INTRODUCTION

Public health policy formulated by governments and informed by stakeholders plays an influential role in changing the trajectory of mortality and morbidity arising from diseases, injuries and their risk factors [1]. While differences exist, the common top considerations by policymakers and stakeholders when formulating health policies are the seriousness of the problem and the number of people affected [2]. However, such is a reactive approach as it often entails that an existing problem grows big enough before any response is taken. The limits of such an approach can be seen in the opioid epidemic in the United States.

The opioid abuse problem in the US was identified by a few researchers in 2006 [3]. However, the US health authorities only started prioritizing the opioid abuse problem in 2015 and 2016 by issuing a brief and publishing a set of guidelines respectively [4]. Meanwhile, the number of deaths per year arising from opioid use more than tripled from an estimated 13,000 to 44,000 deaths from 2006 to 2016, becoming the 7th leading cause of DALYs (which measures burden of disease from mortality and morbidity) in the US in 2016 [5-6].

We can clearly see that reactive intervention in national health crises is too costly and ineffective – a predictive approach towards public health is required. Because there is *no universally recognized solution* to this problem, there are a few challenges as detailed below:

- **Problem with defining predictive.** While it is well-defined in individual healthcare what it means to shift from reactive to predictive care [7], there is no universal definition as to what a predictive approach towards public health entails [8]. Some have approached it as detecting an upward trend in a specific disease, injury or risk factor before it escalates further (i.e. micro-level) [3]. Others have deemed it to be forecasting aggregated health outcomes (i.e. deaths from communicable diseases, non-communicable diseases etc.) for a country to aid setting of current public health priorities (i.e. macro-level) [9].
- **Issue with using threshold as benchmark.** Unless ranked within the current top causes of death or risk factors, there is no benchmark on what constitutes a "public health priority". For example, researchers saw the 91.2% increase in the number of deaths from opioid poisoning between 1999 and 2002 as indicative of an epidemic [3]. However, other stakeholders may not see the same increase as being indicative of a public health priority as no universally agreed threshold benchmark exists.

INTRODUCING LIVSAVER

In this paper, we present LivSaver, a novel system for visualizing data-driven public health priorities that strikes a balance between the micro and macro-level approaches. Besides viewing top health issues for a selected country in a specific year, the user can also interact with LivSaver to view the country's top worrying health trends which is a novel feature not available on major health organization's websites.

LivSaver's major innovative features include:

- **Comprehensive definition of predictive.** Unlike single trend analysis which can only detect escalating trends [3], LivSaver can also detect health issues where the trend is declining too slowly relative to a benchmark (i.e. slow-improving). Detecting escalating trends will allow stakeholders to proactively take measures to slow the rate of increase, while detecting slow-improving trends allows for proactive measures to increase the rate of decline.
- **Uses peer trends to form benchmark.** LivSaver uses comparison to similar peers' trends to determine whether a health trend in a specific country should be a public health priority and visualizes this comparison to the user. Interventions through providing peer information have been shown in domains such as in encouraging small charitable donations to improve behavior towards the peer norm [10].
- **Easily scalable.** LivSaver only requires information about past and current health outcomes of diseases, injuries and risk factors. This is unlike forecasting models which requires additional demographic, economic and educational information to be able to accurately forecast future health outcomes to determine current public health priorities [9]. This makes LivSaver easily scalable to determine public health priorities across multiple diseases, injuries and risk factors, as well as across several countries.

Currently, LivSaver uses the time series population health dataset collected and analyzed under the Global Burden of Disease Study between 1990 and 2017 for 195 countries and territories [11]. The data provides the age-standardized DALYs (disability-adjusted life years) rate per 100,000 population (DALYs rate) for more than 290 diseases and injuries, and more than 65 associated risk factors. DALYs measures the burden of disease from mortality and morbidity (more specifically, it measures the years of healthy life lost) and is an established metric used by international health organizations. The size of the dataset as a tall table is about 9 million rows. However, our general approach is database-agnostic, and could be generalized to work with other time series population health data.

LIVSAVER IN ACTION

LivSaver's Interactive User Interface

Users first interact with LivSaver's user interface (Figure 1) by selecting a country ($country_{sel}$), start year ($year_{start}$), end year ($year_{end}$) in (1). LivSaver's user interface will then display information on three panels: the left panel (which includes a bar and line chart) shows the top current health issues (which can be filtered based on health risks or diseases/injuries) for $country_{sel}$ in $year_{end}$. This is information available on most major health organizations websites. The right panel (bar and line chart) filters to show either the top escalating or slow-improving health issues for the selected time period and represents the major innovative features of LivSaver. The map panel in the middle shows countries with similar health risk profile as $country_{sel}$ in $year_{start}$. The more intricate features are detailed in Figure 1's caption.

Illustrative Scenario: LivSaver in Action

The following illustrates how a user might use LivSaver: Grace is a public health authority official in a developed country and is the team lead in deciding the country's public health priorities for the year. In previous years, the agency had always set the current top diseases/injuries and risk factors as the country's public health priorities. But Grace felt that the approach was too reactive and wanted another approach. Using LivSaver (Figure 1), Grace filters the right panel (1) to display the top escalating diseases/injuries and compares the information to the left panel which displays the top current diseases/injuries. Clicking on the name 'ischemic heart disease' in the bar chart, she noted that while ischemic heart disease is the top current disease/injury for the year, its trend has been stable compared to similar peers. However, for opioid use disorder (which was flagged as the top escalating disease/injury), its line chart (2) clearly showed that the country's trend was escalating quickly relative to similar peers. Grace also noted that DALY rate for opioid use disorders was almost 4 times the average among similar peers (i.e. peer average), while ischemic heart disease was only slightly above. Wondering how LivSaver detected opioid use disorder as an escalating health issue, Grace clicked on the info button (5). To understand the certainty of the data, she hovers around each node in the line chart (3) to obtain the estimate's range. She also hovers over the bar chart (4) to find the top risk factors that contributes to ischemic heart disease. Grace is now ready to decide on the country's public health priorities.

