



## DATA COLLECTION METHODS AND WEIGHTING “NO PATTERN”

Frederick H. Navarro, PhD

This article describes the “No Pattern” outcome of PATH Analysis, its origin, and the need to provide weighting factors to *weight down* those respondents classified as No Pattern and weight *up* respondents identified with one of the nine valid Patterns of Adapting to Health (PATH) when the Adaptive Health Behavior Inventory (AHBI) response data is collected using *online* or *mail surveys*.

### Overview

The Patterns of Adapting to Health (PATH) model was developed from the outcome of a master’s thesis in psychological research. Using a random-digit dialing telephone survey methodology, the research tested and supported the hypothesis that health and health care psychographic segments validated in one geographic region are generalizable to other geographic regions of the United States<sup>1</sup>. Prior to this research, the application of cluster analysis to reveal or construct segments using self-reported health-related variables had only been studied in single populations, markets or single geographic regions<sup>234</sup>.

Commercially available health psychographic segmentation models developed since these early studies have all been based on the analysis of one sample applied to all markets<sup>567</sup>.

Another common outcome of early and contemporary cluster analytic studies to identify segments is the classification of every respondent into a cluster. This is more an artifact of the clustering technique which outputs case classifications without information on the degree of actual fit. This invariably creates reduced homogeneity and inconsistency within each segment resulting in “mushy” segments<sup>8</sup>. To reduce this outcome, the research used to identify the PATH allowed for the possibility that some cases may not have a pattern of responses that fit or conform to patterns within validated segments.

At the final stage of the master’s thesis research when the final segment solution was presented, eleven percent of the sample was un-classified<sup>9</sup>. These were cases whose pattern of responses did not conform to any of the validated segments found to exist across geographic regions. Subsequent attempts to re-cluster these



respondents and validate the clusters using split-half validation failed. No Pattern respondents did not give responses that showed the influence of a latent pattern.

## **Data Collection Methodology and No Pattern**

*Telephone Interviews.* For over a decade telephone survey data collection was the primary methodology used to identify the PATH in community market research. As shown in Table 1, No Pattern rates associated with telephone surveys conducted between 1995 and 2006 in different regions of the U.S. matched the non-classified rate found in the master's thesis research. This supported and validated the reliability of the No Pattern rate.

*Mail Surveys.* The second most frequently used data collection methodology to identify the PATH was mail surveys. Unlike telephone surveys, however, mail surveys produced No Pattern rates as high as 30% and simultaneously under-estimated the size of adults dominated by the other nine PATH. Many issues associated with mail surveys relevant to increasing the No Pattern rate are discussed elsewhere<sup>10</sup>. As a result, PATH results from mail surveys were weighted to match the No Pattern rate obtained by telephone interviewing and correct the under-reported percentage mix of the other nine PATH.

*Interactive Voice Response (IVR).* IVR data collection was used in a 2006 study to identify the PATH mix within a disease management population. The over 6,000 IVR interviews mimicked the experience of a telephone interview in that participants responded to each AHBI item one at a time as if being questioned by an interviewer. Like telephone surveys, IVR data collection produced a No Pattern rate in the 11% range. No sample weighting necessary.

*Online, Web-Based Surveys.* Early in the 2000s AHBI response data started being collected via online data collection including the use of online panels. Early results yielded No Pattern percentage rates at the same level as mail surveys close to the 30% range. Examination of AHBI response data at the case level found similar problems to those encountered with mail surveys as well as others sited in the research investigating the integrity of online data collection including high



volumes of missing data<sup>11</sup> and careless responses given by up to 10 to 12% of respondents<sup>12</sup>.

The benefit of having years of PATH data collected via telephone surveys was the ability to see how the shift from telephone interviewing to online data collection effected the PATH mix and no pattern rate. It was surprising to see PATH mix results obtained through online data collection mimicking that of mail surveys in terms of high No Pattern percentages. As in the case of mail surveys, this required the weighting of PATH results to match the No Pattern rate obtained by telephone interviewing.

