James Hicklin
CPSC 547 Project Proposal

Visualizing Uncertainty: Helping Atrial Fibrillation Patients Understand That Risk Estimates Are Imprecise

Project Member(s)

James Hicklin – hicklin.james@gmail.com

Description of domain, task, and dataset

Uncertainty as a general domain has historically proven to be difficult to visualize in an easy to understand way. While researchers in some fields have made attempts to help users understand the complexity surrounding uncertainty and ambiguity, efforts have been limited to targeting specific groups of people, who are typically experts in their relevant fields, as opposed to more general populations.

Domain

Our domain of focus is in a patient context, where significant tradeoff between the pros and cons of therapy must occur in order for decision making to be optimal. A typical therapy decision consists of a set of distinct options along with a set of attributes related to those options. For example, in atrial fibrillation, patients must decide whether the potential risk of side effects (e.g. increased risk of major internal bleed) outweigh the potential benefits (e.g. stroke risk reduction) when choosing whether to take warfarin as therapy. These risk and benefit estimates always have some degree of ambiguity around them, yet this uncertainty is rarely presented to the patient. Nevertheless, this ambiguity may have an important effect on the patient's decision, especially when the confidence intervals around the provided point estimates overlap. This becomes especially problematic in the case of new therapies, where risk estimates are not well established and wide confidence intervals may exist.

An obvious solution to more generally visualize uncertainty is to incrementally add and encode the additional facets of uncertainty. Kao et. Al[1] do this in 3 dimensions, by deforming the 3D surface based on its standard deviation, encoding the interquartile range by adjusting color hue, and adding an additional line channel to encode the difference between mean and median. While this is an effective approach in the expert use case, our solution needs to be simple and intuitive so that it is understandable from the perspective of a more general population. We therefore restrict our solution to 2 dimensions, and make no attempts to visually encode every aspect of uncertainty on the same plot.

Additionally, we chose to focus our domain of interest to a specific disease process, in order to ensure access to quality data. We aim to implement our project with generalizability in mind, such that the system can be easily adapted to other disease processes.

James Hicklin
CPSC 547 Project Proposal

Task

Given the domain issues identified earlier, the primary high-level goal of the system is to improve patient understanding of uncertainty. We can narrow this down to more concrete tasks:

- Compare different methods of visualizing uncertainty, including:
  - Isotypes
  - Violin plots
  - Gradient plots

  These methods were chosen based off of a review of current methods to visualize uncertainty that identified which methods were superior[2].
- Compare how regions of uncertainty can be visualized differently, and how that may influence understanding of uncertainty (e.g. comparing scalar data uncertainty as confidence intervals vs. probability distributions).
- Manipulate the degree to which uncertainty is represented (e.g. 95% confidence intervals vs. 99% confidence intervals), and the type of point estimate presented.
- Compare best & worst case scenarios.

Dataset

The table below illustrates the raw data and the derived data needed to create the visualizations. Each representation of uncertainty only needs a subset of the derived data. The raw data will come from a meta-analysis of the risks of atrial fibrillation medication side effects[3].

Raw Data	Baseline risk (Bayesian prior) Probability distribution function (PDF) *Relative risk delta *Standard deviation * = Individual sample results, which are what inform the PDF
Derived Data	Minimum Maximum First quartile Third quartile Median Mean 95% CI 99% CI Variance Inverse cumulative probability function Natural Frequencies

