McMaster University

Integrated Biomedical Engineering & Health Sciences Program

IBEHS - 4QZ3
Modelling of Biological Systems

Lecture 7

TAYLOR DEVET MASC.

PHD. CANDIDATE BIOLOGICAL AND BIOMEDICAL ENGINEERING
MCGILL UNIVERSITY

SHRINERS HOSPITAL FOR CHILDREN

Todays Aims…

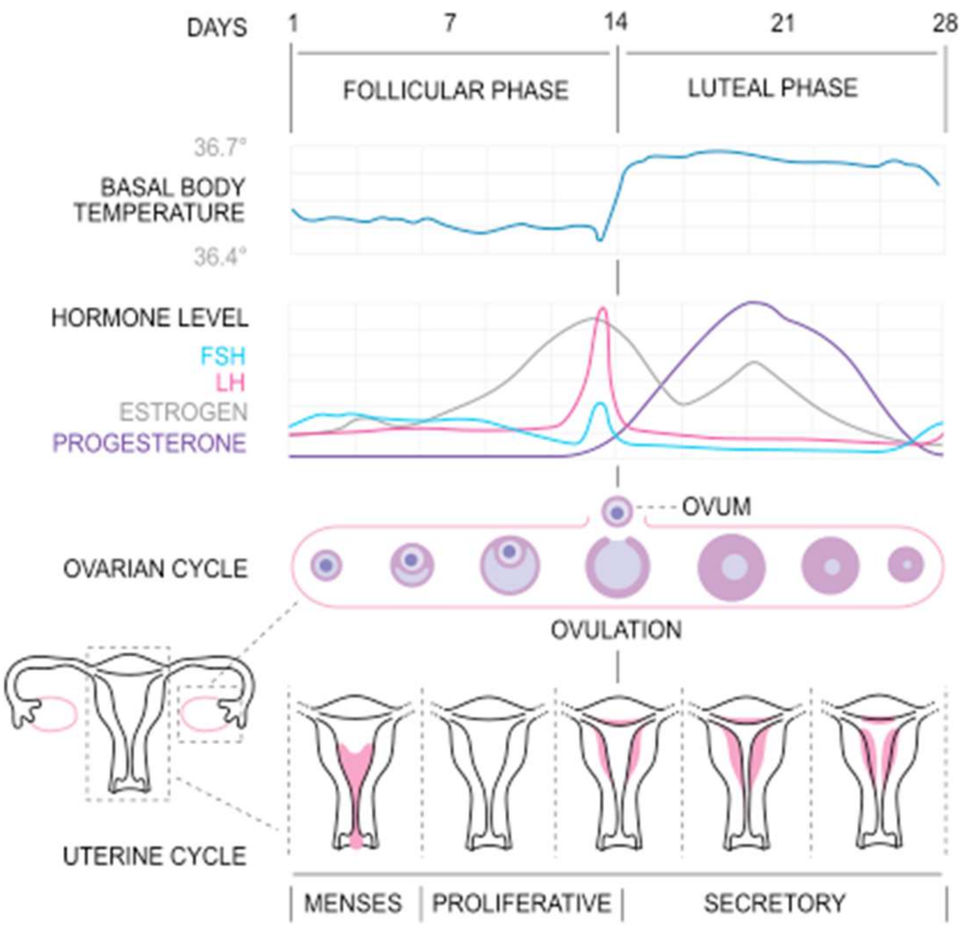
Chronorbiology

Linear Time Invariance

Correlation and Coherence

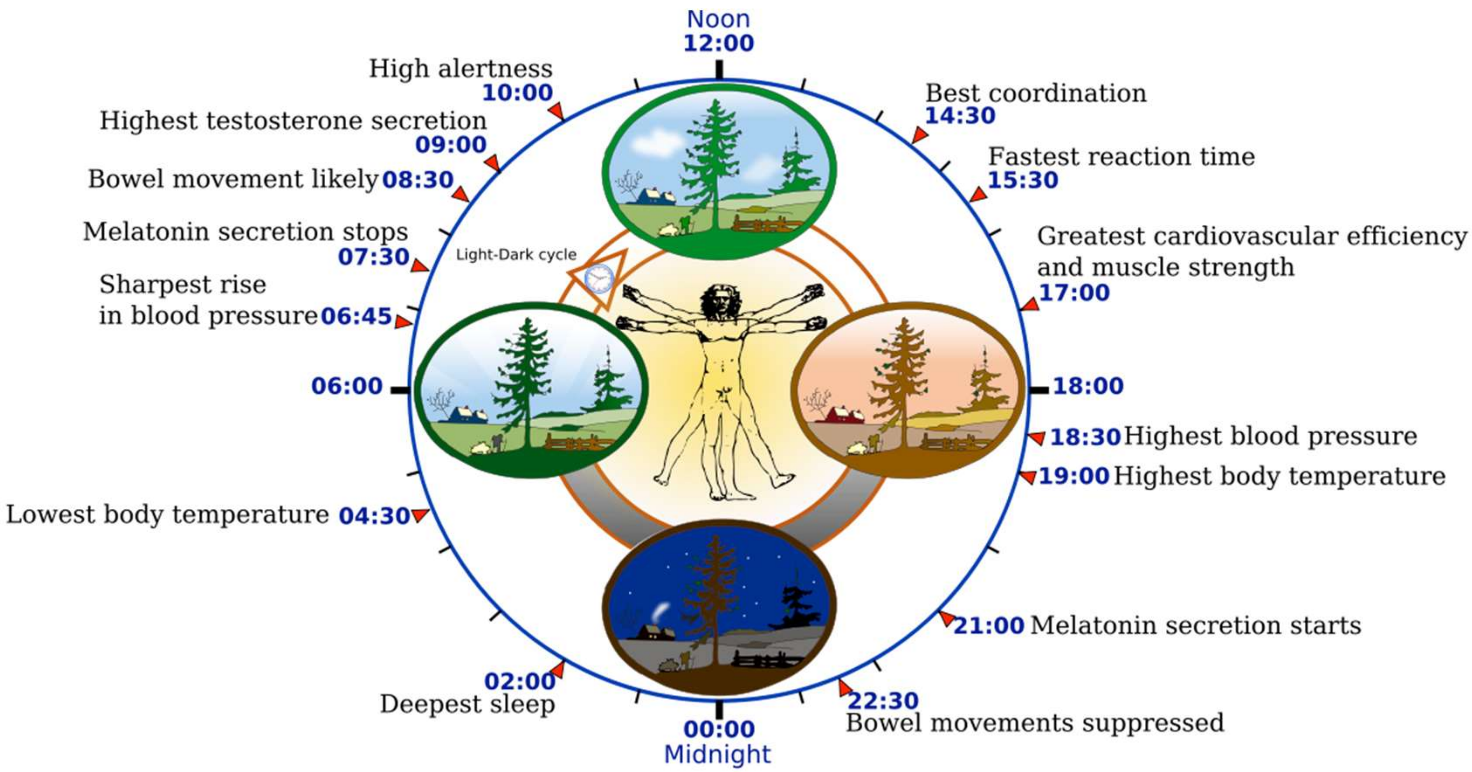
Measuring Physiological Signals in the Time Domain

- linear/multiple regression
- multivariate statistics (PCA/ICA)
- cosinor analysis
- autocorrelation analysis
- Fourier
- Wavelets
- Fractal and Chaos analysis
- etc.



Physiological Signals are Cyclical

Physiological Signals are Cyclical



Chronobiology

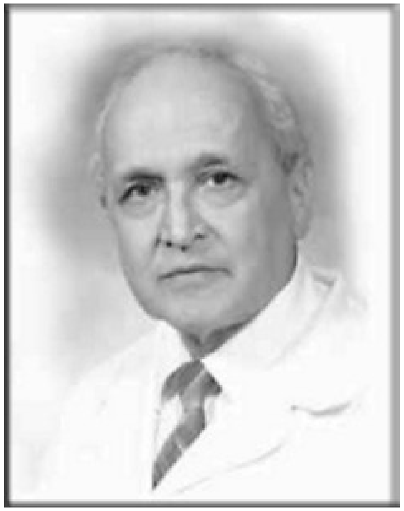
Biology of periodic(cyclic) and other time dependent processes in living organisms

