Visualizing Algorithmic Cause of Death Predictions

Aditya Anerao, Maggie Dorr, Chethan Jujjavarapu, Andrew Teng

Abstract—Determining an individual's cause of death is an important concern, particularly in areas where death commonly occurs outside of hospitals or healthcare facilities. As a result, various algorithms have been developed to predict these causes based on medical information, including symptoms the patient exhibits. Verbal autopsy (VA), a survey with a relative or close contact, is used to identify the leading cause of death in populations without adequate vital registration systems. VA algorithms leverage symptom-cause information (SCI) to associate symptoms with causes of death (CoD). However, these algorithms vary in accuracy, which can be improved by grouping CoD. In collaboration with Tyler McCormick (UW Statistics and Sociology Departments), we created interactive and dynamic visualizations that depict associations between SCI and CoD based on these algorithms to help policymakers and stakeholders in low resource areas visualize uncertainty in predictive models. Furthermore, we designed these visualizations to assist in understanding the cause-symptom relationships and algorithm performances.

Index Terms—Verbal Autopsy; Death Prediction; Cause of Death; Algorithms

1 Introduction

Verbal autopsy (VA) is an important tool to measure cause of death in populations without complete medical certification for cause of death (CoD). A VA gathers data from the relative(s) of the recently deceased or close contain, and contains information such as symptoms and circumstances surrounding the death. This is a common situation in many countries and as a result, building algorithms that can predict the cause of death from a VA is essential to improve healthcare. However, there are copious amounts of data indicating relationships between cause of death and symptoms, and different VA algorithms that can predict varying outcomes for a single patient. These algorithms depend on three distinct components: (1) VA data, (2) symptom-cause information (SCI), and (3) an algorithmic or probabilistic model that merge the previous two components to assign a likely cause of death [1].

There are many VA algorithms currently in use, each with varying levels of accuracy and performance, including InSilico VA, NBC, and InterVA [1, 2]. These algorithms learn SCI through the use of labelled training data where a cause of death has been confirmed with both VA data and an independent expert [1, 2]. In reality, however, training data are hard to obtain and fair comparisons amongst algorithms can only be assessed when the same SCI is fixed across all algorithms.

Clark et al. (2018) explored and compared the different VA algorithms across various sites. They found that the performance of VA algorithms depends heavily on the SCI. Based on VAs in five African and Asian countries, the InterVA algorithm (specifically InterVA-4) had an 83% concordance correlation coefficient for determining cause-specific mortality fractions, which increased to 97% when AIDS and pulmonary TB deaths were joined [3], demonstrating that grouping can be a successful method of increasing correlation.

Here, we aimed to create interactive and dynamic visualizations that efficiently depicts 1) the relationship between true CoD and symptoms and 2) the varying performance among VA algorithms across various regions and groupings. We targeted our visualizations to be used by technical staff or policy-makers in low resource settings who may or may not have formal training.

2 Methods

We obtained data from the Population Health Metrics Research Consortium (PHMRC), which contains 7,841 adult deaths across six distinct locations: (1) Andhra Pradesh, India; (2) Bohol, Philippines; (3) Dar es Salaam, Tanzania; (4) Mexico City, Mexico; (5) Pemba Island, Tanzania; and (6) Uttar Pradesh, India [4]. All recorded deaths have VA data and expert-confirmed CoD. The confirmed CoD are grouped in three levels consisting of 34, 46, and 55 causes.

Within each group, there are two types of data, (1) symptom-causes that indicate which symptoms are related to which CoD for each patient and (2) probabilities for each CoD are determined by three different algorithms. For the first type of data, we calculated the frequency between a cause and symptom and divided by the total number of relationships. These aggregated relationships are presented in Figures 1, 2, and 3. For the second type of data, we averaged probabilities of each cause for each algorithm and present in Figure 4.

