



# A Visual Decision Support Tool for Appendectomy Care

Edward Clarkson<sup>1</sup> · Jason Zutty<sup>1</sup> · Mehul V. Raval<sup>2</sup>

Received: 16 January 2018 / Accepted: 18 January 2018 / Published online: 5 February 2018  
© Springer Science+Business Media, LLC, part of Springer Nature 2018

## Abstract

Appendectomy is the most common abdominal surgical procedure performed in children in the United States. In order to assist care providers in creating treatment plans for the postoperative management of pediatric appendicitis, we have developed a predictive statistical model of outcomes on which we have built a prototype decision aid application. The model, trained on 3724 anonymized care records and evaluated on a separate set of 2205 cases from a tertiary care center, achieves 97.0% specificity, 25.1% true sensitivity, and 58.8% precision. We have also built an interactive decision support tool augmented with simple visualization techniques designed for clinicians to use in the course of making care decisions (e.g., discharge) and in patient/stakeholder communication. Its focus is on end-user ease of use and integration into existing clinician workflows, and is designed to evolve its predictions as more and better data become available.

**Keywords** Decision support · Visualization · Clinical care · Appendicitis · Care pathways · Pediatric surgery

## Introduction

The volume of healthcare data is rising at an exponential rate; harnessing its utility through data-driven prediction of healthcare outcomes has long been touted as a useful tool for the future of medicine. Unfortunately, there are considerable obstacles to widespread adoption of data-driven prediction models and the translation of research in machine learning and similar domains to real world medical care [1, 2]. Focusing on a segment of the healthcare system and patient encounter is necessary to simplify initial attempts within this space.

Surgery represents a finite care experience with three distinct phases. The preoperative phase focuses on establishing the diagnosis and ends with assessment of risk and benefit of

surgical intervention. The intraoperative phase revolves around the surgical procedure taking place, ideally lasting only for the duration of the procedure. The final postoperative phase involves the recovery from surgically treated disease and the stress induced by the surgical intervention, typically ending with discharge from an acute care setting.

While a large body of work with machine learning has focused on diagnosis and the preoperative phase of surgical care, less attention has been placed on optimizing the postoperative phase. The postoperative phase has been the focus of numerous quality improvement efforts to streamline or “fast track” care in order to optimize outcomes such as decreasing length of stay (LOS), but most of these efforts are based on anecdotal experience and are not data driven.

Appendicitis represents the most common surgical emergency for which children undergo abdominal surgery. Approximately 30 to 40% of appendicitis cases are for complicated appendicitis where often the appendix has ruptured [3]. The postoperative management of complicated appendicitis compared to acute, early appendicitis is highly variable with higher rates of complications, longer hospitalizations, higher readmission rates, and higher costs [4]. Any efforts to decrease in hospital LOS for these complex patients must be balanced with concerns over readmission to the acute care setting [5].

The purpose of our study is twofold: to develop a statistical model to provide data-driven recommendations for the

---

This article is part of the Topical Collection on *Systems-Level Quality Improvement*

---

✉ Edward Clarkson  
edward.clarkson@gtri.gatech.edu

<sup>1</sup> Electro-Optical Systems Lab, Georgia Tech Research Institute, 925 Dalney St., Atlanta, GA 30332-0834, USA

<sup>2</sup> Division of Pediatric Surgery, Department of Surgery, Emory University School of Medicine and Children’s Healthcare of Atlanta, Atlanta, GA 30322, USA

postoperative management of complicated appendicitis; and create an interface to this model suitable for use by clinicians (i.e., not computing experts) making care decisions and advising patients. We aimed to identify factors that would optimize LOS while also providing providers and patients with information regarding complications, including readmission.

The result of this work is a prototype smart protocol application driven by an iteratively-refined data model and a visual interface, and suggestions of effective mechanisms for projects in similar domains.

## Method

Creating a mechanism for care providers to incorporate data into their work via evidence-based smart protocols requires a careful marriage of predictive data models and end-user interface (UI); the goal is to allow medical professionals to access rich data models without discouraging their adoption because of complexity or other ease-of-use factors.

