Capstone Project
Module 8 Portfolio Project

Option #1: Capstone Project—Final Report and Slide Presentation: U.S. Organization

Patrick Hauer

MIS581 Capstone Class

Colorado State University – Global Campus

Dr. Steven Chung

March 8, 2020

Accumen Inc.

- Founded in 2011

-Headquarter in Phoenix, AZ

-Private company owned by Arsenal Capital Partners

-Consulting company in the Healthcare Industry
- "Partnering to Deliver Lab Excellence"

-Acquired 4 other companies
- Chi Solutions, Halfpenny Technologies, 3DR Laboratories, and Anemia Management

This company was chosen since the author began working here in December 2019. The author's role is that of a Project Analyst. The duties include internal reporting on the department as a whole as well as doing analysis for individual client engagements. The author has no previous experience in the Healthcare field or working with medical laboratories. One of the main reasons the author got the job is because of this MDA program and the specialization chosen of Population Health. The U.S. organization chosen for this Capstone Project is Accumen Inc. This company was chosen since the author began working here as a Project Analyst in December 2019. Accumen is a consulting company in the Healthcare industry. Accumen works with hospitals and health systems Accumen is owned by a private equity group called Arsenal Capital Partners. Accumen was founded in 2011. Since Acumen's founding it has acquired several other companies including Chi Solutions, Halfpenny Technologies, 3DR Laboratories, and Anemia Management. Accumen is headquartered in Phoenix, Arizona, but also has an office in San Diego, California. In addition, Accumen has employees that work remotely across the country. As this is a private company, it is difficult to find published and/or reliable numbers on the number of employees and revenue of the company. The estimated number of employees is 350. The author works specifically withing a group of about 10 individuals that focus on consulting with medical laboratories. This group advises on

lab and space design, outreach and consolidation of testing, as well as other aspects to increases efficiencies within the lab and lower the cost per test. Accumen's offerings can best be described on their website:
"strategic solutions, services and technology that deliver sustainable performance improvements in operations, clinical services, and data management. These offerings include lab and imaging transformation, consulting, supply chain optimization, lab outreach, patient blood utilization, test utilization, anemia management and clinical data exchange" (Accumen, 2020).

Dataset

-From Definitive Healthcare (DHC)

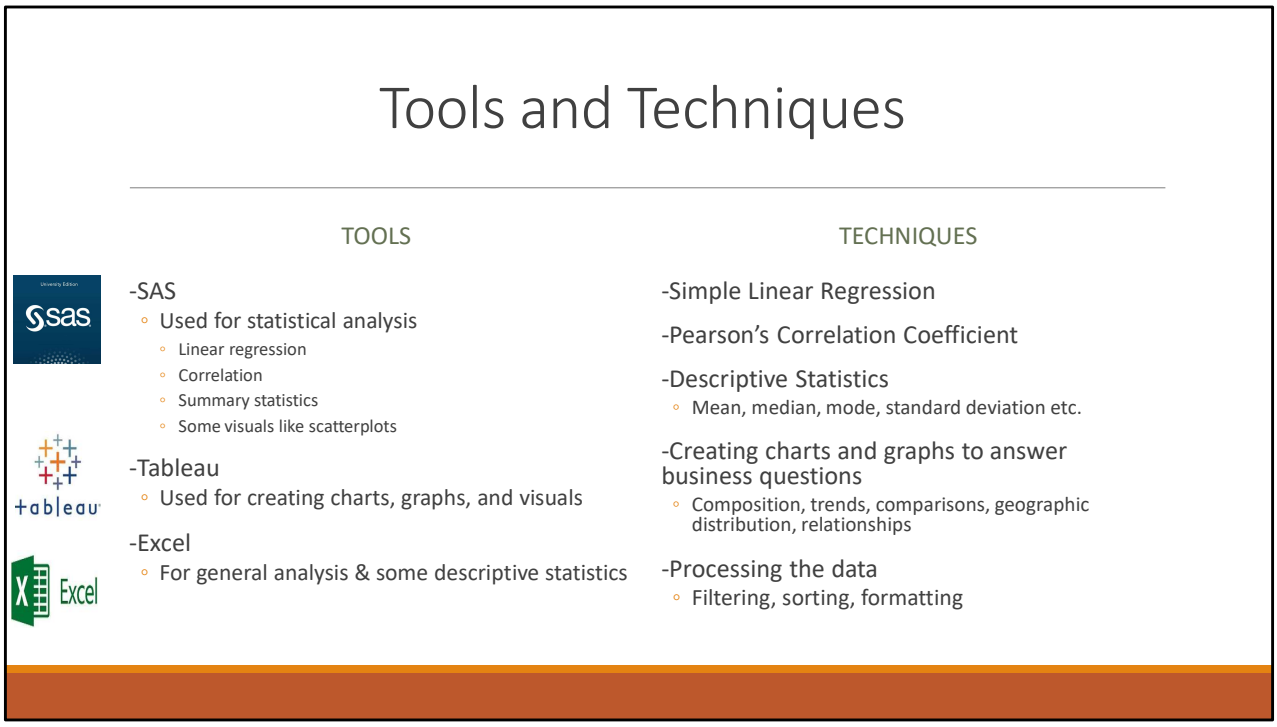
-Mainly focusing on size and financial measures of health systems in the US

