Data-intensive Undergraduate Research Project Informs to Advance Healthcare Analytics

M. J. D'Souza, D. Wentzien, R. Bautista, J. Santana, M. Skivers, S. Stotts, F. Fiedler*

Undergraduate Research Center for Analytics, Talent, and Success, Wesley College, Dover, DE 19901, USA
malcolm.dsouza@wesley.edu (MJD)

***Abstract*—The overarching framework for incorporating informatics into the Wesley College (Wesley) undergraduate curriculum was to teach emerging information technologies that prepared undergraduates for complex high-demand work environments. Federal and State support helped implement Wesley's undergraduate Informatics Certificate and Minor programs. Both programs require project-based coursework in Applied Statistics, SAS Programming, and Geo-spatial Analysis (ArcGIS).**

In 2015, the *State of Obesity* listed the obesity ranges for all 50 US States to be between 21-36%. Yet, the Center for Disease Control and Prevention (CDC) mortality records show significantly lower obesity-related death-rates for states with very high obesity-rates. This study highlights the disparities in the reported obesity-related death-rates (specified by an ICD-10 E66 diagnosis code) and the obesity-rate percentages recorded for all 50 US States. Using CDC mortality-rate data, the available obesity-rate information, and ArcGIS, we created choropleth maps for all US States. Visual and statistical analysis shows considerable disparities in the obesity-related death-rate record-keeping amongst the 50 US States. For example, in 2015, Vermont with the sixth lowest obesity-rate had the highest reported obesity-related death-rate. In contrast, Alabama had the fifth highest adult obesity-rate in the nation, yet, it had a very low age-adjusted mortality-rate. Such disparities make comparative analysis difficult.

Keywords—obesity; mortality; ICD-10; ArcGIS; SAS; CDC; Wesley College; INBRE; EPSCoR

I. INTRODUCTION

Pioneering approaches using resource-based learning for competency in information literacy and for the development of critical qualitative and quantitative thinking skills, are hallmarks to building the proverbial bridge between the basic sciences and the medical research enterprise [1]. Furthermore, when undergraduate research gets integrated in the 1st year of the College curricula, it promotes the value of self-regulated learning, cognition, personal development, and career planning [2]-[6]. Additionally, epidemiology-based research projects gravitate to closing the diversity gap as they promote essential proficiencies in ethical reasoning and navigational techniques for survival in today's digital and political landscape [7]-[9].

To accommodate the profound technological advancement in workforce demands, Delaware's minority-serving Wesley College (Wesley) strengthened diverse academic equity by implanting cheminformatics, chemometrics, data-wrangling, predictive-analytics, and geo-spatial analysis as course-embedded experiences [5], [6], [10]-[31]. Constructed around the best practice of multi-tiered mentoring, Wesley's redesigned (2014) progressive liberal-arts core-curriculum model scaffolds undergraduate course content with practical and realistic data-science research materials [5], [6], [10]-[31].

In Delaware, large State-wide, National Institutes of Health (NIH)-National Institute of General Medical Sciences (NIGMS)-IDeA Networks of Biomedical Research Excellence (INBRE), and National Science Foundation (NSF)-Established Program to Stimulate Competitive Research (EPSCoR), consortia grants provided the financial stimulus for cornerstone data-science research projects [5], [6], [10]-[31]. Through federal support and in conjunction with its liberal-arts core-curriculum revision, Wesley College instituted undergraduate Informatics Certificate and Minor programs [14], [15], [17]. Moreover, in 2017, the College launched the Undergraduate Research Center of Analytics, Talent, and Success (UR-CATS). The UR-CATS team in collaboration with the Office of Academic Affairs, routinely provides professional development training to build dynamic mentor-mentee collaborative partnerships.

The Informatics Minor and Certificate programs require three project-based courses: Applied Statistics, SAS Programming (developed by the SAS Institute), and Geo-spatial Analysis using Geographic Information System (GIS) Mapping Technology (ESRI ArcGIS). The Informatics Certificate is a biological chemistry and mathematics major requirement [14], [17]. However, both certificate and minor programs are available [14] to all majors (in core level-3).

Undergraduate participation in data-science research projects has removed intra-institutional barriers encumbering access to data-resources and expertise for in-depth investigations. Using patient package inserts, Wesley researchers designed and developed a commercial drug-data training tool [26]-[31] and an online cancer drug database [22], [24], [25]. With real-time chemical-use data from Delaware's farmers, students built a smartphone fertilizer app [18]. Furthermore, to streamline procurement documents and hazardous-waste handling workflows, faculty-student engagement allowed for the use of an online platform, resulting in major annual institutional savings [13].

While many Wesley students choose projects directly related to their majors, others tackle unrelated questions with real and direct applications. For example, SAS in-class projects have related Delaware student success outcomes (persistence and degree completion) to demographics, secondary degree zip code, support structures, success in HS science/college gatekeeper courses, and campus engagement. Alternatively, some GIS projects using historical digital elevation models, land-use/land-cover files, and aerial images, identified areas of coastal forest along Delaware's St. Jones River that were vulnerable to inundation.

Nationally, chronic obesity has impacted all segments of the total US population. Its repercussions coalesce around conditions that have negative connotations in public health [10], [12], [20], [32]-[49]. The diagnosis of obesity is classified according to body mass index (BMI) calculations [20], [32]-[44]. Although BMI is a strong predictor of overall mortality [41], its relevance in predicting disease-specific associations is, at best, marginal [20], [34]-[40], [42]-[44].