### Predictive data model

Children's Healthcare of Atlanta (CHOA) serves as a comprehensive regional referral center for pediatric patients in the southeastern United States. Within the CHOA system exist two free-standing children's hospitals that routinely provide surgical care for acute conditions. We collected data from 5929 appendectomy cases from 2009 to 2014 that included more than 120 pre-, intra- and post-operative appendectomy attributes. These data were segmented into training (2009–2010 and 2012–2013) and validation (2011, 2014) sets, consisting of 3724 and 2205 cases, respectively. We focused on *complicated* appendicitis cases, defined as those with LOS of more than two days [11]: 1919 training cases and 1113 validation cases were complicated. We defined a case as having a *negative outcome* if any of the following were true for a particular patient: an occurrence of urinary tract infection (UTI), sepsis, deep/organ-space or superficial surgical site infection (SSI), or pneumonia; a follow up procedure was performed (reoperation); the patient returned to the emergency room (return to the system); or the patient was readmitted to the hospital (readmission). The data not collapsed into the negative outcomes variable were used as features in the following manner:

- All columns in with two choices were treated as Boolean variables (e.g. Gender, Open/Laparoscopic)
- All numeric columns remained as such (e.g. Age, Height)
- All remaining columns were vectorized by making a Boolean column for the presence of each choice (e.g. 1st IV Antibiotic, 1st Procedure)

The requirements for the model were that it be evidence driven, respond to queries within 2 s (a general man-machine “conversational” guideline [6]), require simple input from the user, reflect results for similar patients to the current patient, and report confidence intervals for probability of negative outcome and LOS.

### User interface

Our model operates as follows, starting from the context of all historical patient data:

- Select the column that has the highest correlation with the negative outcomes variable on the remaining data.
- Ask the user to indicate the patient's value for the column.
- Go to the first step after filtering the data based on the previous answer.
  - For numeric answers, the model chooses the 50% of remaining subjects closest to the value the user answered.
  - For non-numeric answers, the model selects the data matching the answer.

By choosing an iterative technique, the model provides at each stage an estimate of the LOS and chance of negative outcome for the patient along with confidence estimates. This allows users to make their own choices about how much effort to produce (i.e., how many questions to answer) while still receiving useful information at every step.

This iterative approach extracts a tree of relevant patient/care variables wherein different questions may be asked depending on prior answers (forming a “pathway” through the tree for a specific patient and time). The tree itself may change over time as data about new patients becomes available; as a result, the model can adapt to changes in care protocols, driven by tools such as this one or by other factors like changes to antibiotic regimens.

### Design alternatives

We considered several machine learning models as alternatives to our approach, which we call “dynamic filtering”. These included traditional techniques such as naïve Bayes, decision trees, nearest neighbors, and random forest classifiers, as well as cascades of the same. These methods are designed to take in a key subset of features deemed to be most predictive of negative outcomes, and predict whether a patient would have a negative outcome. Though we are able to calculate sensitivity, specificity, and positive predictive value, these alternative methods do not yield confidence estimates. Furthermore, the runtime performance of these models was also considerably slower, usually significantly longer than our 10 s threshold for a given step.

## Results

### Predictive data model

Negative outcomes were identified in 455 of the 1919 training (23.7%) and 250 of 1113 validation (22.5%) complicated appendicitis cases. The incidence of negative outcomes in early appendicitis cases were – 125 of 1805 for training (6.9%) and 16 of 1092 for validation (1.46%) cohorts. The initial model from this data achieves 97.0% specificity, 25.1% true sensitivity, and 58.8% precision (or an overall accuracy of 87%: 88%/59% negative/positive predicted value; 75%/3% false negative/positive error rates).

### User interface

Figure 1 displays the opening query of the UI. These questions represent the tension between a strictly data-driven approach and one that is created with an eye to clinician experience and face validity. While these three data items are all prominently featured in the various variable tree pathways, they are not the three most salient variables in all pathways. However, clinicians are likely to know the answers to these questions without having to refer to a reference, lowering the barrier to the first usage set and building investment in a usage session. Moreover, these questions are clinically relevant and build face-validity with clinicians as a tool that asks relevant questions. Finally, there is value in consistency in UI design, which dramatically varying question paths violates.

The model filters the data based on the answers; once the initial questions are answered, subsequent questions are dynamically selected based on which remaining variable has the highest correlation with the combined negative outcomes, based on the current slice of relevant data. The dynamic approach to generating questions asks the most relevant questions based on past answers, which in turn reduces the total number of questions required to generate an accurate prediction for the patient.

At any given point, based on the user’s answered questions, the interface displays an outcomes table augmented with visual representations of its data (Figs. 2 and 3). The table contains the number of similar patients along with their average LOS and the probability of a negative outcome, each of which have 95% confidence intervals.