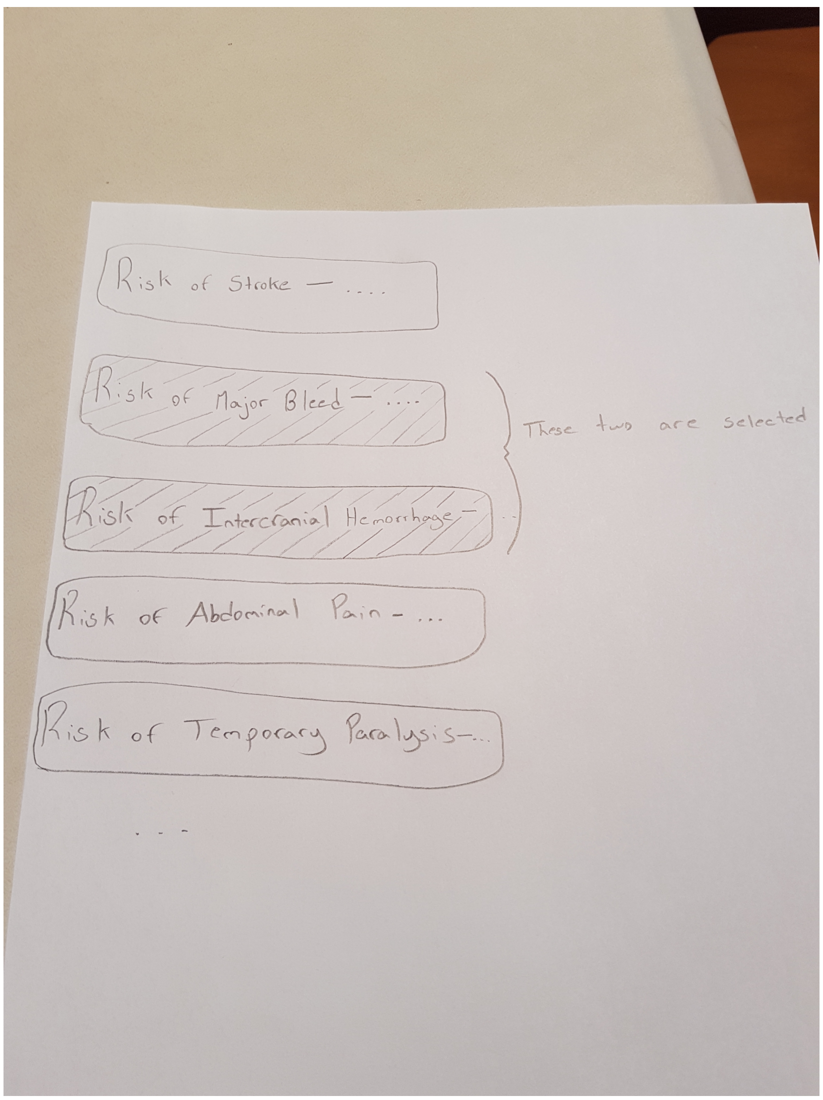
James Hicklin
CPSC 547 Project Proposal

Personal Expertise

My thesis work involves looking at how patients understand second-order, or epistemic uncertainty. We are looking at whether some representations improve knowledge, patient understanding, decisional conflict, and other outcomes with respect to how (and whether) uncertainty is presented, and whether there is an ethical imperative to do so. I have extensive experience in developing software to support shared decision-making.

Proposed Solution

The first step of our solution involves assessing a baseline importance for the items related to atrial fibrillation. This will better inform a default view, as patients are heterogeneous in terms of how they value the various items. We will present each item, and ask users to select the two most important items.



Once we have ascertained important items, we present the default window. This composes of multiple views:

1. The toolbox – this is where settings are changed – it may consist of sliders, dropdown selections, checkboxes, radio buttons, etc.
2. The comparison view – this shows your selected items in the visualization that you have selected

