PERSISTENCE ENTROPY RELATES REST-ACTIVITY DURING PREGNANCY TO UNFAVORABLE HEALTH OUTCOMES IN MATERNAL AND NEWBORN HEALTH

A Thesis

Presented to the

Faculty of

San Diego State University

In Partial Fulfillment

of the Requirements for the Degree

Master of Science in Applied Mathematics

with a Concentration in

Dynamical Systems

---

by

Sashiel Vagus

Spring 2024

SAN DIEGO STATE UNIVERSITY

The Undersigned Faculty Committee Approves the

Thesis of Sashiel Vagus:

Persistence Entropy

Related Rest-Activity During Pregnancy

to Unfavorable Health Outcomes

in Maternal and Newborn Health

Uduak George, Chair
Department of Mathematics and Statistics

Jérôme Gilles
Department of Mathematics and Statistics

Emily Schmied
Department of Public Health

Approval Date

DEDICATION

Dedicated to my parents for all their love and encouragement.

What is mathematics? It is only a systematic effort of solving puzzles posed by nature.

– Shakuntala Devi

ABSTRACT OF THE THESIS

Persistence Entropy
Related Rest-Activity During Pregnancy
to Unfavorable Health Outcomes
in Maternal and Newborn Health
by
Sashiel Vagus
Master of Science in Applied Mathematics with a Concentration in Dynamical Systems
San Diego State University, 2024

Recent studies have suggested the need to understand how rest-activity during pregnancy affect maternal and newborn health. Actigraphy objectively captures rest-activity, but data gathered from actigraphy are very large with irregular visual patterns. To fully understand how the rest-activity relates to maternal and newborn health, novel approaches that allow optimal extraction of patterns and signals from rest-activity data are needed. This study aimed to fill this gap by testing the application of persistence entropy to identify patterns within rest-activity and investigate the correlation between these patterns and maternal and newborn health outcomes. Persistence entropy is a mathematical measure that is able to describe the degree of disorder in temporal data and has potential to identify patterns in rest-activity that may be associated with poor pregnancy outcomes. Actigraphic data was collected from a cohort of pregnant women, with observations made during gestation week 22 (G22; n=41) and 32 (G32; n=44). The daily and weekly actigraphy data for each woman were analyzed using persistent homology and entropy. The vietoris- rips complex filtration was applied to generate persistence diagrams and the entropy of the persistence diagrams were quantified and related to the maternal and newborn health. Women with high entropy values (i.e. above 8.22) during G22 were more likely to have complications during pregnancy. In addition, a significant difference in the variability in daily entropy values during G22 was found between women with complicated versus uncomplicated pregnancy (one-tailed Mann Whitney U test: P-value = 0.027). Findings from this study support the application of persistence homology and entropy to extract patterns of rest-activity data and entropy values of activity rhythms can be related to health outcome. Further studies on a larger sample size is necessary to explore its use as a predictive model for maternal and newborn health.

TABLE OF CONTENTS

PAGE

LIST OF TABLES

LIST OF FIGURES

PAGE

ACKNOWLEDGMENTS

I would like to thank Dr. George for mentoring me on this project. I would also like to thank her for her guidance throughout my undergraduate and graduate years at SDSU. I would have never gotten so far if it wasn't for her patience and support. She made school so much more interesting and enjoyable, and I will miss working and learning alongside her and my colleagues. Thank you for making my experience at SDSU so much better. I will cherish what I've learned from SDSU and you always.

Additionally, I would like to thank Dr. Casey at Purdue University for supplying the data and for all her assistance on this project. It was always a pleasure working with her. I would also like to thank and acknowledge Dr. Gilles and Dr. Schmied for agreeing to be on my thesis committee. Thank you so much for your time and guidance.

Lastly, I would like to thank my family for their never-failing support. Thank you so much for your dedication to my well being and pushing me to strive for higher education. I never would have accomplished what I have so far if it weren't for you all keeping me happy and curious about the world.

CHAPTER 1

Overview and Background

The Centers for Disease Control and Prevention (CDC) found that, in 2020, pregnant women in the United States had the highest mortality risk for live births compared to other industrialized countries with 23.8 deaths per 100,000 live births [1]. France came in second place with 8.7 deaths per 100,000 live births, emphasizing the stark contrast between its rate and that of the United States, which experienced a disproportionately higher number of deaths. The National Sleep Foundation recently released a consensus statement indicating the importance of day-to-day consistency in timing of sleep and wake for health, safety, and performance [2]. Irregular rest-activity affect physiological and neurobehavioral health [3], and is linked to circadian clock disruption. Disruption of the circadian rhythms, including sleep-wake cycles [4], negatively impacts health. Circadian, sleep and reproductive systems are interconnected to each other, meaning that perturbing one system may lead to a negative perturbation of the other systems. Sleep is altered during pregnancy [5], but the importance of regular rest-activity patterns during gestation on maternal-fetal health is understudied.

There is growing evidence that links circadian disruption and sleep disruption patterns during gestation to increased risk of negative maternal and fetal health outcomes. For example, women who work night shifts have an increased risk of delivering low birth weight infants [6]. A cohort study [7], found sleep disordered breathing increased the risk of preeclampsia, hypertensive disorders and gestational diabetes mellitus. Okum et al [8] found a significant association between poor self-reported sleep quality and depression and anxiety symptoms in women at six months postpartum. Therefore, these studies show that irregular sleeping patterns affect both maternal and newborn health, however, many of these results are from self-reports of sleep quality and variability and objective measures and analysis are needed to understand risks and relationships to health.

African American women have about 2.9 times the maternal mortality risk as White women in the United States [1]. Mezick et al [9] found that Black/African American women had shorter and low quality sleep, which puts them at risk for health complications. Additionally, Feinstein et al [10] found that Black pregnant women also had higher short sleep occurrences compared to pregnant White women. Furthermore, Black pregnant women used sleep medication statistically more than White or

Hispanic/Latina women. Compared to most developed nations, no federal law provides a right to paid family or medical leave in the United States [11]. This makes women to increase work hours in order to make ends meet, instead of making time to recover and rest. It is therefore important to study how sleep and rest-activity impact women and fetal health in order to gain knowledge and press for policy changes to advance maternal health.

1.1 Monitoring Sleep and Rest-Activity Levels with Actigraphy

One common tool used to measure rest-activity level is actigraphy. Actigraphy is a low cost method for measuring movement patterns in humans. Its most common use is to study circadian rhythms (i.e. sleep cycle). It is a device worn on the wrist, and contains an acclerometer, battery and memory storage [12]. Wrist actigraphy is a validated measure for activity and an objective measure of sleep [13], with overall actigraphic sensitivity and high accuracy relative to polysomnography, but specificity is low [14]. Researchers often rely on software accompanying actigraphy devices to analyze actigraphic data, but these software tools have limitations in abstracting day-to-day variations. Actigraphic data are often difficult to interpret because they are time series data with irregular visual patterns and multiple days of activity generates large amounts of data, big data, that are difficult to abstract information from. Therefore, novel approaches are needed to analyze rest-activity data gathered from actigraphic devices.

1.2 Existing Methods for the Interpretation of Actigraph Data

Currently, there are a few popular tools that are often used to interpret actigraph data. For example, in Zemet et al [15] they used an activity tracker to measure women's physical activity after an embryo transfer up until they did a pregnancy test. The purpose of this study was to find if there was an association between the level of physical activity and pregnancy rate after an embryo transfer. The manner in which they analyzed the data was to measure the amount of steps taken. Their results found that the women who participated walked significantly less on the day of the transfer compared to the 2 days after. The biggest take away was that there was no significant difference in median steps per day between the pregnant and non-pregnant women after the embryo transfer. Similar to most studies, this study relies on rudimentary methods (i.e. counting the number of steps taken by the women)

to analyze physical activity data. The number of steps taken may not be sufficient to capture important features in their rest and activity patterns.

In Korte et al [16], they compared ultradian and circadian activity-rest rhythms of preterm neonates to full-term neonates using actigraphy. In this study, the neonates used a wristwatch in order to measure activity and rest. To interpret periodic characteristics they used fast Fourier transform (FFT). From this they were able to find that the majority of preterm neonates showed many ultradian frequencies [16]. Similarily, in Korte et al [17] they utilized FFT to analyze the actigraph data of 20 neonates born vaginally, 18 neonates born by Cesarean section (C-section) and 19 born by required C-section after the labor onset. Results showed that the neonates that were born vaginally had a distinguishable circadian frequency in their spectra while the neonates that were born by C-section displayed significantly less distinguishable circadian frequencies.

Slyepchenko et al [18] used cosinor analysis to investigated the association between objective parameters of sleep and biological rhythms with peripartum mood and anxiety. Cosinor analysis is a method that is often used to analyze rhythms in temporal data. Using cosinor analysis, Slyepchenko et al [18] extracted parameters such as the midline estimating statistic of the rhythm (MESOR), amplitude, and circadian quotient (CQ). Using these parameters, they were able to accentuate the importance of biological rhythms and activity. Root mean square of successive differences (RMSSD), a method that is often used to study heart rate variability [19], was used by Casey et al [20] to study sleep variability during gestation. Using RMSSD, Casey et al [20] measured variability in the sleep at gestational week 22 (G22) and 32 (G32). Their study found that high sleep variability in the third trimester of pregnancy is associated with delayed stage II lactogenesis.

1.3 Advantages and Disadvantages of Existing Actigraph Interpretation Methods

The use of FFT does have its advantages. For example, it can be used to extract periodic information by changing time-domain signals into frequency-domain [21]. Additionally, FFT is computationally inexpensive and simple [22]. However, there are some limitations, which include the use of biased interpolations on the data set in order to obtain refined estimates of frequencies and amplitudes with FFTs. Also, the peaks with low amplitudes can be distorted or hidden. According to Li et al [22], if the data has irregular rhythms, then frequency analysis from FFT may not be accurate. Rest and activity data often have irregular rhythms, due to many factors like irregularities in

human behavior, human error or sensor sensibility issues. And this would affect the accuracy of the findings from FFT. In addition, the oscillation and the frequency resolution of the FFT analysis is influenced by the data segment's length and therefore it would need artificial interpolation so that it can meet the need of equal distance, and this artificial interpolation would present biases. Using artificial interpolation would call for making assumptions about the data which in turn would introduce bias. This leads us to question if using FFTs on actigraph data is the best option especially since actigraph data is commonly large and noisy. These constraints are further explained in Henry's [23] where he states that its limitations are that 1) interpolation is often used to obtain improved estimates of peak frequencies and amplitudes, (2) lower amplitude peaks can be distorted or hidden, (3) it is a discrete time approximation of continuous time mathematics. All in all, while FFT is a common tool for actigraph analysis its limitations may justify looking into new methods of analysis.

As for the cosinor analysis, while it is commonly used to illustrate patterns of cyclic activity and detect rhythmic patterns in data [24], a disadvantage it posses is that it can not measure rhythm fragmentation. Fossoin et al [25] describes rhythm fragmentation as the quality of the regular circadian rhythm lowering because of naps taken during the daytime and/or any activity episodes during the night. These random fluctuations are a common characteristic of actigraph data and of human physiological parameters. Additionally, if the data has properties that changes over time, such as having dominant trends and/or amplitudes, frequencies or phases that vary with time then fitting the data to the models is much more difficult [26]. We can also infer that randomness and noise can affect the way we interpret cosinor analysis of actigraph data.

Though RMSSD can be used to compute variance in time series data it has some limitations. In Bourdillon et al [27], they found that RMSSD and standard deviation of the normal to normal inter beat intervals (SDNN) were more sensitive to single artifacts introduced in the data. In other words, when using RMSSD with just one artifact (or irregularity), the RMSSD may produce erroneous interpretation of data. It is often recommended that both time and frequency domains be used to minimize errors in RMSSD. As a consequence, when using RMSSD to analyze purely actigraph data as was done in Casey et al [20] RMSSD may incorrectly interpret the data if irregularities exist in the data.

Researchers often rely on reducing time series activity values to summary statistics because it offers a fast and convenient method to extract information on the data [28]. However, summary statistics may not be able to capture some key features in the activity data. For example, Wang et al [28] showed that if a study were to have two

different groups of patients in which one would have high activity in the morning and low activity levels at night while the other group has a opposite schedule with the same magnitude of activity, then their average activity would be similar and these significant patterns in activity would be lost. Furthermore, if the data is not statistically normally distributed then the summary statistic for mean and standard deviation would not accurately represent the data [29]. As a result, relying only on using summary statistics may be an issue when it comes to correctly interpreting the data.

1.4 Goals for This Thesis

Due to the limitations in the existing methods for interpreting actigraphic data, new methods are needed that can better interpret actigraph data. This thesis introduces topological data analysis as a novel approach for interpreting actigraph data In particular, this thesis aims to test whether the application of persistent entropy, a computational method that is able to capture the topology of temporal data, can provide measures that may allow us to associate rest-activity patterns during pregnancy to maternal and newborn health outcomes. In this thesis, we hypothesize that persistent entropy would be able to succinctly condense the irregular rest-activity patterns during the time of gestation to quantifiable indicators, allowing us to better relate the irregular patterns to adverse health outcomes in maternal and newborn. To test the functionality of the proposed method, we analyzed rest-activity data collected from a previous study involving a cohort of pregnant women, where 24 hour rest-activity was measured during G22 (n=41) and G32 (n=44). These women were selected from a population within an economically disadvantaged group and some of these women worked as shift workers with irregular sleep schedules. Using persistent entropy, we studied how the rest-activity rhythms relate to the health of the mother and newborn based on markers such as BMI, blood glucose levels, maternal disorders and newborn conditions. Findings from this study support the idea that persistent homology and entropy may be used to extract patterns/signals from rest-activity data and has potential to be applied to relate activity rhythms during pregnancy to certain unfavorable health outcomes in maternal and newborn health.

In Chapter 2, we introduce the methods for topological data analysis and describe how to compute persistent entropy. In Chapter 3, we will describe the study design and discuss how we used persistent entropy to describe the topology of rest-activity data for pregnant women. In Chapter 4, we will present the findings from our study. In Chapters 5, we will discuss the importance of the findings from our study. We will conclude the thesis in Chapter 6.

CHAPTER 2

Introduction to Topological Data Analysis

Topology is a discipline of mathematics that studies the properties of geometric bodies that persist through continuous modifications such as stretching, twisting, flattening and other equivalent manipulations. These transformations, however, exclude any tearing, creating holes, or gluing edges together. Topological equivalence is when two topological spaces share a homeomorphism (Definition 2.1). An example of two shapes having topological equivalence is a circle and an ellipse because one could stretch the circle into an ellipse and vice versa. Topology can also be defined in terms of set operations (Definition 2.2.).

Definition 2.1. *If x and y are topologically equivalent, then there is a function h: $x \to y$ such that h is continuous, h is onto, h is one-to-one, and h^{-1}, is continuous. Then h is a homeomorphism [30].*

Definition 2.2. *A topology on a nonempty set X is a collection of subsets of X, called open sets, such that:*

a. *The empty set $\emptyset$ and the set X are open;*
b. *The union of an arbitrary collection of open sets is open;*
c. *The intersection of a finite number of open sets is open.*

A more formal definition of Topology is a collection T of subsets of X is a topology on X if:

a. $\emptyset, X \in T$;
b. *If* $G_\alpha \in T$ *for* $\alpha \in A$, *then* $\bigcup_{\alpha \in A} G_\alpha \in T$;
c. *If* $G_i \in T$ *for* $i = 1,2...,n$, *then* $\bigcap_{i=1}^{n} G_i \in T$.