If the next question is a discrete choice (such as a yes/no, or choice of medication), the interface will display how the choices impact the outcome statistics. An example of this is shown in Fig. 2, where a user is prompted for a second post-operative oral antibiotic choice: it is clear that most patients were not given a second antibiotic.

The system groups similar questions into a single “multiprompt” question when any of the group is dynamically selected: in Fig. 4, all recorded comorbidities are queried at the same time. The multiprompt question accomplishes two useful effects over completely independent questioning. The first is consistency: the question ordering, and consequently the information flow, is the same for multiprompt subparts between patients, and that ordering matches natural expectations

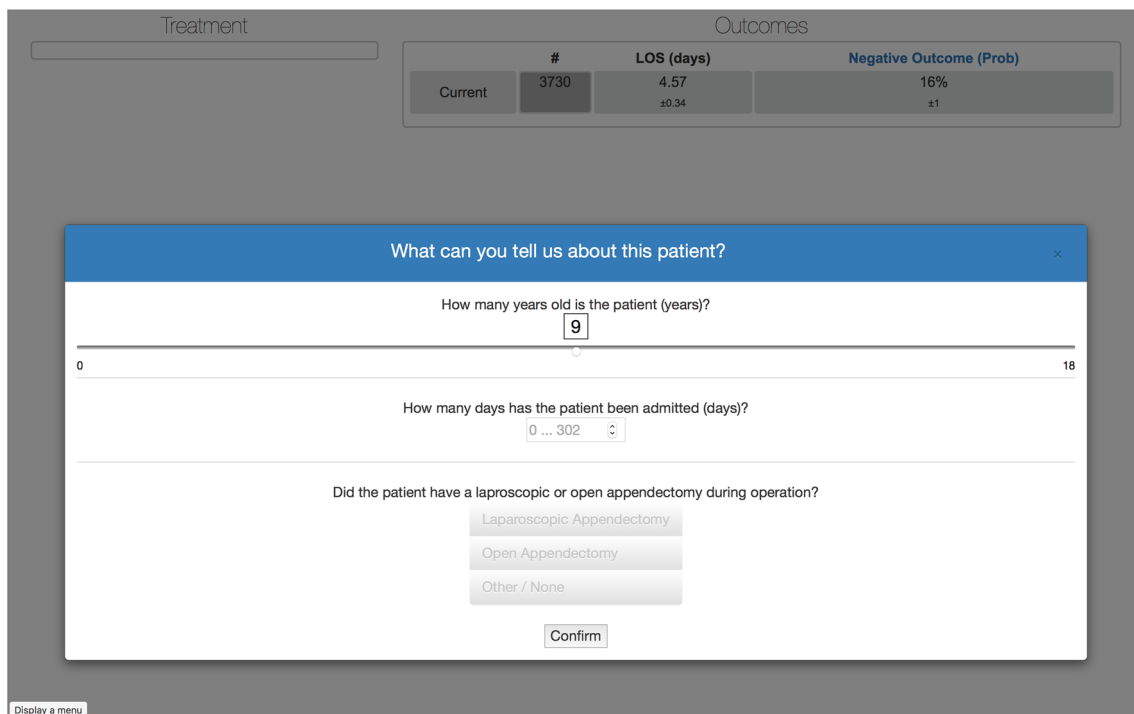
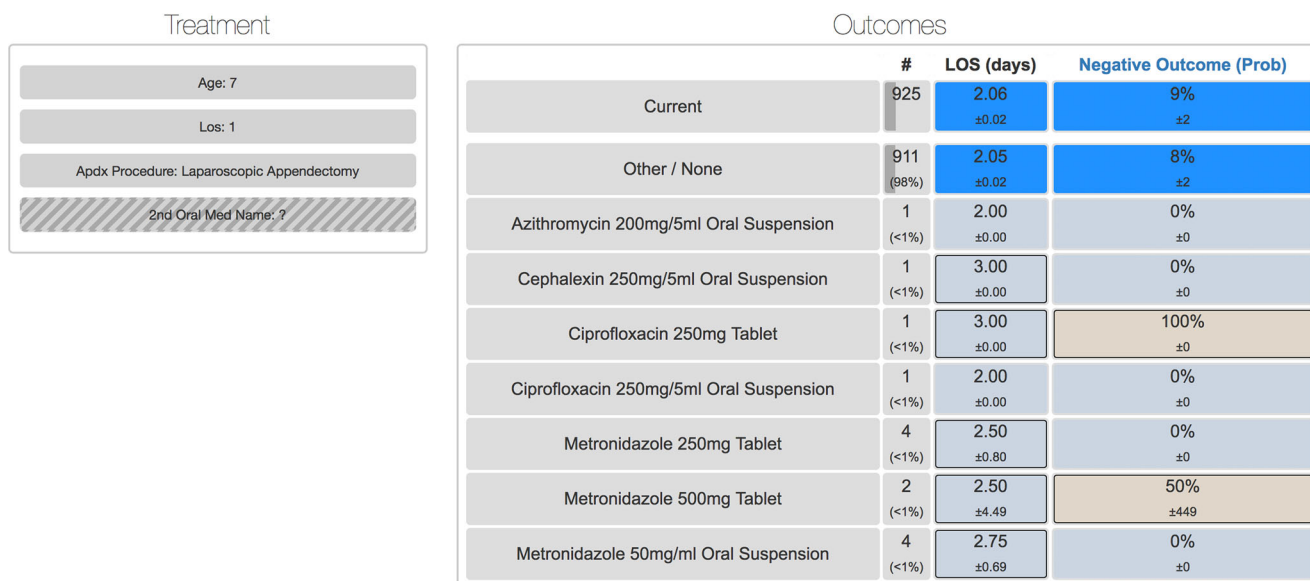