To summarize, of the four major data collection methods used to collect health-related response data from large populations, both mail surveys and online surveys dramatically over-estimate the percentage of those adults with No Pattern and under-estimate the percentage of those adults in the other nine PATH.

Table 1

*Patterns of Adapting to Health (PATH) Assignment by Data Collection Method*

PATH Names	Telephone <sup>a</sup>		Mail <sup>b</sup>		IVR <sup>c</sup>		Online <sup>d</sup>	
	<u>n</u>	%	<u>n</u>	%	<u>n</u>	%	<u>n</u>	%
Critically Discerning	763	2%	631	2%	148	2%	713	2%
Health Contented	2,126	5%	2,459	7%	288	5%	2,917	7%
Wisely Frugal	7,371	19%	5,121	15%	1,126	19%	6,110	15%
Traditionalist	2,290	6%	1,678	5%	215	4%	1,798	4%
Family Centered	5,322	13%	3,312	10%	1,020	17%	3,878	10%
Family Driven	2,747	7%	2,428	7%	622	10%	2,970	7%
Health Care Driven	6,821	17%	3,659	11%	1,001	17%	4,446	11%
Independently Healthy	4,458	11%	2,434	7%	462	8%	3,009	8%
Naturalist	3,189	8%	2,201	6%	512	8%	2,616	7%
No pattern	4,492	11%	10,218	30%	632	10%	11,632	29%
Total	39,580	100%	34,141	100%	6,026	100%	40,090	100%

<sup>a</sup>Data based on telephone surveys conducted from 1995 to 2006

<sup>b</sup>Data based on mail surveys conducted between 2001 and 2003

<sup>c</sup>Data based on interactive voice response (IVR) survey of disease management pop conducted in 2006

<sup>d</sup>Data based on online surveys conducted between 2002 and 2017



### The Need for Sample Weighting.

Ideally, a selected sample is a miniature exact replica of the population it came from. This would make the sample a valid representation of the population across all the variables measured in the survey. Unfortunately, this is typically not the case. Issues such as non-response bias may cause some groups to be over- or under-represented. Self-selection, a problem in online surveys, can also undermine a sample's representativeness as well. When this occurs, conclusions about the population based on the observed survey data are questionable, unless something has been done to correct for the lack of representativeness. This is when sample weighting comes in. Weighting assigns an adjustment weight to each survey respondent. Persons in under-represented segments get a weight larger than 1, and those in over-represented groups get a weight smaller than 1. In the computation of means, totals and percentages, the weighted values are used instead of the original values.

### PATH Weighting Approach

The weighting approach applied to both mail survey and online data collection PATH results is the same: 1) Weight down the percentage of No Pattern to the 11% range and 2) weight up the percentage of mix of the other nine PATH by the percentage difference between the *obtained* No Pattern percentage rate and the 11% percentage rate divided by nine.

#### **Example:**

Table 2 shows PATH results from an online survey of 1,000 adults. The No Pattern percentage rate was 29%.

Table 2

PATH Names	n	%
Critically Discerning	18	1.8%
Health Contented	73	7.3%
Wisely Frugal	152	15.2%
Traditionalist	45	4.5%
Family Centered	97	9.7%
Family Driven	74	7.4%
Health Care Driven	111	11.1%
Independently Healthy	75	7.5%
Naturalist	65	6.5%
No pattern	290	29.0%
Total	1,000	100.0%

**Step 1** - Calculate the percentage point difference between this No Pattern rate and the 11% rate. In this case the difference is  $29\% - 11\% = 18\%$

**Step 2** – Divide the 18% by nine.  $18\%/9 = 2\%$

**Step 3** – Add the 2% to the *obtained* percentage for each of the nine PATH (e.g., Critically Discerning through Naturalist) as shown in Table 3.

