AJUR

American Journal of *Undergraduate Research*

Volume 16 | Issue 2 | September 2019

www.ajuronline.org

Print Edition ISSN 1536-4585
Online Edition ISSN 2375-8732

AJUR
American Journal of
Undergraduate Research

Volume 16 | Issue 2 | September 2019 *www.ajuronline.org*

***American Journal of Undergraduate Research (AJUR)** is a national, independent, peer-reviewed, open-source, quarterly, multidisciplinary student research journal. Each manuscript of AJUR receives a DOI number. AJUR is archived by the US Library of Congress. AJUR was established in 2002, incorporated as an NFP in 2018. AJUR is indexed internationally by EBSCO and Crossref with ISSNs of 1536-4585 (print) and 2375-8732 (web).*

SPECIAL THANKS

AJUR is made possible through the volunteer efforts of our editorial team and assistance of our sponsors.

We thank our sponsors:

OSWEGO
STATE UNIVERSITY OF NEW YORK

Please support undergraduate research
Request sponsorship information from *editor@ajuronline.org*

Classifying Lensed Gravitational Waves in the Geometrical Optics Limit with Machine Learning

Amit Jit Singh[*,a]*, Ivan S.C. Li*[b]*, Otto A. Hannuksela*[a]*, Tjonnie G.F. Li*[a]*, Kyungmin Kim*[a]

[a]*Department of Physics, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong*
[b]*Department of Physics, Imperial College London, London SW7 2AZ, United Kingdom*

https://doi.org/10.33697/ajur.2019.019

*Students: singhamitjit@link.cuhk.edu.hk**, *ivan.li16@imperial.ac.uk*
Mentors: otto.hannuksela@link.cuhk.edu.hk, tgfli@cuhk.edu.hk, kyungmin.kim@cuhk.edu.hk

ABSTRACT
Gravitational waves are theorized to be gravitationally lensed when they propagate near massive objects. Such lensing effects cause potentially detectable repeated gravitational wave patterns in ground- and space-based gravitational wave detectors. These effects are difficult to discriminate when the lens is small and the repeated patterns superpose. Traditionally, matched filtering techniques are used to identify gravitational-wave signals, but we instead aim to utilize machine learning techniques to achieve this. In this work, we implement supervised machine learning classifiers (support vector machine, random forest, multi-layer perceptron) to discriminate such lensing patterns in gravitational wave data. We train classifiers with spectrograms of both lensed and unlensed waves using both point-mass and singular isothermal sphere lens models. As the result, classifiers return F_1 scores ranging from 0.852 to 0.996, with precisions from 0.917 to 0.992 and recalls ranging from 0.796 to 1.000 depending on the type of classifier and lensing model used. This supports the idea that machine learning classifiers are able to correctly determine lensed gravitational wave signals. This also suggests that in the future, machine learning classifiers may be used as a possible alternative to identify lensed gravitational wave events and to allow us to study gravitational wave sources and massive astronomical objects through further analysis.

KEYWORDS
Gravitational Waves; Gravitational Lensing; Geometrical Optics; Machine Learning; Classification; Support Vector Machine; Random Tree Forest; Multi-layer Perceptron

INTRODUCTION
A total of six gravitational wave (GW) signals have been detected by LIGO and Virgo detectors at the time of writing. The source of these signals were believed to originate from merger events of binary systems, such as binary black holes (BBH)[1–5] or binary neutron stars (BNS).[6] With these recent detections, we now have the opportunity to study astronomical objects and events through GW observations in addition to the traditional methods of electromagnetic (EM) wave observations.

When a GW signal passes by massive objects, the incoming wave behaves similarly to light in that the signal becomes gravitationally lensed (as shown in **Figure 1**). This changes the amplitude of the detected signal, and can cause multiple images from the same GW source to be detected at different times.[7] We consider two lensing models in this study: the point mass lens and the singular isothermal sphere (SIS).[8, 9] Our aim is to classify an incoming GW signal as lensed or unlensed through machine learning techniques. By doing so, we may be able to better understand the properties behind the lens mass and GW source through further analysis of the GW signal.

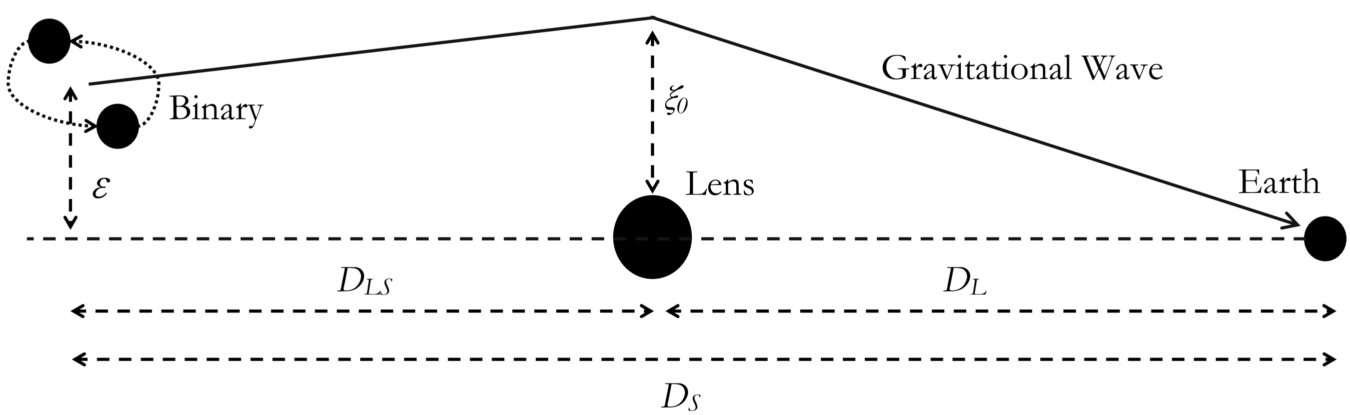


Figure 1. A simplified diagram of gravitational lensing of GWs. The binary system is the source of the gravitational waves. ε is the distance of the binary from the line of sight extending from the lens and Earth. D_{LS} is the distance from the source to the lens and D_L is the distance from the lens to the Earth. ξ_0 is the Einstein radius of the lens.

Currently, the main method of identifying GW signals are through matched filtering techniques with pre-existing post-Newtonian GW waveform models.[1–6, 10] Some studies also suggest that matched filtering techniques can be used to obtain useful information such as the binary parameters of the GW source.[1, 11, 12] However, for this study, we focus on the ability of supervised machine learning classifiers (MLC) to determine if an incoming GW signal is lensed, as a number of difficulties arise with the traditional match filtering techniques. With match filtering, we require accurate templates of the waveforms to compare with detected signals. However, the current lensing models are not exact and thus their templates may not match with actual lensed events, making the method relatively imprecise. Furthermore, unlike unlensed GWs, we would require significantly more templates for lensed GWs due to the different number of lens that exist and the increased number of physical parameters that come with lensing. On the other hand, machine learning does not require us to have an exact model and we do not have to concern ourselves with creating templates with various parameters due the probabilistic nature of the machine learning classifiers. In particular, we choose to test three popular supervised MLC algorithms: support vector machine,[13] random forest,[14] and neural network.[15]

In this study, we generate spectrogram samples of lensed and unlensed GW signals under both the point mass lens and SIS models. We use these spectrogram images as the training and testing data for these MLCs, and we analyze the performance of each MLC in the identification of lensed GWs by comparing the F_1 scores of each classifier.[16]

Gravitational Lensing of Gravitational Wave Signals

Models for lensing of GWs have been studied for decades.[9, 17–23] From these studies, it is known that the lensing effect needs to be treated differently depending on the wavelength, λ, of a GW. We work in the geometrical optics approximation, where the wavelength of the incoming GW is much shorter than the Schwarzschild radius of the lens mass. At longer wavelengths, the diffraction effect becomes dominant and the geometrical optics limit treatment of the GW lensing behaviour becomes invalid.[9] Throughout this study, we will be working in geometrized units (i.e. $G = c = 1$). We also make use of the thin lens approximation, where we assume that the GWs are scattered by a lens with its mass distributed across a two-dimensional plane perpendicular to the line of sight of the observer from the lens mass.[8]

In general terms, we can write the relation between the lensed GW, $h_{\text{lens}}(f)$, and the unlensed GW, $h(f)$, in the frequency domain such that $h_{\text{lens}}(f) = F(f)h(f)$. In this relation, the coefficient $F(f)$ is called the amplification factor and under the geometrical optics limit is defined as[9]

$$F(f) = \sum_j |\mu_j|^{1/2} e^{2\pi i f \Delta t_{d,j} - i\pi n_j}, \qquad \textbf{Equation 1.}$$

where μ_j is the magnification factor of the j-th image of the lensed GWs, $\Delta t_{d,j}$ is the time delay of the j-th image, and

n_j is a discrete number which takes different values depending on the lens model. For a given lens model, the magnifications and time-delays will differ.

In particular, the amplification factor is found by first using the lens surface mass density of a given lensing model to compute a deflection potential for the lens. Finding the solution to an integral related to the deflection potential and various other source parameters would then yield an expression for $F(f)$. The amplification factor is determined not only by the frequency of the GW, but also by the lens position and its mass distribution.

Point-Mass Lens Model

For a point mass lens model, the lens mass is defined as a two-dimensional Dirac-delta function on the thin lens plane. Under this model, two images are detected by the observer, and **Equation 1** reduces to[9]

$$F(f) = |\mu_+|^{1/2} - i|\mu_-|^{1/2}e^{2\pi i f \Delta t_d}. \quad \textbf{Equation 2.}$$

By considering the red-shifted lens mass $M_{Lz} = M_L(1+z)$, we can define the magnification factors of the two interfering GW images as $\mu_\pm = 1/2 \pm (y^2+2)/(2y\beta)$ and the time delay of the images as $\Delta t_d = 4M_{Lz}[y\beta/2 + \ln((\beta + y)/(\beta - y))$ where $\beta = \sqrt{y^2+4}$. The variable y is a value which parameterizes the relative displacements of the source position, observer, and lens position. This is given by

$$y = \frac{\varepsilon D_L}{\xi_0 D_S}, \quad \textbf{Equation 3.}$$

where ε is the distance of the source from the line of sight, D_L is the distance from the lens to the observer, D_S is the distance from the source to the observer, and ξ_0 is the Einstein radius of the lens.

Singular Isothermal Sphere Model (SIS)

Unlike the point mass lens, the singular isothermal sphere (SIS) model assumes that the lens mass is circularly and symmetrically distributed across the thin lens plane. It is generally used to model extended objects such as large stars or galaxy clusters. The SIS model is usually characterized by a velocity dispersion v, which directly relates to the mass distribution of the model.[8, 9] This consequently gives rise to different expressions for the magnification factors and time delay of the interfering GW signals. It should be noted that under the SIS model, there is no second image detected by the observer if the lensing happens outside of the Einstein radius. Hence, the amplification factor for the SIS model is adjusted to become[9]

$$F(f) = \begin{cases} |\mu_+|^{1/2} - i|\mu_-|^{1/2}e^{2\pi i f \Delta t_d} & \text{if } y < 1, \\ |\mu_+|^{1/2} & \text{if } y \geq 1\ , \end{cases} \quad \textbf{Equation 4.}$$

where $\mu_\pm = 1/y \pm 1$ and $\Delta t_d = 8M_{Lz}y$. In this case, M_{Lz} is defined as the mass inside the Einstein radius of the lens mass. The expression for the y parameter in the SIS model is equivalent to that of the point mass lens model, given by **Equation 3**.

Machine Learning Algorithms

We select three supervised classifiers which vary notably in their underlying algorithms for classification and prediction to find the most optimal method for analyzing spectrograms. We choose to utilize the support vector classifier (SVC), the random forest classifier (RFC) and the multi-layer perceptron classifier (MLP). Below we give a general overview of how each algorithm works in order to highlight the significant difference in their method of classification.

Furthermore, we do not focus on the mathematical formulations of each classifier as they are not pertinent to the overall aim of this research. Readers interested in the mathematics of these algorithms are directed towards the references mentioned below.

Support Vector Classifier (SVC)

SVC is based on the the support vector machine algorithm. Each item in our training set is plotted as a point in an n-dimensional space where n is the number of features representing an item and each feature is a coordinate of the item in the n-dimensional space. The algorithm calculates a hyper-plane which divides the two classes of spectrograms and acts as a decision boundary. When we introduce the training set to the classifier, it predicts if a gravitational wave in a spectrogram is lensed based on where it is in the n-dimensional space with respect to the hyper-plane.[13, 24]

The two most important hyper-parameters that we consider are C and γ. The γ parameter determines how the distance of each item in the training set influences the decision boundary. When γ is large, the items closer to the boundary carry a larger weight and influence it heavily while the ones further away do not have much influence on its shape. On the other hand, for a small γ, items nearer to the boundary are given less influence as the ones further away are also given importance. C is responsible for determining the cost of a smooth decision boundary against the cost of misclassification of training points. If C is large, the algorithm focuses on generating a decision boundary which classifies all items in the training set correctly. A small C prioritizes the smoothness of the decision function and allows for a softer boundary such that there are some items in training set that cross and overlap the decision boundary into the other class.[13]

Random Forest Classifier (RFC)

The RFC is based on decision tree algorithms. A decision tree consists of a root node, which would be the training set, which splits into decision nodes based on discriminating features found in the data by the algorithm. These sub-nodes will continue to split until only the terminal node remains which would consist of a homogeneous piece of the original data and thus, cannot be split any further. When a new test item is introduced to the decision tree, it will classify the item according to which terminal node the data is matched to. In a RFC, multiple decision trees are grown. When making a prediction on an item in the test set, the most popular classification given by all the trees is used (i.e. the mode). This reduces the chance of error as multiple random decision trees are used to make the prediction rather than just taking the result from one decision tree.[14]

For this classifier, we consider two hyper-parameters. The first is the number of trees in the forest (N_t). More trees lead to better predictions but also utilizes more computational power. Thus, tuning it is required to avoid using unnecessary computational resources. The second is the minimum number of samples required to allow a node to split (N_{mss}). This is an important factor in controlling over-fitting as a small value could lead to the algorithm selecting highly specific features only occurring in a small set of one class and thus, do not generalize well over the whole class. A large value would then quite clearly lead to under-fitting.[14]

Multi-layer Perceptron Classifier (MLP)

The multi-layer perceptron classifier is a feed-forward neural network. The neural network consists of multi-layers as suggested by its name. The first layer is the input layer which consists of neurons matching the number of features of the input data. After the input layer, there are a number of hidden layers. In each layer, each neuron performs a linear weighted summation and then, a bias is added to avoid zero values. A non-linear activation function is applied to this summation and forwarded to the neurons in the next layer where this is repeated and continues until the outer layer is reached. At the outer layer, the output is given. However, the output often does not match the actual expected output. Thus, the backpropagation is applied to fix this. Backpropagation aims at minimizing the loss function using gradient descent. The loss function calculates the numerical difference between the actual class and the predicted class from the network. The loss is then minimized by adjusting the weights and bias using gradient descent which partially differentiates the loss function with respect to all the weights and bias. Each repetition of feed-forwarding and back-propagation is called an epoch. The MLP algorithm repeats this for a certain number of epochs until efficiency is achieved.[15]

Although the MLP classifier has a number of hyper-parameters which can be optimized, we only consider the number of layers and the number of neurons in each layer ($N_{layer\_neurons}$) and the L2 penalty term α. α is used to regularize over-fitting by adjusting the size of the weights. A larger α leads to smaller weights and creates a smoother decision boundary which reduces over-fitting. On the other hand, a smaller α causes larger weights and a more complex decision boundary.[15] The rectified linear unit (relu) function is used as the activation function.

METHODS AND PROCEDURES

Gravitational Wave Model

The unlensed waveform that we consider is from a binary inspiral source. To simulate the lensed GW signal detected by the observer, we inject two GW strains $h(t)$ with a given amplification factor and time delay depending on the lensing parameters and lensing model used. In particular, we generate a waveform computed up to 0.5 post-Newtonian (PN) order. This approximation suffices for the purposes of this study as we are simply investigating whether the machine learning classifier is able to differentiate between a lensed and unlensed waveform. For this reason, we also choose to omit the post-merger waveform from our model. We use the following expression for the GW waveform[25]

$$h(t) = -8\sqrt{\frac{\pi}{5}}\frac{\mu}{D_S}e^{-2i\varphi(t)}x(t), \qquad \textbf{Equation 5.}$$

where μ is the reduced mass of the binary inspiral source. To 0.5 PN order, $x(t) = \Theta(t)^{-1/4}/4$ and $\varphi(t) = -\Theta(t)^{5/8}/\eta$ are the post-Newtonian parameter and orbital phase respectively. $\Theta(t) = \eta t/5M$ is a surrogate time variable, where M is the total mass of the binary system and $\eta = \mu/M$ is the symmetric mass ratio.

Parameter	M_L	D_L	D_{LS}	m_1, m_2	ε	z
Range	$10 - 10^7 M_\odot$	$10 - 1000\ Mpc$	$10 - 1000\ Mpc$	$4 - 35 M_\odot$	$0 - 0.5\ pc$	$0 - 2$

Table 1. Ranges of randomized parameters of the GW waveforms. m_1 and m_2 are the masses of the binary source, D_{LS} is the distance between the source and lens mass, and z is the redshift parameter. The source and lens masses are sampled from a logarithmic distribution to reduce the bias towards more heavily lensed waveforms being generated.

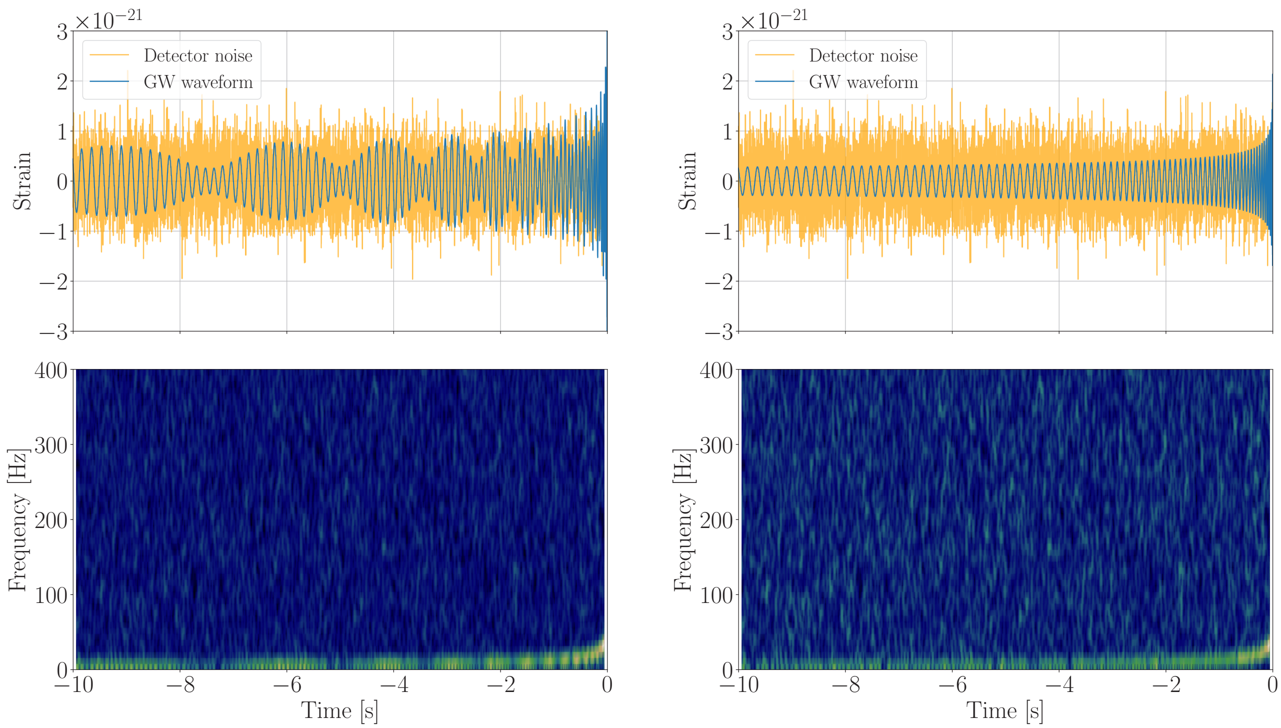
The resultant waveform is shown in **Figure 2**. A beating effect is clearly evident due to the two interfering waveforms arriving at the observer with a time delay between them. In our model, the merger of the two component masses in the binary system occurs at $t = 0$, which can be seen by the peak in the GW waveform. Unlensed waveforms are simply generated by injecting a single GW strain instead of two. We randomize the parameters of the lensed and unlensed GW waveform within the ranges presented in **Table 1**. We also implement Gaussian noise with an amplitude of the order 10^{-21} to test the ability of the classifier to identify a lensed waveform within a noisy background signal.

Signal-to-Noise Ratio (SNR)

The signal-to-noise ratio (SNR) is a measure of the magnitude of a GW signal in relation to the background noise detected. For a given GW waveform $h(f)$, we calculate the SNR using[25]

$$\text{SNR} = \sqrt{4\int_0^\infty \frac{|h(f)|^2}{S_n(f)}df} \qquad \textbf{Equation 6.}$$

where $S_n(f)$ is the power spectral density of the noise profile in the signal. The SNR from the recent six GW detections ranged between 13 to 32.4.[1–6] In our study, we choose to limit the SNR of the generated data to be less than 80. This is to ensure that the data we use for our MLCs are physically valid and plausible, and that the generated waveforms do not dominate the noise in the signal.



(a) Lensed Gravitational Wave **(b)** Unlensed Gravitational Wave

Figure 2. *Top left:* A lensed GW waveform under the point mass lens model (with noise, colored in orange) generated from a binary inspiral source.The two interfering GWs that form the resultant lensed waveform are detected by the observer at different times, causing a beating effect to be seen. *Bottom left:* A spectrogram of the incoming lensed signal is generated by performing a short-time Fourier transform on the signal. Coalescence of the binary system occurs at $t = 0$, as seen from the increase in frequency of the GW over time. The same beating effect due to the lensing of the waveform can be observed. The spectrogram images are then used as the training data for the classifier. *Top right:* An unlensed GW waveform with noise. The amplitude of the signal is smaller due to the lack of any lensing magnification. *Bottom right:* A spectrogram of the unlensed signal. No beating effect is seen, and the smaller waveform leads to the noise being a more prominent feature in the spectrogram.

Spectrogram

Since the GW waveform is generated in the time domain, we perform a short-time Fourier transform (STFT) on the data to extract the frequency information of the signal as a function of time. We then create a spectrogram of the incoming signal with the Gaussian noise added, as shown in **Figure 2**. Depending on the lensing parameters used, the beating effect due to gravitational lensing (seen in **Figure 2**) varies and is not always distinct even though the waveform is lensed. We prepare 2000 spectrogram samples of lensed GWs (1000 each for the point mass and SIS lensing model) and 1000 samples of unlensed GWs, which are used in the training and testing processes for the classifier. Using **Equation 6**, we find that our overall data has SNR values with a mean of 41 and a standard deviation of 19.

Optimization of hyper-parameters

Before we train our classifiers, we perform a grid search with cross-validation to determine the optimal combination of hyper-parameters for each classifier. The result of the grid search are presented in **Table 2**.

Applying the hyper-parameters stated in **Table 2**, we train each classifier. Then, we use the trained classifier to predict the classes of the spectrograms in the test.

Classifiers	Hyper-parameters	Value
SVC	C	1000
	γ	5×10^{-7}
RFC	N_t	1000
	N_{mss}	2
MLP	α	0.72
	$N_{layer\_neurons}$	1000

Table 2. Optimal hyper-parameters for each classifier which are determined using a grid search with cross-validation. The hyper-parameters which gave the best score were chosen.

Evaluation of MLC performance

We choose to assign 75% of our spectrograms as the training set and retain the rest as the test set. We perform two tests for each MLC, each time using different sets of spectrograms: first using the unlensed and point-mass-lensed spectrograms, then using the unlensed and SIS-lensed spectrograms. This results in two sets of data, one for each lensing model. We analyze the results from the classifiers using the classification report and the receiver operating characteristic curve. The classification report provides several figure-of-merits (FOM): the precision, $\mathcal{P}$, recall, $\mathcal{R}$, and the F_1 score of the respective classifiers. These are defined as

$$\mathcal{P} = \frac{\text{TP}}{\text{TP} + \text{FP}} \qquad \textbf{Equation 7.}$$

$$\mathcal{R} = \frac{\text{TP}}{\text{TP} + \text{FN}} \qquad \textbf{Equation 8.}$$

$$F_1 = \frac{2}{1/\mathcal{P} + 1/\mathcal{R}}, \qquad \textbf{Equation 9.}$$

where TP is the number of true positives (i.e. correctly classified lensed waves), FP is the number of false positives (i.e. unlensed waves misclassified as lensed waves) and FN is the number of false negatives (i.e. lensed waves misclassified as unlensed). Precision is a measure of the accuracy of the positive predictions, whereas recall measures the sensitivity of the classifier in identifying lensed signals. Precision and recall can be merged into a single metric called the F_1 score which is their harmonic mean, as shown in **Equation 9.**[24]

The receiver operating characteristic (ROC) curve is also used to find the most optimal classifier. The ROC curve plots the true positive rate against the false positive rate. The true positive rate is the same as recall, while the false positive rate is the ratio of the number of misclassified unlensed spectrograms to the total number of spectrograms classified as lensed. The ROC curve provides us with the information about whether the recall can be increased while keeping the false positive rate low. A theoretically perfect classification would allow us to increase the true positive rate to 1 while maintaining the false positive rate at 0. This also means the larger the area under the curve (AUC), the better the classifier is. By referring to **Figure 4**, it would be clearer what this means.

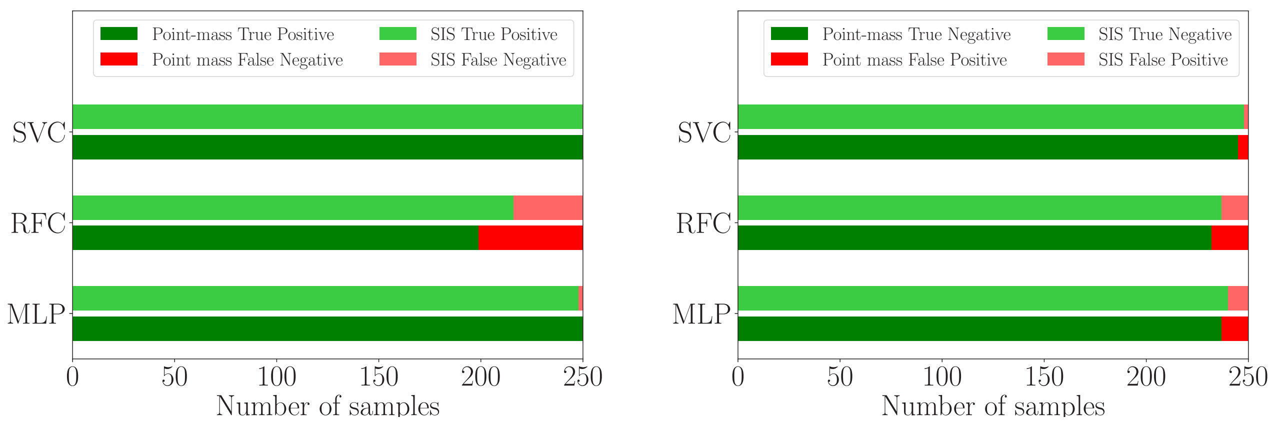
RESULTS

After training and testing each classifier, the output data that we obtain includes a confusion matrix, a classification report, and an ROC curve for both the point-mass lens and SIS models. The number of true positives, true negatives, false positives, and false negatives of each classification test are represented in a confusion matrix. This information is then presented in a bar chart (**Figure 3**) to allow us to compare the performances of each classifier. As mentioned previously, we use 25% of our data to test the MLCs, meaning that the testing data for each lensing model includes 250 unlensed samples and 250 lensed samples. In can be seen that the SVC had a false negative rate of zero when classifying data under both lensing models, as it was able to correctly identify all lensed samples in both lensing cases. The RFC seemed to be able to identify unlensed samples to a good degree of accuracy, but was generally weaker in classifying

Lens Models	Point-mass Lens			SIS		
FOM	Precision	Recall	F_1 score	Precision	Recall	F_1 score
SVC	0.980	1.000	0.990	0.992	1.000	0.996
RFC	0.917	0.796	0.852	0.943	0.864	0.902
MLP	0.951	1.000	0.975	0.961	0.992	0.976

Table 3. Classification report for all three MLCs, including all three figure-of-metrics (FOM). The SVC seemed to perform the best in identifying lensed signals under both lensing models, with F_1 scores of 0.990 and 0.996, while the MLP classifier performed slightly worse, with F_1 scores of 0.951 and 0.975. The performance of the RFC suffered due to low recall ratios of 0.796 and 0.864.

lensed signals as seen from its relatively high false negative and low true positive rates. The MLP classifier performed to a similar degree of accuracy to the SVC, but mistakenly classified two lensed spectrograms as unlensed under the SIS model. Using a larger data set and conducting a longer grid-search to identify the optimal hyper-parameters for each classifier may allow us to further investigate each of their strengths and shortcomings.



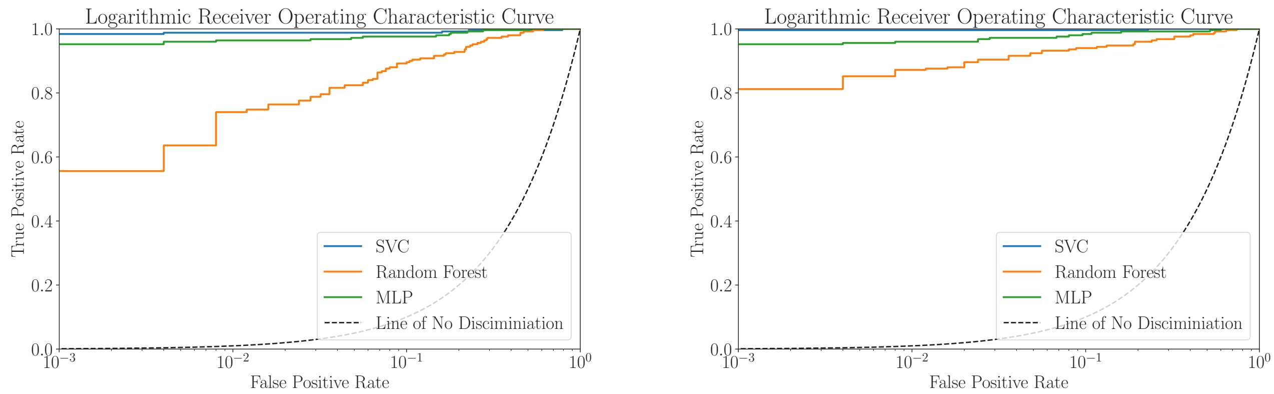
(a) Lensed data **(b)** Unlensed data

Figure 3. Bar charts showing the performance of each MLC in classifying lensed and unlensed waveforms. The charts show the number of correctly classified samples against incorrectly classified samples for data from each lensing model. The RFC seems to be the most prone in misclassifying both lensed and unlensed waves, as seen from its high false positive and false negative rates for both the point mass and SIS model. The SVC and MLP classifier performed much better, with the SVC having a near overall perfect classification accuracy when classifying data from both lensing models.

The classification report gives the precision, recall, and F_1 scores of all three classifiers after fitting, as shown in **Table 3**. The MLC with the best performance seems to be the SVC, as it had F_1 scores of 0.990 and 0.996 when classifying lensed GW signals under the point-mass lens model and SIS model respectively. The RFC had lower F_1 scores as a result of its weaker recall ratios of 0.796 and 0.864. This was most likely due to its higher false negative rate meaning that the RFC was more likely to classify a lensed signal as unlensed. Again, the high F_1 scores of 0.975 and 0.976 for the MLP classifier show that it was able to identify both lensed and unlensed signals to a high degree of accuracy.

The ROC curve for the SVC (**Figure 4**) shows that we are able to increase the true positive rate (TPR) nearly to 1 while maintaining to a false positive rate (FPR) close to 0 for both lensing models. Similarly, the MLP classifier provides a relatively high TPR and low FPR. However, for the RFC, reducing the FPR leads to a TPR below 0.6 for the point-mass lens model and around 0.8 for the SIS model. This indicates that even if we increase the recall of the classifier, the number of misclassifications would increase.