To first explore and understand the relationships between causes and symptoms, we created a force directed network graph. This graph simply encodes causes and symptoms by node color and encodes the relationship between the nodes through line thickness, where a thicker link indicates a higher association and a thinner link indicates a lower association. The network approach allows users to quickly understand the relationship between symptoms and true CoD at a high-level. This approach is ideal for policy-makers who want to get an overview of the data while still being informed of the relations. While the network graph was informative, it lacked a structure to easily understand the data. To address this issue, we developed a parallel coordinate system. With this graph, we now have structure, where causes are on the left y-axis and symptoms are on the right y-axis with relationships represented as lines. We can easily compare different true CoD and symptoms and the associations between the two. With the parallel coordinates graph, higher trained policy-makers will be able to instantly see existing relationships between true CoD and symptoms without the need for further hovering or clicking. Though the parallel coordinates graph was more clear in the relationships between CoD and symptoms, some users may want to also visualize the uncertainty in some relationships. To address this desire, we built a heatmap to allow more trained users to easily observe each cause and symptom along the y- and x-axes, respectively, and at their intersection, view the brightness (i.e., lightness) of the red hue as the level of association. Red and black were used throughout the first three visuals as both colors are often associated with mortality.

Additionally, we had data about the three aforementioned VA algorithms and their probabilities. We created an interactive bar chart that allows users to select the desired grouping to visualize and to hover over the bars to get more information. The graph was

- *All authors are with the University of Washington, Seattle, WA, USA. Emails: {anerao, mldorr, cjujja, akteng} @uw.edu*

simply designed to compare how different algorithms link symptoms to cause of death, and a triple bar chart allows all three algorithms to be simultaneously compared. This visualization shows the average probability of an observed set of symptom-outcome pairings to be matched to a cause of death by a given algorithm. This figure serves as an indirect comparison of the performance of different algorithms. Therefore, the height of different bars can be compared for an individual cause of death.

3 RESULTS

3.1 Network Map

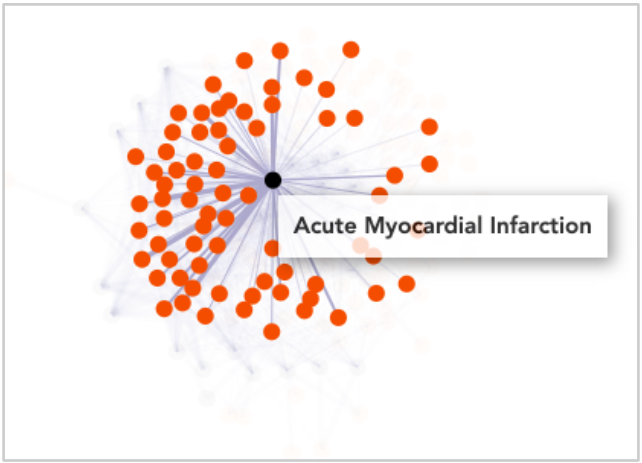

Fig. 1. A view of the force directed network with a cause of death and its associations highlighted.

The force directed network graph encodes true CoD and symptoms through a two-toned node coloring system, where true CoD are black and symptoms are red (**Fig. 1**). The graph is highly interactive and designed for high-level exploration. To explore, one can click on a node of interest, hover over the nodes or links to get various details, or utilize the dropdown on the right-hand side to get a clickable list of true CoD that are displayed in the network. Exploration tips are also included on the right to guide the user in using the graph. The graph is updated every time the dataset (location and grouping) is changed as well as when the user selects a specific node by clicking directly in the graph or through the dropdown. The nodes are dynamic and can be moved anywhere allowing the user to customize the shape of the graph. Furthermore, because there are a lot of nodes and links, the user can also zoom in and out to help assist with accurate selection and hover.

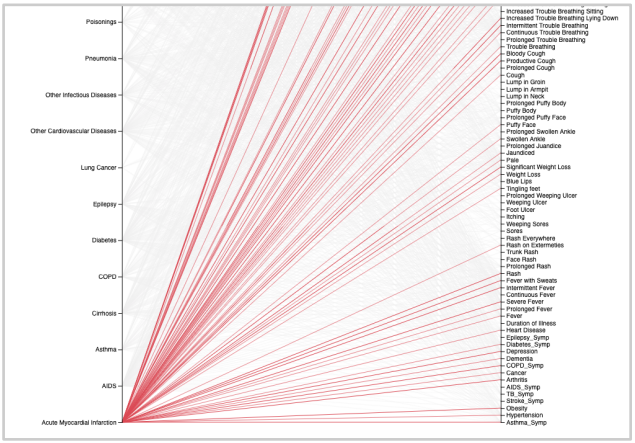

Fig. 2. A partial view of the parallel coordinate system visual with a highlighted cause of death.