Table 3

PATH Names	n	%	Weighted
Critically Discerning	18	1.8%	+ 2% = 3.8%
Health Contented	73	7.3%	+ 2% = 9.3%
Wisely Frugal	152	15.2%	+ 2% = 17.2%
Traditionalist	45	4.5%	+ 2% = 6.5%
Family Centered	97	9.7%	+ 2% = 11.7%
Family Driven	74	7.4%	+ 2% = 9.4%
Health Care Driven	111	11.1%	+ 2% = 13.1%
Independently Healthy	75	7.5%	+ 2% = 9.5%
Naturalist	65	6.5%	+ 2% = 8.5%
No pattern	290	29.0%	11.0%
Total	1,000	100.0%	100.0%

The **Weighted** column in Table 3 shows the *expected* PATH percentage mix. Each *obtained* percentage for the nine PATH was increased by two percentage points and the No Pattern rate decreased to the *expected* 11%.

### Calculating the Weighting Factors

The next step is to calculate the weighting factors that must be applied to respondents dominated by each PATH and the No Pattern outcome.

**Step 4** – Apply the new Weighted PATH percentages to the total sample size. This is simply done by multiplying the total sample size of 1,000 by the Weighted percentages for each PATH (Table 4). This produces new anticipated sample sizes (Weighted n) for each PATH and the No Pattern *after* the weighting factors are applied.



Table 4

PATH Names	n	%	Weighted	Weighted n
Critically Discerning	18	1.8%	+ 2% =	3.8%
Health Contented	73	7.3%	+ 2% =	9.3%
Wisely Frugal	152	15.2%	+ 2% =	17.2%
Traditionalist	45	4.5%	+ 2% =	6.5%
Family Centered	97	9.7%	+ 2% =	11.7%
Family Driven	74	7.4%	+ 2% =	9.4%
Health Care Driven	111	11.1%	+ 2% =	13.1%
Independently Healthy	75	7.5%	+ 2% =	9.5%
Naturalist	65	6.5%	+ 2% =	8.5%
No pattern	290	29.0%		110
Total Sample	1,000	100.0%		1,000

**Step 5** – To calculate the weighting factors, simply divide the Weighted n for each PATH by the original sample size n for each PATH. For example, the weighting factor applied to Critically Discerning respondents is calculated by dividing the Weighted n of 38 by the original subsample n of 18 =  $38/18 = 2.125$ . The weighting factor applied to Health Contented respondents is calculated by dividing the Weighted n of 93 by the original subsample n of 73 =  $93/73 = 1.275$ . Repeating this process for each PATH category results in the weighting factors for all the PATH as shown in Table 5.

Table 5

PATH Names	n	%	Weighted	Weighted n	Weighting Factors
Critically Discerning	18	1.8%	+ 2% =	3.8%	2.125
Health Contented	73	7.3%	+ 2% =	9.3%	1.275
Wisely Frugal	152	15.2%	+ 2% =	17.2%	1.131
Traditionalist	45	4.5%	+ 2% =	6.5%	1.446
Family Centered	97	9.7%	+ 2% =	11.7%	1.207
Family Driven	74	7.4%	+ 2% =	9.4%	1.270
Health Care Driven	111	11.1%	+ 2% =	13.1%	1.180
Independently Healthy	75	7.5%	+ 2% =	9.5%	1.266
Naturalist	65	6.5%	+ 2% =	8.5%	1.307
No pattern	290	29.0%		110	0.379
Total Sample	1,000	100.0%		1,000	



When applying weighting factors, the recommendation is to: 1) figure out all demographic weightings first, 2) calculate the PATH weighting factors, 3) multiply both sets of weighting factors by each other to create “composite weighting factors”, then 4) use the composite weights to run the tables. This will produce a sample that is a better representation of the population.

### **When to Apply Weighting**

As a rule of thumb, if the sample results show a No Pattern rate of 14% or greater, the above methodology should be applied.

### **PATH Weighting Factor Calculator**

The above described calculations to identify PATH weighting factors can be easily done using the **PATH Weighting Factor Calculator** available for download at <http://www.pathinstitute.life/path-reference.html>. It is a standard Excel spreadsheet. Download the file to a folder on your computer. After opening, simply enter the counts for each PATH from your survey if the No Pattern rate is 14% or greater. The spreadsheet will automatically calculate the recommended weight factors.