- For the purpose of identifying potential clients for business development.

-Required processing

- i.e. filtering, sorting, removing of some nulls, and outliers

The dataset being using for this project largely comes from Definitive Healthcare (DHC). Definitive Healthcare is a database online that requires paid access (by monthly subscription). DHC is a “provider of data and intelligence on hospitals, physicians and other healthcare providers” (DHC, 2020). With the scope of this project being centered around business development, using source data from DHC makes sense. Data is very valuable. If used properly it can lead to many great insights, and data-based decisions can lead to success. Data is a lot like oil. It has inherent value, but in its raw form it is not as worthwhile as when refined. Both raw data and oil need to be processed so that they can be used. For these reasons, data can carry a cost to procure and be difficult to work with. It was no different for this project. It is always hard trying to find what data to use and then deciding how to use what one has. The dataset is never perfect, but “enough is as good as a feast”. The dataset had been analyzed in previous modules and had to be further refined to remove some outliers that were skewing the dataset. The variables included in the data set include: Health System Name, Number of Employees, Total Assets, Net Income, Total Operating Expenses, Total Revenues, State, Number of Hospitals, Net Patient Revenue, Estimated Laboratory Spend, Estimated Blood spend, Number of Staffed Beds, and Number of Discharges.



The tools used for this project were SAS University Edition, Tableau, and Excel. R was considered to be used but its capabilities are largely redundant with SAS and Excel. The format of the file is Excel, and it is a very quick way to look at and/or double check different measures like averages and other descriptive statistics. It also has the ability to make charts that look professional and do other calculations like linear regression. It is important to use attractive visuals in today's presentations (Kemp, Santana, & Hude, 2018).

The techniques used for this project were simple linear regression, Pearson's correlation coefficient, and general descriptive statistics (mean, median, mode, variance, etc.). There are 5 main types of business questions that are usually helpful to answer. These questions relate to composition, trends, comparisons, geographic distribution, and relationships. Tableau is a very intuitive and powerful tool to help answer these questions with visuals. A composition question is usually answered with a pie chart. It can show how much of the whole is made up of by a part. A line graph is typically used for showing trends and allows someone to quickly tell how much something is changing over time and whether it is generally increasing or decreasing. Unfortunately, there is not any variables related to time in the dataset, so there will be no trends shown over time. A comparison question will usually use a bar chart to compare values across different categories. Geographic distribution can

show values across a map. It could show revenue levels by state, region, or city. Relationships are typically shown with a scatterplot. It shows how variables move in relation to one another.

Using SAS University Edition, Tableau, and Excel along with descriptive statistics, Pearson's correlation coefficient, and linear regression, the author was able to find a highly correlated variable to lab spend (total revenue) and perform a linear regression equation to predict the lab spend of other health systems going forward.