Where the pair (X, T) is a topological space [30].

For simplicity we refer to these fundamentals as general topology, also known as point-set topology, which consists of topics such as continuity, dimension, compactness, and connectedness [30]. This lays the groundwork for the other branches of topology. These branches include combinatorial topology, algebraic topology and differential topology.

2.1 Simplicial Complexes

In order to portray a topological space one must break it down into low-dimensional geometrical pieces. This is where simplicial complexes come into play. A simplicial complex is a nested family of simplexes which take the form of a point in the 0-dimension, a line also known as an edge in the 1-dimension, a triangle face in the 2nd-dimension and then a tetrahedron in the 3rd-dimension and so on as shown in Figure 2.1.

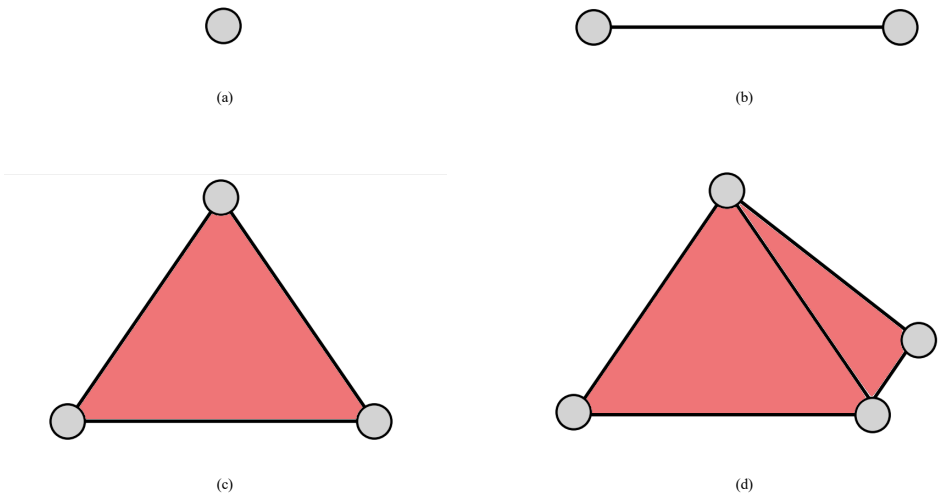


Figure 2.1. Low-dimensional geometrical pieces that simplexes may present as: (a) a point in the 0-dimension, (b) a line also known as an edge in the 1-dimension, (c) a triangle face in the 2nd-dimension and (d) then a tetrahedron in the 3rd-dimension.

In essence, simplicial complexes are used as a way to represent topological spaces. Specifically, a simplicial complex K is a geometrical realization of an abstract simplicial complex K. An abstract simplicial complex is defined formally as:

Definition 2.3. *An abstract simplicial complex K is given by:*

a. *a set V of 0-simplices;*

b. *for each $k \geq 1$ a set of k-simplices $\{\sigma = \{v_0, v_1, ..., v_k\}$, where $v_i \in V\}$;*

c. *each k-simplex has $k + 1$ faces obtained removing one of the vertices;*

d. *if σ belongs to K, then all faces of σ must belong to K.*

These simplicial complexes are an important aspect of Topology because many of its respective areas depend on using simplexes. Such as persistent homology, by

which simplicial complexes are created by applying filtrations to data which ultimately aids in finding persistent features of data.

2.2 Persistent Homology

This brings us to the applications of Topology. As science advances, the amount of data collected has grown both in size and complexity. Mathematicians and computer scientists have produced many new tools for analyzing complex data. One of these tools is topological data analysis (TDA). TDA has been used for the analysis of data from many different areas including the identification of genetic alterations in cancer [31], the biology of cells and organisms [32], and even the spread of contagious diseases [33]. Topology is commonly used as a way to study the shape of data, specifically data points or cluster points. TDA combines data analysis, computer science and topology. This type of analysis allows the data to be studied through persistent homology. Persistent homology is a mathematical technique used in the field of algebraic topology, in which qualitative features are taken into consideration for their persistence in different scales. This type of homology involves a filtration on a data set and as the filtration parameter changes, the user can compute the homology at the different values of that parameter. This is done by creating a filtration of a the simplicial complex which is the nested sequences of increasing subsets [34]. Its formal definition is:

Definition 2.4. *A filtered simplicial complex K is a collection of subcomplexes* $\{K(t) : t \in \mathbb{R}\}$ *of K such that* $K(t) \subset K(s)$ *for* $t < s$ *and there exists* $t_{max} \in \mathbb{R}$ *as such that* $K_{tmax} = K$*. The filtration time (or filter value) of a simplex* $\sigma \in K$ *is the smallest t such that* $\sigma \in K(t)$*.*

Depending on the data, this method can help the user see if the data is noisy and depict which features of the data are important. This data analysis is commonly used as a way to compare data sets.

2.3 Vietoris-Rips complex

A commonly used filtration technique is the Vietoris-Rips filtration also known as the Rips filtration. The Rips filtration is the nested collection of Vietoris-Rips complexes on a metric space which is created by taking the sequence of Vietoris–Rips complexes over an increasing scale parameter [35].

Definition 2.5. *Let* $X \subset \mathbb{R}^D$ *be a point cloud and* $\epsilon \geq 0$*. The Vietoris-Rips complex at scale* ϵ *is defined as* $VR_\epsilon(X) = \{\sigma \subseteq X \mid d(x, x') \leq 2\epsilon, \forall\ x, x' \in \sigma\}$*. In other words, for a given scale* $\epsilon \geq 0$*, if* $d(x, x') \leq 2\epsilon$ *for* $x, x' \in X$*, one adds the p-simplex* $\sigma = [x_0 x_1 ... x_k]$ *to the complex at the largest scale of any of its edges [36].*

In other words, each data point of a point cloud is surrounded by a ball or loop of radius ϵ. The radius of the circle/loop continues to grow larger, and if the pairwise distance between any two points is less than ϵ, then one must build an n-simplex. This results in a simplicial complex X_ϵ. As shown in Figure 2.2. Fundamentally, one could

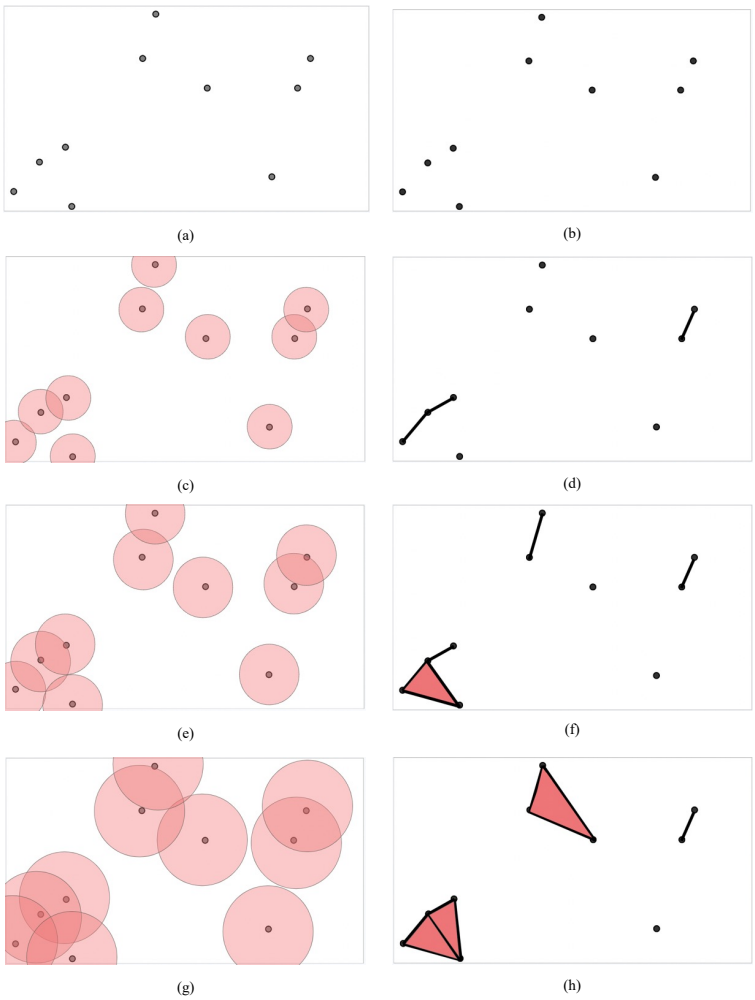


Figure 2.2. Vietoris-Rips complex. In (a), (c), (e) and (g), the red circles show the increasing radius of the loops and the corresponding simplicial complex is created on the right side in the images in (b), (d), (f) and (h). As the radius increases, the simplicial complex begins to develop from points, then to line segments which are also known as edges, and then to triangles.

think of this as keeping record of when these loops touch. Increasing the radius of the circles/loops results in "birth" and "deaths" of loops. To be clear, a "birth" is classified as the moment a loop is made. While a "death" is the moment the loop intersects or touches another loop. After these are computed, the visualization of the persistent homology begins. Essentially, the deaths are recorded on a persistence diagram Figure

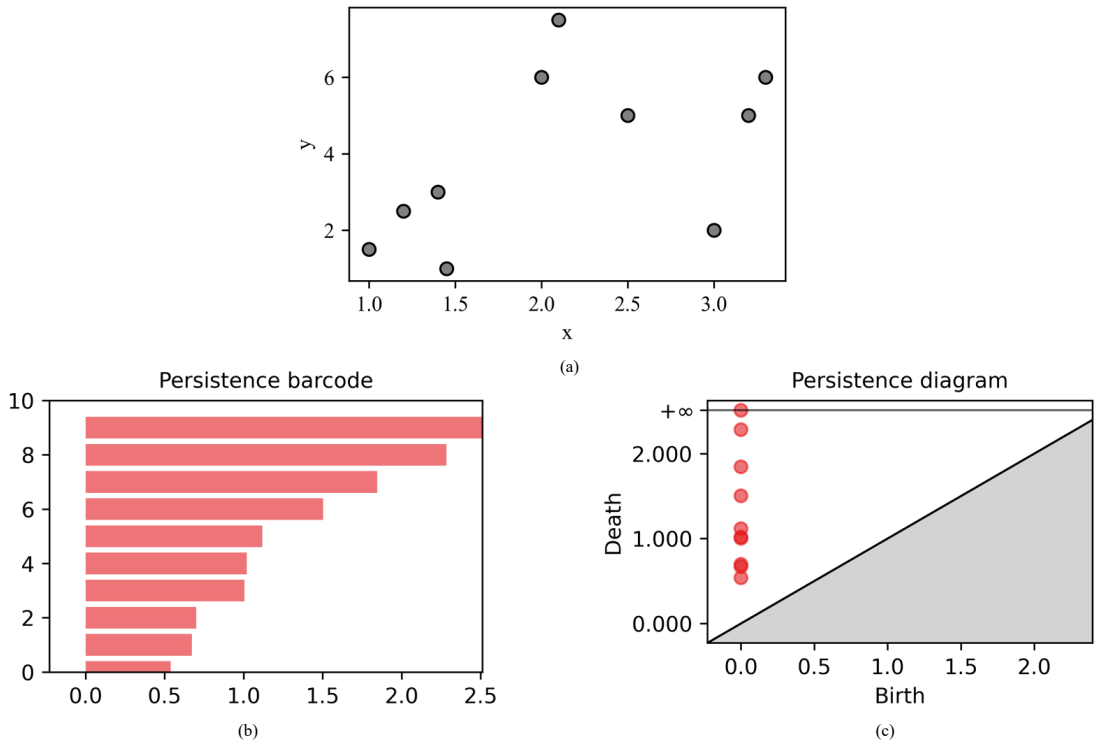


Figure 2.3. (A) Example of the random scatter plot that was displayed earlier in Figure 2.1, and (b) depicts its persistence barcode where the quantification is shown through horizontal bars. It's x-axis corresponds to the length of the radius at the point a death occurred. The y-axis is the dimension that we are in which is the 0-dimension because we are only looking at connected components. (C) Is the data's persistence diagram. This diagram only considers dimension 0 as well. Since this is a persistence diagram, the quantification of the data is portrayed through dots. Each dot represents a death, and it's x-axis represents the moment the loop was born which was at zero. It's y-axis represents the length of radius at which the death occurred.

2.3. A persistence diagram is inherently a recording of when the deaths took place. The points closer to the diagonal line are considered as noise, and those further away from the diagonal indicate a persistent feature.

2.3.1 Visualizing Persistent Homology

There are two main ways of visualizing persistent homology, these are persistence diagrams and persistence barcodes. Persistence diagrams, as explained earlier, are the tracking of "births" and "deaths" of the circles/loops. They are recorded in the form of dots and they are plotted according to their dimension and time of death. While persistence diagrams and persistence barcodes are in essence the same thing there are some distinct characteristics. A persistence barcode is defined as a

graphical representation of the deaths of the loops as bars whose lengths represent the loops life instead of portraying it as dots. The bars represent the length of the lifespan such as in Figure 2.3 where we are in 0-order homology. The y-axis indicates the connected components and the x-axis is the length of the radius. The persistence diagram denotes the x-axis as the moment the loop was born and the y-axis is the length of the loop's life. As one can depict, these two diagrams tell the same story, but they are visually different.

2.3.2 Advantages of the Vietoris-Rips Complex

A similar filtration to Vietoris-Rips complex is the Čech complex. The formal definition of the Čech complex is:

Definition 2.6. *Let V be a (finite) point cloud data in $\mathbb{R}^d$, i.e., a (finite) set of points in $\mathbb{R}^d$. The Čech complex of V and r denoted by $\check{C}_V(r)$ is the simplicial complex whose simplices are formed as follows. For each subset σ of points in V, form a closed ball $B_x(r)$ of radius r around each point x in σ, and include σ as a simplex of $\check{C}_V(r)$ if there is a common point contained in the intersection of the balls $B_x(r)$ for all x in σ, that is: $\check{C}_V(r) = \{\sigma \subseteq V \mid \bigcap_{x\in\sigma} B_x(r) = \emptyset\}$. This structure satisfies the definition of simplicial complex [37].*

Despite its similarity to the Rips filtration, the Čech complex is more computationally expensive. Both filtrations are similar, however, there is a small difference that makes the Čech complex harder to compute. The difference is that the Rips filtration creates a d-simplex as soon as when the loops/circles have pairwise intersections (when they touch), versus the Čech filtration creates a d-simplex when there is a common point of intersection for all the loops/circles (when all loops intersect each other). In other words, one must keep track of numerous intersections for this filtration. For this reason, the Vietoris-Rips complex is more popular when it comes down to computing its simplicial complex. As a result, the Rips complex has more readily accessible software for people to use on data clouds. The difference between filtration's is illustrated in Figure 2.4.