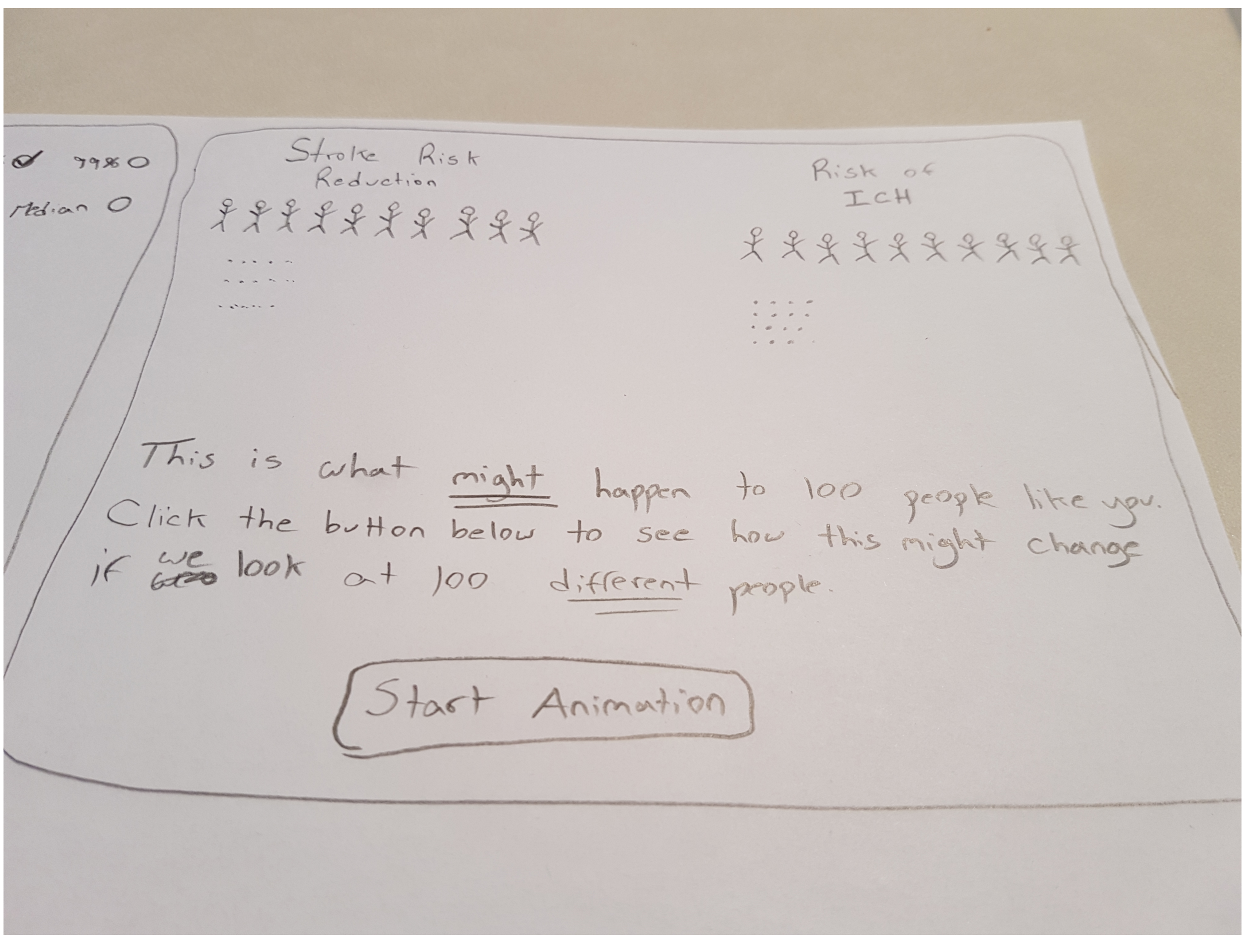
These two views will sit side by side. Depending on which comparison view type is selected, data views may be juxtaposed side by side. Additionally, depending on how implementation goes, we may present multiple comparison views at once. We could then highlight specific aspects of the uncertainty (e.g. the confidence interval), and compare how that specific aspect is presented across different visualizations of uncertainty.

Comparison View:
This is where the majority of uncertainty comparisons will take place. Initially, this will be set to show isotype visualizations, as these are currently preferred in the health domain due to their inherent nature of presenting risk as natural frequencies[4], as opposed to relative percentage changes. The following section will go through each of the potential visualization options.