Security, Privacy, and Ethical Concerns

-Concerns of unsolicited sales attempt to customers

-Concerns of pirating data from a paid service (DHC)

-Concerns that strategic recommendations can ultimately lead to hospitals employees losing their jobs

-Concerns of action being taken based off incorrect data

- i.e. “Garbage in, Garbage out”
- Negative ROI on marketing efforts

There are security, privacy, and ethical concerns involved in any data analytics project (Fisher, & McKendry, 2019). This is especially true when dealing with data from the healthcare industry. For this project, there is no personal/protected health information (PHI) involved which reduces most of the privacy concerns. As far as security, it is important that sensitive materials be encrypted from outsider viewers and protected from competitors. For this project, most of the concerns are ethical in nature. For one, this dataset being used is from DHC which is a paid service. This service is paid for by Accumen and the purpose of the subscription to DHC is for analysis, however, an ethical consideration would be to make sure that all of the data within the DHC is not exported and then uploaded to the internet for others to view for free. This would be unethical and similar to pirating music, movies, or other content. Another ethical consideration is that the hospitals and health systems being analyzed may not want to be contacted by Accumen for lab partnerships. Many people receive enough solicitation as it is and it is an inconvenience to be contacted by so many firms trying to sell products, services, and other offerings. However, the remedy for this is that Accumen engages in mutual beneficial relationships with clients. Also, through this Capstone project, the one of the outcomes should be that Accumen does not reach out to as many companies that would not be interested in their services. Some of the deals that Accumen engages in are profit sharing in nature

where the only payment will be any incremental profits gained from the carrying out Accumen's strategic direction. Regardless of the profile, accepted industry standards such as Six Sigma and LEAN project management are beneficial (Chandra, et al., 2018).

A major ethical concern, by the author, is that the data source is inaccurate, or the methodologies used are flawed in some way and that the company then takes the results and attempts to carry them out but only realizes a poor outcome. This goes along with the common phrase of "garbage in, garbage out". If the data is of poor quality to begin with then any results derived from it will be at best useless and at worst harmful. Especially when there is any processing of the data involved it is important to do quality checks and make sure that all the values are still corresponding to the correct variables. Perhaps a profile is created for the types of health systems that Accumen should market to and the company is ultimately harmed by expending resources to contact health systems that fit the profile and the generates negative returns from the initiative.

Descriptive Statistics

Variable	Mean	Std Dev	Minimum	Maximum	Median	N	Std Error	Variance	Mode	Range	Sum	Coeff of Variation	Lower Quartile	Upper Quartile
Number of Employees	5,980	9,006	13	113,835	3,121	846	310	81,115,421	158	113,822	5,059,432	151	1,355	7,242
Total Assets	$1,998,822,019	$9,438,860,566	-$553,478,452	$191,828,209,144	$703,339,102	843	$325,091,626	$89,092,089,000,000,000,000	$1	$192,381,687,596	$1,685,007,000,000	472	$210,117,865	$1,923,012,872
Net Income	$109,301,189	$557,505,218	-$2,905,261,759	$11,747,323,635	$25,747,002	848	$19,144,808	$310,812,070,000,000,000	.	$14,652,585,394	$92,687,408,619	510	$746,276	$107,126,492
Total Operating Expenses	$1,317,844,084	$2,137,580,898	$5,379,507	$28,682,102,990	$630,509,618	848	$73,404,829	$4,569,252,100,000,000,000	.	$28,676,723,483	$1,117,531,800,000	162	$254,439,989	$1,556,614,552
Total Revenues	$1,437,514,830	$2,413,281,500	-$206,210,915	$30,264,960,475	$677,707,188	848	$82,872,427	$5,823,927,600,000,000,000	$258,686,190	$30,471,171,390	$1,219,012,600,000	168	$256,391,180	$1,689,675,037
# of Hospitals	8	14	1	197	4	848	0	206	2	196	6,429	189	2	7
Net Patient Revenue	$1,341,549,462	$2,281,718,355	-$11,406,479	$28,424,917,132	$621,392,752	848	$78,354,530	$5,206,238,700,000,000,000	$201,269,324	$28,436,323,611	$1,137,633,900,000	170	$234,899,059	$1,579,648,509
Est. lab spend	$46,954,231	$79,860,142	-$399,227	$994,872,100	$21,748,746	848	$2,742,409	$6,377,642,300,000,000	$7,044,426	$995,271,327	$39,817,188,047	170	$8,221,467	$55,287,698
Est. blood spend	$6,707,747	$11,408,592	-$57,032	$142,124,586	$3,106,964	848	$391,773	$130,155,970,000,000	$1,006,347	$142,181,618	$5,688,169,718	170	$1,174,496	$7,898,243
# of Staffed Beds	1,089	1,822	4	21,970	552	848	63	3,319,576	25	21,966	923,743	167	252	1,225
# of Discharges	43,558	75,762	33	962,980	21,588	847	2,603	5,739,897,986	104	962,947	36,893,470	174	8,583	48,586