LIVSAVER KEY ALGORITHMS AND FUNCTIONS

LivSaver detects a country's top worrying health trends by comparing a country's trend for a specific disease, injury or risk factor to trends of a similar

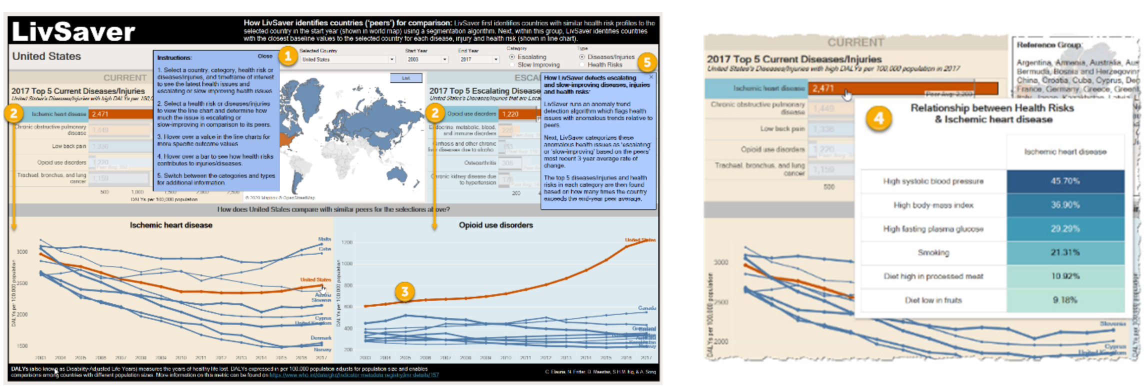


Figure 1. Visualization. (1) Following the instructions, the end-user will first select Country, Start Year, End Year, Category (i.e. Escalating or Slow-Improving), and Type (i.e. Diseases/Injuries or Health Risks). Assuming 'Diseases/Injuries and 'Escalating' is selected, the Top 5 Current Diseases/Injuries will show on the left, and the Top 5 Escalating Diseases/Injuries on the right. (2) The user selects a disease/injury on the bar chart to show the trend of the selected country and its peer group over the selected year range. (3) Hovering over each node in the line chart shows the DALYs rate and its respective range. (4) Hovering over the bar chart shows the relationship between the selected disease/injury and the top health risks from all categories (i.e. current, escalating etc.). (5) Clicking on the info icon will provide the user with information on how LivSaver detects escalating and slow-improving diseases, injuries and risk factors.

peer group. The intuition for defining similar peer groups can be seen in this example: In 1990, Canada and Botswana had similar DALYs for major depressive disorders. In 2017 however, Canada DALYs rate for major depressive disorders had fallen by 10% while Botswana DALYs rate had increased by 20%. As Canada and Botswana had very different socio-economic environments (Canada's national income per capita was 7 times of Botswana's in 1990), it will be difficult to use Canada's trend to persuade Botswana's public health authorities to prioritize major depressive disorder as a public health issue. It will be more appropriate to use the trend of another country who is both similar socio-economically with a similar baseline value. This comparison is achieved using the algorithms and functions detailed below.

Segmentation Algorithm

The segmentation algorithm aims to segment countries into clusters with similar health risk profiles. When the user selects a start year ($year_{start}$), the DALYs rate for 67 risk factors for that year will be used as features in the algorithm. Risk factors were chosen as they largely drive the occurrence of diseases and injuries [12]. As the features are largely linearly correlated with each other, dimension reduction is performed using principal component analysis on the normalized data. Based on the criteria that retains only components that explained at least 5% of the variance in the normalized data, either 2 or 3 components were retained depending on $year_{start}$. The retained components explained at least 79.3% of the total variance across the years.

K-means clustering algorithm is then applied to segment the countries into k clusters based on the retained components. Clustering is used as it can systematically consider information on a large set of indicators when deciding cluster membership, as shown in its use in various studies to segment healthcare systems/ populations for further analysis and targeted response [13-19]. The optimal number of clusters for each $year_{start}$ was found a priori using metrics that examined the quality of clustering results (i.e. Silhouette Coefficient, Calinski-Harabsz Index, Davies-Boudlin Index) and visual inspection. k=5 was found to be the optimal number of clusters across the different years even with different random centroid initializations. The clusters also exhibited short-term temporal stability with only 4 countries changing cluster membership from 1990 to 1992.

Anomaly Detection Algorithm

The anomaly detection algorithm (as seen in Figure 2) aims to detect anomalous trends in a selected comparison group. Part of the algorithm uses Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm to detect anomalous trends [20]. The optimal values for two parameters s (the number of samples in a neighborhood for a point to be considered a core point) and n nearest countries varies based on the K-means cluster size for the user-selected country $country_{sel}$, and are determined a priori through visual inspection of the resulting trend graphs. To ensure that anomalous trends are only singled out for diseases, injuries and risk factors that have moderate to high impact in $country_{sel}$ for a user-selected end year $year_{end}$, the data is pre-filtered to include only diseases, injuries and risk factors where the DALY rate in $year_{end}$ for $country_{sel}$ exceeds 100.

Trend Categorization and Ranking Function

The diseases, injuries and risk factors in which anomalous trends are detected by the anomaly detection algorithm can be categorized into three

Anomaly Detection Algorithm

User-selected parameters: $year_{start}$, $year_{end}$, $country_{sel}$
List of countries in $country_{sel}$ K-means cluster: $cluster_{country}$
Pre-filtered list of diseases, injuries and risk factors: $cause_{list}$
DALY rate for each country, year and cause: $D_{i,j,k}$ where i = country, j = year, k = cause
Initialize epsilon $\varepsilon = 1$
for $cause$ in $cause_{list}$, **do**
  $comp_{group} = n$ nearest countries (including $country_{sel}$) in $cluster_{country}$ to $D_{country_{sel}, year_{start}, cause}$ based on Manhattan distance.
  continue to next $cause$ **if** $\left| D_{median_{comp_{group}}, year_{start}, cause} - D_{country_{sel}, year_{start}, cause} \right| \geq 0.5 * D_{country_{sel}, year_{start}, cause}$
  $DALY_{group} = D_{i,j,k}$ where $i = comp_{group}$, $j = year_{start}, \dots, year_{end}$, $k = cause$
  while True, **do**
    $label$ = Array of labels assigned to $comp_{group}$ by DBSCAN algorithm where data = $DALY_{group}$, epsilon = ε and min_samples = s
    if all($label$) == 0 (i.e. clustered into one cluster), **do**
      break while loop
    $\varepsilon = \varepsilon + 1$
  if $label_{country_{sel}}$ $(at\ \varepsilon - 1)$ is classified as noise element (i.e. $== -1$)
  and number of unique labels in $label$ $(at\ \varepsilon - 1) > 1$, **do**
    Include $cause$ in anomaly trend list for $country_{sel}$ from $year_{start}$ to $year_{end}$