However, since obesity is significantly affecting the quality of life issues [32]-[49], big-data approaches [45]-[57] are utilized to seek corrective solutions. Available obesity prevalence data encompasses interactive databases that report weight/heights, demographic, gender, zip-code, nutritional, and physical activity information [45]-[49]. In addition, the national Centers for Disease Control and Prevention Wide-Ranging Online Data for Epidemiologic Research (CDC WONDER) repository [45] also counts obesity as the underlying cause of death using ICD-10 (10th revision of the International Statistical Classification of Diseases and Related Health Problems) diagnosis coding [10], [12], [45], [58]-[62]. Documentation of the ICD-10 codes on vital records is not without controversy as such quality-assurance based programs are not implemented by the government but by third-party payers [58]-[62]. Despite the differences surrounding the complexity and constraints with the ICD-10 coding, it is definitively used to denote medical significance for the overweight and obesity categories (ICD-10 E66) within the outpatient or inpatient records [62].

Utilizing the CDC WONDER vital records, the US census database, and the Robert Wood Johnson Foundation's *State of Obesity* records [45]-[47], Wesley College students through empirical affiliations definitively showed that Delaware's mortality-rates are heavily influenced by the State's obesity rates and are correlated to its socio-economic indicators [12], [17]. A second retrospective observational study [10] analyzed the 1999-2016 national death records alongside associated ICD-10 (E66) obesity-related diagnostic coding. It found convincing evidence to associate the adverse consequences of obesity-related conditions with the mortality-rates in all segments of the US population. In summary, Wesley data-science project outcomes methodically investigated public interest projects and allowed for coordinated interventions in student well-being and student success [11], [15], [17], [19], [20], [23], [63]-[65].

For analyzing and studying national trends, we are strong proponents in the using of the CDC WONDER records and their affiliated ICD-10 E66 diagnostic coding. However, we believe that some US States significantly under-code and some over-code. The resulting discrepancies are a misrepresentation of facts for healthcare resources, as they establish a high level of false utilization patterns [58]-[62]. To effect corrective action and prevent recurrence, we present the example of five Southern States and five NERIC (North East Regional IDeA Conference) States, where vital record inconsistences can be plainly attributable to errors associated with ICD-10 coding.

In this project, we first outline the digital and communication training that builds fundamental knowledge for collecting, sorting, storing, and analyzing large datasets. Then to validate our coding and reporting comparability concerns, with GIS and SAS tools, we assess the public ICD-10 E66-diagnosis (with any mention of obesity) vital records [45]-[49] for each US State.

II. METHODS

A. Design of the Wesley data-science training program

To increase comprehension of the scientific method, first semester STEM freshmen are enrolled in a scientific process course that includes hands-on science inquiry [14], [17], [19]. A second computer applications course is available to expose them to MATLAB and MAPLE for use in the exploration and discovery phase. In the second and third semesters, the science course load [14], [17], [19], [23] includes organic chemistry coupled with multiple directed research opportunities that explore chemical kinetics, spectroscopic analysis, and the use of quantitative information to correlate chemical activity with molecular structure. In years 2 and 3, advanced approaches to multivariate statistical procedures including hypothesis testing, data-mining techniques, mathematical modeling, and data visualization are emphasized in the advanced statistics, SAS-programming, and geo-spatial courses [5], [6], [10]-[14], [17]-[19]. Finally, to capture the research process in the senior year, a thesis experience allows for independence in qualitative, quantitative, and communication training [5], [6], [14].

B. Teaching meaningful Informatics tools and techniques

Students from a wide range of majors have enrolled in the Wesley Informatics Minor and Certification courses [14]. The majors include biology, biological chemistry, environmental science, environmental policy, medical technology, mathematics, nursing, political science, elementary and physical education, psychology, business, accounting, marketing, liberal studies, history, kinesiology, and pre-professional physical therapy. Graduate students in environmental science, nursing, business, and the Masters-teaching program can also seek informatics certification.

The 200-level Statistics course topics include sampling, frequency distributions, histograms, probability distributions, confidence intervals, hypothesis testing, correlation, regression analysis, chi-square test for independence, and non-parametric hypothesis testing. Assigned projects allow for actual experience in collecting, analyzing, and interpreting data to answer a research question. Data analyses are usually done with manual row and column manipulations in Excel.

For easy-to-follow programming code and to retrieve, manage, and analyze big data, we integrate the Statistical Analysis System software (SAS) suite. SAS is heavily utilized in the finance and life science areas, two dominant industries in Delaware. In the 300-level SAS programming course, students work in teams and use shared data. They begin with data dictionaries for identifying variable names and content. They are introduced to different ways of filtering observations using IF or WHERE clauses and this methodology is linked to filtering in Excel, so students can make connections to their prior Statistics course projects. They are taught to MERGE tables by a (particular) variable, which is possible in Excel, but our students are (usually) unaware of its advanced features. As most non-STEM majors do not have programming practice, we demonstrate the frustrating and time-consuming debugging-of-code experience and the process of utilizing a subset of data for the quick identification and elimination of errors. Furthermore, we address the issue of duplicate observations and missing values and instruct in the design of default handlers for their incorporation in calculations.