Fig. 1 Initial question prompt given to every user



**Fig. 2** A clinical outcomes table with care options, outcome predictions, and visualizations of those outcomes compared to all pediatric appendectomy patients and ones most similar to the patient information provided

of related questions appearing together. Secondly, the multiprompt question helps to reduce decision fatigue on the user: fewer dialogs give the appearance of fewer questions.

The model will continue to generate questions to the user until they have selected a response that causes the number of similar patients to drop below 10, in which case the current statistics will be displayed.

We have implemented this system as a web-based application optimized for mobile platforms, which are the most regularly-used devices among clinical staff; the use of a web-based environment rather than a native application also bypasses application deployment problems.

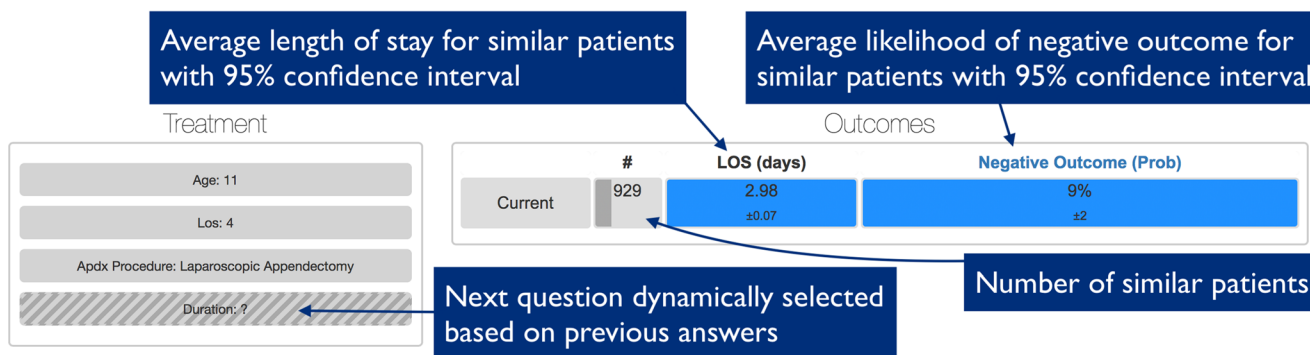
**Design alternatives**

The presentation of the data in a purely tabular form isolates it from important context, such as prediction deviations from overall averages; the relative deviation of prospective choices; and uncertainty estimates. Visual presentations of data are an