Figure 2. Basic pseudo code for anomaly detection algorithm. The anomaly detection algorithm uses DBSCAN with a priori determined optimal min_samples s and optimal n nearest countries which vary depending on size of $cluster_{country}$. Additional lines of code may be added in actual code to hasten the search for the minimum ε value where all(label) is equal to 0.

types: (1) escalating trend - where $country_{sel}$ trend is escalating faster compared to the general increase among similar peers (i.e. $comp_{group}$), (2) slow-improving trend - where $country_{sel}$ trend is declining slower/increasing compared to the general decline among similar peers, and (3) where $country_{sel}$ trend is increasing slower/declining faster compared to the general increase/decline among similar peers. As we are only interested in visualizing (1) and (2), (3) was suppressed by excluding health issues where $country_{sel}$ had the lowest DALY rate among similar peers in $year_{end}$. Anomalous trends were then categorized as 'escalating' or 'slow-improving' based on the peer-averaged recent 3-year rate of change (pa_{rate}). Anomalous trends were considered as 'escalating' if $pa_{rate} > 0.1$, and as 'slow-improving' otherwise. Health issues within each category were then ranked using the formula below where $peer$ is the list of countries in $comp_{group}$ excluding $country_{sel}$:

$$severity = \frac{D_{country_{sel}, year_{end}, cause}}{Mean\ (D_{peer, year_{end}, cause})}$$

Health issues with higher $severity$ will be ranked higher within both type of trends.

VISUALIZATION IMPLEMENTATION

The data output files from LivSaver's algorithms (with Alteryx transformation made) were loaded into Tableau for the visualization layer. This data consisted of the following for each country and year/year range: (1) Top 5: Top five current, escalating and slow-improving diseases/injuries and health risks; (2) Peer Group: The peer group with (a) similar health risk profile and additionally with (b) similar baseline value; (3) DALYs: DALYs rate (including upper and lower); and (4) Diseases/Injuries & Health Risk Relationship: Details risk factor information between the two. The Tableau packaged workbook will be hosted on DropBox for users to download and interact with.

EVALUATION

Evaluation of LivSaver's Algorithms

LivSaver serves to benefit stakeholders (i.e. public health authorities) by detecting worrying public health trends earlier to enable earlier intervention[1]. We evaluated the anomaly detection algorithm against the single time series analysis (rate-of-change) approach to ensure the intended objective is met.

Hypothesis. We hypothesize that LivSaver's anomaly detection algorithm will detect worrying (i.e. escalating and slow-improving) health trends relative to similar peers more accurately when compared to the rate-of-change approach.

Design. 3 countries from different K-means clusters were selected for different year ranges: United States (from year 2003 to 2017), Indonesia (from year 2004 to 2016) and Zimbabwe (from year 2005 to 2017). Each country was subjected to two algorithmic approaches: (1) rate-of-change approach (where diseases/injuries whose most recent 3-year average rate of change is greater than or equal to 2 are flagged), and (2) the anomaly detection algorithm used by LivSaver to flag worrying trends in diseases/injuries (before results were filtered for the top 5 in each category).

Procedure. Results for each country sample were generated using the two approaches. Two independent coders then visually inspected the respective trend graphs for the country and its similar peers for each of the disease/injuries flagged. These trend graphs were either coded as (1) 'positive': the country's trend was either increasing more quickly or declining more slowly relative to peers in the given year range, or (2) 'negative': the country's trend was similar to peers in the given year range. Any discrepancies in coding were resolved through majority consensus by all 5 group members. These ratings formed the ground truth for the trend graphs. A confusion matrix was then tabulated for each approach to compute the true positive rate and false positive rate.

Results. The two approaches flagged a total of 65 health issues (and trend graphs) across all 3 countries. Out of these 65 trend graphs, 32 were rated as 'positive' by coders, while the remaining 33 were rated as 'negative'. LivSaver's algorithmic approach had a much higher true positive rate of 81.25% (versus 25% for the rate-of-change approach), and a lower false positive rate of 39.99% (versus 60.6% for the rate-of-change approach). This shows that LivSaver's algorithm does detect worrying (i.e. escalating and slow-improving) health trends relative to similar peers more accurately than the rate-of-change approach.

Evaluation of LivSaver's User Interface

Besides allowing stakeholders to detect worrying public health trends earlier to enable earlier intervention, LivSaver also serves to allow

[1] Different stakeholders will intervene in different ways to improve public health [21]. For example, healthcare authorities may formulate an appropriate health policy; non-government authorities may engage in advocacy and delivery of targeted programs; healthcare researchers may publish research to shed further light on the health issue; and the general public may raise awareness and volunteer resources [22].

stakeholders to interpret easily the public health evidence to better inform decision-making. We evaluated LivSaver's user interface against Global Burden of Disease Study visualization tool, GBD Compare (GBD-C) to ensure the intended objectives are met [23].

Hypothesis. We hypothesize that compared to GBD-C: (1) users will find LivSaver easier to navigate; (2) users can interpret information shown on LivSaver more easily and accurately; (3) users will also consider similar peers' and trends when deciding on public health priorities.