It can be seen that out of the three MLCs we tested, the SVC is the most accurate in terms of identifying a lensed GW signal. It consistently has the highest precision, recall and F_1 score in both the point-mass lens and SIS model, with scores of over 0.980 in all three metrics. **Figure 3** indicates that the SVC is more likely to misclassify an unlensed sig-



(a) Point-Mass Lens Model **(b)** SIS Model

Figure 4. ROC curves of the three classifiers for both the point-mass lens and SIS lensing model.

Area under the curve (AUC)		
	Point-mass Lens	**SIS**
SVC	0.995	0.999
RFC	0.962	0.976
MLP	0.993	0.993

Table 4. The area under the ROC curve for the three classifiers in the two lensing models. The closer the area is to 1, the better the classifier is for classifying the spectrograms.

nal as lensed than misclassify a lensed signal as unlensed, although this result is statistically insignificant due to the limited data samples and the fact that the number of SVC misclassifications is still comparatively low. On the other hand, the RFC has a higher chance of incorrectly predicting lensed GWs as unlensed.

The ROC curves for both models also indicate that SVC is the most appropriate classifier for our spectrograms under both models due to having a high TPR and a low FPR with the highest AUC (as seen in **Table 4**). The RFC performs the worst because even if we were to increase the recall of the classifier, the number of misclassifications would increase. This means it has poor classification ability which is also reflected in the lowest AUC in both models. This further implies that the SVC is the best classifier tested.

It should be noted that employing a grid-search and training the MLC takes significantly longer on the SVC, while it is the quickest on the RFC due to the nature of the algorithm itself and the fact that we could implement parallel processing when using the RFC. Overall, all three MLCs were able to correctly differentiate between lensed and unlensed GW signals, albeit with varying degrees of accuracy.

DISCUSSION

The use of machine learning classifiers is shown to be viable for identifying lensed GW signals, and can be seen as an alternative to the traditional matched filtering techniques. However, there are many steps we could take to further improve our study and make our findings more rigorous.

One option is to use stricter assumptions in the generation of the GW signal. As we are currently using 0.5 PN order waveforms, we could instead consider using higher PN order waveforms provided by the PyCBC library to improve the accuracy of our GW waveforms. Furthermore, an alternative to using pure Gaussian noise as the signal background is to use the noise power spectral densities provided by LIGO. This would allow us to generate more realistic noise waveforms, as the characteristic noise profile of the LIGO detector has greater contributions at low and high frequencies, due to seismic noise and photon shot noise respectively.[26, 27] Additionally, the SNR of our generated waveforms has a mean of 41, which is arguably high when compared to the recent GW detections. The binary source and lensing parameters may be fine-tuned in the future to reflect a more physically realistic GW source and lens mass, and ultimately produce GW signals with a lower SNR. By reducing the the SNR values of the data to approximately 30 or below, we will be able to test the limits of the MLCs' capability to identify lensed signals when the GW waveform is overwhelmed by noise.

Due to the nature of grid-search processes and the complexity of MLC algorithms, computational and time limitations restricted our ability to use a larger dataset in this study. It may be argued that the sample size we used for the training and testing processes for each classifier was too small to generate statistically significant results. We are considering ways to implement a larger dataset with more spectrogram samples while reducing the size of the data and maintaining a high classification accuracy. Possible methods include principal component analysis and t-distributed stochastic neighbor embedding. Another point of consideration is that we only completed a short grid-search for our hyperparameter selection process for each MLC. If we had performed a more rigorous and exhaustive grid-search, more optimal hyperparameters could be determined and we would be able to further investigate the extent of the MLCs' capabilities.

The implementation of alternative lensing models and MLCs to further test the validity of using machine learning to identify lensed waveforms may also be considered in the future. A potential idea for future investigations may be to use machine learning to not only identify lensed signals, but to predict the properties of the GW source and lens mass associated with a given lensed GW signal through parameter estimation by employing regression algorithms in machine learning.

CONCLUSIONS

In summary, machine learning classifiers were able to identify lensed GW signals (with mean SNR = 41) to a relatively high degree of accuracy. This suggests that MLCs can be considered as a viable alternative to matched filtering techniques in the search for lensing events. Further investigations will be required to test the validity and reliability of the use of machine learning in classifying lensed GW signals, and to understand the limitations of this approach to a greater degree. If subsequent research supports the validity of this method, a possible facet to explore in the future would be to use machine learning on a given GW signal to study and extract useful information regarding the associated GW source and lens mass.

ACKNOWLEDGEMENTS

This project and the students were supported by the 2018 Summer Research Program of the Department of Physics, The Chinese University of Hong Kong and was also partially supported by a grant from the Research Grants Council of the Hong Kong (Project No. CUHK 14306218) and the Direct Grant for Research from the Research Committee of the Chinese University of Hong Kong.

REFERENCES

1. Abbott, B. P., Abbott, R., Abbott, T. D., Abernathy, M. R., Acernese, F., Ackley, K., Adams, C. *et al.* (2016). Observation of gravitational waves from a binary black hole merger. *Phys. Rev. Lett.*, 116:061102. *https://doi.org/10.1103/PhysRevLett.116.061102*
2. Abbott, B. P., Abbott, R., Abbott, T. D., Abernathy, M. R., Acernese, F., Ackley, K., Adams, C. *et al.* (2016). GW151226: Observation of gravitational waves from a 22-solar-mass binary black hole coalescence. *Phys. Rev. Lett.*, 116:241103. *https://doi.org/10.1103/PhysRevLett.116.241103*
3. Abbott, B. P., Abbott, R., Abbott, T. D., Acernese, F., Ackley, K., Adams, C., Adams, T. *et al.* (2017). GW170104: Observation of a 50-solar-mass binary black hole coalescence at redshift 0.2. *Phys. Rev. Lett.*, 118:221101. *https://doi.org/10.1103/PhysRevLett.118.221101*
4. Abbott, B. P., Abbott, R., Abbott, T. D., Acernese, F., Ackley, K., Adams, C., Adams, T. *et al.* (2017). GW170608:

Observation of a 19 solar-mass binary black hole coalescence. *Astrophys. J. Lett.*, 851(2):L35. *https://doi.org/10.3847/2041-8213/aa9f0c*

5. Abbott, B. P., Abbott, R., Abbott, T. D., Acernese, F., Ackley, K., Adams, C., Adams, T. *et al.* (2017). GW170814: A three-detector observation of gravitational waves from a binary black hole coalescence. *Phys. Rev. Lett.*, 119:141101. *https://doi.org/10.1103/PhysRevLett.119.141101*
6. Abbott, B. P., Abbott, R., Abbott, T. D., Acernese, F., Ackley, K., Adams, C., Adams, T. *et al.* (2017). GW170817: Observation of gravitational waves from a binary neutron star inspiral. *Phys. Rev. Lett.*, 119:161101. *https://doi.org/10.1103/PhysRevLett.119.161101*
7. Treu, T. and Ellis, R. S. (2014). Gravitational lensing - Einstein's Unfinished Symphony. *Contemporary Physics*, 56.
8. Narayan, R. and Bartelmann, M. (1996). Lectures on gravitational lensing. In *13th Jerusalem Winter School in Theoretical Physics: Formation of Structure in the Universe Jerusalem, Israel*.
9. Takahashi, R. and Nakamura, T. (2003). Wave effects in the gravitational lensing of gravitational waves from chirping binaries. *Astrophys. J.*, 595(2):1039. *https://doi.org/10.1086/377430*
10. Owen, B. J. and Sathyaprakash, B. S. (1999). Matched filtering of gravitational waves from inspiraling compact binaries: computational cost and template placement. *Phys. Rev. D*, 60:022002. *https://doi.org/10.1103/PhysRevD.60.022002*
11. Cutler, C. and Flanagan, E. E. (1994). Gravitational waves from merging compact binaries: How accurately can one extract the binary's parameters from the inspiral waveform? *Phys. Rev. D*, 49:2658–2697. *https://doi.org/10.1103/PhysRevD.49.2658*
12. Cho, H. S. (2015). Search and parameter estimate in gravitational wave data analysis and the fisher matrix. *Journal of the Korean Physical Society*, 66(10):1637–1641. *https://doi.org/10.3938/jkps.66.1637*
13. Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3):273–297. *https://doi.org/10.1007/BF00994018*
14. Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32. *https://doi.org/10.1023/A:1010933404324*
15. Bishop, C. (2006). *Pattern Recognition and Machine Learning*. Springer-Verlag New York, United States of America.
16. Chinchor, N. (1992). Muc-4 evaluation metrics. In *Proceedings of the 4th Conference on Message Understanding*, MUC4 '92, pages 22–29, Stroudsburg, PA, USA. Association for Computational Linguistics. *https://doi.org/10.3115/1072064.1072067*
17. Ohanian, H. C. (1974). On the focusing of gravitational radiation. *Int. J. Theor. Phys.*, 9(6):425–437. *https://doi.org/10.1007/BF01810927*
18. Bliokh, P. V. and Minakov, A. A. (1975). Diffraction of light and lens effect of the stellar gravitation field. *Ap&SS*, 34:L7–L9. *https://doi.org/10.1007/BF00644818*
19. Bontz, R. J. and Haugan, M. P. (1981). A diffraction limit on the gravitational lens effect. *Ap&SS*, 78:199–210. *https://doi.org/10.1007/BF00654034*
20. Thorne, K. S. (1983). The theory of gravitational radiation: An introductory review. Amsterdam: North Holland.
21. Deguchi, S. and Watson, W. D. (1986). Diffraction in gravitational lensing for compact objects of low mass. *Astrophys. J.*, 307:30–37. *https://doi.org/10.1086/164389*
22. Schneider, P., Ehlers, J. and Falco, E. E. (1992). *Gravitational Lenses*. Springer, New York. *https://doi.org/10.1007/978-3-662-03758-4*
23. Nakamura, T. T. and Deguchi, S. (1999). Wave Optics in Gravitational Lensing. *Prog. Theor. Phys. Suppl.*, 133:137–153. *https://doi.org/10.1143/PTPS.133.137*
24. Géron, A. (2007). *Hands-On Machine Learning with Scikit-Learn and TensorFlow*. O'Reilly Media, United States of America.
25. Creighton, J. D. E. and Anderson, W. G (2011). *Gravitational-wave physics and astronomy: An introduction to theory, experiment and data analysis*. Wiley, Weinheim. *https://doi.org/10.1002/9783527636037*

26. Abramovici, A., Althouse, W. E., Drever, R. W. P., Gürsel, Y., Kawamura, S., Raab, F. J., Shoemaker, D. *et al.* (1992). LIGO: The laser interferometer gravitational-wave observatory. *Science*, 256(5055):325–333. *https://doi.org/10.1126/science.256.5055.325*

27. Abbott, B. P., Abbott, R., Adhikari, R., Ajith, P., Allen, B., Allen, G., Amin, R. S. *et al.* (2009). LIGO: The laser interferometer gravitational-wave observatory. *Rept. Prog. Phys.*, 72:076901. *https://doi.org/10.1088/0034-4885/72/7/076901*

ABOUT THE STUDENT AUTHORS

At the time of writing, Amit Jit Singh began his second year of undergraduate study at the Chinese University of Hong Kong where he is undertaking a BSc in Physics with plans to minor in Computer Science. He intends to attain a Ph.D. in Physics with a focus on data science.

Ivan S. C. Li will graduate in 2020 with an MSci in Physics at Imperial College London, and he hopes to pursue a career in data science. He undertook this research project at the Chinese University of Hong Kong as a summer intern.

PRESS SUMMARY

With the recent gravitational wave detections in the past few years, plenty of research has been focused on methods of examining and analyzing the waveforms to extract useful data. An interesting phenomenon that physicists study is gravitational lensing, which is where gravitational or electromagnetic wave signals can be lensed when they pass by massive astronomical objects. If we are able to identify gravitationally lensed signals, we may be able to learn more about the gravitational wave source and the massive object causing the lensing effect. Traditional methods of identifying lensed signals require highly precise and sophisticated lensing models, which we currently do not have. Our research aims to show that machine learning techniques pose as a possible alternative for performing this task. Through our study, we demonstrate that machine learning classifiers are able to differentiate between gravitationally lensed and unlensed gravitational-wave detections to a relatively high degree of accuracy, which suggests that machine learning techniques should be more frequently considered in gravitational wave analysis. However, further research is required to determine the extent at which machine learning is able to accurately extract useful information regarding the physical parameters of the gravitational wave source and lensing object.

Cost, Quality, and Access of Healthcare in Piura, Peru

Julia B. Griffin & Elaina F. Osterbur*

Department of Clinical Health Sciences, Saint Louis University, St. Louis, MO

https://doi.org/10.33697/ajur.2019.020

*Student: julia.griffin@slu.edu**
Mentor: elaina.osterbur@health.slu.edu

ABSTRACT
The aim of the study is to investigate the patient perceptions on the cost, quality, and access of health care services in Piura, Peru. Although one of the largest cities in Peru, Piura has one of the lowest densities of health care workers in the country which greatly impacts the population's ability to receive medical treatment. Lack of financial resources and health literacy, among other health disparities exist. Modeled after *CAHPS® Health Plan Adult Commercial Survey 5.0* and the *Patient Satisfaction Survey*, a forty-four question English and Spanish survey was created with questions to study healthcare variables. As a correlational study with convenience sampling, the survey was administered to both patients and medical providers in eight city health centers. Over a period of twelve days, 107 surveys were collected. After eliminating subjects who did not meet the study criteria, 92 patients and 13 medical providers were included in the study. Findings from medical providers are not reported because of the small sample size. The results of this study suggests that 32% of subjects do not have health insurance, 24% of subjects rated their healthcare received as average, 18% of participants rated their healthcare as the best possible on a scale of zero to ten, and 29% of subjects had to wait an average of seven days for access to healthcare services when care is urgent. The results of this analysis can be used to better understand the Peruvian healthcare system and educate the Piura community and the Parish Santísimo Sacramento as they continue to improve and expand their health care services.

KEYWORDS
Cost; Quality; Access; Healthcare; Piura; Peru; Satisfaction; Parroquia Santísimo Sacramento; EsSalud; SIS; MINSA

INTRODUCTION

Peru is divided into 25 regions with a total population of 31 million citizens.[1-2] With 34% of the population concentrated in Lima, the remaining 66% of the population reside in other urban and rural areas; three-fourths of the non-Lima residents live in rural areas.[2] Nearly 50% of those who live in rural villages live in poverty.[2] With large socioeconomic and social differences existing between rich and poor groups, the healthcare system of Peru has many social and cultural gaps in its infrastructure which affect access to and quality of health services for its citizens.[1-6]

Peru has a decentralized healthcare system administered by five entities: Ministry of Health (MINSA), EsSalud, Armed Forces, National Police, and the private sector. The Peruvian healthcare system is characterized by fragmentation.[1,4,7] This is evidenced by having a means tested (MINSA), employment-based (EsSalud), military/civil servant system (Armed Forces and National Police), and private sector systems.[1-7] However, there are a number of individuals who do not qualify for SIS, nor do they receive health insurance through an employer and cannot afford private insurance. These individuals represent the gap in the Peruvian health care system and are therefore labeled "No Insurance".[8-9] Similar to many healthcare systems (such as the United States), this fragmentation can lead to inefficiency, ineffectiveness, inequality, commoditization, commercialization, de-professionalization, depersonalization, despair, and discord. As a result, patient's needs may be unmet or mistreated.[10] With high economic disparities, the fragmented health system in Peru is heavily based on one's ability to pay for services.[8-10]

Including both public and private sectors, each system operates independently and attends to its respective populations, with its own rules and network of medical providers.[3,6,11]

MINSA and SIS

The Ministry of Health (MINSA) is the maximum health authority and governing body of the national health system. Directing and managing the country's national health policy, it provides public health preventative and curative services to the greatest part of the population. MINSA provides health insurance; Seguro Integral de Salud (SIS), to about 60% of Peru's residents.[2-3] Similar to Medicaid in the United States, SIS protects the country's vulnerable population, those without health insurance, and individuals in moderate and extreme poverty. Most beneficiaries receive subsidized insurance, do not have to pay for care in MINSA establishments, and obtain medications through MINSA pharmacies free of charge.[2-3] As of 2011, about 60% of citizens are insured through SIS, and about 60% of beneficiaries live in rural zones.[2-3,6]

The MINSA infrastructure functions on a three level system with Level I health posts (puestos de salud) providing the most basic care. While these rural clinics emphasize preventative health measures, Level II health centers (centros de salud) see cases that are more complicated. Level III institutions, hospitals, treat the most complex cases.[3,6] In 2009, there were 3.1 health establishments per 10,000 people in Peru (0.2 hospitals, 0.8 centros de salud, and 2.1 puestos de salud per 10,000 people).[3] **Table 1** displays the number of health establishments by region in Peru according to 2009 data.[3]

Department	**Total**		**Hospitals**		**Health Centers**		**Health Posts**		**Population**	
	Number	%	Number	%	Number	%	Number	%	Number	%
Amazonas	491	5.5	8	1.7	65	2.8	418	6.8	411,043	1.4
Áncash	456	5.1	22	4.7	87	3.7	347	5.6	1,109,849	3.8
Apurímac	334	3.7	9	1.9	54	2.3	271	4.4	444,202	1.5
Arequipa	331	3.7	18	3.8	121	5.2	192	3.1	1,205,317	4.1
Ayacucho	379	4.2	10	2.1	66	2.8	303	4.9	642,972	2.2
Cajamarca	818	9.1	16	3.4	152	6.5	650	10.5	1,493,159	5.1
Cusco	336	3.8	18	3.8	70	3.0	248	4.0	1,265,827	4.3
Huancavelica	346	3.9	2	0.4	63	2.7	281	4.6	471,720	1.6
Huánuco	272	3.0	5	1.1	59	2.5	208	3.4	819,578	2.8
Ica	183	2.0	13	2.8	73	3.1	97	1.6	739,087	2.5
Junín	514	5.7	18	3.8	91	3.9	405	6.6	1,292,330	4.4
La Libertad	337	3.8	37	7.9	103	4.4	197	3.2	1,725,075	5.9
Lambayeque	223	2.5	22	4.7	64	2.8	137	2.2	1,196,655	4.1
Lima	1,195	13.3	158	33.7	574	24.7	462	7.5	8,981,440	30.8
Loreto	375	4.2	11	2.3	60	2.6	304	4.9	970,918	3.3
Madre de Dios	119	1.3	3	0.6	18	0.8	98	1.6	117,981	0.4
Moquegua	69	0.8	5	1.1	30	1.3	34	0.6	169,365	0.6
Pasco	275	3.1	9	1.9	40	1.7	226	3.7	290,483	1.0
Piura	458	5.1	25	5.3	113	4.9	320	5.2	1,754,791	6.0
Prov. Const. del Callao	147	1.6	12	2.6	131	5.6	4	0.1	926,788	3.2
Puno	484	5.4	19	4.1	107	4.6	358	5.8	1,340,684	4.6
San Martín	440	4.9	19	4.1	88	3.8	333	5.4	771,021	2.6
Tacna	95	1.1	3	0.6	35	1.5	57	0.9	315,534	1.1
Tumbes	63	0.7	3	0.6	31	1.3	29	0.5	218,017	0.7
Ucayali	216	2.4	4	0.9	26	1.1	186	3.0	458,177	1.6
Total	**8,957**	**100.0**	**469**	**100.0**	**2,321**	**100.0**	**6,165**	**100.0**	**29,132,013**	**100.0**

Table 1. Listing of health institutions (Hospitals, Health Centers, and Health Posts) in Peru by region.[3]

EsSalud

EsSalud offers health services to the employee population and their families and provides health insurance, Seguro Social de Salud, to about 30% of the population.[3] Workers and their successors are insured within the EsSalud network of hospitals and health centers and receive health coverage and benefits of prevention, promotion, recovery, and rehabilitation. Offering three types of insurance, seguro regular, seguro independiente, and seguro potestativo, beneficiaries represent active workers, pensioners, independent workers, and students.[2-3,12] According to the 2012 data, there are about 9.2 million EsSalud users.[11-12]

Armed Forces, National Police, Private Sector

The Armed Forces (las Sanidades de las Fuerzas Armadas, FFAA), National Police (Sanidad de la Policia Nacional de Perú, PNP), and private sector administer care to about 10% of Peru's residents. The Armed Forces and the National Police provide coverage to about 1.3 million individuals; both departments exclusively serve patients in their own health facilities. Institutions within the private sector include private insurers, hospitals (clínicas), private clinics (consultorios medico), and civil society organizations, such as the American Red Cross.[3]

Approximately 73.6% of people aged 0 to 17 years, 57.9% of people aged 18 to 64 years, and 68.3% of older adults (aged 64+ years) have health insurance in Peru.[3] Despite 72.9% of the Peruvian population being covered by some type of health insurance, there remains about 10.8 million residents who do not have insurance.[2-3] Although Peru passed a Universal Health Insurance Law in 2009, there are economic barriers that prevent universal coverage and still divide the Peruvian healthcare system.[2-3,7-9] While Peru claims to offer coverage for all people, there are individuals who still have "No Health Plan/Insurance." There is segmentation in the Peru health insurance plans based on the socio-economic status of the applicant.[8-9] Individuals who are not considered or recognized as poor have a very difficult time applying for insurance through SIS.[8] However, if one is a self-employed, non-dependent worker then one cannot have access to EsSalud.[3,8] If an individual has a limited income, then they may not be able to purchase private health insurance.[8]

Since Peru only spends 5.5% of its GDP on health care services, about 97% of health services funding is directly from the patients.[2-3,11,12] Approximately 40% of the Peruvian population reported to have purchased medicine from the pharmacy, 43% of individuals paid for services from private providers, and 42% of residents financed the whole medical bill.[3] Thus, those with health insurance are more likely to seek medical attention than those without health insurance.[2-3] Beneficiaries of SIS are more likely to use services at puestos de salud and centros de salud, and beneficiaries of other health insurances are more likely to seek care at hospitals and private clinics.[6,14]

If one cannot afford to pay for services, an individual may not seek medical treatment.[6] In fact, about 12.9% of Peruvians reported that they did not receive medical attention because they did not have money; 43.1% of people in moderate or extreme poverty claimed that money was a barrier.[3] In addition to lack of money and medical education, there are many other disparities which contribute to one's health outcomes and decision of soliciting medical treatment.[2-3,14] For example, many avoid medical appointments since the average wait time is 103 minutes.[2] **Table 2** exhibits a variety of reasons of why Peruvians choose to not seek medical attention.

Health Insurance	Lack of Money (%)	Far Away, Lack of Confidence in Medical Professionals, Delay (%)	Use Home Remedies or Self-Medication (%)	It was not necessary (%)	Other Reason to Not Seek Care (%)
No Insurance	51.0	32.7	35.9	41.3	30.9
EsSalud	6.6	19.2	19.6	21.7	27.9
SIS	41.3	45.6	39.8	31.2	34.9
Other	1.1	2.5	4.8	5.8	6.4
Total	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**

Table 2. Percentage of Peruvians, according to insurance type and reason, who do not use medical services.[3]

Contributing greatly to the health disparities in Peru is the extreme deficiency of medical resources in the communities. Many areas lack specialists and technology, and patients must travel to other regions in order to receive treatment.[1-3] In a country of 31 million citizens, there is only one pediatric hospital, which is in Lima.[1-2,6] There are 15 hospital beds, 11.9 doctors, 12.7 nurses, 2.6 obstetricians, and 1.9 dentists per 10,000 people in Peru.[2-3] Additionally, the government restricts physician and nurse working hours to 48 hours per week, which further reduces the availability of health care workers.[2-3,6] If individuals have the means to pay for a private office visit, then they have better quality and access to health care resources.[2-3,6] However, individuals who do not have the ability to pay may wait days to years to schedule appointments or procedures because of the low density of health care workers.[6]

Despite being home to about 1.8 million Peruvians (6% of the population), Piura, Peru has one of the lowest geographic distributions of health care workers.[3,6] Located in northern Peru and bordering Ecuador, Piura has the lowest availability of hospital beds (3.6 per 10,000 people) and has the fewest obstetricians (1.7 per 1,000 people) in the whole country.[1,3] While 78.4% of the region's population has access to sanitation, only 8.1% have access to safe water, and 42% of residents live in moderate or extreme poverty.[3] The principal causes of death in Piura are circulatory system diseases and transmissible diseases. Additionally, Piura has the lowest per capita healthcare expenditure in the country.[2-3]

In Piura, like many other regions in Peru, one's social position, education level, and economic income are important determinants of health outcomes.[3] In fact according to a Peruvian healthcare analysis, an econometric binomial probability model shows that the higher the level of education, the higher probability that the individual has health insurance.[9] A *Piurano's* respective health insurance plan affects the quality of care and where one can receive treatment. One's health literacy impacts whether he or she pursues medical treatment.[6] While improving home infrastructure, increasing water access, and promoting health education are simple measures to enhance health outcomes in Piura, access to health services is significantly linked to the ability of the individuals to pay for them.[3] The inadequate resources, poor coordination between health institutions, and low health literacy contribute to the inequalities and barriers that Piura's citizens face.[3,6] This correlational research project will address the cost, quality, and access of healthcare to better understand patient perceptions about healthcare in Piura, Peru.

METHODS AND PROCEDURES

The aim of the study was to investigate patient perceptions on the cost, quality, and access of healthcare of patients who receive health care services in Piura, Peru.[A] A forty-four question survey, in both English and Spanish, modelled after the *CAHPS® Health Plan Adult Commercial Survey 5.0* and the *Patient Satisfaction Survey* from the Physician Practice Resource Center was developed. In addition to directly addressing cost, quality, and access to healthcare, the questions covered overall patient satisfaction with provider and office staff, the conditions of the clinic, and the frequency of visiting the doctor, hospital, or specialist. The sections include About You, Your Health Care in the Last 12 Months, Getting Health Care from Specialists, Access to Care, Your Health Plan / Cost, and Quality of Care. Requesting a waiver of written consent, a recruitment statement for the project participants, describing the risks and benefits associated with voluntary participation, was developed. The documents were translated to Spanish and verified by two qualified translators.

The survey was administered during Summer 2018 at la Parroquia Santísimo Sacramento in Piura, Peru. The study locations included la Oficina de Enfermería (nurse's office), Oficina Pro-Vida (pro-life office), Clínica Santa Lucia (health clinic), Hospicio Los Ángeles (hospice center), Beata Margarita de Castello (physical and language therapy center), Vida Nueva en Cristo (rehabilitation center), Centro de San Miguel (psychological counseling center), and Casa María (women's shelter). There was no IRB exception for the interviews collected at the women's shelter.

The subject population included patients and health care providers. The following definition was used to identify health care providers as potential participants: "a health care provider is any individual, institution, or agency that provides health care services to health care consumers."[15] A health care provider could include but not be limited to physicians, medical aids, nurses, and physician assistants. **Table 3** provides a view of the number of patients and medical providers surveyed in each study location. It is important to note that a homogenous population was surveyed. All patients surveyed had free clinic access, and all services were free of charge. Thus, even if they might have had a health plan, patients did not have to use health insurance. Participants could be male or female of Mestizo, Amerindian, European, Asian Peruvian, or Afro-Peruvian descent and living in poverty.

Inclusion Criteria:

Male or female patient (≥ 18 to 90 years) who is seeking or receiving treatment in the study locations. Male or female health care provider (≥ 18 to 90 years) who is employed at the study locations.

Exclusion Criteria:

Male or female patient who is younger than 18 years or older than 90 years. Male or female health care provider who is younger than 18 years or older than 90 years.

Location	Number of Patient Participants	Number of Provider Participants
La Oficina de Enfermería	42	4
Oficina Pro-Vida	15	2
Clínica Santa Lucia	12	0
Hospicio Los Ángeles, Vida Nueva en Cristo, Centro de San Miguel, Casa María	15	5
Beata Margarita de Castello	10	2
Total	**94**	**13**

Table 3. The distribution of participants surveyed in each clinic location.

[A] This project received approval by the Institutional Review Board at Saint Louis University (Protocol #29373).

Adapting a cohort study design, a convenience sample was used and subjects were not randomized. Patients were recruited the day of their office visit. Each patient that entered the clinic was asked if he or she would like to voluntarily participate in the project and answer the survey. Providers were recruited the day the survey was administered. Each health care provider, according to **Reference 15**, who worked in the clinic was asked if he or she would like to voluntarily participate in the project and answer the survey. **Table 4** shows the dates that the surveys were administered and the total number of surveys collected each week.

Date of Data Collection	Number of Surveys Collected
7/1/18 – 7/7/18	16
7/8/18 – 7/14/18	44
7/22/18 – 7/28/18	32
7/29/18 – 8/4/18	15
Total	**107**

Table 4. Dates of survey collections in the health clinics and the number of surveys administered according to the respective week.

Each participant completed a consent process.[B] To minimize potential risk, the survey was administered in a private area with only the researcher and the participant in the room. No identifying information was collected. The participant could choose to write the responses or to have the survey orally read and have answers recorded. The subject could skip any question that he or she chose not to answer.

Measures were taken to minimize risk to potentially vulnerable subjects, as there was a non-English speaking and employee population.[C]

Each survey and recruitment statement were coded (Participant 1, Participant 2, etc.) to ensure confidentiality. The surveys and recruitment statements were locked in a private room, and all of the results are confidential. Data was entered to an Excel spreadsheet on a computer with a password known only to the research team. Only the research team had access to the surveys.

Data was collected from patients (N=94) and medical providers (N=13); 107 subjects were recruited. Two study subjects were eliminated due to incomplete and missing information, so the final study subjects included 105 participants (N=92 patients and N=13 medical providers). Due to the small sample size of medical providers, specific information is not reported. Statistical analysis was completed.[D]

[B] Each participant read a recruitment statement that described the risks and benefits associated with voluntary participation in the survey. In the presence of a witness, verbal consent was obtained and documented for each participant in an Excel spreadsheet; participation in the study assumed consent.

[C] The researcher, who is fluent in Spanish, was prepared to manage communications in Spanish with participants during all phases of study participation. Additionally, the materials (survey and recruitment statement) were translated to subject's native language (Spanish) by the researcher and verified by a qualified translator. There were measures taken to minimize risks to employees or potentially vulnerable subjects (healthcare providers). Employees were informed that their decision to participate in the study would not affect performance evaluations, career advancement, or other employment-related decisions made by peers or supervisors. Lastly, the survey was administered in a private room to respect employee privacy and confidentiality.

[D] Using SPSS Statistical Software and Excel, descriptive, parametric, and non-parametric statistical analysis was completed. Cross-tabulation and frequency techniques were utilized to compare and associate certain variables to obtain means, p-value, and Pearson R values. A one-sample t-test was completed and 95% Confidence Intervals were determined.

RESULTS

Although 105 study subjects were surveyed, the data analysis only includes results from the patients (N=92) due to the small sample size of medical providers (N=13). However, **Table 5** and **Table 6** exhibit patient and provider demographics, respectively.

Patient Demographics	N = 92	%
Gender		
Female	67	73
Male	25	27
Age		
18 – 34	28	30
35 – 44	31	34
45 – 54	11	12
55 and Over	22	24
Education Status		
8th Grade or Less	13	14
Some High School	34	37
High School Graduate	19	21
Some College	16	17
4-Year College or Greater	10	11

Table 5. Demographic information collected from patients.

There were more female participants in this study. The descriptive data shown in **Table 5** indicates that 73% of study participants were female. This was not surprising given that the surveys were collected during the day, while most of the men were at work.[6] As per **Table 5**, 72% of study participants had an education level equivalent to a high school graduate or less.

Provider Demographics	N = 13	%
Gender		
Female	8	62
Male	5	38
Age		
25-34	9	69
35 – 44	3	23
45 – 54	1	8
Education Status		
Some College	2	15
4-Year College	3	23
More Than 4-Year College	8	62

Table 6. Demographic information collected from medical providers.

According to **Table 6**, there were only 13 healthcare providers surveyed (eight females and five males). Only two individuals (15%) had an education status of "Some College" and 11 providers (85%) reported that they had an education status of "4-Year College or Greater." This shows the difference in education status of the providers and the patients visiting the health clinics.

Table 7 shows that the majority of providers are in one of the youngest age categories and more educated than the older providers. Except for two providers who did not provide a response, all providers had health insurance.

Provider Age (in Years) and Education	Some College	4-Year College	More Than 4-Year College	Total (N=13)
25-34	1	1	7	9
35-44	1	1	1	3
45-54	0	1	0	1
Total	**2**	**1**	**8**	**13**

Table 7. Medical provider age and education.