AutoSave (off) File Home Insert Page Layout Formulas Data Review View Developer Tell me what you want to do

C40 A B C D E F G H I J K

**Patterns of Adapting to Health (PATH) Weighting Factor Calculator**

This spreadsheet can quickly calculate PATH weighting factors when the No Pattern rate of a survey is 14% or greater. Simply run a frequency table of PATH counts from the survey. If the count of No Pattern is 14% or greater, enter the counts for each PATH in the "Counts" column. Once entered they should sum to the total sample n. The recommended weighting factors will automatically be calculated and appear in the "Weighting Factors" column. These weighting factors should be added to the data set and multiplied by any demographic weightings to create "composite weighting factors" that can be applied to the data.

Code	The PATH	Counts	%	Adjustments	Weighted %	Weighted Counts	Weighting Factors
1	Critically Discerning		#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
2	Health Contented		#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
3	Wisely Frugal		#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
4	Traditionalist		#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
5	Family Centered		#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
6	Family Driven		#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
7	Healthcare Driven		#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
8	Independently Healthy		#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
9	Naturalist		#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
10	no pattern		#DIV/0!		11.0%	0	#DIV/0!
	Total n	0	#DIV/0!			#DIV/0!	#DIV/0!

Copyright ©2017 by PATH Institute. All Rights Reserved.

You can either save the spreadsheet with a different file name, or simply close it after use.

Here is a quick example of how to use it.

**Step 1:** Run a PATH frequency count from your survey. In this example, the sample size is 1,952 cases.

#### Frequency Distribution of PATH

PATH Code	Count	Percent
1	17	0.9%
2	154	7.9%
3	308	15.8%
4	33	1.7%
5	162	8.3%



6	179	9.2%
7	292	14.9%
8	220	11.3%
9	135	6.9%
10	452	23.2%

The “No Pattern” (code 10) is 23.2%, well over the 14% trigger.

## Step 2: Open the PATH Weighting Factor Calculator

Start entering the “counts” for each PATH in the “Counts” column.

Patterns of Adapting to Health (PATH) Weighting Factor Calculator

This spreadsheet can quickly calculate PATH weighting factors when the No Pattern rate of a survey is 14% or greater. Simply run a frequency table of PATH counts from the survey. If the count of No Pattern is 14% or greater, enter the counts for each PATH in the "Counts" column. Once entered they should sum to the total sample n. The recommended weighting factors will automatically be calculated and appear in the "Weighting Factors" column. These weighting factors should be added to the data set and multiplied by any demographic weightings to create "composite weighting factors" that can be applied to the data.

Enter counts

Calculations start

Code	The PATH	Counts	%	Adjustments	Weighted %	Weighted Counts	Weighting Factors
1	Critically Discerning	17	3.5%	-1.22%	2.3%	11	0.6556
2	Health Contented	154	32.2%	-1.22%	30.9%	148	0.9620
3	Wisely Frugal	308	64.3%	-1.22%	63.1%	302	0.9810
4	Traditionalist		0.0%	-1.22%	-1.2%	-6	#DIV/0!
5	Family Centered		0.0%	-1.22%	-1.2%	-6	#DIV/0!
6	Family Driven		0.0%	-1.22%	-1.2%	-6	#DIV/0!
7	Healthcare Driven		0.0%	-1.22%	-1.2%	-6	#DIV/0!
8	Independently Healthy		0.0%	-1.22%	-1.2%	-6	#DIV/0!
9	Naturalist		0.0%	-1.22%	-1.2%	-6	#DIV/0!
10	no pattern		0.0%		11.0%	53	#DIV/0!
	Total n	479	100.0%		100.0%	479	

Copyright © 2017 by PATH Institute. All Rights Reserved.

Counts are summed

## Step 3: Finish

After entering all the PATH counts, the total sample *n* will sum to the sample size. The full set of PATH weighting factors for the survey are calculated to four decimal places. Associate each weighting factor with the appropriate PATH code (e.g., PATH code 1 {Critically Discerning} gets weighting factor 2.5508).