For a complete and robust toolkit for analyzing big data, we believe that a working knowledge of geospatial analysis is essential. At Wesley, two courses focus on spatial analysis of large data sets, a 300-level Introduction to Geographic Information Systems (GIS) course and a 400-level Spatial Analysis course. Both utilize ESRI ARCMap and the College will fully transition to ESRI ARCPro by spring 2019. The 300-level GIS course is lab intensive, but the instructional approach is structured to accommodate the skillset of the diverse number of majors. Course participants are also required to read and interpret scientific literature regarding the use of spatial analysis of large datasets for group projects to benefit local organizations. The instructional approach for the 400-level Spatial Analysis is much less structured and requires an independent or self-selected group geospatial analysis project. In the two GIS-courses, course instruction is focused on spatial data structure, map design, data queries and joins, data collection, spatial analysis of vector data, and spatial analysis of raster data and we incorporate labs to focus on each skill. At the 400-level, students receive a guided assignment that requires them to download data from the US Census Bureau (American Fact Finder), format the data for input into ESRI's ARCMap, and then, organize and display the data in a meaningful way. This assignment lays the foundation for future independent research projects. To investigate and solve real world problems, the GIS labs and independent research projects rely on high-quality and high-resolution public data-sets from local, county, state, and federal websites.

C. *Analysis of the CDC WONDER vital records*

The CDC WONDER database [45] provides crude and age-adjusted death statistics (per 100,000 US persons). The 95% confidence intervals and standard errors for underlying cause-of-death rates (4-digit ICD code or group of codes) can be obtained by place of death, age, gender, ethnicity, and year. Data query and retrieval with SQL syntax is simple but tedious and can include regression and logistic regression. However, merging data entails a long list of copying and pasting similar code snippets. Hence to create a cleaner code, simple %MACRO programming to loop over the years and %INCLUDE subprograms with the year as a parameter is used. To provide more granular data and to pinpoint inconsistencies with ICD-10 coder interpretation and documentation, the CDC WONDER information was first downloaded into Excel and then imported into ARCMap. The choropleth maps allowed for visualization of the spatial origination as color gradients were matched by scale to give proper comparisons between states.

III. RESULTS AND DISCUSSION

The 1999, 2008, and 2015 illustrations shown in Figures 1 and 2 clearly exhibit contradictions where each state's obesity-rate (Figure 1) burden [45]-[48] is disproportional to the state's obesity-related age-adjusted mortality-rate (Figure 2), that was noted by an ICD-10 vital record code, where obesity was a contributory factor in death [10], [12], [17], [45].

Since 1999, obesity prevalence in the Southern States of Mississippi (MS), Alabama (AL), West Virginia (WV), Louisiana (LA), and Arkansas (AR) were consistently grouped within the top ten fattest-state listings [47]. Furthermore, in 2015, the *State of Obesity*

Report [47] indicated that LA, MS, AR, WV, and AL, all had >35%+ adults who were officially classified as

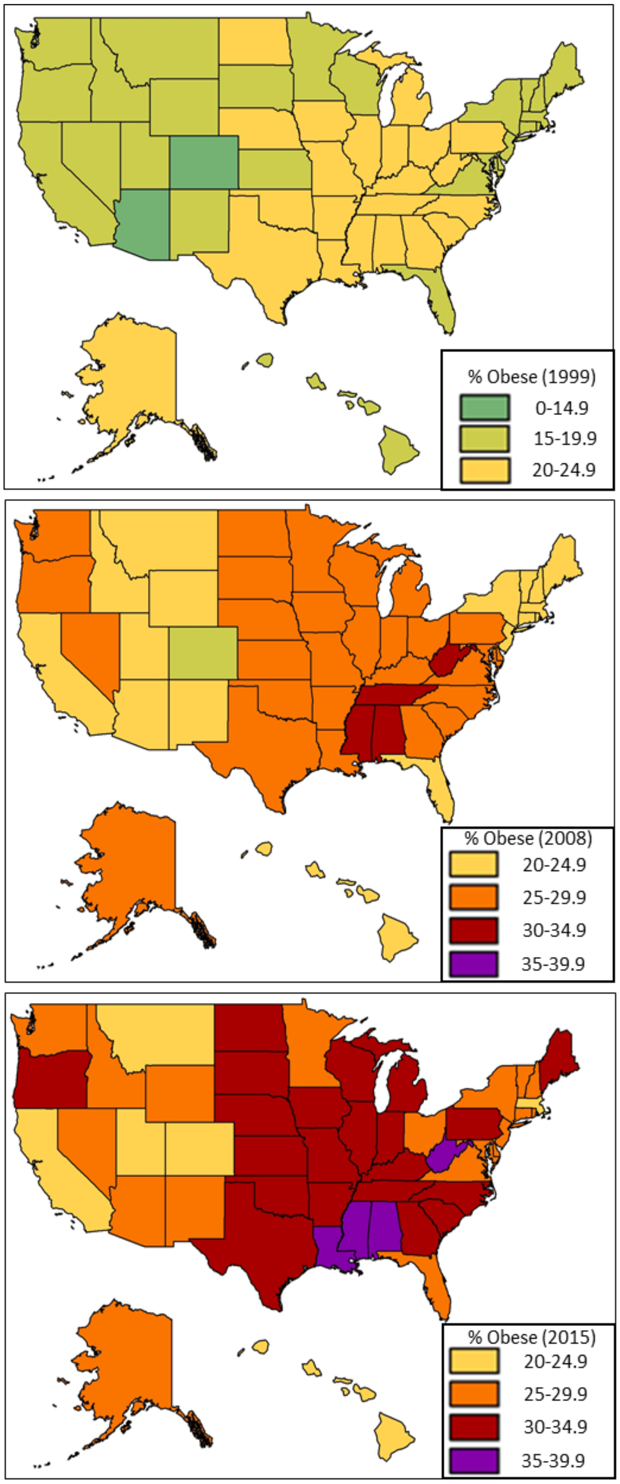

Figure 1. Data from *State of Obesity* was used to create a map of the United States with the percent of the population classified as obese based on BMI for 1999, 2008, and 2015.