biological rhythms

Franz Halberg

In the 1950s, he introduced the word circadian

Father of Chronobiology



Circadian rhythms

Circa – around

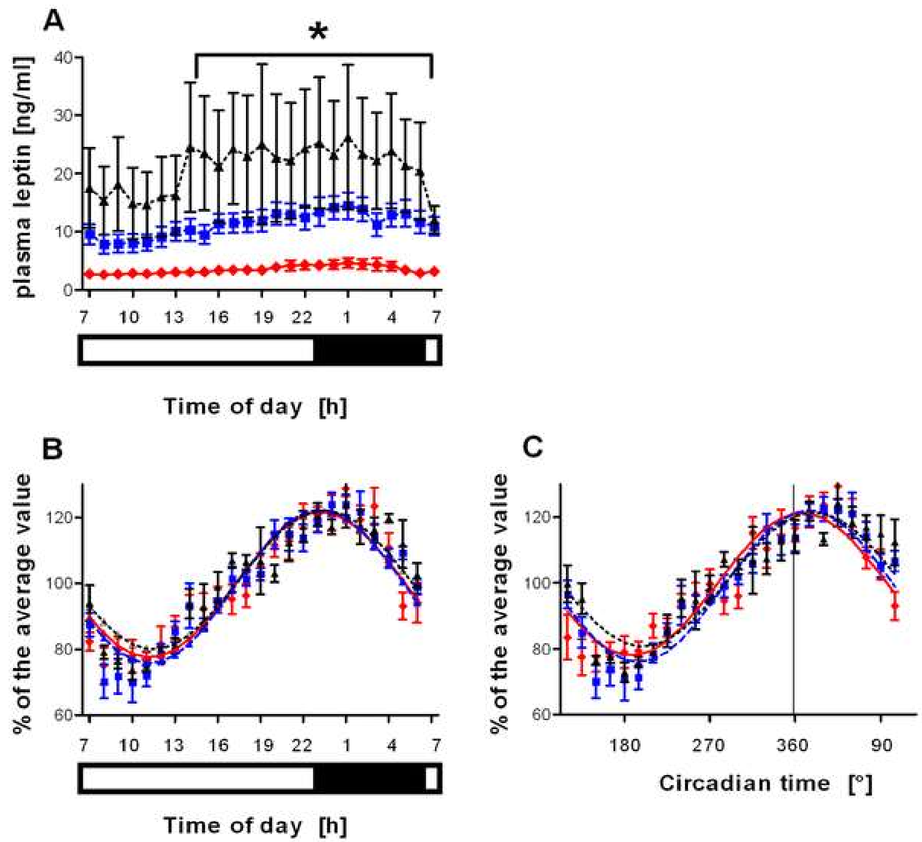
Diem - day

Biological rhythms with durations ~24 hours. Many behavioral and autonomic processes of organisms exhibit circadian rhythmicity