### Patterns of Adapting to Health (PATH) Weighting Factor Calculator

This spreadsheet can quickly calculate PATH weighting factors when the No Pattern rate of a survey is 14% or greater. Simply run a frequency table of PATH counts from the survey. If the count of No Pattern is 14% or greater enter the counts for each PATH in the "Counts" column. Once entered they should sum to the total sample n.<sup>1</sup> Weighting factors recommended weighting factors will automatically be calculated and appear in the "Weighting Factors" column. These weighting factors should be added to the data set and multiplied by any demographic weightings to create "composite weighting factors" that can be applied to the data.

Code	The PATH	Counts	%	Adjustments	Weighted %	Weighted Counts	Weighting Factors
1	Critically Discerning	17	0.9%	1.35%	2.2%	43	2.5508
2	Health Contented	154	7.9%	1.35%	9.2%	180	1.1712
3	Wisely Frugal	308	15.8%	1.35%	17.1%	334	1.0856
4	Traditionalist	33	1.7%	1.35%	3.0%	59	1.7989
5	Family Centered	162	8.3%	1.35%	9.6%	188	1.1627
6	Family Driven	179	9.2%	1.35%	10.5%	205	1.1473
7	Healthcare Driven	292	15.0%	1.35%	16.3%	318	1.0903
8	Independently Healthy	220	11.3%	1.35%	12.6%	246	1.1198
9	Naturalist	135	6.9%	1.35%	8.3%	161	1.1953
10	no pattern	452	23.2%		11.0%	215	0.4750
	Total n	1,952	100.0%		100.0%	1,952	

Copyright © 2017 by PATH Institute. All Rights Reserved.

Counts sum to total sample size

## References

- <sup>1</sup> Navarro, F. H. (1990). Profiles of attitudes toward healthcare: psychographic segmentation. *San Francisco State University (Master's Thesis)*.
- <sup>2</sup> Harris, D. M., & Guten, S. (1979). Health-protective behavior: An exploratory study. *Journal of Health and Social Behavior*, 17-29.
- <sup>3</sup> Walker, S. N., Volkan, K., Sechrist, K. R., & Pender, N. J. (1988). Health-promoting life styles of older adults: comparisons with young and middle-aged adults, correlates and patterns. *Advances in Nursing Science*, 11(1), 76-90.
- <sup>4</sup> Bartko, J. J., & Kasper, S. (1989). Seasonal changes in mood and behavior: a cluster analytic approach. *Psychiatry Research*, 28(2), 227-239.
- <sup>5</sup> Endresen, K. W., & Wintz, J. C. (1988). HealthStyles: a new psychographic segmentation system for health care marketers. *The Journal of Medical Practice Management: MPM*, 3(4), 273-277.
- <sup>6</sup> c2b Solutions (2017). Psychographic segmentation. Retrieved from <https://www.c2bsolutions.com/psychographic-segmentation>
- <sup>7</sup> Deloitte Development, LLC (2012). The U.S. health care market: a strategic view of consumer segmentation. Retrieved from <https://www2.deloitte.com/content/dam/Deloitte/us/Documents/life-sciences-health-care/us-lshc-health-care-market-consumer-segmentation.pdf>
- <sup>8</sup> Riedesel, P. (2008). Applying archetypal analysis in marketing research. *Action Marketing Research*. Retrieved from <http://www.action-research.com/aaa.pdf>
- <sup>9</sup> Navarro, F. H. (1990). Profiles of attitudes toward healthcare: psychographic segmentation. *San Francisco State University (Master's Thesis)*, Table 4, p. 17.
- <sup>10</sup> Adaptive Health Behavior (AHBI) Administration ([http://www.pathinstitute.life/uploads/1/0/7/0/107059763/adaptive\\_health\\_behavior\\_inventory\\_ahbi\\_administration.pdf](http://www.pathinstitute.life/uploads/1/0/7/0/107059763/adaptive_health_behavior_inventory_ahbi_administration.pdf)).
- <sup>11</sup> Cantrell, M. A., & Lupinacci, P. (2007). Methodological issues in online data collection. *Journal of Advanced Nursing*, 60(5), 544-549.



---

<sup>12</sup>Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods*, 17(3), 437.