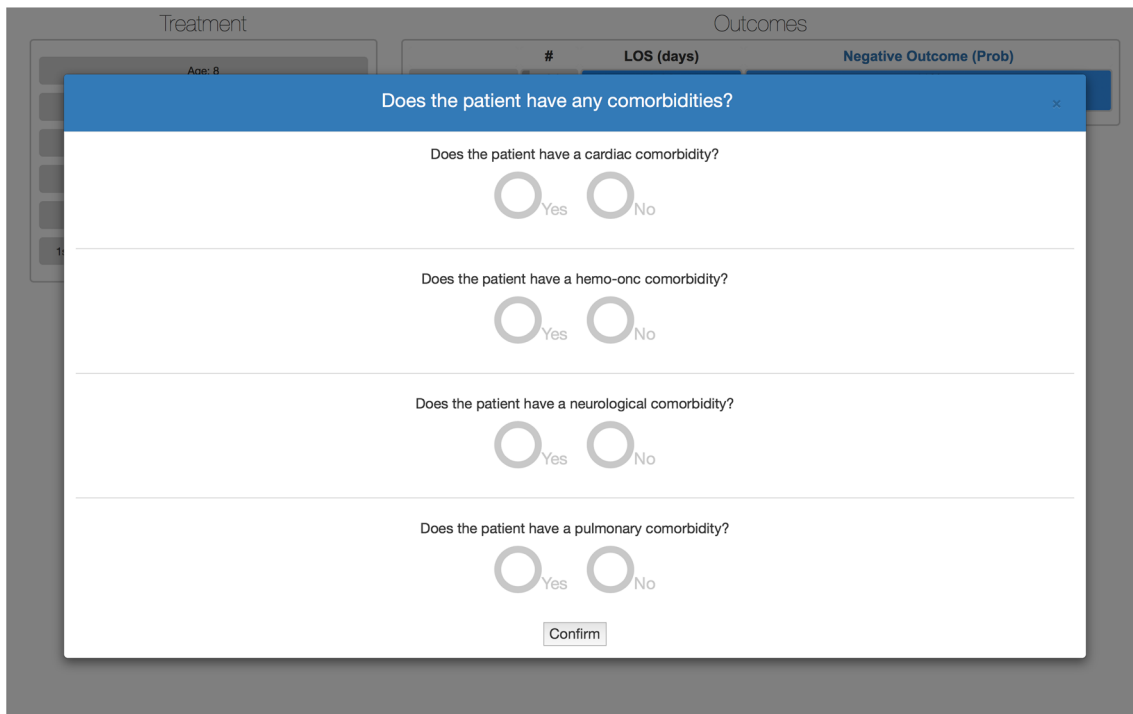
obvious candidate for helping make these aspects more immediately apparent, but we are faced with several challenges:

- Clinical users may access the tool on a range of devices, so the design must both adapt to these screen orientations and resolutions, while also maintaining familiarity across these contexts.
- The performance of the overall system needs to be within “conversational” (~2 s) UI bounds [6].
- Visualization of both the predicted result and the model’s uncertainty are both important components.
- Complex visualizations that require extensive training are unlikely to be adopted.

We therefore chose to insert a few key visual elements for our initial prototype: an outcome cell’s color is assigned according to whether it is better (blue) or worse (red). We then calculate a z-score of the current cell data against the overall patient data and scale the transparency of the color to the score (i.e., a higher probability of differing means corresponds to



**Fig. 3** Current statistics based on user’s responses to the initial question



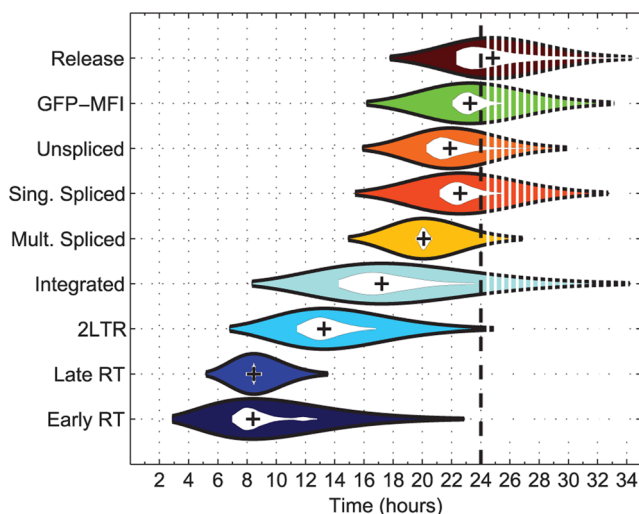
**Fig. 4** A multiprompt question. Here the highest correlated question was one of these yes/no comorbidities, but rather than ask them one at a time, to reduce decision fatigue all are presented together

more opaque colors). The patient count within a row is also represented by a bar within the cell whose width corresponds to the percentage of the overall dataset.

In the course of development, we explored other visualization designs. The relationship between the overall data distribution and the distribution of each of the rows in the outcome table is essentially a box or violin plot comparison (see Fig. 5). However, the space constrained nature – both overall and the

fixed-width space of the tabular presentation – means that kind of straightforward technique is not a suitable choice.

