Table of Contents

Note from the Organizing Committee .......................................................... 2
Note from the MIT IDSS Director ................................................................. 3
Note from the MIT TPP Director ................................................................. 3
Acknowledgments ....................................................................................... 4
About the Hackathon .................................................................................... 8
Challenge: Artificial Intelligence ................................................................. 9
Challenge: Climate ...................................................................................... 25
Challenge: Future of Work ......................................................................... 39
Challenge: Health ....................................................................................... 75
Challenge: Urban Planning ......................................................................... 129
Note from the Organizing Committee

We are the MIT Policy Hackathon organizing committee: a group of students from the Technology and Policy Program (TPP) and the Institute for Data, Systems and Society (IDSS) at the Massachusetts Institute of Technology.

We’re excited to share the proceedings from the second annual event! The hackathon shares goals with TPP and IDSS: bringing policymakers, technical experts and data together so that we can make better informed public policy choices to benefit society. With this event, we aim to create a space where participants from a variety of backgrounds can work together to propose creative policy solutions to real-world problems by combining data science techniques with domain knowledge.

Most importantly, this report contains the work of the more than 120 participants who gave up a weekend to work on critical issues in artificial intelligence, climate, future of work, health, and urban planning. These passionate students and professionals tackled unfamiliar problems that were shared by our challenge partners, who need real answers to support their work.


Our participants came up with excellent solutions, asking great questions and presenting actionable recommendations. The winners impressed the judges with creative analyses, polished presentations, and impactful recommendations. We can’t wait to see how these recommendations are used to impact the world.

-Lama Aoudi, Lawrence Baker, Adrianna Boghozian, Becca Browder, Marie-Laure Charpignon, Nathaniel Fruchter, Sika Gadzanku, Tomas Green, Donovan Guttieres, Nic Rothbacher, and Maryam Shahid
Note from the IDSS Director

The MIT Policy Hackathon has established itself in a short time not only as a fixture of interdisciplinary, student-led collaboration at IDSS, but also as a thriving model of how to address urgent societal challenges with data – a model that others can and should emulate. Enclosed are data-driven policy solutions to challenges spanning climate, urban planning, health, the future of work, and artificial intelligence. Collectively they represent the innovation and hard work of 2019 Policy Hackathon participants, now made available to governments, non-profits, fellow researchers, and the public. I have no doubt that they will prove useful to many, and that they will inspire future collaborations and problem solving.

I hope they inspire you, and that you will join us in 2020 and beyond.

-Dr. Munther Dahleh

Note from the TPP Director

The MIT Policy Hackathon has rapidly emerged as one of the most successful events at MIT’s Technology Policy Program. As a student-run event, the Hackathon reflects the spirit that inspires all of our efforts at TPP — bringing science and engineering to bear on important policy issues affecting society. Through the hackathon, participants partner with stakeholders to bring knowledge and perspectives to difficult problems, integrating data science with politics, law, and economics. By working together with challenge sponsors, hackathon teams have designed policy solutions that are useful, relevant, and action-oriented. I encourage you to engage further with us at TPP in these efforts to bridge policy and practice, and to consider participating in the Hackathon in the future, as a participant or as a challenge sponsor.

-Dr. Noelle Eckley Selin
ACKNOWLEDGMENTS

Financial Sponsors

We would like to thank the MIT Association of Student Activities’ Large Event Fund for their sponsorship, along with many groups both inside and outside MIT. This event would not have been possible without generous support from:

Deloitte
Microsoft
MIT CEEPR
Chevron
MIT Energy Initiative
IDSS
MIT Institute for Data, Systems, and Society
e4Dev
MIT Technology & Policy Program
Real-world policy problems are at the core of the hackathon, so we would like to thank our partners in crafting each of the challenges:

**ACKNOWLEDGMENTS**

**Challenge Partners**

- **ARTIFICIAL INTELLIGENCE**
  - Harvard University
  - Berkman Klein Center for Internet and Society

- **CLIMATE**
  - City of Boston Water and Sewage Commission

- **FUTURE OF WORK**
  - MIT Initiative on the Digital Economy

- **HEALTH**
  - MIT Critical Data

- **URBAN PLANNING**
  - City of Boston Innovation and Technology
ACKNOWLEDGMENTS

Judging Panel

The organizing committee would like to thank the judges for their insight and hard work, as well as for volunteering their Sunday afternoons. They had to make tough decisions between 17 excellent submissions, no easy task!

Dr. Ahmed Mahmoud Abdelfattah, Presidential Scholar and Master of Public Health, Harvard T.H. Chan School of Public Health
Dr. Frank Field, Director of Education, MIT Technology and Policy Program
Inez Frein von Weitershausen, Postdoctoral Associate, MIT Industrial Performance Center
Sharon Gillett, Chief of Staff and Technical Adviser, Microsoft Research
David Goldston, Director, MIT Washington Office
Charlie Jewell, Director of Planning, Boston Water and Sewer Commission
Adam Nagy, Project Coordinator, Harvard’s Berkman Klein Center for Internet & Society
Mycal Tucker, PhD Candidate and Graduate Researcher, MIT Computer Science and Artificial Intelligence Lab (CSAIL)

IDSS

Ed Ballo, TPP Academic Office
Erica Bates, IDSS Finance
Barb DeLaBarre, TPP Academic Office
Beth Milnes, IDSS Academic Office
John Ryan, IDSS Industry Program Manager

Mentors and On-Site Sponsors

Srini Ambati, Microsoft
Hobson Bryan, Deloitte
Mine Kansu, MIT Initiative on the Digital Economy
Charlie Jewell, Director of Planning, Boston Water and Sewer Commission
Brandon Leshchinskiy, MIT Technology and Policy Program
Carmel Levy, MBTA, Jamaica Plain Neighborhood Council
Jose Valencia, Microsoft
Jerome Vigne, Microsoft
ACKNOWLEDGMENTS

Organizing Committee

Lama Aoudi
Lawrence Baker
Adrianna Boghozian
Becca Browder
Marie-Laure Charpignon
Nathaniel Fruchter
Sika Gadzanku
Tomas Green
Donovan Guttieres
Nic Rothbacher
Maryam Shahid

Volunteers

Grace Abuhamad
Joshua Burd
Nolan Hedglin
Yuelin Li
Andrew Mowry
Kevin Shen
Max Vilgalys
Hannah Whisnant

Special Thanks

Saeyoung Rho, TPP, for the Hackathon logo design
Becca Browder, for creating this proceedings packet
Lawrence Baker, Adrianna Boghozian, Nathaniel Fruchter and Nic Rothbacher for editing the proceedings
MIT Facilities
MIT A/V Services
MIT Student Activities Office

Icons used in this publication are licensed under CC-BY, award by Amy Morgan and trophy by ProSymbols from the Noun Project
Photos: Stata Center/MIT Image Library; Computer/Luis Gomes; Water Over Bridge/George Desipris; Medic with Clipboard/rawpixel.com; Workspace/Pixabay; Boston Subway/Thomas Hawk.
ABOUT THE HACKATHON

Overview

The MIT Policy Hackathon, run by students within MIT’s Technology and Policy Program and MIT’s Institute for Data, Systems, and Society, seeks interdisciplinary solutions to societal challenges. Our most pressing modern policy challenges can’t be solved by one discipline alone. In that spirit, the hackathon aims to address some of today’s most relevant challenges while fostering an interdisciplinary spirit. We organized the event to gather participants who would work together in teams to propose creative policy solutions using a combination of robust data analytics and domain knowledge. The 2019 hackathon was the event’s second year and was attended by 120 researchers and students from MIT, 16 other universities, and a variety of companies.

Judging

The judging process is a core part of the hackathon. It aims to be fair and equitable while giving teams a chance to demonstrate the strengths of their policy proposals.

The participants’ work was judged by a panel invited from industry, academia and government, who assessed each proposal. Hackathon participants submitted two work products to the judges and organizers: a three-minute pitch deck, along with a 3-5 page policy memo detailing their proposals and analyses.

The judging panel and audience had the opportunity to hear all of the three-minute pitches during round one of judging. Afterwards, the judges made shortlists for each challenge and took time to deliberate, read, and evaluate the full policy papers. Judges then chose the finalists for each challenge. Each finalist was cold-called up to the panel and had a ten-minute Q&A with the judges. After the Q&A, the judges went through a second round of deliberation to choose the overall winner.

Challenge finalists and the overall winner are marked with a ribbon and a trophy respectively.

We provided the judges with broad guidelines based on the categories of Viability, Originality, Technicality, and Presentation Quality. Judges were asked to leverage their unique perspectives and backgrounds during the process. These guidelines provided a starting point for their extensive discussion and deliberation.
Artificial Intelligence

IN COLLABORATION WITH
Harvard University
Berkman Klein Center for Internet and Society

TEAMS

Transparency AI
"4 for AI": A Four-Fold Policy Initiative to Improve Fairness in Modern Criminal Justice Systems

Fluffy Bunny
Bringing Transparency and Auditability to Recidivism Risk-Assessment and Parole/Pre-Trial Decision-Making Process
**CHALLENGE STATEMENT**

**Artificial Intelligence**

*Translating Mathematical Constructions of Fairness to a Policymaking Audience*

**BACKGROUND**
There has been an explosion of work, particularly in computer science and statistics, discussing mathematical definitions of algorithmic fairness. The technical literature documents incompatibilities between different mathematical constructions of fairness, particularly when mapped onto real-world examples. Of particular concern to the public and their representatives is the disparate impact of risk assessment tools used to inform decisions in pretrial and sentencing on protected classes. Mathematically guaranteed trade-offs have engendered normative debates inside and outside the academy regarding the appropriate fairness criteria for risk assessments - meanwhile, the widespread use of risk assessments across jurisdictions is entrenching a particular approach to fairness, predictive parity, as the de facto national policy. Mathematical fairness is only one element of a robust debate pertaining to the use of risk assessment. Other important lines of inquiry concern the effectiveness of risk assessment tools once implemented, the consequences of optimizing for recidivism (typically defined as any new arrest) as opposed to other metrics of success such as employment, housing, and education, and the societal costs of utilizing data that is heavily influenced by structural inequities and racism. Clarity on the definition of fairness currently used by risk assessment tools and to alternative definitions popular in academic literature is a necessary, but not sufficient, component to an informed policy debate.

**THE CHALLENGE**
The live debate on algorithmic fairness - playing out in academic conferences and journals - is not readily available to individuals without quantitative backgrounds, whether they are policymakers or constituents. The literature is vast, and it can be difficult for an outside observer to distinguish between definitions that are relevant to questions of equity for protected classes, such as race, sex, gender, and disability, in the risk assessment context specifically. How can we best bridge the gap between academic and policy-making discussions on the topic of algorithmic fairness in risk assessment?
"4 for AI": A Four-Fold Policy Initiative to Improve Fairness in Modern Criminal Justice Systems

TEAM
Transparency AI

TEAM MEMBERS
Naveen Arunachalam
Srihari Jayakumar
Frank O.
Brianna White
“4 for AI”: A Four-Fold Policy Initiative to Improve Fairness in Modern Criminal Justice Systems

Brianna White, Naveen Arunachalam, Frank O., Srihari Jayakumar (Team Transparency)

Introduction

AI is being gradually incorporated into our lives and is a critical component of many proprietary systems in the public and private sectors. If current trends continue, the rapid proliferation of AI into new domains can hinder our ability to thoroughly interpret and understand its emergent behavior. Compounding this issue is that most AI algorithms are both built and validated by engineers and other members of the technology community who optimize against performance metrics, meaning that considerations of fairness, justice, and equality are often treated as implicit rather than explicit goals.

One area in which AI is gaining influence is criminal justice, where privately-owned software tools such as COMPAS can have an adverse effect on equality by promoting a narrow interpretation of fairness. For example, the selection of training data and choices made regarding algorithmic implementations in tools such as COMPAS can have a significant impact on the outcomes of trials and parole decisions. If these tools are used to inform decision-making without the aid of additional methods of identifying biases and deficiencies in a human-interpretable way, then they can introduce hidden injustices in a way that is obscure and difficult to combat. Thus, community outcomes could be greatly improved through creation of a transparent, easy-to-use system.

Policy Framework

To identify policy-based solutions to the issue of justice in AI, we take a systems approach to pinpoint areas in which the public sector can have the greatest possible impact on improving community outcomes. The creation of a predictive algorithm usually comprises the following phases (Lipeles 2018):

- Data set acquisition
- Exploratory data analysis
- Model development and training
- Deployment, maintenance, and production
- Use new data to infer predictions

The relationship between these phases is illustrated in Figure 1 below.

When a problem appears or a complaint arises in an AI system, developers will revisit their model, data, and code to fix the issue. This approach is customary in traditional areas of software engineering, and now is being co-opted by the AI field. However, this system of deployment and production comes with a unique set of issues in some areas involving AI; for example, since some AIs (such as those in criminal justice) have the potential to produce wrong decisions and negatively impact others, they can have an adverse impact on real-world outcomes if their underlying errors and biases remain undetected.

Although technological solutions have been attempted to gauge the validity and fairness of AI models, these solutions are constrained to deal with proprietary and complex AIs as black boxes. Naturally, these analyses are limited in their ability to improve implicit performance factors (such as fairness) of critical systems. Because of this limitation, we believe that policy-based solutions are uniquely suited to align AI
performance to stakeholder needs. Looking into the AI development pipeline holistically, we can try to study where exactly policy changes could be implemented in order to improve the outcome. We see a few clear areas where we can introduce specific policy objectives corresponding to stakeholder needs:

- **Data set acquisition.** Developers must choose to train AI models on data that appropriately represent the target property. Although it is difficult to conjure truly unbiased indicators for any given task, a policy-based approach can aid developers by encouraging data producers to curate relevant data, while helping data producers by encouraging developers to use relevant data.

- **Exploratory data analysis.** Because not all data is unbiased, key stakeholders of AI systems should be able to understand and interpret training data to make informed decisions about whether to apply an AI to a particular use case. A policy-based approach that commissions a clear way for all parties to communicate and study a dataset could use approaches such as the data nutrition tool [4] or the data interpretation tool that our team has implemented (see Policy Proposal section).

- **Model development and training.** Training engineers in ethics. Occasionally, the requirements of developers and stakeholders can be at odds during the development of AI models, e.g. when developers optimize for algorithmic measures of performance while stakeholders value other measures of performance. A policy initiative that fosters meaningful exchanges between developers and stakeholders can help bridge the gap between the two groups and result in better outcomes for all parties involved.

- **Use new data to infer predictions.** Currently, individuals from non-technical backgrounds are not always sufficiently empowered to test the inputs and outputs of a commercial/black-box AI model. A policy initiative that commissions approaches such as TuringBox could allow all stakeholders to act as “AI examiners” regardless of technical background by allowing them to provide their own test sets to AI models to better understand their behavior and inherent biases.

![Diagram of the AI development cycle](https://towardsdatascience.com/ai-ml-practicalities-the-cycle-of-experimentation-fd46fcf3835)

**Figure 1.** The AI development cycle, which breaks down the AI development process into components and subsystems. Reproduced from Lipeles, 2018. ([https://towardsdatascience.com/ai-ml-practicalities-the-cycle-of-experimentation-fd46fcf3835](https://towardsdatascience.com/ai-ml-practicalities-the-cycle-of-experimentation-fd46fcf3835))

**Policy Proposal**

The key stakeholders in AI systems for recidivism prediction are public officials, academics, data engineers, and the general public. In this application area, predictive software (and, by extension, the development cycle for such software) is often treated as (or is) a black box, leaving only the last step of the cycle exposed to public inquiry [7] [11]. Because of this, key stakeholders in criminal risk assessment are often unable to critically analyze the underpinnings of the AI models they use, and public-interest
studies are constrained to assess the fairness of risk assessment algorithms based only on input/output behavior.

With this consideration in mind, we utilized the systems framework outlined in the section above to identify three main objectives that a successful policy must achieve:

- **Goal 1:** The policy must improve the well-being of the general public by resulting in AI systems that more accurately and fairly predict recidivism risk.
- **Goal 2:** The policy must increase the transparency and interpretability of recidivism prediction AIs for all stakeholders by providing a means for individuals to analyze dataset features and input-output behavior in a way that is accessible regardless of one's technical background.
- **Goal 3:** The policy must aid public officials and data engineers simultaneously by encouraging exchange of ideas regarding model design and design objectives.

We propose a policy consisting of four action items that addresses the intersection between Artificial Intelligence and public policy agenda. These four items allow key stakeholders (public officials, academics, data engineers, and the general public) to interact with and improve current AI systems. The action items for a 2-year pilot are listed below:

**Action Item 1:** By Month 6 of pilot, publicly release an interactive platform that allows users to test the input/output behavior of all algorithms employed for recidivism risk prediction. Such a platform must allow users to choose their own data to provide to the algorithms, and a corresponding test data set must be provided that is equal in size to the largest training test set. Additionally, the platform must provide a channel of communication through which individuals can indicate how usable they find the platform and alert officials to situations in which the algorithms make unfair decisions.

**Action Item 2:** By Month 6 of pilot, provide an easy-to-use platform to enable the public to visualize and interpret the datasets without any background in field itself or programming. In the case of Massachusetts, hold five separate community meetings within Year 1 of pilot (Harvard Square, Dudley Square, Framingham, Quincy Center, and South Boston) that demonstrate and describe the data to the general public, with a forum for public comments held afterward.

**Action Item 3:** Implement a periodic (i.e. monthly) idea-exchange program during Year 1 and Year 2 of pilot for all developers (public or private) involved in the writing of recidivism-prediction software that brings them together with stakeholders to debug and reduce problems during AI model development. In the case of Massachusetts, the proposed stakeholder cohort includes social service leaders, academic researchers, city officials (e.g. public safety officers), and individuals recently released by MDOC.

**Action Item 4:** Assemble a task force during Year 1 of pilot to investigate more representative factors in AI which are used to predict recidivism. The task force should provide criteria for evaluating the relevance of different factors to recidivism and provide numerical values for these factors in all future training datasets provided to AI developers. An overview illustrating the nature of the issue is provided in the next section.

To gauge progress and public reception of this policy, three measures for the 4 For AI pilot are proposed:

- **Risk prediction accuracy:** By the end of Year 2, improve 1-year recidivism predictability (as measured by a statistic such as binary cross-entropy) beyond a 95% margin of error relative to the average of three years prior to implementation.
- **Usability assessments and feedback rates:** As a result of action items 1 and 2, feedback from the public should be obtained that indicate consistent and favorable usability ratings across all users.
Idea exchange attrition rates: If implemented correctly, stakeholders who experience productive idea exchanges and view the process as worthwhile will continue to show up for idea exchanges. Sustained levels of participation will indicate success in action item 3.

![Image](image.png)

Figure 2. An example data inspection client developed by our team. Such a client can help governments fulfill action items 2 and 4. Our client is built with Metabase and PostgreSQL and is open-sourced (under GNU Affero license).

An Example in Criminal Justice: The Problem with Fairness

Each year, 600,000 people are released from state prisons. Within 3 years, 67% of them have been re-arrested. Five years post-release, 76% have been re-arrested and imprisoned[12]. To improve public safety and lessen the probability of individuals re-entering the criminal justice system, parole boards, prosecutors, and judges across states have employed risk assessment calculators. Risk assessment calculators compute whether an individual that is released has a low, medium, or high probability to be re-incarcerated. With this information, public safety officials decide the length of an individual’s incarceration, and whether or not they should be released. In the discussion of risk assessment, however, there are key elements that have been misused. How public officials (prosecutors, judges, and parole boards) make decisions about pre-trial and sentencing is based around data that has both inherent bias, and disparate impact of the model. However, there are issues with this model. As former Attorney General Eric Holder stated, using the current predictive model: “...may exacerbate unwarranted and unjust disparities that are already far too common in our criminal justice system and in our society” [5].

In the discussion of fairness and criminal justice, there are two lines of logic. The first refers to whether the design of a policy or model is inherently just. Are all individuals within society equally protected in the model’s design? The second line focuses on whether certain groups are disproportionately impacted by the policy.

The current risk assessment models used in criminal justice systems have unintended consequences, with African-Americans being deemed as higher risk in comparison to their white counterparts. We no longer wish to punish those based on their history, identity, and status. Instead our predictive model focuses on individual and contextual factors.
The theoretical relationship between individuals and their risk for recidivism requires both individual and contextual considerations. Gender, educational attainment, and age are characteristics of individuals, while exposure to the mental and substance abuse services, family linkages, and neighborhood quality are the key contextual factors that matter. After conducting interviews and focus groups with former parole board members, social service organizations, and the formerly incarcerated, our team has information on how successful re-entry can occur. Those who are unlikely to be re-arrested have strong ties to the services in their community (substance abuse, mental health programs), have strong relationships with family members, and have achieved educational and vocational training while incarcerated.

References:


https://www.perpetuallineup.org/recommendations


https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm

POLICY BRIEF
Glass Box: Enforcing Algorithmic Transparency Within the U.S. Federal Criminal Justice System

TEAM
Fluffy Bunny

TEAM MEMBERS
Dan Ebner
Cleo Forman
Stephany Palaguachi
Matt Poreda
Shayan Ray
Glass-box
Enforcing Algorithmic Transparency Within the
U.S. Federal Criminal Justice System

Team Fluffy Bunny: Dan Ebner, Cleo Forman, Stephany Palaguachi, Matt Poreda, Shayan Ray

Background

Pre-trial risk assessment tools are increasingly ubiquitous throughout the United States. Traditionally, bail has acted as an insurance policy for the state: a defendant must post bail, typically a monetary investment, that will be refunded upon their return to court for trial. Since the passage of the Bail Reform Act of 1984, defendants determined to be a high flight-risk or a significant and immediate threat to the community may be detained until their trial.

While mathematical risk assessment tools are not a new phenomenon [1], modern machine learning risk assessment algorithms have provoked a vigorous debate regarding the role and repercussions of technology within the criminal justice system. Some proponents of this technology have emphasized its potential to minimize the subjectivity of judges, who have been documented to be biased against black defendants in bail recommendations, by making relatively objective, data-driven guidance [2]. Arguably, these assessment tools could also decrease the number of detained pretrial defendants [4, 13]. In 2009, 62% of the inmates in local prisons had not been convicted of their current charge or were awaiting court action [3]. The number of incarcerated pretrial defendants in Alaska’s court system grew by 81% between 2004 and 2014, placing an increasingly heavy pressure on local resources [4].

However, pretrial risk assessment tools have raised a variety of concerns. The Due Process Clause of the Fifth and Fourteenth Amendment state that no one shall “be deprived of life, liberty or property without due process of law” [5,6]. Recent studies show risk assessment tools are prone to racial bias, violating due process. In 2016, Larson et al. found that Northpointe’s COMPAS algorithm, a common proprietary risk assessment tool, was biased against African-Americans. Black defendants who did not recidivate over a two-year period were twice as likely to be classified by the algorithm to be high risk compared to their white counterparts. Black defendants were also twice as likely to be classified as high risk for violent recidivism. White defendants were 63% more likely to be misclassified as low risk than black defendants. The overall utility of the algorithm was also called into question: only 61% of defendants’ risk scores accurately predicted whether they would recidivate, and the scores only correctly predicted violent recidivism 20% of the time [7]. Given the literature demonstrating inefficacy and bias, risk assessment tools have been criticized as unfair.
However, achieving fair algorithms is difficult as the definition of “fairness” with respect to emerging risk assessment tools is contentious [8]. The machine learning community has recently formulated a variety of mathematical definitions of fairness, a selection of which are particularly relevant to criminal justice risk assessment applications, namely calibration, predictive parity, and error rate balance. However, these mathematical constructions of fairness necessarily cannot align with one another. In a society where crime is differentially detected and enforced between groups, predictive parity and error rate balance are inherently incompatible measures of fairness [8].

In addition to mathematical challenges surrounding algorithmic equity, there is the question of legal recourse when unintelligible and, often, proprietary algorithms are determining the fate of a defendant. U.S. due process relies on the assumption that judgements are transparent and contestable; however, when a judge adheres to the assessment of a proprietary and, potentially, biased algorithm, legal recourse is nearly impossible [9, 10, 11]. While there are various policy routes for amending the due process issues raised by involvement of pre-trial risk assessment algorithms, we work within existing federal legislation to suggest a modification of current judicial practice in a way that will allow for improved legal recourse and increased algorithmic transparency.

**Methodology**

Given the extent to which the publicly-available COMPAS data has been analyzed, we felt that it would be a more effective use of our resources to begin drafting a tool that might mitigate some of the critiques of the system that have been made. We decided to build an end-to-end framework that inputs a dataset, composed of a variety of defendant characteristics. Next, we construct a model using the data. We trained on a variety of different models, including k-NN, random forest, extra trees classifier, XG Boost, and decision trees. We determined that the decision tree model led to the highest accuracy model so a decision tree multi-label classifier was trained, hyper-parameter tuned, and run on the COMPAS training data.
In our model the top-K (‘k’ being the user input) features is determined using Lasso regression on the dataset. Based on the top-K features, the accuracy of the decile score prediction is generated with our machine learning model. Finally, our front-end platform allows the user to swap the top-K variables to/from the entire feature list and determine how the accuracy of the decile score prediction is impacted.

**Component overview:**

Frontend: HTML5, CSS3, JavaScript and Jquery

Backend: python flask

Machine learning API: scikit-learn used for pre-processing, training, hyper-parameter tuning and prediction

![Parameter Analysis](image)

Figure 2. Our tool’s dashboard showing the most influential factors contributing to the accuracy score. Two factors have been removed in this demonstration, leading to the accuracy of 83.57%.

**Our Proposal**

*Aligning with policy*

The Bail Reform Act of 1984 states that both parties may seek appellate review of a release order and, in the case that the defendant is detained, that a magistrate can provide “clear and convincing” evidence that no condition of release would have reasonably assured the safety of
the community [12]. Thus, we propose that a magistrate must explicitly and thoroughly document the influence and role of the pre-trial risk-assessment tool in their decision-making process. Given existing tools, this might be a difficult standard to reach, so we are introducing a new software designed specifically for this purpose.

**Glass-box: transparency and accountability**

The tool we have developed, Glass-box, intakes a dataset made up of a set of questions answered by defendants or pulled from criminal records (similar to the input of comparable proprietary algorithms). Our tool will serve the dual purpose of making the judge more aware of the factors influencing the algorithm’s risk score and easy documentation of agreement with, or deviance from, the existing algorithm. The judge would be required to submit to the judicial record the output of the algorithm. They would also need to explain how they may have used the software to manipulate the different contributing variables. For example, if the algorithm is highlighting a defendants’ past drug use, but the defendant testified convincingly during trial that they were no longer a user or had been to rehab, then the judge might elect to remove that variable from the algorithm’s consideration and explain this decision within their case opinion.

**Analyzing bias**

We do not propose that our algorithm will magically solve the problem of bias. Thus, we need to provide the judge and the public with means of understanding how fair the algorithm is. We propose that the algorithm be continually tested for the three definitions of fairness: calibration, predictive parity, and error rate balance. In a future iteration, this information can be included in our software’s dashboard, so that the judge will be hyper-aware of the potential disparate impact of the risk assessment.

**Score presentation**

We also recommend that scores be presented differently. Currently, most risk assessment algorithms present a “low”, “medium”, or “high” risk score. Statistical methods would suggest that distinctions between scores lying within the “medium” category (from 3-7) are likely somewhat arbitrarily distributed. The difference between a score of 4 and 6 should not determine the result of a bail trial. Instead of describing these scores as “medium”, we present them as “average”. This suggests that the magistrate should rely on his own judgement of the case to discern the specific pre-trial risk. However, if the defendant scores between 1-3 or 7-10, the algorithm supposedly has more confidence in the assessment. In this case, the judge could view the specific scores and be able to interact with the Glass-box dashboard to understand the contributing factors for this assessment. This does not necessarily suggest that the high and low scores are to be trusted blindly (that would need to be evaluated depending on the accuracy and fairness of the algorithm), but algorithmic transparency and documentation is especially important if the algorithm seems to be advocating detention.

**Societal recourse**

Aside from immediate legal recourse, this documentation process will allow increased societal understanding of risk assessment algorithms. As it stands, algorithms are typically deployed
within government without much public awareness. The documentation created by interaction with Glass-box will exist in the public record, which will make it significantly easier for watchdog groups to conduct investigations of the algorithm’s societal impact. An extension of the Glass-box tool could incorporate tracking data that would allow analysis of which factors judge’s most commonly removed or changed the relative weight of. This data could provide valuable, general insight into judicial decision-making and judicial/algorhitmic collaboration.