James Hicklin
CPSC 547 Project Proposal

Isotypes

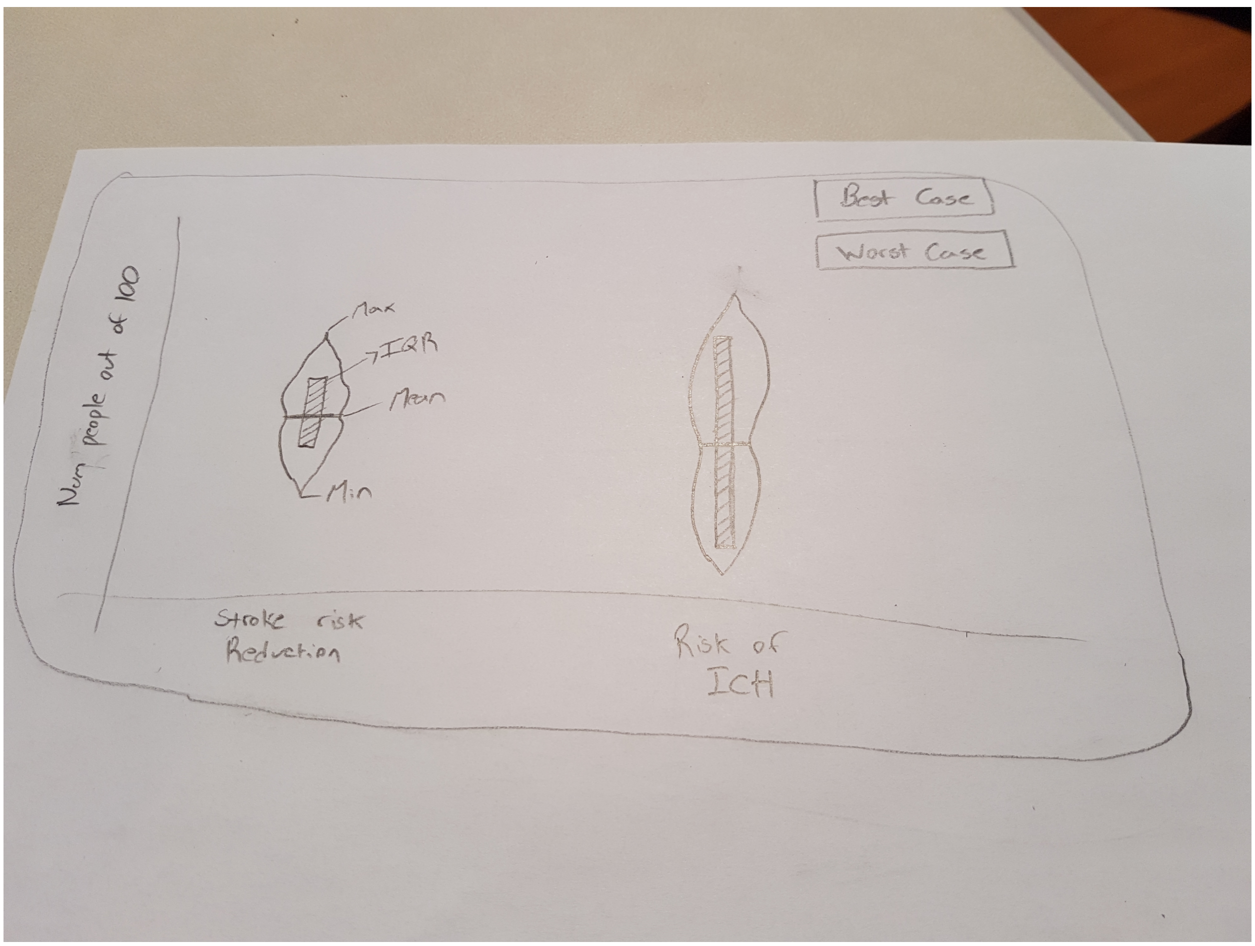
Idiom	Isotype (Icon array)
What: Data	2 quantitative values (baseline risk, relative risk delta) 1 binary value (direction of delta) 1 categorical value (item – e.g. stroke risk, risk of ICH, etc.)
How: Encode	Baseline risk and relative risk delta values encoded by isotype marks (stick figures or faces), direction of delta (ie. positive or negative) encoded by color hue.
Why: Task	Visualizing uncertainty & ambiguity in an icon array requires some form of animation to visualize how different samples might change the risk estimates. We propose randomly sampling the distributions, which will result in the isotype vis changing to show different people and different numbers of people being affected by the risk delta (ie. first-order and second-order uncertainty).