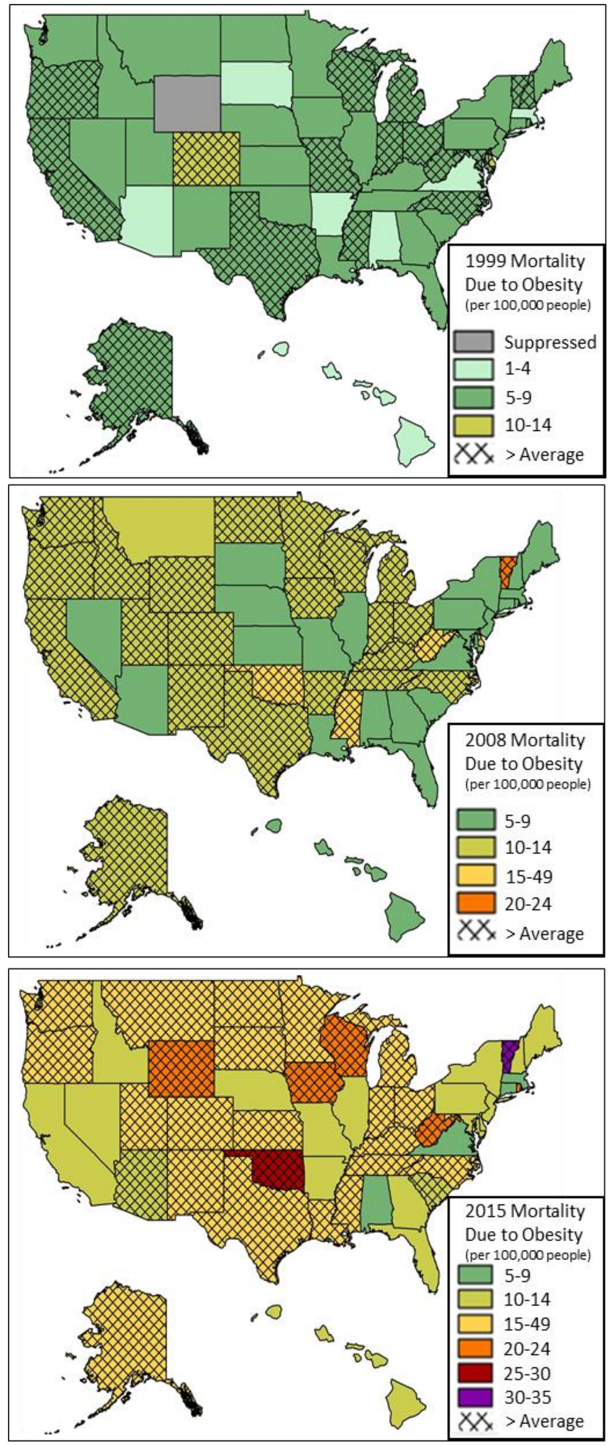

Figure 2. Obesity-related age-adjusted mortality-rates from CDC WONDER were used to create a map of mortality due to obesity per 100,000 people for the United States for 1999, 2008, and 2015.

being obese (with a >30+ BMI reading) and who were residing within their state populations. At the same time, health surveys [49] routinely showed that MS, AL, WV, AR, and LA adults, had one of the lowest compulsions for meeting the basic (federally suggested) physical activity exercise guidelines. Using color gradients to generate quantity maps, Figure 3 exhibits that healthcare access and healthcare quality [47]-[49] are also problematic in these five southern states. However, the age-adjusted ICD-10 mortality codes where the burden of obesity as a contributing mortality factor (Figure 2) is distinctly underestimated in the CDC WONDER vital records [45]. In fact, in 2015, AL is shown to have one of the smallest obesity-related age-adjusted death-rates (Figure 2).

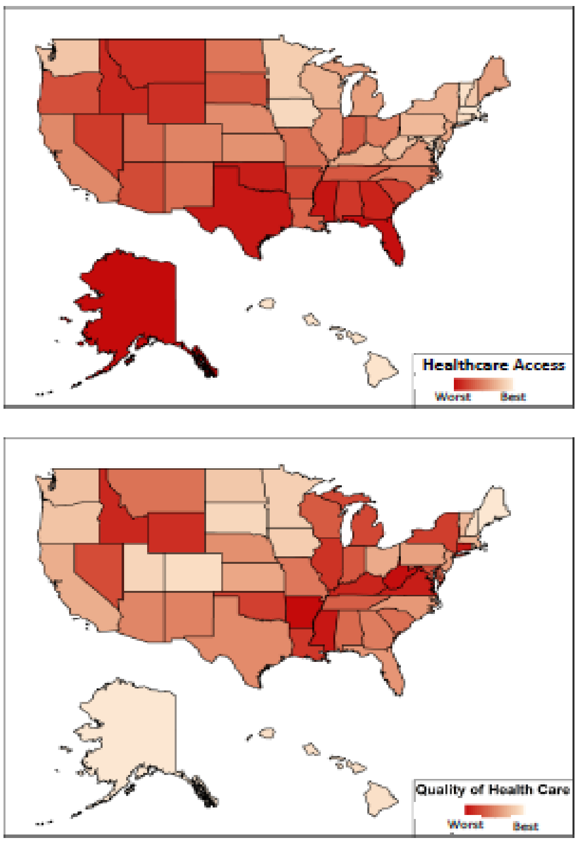

Figure 3. The top map shows 2015 healthcare access, and the bottom map shows 2015 healthcare quality, with red being the worst and lighter-shade of colors being the best.