Another common filtration is the Alpha filtration defined as:

Definition 2.7. *Let $X \subset \mathbb{R}^D$ be a point cloud. Let $\varepsilon \geq 0$ and let $S_x(\varepsilon) := V_x \cap B_x(\varepsilon)$, where $B_x(\varepsilon)$ is the D-dimensional ball of radius ε centered on $x \in X$. The Alpha complex at scale $\varepsilon \geq 0$ is $Alpha_\varepsilon(X) = \{\sigma \subseteq X \mid \bigcap_{x\in\sigma} S_x(\varepsilon) \neq \varnothing\}$ [36].*

The alpha filtration can also be described as the subcomplex of the Delaunay complex [38]. The Delaunay complex is defined as:

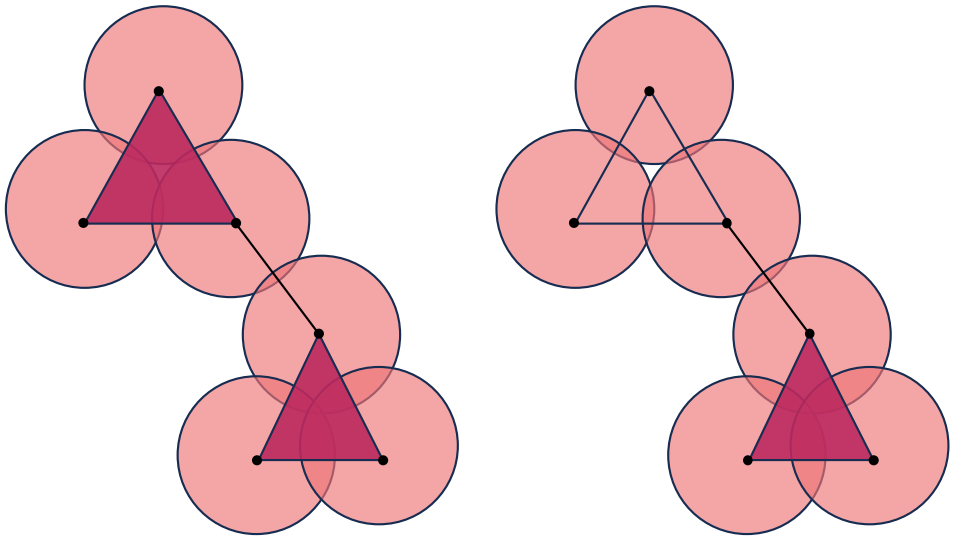


Figure 2.4. (A) is an example of the implementation of the Vietoris-Rips complex. In the instances where three circles touched, a triangle simplex is formed. In this case, two triangle simplexes were formed. However, (b) is an example of the implementation of the Čech complex. If we were to use the Čech complex then we would have to wait until all the circles overlap to form the triangle simplex. The gap between the three loops at the top halts the Čech complex from forming a simplex in (b) so it therefore has only one triangle simplex.

Definition 2.8. *The Delaunay complex of a finite set $S \subseteq \mathbb{R}^d$ is isomorphic to the nerve of the Voronoi diagram, Delaunay* $= \{\sigma \subseteq S \mid \bigcap_{u\in\sigma} V_u \neq \emptyset\}$. *[39]*

Here isomorphic is defined as when mathematical objects have the same structural properties [40]. A nerve is defined as the simplicial complex formed from a family of objects by taking sets that have nonempty intersections [41]. Lastly, the Voronoi diagram refers to the Dirichlet tessellation [42]. The formal definitions are provided below:

Definition 2.9. *Two sets are isomorphic to each other if they share an isomorphism. A one-to-one correspondence (mapping) between two sets that preserves binary relationships between elements of the sets. [43]*

Definition 2.10. *Let $U = \{U_i\}_{i\in I}$ be a non-empty collection of sets. The nerve of U is the simplicial complex with set of vertices given by I and k-simplices given by $\{i_0, ..., i_k\}$ if and only if $U_{i0} \bigcap ... \bigcap U_{ik} \neq \emptyset$. [38]*

Definition 2.11. *The Voronoi diagram is created by scattering points at random on a Euclidean plane. The plane is then divided up into tessellating polygons, known as*

cells, one around each point, consisting of the region of the plane nearer to that point than any other. [44]

Nevertheless, users frequently choose the Vietoris-Rips filtration because of its simplicity. Similarily to the Čech complex, the Alpha filtration must satisfy a specific condition in order to form a complex. On the otherhand, the Vietoris-Rips complex is all based on whether the loops intersect each other. Furthermore, according to Otter et al [38] they found through testing data that the Ripser package, which was used for this study, is the best-performing library available at this moment for computing the persistent homology of the Vietoris-Rips complex.

2.4 The Soundness of the Filtrations

When building simplicial complexes on a point cloud sampled from a space, the homology of the simplicial complex has to approximate the homology of the space. Both the Čech complex and the Alpha complex are theoretically guaranteed by the Nerve theorem [38]. The Vietoris-Rips complex is theoretically guaranteed because it approximates the Čech complex [38].

The Nerve Theorem implies that the nerve of the cover and the space X have the same homology [38]. The theorem is as follows:

Theorem 2.12. *Let F be a finite collection of closed, convex sets in Euclidean space. Then the nerve of F and the union of the sets in F have the same homotopy type. [39]*

The Vietoris-Rips complex approximates the Čech complex because the edges/lines formed in the Vietoris-Rips complex are the same as in the Čech complex. Furthermore, Čech(r) $\subseteq$ Vietoris-Rips(r) [39]. As a result, the effectiveness of the Vietoris-Rips complex makes it highly attractive. The persistence diagrams formed from the filtration's are also theoretically stable. According to Cohen et al, [45] they referred to the Bottleneck Stability Theorem for Persistence Diagrams.

Theorem 2.13. *Let X be a triangulable space with continuous tame functions f,g :* $\mathbb{X} \implies \mathbb{R}$. *Then the persistence diagrams satisfy* $d_B(D(f), D(g)) \leq \| f - g \|_\infty$. *Where* $D(f)$ *and* $D(g)$ *correspond to their persistence diagrams and* $d_B(D(f), D(g))$ *corresponds to the bottleneck distance between the two diagrams. [45]*

Essentially, what this theorem infers is that the persistence diagrams are stable even when small perturbations are introduced. It is important to have these stability properties in physical problems because if small perturbations are creating big discrepancies then the results will not be accurate. In other words, stability is needed for reliable predictions and small changes in data should not be resulting in very different results.

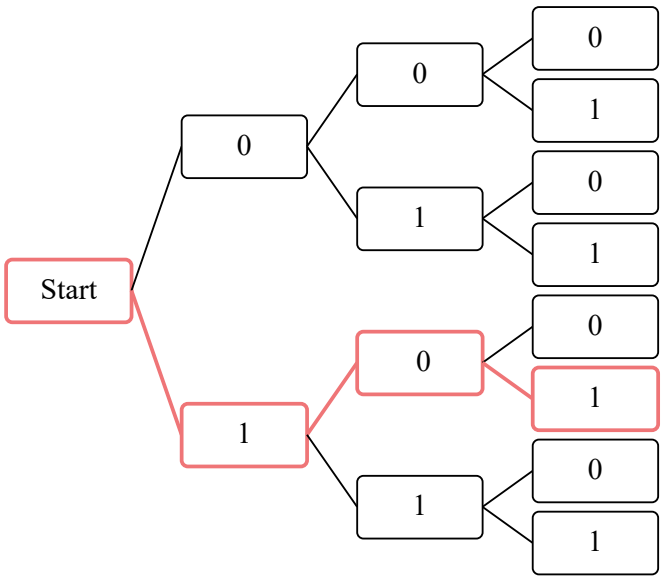


Figure 2.5. The "traveler" started at the leftmost pink outlined tile. The pink lines indicate the roads he chose. One could describe his path as taking a right, left then right again. Or in binary as 101. Shows that the amount of possibilities to choose from are exponential.

In addition, for this project it was decided to stay in dimension 0 since the diagrams were able to depict the important features of the data and there was no need to increase its dimension. The 0-dimension keeps track of the number of path components (connected components), the 1-dimension keeps track of the number of holes and the 2-dimension keeps track of the number of voids [38].

2.5 Information Theory

Information Theory is in the field of applied mathematics and is a combination of statistics, computer science and engineering. It is the mathematical study of information and its quantification. Mathematician Claude Shannon published a paper in 1948 called "A Mathematical Theory of Communication" [46] where he revolutionized the way we understood information. In specific, he explained that information is measurable and not something ambiguous as we originally believed. Since then, he is referred to as the father of information theory [47]. One must understand that information usually is measured in bits. One bit of information gives you the opportunity to choose between two equally probable alternatives and a bit is defined as an amount of information [48]. A common way of explaining bits is the traveler example in Figure 2.5.

In this example a traveler must choose between forks in the road. At the first fork he has two options to choose from, at the second option he has had 4 options, at

the third fork he has had 8 options and so on. This is where one could think of these decisions as the equation $m = 2^n$ where m is the number of final destinations and n is the amount of forks encountered. Which is also equal to $n = \log_2 m$ [48].

Shannon information is measured in bits, which is essentially measuring the unexpectedness. Shannon information = $\log \frac{1}{p(x)}$ bits, which is the same as = -log($p(x)$) bits where $p(x)$ is the probability of x happening. This allows us to segue into Shannon entropy which is the average level of uncertainty related to a variable's outcome.

2.5.1 Shannon Entropy

Shannon entropy is defined as a probability distribution:

Definition 2.14. *$p = (p_1, ..., p_n)$: $H(p) = -\sum_{i=1}^{n} p_i \log_2(p_i) = \sum_{i=1}^{n} p_i \log_2 \frac{1}{p_i}$ where for $p_i = 0$, $p_i \log_2(\frac{1}{p_i})$ is defined as 0 [49].*

Essentially, the equation sums up the variable's possible outcomes or in other words its uncertainty. To be specific, the $H(p)$ is equal to the final entropy. The p_i is the probability of an event and the $\log_2(p_i)$ is the logarithm of the events probability. As for its applications, Shannon entropy has been applied to many modern day issues because of it's usefulness to quantify the amount of entropy in a variable. For example, in Maasoumi's article, he used metric entropy to calculate the predictability of the stock market [50]. It has also been applied to electroencephalogram (EEG) signals to detect abnormalities in concussed individuals [51]. As the applications of Shannon entropy grow it has evolved into new forms such as Persistent entropy which is what was utilized to analyze our data.

2.5.2 Persistent Entropy

Persistent entropy stems from Shannon entropy and it measures the amount of entropy from the persistence diagram created when applying persistent homology. However, in order to apply it to persistent homology, the Shannon entropy formula was modified into the Persistent entropy formula which is defined as:

Given a filtration $F = \{ K(t) \mid t \in \mathbb{R} \}$ and the corresponding persistence diagram $dgm(F) = \{ a_i = (x_i, y_i) \mid 1 \leq i \leq n \}$ (being $x_i < y_i$ for all i), let L = $\{l_i = y_i - x_i \mid 1 \leq i \leq n\}$. The persistent entropy $E(F)$ of F is calculated as follows:

$E(F) = -\sum_{i=1}^{n} p_i \log(p_i)$ where $p_i = \frac{l_i}{S_L}, l_i = y_i - x_i$, and $S_L = \sum_{i=1}^{n} l_i$ [34].

Note that l_i is basically measuring the lengths of the barcodes. One of the steps in this project was to interpret the results from the persistence diagrams. In order to interpret the results, we used the persistence diagram's entropy. The persistence diagram's entropy consists of the main ideas from Shannon entropy, which quantifies

the amount of information contained in an object. In other words, persistent entropy calculates the Shannon entropy of the persistence diagrams.

A barcode with uniform lengths has small entropy. However, the more different the barcodes are in length the greater the entropy. All in all, computing the entropy of persistence diagrams can provide information about the shape of the data. Usually, persistence entropy values are used to analyze two separate groups. For example, persistence entropy has been used by Ruccoa et al [52] to compare good versus faulty motors. It was found that the motors with an entropy value over 7.5 was more likely to be good motors [52]. Therefore, they were able to evaluate what level of entropy could lead to optimal results. In this project, entropy played a similar role and it was used to compare the complicated pregnancy versus uncomplicated pregnancy groups. In specific, we used high entropy values to relate the likelihood for a woman to experience a complicated pregnancy. However, instead of looking only into which high entropy value relates to undesirable outcomes, we explored alternative approaches including how variability in entropy values through the days for G22 and G32 related to pregnancy outcome.

CHAPTER 3

Modeling the Rest-Activity During Pregnancy with Persistent Entropy

This project used data from a previous study conducted between August 2014 to October 2015 at Purdue University and approved by the Institutional Review Boards (IRB), Indiana University, and Eskenazi Health (1405014885).

3.1 Study Design

In this section, we will describe the data that was collected from the pregnant women in this study. The women who participated in this study had their age, race, income, and highest level of education collected from them. It was also noted if they had preeclampsia, hypertension (high blood pressure), and/or gestational diabetes. Preeclampsia is described as something that happens during pregnancy where the mother has high blood pressure, excessive amount of protein in their pee and swelling of their feet [53].

These complications not only affect the mother but it also affects the fetus. A common effect of preeclampsia on the fetus is deficiency of oxygen and nutrients which can hinder fetal growth, preterm birth, stillbirth, and infant death according to the National Institute of Child Health and Human Development (NICHD)[54]. Hypertension in pregnant women causes preterm delivery, and low birth weight. This is because less oxygen and nutrients to get to the baby according to the CDC [55]. Thirdly, gestional diabetes can cause the fetus to gain weight, making the pregnancy harder on the mother. It is also at risk for preterm birth, having low blood sugar, and developing type 2 diabetes later in life according to the CDC [56].

In addition, their blood glucose level was measured after 50 g glucose challenge test. The 50 g glucose challenge test was performed between 24–28 weeks pregnancy. The women were diagnosed with preeclampsia if their blood pressure was higher than 140/90 mm Hg from 20 weeks on during pregnancy with increased protein in their urine. Similarly, hypertension is diagnosed when their blood pressure is consistently measuring 140/90 mm Hg or greater after 20 weeks of gestation. Additionally, a survey was used to get the demographic data which was retrieved with their verbal and written consent. The participants also provided written consent to abstract protected health information (PHI). Also, their PHI was collected and this included their

pre-pregnancy body mass index (BMI). BMI is computed using a person's height and weight and is used as an index of an individual's body fat [57, 58]. Furthermore, the women were recruited from clinics associated with the Eskenazi hospital in downtown Indianapolis. The goal of the hospital is to serve the economically disadvantaged neighborhood it is placed in. As for this study, women with low income and African American backgrounds were recruited in efforts to give these women more representation. Participants were given $150 or $300 in gift cards as compensation. In order for the participants to receive compensation, they needed to complete all data collection. The greater amount was given to those who completed saliva sampling every 4h for a 24h period after actigraphy data was collected. The saliva samples were not used in this study specifically, but they were analyzed in [59]. As for the fetus, the information collected from them was the following: the number of days they stayed at the hospital after birth, their gestational age, birth weight/length and if they had gestational disease. These variables are shown in Figure 3.1.