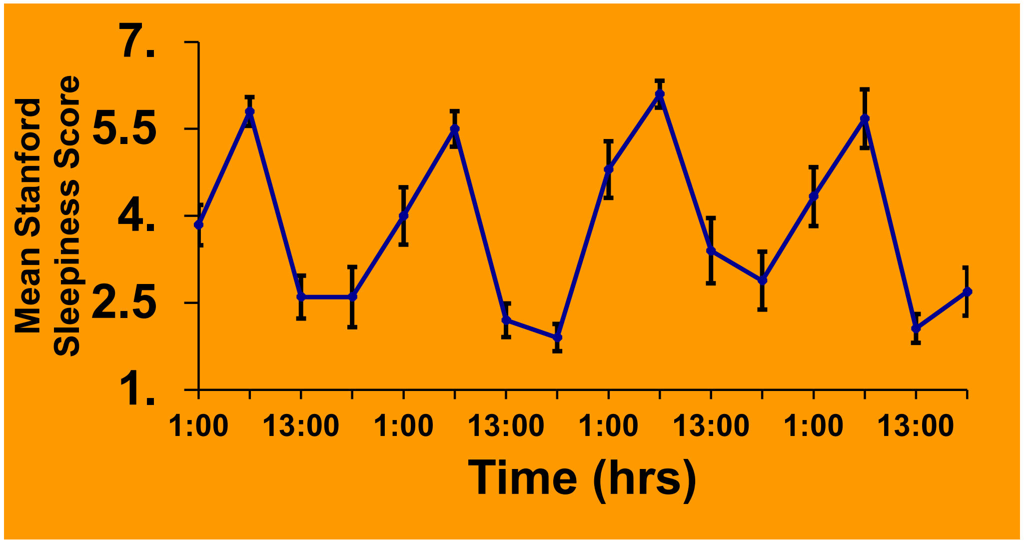
Diurnal rhythm

- a cycle that is synchronized/correlated with the day/night cycle
- Diurnal rhythms are circadian but not all circadian are diurnal

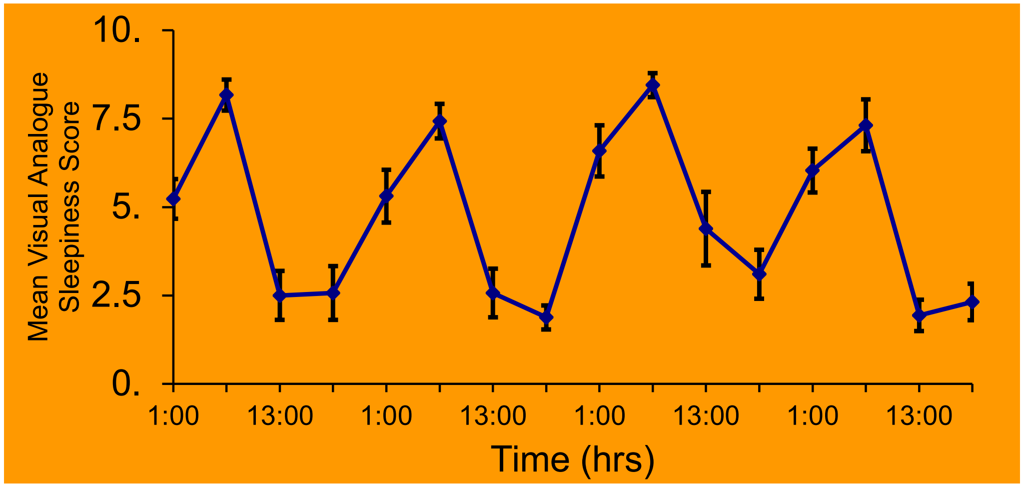
Diurnal rhythms of plasma leptin concentrations



Mäntele S, Otway DT, Middleton B, Bretschneider S, et al. (2012) Daily Rhythms of Plasma Melatonin, but Not Plasma Leptin or Leptin mRNA, Vary between Lean, Obese and Type 2 Diabetic Men. PLoS ONE 7(5): e37123. doi:10.1371/journal.pone.0037123
http://www.plosone.org/article/info:doi/10.1371/journal.pone.0037123