To further explore the ICD-10 obesity-related reporting discrepancies, we evaluated the obesity-rates and obesity-related death records (Figures 4 & 5, and Tables 1 & 2) [10], [12], [17], [45], [47]-[49] for the five NERIC States (Delaware, Maine, New Hampshire, Rhode Island, and Vermont). The lighter color gradients in Figure 3 imply healthcare access and healthcare quality [47]-[49] are reasonably good in Vermont (VT), Maine (ME), Delaware (DE), New Hampshire (NH) and Rhode Island (RI).

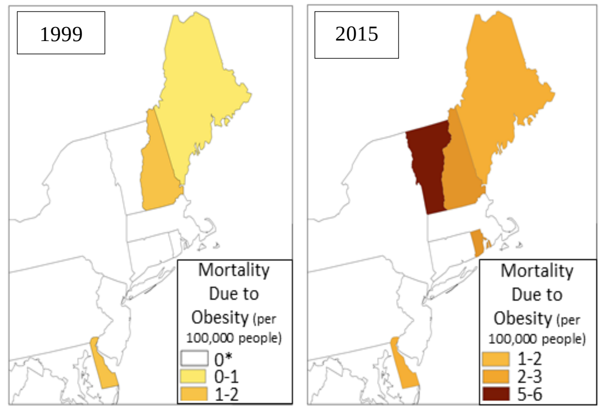

Figure 4. Obesity-related deaths per 100,000 population (not adjusted for age) within the five NERIC States, with the lighter-shade of colors being the smallest amount and the darker shade of colors being the greatest amount of obesity-related deaths. *Only Rhode Island has a value of zero. The other states are not included in this map.

Also, the higher than (national) average physical activity undertaken by adults residing in the five NERIC States demonstrates that the residents more than satisfy the minimum federal guidelines for aerobic and muscle strengthening activities [49]. Understandably in 2015, Vermont (VT) had the sixth lowest obesity rate [47], but the ICD-10 coding for obesity as a secondary cause-of-death record placed VT as having the highest obesity-related mortality-rate (Figures 2, 4 & 5) [45]. In addition, over the sixteen-year (1999-2015) time-period, VT was shown to have an exponential (419%) obesity-related age-adjusted mortality-rate increase (Figure 5 and Table 1) while its obesity rate (only) increased by 47.6% (Table 2).

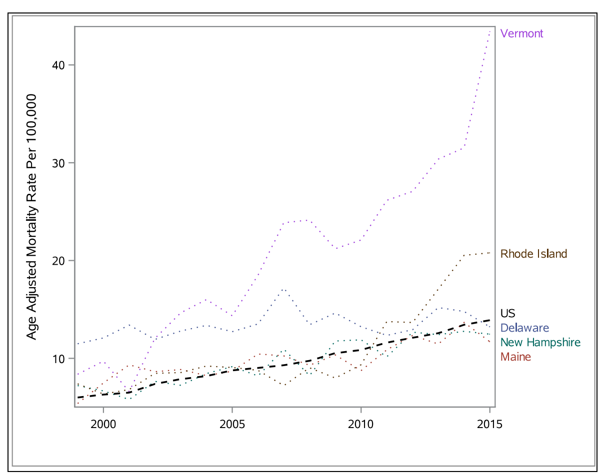

Figure 5. Obesity-related mortality-rates per 100,000 population within the five NERIC States.

Table 1. The 1999-2015 obesity-related mortality-rates of the NERIC States.

Obesity-Related Mortality-Rates			
	1999 Rate	**2015 Rate**	**%Change**
Maine	5.38	11.69	117.29
Vermont	8.37	43.44	419.00
New Hampshire	7.21	12.47	72.95
Rhode Island	7.39	20.79	181.33
Delaware	11.49	13.21	14.97
National Obesity-Related Mortality-Rate	5.99	13.90	132.05

Source: Rates are age-adjusted from CDC Wonder

Table 2. The 1999-2000 and 2015 obesity-rates of the NERIC States.

Obesity Rates			
	1999-2000 Rate	**2015 Rate**	**%Change**
Maine	18.9%	30.0%	58.7%
Vermont	17.0%	25.1%	47.6%
New Hampshire	16.1%	26.3%	63.4%
Rhode Island	16.9%	26.0%	53.8%
Delaware	17.1%	29.7%	73.7%
National Obesity-Rate	30.5%	39.6%	29.8%

Source: Rates are from the stateofobesity.org and the National Center for Health Statistics (NCHS) National Health and Nutrition Examination Survey.

IV. CONCLUSIONS

Wesley College's revised core-curriculumn incorporates informatics and fosters a student-community. This community has crucial quantitative, qualitative, and participatory credentials required to directly move into graduate programs or research and development centers to further explore public interest projects. This undergradute cross-sector collaborative training in information analyses and solutions has distinctly helped address and advance population surveillance projects.

The CDC WONDER database (including its ICD-10 coding data) is an excellent source for historical (national) trend analysis. However, there are a few states where inconsistencies in ICD-10 documentation protocols clearly lead to rampant reporting errors. To help guide the coder to correct (specificity) diagnosis codes, coder professional training should incorporate the use of queries, for determining cause-and-effect associations (documentation elements) that link related conditions where there is a lack of supporting clinical/provider documentation or lack of familiarity.

ACKNOWLEDGEMENTS

Wesley College acknowledges support from an IDeA award from an NIH-NIGMS IDeA award (INBRE grant no. P20GM103446, DE-INBRE program), an NSF-EPSCoR award (Grant No. IIA-1301765, OIA-1757353), an NSF (DUE) S-STEM grant 1355554 (Cannon Scholar program), the NASA Delaware Space Grant Consortium grant program (NNX15AI19H), and the State of Delaware.