We attempted a series of variations bar graph-esque cell colorings, which presented the mean of the distribution by bar width, but also indicated the spread of the distribution and how it overlapped (or not) with the overall mean (see Fig. 6). We used various color mappings and gradients to show where distributions crossed over the overall confidence interval thresholds, and tried a variety of mappings of the distribution widths into fixed column widths. However, none of the attempts yielded easily-intelligible results.

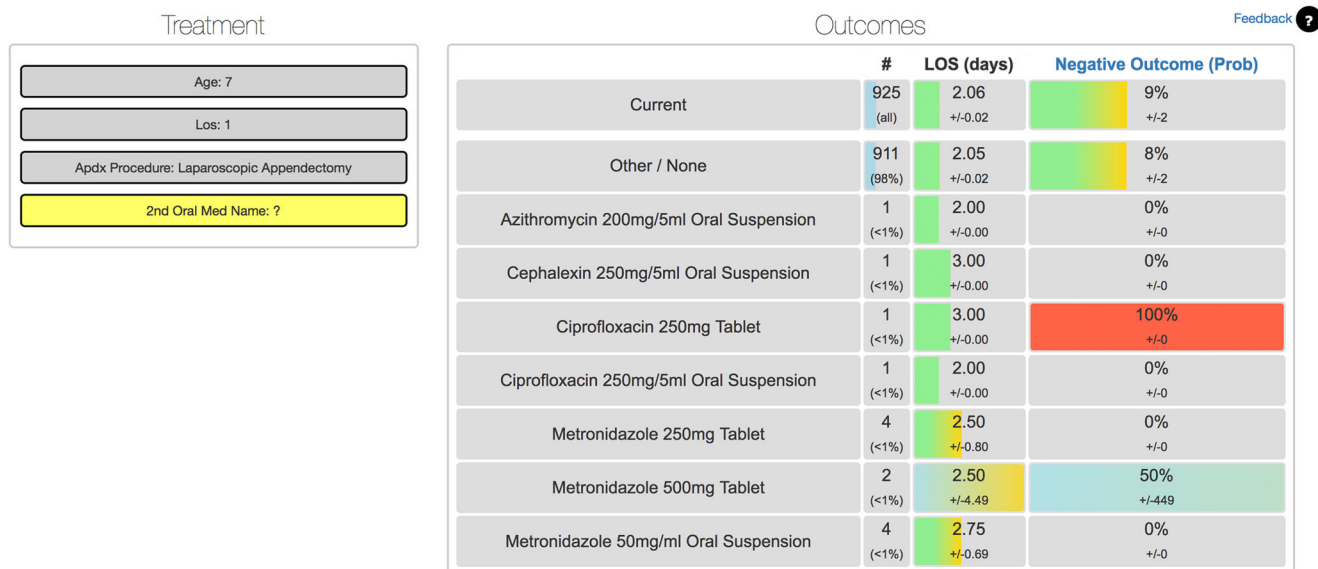


**Fig. 5** Violin plot example, reused from *PLoS Pathogens* via <https://commons.wikimedia.org/wiki/File:Violinplot-hiv-paper-plot-pathogens.png>

## Discussion

We developed a statistical model and interface to provide data-driven recommendations for the postoperative management of complicated appendicitis in children. Our models identify factors that would optimize LOS and provide clinicians and patients with information about complications and readmissions. The result of this work is a prototype smart protocol application driven by an iteratively-refined data model and under evaluation by pediatric appendectomy care providers. Efforts such as this are needed in healthcare. Tian, et al. [11] analyzed care in children’s hospitals for appendicitis and found significant variations in standards of care existed, and most likely contribute to higher costs. Raval et al. [12] identify that there is room for improvement in hospitals with high volume for





**Fig. 6** Rejected design alternative mapping box/violin distribution data onto fixed-width space

appendicitis care. There has also been a significant amount of research in using machine learning [9] and more traditional statistical techniques [10] in order to predict outcomes of various procedures. By combining these types of predictive approaches with a smart protocol we expect to both decrease negative outcomes and LOS for complex appendicitis.

Several themes are prominent within the design of computing systems for use by health care professionals. Carefully considering design choices from a user-centered perspective and iteratively developing systems are considered key methods for successfully-adopted systems [13, 14]. Developing algorithmic solutions in isolation of the social constructs of care is a common failing and reason that systems useful from a purely data or technical perspective are not adopted by care professionals [14, 15].