Measuring Impact

Of course, we are merely hypothesizing the positive effects of our proposed model of pre-trial risk assessment. At its best, this algorithm would increase both transparency and efficiency of pre-trial risk assessment. However, we would seek to implement this algorithm incrementally within the federal court system, explicitly tracking the positive and negative effects. We would want to measure how the software’s implementation affected the number of detained defendants, compared to both pre-trial hearings that don’t use a risk assessment tool and those that employ a current proprietary implementation of risk assessment algorithms. We would also want to measure whether it affected the number of appealed cases, the number of correctly predicted risk scores (including false positives, false negatives, true positive and true negatives), and the software usability. Depending upon the results of this initial implementation, we could narrow down best practices regarding civic algorithm use. If our software’s deployment was successful, it could be iterated upon and deployed more widely. Additionally, its success would provide guidance for local and state-level risk assessment policies.

Conclusion

Pre-trial risk assessment algorithms are deployed widely across the U.S., and, if used conservatively and transparently, might decrease judicial bias. However, as currently implemented, these tools are too unregulated to ensure that a defendant’s right to due process is protected. Thus, we introduce a selection of new policy ideas via the software we developed: Glass-box. The primary contributions of Glass-box are:

- Increased transparency of the weight of the variables that lead to the risk assessment score
- Documentation of the judge’s consideration and interaction with the risk assessment
- Increased legal recourse of algorithmic determination due to documentation
- Contrasting fairness assessments
- Augmented auditability for use by watchdog groups
Works Cited


Climate

IN COLLABORATION WITH
City of Boston Water and Sewage Commission

TEAMS

GAMMDRYL
Supplement Hyper-Local Data to Improve Flood Resiliency
CHALLENGE STATEMENT

Climate

Using Precipitation Data to Inform City Resilience Plan in the Face of Climate Change

BACKGROUND

For the climate challenge, we are partnering with the City of Boston to help the city develop its climate action plan. In December of 2016, Mayor Marty Walsh unveiled Climate Ready Boston, a set of plans and actions to increase the city’s climate resilience. The plan consists of 11 strategies:

1. Maintain up-to-date projections of future climate conditions to inform adaptation.
2. Expand education and engagement of Bostonians on climate hazards and action.
3. Leverage climate adaptation as a tool for economic development.
4. Develop local climate resilience plans to coordinate adaptation efforts.
5. Create a coastal protection system to address flood risk.
6. Coordinate investments to adapt infrastructure to future climate conditions.
7. Develop district-level energy solutions to increase decentralization and redundancy.
8. Expand the use of green infrastructure and other natural systems to manage stormwater, mitigate heat, and provide additional benefits.
9. Update zoning and building regulations to support climate readiness.
10. Retrofit existing buildings against climate hazard.
11. Insure buildings against flood damage.

There are numerous projects that are either existing, under construction, or planned for the short or long term. The projects relate to coastal flooding risks and protecting key infrastructure. Currently the plan does not address how these projects relate to the potential for heavier rain and more extreme weather.

THE CHALLENGE

Use 20 years of site-specific precipitation data to evaluate the impact of climate change on the City of Boston and how that should inform the climate resilience plan and urban planning.

Questions

- What trends in precipitation do we see over time?
- How can these patterns inform the flooding preparedness plans for the city?
- How do threats of extreme precipitation compound the risk of coastal flooding?
- How will extreme precipitation affect the city’s climate resilience projects?
POLICY BRIEF
Supplement Hyper-Local Data to Improve Flood Resiliency

TEAM
GAMMDRYL

TEAM MEMBERS
Marco Montalto Monella
Minghao Qiu
Lauren Yee
Arthur Yip
Climate Challenge Policy Brief - Team GAMMDRYL

Executive Summary
The Boston Water & Sewer Commission (BWSC) requested us to identify the flooding risk and policy implications on Boston’s climate resiliency and urban planning, using 20 years of data from BWSC’s rain gauges. We find that (a) there were a limited number of trends that could be identified over the time period from BWSC’s hyper-local rain measurements, (b) both the timing and magnitude of rainfall differed significantly between the different measurement locations during the extreme precipitation events, and (c) the existing precipitation data is not enough for a comprehensive evaluation of the flooding risks in Boston. We recommend that BWSC continues to support and collaborate with existing partners such as Climate Ready Boston and provide additional resources for mitigating multiple climate effects. Additionally, BWSC should join the Community Collaborative Rain, Hail and Snow (CoCoRaHS) data network, a citizen science initiative that crowdsources rainfall data. In addition, BWSC should take the lead in encouraging Bostonians in collecting their own rainfall data in their own backyard and also submitting it to CoCoRaHS.

Goals and Objectives:
(1) Examine trends of total, intensity, and frequency of precipitation events in Boston.
(2) Investigate local differences in precipitation events at different gauge locations.
(3) Understand local patterns of flooding risk and vulnerability related to historical and future extreme precipitation events.
(4) Assess the value of BWSC’s hyper-local rain gauges and data.

Data Analysis and Findings:
The BWSC’s rain gauge data contained precipitation data from 1999 to 2019 for ten stations [1]. First, we examined data quality, and temporal and spatial coverage. Figure 1 shows the data coverage for each site from 1999 to 2019. Due to missing and faulty data issues, including mechanical failure, human interaction, and delayed software patches, we focused on the six sites on the period which we have the most data available, from 2002 to 2014.

Through this preliminary analysis we found that although local variation was present in the rain gauge data it was not entirely comprehensive or representative when compared to noted extreme weather events in Boston. This data was supplemented with Boston weather station data from the NOAA [2] and information from news sources regarding extreme events, BWSC documents, and Climate Ready Boston (CRB) reports.

Figure 1: Data coverage for all sites

Figure 2 shows the monthly total precipitation for the six sites. There is no clear trend from the total precipitation for all six sites and there is a high degree of correlation. We also found few differences in the annual total precipitation across different sites (<10%), likely because all six sites were
influenced by similar weather systems in general. Previous BWSC analysis using long-term NOAA data suggested a slight upward trend in total annual precipitation. We find that the BWSC local rain gauge data is insufficient to confirm or deny that finding. In addition, total annual or monthly precipitation data analysis may be convenient, but is of limited value due to enormous variability in precipitation and its impacts at much finer time scales.

![Trends From 2002 to 2014](image)

*Figure 2: Monthly total precipitation at six BWSC sites, in inches per month*

We focused on capturing the extreme precipitation events for each year and determined their annual frequency, based on a fixed intensity criterion. We obtained baseline BWSC criteria for the intensity of extreme events e.g. the rainfall over 24 hours for a “10-year event,” which is what the sewer system is designed and sized to handle [3]. These were derived from historical data (1948-2001). Analysis of extreme rainfall event occurrences is presented in Figure 3.

Similar to the time series of total precipitation, we find that the BWSC gauges do not have enough historical data to produce an appropriate trend. Many of the data points suffer from high leverage, meaning they disproportionately pull up/down the linear trendline. It is tempting to conclude increased frequency of extreme events based on the trendlines, but we find the trends weak and inconclusive. For robustness, we tested several other criteria and time intervals, including 6 hr and 1 hr and their respective criteria, for all locations (plots in appendix). Notably, the trend lines are not consistently upward.

Although we did not find significant trends in localized risks of extreme events, we found large spatial and temporal heterogeneity of precipitation across different sites during the extreme precipitation events. On May 14, 2010, the most extreme rainfall event of the past 20 years to hit Boston, Gov. Patrick declared a state of emergency due to flooding [4]. Many roads and basements flooded, schools and subways closed, and eight million gallons of raw sewage were released into the Bay due to over capacity facilities. Figure 4
shows the hourly rainfall during this event.

*Figure 3: Annual frequency of extreme rainfall events based on BWSC criteria, for NOAA-BOSTON weather station and 6 BWSC locations. Linear fits are in grey and intentionally de-emphasized due to their poor fit.*

*Figure 4: Hourly precipitations at six BWSC sites and NOAA site during the 2010 extreme rainfall event.*

We find that both the timing and magnitude of rainfall differed between the different measurement
locations. Particularly, the official NOAA-Boston was significantly different than the measurements at different sites. This could mean that certain areas were subjected to more flooding earlier than others. Figure 5 shows the spatial heterogeneity in the total 24-hr precipitations of different sites of the extreme event of March 14, 2010. The second event, July 6, 2005, is shown in the appendix. The regional difference is large, which has implications for storm management at the local level. It’s also important to note that the regional patterns and correlations are different for the two events as well, which suggest more analysis is required to conclude any consistent regional pattern out of this.

Spatial heterogeneity in precipitation pattern at the extreme precipitation events suggest that different demographic groups could be exposed to different level and timing of flooding risks purely due to the spatial heterogeneity of the precipitation. We observe that some low-income regions (in red) were exposed to extreme precipitations disproportionately during some of the 10 most extreme events that we examined. Considering the lack of private investment in preventive measures, our finding might suggest a higher flooding risk for the lower income group living in the city of Boston.

Figure 5: Spatial heterogeneity of precipitation during extreme precipitation event and its interaction with regional demographic information. The number on the plot is the 24-hr total precipitation (unit: inches). The colormap shows the median household income of the region in 2018.

We also examined how the potential flooding risk from extreme precipitation events compound the risk of coastal flooding. We analyze the most extreme once-in-ten-years event in our study period: the event happened on March 14, 2010. The peak hours for both events happened to be overlapped with the high tide hours. More analysis need to be done to ensure how would the coastal risks interact with the flooding risks due to extreme precipitation.

We find that the existing precipitation data is not enough for a comprehensive evaluation of the flooding risks in Boston. When we compare the rain gauge data with previous analyses on flooding risk such as in the climate resilience plans, they do not overlap with each other. We believe this is because flooding risk is a function of both the precipitation, the capacity of the storm-water system, and the topography. Other
local non-rain factors would need to be considered for policymaking on storm management.

Stormwater flooding in Boston is a serious concern. The Climate Ready Boston vulnerability assessment estimates that 7 percent of the total land area of Boston will likely be exposed to the 10-year, 24-hour event as soon as the 2050s and 9 percent by the end of the century [5]. Further, this report identified neighbourhoods (shown in red in figure 6) that have the largest areas of land affected by stormwater flooding or sea level rise and extreme precipitation events. We note that some of these areas do have local rain gauges (Allston, Logan Airport), but others do not.

We further note that the South Boston waterfront is a low-lying area that faces the greatest exposure and potential losses to coastal flooding across all sea level rise conditions and flood events, however, this area has less socially vulnerable groups than other neighbourhoods. [6]

Similar to our findings, previous analysis also identified the storming flood risk on the low income neighborhoods. Climate Ready Boston neighborhood assessment found that East Boston could be subject to a five foot rise of water with sea-level rise. This neighbourhood has a third of people who are low income. Additionally, this neighbourhood is a vital part of the economy for the city of Boston and generates $4.6 in economic value and 50,000 jobs. [7]

**Figure 6: climate vulnerability regions identified in Climate Ready Boston vulnerability assessment report.**

**Policy recommendations**

Climate Ready Boston (CRB) aims to increase the resiliency of the city of Boston for future climate change impacts through projects and mitigation procedures that benefit multiple sectors at once. The Climate Projection Consensus by CRB states that: an increase in extreme precipitation events is expected to continue, however, due to the complexity of the processes underlying precipitation and natural variability, the magnitude of this increase is not clear. Through our analysis with the rain gauge data, these relationships were also not clear, therefore, specific recommendations were difficult to determine. The policy recommendations provided by CRB would benefit stormwater flooding in conjunction with providing multiple benefits for other climate related events and effects. The recommendations include: increasing tree cover to slow infiltration rates, increasing community resilience, requiring developers to retain stormwater on site.

An increasing amount of literature utilizes the public good and citizens to collect scientific data in their own backyards. We recommend that BWSC join the NOAA, NWS, and NSF-supported Community Collaborative Rain, Hail and Snow Network (CoCoRaHS) [8], by submitting their local rain gauge data to these organizations BWSC could take the lead in encouraging citizens from all over Boston to report rain measurements to CoCoRaHs, at schools and homes. This additional data could supplement the spatial gaps in coverage for rain gauge data in at risk neighbourhoods (Figure 6). Collecting data in this way mobilizes citizens to care about increasing extreme precipitation events and adaptation strategies in Boston. The main objective would not necessarily be high-quality data, but it would be to spread awareness about local differences in meteorological data, data quality, and the data-generating process,
and to inspire thought and discussion about weather and climate issues. Climate change is a long-term issue and we suggest that an indirect but powerful mechanism for impact would be for BWSC to use its local rain gauge program to support education and citizenship. Potential partners already involved with CoCoRaHS include the local National Weather Service Offices in Albany and Boston/Norton, the Massachusetts Department of Conservation and Recreation (DCR), and the American Meteorological Society (AMS) Headquarters in Boston.

Data recommendations
For long-term trend information and analysis, we recommend BWSC supplement missing data with alternative data sets, such as NOAA weather stations, additional gauges, radar/remote sensing or pre-processed precipitation data sets (Daymet), 311 calls for flooded properties or sewage back-up to further validate and examine vulnerability. However, there is value in local rain gauges and a wide variety of data users such as: weather professionals and scientists, water and emergency managers, gardeners and urban farmers, parks and recreation managers, public health professionals (mosquito control), insurance adjusters, engineers, and educators. [9]

References:


[8] https://twitter.com/NWSBoston/status/1103096080751824899

Human Health: Sewer backups due to stormwater flooding can contribute to pathogen spread through contact with sewer waste which contains bacteria, viruses, mold and other mobilize pollutants from the environment. Floodwaters (not including sewage backups) are also known to mobilize pollutants in the environment. Often these impacts are mitigated by citizens installing back-flow prevention valves to halt the flow of sewage back-up into their homes.
R Code:
library(readr); library(tidyverse); library(magrittr); library(lubridate); library(cowplot); library(readxl)
source("mit_hackathon_fns.R")

# read and initial process of BWSC rain gauge data ####
n rainfall_all_sites <- read_csv("mit policy hackathon rain/rainfall_all_sites_raw.csv",
  col_types = cols(date_final = col_character(),
  charleston = col_double(),
  dorchester-adams = col_double(),
  dorchester-talbot = col_double(),
  east-boston = col_double(), hyde_park = col_double(),
  longwood = col_double(), roxbury = col_double())) %>%
machine= date = ymd_h(paste(year, month, day, hour, sep=" ")) %>%
gather(key = "location", value = "rain", 6:15)
#filter(!is.na(rain)) # do not filter out NAs yet

# inspect rainfall data ####
rainfall_all_sites %>%
  filter(!is.na(rain)) %>%
group_by(year, location) %>%
tally() %>%
spread(key = "location", value = "n") %>%
machine= at(-1, divide_by, 8760) %>%
machine= at(-1, round, 2) %>%
View()

# read noaa data ####
boston_rain <- read_excel("mit policy hackathon rain/boston_rain.xlsx") %>%
machine= (date = ymd_h(paste(year, month, day, hour, sep=" ")) ) %>%
rename(rain = inches,
  location = STATION_NAME)

# inspect boston weather station data - only hours with rain were included.
# BOSTON MA US looks complete 1948 - 2013
boston_rain %>%
  filter(!is.na(rain)) %>%
group_by(year, STATION) %>%
tally() %>%
spread(key = "STATION", value = "n") %>%
machine= at(-1, divide_by, 8760) %>%
machine= at(-1, round, 2) %>%
View()

# simple plot ####
ggplot(rainfall_all_sites) +
gem_point(aes(x=date, y=rain)) + #, col = location)
   facet_wrap(vars(location))
```r
# preparing data for analysis and visualization (merge/join, complete) ####
gaugedata <- rainfall_all_sites %>%
  filter(location %in% c("allston", "dorchester-adams", "dorchester-talbot", "hyde_park", "roslindale",
    "union_park") &
    year > 2001 &
    year < 2015)

noaadata <- boston_rain %>%
  # filter(location %in% c("BOSTON MA US") &
  #       year > 1948 &
  #       year < 2014) %>%
  complete(date = seq(min(date), max(date), by = "hour"), fill = list(rain = 0, location = "BOSTON MA US")) %>%
  mutate(year = year(date),
    month = month(date),
    day = day(date),
    hour = hour(date),
    rain = rain * 100) %>%
  mutate(rain = case_when(rain > 100 ~ rain/1000,
                          TRUE ~ rain))

full_mit_rain_data <- nooadata %>%
  bind_rows(rainfall_all_sites)

# sum up per period and count extreme events with summed_rain over threshold ####

# run first with BWSC data
data1 <- extreme_count_setup(gaugedata, 24, 2.6)
data2 <- extreme_count_setup(gaugedata, 24, 5.2)
data3 <- extreme_count_setup(gaugedata, 6, 1.6)
data4 <- extreme_count_setup(gaugedata, 6, 3.4)
data5 <- extreme_count_setup(gaugedata, 1, 0.8)
data6 <- extreme_count_setup(gaugedata, 1, 1.8)

# // run separately for nooadata
data1 <- extreme_count_setup(nooadata, 24, 2.6)
data2 <- extreme_count_setup(nooadata, 24, 5.2)
data3 <- extreme_count_setup(nooadata, 6, 1.6)
data4 <- extreme_count_setup(nooadata, 6, 3.4)
data5 <- extreme_count_setup(nooadata, 1, 0.8)
data6 <- extreme_count_setup(nooadata, 1, 1.8)

# //
write.csv(x = data1, file = "data1.csv")
write.csv(x = data2, file = "data2.csv")
```
write.csv(x = data3, file = "data3.csv")
write.csv(x = data4, file = "data4.csv")
write.csv(x = data5, file = "data5.csv")
write.csv(x = data6, file = "data6.csv")

plot1 <- data1 %>%
  Extreme_events_trend_plot() + ggtitle("number of events with 2.6+ inches of rain / 24 hrs, in a 1-year rainfall event based on 1948-2011")
ggsave(filename = "plot1.png", plot1)

plot2 <- data2 %>%
  Extreme_events_trend_plot() + ggtitle("number of events with 5.2+ inches of rain / 24 hrs, in a 10-year rainfall event based on 1948-2011")
ggsave(filename = "plot2.png", plot2)

plot3 <- data3 %>%
  Extreme_events_trend_plot() + ggtitle("number of events with 1.6+ inches of rain / 6 hrs, in a 1-year rainfall event based on 1948-2011")
ggsave(filename = "plot3.png", plot3)

plot4 <- data4 %>%
  Extreme_events_trend_plot() + ggtitle("number of events with 3.4+ inches of rain / 6 hrs, in a 10-year rainfall event based on 1948-2011")
ggsave(filename = "plot4.png", plot4)

plot5 <- data5 %>%
  Extreme_events_trend_plot() + ggtitle("number of events with 0.8+ inches of rain / 1 hrs, in a 1-year rainfall event based on 1948-2011")
ggsave(filename = "plot5.png", plot5)

plot6 <- data6 %>%
  Extreme_events_trend_plot() + ggtitle("number of events with 1.8+ inches of rain / 1 hrs, in a 10-year rainfall event based on 1948-2011")
ggsave(filename = "plot6.png", plot6)

# finding and examining the extreme events ####
gaugedata %>%
  sum_rain(24) %>%
  flag_storms(period = 24, threshold = 5.2, keep_only_storms = FALSE) %>%
  write.csv(file = "rain52in24.csv")
gaugedata %>%
  sum_rain(6) %>%
  flag_storms(period = 1, threshold = 1.8, keep_only_storms = FALSE) %>%
  write.csv(file = "rain18in1.csv")

noaadata %>% filter(year == 2010 & month == 3 & day == 14) %>% View()
noaadata %>% filter(year == 2010 & month == 3 & day == 14) %>% summarize(sum(rain))
```r
noaadata %>% filter(year == 2010 & month == 3 & day == 13) %>% summarize(sum(rain))
noaadata %>% filter(year == 2010 & month == 3 & day == 15) %>% summarize(sum(rain))
noaadata %>% filter(year == 2009 & month == 11 & day == 28) %>% View()
noaadata %>% filter(rain > 1) %>% View()

full legislature data$location %>% as.factor() %>% unique()

storm_graph <- full legislature data %>%
  mutate(Data_Source = ifelse(location == "BOSTON MA US", "NOAA", "BWSC")) %>%
  filter(year == 2010 & month == 3 & day >= 13 & day <= 15) %>%
  ggplot() +
  geom_line(aes(x=date, y=rain, color=location, size=Data_Source)) +
  scale_x_datetime(name = "Date and Time",
                   limits = c(as.POSIXct("2010-03-13 0:00"), as.POSIXct("2010-03-16 0:00")),
                   breaks = seq.POSIXt(from = as.POSIXct("2010-03-13 1:00"),
                                  to = as.POSIXct("2010-03-16 1:00"),
                                  by = "6 hours"),
                   date_labels = "%Y-%m-%d %H:%M") +
  scale_y_continuous(name = "Rain (inches per hour)", limits = c(0, 0.7)) +
  scale_color_hue(drop = TRUE) +
  scale_size_manual(values = c(1, 4)) +
  theme(legend.position = "bottom",
         axis.text.x = element_text(angle = 45, vjust = 1, hjust = 1))
  # ggrepel::geom_label_repel(data = tibble(x = as.POSIXct("2010-03-14 16:15"), y = 0.45, label = "NOAA - BOSTON"),
  #   mapping = aes(x=x, y=y, label=label))

ggsave(plot = storm_graph, filename = "storm.png", width = 12, height = 6)
```
CHALLENGE
Future of Work

IN COLLABORATION WITH
MIT Initiative on the Digital Economy

TEAMS
Future HR
Resilient Workforce Initiative

No Worker Left Behind
Skill- and Preference-Based Job Retraining

StrawHats
Automation Displacement Bill and Skillmatcher App

We're Gonna be Okay
FutureSkills: A Three-Cohort Approach
**CHALLENGE STATEMENT**

**Future of Work**

*Balancing the Benefits of Automation and the Concerns of Workers*

**BACKGROUND**

Machine learning (ML) has the potential to disrupt the global economy and dramatically change the labor market by automating many tasks which have traditionally been labor-intensive. The effects of automation are both uncertain and uneven. It is important that policy makers can predict who will be affected by advances in automation and how it will impact their jobs, but this relationship is not always obvious. In some cases innovations which would appear to substitute for a role can, instead, compliment it. In other cases, the introduction of automation led to a dramatic reduction in the number of jobs.

Jobs can be modeled as a collection of tasks, some of which can be performed by a computer and some of which require human input. As machine learning capability evolves, the range of tasks that can be undertaken by computers expands. Those jobs in which all or most of the primary tasks can be performed by a computer are most at risk of automation. The provided dataset included metrics for how susceptible different occupations are to automation, generating a ‘suitability for machine learning’ (SML) metric.

**THE CHALLENGE**

Use the provided dataset to analyze the suitability of different jobs for ML to make policy suggestions designed to mitigate the negative consequences and enhance the positive effects of ML on the labor force.

**Questions**

- What policy interventions would you recommend for different groups of workers?
- Which occupations with high SML scores could easily transfer to other jobs?
- What changes would you make the SML score to make it more robust? How would you change the rubric to gain more information or reduce the number of questions?
- What tasks are both considered valuable and appear to be unsuitable for machine learning?
- Which occupations under threat represent the largest proportion of workers? How would you prioritize interventions in specific sectors?
- Which jobs would you expect to be similarly affected by automation?
Resilient Workforce Initiative

TEAM
Future HR

TEAM MEMBERS
Eric Magliarditi
Dylan Muramoto
Lewis Won
Lydia Zhang
WORKFORCE RESILIENCE INITIATIVE

Meet Jane
Jane works as a concierge at a 3-star hotel in Arizona. She has had this job for 16 years and has enjoyed it, but there are times where she gets bored due to the repetitive nature of the tasks involved. Although her main role is to work with hotel clients, much of her work involves clerical duties and completing requests from guests such as making restaurant reservations and relaying messages. Given all the coverage on new technologies such as machine learning, Jane has recently begun to question if she will still have her job in the next 5 years. Around town, she has seen businesses replace certain jobs like cashiers with computers. Jane is worried that soon her job will also be replaced by automation. How will she support herself and her two children then?

Automation
Jane is not alone. All around the U.S., more jobs are starting to be replaced by automation. For businesses, the incentive to automate tasks is to increase profit and improve efficiency, but at the cost of potentially laying-off workers. Faced with this situation, how can the government prepare workers in job categories that are at risk of being automated to be more resilient? Our main objective is to create a more resilient workforce. To do this, we recommend state governments incentivize key stakeholders to examine and promote the rebranding of tasks within similar jobs categories to ease the transition of workers into the future economy.

Data Analysis
The data provided to us in the competition classified an individual job using a six digit number that represents the Standard Occupational Classification (SOC) code or OES-specific code for an occupation provided by the Bureau of Labor Statistics (BLS). The first two digits of the number represent the highest level of job classification, which we define as the job category. To get a better understanding of the SOC code, Jane’s identifier is listed in Appendix A. We then determined which job categories had a “high risk of automation.” To measure high risk of automation, we utilized the suitability for machine learning (SML) index. SML score is defined as the perceived likelihood that an activity is to be automated by machine learning (ML) algorithms, rated from 1 to 5. We assume that a higher SML score is an indicator of a higher likelihood that a job will be replaced by ML capacity at the current time.

![Distribution Plot of Task SML](image1.png)

![Distribution Plot of Job SML](image2.png)
Tasks were classified as at high risk of being automated if their corresponding SML score was greater than 3.67. This represents the 75% quartile of the task level SML data. Task level data was utilized because it has more variance than job level SML. This can be seen in the two figures above.

We then classified each job into its corresponding job category, grouped the data according to the job category, and took the mean of the data during grouping. This allowed us to understand the average number of tasks, average number of high risk tasks, and average percentage of high risk tasks for a particular job category. See Appendix B for further breakdown.

Using the percentage of high risk tasks as a proxy of likelihood of automation, the top 6 high risk job categories are listed in Appendix C, and the bottom 5 job categories are listed in Appendix D. Jane works in the personal care & services job category, which was ranked 6th because 31% of tasks within the job category are high risk.

We then incorporated state-level data provided by the BLS. This data showed which states had a large percentage of their population that work in the at risk-for-automation job categories. The figure to the right shows a heat-map of the United States. The darker the state, the higher percentage of employees who work within the top 5 riskiest job categories.

**Rebundling**

Some job categories have a large proportion of high SML tasks which indicates a high susceptibility to automation, but that does not mean every task can be replaced by a computer or robot. To maximize efficiency, automatable tasks should be removed from the human worker’s responsibility and bundled with other high SML tasks to be completed by an automated system. The human worker should be assigned a set of rebundled tasks with low SML ratings. With this restructuring, workers and automation can focus on their comparative advantages and the overall workflow can be optimized. Rebundling is an important concept because instead of automation displacing workers, it can work in conjunction with humans as each can focus on tasks that are rebundled to utilize their respective advantages.

By examining the mean percentage of tasks that are at low risk for automation, we can get a quick proxy estimate of rebundling potential. For data purposes, we defined low risk tasks as those with a SML score lower than 3.25, or the 25% quartile of the task SML data. The top 5 low risk SML job categories can be seen in Appendix D.
**Policy Implementation**

Jane is notified that corporate may remove her job and replace services with an Internet-connected touchscreen computer. As she sits at home worried about unemployment, she sees a new ad by the Arizona State Government for Workforce Resilience Initiative. In the ads, Arizona state government promises to take three main actions to help workers like Jane to become more resilient:

1. Provide financial support to at-risk workers in order to encourage rebranding for retraining.
2. Provide financial support to educational institutions to encourage more affordable rebranding education and increase enrollment.
3. Incentivize companies and company consortiums with taxes and grants to encourage in-house trainings, apprenticeship programs, and retraining support for at-risk workers.

Before specifying each policy item, we first take a look at the key stakeholders in our policy and analyze their relationships, as seen in the figure above.