Design. All users assumed the persona of a public health official in response to two scenarios (Scenario US and Scenario Chile). A within-subjects design was used where all users used both LivSaver and GBD-C to respond to the scenarios (i.e. if user had used LivSaver for Scenario US, they would use GBD-C for Scenario Chile.). The order in which the scenarios was completed, and the interface tool used for each scenario were randomized across all users to avoid systematic bias. After completing each scenario, the users will complete a questionnaire about the user interface that they had just interacted with.

Users. 10 users were recruited through convenience sampling. All users had no prior experience with using GBD-C or LivSaver. Users either directly or dialed in remotely to use the tester's computer to complete the experiment. All experiments were conducted as a 1-to-1 session. Participation in this experiment was voluntary and no compensation was provided.

Materials. *Scenario Questionnaires (See Appendix A and B).* Two scenarios – Scenario US and Scenario Chile, were developed for this experiment. In each scenario, the user will answer two types of questions based on information found in the user interface: (1) FACT questions – where the user provides specific facts and figures, and (2) JUDGMENT questions – where the user recommends the public health priority that the country should pursue. FACT questions are multiple-choice or checkbox, and are timed and scored. JUDGMENT questions are multiple-choice or open-ended, with no right or wrong answers. All questions for each scenario could be answered based on information found in each user interface. Questions for both scenarios are similar, with answer options adjusted to incorporate country-specific facts.

User Interface Questionnaire (See Appendix C). Users' perceptions about the ease of navigating and interpreting the information on the user interface are measured based on their level of agreement with certain statements based on a Likert scale which ranged from 1 (Strongly Disagree) to 5 (Strongly Agree). For LivSaver, users are additionally queried in an open-ended format about the features in the user interface that they found more useful or could be improved on.

Procedure. Before the experiment, the tester pre-loaded GBD-C and LivSaver in the internet browser and Tableau Reader respectively. All questionnaires that the users were required to complete during the experiment were opened in internet browser. All windows were minimized before the start of the experiment.

At the start of the experiment, the tester briefed the user using a standardized script (see Appendix D) about the purpose of the experiment, the scenarios, and the type of questions that they will complete for each scenario. The tester also informed the user that for FACT questions, (1) these questions are timed and scored, (2) hints (which will be measured) may be requested, and (3) they will be stopped after 20 minutes. The user then proceeded to complete the FACT questions for the first scenario assigned to them using the given visualization tool. After 20 minutes or on completion of the FACT questions, tester informed the user to answer the JUDGMENT questions. After user completed the JUDGMENT question, the tester directed the user to complete the user interface questionnaire for the user interface that they had just interacted with. The user completed the same process for the second scenario using the other visualization tool.

At the end of the experiment, the user was thanked for their time and a debrief regarding the purpose of the study was provided if requested.

Results (see Figure 3). *Ease of use.* The ease of navigating the user interface was measured using (1a) time spent by user on answering the FACT questions in the scenario, (1b) number of hints given to user, and (1c) user's perception of the ease of navigating the user interface as measured in the user interface questionnaire. Results using two-tailed paired-sample t-test showed that this difference was statistically significant for (1a) and (1c), and close to significance for (1b). This showed that users found LivSaver easier to navigate than GBD-C.

Ease and accuracy of interpreting information. The ease and accuracy of interpreting information on the user interface was measured using (2a) user's score for FACT questions, and (2b) user's perception of the ease of interpreting information as measured in the user interface questionnaire. Users' scores for FACT questions had to be adjusted accordingly as some users did not manage to complete all questions due to the time limit. Results showed that this difference was statistically significant for (2a) and (2b). This showed that users could more easily and accurately interpret information on LivSaver compared to GBD-C.

Consider similar peers and trends. Whether the user considered similar peers and trends when deciding on public health priorities was measured using (3a) the justification provided by the user in recommending a certain public health priority. The justification provided by the user was coded as '1 if the user made references to peers and trends and '0' if no such references were made. Results showed that this difference was statistically significant for (3a). This showed that users were more likely to consider similar peers' and trends when deciding on public health priorities when using LivSaver versus GBD-C. In Scenario US, this also led to more users opting to recommend opioid use disorder as a public health priority (which had a lower DALY rate but had an escalating trend relative to similar peers) over ischemic heart disease.

Users' feedback on LivSaver features. Users generally found the bar charts, line charts, and the concise drop-down menus useful. Some users felt that more directions could have been given on where to find the instructions and how to interpret the charts. Users also mentioned that the map could have included more information.

	GBD-C		LivSaver		Two-tailed paired-sample t-test p-value
	Mean	SD	Mean	SD	
Ease of use					
Minutes spent answering FACT questions	18.85	2.74	13.22	4.97	<0.05*
Number of hints given	3.70	2.11	2.30	1.83	<0.1
Perceived ease of navigating UI (out of 10)	4.10	2.02	8.30	1.49	<0.05*
Ease and accuracy of interpreting information					
Adjusted score for FACT questions (%)	72.87	16.29	92.46	10.59	<0.05*
Perceived ease of interpreting info (out of 10)	5.00	2.40	7.60	2.07	<0.05*
Consider similar peers and trends					
Justification included peers and trends (0 = excluded, 1 = included)	0.2	0.42	0.6	0.52	<0.05*
No. of users (out of 5) who recommended opioid use disorder in Scenario US		2		4	n.a.
No. of users (out of 5) who recommended ischemic heart disease in Scenario US		2		1	n.a.

Figure 3. Results of the two-tailed paired-samples t-test. Significant results are denoted using *. Significant results were observed for all measures (except number of hints given).

DISCUSSION

Our study showed that with LivSaver, users took into consideration similar peers' and trend information when deciding on public health priorities. In Scenario US, this led to more users opting to recommend opioid use disorder as a public health priority over ischemic heart disease. LivSaver's algorithmic approach was also superior to single trend analysis in detecting worrying health trends relative to similar peers.

LivSaver overcame problems noted with existing micro-level and macro-level approaches towards public health by incorporating peer trends. However, this does not make LivSaver superior to these other approaches under all circumstances. Rather, LivSaver thrives in situations where the selected country has similar peers that are close in baseline values for a specific disease, injury or risk factor.