REFERENCES

[1] R. L. Houlden, J. B. Raja, C. P. Collier, A. F. Clark, and J. M. Waugh, "Medical Students' Perceptions of an Undergraduate Research Elective," *Medical Teacher*, 26(7), pp. 659-661, 2004.

[2] E. Goldmann, J. H. Stark, F. Kapadia, and M. B. McQueen, "Teaching Epidemiology at the Undergraduate Level: Considerations and Approaches," *American Journal of Epidemiology*, 187(6), pp.1143-1148, 2018.

[3] N. A. Bowman, and J. M. Holmes, " Getting off to a Good Start? First-year Undergraduate Research Experiences and Student Outcomes," *Higher Education*, pp.1-17, 2017.

[4] E. L. Dolan, "Undergraduate Research as Curriculum," *Biochemistry and Molecular Biology Education*, 45(4), pp.293-298, 2017.

[5] M. J. D'Souza, K. L. Curran, P. E. Olsen, A. P. Nwogbaga, and S. Stotts, "Integrative Approach for a Transformative Freshman-level STEM Curriculum," *Journal of College Teaching and Learning*, 13(2), 47-64, 2016.

[6] M. J. D'Souza, P. Dwyer, J. M. Miller, J. Drohan, and B. E. Allison, "Wesley College Ignites Potential with Undergraduate Student Research Program," *Council of Undergraduate Research Quarterly*, 32, pp. 41-46, 2011.

[7] S. Grineski, H. Daniels, T. Collins, D. X. Morales, A. Frederick, and M. Garcia, "The Conundrum of Social Class: Disparities in Publishing among STEM students in Undergraduate Research Programs at a Hispanic Majority Institution," *Science Education*, 102(2), pp.283-303, 2018.

[8] E. Goldmann, J. H. Stark, F. Kapadia, F. and M. B. McQueen, "Teaching Epidemiology at the Undergraduate Level: Considerations and approaches," *American Journal of Epidemiology*, 187(6), pp.1143-1148, 2018.

[9] C. S. Lupi, M. Ward-Peterson, and J. M. Lozano, "The Impact of a Structured Research Scholarship Course on Evidence-Based Medicine Skills among Undergraduate Medical Students," *Medical Science Educator*, pp.1-7, 2018.

[10] M. J. D'Souza, R. C. Bautista, and D. E. Wentzien, "Data Talks: Obesity-Related Influences on US Mortality Rates," *Research in Health Science*, 3(3), pp. 65-78, 2018.

[11] M. J. D'Souza, K. E. Shuman, D. E. Wentzien, D.E. and K. P. Roeske, "Working with the Wesley College Cannon Scholar Program: Improving Retention, Persistence, and Success," *Journal of STEM Education*, 19(1), pp.31-40, 2018.

[12] M. J. D'Souza, D. E. Wentzien, R. C. Bautista, and C. E. Gross, "Investigation Of Obesity-Related Mortality Rates In Delaware," *American Journal of Health Sciences*, 8(1), pp.19-32, 2017.

[13] M. J. D'Souza, K. P. Roeske, and L. S. Neff, "Free Inventory Platform Manages Chemical Risks, Addresses Chemical Accountability, and Measures Cost-Effectiveness," *International Journal of Advances in Science, Engineering and Technology*, 5(3), pp.25-29, 2017.

[14] M. J. D'Souza, E. A. Brandenburg, D. E. Wentzien, R. C. Bautista, A. P. Nwogbaga, R. G. Miller, and P. E. Olsen, "Descriptive and Inferential Statistics in Undergraduate Data Science Research Projects," In *Advances in Statistical Methodologies and Their Application to Real Problems*. Tsukasa Hokimoto, Editor, InTech, pp. 295-315, 2017.

[15] P. M. Dwyer, "Transforming a Core Curriculum--and Minimizing the Battle Scars," *Liberal Education*, 103(1), p.n1., 2017.

[16] M. J. D'Souza, and D. N. Kevill, "Mechanistic Studies of the Solvolyses of Carbamoyl Chlorides and Related Reactions," *International Journal of Molecular Sciences*, 17(1), p.111, doi.org/10.3390/ijms17010111, 2016.

[17] M. J. D'Souza, R. J. Kashmar, K. Hurst, F. Fiedler, C. E. Gross, J. K. Deol, and A. Wilson, "Integrative Biological Chemistry Program Includes the Use of Informatics Tools, GIS And SAS Software Applications," *Contemporary Issues in Education Research*, 8(3), pp.193-214, 2015.

[18] M. J. D'Souza, B. Barile, and A. F. Givens, "Evolution of a Structure-searchable Database into a Prototype for a High-fidelity SmartPhone App for 62 Common Pesticides Used in Delaware," In *2015 International Conference on Industrial Instrumentation and Control (ICIC)*, Pune, India, IEEE, 2015, pp. 71-76.

[19] M.J. D'Souza, W. K. Kroen, C. B. Stephens, and R. J. Kashmar, "Strategies and Initiatives that Revitalize Wesley College STEM Programs," *Journal of College Teaching and Learning*, 12(3), pp.195-208, 2015.

[20] M. J. D'Souza, K. J. E. Walls, C. Rojas, L. M. Everett, and D. E. Wentzien, "Effect of Gender and Lifestyle Behaviors on BMI Trends in a Sample of the First State's Undergraduate Population," *American Journal of Health Sciences*, 6(1), pp. 59-74, 2015.

[21] M. J. D'Souza, and D. N. Kevill, "Influence of Sulfur for Oxygen Substitution in the Solvolytic Reactions of Chloroformate Esters and Related Compounds," *International Journal of Molecular Sciences*, 15(10), pp.18310-18332, 2014.