With targeted training and education, state governments can help create resilient workers by assisting workers to transit to new tasks within the same job categories. Resilient workers might also experience higher wages as their productivity increases. Our policy will create a positive feedback loop by increasing tax revenues for local governments to provide more subsidies towards training more resilient workers. The local governments can also provide tax incentives in the form of tax credits to companies which provide tuition support to workers to become more resilient, while levying a machine tax to make labor relatively cheaper to capital. Tax incentives can also be used by companies to provide in-house training for workers on tasks with low SML scores. We propose that state governments kickstart this positive feedback loop by implementing the following three policy pieces:
Goal 1. Provide financial support to at-risk workers in order to encourage retraining.

We will work with federal or local government agencies to rework existing grants, such as the Workforce Opportunity for Rural Communities Grants. Grants will be offered to workers to take up trainings in preparation for rebundled jobs. We also recommend that state governments offer income contingent loans (ICL) to augment grants such that trainees will be taxed a fixed percentage of future income gains accrued due to training. ICL allows the government to recoup a major portion of the initial loan, while ensuring equity by having a worker with a higher future income pay a larger absolute sum as compared to a worker who a lower increase in future income. Also, establishing portable lifelong learning accounts (LiLAs) to facilitate lifelong learning will be critical. These are tax-advantaged accounts that workers can use for skill development that stays with workers regardless of employer.

Goal 2. Provide financial support to educational institutions to encourage more affordable rebundling education and increase enrollment.

We propose that state governments also provide direct financial support to incentivize the restructuring of training programs in community colleges and online courses. Community colleges such as the Des Moines Area Community College have been providing retraining for workers aged 40 and above, with costs ranging from as little as $89 to $1000. These programs are usually completed within 2 years with an average completion rate of 85%. Such highly effective programs can be made more relevant by incorporating knowledge of what we already know about job rebundling according to SML scores. We can also work with online learning platforms such as Coursera to provide SML score rankings to their courses, which will provide more information to users to decide what lessons to take on. We will also work with information providers such as the US Department of Labor’s Employment and Training Administration (ETA) to guide workers towards taking on trainings in lower SML tasks.

Goal 3. Incentivize companies and company consortiums with taxes and grants to encourage in-house trainings, apprenticeship programs and retraining support for at-risk workers.

We propose that government may use tax credit akin to the Federal Research and Development Tax Credit to incentivize companies or company consortiums to provide inhouse training for employed workers. Given workspaces provide tremendous learning opportunities for workers. We also propose a tax credited apprenticeship program by companies to train external at-risk workers. In addition to tax credit, an existing case of reimbursable grant has also proven effective. For example, the Arizona Job Training Program awards up to 75% of eligible training expenses for employers who create net new jobs.

Discussion

Our policy recommendation takes a proactive stance to prepare workers for a ML-enabled labour market. We believe that this future could be a bright one for all American workers, but only if we put in the resources and kick-start a movement that helps workers to retrain and take on tasks which complement machines, and prevent workers from being displaced by machines.

Our policy recommendation is actionable and will generate practical, measurable results. Guided by the data, we can work with state governments to incentivize educational institutions, workers, and
companies to proactively think about optimizing the rebundling of tasks. We will encourage governments to actively measure the effectiveness of our rebundling recommendations in helping workers to meet the challenges of automation by actively collecting data and employing methods such as econometrics to evaluate policy impact.

Our policy recommendation helps rebuild trust of the American public in the government. Political Scientist Katherine Cramer at the University of Wisconsin-Madison showed that voters are distrustful of the government, partly because many voters believe that politicians are only interested in growing the economy even at the cost of workers losing their jobs. While some jobs may be lost due to offshoring, economists in general agree that automation is a much greater culprit for the decline of the manufacturing sector in the US. We believe that instead of resigning to the fact that public trust in the American government is at a historical low of 35%, the government can work to help regain the trust of the American worker by showing how government actions can help workers to not only keep their jobs, but also to be more productive and earn higher wages.

Conclusion

After Jane read the newspaper ad, she called the ETA to learn more about the new policy she just read about. She talked with an agent and learned more about task rebundling, and realized it was the key to not only keeping her job, but also increasing her productive potential in order to get that raise she has been working extremely hard for. Jane met with her manager, and discussed her future with the hotel. She explained that she did not have time to learn a new skill, but rather than lose employment, she could rebrand her skills to focus more on the client facing roles that have a much lower risk of automation. She pitched a new role to the manager, the corporate liaison, whose job it is to reach out to businesses throughout the U.S. who send workers for short and long term stays in Arizona. Given her background, she had the skills and the knowledge to succeed in this role, which would not only help her keep a job, but help the hotel bring in more revenue each year.

Jane’s story is just one example of what can happen when policy makers focus on the idea of rebundling. Although complete retraining of workers is something that can help ease the pain of automation, we view that a more realistic approach is to have people work within their skillset and industry, rather than make a complete career move when they have to worry about bills, children, etc. Our policy recommendation focuses on helping state governments to use financial instruments and tax benefits to incentivize key stakeholders to create a positive feedback loop. By doing so, we encourage the labor market to rebundle jobs with tasks to help workers become more resilient and survive the uncertain workplace future.
References


Appendix

Appendix A: Jane’s Profession SOC Code
- Personal Care & Service Occupation: 39
  - Baggage Porters, Bellhops, & Concierges: 39-6000
  - Baggage Porters, Bellhops, & Concierges: 39-6010
  - Concierge Unique SOC: 39-6012

Appendix B: Breakdown of Tasks & High Risk Task Percentage by Sector

<table>
<thead>
<tr>
<th>Sector</th>
<th>Total Number of Tasks</th>
<th>High Risk Tasks</th>
<th>Percent High Risk Tasks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Architecture and Engineering Occupations</td>
<td>18.467143</td>
<td>6.714385</td>
<td>0.298303</td>
</tr>
<tr>
<td>Arts, Design, Entertainment, Sports, and Media Occupations</td>
<td>18.203256</td>
<td>5.139585</td>
<td>0.268893</td>
</tr>
<tr>
<td>Business and Financial Operations Occupations</td>
<td>18.498000</td>
<td>6.182000</td>
<td>0.328042</td>
</tr>
<tr>
<td>Community and Social Service Occupations</td>
<td>18.714286</td>
<td>6.214286</td>
<td>0.330003</td>
</tr>
<tr>
<td>Computer and Mathematical Occupations</td>
<td>20.587300</td>
<td>6.093000</td>
<td>0.298573</td>
</tr>
<tr>
<td>Construction and Extraction Occupations</td>
<td>18.282992</td>
<td>3.061907</td>
<td>0.170919</td>
</tr>
<tr>
<td>Education, Training, and Library Occupations</td>
<td>20.300000</td>
<td>8.769667</td>
<td>0.348816</td>
</tr>
<tr>
<td>Farming, Fishing, and Forestry Occupations</td>
<td>16.523529</td>
<td>3.523412</td>
<td>0.203109</td>
</tr>
<tr>
<td>Food Preparation and Serving Related Occupations</td>
<td>18.547039</td>
<td>2.117847</td>
<td>0.104755</td>
</tr>
<tr>
<td>Healthcare Practitioners and Technical Occupations</td>
<td>18.357907</td>
<td>3.523256</td>
<td>0.187325</td>
</tr>
<tr>
<td>Healthcare Support Occupations</td>
<td>16.944444</td>
<td>3.833333</td>
<td>0.213391</td>
</tr>
<tr>
<td>Installation, Maintenance, and Repair Occupations</td>
<td>14.552256</td>
<td>2.898089</td>
<td>0.152014</td>
</tr>
<tr>
<td>Legal Occupations</td>
<td>14.200000</td>
<td>3.125000</td>
<td>0.214705</td>
</tr>
</tbody>
</table>

Appendix C: Top 6 Job Categories and Associated Mean High Risk Task Percentage

<table>
<thead>
<tr>
<th>Job Category</th>
<th>Mean High Risk Task Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Office &amp; Administrative Support</td>
<td>42%</td>
</tr>
<tr>
<td>Education, Training, &amp; Library Occupations</td>
<td>35%</td>
</tr>
<tr>
<td>Sales &amp; Related Occupations</td>
<td>34%</td>
</tr>
<tr>
<td>Community &amp; Social Service</td>
<td>33%</td>
</tr>
<tr>
<td>Business &amp; Financial Operations</td>
<td>33%</td>
</tr>
<tr>
<td>Personal Care &amp; Services</td>
<td>31%</td>
</tr>
</tbody>
</table>
Appendix D: Bottom 5 Job Categories and Associated Mean High Risk Task Percentage

<table>
<thead>
<tr>
<th>Job Category</th>
<th>Mean High Risk Task Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Food Preparation &amp; Serving</td>
<td>10%</td>
</tr>
<tr>
<td>Installation, Maintenance, &amp; Repair</td>
<td>15%</td>
</tr>
<tr>
<td>Construction &amp; Extraction</td>
<td>17%</td>
</tr>
<tr>
<td>Healthcare Practitioners &amp; Technicians</td>
<td>19%</td>
</tr>
<tr>
<td>Transportation &amp; Material Moving</td>
<td>19%</td>
</tr>
</tbody>
</table>

Appendix E: US Map for Most High Risk Job Category: Office & Administrative Support

US Map Breakdown for Most High Risk Job Category: Office and Administrative Support Occupations
POLICY BRIEF
Skill- and Preference-Based Job Retraining

TEAM
No Worker Left Behind

TEAM MEMBERS
Zhen Dai
Judy Shen
William Yuan
David Zhang
No Worker Left Behind

David Zhang, Zhen Dai, William Yuan, Judy Shen*  

April 7, 2019

1 Introduction and Problem Definition

The well-being of workers displaced by technological advancement is an increasing area of policy concern due to the rise of artificial intelligence (AI) and automation. To address the problem of labor displacement effectively, these policies should focus on both the right areas and the right intervention strategies. The focus of our project is to provide a recommendation engine that directs job re-training efforts to take advantage of individuals' unique skills, knowledge, and experiences.

AI is often quoted as the culprit for decreasing wages and increasing unemployment, so policy makers may be tempted to focus on jobs that are most susceptible to AI and automation. However, increased worker efficiency and generation of complimentary jobs often belie the naïve assumption that AI displaces human workers from their jobs. It is therefore important to direct policies to address sectors where AI/automation most adversely affect employment. We explore this relationship by looking at projected employment growth in each sector. We find that while jobs with a greater susceptibility to machine learning have lower projected growth, the correlation is two times bigger for automation. Therefore, in the near term, automation may be a greater policy concern than AI. Furthermore, many jobs are predicted to grow despite automation and AI susceptibility, suggesting that policies should target jobs based on more holistic considerations.

In addition, policy interventions often assume that worker retraining programs should focus on increasing computer literacy among displaced workers. This ignores the unique experiences and skills of these workers, whose occupational placements often reflect their advantages in or preferences for knowledge and skill other than computer programming. Therefore, instead of instituting policy interventions that place a singular emphasis on computer literacy, policy makers should look for the most efficient career transformation pathway catered to individual workers.

To that end, we developed a recommendation engine for workers worried about job displacement by AI. We study the relationship between job growth and tendencies of jobs to be replaced by AI and automation to identify the critical area that policies should focus on. In addition, we investigate efficient ways to retrain workers whose jobs have high risks of being displaced by AI/automation. Specifically, we consider the difficulty of job transition based on the skill/knowledge requirement as well as the likelihood of job transition based on historical data. By considering these two factors, the recommendation engine output efficient career-transition strategies for workers in each occupation. These strategies recommend jobs less threatened by AI. In addition, based on historical patterns of job flow, the recommended jobs require skills and knowledge that workers are most willing and able to acquire. Policy-makers can use the recommendation engine to guide their efforts in funding and/or designing occupational re-training programs. This engine can also be easily distributed to stakeholders in the private sector to help employers target potential hires.

2 Data Analysis

2.1 Data sources

We used the following data in our analysis:

*The order of the authors reflect the probability that their field will be automated according to the Brynjolfsson et al. [2018] SML score.
• The provided susceptibility to machine learning (SML) scores by job and task, as well as skills, knowledge, and wages by job, from Brynjolfsson et al. [2018],

• Job transitions by individual, from IPUMS-CPS in Flood et al. [2018],

• Susceptibility to automation, from Frey and Osborne [2017],

• Projected job growth for each job, from the Bureau of Labor Statistics (BLS).

The datasets define jobs at different levels, from O*NET-SOC (most specific) to SOC to CPS (most general). We use the BLS crosswalk\(^2\) when merging data and take the means of skill/knowledge scores when we move to a more general level.

2.2 Which workers are at risk of displacement?

Jobs that are most susceptible to machine learning and automation may not be the ones where workers are at a greater risk of displacement. As explored in Autor [2015] and Acemoglu and Restrepo [2018], automation of tasks within jobs could free up the worker to do more valuable tasks, leading to demand for more workers in the task. Therefore, for policy makers, it is not enough to simply look at jobs with a high susceptibility to AI and automation displacement, but rather at jobs with anticipated declines in employment. We investigate the correlation between employment growth and susceptibility to AI and automation, where employment growth is from BLS data on projected employment growth by job (2016 to 2026).

![Figure 1: Employment growth and susceptibility to ML/Automation](image)

Figure 1 shows the correlation between employment growth and AI/Automation. The first subplot illustrates that the SML score and the BLS projected 10 year growth has a weak correlation of \(r = -0.14\) (p-value = 3.9e-5). The second plots projected growth against our own measure of susceptibility to machine learning: the fraction of tasks (weighted by task frequency) that are susceptible to machine learning (with a task-level SML score greater than the 75th percentile). There is no significant correlation between this score and projected growth. Finally, the third subplot illustrates a much stronger correlation between automation likelihood defined in Frey and Osborne [2017] and projected 10 year growth at \(r = -0.38\) (p-value = 2.2e-16).

We draw the following conclusions from our correlation analysis.

1. In the near term, automation appears to be a greater threat to 10-year job growth than machine learning.

2. Many jobs are predicted to grow despite automation and machine learning susceptibility. Therefore, when deciding which workers to target, policy makers should consider the productivity effect of automation and the creation of new tasks which could have a countervailing effect on the demand for jobs.

\(^2\)https://www.bls.gov/cmp/documentation/crosswalks.htm
2.3 A recommendation system which takes each individual’s unique skills, knowledge, and experiences into account

For each job, the Euclidean distance with every other job was computed using the skills and knowledge importance ratings. That is, for jobs u and v with the average level of skills and knowledge from Brynjolfsson et al. [2018],

\[ d(u, v) = \sqrt{\sum_i \beta_i(u)(u_i - v_i)^2} \]  

(1)

Where \{a_i, v_i\} are the skills and knowledge associated with jobs u and v, and \beta_i(u) are job-specific skill acquisition weights. The weights \beta_i(u) were computed based on a linear regression of job transition probability against the jobs that were historically transitioned to from IPUMS-CPS in Flood et al. [2018]. That is,

\[ \delta(u) = (X^tX)^{-1}X^ty \]  

(2)

Where X is a matrix containing the difference in skill and knowledge between a potential new job and a job u, and the outcome is defined as \( y_u = \frac{p_u(u)}{\bar{p}_u} \), where \( p_u(u) \) is the proportion of job switches from u to v and \( \bar{p}_u \) is the prevalence of job u in the population. Then, we define the weights \( \beta(u) = const - \delta(u) \), so that the more negative the regression coefficient on a skill or knowledge, the greater the weight we put when computing the distance. The intuition is that the regression coefficients reflect the historical likelihood for someone in job u to improve upon a skill/knowledge when switching to a new job, and the greater that likelihood the less weight we want to place on it when computing the distance.

Figure 2: Skill/Knowledge-Job Heatmap

Figure 2 shows the clustering of jobs based on their skill/knowledge components. The intensity represents the importance of a given skill to a given job. As an example, the highlighted region corresponds to medical skills and medical professionals- medical skills are not found to be strongly important to other kinds of jobs. The low importance that other jobs assign to medical skills is caused by three factors: i) other jobs assign a low importance to medical...
skills, ii) individuals from jobs with medical skills tend to switch to other jobs with medical skills, and iii) a high skill/knowledge-based barrier to entry is observed in jobs that require medical skills.

These metrics were used to construct a unified job distance matrix. We utilize Truck Drivers as an example. Figure 3 describes the relationship between the automation likelihood of a job, and its distance to Truck Driving. We plot a Pareto frontier along the top of this distribution, describing the optimal jobs for a given Automation Likelihood in terms of Distance. Only jobs above a wage (90% of the current wage) and growth floors (the median of growth of all sectors) relative to truck driving are used in this analysis.

The job of computer programmer represents a convenient comparison in this case. The jobs in the green shaded region represent transition opportunities for a truck driver that are more Pareto efficient than computer programmer. Power line installer/repairer has been highlighted as a particular job of interest: relative to programmers, it i) has a comparable wage, ii) a higher growth potential, iii) lower chance of automation, and iv) greater ease of retraining for truck drivers.

3 Recommendations

We suggest that policy efforts to alleviate the labor force disruptions brought on by AI and automation be directed in a manner is sensible and cost effective. To that end, we suggest that policy makers consider the effect of AI and automation on job growth, and if potential disruption is identified, to use our recommendation engine to direct retraining/suggestions in an effective manner.

More specifically, our recommendation engine can be provided to both private and public sector stakeholders to focus their career guidance efforts. In the public sector, policy makers in career counseling offices can use the engine to guide job-seekers that come to them for advice, and recommend training programs that are on the Pareto frontier. Before fully implementing the program, it is advisable to set up a pilot program that are limited in region and/or the number of jobs. We utilized our recommendation engine to nominate several candidate pilot projects based on the automation risk of the starting jobs, the ease of transition to candidate jobs, and the resistance of the candidate job to automation in the future.

- Truck Driver → power line installer and repairers, drill operators, or captains of water vessels.
- Secretary → credit counselor, loan interviewer, or legal assistance.
- Retail Sales → home health aides, cargo and freight agents, tax preparers.
The pilot program can be evaluated based on a post-placement job satisfaction survey. If the pilot program is successful, funding can be re-appropriated to training programs that have a high demand based on our recommendation analyses. Importantly, because this program would be refocusing the efforts of existing job retraining programs, it would be cost-neutral.

For the private sector, the recommendation engine can be provided to employers to advise them on where to look when to hire or direct internal re-training efforts.

References


POLICY BRIEF
Automation Displacement Bill and Skillmatcher App

TEAM
StrawHats

TEAM MEMBERS
Godha Bapuji I.
Mahiti Bapuji I.
Harrison Hur
Shriank Kanaparti
Kane Magnuson
Policy Paper - Challenge: Future Of Work
Authors: Kane Magnuson, Godha Bapuji I., Harrison Hur, Mahiti Bapuji I., Shriank Kanaparti

Introduction

Individual states are far more effective at helping unemployed job seekers displaced by automation to acquire jobs than federally managed career training programs. The effectiveness of state-funded job training programs are important to note because states have disparate levels of needs, which federally-managed programs are unable to satisfy due to their centralized approaches. It is also significant to note that data shows that Federal job reskilling programs do not achieve the expected success rates, due to lack of intimate familiarity with state-level socioeconomic conditions. Therefore, the implementation of job reskilling programs in individual states is more effective at mitigating unemployment caused by automation.

The aforementioned argument is grounded largely within the definition of machine learning (ML). ML is defined as the ability in computers to react, learn, accomplish objectives and predict scenarios to allow humans to make more-efficient real-time decisions.

Methods & Results

A 2018 study by Brynjolfsson, Rock and Mitchell aimed to quantify the vulnerability of a variety of occupations to automation by conducting a detailed survey. The survey consisted of 23 questions relating to over 18,000 tasks deriving from occupations found in the O*NET database. For a more in-depth explanation of the methods used by the study, please refer to the original paper.

In regards to original data analysis, we initially examined the outliers and the general characteristics of figure 1.1. The more leftwards the data points travel, the less susceptible the corresponding occupations are to machine learning. The figure’s lower-left quadrant contains an outlying blue dot, which is occupations in the healthcare support occupations. A majority of positions paying greater than $75/hour are also in the healthcare industry. Moreover, the concentration of data points are clustered around suitability for machine learning (SML) = 3.45 and an hourly wage below $35/hour. These are primarily sales and office/administrative support. As shown in figures 3 and 4, the work activities associated with these jobs involve computer-readable information and machine input.

These types of tasks are ideal for machine learning, as they are also repetitive and deal with data processing and analysis. While the data in figure 2.1 for questions 1 and 2 point to the same conclusion, questions 15 through 18 tell a different story. These four questions ask whether the person’s daily tasks involve data processing with video, speech, and audio. Although these are areas machine learning would be expected to excel in, analysis of job zone data suggests otherwise. This relation can be attributed to the likelihood that machine learning cannot yet understand more intricate data, such as a high-level project execution plan at a business. As mainstream machine learning technologies become more capable of handling such data, it can be expected that higher-level positions dealing with these tasks will be affected. As such, when

\[\text{Reference}\]

1 https://www.onetonline.org/
2 http://faculty.chicagobooth.edu/chad.syverson/research/aiparadox.pdf
3 See Appendix A
training human workers displaced by such technologies, it will be important to consider how data processing can be “re-skilled” into new positions.

On the contrary, questions 7, 10, 14, and 20 have exceptionally negative percent changes as the position requires more training. These occupations each require abstract, irregular, and long-term planning. Examples include surgeons and C-suite management; jobs most immune from machine learning’s effects in the labor market.

We have also categorised the most “replaceable” of the O*Net jobs by ones with SML>3.6 and largest resource pools, which are the occupation’s total employment multiplied by their annual salaries. This assumes however, that companies will replace employee groups using the most amount of company resources and ones that can also be easily automated. Under this concept, the most susceptible would be office clerks, sales representatives, secretaries, administrative assistants, and accountants. A few occupations we have determined to be the most resistant to machine learning change are chemical engineers, pediatricians, physicists, personal financial advisors, and software developers.

Through these data sets, the types of employees that machine learning will most likely uproot are made clearer. Tasks that involve manual computer input and retrieval or processing of basic information will be affected early on. Occupations that are not upper-level with high levels of training, or do not involve complex, abstract thinking will also begin to see more widespread use of automation in their workplaces, if not so already.

Policy Recommendations

The dataset provided does not contain a geographic distribution but in order to contain the scope of the policy to manageable chunks, we extrapolated information based on levels training required for each type of occupation, and extending that to see which regions of the US contain the at-risk occupations with a high smr score. Office clerks, general sales representatives, and administrative assistant represent a relatively small portion of Texas’s population but account for a large share of the workers at risk for losing their jobs to automation. Therefore, it makes sense to focus policy on this category of high-risked displaced workers. Based on the presented analysis, we formulate the following three categories of policy recommendations, aiming to balance ambition with feasibility.

Automation Displacement (AD) Bill to U.S. Congress

- Regulates rapid waves of automation into timed phases over a period of time so that the work force can adapt to changes easier via the top-down governance support model.
- Because bottom-up policy implementation is both effective and efficient, AD bill provides federal grant money to states to retrain the unemployed displaced by automation via state-organized programs.
- Outlines requirements that grant recipients must meet in order to receive grant money. For example, KPIs are used to measure the success rate of programs receiving the grant.

Expansion of Texas’s JET Grant Program

- Expand Jobs & Education for Texans (JET) Grant Program to include all of Texas’s public schools due to past successes in equipping the workforce with in-demand skills.
Facilitate pilot state-sponsored, proposed SkillsMatcher app, in equipping unemployed workers displaced by machine learning-based automation with harder to displace skills and work. While a simple keyword search returns several job hunting websites and apps, results are scarce for literature on why despite several such tools, the labour market is unable to fill in the demand and supply gaps. One theory is outlined in a recent paper by the MIT Media Lab\(^4\). We base our policy proposal in line with their observations.

**Pilot State-sponsored SkillsMatcher App to work in line with SkillsScape**

Texas has been chosen as our pilot program candidate due to its population size and wide range of industries. Studies of the state’s labor force show that there are approximately 335,000 office clerks. The Bureau of Labor Statistics also estimates there to be a decrease of 31,800 office clerks in the United States between 2016 and 2026, an indication that automation has already begun to replace human office clerks\(^5\).

- Incentivize Texas-based industries with WOTC and other state-based tax credits to both sign up on the state’s SkillsMatcher to maintain accountability while empowering companies to hire the unemployed, displaced workers from the database.
- Automatically register unemployed office clerks, who are displaced due to automation driven by machine learning, on the state-owned SkillsMatcher app.
- Collect and analyze results from registered unemployed, displaced workers’ SkillsScape tests to either connect them with jobs that enable them to use their transferable skills, or provide them training via JET grant program to prepare them for a new career field. The aims to close the gap between the talent, skills, and job preferences of an individual thereby providing a closer match to the skills of the individual while giving them the opportunity to upskill or reskill suitably at their convenience, keeping in mind their socio-economic limitations and barriers prevailing at the time due to unemployment.

**Counter Argument**

One could object to the Automation Displacement Bill by asserting that some states may lack the resources to fulfill the grant’s requirements because of their labor force. It is important to note that although “automation phasing” may be introduced by the bill, some states such as Michigan could be totally devastated by a phase of automation due to the vast amount people working in automatable jobs there. Such an occurrence would make it hard for the AD bill to succeed in helping people regain a job that isn’t at risk of being automated any time soon. Thus, some states may have struggle to replace the labor due to a lack of jobs within the state.

**Response**

---

\(^4\) Unpacking the polarization of workplace skills
BY AHMAD ALABDULKAREEM, MORGAN R. FRANK, LIJUN SUN, BEDOOR ALSHEBLI, CÉSAR HIDALGO, IYAD RAHWAN
SCIENCE ADVANCES 18 JUL 2018 : EAAO6030

Although the objector makes a good point, state governments have total control over the phasing element within their own state, which enables certain states to slow down wide-scale automation if it threatens the state’s labor force. Thus, states are able to control the momentum and strength of the aforementioned events. States are able to still find ways to fulfill KPI requirements because “automation phasing” provides state policymakers and governors the ability to halt mass-waves of automation for short periods of time so that they can find solutions.

**Conclusion**

Individual states are far more effective at helping unemployed job seekers, displaced by automation, acquire jobs than federally managed career training programs. It is significant to consider the data, which shows that Federal job reskilling programs do not achieve the expected success rates due to lack of intimate familiarity with state-level socioeconomic conditions. Thus, individual states implementation of job reskilling programs are more effective at mitigating unemployment caused by automation. Therefore, the aforementioned three policy objectives will succeed in effectively mitigating the deleterious effects machine learning-driven automation will have.
References

2. https://static1.squarespace.com/static/5a34214132601ed822bc4f1a/t/5b33f3ef352f5361ae969f16/1530133762151/MIT+Policy+Hackathon+Proceedings_finalWEB.pdf
17. https://thebestschools.org/magazine/future-meaningful-work/
25. https://learning.linkedin.com/blog/future-skills/a-list-of-the-jobs-that-are-most--and-least-likely-to-be-replaced
39. https://blog.opportunityatwork.org/2yearanniversary-40bd3972ba64
42. https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-will-redesign-not-replace-work
44. https://knowledge.wharton.upenn.edu/article/the-forces-shaping-the-future-of-work/
46. https://www.simplypsychology.org/maslow.html
Appendix A

Figure 1.1: Bubble chart depicting occupations graphed by SML score and hourly wage.
Industry
- Architecture and Engineering Occupations
- Arts, Design, Entertainment, Sports, and Media Occupations
- Building and Grounds Cleaning and Maintenance Occupations
- Business and Financial Operations Occupations
- Community and Social Service Occupations
- Computer and Mathematical Occupations
- Construction and Extraction Occupations
- Education, Training, and Library Occupations
- Farming, Fishing, and Forestry Occupations
- Food Preparation and Serving Related Occupations
- Healthcare Practitioners and Technical Occupations
- Healthcare Support Occupations
- Installation, Maintenance, and Repair Occupations
- Legal Occupations
- Life, Physical, and Social Science Occupations
- Management Occupations
- Office and Administrative Support Occupations
- Personal Care and Service Occupations
- Production Occupations
- Protective Service Occupations
- Sales and Related Occupations
- Transportation and Material Moving Occupations

Figure 1.2: Legend for figure 1.1 (By Industry)

Figure 2: Questions whose responses indicated less autonomy as the O-Net job zone increased from one (minimal preparation) to five (extensive preparation).
Figure 2.1: The percent differences from job zone 1 (minimal preparation) to job zone 5 (extensive preparation) for questions 1~21.