Visualization has been applied to many aspects of health-related problems. The CareFlow system [16] also addresses the issue of helping clinicians understand the outcomes of past patients along different care pathways, using Sankey-style flow diagrams with temporal care data mined from EMRs. Like our own, the MIVA system [17] targets a well-defined and specific care scenario (ICU care), and is a real-time biometric visual dashboard designed to reduce cognitive load, with prototype evaluations showing improvements over paper charts.

This work rests between these two in terms of the “criticality” of the work: it is more tailored and more integrated into existing care workflows than the CareFlow system, but does not address real-time data nor time-sensitive care decisions like MIVA. Furthermore, there has been a significant amount of research in the concept of fast track care [7, 8], an

evidenced-based protocol to improve care outcomes and reduce hospital LOS. Making data-driven predictions based on similar patients for the post-operative care of appendicitis available provides a mechanism to create data-driven decisions. Our work therefore offers another data point along the spectrum of health care model and software design.

Currently, the UI is being utilized by clinicians and shared with patients and families. Feedback for the users are being collected and patient satisfaction with the decisions surrounding discharge and likelihood of complications is being collected. Furthermore, outcomes such as adherence with postoperative management guidelines, LOS, and complications are being tracked. Initial anecdotal feedback has been positive from both the clinician and patient perspective as the tool facilitates communication and transparency between stakeholders.

Several items have been identified as limitations and opportunities for future refinement. Most prominently, many care variables comprise dozens of possible choices (e.g., drug/ dosage/delivery combinations), which presents a usability barrier if users are forced to choose a single item from a long scrolling list or menu. We achieved a small amount of aggregation by combining clearly identical elements (e.g., “CIPROFLOXACIN 750 MG TABLET” vs. “CIPROFLOXACIN 750 MG TAB”), but not enough to reduce many variables to a dozen or fewer choices.

The core challenge is to form meaningful categories that are statistically coherent (i.e., result in an acceptable amount of information loss from a data modeling perspective) but are also semantically coherent (i.e., an end user instantly knows

that item X belongs in category A). Work on characterizing this information loss is underway. Forming semantically coherent categories is more formidable. Algorithmic approaches (e.g., topic modeling) are unlikely to be useful; ad hoc expert judgement may be the only feasible method, though automated heuristics for similar data variables may be possible.

On the visualization side, when users are answering the question prompts, the existing visual representations of the alternatives are hidden by the prompts. Either migrating our existing data to the prompt dialog or developing other mechanisms for visual representations (e.g., Sankey diagrams to show how patients “flow” between questions, similar to CareFlow [16]).

## Conclusion and future work

In order to assist care providers in creating treatment plans for the postoperative management of pediatric appendicitis, we have developed a predictive statistical model of outcomes on which we have built a prototype decision aid application. Our ultimate goal is to move from a set of interactive questions to a full smart protocol by focusing on key decision points and tests, which can give more illumination on a patient’s best course of treatment. The visual representation of these decision points and available alternatives is critical to this goal. As we improve our access to historical data – for example, enabling the analysis of hospital discharge criteria based on body temperature by collecting daily temperature measurements – new opportunities for visualization and visual analytics may present themselves.

Through this and future work we demonstrate both the utility of such a project and suggest effective mechanisms (both technical and non-technical) for projects in similar healthcare scenarios to employ visual analytic techniques for clinicians to harness practice data and outcomes to guide clinical care.

**Acknowledgements** This work was conducted under Georgia Tech Institutional Review Board (IRB) Protocol H15340 and Emory IRB Protocol 00077519 and supported by the joint Georgia Tech/Children’s Healthcare of Atlanta Quick Wins grant program.

**Funding** This study was funded by a grant from the Emory+Children’s Healthcare of Atlanta Pediatric Research Alliance to Mehul Raval.

## Compliance with Ethical Standards

**Conflict of Interest** Edward Clarkson and Jason Zutty have received research support from the United States Centers of Disease Control and Department of Defense in addition to the Emory+Children’s Healthcare of Atlanta Pediatric Research Alliance. Mehul Raval has research support from the Emory+Children’s Healthcare of Atlanta Pediatric Research

Alliance, the Emory University Department of Surgery, and the Agency for Healthcare Research and Quality.

**Ethical Approval** This article does not contain any studies with animals performed by any of the authors. All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. The data analysis was conducted according to and approved by Emory IRB protocol #00077519 and Georgia Tech IRB H15340. For this type of study formal consent is not required.