A key limitation in this study was the data limit in Tableau for the visualization created. This led to a smaller dataset ("toy dataset") being used for the visualization, which accounted for only ~10% of the original results. Nevertheless, even with the toy dataset, the strengths of LivSaver are still apparent.

Future work on LivSaver could be on refining LivSaver's algorithm approach to improve its accuracy in detecting worrying health trends relative to similar peers. Future work on LivSaver's visualization and overall user experience would focus on dashboard enhancements and website launch that incorporates background information and additional job-aids on how to interpret the data. In addition, a user group forum can be set up for stakeholders to share insights amongst each other.

CONCLUSION

This paper introduced LivSaver, which is a novel system for visualizing data-driven public health priorities. Validation studies have shown that LivSaver is effective in getting users to tackle public health issues more proactively through incorporation of peer trends. Future work can focus on refining LivSaver's algorithm approach and user interface. The timeline (Appendix E) shows the tasks completed by each team member. Each member has contributed similar amount of effort throughout the life of this project.

REFERENCES

[1] Mackenbach, J. P., & McKee, M. (Eds.). (2013). *Successes and failures of health policy in Europe: Four decades of divergent trends and converging challenges*. Maidenhead, Berkshire: Open University Press/McGraw-Hill Education.

[2] Vogel, J. P., Oxman, A. D., Glenton, C., Rosenbaum, S., Lewin, S., Gülmezoglu, A. M., & Souza, J. P. (2013). Policymakers' and other stakeholders' perceptions of key considerations for health system decisions and the presentation of evidence to inform those considerations: an international survey. *Health Research Policy and Systems*, *11*(1), 11-19.

[3] Paulozzi, L. J., Budnitz, D. S., & Xi, Y. (2006). Increasing deaths from opioid analgesics in the United States. *Pharmacoepidemiology and Drug Safety*, *15*(9), 618–627.

[4] Gross, J., & Gordon, D. B. (2019). The Strengths and Weaknesses of Current US Policy to Address Pain. *American Journal of Public Health*, *109*(1), 66–72.

[5] Institute for Health Metrics and Evaluation (IHME). (2017). *GBD-C*. Seattle, WA: IHME, University of Washington. Retrieved February 25, 2020, from http://vizhub.healthdata.org/gbd-compare.

[6] US Burden of Disease Collaborators. (2018). The State of US Health, 1990-2016: Burden of Diseases, Injuries, and Risk Factors Among US States. *JAMA*, *319*(14):1-33.

[7] Ghorbani, R., & Ghousi, R. (2019). Predictive data mining approaches in medical diagnosis: A review of some diseases prediction. *International Journal of Data and Network Science*, 47–70.

[8] Mählmann, L., Reumann, M., Evangelatos, N., & Brand, A. (2017). Big Data for Public Health Policy-Making: Policy Empowerment. *Public Health Genomics*, *20*(6), 312–320.

[9] Hughes, B. B., Kuhn, R., Peterson, C. M., Rothman, D. S., Solórzano, J. R., Mathers, C. D., & Dickson, J. R. (2011). Projections of global health outcomes from 2005 to 2060 using the International Futures integrated forecasting model. *Bulletin of the World Health Organization*, *89*(7), 478–486.

[10] Frey, Bruno S., and Stephan Meier. (2004). Social comparisons and pro-social behavior: Testing 'conditional cooperation' in a field experiment. *American Economic Review*, 94, 1717-1722.

[11] Global Burden of Disease Collaborative Network (2018). *Global Burden of Disease Study 2017 (GBD 2017) Results*. Institute for Health Metrics and Evaluation (IHME): Seattle, United States. Retrieved February 25, 2020 from http://ghdx.healthdata.org/gbd-results-tool.

[12] Stanaway, J., Afshin, A., Gakidou, E., Lim, S., Abate, D., Abbafati, K., ..., .. (2018). Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks for 195 countries and territories, 1990–2017 : A systematic analysis for the Global Burden of Disease Study 2017. LANCET, LANCET, 2018.

[13] Reibling, N., Ariaans, M., & Wendt, C. (2019). Worlds of Healthcare: A Healthcare System Typology of OECD Countries. *Health Policy*, *123*(7), 611–620.

[14] Rabel, M., Laxy, M., Thorand, B., Peters, A., Schwettmann, L., & Mess, F. (2019). Clustering of Health-Related Behavior Patterns and Demographics. Results From the Population-Based KORA S4/F4 Cohort Study. *Frontiers in Public Health*, *6*. doi: 10.3389/fpubh.2018.00387

[15] Malembaka, E. B., Karemere, H., Balaluka, G. B., Lambert, A.-S., Muneza, F., Deconinck, H., & Macq, J. (2019). A new look at population health through the lenses of cognitive, functional and social disability clustering in eastern DR Congo: a community-based cross-sectional study. *BMC Public Health*, *19*(1). doi: 10.1186/s12889-019-6431-z

[16] Austregésilo, S. C., Goes, P. S. A. D., Júnior, M. R. D. S., & Pazos, C. T. C. (2019). Clustering of oral and general health risk behaviors among adolescents. *Preventive Medicine Reports*, *15*, 100936. doi: 10.1016/j.pmedr.2019.100936

[17] Vuik, S. I., Mayer, E., & Darzi, A. (2016). A quantitative evidence base for population health: applying utilization-based cluster analysis to segment a patient population. *Population Health Metrics*, *14*(1). doi: 10.1186/s12963-016-0115-z

[18] Soedamah-Muthu, S. S., Chaturvedi, N., Witte, D. R., Stevens, L. K., Porta, M., & Fuller, J. H. (2008). Relationship Between Risk Factors and Mortality in Type 1 Diabetic Patients in Europe: The EURODIAB Prospective Complications Study (PCS). *Diabetes Care*, *31*(7), 1360–1366

[19] Chang, H.-C., Wang, M.-C., Chen, M.-H., Liao, H.-C., & Wang, Y.-H. (2018). The Impact of the

Hazard Correlation between Risk Factors and Diabetes. *International Journal of Environmental Research and Public Health*, *15*(10), 2213. doi: 10.3390/ijerph15102213