Figure 2.2: Legend (By Question)
**Figure 3:** The five activities valued most by a market analyst. In descending order: Getting Information, Analyzing Data or Information, Interpreting the Meaning of Information for Others, Processing Information, and Interacting with Computers.

**Figure 4:** The five most valued work activities categorized by top susceptible occupations. In descending order, the top five work activities are: Getting Information, Communicating with Supervisors, Peers, or Subordinates, Interacting with Computers, Establishing and Maintaining Interpersonal Relationships, and Processing Information.
POLICY BRIEF

FutureSkills: A Three-Cohort Approach

TEAM
We're Gonna be Okay

TEAM MEMBERS
Joseph Keller
Chris Kreis
Victoria Pham
Sam Sands
Rebecca Xiong
FutureSkills: Uniting Human Intelligence to Mitigate AI-Driven Workforce Changes

By Chris Kreis, Rebecca Xiong, Joseph Keller, Sam Sands, and Victoria Pham

Abstract

The rise of machine learning poses a novel threat to the future of work in the United States, where the realm of decision making, formerly an exclusively human realm, is now open to automation or computerization. This new frontier of disruptive reformation spurs a heightened level of public anxiety in the economic implications of machine learning, through concerns raised in how these impacts will reshape the labor force that has, until now, been relatively unaffected by rising automation.

Historically, automation has been a threat to jobs, but machine learning appears to be causing particular concern among workers and policymakers. Machine learning promises a new mode of machine-based, computational behaviors, designed with human thinking like capabilities. While still in its infancy, this technology already proves to best humans at many tasks, including certain strategy games, image recognition and categorization, and predictive analytics. What people in the general public find uncomfortable about machine learning is the machine’s inherent ability to learn without explicit instructions; directly through observation, the machine infers patterns that humans might not consciously observe or process. This unique and powerful adaptive feature within ML is, thus, perceived as a potential threat, when humans view its behavior as uncontrollable.

We examined the suitability for machine learning (SML) score created by Brynjolfsson et al., 2018 to analyze for potential impact derived from the perceptions of those currently employed in the workforce, and provided critiques and suggested improvement. We examined the job population at high and medium risks of displacement by machine learning as measured by combined SML scores over their composite tasks. The resulting 22 million represents a significant percent (17%) of the workforce.

Our proposed solution looks at three cohorts: those displaced and are aging out, those displaced and would like to retrain for new jobs, and those who are newly entering the workforce and would no longer have the options of those jobs.

A central piece of our proposal is the National FutureSkills Fund, to enable companies, cities, or unions to apply for funding for their proposed re-skilling plans. We want to enable those at the frontline of machine-learning changes to innovate and incorporate the advances of ML in their re-skilling plan. We further believe in more quickly share successes of these training with other companies, cities, and unions by establishing a FutureSkills Knowledge Base, a platform where re-skilling entities can report on their successes, and job transitioners can share their
own experience and learn from and support each other. We believe in our collective human intelligence and creativity to better harness machine intelligence for common gains.

SML and improvements

The Brynjolfsson et al. study has a narrow scope of analysis which may limit broader generalization. The methodological approach focuses on technical feasibility. Consequently, the rubric lacks consideration for the economic, cultural, societal, and legal elements that could determine potential adoption and implementation of these technologies. In addition, this examination assumes a short-term timeline for this transition, despite uncertainty surrounding how this disruption may influence labor markets over a longer timescale. Nevertheless, there remain substantial barriers that inhibit scientists from accurately measuring artificial intelligence and automation on the future of work (Frank et al., 2019).

One way that we may be able to increase the accuracy of the data would be to add a few of questions to the survey that address societal or cultural concerns. An example we have referred to before is that of the mortician or undertaker. These were jobs that scores very high on the technical feasibility for SML, but culturally or socially may, in fact, be much less feasible. How many people would really be comfortable with a computerized robot preparing or embalming their father, mother, or child? We may be able to include questions such as

- How much customer interaction does the job require?
- What’s the comfort level people would have with the specific task or job being performed by a machine?
- What degree of empathy/emotional intelligence does the task require?

Jobs most at-risk based on SML score

In order to determine the validity of these concerns and their ability to be addressed, we need to better understand the scope and scale of the problem. To do this we looked at data provided by MIT from Brynjolfsson. SML scores were generated on a crowd-sourcing platform with responses from study participants. Each participant shared their own work history and was shown a job in the economy and a description of its functions. Next, individuals are provided with a task that is commonly performed by a specific worker in a given sector. Assuming an adequate comprehension of the task, each participant states his/her agreement with a statement regarding SML. A higher SML value suggests a task where ML has a greater potential to transform that corresponding role.

Understanding the limitations of the data, we decided to look at jobs that we defined as a moderate to high susceptibility to substitution by machine learning to computing. We decided to look at jobs with an SML score of at least 3.6 as moderate risk and scores of 3.7 or higher as high risk. On the chart below we graphed the jobs by their SML score and their median annual income, with a dropline indicating our cutoff point for moderate and high-risk jobs.
Figure 1. SML scores are shown relative to each job’s median income. The size of each circle reflects the number of workers in that field. The orange reference line indicates the cutoff for moderate risk, while the red reference line is the cutoff for high risk.

High Risk Jobs

There are about 3.5 million people or 2.5% of the total labor force working in jobs we have classified as high risk, we can see this population broken into the most prevalent jobs below. The largest jobs affected were bookkeeping, accounting, and auditing jobs, Human resource specialists, tellers, and file clerks.

Medium Risk Jobs

In addition to the highest risk jobs, we also included jobs we determined to expect a moderate impact by machine learning. There are 18.5 million people or ~14% of the labor force working in jobs categorized as moderate risk. The biggest jobs affected were cashiers, office clerks, secretaries and administrative assistants, and sales.

Combined between the low and moderate risk, we are looking at roughly 22 million people who may experience either a shifting of their roles or even replacement at their job with the advent of machine learning. These 22 million people form 16.5% of the labor market, which indicated to that issue facing us stretches beyond the 22 million that are currently at risk of seeing changes or loss of jobs, it means that we have to train the upcoming generation to fill these new and changed roles. With 16.5% of the labor market affected, we will need to provide vocational training and technical education options to students on the cusp of entering that workforce.
Proposed solutions

It makes sense to develop policy recommendations which cover three anticipated cohorts of workers.

1. **Displaced employees**: Employees with only a few more years in the workforce, when faced with job obsolescence, might find that they do not wish to make the efforts required for transitioning to a new job. Additionally, the financial investment required to retrain these employees might be better given to them in the form of a pension or early retirement plan.

2. **Transitioning employees**: For employees who still face significant time in the workforce, the goal should be to support these employees while they undergo job retraining. Support would include both a living wage stipend and tuition assistance for retraining. These stipends might take the form of grants or low-interest loans, perhaps with liberal payback policies.

3. **New entries to the job market**: For employees newly entering the job market out of high school or college, the educational institutions must be responsible for training them for the future, not the past. Schools must anticipate, based on data much like that in the Brynjolfsson study, which jobs are obsolescent, and which jobs will resist automation. Additionally, institutions should instruct students when possible on how to utilize new
technologies themselves, to form new types of jobs. Finally, institutions should encourage critical thinking and problem solving, and, promote lifelong learning initiatives.

We believe that public and private initiatives should address these cohorts:

1. **Tax-deferred accounts** for Displaced Employees for transition support (which might roll to retirement if never used). And/or allowing withdrawal from retirement accounts for the purposes of job retraining.

2. **Re-Skilling Subsidies & National FutureSkills Fund** for Transition Employees - Create a fund to subsidize corporate, municipal, or union reskilling programs. These funds would be allocated based on application process similar to standard grant applications. Allowing companies to take the initiative on training allows them to customize that training to the skills that they know they will need moving forward. This allows them to be drivers of innovation in their respected fields, being better able to take advantage of the new possibilities machine learning offers.
   - The most recent prominent example of this is AT&T’s Future Ready program, where the company currently spends $1 billion to ‘reskill’ 100,000 of its workers by 2020.
   - If AT&T’s model is accurate, we estimate a cost of $10,000 per worker to adequately reskill him or her with the competitive edge needed in a world with more machine learning. While this may vary widely across fields, it serves as a cornerstone for estimating the costs of large scale training programs across the nation.
   - If we want the response to be both rapid and comprehensive, policymakers must subsidize this cost so that companies who do not have the available capital of AT&T are able to begin their own initiatives. If we assume a rough cost of $10,000 per reskill, we will need about $35 billion to subsidize the cost of reskilling the 3.5 million workers in high-risk jobs and $220 billion to cover the re-skilling of all workers in the moderate to high-risk jobs (over a number of years). We call this the **National FutureSkills Fund**. This money would be set aside, with an agency created to evaluate applications and distribute money accordingly.
   - While this is a substantial investment, it allows us to avoid the massive cost of millions losing full or partial employment and ensures that newly re-skilled workers are able to keep gainful employment and continue to pay taxes.

3. **Strategic investments upstream in the US education and training pipeline.**
   - Apprenticeships, internships, and cooperative programs can provide a desired entry point to jobs for younger workers. This may include broad career exposure to relevant fields, opportunities for applied learning, and cohort approaches that may improve matriculation into ML-friendly roles. These initiatives may be deployed at the state level and scaled for wider adoption.
   - Transform curriculum in final high school years and initial college years so that students can prepare for an automated workforce.
Complement technical skills and training with a focus on critical thinkings and communication skills, and promote flexibility and lifetime learning that improves workforce satisfaction and dignity.

Conclusion

As we move forward, it is very hard to predict exactly what changes the normalization of machine learning will cause in our workforce. The question about how machine learning will affect a particular job or set of tasks is based on much more than simply technical feasibility, but for now, that gives us an idea of the scope going forward. We do know there will be some effect, and we can estimate a range for this effect by looking both at this data and the actions and studies of companies in fields that use machine learning. Therefore, our best action is to create a flexible solution that can adapt to the changing landscape as it comes. The National FutureSkills Fund allows us to have a flexible, solutions that can meet needs as they develop. While a significant investment, the one-time reskilling cost of $10,000 per employee is a drop in the bucket compared to the ongoing costs of supporting these workers if they find themselves with a lack of employable skills. Additionally, these re-trained workers will be filling high demand jobs and contributing taxes that they would have otherwise been unable to contribute. We will be paying for this one way or another, it is imperative that we be proactive instead of reactive.
References


https://www.brookings.edu/research/meeting-the-automation-challenge-to-the-middle-class-and-the-american-project/

https://www.brookings.edu/blog/the-avenue/2019/02/25/automation-and-ai-will-disrupt-the-american-labor-force-heres-how-we-can-protect-workers/


**CHALLENGE**
**Health**

**IN COLLABORATION WITH**
MIT Critical Data

**TEAMS**

**Dream ER**  
Three Pathways for A Smoother Flow

**The Emergency Crew**  
Optimizing Emergency Department (ED) Utilization

**ER-Flash**  
Improving ER Efficiency, Effectiveness, and Accessibility

**The Impatients**  
Reducing Emergency Room Congestion: An Adaptive Priority Queueing Model

**Marshmallow**  
Patient Allocation Policy Simulation

**QueueBusters**  
Clearing the Queue: Optimizing Emergency Services for Everyone
CHALLENGE STATEMENT

Health

Evaluating Emergency Department Utilization to Optimize Hospital Design and Workflow

BACKGROUND
Optimizing Emergency Department (ED) utilization is an important aspect of managing a health care establishment. Emergency departments are known to be the catch-all and first line of contact with the hospital for many patients and must be prepared to handle varying patient volumes throughout the day and all through the year.

By evaluating the ED utilization, patient wait times, volume and flow, health care facility managers will be better prepared to optimize the design and workflows within their EDs, and will be able to anticipate periods of high demand, in order to preemptively prepare through appropriate staffing and inpatient bed utilization.

Emergency Queue Management System (EQMS) is a queue management and automation software built for an emergency department at a hospital. The application enables staff to perform patient triaging functions and patient queue management, in addition to collecting patient traffic and outcome data for an average of 120 patient visits a day.

THE CHALLENGE
We aim to analyze the patient volume trends, how they change over the months, and fluctuate in response to special events like national holidays, inclement weather, and religious events like the month of Ramadan. Additionally, we may perform comparisons of the visit outcomes by diagnosis and treatment status, waiting time and admission stand-by duration and how they respond to patient volume.

Questions
- How well does the triage score reflect the urgency of treatment for patients?
- What strategies could the hospital choose to use to avoid moral hazard, without rescinding on its promise to offer ‘free care for all, for life’, offending the public who finance the hospital, or decreasing quality of care for patients?
POLICY BRIEF
Three Pathways for A Smoother Flow

TEAM
Dream ER

TEAM MEMBERS
Kyle Heuton
Ruochen Sun
Steven Susana-Castillo
Tingting Xu
Ruochen Zhang
**Dream ER: Three Pathways for A Smoother Flow**

**Executive Summary**

Emergency departments (ED) are having difficulty managing their patient flow due to patients using the ED as their go-to spot for medical treatment. Many ED, including at Hope Hospital in the Middle East, has implemented emergency management systems (EMS) to improve their patient flow which would lead to better quality of care for patients. Even with these systems in place, there are still many opportunities for improvement given the data collected from the EMS. Not only would an improved patient flow save the hospitals money, but clinicians would also have lower stress managing increased workload, and patients would have a much better experience in hospitals by waiting less to get the care they need (Wilson & Nguyen, 2004). By identifying three main problems in the current ED system, we provide policy options for each of the problem.

The first problem is the major delays in the flow of patients in the ED in the month of July 2018 which follows the month-long holiday, Ramadan. Out of all delays between April 2018 and March 2019, 50% of delays were in the months of July and August. Given the stark density in July compared to the other months, this is a potential period of focus for critical intervention to reduce the delay for patients visiting the ED.

The second problem is to characterize patients who will be more likely to be admitted in the final decision, which is an important urgency indicator of treatments for patients. With only information from the assessment stage, (‘complaint’, treatment status’, ‘refer; ‘recent discharge’, ‘diagnosis’, ‘triage category’, we aim to predict the probability of a patient to be admitted in the final decision. By identifying the high-risk patients who need urgent treatment, the ED resources can be utilized more efficiently.

The third problem is the existing of frequent flyers in ED. Frequent flyers, which are patients who go to the ED more times than is considered necessary, are a burden to the ED and accrue additional costs for the hospitals. Many patients who wait too long may eventually leave without being seen, and as a result, come back at a later date with a worse health condition (Studer et al., 2013). These patients can be flagged and targeted for more involved care to address their needs outside of the ED to prevent further visits to the ED.

Based on the analysis, we propose three policy recommendations to deal with the problem.

1. Establish an independent rapid clinic unit for post-chemotherapy infections to deal with the peak demand in July.
2. Create non-clinical navigators to enhance access and care continuity, and stratify the high-risk patients of being admitted and independently conduct interventions on them to efficiently allocate care resources.
3. Flag frequent flyer patients for supplement care coordination to reduce their frequent overuse of the ED.

The effect of implementing the policies above is significant. By implementing the above policies, the general waiting time is estimated to reduce by 27.31% in total, respectively by 12.34%, 7.85%, and 7.12% for each of the policies. We suggest the hospitals to use the analytical scheme of our policy recommendations to dynamically optimize the operation both inside and outside of ED.

**Problem 1: Post-Chemotherapy Hump**

**Description**

Figure 1 shows the number of delayed the patient, whose waits longer than the expected waiting time suggest by their triage score, in each month and weekdays. It shows that there is a hump during July and August following the Ramadan, which accounts for more than half of the delays in total. This delay could be explained by the decreased immunity among the general population during fasting and the increased influx of chemotherapy patients following the week-long holiday following Ramadan. The infection rate could be increased following the chemotherapy since the immune system was suppressed, thus more patients are coming to the clinic during July and August. This over-influx of patients lead to the increased waiting time in the Emergency Department. The extra waiting time for the patients not only delayed the urgent cases which should be treated promptly but also keep some patients away from waiting in our ED, decreasing the revenue for the hospital.

*Figure 1: The Distribution Of Delayed Patients In Each Month And Weekdays*

**Policy Options**
To deal with the peak of demand for ED visits in middle July, we propose to set up an independent clinical service to treat the large numbers of patients who get re-admitted to the hospital during this time window around two weeks after the chemotherapy. The patients who got chemotherapy during the last month will be directly sent to the rapid post-chemotherapy clinical service, rather than waiting for the ED with the normal procedure. The post-chemotherapy infectious patients could have their history taken and do all physical examinations, based on which physicians will decide patients need further treatment in the inpatient department.

This policy option could be generalized to broader situations to deal with different seasonal diseases. The core concept is when the hospital could detect a time pattern ED visits, such as pandemic flu season, setting up a special clinic service and centralizing patients with the same diagnosis is an effective option to reduce the waiting time and increase the efficiency in ED.

Policy Effect and Trade-offs

Setting up the rapid post-chemotherapy clinical service is beneficial in reducing waiting time by separating the patient with post-chemotherapy infections to a special clinic unit. A program set up in Taiwan during the pandemic flu season shows that this type of special program could reduce 17% of the total waiting time (Thomas Bodenheimer, 2013). By applying this parameter, we estimate our decrease in waiting time to be 17% in July and August, which leads a 12.34% decrease in waiting time for the whole year. While one thing we have to pay attention is that the special clinic unit occupies the other medical resource in the hospital which may have a negative impact on the hospital’s operation in other departments. Since the waiting time for the Emergency room is a more urgent and crucial problem compared with waiting time in other departments, this policy is still worth putting into effect.

Problem 2: Moving Beyond Triage Score

Description

To make the fullest use of limited care resources, natural strategies are either prioritizing patients with the most urgency or removing patients with the least urgency from the systems. This section targets on prioritizing high-risk patients and providing the services in a shorter time, the current way to identify the high-risk patients is only by triage score. However, the triage score, a rough estimate using five categories, could be not valid to represent the emergency level. A better indicator should be able to identify highly risky patients more accurately, which is a critical tool for hospital administrators to efficiently allocate the resource in ED.

We assume that people with more severe conditions are more likely to be admitted in the final decision based on the limited variables in the dataset. Inspired by the triage score which assesses the urgency of patients’ visits with 5 levels, we quantify the risk of a patient being admitted in the final decision as a probability score between the scale [0, 1]. The probability score would combine information from all available variables including triage score. With all available variables from the assessment stage, we build reliable logistic regression predictive models that achieve high prediction accuracy **AUC =0.81**,
which is significantly higher than the prediction with only the triage score (AUC=0.71). This implies that although triage score is already a good indicator, we can still make an improvement with more information and get a highly accurate probability estimate for the patients’ final decision (admitted or not).

Table 1: The prediction accuracy of patients to be admitted

<table>
<thead>
<tr>
<th>Features</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Only triage scores</td>
<td>0.71</td>
</tr>
<tr>
<td>All available features from the assessment stage</td>
<td>0.81</td>
</tr>
</tbody>
</table>

To ensure practical usage of the predictive model to maximize the use of limited resources, it is critical to identify the most important factors for high-risk patient prediction. We rank the feature importance by the absolute value of L2-regularized logistic regression model coefficients. Higher coefficient means a higher probability of being admitted with that variable occurrence. “Status on Treatment”, “status New Case”, “complaint with Fever” and “category Referred” are among the most important features that make patients more likely to be admitted. And “complaint” of “Tooth pain and Pain”, “category” of “4 and 5”, and “diagnosis ALL” are the most important features that make patients less likely to be admitted.

Policy Options

The fact that the most important feature for the above prediction model is “status On Treatment” implies that patients receiving cancer treatment are different from who not. The paper (Handley, Schuchter, & Bekelman, 2018) shows that by deploying a nonclinical navigators program on the patients with cancer by the Patient Care Connect Program, the ER visits are reduced by 20% in 30 days. Inspired by their work, we propose the following policies: (1). Create non-clinical patient navigators to enhance access and care continuity, such as patients’ self-monitoring of biomarkers and vital signs, and proactive use of STAR (symptom tracking and reporting) triggers patient report results, etc. (2). Stratify the high-risk patients of being admitted and independently conduct interventions on them, the scarce care resources can be allocated more efficiently. (Handley, Schuchter, & Bekelman, 2018)

Policy Effect and Trade-offs

The predictive analytics technique has the capability of accurately identifying high-risk patients, and identifying the most important high-risk factors allows administrators to focus on the most urgent high-risk subgroups to maximize; the whole framework is generalizable to other disease prediction even in large scale. The proposed policy to create non-clinical patient navigators will lead to more accurate tracking of patients’ disease conditions, hence doctors can provide more personalized treatment options for patients and improve treatment efficiency, which could reduce patients’ ER visits and save costs in long term. On the other hand, we also need to consider the cost of motivating patients to continuously
monitor their physical condition and the tracking expense of medical staff. For the policy of risk stratification with predictive analysis, despite the fact it requires data management and collaboration of the technical data analytics team, it recognizes high-risk patient patterns from large datasets and targets on highest-risk patients who are most likely to be admitted, which can take the most advantage of the medical care resource.

**Problem 3: Frequent Flyers**

**Description**

Frequent flyers are oftentimes a big problem for the management of the ED, but their visits are seen as avoidable. Frequent flyers have many downstream effects from excess wait time for other patients waiting in the ED, increased costs for the hospitals providing the care, and an increased workload for the clinical staff. Based on the data from Hope Hospital, different thresholds were explored in our model to demonstrate that targeting different degrees of overuse of the ED accounted for many delays (Table 2). A reduction in the behavior of frequent flyers can lead to dramatic reductions in the volume of patients in the ED at any given time and will reduce costs for the hospitals.

Table 2: Frequent Flyer Threshold and the corresponding percentage of all encounters, admissions, and delay causes.

<table>
<thead>
<tr>
<th>Frequent Flyer Threshold</th>
<th>% of All Encounters</th>
<th>% of All Admissions</th>
<th>% of Delays Causes</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>60.1</td>
<td>63.8</td>
<td>74.8</td>
</tr>
<tr>
<td>7</td>
<td>46.4</td>
<td>47.7</td>
<td>68.4</td>
</tr>
<tr>
<td>10</td>
<td>30.6</td>
<td>29.0</td>
<td>57.23</td>
</tr>
<tr>
<td>12</td>
<td>21.1</td>
<td>17.6</td>
<td>47.8</td>
</tr>
<tr>
<td>20</td>
<td>7.8</td>
<td>4.9</td>
<td>23.5</td>
</tr>
<tr>
<td>30</td>
<td>3.0</td>
<td>1.2</td>
<td>10.3</td>
</tr>
<tr>
<td>40</td>
<td>1.3</td>
<td>0.2</td>
<td>4.4</td>
</tr>
<tr>
<td>50</td>
<td>0.6</td>
<td>0.1</td>
<td>2.5</td>
</tr>
</tbody>
</table>

*N = 18,723 delays at all stages of the care continuum*

**Policy Options**

Various ED has implemented strategies to address the issue of frequent flyers. Reducing the number of frequent flyers decreased costs, decreased morbidity and mortality, and reduced potential risks for hospital-associated hazards (Podolsky et al., 2017). Seven sites with between 29,000 to 79,000 patients
piloted the use of Individualized Care Plans (ICP). Identified frequent flyers were reviewed by an interdisciplinary team of clinicians and a social worker or care coordinator twice per month. This pilot resulted in a 43% reduction in ED visits, a 52% reduction in inpatient hospital days, and a 43% reduction in direct and indirect costs.

Policy Effect and Trade-Offs

The implementation of ICP at Hope hospital can greatly reduce the overall number of patients in the ED at any given time as well as reduce costs for the hospitals. Given the proportion of delays for patients in the ED attributed to frequent flyers, reducing their use of the ED can have a compounding positive effect.

However, the additional coordinated care by the hospital may require some additional monetary investments as well as changes to the workflow. Care coordination for the ICP would require that staff are trained and are able to integrate this workflow into their existing workflow. The effort to implement ICP at Hope Hospital outweighs the massive and long-term savings in workload and costs. In addition, patients are given better care by being supported in their care and other patients having a shorter wait time in the ED.

Conclusions

Emergency departments (ED) are facing big challenges of dealing with patients visiting as their go-to spot for medical treatment. Based on the three problems identified from the data, we propose to increase the capacity of ED by setting up a special clinic service to deal with the jump of demand in a specific time window, create non-clinical navigators to enhance access and care continuity by better estimating patient’s risk, and provide multidisciplinary case management for the frequent flyer to reduce the unnecessary ED visits. By implementing the policy interventions, the ED waiting time will reduce by 27.33% in total. We suggest the hospital implement dynamic schemes in our policy recommendations and to interactively optimize their operation of ED. Although the current recommendations laid out are specific to Hope Hospital, the same strategic approach can be used by any hospital to improve its patient flow to reduce clinician workload.

Reference


Studer, Q., & Studer Group. (n.d.). The Hospital CEO’s Ultimate Dashboard: What to Check Daily, Quarterly and Yearly. So what are the big issues a CEO should focus on? In my experience, they fall into four. Retrieved from


**Appendix: Experimental Model of Predicting Patient’s Risk**

**Objective/Problem description:** In this section, we aim to identify the highly risky patients who will be most likely to be admitted in the final decision, which can be provided together with triage score to the hospital administrators for decision making (e.g., reducing waiting time). Inspired by the triage score which assesses the urgency of patients’ visits with 5 levels, we quantify the risk of a patient being admitted in the final decision as a probability score between the scale [0, 1].

**Dataset:** The data in the model include all features available before the examination, which includes the following 6 features ‘complaint patient arrived with’, ‘cancer treatment status’, ‘referlocation’, ‘recentdischarge’, ‘primary patient diagnosis’, ‘triage category’.

**Model & Accuracy Metric:** We use L2-regularized Logistic Regression (LR) is a linear classifier widely used in binary-classification problems, which has gained popularity in medical research since it can infer the probabilities of samples being classified into different categories. The data set is randomly divided into a training set (80%) and a test set (20%), respectively. The classification models are trained using the patients’ features and labels in the training set. Based on the patients’ features in the test set, the classification models provide corresponding predicted labels, which can be compared with the ground truth to calculate the prediction accuracy on the test set. We report accuracy metrics AUC, which is the areas under the receiver operating characteristic (ROC) curve. A random binary classifier can achieve an AUC of 0.5.

**Experiment Model:** Logistic Regression (LR) is a linear classifier widely used in binary-classification problems, which has gained popularity in medical research since it can infer the probabilities of samples being classified into different categories. In the classic logistic regression model, the posterior probability that a patient is predicted as pregnant can be expressed as a logistic function of a linear function of input features, with parameters “w” as weights of input features and “b” as an offset. In
order to increase the interpretability of the classifier, we add an extra penalty term proportional to $\|w\|_q$, which induce a sparse model that selects only a subset of informative features. Given the training set containing $n_{\text{train}}$ patient data and pregnant labels $y_{\text{train}}$, we can obtain the model parameters by maximum likelihood estimation. Specifically, we aim to solve the following regularized sparse logistic regression problem:

$$\min_{w, b} \sum_{i=1}^{n_{\text{train}}} (-\log p(y_i | x_i; w, b)) + \lambda \|w\|_q,$$

where $q=2$ represents L2-regularized logistic regression, respectively. The parameter $w$ represents the weight of the input features. This is a convex optimization problem that can be solved with a linear convergence rate by the incremental gradient method. In the test session, for a new test sample $x_{\text{test}}$, the probability that a patient is predicted as pregnant is:

$$p(y_{\text{test}} = 1|x_{\text{test}}; w, b) = \frac{1}{1 + e^{-w - b'x_{\text{test}}}}$$

**Experimental Result:** The probability score would combine information from all available variables including triage score. We build reliable predictive models that achieve high accuracy. With all available variables from the assessment stage, the prediction accuracy is **AUC=0.81**, which is much higher than the prediction with only the triage score (AUC=0.71). This implies that although triage score is already a good indicator, we can still improve the prediction with more information and get a highly accurate probability score of the final admission results.