First step of the analysis was looking at some descriptive statistics of the dataset in SAS . The slide looks are descriptive statistics for the numeric variables in the dataset. Some of the ranges and averages are interesting. There is quite a large range within some of the variables and it is obvious that there is a lot of money at play here. This is a big industry and market, and there are lots of opportunities to gain revenue.

Pearson's Correlation Coefficient

Pearson Correlation Coefficients Prob > |r| under H0: Rho=0										
Variables	Estimated blood spend	Net Patient Revenue	Total Revenues	Total Operating Expenses	Number of Employees	Number of Discharges	Number of Staffed Beds	Number of Hospitals	Net Income	Total Assets
Correlation to Lab Spend	1	1	0.99724	0.96879	0.92786	0.89852	0.85114	0.68977	0.58037	0.35552
P-value	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001

First thing is to look at all the variables to determine what is the most highly correlated with the variable of lab spend since that is what the project is most interest in. The estimated blood spend, and Net Patient revenue will be disregarded as the original estimations for lab spend and blood spend came from an assumption based off net patient revenue. Given that, it appears that total revenues and total operating expenses are the next highest correlations, and all the correlations meet the threshold for significance (p-value is less that alpha). With this direction, we can use these variables along with regression and try to get to the point that of predicting lab spend based off this more readily available information.

Hypotheses

-**Alternative Hypothesis** (HA): There will be a significant prediction of laboratory spend (y variable) by total revenue of the health system (x variable)

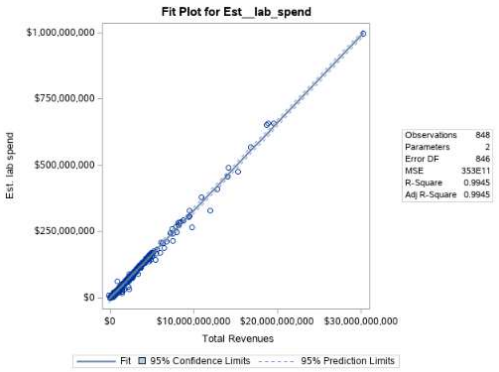
-**Null Hypothesis** (HO): There will be no significant prediction of laboratory spend (y variable) by total revenue of the health system (x variable)

For this project, the goal is to figure out the best way to estimate the laboratory spend of a health system. Since Accumen consults on improvements for a medical lab, it is logical to want to pursue those health systems that have large expenditures related to labs as these would be the places that could benefit the most from Accumen's services. The issue is that lab spend information is hard to find. Thus, linear regression can be used to create a formula to predict lab spend given some predictor variable. The predictor variable will be discovered using Pearson's correlation coefficient. The confidence interval will be at 95%. The hypotheses for this linear regression are as follows:
Alternative Hypothesis (HA): There will be a significant prediction of laboratory spend (y variable) by total revenue of the health system (x variable)
Null Hypothesis (HO): There will be no significant prediction of laboratory spend (y variable) by total revenue of the health system (x variable)

Linear Regression - SAS

Parameter Estimates								
Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Standardized Est.	Squared Partial Corr. Type 1	Squared Partial Corr. Type 2
Intercept	1	-484405	237363	-2.04	0.0416	0	.	.
Total Revenues	1	0.033	8.454E-05	390.36	<.0001	0.99724	0.99448	0.99448

Linear Equation:
Y= .033x – 484,405



Next, linear regression was used to try and discover a way to predict lab spend of a health system. This was performed in SAS (Figure 2), Tableau (Figure 3), and Excel (Appendix Figure 4). All the results were similar. As the p value was less than alpha, the null hypothesis was rejected. This allows the linear equation to be used with more confidence to predict the laboratory spend of health systems. Using Tableau, other visuals were created to help answer questions about the data. These are available in the Appendix. Figure 5 shows a filtered view of estimated lab spend by health system with a bubble chart, and Figure 6 shows estimated lab spend by state. It is difficult creating visuals for the data because it is hard to show all the different health systems (approximately 850).

