

Paper 225-2011

## Investigating Web Search Engine Retrieval of Medical Information on Antidepressant Discontinuation Syndrome Using SAS® Text Miner

Barry W. Row, University of Louisville, Louisville, KY, USA

### ABSTRACT

Availability of information on medical topics is increasingly important for patients to be able to make informed choices regarding their own healthcare. Given that Web search engines are now a primary information resource, the present study used SAS Text Miner and the SAS %tmfilter Web crawling macro to retrieve and analyze documents from the Web search engines Bing, Ask.com, and Yahoo related to antidepressant discontinuation syndrome, a common but little-known side-effect of drugs used to treat depression. Cluster- and concept-link analyses revealed that distinct differences were found between search engines, with the documents retrieved from the Bing search engine related to specific drug treatments, while the documents retrieved from the Ask.com and Yahoo searches were related to more general information on antidepressant discontinuation syndrome. Our findings illustrate the utility of SAS Text Miner and the SAS %tmfilter Web crawling macro in evaluating the information readily available to patients through Web searches, since the search engine used may impact the type of information retrieved.

### INTRODUCTION

The exponential increase in health-related information over the last few decades has made it increasingly imperative that individuals be able to act as their own healthcare advocate in order to have full knowledge of healthcare treatments and potential side effects that may not be supplied by their healthcare provider, either due to internal bias, or simply lack of knowledge. One such side effect that is not widely known is antidepressant discontinuation syndrome. Antidepressant discontinuation syndrome is a constellation of symptoms that occurs in some patients upon termination or interruption of antidepressant treatment, and includes symptoms such as anxiety, insomnia, headaches, dizziness, fatigue, nausea, and hyperarousal (Warner, 2006). Although rarely severe enough to require hospitalization, many patients still report significant discomfort and work loss. Furthermore, these symptoms can be misdiagnosed as a relapse of the depression, given that they resolve quickly with the reinstitution of antidepressant treatment (Blier and Tremblay, 2006). The wide use of antidepressants by non-specialists and for off-label use, in addition to the fact that antidepressants are one the most prescribed class of drugs in the U.S., has made it important for patients to be aware of such potential problems (Olfson and Marcus, 2009). Given that web search engines are now one of the primary sources of information for most people, we sought to examine the availability of information on antidepressant discontinuation syndrome that is easily available to patients, using SAS Text Miner and commonly used web search engines. Cluster and concept link analysis revealed a large amount of information on antidepressant continuation syndrome is now readily available to the lay public. Furthermore, we also found that the web search engine used plays an important role in the final composition of the information available.

## METHODS

Documents were retrieved and imported into SAS data sets from the search engines, Bing (<http://www.bing.com/>), Ask.com (<http://www.ask.com/>), and Yahoo! (<http://www.yahoo.com/>) [6] using a keyword search of the term “antidepressant discontinuation syndrome”, and the web-crawling SAS %tmfilter macro as follows:

```
%tmfilter
(url=%NRSTR(http://search.yahoo.com/search;_ylt=AmQjsrOyUz4DqcxXSaVbfOWbvZx4?p=anti
depressant+discontinuation+syndrome&toggle=1&cop=mss&ei=UTF-8&fr=yfp-t-701),
depth=2,
dir=c:\BarryRow\YahooADCSWebcrawl800\dir,
destdir=c:\BarryRow\YahooADCSWebcrawl800\destdir,
norestrict=1, dataset=BR.YahooADCSWebcrawl800, numbytes=800)
```

```
%tmfilter
(url=%NRSTR(http://www.ask.com/web?q=antidepressant+discontinuation+syndrome&search
=&qsrc=0&o=0&l=dir), depth=2,
dir=c:\BarryRow\YahooADCSWebcrawl800\dir,
destdir=c:\BarryRow\YahooADCSWebcrawl800\destdir,
norestrict=1, dataset=BR.YahooADCSWebcrawl800, numbytes=800)
```

```
%tmfilter
(url=%NRSTR(http://www.bing.com/search?q=antidepressant+discontinuation+syndrome&go
=&form=QBLH&qsn=&sk=&sc=8-12), depth=2,
dir=c:\BarryRow\BingADCSWebcrawl800\dir,
destdir=c:\BarryRow\BingADCSWebcrawl800\destdir,
norestrict=1, dataset=BR.BingADCSWebcrawl800, numbytes=800)
```

