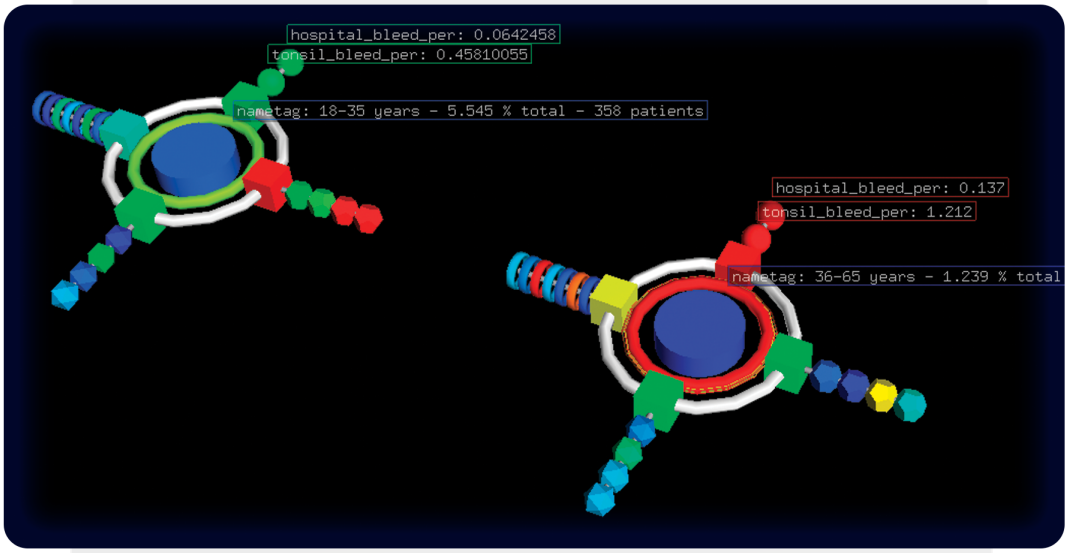
Duke University School of Medicine

CHEER APPLIED

SYNGLYPHX

ON YOUR MARK, GET SET, GO!



Using data visualization software to get there quicker and without missing any landmarks – a case study in healthcare research.

TRANSFORM DATA TO KNOWLEDGE. FASTER.

(571) 733-9470 • INFO@SYNGLPYX.COM

Kristine A. Schulz, DrPH[a], Mark Sloan, MBA[b], Alex Ulvila, JD[b], David L. Witsell, MD[a]

[a]Duke University Medical Center, Durham, NC, USA

[b]SynGlyphX, Arlington, VA, USA

Corresponding Author:

Kristine A. Schulz,

Head and Neck Cancer & Communication Sciences,

Duke University Medical Center

Box 3805, Durham, NC 27710, USA

kristine.schulz@duke.edu

Duke University School of Medicine

CHEER APPLIED

SYNGLYPHX

ABSTRACT:

Introduction: A traditional approach to healthcare research is to establish the hypotheses or set of research questions to be explored. When research is conducted using large databases, the statistical plan is set forth in advance to address these hypotheses and/or questions. By convention, the investigator develops the analysis plan based on prior evidence or their specified study goals. With the advent of Big Data, investigators and organizations need to enhance their capacity to glean value and insight from mass quantities of data.

Materials and Methods: In our case study, we used the SynGlyphX visualization platform to re-"visualize" our tonsillectomy study database for possible new findings. SynGlyphX allows anomalies and opportunities in datasets to quickly be visualized. Our research study on national tonsillectomy bleed rates was conducted by the CHEER practice-based research network in 2015. The data from this study was re-analyzed with the SynGlyphX platform.

Results: The SynGlyphX software highlighted the same key findings as the previous statistical analyses and additionally highlighted several anomalies in the data not explored in the original statistical analyses. The re-analysis was noted to be efficient and quickly provided insights within the study data. In presentation, the visualization software mode was also found to be more compelling than traditional oral presentation.

CONCLUSIONS:

1. **Interactive data visualization and discovery software, in this case SynGyphX, can be used as a complement to statistical analyses to quickly hone in on both anomalies and opportunities.**
2. **In today's information overload environment, methods to distill big data into digestible components without missing key areas of concern or action is imperative.**
3. **Additionally, an approach to present data in ways that captures one's audience is equally important.**

Duke University School of Medicine

CHEER APPLIED

SYNGLYPHX

INTRODUCTION

Big Data is not only here but it keeps coming. And it's not all the same. While there are a multitude of efforts underway to develop common data standards and definitions, we are far from a solution. In healthcare, as federal research funding continues to be tight, research and analyses using Big Data are more and more frequent (versus human subjects research). However, using Big Data is not easy. It arrives as a heterogeneous, seemingly chaotic meshwork of data elements that make insightful assumptions or assertions impossible. Many healthcare researchers spend hours and days learning about their dataset before any questions can begin to be explored.