Linear Regression - Tableau

Describe Trend Model

Trend Lines Model

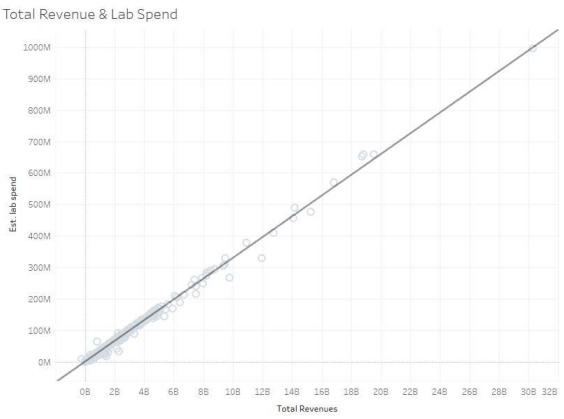
A linear trend model is computed for Est. lab spend given Total Revenues. The model may be significant at p <= 0.05.

Model formula:	(Total Revenues + intercept)
Number of modeled observations:	847
Number of filtered observations:	0
Model degrees of freedom:	2
Residual degrees of freedom (DF):	845
SSE (sum squared error):	2.98237e+16
MSE (mean squared error):	3.52944e+13
R-Squared:	0.994477
Standard error:	5.94091e+06
p-value (significance):	< 0.0001

Individual trend lines:

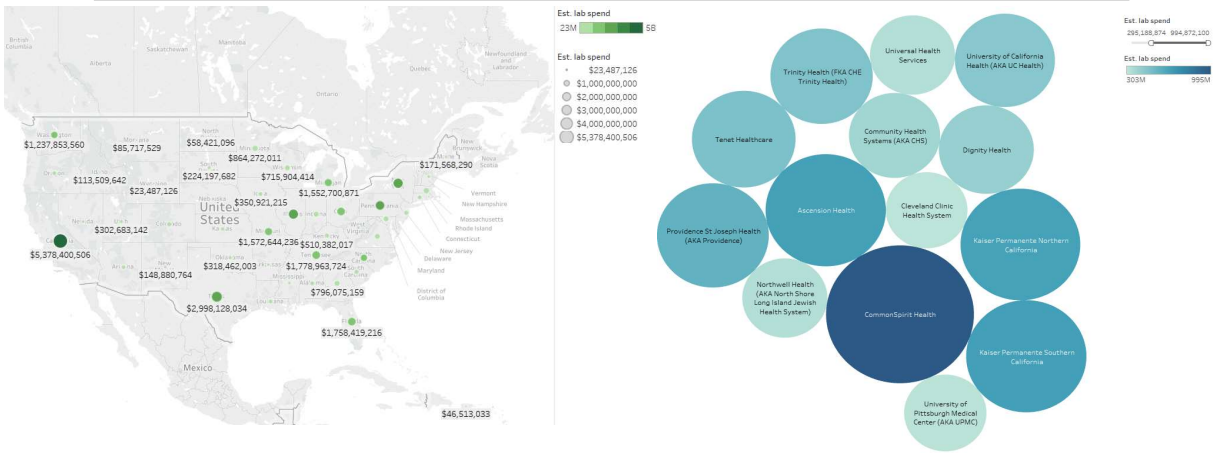
Panes		Line		Coefficients				
Row	**Column**	**p-value**	**DF**	**Term**	**Value**	**StdErr**	**t-value**	**p-value**
Est. lab spend	Total Revenues	< 0.0001	845	Total Revenues	0.0330002	8.46e-05	390.079	< 0.0001
				intercept	-482868	237672	-2.03166	0.0425008

Linear Equation: y=.033x – 482,868



The results in SAS were verified with Tableau and SAS. They are nearly identical. The only difference is in SAS there appears to be 848 observations whereas in Tableau there were 847. This is immaterial, however, since the equations are nearly identical. Also, the level of significance is below the threshold (alpha). This allows the null to be rejected. As the saying goes “if the p-value is low, the null must go”. There was also a regression run in Excel which produced the same results as SAS.