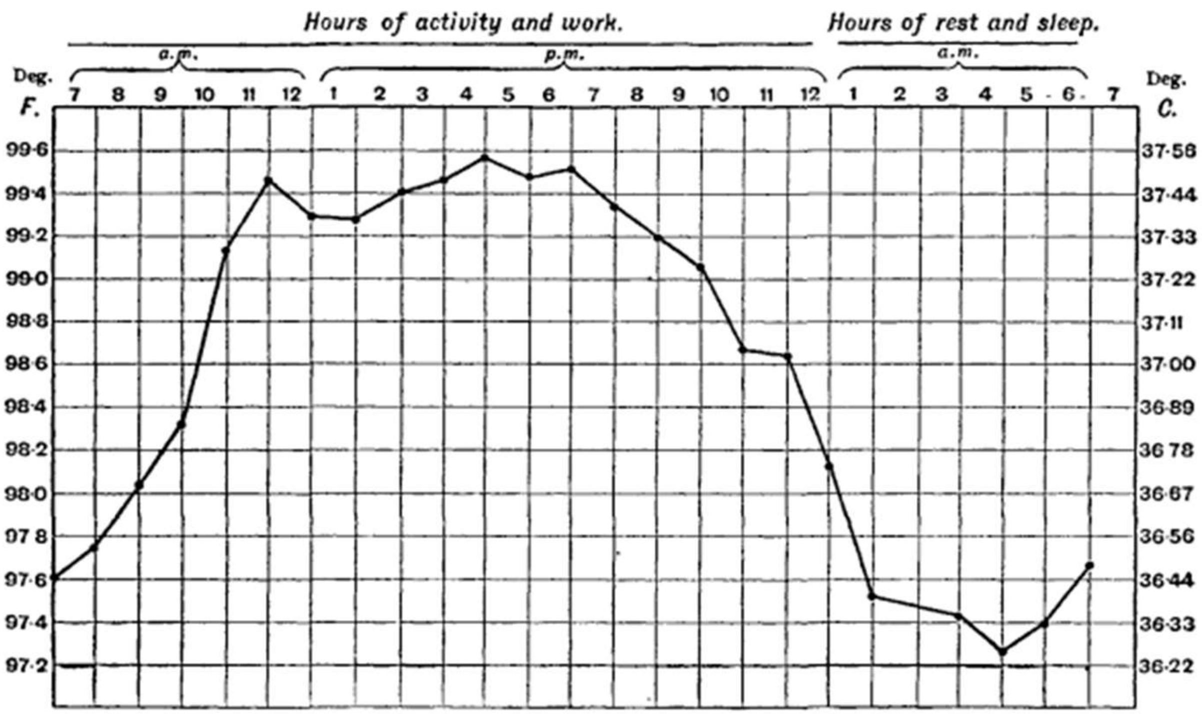


7.
5.5
4.
2.5
1.
Mean Stanford
Sleepiness Score
1:00
13:00
1:00
13:00
1:00
13:00
1:00
13:00
Time (hrs)



10.
7.5
5.
2.5
0.
Mean Visual Analogue
Sleepiness Score
1:00
13:00
1:00
13:00
1:00
13:00
1:00
13:00
Time (hrs)

Basal body Temperature: Daily Rhythm



Ultradian rhythms

Biological rhythms with durations shorter than circadian rhythms (i.e., < ~19 hours)