[20] Thang, T. M., & Kim, J. (2011). The Anomaly Detection by Using DBSCAN Clustering with Multiple Parameters. *2011 International Conference on Information Science and Applications*. doi: 10.1109/icisa.2011.5772437

[21] Zahner, S. J., Oliver, T. R., & Siemering, K. Q. (2014). The Mobilizing Action Toward Community Health Partnership Study: Multisector Partnerships in US Counties with Improving Health Metrics. *Preventing Chronic Disease*, *11*. doi: 10.5888/pcd11.130103

[22] Kenworthy, N. J. (2019). Crowdfunding and global health disparities: An exploratory conceptual and empirical analysis. *Globalization and Health*, *15*(S1). doi: 10.1186/s12992-019-0519-1

[23] Institute for Health Metrics and Evaluation (IHME). (2020). GBD-C. Seattle, WA: IHME, University of Washington. Retrieved April 17, 2020 from http://vizhub.healthdata.org/gbd-compare.

APPENDIX A:
Scenario US: https://forms.gle/kDJVybSU7PGgszrV7

Scenario US

You are a public health official in the United States. The president has given a directive to the department to look for ways to decrease the amount of morbidity and mortality in the US. In order to do that, you must first gather health data on not only the US but also the rest of the world to compare with. We would like you to navigate the user interface provided to you for information that would answer the specific questions that have been directed to you. Find the best answer to the FACT questions (i.e. questions where you are required to provide specific facts and figures that can be found the user interface below). Please remember that you are being timed for these questions.

* Required

(Tester to complete) Tester's Name and Participant No. *

Your answer

(Tester to complete) Interface Tool used *

Choose

In 2017, what were the Top 5 diseases/injuries (sometimes referred to as 'causes') with the highest death and disability combined (i.e. DALYs rate)? * — 1 point

- Ischemic heart disease, Alzheimer's disease, Lung cancer, Chronic obstructive pulmonary disease (COPD), Lower respiratory infection
- Ischemic heart disease, Chronic obstructive pulmonary disease (COPD), Low back pain, Opioid use disorders, Lung cancer
- Low back pain, Migraine, Diabetes Type 2, Chronic obstructive pulmonary disease (COPD), Anxiety disorders

Chronic obstructive pulmonary disease (COPD) was identified as one of the top 5 diseases/injuries. What is the value of the DALY rate and its corresponding range in 2017? * — 1 point

- Value: 2470 DALYs per 100,000; Range 2397 to 2546
- Value: 1336 DALYs per 100,000; Range: 963 to 1758
- Value: 1449 DALYs per 100,000; Range: 1338 - 1541

In 2017, what were the Top 3 risk factors with the highest death and disability combined (i.e. DALYs rate)? * — 1 point

- High body-mass index, Smoking, High fasting plasma glucose
- High systolic blood pressure, Smoking, High fasting plasma glucose
- High body-mass index, High systolic blood pressure, High fasting plasma glucose

Among the top 3 risk factors, which risk factors contributed to chronic obstructive pulmonary disease (COPD) in 2017? (You may select more than one option) * — 1 point

- [ ] High body-mass index
- [ ] High systolic blood pressure
- [ ] High-fasting plasma glucose
- [ ] Smoking

The President is interested to know how opioid use disorder has compared to other countries ("peers") since 2003. In 2003, did United States have a higher DALY rate than Canada and Australia? * — 1 point

- Higher than Canada, but lower than Australia
- Higher than Canada and Australia
- Higher than Australia, but lower than Canada
- Lower than Australia and Canada

Has opioid use disorder in the United States increased, decreased or stayed constant relative to peers (i.e. Canada and Australia) in terms of DALY rate since 2003? * — 1 point

- Increased relative to peers
- Decreased relative to peers
- Stayed constant relative to peers

Among the diseases/injuries (sometimes referred to as "causes") listed below, which causes have United States been increasing/not declining as much relative to Finland from 2003 to 2017 in terms of DALY rate? (You may select more than one option) * — 1 point

- [ ] Osteoarthritis
- [ ] Other drug use disorders
- [ ] Self-harm by firearm

Choose two countries that are close to United States in baseline values (i.e. within 20%) in 2006 for the risk factor 'Diet low in fruits') * — 1 point

- [ ] South Korea
- [ ] Sweden
- [ ] Norway
- [ ] Australia

You have finished answering the FACT questions. Please inform your tester to stop the timer and proceed to the JUDGMENT question below. For JUDGMENT question, you are required to provide your own conclusion based on the information found in the user interface. Note that for JUDGMENT questions, there are no right or wrong answers. JUDGMENT questions are not timed, but note that you will be stopped after 5 minutes.

Which of the following disease or injury would you recommend to the President to tackle as a public health priority? Feel free to use the tool to explore for more information. *

- Ischemic heart disease
- Opioid use disorder
- Chronic obstructive pulmonary disease (COPD)
- Low back pain

Please explain the rationale behind your answer above in one or two sentences. *

Your answer

You have completed all questions for this scenario. Please let your tester know. Your tester will then proceed to complete the section below.

(Tester to Complete) Time taken by participant to answer FACT question (give time in minutes, up to two decimal places) *

Your answer

(Tester to Complete) Number of hints given to participant *

Your answer

(Tester to Complete) If the participant did not complete all FACT questions before the timer stop, indicate below the last question number completed by the user.

Your answer

Submit

APPENDIX B:

Scenario Chile: https://forms.gle/zG7Vv5oC1ig3H3Fa6

Scenario Chile

You are a public health official in Chile. The president has given a directive to the department to look for ways to decrease the amount of morbidity and mortality in Chile. In order to do that, you must first gather health data on not only Chile but also the rest of the world to compare with. We would like you to navigate the user interface provided to you for information that would answer the specific questions that have been directed to you. Find the best answer to the FACT questions (i.e. questions where you are required to provide specific facts and figures that can be found the user interface below). Please remember that you are being timed for these questions.