James Hicklin
CPSC 547 Project Proposal

Violin Plots

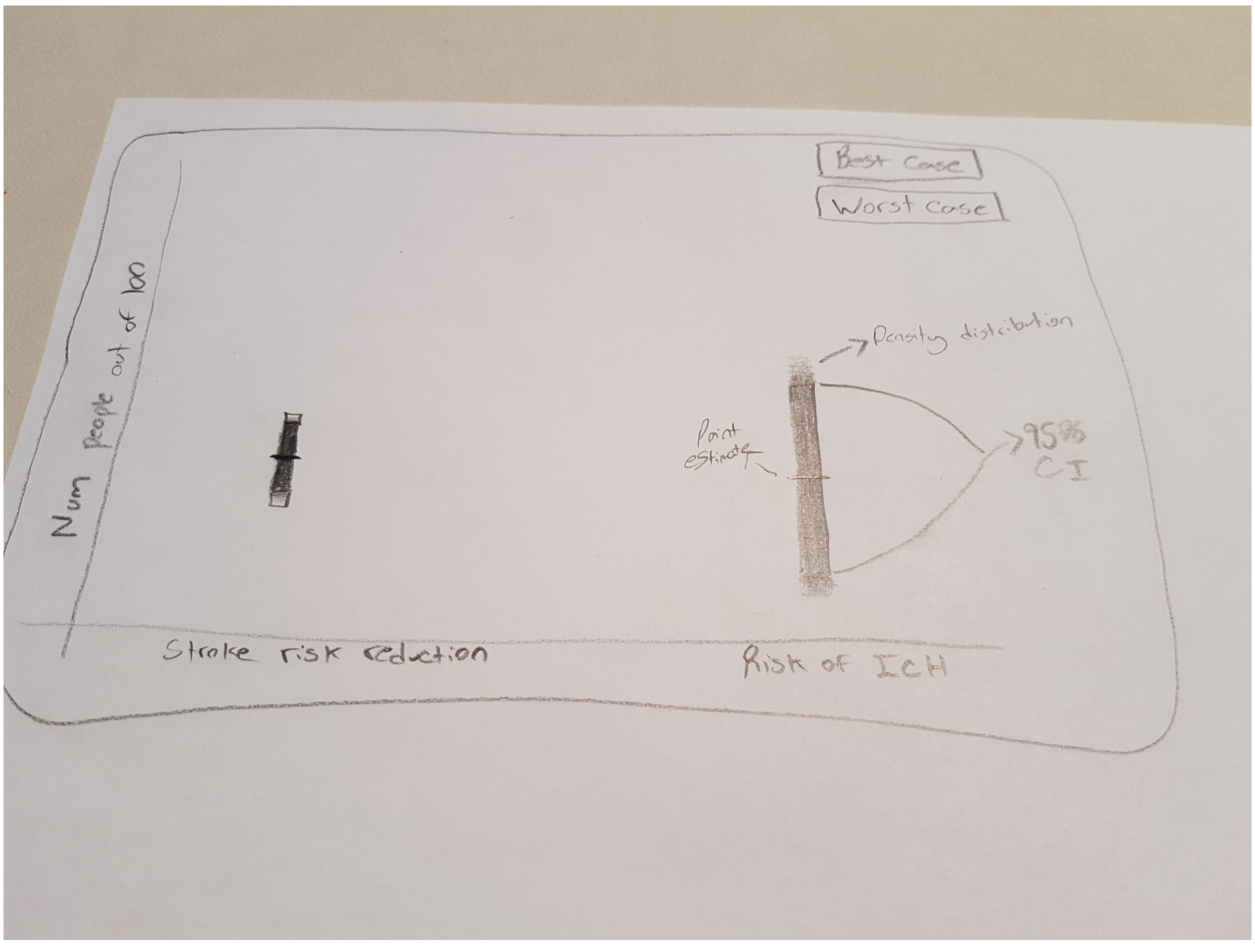
Idiom	Violin plot
What: Data	5 quantitative values (Minimum, maximum, interquartile range, mean, distribution) 1 categorical value (item – e.g. stroke risk, risk of ICH, etc.)
How: Encode	Minimum and maximum encoded by mark position on y axis, IQR encoded by length of line, mean encoded by mark position on y axis, distribution encoded by area.
Why: Task	Compare distributions, best & worst case scenarios (max and min), variability (IQR)