## References

1. Belle, A., Thiagarajan, R., Soroushmehr, S. M., Navidi, F., Beard, D. A., and Najarian, K., Big data analytics in healthcare. *BioMed Research International*, 2015. <https://doi.org/10.1155/2015/370194>.
2. Beam, A.L., and Kohane, I.S., Translating artificial intelligence into clinical care. *JAMA*. 316(22):2368–2369, 2016.
3. Ponsky, T.A., Huang, Z.J., Kittle, K., Eichelberger, M.R., Gilbert, J.C., Brody, F., and Newman, K.D., Hospital-and patient-level characteristics and the risk of appendiceal rupture and negative appendectomy in children. *JAMA*. 292(16):1977–1982, 2004.
4. Rice-Townsend, S., Barnes, J.N., Hall, M., Baxter, J.L., and Rangel, S.J., Variation in practice and resource utilization associated with the diagnosis and management of appendicitis at freestanding children’s hospitals: implications for value-based comparative analysis. *Annals of surgery*. 259(6):1228–1234, 2014.
5. Short, H.L., Parakati, I., Heiss, K.F., Wulkan, M.L., Sweeney, J.F., and Raval, M.V., Challenge of balancing duration of stay and readmissions in children’s operation. *Surgery*. 162(4):950–957, 2017.
6. Miller, R. B., Response time in man-computer conversational transactions. In *Proceedings of the December 9-11, 1968, Fall Joint Computer Conference, Part I*. ACM, New York, pp 267–277, 1968. <https://doi.org/10.1145/1476589.1476628>
7. Kehlet, H., and Wilmore, D.W., Evidence-based surgical care and the evolution of fast-track surgery. *Annals of Surgery*. 248(2):189–198, 2008.
8. Vlug, M. S., Wind, J., Hollmann, M. W., Ubbink, D. T., Cense, H. A., Engel, A. F., ... and Sprangers, M. A., Laparoscopy in combination with fast track multimodal management is the best perioperative strategy in patients undergoing colonic surgery: a randomized clinical trial (LAFa-study). *Ann. Surg.* 254(6), 868–875, 2011.
9. Cruz, J.A., and Wishart, D.S., Applications of machine learning in cancer prediction and prognosis. *Cancer Informatics*. 2:59, 2006.
10. Armon, C., Radtke, R.A., Friedman, A.H., and Dawson, D.V., Predictors of outcome of epilepsy surgery: multivariate analysis with validation. *Epilepsia*. 37(9):814–821, 1996.
11. Tian, Y., Heiss, K.F., Wulkan, M.L., and Raval, M.V., Assessment of variation in care and outcomes for pediatric appendicitis at children’s and non-children’s hospitals. *Journal of Pediatric Surgery*. 50(11):1885–1892, 2015.
12. Raval, M.V., Deans, K.J., Rangel, S.J., Minneci, P.C., Nwomeh, B.C., Kelleher, K.J., and Moss, R.L., Assessing quality in pediatric surgery—the limited role of appendectomy as the optimal target. *Journal of Pediatric Surgery*. 48(11):2313–2319, 2013.
13. Horsky, J., Schiff, G.D., Johnston, D., Mercineavage, L., Bell, D., and Middleton, B., Interface design principles for usable decision support: a targeted review of best practices for clinical prescribing interventions. *Journal of Biomedical Informatics*. 45(6):1202–1216, 2012.

14. Yang, Q., Zimmerman, J., Steinfeld, A., Carey, L., and Antaki, J. F., *Investigating the Heart Pump Implant Decision Process: Opportunities for Decision Support Tools to Help*. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16). ACM, New York, NY, USA, pp 4477–4488. <https://doi.org/10.1145/2858036.2858373>
15. Yang, Q., Zimmerman, J. and Steinfeld, A., Review of Medical Decision Support Tools: Emerging Opportunity for Interaction Design. In: *Proceedings of the International Association of Societies of Design Research (IASDR)*. Congress (Interplay 2015) pp 2366–2382, 2015.
16. Perer, A., and Gotz, D., Data-driven exploration of care plans for patients. In: *CHI'13 Extended Abstracts on Human Factors in Computing Systems* (pp. 439–444). ACM, 2013.
17. Faiola A., Newlon C. Advancing critical care in the ICU: A human-centered biomedical data visualization systems. In: Robertson M. M. (eds) *Ergonomics and Health Aspects of Work with Computers* (pp. 119-128). Lecture Notes in Computer Science, vol 6779. Springer, Berlin, Heidelberg, 2011. [https://doi.org/10.1007/978-3-642-21716-6\\_13](https://doi.org/10.1007/978-3-642-21716-6_13)