[22] M. J. D'Souza, G. J. AlAbed, M. Earley, N. Roberts, and F. J. Gerges, "Manipulating In-House Designed Drug Databases For The Prediction of pH-Dependent Aqueous Drug Solubility," *American Journal of Health Sciences*, 4(3), pp. 137-150, 2013.

[23] M. J. D'Souza, and Q. J. Wang, "Inter-institutional Partnerships Propel a Successful Collaborative Undergraduate Degree Program in Chemistry," *Journal of College Teaching and Learning*, 9, pp. 245-252, 2012.

[24] M. J. D'Souza, G. J. AlAbed, J. M. Wheatley, N. Roberts, Y. Veturi, X. Bi, and C.H. Continisio, "A Database Developed with Information Extracted from Chemotherapy Drug Package Inserts to Enhance Future Prescriptions," In Conference on Computer Vision and Pattern Recognition Workshops IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Workshops, 2011, pp. 219-226.

[25] M. J. D'Souza, A. J. Ghada, "Deficiencies in the Reporting of V_D and $t_{1/2}$ in the FDA Approved Chemotherapy Drug Inserts," *Pharmaceutical Reviews,* 8(1), 2010, ISSN 1918-5561.

[26] M. J. D'Souza, F. J. Gerges, "Raising a Red Flag: Deficiencies found in the Reporting of Important Parameters in FDA Approved Drug Profiles," *Pharma IQ*, July issue, 2010.

[27] M. J. D'Souza, F. Koyoshi, F. "Extracting Relevant Information From DFA Drug Files to Create a Structurally Diverse Drug Database Using KnowItAll®," *Pharmaceutical Reviews,* 7(3), ISSN 1918-5561, 2009.

[28] M. J. D'Souza, F. Koyoshi, L. M. Everett, "Structure Activity Relationship (SAR) Patterns Observed Within a Series of Unrelated Common Consumer Drugs," *International Conference on Bioinformatics, Computational Biology, Genomics, and Chemoinformatics,* 2009, 1-6, ISBN: 978-1-60651-009-4.

[29] M. J. D'Souza, F. Koyoshi, L. M. Everett, "Structure Activity Relationships (SARs) Using a Structurally Diverse Drug Database: Validating Success of Predictor Tools," *Pharmaceutical Reviews,* 7(5), ISSN 1918-5561, 2009.

[30] M. J. D'Souza, "KnowItAll® U System - Software Reviews," *Chemistry World,* 4(11), 70-72, 2007.

[31] M. J. D'Souza, "KnowItAll® - Software Reviews," *Chemistry World,* 2(9), 2005, 70-71.

[32] C. M.Hales, C. D. Fryar, M. D. Carroll, D. S. Freedman, and C. L. Ogden, "Trends in Obesity and Severe Obesity Prevalence in US Youth and Adults by Sex and Age, 2007-2008 to 2015-2016," *JAMA*, 319(16), pp.1723-1725, 2018.

[33] C. L. Ogden, M.D. Carroll, C. D. Fryar, and K. M. Flegal, K.M., *Prevalence of obesity among adults and youth: United States, 2011-2014*. Washington, DC: US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics. 2015, pp. 1-8.

[34] U. Ladabaum, A. Mannalithara, P. A. Myer, and G. Singh, "Obesity, Abdominal Obesity, Physical Activity, and Caloric Intake in US Adults: 1988 to 2010," *The American Journal of Medicine*, 127(8), pp.717-727, 2014.

[35] E.S. Ford, L. M. Maynard, and C. Li, "Trends in Mean Waist Circumference and Abdominal Obesity Among US Adults, 1999-2012," *JAMA*, 312(11), pp.1151-1153, 2014.

[36] L. Abarca-Gómez, Z. A. Abdeen, Z. A., Hamid, N. M. Abu-Rmeileh, B. Acosta-Cazares, C. Acuin, R. J., Adams, W. Aekplakorn, K. Afsana, C. A. Aguilar-Salinas, and C. Agyemang, C. " Worldwide Trends in Body-mass Index, Underweight, Overweight, And Obesity From 1975 to 2016: A Pooled Analysis of 2416 Population-Based Measurement Studies in 128.9 Million Children, Adolescents, And Adults." *The Lancet*, 390(10113), pp. 2627-2642, 2017.

[37] B. Kennedy, C. J. Lavie, and S. N. Blair, S. N, "Fitness or Fatness: Which Is More Important?," *JAMA*, 319(3), 231-232, 2018.

[38] M. Hobbs, C. Griffiths, M. A. Green, H. Jordan, J. Saunders, and J. McKenna, "Neighbourhood Typologies and Associations with Body Mass Index and Obesity: A Cross-Sectional Study," *Preventive Medicine*, 111, pp.351-357, 2018.

[39] C. Pozza, and A. M. Isidori, "What's Behind the Obesity Epidemic," In *Imaging in Bariatric Surgery*, *Springer, Cham.* 2018, pp. 1-8.

[40] S. P. Hung, C. Y. Chen, F. R. Guo, C. I. Chang, and C. F. Jan, "Combine Body Mass Index and Body Fat Percentage Measures to Improve the Accuracy of Obesity Screening in Young Adults," *Obesity Research & Clinical Practice*, 11(1), pp.11-18, 2017.

[41] Berrington de Gonzalez, P. Hartge, J. R. Cerhan, A. J. Flint, L. Hannan, R. J. MacInnis, S. C. Moore, G. S. Tobias, H. Anton-Culver, L. B. Freeman, and W. L. Beeson, "Body-Mass Index and Mortality Among 1.46 Million White Adults," *New England Journal of Medicine*, 363(23), pp. 2211-2219, 2010.