Cost
As shown in **Figure 1**, 46 study participants had SIS, 22 individuals had EsSalud, one person had Fondo Seguro Policial (National Police insurance), and two community members had Private Insurance; 34 individuals indicated that they did not have any health plan.

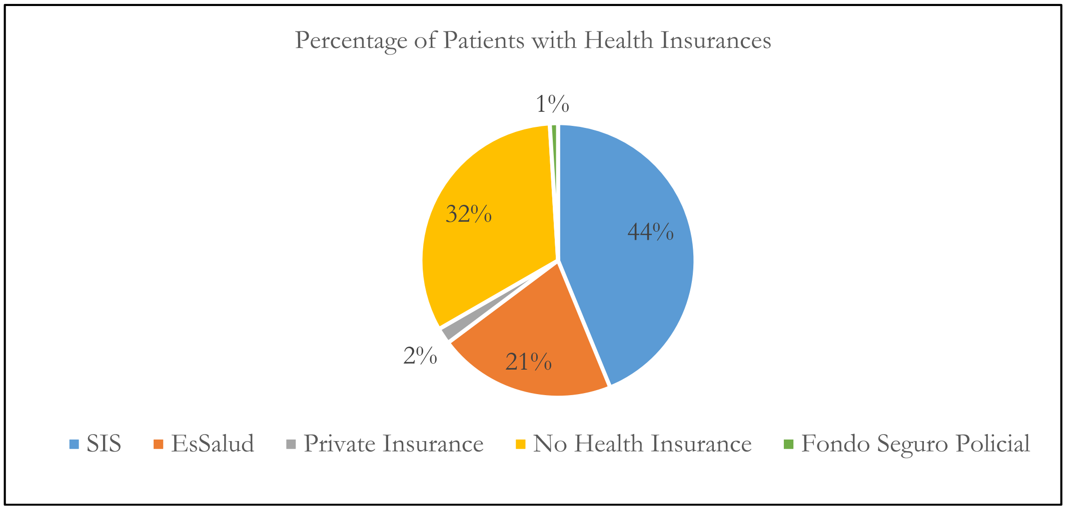


Figure 1. Percentage of patients according to individual health insurances.

Figure 2 displays patients' responses when asked how they would rate their health insurance on a scale of zero to ten, with zero being the worst health plan possible and ten being the best health plan possible. Interestingly, about 27% of participants rated their health insurance plan as a five on the scale; 19 participants rated SIS as a five on the scale. Beneficiaries of SIS tended to rate their health plan lower on the scale, while beneficiaries of EsSalud more frequently rated their insurance plan higher on the scale.

In **Figure 2**, note that "No Health Plan" scores were rated 5, 8, 9, and 10. One would expect these participants to give a zero rating. Certain participants indicated that they had "No Health Plan," but then they also rated a health plan. They answered both questions even though they should not have. In addition, individuals may have become confused about how to answer because "No Health Insurance" may have been correlated as utilizing the free clinic access and parish services; thus, subjects rated their responses based on clinic treatment rather than experiences at public or private health institutions.

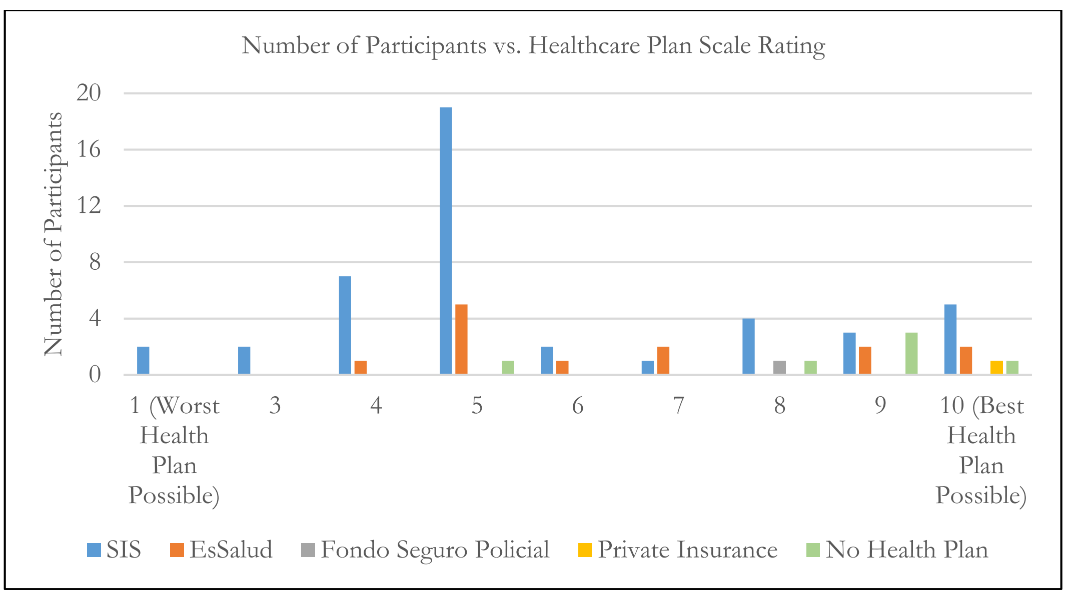


Figure 2. Study participants' ratings of their respective health insurance plan on a scale of zero to ten.

Participants were questioned whether they were satisfied with their health insurance provider. While 32% of patients did not have a health insurance provider, per **Figure 1**, about 32% of patients indicated that they were happy with their health plan, and 33% stated that they were not content with their health insurance, as shown in **Figure 3.**

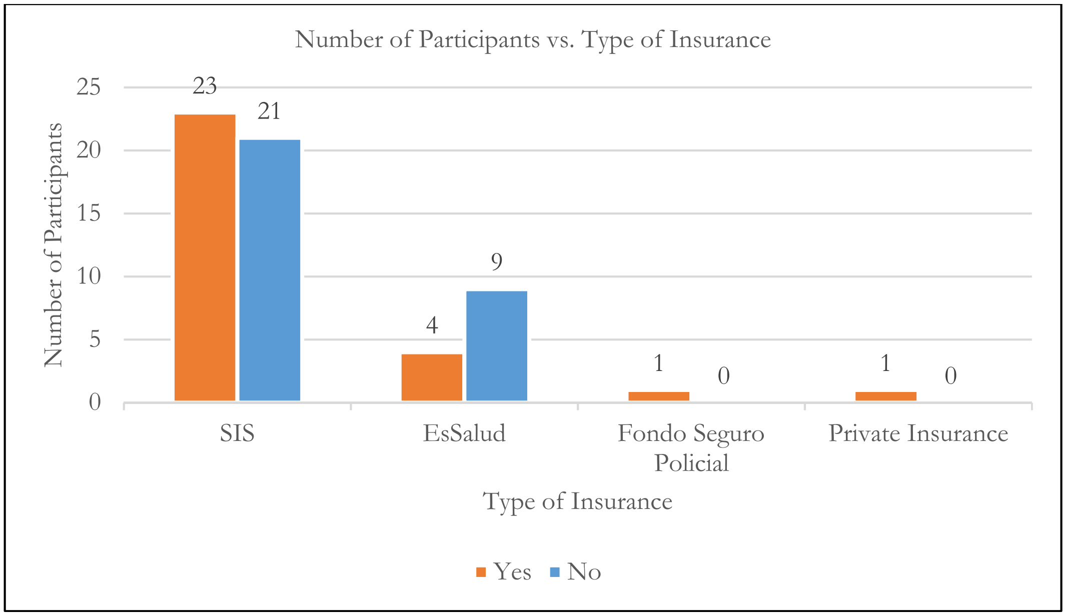


Figure 3. Patient responses when asked if they were satisfied with their health insurance provider.

Quality

Figure 4 demonstrates the diverse patient responses when they were asked to rate the healthcare that they have received in the last 12 months, with zero being the worst healthcare possible and ten being the best healthcare possible. While 24% indicated an average rating of healthcare received (a score of five), 19% stated that they have received the best healthcare possible in the past twelve months (a score of ten). Responses may be varied because individuals seek health treatment at different health institutions.

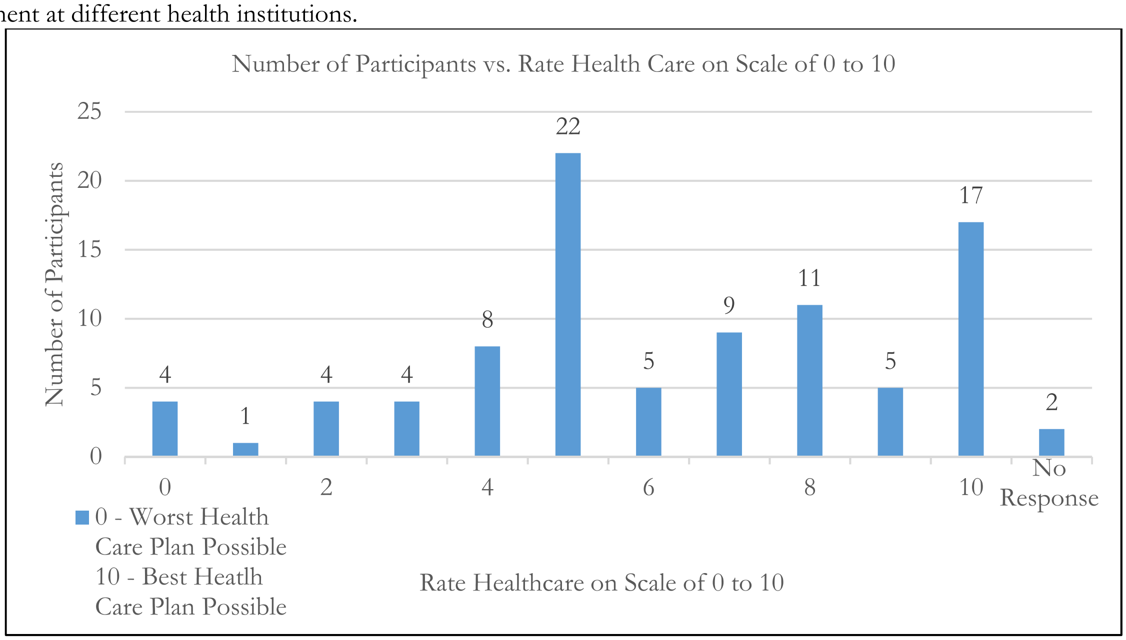


Figure 4. Participant's ranking, on a scale of one to ten, of all healthcare received in the last 12 months.

All patients were surveyed using eleven questions about their perceptions of the quality of care received. Answers may be surprising for a variety of reasons such as never receiving any other quality of medical attention and knowing no differently, lack of medical education, financial barriers, or location of medical services. Answers are fascinating because 77% of subjects reported their quality of care to be a five or better, and 19% of responders perceive their care as "Excellent." This is interesting because these responses represent an area where care would be thought to have poor satisfaction overall.

As explained in the *Cost* section, participants were asked if they were happy with their health insurance provider. Among the 33% of patients that were not satisfied with their health plan, they were asked to describe their reason why they were not content. **Table 8** shows the patient responses.

Limitation of Physicians and Specialists	30%
There is No Medicine	24%
Wait Time	30%
Poor Quality of Medical Attention	38%
Negative Perception of Physician	15%

Table 8. Percentages of patient explanations as to why they were not satisfied with their health insurance provider.

Access

Health disparities such as basic medical education, ability to pay, coverage of health insurance, and transportation to appointments are just a few ways that access to healthcare may be limited. Access to healthcare is closely linked to one's ability to pay for medical treatment.[3,6,8-10] If an individual cannot pay for services at a private office, one must wait to receive attention from the health establishments according to the respective insurance provider.[3,6] Peruvians wait days to years to complete medical testing, operations, or appointments.[6] It is notable that 29% of patients report to waiting more than seven days to receive medical attention when care is "urgent." Additionally, 15% of individuals indicated that they wait more than 30 days for a routine appointment. **Figures 5 and 6** demonstrate the number of days citizens in Piura must wait for medical care when it is urgent or for a routine appointment.

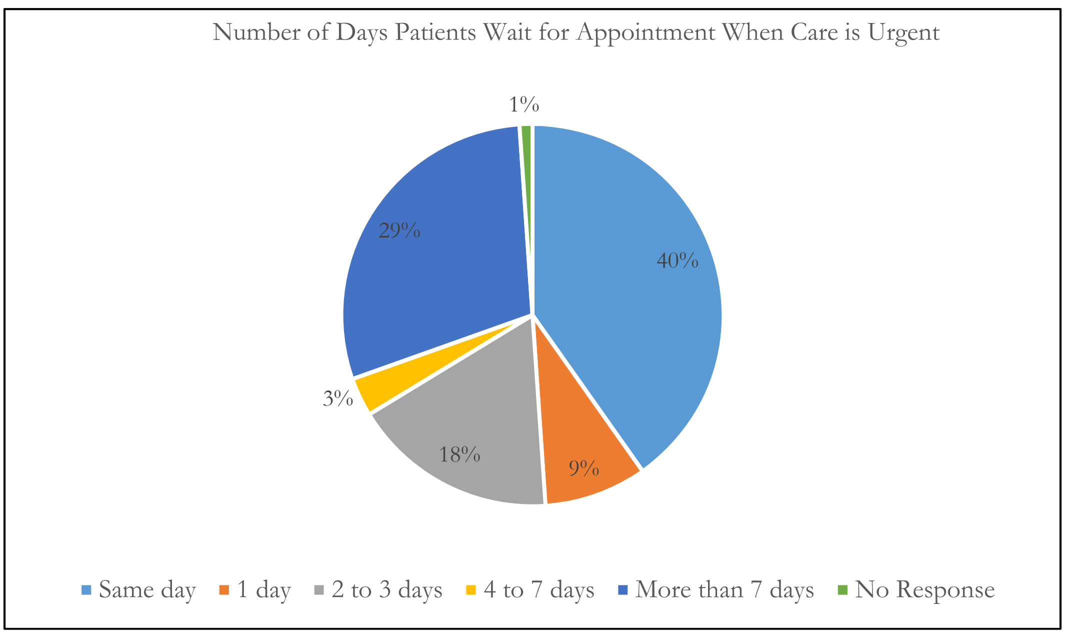


Figure 5. Number of days patients wait to receive urgent care.

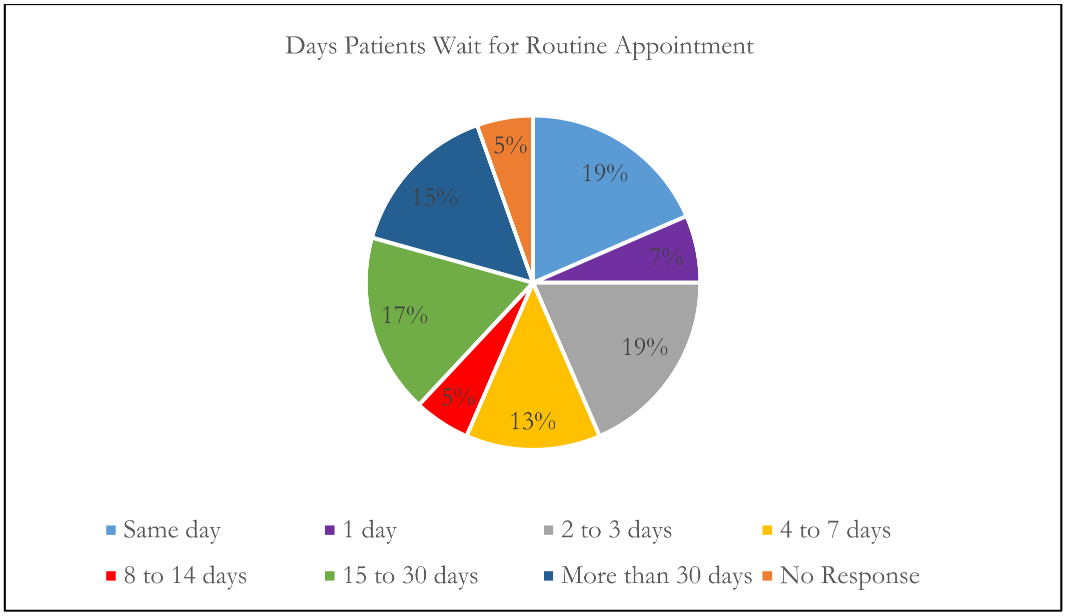


Figure 6. Number of days patients wait for routine appointment.

Figures 5 and 6, participants rated their overall health as "Good," "Fair," or "Poor." Ninety-three percent of patients rated their health as "Good," "Fair" or "Poor" while six participants rated their overall health as "Very good." Zero subjects suggested their overall health as "Excellent".

DISCUSSION

The primary purpose of this project was to examine patient's perceptions of healthcare in Piura, Peru and to understand the relationship between cost, quality, and access to healthcare.

Cost

There is a strong and significant relationship between the respective health insurance plan and ratings of their health plans (on a scale of zero to ten, zero = worst plan to ten = best plan) (p-value ≤ 0.01). As shown in **Figure 2**, 67% of those with SIS insurance rated their plan as a five or lower. Thus, 33% of SIS subjects rated their plan six or more. Forty-six percent of EsSalud subjects rated their plan as a five or less, and 64% of EsSalud subjects rated their health plan as six or more. SIS subjects did tend to rate their health plan lower. 100% of individuals who have Fondo Seguro Policial, insurance for individuals associated with the National Police, or private insurance indicated that they had the best plan possible. The results may be biased because of the small proportion of EsSalud participants or the significant proportion of those subjects who decided to not answer the question (28%). Additionally, there is a strong positive relationship between subjects' satisfaction with the health coverage and insurance provider (p-value ≤ 0.01), as demonstrated in **Figure 3**. Those who do not live in poverty (not be a beneficiary of SIS) are more likely to have greater satisfaction with their health insurance plan (see **Table 9**).

Cross-Tabulated Variables	95% CI	P-value	Pearson R
Rate Health Plan vs. Name of Health Plan	(0.76, 0.89)	0.000	0.84
Satisfaction with Health Insurance Provider vs. Name of Health Plan	(0.63, 0.82)	0.000	0.74

Table 9. Statistical analysis of Cost variables.

Approximately ten subjects commented that they rated their health plan higher than it actually was because they felt blessed to be receiving some health coverage, as opposed to none. Additionally, subjects indicated a higher satisfaction with their provider because they do not know the amount of coverage or access to care that other insurance plans offer. Lack of medical education and financial resources may contribute to the subject's perceptions of health insurance coverage.[2-3,6,8-10, 14]

Quality

According to **Figure 4**, 47% subjects reported they received average to the worst healthcare possible (scores of zero to five) in the past twelve months. As exhibited in **Table 10**, the variables of Quality and Convenience, Accessibility, etc. are being correlated; the quality of care in Piura is significant (p-values ≤ 0.01).

Quality of Care You Received	95% CI	P-value	Pearson's R
Convenience	(-0.04, 0.36)	0.119	0.16
Accessibility	(-0.01, 0.39)	0.058	0.20
Reception Area	(0.52, 0.76)	0.000	0.65
Comfort of Exam Room	(-0.03, 0.47)	0.097	0.17
Cleanliness	(0.14, 0.50)	0.001	0.33
Telephone Promptness	(-0.01, 0.38)	0.067	0.19
Courtesy of the Staff	(0.57, 0.78)	0.000	0.69
Time Spent with Physician	(0.56, 0.78)	0.000	0.68
Communication with Physician	(0.69, 0.86)	0.000	0.79
Privacy of Exam Room and Consult Room	(0.55, 0.78)	0.000	0.68
Overall Experience	(0.53, 0.77)	0.000	0.67

Table 10. Correlation between Quality of Care and Quality variables.

As reported in **Table 8**, 38% of subjects directly report poor quality of medical attention. When only 5.5% of the country's GDP is spent on healthcare, Peru lacks certain medical technology, and there is a deficiency of specialists and medical providers.[2-3, 11, 13] It is not uncommon to find small, unclean exam rooms in public clinics and hospitals along with missing toilet seats, chipped paint, heaps of trash, and broken floor tiles.[6] There were 75% of individuals who reported that the promptness with which professionals answer the phones, if ever, is "Fair" or "Poor." In fact, since Peru is still a developing country, many stated that the puestos de salud, centros de salud, and clinics do not even have telephones.[6] Again, if one can pay for care in a private clinic or has private health insurance, care will be quicker and of higher quality, and there is more likely to be greater patient satisfaction.[6]

Access

The data reported in **Table 11**, suggests a statistically significant relationship between how individuals rate their healthcare and their access to care (p-values ≤ 0.01). As reported in **Tables 2 and 8** and explained in **References 2, 3, 8, 9, and 14**, health disparities, such as health literacy and financial resources, may affect health outcomes and access to services.

Access to Care in the Last 12 Months	Mean	95% CI	T	P-value	Interpretation
Days Wait for Appointment Urgent	2.79	(2.14, 3.17)	14.78 (df=91)	0.000	2.79 = 1 to 3 Days
Days Wait for Appointment Routine	4.26	(3.77, 4.75)	17.30 (df=91)	0.000	4.26 = 4 to 7 Days
How Often Get Care During Holidays	2.26	(1.98, 2.54)	16.07 (df=91)	0.000	2.26 = Sometimes
How Often Get Care After Hours	1.93	(1.65, 2.22)	13.44 (df=91)	0.000	1.93 = Sometimes
How Often Wait Time Within 15 Minutes	1.90	(1.65, 2.16)	14.92 (df=91)	0.000	1.90 = Sometimes
How Often Get Answers to Medical Question	2.84	(2.58, 3.09)	21.90 (df=91)	0.000	2.84 = Mostly Usually
Routine Care as Soon as Needed	2.86	(2.50, 3.21)	15.77 (df=91)	0.000	2.86 = Mostly Usually
Ease of Getting Care	2.29	(2.12, 2.47)	26.42 (df=91)	0.000	2.29 = Sometimes

Table 11. Correlation between Access to Care in the Last 12 Months and Access variables.

Access to healthcare is not easy for many people according to the **Figures 5 and 6,** which exemplify the difficulties of obtaining medical appointments. Twenty-nine percent of patients who needed urgent care wait more than seven days for attention while 32% of patients indicated that they often wait 15 or more days to be seen for a routine appointment. The limited number of specialists in Piura and low number of hours that physicians work make it difficult to receive attention.[2-3,6]

According to **Reference 6**, each health establishment operates independently with its own set of rules. In order to obtain a medical appointment at some institutions, an individual must stand in line early in the morning (such as 3:00 AM) to try to procure an appointment slot for the same day. Appointments for respective specialties are distributed (usually around 5:30 AM). If the individual does not receive a time spot, the only way to obtain an appointment with the respective health insurance provider is to try the following day. Thus, some individuals attempt for days and weeks to acquire appointments. If financially able, one could obtain attention immediately in a private clinic. However, this method is the only way to obtain medical appointments at some health establishments.[6]

Gender, Education, and Age

To view the relationship between gender, education, and age and health plans the following linear models were created and correlated in **Table 12**.

Formula = HEALTHPLAN ~ GENDER	p-value: 0.2227
Formula = HEALTHPLAN ~ EDUCATION	p-value: 0.4566
Formula = HEALTHPLAN ~ AGE	p-value: 0.4851
Formula = HEALTHPLAN ~ AGE + GENDER	p-value: 0.4119
Formula = HEALTHPLAN ~ AGE + GENDER + EDUCATION	p-value: 0.4997

Table 12. P-values of linear models crossed with certain variables.

According to **Table 12,** there is no significant association in any of these findings. Intuitively, education, age and gender should have some bearing on the ability to afford health insurance.[9] However it should be noted that 72% of the sample had less than a high school education. The sample is also biased toward women. In the future, adding the variables "income" and "employment status" might assist in providing a better view of the data.

Based on **Table 12**, there are a variety of observations to be made. All of the clinics were "free clinics" that were either more accessible and/or more affordable. Most of the subjects were either SIS qualified (45) or had No Insurance (32). Based on the Health plan status it can be assumed that all participants are in the low to low-end of the middle-income bracket.

Furthermore, **Table 13** shows the correlation between Education and Health plan.

HEALTHPLAN	<8TH Grade	Some High School	High School Graduate	Some College	>4-Year College	Total
SIS	8	18	10	6	3	45
EsSalud	1	3	3	5	1	13
Armed Forces / National Police	0	0	0	1	0	1
Private Insurance	0	0	0	0	1	1
No Insurance	4	13	6	4	5	32
Total	**13**	**34**	**19**	**16**	**10**	**92**

Table 13. The correlation between Education and Health plan variables.

In a test for independence between EDUCATION and HEALTHPLAN the p-value is 0.1956; thus, there is no connection between EDUCATION and HEALTHPLAN. In a test for correlation, the p-value is 0.4566; there is no significance between Education and Health plan. According to the variance, less than 1% of the variance of EDUCATION can be accounted for by HEALTHPLAN.

Typically, as education increases, so does insurance status.[9] However, as noted in **Table 13** education does not influence the ability to afford insurance.

CONCLUSIONS

This research study provides interesting insights and an understanding of the Peruvian healthcare system. Findings help us comprehend how closely the variables cost, quality, and access to healthcare are intertwined with one's ability to pay for health services in Piura, Peru.[3,6,8-10] The fragmented health care system is navigated based on one's ability to pay for services and can lead to poor quality of patient care.[8-10] While administering the surveys to patients, disparities (such as financial resources, access to transportation, and education levels) became evident among citizens and how these may affect health outcomes or one's level of satisfaction with the services provided. For this reason, there may have been differing views and responses.

This research deviates from minimal, previous research about the Peruvian healthcare system. These cases are specific to the Piura area, where Piura has one of the largest populations but lowest density of healthcare providers.[2-3, 6]

A limitation of this project was the study locations. There could have been selection bias because all patients interviewed at the health facilities of la Parroquia Santísimo Sacramento; all patients were receiving the parish services free of charge. Thus, it is important to note that there may be variability in the responses. Despite instructing the participants to answer the questions as if they were receiving services in other health establishments in Piura, patients could have answered the questions based on the high quality of care that the talented medical professionals provided at the parish.

This sample is highly skewed towards women (+70% of subjects) and results could be considered as potentially biased. Women represent nearly half of the population in Peru, which is not represented in the sample.[3] Thus, the sample is not representative of the gender distribution in the Peruvian population. Additionally, only 28% of the sample had more than a high school education. Usually, education is categorized an explanatory variable or as a determinant to health outcomes. Normally, as education increases, so does insurances status.[9] However, this sample proved differently as education does not influence one's ability to afford insurance, per **Table 13.**

Another limitation was the survey instrument that was used. When creating questions about Access, questions should be included that address obstacles for patients to see a physician: how far individuals live from the nearest provider, transportation costs, the ability to take time off from work, access to childcare, etc. Additionally, the survey may not have used the most current and applicable terminology that the Peruvians understood. In addition, the subjects' responses may be biased if they did not understand certain questions (premium, copay, shared payment, etc.). Thus, this survey, which was adapted from two national surveys designed to be administered in the United States, is a poor tool to measure Peruvians' responses because the healthcare system is significantly different. In future studies, this must be addressed and the survey tool needs to be revised to be more culturally sensitive, ask more specific questions, use more relevant terminology, and address these concerns.

Lastly, there was an error in a sentence overlooked by the IRB on the recruitment statement. On Number 4, there is no loss of confidentiality. Even though it states that confidentiality may be lost, there was no loss of confidentiality.

Opportunities abound to further study the targeted population. In future studies, more questions must be addressed that are specific to women's issues that influence perceptions of healthcare and access to healthcare. A future research project should focus on surveying patients at public and private health institutions in Piura where individuals do pay for health care services. While this project was completed in city and region of Piura, individuals living in other cities (but still in the Piura region) could be surveyed. Future work needs to be conducted to examine the differences between the EsSalud and MINSA health establishments in Piura. Additionally, it would be interesting to investigate the number of specialists in Piura and categorize them based on their work location. Lastly, a future project could study the relationship between a patient's income level and type of health insurance.

This project is important for the Parish community of Santísimo Sacramento because it highlights basic demographic information of the patients populating the parish health clinics. It will help the parish better allocate human and financial resources to its health clinics. Additionally, it will assist the parish in determining what future improvements or additions can be made to benefit the health of the Piura community. This project is beneficial for any aspiring or current medical professional because it provides the opportunity to compare the United States and Peruvian healthcare systems.

APPENDICES

Appendices for this manuscript can be found at http://www.ajuronline.org/ajur-volume-16-issue-2-september-2019/

ACKNOWLEDGEMENTS

The authors thank Dr. Christina Garcia and Ms. Sarah Bauer for reviewing the translated documents. Authors also thank Dr. Jesús García-Martínez and anonymous reviewers for reviewing the IRB application and manuscript. This project was made possible with financial assistance from the University Honors Program Investigative Learning Experiences Grant and the Alpha Epsilon Delta Summer Scholarship Award. Authors are thankful for support from la Parroquia Santísimio Sacramento, the survey participants, and the respective site coordinators.

REFERENCES

1. Central Intelligence Agency (2019) South America: Peru *https://www.cia.gov/library/publications/the-world-factbook/geos/pe.html* (accessed January 2019)
2. Lazo-Gonzales, O., Alcalde-Rabanal, J., Espinosa-Henao, O. (2016) El Sistema de Salud en Perú situación y desafíos (REP, Ed.) 1st ed., 1-84, Colegio Médico del Perú, Lima.
3. Cetrángolo, O., Bertranou, F., Casanova, L., Casalí, P. (2013) El Sistema de Salud del Peru: situación actual y estrategias para orientar la extension de la coberatura contributiva. 1st ed., 1-184, Oficina de la OIT, Lima.
4. Petrera, M., Valdiva, M., Jimenez, E., Almeida, G. (2013) Equity in health and health care in Peru, 2004-2008, *Revista Panamericana De Salud Publica = Pan American Journal of Public Health,* 33(2), 131. *http://doi.org/10.1590/s1020-49892013000200008*
5. World Health Organization (2015) Success Factors for Women's and Children's Health: Peru. *http://www.who.int/pmnch/knowledge/publications/peru_country_report.pdf* (accessed January 2019)
6. V. Garay, personal communication, August 2, 2018.
7. World Health Organization (2019) Peru *http://www.who.int/workforcealliance/countries/per/en/* (accessed January 2019)
8. Francke, P. (2013) UNICO studies series 11 Peru' comprehensive health insurance and new challenges for universal coverage *http://documents.worldbank.org/curated/en/371851468086931725/Perus-comprehensive-health-insurance-and-new-challenges-for-universal-coverage* (accessed August 2019)
9. Seinfeld, J., Montañez, V., & Besich, N. (2013) The health insurance system in Peru: towards a universal health insurance *https://www.academia.edu/33391064/The_Health_Insurance_System_in_Peru_Towards_a_Universal_Health_Insurance* (accessed August 2019)
10. Stange, K (2009) The problem of fragmentation and the need for integrative solutions, *Annals of Family Medicine*, 7(2), 100-103. *http: / / doi.org/10.1370/afm.971*
11. Vammalle, C., Ruiz Rivadeneira, A., James, C., Prieto, L., Montañez, V. (2018) Financing and budgeting practices for health in Peru, *OECD Journal on Budgeting*, 33(2), 25-64. *https://doi.org/10.1787/budget-17-5j8v16g3czth*
12. Government of Peru (2019) Seguro Social del Perú – EsSalud *https://www.gob.pe/194-seguro-social-del-peru-essalud* (accessed January 2019)
13. Shafrin, J. (2018) Health care in Peru *https://www.healthcare-economist.com/2018/02/11/health-care-in-peru/* (accessed January 2019).
14. Ypanaqué-Luyo, P., Martins, M. (2015) Uso de los servicios de salud ambulatorios en la población peruana, *Rev Peru Med Exp Salud Pública,* 32(3), 464-470. *https//doi.org/ 10.17843/rpmesp.2015.323.1675*
15. Health care Provider in *Mosby's Medical Dictionary, 8th ed.* Mosby/Elsevier, St. Louis, MO.
16. Physician Practice Resource Center (2004) Patient satisfaction surveys *http://www.massmed.org/physicians/practice-management/patient-satisfaction-surveys-(pdf) (accessed August 2019)*
17. CAHPS (2012) CAHPS® Health Plan Surveys Adult Commercial Survey 5.0 *https://cahpsdatabase.ahrq.gov/files/HPGuidance/Adult%20Core%20Commercial%20Survey%205.0.pdf (accessed August 2019)*

ABOUT STUDENT AUTHOR

Julia Griffin is a fourth-year student pursuing an Honors Bachelor of Science in Health Sciences, an Honors Bachelor of Art in Spanish, and a minor in Biology at Saint Louis University, St. Louis, MO. Griffin conducted this independent research project in Summer 2018 in Peru. Fluent in Spanish, Griffin has studied the Spanish language for ten years. Additionally, she has spent time in Honduras, Spain, Costa Rica, and the Dominican Republic. She is familiar with the culture and was prepared to manage communications in Spanish with subjects during all phases of study participation. She will graduate in May 2020 and plans to attend medical school.