The traditional approach is to establish a set of hypotheses or research questions to explore. The associated analyses are typically dictated by the evidence and/or interests and assumptions of the investigator. With Big Data, there may be more data fields available to the investigator and more angles potentially to explore beyond the initial hypotheses. However, it is difficult for an individual or even team of individuals to quickly determine where the data might lead without the ability to "see" the data as they were formulating their approach. MIT Sloan Management Review and the IBM Institute for Business Value conducted a survey of 3,000 executives, managers and analysts regarding Big Data and one of the key findings was that those that will reap success from Big Data need to have the accompanying tools and analytics to transform data to insights to value[1].

In this case study, investigators replicated a database study previously conducted via a traditional approach using SynGlyphX's GlyphIT™ visualization software. SynGlyphX software simultaneously visualizes complex, heterogeneous data allowing anomalies and opportunities in datasets to quickly become apparent. A research study on tonsillectomy bleed rates conducted using the RDC dataset from the CHEER practice-based research network was the basis for the replication in the software.

CHEER (Creating Healthcare Excellence through Education and Research) is a practice-based research network in otolaryngology-head and neck surgery, funded through the National Institute on Deafness and other Communication Disorders (NIDCD). CHEER encompasses 30 sites that span 19 states. As a vital component, the retrospective data collection (RDC) project serves as a capacity descriptor and holds more than 277,519 unique patients with demographics, ICD-9, and CPT codes. The overall objective of the study utilized for replication in SynGlyphX was to assess the incidence of post-tonsillectomy bleeding in patients at CHEER sites, as recommended by national guidelines. In the study utilized for replication, the authors postulated that bleeding rates differed among CHEER sites due to variability among site types, demographics, clinical diagnosis, and patient ages. The statistical plan was established by the team of authors and run by a statistician. The elapsed time to create the statistics plan and run the statistics plan was two weeks, andanalysis of the data and summary into the initial draft manuscript took approximately one week. These timeframes are estimated for later comparison to SynGlyphX. The manuscript from the study utilized for replication has been submitted and accepted for publication[2].

Duke University School of Medicine

CHEER APPLIED

SYNGLYPHX

MATERIALS AND METHODS

The Duke University Health System IRB evaluated and deemed studies utilizing the RDC to be exempt research. The RDC database includes de-identified patient data of at least one calendar year (2011-2012 or 2012-2013) from participating CHEER sites. These data include administrative, demographic (i.e. age, gender), diagnostic, and procedural data[3].

The specifications[2] for replicating the tonsillectomy data pull were provided to the SynGlyphX analyst. The specifications included ICD-9 diagnosis codes for tonsillectomy and associated conditions; ICD-9 bleed codes; CPT procedure codes for tonsillectomy and adeno-tonsillectomy, demographics; and site descriptors. SQL was utilized to replicate the creation of the study subset. At this stage, as an alternative to traditional statistical testing, analysis and interpretation, the SynGlyphX software was utilized.

SynGlyphX visually represents data in three dimensional (3D) "glyphs." A glyph is a collection of geometric objects that have data mapped to properties of color, shape, size, transparency, rotation, position, and other attributes to communicate and convey meaning to the user. The system optimizes human comprehension of rich, high volume data and can account for such factors as geographic location and temporality. (See Figure 1)