* Required

Tester's Name and Participant No. (Tester to complete) *

Your answer

Interface Tool used (Tester to complete) *

Choose

In 2017, what were the Top 5 diseases/injuries (sometimes referred to as 'causes') with the highest death and disability combined (i.e. DALYs rate)? * (1 point)

- ◯ Ischemic heart disease, stroke, low back pain, depression, anxiety disorders
- ◯ Low back pain, ischemic heart disease, stroke, depression diabetes type 2
- ◯ Low back pain, ischemic heart disease, diabetes type 2, anxiety disorders, migraine

Low back pain was identified as one of the top 5 diseases/injuries. What is the value of the DALY rate and its corresponding range in 2017? (1 point)

- ◯ 1910 DALYs per 100,000; Range 1357.72 – 2553.33
- ◯ 2910 DALYs per 100,000; Range 1909.90 – 2739.66
- ◯ 1176 DALYs per 100,000; Range 1060.66 – 1312.19

In 2017, what were the Top 3 risk factors with the highest death and disability combined (i.e. DALYs rate)? * (1 point)

- ◯ High body-mass index, high fasting plasma glucose, high blood pressure
- ◯ High blood pressure, diet high in processed meat, high fasting plasma glucose
- ◯ High body-mass index, high blood pressure, smoking

Among the top 3 risk factors, which risk factors contributed to low back pain in 2017? (You may select more than one option) * (1 point)

- ☐ High fasting plasma glucose
- ☐ High body-mass index
- ☐ Smoking
- ☐ High systolic blood pressure

The President is interested to know how "interstitial lung disease and pulmonary sarcoidosis" has compared to other countries ("peers") since 2004. In 2004, did Chile have a higher DALY rate than Greenland and Canada? * (1 point)

- ◯ Higher than Greenland and Canada
- ◯ Lower than Greenland and Canada
- ◯ Higher than Greenland, lower than Canada
- ◯ Higher than Canada, lower than Greenland

Has "interstitial lung disease and pulmonary sarcoidosis" in Chile increased, decreased or stayed constant relative to peers (i.e. Greenland and Canada) in terms of DALY rate since 2004? * (1 point)

- ◯ Increased relative to peers
- ◯ Decreased relative to peers
- ◯ Stayed constant relative to peers

Among the diseases/injuries (sometimes referred to as "causes") listed below, which causes have Chile been increasing/not declining as much relative to Greenland from 2004 to 2017 in terms of DALY rate? (You may select more than one option) * (1 point)

- ☐ Interstitial lung disease and pulmonary sarcoidosis
- ☐ Congenital heart anomalies
- ☐ Anxiety disorders

Choose two countries that are close to Chile in baseline values (i.e. within 20%) in 2006 for the risk factor 'High systolic blood pressure') * (1 point)

- ☐ Guam
- ☐ Switzerland
- ☐ Mexico
- ☐ Canada

You have finished answering the FACT questions for Scenario 1. Please inform your tester to stop the timer and proceed to the JUDGMENT question below. For JUDGMENT question, you are required to provide your own conclusion based on the information found in the user interface. Note that for JUDGMENT questions, there are no right or wrong answers. JUDGMENT questions are not timed, but note that you will be stopped after 5 minutes.

Which of the following disease or injury would you recommend to the President to tackle as a public health priority? Feel free to use the tool to explore for more information. *

- ◯ Low back pain
- ◯ Ischemic heart disease
- ◯ Diabetes mellitus type 2
- ◯ Anxiety disorders

Please explain the rationale behind your answer above in one or two sentences. *

Your answer

You have completed all questions for this scenario. Please let your tester know. Your tester will then proceed to complete the section below.

(Tester to Complete) Time taken by participant to answer FACT question (give time in minutes, up to two decimal places) *

Your answer

(Tester to Complete) Number of hints given to participant *

Your answer

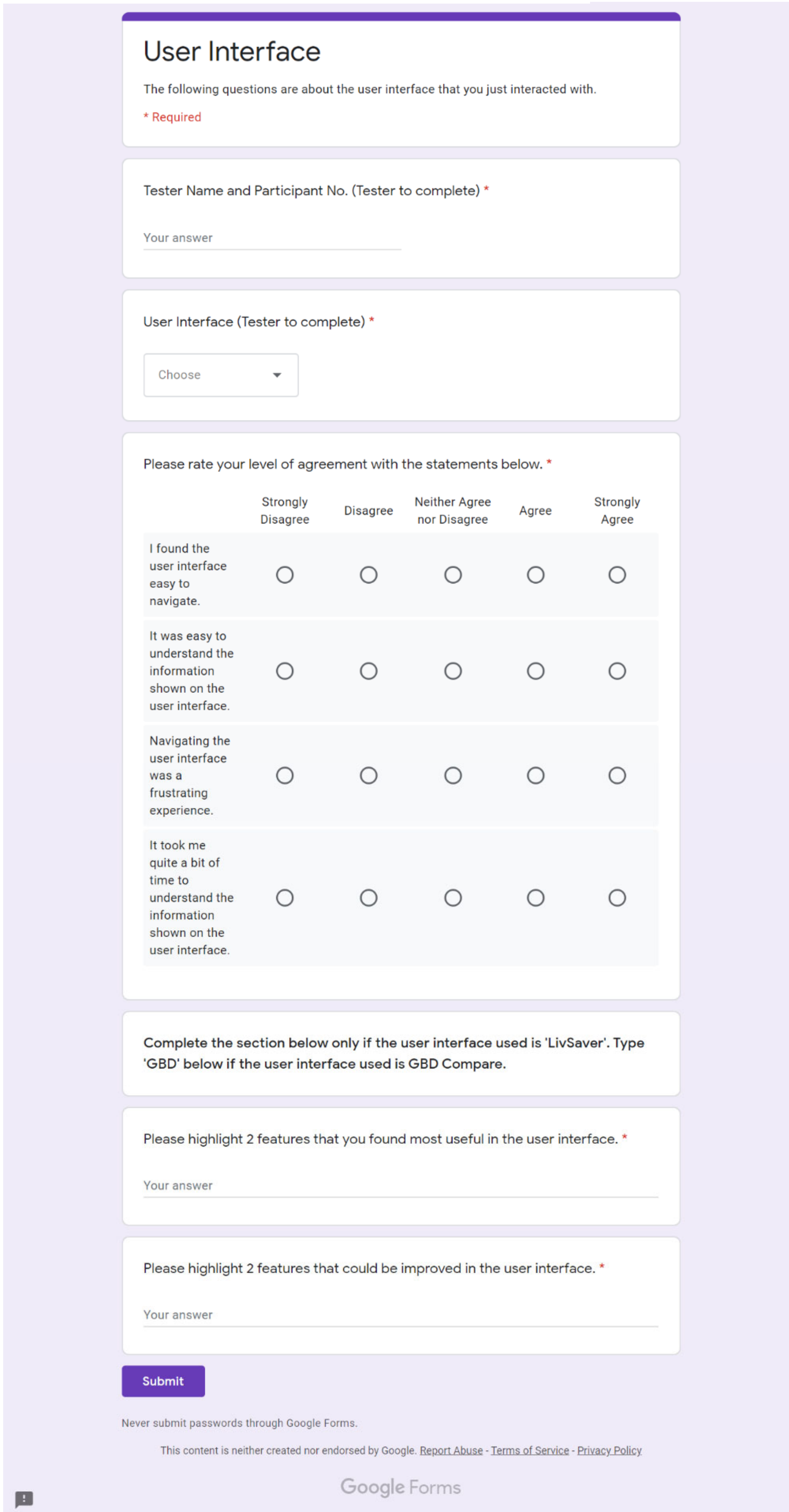
(Tester to Complete) If the participant did not complete all FACT questions before the timer stop, indicate below the last question number completed by the user.