PRESS SUMMARY

Possessing financial resources affects cost, quality, and access of healthcare in Piura, Peru. One's health literacy and health insurance provider can influence health outcomes and patient satisfaction of health care services. Patient perceptions of cost, quality, and access of healthcare in Piura, Peru were investigated as 92 patients and 13 medical providers were surveyed with a Spanish survey in eight city health centers. Upon completion of the quantitative and qualitative data collection, statistical analysis was performed which yielded relationships among healthcare variables. The data captured will educate the Piura community and the Santísimo Sacramento Parish as they continue to improve health institutions and services.

The Effectiveness of Debt Relief: Assessing the Influence of the HIPC Initiative and MDRI on Tanzania's Health Sector

Fernando Lopez Oggier
Center for American Politics and Public Policy, University of California at Los Angeles, Los Angeles, CA

https://doi.org/10.33697/ajur.2019.021

Student: fernando.lopezoggier@ucla.edu, flopezoggier@gmail.com*
Mentor: desveaux@ucla.edu

ABSTRACT
Debt relief initiatives have been part of the international development sphere since the early 1990s. With the launch of the Heavily Indebted Poor Country (HIPC) Initiative in 1996 and the Multilateral Debt Relief Initiative (MDRI) in 2005 many countries have been able to successfully qualify for debt relief. Tanzania has been one of the primary beneficiaries of debt relief over the years. While empirical evidence demonstrates that the country's economic growth has been positively impacted by debt relief initiatives, other aspects of human development need to be analyzed to ensure a comprehensive assessment of the HIPC Initiative and the MDRI. This study compiles Tanzania's health data into a composite indicator to perform a graphical analysis to compare the trends between health outcomes and external debt. The graphical analysis is contextualized through a qualitative analysis of political, economic and health financing literature from the Bank of Tanzania, UNICEF and USAID. The results indicate that health outcomes improved throughout the whole study's time period particularly after the HIPC Initiative. The health financing literature also points to increased development expenditure during this period. Nonetheless, the effects of debt relief seem to diminish in the long-term due to fluctuations in external donors and logistical barriers to budget execution. Tanzania also continues to face socio-economic and geographic disparities in health outcomes and funding. Some of the literature also states that the country's weak system of checks and balances and the lack of robust institutions could cause opportunistic policy preferences that might not necessarily improve Tanzania's health sector.

KEYWORDS
Child Mortality Rate; Debt Relief; External Debt; Heavily Indebted Poor Country Initiative; International Monetary Fund; Life Expectancy; Maternal Mortality Rate; Multilateral Debt Relief Initiative; Official Development Aid; Prevalence of Undernourishment

INTRODUCTION

Debt relief for heavily indebted countries has been a pressing subject since the 1990s. The World Bank and the International Monetary Fund (IMF) have established two primary debt relief programs over the years: the Heavily Indebted Poor Country (HIPC) Initiative in 1996 and the Multilateral Debt Relief Initiative (MDRI) in 2005. The HIPC Initiative was designed to ensure that the world's poorest countries weren't saddled with unmanageable debt. In order to qualify for debt relief countries implemented a set of socio-economic reforms to diminish poverty and encourage economic growth.[1] The MDRI, established in 2005, is an extension of the HIPC Initiative providing 100% cancellation of multilateral debts owed by HIPC countries to the World Bank, IMF and African Development Bank.[2]

The goal of these initiatives, and debt relief as a whole, is to increase a government's fiscal space. Fiscal space is defined by Peter Heller as "room in a government's budget that allows it to provide resources for a desired purpose without jeopardizing the sustainability of its financial position or the stability of the economy."[3] The fiscal space created by debt relief programs should allow recipient countries to make meaningful progress in various sectors of society, namely infrastructure, health and education. This should in turn increase the country's human development attainment.

As of 2013, 35 countries have reached the "completion point" of these debt relief initiatives. Tanzania has been one of the primary beneficiaries of debt relief over time, qualifying for assistance in 2000 under the HIPC Initiative and 2005 under the MDRI. Various studies have demonstrated that these initiatives have positively impacted Tanzania's economic growth.[4]

However, economic growth and GDP figures are not exact proxies for human development and don't form a comprehensive analysis of a country's socioeconomic progress. While these countries were able to decrease their debt service payments and increase their fiscal space, it's important to assess how well this translated into tangible improvements in society. This assessment can be done through the lens of Tanzania's progression through debt relief. Instead of attempting to evaluate every area that qualifies as "human development", the impacts on the country's health sector can be specifically analyzed in order to narrow the purview of the research. The World Health Organization (WHO) has outlined in several of its reports the fundamental role that better health plays in human development as well as the link between health and poverty reduction.[5] Hence, the question to be addressed is what influence did the HIPC Initiative and the MDRI have on Tanzania's health sector?

Official development aid (ODA) and debt relief are some of the most important sources of development assistance. For most low income countries, ODA and debt relief are their largest foreign financial inflows, especially given their restricted access to international markets. Thus, when large-scale development programs like the HIPC Initiative and the MDRI are established, assessing their effects on different sectors of human development is crucial. These initiatives tend to encourage and foster economic growth, theoretically allowing governments to allocate more funding towards other sectors of society instead of debt service payments. Nonetheless, it's important to evaluate whether this is resulting in meaningful improvements. GDP and economic figures alone cannot map these outcomes. If increased fiscal space through debt relief is not allowing for financial resources to trickle down to fundamental sectors of society, such as the health sector, debt relief isn't really a meaningful form of development assistance. However, if there are substantial improvements in Tanzania's health sector post-debt relief as opposed to the years prior to receiving assistance, then there is a strong argument to be made about the success of debt relief programs.

Background

Debt relief was brought to the international stage in the 1990s through movements like the Jubilee 2000 campaign, an international coalition of NGOs and other organizations that called for the cancellation of debt in developing economies by the year 2000. The campaign was quite successful in pushing debt relief onto the agenda of Western governments and international organizations.[6] Ultimately, the World Bank and the IMF established their own debt relief programs starting with the HIPC Initiative in 1996. However, after receiving criticism that highlighted the lengthy and restrictive requirements to qualify for debt relief, the World Bank and the IMF "enhanced" the HIPC Initiative by loosening the qualifying thresholds in 1999.[4] Tanzania qualified for debt relief under the enhanced HIPC Initiative in 2000 after demonstrating commitment to poverty reduction and macroeconomic stability.[1]

Literature Review

The majority of the literature surrounding the analysis of debt relief, namely the HIPC Initiative and the MDRI, focuses on the economic and fiscal ramifications of such initiatives. Most studies analyze the relationship between debt relief initiatives and economic growth or the fiscal responses that ensue.

Gomera (2014) cites that there is a positive correlation between debt relief and economic growth, stating that this correlation is especially significant amongst primary benefactors of debt relief like Tanzania. The empirical evidence demonstrates that on average debt service payments from the recipient countries have declined by 2% of their GDP. From 1990 – 2010 Tanzania received US$7,573,760,000 in the form of debt relief; during this period Tanzania's annual GDP also grew by an average of almost 6%.[4] Moreover, Gomera further states that debt relief recipients have demonstrated improved debt sustainability and economic growth compared to other low income countries that were not recipients of debt relief assistance.[4]

Other research also determines that debt relief programs allow countries to improve their public financial behavior in a desired way[7] by offering opportunities to improve their investments and domestic resources.[8]

Another study finds that debt relief initiatives increase average income per household and decrease the percentage of people below the poverty line.[9] Moreover, the study also states that "the proportion of income needed to transfer the poor above the poverty line… significantly reduced over the [time] period [analyzed]."[9]

Nonetheless, some literature finds different results. Dijemu (2018) investigated the trends in public investment at the decision and post-completion points of the enhanced HIPC Initiative and MDRI. On one hand the study found that there are increases in public investment of around 2 and 3 percent at the decision and post-completion points of the enhanced HIPC Initiative respectively.[10] However, when it comes to the MDRI, Dijemu doesn't find any impacts on growth or public investment. In addition, this research also didn't find heterogeneous impacts on growth as a result of the enhanced HIPC Initiative and argues that improvements in institutional quality need to be made for debt relief to have meaningful effects on growth and investment.[10] Thus, there is literature that points towards debt relief not having a substantial effect on increasing growth like some of the other studies do.

Additionally, some previous research has also been conducted surrounding the effects of multilateral debt on child health across multiple countries. Welander (2016) assessed the impacts of the HIPC Initiative on infant mortality across 56 countries and found that after debt relief a country's infant mortality rate decreases by an average of 0.5%.[11] This correlates with around 3,000 less infant deaths per heavily indebted poor country. Welander states that the empirical evidence indicates that health outcomes for infants born to poor mothers improve even more after debt relief assistance. The study found a 0.9% drop in deaths amongst infants born to poor mothers. Nonetheless, Welander's results found that there are no statistically significant child health effects from graduating the program and receiving full debt relief.[11]

This study seeks to build on the foundation laid out by the current debt relief literature by taking the assessment of the HIPC Initiative and the MDRI one step further. Thus, this study evaluates whether the improvements in economic growth and fiscal response provided by these debt relief initiatives are resulting in tangible improvements in the health sector as a whole. As stated previously, if positive outcomes in human development aren't present in recipient countries despite increases in fiscal space, these initiatives aren't meaningful and should be amended.

Hypothesis

Theoretically, debt relief initiatives should provide governments with increased fiscal space for investment into social expenditure. Evidently, debt relief isn't the only way for governments to create fiscal space. For instance, governments can raise revenues through tax reforms, re-prioritize expenditure and improve government efficiency which would allow for increased government spending. Nonetheless, debt relief is still a very direct way of creating fiscal space for impoverished countries that are saddled with unmanageable debt. Furthermore, since developing economies are in greater need of increased social expenditure and often lack the ability to respond to fiscal challenges, having more fiscal space is a pressing issue.[12]

Given that increased social expenditure should allow for tangible improvements in Tanzania's health sector, I hypothesize that the HIPC Initiative and the MDRI have facilitated the ability for Tanzania to make considerable advances in health performance. Nevertheless, there are other variables that can affect the country's ability to perform. Inefficient allocation of funds and lack of general guidance need to be taken into account as they can always impact the final outcomes of debt relief programs.

METHODOLOGY AND PROCEDURES

Case Study

As stated previously, Tanzania has been one of the main recipients of debt relief over the years. Tanzania qualified for $3 billion in debt service relief under the enhanced HIPC Initiative after implementing its own set of economic reforms.[13] By 2005, Tanzania had further qualified to receive 100% debt relief under the MDRI which amounted to $297 million.[14] As outlined in the literature review, studies demonstrate that this assistance positively impacted Tanzania's economy. Debt relief was sacrosanct in helping the country fulfil Goal 8D of the Millennium Development Goals relating to debt sustainability and management.[3] Since Tanzania is one of the more successful beneficiaries of debt relief, the country is an exemplary case study to identify whether debt relief programs result in meaningful improvements in human development. Additionally, there is substantial availability of health-related data for Tanzania in order to map out the country's health performance over time.[15]

Data Calculations

The purview of the research spans from 1995 – 2015 in order to analyze Tanzania's health sector performance before and after being impacted by debt relief assistance. In order to assess the effects of the HIPC Initiative and the MDRI on Tanzania's health sector, data from the World Bank and different United Nations agencies is compiled to create a model that acts as a proxy for health performance. This Health Index is a geometric mean of various health statistics to create a composite indicator similar to how the Human Development Index (HDI) is set up. The HDI compiles life expectancy, mean years of schooling, expected years of schooling and GNI per capita to produce a value from 0 – 1. This value is representative of that country's performance in human development. A value closer to 0 represents poorer human development while a value of 1 represent excellent human development attainment. In order to convert the indicators into indices with a score from 0 – 1 minimum and maximum values (goalposts) are established.[16] The maximum values are the highest observed values for that time series. The United Nations Development Programme (2015) states that the rationale behind minimum values is based on the values that society needs to survive over time. Once these values are defined, sub-indices are calculated for each indicator using the following formula:

$$Dimension\ index = \frac{actual\ value - minimum\ value}{maximum\ value - minimum\ value}$$ [17]

This study will compile the following statistics to calculate the Health Index: life expectancy (LE)[A], child mortality ratio per 1,000 births (CMR)[B], maternal mortality ratio per 100,000 births (MMR)[C], and prevalence of undernourishment amongst the population (PoU)[D]. However, the formula for the dimension index presented above that the UN uses to calculate the HDI only works for statistics where a higher value represents a better outcome, such as with life expectancy. When it comes to child mortality, maternal mortality and undernourishment, higher figures represent poorer health sector performance. Hence, the formula must be truncated so that the score of 0 – 1 assigned to each statistic is consistent in mapping out health performance. Therefore, the truncated dimension index formula is as follows:

$$Truncated\ Dimension\ index = 1 - \left(\frac{actual\ value - minimum\ value}{maximum\ value - minimum\ value}\right)$$

The geometric mean is then found by multiplying all the sub-indices together and finding the x root of the product where x is the number of sub-indices used:[17]

$$\sqrt[x]{Product\ of\ x\ subindicies}$$

Hence, the geometric mean for the Health Index is calculated using the following formula:

$$\sqrt[4]{LE\ subindex \times CMR\ subindex \times MMR\ subindex \times PoU\ subindex}$$

Justification for Maximum and Minimum Values

The UNDP states that the minimum value for life expectancy is 20 years because historical evidence indicates that if life expectancy in a society drops below the typical age of reproduction the society would die out.[15] Moreover, from 2014 onwards the maximum value for life expectancy has been 85 according to the UNDP and that's what has been used for this model.[15]

When it comes to a country's maternal mortality ratio per 100,000 births (MMR), the UNDP states that countries where the ratio exceeds 1,000 deaths do not differ in their inability to create supportive conditions for maternal health. Countries with a ratio of 10 or fewer deaths per 100,000 births are performing at a level where any differences in their ratios are random.[17] Hence, the MMR's maximum and minimum values for the dimension index calculations will be 1,000 and 10 respectively.

FAO's prevalence of undernourishment indicator (PoU) measures "the share of the population that has a caloric intake which is insufficient to meet the minimum energy requirements necessary for a given individual".[18] The scale does not include countries below 5% so it can be argued that there's not much difference in a country's ability to feed its population if values drop below 5%. Moreover, the scale goes up to 60% so it is also justifiable that if a country's PoU goes over 60% there's not much difference in its persistence of food insecurity. Thus, 60 and 5 will be the PoU's maximum and minimum values respectively.

Based on the minimum and maximum value rationales by different UN agencies for these statistics, similar justifications can be made for child mortality per 1,000 births (CMR). The CMR has been under 200 deaths in all countries since 1980 according to the UN WPP.[19] Hence, it can be inferred that comparing CMRs above 200 wouldn't indicate much of a difference in a country's health performance. Moreover, in the same way that a MMR below 10 doesn't indicate any substantial dissimilarity in health performance, a CMR below 5 deaths per 1,000 births doesn't make much of a difference in health attainment either. Therefore, the maximum and minimum values for CMR are 200 and 5 respectively.

Graphical Analysis

Using Tanzania's Health Index scores for each year within the stated time frame, the country's health performance can be mapped out. In order to compare the country's performance with the HIPC Initiative and the MDRI, Tanzania's External Debt

[A] The Human Development Index uses life expectancy as one of its subindicies

[B] UN WPP (2017) provides annual data for child mortality rate for all countries from 1950 onwards

[C] The World Bank (2015) provides data for maternal mortality starting in 1990

[D] FAO, IFAD, UNICEF, WFP and WHO (2018) provide undernourishment prevalence data for countries between 5% and 60% undernourishment in billions[E] is mapped out alongside health performance. The health index scores are multiplied by 10 for the graphical analysis so that the highest score is 10 rather than 1. This makes the scale for the health index scores more similar to the scale for external debt (in $bn) which allows for a clearer comparison. The years where HIPC and MDRI were granted are also marked out.

Qualitative Analysis
Finally, in order to take into account Tanzania's socioeconomic and political context, a qualitative analysis using literature from the Tanzanian government, UNICEF, USAID as well as key political scientists and economists is presented. This analysis offers internal and external public finance information surrounding debt management as well as the policies Tanzania has in place to allocate funds. Moreover, the qualitative analysis examines both temporal and fixed factors that affect Tanzania's health expenditure and health outcomes. Finally, overarching explanations for the potential inefficiency of resource allocation are also discussed in the qualitative analysis. Literature from organizations like UNICEF provides a third party perspective on health financing and a more objective analysis that can further contextualize the graphical analysis.

Results and Discussion
Appendix A includes all the statistical data compiled to make the Health Index. Appendices B – E include the results of the calculations for each subindex using the dimension index formula as well as the truncated dimension index formula from the Methodology.

The tables below include Tanzania's Health Index score and External Debt in billions (USD)[20, 21] for the study's time frame. The scores range from 0 – 1. A value closer to 0 represents poor health performance whereas a score closer to 1 represents exemplary health performance.

Year	Health Index Score
1995	0.202
1996	0.223
1997	0.246
1998	0.272
1999	0.301
2000	0.327
2001	0.352
2002	0.378
2003	0.401
2004	0.424
2005	0.449
2006	0.476
2007	0.505
2008	0.529
2009	0.539
2010	0.553
2011	0.568
2012	0.586
2013	0.602
2014	0.617
2015	0.627

Table 1. Health Index Score over time.

Year	External Debt in billions (USD)
1995	8.0
1996	7.5
1997	7.6
1998	7.4
1999	7.7
2000	7.7
2001	6.8
2002	6.8
2003	6.8
2004	6.6
2005	7.3
2006	8.2
2007	4.6
2008	4.4
2009	5.4
2010	6.9
2011	7.6
2012	9.5
2013	11.6
2014	13.8
2015	15.9

Table 2. External Debt in billions (USD) [20, 21] over time.

[E] Found through Bank of Tanzania (2014) and CIA World Factbook (2015) databases

Graphical Analysis
In order to analyze debt service and health performance fluctuations over time, both variables have been mapped in a linear graph. As mentioned previously, the health index scores have been multiplied by 10 in order to facilitate the graphical comparison in **Figure 1.**

Figure 1: Tanzania's External Debt ($bn) and Health Index Score (out of 10) throughout the study's time frame

As depicted in Figure 1, health performance has steadily increased whereas total debt service has wavered but has still considerably decreased since 1995. It is important to note that Tanzania's total debt service was already significantly decreasing before receiving debt relief assistance in 2000. However, from the period following the HIPC Initiative (2000 – 2005) Tanzania's Health Index score increased by 1.22 (0.122) whereas during the pre-debt relief period (1995 – 1999) it increased by 0.99 (0.099), meaning that the increase in health after the HIPC Initiative is slightly higher. Additionally, the period after the MDRI (2005 – 2010) also saw a slightly higher score increase than the pre-debt relief period of 1.04 (0.104). Nevertheless, the long-term post-debt relief period (2011 – 2015) only saw an increase of 0.59 (0.059) in the country's Health Index score. However, Tanzania's external debt has increased quite steadily since 2009, 4 years after the MDRI. This substantial increase in external seems to overlap with the slower improvements in health outcomes. Despite this, Tanzania's Health Index score is still slowly increasing over time.

Qualitative Analysis
The literature on Tanzania's health expenditure can contextualize the graphical analysis by identifying temporal factors that affect how the government is allocating its resources and whether debt relief assistance was significant in impacting health outcomes.

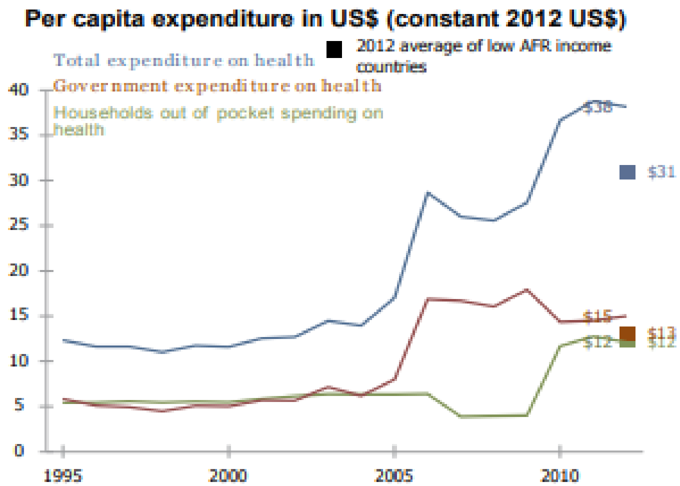


Figure 2: World Health Organization (2014) depicting Tanzania's per capita health expenditure from 1995 – 2012

These temporal factors include: per capita health expenditure trends, levels of foreign donors over time, net ODA received and comparing recurrent versus development spending. Moreover, the research also indicates fixed factors, namely geographic and socioeconomic disparities, that provide health financing challenges which increased fiscal space might not be addressing. Finally, the literature also provides overarching explanations for why African states might allocate resources towards public sectors less effectively.

As part of USAID's Health Policy Project, a five year cooperative to strengthen developing countries' health programming and governance, USAID published Health Financing Reports providing an overview of how different countries invested in their health sectors.[22] Tanzania's Health Financing Report states that from 2002 – 2013 per capita expenditure has been slowly increasing, reaching US$49 in 2013 which compares quite favorably to the other countries in the region.[22] As seen in Figure 2,[23] there has been some stagnation in total health expenditure in recent years, but the increase has been steady since the late-2000s. The graph depicts a spike in government expenditure on health right after 2005 which is the year that Tanzania qualified for MDRI assistance. However, since this spike the government's expenditure has stagnated and fluctuated, decreasing from 2009 onwards. This is also when Tanzania's total debt service began to stagnate, eventually increasing in 2013.

Another temporal element of Tanzania's health financing is foreign donors. A significant portion of Tanzania's health expenditure comes from foreign donors which are not depicted in Figure 2. UNICEF reports that over one third of total health care spending is accounted for by foreign resources which is far above the sub-Saharan average.[24] Since 1995 Tanzania's foreign resource contribution has consistently been higher than the sub-Saharan mean. While there have been fluctuations over the years, the margin has increased on average, bringing foreign contributions to 35.9% by 2014. However, from 2010 – 2014 these external resources have decreased by almost 4%.[24] Referring back to Table 1, this drop in external resources correlates with a slower increase in Health Index scores. During the years when external resource contributions were increasing, there was a slightly steeper increase in Health Index scores.[24]

It is important to note that other forms of aid can influence health outcomes. Certain fluctuations in health index patterns might be explained by tracking ODA (Official Development Aid) in Tanzania over the study's time period. As depicted in Figure 3, Tanzania's net ODA increased relatively steadily from 1995 to 2007 by about $2bn.[25] According to health index scores, health outcomes also increased steadily during this period. However, since 2007 ODA has fluctuated more with a recent drop in 2013.[25] During these recent fluctuations the growth in health index scores also slowed. In addition, from 2013 to 2015 the score growth was even slower. Thus, it could be argued that ODA contributions played a role in health outcomes, possibly even more so than debt relief. Nonetheless, it's also important to note that some forms of debt relief are accounted for in ODA.

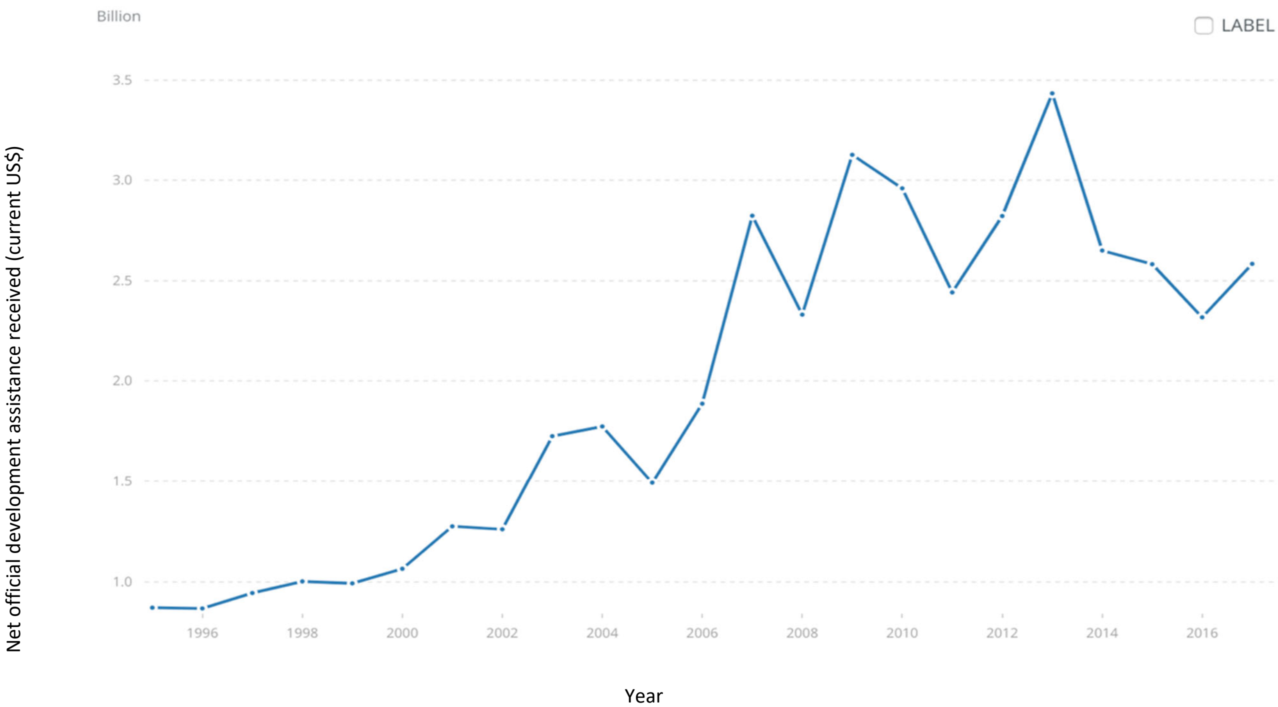


Figure 3: World Bank Net ODA received in Tanzania over time[25]

Moreover, another key distinction that needs to be made to assess health expenditure trends over time is the difference between recurrent and development spending. Tanzania allocates the majority of its health care budget to recurrent spending (salaries, personal emoluments, commodities and other charges) rather than development spending (capital, health technology and other additions). According to Tanzania's Ministry of Finance, since 1995 recurrent expenditure has made up the majority of the government's total expenditure. Nonetheless, development expenditure increased relatively steadily from 1999 – 2010 by 18.5% whereas the preceding years saw (1995 – 1999) saw inconsistent fluctuations.[26] This could also explain the slightly faster improvement in health outcomes following the HIPC Initiative: the Tanzanian government was allocating more of its health care budget to development expenditure while debt service payments were decreasing. Nevertheless, the portions of recurrent and development spending have been fluctuating recently. Development spending significantly decreased from 2013 to 2015. The average recurrent budget was 72.3% in recent years whereas the development budget was 27.7%.[26] Hence, Tanzania's health expenditure has been partly exaggerated by recurrent spending leading to less funding for capital improvements and other additions.[27] This also means that the fluctuations in recent years could have had an influence over the 2011 – 2015 trends in health outcomes.

In addition to these temporal elements, the government of Tanzania also faces challenges that are more fixed over time. Firstly, there are many logistical barriers to budget execution. The health budget's development spending execution is only 81% compared to the 91% execution of the recurrent budget in the 2013/14 fiscal year. The development budget execution was even lower in the 2012/13 fiscal year at 69%.[27] In previous years the recurrent execution of the health budget has consistently been higher than the development execution.[28] The Tanzanian government's Health Sector Public Expenditure Review states that budget performance continues to be hindered by "the low absorption capacity of the spending units; delays in the release of funds; non-release of the funds; over-ambitious budgeting… and lengthy and cumbersome procurement processes".[28] Another logistical error that the government has made is the allocation of funds to different districts. In Tanzania, health sector resources are distributed from the central government to the local governments based on an allocation formula that takes into account district demographics. The implementation of this formula lacks thoroughness given that there is considerable variance between the allocation predicted by the formula and the actual allocation. This has led to funding disparities between districts.[24] These logistical barriers make the implementation of development projects less efficient and can affect health outcome data.

Another fixed barrier to improving health outcomes in Tanzania is the country's socio-economic and geographic disparities which pose a challenge to efficiently tackling health. UNICEF's budget brief states that "child mortality, nutrition status, vaccination coverage and likelihood of attending a health facility to deliver a baby are still heavily determined by; location, wealth and the education of a mother".[29] This creates geographic disparities in health care accessibility. Central Tanzania has a vaccination coverage for children of 83% whereas the Southwest Highland only reaches 67%. Moreover, in rural areas 44.2% of births are at home compared to 12.8% in urban areas.[29] Thus, it is evident that even if development expenditure was increasing due to increased fiscal space from 2000 – 2010, Tanzania still continues to see substantial inequities in the availability and accessibility of health resources.

In addition to the temporal and fixed elements that might affect the quantitative patterns, there is also substantial literature that can provide more comprehensive explanations for the reasons why Tanzania and other African states struggle to effectively invest in social services.

Some research indicates that many African countries do not provide social services to their citizenry due to preferences or resource constraints. Robert Bates discusses how the characteristics of African governments, namely their economies and political institutions, can influence their policy decisions. Humphreys and Bates discuss that states that have immobile economic bases have a higher likelihood of engaging in predation.[30] Moreover, weak institutions and political instability can also affect policy choice. Many African governments that suffer from political instability exhibit opportunistic behavior.[30] For instance, Robert Bates discusses how various African states use agricultural policies to recompense political supporters. In many cases, rather than letting market forces determine price points, the government acts as a monopsony and buys all the crops from farmers at a fixed price that's lower than the market price. The government provides subsidies and loans in order to keep production at these artificially low prices. Although all producers stand to benefit, this still allows the government to pick which farmers win and which don't.[31] In addition, governments that are less constrained also exert more freedom over policy choices than governments that have robust checks and balances. Tanzania's political system has a weak structure of checks and balances which allows for a system that's dominated by the Presidency, the Executive and the Chama Cha Mapinduzi (CCM) party (the dominant ruling party in Tanzania). Although the country's institutional framework does include some checks and balances, the Legislature and other official entities have a limited formal capability to restrain the Executive.[32] This facilitates the government's ability to engage in opportunistic behavior and follow its own set of policy preferences.

Another overarching theory for the weakness of Tanzania's institutions is Jeffrey Herbst's work on the power of African states. Herbst theorizes that the lack of solid institutions in Africa and the prevalence of state failure is due to the continent's state-building experience. Africa did not have the same international pressures for war-making that medieval Europe experienced. European state-building occurred under systematic pressures, such as scarcity of land and the presence of densely populated regions, which promoted state consolidation.[33] On the other hand, Africa was a sparsely populated continent and precolonial African states did not have to face many of the survival imperatives that Europe did. They did not have the need to amass state power over rural terrain. The geographic features of Africa encouraged shared distribution of power and states did not have to exert strict control over their territories. African farmers invested in small slices of land due to the continent's low population density. If harsh rulers emerged farmers could easily flee rather than fight. This meant that trying to impose complete control over a particular region was very difficult so shared sovereignty became the standard. Thus, unlike Europe, Africa never developed a Westphalian notion of sovereignty and state borders.[33] Additionally, during colonialism the European powers that divided Africa had little incentives to develop the region's institutions. Their focus was on resource exploitation. After independence, the enforcement of arbitrary colonial borders on Africa made it even more difficult for them to adopt a survival mandate. Herbst argues that this caused African states to develop without having the strong and responsive economic, physical and political infrastructure that European states were able to develop.[33]

CONCLUSIONS

An important consideration for any type of economic-related development program is whether tangible advancements are being made in society beyond improvements in economic figures. In assessing how debt relief has impacted health outcomes in Tanzania, this study used graphical and qualitative analyses for the years prior to debt relief assistance as well as the years following it.