The SAS %tmfilter macro imported 2322, 2133, and 2219 web pages from the Bing, Ask.com, and Yahoo! search engines, respectively. Once imported into SAS, all analyses were performed using the Text Miner 4.2 component of SAS Enterprise Miner 6.1. Text parsing was performed with the default parameters, with the exception of the Different Parts of Speech option, which was set to no. Text transformation, and clustering was performed using the default parameters with the exception of the Automatically Cluster option set to Yes, and the Number of Clusters and Descriptive Term options set to 10 and 20, respectively. Cluster analyses were then performed using default SVD weights and the Expectation-Maximization method. Parameters were kept consistent between all data sets and the standard SAS stoplist was used for all data sets to allow meaningful comparisons between web search engine results.

## RESULTS

Because web-crawling macros will typically retrieve a large number of documents unrelated to the topic, an initial clustering was performed on each data set to filter out irrelevant web sites as outlined in Methods.

### CLUSTERING

Preliminary analyses of the data set from the search engine Bing revealed 6 clusters, as shown in Table 1. Inspection of the clusters revealed that only Clusters 2 (480 documents) and 4 (763 documents) contained terms relevant to our topic of interest.

**Table 1. Preliminary Clusters identified from Bing Web Search.**

#	Descriptive Terms	Freq	Percentage	RMS Std.
2	+ publish, + result, + much, + have, discontinuation, + turn, + antidepressant, + advance, clear, trazadone, effexor, bing, + bing web image video, web, + sign, + see, history, trazodone, + shop, withdrawal	763	0.3285960378983635	0.15913724616321
4	+ review, + content, navigation, depression, as, + medication, + page, information, + condition, + treatment, free encyclopedia, + disease, + article, + disorder, health, with, bipolar, medical, + drug, jump	480	0.20671834625322996	0.13979161405395
6	music, + ad, windows, legal, + help, + advertise, privacy, + live, microsoft, + support, + no., + movie, + university, louisville, kentucky, hotmail, bing, msn, web, + setting	478	0.2058570198105082	0.16368580457281
5	+ warn, + not, + resource, + gene, ncbi, + reference, + require, + request, blast, popset, sra, cache, + site, + find, + journal, + database, + sequence, javascript, ncbi web site, + can	333	0.1434108527131783	0.15025534334981
3	privacy, + medium, south, america, sprint, wimbledon, microsoft, contest, + schedule, + preference, ufc, + advertise, tournament, thom, stone buddha head, stone, mvps, + help, tree, united	165	0.0710594315245478	0.14375659806364
1	white, real, weather, today, + edition, + career, + game, + show, + science, estate, + business, + travel, yellow, + horoscope, + page, + quote, msnbc, + yellow page, wonderwall, personals	103	0.044358311800172266	0.13691056222979

Documents matching these clusters were filtered from the original data set and reclustered as shown in Table 2. Documents were again filtered to concentrate on those clusters most closely related to our search topic (Clusters 2 and 4), "antidepressant discontinuation syndrome".

**Table 2. Filtered and Reclustered Documents from Bing Search.**

#	Descriptive Terms	Freq	Percentage	RMS Std.	Label
4	+ cause, + live, with, + site, + journal, + short, + physician, qt, + child, + page, diagnosis, ncbi, ekg, ecg, + help, information, + advance, + article, on, + search	166	0.3472803347280335	0.13817232985452	Medical References
3	with, + treatment, bipolar, + drug, + support, on, diagnosis, side, + medication, + share, depression, + home, + effect, + disorder, + review, information, health, + symptom, free, + do	263	0.5502092050209205	0.13386019152469	Mental Health Disorders and Information

#	Descriptive Terms	Freq	Percentage	RMS Std.	Label
4	+ cause, + live, with, + site, + journal, + short, + physician, qt, + child, + page, diagnosis, ncbi, ekg, ecg, + help, information, + advance, + article, on, + search	166	0.3472803347280335	0.13817232985452	Medical References
3	with, + treatment, bipolar, + drug, + support, on, diagnosis, side, + medication, + share, depression, + home, + effect, + disorder, + review, information, health, + symptom, free, + do	263	0.5502092050209205	0.13386019152469	Mental Health Disorders and Information
1	fluoxetine, + discontinuation syndrome, withdrawal, zoloft, effexor, prozac, + trazodone, syndrome search history, serotonin, + have, + search, free encyclopedia, navigation, wikipedia, encyclopedia, ssri, cymbalta, jump, + taper, free	358	0.7489539748953975	0.13518379109301	Antidepressant treatments and Discontinuation
2	+ brand, web, oxford, news, + food, + publish, + map, wellness, health, wiley, + browse, + setting, + result, + sport, + compute, software, + browse all product, sons, antidepressant discontinuation, + much	456	0.9539748953974896	0.13413569269821	Health and Wellness