<table>
<thead>
<tr>
<th>Features</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Only triage scores</td>
<td>0.71</td>
</tr>
<tr>
<td>All available features from the assessment stage</td>
<td>0.81</td>
</tr>
</tbody>
</table>

The interpretability of the results is critical to ensuring practical use. We rank the feature importance by the absolute value of L2-regularized logistic regression model coefficients (Table II and Figure I). By listing the coefficients corresponding to the most significant 10 features in the following table. Higher coefficient means a higher probability of being admitted with that variable occurrence. We can see that “Status on Treatment”, “status New Case”, “complaint with Fever” and “category Referred” are among the most important features that have positive effects for patients to be admitted. And “complaint” of “Tooth pain and Pain”, “category” of “4 and 5”, and “diagnosis ALL” are the most important features that have negative effects for patients to be admitted.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Features</th>
<th>Coefficient</th>
<th>abs(Coefficient)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>status_On Treatment</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>2</td>
<td>complaint_Tooth pain</td>
<td>-0.46</td>
<td>0.46</td>
</tr>
<tr>
<td></td>
<td>Feature</td>
<td>Value 1</td>
<td>Value 2</td>
</tr>
<tr>
<td>---</td>
<td>-------------------------------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>3</td>
<td>category_5</td>
<td>-0.43</td>
<td>0.43</td>
</tr>
<tr>
<td>4</td>
<td>complaint_Common Cold</td>
<td>-0.43</td>
<td>0.43</td>
</tr>
<tr>
<td>5</td>
<td>status_New Case</td>
<td><strong>0.34</strong></td>
<td>0.34</td>
</tr>
<tr>
<td>6</td>
<td>complaint_Pain</td>
<td>-0.31</td>
<td>0.31</td>
</tr>
<tr>
<td>7</td>
<td>complaint_Fever</td>
<td><strong>0.29</strong></td>
<td>0.29</td>
</tr>
<tr>
<td>8</td>
<td>category_4</td>
<td>-0.29</td>
<td>0.29</td>
</tr>
<tr>
<td>9</td>
<td>diagnosis_ALL</td>
<td>-0.27</td>
<td>0.27</td>
</tr>
<tr>
<td>10</td>
<td>category_R</td>
<td><strong>0.27</strong></td>
<td>0.27</td>
</tr>
</tbody>
</table>

Figure 1: The most important features on the admission prediction
POLICY BRIEF

Optimizing Emergency Department (ED) Utilization

TEAM
The Emergency Crew

TEAM MEMBERS
Majid Ahmadi
Vahid Khatami
Fatemeh Mehrabi
Fereshteh Razmi
Mohsen Salari
**Issue**

Healthcare costs are increasing at an alarming rate. U.S. healthcare spending grew 3.9% in 2017, reaching $3.5 trillion or $10,739 per person. Health spending amounted to 17.9% of the nation’s Gross Domestic Product (GDP). [1] Under the current law, national health spending is projected to grow at an average rate of 5.5 percent per year for 2018-27 and to reach nearly $6.0 trillion by 2027. Health spending is projected to grow 0.8% point faster than GDP per year over the 2018-27 period; as a result, the health share of GDP is expected to rise from 17.9 percent in 2017 to 19.4 percent by 2027. [2] One of the biggest sources of increase in healthcare costs are Emergency Department visits. It is also really important that EDs be efficient as healthcare costs rise.

The ED is often the first line of patient care for many hospitals and is often a catch-all for patients with serious illnesses. Given this important role, it is vital that EDs run efficiently and as smoothly as possible. Inefficient processes, the misallocation of resources, and "Frequent Flyers" all contribute to an unnecessary increase in care times and a reduced quality of care for patients in the ED.

The issue we sought to address predominantly centered around the development of strategies that hospitals can choose to avoid moral hazard without rescinding on their promise to offer “free care for all, for life”, offending the public who finance the hospital, or decreasing the quality of care for patients. If hospitals are able to properly perceive and predict the issues patients have when they come to the ED, they can formulate better policies for resource allocation, which in turn will bring down costs while increasing the quality of patient care.

**The ED Process**

When a patient enters the ED, they are first admitted to the triage area. They are given a triage score; the lower the category, the more severe the perceived symptoms/issues. Then, they are called in to the physician workspaces, where they are visited by a physician and undergo specific tests (as determined by the physician). Subsequently, the patient’s information is sent to the lab and/or the patient undergoes further tests (MRI, CT scan, etc.). Patients who can be treated are treated while those who are determined to need more care are prepared for admission to the hospital. Finally, the patient is either discharged (after being treated) or is admitted to the hospital for further care.

![Patient Flow Diagram](image)

**Figure 1. Overview of the patient intake and flow process**

**Data Analysis and Insights**

Looking at the data, triage scores seemed to be reasonable given that the majority of patients with lower categories (more severe issues) were ultimately admitted to the hospital, while those with higher categories (less severe issues) were overwhelming discharged at higher rates.
Another trend in our data becomes apparent when we map the number of visits across different times. As illustrated in the figure below, two different demand patterns exist during the weekdays and weekends. On initial thought, policy planners may be inclined to simply add more resources to alleviate patient demands. In fact, this exact action was taken by the Hope Hospital, where management increased the number of work spaces (and physicians) from 2 to 3 in hopes of reducing patient wait times. (Note: an additional room was used in an ad hoc manner, as a first response to times of increased demand.)

![Graphs showing demand trends and expected DLOS during weekdays and weekends.](image)

Upon analyzing the data, however, it became apparent that the increase in workspace capacities (workspace and doctor) had little effect on the patient experience. Specifically, a 33% increase to resource capacity resulted in only a 2% reduction in door to discharge times.

Notably, the expected DLOS remains fairly constant in the face of this changing demand. This is evidence that the most important problem does not lie in the fluctuation of the demand, and hence the responses should not be limiting to strictly addressing this problem. Thus, there are bottlenecks elsewhere in the process that should be identified. This can be done so through lean processing.

**Lean Methodology**

Through the use of lean methodology, hospitals can identify potential problem areas and bottlenecks with regards to inefficient systems and processes. Such implementations have been very successful throughout the United States, including in major cities like Houston, Atlanta, San Francisco, and Orlando. [3] For example, the Mercy Medical Center in Springfield, MA reported
a savings of $7.9 M in annual financial benefit through the implementation of lean process systems. [3]

Given our limited data, we were able to identify more pertinent sources of bottlenecking in this process.

![Average Time Spent per Category](image)

**Figure 4.** Average time spent at each stage per category

Consultation times were about the same across different categories (~9-11 minutes), while triage waiting times only slightly increased as the acuity level decreased. Triage increases, however, were significantly less than recommended benchmarks in Categories 3-5 (who were not at risk if wait times were longer). A particularly alarming trend appeared with regards to intervention wait times in which patients with lower category scores (and thus, more severe problems) were forced to wait longer for their results.

When mapping the amount of resources needed per category, a paradox immediately becomes apparent: patients in Category 1 are forced to wait significantly longer for fewer resources, despite needing them more urgently.

![Resource needs per category](image)

**Figure 5.** Resource needs per triage category

Moreover, when assessing patient needs, the issue of “Frequent Flyers” comes to light. Individuals who visit the hospital 4 times or more in a single year are commonly known as “Frequent Flyers,” or, in other words, individuals who highly utilize ED services. [4]. As apparent in our data, these individuals are only 7% of patients but make up 19% of all ED visits. Such patients tax important ED resources for issues that can be otherwise solved outside of the ED (and are overwhelmingly not admitted to the hospital at the end of their visit).

This is a particularly important and costly issue; under the Hospital Readmissions Reduction Program of the Affordable Care Act (§ 3025 of the Affordable Care Act), the federal government charges hospitals for 30-day readmission. Specifically, it will reduce payments to Inpatient Prospective Payment System hospitals for excess readmissions beginning Fiscal Year 2013. [5] Additionally, the 21st Century Cures Act allows for penalties based on a hospital’s performance relative to other hospitals with a similar proportion of patients who are dually
eligible for Medicare and full-benefit Medicaid beginning in FY 2019. Thus, if there are fewer
Frequent Flyers at other hospitals, the hospital in question can be penalized.

Additionally, as indicated above, the variation of the expected DLOS is too high during
the night shift, particularly with regards to patients in Categories 1-3. This problem is amplified
during the weekends.

Policy Recommendations

In light of the above-noted inefficiencies, we arrive at the following policy
recommendations:

Policy Recommendation #1: Address the inefficient allocation of resources, specifically (1)
Patients in Categories 1-3 are left waiting too long for admission/discharge and (2) Patients in
Categories 4 & 5 are being processed faster than the prescribed benchmarks through the stricter
prioritization of resources.

A better prioritization of resources for patients of Categories 1-3. Specifically, Category
1-3 patients should be prioritized with respect to diagnostic resources (lab and imaging) while
Category 4 & 5 patients are within the first hour of their recommended benchmarks. After the
first hour, patients are served on a first come, first served basis.

Outcome: Analyzing our data, we realized that on Category 1 patients use 2.4 times the
number of services (Lab and Imaging) Category 5 patients use. In addition, research shows that
the allocated lab times should be 57 minutes; yet, most patients spend 195 minutes waiting for
results [6]. Our analysis shows that simply prioritizing the required resources and services for
Category 1 patients over Category 5 can reduce the average intervention time by 24% while only
increasing the intervention time for Category 5 by 19%. As more resources are allocated to
patients in Categories 1-3, patients in Categories 4 & 5 will face wait times closer to the
appropriate benchmark. This will discourage patients from arbitrarily visiting the ED. While we
understand that some hospitals can only afford prioritization of resources for Category 1 patients
over Category 5 patients. The impacts of these two scenarios are shown below:

![Average Time Spent per Category](image)

Figure 7. Resource counts per triage category

Moreover, the hospital can publish the increased wait times on its website. Patients who
do not have true emergencies can consult this wait time and may be discouraged from coming to
the ED in light of increased wait times. As such, moral hazards will be reduced.

Additionally, patients in Categories 4 & 5 should be routed to a different geographic area
of the ED. Less resources are allocated to these patients, as is commensurate with their needs
relative to patients Categories 1-3. It is important to note that because these patients will be in a different area of the ED, they will be unaware of the processing times of Category 1-3 patients.

**Policy Recommendation #2:** Creation of a specific group of caretakers/network that can proactively address the problems of Frequent Flyers, thus reducing the number of people who come to the ED without pressing problems.

**Expected cost savings:** Optimally, total number of visits to the ED by Frequent Flyers will be so reduced by at least 9%, so that they are no longer deemed Frequent Flyers. This means that for this particular hospital, the number of visits for Category 4 and 5 would be reduced by 922 and 2005 visits, respectively. This scenario can also open up 416,000 hours of hospital’s time resources new available for Category 1-3 patients.

In addition, instead of visiting the ED, these patients will instead go to Urgent Care Clinics (UC), specialized help centers, and/or schedule a visit with their doctors (DV). Using the BCBS costs for medical costs of a hospital in the U.S., this policy translates into a total savings of $1.87M in ED visits for each hospital. [7]

<table>
<thead>
<tr>
<th>Category</th>
<th># of reduced visits</th>
<th>ED Savings</th>
<th>DV Savings</th>
<th>UC Savings</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>922</td>
<td>590,080</td>
<td>447,170</td>
<td>491,426</td>
</tr>
<tr>
<td>5</td>
<td>2005</td>
<td>1,283,200</td>
<td>972,425</td>
<td>1,068,665</td>
</tr>
<tr>
<td>Total</td>
<td>2927</td>
<td>1,873,280</td>
<td>1,419,595</td>
<td>1,560,091</td>
</tr>
</tbody>
</table>

Figure 7. Resource counts per triage category

Moreover, in the same vein of effort, hospitals can begin community-wide public health campaigns to educate potential patients as to the costs of unnecessary ED visits. This includes costs to the overall healthcare infrastructure as well as the fact that such ED visits take away resources from those in the community that actually need it.

**Policy Recommendation #3:** There are still issues to be solved, including a considerable number of patients coming into the hospital but leaving the ED before seeing the doctor. In our case, this number is 5% which requires a deep root cause analysis to avoid wasting resources. In light of our limited data, our recommendation is that the hospital undergo a pilot program in which it creates an interdepartmental task force dedicated to gathering more data. One reason for the variance in Category 1 expected DLOS times is the small sample size. The hospital should continue to gather data with regards to this specific group of patients and revisit this issue when it has the appropriate data. Moreover, the taskforce should identify specific steps in the ED processes that can be optimized according to the following metrics: a reduction in the: (1) Expected DLOS for Category 1 cases, (2) Standard deviation in expected DLOS for Category 4-5 cases, (3) Standard deviation in expected DLOS throughout the week (weekend vs weekday), (4) Mean time to be discharged from the ED, (5) Overall ED volume of patient visits, and (6) Patient satisfaction scores. Once the recommendations have been put into place, the hospital should gather the relevant data for a period of one year, performing an evaluation every six months.
References:


POLICY BRIEF

Improving ER Efficiency, Effectiveness, and Accessibility

TEAM
ER-Flash

TEAM MEMBERS
Victor Agbafe
Shweta Chopra
Rahul Chowdhury
Chun-Chieh Liang
Shiyin Wang
A Proposal to Improve ER Efficiency, Effectiveness, and Accessibility

ER-Flash: Victor Agbaje, Rahul Chowdhury, Shiyin Wang, and Shweta Chopra

1 Introduction

Emergency medicine involves the care and treatment of patients who present serious, life threatening injuries. These cases require speedy evaluation and treatment, and often non-elective admission, that is provided as a 24/7 service by hospitals. The demand for emergency care is rising worldwide due to an aging population in most countries, and rising survival rates in acute conditions\(^1\). Health systems that are unable to expand services or manager resources to meet rising demand, are plagued by delays\(^2\). Emergency medicine delays are of particular concern due to the urgency of cases presented. Additionally, the provision of emergency care is costlier than all other forms of care, with an overburdening of the department, impacting overall financial soundness of a hospital.

The problem presents in the following forms:

- Patient Queuing - Queuing is the phenomenon of long lines of patients boarding up outside emergency rooms(ER), waiting to be treated, due to the inability of the hospital staff to adequately meet patient demand. Overburdened, understaffed ERs result in long queues, that may lead patient deterioration or death\(^3\). The strategy employed by a hospital to manage these queues determines efficiency outcomes. The usual pattern of pooled queuing of patients may be altered by accounting for the urgency of a case.

- Moral Hazard - Even in countries that do not provide universal healthcare, ERs cannot refuse care on the basis of one's ability to pay or insurance status\(^1\). Further, unlike other forms of care, there is no requirement to schedule advance appointments. This makes an ER more directly accessible to patients, than other forms of care. These two phenomena make individuals more likely to pursue excessive ER care\(^5\).

- Triaging Errors - In simple terms, triaging is the process of sorting patients according to their need for emergency medical attention. Historically, triaging has been used to tackle the problem of excessive demand and is essential to both, the effective management of a hospital ER and the provision of clinical justice to patients\(^6\). Triaging standards vary slightly across countries, but the process usually involves the assessment of a short questionnaire and patient vitals to determine emergency need\(^7\). Patients are categorized into triage categories ranging from 1 - most urgent, to 5 - non-urgent\(^8\). This process, however, isn't always accurate as care is sometimes over or under provided due to triaging errors. Improving triaging accuracy is an essential step towards equipping hospitals to meet patient demand.

In order to restore the efficacy of emergency services across hospitals, the optimization of emergency department utilization and the reduction of patient flow into ERs must be addressed.

2 Data Findings

Data from the Emergency Management System (EMS) deployed at Hope Hospital, a tertiary-level cancer hospital located in the Middle East, provides details regarding 31,000 patient visits from April, 2018 to April, 2019. Insights from this data helps us identify focal points for emergency department optimization.

\(^{3}\)https://www.theguardian.com/society/2014/sep/18/ambulance-queue-death-nhs-cuts
\(^{5}\)https://www.newyorker.com/magazine/2015/05/11/overkill-atul-gawande
\(^{8}\)The Canadian Triage and Acuity Scale: Education Manual
The data reveals patterns in emergency care demand and service. 36% of patient visits are by patients categorized as non-urgent (category 5), while only 2% of visits are categorized 1.

![Patient Arrival Distribution](image1.png)

Figure 1: Patients Arrival Distribution

Time trends show that 9am to 12pm are peak hours, with maximum patient in flow, with numbers dipping in the evenings and at night. Patient numbers also drop during the weekend and peak right after. It is important to note that since this data comes from a cancer hospital, the nature of cases presented tend to be medical in nature, so these trends may not be easily generalizable to general hospitals where trauma care forms a larger proportion of care demand\(^a\). The important insight here is that emergency care demand varies over time.

Wait time, measured as the time elapsed between the patients arrival at the ER, and the patients examination by a doctor, is an important measure of department effectiveness. Average waiting times followed the same pattern as patient demand, across hours, and days of the week. Delays can be defined as cases when a patient experiences a wait time duration that is longer than the wait threshold defined for his or her triage category. By examining the data, we found that approximately 17.3% of patient visits resulted in delays in care provision. An important insight was that delay proportion was found to be the highest for Category 1 patients, at 73%, and fell as the patient category number increased, down to 5% for category 5.

This is worrying as patients with lower triage categories require treatment without delay. However, this pattern may indicate that the Triage thresholds for wait time that have been set may be too strict for lower priority patients (category 4 and 5). This also indicates the urgent need to optimize workload by employing alternative patient queuing strategies that incorporate patient category. Additionally, an exploration of high frequency visitors revealed that approximately 1200 patients who visited the ER more than once a month were categorized 4 or 5. There were a few cases of frequent flyers, defined by at least 9 visits a month, who make unnecessary and frequent visits to the ER for care which exposes a moral hazard problem. In summary, the data suggests that staffing interventions, reduction of low urgency patients, transformation of triage thresholds, and alternative queuing strategies, are all viable interventions to tackle ER problems.

3 Patient Flow Model

3.1 Busy Hours

The number of arriving patients variates during every day. We believe by discovering the pattern of customers can help the hospital better prepare. Therefore, I estimate the kernel density of the average arrival distribution every day.

We have also done the same visualization for each of the Triage Category, which appear basically in the same pattern. Therefore, we characterize: “Peak” 9:00-12:00, average 8.29 visitors; “Non-Peak” 8:00-9:00 and 12:00-15:00, average 6.05 visitors; “Medium” 6:00-8:00, average 3.21 visitors, “Extremely Low” 0:00-6:00 and 23:00-24:00, average 1.28 visitors.

![Delay Analysis for Triage Categories](image2.png)

Figure 3: Delay Analysis for Triage Categories

![Average Arrival Time](image3.png)

Figure 4: Average Arrival Time

3.2 Maximum Likelihood Time Estimation

The estimated percentage and average number of visitors are used in our simulation model.

The differences between the completedtime and scetime are used for conducting several lab tests and waiting the doctors to make a decision. Unlike we can estimate the time for testing time by directly calculating the difference between

**Table 1:“encounters”**

<table>
<thead>
<tr>
<th>encounter id</th>
<th>room number</th>
<th>encounter time</th>
<th>call time</th>
<th>seen time</th>
<th>completed time</th>
<th>lab</th>
</tr>
</thead>
<tbody>
<tr>
<td>406</td>
<td>1</td>
<td>4/17/18</td>
<td>4/17/18</td>
<td>4/17/18</td>
<td>4/17/18</td>
<td>CBC; Blood Chemistry</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8:36 AM</td>
<td>8:43 AM</td>
<td>8:57 AM</td>
<td>12:52 PM</td>
<td>Blood Chemistry</td>
</tr>
</tbody>
</table>

If we want to simulate the total time used in each step, we don’t have enough information to calculate the real-time. So we relax the problem setting to the following linear model:

\[
\text{completed}\_\text{time}(i) - \text{seen}\_\text{time}(i) = \sum_{t_{\text{lab}(i)}} t_{t} + r_{\text{room}(i)} + \epsilon_{\text{mean}}
\]  

(1)

For each encounter \(i\), the total time is the sum of the time used in each procedure. Here I made the assumption that the time used for each test have no correlations with each other. Then we can add them to represent the total testing time. The error is modelled as identical Gaussian error.

Under the assumption of uncorrelation, the variance of each test time can be added to get the total testing variance.

\[
\text{Var}(\text{completed}\_\text{time}(i) - \text{seen}\_\text{time}(i)) = \sum_{t_{\text{lab}(i)}} \sigma_{t} + \sigma_{\text{room}(i)} + \epsilon_{\text{std}}
\]  

(2)

Our problem has been reduced into linear equations. Where we can estimate the mean and variance of each different \(\text{lab, room number}\) pairs, the total number of which is 361.

\(X\) is a 0-1 matrix represent the patterns in rows. The columns correspond to tests, room number decision time, and inception. The number of functions is 361, while the number of parameters is 18. In this case, we can not use direct linear algebra techniques to solve it. Instead, we apply a maximization likelihood approach to optimize the parameters by gradient descent method.

\[
X_{t} = y_{\text{mean}}
\]

(3)

\[
X_{\sigma} = y_{\text{var}}
\]

(4)

Finally, we apply a gradient descent method using tensorflow package. Because the variance of the testing time can not be negative, here we assign a minimal variance as 0.1 for each test.

**Table 2: Results**

<table>
<thead>
<tr>
<th>Test</th>
<th>Mean</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Urine Analysis</td>
<td>0.62131184</td>
<td>0.3841998</td>
</tr>
<tr>
<td>Urine Culture</td>
<td>-0.52856034</td>
<td>1.3210369</td>
</tr>
<tr>
<td>Bleeding Profile</td>
<td>-1.5207993</td>
<td>0.82174927</td>
</tr>
<tr>
<td>CBC</td>
<td>-1.3526688</td>
<td>0.1</td>
</tr>
<tr>
<td>Blood Chemistry</td>
<td>-1.3425401</td>
<td>0.1</td>
</tr>
<tr>
<td>Room 1</td>
<td>1.3596053</td>
<td>1.8585489</td>
</tr>
<tr>
<td>Other Labs</td>
<td>-0.3192535</td>
<td>1.0111718</td>
</tr>
<tr>
<td><strong>...</strong></td>
<td><strong>...</strong></td>
<td><strong>...</strong></td>
</tr>
</tbody>
</table>

The results show interesting patterns\(^{10}\). CBC is usually a sign of admission, so the contribute of them is negative to the testing time. Bleeding Profile occurs in emergent accidents, which will have high priority accordingly, so this test can actually reduce the testing time. Some of the tests have fixed duration, but there are time variation(high variance) regarding to the tests like Bleeding Profile, Other Labs, Urine Analysis, Urine Culture. The efficiency of doctors in Room 1 is the worst among all the three rooms.

### 3.3 Simulation

Having parameter estimated, we can use AnyLogic to simulate the queueing situation in a average day setting.

After the simulation, we have come to the following procedure design.

\(^{10}\)Whole results refer to Appendix A.
policy can work as a behavioral nudge to prevent excess use of hospital resources due to the effect of loss aversion. This is because patients would avoid using the ER as a place to come and get their social needs addressed and would be more likely to try and seek a primary care physician if the patient knows that he or she may lose on some possible rebates if one isn't prudent with one's ER visits.

4. **Strategic Queuing and Triage Adjustments** - Our final policy proposal involves segmenting the ER based on the triage number that a patient is given. These patients are then assigned either to a specialist (category 1 and 2) or a generalist physician (categories 3,4 and 5) based on their category. Lower priority patients are referred to a specialist upon need. This queue model has been noted to bring about efficiency gains by inculcating a sense of ownership amongst physicians. From a physician standpoint those who have 1s and 2s would be seated closer to the operating area and those with 3s, 4s and 5s would be seated further and further away. In addition to this, the triage wait times for 4s and 5s would be doubled in order to put more priority on those who fall within triage levels 1, 2, and 3. This would lower delays for those who fall within triage levels 1, 2, and 3. In addition to this, nurses would be notified a minute before the triage wait time expires for a certain patient. This would help nurses remember the exact order of patients to prioritize for admittance within a certain triage. From a behavioral standpoint this would serve to address the cognitive limits that people have and would help the nurse get the high priority ER patients as quick as possible. It would also help the nurse remember the order of patients that were admitted but may reside in the same triage.

5. **Staffing interventions** - Staffing Interventions of the nature of staffing to demand, and strategic hiring, would be employed to better equip hospitals to address patient demand. By tracking patient demand over time, peak hours and days can be identified to alter shift timings, and number of staff employed at a particular time. The high proportion of low urgency cases in the ER also indicates that strategic hiring of less specialized staff, is well equipped to handle non-life threatening cases, may be done. Such staff is likely to be relatively inexpensive compared to specialized emergency trained staff, and could help ease the burden on the latter, allowing them to treat a large proportion of urgent cases.

5 **Conclusion**

These recommendations could go a long way towards addressing the problems of queuing, moral hazard and triage errors. Though they have been derived from the data in this case, these recommendations are generalizable to hospitals across geographies, that are facing similar challenges in the provision of emergency care. Though these improvements require some investment from the hospitals end, the potential for eventual cost savings is large due to the high cost of providing emergency care. One can look at the costs savings, about 2.3 million that BlueCross BlueShield of Illinois, Montana, New Mexico, Oklahoma and Texas achieved through an ACO model. I assume with the high costs of ER services, about 150–3,000 depending on the extent of the injuries sustained that one can expect even larger cost savings from these reforms. Thus, these steps would help reduce the burden on the hospital ER, improving the quality of care provided, reducing the probability of mishaps due to excessive delays, and providing overall cost savings to hospitals. Despite the expanded provision of healthcare through the Affordable Care Act, many still lack health coverage which makes it imperative that those who go to the ER get the best care possible.
Appendices

A Exploratory Data Analysis

A.1 Triage Category

Figure 8: Triage Category

```python
sns.countplot(encounters.category, order=['1', '2', '3', '4', '5', 'R'])
plt.xticks(rotation=0)
plt.show()
```

Figure 9: Higher Triage Category, Longer Seeing Time

```python
sns.boxplot(y="see_min", x="category", data=encounters, order=['1', '2', '3', '4', '5', 'R'])
plt.xlabel("category")
plt.ylabel("duration")
plt.ylim(0, 70)
plt.show()
```
A.2 Frequency Long-Tail Distribution

![Figure 10: Number of Visits](image)

```python
count = encounters["Rash"].value_counts()
sns.countplot(count)
plt.xticks(rotation=45)
plt.yscale('log')
plt.xscale('log')
plt.xlabel("Frequency")
plt.ylabel("Number of Visits")
plt.savefig("num_visits.pdf")
plt.show()
```

A.3 Complaint

The size of each word is processed in $\log(n)$ scale.²⁸

![Figure 11: Complaint Word Cloud](image)

```python
" ".join(list(encounters["complaint"].dropna()))
```

²⁸https://www.jasondavies.com/wordcloud/
A.4 Refer From

![Referred From Diagram]

```python
plt.figure(figsize=(7, 5))
sns.countplot(data=encounters, order=['Walk-in', 'Referred', 'Multi-specialty Clinics',
                                      'O2D Assessment', 'Oncology Clinic', 'Radiology',
                                      'Radiation Therapy', 'Tanta Hospital', 'Other'])
plt.xticks(rotation=18)
plt.xlabel('In')
plt.yscale('log')
plt.savefig('referredfrom.pdf')
plt.show()
```

B Discriminate Busy Hours

B.1 Method

The number of arriving patients varies during every day. We believe by discovering the pattern of customers can help the hospital better prepare. Therefore, I estimate the kernel density of the average arrival distribution every day.