Now we will go over the demographics of the women. Of the women who were eligible to participate in the study, 50 participated in week 22 and 46 participated in week 32. However, for this study, the sample size was 41 women in G22 and 44 women in G32. When they were recruited they had less than 22 weeks of gestation at the time. The loss of the fetus or any other life event resulted in withdraw and if they did not have at least 5 days of data then they were withdrawn as well. Only women with a single fetus were used for this study. Specifically, the women in this study were 65.8% African American for G22 and 63.6% for G32. They were also between the ages of 18-40. Furthermore, these women were from an economically disadvantaged population. Their highest level of education was high school or GED for 68.2% of them in G22 and G32. Additionally, 65.8% of them only made up to $24,999 yearly for G22. Only 9 out of 41 women were out of work for G22. For G32, their highest salary was $24,999 yearly for 63.6% and only 10 out of 44 were out of work. No difference in demographic variables between recruited, retained and withdrawn populations. These characteristics are displayed in Figure 3.2 and in Table A.5.

3.2 Methodology for Labeling Complicated and Uncomplicated Pregnancies

The women were split into two groups complicated and uncomplicated. Specifically for this study, it was decided that a woman had a complicated pregnancy if the newborn had any of the following: irregular birth weight, irregular birth length, abnormal gestational age, disease, prolonged number of days at the hospital after birth.

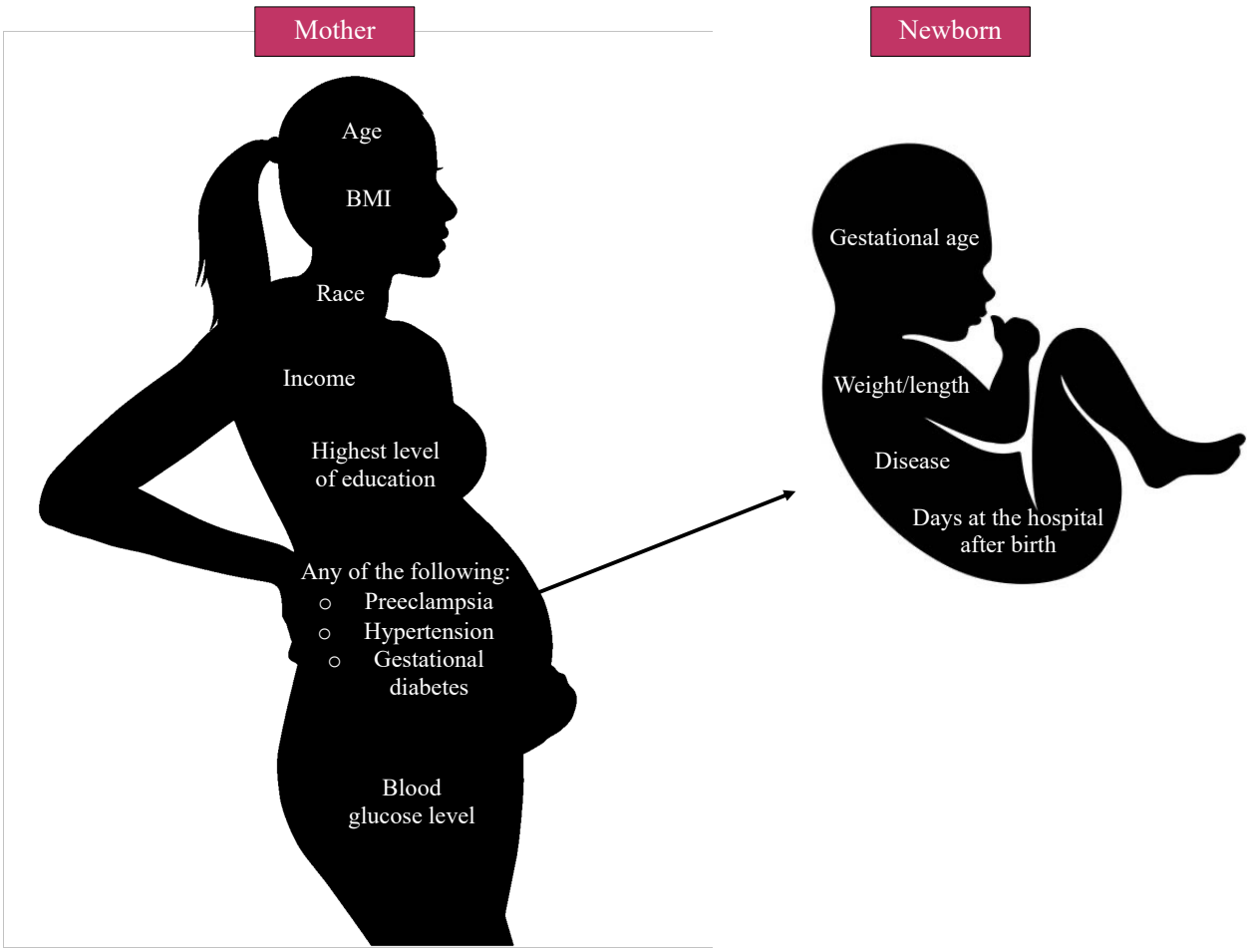


Figure 3.1. The pregnant woman on the left shows the information that was collected for women the study. The fetus on the right shows the features collected for the newborns in the study.

Or if the mother had any of the following: preeclampsia, hypertension and/or gestational diabetes. Specifically for the fetus, if their birth weight was not between 2.5-4.0 kg, if their birth length was not between 47-53 cm, if their gestation period was not between 38-42 weeks, and lastly if they were discharged after more than 4 days from the hospital after they were born (See Table A.4). These were decided upon based on average measurements of newborns. According to the University of Rochester Medical Center, and the World Health Organization the average weight for newborn baby born between 37-41 weeks is 3.2 kg [60, 61]. For this study we decided to use a range from 2.5-4.0 kg because according to Medical News Today, the average is 2.5-4.0 kg for those baby born full term [61]. As for their average length, if a newborn's average length is 47-53cm that is considered to be normal according to the National Library of Medicine [62]. Furthermore, normal pregnancies last between 37-41 weeks [63], anything below 37

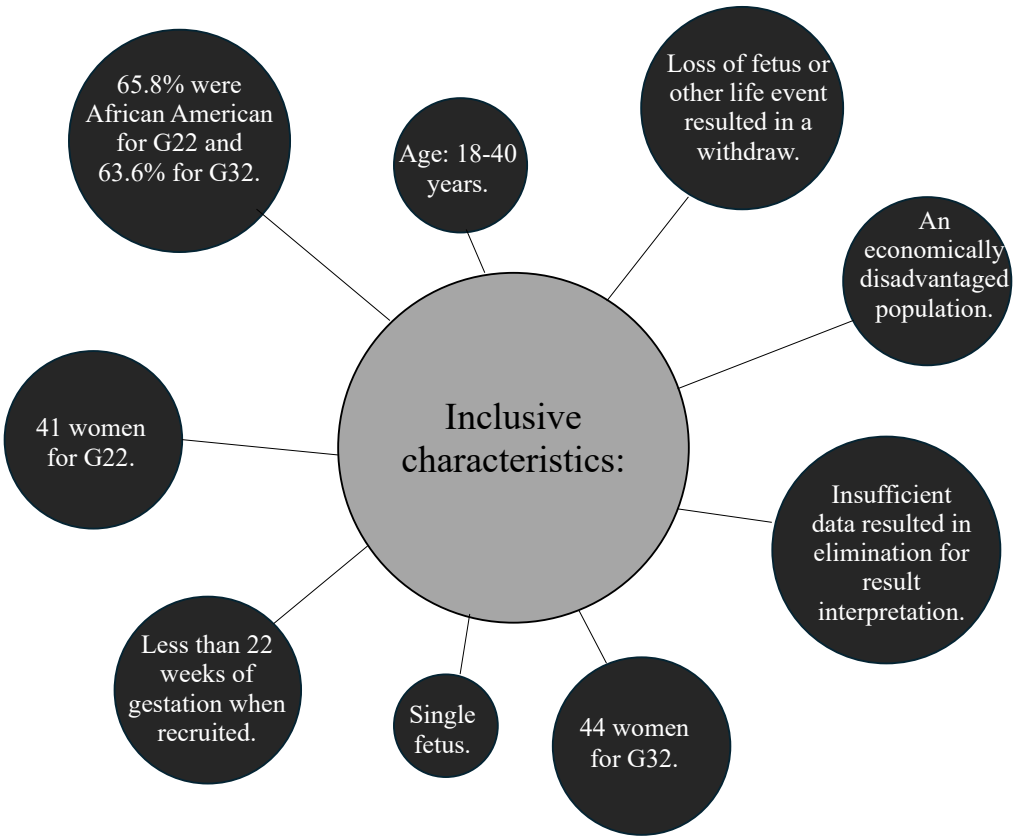


Figure 3.2. The characteristics of the women for this study.

is considered premature, and babies born before 37 weeks have higher risk of death or disability [64]. On the other hand, giving birth after 42 weeks gives you a higher chances of a stillbirth or other complications and doctors may need to induce labor [65]. Lastly, the average length of stay of newborns in the hospital is 1.6 days [66]. Those born by cesarean were discharged between 2-4 days [67]. Babies that stayed longer in the hospital were associated with having more birth complications [68]. From this research we were able to create guidelines for which group the women would be assigned to.

3.3 Actigraphic Data Collection

The participants of this study were asked to wear an actigraph wrist device (Actiwatch Spectrum, Philips Respironics, Andover, MA) on their non-dominant wrist during G22 and G32. There data was recorded on an excel sheet for both gestation weeks. Before G22 and G32, there was a scheduled time for picking up the actiwatch. At this pick up time the instructions on how to use the device were given. There was also a scheduled time for dropping off the watch after the data collection. The actigraphic data was collected in 30-sec epochs and Actiware software was used to export the data to excel.

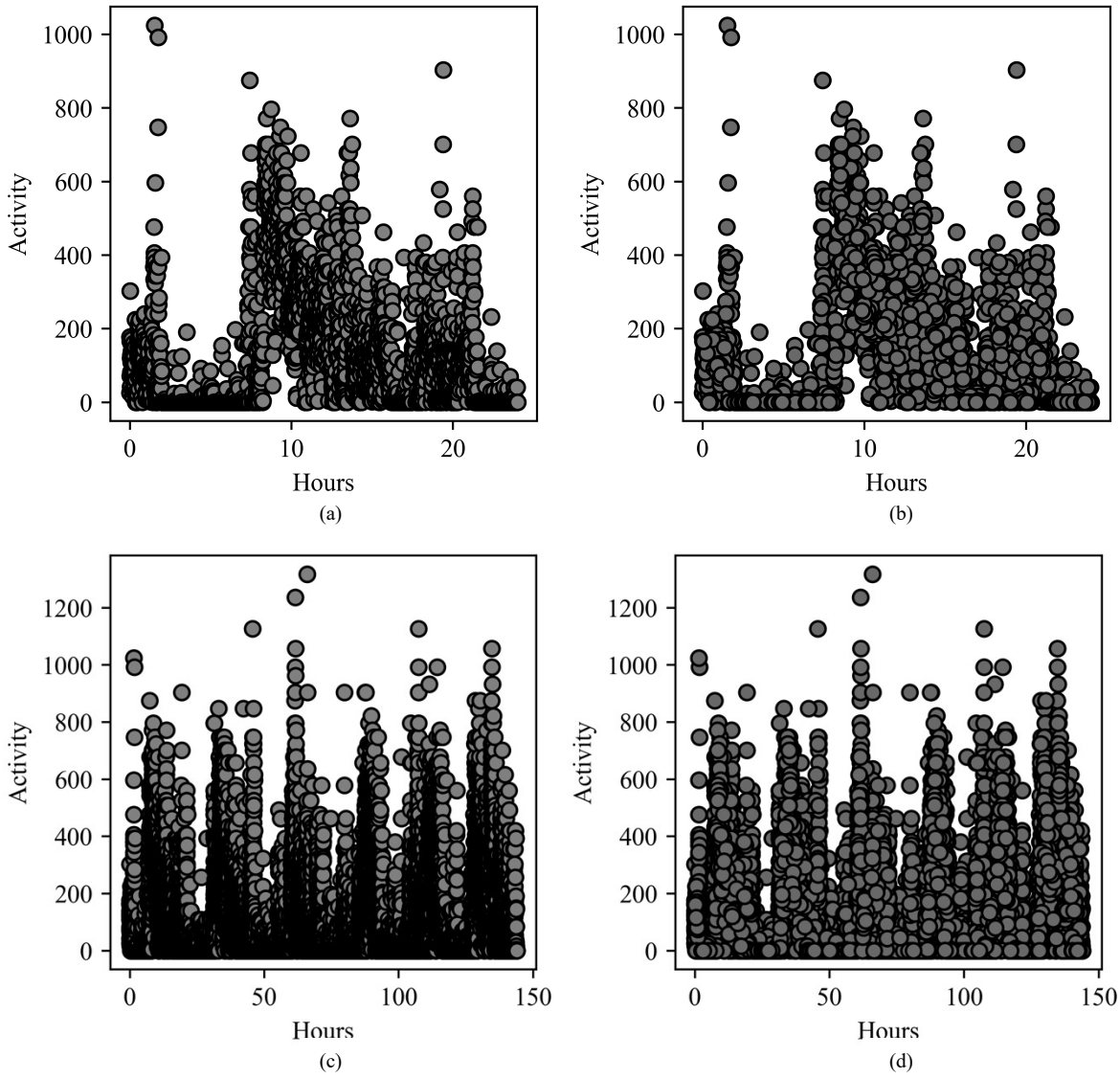


Figure 3.3. Greedy Permutation re-sampling. (A) is the original point cloud of woman 5 from the study, but for the first day of G22. (B) is what the point cloud looks like after the greedy resampling. The greedy resample reduced the original point cloud to 2000 points and the scatter plots looks practically the same as the original. (C) is the original point cloud of 17254 points of woman 5 from G22 versus (b) is what the point cloud looks like after the greedy resampling. In this case we cut it down to 8640 points. The differences are almost unnoticeable. Overall, the greedy resample maintains the shape of the data while reducing the number of data points so that the data is more manageable.

Once these women were properly labeled, their actigraph data was recorded during week 22 and 32 of gestation. It was recorded over consistent intervals of time

producing time series data. This data was all recorded and saved as an excel sheet so that it could later be analyzed with persistent homology and entropy.