The findings from the graphical analysis suggested that debt relief had a beneficial impact on health outcomes in the short-term compared to the pre-debt relief period. The qualitative analysis partly corroborates these findings given that development expenditure did increase during the short-term period (2000 – 2010). However, the improvements in health outcomes during the short-term could have also been influenced by foreign donors that aren't affected by debt relief. Since foreign donor contributions increased in the mid-2000s this might have played a role in improving health outcomes during this time.[24] Moreover, the health financing literature suggests that there were factors not accounted for in debt relief assistance that influenced health outcomes in the long-term. These include the 4% drop in foreign donor contributions from 2010 – 2014 as well as the logistical barriers to budget allocation and execution. It should also be noted that there were significant fluctuations in development expenditure in recent years.[26] Moreover, there are also some explanations for why African states might be less efficient at allocating resources to the public sector. Firstly, political instability and weak institutions can influence policy choice. When there aren't robust checks and balances, as is the case in Tanzania, the government can more easily engage in opportunistic behavior and follow its own policy preferences.[30] Additionally, the lack of state-building experience that African states had in the pre-colonial era as well as the challenges that colonialism imposed on such states impaired their ability to develop robust institutions.[33] In general, the lack of efficiency in Tanzania's public spending is most likely an amalgamation of the previously stated explanations as well as the temporal and fixed elements described.

These findings aren't completely consistent with the hypothesis, which predicted that debt relief would increase fiscal space and allow for more health care spending in turn improving health outcomes. Although debt relief does increase fiscal space, it's not always the case that this fiscal space will be invested in improving health outcomes. The graphical analysis does indicate that debt relief had a positive impact on health outcomes in the period after the HIPC initiative. The qualitative analysis also points to substantial increases in development expenditure during this period. However, the results indicate that in the long term the effects of debt relief are diminished. As stated previously, from 2011 – 2015 development expenditure fluctuated and there were smaller increases in Tanzania's Health Index Score. The hypothesis also predicted that logistical barriers could affect the efficiency of increased spending which was supported by the health financing literature.

While debt relief programs increase fiscal space and might allow for some tangible socio-economic improvements, they can still be amended to increase their efficacy. Tanzania is regarded as the poster child of debt relief; however, the country still faces lots of challenges and doesn't always see satisfactory outcomes. Tanzania experiences considerable logistical barriers to its health budget execution and is prone to fluctuations in expenditure. If Tanzania were to receive logistical support and guidance along with debt relief assistance the country could make the most of its increased fiscal space. It's important to note that in order to qualify for the enhanced HIPC Initiative countries needed to demonstrate a commitment to poverty reduction and a stable macroeconomic environment.[1] However, these commitments should be maintained post-debt relief as well. Additionally, logistical assistance in budget execution and allocation of funds would also allow Tanzania to reduce some of the geographic disparities and underperformance it currently sees. Although debt relief is an important form of assistance for developing economies, it does not generate new resources. Thus, it is important to maximize and correctly allocate the resources that debt relief can free up. If nation-wide development projects provided more logistical assistance and guidance, developing countries might see more consequential outcomes, and national inequities in resource availability and accessibility could be significantly curtailed.

ACKNOWLEDGEMENTS

The author thanks Professor James Desveaux, Andrea Vilán, PhD and anonymous reviewers for crucial suggestions, edits and revisions to the content of the manuscript as well as their support throughout the research process. The author also thanks the Center for American Politics and Public Policy at UCLA and their Quarter in Washington program for making this research possible.

REFERENCES

1. International Monetary Fund, (2000) Press Release: IMF and IDA Support Debt Relief for Tanzania. 5 April. *https://www.imf.org/en/News/Articles/2015/09/14/01/49/pr0026* (Accessed September 30, 2018)
2. International Monetary Fund. (2017) Debt Relief Under the Heavily Indebted Poor Countries (HIPC) Initiative. Factsheet, IMF & World Bank, 1-4.
3. Heller, Peter (2005) Understanding Fiscal Space. IMF Policy Discussion Paper. Fiscal Affairs Department, International Monetary Fund.
4. Gomera, Jumanne Said. (2014) Impact of Debt Relief on Economic Growth in Heavily Indebted Poor Countries.Masters Thesis. Seoul National University.
5. World Health Organization. (2006) Health and development. 30 March. *https://www.who.int/hdp/en/* (Accessed September 25, 2018)
6. Johnson, Harold J. (1998) Developing Countries: Status of the Heavily Indebted Poor Countries Hipc Debt Relief Initiative. Washington, D.C.: United States General Accounting Office.
7. Cassimon, Danny, and Bjorn Van Campenhout. (2008) Comparative Fiscal Responce Effects of Debt Relief: an Application to African HIPCS. *South Africal Journal of Economics* 76 (3): 427-442.
8. Cassimon, Danny, Marin Ferry, Marc Raffinot, and Bjorn Van Campenhout. (2013) Dynamic Fiscal Impact of The Debt Relief Initiatives on African Highly Indebted Poor Countries (HIPCs). *Working paper, DIAL (Développement, Institutions et Mondialisation)*, Paris, France, 24

9. Osei-Fosu, Anthony Kofi. (2010) Three Essays on Economic Impact Assessment of the Use of the HIPC intiative for Poverty Reduction in Ghana. *Thesis*, Department of Economics, Kwame Nkrumah University of Science and Technology, Kumasi, Kumasi, Ghana.
10. Dijmeu, Eric W. (2017) The impact of the Heavily Indebted Poor Countries initiative on growth and investment in Africa. *Journal Article, World Development 104: 108-127.*
11. Welander, Anna. (2016) Does Debt Relief Improve Child Health? Evidence from Cross-Country Micro Data. *Working Paper,* Department of Economics, School of Economics and Management Lund University, Lund, Sweden.
12. Heller, Peter. (2005) Back to Basics -- Fiscal Space: What It Is and How to Get It. *Finance and Development* 42 (2).
13. International Monetary Fund. (2001) Press Release: IMF and World Bank Support US$3 billion In Debt Service Relief For Tanzania Under Enhanced HIPC Initiative. 27 November. *https://www.imf.org/en/News/Articles/2015/09/14/01/49/pr0148* (Accessed September 30, 2018)
14. International Monetary Fund. (2005) Press Release: IMF to Extend 100 Percent Debt Relief to Tanzania Under the Multilateral Debt Relief Initiative. 23 December. *https://www.imf.org/en/News/Articles/2015/09/14/01/49/pr05304* (Accessed September 2018, 2018)
15. United Nations Development Programme. (2010) What is the rationale behind the minimum values for indicators? 5 November. *http://hdr.undp.org/en/content/what-rationale-behind-minimum-values-indicators* (Accessed October 3, 2018)
16. United Nations Development Programme. (2015) Human Development Report. Washington, D.C.
17. United Nations Development Programme. (2015) Human Development Report - Technical notes. Washington, D.C.
18. FAO, IFAD, UNICEF, WFP and WHO. (2018) The State of Food Security and Nutrition in the World 2018. Rome, Italy.
19. United Nations Population Division. (2017) World Population Prospects: The 2017 Revision. (Department of Econonics and Social Affairs).
20. Bank of Tanzania (2014) Publications and Statistics. June 15. *https://www.bot.go.tz/Publications/publicationsAndStatistics.asp* (Accessed June 21, 2019)
21. CIA World Factbook (2015) Debt: External Billion (US$). 30 June. *https://www.indexmundi.com/g/g.aspx?c=tz&v=94* (Accessed June 21, 2019)
22. USAID. (2016) Health Financing Profile Tanzania. *Factsheet*, Health Policy Project, Washington, DC, 2.
23. World Health Organization. (2014) WHO Global Health Expenditure Atlas. Geneva, Switzerland: WHO Press.
24. UNICEF. (2016) Health Budget Brief Tanzania FY 2011/12–FY 2015/16. Summary Report, 1-12.
25. World Bank. (2017) Net official development assistance received (current US$). April 22. *https://data.worldbank.org/indicator/DT.ODA.ODAT.CD* (Accessed July 4th 2019)
26. Ministry of Finance and Economic Affairs. (2010) Budget Digests 2004/5 - 2010/11. Digest, Dodoma, Tanzania: Government of Tanzania.
27. USAID. (2015) Prospects for Sustainable Health Financing in Tanzania. Baseline Report, Health Policy Project, Washington, D.C.: U.S. Government, 1-27.
28. Ministry of Health and Social Welfare. (2012) *Health Sector Public Expenditure Review.* Dar-es-Salaam, Tanzania: Government of Tanzania.
29. UNICEF. (2018) Health Budget Brief 2018 Tanzania. Summary Report, 1-12.
30. Humphreys, Macartan, and Bates, Robert. (2005) Political Institutions and Economic Policies: Lessons from Africa. *British Journal of Political Science 35 (3): 403-428.*
31. Bates, Robert, and Lofchie, Michael. (1980) Agricultural Development in Africa: Issues of Public Policy. New York: University of California, Los Angeles.
32. Lawson, Andrew and Rakner, Lise (2005). Understanding Patterns of Accountability in Tanzania. *Synthesis Report,* Oxford Policy Management, Chr. Michelsen Institute, Research on Poverty Alleviation.
33. Herbst, Jeffrey. (2000) African States and Power. Princeton University Press.
34. World Bank. (2015) Maternal mortality ratio (modeled estimate, per 100,000 live births). *https://data.worldbank.org/indicator/SH.STA.MMRT?locations=TZ.* (Accessed September 25, 2018)

ABOUT THE STUDENT AUTHOR

Fernando Lopez Oggier is an undergraduate student at the University of California, Los Angeles (UCLA) that will graduate in June 2020. He is pursuing a Bachelor of Arts in Political Science and a Bachelor of Arts in Geography. Fernando aims to pursue a career in international affairs and hopes to attend graduate school.

PRESS SUMMARY

In 1996 The World Bank and the International Monetary Fund established for the first time a debt relief program for countries that had been saddled with unmanageable debt. High levels of debt can inhibit sustainable development and prosperity. These debt relief programs aimed to reduce debt service payments and allow countries to allocate their budgets to other sectors of society. Previous research demonstrates that a country's economic growth is positively impacted by debt relief initiatives. However, this study aimed to analyze whether freeing up a government's budget through debt relief allows for prosperity in other sectors of society. The case study that this research focused on is Tanzania given that the country has been a primary recipient of multiple debt relief initiatives. Instead of attempting to evaluate every area that qualifies as "human development", the impacts on the country's health sector can be specifically analyzed in order to narrow the purview of the research. Increased health attainment is a critical aspect of human development and plays a fundamental role in poverty reduction. Therefore, this study compiled Tanzania's health data into a composite indicator to compare health outcome trends with the country's external debt. Health financing information from the Bank of Tanzania, UNICEF and USAID as well as political science literature contextualized this analysis. The study found that health outcomes improved in the short term period after the HIPC Initiative; the health financing literature also pointed to increased development expenditure during this time period. However, the effects of debt relief seemed to diminish in the long-term due to fluctuations in external donors and logistical barriers to budget execution. Some overarching explanations also stated that the country's weak system of checks and balances and the lack of robust institutions could cause opportunistic policy preferences that might not necessarily improve Tanzania's health outcomes. The country also continues to face socio-economic and geographic disparities in health outcomes and funding.

APPENDICES

Appendix A: The statistics compiled to calculate Tanzania's Health Index: life expectancy[16], maternal mortality[34], child mortality[19], and undernourishment[18]

Year	Life Expectancy	Maternal Mortality Ratio (per 100,000 births)	Child mortality (per 1,000 births)	Percentage of population that is undernourished
1995	49.5	961	163.9	32.1
1996	49.6	944	159.5	34.5
1997	49.9	924	154.1	35.6
1998	50.3	900	147.6	36.3
1999	50.8	870	140.1	36.3
2000	51.5	842	131.8	36.8
2001	52.3	813	123.1	37.1
2002	53.2	777	114.7	37.8
2003	54.1	747	106.8	37.7
2004	55.2	717	99.8	37.7
2005	56.2	687	93.7	36.7
2006	57.2	652	88.5	35.4
2007	58.2	608	83.7	34.2
2008	59.2	566	79.2	33.8
2009	60	542	75.7	34.5
2010	60.9	514	71.9	34.7
2011	61.7	483	68.5	34.7
2012	62.5	464	65.8	33.4
2013	63.3	438	63.4	32.7
2014	64.2	418	60.9	32.1
2015	65	398	58.8	32.1

Appendix B

Year	Life Expectancy Sub index
1995	0.454
1996	0.455
1997	0.460
1998	0.466
1999	0.474
2000	0.485
2001	0.497
2002	0.511
2003	0.525
2004	0.542
2005	0.557
2006	0.572
2007	0.588
2008	0.603

2009	0.615
2010	0.629
2011	0.642
2012	0.654
2013	0.666
2014	0.680
2015	0.692

Appendix C

Year	Child Mortality Sub index
1995	0.185
1996	0.208
1997	0.235
1998	0.269
1999	0.307
2000	0.350
2001	0.394
2002	0.437
2003	0.478
2004	0.514
2005	0.545
2006	0.572
2007	0.596
2008	0.619
2009	0.637
2010	0.657
2011	0.674
2012	0.688
2013	0.701
2014	0.713
2015	0.724

Appendix D

Year	Maternal Mortality Sub index
1995	0.039
1996	0.057
1997	0.077
1998	0.101
1999	0.131
2000	0.160
2001	0.189
2002	0.225
2003	0.256
2004	0.286
2005	0.316
2006	0.352
2007	0.396
2008	0.438
2009	0.463
2010	0.491
2011	0.522
2012	0.541
2013	0.568
2014	0.588
2015	0.608

Appendix E

Year	Undernourishment Sub index
1995	0.507
1996	0.464
1997	0.444
1998	0.431
1999	0.431
2000	0.422
2001	0.416
2002	0.404
2003	0.405

2004	0.405
2005	0.424
2006	0.447
2007	0.469
2008	0.476
2009	0.464
2010	0.460
2011	0.460
2012	0.484
2013	0.496
2014	0.507
2015	0.507

The Association of *TAS1R2* to Dental and Cardiovascular Health

Joseph Keryakos, Annika van Oosbree, & Alyssa Marie Kiesow*

Department of Biology, Northern State University, Aberdeen, SD

https://doi.org/10.33697/ajur.2019.022

Students: joseph.keryakos@wolves.northern.edu, annika.vanoosbree@wolves.northern.edu*
Mentor: alyssa.kiesow@northern.edu

ABSTRACT
Despite brushing and/or flossing their teeth twice daily, many people are still susceptible to dental cavities and tooth decay. This research investigates the genetic and cardiovascular health behind this phenomenon. Two gene variants related to taste pathways, taste 1 receptor member 2 (*TAS1R2*) and taste 2 receptor member 38 (*TAS2R38*), were tested on the DNA of 20 students at Northern State University (10 males and 10 females). In concert with genetic screening, tooth impressions were taken of the participants' upper and lower jaws along with salivary pH, heart rates, and blood pressures. Participants' cavities and fillings were counted and their gums examined for inflammation. Results showed that seven out of 10 males and two out of 10 females had the gene variant (*TAS1R2*). Students with this gene variant had an average salivary pH of 5.22—significantly lower than the salivary pH for the other non-carrier students ($p < 0.05$). These students also had smaller-sized tooth enamel, with none showing a size greater than one millimeter ($\bar{x} = 0.84$ millimeters). Students not expressing the gene variant had fewer cavities than those expressing the *TAS1R2* gene variant (*i.e.,* one of the regions amplified). Four of the males and both of the females that carried the gene variant also showed signs of swollen gums, possibly contributing to heart disease in the future. Blood pressures and heart rates for the carriers were statistically significant ($p < 0.05$), showing higher pressures and faster rates compared to non-carriers; meanwhile, all of the non-carriers had normal pressures and rates. Further, body mass index was lower among individuals without the gene variant. The results this limited study indicate that the *TAS1R2* gene variant may play a role in cavity development and impact (or indicate poor) cardiovascular health, highlighting the importance of understanding the role of gene variants with regard to risk of tooth decay and gum and heart disease.

KEYWORDS
Dental cavities; Tooth decay; Gum and heart disease; Taste pathway gene; Gene variant; Blood pressure; Heart rate; Salivary pH; Tooth enamel

INTRODUCTION

Many people are still susceptible to dental cavities and tooth decay despite brushing and/or flossing their teeth twice daily. While dentists advise all people to brush their teeth and floss every day, this will not stop cavities from forming on their teeth; rather, it will just make it less likely to occur. Gene variants play a major role in the formation of dental caries. For example, a taste pathway gene variant can diminish oral and/or cardiovascular health.[1] Furthermore, sweet taste preferences, also linked to taste pathway gene variants, can lower salivary pH, which decreases the thickness of tooth enamel, making teeth more prone to cavities and decay. This subsequently leads to gum disease that could, later on, provoke heart disease.

Past studies examined the effects of four taste pathway gene variants associated with tooth decay and cavities. Wendell et al. (2010) researched the human taste receptor gene variants and showed that there were "statistically significant associations in taste pathway gene variants for caries risk."[1] They were able to find two genotypes that make the individual's diet rich in sugar and sweets – taste 2 receptor member 38, *TAS2R38* (7q34, 607751) and taste 1 receptor member 2, *TAS1R2* (1p36.13, 606226).[1] The genotypes transfer a signal to the brain notifying it of the sweetness of the food, and this leads the brain to favor sweet-flavored food. Hollá (2015) had a similar finding, determining that specific taste pathway gene variants lead an individual to prefer a diet rich in sugar.[2] Each one of these gene variants causes the person to prefer sweet-tasting foods and can lead to higher chances of forming oral cavities.

Preferring sweet-tasting food is closely linked to lower salivary pH. Food rich in sugar causes both an increase in oral bacteria that feed on the sweet food debris in the human mouth and a decrease in salivary pH.[3] Oral bacteria living in the human mouth favor the food debris that remains on the person's teeth and gums. Once the bacteria starts feeding on this debris, its product is acidic, thereby lowering salivary pH due to the process of sucrose fermentation.[4]

Acidic saliva caused by oral bacteria feeding on the debris of sweet food alters tooth enamel depth. Tooth enamel, the hard outer covering of the tooth, is mainly composed of minerals such as hydroxyapatite, an acid-soluble compound.[5] Acidic saliva dissolves tooth enamel,[6] and thinner enamel is associated with a higher risk of developing tooth caries.[7] The acidic saliva in the mouth combines with the oral bacteria and food debris, forming a plaque that clings to the teeth. Further, thinner enamel with the presence of plaque causes caries that eventually lead to tooth decay.[8]

Dental caries are more prevalent in people carrying one of the aforementioned taste pathway gene variants because they are prone to the formation of plaque. Dental plaque remaining on teeth for more than three months hardens and cannot be removed by brushing, instead requiring scaling or planing.[9] Rostami et al. (2017) concluded that, because of their propensity to develop/retain plaque, people expressing taste pathway gene variants should visit the dentist more than two times per year to prevent dental caries.[9]

Oral plaque causes three major damages: it reduces the size of the tooth enamel, induces the formation of dental caries, and instigates gingiva inflammation.[10] Dental plaque contains bacteria that produce toxins, like butyrate and propionate, which can cause gum tissue to swell and bleed, or induce symptoms of gingivitis and periodontal disease.[11] Additionally, there seems to be a link between the human immune system and the swollen gums.[12] Once toxins from plaque contact the gums, the toxins can enter the bloodstream. Once the toxins are detected, the immune system sends phagocytes to ingest them, causing gums to swell and bleed.

Unfortunately, swollen gums are often a symptom of gum disease. Gum disease not only negatively affects oral health, but it may also disturb cardiovascular health.[13] Once gum tissue is penetrated, bacteria can enter the bloodstream, circulate through the body into the heart, and accumulate in the carotid arteries, elevating blood pressure and heart rate and leading to heart diseases and stroke later on.[14]

The aforementioned research examines many aspects of dental health (i.e., taste gene variants in relation to tooth decay/plaque build-up, plaque build-up and gum health/immune system response, and oral health in relation to cardiovascular health); however, while all researched topics reported correlations, none sought to link all components together. Our research reported herein builds on existing data while gaining additional insight on correlations and relationships among taste pathway gene variants and physiological conditions in certain age groups. More specifically, our study examines the acidity of the saliva on teeth, screening for variants in two taste pathway gene variants, *TAS1* and *TAS2*, to identify specific physiological limitations caused by these variants in male and female individuals ranging in age 18-23. Further, we seek to examine the link (if any) between tooth enamel depth and cavities as well as tooth decay, gum disease, and heart disease.

METHODS AND PROCEDURES

Subjects

Ten females and ten males (18 to 23 years old) of seemingly good health based on hair shininess, skin appearance, and weight, were asked to participate in this study. The Institutional Review Board of Northern State University approved our research on October 10, 2016 (IRB Protocol #: 2016-09-28A) to examine a variety of physical attributes on human subjects. Participants voluntarily signed and submitted an informed consent stating the description of the research project, its risks and benefits, the time involvement, participants' rights, and the contact information of the research supervisors.

Determining Gene Variants of Taste Pathway Genes

Genomic DNA was isolated from at least 1 mL of saliva collected from each participant following protocol from previous research,[15] albeit using AB lyses buffer (1 M Tris [pH 8.0], 5 M NaCl, 0.5 M EDTA [pH 8.0], 10% SDS) as a substitute for the procedural lyses buffer. Confirmation of DNA was conducted on a 1.0% agarose 1X TBE gel using EtBr (Carolina®) as a nucleic acid stain in each sample and a water blank to test for contamination of products. Taste 2 receptor member 38, *TAS2R38* (7q34, 607751), and taste 1 receptor member 2, *TAS1R2* (1p36.13, 606226), two known taste pathway gene variants,[1] were selected for further analyses to identify amplifiable variants of these genes. Using Primer3 from IDT SciTools (*http://www.idtdna.com/SciTools/SciTools.aspx*) we designed four primer pairs targeting gene variant regions of *TASR38* and *TAS1R2* taste pathway genes, then estimated annealing temperatures and levels of self-priming. IDT SciTools uses a suite of tools to adequately characterize sequences and customize primers. Four primer sets were developed and compared to target sequences using BLAST (reference) for specificity to detect a specific gene variant of two separate genes associated with taste perception (**Table 1**).

DNA Amplification

PCR was performed with 15 µL reactions containing 50 ng genomic DNA, 10 µM of each primer (forward and reverse, **Table 1**), 400 µM of each dNTP (dATP, dGTP, dCTP, and dTTP), 0.50 units of Taq polymerase in buffer (Promega), and 3 mM of $MgCl_2$.

PCR started with initial denaturing at 94 °C for 4 minutes followed by 30 cycles of 94 °C for 1 minute, annealing at 53 °C for 1 minute, and extension at 72 °C for 1 minute. Confirmation of PCR amplification was conducted by running a sample on a 1.5% agarose 1X TBE gel using EtBr (Carolina®) as a nucleic acid stain in each sample and a water blank to test for contamination of products in every PCR run. Even though real-time quantitative PCR could have quantified the gene variant expression, whether homozygous or heterozygous, presence of the gene variants was determined and compared to a negative control then quantified with an Invitrogen® Low DNA Mass ladder.

Health Screening

Several physiological attributes were tested to determine dental and cardiovascular health of participants in concert with genetic data. Salivary pH was measured with pH test strips. With an ultraviolet light, we examined the mouth of each participant for cavities, fillings and swollen gums. Finally, in order to measure the teeth enamel depth, we made tooth impressions of the upper and lower jaws using alginate and dental trays. Plaster was used to make the cast of the dental impressions and a Vernier® Caliper ruler was used to measure the width of the central incisors of each participant. Blood pressure and heart rates were measured twice a day (once in the morning before eating and once in the afternoon). Two readings of both heart rates and blood pressures were taken from each participant in order to get accurate results. Height and weight were also collected to determine the body mass index (BMI) of each participant.

Data Analyses

Data were analyzed using JMP® software (SAS Institute Inc.) to conduct comparisons via analysis of variance (ANOVA), where P-values less than 0.05 were considered significant.

RESULTS

Gene Variants of Taste Pathway Genes

Four primer sets were created for the purposes of this research (**Table 1**). The differences between these TAS gene variants were the product length and the variety of exons. For example, TAS2R38 had few but large exons while TAS1R2 had many but small exons. The *TAS1R2.F1* and *TAS1R2.R1* set had successful amplification, amplifying a gene variant of taste receptor 1 member 2 (*TAS1R2,* 1p36.13, 606226). This particular gene variant was found in seven out of 10 male participants and two out of 10 female participants. This gene variant was measured/quantified using low DNA mass ladder, and used hereafter for identifying physiology traits in relation to dental traits.

LOCI	PRIMER SEQUENCES (5'-3')	REGION	PRODUCT LENGTH
TAS2R38.F1	F: AGG CCC ACA TTA AAG CCC TC	Start: 803 Stop: 822	204 bp
TAS2R38.R1	R: CAG CTC TCC TCA ACT TGG CA	Start: 1006 Stop: 987	
TAS2R38.F2	F: GCA TCC GCA CTG TGT CCT AT	Start: 101 Stop: 120	355 bp
TAS2R38.R2	R: AAA CGG ATG AGC TTG GAG CA	Start: 455 Stop: 436	
TAS1R2.F1	**GGC TGT GAC AAA AGC ACC TG**	**Start: 1211 Stop: 1230**	**750 bp**
TAS1R2.R1	**GAA AGA ACG CAC GGC GAT AC**	**Start: 1960 Stop: 1941**	
TAS1R2.F2	CAG ATC ACC TAC AGC GCC AT	Start: 482 Stop: 501	824 bp
TAS1R2.R2	TGG TCC AGG AGA GTG AA	Start: 1305 Stop: 1286	

Table 1. A specification table of the four primer sets (F=Forward, R=Reverse) designed during this research to isolate and amplify gene variants related to dental caries affiliated with taste receptors 1 and 2. We showed successful amplification of taste receptor 1 member 2 with primers TAS1R2.F1 and TAS1R2.R1 (in bold).

Health Screening

The average salivary pH of the participants with the gene variant *TAS1R2* was 5.22 (SD ± 0.45), which is significantly lower than the average of the participants without the gene variant ($\bar{x}$ = 6.72 [SD ± 0.46]; t = 7.4, df = 18, $p < 0.05$). Median values are similar, where salivary pH is slightly acidic in those with the gene variant and salivary pH is neutral in those without the gene variant (**Figure 1**). Six out of the seven males with the gene variant *TAS1R2* and both of the females with the gene variant had thin enamel (0.84 mm [SD ± 0.09]), while all of the individuals without the gene variant had an average thickness of 1.17 mm (SD ± 0.1), which is considered normal and healthy. Enamel thickness is significantly different between those with and without the gene variant (t = 7.3, df = 17, $p < 0.05$; **Figure 1**)

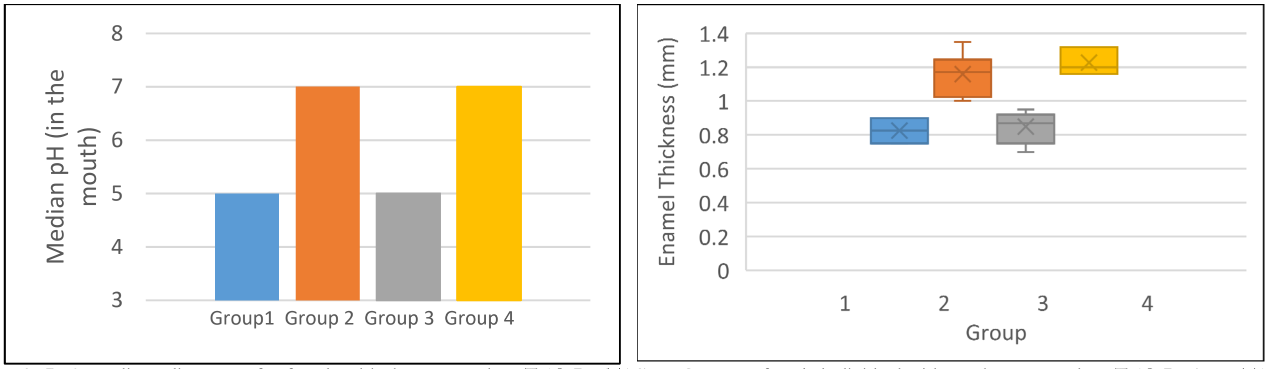


Figure 1. ***Left:*** Median salivary pH for female with the gene variant *TAS1R2* (TAS) or Group 1, female individual without the gene variant *TAS1R2* (Non-TAS) or Group 2, male with the gene variant *TAS1R2* (TAS) or Group 3, male individual without the gene variant *TAS1R2* (Non-TAS) or Group 4. ***Right:*** Enamel thickness (in mm) for female with the gene variant *TAS1R2* (TAS) or Group 1, female individual without the gene variant *TAS1R2* (Non-TAS) or Group 2, male with the gene variant *TAS1R2* (TAS) or Group 3, male individual without the gene variant *TAS1R2* (Non-TAS) or Group 4. On the box and whisker plots, the line indicates median and X indicates mean. For both data figures, there is a difference between these attributes among these groups ($p<0.05$).

The seven male participants with the gene variant *TAS1R2* had an average of 11.86 cavities (SD ± 7.14) with a median of 12, while the three males without the gene variant *TAS1R2* had a median of four cavities. The two female participants with the gene variant had an average and median of 6.5 cavities; one female had seven cavities while the other had six. The eight females without the gene variant *TAS1R2* had a median of three cavitites. There is a significant difference in prevalence of cavities between individuals with the gene variant and those without the gene variant ($t = -4.6$, df = 10, $p < 0.05$; **Figure 2**).

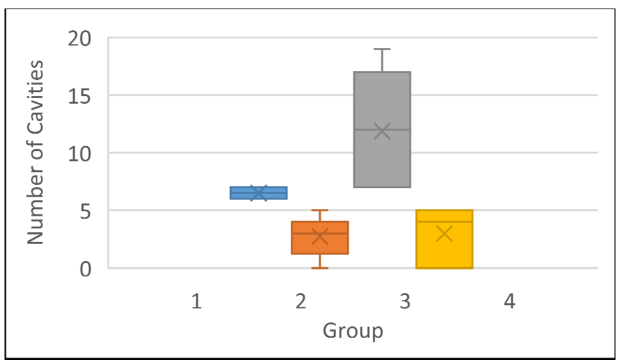


Figure 2. Number of cavities for female with the gene variant *TAS1R2* (TAS) or Group 1, female individual without the gene variant *TAS1R2* (Non-TAS) or Group 2, male with the gene variant *TAS1R2* (TAS) or Group 3, male individual without the gene variant *TAS1R2* (Non-TAS) or Group 4. There is a difference between these attributes among these groups ($P<0.05$). On the box and whisker plots, the line indicates median and X indicates mean.

Four out of the seven male participants with the gene variant *TAS1R2* showed some gum swelling and were symptomatic for gum disease. Both female participants with the gene variant had swollen gums. None of the participants without the gene variant had a sign of gum disease or inflammation.

The seven males with the gene variant *TAS1R2* had slightly high blood pressure levels ($\bar{x}$ = 127.14/82.86 [SD ± 6.9]) and heart rates ($\bar{x}$ = 87 [SD ± 6.1], $\tilde{x}$ = 77). Both females with the gene variant had slightly high blood pressure levels ($\bar{x}$ = 125/90 [SD ± 6.5]) and heart rates ($\bar{x}$ = 86 [SD ± 6.01], $\tilde{x}$ = 86). In all instances, participants without the gene variant showed normal blood pressure levels ($\bar{x}$= 115.45/ 72.72 [SD ±6.86]) and heart rates ($\bar{x}$ = 62.90 [SD ± 6.11], $\tilde{x}$ = 62 [females], $\tilde{x}$ = 60 [males]). Heart rates ($t = -6.1$, df = 13, $p < 0.05$, **Figure 3**) and blood pressure levels ($t = -3.5$, df = 14, $p < 0.05$) were significantly different between those with and without the gene variant.