3.2 Parallel Coordinate System

The parallel coordinate system describes the relationship between CoD and symptoms (**Fig. 2**). This graph uses vertical axes to encode the CoDs and symptoms with lines to encode their associations. This interactive graph allows users to better understand the data by exploring the differences among CoDs and symptoms. Users can click on a CoD or symptom and the graph will highlight all relevant associations. In addition, users can hover over a specific line for a tooltip to pop up that provides a numerical value that represents the strength of the CoD-symptom association of interest. As mentioned in 3.1, the parallel coordinate system also updates every time the dataset (location and grouping) is changed.

3.3 Heatmap

The heatmap (**Fig. 3**) is able to simultaneously visualize all connections between CoD and symptoms, as well as the strength of their association. Users can see potential connections and strength of connection. Hue brightness is used on a continuous gradient from very pale red (meaning almost zero) to dark red (1), but if they would like more information, a tooltip provides the exact value. Due to the large number of symptoms, the graph must be very wide. Therefore, the y-axis moves when the user scrolls to the right, so it continues to be readable. Users are also able to click each square to turn on a highlight feature to enhance readability. Unfortunately, the moving y-axis means that the axis labels are not clickable.

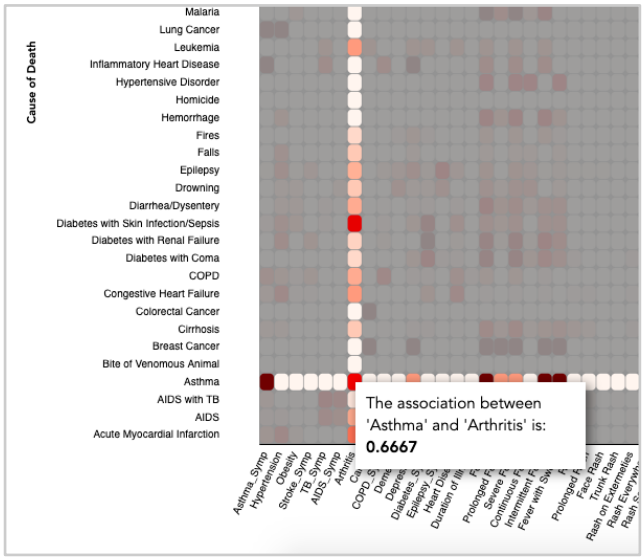

Fig. 3. A partial view of the heatmap with a highlighted row and column.

3.4 Outcome Probability Chart

Figure 4 shows a bar chart of each algorithm's average probability of predicting the correct CoD. The probabilities of all symptoms-CoD observations within a cause grouping (ie. 34, 55, 65) were grouped based on the CoD gold label. Within each CoD gold label group, the average probability for the correct CoD was calculated for each algorithm. Probabilities close to 1 indicates the algorithm predicted the correct CoD with confidence. Each color bar corresponds to one of the three algorithms used; blue for InterVa, yellow for NBC and red for InSilicoVA. Different outcome groupings can be selected using the dropdown menu.

This figure was designed to compare how different algorithms link symptoms to cause of death, and a triple bar chart allows all three algorithms to be simultaneously compared. The relative performance of different algorithms can be observed by comparing the height of different bars for an individual cause of death. For example, when looking at Figure 4 with the default 34 grouping, the bars corresponding to Breast Cancer have similar heights, indicating a similar performance (0.65-0.72) of all three algorithms. In contrast, all three algorithms performed poorly at predicting Asthma (0.07-0.15). These results indicate that these

algorithms vary in performance across CoDs. Follow-up work could involve experimenting with different algorithms, such as neural networks to rescue these difficult to predict CoDs.