3.4 Greedy Permutation Resampling

Multiple scatter plots were made for each of the women in the study to visualize the daily and weekly temporal trends in rest-activity. The data was very large (>17,000 data points for the majority of the women in the study). It was computationally expensive to use the full data set to compute the persistence diagram and entropy. Therefore the data was resampled to 8,640 points per week and to 2,000 points per day using a *Greedy Permutation* algorithm [69] (Figure 3.3). The data was able to be reduced to around half the original size, and still get significant results. The number of points were deduced through trial and error, while keeping the computational cost in mind. The purpose for using the Greedy Permutation resampling method was to pick out points that maintain the overall shape and keeps the key features of the data intact in order to make it easier (i.e. by reducing the computational cost) to compute the persistence diagram and entropy for the women. The Greedy Permutation resampling was able to keep the shape of the data, while significantly reducing the amount of time it would take to complete the computation. Figure 3.3 shows an example of the data points before and after being resampled and demonstrates the ability of Greedy Permutation algorithm to keep the essential visually characteristics of the original data points in the re-sampled data. The Greedy Permutation is based on the Cavanna et al [70] definition:

Definition 3.1. *Let P be a set of points in some metric space with distance d. We say that $P = \{p_1, \ldots, p_n\}$ is ordered according to a greedy permutation if each p_i is the farthest point from the first i - 1 points. We let p_1 be any point. Formally, let $P_i = \{p_1, ..., p_i\}$ be the ith prefix. Then, the ordering is greedy if and only if for all $i \in \{2, \ldots, n\}$, $d(p_i, P_{i1}) = \max_{p \in P} d(p, P_{i-1})$. For each point p_i, the value $\lambda_i := d(p_i, P_{i-1})$ is known as the insertion radius. By convention, we set $\lambda_1 = \infty$. It is well-known (and easy to check) that Pi is a λ_i-net in the sense that it satisfies the conditions: for all distinct $p, q \in Pi$, $d(p, q) \geq \lambda_i$ (packing) and $P \subseteq P_i^{\lambda_i}$ (covering).*

A more simple explanation is that the resampling works by picking a first arbitrary data point and then selecting the next furthest data point that has not been selected yet. By doing this the greedy permutation allows the shape of the data to remain while reducing noise/excessive points. Observe that there is not much visual difference between the original and re-sampled data (Figure 3.3). After resampling the data, the daily and weekly persistence diagram and entropy were computed for each

woman. We compared the weekly persistent entropy values for women in the complicated pregnancy group to those in the uncomplicated pregnancy group. We calculated the daily variability in entropy values for each woman and then compared the daily variability values for women in the complicated to those for women in the uncomplicated pregnancy group. Statistical analysis to test for differences between women in complicated and uncomplicated pregnancy where carried out using the Fisher test and Mann Whitney U test.

3.5 Calculating the Persistence Diagrams for the Actigraphy Data for Each Pregnant Woman

After the data points were resampled using Greedy Permutation, the data points were run through a Python code to obtain the Vietoris-Rips filtration and persistence diagram. The Vietoris-Rips filtration and persistence diagram was computed using the *Ripser Python package* [71]. This was done for each women in both G22 and G32. Each diagram for each woman was recorded specifically for daily and weekly.

3.6 Calculating the Persistent Entropy in Rest-Activity for Each Pregnant Woman by Using the Actigraphy Data

Once the persistence diagrams were obtained, the next step was to calculate the entropy for each woman per individual day in G22 and G32 and for the whole G22 and G32 weeks. The entropy value of each persistence diagram was calculated using the persistent entropy formula using *Persim Python package* [72]. These values were stored and later used for statistical analysis.

3.7 Calculating the Average Daily Entropy Value for Each Woman

The average daily entropy for each woman's activity was computed for G22 and G32 using root mean square formula (RMS). The RMS is often used as a way to measure distinctiveness between samples. The RMS formula is RMS = $\sqrt{(x(1)^2 + x(2)^2 + ... + x(n)^2)/n}$, where $x(i)$ denotes the activity of a given woman at time i and n is the number of time points.

3.8 Calculating the Variability of Daily Entropy Values for Each Woman

The variability of their daily entropy was computed for each women using median absolute deviation (MAD). MAD is preferred over other measures of variability in samples that may have outliers. The formula is MAD = median($|X_i - \tilde{X}|$). Where X_i is our data set, and $\tilde{X}$ = median(X).

CHAPTER 4

Results

The entropy value for each woman's rest activity was computed for each day separately and for the whole 6 days of the week that the data was collected. This was done for both G22 and G32, however, the most significant results came from week 22. The data was analyzed using Mann Whitney U test, Fisher tests, RMS and MAD, which is further explained in the next sections.

4.1 Average Daily Entropy Value for the Women's Rest-Activity is Slightly Associated with Pregnancy Outcome

The entropy values were first visualized using boxplots. Figure 3.3b shows a re-sampled rest activity for a 24-hours duration, representing one day of the week. The entropy value for the rest activity was computed for each day, which we refer to as the daily entropy value. Figure 3.3d shows a re-sampled rest activity for a 144-hours duration, representing six days of the week. This was used to estimate the weekly entropy value for each woman's rest-activity. We show the daily entropy values for each woman's rest activity in Figure 4.1 using box plots. Figure 4.1a and b are the box plots for the daily entropy values for each woman in G22 and G32 respectively. The box plot for each woman shows the distribution of the daily entropy values for the six different days. As we would expect, the women's entropy values were different for each day, because the level of human activity often vary daily. Also, the distribution of the daily entropy values where mostly different for each woman in G32 compared to G22. The box plots that are colored in magenta denote women who had complicated pregnancies and the ones that are colored in blue denote women that had no complications.

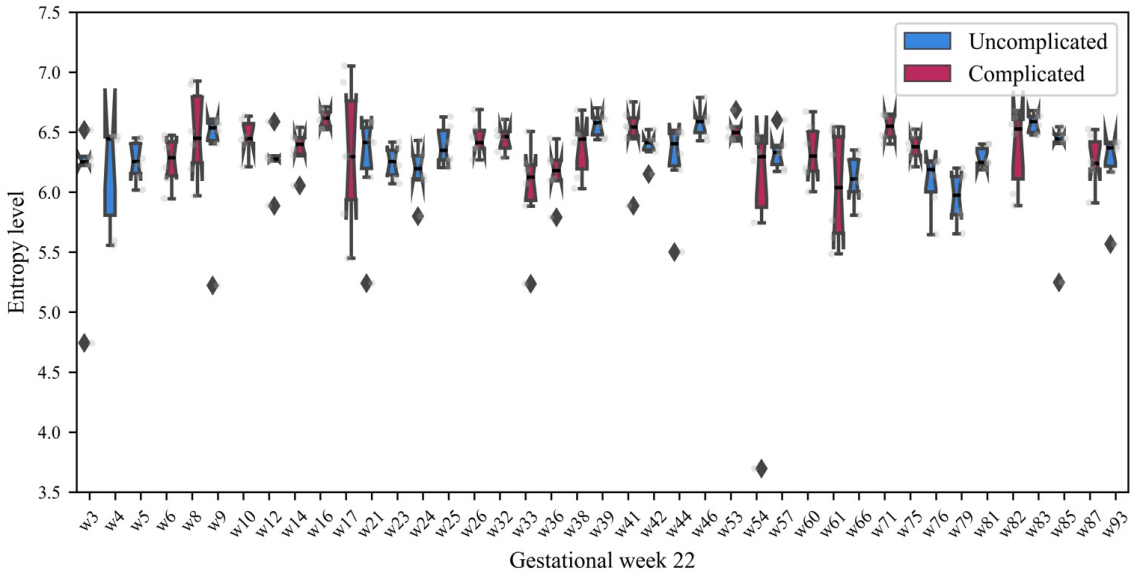
We investigated if the average entropy values were associated with pregnancy outcomes in G22 and G32. We estimated the average daily entropy values by using the RMS formula. The box plot in Figure 4.2, shows the average daily entropy values for each woman in G22 and G32. We compared the average daily entropy values for women with complicated pregnancy to those with uncomplicated pregnancy in G22 and also in G32. We found there was a difference in average daily entropy values for women with complicated pregnancy versus uncomplicated pregnancy in G22 (P-value = 0.048 for one-tailed Mann Whitney U test; P-value = 0.095 for two-tailed Mann Whitney U

test). We also found that there was a significant difference in average daily entropy values for women with complicated pregnancy versus uncomplicated pregnancy in G32 (P-value = 0.067 for one-tailed Mann Whitney U test, P-value = 0.133 for two-tailed Mann Whitney U test). These differences where either close to or statistically significant for both one-tailed and two-tailed Mann Whitney U test, which implies that the average daily entropy values in G22 and G32 may potentially be associated with women's pregnancy outcome. Further studies are needed on a larger sample of pregnant women in order to be able to draw reasonable conclusion.

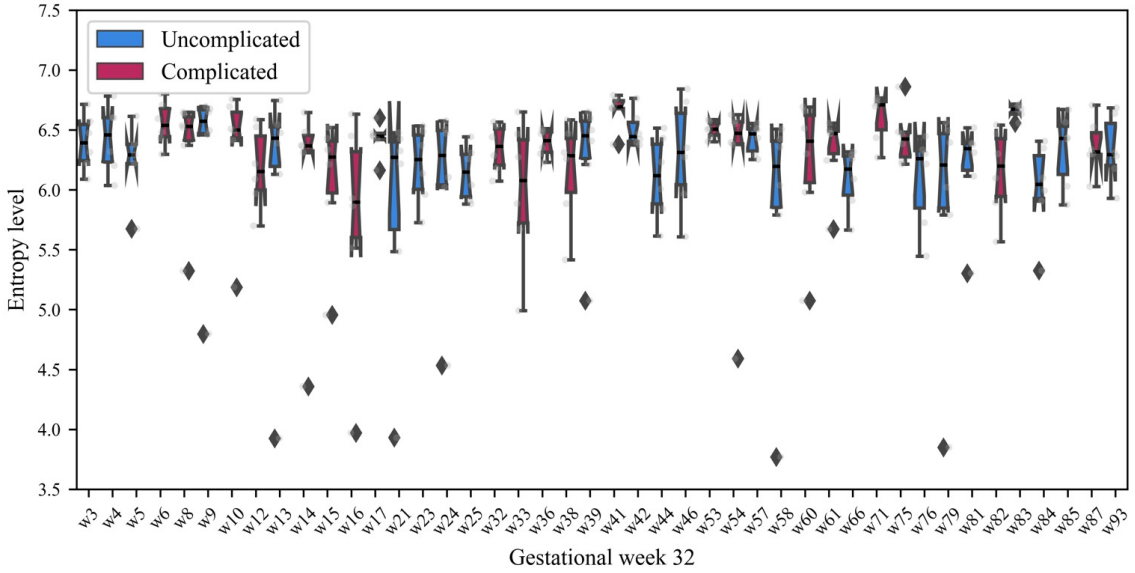
4.2 Variability in G22 Daily Entropy Values is Associated with Complicated Pregnancies

One could see that the box plot for the women often had large variations in daily entropy values (Figure 4.1). MAD was used to compute the degree of variability in each woman's daily entropy values over the six days. Using MAD, we determined if the variability in the women's daily entropy value is associated with pregnancy outcomes by comparing the median absolute deviation for women in the complicated pregnancy group to those in the uncomplicated pregnancy group for both G22 and G32. The variability in daily entropy values for each woman was measured by computing the MAD of the daily entropy values for the six days. We found that the MAD was significantly different for women in complicated versus uncomplicated pregnancy groups for G22 (one-tailed Mann Whitney U test: P-value=0.028 for G22, P-value= 0.881 for G32. Two-tailed Mann Whitney U test: P-value = 0.055 for G22, P-value = 0.248 for G32).

The variability in daily entropy values increased significantly in G32 for women in the uncomplicated pregnancy group when compared to G22. The results in Figure 4.3 implies that the women have higher activity levels in some days than others in G32 compared to G22. It is remarkable that the variability in the values of the persistent entropy shows differences in rest-activity patterns in G22 versus G32 for women in uncomplicated pregnancy group. This would be expected since the consistency in rest-activity levels may have a higher tendency to change because the women get heavier in G32 since they are nearer to full-term pregnancy. At this point in the pregnancy (i.e. G32), the baby is far along the developmental stage so there could be a chance that the irregular activity affects the child during earlier stages of pregnancy and less during late stages of pregnancy. For instance, according to the National Institutes of Health (NIH), many important developmental changes occur during the first trimester of pregnancy, and therefore harmful exposures during this time cause the



(a)



(b)

Figure 4.1. Daily entropy values for G22 and G32. (A) is the distribution of daily entropy values in G22 and (b) depicts the same but for G32. The x-axis denotes the women in the study and the y-axis denotes the daily entropy value.

greatest birth defects [73]. Therefore, the variability in daily rest activity in G22 is more associated with pregnancy outcome. Nevertheless, more research is needed to see how irregular activity impacts the fetus during early gestation versus the later stages of pregnancy.

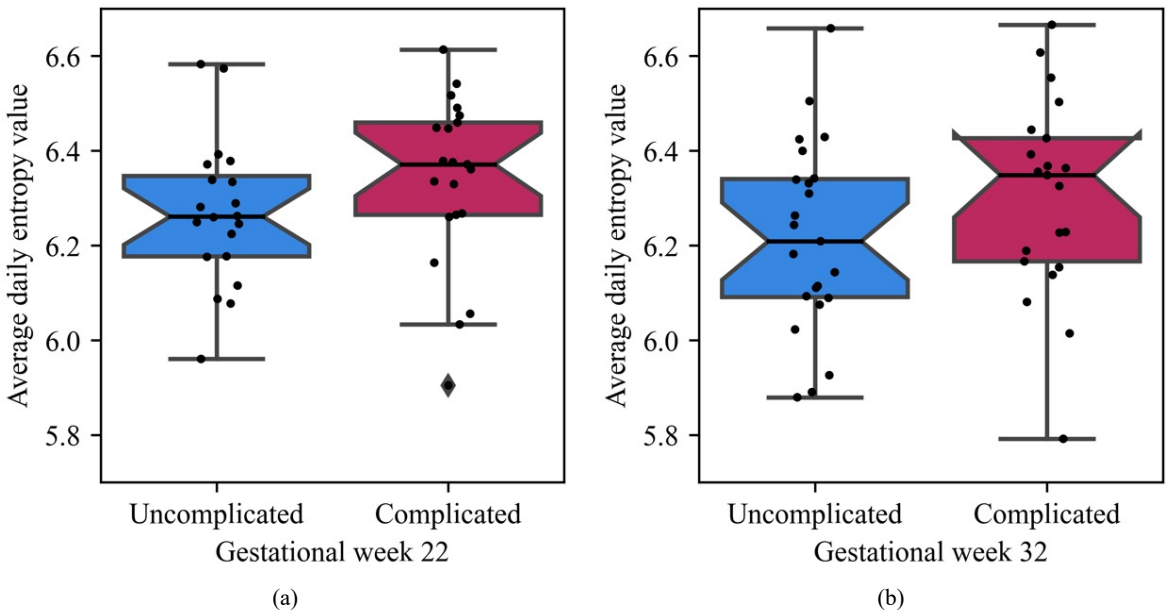


Figure 4.2. Average daily entropy value for (a) G22 and (b) G32. The average daily entropy is computed using the RMS formula.