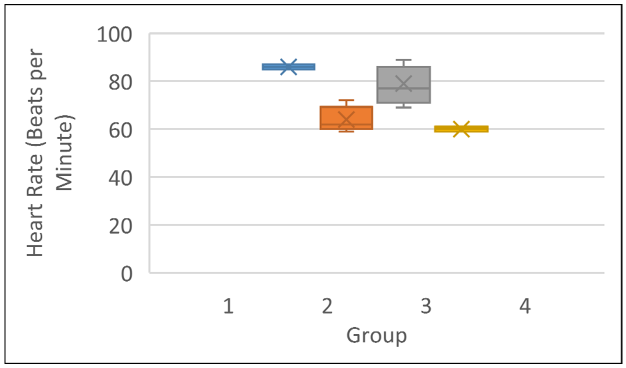


Figure 3. Heart rates for female with the gene variant *TAS1R2* (TAS) or Group 1, female individual without the gene variant *TAS1R2* (Non-TAS) or Group 2, male with the gene variant *TAS1R2* (TAS) or Group 3, male individual without the gene variant *TAS1R2* (Non-TAS) or Group 4. There is a difference between these attributes among these groups ($p<0.05$). On the box and whisker plots, the line indicates median and X indicates mean.

BMI values ranged from 16.5 to 24 in males and females. Males and females with the gene variant *TAS1R2* had a slightly higher BMI (females, $\bar{x}$ = 24 [SD ± 1], $\tilde{x}$ = 24; males, $\bar{x}$ = 24.88 [SD ± 1.9], $\tilde{x}$ = 25.2) than those without the gene variant. Male and female participants without the gene variant showed normal BMI (females, $\bar{x}$ = 17 [SD ± 0.6], $\tilde{x}$ = 17; males, $\bar{x}$ = 18.77 [SD ± 1.6], $\tilde{x}$ = 18.3). BMI was significantly different between those with and without the gene variant, where individuals without the variant showed lower BMI (**Figure 4**).

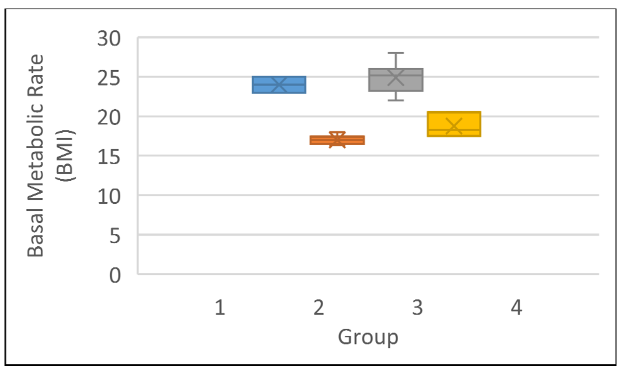


Figure 4. Body mass index (BMI) for female with the gene variant *TAS1R2* (TAS) or Group 1, female individual without the gene variant *TAS1R2* (Non-TAS) or Group 2, male with the gene variant *TAS1R2* (TAS) or Group 3, male individual without the gene variant *TAS1R2* (Non-TAS) or Group 4. There is a difference between these attributes among these groups ($p<0.05$). On the box and whisker plots, the line indicates median and X indicates mean.

DISCUSSION

Here, we provide evidence for an association between a specific gene variant, *TAS1R2*, and state of dental and cardiovascular health. Our data suggest that the presence of certain alleles for taste pathway genes, most notably the *TAS1R2* gene variant, has an effect on dental and cardiovascular health. All participants at Northern State University presenting the *TAS1R2* gene variant had poor dental health (*i.e.*, low salivary pH and high numbers of dental caries) and subpar cardiovascular health (*i.e.,* high blood pressure and heart rate) compared to individuals without the presentation of *TAS1R2*. Further, BMI was slightly higher in individuals with the *TAS1R2* gene variant.

Specific single nucleotide polymorphisms (SNPs) of *TAS1R2* have been linked with a susceptibility of developing dental caries.[2] Specifically, the Val allele of rs35874116 (Ile191Val) was linked to overall dental development of dental caries. A study conducted on a specific population in West Mexico found that the Val191Val polymorphism in *TAS1R2* is linked to high carbohydrate intake and dyslipidemia.[16] Ramos-Lopez et al. (2016) hypothesized that "the Ile191Val polymorphism resides in the predicted first large extracellular domain of the *TAS1R2* receptor, which hypothetically contains the ligand-binding site for carbohydrates and dipeptide sweeteners."[16] Additionally, compared with both Ile/Ile and Ile/Val polymorphisms, Val/Val convened a significantly higher prevalence of hypertriglyceridemia—a disorder, which is associated with atherosclerosis and may predispose a patient to developing heart disease.[16] This finding, paired with those stating this same polymorphism is linked to high carbohydrate intake, correlates with the findings of this study. Carriers of the *TAS1R2* gene variant have lower salivary pH, thinner enamel, more cavities, and higher blood pressure and heart rate (all with statistical significance).

The presence of the *TAS1R2* gene variant seems to initiate a cascade of events that can lead to tooth decay. This genotype facilitates the transfer of a signal to the brain indicating that the consumed food is very sweet.[1] Based on the results of the present study involving salivary pH, it is likely that the signal increases the brain's affinity for that particular sweetness of food, making future consumption of sweet food more likely. To our knowledge, this is the first study to study and identify a statistically significant link between the *TAS1R2* variant and salivary pH, potentially highlighting how this gene variant initiates this cascade that leads to tooth decay. Salivary pH lowers when sweet food is consumed due to bacteria feeding on the debris of the sweet food;[3] therefore, a significantly lower salivary pH in the present study indicates higher consumption of sweet food. Acidic saliva eats away at the enamel,[6] and, not surprisingly, the carriers of the *TAS1R2* gene variant had significantly thinner enamels. Moreover, as chronic consumption of sweet food occurs, cavity development becomes more likely, as evidenced by the significantly higher number of cavities in those with the gene variant of *TAS1R2* versus those without. This is in part due to the thinner enamel from the lower salivary pH, as thinner enamel has been connected with a higher risk of developing caries.[7] Ultimately, this could lead to gum disease or periodontitis, as both females and four out of the seven males with the gene variant show signs of one or both of these oral issues. Overall, a simple inherent affinity for sweet-flavored foods due to the *TAS1R2* gene variant can lead to a cascade of events eventually ending in gum disease; the lower salivary pH subsequently leads to thinner enamel and more cavities.

This cascade can continue on to affect cardiovascular health. A strong relationship has been previously shown to exist between periodontitis and atherosclerotic vascular disease (ASVD), with patients with cardiovascular diseases having a higher percentage of caries, tooth loss, and orofacial pain.[17] Heart rates and blood pressures collected during the present study demonstrated a close relationship between caries and elevated blood pressures and heart rates, with students presenting symptoms of gum disease having higher heart rates and blood pressures than participants with healthy gums ($p < 0.05$). This finding is likely due to the heavier prevalence of gum inflammation and evidence of future gum disease found at significantly higher levels in individuals with the gene variant. Once the gum tissue is penetrated, bacteria can easily enter the bloodstream and eventually lead to the accumulation of plaque in the lumen of the carotid arteries.[14] This then causes the elevation in the patient's heart rate and blood pressure level—an elevation that could lead to ASVD and provoke a stroke later on. Eventually, the presence of the *TAS1R2* gene variant creates an oral environment with a lower salivary pH due to an increased consumption of sweet food, after which a cascade of physiological events (due to lower salivary pH) can lead to cardiovascular issues. The significantly higher heart rates and blood pressure levels found in those with versus without the gene variant confirm this relationship.

BMI plays a role in cardiovascular health. Research has shown that as BMI increases the risk of cardiovascular disease increases in both adults and children.[18] Our research links dental health and the presence of the gene variant to cardiovascular health. Thus, the presence of the *TAS1R2* gene variant may increase the chances of poor dental health and cardiovascular disease.

Regions attempted for amplification of *TAS2R38* contain a large continuous exon, while *TAS1R2* is comprised of many smaller exons. Successful amplification of *TAS1R2* captured multiple exons with only one of the two primer sets. It is likely that the second set of primers did not amplify the *TAS1R2* because the primers targeted a larger exon region with minimal variation much like the primer sets for TAS1R38. We suggest targeting regions with more variation in exon length. Also, we amplified each DNA sample at least twice to ensure accuracy of identifying mutant versus wild-type (or presence versus absence) of the *TAS1R2* gene variant.

Despite our study's statistically significant results, it had some limitations. The small sample size of 20 students was the major limitation of this study. In addition, the age demographic of college student participants was skewed to represent a younger population. Also, our study had some technical challenges as a limitation that could have interfered with the validity of our results. College students tend to have poorer oral health due to their diet and their failure to maintain good oral hygiene habits. In addition, tobacco use and caffeine intake were not assessed in this study. Another technical challenge could have been the blood pressures that were taken. The high blood pressures could have resulted from different factors not related to the gene variants such as caffeine consumption, physical exercise, medications, or from the White Coat Syndrome. Finally, teeth impressions, caries counting, and enamel measuring were not performed by a professional dental specialist. Therefore, the number of cavities and the enamel sizes could have been somewhat skewed. In future studies, a dentist could be employed to perform a professional dental exam.

CONCLUSION

To our knowledge, no studies have been conducted to date on whether a connection exists between *TAS* receptors and cardiovascular health. The present study offers preliminary data, albeit on a small sample, on the potential link between the *TAS1R2* variant and signs of future heart disease. Further, our findings seem to support previous findings where females have a lower prevalence of periodontitis than males,[19] which may suggest a link between gene variant expression, sex, and cardiovascular health. Perhaps future research could target the potential mechanisms behind the difference in rates of periodontitis between the sexes further to determine if this is primarily due to genetic reasons, *e.g., TAS1R2* variant, or some other external factors.

In the future, this research could lead to further experiments on the taste pathway gene variants that are specifically associated with dental caries, gum disease, and heart disease. Variants of *TAS1R2* could be genotyped and used to characterize large datasets according to expected and observed values as it relates to Hardy-Weinberg probabilities. As mentioned earlier, this research highlights the relevance of future studies on *TAS1R2* and cardiovascular disease. In addition, genetic screenings for *TAS1R2* gene variants could be conducted to assist with preventative dental measures and give indication of pre-existing health conditions (cavities, gum disease, and heart disease). Rostami et al. (2017) stated that people presenting the gene variant have to visit the dental office more often than non-expressers in order to eliminate the plaque on their teeth that leads to the formation of cavities.[9] Screening kits could be assembled and sold to dental clinics, or a screening service could be provided to dental clinics, dentists, dental hygienists, and dental labs. These screenings could help practitioners and insurance companies identify patients that are more prone to cavities, gum disease, and heart disease, and help create more preventative, proactive dental plans for these individuals.

ACKNOWLEDGEMENTS
Authors thank Northern State University for the resources and assistance in conducting this research project and the Northern State University Institutional Review Board for approving this project. Additionally, authors thank the twenty participants for volunteering their time, saliva, and any inconvenience experienced while getting their teeth impressions taken for the benefit of this research.

REFERENCES

1. Wendell, S., Wang, X., Brown, M., Cooper, M. E., DeSensi, R. S., Weyant, R. J., Crout, R., McNeil, D. W., and Marazita, M. L. (2010) Taste genes associated with dental caries, *J Dent Res* 89, 1198–1202. *https://doi.org/10.1177/0022034510381502*
2. Hollá, I. L., Borilova, L. P., Lucanova, S., Kastovsky, J., Musilova, K., Bartosova, M., Kukletova, M., Kukla, L., and Dusek, L. (2015) GLUT2 and *TAS1R2* polymorphisms and susceptibility to dental caries, *Caries Res* 49, 417–424. *https://doi.org/10.1159/000430958*
3. Gupta, P., Gupta, N., Pawar, P. A., Birajdar, S. S., Natt, S. A., and Singh, P. H. (2013) Role of sugar and sugar substitutes in dental caries: a review, *International Scholarly Research Notices: Dentistry* 2013, 1–5. *https://doi.org/10.1155/2013/519421*
4. Selwitz, H. R., Ismail, I. A., and Pitts, B. N. (2007) Dental caries, *The Lancet* 369, 51-59. *https://doi.org/10.1016/S0140-6736(07)60031-2*
5. Bartlett, D. W., Lussi, A., West, N. X., Bouchard, P., Sanz, M., and Bourgeois, D. (2013) Prevalence of tooth wear on buccal and lingual surfaces and possible risk factors in young European adults, *J Dent* 41, 1007–1013. *https://doi.org/10.1016/j.jdent.2013.08.018*
6. Zawaideh, F. I., Owais, A. I., and Mushtaha, S. (2017) Effect of CPP-ACP or a potassium nitrate sodium fluoride dentifrice on enamel erosion prevention, *J Clin Pediatr Dent* 41, 135–140. *https://doi.org/10.17796/1053-4628-41.2.135*
7. Curson, M. E. J., and Duggal, M. S. (2003) Dental Disease - Structure of Teeth, in *Encyclopedia of Food Sciences and Nutrition* (Caballero, B., Ed.) 2nd ed., 1743–1746, Academic Press, New York. *https://doi.org/10.1016/B0-12-227055-X/00324-2*
8. Marsh, P. D. (2010) Microbiology of dental plaque biofilms and their role in oral health and caries, *Dent Clin North Am* 82*,* 4–8. *https://doi.org/10.1016/j.cden.2010.03.002*
9. Rostami, N., Shields, R. C., Yassin, S. A., Hawkins, A. R., Bowen, L., Luo, T. L., Rickard, A. H., Holliday, R., Preshaw, P. M., and Jakubovics, N. S. (2017) A critical role for extracellular DNA in dental plaque formation, *J Dent Res* 96, 208–216. *https://doi.org/10.1177/0022034516675849*
10. Hujoel, P. P., and Lingström, P. (2017) Nutrition, dental caries and periodontal disease: a narrative review, *J Clin Periodontol* 18, S79–S84. *https://doi.org/10.1111/jcpe.12672*
11. Singer, R. E., and Buckner, B. A. (1981) Butyrate and propionate: important components of toxic dental plaque extracts, *Infect. Immun* 32, 458–463.
12. Jepsen, S., Blanco, J., Buchalla, W., Carvalho, J. C., Dietrich, T., Dörfer, C., Eaton, K. A., Figuero, E., Frencken, J. E., Graziani, F., Higham, S. M., Kocher, T., Maltz, M., Ortiz-Vigon, A., Schmoeckel, J., Sculean, A., Tenuta, L. M., van der Veen, M. H., and Machiulskiene, V. (2017) Prevention and control of dental caries and periodontal diseases at individual and population level: consensus report of group 3 of joint EFP/ORCA workshop on the boundaries between caries and periodontal diseases, *J Clin Periodontol* 18, S85–S93. *https://doi.org/10.1111/jcpe.12687*
13. Dhadse, P., Gattani, D., and Mishra, R. (2010) The link between periodontal disease and cardiovascular disease: How far we have come in last two decades? *J Indian Soc Periodontol* 14, 148–154. *https://dx.doi.org/10.4103%2F0972-124X.75908*
14. Humphrey, L. L., Fu R., Buckley, D. I., Freeman, M., and Helfand, M. (2008) Periodontal Disease and Coronary Heart Disease Incidence: A Systematic Review and Meta-analysis, *J Gen Intern Med* 23, 2079–2086. *https://doi.org/10.1007/s11606-008-0787-6*

15. Hearn, R. P., and Arblaster, K. E. (2010) DNA Extraction techniques for use in education, *Biochemistry and Molecular Biology Education* 38, 161–166. *https://doi.org/10.1002/bmb.20351*
16. Ramos-Lopez, O., Panduro, A., Martinez-Lopez, E., and Roman, S. (2016) Sweet taste receptor TAS1R2 polymorphism (Val191Val191) is associated with a higher carbohydrate intake and hypertriglyceridemia among the population of West Mexico, *Nutrients* 8, 101. *https://doi.org/10.3390/nu8020101*
17. Dietrich, T., Webb I., Stenhouse, L., Pattni, A., Ready, D., Wanyonyi, K. L., White, S., and Gallagher, J. E. (2017) The relationship between oral and cardiovascular disease, *Br Dent J* 222, 381–385. *https://doi.org/10.1038/sj.bdj.2017.224*
18. Friedemann, C., Heneghan, C., Mahtani, K., Thompson, M., Perera, R., and Ward, A. (2012) Cardiovascular disease risk in healthy children and its association with body mass index: systematic review and meta-analysis, *BMJ* 345, 1-16. *https://doi.org/10.1136/bmj.e4759*
19. Shiau, H. M., and Reynolds, M. A. (2010) Sex differences in destructive periodontal disease: a systematic review, *J Periodontol* 81, 1379–1389. *https://doi.org/10.1902/jop.2010.100044*

ABOUT THE STUDENT AUTHORS

Joseph Keryakos graduated from Northern State University in Aberdeen, SD, on August 18, 2017, with a Bachelor of Science Degree in Biology *In Honoribus* and with a minor in Chemistry. Joseph is currently a third year Doctor of Pharmacy (Pharm.D.) candidate at South Dakota State University in Brookings, SD. He is also working as a clinical pharmacy intern at Avera McKennan Hospital & University Health Center in Sioux Falls, SD. Annika van Oosbree graduated from Northern State University in May 2018 with a Bachelor of Science in Biology *In Honoribus* and in August 2018 with a Bachelor of Arts in English with a literature emphasis. She is currently a first year MD/PhD candidate at the University of South Dakota Sanford School of Medicine in Vermillion, SD.

PRESS SUMMARY

Despite brushing and/or flossing their teeth twice daily, many people are still susceptible to dental caries and tooth decay. This research investigated the genetics behind this phenomenon. Two gene variants were tested, taste 1 receptor member 2 (*TAS1R2*) and taste 2 receptor member 38 (*TAS2R38*). One gene variant, *TAS1R2*, may show a link between its presence and dental and cardiovascular health. Though we only sampled 20 individuals (10 males and 10 females), our results indicate that the *TAS1R2* gene variant may play a role in cavity development and cardiovascular health.

Body Image and Self-Esteem in Female College Students of Healthy Weight and Excess Weight: The Mediating Role of Weight Stigma

Carlie Smith, Jennifer Becnel, & Amanda Williams*

School of Human Environmental Sciences, University of Arkansas, Fayetteville, AR

https://doi.org/10.33697/ajur.2019.025

Students: crs020@uark.edu
*Faculty Advisors: *becnel@uark.edu, alw077@uark.edu*

ABSTRACT
Emerging adulthood is an important transition in which the development of lifelong behaviors emerge. Recent research suggests that women in college are particularly vulnerable to poor body image and low self-esteem. This is also a time of possible weight gain as individuals learn to eat and exercise on their own. These are important as body mass index (BMI) influences how women feel about themselves and how others view them. Thus, the purpose of the present study was to examine the associations between body image, self-esteem, and weight stigma among female college students of healthy weight and excess weight. Participants (n=124; 83% White) were recruited to take a short survey administered online through Facebook advertisements and snowball sampling. Results reveled poorer body image and more experiences with weight stigma among individuals with excess weight. Additionally, weight stigma fully mediated the relationship between BMI and self-esteem as well as BMI and body image. Results were consistent with previous research noting the stigma and stereotypes associated with excess weight. Future work should examine these relationships in more diverse groups to identify those at greatest risk for negative self-concept for intervention.

KEYWORDS
Weight; Weight Stigma; Self-Esteem; Body Image; College Students; Women

INTRODUCTION

Emerging adulthood is a time full of transitions. Although this is a relatively new developmental period, it is a time that is crucial for the development of an individual.[1] An emerging adult is no longer an adolescent, but they are not yet a full adult either.[1] Between the ages of 18 to 25, emerging adults are just getting to know themselves and figure out how they will spend the rest of their lives.[2] Approximately one-third of people attend college right after high school. It is known that college students are exposed to a variety of stressors.[3,4] Considering that college students face increased stressors in the academic atmosphere, there is evidence to suggest that emerging adults are a group of people who are considered "at risk" for maladaptive behaviors including drinking and other risk taking behaviors as well as and internalizing and externalizing difficulties such as depression, anxiety, and risk-taking behavior.[5] Depression, anxiety, binge drinking, drug use, casual sex without protection, eating disorders, weight gain, and drunk driving are examples of risky behavior that are prevalent during the college years.[5] These are particularly problematic considering that the behavioral patterns developed in the emerging adult years have the potential to impact health in adulthood.[6]

Theory: Emerging Adulthood and Identity Development

Moving from adolescence to adulthood is a significant developmental transition. This transition was well-structured for most until the mid-20th century but has recently shifted to being unstructured.[5] Many factors have contributed to a lack of transitional structure, including delayed age at marriage and parenthood.[1] In other words, instead of early adulthood being a time of settling down and commitment in the form of marriage, it has become a time that is full of exploration.[2] This exploration and change has made the late teens to early twenties a distinct developmental period in the lifespan, called emerging adulthood.[1] Not only do young adults experience independence from normal social roles and expectations, but they also engage in identity exploration.[7] Specifically, identity exploration becomes more distinct relative to adolescence. Erik Erikson has proposed the theory of psychosocial development where identity development is crucial for the transition from adolescence into adulthood.[8] For long term success, emerging adults need to be confident and firm in their identity development and to form their identity as they explore of various paths.[1] Identity exploration focuses on domains such as love, work, and worldviews and mainly takes place during emerging adulthood.[9] For example, those that struggle with their own identity may be at risk for low self-esteem, depression and anxiety, substance use, and risky sexual behaviors.[5,10] Additionally, other psychosocial characteristics, like self-efficacy, developed during this time and become important in the establishment of healthy behavioral patterns.[6] Thus, the unique nature of this developmental period makes it an important age group to study.

Weight Status
Currently, the United States is going through an obesity epidemic.[8] Since 2014, 70.7% of adults aged 20 or older were overweight or obese while obesity in emerging adults has doubled in the past 30 years.[11] Although obesity is becoming common with increased prevalence rates, those who are overweight experience a myriad of negative outcomes.[11] These range from physical (*i.e.* Type 2 diabetes, hypertension, stroke) to emotional (*i.e.,* higher depression, anxiety, low self-esteem, poor body image), and social (*i.e.,* peer rejection, weight-based discrimination).[12]

Weight and Body Image and Self-Esteem
One domain that influences emerging adults well-being is body image.[13] Body image satisfaction is one's overall opinion or attitudes toward their body figure.[14] Unfortunately, not all college students feel positive about their bodies. Since the 1980s, body image dissatisfaction has become common among college students; in fact, up to 90% of college students say they are concerned about their own body image and these body image concerns are considered normal for college women.[13] Body image dissatisfaction is associated with symptoms of depression, anxiety, lower self-esteem, and eating disorders.[13] It has been suggested that society's emphasis on a thin figure contributed to body dissatisfaction among women.[15] Magazines, television, music videos, and movies are a few ways that society contributes to an unrealistic ideal body image, and constant exposure to these socialized ideals adds to women's struggles for body satisfaction.[15] More often than not, those that were concerned with their weight had an unrealistic body image for themselves.[14] Typically, individuals compare their bodies to the ideal body, as well as the bodies of those around them.[15] Those who constantly compare themselves to others may have more negative mental health outcomes (*e.g.*, lower self-worth/self-esteem).[7,13] Not only is body image perception subjective, but it can also be manipulated through the influence of others through weight stigma.[13]

The pressure college women face to be thin and attractive contributes to a lower self-esteem among these students.[13] There are multiple domains of self-esteem, but one that is distorted the most is physical self-worth. This domain of self-esteem is constructed by physical condition, weight concerns, and attractiveness.[13] Physical self-worth and dissatisfaction is heavily influenced by social ideals and experiences of objectification (called an "internalized third-person perspective" by Mercurio and Landry; 2008, p. 458) of their appearance.[16] College students who have a weight problem face additional issues that can lower their self-esteem, including peer rejection, depression and anxiety, and weight stigma.[14] Experiencing body dissatisfaction during this sensitive developmental period can lead to negative health outcomes in adulthood through unhealthy dietary and physical behaviors and substance use.[17]

Weight and Weight Stigma
With the focus on obesity from the medical and public health fields, the media has also focused on obesity and contributes to weight stigma.[8] Weight stigma is defined as the social devaluation of individuals perceived to carry excess weight and leads to prejudice, negative stereotyping, and discrimination toward those individuals.[18] Mass media contributes to anti-fat stereotypes of individuals with obesity through negative portrayals that people with larger bodies are less sophisticated, less intelligent, lazy, weak, and self-indulgent.[8,19,20] This further perpetuates assumptions that those with excess weight are lonely, slow, unhappy, ashamed, and sloppy.[21] In contrast, thin figures are portrayed more positively and assumed to be healthy due to self-control and hard work; therefore perpetuating the assumption that people with an ideal, thin figure are more confident, more attractive, and happier.[21]

Size-based and anti-fat attitudes do not begin in adolescence or adulthood, rather they begin during childhood.[21] Overweight children are often rejected by friends because of their excess weight.[21] Those who are overweight experience teasing and are rated as unpopular which can lead to low self-esteem and body dissatisfaction.[21] Experiences with weight stigma have negative consequences for physical and psychological well-being that can continue into adulthood including, increased weight and unhealthy weight management behaviors driven by a desire to be thin.[21]

Weight stigma and anti-fat attitudes contribute to poorer physical and mental health. Individuals can experience weight stigma from a variety of sources including medical professionals, employers, police officers, friends, and family.[18] When women who are overweight or obese experience discrimination because of their weight, they are less likely to engage in physical activity. There are also multiple unfavorable eating behaviors that arise due to weight stigma including overeating, dieting, and eating disorders.[18] When an individual experiences weight stigma, they are likely to have increased depression, body dissatisfaction, and lower self-esteem.[14] Examining college students' weight stigma will aid in understanding how to avoid the negative effects of those stereotypes and prevent poor body image and low self-esteem.[21]

Present Study
Although weight stigma has garnered attention in the medical fields and has been shown to be related to obesity and self-esteem, research has not yet examined weight stigma as a mediator between weight, self-esteem, and body image. This study utilized a non-experimental correlational design to examine the associations between weight stigma, body image, and self-esteem in female college students of healthy weight and excess weight. The primary aims and hypotheses of this study are:

1. Examine the relationship between body image, self-esteem, weight stigma by weight status. It is hypothesized that body image, self-esteem, and weight stigma will be correlated for women with excess weight, but will not be correlated for women of healthy weight.
2. Examine if there are differences in body image, self-esteem, and weight stigma by weight status (i.e., healthy weight and excess weight). It is hypothesized that women with excess weight will have poorer body image, lower self-esteem, and report greater weight stigma compared to women of healthy weight.
3. Examine if weight stigma mediates the relationship between BMI and self-esteem as well as BMI and body image. It is hypothesized that weight stigma will mediate the relationship between BMI and self-esteem and BMI and body image.

METHODS AND PROCEDURES
Participants
The participants in this study were 125 women between the ages of 18 and 25 (M_{age} = 21.75; *SD* = 4.48) attending various universities across the United States. Of the individuals who completed this study, 5.6% were Black or African American, 3.2% Asian or Pacific Islander, 4.8% Hispanic or Latino, and 83.2% White or Caucasian. Of the participants, 13% were freshmen, 24% were sophomores, 25% juniors, 11% seniors, 12% graduate students, and 15% were other student status. Eighty percent of the sample identified as heterosexual and 56% identified as single and not dating.

Procedures
After obtaining IRB approval, participants were recruited through advertisements in Facebook groups and pages as well as through snowball sampling. The data for this study was gathered in the August to December 2018. Data was collected using a questionnaire created in Qualtrics.

Measures
1. Demographics. Participants in this study were asked background questions including age, gender, height, weight, race/ethnicity, and year in school. To measure gender, participants were asked which gender they identified most with (1 = male; 2 = female). Participants were asked what year they currently are in college (1 = Freshman; 2 = Sophomore; 3 = Junior; 4 = Senior; 5 = Graduate Student; 6 = Non-Degree Student; 7 = Other). Racial group was established by asking participants how they would describe their race/ethnicity (1 = Black or African American; 2 = American Indian or Native American; 3 = Asian or Pacific Islander; 4 = Hispanic or Latino; 5 = White or Caucasian; 6 = Multicultural). Participants also reported their sexual orientation (1 = heterosexual; 2 = lesbian; 3 = gay; 4 = bisexual; 5 = questioning; 6 = queer; 7 = other) and relationship status (1 = single, not dating; 2 = single, dating multiple people; 3 = single, but in a committed relationship; 4 = married; 5 = divorced; 6 = separated; 7 = widowed; 8 =other).

2. Body Mass Index (BMI). Height was self-reported in feet and inches and weight was also self-reported in pounds. Height and weight were used to calculate BMI for each participant (kg/m^2).

3. Self-Esteem. Self-esteem was assessed using a one-item question that stated "I have high self-esteem" and asking participants to rank whether they agree with that statement (1 = Strongly agree, 2 = Agree, 3 = Neutral, 4 = Disagree, and 5 = Strongly disagree). This single item has been shown to be as valid and reliable in men and women of various racial/ethnic backgrounds and in college students.[22]

4. Body Shape Questionnaire Short Form (BSQ). The Body Shape Questionnaire was used to evaluate participants' body image perception with emphasis on "feeling fat."[15, 16] This 16-item questionnaire was developed and validated in 1993 and was adapted from the original 34-item questionnaire developed and validated in 1987. The results are found by summing all the items in the questionnaire (1 = Never, 2 = Rarely, 3 = Sometimes, 4 = Often, 4 = Very Often, 5 = Always; α = 0.95).[23]

5. Weight Self-Stigma Questionnaire (WSSQ). The Weight-Self Stigma Questionnaire was used in this study to measure the level of stigma associated with obesity. This twelve-question scale will also be used to measure how one perceives their own body image when exposed to experiencing weight discrimination/stigma and internalization of stigma (1 = Completely Disagree, 2 = Mostly Disagree, 3 = Neither Agree nor Disagree; 4 = Mostly Agree, 5 = Completely Agree; α

= 0.90).[18] In previous research conducted to validate the scale, it was found that the WSSQ correlates with BMI, psychological distress, and gender.[18] Total subscale was the sum of all the items. There were two subscales, fear of enacted stigma and self-devaluation. The fear of enacted stigma, referring to external stigma or discrimination experienced by others because of one's weight, was the sun of 1-6. The self-devaluation subscale, referring to attributing negative qualities to oneself in regards to one's weight, was the sum of items 7-12.

Data Analyses
Before testing the study aims, BMI was divided into healthy weight (18.5 $\leq$ BMI $\leq$ 24.9) and excess weight (BMI $\leq$ 25). Those that reported a low body weight were led to the exit of the survey and thus not included in analyses. To test the first aim examining the relationships between body image, weight stigma, and self-esteem, a series of correlations were run. Correlations were run separately for healthy weight and excess weight. To test the second aim and examine if there were differences in the study variables by weight status, a series of independent samples t-tests were run comparing women of healthy weight and excess weight. To test the third aim examining if weight stigma mediated the relationship between BMI and self-esteem, Hayes' PROCESS, via SPSS v.24 was used.[24] After specifying a simple mediation model (model 4), Hayes' PROCESS uses ordinary least squares regression and estimates each path's significance, coefficients, and confidence intervals while also providing direct and total effects of the model.

RESULTS

Before testing the aims, participants were divided into groups based on BMI. Healthy (n = 55) and excess weight (n = 62) BMIs were determined using Centers for Disease Control defined criteria. To test the first aim, correlations were run separately for each group. As shown in **Table 1**, for women of healthy weight, self-esteem was negatively associated with body image dissatisfaction and total weight stigma and weight stigma subscales. Indicating that, as body image dissatisfaction and weight stigma increased, self-esteem decreased. Body image dissatisfaction was also significantly positively correlated with weight stigma. As body image dissatisfaction increased so did experiences with weight stigma. Interestingly, the associations between the study variables were in the same directions and of similar magnitude for women with excess weight.

Indicator	1.	2.	3.	4.	5.
1. Self-Esteem	-	-.63**	-.53**	-.51**	-.46**
2. BSQ	-.57**	-	.67**	.59**	.62**
3. WSSQ Total	-.44**	.73**	-	.88**	.92**
4. WSSQ Fear	-.40**	.69**	.93**	-	.62**
5. WSSQ Self	-.43**	.68**	.94**	.75**	-

Table 1. Correlations among Study Variables with Healthy Weight below the Diagonal and Excess Weight Above the Diagonal.

To test the second aim examining differences in self-esteem, body image, and weight stigma between women of healthy weight and excess weight, independent samples t-tests were run. As shown in **Table 2**, there were no significant differences in self-esteem. However, women with excess weight did report significantly greater body dissatisfaction (t [115] = -3.45, p < .001) and weight stigma (t [115] = -3.19, p < .01). Women with excess weight were also significantly higher on both subscales of the WSSQ, fear (t [115] = -3.15, p < .01) and self-devaluation (t [115] = -2.735, p < .01). Women of healthy weight had significantly higher dislike of fat (t [115] = 2.11, p < .05), compared to women's of excess weight.