![Patient Arrival Distribution]

I have also done the same visualization for each of the Triage Category, which appear basically in the same pattern.
Algorithm 1: Arrival Time Period Design

```python
sns.kdeplot(encounters['arrival_time'])
plt.xlim(right=24, left=0)
plt.xticks(np.arange(0, 25, 2))
plt.ylabel('Hours')
plt.xlabel('Density')

# peak
plt.axvspan(9, 12, facecolor=[0.95019603784313725, 0.101960378431372549, 0.103980392156862745], alpha=0.5)

# non-peak
plt.axvspan(0, 9, facecolor=[1.0, 0.4980392156862745, 0.0], alpha=0.3)
plt.axvspan(12, 15, facecolor=[1.0, 0.4980392156862745, 0.0], alpha=0.3)

# medium
plt.axvspan(15, 23, facecolor=[0.41568627450980394, 0.23921568627450981, 0.6339215686274509], alpha=0.3)
plt.axvspan(6, 8, facecolor=[0.41568627450980394, 0.23921568627450981, 0.6339215686274509], alpha=0.3)

# extremely low
plt.axvspan(0, 6, facecolor=[0.12156062745098039, 0.47058823529411764, 0.70588235294117645], alpha=0.3)
plt.axvspan(23, 24, facecolor=[0.12156062745098039, 0.47058823529411764, 0.70588235294117645], alpha=0.3)

plt.title('Arrival Time Density in a Day')
plt.savefig('arrival_time.png')
plt.show()
```

![Histograms of Patient Arrival Distribution](arrival_time.png)

Figure 14: Patient Arrival Distribution

```python
g = sns.FacetGrid(encounters, col='category', col_order=['1', '2', '3', '4', '5', '6'], hue='category', xlim=[0, 24], palette='inferno')
g = g.map(sns.kdeplot, "arrival_time", shade=True)
g.savefig('arrival_timeカテゴリー.png')
g.show(g)
```

B.2 Results

<table>
<thead>
<tr>
<th>Name</th>
<th>Period</th>
<th>Table 3: Results</th>
<th>Average Number of Visits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peak</td>
<td>9-12</td>
<td>0.27613536294</td>
<td>8.2908401549</td>
</tr>
<tr>
<td>Non-Peak</td>
<td>8-9, 12-15</td>
<td>0.26861869418</td>
<td>6.0488685552</td>
</tr>
<tr>
<td>Medium</td>
<td>6-8, 15-23</td>
<td>0.35583092213</td>
<td>3.20509915014</td>
</tr>
<tr>
<td>Extremely Low</td>
<td>0-6, 23-24</td>
<td>0.09941502076</td>
<td>1.27923917442</td>
</tr>
</tbody>
</table>

The estimated percentage and average number of visitors are used in our simulation model.
C Estimation of Lab Time

C.1 Problem Statement

The differences between the \textit{completen_time} and \textit{seen_time} are used for conducting several lab tests and waiting the doctors to make a decision. Unlike we can estimate the time for testing time by directly calculating the difference between \textit{seen_time} and \textit{completen_time}, this period of time is entangled by a set of several actions together. For example, for the encounter 406, the record just shows that it took 233 minutes on three parts: CBC, Blood Chemistry, and doctor decision making.

<table>
<thead>
<tr>
<th>encounter id</th>
<th>room number</th>
<th>encounter call time</th>
<th>seen time</th>
<th>completen time</th>
<th>lab</th>
</tr>
</thead>
</table>

C.2 Model

If we want to simulate the total time used in each step, we don’t have enough information to calculate the real-time. So I relax the problem setting to the following linear model:

\[
\text{completed_time}(i) - \text{seen_time}(i) = \sum_{t \in \text{test}(i)} t_t + r_{\text{room}(i)} + r_{\text{error}}
\]  

(5)

For each encounter \( i \), the total time is the sum of the time used in each procedure. Here I made the assumption that the time used for each test have no correlations with each other. Then we can add them to represent the total testing time. The error is modelled as identical Gaussian error.

Under the assumption of uncorrelation, the variance of each test time can be added to get the total testing variance.

\[
\text{Var}(\text{completed_time}(i) - \text{seen_time}(i)) = \sum_{t \in \text{test}(i)} \sigma_t + \delta_{\text{room}(i)} + \sigma_{\text{error}}
\]  

(6)

C.3 Solution

Our problem has been reduced into linear equations. Where we can estimate the mean and variance of each different lab, room number pairs, the total number of which is 361.

\( X \) is a 0-1 matrix represent the patterns in rows. The columns correspond to tests, room number decision time, and inception. The number of functions is 361, while the number of parameters is 18. In this case, we can not use direct linear algebra techniques to solve it. Instead, I apply a maximization likelihood approach to optimize the parameters by gradient descent method.

\[
X_t = y_{\text{mean}}
\]  

(7)

\[
X \sigma = y_{\text{var}}
\]  

(8)

Algorithm 2: Calculating \( X \), \( y_{\text{mean}} \), and \( y_{\text{var}} \)

```python
# The code snippet is not visible in the image.
```
Finally, I apply a gradient descent method using tensorflow package. Because the variance of the testing time can not be negative, here I assign a minimal variance as 0.1 for each test.

Algorithm 3: Optimization

def train(X_input, y_input, lr, num_epochs, relu=True):
    tf.reset_default_graph()

    # define input placeholders
    X_in = tf.placeholder("float", [n, 18], name="X_in")
    y_in = tf.placeholder("float", [n, 1], name="y_in")
    t = tf.Variable(tf.truncated_normal([18, 1]))
    if relu:
        t = tf.nn.relu(t)
    L = tf.maximum(t, tf.constant(0.1))
    e = y_in - tf.matmul(X_in, t)
    loss = tf.reduce_sum(tf.square(e), name="12_mean")

    # optimizer
    optimizer_op = tf.train.GradientDescentOptimizer(lr).minimize(loss)

    best_loss = 1e6
    best_t = None

    with tf.Session(config=tf.ConfigProto(log_device_placement=True)) as sess:
        init = tf.global_variables_initializer()
        sess.run(init)
        start_time = time.time()

        for step in range(num_epochs):
            sess.run(optimizer_op, feed_dict={X_in: X_input, y_in:y_input})
            a_cur = sess.run(c, feed_dict={X_in: X_input, y_in:y_input})
            loss_cur = sess.run(loss, feed_dict={X_in: X_input, y_in:y_input})
            t_cur = sess.run(t, feed_dict={X_in: X_input, y_in:y_input})

            if best_loss>loss_cur:
                best_loss = loss_cur
                best_t = t_cur
            if((step+1)%10==0):
                print("step {} / {}, loss {}, time {}").format(step+1, num_epochs, loss_cur, time.time()-start_time)

        return best_loss, best_t
C.4 Results

<table>
<thead>
<tr>
<th>Test</th>
<th>Mean</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Urine Analysis</td>
<td>0.62131184</td>
<td>0.3841998</td>
</tr>
<tr>
<td>Urine Culture</td>
<td>-0.52856034</td>
<td>1.3210369</td>
</tr>
<tr>
<td>VIRAL SWAB</td>
<td>0.38158008</td>
<td>0.1</td>
</tr>
<tr>
<td>ABG</td>
<td>-1.1003572</td>
<td>0.1</td>
</tr>
<tr>
<td>Blood Culture</td>
<td>1.160526</td>
<td>0.1</td>
</tr>
<tr>
<td>Other Labs</td>
<td>-0.3192535</td>
<td>1.0111718</td>
</tr>
<tr>
<td>PT</td>
<td>0.14623423</td>
<td>0.1</td>
</tr>
<tr>
<td>Stool Panel</td>
<td>-0.4725617</td>
<td>0.1</td>
</tr>
<tr>
<td>CRP</td>
<td>0.034993492</td>
<td>0.1</td>
</tr>
<tr>
<td>Bleeding Profile</td>
<td>-1.5207993</td>
<td>0.82174927</td>
</tr>
<tr>
<td>CBC</td>
<td>-1.3526688</td>
<td>0.1</td>
</tr>
<tr>
<td>Other</td>
<td>-1.1311041</td>
<td>0.1</td>
</tr>
<tr>
<td>PTT</td>
<td>-1.357972</td>
<td>0.1</td>
</tr>
<tr>
<td>Blood Chemistry</td>
<td>-1.3425401</td>
<td>0.1</td>
</tr>
<tr>
<td>Room 1</td>
<td>1.3596053</td>
<td>1.8585489</td>
</tr>
<tr>
<td>Room 2</td>
<td>1.0736116</td>
<td>0.1</td>
</tr>
<tr>
<td>Room 3</td>
<td>0.20024449</td>
<td>0.1</td>
</tr>
<tr>
<td>constant</td>
<td>0.7303601</td>
<td>0.39519298</td>
</tr>
</tbody>
</table>

The means and variances are used in our simulation model. I have also found some interpretation regarding the resulting testing durations. CBC is usually a sign of admission, so the contribute of them is negative to the testing time. Bleeding Profile occurs in emergent accidents, which will have high priority accordingly. The time of most of the tests are fixed, but there are uncertain(high variance) regarding to the tests like Bleeding Profile, Other Labs, Urine Analysis, Urine Culture. The efficiency of doctors in Room 1 is the worst among all the rooms.
Reducing Emergency Room Congestion: An Adaptive Priority Queueing Model

TEAM
The Impatients

TEAM MEMBERS
Abhishek Bhatia
Stephanie Lanius
Hillary Rodgers
Bradly Stone
Reducing Emergency Room Congestion: An Adaptive Priority Queueing Model

Policy Brief

Authors:
Abhishek Bhatia, Stephanie Lanius, Hillary Rodgers, Bradley Stone

MIT Hackathon
April 7, 2019

Emails:
abhi@mail.harvard.edu,
steph.phone1234@gmail.com,
hrodgers@brandeis.edu,
bradlytstone@brandeis.edu
# Table of Contents

*Executive Summary* .................................................................................................................. 2
  - Current Policy ..................................................................................................................... 2
  - Need for Change ................................................................................................................ 2
  - Policy Options .................................................................................................................. 2
  - Recommended Policy ......................................................................................................... 2

*Policy Proposal* ...................................................................................................................... 3
  - Introduction ....................................................................................................................... 3
  - Policy Summary ................................................................................................................ 3
  - Approach ........................................................................................................................... 3
  - Patients .............................................................................................................................. 5
  - Healthcare Providers and Facilities ................................................................................ 5
  - Payors ................................................................................................................................. 6
  - Policymakers .................................................................................................................... 6

*References* ............................................................................................................................ 7
Executive Summary

Current Policy

The Emergency Department (ED) at Hope Hospital, along with many EDs globally, face several issues with respect to finances, personnel, and medical resources, as well as patient satisfaction. These setbacks have been attributed to the exacerbated times patients spend waiting in ED intake rooms. The current policy attempts to ameliorate said issues through allocating more resources with respect to medical care professionals but has been unsuccessful in maintaining set goals of reducing wait times and honoring agreements made with stakeholders.

Need for Change

Increasing medical care professionals not only is cost invasive (cost to hire a physician can amount close to $250,000 [1]) but is unsustainable. This type of temporary remedy introduces a cyclic problem whereby more professional care may draw a larger patient population, which further necessitates the need for more staff.

Policy Options

Alternative solutions to this issue have been put into effect at an international level, through the employment of incentivizing programs [2, 3], statistical modeling platforms [4], and changing physical frameworks by placing patients in an operative role through an electronic intake process [5]. Though each proposed solution did show impact in minimizing wait times, a gap still remains in the feasibility of maintaining these reductions while simultaneously reducing costs.

Recommended Policy

This policy aims to address overutilization for non-urgent cases and the resultant bottlenecks in health service delivery that negatively impact patients, providers, and health financing. Our proposal applies queuing theory to establish a dynamic, threshold-sensitive patient flow management system in EDs. By incorporating a fast-track to care for non-urgent cases, with a built-in education and referral pathway to primary care for identified “frequent flyers”, we aim to improve patient satisfaction and health outcomes, health care professionals’ ability to provide a quality continuum of care, and health system financial sustainability.
Policy Proposal

Introduction. The current environment surrounding the optimization of Emergency Department (ED) resources and flow is a critical aspect of ensuring that healthcare systems are as productive in provision of quality services as they can be. EDs face a dilemma with respect to the aforementioned goals due to the nature of being the first point of contact with the hospital and health system for many patients. For this reason, it is imperative that EDs are prepared to handle the panoply of patient demands and volumes regularly.

On the frontline of such efforts is a queue management dashboard software (Emergency Management System; EMS), built for the ED at Hope Hospital (an oncology facility) in the Middle East, which provides detailed patient visitation data. Hope Hospital’s ED faces the ever-prominent healthcare issue: waiting room congestion. The readily accessible nature of healthcare at EDs, regardless of prior history with the health system, often results in patient overcrowding to seek care, long wait times, prolonged morbidity, decreased quality of patient-provider interaction, patients leaving the emergency department without being seen, and/or decreased physician productivity [6, 7] Taken together, these potential byproducts of the walk-in nature of ED facilities lead to patient dissatisfaction [7] and increased costs to related stakeholders (e.g. healthcare facility, professionals, and financial institutions;[8]). As of yet, the current policy in place to tackle this problem relies on the increase of resources through the hiring of health care professionals. This is an ephemeral solution and economically disastrous for the Hospital and associated funding agencies.

This proposal, designed by the co-founders at The Impatient’s (Massachusetts, USA), evaluates ED utilization, patient wait times, and volume to optimize overall patient wait-time during high volumes while conserving physical, financial and medical care professional resources. This solution employs priority queueing, an operable shift in resource allocation, alongside patient education and referral mechanisms to primary care for non-urgent patients, to not only reduce the bottleneck effect that occurs at varying times/days but also improve patient self-efficacy and provide an alternative continuum of care. The latter component to this policy aims to reduce unnecessarily overutilization of emergency services by promote conscious decision-making for patients to determine whether their symptoms are indicative of emergency needs or whether they should opt for a more resource-friendly choice such as a primary care physician visit—ultimately reducing the number of non-urgent patients and improving care coordination.

Policy Summary

- **Recommendation 1**: Patient Flow management through queue theory thresholding;
- **Recommendation 2**: Education and referral of non-urgent cases to primary care;
- **Recommendation 3**: Identification of ED “frequent fliers” to reduce bottleneck effect

Approach

Data. From the EMS database, a total of 33,968 encounters (8120 unique patients) were admitted to Hope Hospital between 2018-04-17 and 2019-04-02. This database was filtered based on outside expertise, resulting in the removal of all encounters which were not conclusive (‘admission’ or ‘discharge’ information missing): 31,763 encounters for 7939 unique patients remained. Triage scores are computed and assigned to each patient according to the Emergency Severity Index (ESI). These scores range from most (Category 1) to least severe (Category 5) in incremental numbers.
Descriptive Statistics. When comparing patient triage scores with respect to the daytime, respectively more patients with a triage score of 1 came during the night than did those with a triage score of 5 (compared to regular office hours from 9am to 5pm; Table 1). Furthermore, the admit rate within the patients with a triage score of 1 was much higher than the admit rates in any of the other score classifications. However, as the general prevalence of this group (triage score 1) is very low (2.1%), it is worth noting that patients in the other categories contribute greatly to the total numbers of admits (Table 1).

In a subsequent analysis (Appendix 1), we exposed an alarming trend which likely was a cause in the bottleneck experienced at Hope Hospital; a skewed distribution exists between the number of patients arriving at the hospital that had previously received services ('Frequent flyers'). For our analyses and in agreement with previous literature, Frequent Flyers are classified as any patient who returned to the hospital more than five times in a given year. Of the 7,939 unique patients, 1,818 were Frequent Flyers. However, it is necessary to note that the population at hand are checking into an oncology facility, which in and of itself poses discontinuity in reference to general hospital's: it is common that those suffering from cancer-related issues visit the ED more often than others with non-cancerous problems. We stratified these Frequent Flyers further based on commonality of intake complaint, which resulted in a bifurcation based on non-urgent (classified with a 4 or 5 triage score on each visit) and urgent criteria (Table 2). This further accentuated the impact that Frequent Flyers have on waiting room congestion, which prompted a more descriptive analysis aimed at targeting when a threshold (physicians present; N =3) was surpassed during a given day.

Table 2: Odds Ratios Across Triage Scores

<table>
<thead>
<tr>
<th></th>
<th>Non-frequent Encounters</th>
<th>Frequent Flyers urgent</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALL</td>
<td>1.27</td>
<td>0.52</td>
</tr>
<tr>
<td>Abdominal Pain</td>
<td>0.63</td>
<td>0.70</td>
</tr>
<tr>
<td>Blood Chemistry</td>
<td>4.55</td>
<td>3.24</td>
</tr>
<tr>
<td>Blood Culture</td>
<td>5.66</td>
<td>4.67</td>
</tr>
<tr>
<td>CBC</td>
<td>3.52</td>
<td>2.85</td>
</tr>
<tr>
<td>CRP</td>
<td>5.40</td>
<td>4.32</td>
</tr>
<tr>
<td>CT</td>
<td>2.82</td>
<td>2.41</td>
</tr>
<tr>
<td>Common Cold</td>
<td>0.09</td>
<td>0.20</td>
</tr>
<tr>
<td>Cough</td>
<td>0.25</td>
<td>0.44</td>
</tr>
<tr>
<td>Ended more than 6 months</td>
<td>0.09</td>
<td>0.36</td>
</tr>
<tr>
<td>Fever</td>
<td>2.02</td>
<td>2.85</td>
</tr>
<tr>
<td>NB</td>
<td>1.21</td>
<td>1.54</td>
</tr>
<tr>
<td>No</td>
<td>0.52</td>
<td>0.74</td>
</tr>
<tr>
<td>On Treatment</td>
<td>3.29</td>
<td>2.61</td>
</tr>
<tr>
<td>Pain</td>
<td>0.55</td>
<td>0.45</td>
</tr>
<tr>
<td>Walk-in</td>
<td>0.17</td>
<td>0.28</td>
</tr>
</tbody>
</table>

Figure 1: Queuing Rule of Thumb demonstrates that staffing need is above resource allocation.
wait time requirements, $s$, the number of service providers, needs to be greater than the product of the number of patients ($N$) and the time for service ($r$) divided by the maximum time to serve all patients in the queue ($T$; Eq. 1). Not surprisingly, our threshold calculation illustrates the well-known fact that the emergency department is understaffed, particularly during the peak hours of 9 am to 3 pm (Figure 3).

We propose that a priority queue follow the course of action detailed in Figure 2 which begins with the arrival of the patient at the ED. With exception to those arriving via ambulance or in immediate risk of loss of limb, or death (all of which will pass assortment and proceed immediately to an attending physician), patients continue into the ED, leading to registration and assortment. Based on the QT model, we established that regardless of month, a consistent bottleneck occurs between assortment (Figure 2c) and being called in to see the physician; time of day 9:00AM-12:00PM.

**Policy Implications for Stakeholders**

**Patients.** At the center of any equitable healthcare system lies the focal point, the patient. This proposal endorses a framework that emphasizes patient satisfaction through reduced wait time and facilitated patient-provider interaction. As consumers, patients may exercise purchase discretion regarding their choice of providers and payors. Regardless of the context with which a hospital operates (non-profit, for-profit, donation-based), it is imperative that the patient-provider relationship be revered given that patients can impact/vote policy in/out of a given system. To this end, a reduction in wait times has been shown to increase consumer satisfaction, which further emphasizes the circular process whereby providing quick, reliable care begets better patient-provider correspondences. Similar versions of interventions implemented in large urban hospitals to establish a fast-track for low triage score patients to reduce wait times and clear bottlenecks for non-urgent patients have also shown to have no effect on wait times and quality of care for critical patients in the ED. Notably, however, ascribing a low priority score/track to patients may negatively impact their psyche, and feelings regarding the healthcare they are receiving, and likeness to pursue future care at the given ED [6]. Our current proposal stresses to minimize such feelings through education-based treatment and referral mechanisms to comprehensive primary care -- placing access to quality medical care back in the hands of the patient.

**Healthcare Providers and Facilities.** Hospital EDs face regulatory pressure to reduce costs, improve efficiency, and increase throughput capacity to positively impact hospital revenue, and patient satisfaction with care. While there is a need for initial investment in surveying resources and operational changes, our proposal will benefit healthcare facilities in both the short and long term through three validated mechanisms that appeal to established incentive structures [13]:

**Treating “LWBS” patients.** Improved wait times during high volume periods capture patients who would have otherwise been classified as “Left without being seen (LWBS)”. Decreased LWBS rates captures additional revenue by treating both patients who would have otherwise left through both added discharges and added in-patients. An additional downstream effect of treating this patient population is the resultant improvement in perception from decreased wait time and improved access to quality care.

**Reduced patient “Boarding.”** The inability to move admitted patients to inpatient bed efficiently [14], typically occurs during flow backups, and results in increased length of stay (LOS), increased morbidity, and
significantly increased mortality. Mathematically, boarding occurs when bottlenecks form as a result of the hospital operating above 85-90% capacity, decreasing overall efficiency [15]. Our proposal, grounded in queuing theory, recommends a switch in patient flow management at this calculated occupancy threshold, with positive downstream effects for the health facility; decreased LOS, and improved patient health outcomes both decreases financial penalties accrued, and improves provider efficiency, and patient and community satisfaction.

**Continuity of care.** Reduced ED visits for non-urgent cases coupled with increased primary care referrals has proven to improve the uptake of preventative care, adherence to therapy, and reflect positively on patient and physician satisfaction [16]. Furthermore, creating a continuity of patient care through this channel has also been shown to improve the increasing burden of chronic disease outcomes across both developed and developing contexts.

**Payors.** Payors within a health system operationalize financial elements under established policy frameworks to procure health services for their enrolled beneficiaries. The payor-provider relationship rests on procuring efficient, high-quality, cost-effective services for enrolled patients while minimizing patient costs. In the context of Hope Hospital's mandate to offer services free-of-cost, payors are incentivized to reduce the cost to the health facility while maintaining a high quality of service provision. With regards to this, we identify two key payors:

**Accountable Care Organizations (ACOs)** have increasingly worked with hospitals to reduce needless spending and facilitate care coordination. ED care for low-acuity conditions result in high charges across the system in comparison to similar diagnoses seen in primary care and are a key point of intervention to reduce unnecessary health expenditure. Interventions like the one we propose to improve managed care through linkages to primary care and patient education have shown the greatest reduction in the magnitude of ED visits. In an effort by the PACT Program, a centerpiece of the Mount Sinai ACO program, EHR data was mined to flag “frequent flyers” in emergency care resulting in a reduction in emergency room visits by 51%, and reduced rates of 30-day readmission. Our proposal to optimize the ED for quality care for patients with urgent needs and improve the referral link to primary care for non-urgent cases, while taking similar data-driven approaches to address ED frequent flyers, appropriately complements existing ACO efforts globally.

**Health insurance** aims to safeguard beneficiaries against catastrophic health expenditures. Generalizing our proposal beyond Hope Hospital to non-donor funded health facilities introduces the influence of health coverage on patient health-seeking behavior, and the corresponding reimbursement pipeline to hospitals that provide services. As countries globally move towards health targets set by the Sustainable Development Goals, the aim to extend Universal Health Coverage will incentivize a shift from a fragmented private health insurance system, to a larger base population with public health insurance, which has shown to only increase ED utilization rates. As a response to this phenomenon in the United States, federal agencies levy penalties on hospitals for excess readmissions through value-based purchasing programs like the Hospital Readmissions Reduction Program (HRRP) established by Section 3025 of the Affordable Care Act [17]. In an effort to advance health equity through access to affordable healthcare, our proposal preempts pushback from insurance agencies subject regulatory checks, by facilitating a shift toward accessible primary care for non-critical, “frequent flyer” ED patients that would be the primary drivers of readmission with extended health insurance.

**Policymakers.** Policymakers set the context within which the health system operates by establishing frameworks to provide quality, affordable healthcare to their citizens. Within the scope of this proposal, the State is incentivized to maximize population health while staying within a country’s financial and resource constraints. The policy lever we use allows policymakers to support the improvement of patient-provider health service delivery by optimizing ED utilization while concurrently extending preventative public health measures, addressing physical and human resource constraints, and regulating costs with payors through a continuum of coordinated care along the health system.
References

### Appendix 1

<table>
<thead>
<tr>
<th></th>
<th>Non-frequent Encounters</th>
<th>Frequent Flyers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>non-urgent</td>
</tr>
<tr>
<td>Cough</td>
<td>15.1</td>
<td>16.2</td>
</tr>
<tr>
<td>Other Solid Tumor</td>
<td>8.8</td>
<td>10.1</td>
</tr>
<tr>
<td>NB</td>
<td>6.9</td>
<td>15.5</td>
</tr>
<tr>
<td>Palliative</td>
<td>2.0</td>
<td>2.3</td>
</tr>
<tr>
<td>Blood Chemistry</td>
<td>12.9</td>
<td>11.3</td>
</tr>
<tr>
<td>Common Cold</td>
<td>15.4</td>
<td>16.6</td>
</tr>
<tr>
<td>Fever</td>
<td>25.7</td>
<td>23.6</td>
</tr>
<tr>
<td>CRP</td>
<td>20.4</td>
<td>19.1</td>
</tr>
<tr>
<td>Blood Culture</td>
<td>21.1</td>
<td>19.6</td>
</tr>
<tr>
<td>CBC</td>
<td>28.9</td>
<td>28.0</td>
</tr>
<tr>
<td>On Treatment</td>
<td>45.1</td>
<td>49.1</td>
</tr>
<tr>
<td>Ended more than 6 months</td>
<td>41.2</td>
<td>39.6</td>
</tr>
<tr>
<td>ALL</td>
<td>29.3</td>
<td>12.7</td>
</tr>
</tbody>
</table>

We split all patients in two groups, Frequent Flyers (with six or more admits during one year) and non-frequent Flyers. Furthermore, to acknowledge that there are two kind of Frequent Flyers, we divided this group further in urgent (at least once triage score 1, 2 or 3) and non-urgent (patients that always have a triage score of 4 or 5). The latter group might have a low threshold to visit the emergency room.
POLICY BRIEF
Patient Allocation Policy Simulation

TEAM
Marshmallow

TEAM MEMBERS
Gabriel Chua
Elizabeth Lim
Jiongwei Lua
Summary
This policy brief summarizes our proposal on how to improve the operational efficiency of an emergency department by reducing waiting time. In particular, we propose two patient allocation policies - policies to determine which patient goes first when there is queue. Both policies have the commonality of always allowing the Triage 1 Patients to go first. Through our simulation analysis, we find potential opportunities to reduce wait times for Triage 1 Patients by 13% and variability by 46% on average.

Using a unique dataset of visits from one year, we analyzed utilization rates across different patient triage groups. In doing so, we identified potential efficiency gains for patients in the “Status 1” category (i.e. the most severe group).

While the data was for the hospital described in the case, we believe this policy brief contributes in two ways:

1. We demonstrate how a simple heuristic of always prioritizing the most severe group can benefit all patients in the system, regardless of severity.
2. We demonstrate the value of a simulation system in guiding policy. Our simulation system currently allows hospital administrators to test different allocation policies. However, it also has the potential to scale to also examine the impact of different triage assignment policies.

1. Introduction
The hospital in this challenge, along with all other hospitals around the world, wants to be more efficient by reducing the patient waiting time. This is therefore the policy objective guiding the entire analysis. For the purposes of this hackathon, the analysis of this policy brief will be contextualized in the case of the aforementioned specific hospital which is the source of the data.

2. Relationship between Triage Scores, Reducing Wait Times and Moral Hazard
Triage scores are one solution to reducing wait time/improving operation efficiency as they reflect the severity of the patient. One could describe them as playing the analogous role of prices in a market. With a particular patient’s score, hospital administrators can determine the maximum waiting time and nurses can allocate patients to the doctors. In fact, the two patient allocation policies we propose in this brief also use the triage scores. Therefore, as part of our analysis, we dedicated some time examining the validity of the assumption that triage scores adequately reflect severity.

As seen in the two graphs below, patients allocated a lower triage score (i.e. deemed more severe) did receive more tests (see Fig 1) and also were eventually more likely to be admitted to the hospital (see Fig 2). We also find that patients with Triage Score 1 were less predictable relative to those with Triage Score 5 (see Fig 3 & 4).

A related concept is the problem of moral hazard, especially in the case of the challenge’s hospital which provides free healthcare. We find evidence that lower waiting times is associated with more frequent visits, even after accounting for confounders like severity of the illness (see Fig 5). If triage scores do adequately reflect the risk and yet evidence of possible moral hazard does exist, the issue lies with how the triage scores actually affect the assignment policy. Our data analysis finds that waiting times are uniform across triage scores (and hence the relative cost for those who might “abuse” the system is low - i.e. instances of moral hazard). Our proposed patient allocation policies therefore also indirectly address the problem of moral hazard by ensuring those who really need these healthcare services get it first.
Figure 1: Mean Number of Tests Per Category

Figure 2: Distribution of Decisions by Priority Group

Figure 3: Scatter Plot of Median Waiting Times and Number of Visits, with fitted LOESS line

Figure 4: Triage 1 Case Heatmap distributed by time of day (y-axis), day of week (x-axis), and month (column)

Figure 5: Triage 5 Case Heatmap distributed by time of day (y-axis), day of week (x-axis), and month (column)
standards. We also emphasize that contributions to higher aggregate variability are driven mainly by the huge jump in the standard deviation for patients in Triage 5 from 36.4 minutes to 71.0 minutes. This may not necessarily be a bad thing, as it may reduce moral hazard by patients as patients with less severe conditions are less incentivised to come to the emergency room.