Preliminary analyses of the data set from the search engine, Ask.com, revealed 5 clusters, as shown in Table 3. Inspection of the clusters revealed that only Cluster 1 (605 documents) contained terms relevant to our topic of interest.

**Table 3. Preliminary Clusters identified from Ask.com Web Search.**

#	Descriptive Terms	Freq	Percentage	RMS Std.
1	+ antidepressant, + symptom, health, anxiety, mental, depression, register, + guide, + forget, psychiatric, information, + treatment, remember, + relate, + medication, + disorder, + no., + suicide, + community, + definition	605	0.28363806844819506	0.15480558907654
2	+ service, + set, + create, + business, out, + title, + code, + state, + city, + search, + result display save location, + result, + list, + feature, + location, + sign, ask.com, + filter, customize, + save	391	0.18330989217065166	0.14116067399782
3	+ page, + link, + search, + display, + tv, + shop tv listing event, + result display save location, + result, + save, + list, + browse, ask.com, + will, + feature, + event, email, + deal, + have, + shop, + recipe	674	0.31598687294889827	0.14076425271668
4	+ print, + position, + right, font-size, + clear, + top, + leave, margin, height, + float, text-decoration, 1px, background-color, :none, font-weight, width, + medium, + pad, color, li	122	0.05719643694327239	0.1180489304012
5	wikipedia, + encyclopedia more map, + article, + other, on, jump, free encyclopedia, original, + redirect, navigation, + see, encyclopedia, last, + may, free, + modify, + can, with, + not, + shop tv listing event	341	0.15986872948898265	0.12139130091193

Documents matching Cluster 1 were filtered from the original data set and reclustered as shown in Table 4. Documents were again filtered to concentrate on those clusters most closely related to our search topic (Clusters 4, 1, and 3), "antidepressant discontinuation syndrome".

**Table 4. Filtered and Reclustered Documents from Ask.com Search.**

#	Descriptive Terms	Freq	Percentage	RMS Std.	Label
---	-------------------	------	------------	----------	-------

#	Descriptive Terms	Freq	Percentage	RMS Std.	Label
4	bipolar, + forget, healthyplace, + bookmark, anxiety, remember, personality, information, + tv, psychiatric, mental, news, + site, + eat, + alternative, + no., mental health, abuse, + video, + community	169	0.2793388429752066	0.12609596562239	Psychiatric disorder and Mental health
6	+ condition, feedback, about.com, + review, + cope, + member, + share, login, + article, + story, + newsletter, health, + treatment, diagnosis, + disease, + search, + profile, + forum, + content, + home	149	0.2462809917355372	0.13879850419698	Disease Information
1	ssri, selective, navigation, encyclopedia, wikipedia, discontinuation, jump, + syndrome, reuptake, + inhibitor, serotonin, + article, + author, + forum, + search, free, + antidepressant, + list, + drug, medical	112	0.18512396694214875	0.14661217532145	SSRI Information
3	+ have, + answer, + stop, + feel, + not, + drug, most, + find, + may, + include, + antidepressant, + withdrawal, + content, + patient, with, as, + help, + can, discontinuation, + syndrome	82	0.13553719008264462	0.14589797942907	Antidepressant Withdrawal
5	web, all, central, + archive, + resource, + database, + protein, workbench, virus, stand-alone, sra, splign, sequence, rna, refseq, prosplign, primer-blast, popset, nucleotide, ncbi web site	53	0.08760330578512397	0.13998646894886	Genetic Searches
2	+ require, + error, on, disclaimer, information, + service, unexpected, merchant, pronto, full disclaimer, unexpected error, product related information, full, + product, responsible, ask.com, + read, + not, + relate, + in	40	0.06611570247933884	5.6623876789493E-15	Not Applicable

Preliminary analyses of the data set from the search engine Yahoo! revealed 6 clusters, as shown in Table 5. Inspection of the clusters revealed that only Cluster 1 (830 documents) and 2 (568 documents) contained terms relevant to our topic of interest.