	Total Sample n = 124	**Healthy Weight n = 55**	**Excess Weight n = 62**	***t*-test**
Variable	**M(SD)**	**M(SD)**	**M(SD)**	
BMI	26.93(6.53)	21.81(1.51)	31.47(5.88)	--
Self-Esteem	3.22(1.10)	3.40(1.05)	3.05(1.11)	ns.
BSQ	32.17(17.31)	27.06(15.89)	37.68(16.75)	-3.45***
WSSQ	20.45(10.37)	17.45(10.77)	23.61(8.99)	-3.19**
Fear	10.75(5.38)	9.20(5.61)	12.35(4.63)	-3.15**
Self-Dev	9.69(5.81)	8.25(5.91)	11.25(5.36)	-2.73**

Table 2. Means, Standard Deviations, and t-test Results Comparing Women of Healthy Weight and Excess Weight on the Study Variables.

To test the third aim examining if weight stigma mediates the relationship between BMI and self-esteem and BMI and body image, Hayes' mediation model via PROCESS was used for the total sample. Mediation was tested separately for self-esteem and body image for the entire sample. The results are shown in **Figure 1, Table 3,** and **Table 4**. Hayes' PROCESS estimates pathways A (the independent variable to the mediator), B (the mediator to the outcome), C (the independent variable to the outcome), and C' (the independent variable to the outcome through the mediator).

For the self-esteem, pathways A and B were significant. The total effect of BMI on self-esteem, pathway C, was significant (b = .97, p<.001), while the direct effect, pathway C' (accounting for the mediator), was non-significant (b = -.01, p= .83). Similar results were found in the body image model; pathways A and B were significant. The total effect of BMI on body image, pathway C, was significant (b = -.03, p= .02), while the direct effect, pathway C', was non-significant (b = .28, p= .13). Both models showed significant indirect effects (unstandardized total indirect effect = -.03, Bootstrapped SE = .01, 95% CI = -.05, -.01 and unstandardized total indirect effect = .67, Bootstrapped SE = .16, 95% CI = .31, .97 for self-esteem and body image, respectively).

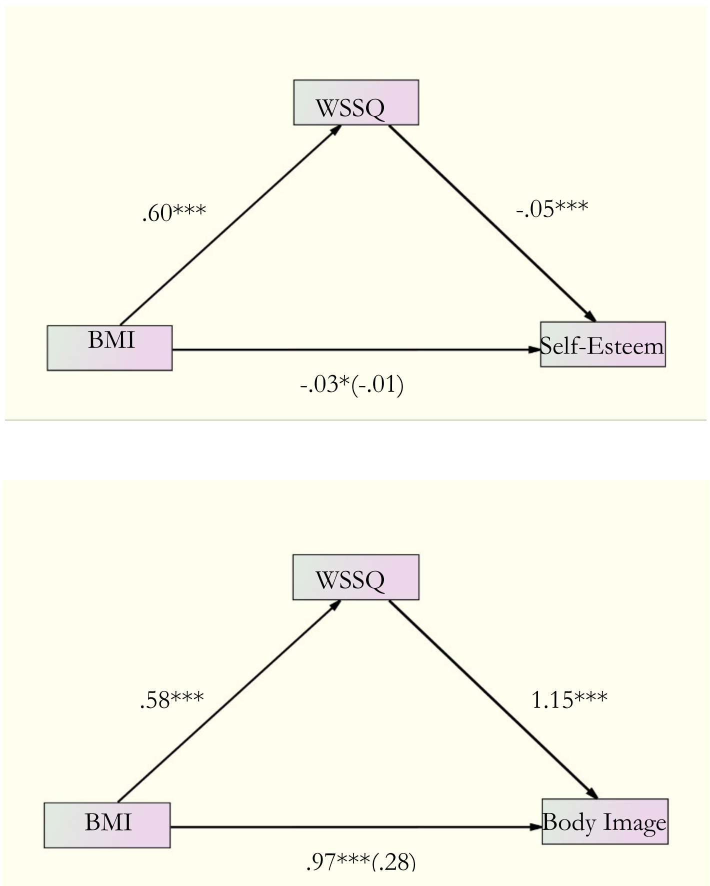


Figure 1. Standardized Regression Coefficients for the Mediation Models Testing Weight Stigma as a Mediator between BMI and Self-Esteem and BMI and Body Image.

Predictors	*b*	SE	t	CI_{lower}	CI_{higher}
Outcome: WSSQ (Path A)					
BMI	**.60*** **	.14	4.24	.32	.87
Outcome: Self-Esteem (Path B)					
	b	SE	t	CI_{lower}	CI_{higher}
WSSQ	**-.05*** **	.01	-5.43	-.07	-.03
Outcome: Self-Esteem (Path C)					
	b	SE	t	CI_{lower}	CI_{higher}
BMI	**-.03* **	.01	-2.21	-.06	-.01
Outcome: Self-Esteem (Path C')					
	b	SE	t	CI_{lower}	CI_{higher}
BMI	-.01	.02	-.20	-.03	.03
WSSQ	**-.05*** **	.01	-5.43	-.07	-.03

Significance is bolded. *p<.05, **p<.01, ***p<.001. All values are unstandardized.

Table 3. Mediation Analysis Using Self-Esteem.

Predictors	*b*	SE	t	CI_{lower}	CI_{higher}
Outcome: WSSQ (Path A)					
BMI	**.58*** **	.13	4.18	.31	.86
Outcome: Body Image (Path B)					
	b	SE	t	CI_{lower}	CI_{higher}
WSSQ	**1.15*** **	.12	9.38	.91	1.39
Outcome: Body Image (Path C)					
	b	SE	t	CI_{lower}	CI_{higher}
BMI	**.97*** **	.23	4.05	.49	1.42
Outcome: Body Image (Path C')					
	b	SE	t	CI_{lower}	CI_{higher}
BMI	.28	.18	1.51	-.08	.65
WSSQ	**1.15*** **	.12	9.38	.91	1.39

Significance is bolded. *p<.05, **p<.01, ***p<.001. All values are unstandardized.

Table 4. Mediation Analysis Using Body Image.

DISCUSSION

Developing positive body image and self-esteem are important when it comes to young women's overall well-being.[13] This study examined the associations between self-esteem, body image, and weight stigma among women of healthy weight and excess weight. Overall, there were strong relationships between self-esteem, body image, weight, and weight stigma. Specifically, lower self-esteem was associated with poorer body image and experiences with weight stigma. Additionally, body image dissatisfaction was associated with more experiences of weight stigma. Interestingly, these relationships were in the same direction and of a similar magnitude for both women of healthy and excess weight. This speaks to everyone feeling the pressure to be thin and even those that are healthy weight are susceptible to unrealistic body image expectations and report experiencing weight stigma. It is interesting that having a lower BMI and being of healthy weight does not exclude an individual from having issues with weight stigma or body shape issues. This is consistent with previous work noting that pressures to be thin or have an ideal body size can have a negative influence on one's self-esteem.[5] Physical self-worth is constructed by weight concerns and attractiveness and, therefore, can be built by one's level of self-esteem and body image perception. All of these variables have some sort of influence on one another and can have an impact on one's overall mental and physical health. This is important given that emerging adulthood is a critical time to develop positive body image and self-worth as these will be carried forward into adulthood.[5,21] These results suggest that more research should be done on the effect of weight stigma on not just unhealthy weight individuals, but healthy weight individuals as well.

Although the correlations were similar for women of healthy weight and excess weight, there were significant differences in the study variables by weight status when examined using t-tests. However, there was no difference in self-esteem between women of healthy weight and excess weight. These results are slightly divergent from other studies using adults or children that have noted self-esteem differences between women of healthy weight and excess weight. One possible explanation is that emerging adulthood is a unique time of development – and likely a critical turning point in the life course where behaviors and associated psychosocial well-being have long-term implications.[25] Emerging adults often struggle with self-esteem, particularly women. Emerging adulthood is also a time marked by heightened identity exploration when women are trying out different behaviors, careers, relationships, and ideologies as they figure out who they want to be.[1,9] Emerging adult women are particularly vulnerable to issues with self-esteem as they develop their identity and figure out what is important in terms of their overall self-concept.[13,15] Present study results further suggest that young women are susceptible self-esteem issues - regardless of weight status. Both groups of women (those with excess weight and those of average/healthy weight) showed similar self-perception patterns, suggesting that thin-ideal messages affect all women and are ubiquitous in our media and society. [8,19,20]

There were significant differences in body image and weight stigma for women of healthy weight and excess weight. Specifically, women with excess weight had greater body image dissatisfaction and more experiences with weight stigma. These results are consistent with previous work noting poorer psychosocial health of women with excess weight and the stigma and stereotypes associated with excess weight.[13] Given the emphasis from society on a thin body shape, it is unsurprising that women with excess weight have greater body image dissatisfaction and experience more weight stigma. Women with excess weight would be uniquely aware that they do not fit with the idealized body through the comparisons women often make to others. The greater body dissatisfaction experienced by women with excess weight is likely due to experiences with weight stigma.

This study also examined if weight stigma mediated the relationship between BMI and self-esteem and BMI and body image. Results showed significant full mediation, meaning that experiencing weight stigma completely accounted for the relationship between BMI and self-esteem and body image. These results support the idea that it is not necessarily an individual's weight that results in their self-esteem but rather it is the stigma society puts on being overweight. Overall, society places an emphasis on a thinner figure for women stating that the thinner you are, the better you are.[14] Specifically, the stereotype of people with excess weight is that they are lazy, lonely, slow, unhealthy, and sloppy.[8] If you are thin, you are happy, but if you are overweight, you are unhappy. Women receive these messages from a variety of sources including medical professionals, family and friends, and even from strangers and acquaintances; they then internalize these messages, which impacts the way they feel about themselves, resulting in poor body image and low self-esteem.[27]

Limitations

In this study, there were various limitations encountered by the researchers. One of the largest limitations of this study was the sample size. Participants were recruited through social media posts and snowball sampling in order to obtain the sample. Although there were plenty of White women, minority groups were lacking. Additionally, snowball sampling limits the ability of this study to generalize to the population and may introduce sample bias since participants were sharing the survey with their friends who are likely similar to them. Another limitation of the study was that this questionnaire could not be distributed to colleges with a more diverse student population. Had this questionnaire been able to reach other areas, the sample size would have had the potential to be more diverse. Additionally, this study used a correlational design, so temporal ordering of effects cannot be established. Finally, this study utilized a single item for self-esteem. While this item has been established as reliable and valid using other studies,[22] use of a single item does limit variability.

CONCLUSIONS

Emerging adulthood is a particularly important time for identity development and development of life long physical and psychological health. [5,10] With greater awareness of increasing trends in overweight and obesity, significant prevention and intervention efforts have been targeted for children, adolescents, and adults.[11] However, given the vulnerable nature of emerging adulthood as a life course turning point,[25] combined with the health and weight implications of transitioning out of the family home and into independent living (in college or otherwise),[26,27] focusing these efforts on women between the ages of 18-25 in ways that address unhealthy youth behaviors and shape adult health practices might have the most impact on national obesity trends and correlated health outcomes. Women are particularly vulnerable to experiencing low self-esteem and poor body image.[15] Additionally, society places pressures on young women to be thin and shuns women for the perception of being overweight.[8,19,20] These stigmas of having excess weight become internalized by young women and lead to poor self-esteem.[21] The present study supports these associations and found significant full mediation. That is, weight stigma completely accounted for the relationship between BMI and self-esteem and BMI and body image. Thus, it appears that it is not an individual's weight that influences how they feel about themselves, but the way they are treated by others.

ACKNOWLEDGEMENTS

The authors thank Dr. Kristin Jozkowski for her input and assistance in writing the thesis and serving on the Honors Committee. The authors also thank the Dale Bumpers College Honors Program and well as the University of Arkansas Honors College for the opportunity to pursue research. The authors thank everyone who made this manuscript possible, and for all the hard work put into it from all parties.

REFERENCES

1. Arnett, J.J. (2000) Emerging adulthood: A theory of development from the late teens through the twenties, *Am Psychol* 55, 469-480. *https://doi.org/10.1037//0003-066x.55.5.469*
2. Arnett, J.J., and Mitra, D. (2018) Are the features of emerging adulthood developmentally distinctive? A comparison of ages 18–60 in the United States, *Emerg Adulthood* 70, 569-576. *https://doi.org/10.1177/2167696818810073*
3. National Center for Education Statistics, The Condition of Education 2018, *https://nces.ed.gov/pubs2018/2018144.pdf* (accessed Mar 2019)
4. van der Zanden, P.J., Denessen, E., Cillessen, A.H., and Meijer, P.C. (2018) Domains and predictors of first-year student success: A systematic review, *Educational Research Review* 23, 57-77. *https://doi.org/10.1016/j.edurev.2018.01.001*
5. Schwartz, S.J., Hardy, S.A., Zamboanga, B.L., Meca, A., Waterman, A.S., Picariello, S., Luyckx, K., Crocetti, E., Kim, S.Y., Brittian, A.S., Robert, S.E., Whitbourne, S.K., Ritchie, R.A., Brown, E.J., and Forthun, L.F. (2015) Identity in young adulthood: Links with mental health and risky behavior, *J Appl Dev Psychol* 36, 39-52. *https://doi.org/10.1016/j.appdev.2014.10.001*
6. Nelson, M.C., Story, M., Larson, N.I., Neumark-Sztainer, D., and Lytle, L.A. (2008) Emerging adulthood and college-aged youth: an overlooked age for weight-related behavior change, *Obesity* 16, 2205-2211. *https://doi.org/10.1038/oby.2008.365*
7. Grossbard, J.R., Lee, C.M., Neighbors, C., and Larimer, M.E. (2009) Body image concerns and contingent self-esteem in male and female college students, *Sex Roles* 60, 198-207. *https://doi.org/10.1007/s11199-008-9535-y*
8. Major, B., Hunger, J.M., Bunyan, D.P., and Miller, C.T. (2013) The ironic effects of weight stigma. *J Exp Soc Psychol* 51, 74-80. *https://doi.org/10.1016/j.jesp.2013.11.009*
9. Vosylis, R., Erentaitė, R., and Crocetti, E. (2018) Global versus domain-specific identity processes: Which domains are more relevant for emerging adults? *Emerg Adulthood* 6, 32-41. *https://doi.org/10.1177/2167696817694698*
10. Persike, M., Seiffge-Krenke, I., Cok, F., Glogowka, K., Pavlopoulos, V., Tantaros, S., Perchec, C., Rohail, I., and Saravia, J.C. (2018) Emerging adults' psychopathology in seven countries: the impact of identity-related risk factors, *Emerg Adulthood* 67, 1-16. *https://doi.org/10.1177/2167696818791108*
11. Centers for Disease Control, Obesity and Overweight, *https://www.cdc.gov/obesity/index.html* (accessed Mar 2019)
12. Djalalinia, S., Qorbani, M., Peykari, N., and Kelishadi, R. (2015) Health impacts of obesity, *Pakistan Journal of Medical Sciences* 31, 239-245. *https://doi.org/10.12669/pjms.311.7033*
13. Lowery, S.E., Kurpius, S.E.R., Befort, C., Blanks, E.H., Sollenberger, S., Nicpon, M.F., and Huser, L. (2005) Body image, self-esteem, and health-related behaviors among male and female first year college students, *J Coll Stud Dev* 46, 612-623. *https://doi.org/10.1353/csd.2005.0062*
14. Chang, H.J., and Suttikun, C. (2017) The examination of psychological factors and social norms affecting body satisfaction and self-esteem for college students, *Fam Consum Sci Res J* 45, 422-437. *https://doi.org/10.1111/fcsr.12220*

15. Fitzsimmons-Craft, E.E., Harney, M.B., Koehler, L.G., Danzi, L.E., Riddell, M.K., and Bardone-Cone, A.M. (2012) Explaining the relation between thin ideal internalization and body dissatisfaction among college women: The roles of social comparison and body surveillance, *Body Image* 9, 43-49. *https://doi.org/10.1016/j.bodyim.2011.09.002*
16. Mercurio, A.E., and Landry, L.J. (2008) Self-objectification and well-being: The impact of self-objectification on women's overall sense of self-worth and life satisfaction, *Sex Roles* 58, 458-466. *https://doi.org/10.1007/s11199-007-9357-3*
17. Neumark-Sztainer, D., Paxton, S.J., Hannan, P.J., Haines, J., and Story, M. (2006) Does body satisfaction matter? Five-year longitudinal associations between body satisfaction and health behaviors in adolescent females and males, *J Adolesc Health* 39, 244-251. *https://doi.org/10.1016/j.jadohealth.2005.12.001*
18. Vartanian, L.R., and Novak, S.A. (2011) Internalized societal attitudes moderate the impact of weight stigma on avoidance of exercise, *Obesity* 19, 757-762. *https://doi.org/10.1038/oby.2010.234*
19. Langdon, J., Rukavina, P., and Greenleaf, C. (2016) Predictors of obesity bias among exercise science students, *Adv Physiol Educ* 40, 157-164. *https://doi.org/10.1152/advan.00185.2015*
20. Pearl, R.L., Puhl, R.M., and Brownell, K.D. (2012) Positive media portrayals of obese persons: Impact on attitudes and image preferences, *Health Psychol* 31, 821-831. *https://doi.org/10.1037/a0027189*
21. Greenleaf, C., Chambliss, H., Rhea, D.J., Martin, S.B., and Morrow, J.R. (2006) Weight stereotypes and behavioral intentions toward thin and fat peers among White and Hispanic adolescents, *J Adolesc Health* 39, 546-552. *https://doi.org/10.1016/j.jadohealth.2006.01.013*
22. Robins, R.W., Hendin, H.M., and Trzesniewski, K.H. (2001) Measuring global self-esteem: Construct validation of a single-item measure and the Rosenberg Self-Esteem Scale, *Pers Soc Psychol Bull* 27, 151-161. *https://doi.org/10.1177/0146167201272002*
23. The body shape questionnaire (BSQ) and its shortened forms, Revisited, 2011. *https://www.psyctc.org/tools/bsq/* (accessed Dec 2018)
24. Hayes, A.F. (2017) *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach.* Guilford Publications, New York.
25. Elder, G.H. (1998) The life course as developmental theory, *Child Dev* 69, 1-12. *https://doi.org/10.2307/1132065*
26. Becnel, J.N., Zeller, M.H., Noll, J.G., and Peugh, J. (2017) Romantic, sexual, and sexual risk behaviours of adolescent females with severe obesity, *Ped Obesity* 12, 388-397. *https://doi.org/10.1111/ijpo.12155*
27. Gillen, M.M., and Lefkowitz, E.S. (2011) Body size perceptions in racially/ethnically diverse men and women: Implications for body Image and self-esteem, *N Am J Psychol* 13, 1-5. *https://doi.org/10.1016/j.bodyim.2013.03.002*

ABOUT THE STUDENT AUTHOR

Carlie Smith is from Meade, Kansas and will graduate in May 2019 from the University of Arkansas with a Baccalaureate degree in Human Development and Family Sciences with a concentration in Child Development. Following graduation, Carlie will be attending Rockhurst University to pursue a Master's Degree in Occupational Therapy.

PRESS SUMMARY

This article examines the relationships between self-esteem, body image, and weight stigma in female college students. Findings suggest that women that are overweight experience weight stigma that contributes to low self-esteem and poor body image.

After Hurricanes Irma and Matthew: Living Shorelines Stabilize Sediments

Taryn Chaya, Jessica Veenstra, & Melissa Southwell*

Flagler College Department of Natural Sciences, St. Augustine, Florida

https://doi.org/10.33697/ajur.2019.024

*Student: tarynchaya@outlook.com**
Mentor: jveenstra@flagler.edu

ABSTRACT
Constructed intertidal oyster reefs, an example of a "living shoreline", can protect against erosion and loss of habitat, but can they prevent erosion during high-energy storm events such as hurricanes? Oyster reefs were constructed in 2012 within the Guana Tolomato Matanzas National Estuarine Research Reserve in Northeast Florida to stabilize the shoreline sediment and prevent erosion of an archeological site. Sediment cores were collected behind constructed oyster reefs before and after hurricanes Matthew (2016) and Irma (2017) to study changes in sediment particle size due to these high-energy storms. Pre-hurricane data were collected in 2016 from three different constructed reefs, as well as three control sites where no reef was present. Pre-hurricane sediment profiles behind the constructed reefs consisted of finer sediments, ~36% silt and clay, in the surface ~10-12 cm, with decreasing silt and clay and increasing sand content as depth increased. This was different than the sediment from the control sites with ~4% silt and clay in all depths sampled. Like the sediment profiles before the high energy storms, the post-hurricane sediment data showed a clear layer of finer sediment ~10-12 cm over coarser sediment. Although they were high-energy storms, the storms did not appear to significantly affect the sediment behind the constructed oyster reefs. Sediment profiles remained consistent after these storms but may not remain undisturbed during the next storm without some intervention because the oyster reefs have been degrading.

KEYWORDS
Oyster Restoration; Living Shorelines; Hurricanes; Coastal Erosion; Sedimentation; Salt Marsh

INTRODUCTION

Salt marshes are low-energy coastal wetland ecosystems that provide several ecosystem services, such as erosion control, water filtration, and nutrient cycling.[12,33] Salt marshes are naturally inhabited by organisms, such as cordgrasses (*Spartina alterniflora*) and eastern oysters (*Crassostrea virginica*), that trap sediment, prevent its resuspension from waves, and encourage sediment accretion.[15,32] In recent years, these habitats have been threatened due to the impacts of human development and climate change—sea-level rise, intense storms, and human construction can be detrimental to wetlands. A natural solution becoming more widely accepted—the construction of living shorelines—is effectively stabilizing sediment.[15,39] Living shorelines built along wetland shores, such as the construction of oyster reefs, replenishment of shoreline vegetation, or the use of other living materials, are designed to stabilize the shoreline from storms and boat wakes by providing a natural barrier, or breakwater.[3,23,39] Oysters promote biodiversity, healthier water quality, and shoreline stabilization[6,11]; however, 85% of oyster reef ecosystems have been lost globally due to overharvesting, disease, and habitat destruction.[2,18]

Other methods of preventing coastal erosion are particularly popular in areas of coastal development. Commonly used to prevent erosion, built infrastructure such as sea walls, levees, and other hardening techniques can be difficult to maintain and can potentially lead to worsening shoreline conditions. [1,38] Declining water quality and species diversity can result from the construction of these unnatural shoreline hardening techniques.[1,13] Furthermore, shoreline hardening can increase erosion rates as waves and storm surge scour the structures and promote sediment resuspension.[13] Despite mounting evidence showing the ineffectiveness of armored shorelines,[1,8,13] 14% of the continental U.S.' tidal coastlines are unnaturally hardened.[13] By using native plants and animals, living shorelines have been found to effectively attenuate waves, [15,23] encourage local biodiversity,[39] and recycle nutrients[14] more efficiently than other hard armoring techniques such as bulkheads. Living shorelines can mitigate the effects of storms without damaging the ecosystem in the long run.

Shorelines respond geomorphologically to several factors—a combination of storms, human development, and sediment supply can significantly change the shape and profile of a shoreline. [7,8,16,28] The combination of these conditions can influence whether a shoreline erodes or accretes sediment over time.[19] Tropical storms and hurricanes can produce strong winds, storm surges, and large quantities of rain that tend to erode sediment.[33] Extreme waves coupled with fluctuating water-levels influence the magnitude of shoreline change.[34] Sediment is often naturally taken from shorelines during storm surge events and deposited elsewhere, known as allogenic sedimentation;[31,41] however, exposed shorelines with little barriers allow large amounts of sediment to be resuspended and eroded away in the event of a storm.[26]

For a shoreline to be stable, sediment accretion must occur at a rate that equals or exceeds the rates of sea-level rise and erosion.[25] Salt marshes with low sediment accretion rates succumb to sea level rise, resulting in shoreward migration of plant and animal species.[41] In response to rising sea level, however, oyster reefs can vertically and laterally accrete, as changes in inundation time and aerial exposure provide the conditions for continued growth.[31] Furthermore, constructed oyster reefs can potentially outpace sea level rise at effectively higher rates than vegetation, such as cordgrasses and mangroves.[31] Although vegetation can trap sediment and prevent erosion like oysters, high energy waves can uproot vegetation.[12,21] Therefore, vegetation replenishment is a more effective living shoreline technique in areas of minimal boat wake and wave energy. On high energy coastlines, vegetative planting is more effective for buffering coastal erosion when it is in conjunction with other shoreline protection methods such as oyster reefs.[42]

The impacts of living shorelines in coastal habitats have been widely studied to gauge their usefulness in restoring ecosystem services.[1,14,38,39] However, there has been limited research done regarding how living shorelines affect sediment distribution after storm events and their short and long-term effects after those storms.

In June 2012, the Guana Tolomato Matanzas National Estuarine Research Reserve (GTM Research Reserve) in Ponte Vedra Beach, Florida, USA constructed 28 eastern oyster reefs to protect an archeological site and prevent shoreline erosion. The creation of oyster reefs along the Tolomato River (the Intracoastal Waterway) was intended to attenuate wave energy from wind, storms, and boat wakes. Following their construction, previous studies estimated sediment accretion rates to be 5 cm/yr directly behind the constructed reefs[35] based on the thickness of the fine sediment layer (surface depths ~0-10 cm in all reefs). One year after their restoration, the fine sediment reached a volume of 0.36 m^3 behind the created reefs that were tested, compared to a fine sediment volume of 0.12 m^3 at the control sites.[35] The shift in particle size distribution in the surface depths indicated the created oyster reefs' ability to accumulate fine sediment and alter sediment profiles.

In October 2016 and September 2017, storms Matthew and Irma, a category two hurricane and tropical storm respectively when they reached the study site,[5,36] caused considerable damage to man-made structures in the community through storm surge and high winds. The GTM Research Reserve's System Wide Monitoring Program (SWMP) database showed turbidity and water levels increasing significantly during both storm events. Typically, turbidity values range from 4-20 NTU. During Matthew, turbidity was approximately 100 NTU and during Irma it was approximately 144 NTU. Water level typically ranges between approximately 2.5-4 m. During Matthew, water level reached approximately 4.5 m and during Irma water level reached approximately 5.2 m.[27] The study site experiences frequent boat wakes and an average 1.6 m tidal range.[17] The climate is temperate year-round, and the site's monthly average salinity ranges between 25-32 ppt.[27]

Our goal in this study was to investigate whether the constructed oyster reefs could effectively stabilize sediment following two major storms. If the oyster reefs were functioning as predicted, we would see (1) no change in the depth of fine sediment or (2) no deposition of coarse sediment at the surface or (3) no significant differences in the particle size distribution behind oyster reefs before and after the storms.[4,30]

METHODS AND PROCEDURES

Recycled oyster shells were collected, put into plastic net bags, and placed in 5.3 m x 1.8 m x 0.5 m groups six meters apart along 315 m of the eastern shoreline of the Tolomato River.[35]

Sediment cores (22 cm depth; 7 cm diameter) were collected with a push corer before and after the storms to compare the sediment profiles and determine if the sediment had become disturbed due to these storms. Before the storms, in June 2016, one sediment core was taken from behind/landward each of three oyster reefs (2, 17, 21) and at three control sites, where no reefs were present (**Figure 1**). Post-storms, in December 2017, sediment cores were collected from behind three different reefs (15, 16, 18). Getting to the site and taking the core mixes the sediment. We did not sample the same reefs twice in order to avoid artifacts due to disturbance from previous coring.

Figure 1. A map of the study site, along the Tolomato River in the GTM Research Reserve, showing the locations of the constructed oyster reefs. Sediment cores before the storms were taken from behind reefs 2, 17, and 21. Sediment cores taken after the storms were collected from behind reefs 15, 16, and 18. Pre-storms control samples were taken at the same elevation at three sites south of the constructed reefs along the same shoreline as the reef samples.

Sediment cores were sectioned into layers 2 cm thick, oven dried at 80°C for 1-3 days and then sieved to allow sediment particles >2 mm to be removed. Thirty grams of each sieved sediment sample was added to a pre-weighed 1-L Nalgene bottle. Drawing from a method created by Erftemeijer and Koch (2001) for studying sediment particle size in seagrass habitats, each sample was treated with increments of 15 mL of 30% concentrated hydrogen peroxide every two hours to remove organic matter for 48 hours to two weeks.[10] Once visible chemical changes ceased (i.e. bubbling), organic matter was considered removed. Samples free of organic matter were placed into the oven overnight at 80^{0}C. The cooled, dry samples were weighed in grams to the nearest tenth—this value being the weight of the oven-dried sediment. One hundred milliliters of sodium hexametaphosphate, a dispersal chemical, was added to each sediment sample to separate the sediment particles before being shaken on an orbital shaker for two hours. The sediment samples were then wet sieved through a 63 μm sieve to separate the sand (>63 μm) from the silt (63 μm- 4 μm) and clay (<4 μm). The suspension of silt and clay particles sat undisturbed for at least an hour and a half in an 800 mL beaker (six hours maximum), while the 100 mL beaker of sand from the sediment sample could immediately be placed into the oven overnight. After the required hour and a half of complete rest, the clay suspended in the distilled water above the silt was poured off to allow the silt particles to be dried overnight. The following day, the oven-dried sand and silt samples were weighed. Following the simplified method for soil particle-size determination method synthesized by Kettler et al. (2001), the sand was then dry sieved in a sieve shaker composed of four layers: coarse (2 mm-500 μm), medium (500 μm-250 μm), fine (250 μm-125 μm), and very fine (125 μm-63 μm) sands, for fifteen minutes. Each layer of sediment was weighed in grams, allowing us to determine the percentages of each particle size by dividing each particle size weight by the total oven-dried sediment mass from each sample.[9,20]

Using the statistical software R, the Wilcoxon Ranked Sum Test determined if the percentages of the various particle size classes from cores collected before and after the storms were significantly different ($\alpha < 0.05$). This test was used because of our small sample size (n=3). Although several sand particle size percentages were calculated (i.e. very fine sand, fine sand, medium sand, coarse sand), fine sand and silt and clay percentages were primarily used to convey our data because they composed the majority of the sample. Silt and clay percentages were added together to account for pouring and weighing errors that occurred in the procedure.

We determined the thickness of the fine sediment layer for each reef sampled. We defined the lower depth boundary of the fine sediment layer as (1) a silt and clay percentage decrease of at least 15% or (2) an increase in fine sand percentage of at least 15%.

RESULTS

A fine sediment layer 10-12 cm deep was observed behind two out of the three reef sites in both the pre- and post-storm cores. When present, the fine layer consisted of ~36% silt and clay in the surface 10-12 cm, with a decrease in silt and clay (~18%) and increase in sand content in the deeper depths. The sediment cores from the control sites contained ~4% silt and clay at most depths sampled (**Figure 2**). The pre-storm control samples are presented to emphasize the differences between sites with oyster reefs and those without—the implementation of the oyster reefs increased the amount of fine sediment (**Figure 2**).

We used the Wilcoxon Ranked Sum Test to test the null hypothesis that the pre-storm thickness of the fine sediment layer was equal to the post-storm thickness of fine sediment layer. The average fine sediment layer was not significantly different after the storms (p-value=0.45). However, our sample size is small (n=3), and one reef from each pre-storm and post-storm sample did not exhibit the fine sediment layer at all (Reef 2 in the pre-storm samples and Reef 15 in the post-storm samples, **Figures 3 and 4**).

To determine if the storms deposited coarse sediment on the surface, we used the Wilcoxon Ranked Sum Test to test the null hypothesis that the pre-storm percent fine sand in the 0-4 cm depths was equal to the post-storm percent fine sand in the 0-4 cm depths. The average percentage of coarse sediment on the surface was not significantly different after the storms (p-value=0.90). Although the differences were not statistically significant, there were lower percentages of silt and clay, and higher percentages of fine sand in the top 0-6 cm in two of the three post-storm reefs vs. the pre-storm reefs. Unfortunately, we had samples from the surface 0-2 and 2-4 cm where we did not have enough sample to process, so there were missing data points for reefs 16, 18, and 21.

To examine differences in the particle size distribution, we compared depth profiles of averaged percentages of fine sand and silt and clay between the pre-storm data, and post-storm data. P-values for Wilcoxon Ranked Sum Test are reported for each depth sampled. Post-storm percentages of fine sand and silt and clay was not significantly different from the pre-storm data (**Figure 5**).