James Hicklin
CPSC 547 Project Proposal

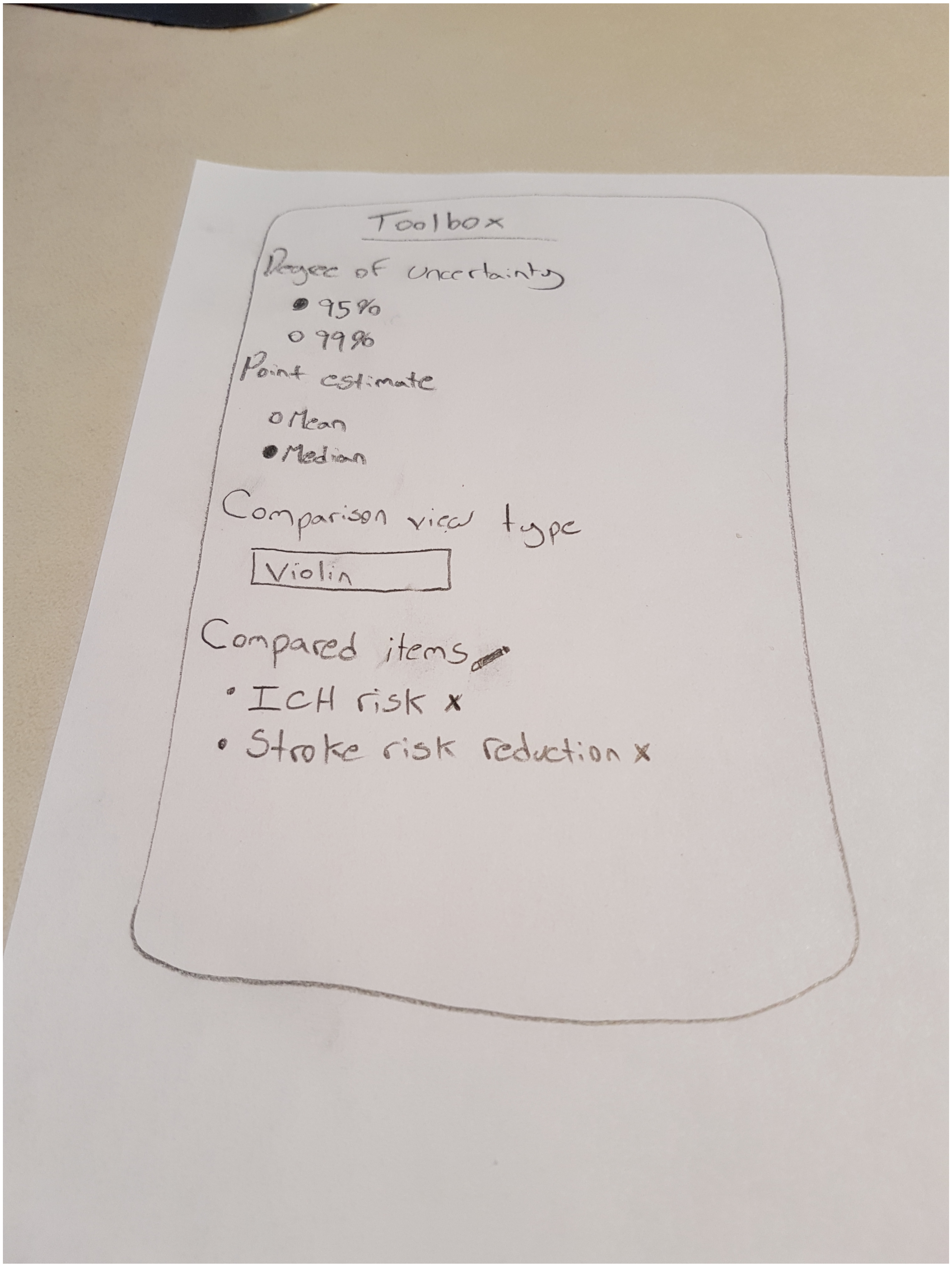
Gradient Plots

Idiom	Gradient Plot
What: Data	3 quantitative values (mean, distribution, 95% confidence interval) 1 categorical value (item – e.g. stroke risk, risk of ICH, etc.)
How: Encode	Mean encoded by mark position on y axis, 95% confidence interval encoded by length of line, distribution outside 95% CI encoded by color saturation
Why: Task	Compare distributions and how likely true mean lies outside 95% CI



