371

Miller J[1], Guo Y[1], Rodseth-Becker S[1], DeLancey J[1], Fenner D[1]
1. University of Michigan

DIARY DATA SUBJECTED TO CLUSTER ANALYSIS OF INTAKE/OUTPUT/VOID HABITS WITH RESULTING CLUSTERS COMPARED BY CONTINENCE STATUS, AGE, RACE

Hypothesis / aims of study

Clinicians attempt to reduce and interpret the multi-factorial information contained on a voiding diary by subjectively reducing the differing aspects into a single, unified pattern presumed to be associated with a particular bladder health state. This process, however, has relied on subjective rather than objectively driven analysis. We analyzed diary data, from both continent and incontinent women, to identify simple, discrete, patterns that emerge from the full spectrum of intake/output variables when subjected to a statistical data-driven process. We: 1) used cluster analysis to look for naturally occurring groupings, 2) described and labelled the resulting clusters, and 3) profiled the clusters by comparing them across variables commonly associated with bladder health but external to the cluster analysis variables.

Study design, materials and methods

Participants included 352 community dwelling women of Black (n =196) and White (n = 156) race who agreed to participate in the clinical portion of a parent population-based study on prevalence of urinary incontinence [1, 2]. The clinical portion of the study oversampled for incontinence. All interpretable 3-day intake/output/continence status diaries were evaluated. Six variables (void frequency during daytime hours, void frequency during night-time hours, modal output, total output, beverage intake, and BMI) were used for the cluster analyses. Cluster analysis requires the investigator to specify the number of clusters to extract. We selected to evaluate 3, 4, and 5 cluster models to compare classification results for best fit (most parsimonious yet distinct) using the six variables subjected to the analysis. The hierarchical tree clustering statistical technique is designed to identify homogenous groups within sets of data [3]. Once cluster groupings were revealed, we compared other diary variables of interest including proportion and amount of beverage "irritants" (coffee, tea, colas, alcohol, and juice), continence status, incontinence episodes; we also compared clusters by race (Black/White), education, and age.

Results

The 3 clusters model demonstrated best fit as compared to the 4 or 5 clusters models. Each of the three clusters that emerged differed significantly ($p < .001$) on all of the 6 variables used to construct them except BMI, as shown in Table 1. We descriptively labelled the three clusters "Conventional," "Benchmark," and "Supersize."

Table 1 shows descriptive results of all clustering variables for all subjects. Results are expressed as Mean ± standard error; P values are computed by ANOVA test. Cluster comparison significance at 0.05 level is indicated by *. Bonferroni correction is used to adjust for multiple comparisons among the clusters

	Cluster 1: Conventional (N=233)	Cluster 2: Benchmark (N=96)	Cluster 3: Supersize (N=23)	P value	R^2	Cluster comparison
Total intake per Daytime (ml)	1320 ±375	2444 ±542	3773 ±912	<0.001	0.626	3-2* 2-1* 1-3*
Total output during 24 Hr (ml)	1068 ±434	1907 ±437	3280 ±1066	<0.001	0.689	3-2* 2-1* 1-3*
Modal output at daytime (ml)	289 ±143	444 ±201	646 ±209	<0.001	0.278	3-2* 2-1* 1-3*
Num of voids daytime (count)	5.2 ±1.8	6.9 ±2.3	10.5 ±5.3	<0.001	0.270	3-2* 2-1* 1-3*
Num of voids at night-time (count)	1.1 ±0.8	1.4 ±0.9	1.9 ±1.2	<0.001	0.056	3-2* 2-1 1-3*
BMI (points)	30.3 ±7.1	30.7 ±7.6	30.0 ±9.3	0.81	0.001	3-2 2-1 1-3

Table 2 shows comparison by clusters of clinical and demographic variables external to those used for the cluster analysis. For continuous variables (i.e., age), p values were computed by ANOVA test. Results are expressed as mean ± standard error; for categorical variables (i.e., race). P values were computed by chi-square test and results are expressed as numbers (percentages).

	Conventional (N=233)	Benchmark (N=96)	Supersize (N=23)	P value

Amount of Bladder irritants intake (ml)	1310 ±373	2417 ±559	3318 ±110	<0.01
Proportion of bladder irritants	0.64 ±0.3	0.58 ±0.3	0.65 ±0.3	0.27
Incontinence episodes/ day (average)	0.9 ±1	1.2 ±2	2.0 ±3	0.02
Incontinence episodes/day (n & %)				0.05
0 leakages (n=164)	114(41.7%)	40(48.9%)	10(43.5%)	
0< leakages = 1 (n=97)	66(30.2%)	29(28.3%)	2(8.7%)	
leakages > 1 (n=91)	53(28.1%)	27(22.8%)	11(47.8%)	
Race				<0.01
White	82(35.2%)	58(60.4%)	16(69.6%)	
Black	151(64.8%)	38(39.6%)	7(30.4%)	
Age (years)	49.7±8.1	49.2±7.7	53.3±7.4	0.08
Education (years)	14.2±2.0	14.4±2.1	14.0±2.2	0.60

Table 2 shows that the proportion of beverage irritant to non-irritant intake did not differ significantly by clusters, but the absolute amount of irritants did differ with the Benchmark cluster nearly double, and Supersize cluster nearly triple, compared with the Conventional cluster ($p<.001$). The Supersize cluster had greatest incontinence severity and was predominantly of White race. Age and education did not differ across the clusters.

Interpretation of results

There is no agreed-upon standard, simple, yet comprehensive classification system for voiding diaries. Our results offer an objective and data-derived 3-cluster schema with labels of “Conventional,” “Benchmark,” and “Supersize” paralleling the descriptive parameters that emerged. This 3-cluster categorization schema solves a difficult data management issue for both clinicians and researchers, and opens the door to scrutiny of whether pre-conceived notions of “healthy bladder behaviors” are supported by actual data. For instance, the advice to drink at least 8 glasses of water per day does not appear to be protective against leakage.

Concluding message

Identification of three discrete clusters of diary data provides a parsimonious but data-driven means of classifying individuals for additional epidemiological or clinical study. The research and clinical utility rests with the simplicity of classification, and the potential for intervening to move an individual from high risk to low risk cluster with regards to incontinence.

References

1. Fenner, D.E., Trowbridge, E.R., Patel, D.L., Fultz, N.H., Miller, J.M., Howard, D. & DeLancey, J.O.L. (2008). Establishing the prevalence of incontinence study: Racial differences in women’s patterns of urinary incontinence. The Journal of Urology, 179, 1455-1460.
2. Delancey J.O., Fenner D.E., Guire K., Patel D.A., Howard D., & Miller J. Differences in continence system between community dwelling black and white women with and without urinary incontinence in the EPI study. American Journal of Obstetrics & Gynecology, in press.
3. Hair, J.F., Anderson, R.E., Thatham, R.I., Black W.C. (1992). Multivariate data analysis. 3rd ed. New York, NY: Maxwell MacMillan International Editions.

Specify source of funding or grant	**We gratefully acknowledge the financial support of U.S. Public Health Service grants R01 HD/AG41123, P30NR009000, P50 HD44406.**
Is this a clinical trial?	**No**
What were the subjects in the study?	**HUMAN**
Was this study approved by an ethics committee?	**Yes**
Specify Name of Ethics Committee	**University of Michigan Institutional Review Board**
Was the Declaration of Helsinki followed?	**Yes**
Was informed consent obtained from the patients?	**Yes**