**Table 5. Preliminary Clusters identified from Yahoo! Web Search.**

#	Descriptive Terms	Freq	Percentage	RMS Std.
1	+ disorder, + book, + symptom, wikipedia, + image, discontinuation, + syndrome, on, all, + program, + antidepressant discontinuation syndrome, free, + much, directory, + show, + video, + antidepressant, guest, + shop, + result	830	0.3740423614240649	0.15039991656853
2	+ treatment, login, + journal, + may, + family, + physician, + click, + article, + content, with, + do, email, + home, + drug, information, + contact, health, + not, + have, + other	568	0.25597115817936006	0.13513518688554
3	+ match, + learn, + may, below, + have, + will, smart, + location, + page, current, + policy, + other, + city, + update, + service, + provide, privacy, + much, email, + content	173	0.07796304641730509	0.14722404743322

#	Descriptive Terms	Freq	Percentage	RMS Std.
4	news finance sports search, + politic, + point, + science, out, education, entertainment, + work, + business, + sport, + family, + game, + sign, + category, internet, + new, + consumer, + art, + browse, + finance	237	0.106804867057233	0.14515349950133
5	internet, online, + user, back, explorer, + mail, news finance sports search, + replace, firefox, + article, + sign, + topic, + account, with, + security, + do, + late internet, + new, email, + sport	203	0.0914826498422713	0.13951031901705
6	+ set, + result, + see, + select, + suggestion, + advance, + pad, + image, + option, + key, up, + show, local, + start, + assist, + query, directory, + program, + advertise, + preference	208	0.09373591707976565	0.1402720542828

Documents matching Clusters 1 and 2 were filtered from the original data set and reclustered as shown in Table 6. Documents were again filtered to concentrate on those clusters most closely related to our search topic (Clusters 2 and 3), "antidepressant discontinuation syndrome".

**Table 6. Filtered and Reclustered Documents from Yahoo! Search.**

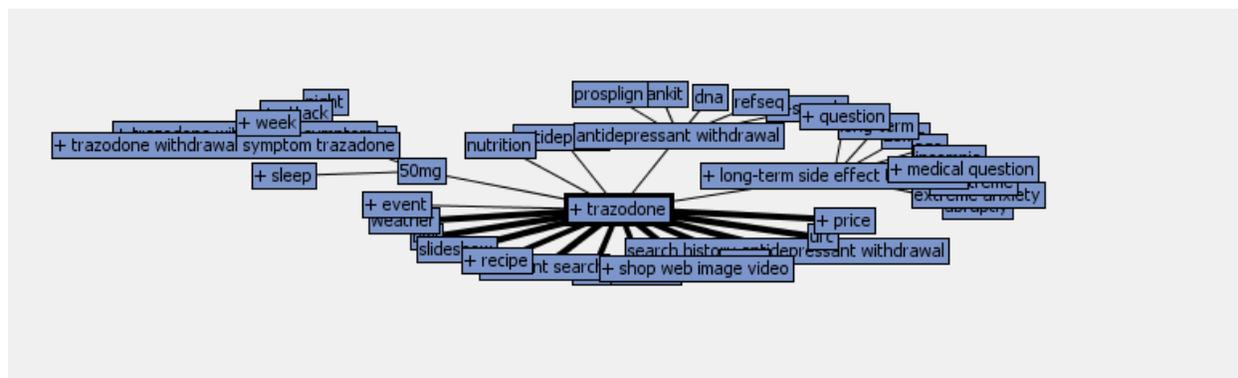
#	Descriptive Terms	Freq	Percentage	RMS Std.	Label
1	central, + analysis, pubmed, ncbi, javascript, + journal, workbench, stand-alone, sra, splign, refseq, prosplign, primer-blast, popset, nucleotide, ncbi web site, genome, genbank, dna, clusters	26	0.01859799713876967	0.00765678544929	Genomic Information
2	+ symptom, tour, news, + advance, + program, + mail, directory, guest, + job, + antidepressant, + video, discontinuation, + page, + help, + make, + answer, all, + preference, + query, antidepressant disco	273	0.19527896995708155	0.11164517151164	Antidepressant symptoms and discontinuation
3	+ forget, + remember, depression, free, + medication, jump, + community, + suicide, cancel, navigation, information, + test, + password, + addiction, healthyplace, bipolar, free encyclopedia, encyclopedia, panic, + other	436	0.31187410586552217	0.15122793618107	Mental Health Information and Treatments
4	+ user, + treatment, + live, + game, news finance sports search, + feedback, + computer, health, + garden, + article, + deal, + new, + register, + assist, + home, + have, music, with, + search, off	568	0.4062947067238913	0.15383749953677	Not Applicable
5	internet, earth, music, daily, government, + movie, poor, tech, most, nympho, hot, palin, noble, police, mcsweeney, losing, immigration, high price, gop, flooded	95	0.06795422031473533	0.00466502428329	Not Applicable