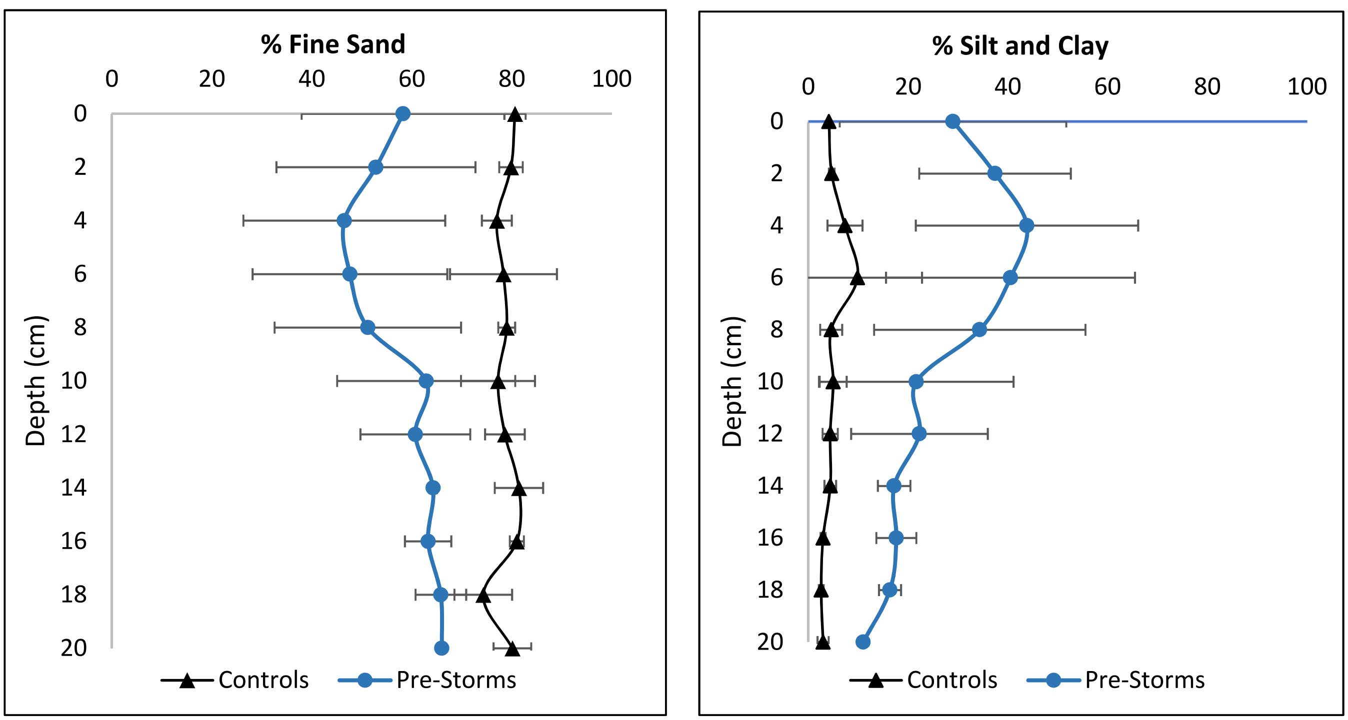


Figure 2. Percentages of fine sand (left), and silt and clay (right) plotted throughout the depths for control (black) and pre-storms (dark blue) data. The oyster reef construction increased the amount of finer sediment behind the reef.

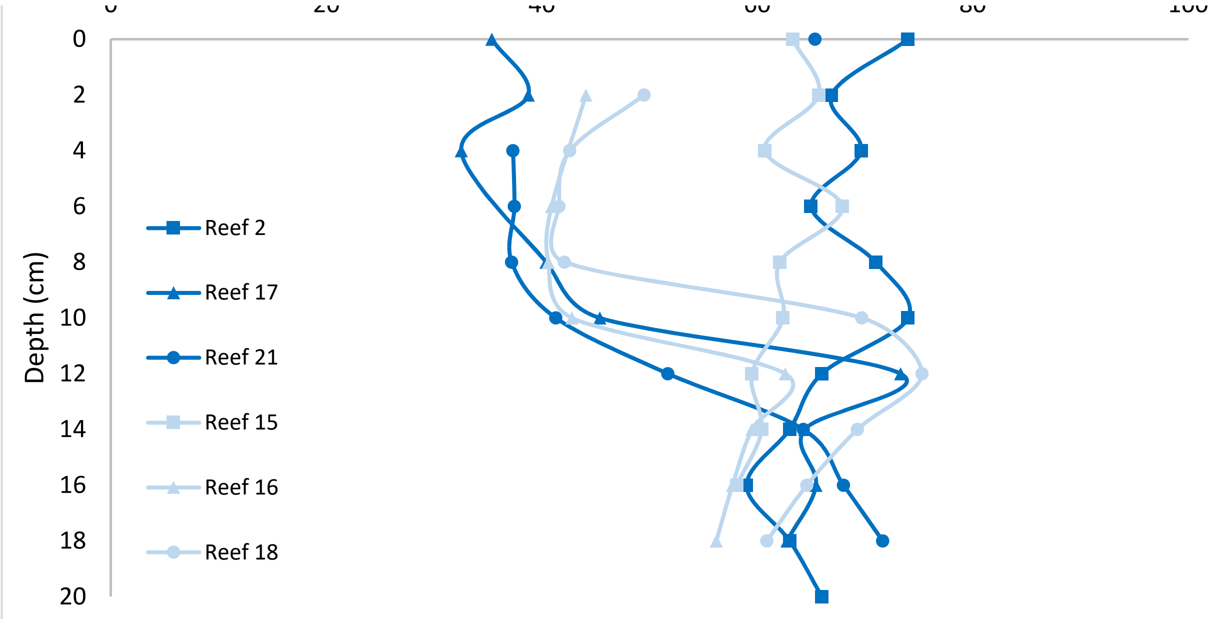
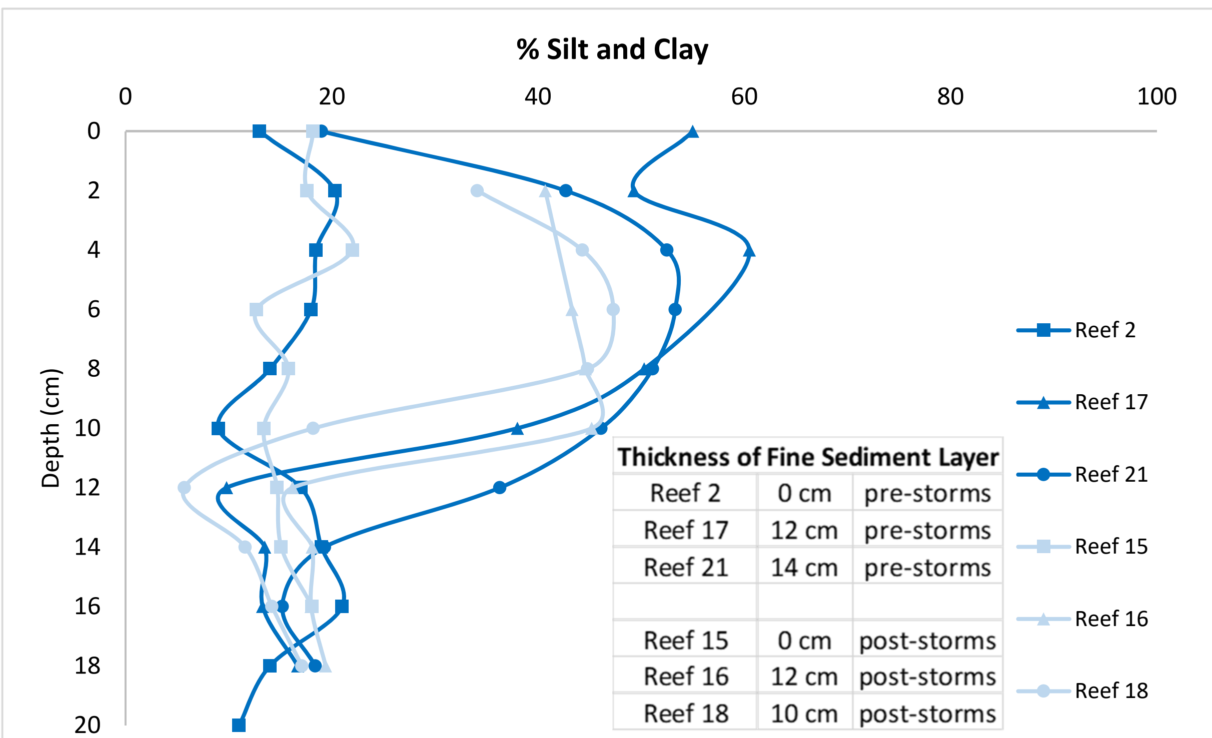


Figure 3. Individual percentages of fine sand for pre-storm (dark blue) and post-storm (light blue) sediment cores from depths of 0-20 cm.



Thickness of Fine Sediment Layer		
Reef 2	0 cm	pre-storms
Reef 17	12 cm	pre-storms
Reef 21	14 cm	pre-storms
Reef 15	0 cm	post-storms
Reef 16	12 cm	post-storms
Reef 18	10 cm	post-storms

Figure 4. Percentages of silt and clay in pre-storm (dark blue) and post-storm (light blue) individual sediment cores. The table shows the depths where the profile shifted from fine sediment to coarser sediment.

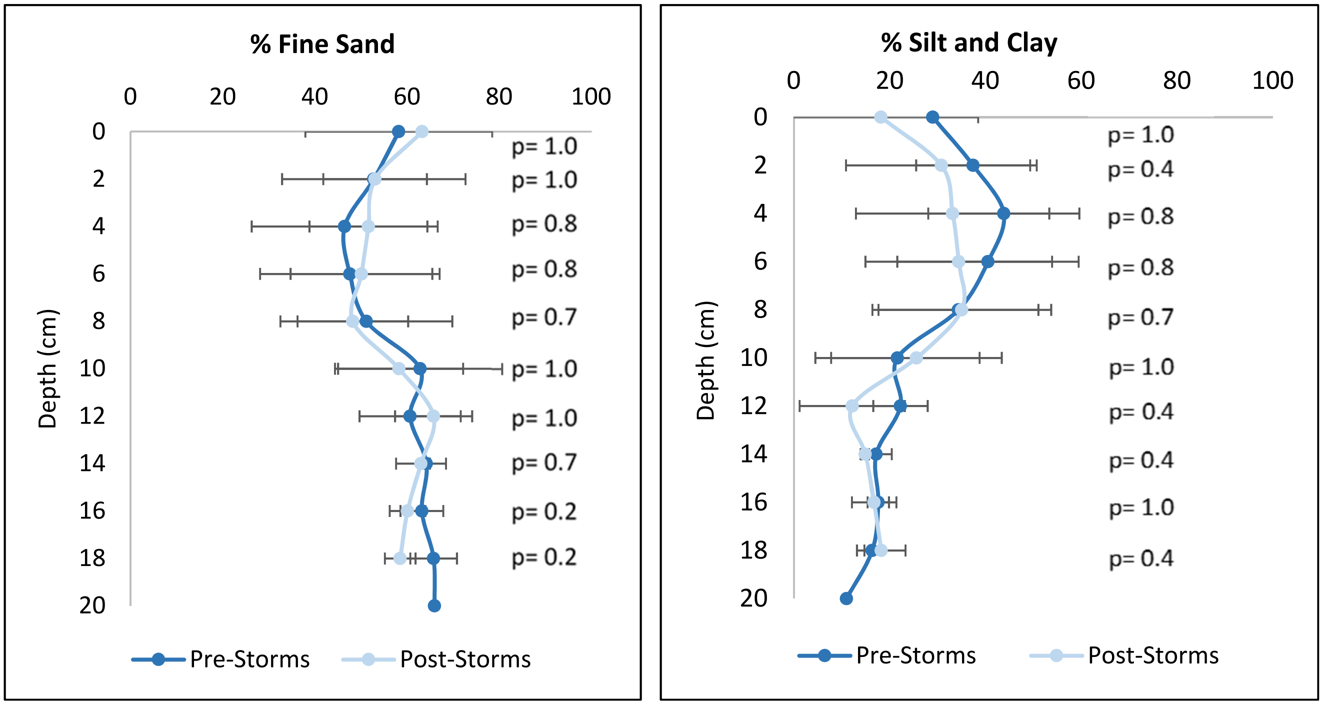


Figure 5. Average percentages of fine sand (left), and silt and clay (right) plotted throughout the depths for pre-hurricane and post-hurricane data with respective standard deviations. The overlapping standard deviations emphasize the similarities between the sediment profiles. Values on the right of each graph show the resulting p-values (n=3) when comparing the pre- and post-hurricane data.

DISCUSSION

Our findings show no statistically significant differences in (1) the thickness of fine sediment layer, (2) the proportion of coarse sediment at the surface (percent fine sand in the 0-4 cm depth), or (3) the profile distribution of fine sand or silt and clay when comparing the pre- and post-storm reef cores. This could suggest that the oysters stabilized the shoreline sediment despite two strong storms.

Our small sample size, lack of post-storm control cores, and the potential to overestimate or underestimate the storms' impacts limit the strength of our conclusions. One core was collected from behind each of three reefs during each sample period (pre- and post-storms; pre-storm controls). One reef out of three showed variances; pre-storms' Reef 2 and post-storms' Reef 15 did not exhibit a surface fine sediment layer. Characteristic differences amongst the individual reefs, such as height and oyster density, may account for the variation in these sediment profiles.

Secondly, the strength of our conclusion is limited by the fact that we did not analyze post-storm control cores. After the storms, visual and texture-by-feel analysis[40] indicated that the control sites continued to have a predominance of fine sand, and very little silt and clay, throughout the depth profile. Thus, even if we had analyzed sediment samples from the control sites, it would have been impossible to determine if erosion had removed sediment from the profile, if similarly textured, fine-sand sediments were deposited, or if the storms had not affected the sediments there at all.

Therefore, although we have evidence of the storms impacting the wider area through increased turbidity and high rainfall, the storm surge and higher water levels may mean that there was little impact on the sediment at the elevation of our oyster, and instead the storms may have had a bigger impact on the sediments at higher elevations. On the other hand, we could have underestimated impacts of the storms, given that we collected samples several months after the second storm, which could have given the sites time to reequlibrate. For example, sediment could have been eroded during one storm, and then been deposited during another.

Typically, storms decrease the thickness of fine sediment on the surface because of sediment resuspension and transport.[4,30] It does not appear that this occurred at our sites, because the thickness of the fine sediment layer was not significantly different in the pre-storm and post-storm cores. In low-energy environments, constructed oyster reefs have been found to be a viable option for preventing coastal erosion.[22,29,39]A study in North Carolina, measuring sediment stabilization through accretion rate, determined sites with oyster shell had an average accretion rate of +2.9 cm versus a loss of -1.3 cm in sites without oyster shell.[24]

The oyster shell sites accreted a significantly larger amount of sediment compared to the areas without shell in each of their study sites.[24] A significant storm event occurred during their same study—one site without oyster shell experienced heavy damage leading to a loss of -3.2 cm, versus an accretion rate of +6.3 cm in the site with oyster shell during the storm. It was concluded that the oyster shell, coupled with vegetation, stabilized the sediment during the storm.[24] Furthermore, differences in accretion rates between natural and constructed reefs have been investigated. In a study from the Grand Bay National Estuarine Research Reserve in Mississippi, all reefs prevented shoreline retreat, but constructed reefs were equally or more effective at doing so than natural reefs, due to increased oyster coverage.[37]

Sedimentation is a significant governing force of an oyster reef's health, as too much or too little sediment can bury or provide minimal substrate for further growth, respectively.[31] Oysters themselves can alter the sediment surrounding them as they deposit organic matter and collect incoming sediment.[42] Allogenic sediment, supplied by processes of sediment resuspension and erosion, contribute to oyster reef sediment accretion; however, coastal storms can alter the quantity of allogenic sediment available.[25,36] Although we did not detect significant sediment changes behind the reefs, the oysters themselves were found to be partially buried in sediment before the storms.[35] In addition to storm surge during hurricane season, oyster reef sites along the Tolomato River are also exposed to frequent boat wakes and wind-driven waves. After the storms, the reefs were considerably damaged, and their dead shell was spread over a large area of shoreline (**Figure 6b**). We cannot attribute the deterioration of the reefs to the storms however, because the plastic mesh bags holding the oysters together began breaking apart before the storms.

Constructed oyster reefs in North Carolina significantly decreased shoreline retreat in low-energy areas but were relatively ineffective at preventing erosion in high-energy environments, particularly after two storms.[28] Shoreline retreat or lateral accretion was not measured in our study, but the stabilization of the sediment is evidence of prevented coastal erosion. The ineffectiveness in relatively high-energy environments found in other studies[28,29,39] and the consequent damage of our constructed oyster reefs, suggests the need for more than one coastal erosion prevention method.[23,38] Furthermore, effectiveness against storms and high wave energy is hindered if oysters cannot successfully accrete in pace with sea level rise, so shoreline protection from coastal erosion decreases as a result.[24,31] Due to the current physical state of the oyster reefs following the two storms, it is unlikely that the reefs will keep pace with sea-level rise, especially in the high energy environment.

.

Figure 6a. The newly constructed oyster reefs at the study site in 2012. The 28 reefs are made of net bags of recycled oyster shells.

Figure 6b. The deteriorating state of the study site after Matthew and Irma.

CONCLUSIONS

The particle size distribution of sediment behind the constructed oyster reefs seems to have remained relatively consistent despite the occurrence of two strong storms. Fine sand and silt and clay percentages behind the oyster reefs were relatively similar before and after the storms, there was no statistically significant evidence of erosion of fine sediment or deposition of coarser sediment. Living shorelines can protect shorelines from erosion by attenuating high wave energy and trapping sediment. Unfortunately, significant damage has occurred on the oyster reefs at the GTM Research Reserve (**Figure 6b**). Although the reefs were able to attenuate the waves from the storms, intervention is needed to ensure they will protect the shoreline in the future. Further research could help identify if there are subsequent long-term effects to the sediment stabilization surrounding living shorelines at the GTM Research Reserve, such as stabilized carbon, increased elevation, and increased biodiversity.

ACKNOWLEDGEMENTS
The authors thank Dr. Nikki Dix and the GTM Research Reserve, along with John Wooldridge the Flagler College Lab Manager, Colin Herbert, Chris Blanco, Dani Cayouette, Tim Mullen, Cody Burns, Aubrey Kemper, Courtney Cunningham, and the Flagler College Natural Sciences Department.

REFERENCES

1. Arkema, K. K. Guannel, G., Verutes, G., Wood, S. A., Guerry, A., Ruckelshaus, M., Kareiva, P., Lacayo M., and Silver, J. M. (2015) Coastal habitats shield people a property from sea-level rise, *Nature Climate Change. doi:10.1038/nclimate1944*
2. Beck, M. W., Brumbaugh, R. D., Airoldi, L., Carranza, A., Coen, L. D., Crawford, C., Defeo, O., Edgar, G. J., Hancock, B., Kay, M. C., Lenihan, H. S., Luckenbach, M. W., Toropova, C. L., Zhang, G., and Guo, X. (2011) Oyster Reefs at Risk and Recommendations for Conservation, Restoration, and Management, *Bioscience* 61(2), 107-116. *doi:10.1525/bio.2011.61.2.5P.*
3. Bradley, M., Morris, J. T. (1990) Physical characteristics of salt marsh sediments: ecological implications, *Mar. Ecol. Pro. Ser* 61, 245-252. *doi: 0.3354/meps061245*
4. Cahoon, D., Reed D. J., Day W. J., Steyer, G. D. Bouman, R., Lynch, J., McNally, D., Latif, N. (1995) The influence of Hurricane Andrew on Sediment Distribution in Louisiana Coastal Marshes, *Journal of Coastal Research* Special Issue (21), 280-294.
5. Cangialosi, J.P., Latto, A.S., and Berg, R. (2018) National Hurricane Center Tropical Cyclone Report: Hurricane Irma. National Oceanic and Atmospheric Administration*,* Retrieved from *https://www.nhc.noaa.gov/data/tcr/AL112017_Irma.pdf* (accessed Sep 2018)
6. Coen L.D., Brumbaugh R.D., Bushek D., Grizzle R., Luckenbach M.W., Posey M.H., Powers S.P., Tolley S.G. (2007) Ecosystem services related to oyster restoration, *Marine Ecology Progress* 41, 303–307. *doi:10.3354/meps341303*
7. Cowart, L., Walsh, J.P., and Corbett, D.R. (2010) Analyzing estuarine shoreline change: a case study of Cedar Island, North Carolina, *Journal of Coastal Research* 26(5), 817-830. *https://doi.org/10.2112/JCOASTRES-D-09-00117.1*
8. Currin, C., Davis, J., Baron, C.L., Malhotra, A., Fonseca, M. (2015) Shoreline Change in the New River Estuary, North Carolina: Rates and Consequences, *Journal of Coastal Research* 31(5), 1069-1077. *https://doi.org/10.2112/JCOASTRES-D-14-00127.1*
9. Dalsgaard, Jensen, and Sorensen (2007) Small sample sieving and calibration. In Principles, Methods, and Application of Particle Size Analysis. Cambridge University Press. *doi: 10.1017/CBO9780511626142*
10. Erftemeijer, P.L.A. and Koch, E.W. (2001) Chapter 18 Sediment geology methods for seagrass habitat. *In* Global seagrass research methods. Ed. F.T. Short an R. G. Coles. *Elsevier. doi: 10.1016/B978-044450891-1/50019-0*
11. Garvis, S. K., Sacks, P. E., and Walters L. J. (2015) Formation, Movement, and Restoration of Dead Intertidal Oyster Reefs in Canaveral National Seashore and Mosquito Lagoon, Florida, *Journal of Shellfish Research* 34(2), 251–258. *https://doi.org/10.2983/035.034.0206*
12. Gedan, K. B., Kirwan, M. L., Wolanski, E., Barbier, E. B., Silliman, B. R. (2010) The present and future role of coastal wetland vegetation in protecting shorelines: answering recent challenges to the paradigm, *Climatic Change* 106, 7–29. *https://doi.org/10.1007/s10584-010-0003-7*
13. Gittman, R. K., Fodrie, J. F., Popowich, A. M., Keller, D. A., Bruno, J. F., Currin, C. A., Peterson, C. H. and Piehler, M. F. (2015) Engineering away our natural defenses: an analysis of shoreline hardening in the US, *Frontiers in Ecology and the Environment* 13(6), 301-307. *https://doi.org/10.1890/150065*
14. Gittman, R. K., Peterson, C. H., Currin, C. A., Fodrie, J. F., Piehler, M. F., and Bruno, J. F. (2016) Living shorelines can enhance the nursery role of threatened estuarine habitats, *Ecological Applications* 26(1), 249-263. *doi: 10.1890/14-0716.1*
15. Grabowski, J. H. and Peterson, C. H. (2007) Restoring Oyster Reefs to Recover Ecosystem Services, *Theoretical Ecology Series* 4, 281-298. *doi: 10.1016/S1875-306X(07)80017-7*
16. Grizzle, Raymond E., Jamie R. Adams, and Linda Walters. (2002) Historical changes in intertidal oyster (Crassostrea virginica) reefs in a Florida lagoon potentially related to boating activities, *Journal of Shellfish Research* 21(2), 749-756.
17. Herbert, D., Astrom, E., Bersoza, A. C., Batzer, A., McGovern, P., Angelini, C., Dix, N., and Sheremet, A. (2018) Mitigating Erosional Effects Induced by Boat Wakes with Living Shorelines, *Sustainability* 10(2), 436. *https://doi.org/10.3390/su10020436*
18. Jackson, Jeremy B. C. Kirby, M. X., Berger, W. H., Bjorndal, K. A., Botsford, L. W., Bourque, B. J., Bradbury, R. H., Cooke, R., Erlandson, J., Estes, J. A., Hughes, T. P., Kidwell, S., Lange, C. B., Lenihan, H S., Pandolfi, J. M., Peterson, C. H., Steneck, R. S., Tegner, M. J., Warner, R. R. (2001) Historical Overfishing and the Recent Collapse of Coastal Ecosystems, *Science* 293(5530), 629-637. *doi: 10.1126/science.1059199*
19. Kish, Stephen K. and Joseph F. Donoghue (2013) Coastal Response to Storms and Sea-Level Rise: Santa Rosa Island, Northwest Florida, U.S.A., *Journal of Coastal Research* 63, 131-140. *https://doi.org/10.2112/SI63-012.1*
20. Kettler, T.A, Doran, J.W., and Gilbert, T.L. (2001) Simplified Method for Soil Particle-Size Determination to Accompany Soil-Quality Analyses. *Soil Science Society of America Journal* 65, 849-852. *doi: 10.2136/sssaj2001.653849x*

21. Knutson, P. L., Ford, J. C., Inskeep, M. R., and Oyler, J. (1981) National survey of planted saltmarshes (vegetative stabilization and wave stress), *Wetlands* 1, 129–157. *doi: 10.1007/BF03160460*

22. La Peyre, M. K., Serra, K., Joyner, T. A. and Humphries, A. (2015) Assessing shoreline exposure and oyster habitat suitability maximizes potential success for sustainable shoreline protection using restored oyster reefs, *PeerJ* 3:e, 13-17. *doi:10.7717/peerj.1317*

23. Manis, J. E., Garvis, S. K., Jachec, S. M., and Walters, L. J (2015) Wave attenuation experiments over living shorelines over time: a wave tank study to assess recreational boating pressures, *Journal of Coastal Conservation* 19, 1-11. *doi: 10.1007/s11852-014-0349-5*

24. Meyer, D.L., Townsend E.C., Thayer G.W. (1997) Stabilization and erosion control value of oyster cultch for intertidal marsh, *Restoration Ecology* 5, 93-99. *doi: 10.1046/j.1526-100X.1997.09710.x*

25. Morris, J. T., Sundareshwar, P. V., Nietch, C. T., Kjerve, B., and Cahoon, D.R. (2002) Responses of Coastal Wetlands to Rising Sea Level, *Ecology* 83(10), 2869-2877. *doi: 10.1890/0012-9658(2002)083[2869:ROCWTR]2.0.CO;2*

26. Morton, R. A. and Barras, J. A. (2011) Hurricane Impacts on Coastal Wetlands: A Half-Century Record of Storm-Generated Features from Southern Louisiana, *Journal of Coastal Research* 27(6A), 27-43. *doi: 10.2307/41315913*

27. NOAA National Estuarine Research Reserve System (NERRS). System-wide Monitoring Program. Data accessed from the NOAA NERRS Centralized Data Management Office website: http://www.nerrsdata.org/; *accessed* 31 July 2019.

28. Phillips, J. D. (1999) Event Timing and Sequence in Coastal Shoreline Erosion: Hurricanes Bertha and Fran and the Neuse Estuary, *Journal of Coastal Research* 15(3), 616-623.

29. Piazza, B. P., Banks, P. D., La Peyre, M. K. (2005) The Potential for Created Oyster Shell Reefs as a Sustainable Shoreline Protection Strategy in Louisiana, *Restoration Ecology* 13(3), 499-506. *https://doi.org/10.1111/j.1526-100X.2005.00062.x*

30. Reese, C. A., Strange, T. P., Lynch, W. D., Liu K. (2008) Geologic Evidence of Hurricane Katrina Recovered from the Pearl River Marsh, MS/LA, *Journal of Coastal Research* Vol. 24(6), 1601-1607. *https://doi.org/10.2112/07-0956.1*

31. Rodriguez, A. B., Fodrie, J., Ridge, J., Lindquist, N. L., Theuerkauf, E. J., Coleman S. E., Grabowski, J. H., Brodeur M. C., Gittman, R. K., Keller, D. A., and Kenworthy, M. D. (2014) Oyster reefs can outpace sea-level rise, *Nature Climate Change* 4, 493–497. *doi: 10.1038/NCLIMATE2216*

32. Schuerch, M., Dolch, T., Bisgwa, J. and Athanasios T. Vafeidis (2018) Changing Sediment Dynamics of a Mature Backbarrier Salt Marsh in Response to Sea-Level Rise and Storm Events, *Frontiers in Marine Science* 5(155), 1-14. *doi: 10.3389/fmars.2018.00155*

33. Scott, D. B., Frail-Gauthier, J., and Mudie, P. J. *Coastal Wetlands of the World.* New York: Cambridge University Press, 2014. Print.

34. Scyphers S. B., Powers S. P., Heck K. L. Jr, Byron D. (2011) Oyster Reefs as Natural Breakwaters Mitigate Shoreline Loss and Facilitate Fisheries. *PLoS ONE* 6(8): e22396. *doi: 10.1371/journal.pone.0022396*

35. Southwell, M. W., Veenstra, J. J., Adams, C. D., Scarlett, E. V. and Payne, K. B. (2017) Changes in Sediment Characteristics upon Oyster Reef Restoration, NE Florida, USA, *Journal of Coastal Zone Management* 20(1), 1-7. *doi: 10.4172/2473-3350.1000442*

36. Stewart, S. R. (2016) National Hurricane Center Tropical Cyclone Report: Hurricane Matthew. National Oceanic and Atmospheric Administration, Retrieved from *https://www.nhc.noaa.gov/data/tcr/AL142016_Matthew.pdf.* (accessed Sep 2018)

37. Stricklin, A. G., M. S. Peterson, J. D. Lopez, C. A. May, C. F. Mohrman and Woodrey, M. S. (2010) Do Small, Patchy, Constructed Intertidal Oyster Reefs Reduce Salt Marsh Erosion As Well As Natural Reefs?, *Gulf and Caribbean Research* 22(1), 21-27. *doi: 10.18785/gcr.2201.03*

38. Sutton-Grier, A. E., Wowk, K., and Bamford, H. (2015) Future of our coasts: The potential for natural and hybrid infrastructures to enhance the resilience of our coastal communities, economies, and ecosystems, *Environmental Science and Policy* 51(1), 137-148. *https://doi.org/10.1016/j.envsci.2015.04.006*

39. Swann, L. (2008) The Use of Living Shorelines to Mitigate the Effects of Storm Events on Dauphin Island, Alabama, U.S.A., *American Fisheries Society* 64(1), 1-12.

40. Vos, C., Don, A., Prietz, R., Heidkamp, A., Freibauer, A. (2016) Field-Based Soil Texture Estimates Could Replace Laboratory Analysis. *Geoderma* 267, 215-219.

41. Warren, S. R. and Niering, W. A. (1993) Vegetation Change on a Northeast Tidal Marsh: Interaction of Sea-Level Rise and Marsh Accretion, *Ecology* 74(1), 96-103. *doi: 10.2307/1939504*

42. Ysebaert T., Walles B., Haner J., Hancock B. (2019) Habitat Modification and Coastal Protection by Ecosystem-Engineering Reef-Building Bivalves. In: Smaal A., Ferreira J., Grant J., Petersen J., Strand Ø. (eds) Goods and Services of Marine Bivalves. Springer, Cham. *https://doi.org/10.1007/978-3-319-96776-9_13*

ABOUT STUDENT AUTHOR

Taryn Chaya graduated from Flagler College in May 2019 with a degree in Coastal Environmental Science. Taryn is pursuing a master's degree in Soil and Water Science at the University of Florida. She hopes to use this degree to remediate or prevent the negative effects climate change and anthropogenic factors place on wetland ecosystems.

PRESS SUMMARY

Shoreline loss due to storms, sea level rise, and human activities is becoming increasingly prevalent. Coastal erosion prevention practices are gaining traction as shoreline loss continues to threaten economic and natural systems in coastal cities—living shorelines have been considered a long-term solution to coastal erosion in comparison to hard structures like seawalls. In St. Augustine, FL, two major storms, hurricane Matthew and tropical storm Irma, caused significant infrastructural and natural damage. Living shorelines, such as the construction of oyster reefs or replenishment of vegetation can buffer waves and trap sediment during storms. Oyster reefs were constructed at the Guana Tolomato Matanzas National Estuarine Research Reserve (GTM Research Reserve) in Northeast Florida to stabilize the shoreline sediment and prevent the erosion of an archeological site. Sediment cores were collected behind constructed oyster reefs at the GTM Research Reserve before and after the two storms, and although they were high-energy storms, they did not significantly change the sediment behind the constructed oyster reefs. This suggests that the constructed oyster reefs adequately protected the sediment from erosion during the storm events.

Print Edition ISSN 1536-4585
Online Edition ISSN 2375-8732