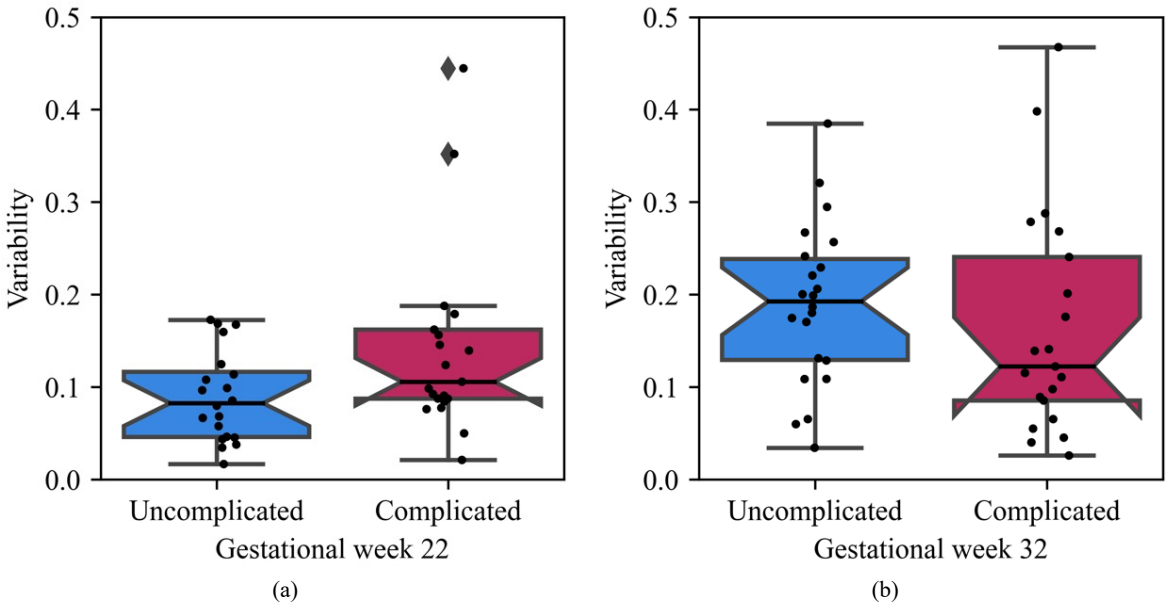


Figure 4.3. Variability in daily entropy values for (a) G22 and (b) G32. Variability was measured using MAD formula.

4.3 Number of Pregnancy Complications is Positively Associated with Number of High Daily Entropy Days in G22

Another way in which the data was inspected was by looking at the women that had more complications than the others and then relate it to entropy values. In order

to do this, we counted the number of days each woman had an entropy value greater than 6.5 per day (i.e. high entropy days). Next, we counted the number of pregnancy complications/abnormalities that occurred in each woman and then we relate these to the number of high entropy days (Figure 4.4). The box plot in Figure 4.4 shows that the number of pregnancy complications increases with the number of high entropy days in G22. In other words, the more days of high entropy values recorded the more the number of complications/abnormalities a woman exhibited. For example, for both G22 and G32, women who had 3 high entropy days out of that week had an average of 3 to 4 complications. However, G32 did not have such a well-defined increasing trend as G22 did. Which is consistent with G22 rest activity being more predictive of pregnancy outcomes than G32.

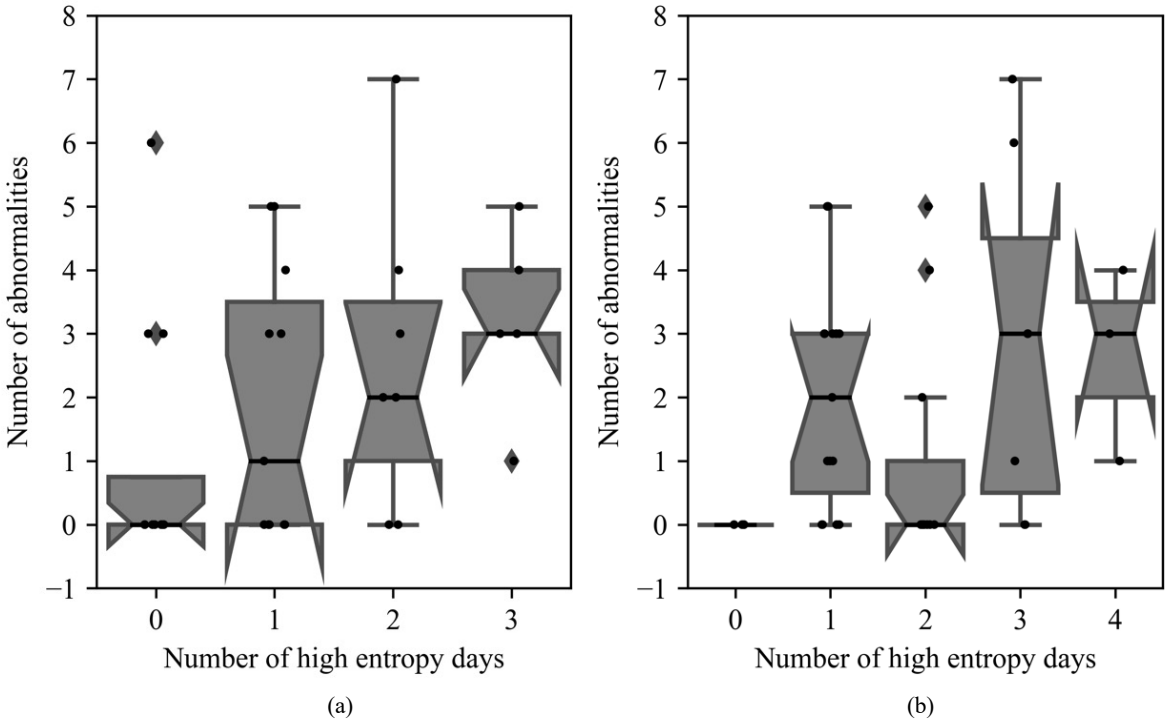


Figure 4.4. Comparison of the number of days women had high persistent entropy values to the number of complications/abnormalities that occurred during pregnancy. (A) is the number of days women had persistent entropy values over 6.5 versus the number of diseases for G22. (B) had the same process done but for G32.

4.4 Persistent Entropy Values for the Whole Duration of G22 Was Associated with Pregnancy Outcome

In addition to analyzing the daily entropy values for the rest activity, we also analyzed the entropy value for the whole G22 and G32 rest activity for each woman. Figures 4.5 and 4.6 show the entropy values for the rest activity for each woman for G22 and G32 respectively. Visually, we see that for G22 (Figures 4.5) the persistent entropy of women in the complicated pregnancy group where on average higher than those in uncomplicated pregnancy group. Through visual inspection of the plot in Figure 4.5, we found that women that had persistent entropy values above 8.22 in G22 were more susceptible to having a complicated pregnancy compared to women with entropy values below 8.22 (Fisher test, P-value = 0.003, Table A.1). Specifically, the majority of the complicated pregnancies resided above the entropy value of 8.22 in G22. As one can see in Figure 8, only 3 magnenta colored dots were under the threshold of 8.22 entropy value. This implies that pregnancy outcomes may be influenced by activities that occur early in pregnancy (i.e. at G22). On the other hand, for the week 32 entropy values (Figure 4.6), there was no significance difference between complicated and uncomplicated women groups based on the 8.22 threshold value or any entropy value. The two groups were distributed equally above and below the 8.22 threshold. This could be because at G32, the pregnancy is far along and the women have become larger resulting in less mobility for most of the women in both complicated and uncomplicated pregnancy groups. This assumption is supported by our observation that G22 has lower entropy values than G32 (Figures 4.5 and 4.6), but more research is needed to confirm why there wasn't enough distinctive features in G32.

4.5 Women with High BMI (i.e. BMI >25) and High Persistent Entropy at G22 Were More Likely to Have a Complicated Pregnancy

BMI is often linked to chronic medical conditions and poor prognosis in pregnant women [74, 75]. After we found that G22 entropy values was related to pregnancy complications/outcome, we decided to investigate how BMI relates with the G22 entropy values and pregnancy outcomes. Usually, a person with BMI between 15-19.9 is said to be underweight, 20-24.9 is normal weight, 25-29.9 is overweight and >30 is obese [57]. We studied two different groups, those with BMI above the normal range (i.e. > 25, overweight and obese women) and those below 25 (i.e. underweight

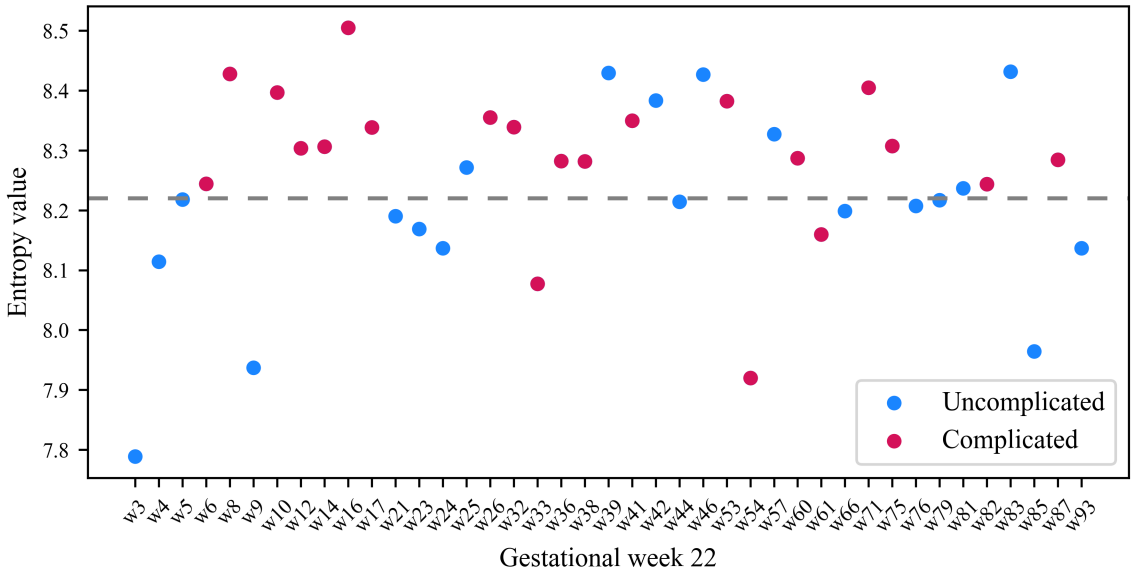


Figure 4.5. Persistent entropy values for G22. The x-axis represents the different women and the y-axis represents their entropy values. The dashed line denotes where the entropy value equals 8.22.

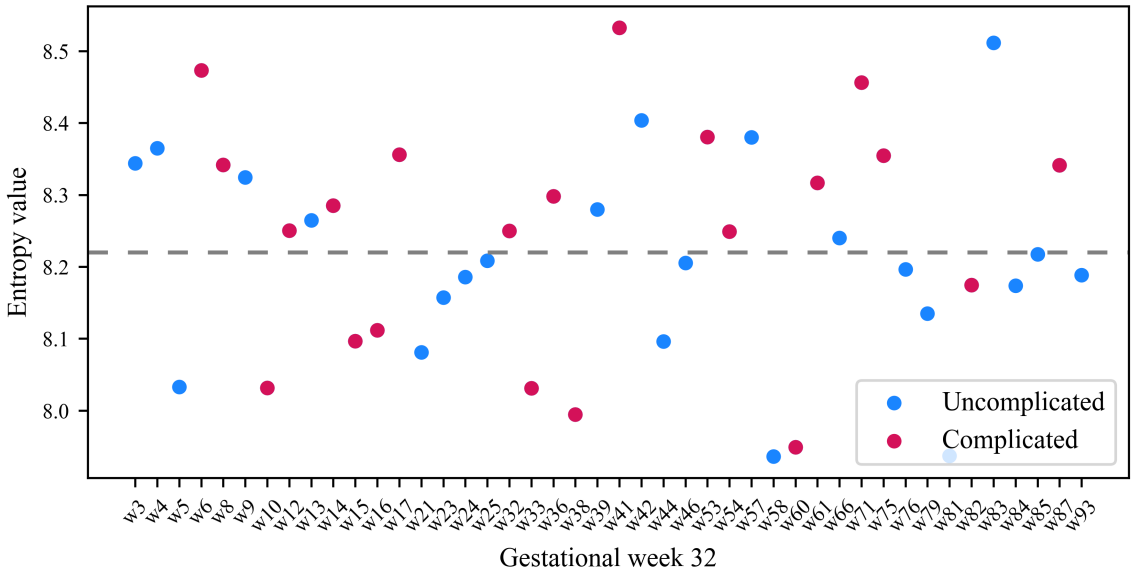


Figure 4.6. Persistent entropy values for G32. The x-axis represents the different women and the y-axis represents their entropy values. The dashed line denotes where the entropy value equals 8.22.

and normal weight). We compared how entropy values relate to complicated versus uncomplicated pregnancy in these two different groups. Figure 4.7 is a plot of the persistent entropy values against BMI for G22. Visually, we see that those who had

both an entropy value larger than 8.22 and a BMI > 25 were more likely to have a complicated pregnancy than those with an entropy values less than 8.22 (top rightmost quarter versus bottom rightmost quarter). In particular women with BMI>25 where more likely to be in the complicated pregnancy group if their entropy values were higher than 8.22 (Fisher test: P-value = 0.0198, Table A.2. In order words, entropy values is a greater predictor of pregnancy outcomes than BMI in G22 for those with BMI>25. We note that women with BMI <25 had a higher tendency to have complicated pregnancy if they had persistent entropy values greater than 8.22, however, this was not statistically significant (Fisher test: P-value = 0.099, Table A.3).

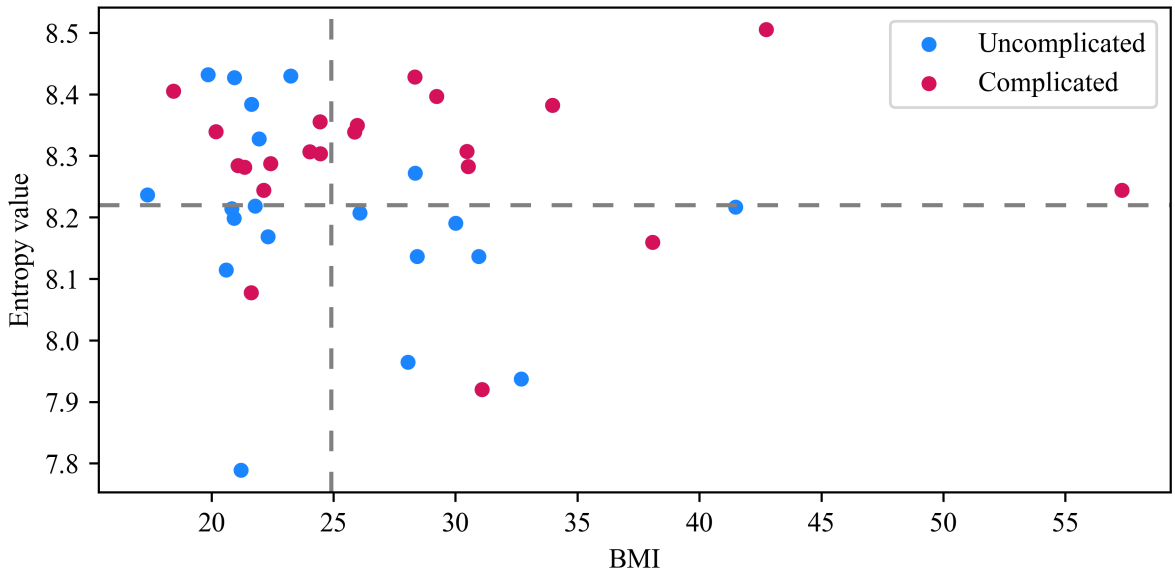


Figure 4.7. G22 entropy values were plotted against BMI. The vertical dashed line (i.e. BMI = 25), separates healthy versus unhealthy BMI. The horizontal dashed line represents an entropy value of 8.22.