<table>
<thead>
<tr>
<th>Triage Category</th>
<th>Proportion of Cases</th>
<th>Current Policy Average Waiting Time (SD)</th>
<th>Policy Option 1 Average Waiting Time (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>100%</td>
<td>27.4 (28.1)</td>
<td>23.7 (54.3)</td>
</tr>
<tr>
<td>1</td>
<td>2.06 - 2.29%</td>
<td>13.8 (21.5)</td>
<td>2.6 (5.9)</td>
</tr>
<tr>
<td>2</td>
<td>17.5% - 18.2%</td>
<td>14.9 (14.0)</td>
<td>8.1 (29.4)</td>
</tr>
<tr>
<td>3</td>
<td>26.3% - 30.1%</td>
<td>19.2 (14.5)</td>
<td>15.3 (40.6)</td>
</tr>
<tr>
<td>4</td>
<td>15.3% - 15.9%</td>
<td>27.7 (21.1)</td>
<td>20.6 (47.8)</td>
</tr>
<tr>
<td>5</td>
<td>35.1 - 37.4%</td>
<td>43.4 (36.4)</td>
<td>41.4 (71.0)</td>
</tr>
</tbody>
</table>

4.2 Option 2: “Absolute Priority” + First Come First Serve

This patient allocation strategy is a modification of the one proposed above. Triage 1 Patients continue to be given Absolute Priority and can jump the queue upon arrival at the Hospital. However, instead of allocating Triage 2 - 5 Patients based on their Arrival Time plus a category-specific benchmark waiting time, we allocate patients based on a first come first served policy. At any point, the first person in the queue will be the patient has waited the longest amount of time.

Quantified Improvements:
Under this alternative policy, we find broadly similar overall results to Policy 1: reduced waiting times but greater variability.

Notably, under Option 2, Triage 2 and 3 patients are unambiguously worse off relative to the status quo - they experience both longer average waiting times and greater variability.
3. Limitations of Current Policy

From our analysis of the empirical distribution of wait times by patients across categories, we found that in the current patient allocation strategy, the hospital was consistently meeting the response time benchmarks for patients from Triage 2 - 5 but falls short for Triage 1 Patients.

In the current allocation strategy, patients are attended to according to their Arrival Time plus a benchmark waiting time which depends on their triage category. Patients deemed to have a more severe condition have shorter waiting times. However, in high-volume situations such as the one illustrated below in Table 2, Triage 1 patients in urgent need of assistant may be crowded out by other patients.

<table>
<thead>
<tr>
<th>Triage Category</th>
<th>Arrival Time</th>
<th>Target Response Time</th>
<th>Order</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.15pm</td>
<td>1.20pm (+05 minutes)</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>1.03pm</td>
<td>1.18pm (+15 minutes)</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>12.32pm</td>
<td>1.02pm (+30 minutes)</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>12.08pm</td>
<td>1.08pm (+60 minutes)</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>1.02pm</td>
<td>3.02pm (+123 minutes)</td>
<td>5</td>
</tr>
</tbody>
</table>

4. Proposed patient allocation policies

4.1 Option 1: “Absolute Priority” for Triage 1 Patients

Recognising that Triage 1 patients are in urgent need of medical care, we propose an absolute priority system where Triage 1 patients are placed first in the existing queue.

**Quantified Improvements:** Overall, across all categories, we find that with our Triage 1 Absolute Priority policy, we have a 13% improvement in the overall average waiting time, although there is a corresponding tradeoff of greater variability.

As expected, there is a significant improvement in both the average waiting time and variability in waiting time faced by Triage 1 patients when they are enabled to always jump the queue. This can be a matter of life and death in emergency situations.

Across patients in the Triage 2 - 5 categories, the average waiting time remains well under the benchmark.
<table>
<thead>
<tr>
<th>Triage Category</th>
<th>Proportion of Cases</th>
<th>Current Policy Average Waiting Time (SD)</th>
<th>Policy Option 2 Average Waiting Time (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>100%</td>
<td>27.4 (28.1)</td>
<td>24.3 (51.3)</td>
</tr>
<tr>
<td>1</td>
<td>2.98 - 2.20%</td>
<td>12.6 (21.5)</td>
<td>2.8 (5.5)</td>
</tr>
<tr>
<td>2</td>
<td>17.5% - 18.2%</td>
<td>14.9 (14.0)</td>
<td>17.1 (44.6)</td>
</tr>
<tr>
<td>3</td>
<td>20.3% - 30.1%</td>
<td>16.2 (14.5)</td>
<td>26.9 (54.1)</td>
</tr>
<tr>
<td>4</td>
<td>15.3% - 15.9%</td>
<td>27.7 (21.1)</td>
<td>24.3 (51.3)</td>
</tr>
<tr>
<td>5</td>
<td>35.1 - 37.4%</td>
<td>42.4 (36.4)</td>
<td>27.6 (52.8)</td>
</tr>
</tbody>
</table>

5. Conclusion
Through our work on this project, we have demonstrated that a combination of data science and simulation based approaches can identify opportunities to improve and optimize hospital operations management. While our simulation tool has explored two potential allocation policies, we believe that it has the potential to scale and examine the impact of different triage assignment policies on a wider range of operational metrics.
Clearing the Queue: Optimizing Emergency Services for Everyone

TEAM
QueueBusters

TEAM MEMBERS
Baker Flagg
Josh Mora
Backtosch Mustafa
Kevin Payumo
Vidur Sharma
Clearing the Queue – Optimizing Emergency Services for Everyone
Baker Flagg, Josh Mora, Backtosch Mustafa, Kevin Payumo, Vidur Sharma

EXECUTIVE SUMMARY

Unnecessary visits by low-acuity patients are a key contributor to difficulties faced by Emergency Departments (EDs) in providing timely care to urgent patients. The moral hazard that causes this phenomenon can be addressed by providing low-acuity patients with care alternatives that are superior (in terms of accessibility, convenience, and availability) to the ED while providing comparable levels of trust and expertise found at the ED. A method of achieving this is a web- or app-based check-in service that identifies low-acuity patients and “nudges” them to choose an alternative to the ED by presenting and facilitating access to other options (urgent care, tele-health, virtual care) that are manifestly superior to an ED visit (communicated by a comparison of effective wait-times).

BACKGROUND

Patients with Semi- and Non-Urgent conditions in Emergency Departments.

An emergency department (ED), also known as accident & emergency department (A&E), provides acute care for patients who attend hospital without prior appointment. The EDs of most hospitals customarily operate 24 hours a day, 7 days a week. Nevertheless, overcrowding in EDs is an increasing problem in countries around the world. According to a ProPublica analysis of data by the Centers for Medicare and Medicaid Services (CMS), ED wait times can reach nearly 50 minutes in the United States (https://projects.propublica.org/emergency/) and even higher in other countries (1-2).

Frequent crowding in hospital EDs has resulted in (2-6):

- Longer patient waiting times
- Decreased protection of patient privacy and confidentiality
- Impaired patient evaluation and treatment
- A higher probability of being seen which results in poorer outcomes for the patient
- Potential of poorer performance and adverse clinical outcomes including mortality and hospital lengths of stay (including a linear relationship between an overcrowded ED and deaths)

A major contributor to the supply-demand mismatch and resultant crowding are patients who come to the emergency room with semi-urgent or non-urgent conditions.

Moral Hazard in the Context of Emergency Departments

Improved accessibility of EDs and pricing raises the question of moral hazard as a potential driver of demand. Moral hazard describes the tendency to use more health services when people are covered by some form of insurance and therefore their out-of-pocket expenses at the time of service are lower or non-existent. Provision of non-urgent medical care in the ED is a suboptimal use of hospital EDs and results in several negative consequences, including crowding, higher cost of medical care, and reduction in its quality.

According to Houston emergency physician Tim Seay, “the ER’s a bad place to get your primary care” (7). Several common features of ED care may contribute to lower-quality care for nonurgent conditions.
Because patients and physicians in the ED are typically strangers to one another and treatment is provided on a one-time basis with little or no follow-up, there is essentially no continuity of care and no opportunity to develop an ongoing therapeutic relationship. Neither is there an opportunity to gather detailed information about the patient’s values and goals to monitor chronic medical conditions over time, or to adjust treatments accordingly.

The main reasons for non-urgent care patients using the ED were difficulty to get an appointment with a general practitioner, the availability of medical services in the ED (like imaging, laboratory tests), as well as the knowledge of the location of a hospital (8-9).

**PROBLEM**

*Queuing Issues*

In the ED we analyzed, a significant percentage of patients in urgent or serious condition that had to wait longer than the maximum goal time established by the hospital based on their condition to be seen by a physician (69.6% of Category 1, 45.8% of Category 2, and 24.4% of Category 3). Slightly more than half of the encounters at this ED (50.7%) involved low-acuity patients belonging to Categories 4 and 5 who could have received treatment elsewhere or by other means.

We hypothesized that reducing avoidable visits by these low-acuity patients could address the problem of excessive wait times for more urgent patients.

*Moral Hazard*

We identified moral hazard as the main contributing factor to this high number of avoidable visits by low-acuity patients. Therefore, the key problem for the hospital is how to dissuade low-acuity patients from seeking care at the ED out of convenience without reneging on their promise to offer "free care for all, for life," offending the public, or decreasing the quality of care.

**METHODOLOGY**

For our analysis, we created a variable, "crowd factor," representing the number of low-acuity patients in the ED at the time an encounter began for a more urgent patient. We determined crowd factor by counting, for each encounter ("encounter x"), the number of other encounters for which the following conditions were true: encountertime < encountertime_x, completedtime > encountertime_x, and category = 4 or 5. We then used logarithmic regression to determine the relationship of this variable to the resulting wait time for the higher acuity patient (category = 1, 2, or 3). This enabled us to determine the "crowd factor" at which a higher acuity patient’s wait time would exceed the maximum prescribed by their category.

**RESULTS**

We found crowd factor to be positively correlated with wait time for both Category 2 and 3 patients, indicating that we can effectively reduce wait times by reducing the number of low-acuity patients that seek treatment at the ED (See appendix for graphs of crowd factor vs. wait time for Categories 1, 2, and 3). The trendline for Category 1 suggested crowd factor was negatively correlated with wait time, but we are not confident in the trend indicated here for two reasons. First, we believe this is because the max
wait time for Category 1 patients (5 minutes) is so short that operational inefficiencies (slow patient processing time, etc) account for most of the delays for this category. Second, we believe Category 1 patients are potentially being miscategorized by clinical staff because a high percentage, 57%, of Category 1 patients are not admitted into the hospital post-ED visit. Given the severity of these patients, one would expect a high percentage of Category 1 patients to be admitted post-ED visit.

For Categories 2 and 3 we found the crowd factor at which wait times would exceed the maximum allowable time to be 2.14 and 9.8 patients respectively. The average crowd factors for those categories are 1.57 and 1.71, but high crowd factor standard deviations (2.2 and 2.3 respectively) attest to significant volatility. This indicates that, in addition to reducing the total number of visits, a more even distribution of low-acuity patient visits to the ER would reduce this volatility and therefore reduce the incidents of urgent patients exceeding their maximum wait times.

We also determined that low acuity patients make up 30% of laboratory tests in the ED. While only 0.26% of all low-acuity patients are actually found to have a condition that requires admission to inpatient care, physicians still carry out further tests (e.g. laboratory tests) on almost a third of this patient population in order to eliminate the extremely negative repercussions of misdiagnosing one of these individuals. This not only costs more money, but also takes time, as the people remain in the ER until you get the laboratory results, which then have to be evaluated and discussed with the patient. This increases the crowd factor by increasing the amount of time that low-acuity patients spend in the ED.

**RECOMMENDED SOLUTIONS**

*Health Systems Should Offer Alternatives to the Emergency Department that are as Convenient, Accessible, Transparent, and Responsive to Non-Emergent, Low-Acute Needs*

Patients of low-acuity are drawn to emergency departments because of their convenience, accessibility, and the responsiveness of providers. However, the emergency department is also the most expensive site of care. There is a significant opportunity to shift the treatment of low-acuity patients to non-emergency department acute care venues such as urgent care or telemedicine (10). Given the choice of alternatives, patients with low-acuity needs have preferred non-emergency department acute care venues over a hospital-based emergency department in recent years.

Health systems, particularly those with significant brand value and loyalty from patients, should offer branded alternatives to a hospital-based emergency department for patients with low-acuity needs.

- **Telehealth Visits:** In accordance with local laws and regulations, health systems can offer telehealth visits that use video, audio, and/or text to consult with patients on their low-acuity needs instead of requiring their visit to the emergency department. This is convenient for patients and allows hospitals to reserve their emergency department resources for high-acuity patients with emergent needs.

- **Virtual Care Center:** Health systems can establish virtual care centers which are facilities staffed by advanced practice clinicians (APCs). Depending on the jurisdiction, APCs are able to evaluate low-acuity patients, establish a treatment plan, and write a prescription under the supervision of a physician virtually.

- **Urgent Care Centers:** Health systems can also establish a physical non-emergency department acute care venue. Depending on the services offered, it may be staffed by physicians and APCs
Aloke. Urgent care centers are typically located in common and convenient community locations such as near grocery stores, residential neighborhoods, and strip malls.

To encourage low-acuity patients with non-emergent conditions to utilize these services over a hospital-based emergency department, health systems can offer a web- or app-based check-in service for their hospital-based emergency departments which could offer a wait-time based on inputs offered by the patient and the current wait-time at the ED. The patient would also be shown the wait time to reach an APC by text, phone, and/or video alongside the hospital-based ED’s wait time. The ability to seek an equivalent level of care in a shorter amount of time and in a more convenient setting would likely “nudge” the patient to choose a non-emergency department acute care venue over a hospital’s emergency department. If the patient still elects to be seen at the hospital-based ED, the application can recommend a time-window for the visit that will minimize the patient’s wait time. This further “nudges” the patient to arrive at a time when volume is predicted to be low enough that a low-acuity patient will not negatively impact the wait-times of more urgent patients.

A Dashboard Should Provide Hospital Administrators with Meaningful Information that Assists with Daily Operations and Future Planning

Hospital administrators should use performance and clinical data in their decision-making processes, particularly for managing daily operations and predicting and planning for future demand. Emergency department managers should receive data from the aforementioned patient-facing platform, particularly the number of patients who have checked in advance, the predicted wait time given to patients, and the time-windows for their visits. The dashboard should also include wait time, volume by triage determination, and a visualization of room utilization which includes who is in the room, staffing assignments, and elapsed time in the room. After a patient encounter has been concluded, reporting resources used in that encounter (nurse/physician time, equipment usage, and supplies consumed) can aid in resource management.

The dashboard should also provide predictive data such as expected volume for selected future time periods based on historical data and previously observed trends. The tool should also provide alerts in advance of predicted significant influxes (i.e. in advance of flu season, major holidays, or other known dates requiring different resources than normal). The emergency department manager can use this information to plan resources -- supplies, staffing, space, for example -- for these periods.

With limited customization, this tool could be deployed across the entire health system.

CONCLUSION

While we have shown that reducing the number of unnecessary ED visits and also distributing those visits more equally across time can reduce instances of high-acuity patients waiting too long to receive care, additional testing and research can be conducted to determine the ultimate effectiveness of these measures. It is possible that increasing ease and access to care through the app could induce demand and ultimately place a greater strain on the hospital or healthcare system as a whole while reducing strain on the ED. Additionally, our discovery of the high percentage of potentially mis-categorized high-acuity patients indicates the triaging process needs revision and further assessment. We believe these topics constitute the next steps required in our research.

REFERENCES
1.) Groeger LV, Tигas M, Wei S. ER Wait Watcher. ProPublica. 2019
2.) OECD Waiting Times Study. Organisation for Economic Co-operation and Development. 2013
4.) Hospital-Based Emergency Care: At the Breaking Point. Washington, DC: National Academies Press; 2006
5.) The Effect of Emergency Department Crowding on Clinically Oriented Outcomes, Steven L. Bernstein MD et. al.
6.) Adam J. Singer MD Henry C. Thode Jr PhD Peter Viccellio MD Jesse M. Pines MD, MBA, MSCE The association between hospital overcrowding and mortality among patients admitted via Western Australian emergency departments.
7.) Greene J. Sending low-acuity patients away from the ED: closing the door or stemming the tide?. Ann Emerg Med. 2007;49(3):317-319
8.) Nonurgent patients in the emergency department? A French formula to prevent misuse, Stéphanie Gentile et.al
APPENDIX

Effect of Crowd Factor on Category 2 Wait Times

Effect of Crowd Factor on Category 3 Wait Times
CHALLENGE
Urban Planning

IN COLLABORATION WITH
City of Boston Innovation and Technology

TEAMS
Parsonites
Regulating Super-Duper Hosts

The Rent Is Too Damn High Partnership
Transcendental Short-Term Rentals

Rutgers
Tax-Based Solution to Commercial Airbnbs

Urban Development Team One
Restoring Communities: How to Use Short Term Rental Regulation to Fight Displacement
CHALLENGE STATEMENT

Urban Planning

Analyzing the Economic Impact of Short-Term Rentals on Local Communities

BACKGROUND

Short term rentals (STRs) have exploded in recent years primarily thanks to online platforms such as Airbnb and HomeAway. STRs have seen an increase in popularity as tourists prefer a more “authentic, local experience” over traditional hotels, while property owners enjoy the flexibility and extra income that an online platform provides. With landlords finding greater profitability keeping long-term rentals in the short-term market and out-of-state investors running entire apartment buildings as hotels, metropolitan areas across the country are grappling with the potential economic impacts on existing communities and neighborhoods. Cities are forced to weigh the potential loss of revenue from tourism, particularly in previously under-appreciated neighborhoods, with protecting its long-term residents against rising housing costs.

Many cities have taken action against Airbnb to mixed effect. While there is some early understanding of the impact on the housing market (but no clear consensus), there is even less known about STR’s impact on the local economy.

THE CHALLENGE

Investigate Airbnb’s economic impact on Boston’s local communities. The city is additionally interested in analyzing attempts by other cities to address STRs, specifically regarding their success in promoting affordable housing and economic vitality of local communities.

Questions

- Can we assess the number of STR units that should go back to the long-term rental registration market in order to address the potential affordable housing crisis?
- What, if any, economic impact do STRs in your neighborhood have on local businesses?
- Are there other dimensions of impact that should be considered for each neighborhood - e.g. spirit, culture, percent of population who moves further away from their original neighborhoods? What other indicators should the city be collecting to measure this impact?
POLICY BRIEF
Regulating Super-Duper Hosts

TEAM
Parsonites

TEAM MEMBERS
Cherry Gao
Craig McLean
Regulating Super Duper Hosts: A Data-Driven Approach to Identify and Regulate Investor Units and Redistribute Demand on Airbnb in Boston, MA

Craig McLean, Cherry Gao
PhD Students, Massachusetts Institute of Technology, Boston, MA

Introduction

Short-term rentals (STR) have become a major influencer of housing markets in many major cities in the United States, especially with the rise in popularity of Airbnb. Since the inception of Airbnb in 2009, home sharing has become an attractive alternative to traditional hotels for travelers, and have financially empowered hosts, some of whom now rely on Airbnb as a major source of income.

Recently, Airbnb has come under fire for their deleterious effects on rent increases by because of the demand for STRs. Incentivized by the possibility of making the equivalent of one year’s rent in ~80 days of STRs, landlords have begun to remove housing from the rental market, or raise rental prices. In the face of rising demand for housing, the City of Boston faces a challenge in balancing demands for STRs with that of residents’ need for affordable housing.

It has been found that a small number of individuals, whom we call “Super Duper Hosts”, purchase numerous units in Boston, sometimes over a hundred, and offer them for STRs on Airbnb. Known as “investor units”, these apartments are purchased for the sole purpose of renting on Airbnb and cause the most damage to the Boston housing market by removing rentals from the housing market and displacing potential residents. In order to curb the damage of “investor units” on the Boston housing market, the City of Boston passed the Short-Term Rental Ordinance, creating a set of criteria for short-term rental eligibility, in January 2019. Although designed to alleviate problems related to “investor units”, enforcement of these new regulations have been and will be, a challenge.

Here, we present a detailed analysis of the “Super Duper Hosts” who profit from purchasing and renting out “investor units” on Airbnb in Boston. First, we discuss the magnitude of the influence of “Super Duper Hosts” on each neighborhood in Boston. Second, we identify the characteristics of neighborhoods that are most correlated with the enrichment of these “investor units” in certain neighborhoods in Boston. Based on our data analysis results, we propose a novel regulation strategy that dampens the demand for STRs in neighborhoods that are enriched in “Super Duper Hosts”, and increases the financial attractiveness of STRs in non-traditional neighborhoods. This novel strategy is designed to add an economic layer to the Short-Term Rental Ordinance that would increase the effectiveness of this new regulation aimed at minimizing the impact of STRs on Boston’s housing market.

Methods

Data

From a complete list of Airbnb listings scraped from February 9, 2019, we defined active listings for each year as those that had a guest review in that year. Of these active hosts, we defined “Super Duper Hosts” as having 5 or more active units listed on Airbnb during 2018. Identification of features (types building usage) that are major differentiators of neighborhoods was performed with building energy and water data reported to BERDO in 2015-2018. Historical census data provided estimates of the percentage of residents rented or owned housing within each neighborhood.
Model
Here, we asked the question, can certain features of neighborhoods (types of building usage) predict areas of enrichment of “investor units”? First, we performed a principal component analysis (PCA) on count data on different types of municipal buildings in each neighborhood. Using k-means clustering with 4 clusters (identified through gap statistic), we identified predictor variables to design the model (Fig. 2). A representative feature from each cluster (Office, Hotel, College/University, and Multi-family Housing) was used in our multi-linear regression predictive model. The variables were tested using an ANOVA to determine which contributed most significantly to the prediction of the model. The model was validated using 10-fold cross-validation.

Results
“Super Duper Hosts” have a skewed influence on the Boston housing and STR market
We define “Super Duper Hosts” as hosts who have 5 or more active listings on Airbnb in Boston in 2018. Despite making up only 3.85% of the total number of Airbnb hosts in Boston, “Super Duper Hosts” rent up to 80% of properties in some neighborhoods (Fig. 1). The neighborhoods in which “Super Duper Hosts” have a disproportionately large share of Airbnb properties, many of which are for “Entire Homes”, are also those that have the highest rental prices in Boston (West End, Chinatown, Downtown, Fenway, Back Bay, North End)⁴, suggesting that the large presence of “Super Duper Hosts” is correlated with neighborhoods whose housing is financially inaccessible to many Boston residents (Table 1).

Figure 1: Number and ratio of housing rented by “Super Duper Hosts” on Airbnb

<table>
<thead>
<tr>
<th>Number of properties owned by “Super Duper Hosts”</th>
<th>Ratio of total housing rented by “Super Duper Hosts” on Airbnb</th>
</tr>
</thead>
<tbody>
<tr>
<td>West End</td>
<td>0.80</td>
</tr>
<tr>
<td>Chinatown</td>
<td>0.75</td>
</tr>
<tr>
<td>Downtown</td>
<td>0.60</td>
</tr>
<tr>
<td>Roxbury</td>
<td>0.50</td>
</tr>
<tr>
<td>South Boston Waterfront</td>
<td>0.40</td>
</tr>
<tr>
<td>Jamaica Plain</td>
<td>0.30</td>
</tr>
<tr>
<td>Mattapan</td>
<td>0.20</td>
</tr>
<tr>
<td>Mattapan</td>
<td>0.10</td>
</tr>
</tbody>
</table>

Table 1: Impact of “Super Duper Hosts”

| % of Super Duper Hosts       | 3.85% |
| % of hosts with more than 50 properties | 0.5% |
| % of total properties (n = 3798) listed by Super Duper Hosts | 49% |
| Mean monthly rent in top 5 neighborhoods enriched in Super Duper Hosts | $2700 (±$400) |
| Mean monthly rent in bottom 5 neighborhoods enriched in Super Duper Hosts | $1987 (±$344) |
Figure 2: A Principal Component Analysis of features (types building usage) that are major differentiators of neighborhoods

Table 2: Identification of features (types building usage) that are major differentiators of neighborhoods

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Error</th>
<th>T-value</th>
<th>Coef P-value</th>
<th>Model Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Office</td>
<td>1.68</td>
<td>0.78</td>
<td>2.16</td>
<td>0.0465</td>
<td>P &lt; 0.01</td>
</tr>
<tr>
<td>Hotel</td>
<td>8.66</td>
<td>3.88</td>
<td>2.23</td>
<td>0.0406</td>
<td>$R^2 = 0.489$</td>
</tr>
</tbody>
</table>

The magnitude of coefficients denotes how important each feature (type of building) is in identifying neighborhoods enriched with Super Duper hosts. The variables making up the model were tested using an ANOVA (Office: $P < 0.001$, Hotel: $P < 0.05$), which showed that these variables contributed significantly to the prediction. Ten-Fold cross-validation demonstrates that the model had a mean absolute percent error of 13.4.

Figure 3: Certain neighborhoods have a high percentage of home-owners who live in the homes that they own.
Policy Recommendations

Redistribute STR demand from hotspots to nontraditional neighborhoods through a differential taxation strategy.

We showed that certain factors, such as number of offices and colleges/universities, highly predict neighborhoods in Boston that enrich in “investor units”. As the City of Boston works to regulate the STR industry, our predictive model can be used to continue to monitor these areas in the future for new “investor unit” hotspots.

The neighborhoods we identified as hotspots for “investor units”, such as Downtown, Fenway, Back Bay, West End, and Chinatown, are those that the City of Boston should focus heavily on to regulate due to their high potential in displacing residents. These neighborhoods are also those that have the highest rental prices in Boston (West End, Chinatown, Downtown, Fenway, Back Bay, North End)\(^5\), suggesting that the large presence of “Super Duper Hosts” may be correlated with neighborhoods whose housing is financially inaccessible to many Boston residents. This was not the case with neighborhoods depleted of super duper hosts.

However, recognizing that facilitating home sharing through Airbnb can have a positive economic effect (by facilitating financial independence of hosts and increasing visits to local businesses through tourism), we recommend that STRs located in nontraditional neighborhoods (especially low-income neighborhoods such as Mattapan, East Boston, and Roxbury) be less financially penalized through state hotel tax. These neighborhoods often lose STR bookings due to their lack of connection with public transportation. These neighborhoods also have high percentages of home-owners who actually live in the homes that they own, which is the demographic of hosts that would have the least effect on the Boston’s housing market according to City of Boston’s Short-Term Rental Ordinance (Figure 3). Indeed, the rate of increase in Airbnb in these areas has been slowest over time relative to those enriched in Super Duper Hosts (Figure 4). By charging a
smaller tax on STRs in these neighborhoods and decreasing the overall cost of staying in these neighborhoods significantly, we could financially incentivize the redistribution of STR demand from the center of the city to these nontraditional neighborhoods. Neighborhoods with similar characteristics as those predicted by the modeling would instead be taxed more heavily. With this differential taxation strategy, we maximize the economic benefits from Airbnb in nontraditional neighborhoods, while minimizing the effect on the housing market in neighborhoods with Airbnb hotspots. Ultimately, the aim of this policy is designed on consolidating profits from Airbnb in the hands of local citizens rather than individuals living abroad, ensuring a higher percentage of the earnings returns to the local economy.

References
POLICY BRIEF
Transcendental Short-Term Rentals

TEAM
The Rent Is Too Damn High Partnership

TEAM MEMBERS
Kevin Fei
Kirby Ledvina
Eli Rider
Celina Scott-Buechler
Transcendental Short-term Rentals
Celina Scott-Buechler, Kevin Fei, Kirby Ledvina, and Eli Rider

Boston’s Housing Challenge

Boston’s rising housing costs have had a disproportionate impact on renters. While single-family home and condo prices risen since 2009 by 32% and 44%, respectively, renters saw the largest price increase of nearly 55\%\(^1\). These renters include college students and young professionals, as well as low and middle income families. While rents increase and wages have generally remained stagnant\(^2\), online booking platforms such as AirBnB and HomeAway have created an opportunity for short-term rentals to proliferate across the city and impact the equity and accessiblity of longer-term housing. Multiple studies have found that the presence of short-term rental activity can exacerbate housing issues as evidenced by increased rents and decreased rental availability\(^3\), an expanding rent gap, gentrification\(^4\), and increased evictions\(^5\).