James Hicklin
CPSC 547 Project Proposal

Toolbox View:
This view is where most user interaction will occur. The comparison view type can be changed here, along with the degree of uncertainty, mean vs. median point estimate comparisons,



James Hicklin
CPSC 547 Project Proposal

Use Case

Imagine you are a patient who just left your physician's office with a diagnosis of atrial fibrillation. Your doctor has given you a link to a website and sent you home to look at the site, and try to decide for yourself whether you think you should take warfarin or not. Upon opening your browser and navigating to the URL, the first thing you see is a list of items associated with atrial fibrillation. You are asked to choose those that you think are most important. You choose "Reducing my risk of stroke" and "Possibility of major internal bleeding" as the two most important items, and after clicking "Begin", you are presented with the vis tool. The default view corresponds to those items that you selected as being important, but you can optionally add or remove items from those being presented. The default presentation uses isotype visualizations, as those are dominant in health.

You decide that the isotype visualization is difficult to understand, as the point estimates appear to be the same. The probability of decreasing your risk of stroke is the same as your probability of experiencing major internal bleeding if you start taking warfarin, and you are therefore conflicted. Using the toolbox, you click on the "Graphic type" dropdown and choose "Violin" instead of "Isotype". The comparison view isotypes animate out of the comparison, and a new violin plot animates into the comparison view. You take a look at the point estimates, and again notice that they are the same. But the violins look quite different, so you take a close look at their shape and notice that the interquartile range of stroke risk reduction is three times as wide as that for the risk of internal bleeding.

You are starting to feel that the stroke risk reduction estimates you have been given are not very accurate, and are a bit of a worrywart, so you go back to the toolbox and click on the "Worst Case Scenario" button. The comparison view changes back to "Isotype", but this time, the natural frequency values are vastly different. Your risk of major bleed is showing as much greater than the probability of stroke risk reduction, based off the 95% or 99% confidence interval of the input data. You decide that taking warfarin may be quite risky, and write down some discussion points for the next time you meet with your doctor to better inform your final decision.

Implementation Approach

We propose using D3 within an AngularJS application to implement our solution. These are being chosen because the project creator has experience with both these technologies, and can therefore focus more on the data visualization itself as opposed to learning new technologies. The system can be deployed to my personal website, jameshicklin.com, for demo purposes.

James Hicklin
CPSC 547 Project Proposal

Schedule

Task	Completion date
Extract input data from meta-analysis and convert to generic CSV that system will use	March 17th
Implement initial item screen that informs default comparison in main vis window	March 24th
Implement violin plots and isotype visualizations	March 31st
Implement gradient plots	April 7th
Complete all plot & isotype implementations	April 7th
Implement toolbox and work on visualization animation as parameters are changed within toolbox	April 14th
Unit test and fix remaining issues, begin work on presentation and paper	April 21st
Presentation	April 25th
Paper due	April 28th

Previous Work

A variety of work has been done in the field of uncertainty visualization, but few attempts have been made to assess how average users understand uncertainty. One study that did assess the general population was that done by Correll and Gleicher, in which they noted that traditional error bars often yield to a binary interpretation of uncertainty, in which values either lie within a range or do not. They concluded that it is important for encodings to be both visually symmetric and visually continuous[2]. Violin plots and gradient plots are examples of visualizations that satisfy both these constraints.

References

1. Kao, D., J.l. Dungan, and A. Pang. "Visualizing 2D probability distributions from EOS satellite image-derived data sets: a case study." *Proceedings Visualization, 2001. VIS '01.* (n.d.): n. pag. Web.
2. Correll, Michael, and Michael Gleicher. "Error Bars Considered Harmful: Exploring Alternate Encodings for Mean and Error." *IEEE Transactions on Visualization and Computer Graphics* 20.12 (2014): 2142-151. Web.
3. Benavente, O., R. Hart, P. Koudstaal, A. Laupacis, and R. Mcbride. "Oral anticoagulants for preventing stroke in patients with non-valvular atrial fibrillation and no previous history of stroke or transient ischemic attacks." *The Cochrane Database of Systematic Reviews* (1999): n. pag. Web.
4. Gigerenzer, G. "Simple tools for understanding risks: from innumeracy to insight." *Bmj* 327.7417 (2003): 741-44. Web.