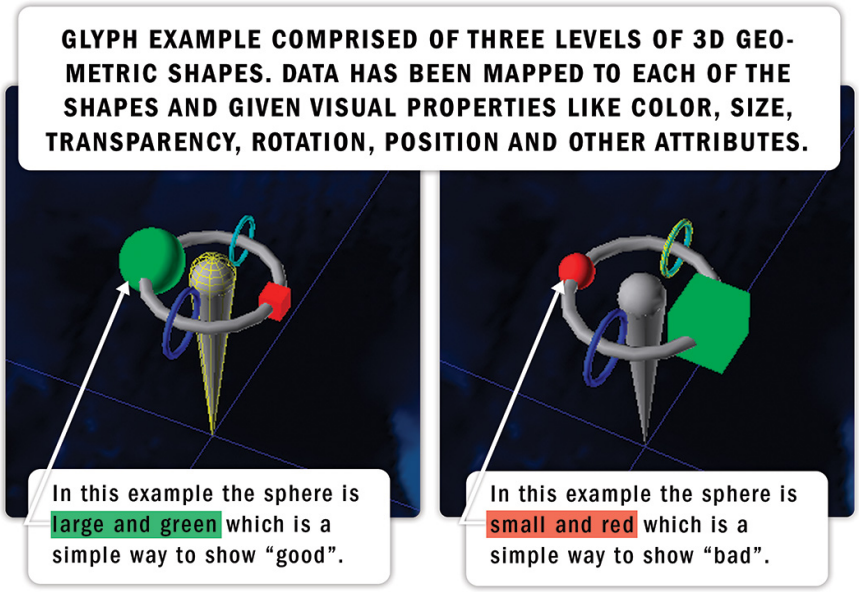


FIGURE 1

For this study, several glyphs were developed based on the hypotheses outlined in the study. The site was used as the primary "pill" or grouping variable by which the data would be visually compared. Demographics and other variables such as associated conditions (infectious and non-infectious) were included as "rings" on arms of the "pill" and the outcome variable, or "tonsillectomy bleed rates" was identified as a box on the "pill". This glyph based approach to data visualization is a novel approach; however, after some cursory training, a typical analyst can design and build a glyph quickly (i.e., less than one hour – often much quicker). Once built, the resulting visualization can be re-used to explore data and discover insight.

As previously mentioned, the user gains an understanding of the data being visualized through a visual analysis of scale, color, transparency and other properties. This is the same process that users of a traditional bar chart or x-y chart use when interpreting: a line in an x-y chart moving up and to the right conveys an increase. In SynGlyphX, increasing scale can be configured to mean "increase" or "more"; transparency can be configured to communicate "confidence levels"; color can be configured to convey "green = good" or "red = bad".

SynGlyphX software is fundamentally unique because it is designed to visually integrate spatial, temporal, and relational data – handling high volumes of multi-dimensional and multi-source data. Said more simply, SynGlyphX software helps the user see WHAT is happening WHERE, WHEN it happened, and additional CONTEXT about the event. This enables the user to deduce more about what is happening – and to identify questions not yet being asked.

RESULTS

25 variables were easily managed within the SynGlyphX framework while only 3 primary variables had been part of the previous analysis plan. Creating glyphs of the tonsillectomy data in the SynGlyphX software enabled us to: 1) replicate the findings from more traditional data analysis – in less time; 2) discover previously undiscovered insights in analyzed data; and 3) more efficiently and effectively communicate findings to researchers at a national conference.

Using SynGlyphX software, investigators quickly and visually identified the same findings reaped from the statistics run, analysis and interpretation process. In brief, the investigators found that bleed rates were significantly higher in patients aged 12 and older; that there were no significant differences in bleed rates between setting (academic versus community) or in those that had associated infectious versus non-infectious diagnoses. In the traditional approach, the elapsed time from study data subset to interpretation (or the list of key findings to incorporate in the manuscript) was about three weeks. In the SynGlyphX glyph based approach, it was about one day.

As, if not more importantly: ***there were several findings that were identified in the SynGlyphX visualization that were not explored in the statistical plan that would have been interesting to explore further having had the hindsight***. For example, those with Hispanic ethnicity had a bleed rate that was 312% higher than those of non-Hispanic ethnicity. Ethnicity had not been included in the research plan. Additionally, in the research plan, the age group had been pre-determined (<12 and ≥12) for the analyses. In the SynGlyphX visualizations, it was clear that while age did drive the results, there was one specific age group (36-65 years) with a bigger impact on the older age range (**see figure 2 and figure 3**). There were other anomalies, such as specific sites, and variations in association conditions by age group, that may have been interesting to explore statistically if the glyph visualizations had preceded or complemented the statistical plan.

Finally, when results were shared via the glyph as a video at a meeting of researchers during a national conference, the qualitative feedback was that ***the SynGlyphX glyph based approach attracts and keeps attention and inspires more discussion, more so than the traditional oral presentation of results***.