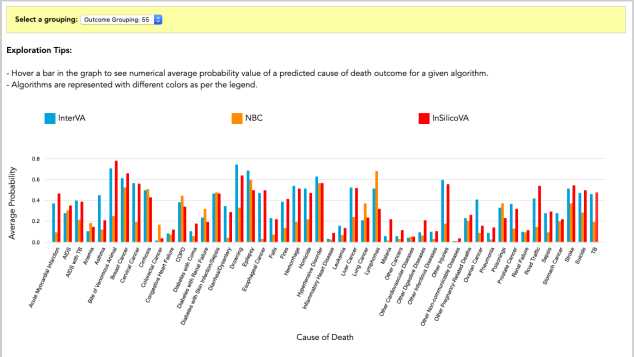

Fig. 4. Bar Graph

3.5 Case Study

Take, for example, a user interested in Acute Myocardial Infarction (AMI) in Andhra Pradesh, India with 34 possible causes. This user is able to first select the location and grouping from the top menu bar, which will correspond to the first three figures (and scroll with the figures, in case the user wishes to look at other locations/groupings). With the first visual, they could search for the point individually, or they are able to easily select AMI from the right hand menu to see that it has many associated symptoms with varying levels of strength. The user can hover for more information over highlighted nodes and links. However, if they wished to view all symptoms simultaneously, the user can then scroll down to the second visual. By clicking on the AMI label on the right y-axis, all the connections are highlighted, and the user can view all symptom labels at once. If the user wishes to simultaneously view all the connections and know the association strengths, then they can scroll to Figure 3 and click on a square associated with AMI. As seen in the previous two graphs, many connections exist, but at the same time, many of those connections are weak. Scrolling to the right, reveals that it is symptoms such as Trouble Breathing and Chest Pain Month before Dying which are most associated with AMI. With the bottom graph, the user selects Outcome Grouping: 34 and observes that the average probability value for predicting AMI based on symptoms is low and quite different between the three algorithms: NBC - 0.13, InterVA - 0.32, and InSilicoVA - 0.42. One could theorize that because AMI is associated with 71 symptoms, it may share enough symptoms with other serious causes that makes it difficult to predict with confidence. For example, AMI and AIDS share 69 symptoms. In addition, symptoms that are highly associated with AMI are also associated with other causes, such as the symptom Continuous Trouble Breathing is also highly associated to the Cancer-related CoDs. This may make it difficult for the algorithms to find features that are strongly associated to AMI specifically.

4 Discussion

Our visualization tool is used primarily for exploring the data and comparing the performance of different CoD prediction algorithms. Our audience spent time exploring our Figures 1-3 to understand how specific symptoms and CoDs relate to each other. Our visualization tool allowed users to discover new relationships that could not have been observed with the original table format of the data. In addition, Figure 4 showed our audience that performance across algorithms was not consistent, which prompted a discussion on what could be done to improve the results, which included applying specific algorithms, such as neural networks.

4.1 Future Work

In the future, to assess our visualizations we could interview policy makers and stakeholders from each region to understand what details and information they want to see. Furthermore, we can obtain feedback from said policy makers and stakeholders on usability and design. Once we obtain their feedback, we would iterate our approach and design. Based on the feedback, we could potentially customize visualizations to each region for optimal dissemination of information.

5. Acknowledgements

We are thankful to our collaborators Dr. Tyler H. McCormick from the University of Washington Department of Sociology and Department of Statistics, and Dr. Zehang Richard Li from the Yale University Department of Biostatistics, whose research generated the datasets and whose guidance inspired this visualization project. We also thank the teaching staff of CSE 512 at the University of Washington, led by Dr. Jeffrey Heer for guidance throughout this project and the course.

6. References

[1] S. J. Clark, Z. Li, and T. H. McCormick, "Quantifying the Contributions of Training Data and Algorithm Logic to the Performance of Automated Cause-assignment Algorithms for Verbal Autopsy," *arXiv,* Nov 2018, doi: 1803.07141v4

[2] S. J. Clark, T. McCormick, Z. Li, and J. Wakefield, "InSilicoVA: a method to automate cause of death assignment for verbal autopsy," *arXiv preprint arXiv*, 2015, doi: 1504.02129.

[3] P. Byass, et al, "Comparing Verbal Autopsy Cause of Death Findings as Determined by Physician Coding and Probabilistic Modelling: a Public Health Analysis of 54 000 Deaths in Africa and Asia." *J Glob Health*, vol. 5, no. 1, Jun 2015, pp. 010402, doi:10.7189/jogh.05.010402

[4] C. J. L. Murray, et al, "Population Health Metrics Research Consortium gold standard verbal autopsy validation study: design, implementation, and development of analysis datasets," *Popul Health Metr,* vol. 9, no. 27, Aug 2011, doi:10.1186/1478-7954-9-27