Examination of the primary clusters revealed distinct differences in the information provided by each of the three search engines on our search topic. The Bing search found 1 primary cluster related to discontinuation of antidepressant treatment (Cluster 1 in Table 2.), which was described by the generic or brand names of specific antidepressants, such as fluoxetine, Zoloft, Effexor, and trazodone. The Ask.Com search found 2 primary clusters related to antidepressant withdrawal (Cluster 3 in Table 4.), which were described by more general terms, such as antidepressants, and selective serotonin reuptake inhibitors (Cluster 1 in Table 4.), a common class of antidepressants. Similar to the Ask.com search, the Yahoo! search found one primary cluster related to antidepressant discontinuation syndrome, which was described by general terms, such as antidepressant and discontinuation (Cluster 2 in Table 6).

Comparison of secondary clusters also illustrated differences in the results provided by each of the three search engines. The Bing search found 2 secondary clusters related to general medical references (Cluster 4 in Table 2.) and general information on mental health disorders (Cluster 3 in Table 2.) The Ask.Com search found 2 secondary clusters related to general disease information and news (Cluster 6 in Table 4.) and psychiatric disorders and mental health (Cluster 4 in Table 4), while the Yahoo! search identified 1 large secondary cluster related to general mental health information and treatments (Cluster 3 in Table 6). Interestingly, both the Ask.com and Yahoo! Searches identified clusters related specifically to genetic research and information (Cluster 5 in Table 4, and Cluster 1 in Table 6) that would likely be of limited utility to an individual without a substantial background in biology and/or biomedicine. No such specific cluster was identified in the Bing search.

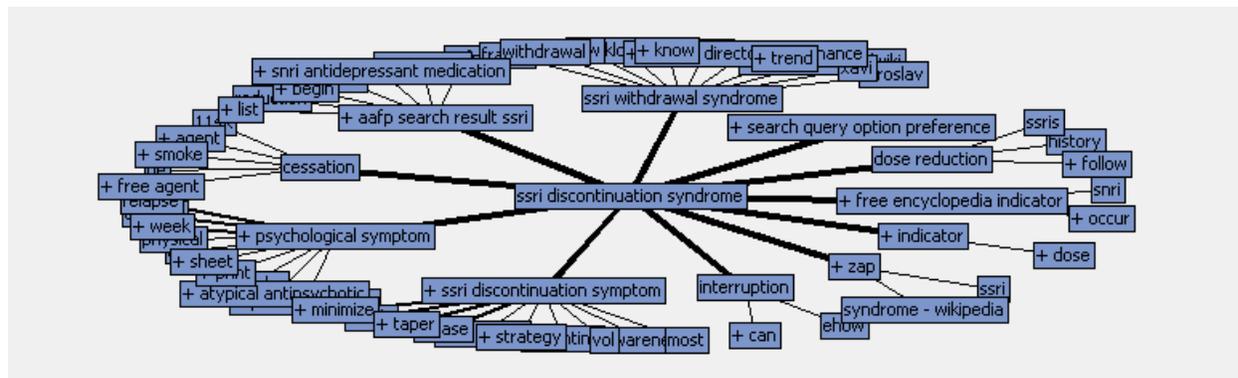
### CONCEPT LINKS

Using concept links, we can visually examine the relationships between terms related to our search. Concept link analysis identified key grouping differences in the data sets from the different search engines. For the Bing search, the strong concept links were identified with individual drugs used to treat depression. For example, Figure 1 shows the concept linkage map for the drug Trazodone, a serotonin modulator used to treat depression.