CHAPTER 5

Discussion

The findings from this study are novel and significant. It presents a new method to predict future maternal and newborn outcomes by using data obtained at G22. This study contributes to the body of work that demonstrate that topology can be a great tool for analyzing data. Topology is a original way to pick apart noisy features and allows us to focus on the important ones. Figure 4.1, allowed us to understand how the women with complicated pregnancies had longer boxplots in G22. Longer box plots portrays that is has a larger interquartile range, which means more variability. However, in G32, both the complicated (magenta) and uncomplicated (blue) had long boxplots. This lead us to measure the variability for both weeks.

We computed the daily average of entropy values for G22 and G32. In Figure 4.2, we found that the difference between the two groups were only statistically significant for a one-tailed Mann Whitney test. This indicates that in G22 the complicated group had significantly higher average daily entropy values than the uncomplicated group (P-value = 0.048; one-tailed Mann Whitney test). This finding coincides with the proposition that having more irregular/random activity during pregnancy could lead to negative impacts for the women and newborns. On the other hand, G32 was not statistically significant at all with either the one-tailed or two-tailed test. This means that the average entropy values were not significantly different between the complicated and uncomplicated groups in G32. In other words, this result could imply that the two groups had similar rest-activity patterns for G32. A reason for this could be that at this point the women are very large and that may decrease their mobility. This in turn makes their rest-activity data more similar to each other, and the persistent homology/entropy can't distinguish the differences between the two groups. A future alternative for this could be to look at their night time sleep only instead of the whole day. Topology may be able to extract more differences during their sleep hours.

In Figure 4.3, we compared the variability in daily entropy values between the complicated versus uncomplicated groups. We found a close to significant difference in the variability in daily entropy values for G22 between women with complicated versus uncomplicated pregnancy (P-value = 0.055; two-tailed Mann Whitney test). Additionally, we found that the entropy variability was significantly higher in the complicated group with a one-tailed test (P-value = 0.028; one-tailed Mann Whitney

test). This result implies that the women in G22 who had a complicated pregnancy had more variability in their entropy and therefore had more random/unpredictable rest-activity data and less consistency. This inconsistency is what we believe could be one of the causes for their complications. We did not observe any differences between the two groups for G32.

After, we looked at how the having many high entropy days related to the complications they had in Figure 4.4. Due to a small sample size we could not test significance, however, it is interesting to see how the women with frequent entropy days had increasing complications. However, we can not say that the high entropy values in their rest- activity caused these complications. More research would have to be completed in order to know the true reason behind this result.

In Figure 4.5, we measured the entropy of the whole week and found that the women with a week's persistent entropy value of above 8.22 were more likely to have a complicated pregnancy. However, this significance is lost in Figure 4.6. Again, we believe that this is because the women at the time for G22 had more mobility and were taking advantage of that and as a result their rest-activity data is more random and scattered. In Figure 4.7, we can see that the women were more likely to have a complicated pregnancy if they had a higher entropy value than 8.22 and a BMI higher than 24.99. This is a major finding from our study. The Fisher test was highly significant, so it would be beneficial for future research to investigate how the BMI and high persistent entropy relate to each other and if they are key components in affecting the pregnancies. Furthermore, this result may be portraying that the women's prepregnancy BMI does play an important role in what the outcome of the pregnancy will be.

One key take away from all these results is that a disadvantage about this method is that it may not work as well for data that has no distinct features. That is, when comparing data sets with persistent entropy, if there are no distinct features present that can set the samples of data apart, then one might not be able to successfully contrast the sets of data. We could see this happen in G32. That week's results did not have enough features to make it significantly different between the complicated and uncomplicated pregnancy groups. Which we believe one reason behind this result could be because of the lack of differences in mobility between G22 and G32 due to their late pregnancy stage. In other words, because of the women's large bellies, their mobility is limited and therefore their activity levels for that week may all look similar to each other with no distinctive features. However, there may be other explanations as to why there was no significant distinctions between the two groups.

Furthermore, it is important to point out that in this day and age of big data, persistent homology and entropy can be a great way to analyze and compare groups because studying the shape of data is relatively simple and can indicate particular important features. An example is in the current study, where we were able to determine that the majority of women who had a higher persistent entropy value for rest-activity patterns had a complicated pregnancy.

CHAPTER 6

Conclusions

Overall, one of the most important outcomes of this research was that by using persistent entropy, a method in topology, we were able to relate rest-activity patterns to maternal health and outcome based on simply just the shape of the data. Furthermore, the study's findings support the idea that there may be a relationship between persistent entropy of rest-activity data and complicated (or uncomplicated) pregnancy. In particular, we found that having a high entropy rest-activity in G22 is associated with a higher tendency to have pregnancy complications. We also found that the rest-activity entropy values in G22 maybe a greater predictor of pregnancy outcomes than BMI for those with BMI>25. In order to reap the benefits of using persistent entropy, this study needs to be replicated in a larger representation of pregnant women to access the predictability of pregnancy outcomes using persistent entropy. The findings from this projects contributes to studies in the literature that highlight the potential of using techniques in topology for predicting health outcome.

This study was able to show how just analyzing the shape of the rest-activity data was enough to relate physical activity with the well-being of pregnant women. Of course, one must keep in mind that if the data is too large it may be necessary to find ways to make the data more manageable. In the current study, the greedy resampling allowed us to reduce the data and noise, which as a result allowed us to compute the persistence diagrams for each woman in the study in a reasonable computational time. Before resampling the code would take numerous hours to run, however, after the greedy resampling it takes a few minutes per woman. This of course also applies to data that has missing values which often affect the computation of the persistence diagrams. Using greedy resampling, data sets may be resampled to have similar size/number of entries across data sets, before computing the persistence entropy. This study demonstrates the potential of using rest-activity during early pregnancy together with persistent entropy to identify individuals that have greater odds of having a complicated pregnancy. As for future research, this study needs to be replicated in a larger cohort of pregnant women to validate results and access the predictability of pregnancy outcomes using persistent entropy. The amount of women in this study was too small for some of the analysis and increasing the sample size would be beneficial.

Nevertheless, due to the strong statistical results, extending the analysis could lead us to further understand the way rest-activity patterns affect maternal and fetal health.

Additionally, exploring this topic further is important especially for economically disadvantaged women who may need to work more hours and have unfavorable work schedules. These women will be more prone to irregular sleep patterns and if continued for long periods of time it could potentially lead them to dangerous health outcomes for both her and their baby. Future investigation on this subject needs to focus on how the disruption of sleep can negatively affect metabolic-hormonal systems in pregnant women. That way we can know exactly how rest-activity and sleep schedules impact them, since studies have already suggested that it has a great impact on people who are not pregnant. All in all, based on this study and past studies, pregnant women may need consistent rest-activity levels and a good sleep schedule for the sake of their health and the newborns' health. As for practice recommendations, it would be advantageous to look into women who have irregular work shifts/schedules versus women who can take time off and see if there is a correlation between this and their health outcomes. If it is deduced that their irregular schedules is affecting women then policies can be implemented to help prevent this.

BIBLIOGRAPHY

[1] Donna L Hoyert. Maternal mortality rates in the United States, 2020. 2022.

[2] Tracey L. Sletten, Matthew D. Weaver, Russell G. Foster, David Gozal, Elizabeth B. Klerman, Shantha M.W. Rajaratnam, Till Roenneberg, Joseph S. Takahashi, Fred W. Turek, Michael V. Vitiello, Michael W. Young, and Charles A. Czeisler. The importance of sleep regularity: a consensus statement of the national sleep foundation sleep timing and variability panel. *Sleep Health*, 9(6):801–820, 2023.

[3] Stephen M James, Kimberly A Honn, Shobhan Gaddameedhi, and Hans PA Van Dongen. Shift work: disrupted circadian rhythms and sleep—implications for health and well-being. *Current sleep medicine reports*, 3:104–112, 2017.

[4] Steven H Strogatz. Human sleep and circadian rhythms: a simple model based on two coupled oscillators. *Journal of mathematical biology*, 25(3):327–347, 1987.

[5] Helen S Driver and Colin M Shapiro. A longitudinal study of sleep stages in young women during pregnancy and postpartum. *Sleep*, 15(5):449–453, 1992.

[6] Shengkui Zhang, Yongbin Wang, Qinglin Li, Zhende Wang, Han Wang, Chao Xue, Ying Zhu, Weijun Guan, and Juxiang Yuan. Different exposure metrics of rotating night shift work and hyperhomocysteinaemia among chinese steelworkers: a cross-sectional study. *BMJ open*, 10(12):e041576, 2020.

[7] Francesca L Facco, Corette B Parker, Uma M Reddy, Robert M Silver, Matthew A Koch, Judette M Louis, Robert C Basner, Judith H Chung, Chia-Ling Nhan-Chang, Grace W Pien, et al. Association between sleep-disordered breathing and hypertensive disorders of pregnancy and gestational diabetes mellitus. *Obstetrics and gynecology*, 129(1):31, 2017.

[8] Michele L Okun, Roberta A Mancuso, Calvin J Hobel, Christine Dunkel Schetter, and Mary Coussons-Read. Poor sleep quality increases symptoms of depression and anxiety in postpartum women. *Journal of behavioral medicine*, 41:703–710, 2018.

[9] Elizabeth J Mezick, Karen A Matthews, Martica Hall, Patrick J Strollo Jr, Daniel J Buysse, Thomas W Kamarck, Jane F Owens, and Steven E Reis. Influence of race and socioeconomic status on sleep: Pittsburgh sleepscore project. *Psychosomatic medicine*, 70(4):410–416, 2008.

[10] Lydia Feinstein, Ketrell L McWhorter, Symielle A Gaston, Wendy M Troxel, Katherine M Sharkey, and Chandra L Jackson. Racial/ethnic disparities in sleep duration and sleep disturbances among pregnant and non-pregnant women in the united states. *Journal of sleep research*, 29(5):e13000, 2020.

[11] Gretchen Livingston and Deja Thomas. Among 41 countries, only us lacks paid parental leave. 2019.

[12] Piotr Biegański, Anna Stróż, Marian Dovgialo, Anna Duszyk-Bogorodzka, and Piotr Durka. On the unification of common actigraphic data scoring algorithms. *Sensors*, 21:6313, 09 2021.

[13] Nicola Cellini, Matthew P Buman, Elizabeth A McDevitt, Ashley A Ricker, and Sara C Mednick. Direct comparison of two actigraphy devices with polysomnographically recorded naps in healthy young adults. *Chronobiology international*, 30(5):691–698, 2013.

[14] Miguel Marino, Yi Li, Michael N Rueschman, John W Winkelman, JM Ellenbogen, Jo M Solet, Hilary Dulin, Lisa F Berkman, and Orfeu M Buxton. Measuring sleep: accuracy, sensitivity, and specificity of wrist actigraphy compared to polysomnography. *Sleep*, 36(11):1747–1755, 2013.

[15] Roni Zemet, Raoul Orvieto, Hadel Watad, Eran Barzilay, Eran Zilberberg, Oshrit Lebovitz, Shali Mazaki-Tovi, and Jigal Haas. The association between level of physical activity and pregnancy rate after embryo transfer: A prospective study. *Reproductive BioMedicine Online*, 42(5):930–937, 2021.

[16] Janou Korte, Katharina Wulff, Claudia Oppe, and Renate Siegmund. Ultradian and circadian activity-rest rhythms of preterm neonates compared to full-term neonates using actigraphic monitoring. *Chronobiology International*, 18(4):697–708, 2001.

[17] Janou Korte, Thomas Hoehn, and Renate Siegmund. Actigraphic recordings of activity-rest rhythms of neonates born by different delivery modes. *Chronobiology international*, 21(1):95–106, 2004.

[18] Anastasiya Slyepchenko, Luciano Minuzzi, James P Reilly, and Benicio N Frey. Longitudinal changes in sleep, biological rhythms, and light exposure from late pregnancy to postpartum and their impact on peripartum mood and anxiety. *The Journal of Clinical Psychiatry*, 83(2):39211, 2022.

[19] Fred Shaffer and Jay P Ginsberg. An overview of heart rate variability metrics and norms. *Frontiers in public health*, page 258, 2017.

[20] Theresa Casey, Hui Sun, Helen J Burgess, Jennifer Crodian, Shelley Dowden, Shelby Cummings, Karen Plaut, David Haas, Lingsong Zhang, and Azza Ahmed. Delayed lactogenesis ii is associated with lower sleep efficiency and greater variation in nightly sleep duration in the third trimester. *Journal of Human Lactation*, 35(4):713–724, 2019.

[21] Xinyue Li, Michael Kane, Yunting Zhang, Wanqi Sun, Yuanjin Song, Shumei Dong, Qingmin Lin, Qi Zhu, Fan Jiang, and Hongyu Zhao. Circadian rhythm analysis using wearable device data: Novel penalized machine learning approach. *Journal of Medical Internet Research*, 23(10):e18403, 2021.

[22] Kai Li, Heinz Rüdiger, and Tjalf Ziemssen. Spectral analysis of heart rate variability: time window matters. *Frontiers in Neurology*, 10:545, 2019.

[23] Manus Henry. An ultra-precise fast fourier transform. *Science Talks*, 4:100097,

2022.

[24] Margaret M Doyle, Terrence E Murphy, Brienne Miner, Margaret A Pisani, Elizabeth R Lusczek, and Melissa P Knauert. Enhancing cosinor analysis of circadian phase markers using the gamma distribution. *Sleep medicine*, 92:1–3, 2022.

[25] Ruben Fossion, Ana Leonor Rivera, Juan C Toledo-Roy, Maia Angelova, and M El-Esawi. Quantification of irregular rhythms in chrono-biology: A time-series perspective. In *Circadian Rhythm: Cellular and Molecular Mechanisms*, pages 33–58. InTech, 2018.

[26] Ruben Fossion, Ana Leonor Rivera, Juan C Toledo-Roy, Jason Ellis, and Maia Angelova. Multiscale adaptive analysis of circadian rhythms and intradaily variability: Application to actigraphy time series in acute insomnia subjects. *PloS one*, 12(7):e0181762, 2017.

[27] Nicolas Bourdillon, Sasan Yazdani, Jean-Marc Vesin, Laurent Schmitt, and Grégoire P Millet. Rmssd is more sensitive to artifacts than frequency-domain parameters: implication in athletes' monitoring. *Journal of Sports Science & Medicine*, 21(2):260, 2022.

[28] Jia Wang, Hong Xian, Amy Licis, Elena Deych, Jimin Ding, Jennifer McLeland, Cristina Toedebusch, Tao Li, Stephen Duntley, and William Shannon. Measuring the impact of apnea and obesity on circadian activity patterns using functional linear modeling of actigraphy data. *Journal of circadian rhythms*, 9(1):1–10, 2011.