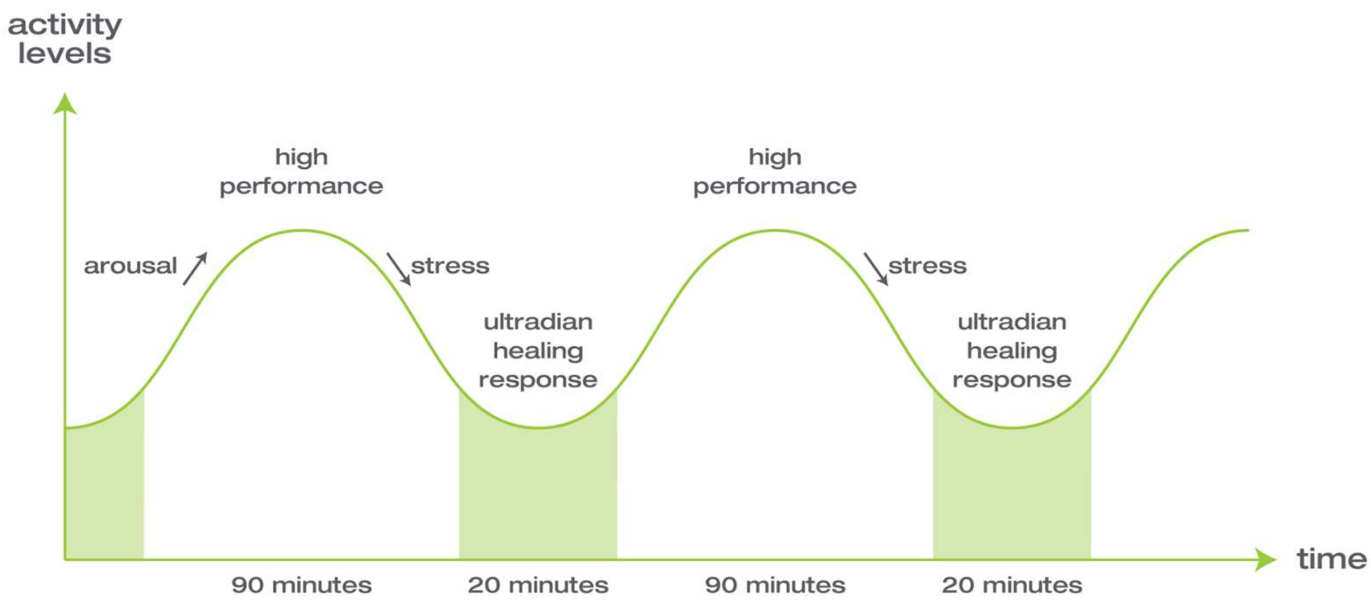
Rhythms longer than an hour

e.g. cardiac, respiratory, neuroendocrine, gastrointestinal, tidal, and other rhythms.

Many are endogenously generated by some sort of pacemaker

only tidal rhythms are regularly synchronized to environmental cycles.

General Human Performance



https://www.kosmotime.com/ultradian-rhythm/

Infradian rhythms

Biological rhythms with durations greater circadian rhythms (i.e. > ~28 hours).

e.g. estrous, weekly, lunar, annual

Many infradian rhythms are endogenously generated by some sort of pacemaker

only lunar and annual rhythms can be fully synchronized to environmental cycles.

Infradian rhythms

THE EFFECTS OF THE FULL MOON ON HUMAN BEHAVIOR*

Edgecliff College

JODI TASSO AND ELIZABETH MILLER[1]

SUMMARY

Data were gathered in a large metropolitan area over a period of one year as to nine categories of 34,318 criminal offenses committed during the phases of the full moon and non full moon. It was found that the eight categories of rape, robbery and assault, burglary, larceny and theft, auto theft, offenses against family and children, drunkenness, and disorderly conduct occurred significantly more frequently during the full moon phase than at other times of the year. Only the category of homicide did not occur more frequently during the full moon phase. The results support further exploration and research related to cosmic influences on man's behavior.

Infradian rhythms

ELSEVIER

The American Journal of Emergency Medicine
Volume 14, Issue 2, March 1996, Pages 161-164