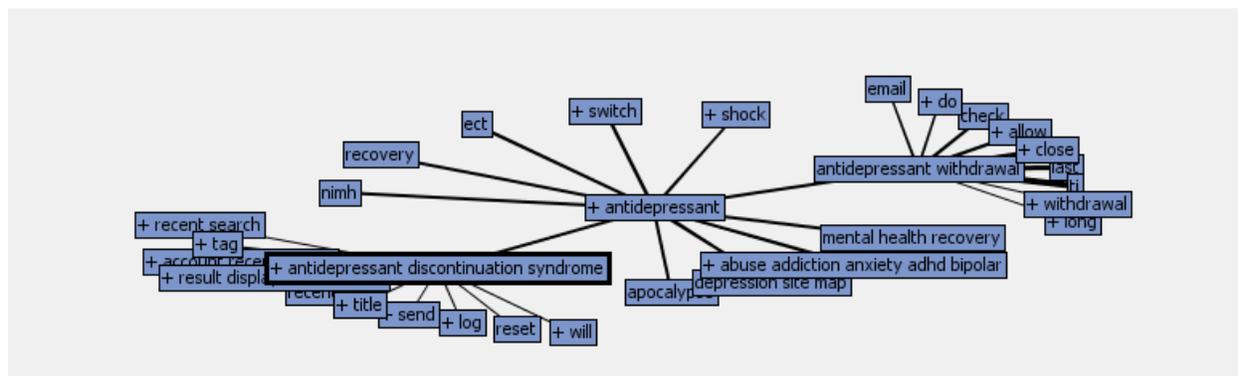
**Figure 1. Concept Linkage Map for Trazodone with expansion on terms related to Antidepressant Discontinuation Syndrome.**

In contrast, the strongest concept links from the Ask.com search related to antidepressant discontinuation syndrome were terms related to the general class of selective serotonin re-uptake inhibitors, as shown in Figure 2.



**Figure 2. Concept Linkage Map for SSRI Discontinuation Syndrome with expansion on terms related to Antidepressant Discontinuation Syndrome.**

As in the preceding example, the strongest concept links in the Yahoo search were for the general term, antidepressant, and the secondary links antidepressant discontinuation syndrome and antidepressant withdrawal, as shown in Figure 3.



**Figure 3. Concept Linkage Map for Antidepressant with expansion on terms related to Antidepressant Discontinuation Syndrome.**

It is important to note that these results are consistent with the clusters identified in each of the respective searches. As can be seen in Table 2, Cluster 1 from Bing search is clearly associated with specific drugs used in the treatment of depression. Cluster 1 from Table 4 (Ask.com) contains terms related to information on the general SSRI class of antidepressants. Finally, Clusters 2 and 4 in Table 6 (Yahoo!) contain terms related to depression and antidepressants.

## CONCLUSION

SAS Text Miner, in conjunction with the %tmfilter macro and a search engine, represent a powerful tool for the evaluation of publically available health information. It is important to note that despite the strengths these tools, it is apparent from our results that any examination of a data set obtained in this manner will still require substantial preprocessing. For example, only 22% of the initial Bing search, 17% of the initial Ask.com search, and 32% of the Yahoo search were ultimately clustered in relation to our original search term (524, 363, and 749 documents, respectively). Furthermore, our results also show that the search engine selected plays an important role in the final composition of the data set as well. Results from all three searches yielded conceptually valid clusters, albeit with varying degrees of conceptual organization ranging from specific drugs (Bing) to more general classes of drugs (Ask.com and Yahoo). Irrespective of the web search engine differences, our findings clearly show that a substantial subset of information on antidepressant continuation syndrome is now readily available to the lay public.

## REFERENCES

1. Warner, C.H., et al., Antidepressant discontinuation syndrome. *Am Fam Physician*, 2006. 74(3): p. 449-56.
2. Blier, P. and P. Tremblay, Physiologic mechanisms underlying the antidepressant discontinuation syndrome. *J Clin Psychiatry*, 2006. 67 Suppl 4: p. 8-13.
3. Olfson, M. and S.C. Marcus, National patterns in antidepressant medication treatment. *Arch Gen Psychiatry*, 2009. 66(8): p. 848-56.

## ACKNOWLEDGMENTS

This project was conducted as part of an online graduate course with a focus on group learning at the University of Louisville. The help of fellow classmates was greatly appreciated, as was the input and support of the instructor, Dr. Patricia Cerrito.

## CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Name: Barry W. Row  
 Phone: 502-419-8962  
 E-mail: row.barry@gmail.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.