43:1947–58.

70. Liaw A, Wiener M. Classification and regression by randomForest. R News. 2002;2(3):18–22.

71. Haley R. Benefit segmentation: a decision-oriented research tool. J Mark. 1968;32(3):30–5.

72. Zapert K, Spears D. Reengineering a US-based diabetes patient segmentation for Japan: lost in translation. Presented at 2011 Annual National Conference of the Pharmaceutical Marketing Research group; 2011.

73. Bogle A, Simpson SL, Mills TM. Segmentations that work. First Annual Meeting of the Pharmaceutical Marketing Research Group; 2007.

74. Ross C, Steward CA, Sinacore JM. The importance of patient preferences in the measurement of health care satisfaction. Med Care. 1993;31(12):1138–49.

75. Magidson J, Eagle T, Vermunt JK. New developments in latent class choice models. In: Sawtooth Software Conference Proceedings; 2003: p. 89–112.

76. The CBC Latent Class Technical Paper. Version 3. Sawtooth Software Technical Paper Series; 2004.

77. McCullough PR. Comparing hierarchical Bayes and latent class choice: practical issues for sparse data sets. In: 2009 Sawtooth Software Conference Proceedings, Delray Beach (FL); Mar 2009.