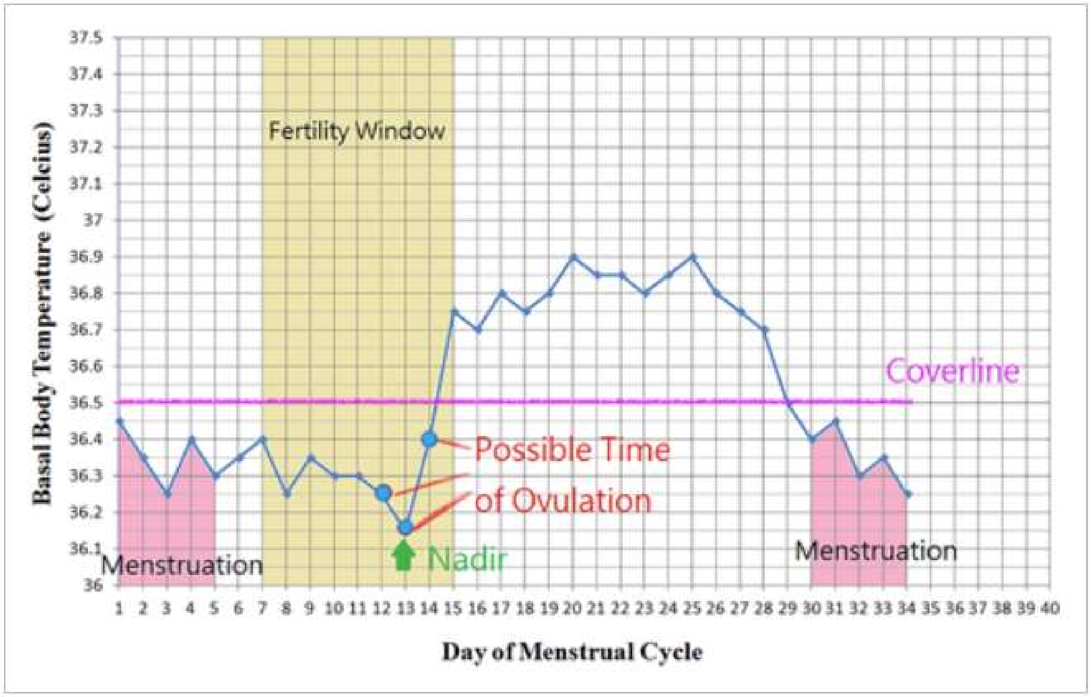
Original contribution

The full moon and ED patient volumes: Unearthing a myth

David A Thompson MD *, Stephen L Ad

ambulance. A total of 35,087 patients was admitted to the hospital and 11,278 patients were admitted to a monitored unit. No significant differences were found in total patient visits, ambulance runs, admissions to the hospital, or admissions to a monitored unit on days of the full moon. The occurrence of a full moon has no effect on ED patient volume, ambulance runs, admissions, or admissions to a monitored unit.

Infradian Rhythms



https://aiche.onlinelibrary.wiley.com/doi/full/10.1002/btm2.10058

Annual rhythms

A class of infradian rhythms that fluctuate on a yearly basis

e.g. body mass, cold-induced thermogenesis, food intake, heterothermy, melatonin secretion, molting, and reproductive capacity.

Many, but not all, are synchronized to annual environmental cycles (called circannual rhythms).

Migration of birds

Spring babies

Other Rhythms

- other cycles having periods of ~ 1week, ~1 month, and ~1 year are also ubiquitous, as are some other newly discovered cycles with periods of about 5 and 16 months, and much longer periods.

- biological cycles are typically synchronized by environmental cycles (e.g. lighting and feeding schedules).

- More generally, environmental geophysical cycles such as the day-light cycle, tidal motions, moon phases, seasonal variation (circanual) are the dominant rhythms

Chronobiology Study Designs:

Longitudinal sampling

- obtaining data on the same individual (experimental unit) as a function of time
- (e.g. 24hr blood pressure monitoring of blood pressure every minute intervals for 7 days).

Transverse (cross-sectional) sampling

- obtaining only one value per individual (experimental unit),
- with different individuals providing data at the same or different sampling times.

Hybrid (linked cross-sectional) sampling

- taking serial measurements from several individuals (experimental units).
- e.g. circulating prolactin determined at 20-minute

intervals for 24 hours in women at low or high familial risk of developing breast cancer later in life.

Chronobiology Study Design

- when sampling is performed on more than a single individual, it is important that they are synchronized

- Synchronizers are often environmental periodicities determining the temporal placement of biological rhythms.

- e.g. rest-activity or light–dark schedules (photoperiod) can be used to determine a reference time

Cosinor Analysis

- used in chronobiology

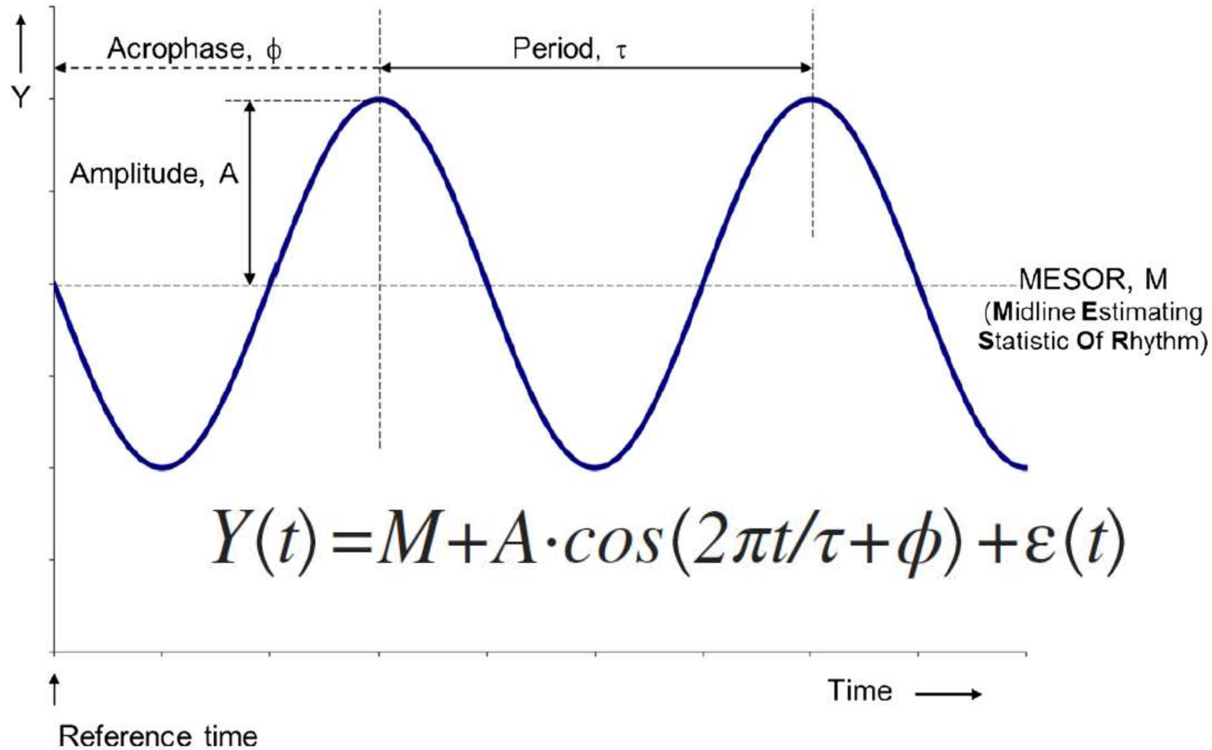
- a type of regression problem (i.e. least squares minimization)