**Figures A1 and A2** in the Appendix illustrate the increasing rent as well as rent disparities across Boston.

To date, the City of Boston has enacted various ordinances to provide affordable housing and protect households most at risk\(^6\). We propose a data-informed approach to better utilize this existing structure and redistribute the costs and benefits of transactions in the short-term housing market. Specifically, we (1) identify the existing fiscal policies impacting the rental market, (2) utilize rental and demographic data to highlight areas of Boston most impacted by short- and long-term rental market interactions, and (3) propose a pilot program framework to engage community organizations and determine effective and fair allocations for affordable housing funds.

**Important Definitions**

*Short-Term Rental.* A property in which one or more rooms or units is rented for no more than 31 consecutive days. Short-term rentals are based on advanced reservation, and do not refer to registered hotels, motels, lodging houses or bed and breakfasts. The most prominent example is AirBnB.

*Rent.* In the case of short-term rentals, rent refers to the sum of all charges levied in exchange for occupancy, including insurance, booking fees, cleaning fees, among others. Any rent that exceeds $15 per night is subject to taxation.

*Affordable Housing.* Generally, housing that costs less than 30% of household gross income though some planning documents in Boston use a cutoff of 35%.

\(^5\) Budget and Legislative Analyst’s Office (2015). Analysis of the impact of short-term rentals on housing. Memo to Board of Supervisors, City and County of San Francisco. [https://sboos.org/sites/default/files/FileCenterDocuments/52651-BLA_ShortTermRentals_051315.pdf](https://sboos.org/sites/default/files/FileCenterDocuments/52651-BLA_ShortTermRentals_051315.pdf)
Existing Policies

Both the Commonwealth of Massachusetts and the City of Boston have enacted legislation addressing the prevalence of short-term rentals and their impact on the general Boston housing market. Most recently, Mayor Walsh signed a city-wide ordinance specifically requiring the classification and registration of short-term rentals7. Effective as of January 2019, this new ordinance also imposes restrictions on the number of nights a unit can be booked per year; requires an annual license fee for short-term rental operators; extends the room excise tax to short-term rentals; and bans “investor units” wherein operators register and list short-term rentals in units other than their primary residence8.

The city’s new ordinance comes in tandem with a law passed by the Commonwealth of Massachusetts to extend the state room occupancy tax (5.7%) and convention center financing fee in relevant areas (2.75%) to short-term rentals. This law, G.L. c. 64G Room Occupancy Excise, also regulates housing units that exceed 14 days on short-term rental markets by requiring the landlord to apply for a rental contract. The policy takes effect July 1, 2019 for contracts granted on January 1, 2019 or later. It also gives the option for local municipalities to impose local room excise taxes and additional community impact fees9. The key financial regulation affecting short-term rentals and affordable housing in Boston is summarized in Table 1.

<table>
<thead>
<tr>
<th>Table 1. Key financial programs relevant to Boston housing</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Name</strong></td>
</tr>
<tr>
<td><strong>Taxes on Short-Term Rentals</strong></td>
</tr>
<tr>
<td>Local Room Option Excise Tax</td>
</tr>
<tr>
<td>State Excise Tax</td>
</tr>
<tr>
<td>Community Impact Fee</td>
</tr>
<tr>
<td>Convention Center Finance Fee</td>
</tr>
<tr>
<td><strong>Funding for Affordable Housing</strong></td>
</tr>
<tr>
<td>Linkage Fee</td>
</tr>
<tr>
<td>Community Impact Fee</td>
</tr>
</tbody>
</table>

---

While these efforts are to be commended as progress for the City of Boston in achieving its Imagine Boston 2030 Housing Goals to preserve existing affordable housing and protect households most at risk, questions remain as to how one might assess and identify short-term rental units that should re-enter the long-term rental registration market; how the city ordinance’s requirements of registration will be enforced; and how the local revenue from both the room occupancy tax and community impact fee will be allocated.

**Our Approach**

Our investigation focuses on one of the above questions, specifically how the City of Boston could allocate its future revenue from the room occupancy tax and community impact fee. The analysis and our subsequent policy recommendation include three overarching components:

1. **Identify neighborhoods** most in need of affordable housing due to short-term rental trends
2. **Collaborate with neighborhood leadership** to identify community priorities and needs
3. **Establish a pilot program** supporting key development initiatives specific to each neighborhood and leveraging existing structures

As part of our recommendation, we provide an initial assessment of which neighborhoods are most impacted by short- and long-term rental market interactions by computing a “Neighborhood Vulnerability Index.” Below we describe the datasets used, our methodology, and our findings.

**Quantifying Sensitivity to Short-Term Rentals**

We need to understand the relative situation of each of Boston’s 25 neighborhoods to decide how to allocate affordable housing funding. As a first assessment, we use an Airbnb price dataset for Boston compiled by Inside Airbnb to calculate the average price of each one-bedroom listing over 2018, and then compare the median price in each neighborhood to the city median of $156 per night. Similarly, we compared median long-term rental rates in each neighborhood to the city median of $2,250 per month using rental rate statistics on one-bedroom apartments published by Zumper. Comparing each neighborhood's deviation from the citywide median highlights potential hotspots for exceptional high Airbnb prices and/or rental rates. **Figure 1a** illustrates each neighborhood's deviation in Airbnb prices; we see central Boston (with Fenway, Back Bay, and Downtown) drives up the citywide median price while values in Boston's extremity neighborhoods fall below the citywide median.

---


However, to better link Airbnb prices and rental rates, we need to consider other neighborhood statistics. Thus, we separately compute the Neighborhood Vulnerability Index based on each region’s median one-bedroom rent, poverty rate\textsuperscript{16}, share of total units that are rented long-term, share of total units deemed affordable (public and private), share of total units listed for short-term rental, and proportion of vulnerable affordable housing\textsuperscript{17}. Synthesizing previously used measures of gentrification and housing change\textsuperscript{18}, we reweight the indices to better reflect the specific impact of short-term rentals. Figure 1\textit{b} maps the Neighborhood Vulnerability Index across Boston with possible scores ranging from 0 (least vulnerable) to 100 (most vulnerable). The least vulnerable neighborhoods include Hyde Park, West Roxbury, and Roslindale, while more vulnerable neighborhoods include Fenway, Downtown, South End, and Chinatown. This highlights the outsized effect Airbnbs have on gentrifying neighborhoods such as Fenway and Chinatown, avoiding those that have less cultural significance to tourists and protected with abundant affordable housing.

\textbf{Figure 1. a}, right: Price Deviations in Nightly Air\textregistered{}BnB Rates. \textit{b}, right. Vulnerability metric, by neighborhood.

We suspect that other variables like rates of change and supply elasticities in the housing markets are also important factors in vulnerability, and we intend that the method for calculating the index will evolve through conversations with domain experts, neighborhood leaders, and housing and development teams in other cities. We also want to emphasize that the overarching purpose of a neighborhood-level index is to identify areas of the city that could most benefit from targeted affordable housing funding and to provide a bigger picture method to monitor how regions change over time.


Pilot Program for Revenue Redistribution

Having identified priority regions for targeted funding, we propose the City of Boston establish a pilot program on the distribution of new tax-revenue funds received. These funds would be received beginning July 2019 through the room excise and community impact taxes on short-term rental transactions.

The overall aim of the pilot is to support key development initiatives specific to each neighborhood with the goal of supporting affordable housing. Community funding recipients could be identified in conversation with neighborhood associations. Examples of candidate programs are

- Supplementing insurance for landlords who rent to the formerly homeless (e.g., Landlord Guarantee Pilot)
- Subsidizing legal representation for tenants facing eviction (e.g., Jim Brooks Community Stabilization Act)
- Supporting jobs trusts (e.g., Neighborhood Housing Trust, Neighborhood Jobs Trust)
- Supporting compact or co-living housing developments (e.g., Compact Living Pilot)

Neighborhood associations would provide periodic updates on the progress of supported programs, and at the end of two years, the City would perform a quantitative analysis of the program’s impact on affordable housing availability in the neighborhood.

Overall, our policy proposal is designed to be realistic, rigorous, effective, and fair. We have identified an already upcoming revenue source, which in the case of the community impact fee, already has a requirement that it support affordable housing. We suggest decision-makers adopt a data-informed approach on how to best allocate those funds to the regions most sensitive to short-term rental prices. Finally, we suggest a framework for a pilot program that supports a variety of candidate programs promoting affordable housing. Diverse angles are encouraged, and funding decisions are made in collaboration with neighborhood associations to ensure the local community is enhanced for its permanent residents.
Appendix

Figure A1. Median one-bedroom rent in Spring 2018, by neighborhood.

Figure A2. Average monthly rent in Boston, by apartment size for 2011-2018
POLICY BRIEF
Tax-Based Solution to Commercial Airbnbs

TEAM
Rutgers

TEAM MEMBERS
Aneesh Deshpande
Jaidev Phadke
Abdullah Shareef
Aditya Shastri
**Tax-Based Solution to Commercial Airbnbs**

**Introduction**

Airbnb provides people with the ability to experience many different areas and cultures through a home sharing platform. It comes with increased profits for the renter as well as a convenience of listing the property online and allowing the website to handle the booking. Due to these incentives, many homeowners and landlords are choosing to rent properties as Airbnbs, which are one type of Short Term Rental (STR) rather than offering Long Term Rentals (LTR). Across the United States, many cities view Airbnb as a business disrupting the housing market and increasing housing prices for the community [1].

**Problem**

**Vacancies**

Airbnb brings about potential inefficiencies and issues in the market that would otherwise not arise with long term rentals. One particular factor we chose to examine was occupancy: the ratio of the time the STR was booked to the time the STR is listed as available. The data on occupancy is incomplete because it doesn't contain the exact days that the STR is booked. Therefore, we used the Occupancy Model recommended by Inside Airbnb, the independent and noncommercial organization that provides this data [2]. We replicated the Occupancy Model for the Airbnbs listed in Boston from April 14, 2018, to the present day. Some key assumptions of this model are that approximately half the tenants leave reviews and that the average stay is 3 nights unless the listing imposes a longer minimum stay. Using these assumptions, we calculated how many days the unit was booked over the course of the same time period. As the figure shows, most Airbnbs are booked for shorter portions of their availability as listed online, while the number that are always full is much fewer. Notice that over half of Airbnbs aren’t occupied for a majority of their availability.

![Occupancy Rate of Airbnbs](image-url)
We view vacancy as lost potential for a resident to live in the particular neighborhood as well as some lost revenue. Hence, we partially target our policy recommendation to incentivize more long term rentals so that there are fewer vacancies and more long term residents.

**Ordinance Enforcement**

On paper, the City of Boston, through their January ordinance, has created a solid set of rules on STRs that could increase the amount of housing available for LTRs. Namely, the ordinance requires that STRs much be owner-occupied or owner-adjacent, thus eliminating the presence of commercial enterprises parading as individual landowners on AirBnB within Boston. Despite this step forward, the broad-reaching ordinance is difficult to enforce in practice.

For starters, of the 400,000 properties in Boston, there are 63,000 homes that are eligible to be STRs under the detailed criteria of the ordinance [3]. We parsed through this data by looking at the factors such as whether the unit was occupied or not, and whether the units were properly owner-adjacent. Clearly, this is a huge population of potential STRs to regulate.

On top of this, of the 6,155 AirBnB’s listed in Boston (which we counted from the AirBnB dataset) only 12 are currently registered as STRs under the rules of the new ordinance. This exemplifies that the current ordinance, perhaps because it is in its infancy, has yet to push landlords to use the registration system.

Most importantly, though the rules requiring STRs to be owner-occupied and owner-adjacent have great potential, they have failed in practice. We concluded this by identifying the number of AirBnB hosts who have multiple addresses listed on the platform at the same time. Because these different-address properties are clearly not “within owner-occupied two- or three-family buildings,” they don’t fit the rules for owner-adjacent (and certainly not owner-occupied) buildings. As a result, they are illegal listings under the new ordinance - but they are still thriving online! This population of STRs, which we call ‘multi-home’ properties, is the one we will target for the remainder of our paper. As the Boston Area Research Initiative points out, the vast majority of ‘multi-home’ properties are “quasi-hotel agencies and commercial enterprises” that don’t spend their revenue in local environments [3.5].

**Risks of Mis-implementing New Taxes**

As many know, Massachusetts has recently passed legislation authorizing taxes on STRs, with special rules for Boston [5]. Generally, one would expect such a tax to strictly be used to gain revenue for the city. However, because of how onerous it is to enforce Boston’s ordinance over multi-home properties, we recommend using this
new taxing power as a market-based incentive for landlords to shift from STR to LTR. In essence, because cracking down on illegal rentals is so challenging (especially considering the ongoing negotiations with the enforcement vendor), we want landlords to make that switch on their own.

Using the tax as an incentive for multi-home landlords to shift from STR to LTR must be done in a way that avoids hurting homeowners that are renting out their own homes for short periods of time. While the new state law creates a tax exemption for those renting their homes for under 14 days a year [5], we believe that there should be more protections for this type of healthy STR business. As one researcher points out:

“Regulators should not restrict home sharing for people who would not have made their homes available to long-term renters anyway, such as: owner-occupiers; using Airbnb to share unused rooms or rent their home while they are away.” [6]

These people are clearly members of local communities, and as local spenders, they bring large economic benefits to the area. With this in mind, we believe that applying the tax in a blanket fashion across all AirBnB listings in Boston, or even in a given neighborhood, is critical to avoid. Our primary recommendation will focus on ensuring that the forthcoming tax plan for the City of Boston maximizes the shift from STR to LTR, without hurting the positive externalities of STR for local communities. By focusing on ‘multi-homes’ we will address both of these goals.

**Recommendations**

We look to build on top of the STR ordinance that has been put in place to make it more potent as well as provide alternatives where the STR ordinance fails to be effective. By recognizing both the ordinance and the recently authorized STR taxation abilities available to the Boston government, we believe we are propose a practical solution that is in line with the current plan and approach of the city. As our keynote speaker, Boston’s Chief Data Officer, mentioned - data analytics means being practical first, and fancy second. We want to ensure our analytics are insightful, but well tied to the city’s priorities.

*Tax on Multi-Home Airbnb Units (“Multi-home tax”)*

Clearly, we want Airbnb to continue to flourish as a business as long as it benefits the housing economy of Boston. While owner-occupied STRs have mixed effects, multi-home units - generally commercial enterprises - pose a clear harm. This is because, in addition to breaking the law, they are most likely to take up space from potential LTR housing stock, and accrue profits that are not reinvested into local communities.
Our primary solution to this is to narrow the population of STRs over which the tax stated above is enforced. Instead of applying a flat tax rate across all STRs, we will focus the tax on the highly commercial multi-home properties. The law indicates that a 5.7% state tax will be applied to all properties, and that Boston can implement a 6.5% tax as well (this is the tax that is relevant to us).

The goal of this tax is to disincentivize landlords from putting up property on Airbnb that can be rented as a LTR. Because a vast majority of Boston's Airbnb marketplace is multiple homeowners [4], this tax would be effective in reducing the number of these long term commercial Airbnbs being rented as STRs and put them into the LTR market.

Using the same Occupancy Model, the revenue generated from this tax can be estimated to be between $6.08 M and $8.41 M per year. This was calculated by multiplying the price per night with the approximate booking time for the listings that fit the criteria for the tax.

Importantly, to avoid the enforcement troubles experienced with the ordinance plan, we will consider this tax an opt-out measure. In other words, if a landlord can show proof of residency at their property sometime during the previous year, or if they can show that listings are for less than 28 days, they meet the ordinance requirements and are free from the tax. These protect “true” short term rentals by people who are simply sharing their home.

**Tax Incentives**

In order to incentivize switching over a unit from STR to LTR, we can provide tax relief to those landlords to do so. The income they get from their new rental will be tax free (state income tax) for the first year and they will not pay property taxes for the first year (local tax).

**3%-Scaled Community Benefit Tax**

In addition to 6.5% tax, the state legislation allows towns and cities to impose a “community impact fee” of up to 3 percent. According to the law, at least 35% of revenues from this fee (or tax) must be dedicated toward affordable housing or local infrastructure projects [5]. We developed a heat map to demonstrate the frequency of multi-home owners as a fraction of total Airbnbs in various neighborhoods neighbourhood. Our tax would target neighborhoods with higher frequencies of multi-home owners to help decentivize this commercial practice, and subsidize revitalization in these same neighborhoods (because they're most likely to have housing shortages and rent increases).
Affordable Housing Investment

The revenue from the multi-home tax can be used to finance and subsidize housing projects around the city of Boston, especially targeting neighborhoods affected the most by increasing rental prices.

Local Community Incentives

The revenue from the community benefit tax can be used to form a public-private partnership with Airbnb to provide coupons and discounts for local businesses. This incentivizes tourists to explore the local community rather than just living there in an Airbnb and leaving for more popular places. This would be implemented through the “Experiences” tab, and would show up with coupon codes in the early suggestions area.

Conclusion

Given how we’ve observed the Airbnb impact on communities, we strongly feel that our data analysis helps points toward creating an incentives-based tax structure to help establish market equilibrium of STRs and LTRs. Modern tech must be reflected by modern policy, and the revenue that can be generated from taxing STRs at the proposed rate through the Boston Ordinance Law will surely have a positive effect on the relative supply of LTR units, thereby making LTR more affordable. Our data driven models provided us with a revenue stream to base our proposals off of in terms of budgeting, and we were useful in creating visual tools to truly realize the effects of short term home rental services in the area. A series of internal tools in terms of coupons, a community-beneficiary tax, and public-private partnerships all helped support our key policy measure. Ultimately, we believe that our solution takes strong heed of the existing direction and capabilities of Boston’s city government on this issue. When applying data analytics to governance, the primary challenges are keeping recommendations feasible and relevant to a department’s existing work. Our pragmatic approach is a direct response to this issue.
[3.5] https://www.youtube.com/watch?v=wRMnwe3fsF8&feature=youtu.be&t=1497
POLICY BRIEF

Restoring Communities: How to Use Short Term Rental Regulation to Fight Displacement

TEAM
Urban Development Team One

TEAM MEMBERS
Adam Beach
Natalie Lau
Hardy Montoya
Jen Zhang
Title: Restoring Communities: How To Use Short Term Rental Regulation To Fight Displacement

1. Introduction

The gentrification of Boston is well documented. Research indicates that since 2000, 21.1% of eligible tracts in the city are gentrifying, compared to a rate of 6.7% between 1990 and 2000 (Governing.com)\(^1\). The process brings with it the gradual pricing-out of a community’s less well off residents and, by extension, evictions.

While the debate surrounding gentrification is not new, one of its most recent fronts has centered on the regulation or non-regulation of Short Term Rental (STR) services: the sharing-based lodging service popularized by platforms like Airbnb and HomeAway\(^2\). The question this paper will seek to answer is then: how does the prevalence of STR services impact the local housing market, including its relevance to gentrification, and if a link exists, how should local governments respond?

First, research suggests that there is a causal link between the proliferation of investment-centered STRs and climbing rent prices. Figure 1 shows the concentration of Airbnb in the Boston area and Figure 2 shows the percentage of renter households that pay 30% of the household income rent (Inside Airbnb, 2019; Boston Displacement Mapping Project, 2019). A side-by-side comparison between the two graphs present a staggering correlation between the increase in STRs and rent burden which presents a strong displacement and gentrification risk in the community.

---

\(^1\) Gentrification here understood as tracts that go from the bottom 40\(^{th}\) percentile of median household income to the top 3\(^{rd}\) percentile in both median income and percentage of adults with a bachelor’s degree.

\(^2\) the City of Boston defines STRs as residences that rent to visitors for periods of thirty calendar days or less.
**Title:** Restoring Communities: How To Use Short Term Rental Regulation To Fight Displacement

![Figure 1. Concentration of Airbnb in Boston](image1)

Figure 1. Concentration of Airbnb in Boston

![Figure 2. Percentage of Renter Households that Pay 30% of the Household Income Rent](image2)

Figure 2. Percentage of Renter Households that Pay 30% of the Household Income Rent

Data from Boston indicates that there is a .018% increase in rent prices per 1% increase in available AirBnB listings in a given zip code (Barron et al 2018). Second, the same research suggests that there is a causal link between an increase in Airbnb and a decrease in the supply of vacant units. Consider also that in 2017 available Airbnb listings increased approximately 35% over the previous year (CNBC, Kharpal, 2017), data from which we can infer that rent prices and housing availability will continue to be adversely affected by the rise of STRs.
Title: Restoring Communities: How To Use Short Term Rental Regulation To Fight Displacement

A serious assessment of gentrification in Boston must also include analysis of how the process impacts certain intangible indicators, such as the culture and social spirit. According to the Chinese Progressive Association, 90% of respondents indicated that living in Chinatown kept them connected to the larger Chinese community. Respondents indicated that living in Chinatown made it feel “...easier to communicate...” and that they would like their children to be around other Chinese people to “understand the culture”.

It is clear from these responses that Chinatown residence fear for a slow erosion of their Chinese identity, and the complete annihilation of Chinatown as a social unit. While these metrics may be considered “soft” policy outcomes, it is clear that Boston’s diversity of culture, language, ethnicity, and religion faces an existential threat in the advance of gentrification and displacement, brought about in part by the expansion of the STR market.

2. Background of Existing Regulations
The negative effects that the STR market boom has had on the availability of affordable housing led Boston to regulate the industry. Part of the ordinance proposed has been passed and is now in effect, while other aspects are currently held in court by a lawsuit from Airbnb. Current policy limits the number of days that Investor Units can be offered as STR, and the number of days the days that Home Share Units can while the operator is not present. Current estimates indicate that potentially more than 2000 units would return to the long term rental market at in the near future after implementation of this policy (Sullivan, 2018). While this initial measure reduces the desirability of incrementing the number of units offered by single operator, it does not do enough to revert the currently problematic state already created by the overabundance of STRs in certain areas, which has caused the gentrification of neighbors with a majority non-white demographic, and placed its populations at risk of being displaced.

3. Policy Proposal and Analysis
Community building, increasing housing unit supply, and housing affordability are critical to the urban development of Boston. To mitigate the negative impacts of STRs, we have come up with three policy solutions, including (1) Anti-Displacement Policy, (2) Progressive Tax on Multi-listings and (3) Subsidies for Affordable Housing. In the following sections, we would discuss the objectives, case examples, and the implementation plan:

3.1 Anti-Displacement Policy
As discussed in 1.2, neighborhoods like Chinatown have already experienced the impacts of STRs where families and businesses are forced out of the neighborhood that they resided in for years. In order to rebuild the community, we are proposing an Anti-Displacement Policy where residents who were displaced by the rise of STRs would be given priority status in affordable housing lotteries in the district they resided in[1].
Title: Restoring Communities: How To Use Short Term Rental Regulation To Fight Displacement

While we recognize the importance of the Short-term Rental Ordinance in limiting the supply of STRs, we think that it is equally important to produce a policy that provides relief for residents and communities who have been severely impacted by the emergence of STRs. And we are certainly not the first to propose the solution.

To combat the threat of gentrification, the City and County of San Francisco implemented an ‘Anti-Displacement Housing Preference (ADHP)’ policy to help residents who face extreme displacement pressure to claim a preference for public housing application (City & County of San Francisco, 2019). Similarly, in 2015, the Portland Housing Bureau implemented the “Right to Return” initiative that gives preferences to displaced families with historic roots in inner North and Northeast Portland for rental housing and home ownership programs (Portland Housing Bureau, 2015).

3.2 Progressive Tax on Additional STR Units
Considering the severity of STRs on housing supply, we suggest imposing a progressive tax on STR operators who rent out more than one STR Investor Unit. Under the new legislation, STR operators would pay a local excise tax of 6.5%, the same rate that Massachusetts taxes hotels and other lodging establishment (Commonwealth of Massachusetts, 2019).

While we agree that this is a strong fiscal policy to reduce the supply of short-term investment rental units, the ordinance does not necessarily target operators who rent out multiple listings. Currently, 64% of the STR operators have more than one listing, which, often times, have a higher vacancy rate. To reduce the number of multi-listed investment units, we believe that it would be more effective to increase the marginal cost of listing additional investment units via a scaling tax that increases incrementally per unit operated.

3.3 Affordable Housing Trust Fund
With a significant increase in tax revenue from the newly imposed excise tax, we suggest that the City of Boston redistribute the tax revenue through the City Affordable Housing Trust Fund to support the production and preservation of affordable housing units (Housing Trust Fund Project, 2019).

In 2015 State of the City address, Mayor Eric Garcetti committed to negotiating a TOT contract with Airbnb and injecting $5 million from those revenues into the City’s Affordable Housing Trust Fund. The Los Angeles City Council approved the $5 million allocation for the Affordable Housing Trust Fund, for the 2015-16 Fiscal Year (Carbajal, 2015).
4. Implementation

4.1 Anti-Displacement Program

The development and implementation of our proposed Anti-Displacement Program will be best served under the direction of the City of Boston’s Office of Housing Stability. The Office’s mission of keeping communities intact (“New Tools”, 2018) aligns with the restorative goals of the program. Key stakeholders include the Boston Fair Housing Commission, Boston Housing Authority, Boston City Council, Mayor’s Office, Boston Planning and Development Agency, local housing advocacy groups—such as CityLife/Vida Urbana—as well as neighborhood councils. For Boston’s Chinatown neighborhood—an area that has been disproportionately affected by STRs and housing displacement—community buy in from members such as the Chinese Progressive Association and Boston City Councilor Ed Flynn is essential for the effectiveness of the program.

The Anti-Displacement Policy would apply to residents who have been displaced over the past nineteen years. These past two decades have seen accelerated depopulation of the local community (CPA, 2019) coinciding with rapidly increasing median rent prices. To counteract this trend, we propose to attach weighted point values (“preferences”) to the affordable housing applications of formerly displaced residents now residing outside of the City. The aim would be to aid previously priced out residents in the affordable housing lottery system run by the Boston Fair Housing Commission. We also aim to substantially increase the vacancy rates for affordable units in certain low-income neighborhoods; namely those that have experienced high rates of displacement since 2000. Operating in tandem with our proposed policy, the STR ordinance passed by the City Council disincentivizes the proliferation of investor unit STR listings in low-income neighborhoods. This negative pressure, coupled with the scaling taxation on operators who own multiple investment STR units, will cause a substantial return of STR units to the long-term rental market.

In order to ensure that a substantial number of units returning to the long-term rental market are affordable, we propose a rent subsidization program in tandem with the proposed anti-displacement policy to incentivize operators to list a higher proportion of their property at affordable prices in historically low-income communities. This would be funded by the Affordable Housing Trust, which is in turn funded by the proposed scaling taxes on multi-unit operators.

4.2 Performance metrics and assessments

In order to ensure the success of the policy it will be critical to build in performance standards and assessment mechanisms so that policymakers can continue to evaluate the legislation’s returns. As such, we propose the following performance indicators as the most prominent points of the policy’s success:
Title: Restoring Communities: How To Use Short Term Rental Regulation To Fight Displacement

1. Percentage of newly vacant housing that is listed as affordable,
2. Percentage of displaced persons that choose to return to a gentrified community and,
3. Rate of resident retention in newly vacant housing.

We believe that these indicators are the most appropriate because the policy’s goals can be broadly defined as returning the largest possible amount of displaced persons to their newly-vacant homes, and promoting conditions that allow them to remain there. In short, assessments of the policy’s efficacy should quantify how many people are returning, and how many people are staying.

Secondly, the policy must contain provisions for bi-annual neighborhood surveys to assess the legislation’s impact along the aforementioned outcomes. We believe that a bi-annual assessment is appropriate in light of the rapid gentrification of predominantly minority communities (e.g. Chinatown) over the course of the past two decades. A bi-annual assessment is necessary to provide a detailed statistical picture of the policy’s success.
Title: Restoring Communities: How To Use Short Term Rental Regulation To Fight Displacement

References


Title: Restoring Communities: How To Use Short Term Rental Regulation To Fight Displacement