[42] L. R. M. D. Lima, R. P. D. A. Freitas, L. R. D. Silva, and A. C. Q. D. Medeiros, "Estimation of Body Mass Index From Self-Reported Measures: What is the Validity?," *Journal of Physical Education*, 29, e2907, 2018.

[43] P. B. Júlíusson, M. Roelants, B. Benestad, S. Lekhal, Y. Danielsen, J. Hjelmesæth, and J. K. Hertel, "Severe Obesity is a Limitation for the Use if Body Mass Index Standard Deviation Scores in Children and Adolescents," *Acta Paediatrica*, 107(2), pp.307-314, 2018.

[44] R. Kim, I. Kawachi, B. A. Coull, and S. V. Subramanian, "Patterning of Individual Heterogeneity in Body Mass Index: Evidence From 57 Low-And Middle-Income Countries," *European Journal of Epidemiology*, pp.1-10, 2018.

[45] Friede, J. A. Reid, and H. W. Ory, "CDC WONDER: A Comprehensive On-Line Public Health Information System of the Centers for Disease Control and Prevention," *American Journal of Public Health*, 83(9), 1289-1294, 1993.

[46] US Census Bureau, "Products and Tools Including Data Visualizations, Mobile Apps, Interactive Web Apps and Other

Software," [Online]. Available: https://www.census.gov/data.html [Accessed: May-July 2017].

[47] A Project of the Trust for America's Health and the Robert Wood Johnson Foundation, "The State of Obesity: Better Policies for a Healthier America," [Online]. Available: https://stateofobesity.org/adult-obesity/ [Accessed: Jan 2017-June 2018]

[48] US News, "Public Health Rankings: Measuring the Health of State Populations," [Online]. Available: https://www.usnews.com/news/best-states/rankings/health-care/public-health [Accessed, July 15, 2017].

[49] CBS News, "States With the Highest and Lowest Exercise Rates," June 28, 2018. [Online]. Available: https://www.cbsnews.com/news/states-with-the-highest-and-lowest-exercise-rates/ [Accessed, June 30, 2018].

[50] L. Beam, and I. S. Kohane, "Big Data and Machine Learning in Health Care," *JAMA*, 319(13), 1317-1318, 2018.

[51] M., Vo, A. Plachkinova, R. Bhaskar, and B. Hilton, "A Conceptual Framework for Quality Healthcare Accessibility: A Scalable Approach For Big Data Technologies," *Information Systems Frontiers*, 20(2), 2018, pp.289-302.

[52] G. Manogaran, and D. Lopez, "A Survey of Big Data Architectures and Machine Learning Algorithms in Healthcare," *International Journal of Biomedical Engineering and Technology*, 25(2-4), pp.182-211, 2017.

[53] Y. Jin, and O. Yu, "Understanding Obesity through a Big Data Analysis," *The FASEB Journal*, 30(1 Supplement), 907-15, 2016.

[54] M. J. Khoury, and J. P. Ioannidis, "Big Data Meets Public Health," *Science*, 346(6213), pp. 1054-1055, 2014.

[55] T. B. Murdoch, and A. S. Detsky, "The Inevitable Application of Big Data to Health Care," *JAMA*, 309(13), 1351-1352, 2013.

[56] H. Chen, R. H. Chiang, and V. C. Storey, "Business Intelligence and Analytics: from Big Data to Big Impact," *MIS Quarterly*, pp.1165-1188, 2012.

[57] M. H. Kuo, T. Sahama, A. W. Kushniruk, E. M. Borycki, and D. K. Grunwell, "Health Big Data Analytics: Current Perspectives, Challenges and Potential Solutions," *International Journal of Big Data Intelligence*, 1(1-2), pp.114-126, 2014.

[58] M. Liebovitz, and J. Farhenbach, "Counterpoint: Is ICD-10 Diagnosis Coding Important in the Era of Big Data? No," *Chest*, 153(5), pp.1095-1098, 2018.

[59] D. Vandyk, I. D. Graham, E. G. VanDenKerkhof, and M. B. Harrison, M. B. "Accuracy of ICD codes for Persons with Considerable Emergency Department Use for Mental Health Complaints," *International Journal of Emergency Mental Health and Human Resilience*, 17(3), pp. 600-601, 2015.

[60] S. E. Bowman, "Why ICD-10 is Worth the Trouble," *Journal of AHIMA*, 79(3), pp. 24-29, 2008.

[61] K. J. O'Malley, K. F. Cook, M. D. Price, K. R. Wildes, J. F. Hurdle, and C. M. Ashton, "Measuring Diagnoses: ICD Code Accuracy," *Health Services Research*, 40(5p2), pp. 1620-1639, 2005.

[62] J. Verhovshek, "Obesity: ICD-10-CM Code Assignment," *AAPC-Advancing the Business of Healthcare*, July 15, 2015,[Online]. Available: https://www.aapc.com/blog/31368-obesity-icd-10-cm-code-assignment/.

[63] H. Quinn, "How a Wesley College Study on Obesity Brought Wellness Programs to Campus," *Technical.ly Delaware*, August 9, 2017, [Online]. Available: https://technical.ly/delaware/2017/08/09/wesley-college-study-obesity-brought-wellness-programs-campus/.

[64] J. Cobourn, "History Has 'Not Told the Complete Story of All Scientists': Leaders Talk Equality In STEM," *Technical.ly Delaware*, November 21, 2016, [Online]. Available: https://technical.ly/delaware/2016/11/21/stem-council-symposium/.

[65] S. Edwards, "Wesley Cannon Scholars Program Provides Support for Low-Income STEM Majors," *Insight into Diversity*, pp. 40-42, 2015.