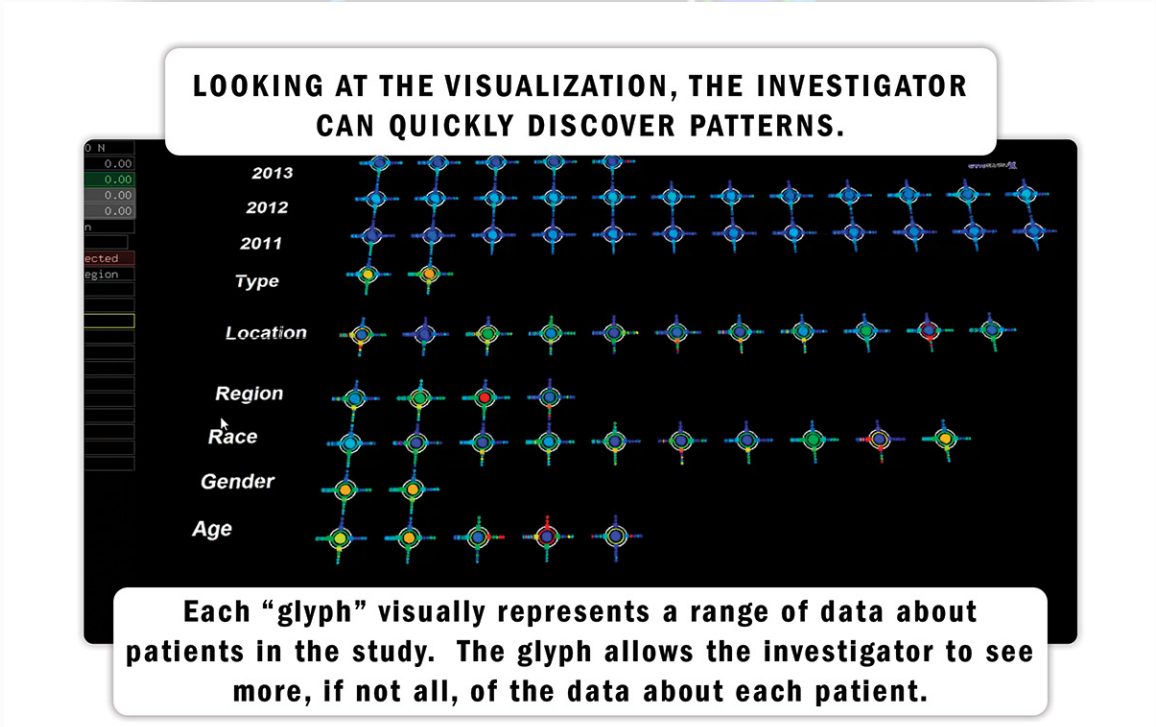


FIGURE 2

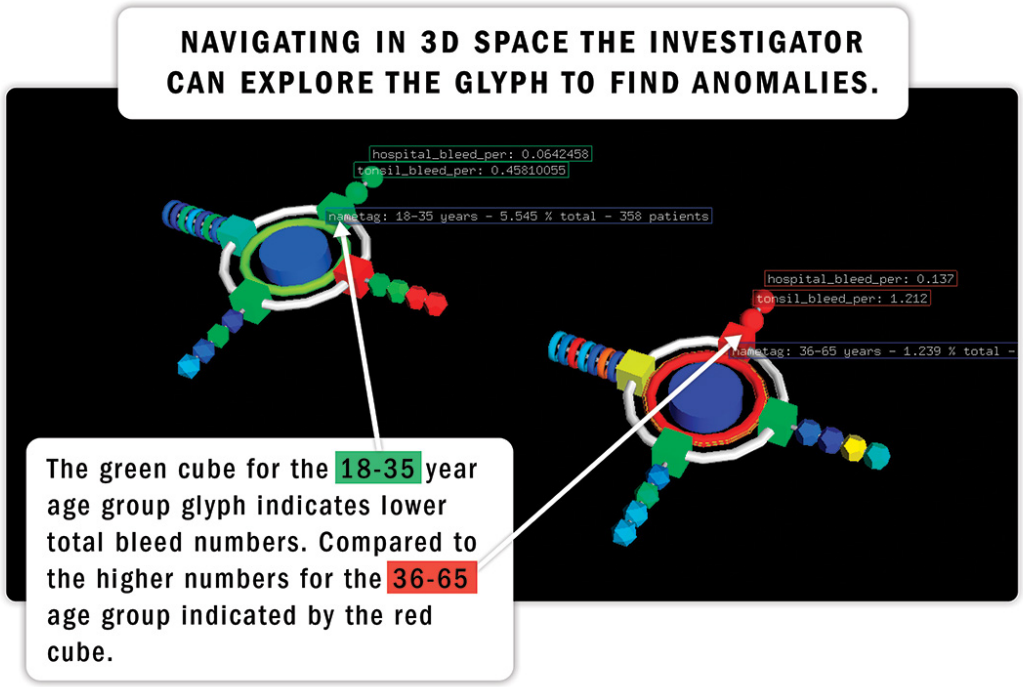

NAVIGATING IN 3D SPACE THE INVESTIGATOR CAN EXPLORE THE GLYPH TO FIND ANOMALIES.
hospital_bleed_per: 0.0642458
tonsil_bleed_per: 0.45810055
nametag: 18-35 years - 5.545 % total - 358 patients
hospital_bleed_per: 0.137
tonsil_bleed_per: 1.212
The green cube for the 18-35 year age group glyph indicates lower total bleed numbers. Compared to the higher numbers for the 36-65 age group indicated by the red cube.

FIGURE 3

Duke University School of Medicine

CHEER APPLIED

SYNGLYPHX

DISCUSSION

Big Data has to be used judiciously, especially in healthcare where either incorrect conclusions or missed opportunities can be potentially harmful. Investigators using Big Data now have an additional responsibility to learn about the data they are using and/or work with the right experts to mine and aggregate it appropriately. As stated at the beginning, to be successful with Big Data, those that will reap success need to have the accompanying tools and analytics to transform data to insights to value[1].

SynGlyphX allowed many more variables to be screened for associations and outliers than could be managed or specified within the original study. While SynGlyphX is not intended to completely replace statistical analyses and software packages, investigators have seen how it can add efficiency to the front-end of the process and be used along the way to provide insight and complement one's research and statistical planning. In this case study, use of SynGlyphX as an accompanying tool proved successful in both identifying the same key findings as a traditional process and further identified additional areas that may have been interesting to explore. The investigators were surprised, in a good way, to be able to find data variable associations for further analysis in such an efficient and timely matter.

Additionally, an interactive visual mode of presenting results of this study proved powerful. In a time where we are inundated with information, electronically and otherwise, the visual representation of results is compelling and different. In a 2012 New York Times article, Stanford statistics professor Trevor Hastie was quoted on the perils of Big Data – "The trouble with seeking a meaningful needle in massive haystacks of data, is that many bits of straw look like needles."[4] Visualizing data helps eliminate the bits of straw and allows for targeted statistical analyses on the needles.