Analysis Outcomes and Explanation



Using Tableau, other visuals were created to help answer questions about the data. These are available in the Appendix. Figure 5 shows a filtered view of estimated lab spend by health system with a bubble chart, and Figure 6 shows estimated lab spend by state. It is difficult creating visuals for the data because it is hard to show all the different health systems (approximately 850).

Conclusion

-The null was rejected
- given this sample, the position is rejected that there is no significance prediction of laboratory spend by total revenue

-This is very helpful since total revenue is a relatively easy metric to find whereas laboratory spend figures per health system are difficult to find
- With this new linear equation, Accumen can estimate lab spend of a health system

--This will provide direction as to what health systems specifically Accumen should reach out to as potential clients
- will reduce wasted resources spent on marketing and advertising for Accumen

-New info. can also serve as a vetting process for new clients

-Provides Accumen direction for further research to be done going forward on business development protocols

This study revealed several points that will make it easier for Accumen to predict and find health systems with high levels of laboratory expenditures. The balance between benefits and costs must be considered (Grover, et al., 2018). Given Accumen's business, it is important to find health systems that have high levels of lab spend as these would make for the best potential clients. The lab spend figure is not widely reported and hard to find. Accumen has been struggling to know which health systems to reach out to. By using Pearson's correlation coefficient, it was discovered that total revenue has the highest correlation to the estimated lab spend. Then using linear regression to create a formula to predict lab spend in the future. Since the null hypothesis was rejected, the linear equation can be used for lab spend estimates. The total revenue is easier to uncover than the laboratory expenditure total. This will help direct further analysis and begin to create a profile of Accumen's target demo. There are other benefits to this analysis as well. Not only will this help Accumen with business development and targeting of potential clients, but it also can serve as a vetting process for onboarding clients and will provide direction for further analysis on client profiles to be done going forward.

References

Accumen. (2020) Accumen information. Retrieved from https://accumen.com/

Chandra, A., Finkelstein, A., Sacarny, A. and Syverson, C. (2016). Health care exceptionalism? Performance and allocation in the US health care sector. *American Economic Review*, 106(8), 2110–2144. https://doi.org/10.1257/aer.20151080

Chi Solutions. (2016). Accumen, Inc. acquires chi solutions for continued growth. Retrieved from https://accumen.com/accumen-inc-acquires-chi-solutions-for-continued-growth/

Halfpenny Technologies. (2019). Accumen acquires halfpenny technologies

CSU Global. (2019). MIS581 Interactive Lecture.

DHC. (2020). Definitive health care. Retrieved from https://www.definitivehc.com/

Fisher, J. and McKendry, K. (2019). Health care data Privacy trends: Risks and costs of third-party data collection and cyberattacks. *Journal of Health Care Compliance*, 21(2), 41–44. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=her&AN=135700650&site=eds-

References

Github. (2019). Retrieved from https://github.com/hauerpc1/capstone

Grover, V., Chiang, R. H. L., Liang, T.-P., & Zhang, D. (2018). Creating strategic business value from big data analytics: A research framework. *Journal of Management Information Systems*, 35(2), 388–423. https://doi.org/10.1080/07421222.2018.1451951

Oh, J. Kang, J. & Park, J. (2019). A study of smart healthcare service model based on cloud platform: Focus on small and medium sized hospitals. *Medico-Legal Update*, 19(2), 434–440. https://doi.org/10.5958/0974-1283.2019.00216.0

Kemp, K., Santana, M., & Hude, Q. (2018). Interactive data visualization of patient experience and inpatient datasets using tableau desktop. *International Journal of Population Data Science*, (4). https://doi.org/10.23889/ijpds.v3i4.632

Rogers, E. M. (1966). *Physics for the inquiring mind.* Princeton, NJ: Princeton University Press.

Sanders, B. (2020). Data literacy and data analytics. *RocktheViz.* Retrieved from https://rockthevizcomm.com/feature/data-literacy-and-data-analytics/

Snyder, J. (2019). Data cleansing: An omission from data analytics coursework. *Information Systems Education Journal*, 17(6), 22–29. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=EJ1224578&site=eds-live

Yoo, J. (2019). Housekeeping tips to mitigate data security risk. *Journal of Accountancy*, 228(3), 1–4. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=138492949&site=eds-live