[29] Elizabeth B Klerman, Allison Brager, Mary A Carskadon, Christopher M Depner, Russell Foster, Namni Goel, Mary Harrington, Paul M Holloway, Melissa P Knauert, Monique K LeBourgeois, et al. Keeping an eye on circadian time in clinical research and medicine. *Clinical and Translational Medicine*, 12(12):e1131, 2022.

[30] John K Hunter and Bruno Nachtergaele. *Applied analysis*. World Scientific Publishing Company, 2001.

[31] Raúl Rabadán, Yamina Mohamedi, Udi Rubin, Tim Chu, Adam N Alghalith, Oliver Elliott, Luis Arnés, Santiago Cal, Álvaro J Obaya, Arnold J Levine, et al. Identification of relevant genetic alterations in cancer using topological data analysis. *Nature communications*, 11(1):3808, 2020.

[32] Erik J Amézquita, Michelle Y Quigley, Tim Ophelders, Elizabeth Munch, and Daniel H Chitwood. The shape of things to come: Topological data analysis and biology, from molecules to organisms. *Developmental Dynamics*, 249(7):816–833, 2020.

[33] Derek Lo and Briton Park. Modeling the spread of the zika virus using topological data analysis. *PloS one*, 13(2):e0192120, 2018.

[34] Matteo Rucco, Rocio Gonzalez-Diaz, Maria-Jose Jimenez, Nieves Atienza, Cristina Cristalli, Enrico Concettoni, Andrea Ferrante, and Emanuela Merelli. A new

topological entropy-based approach for measuring similarities among piecewise linear functions. *Signal Processing*, 134:130–138, 2017.

[35] Wikipedia contributors. Vietoris–rips filtration — Wikipedia, the free encyclopedia. `https://en.wikipedia.org/w/index.php?title=Vietoris%E2%80%93Rips_filtration&oldid=1196212548`, 2024. [Online; accessed 16-February-2024].

[36] Amish Mishra and Francis C Motta. Stability and machine learning applications of persistent homology using the delaunay-rips complex. *arXiv preprint arXiv:2303.01501*, 2023.

[37] Nieves Atienza, Rocio Gonzalez-Diaz, and Matteo Rucco. Persistent entropy for separating topological features from noise in vietoris-rips complexes. *Journal of Intelligent Information Systems*, 52:637–655, 2019.

[38] Nina Otter, Mason A Porter, Ulrike Tillmann, Peter Grindrod, and Heather A Harrington. A roadmap for the computation of persistent homology. *EPJ Data Science*, 6:1–38, 2017.

[39] Herbert Edelsbrunner and John L Harer. *Computational topology: an introduction*. American Mathematical Society, 2022.

[40] Eric W. Weisstein. Isomorphic. `https://mathworld.wolfram.com/Isomorphic.html`, 2000.

[41] Eric W. Weisstein. Nerve. `https://mathworld.wolfram.com/Nerve.html`, 2000.

[42] Eric W. Weisstein. Voronoi diagram. `https://mathworld.wolfram.com/VoronoiDiagram.html`, 2000.

[43] William L. Hosch. Isomorphism. `https://www.britannica.com/science/isomorphism-mathematics`, 2024.

[44] What is a voronoi diagram? `https://www.bristol.ac.uk/maths/fry-building/public-art-strategy/what-is-a-voronoi-diagram/`.

[45] David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. Stability of persistence diagrams. In *Proceedings of the twenty-first annual symposium on Computational geometry*, pages 263–271, 2005.

[46] Claude Elwood Shannon. A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423, 1948.

[47] John Horgan. Claude shannon: Tinkerer, prankster, and father of information theory. *IEEE Spectrum Retrieved from https://spectrum. ieee. org/tech-history/cyberspace/claude-shannon-tinkerer-prankster-and-father-of-information-theory Horgan J*, 2016.

[48] James V Stone. Information theory: a tutorial introduction. 2015.

[49] David Ellerman. Introduction to logical entropy and its relationship to shannon entropy. *arXiv preprint arXiv:2112.01966*, 2021.

[50] Esfandiar Maasoumi and Jeff Racine. Entropy and predictability of stock market returns. *Journal of Econometrics*, 107(1-2):291–312, 2002.

[51] Cheng Cao and Semyon Slobounov. Application of a novel measure of eeg non-stationarity as 'shannon-entropy of the peak frequency shifting'for detecting residual abnormalities in concussed individuals. *Clinical Neurophysiology*, 122(7):1314–1321, 2011.

[52] Matteo Ruccoa, Rocio Gonzalez-Diazb, Maria-Jose Jimenezb, Nieves Atienzab, Cristina Cristallic, Enrico Concettonic, Andrea Ferrantec, and Emanuela Merellia. A new topology-based approach for measuring similarities among discrete real noisy signals. *arXiv preprint arXiv:1512.07613*, 2015.

[53] Better Health Channel. Pregnancy - preeclampsia. *Better Health Channel*, 2024.

[54] National Institutes of Health (NIH). What are the risks of preeclampsia eclampsia to the fetus? `https://www.nichd.nih.gov/health/preeclampsia-and-eclampsia`.

[55] Centers for Disease Control and Prevention. High blood pressure during pregnancy. 2019.

[56] Centers for Disease Control and Prevention. Gestational diabetes. 2022.

[57] Frank Q Nuttall. Body mass index: obesity, bmi, and health: a critical review. *Nutrition today*, 50(3):117, 2015.

[58] Magdalena Nowak, Maria Kalwa, Piotr Oleksy, Katarzyna Marszalek, Malgorzata Radon-Pokracka, and Hubert Huras. The relationship between pre-pregnancy bmi, gestational weight gain and neonatal birth weight: a retrospective cohort study. *Ginekologia polska*, 90(1):50–54, 2019.

[59] Theresa Casey, Hui Sun, Aridany Suarez-Trujillo, Jennifer Crodian, Lingsong Zhang, Karen Plaut, Helen J Burgess, Shelley Dowden, David M Haas, and Azza Ahmed. Pregnancy rest-activity patterns are related to salivary cortisol rhythms and maternal-fetal health indicators in women from a disadvantaged population. *PLoS One*, 15(3):e0229567, 2020.

[60] Newborn measurements - health encyclopedia - university of rochester medical center. `https://www.urmc.rochester.edu/encyclopedia/content.aspx?contenttypeid=90&contentid=P02673#:~:text=The%20average%20newborn's%20head%20measures,centimeters%20(18%20x%202.54).` Accessed: 2023-10-20.

[61] What is the average baby weight by month? `https://www.medicalnewstoday.com/articles/325630`. Accessed: 2023-10-20.

[62] Saira Jamshed, Farah Khan, Sarwat Khalid Chohan, Zakia Bano, Shizra Shahnawaz, Adnan Anwar, and Atif A Hashmi. Frequency of normal birth length and its determinants: A cross-sectional study in newborns. *Cureus*, 12(9), 2020.

[63] Health - adverse birth outcomes. `https://www.epa.gov/americaschildrenenvironment/health-adverse-birth-outcomes`. Accessed:

2023-10-20.

[64] Premature birth. https://www.cdc.gov/reproductivehealth/features/premature-birth/. Accessed: 2023-10-20.

[65] Anita CJ Ravelli, Joris AM van der Post, Christianne JM de Groot, Ameen Abu-Hanna, and Martine Eskes. Does induction of labor at 41 weeks (early, mid or late) improve birth outcomes in low-risk pregnancy? a nationwide propensity score-matched study. *Acta obstetricia et gynecologica Scandinavica*, 102(5):612–625, 2023.

[66] Rawad Farhat and Mariam Rajab. Length of postnatal hospital stay in healthy newborns and re-hospitalization following early discharge. *North American journal of medical sciences*, 3(3):146, 2011.

[67] Katie Harron, Ruth Gilbert, David Cromwell, Sam Oddie, and Jan van der Meulen. Newborn length of stay and risk of readmission. *Paediatric and perinatal epidemiology*, 31(3):221–232, 2017.

[68] Niranjana M Kowlessar, HJ Jiang, and Claudia Steiner. Hospital stays for newborns, 2011. 2013.

[69] Ripser Project Developers. Greedy permutation. https://ripser.scikit-tda.org/en/latest/notebooks/Greedy%20Subsampling%20for%20Fast%20Approximate%20Computation.html.

[70] Nicholas J. Cavanna, Mahmoodreza Jahanseir, and Donald R. Sheehy. A geometric perspective on sparse filtrations, 2015.

[71] Christopher Tralie, Nathaniel Saul, and Rann Bar-On. Ripser.py: A lean persistent homology library for python. *The Journal of Open Source Software*, 3(29):925, Sep 2018.

[72] persim: Python tools for persistence data analysis. https://persim.scikit-tda.org/en/latest/reference/stubs/persim.persistent_entropy.html.

[73] Critical periods of development. https://www.ncbi.nlm.nih.gov/books/NBK582659/.

[74] Rebecca F Goldstein, Sally K Abell, Sanjeeva Ranasinha, Marie Misso, Jacqueline A Boyle, Mary Helen Black, Nan Li, Gang Hu, Francesco Corrado, Line Rode, et al. Association of gestational weight gain with maternal and infant outcomes: a systematic review and meta-analysis. *Jama*, 317(21):2207–2225, 2017.

[75] Yi Huang, Yan-Qiong Ouyang, and Sharon R Redding. Maternal prepregnancy body mass index, gestational weight gain, and cessation of breastfeeding: a systematic review and meta-analysis. *Breastfeeding Medicine*, 14(6):366–374, 2019.

APPENDIX
SUPPLEMENTAL MATERIAL

SUPPLEMENTAL MATERIAL

Table A.1 shows the Fisher Test use to determine if high entropy for the whole G22 rest activity was associated with complicated/uncomplicated pregnancy outcomes. Table A.2 - A.3 shows the Fisher Test that was use to determine if high entropy was associated with BMI in G22.

Table A.1. Fisher Test for G22

	High Entropy Value	Low Entropy Value	Total
Uncomplicated Pregnancy	7	13	20
Complicated Pregnancy	18	3	21
Total	25	16	41
P-value: 0.0013			

Table A.2. Fisher Test for G22: High BMI (BMI $>$ 25))

	High Entropy Value	Low Entropy Value	Total
Uncomplicated Pregnancy	1	7	8
Complicated Pregnancy	9	2	11
Total	10	9	19
P-value: 0.0055			

Table A.3. Fisher Test for G22: Low BMI (BMI $<$ 25)

	High Entropy Value	Low Entropy Value	Total
Uncomplicated Pregnancy	6	6	12
Complicated Pregnancy	9	1	10
Total	15	7	22
P-value: 0.0743			

Table A.4. List of the Pregnancy Complications Experienced by the Study Participants

Participant ID	Status	Newborn's Complications	Mother's Complications
w3	Healthy	None	None
w4	Healthy	None	None
w5	Healthy	None	None
w6	Unhealthy	Disease	Hypertension, Preeclampsia
w8	Unhealthy	Age Of Discharge, Birth weight, Birth Length, Gestational Weeks	None
w9	Healthy	None	None
w10	Unhealthy	Disease, Birthweight, Birth Length, Gestational Weeks	Diabetes, Hypertension, Preeclampsia
w12	Unhealthy	Disease, Age Of Discharge, Gestational Weeks	Diabetes, Hypertension
w13	Healthy	None	None
w14	Unhealthy	Disease, Gestational Weeks	Hypertension
w15	Unhealthy	Disease, Gestational Weeks	Diabetes
w16	Unhealthy	Age Of Discharge, Gestational Weeks	None
w17	Unhealthy	Age Of Discharge	None
w21	Healthy	None	None
w23	Healthy	None	None
w24	Healthy	None	None
w25	Healthy	None	None
w26	Unhealthy	Age Of Discharge, Birthweight, Birth Length, Gestational Weeks	None
w32	Unhealthy	Age Of Discharge, Gestational Weeks	None
w33	Unhealthy	Disease, Birthweight, Birth Length, Gestational Weeks	Hypertension
w36	Unhealthy	Disease, Gestational Weeks	Hypertension
w38	Unhealthy	Gestational Weeks	None
w39	Healthy	None	None
w41	Unhealthy	Disease, Age Of Discharge, Birth Length, Gestational Weeks	Diabetes, Preeclampsia
w42	Healthy	None	None
w44	Healthy	None	None
w46	Healthy	None	None
w53	Unhealthy	Gestational Weeks	None
w54	Unhealthy	Disease, Age Of Discharge, Birthweight, Birth Length, Gestational Weeks	Hypertension
w57	Healthy	None	None
w58	Healthy	None	None
w60	Unhealthy	Disease, Gestational Weeks	Hypertension
w61	Unhealthy	Age Of Discharge, Birthweight, Birth Length, Gestational Weeks	None
w66	Healthy	None	None
w71	Unhealthy	Disease, Gestational Weeks	Preeclampsia
w75	Unhealthy	Disease, Gestational Weeks	Hypertension
w76	Healthy	None	None
w79	Healthy	None	None
w81	Healthy	None	None
w82	Unhealthy	Disease, Gestational Weeks	Hypertension
w83	Healthy	None	None
w84	Healthy	None	None
w85	Healthy	None	None
w87	Unhealthy	Disease, Gestational Weeks	Hypertension, Preeclampsia
w93	Healthy	None	None

Table A.5. Participant Characteristics and BMI Status Stratified by Pregnancy Outcome of Overall Sample (n = 45), Complicated (n = 22), Noncomplicated (n = 23)

Characteristics	**Overall** ***n(%)***	**Complicated** ***n(%)***	**Noncomplicated** ***n(%)***
Race:			
Black or African American	29 (64.44)	17 (77.27)	12 (52.17)
White	8 (17.77)	1 (4.54)	7 (30.43)
Asian	1 (2.22)	0 (0)	1 (4.34)
More Than One Race	5 (11.11)	3 (13.63)	2 (8.69)
Unknown	2 (4.44)	1 (4.54)	1 (4.34)
Salary:			
Less Than $10,000	19 (42.22)	8 (36.36)	11 (47.82)
$10,000-24,999	10 (22.22)	6 (27.27)	4 (14.39)
$25,000-49,999	10 (22.22)	7 (31.27	3 (13.04)
$50,000-74,999	2 (4.44)	1 (4.54)	1 (4.34)
Greater Than $74,000	2 (4.44)	0 (0)	2 (8.69)
Unknown	2 (4.44)	0 (0)	2 (8.69)
Education Level:			
Less Than 7th Grade	1 (2.22)	0 (0)	1 (4.34)
9th Grade	1 (2.22)	0 (0)	1 (4.34)
High School Graduate or GED	29 (64.44)	14 (63.63)	15 (65.21)
Bachelor's Degree	5 (11.11)	5 (22.72)	0 (0)
Associate's Degree	3 (6.66)	2 (9.09)	1 (4.34)
Graduate Degree	6 (13.33)	1 (4.54)	5 (21.73)
BMI:	Mean ± SD	Mean ± SD	Mean ± SD
	26.75 ± 7.67	28.55 ± 8.92	25.03 ± 5.95
Pregnancy Outcome			
Complicated	22 (48.88)		
Noncomplicated	23 (51.11)		