Your answer

Submit

APPENDIX C:

User Interface: https://forms.gle/TtBrP8m9wijmqAKu6



User Interface

The following questions are about the user interface that you just interacted with.

* Required

Tester Name and Participant No. (Tester to complete) *

Your answer

User Interface (Tester to complete) *

Choose

Please rate your level of agreement with the statements below. *

	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
I found the user interface easy to navigate.	○	○	○	○	○
It was easy to understand the information shown on the user interface.	○	○	○	○	○
Navigating the user interface was a frustrating experience.	○	○	○	○	○
It took me quite a bit of time to understand the information shown on the user interface.	○	○	○	○	○

Complete the section below only if the user interface used is 'LivSaver'. Type 'GBD' below if the user interface used is GBD Compare.

Please highlight 2 features that you found most useful in the user interface. *

Your answer

Please highlight 2 features that could be improved in the user interface. *

Your answer

Submit

Never submit passwords through Google Forms.

APPENDIX D:
User Script

(Ensure the visualization links and user interface questionnaire links are all opened and minimized before the user study starts. Make sure 2 user interface questionnaires are open. Also ensure that the starting 'tester-to-complete' portions for all questionnaires are completed before the study begins.)

"Hello and thank you for participating in our study. In this study, you will be using two different user interfaces assuming the persona of a public health official in response to two scenarios.

The scenarios will be shown on the questionnaire given to you *(Direct participant to refer to online questionnaire in Scenario 1)*. In each scenario, you will be required to answer specific questions that have been directed to you in your capacity as a public health official. Two types of questions will be asked. The first type are FACT questions where you are required to provide specific facts and figures that can be found using the user interface provided to you. Your responses to these questions will be timed and scored. You should aim to find the correct information for these questions as quickly as possible. If you feel stuck at any point in time when answering the FACT questions, you may ask for a hint from me. However, note that the number of hints you asked for will be measured.

The second type are JUDGMENT questions where you are required to provide your own conclusion based on the information found in the user interface. JUDGMENT questions are not timed. Note that for JUDGMENT questions, there are no right or wrong answers so long as a reasonable justification is provided.

After you have completed the FACT questions for the first scenario, please let me know so that I can record your timing. If the time exceeds 20 minutes, you will be asked to stop. You will then be asked to complete the JUDGMENT question. After completion of the JUDGMENT question, you will be asked to rate your user experience with the user interface that you have just interacted with before proceeding to the next scenario.

You may now begin answering the FACT questions for the first scenario. When you are ready to start, click on the <browser/Tableau> icon and I will begin the timer. *(Tester to begin timing)*

User completes FACT questions for Scenario 1.

You may now begin answering the JUDGMENT question. While this question is not timed, you will be asked to stop after 5 minutes. You may begin.

User completes JUDGMENT question for Scenario 1.

(Tester to complete the sections on time taken and number of hints given before submitting the questionnaire).

You have completed Scenario 1. You will now be asked to complete a questionnaire about the user interface that you just interacted with.

(Direct participant to refer to user interface questionnaire).

User completes user interface questionnaire.

We will now move on to Scenario 2. You may now begin answering the FACT questions for the second scenario. Please remember that you are being timed and will be asked to stop if the time exceeds 20 minutes. When you are ready to start, click on the <browser/Tableau> icon and I will begin the timer. *(Tester to begin timing)*

User completes FACT questions for Scenario 2.

You may now begin answering the JUDGMENT question. While this question is not timed, you will be asked to stop after 5 minutes. You may begin.

User completes JUDGMENT question for Scenario 2.

(Tester to complete the sections on time taken and number of hints given before submitting the questionnaire).

You have completed Scenario 2. You will now be asked to complete a questionnaire about the user interface that you just interacted with.

(Direct participant to refer to user interface questionnaire).

User completes user interface questionnaire.

Thank you for completing the study.

<The debrief below is optional. Feel free to share with the participant if he/she requests for it.>

To share, the objective of the user-study is to evaluate user interfaces that enable public health authorities and other stakeholders to deal with public health issues more proactively. The solution developed in our project is LivSaver which uses global health data from the Global Burden of Disease (GBD) study. In the study that you just did, we evaluated it against another visualization tool - GBD compare tool which also uses the same dataset. If you are interested in the results of the study, please let me know and I will update with the results when the analysis is done.

APPENDIX E:

Project Timeline – Planned and Actual Dates