Duke University School of Medicine

CHEER APPLIED

SYNGLYPHX

CONCLUSIONS

SynGlyphX can be used as a complement to statistical analyses to quickly hone in on both anomalies and opportunities. In today's information overload environment, methods to distill big data into digestible components without missing key areas of concern or action is imperative. Additionally, an approach to present data in ways that captures one's audience is equally important.

To see a video of the interactive visualization of Cheer Applied data, please visit:
HTTP://CHEERAPPLIED.ORG/PROJECTS/SYNGLYPHX-DATA-VISUALIZATION/

Acknowledgments

We acknowledge the CHEER Practice-based Research Network and the resources they have built through their NIH/NIDCD funding such as the RDC database, as well as access to the protocol and findings for replication of the tonsillectomy study in the SynglyphX software provided from the Task Force on Practice-based Research.

Author Disclosure Statement

Schulz/Witsell have one license for the SynGlyphX software.

Reference

LaValle S., Lesser E., Shockley R. et al. Big data, analytics and the path from insights to value. Sloan Manage Rev 2011; 21:e1.

Lee W.T., Witsell D.L., Parham K. et al. Tonsillectomy bleed rates across the CHEER practice research network: Pursuing guideline adherence and quality improvement. J Otolaryngol Head Neck Surg [accepted; date TBD].

Witsell D.L., Rauch S.D., Tucci DL, et al. The Otology Data Collection project: report from the CHEER network. Otolaryngol Head Neck Surg. 2011;145(4):572-580.

Lohr, S. The age of big data. New York Times. 2011;11.

Schulz, Kristine, and Mark Sloan. "Discovering Insight In Complex Data Using Interactive Data Visualization". 2016.

Presentation: Data4Decisions, Raleigh, NC. (HTTP://DATA4DECISIONSEVENT.COM/SESSION/DISCOVERING-INSIGHT-IN-COMPLEX-DATA-USING-INTERACTIVE-DATA-VISUALIZATION/)

TO LEARN MORE ABOUT SYNGLYPHX, VISIT WWW.SYNGLYPHX.COM.