- the biggest advantage over other time based analytical methods is this is applicable to non-equidistant data

- Assessment of non-random variations as a function of time at the cellular level, in tissue culture, as well as in multi-cellular organisms at different levels of physiologic organization

- biological time structure covers many different ranges from fractions of seconds in single neurons to seconds in the cardiac and respiratory cycles, and a few hours in certain endocrine functions.

Cosinor Analysis



Important Components

M = MESOR (Midline Statistic Of Rhythm), a rhythm-adjusted mean

A = amplitude (a measure of half the extent of predictable variation within a cycle)

φ = acrophase (a measure of the time of overall high values recurring in each cycle)

τ = cycle period

ε(t) = error term

$$Y(t) = M + A \cdot cos(2\pi t/\tau + \phi) + \varepsilon(t)$$

Cosinor Forms

$$Y(t) = M + A \cdot cos(2\pi t/\tau + \phi) + \varepsilon(t)$$

- can be re-written as:

$$Y(t) = M + \beta x + \gamma z + \varepsilon(t)$$

- where:

$$\beta = A \cdot cos\phi \qquad x = cos\left(\frac{2\pi t}{\tau}\right) \qquad z = sin\left(\frac{2\pi t}{\tau}\right)$$

$$\gamma = -A sin\phi$$

Least Squares Model

This is a least squares minimization problem.

So, minimize the residual sum of squares (RSS):

i.e. minimize the squared differences between measures (Yi) taken at times ti

- Estimates for M, **β**, and **γ** are obtained by solving the normal equations, obtained by expressing that RSS is minimal when its first-order derivatives with respect to each parameter are zero.

In Matrix Form

Therefore the normal equations are:

$$\sum Y_i = MN + \beta\sum x_i + \gamma\sum z_i$$
$$\sum Y_i x_i = M\sum x_i + \beta\sum x_i^{\,2} + \gamma\sum x_i z_i$$
$$\sum Y_i z_i = M\sum z_i + \beta\sum x_i z_i + \gamma\sum z_i^{\,2}$$

These normal equations in matrix form:

$$\begin{pmatrix} \sum Y_i \\ \sum Y_i x_i \\ \sum Y_i z_i \end{pmatrix} = \begin{pmatrix} N & \sum x_i & \sum z_i \\ \sum x_i & \sum x_i^{\,2} & \sum x_i z_i \\ \sum z_i & \sum x_i z_i & \sum z_i^{\,2} \end{pmatrix} \begin{pmatrix} M \\ \beta \\ \gamma \end{pmatrix}$$

In Matrix Form

$$\underbrace{\begin{pmatrix} \sum Y_i \\ \sum Y_i x_i \\ \sum Y_i z_i \end{pmatrix}}_{W} = \underbrace{\begin{pmatrix} N & \sum x_i & \sum z_i \\ \sum x_i & \sum x_i^{\,2} & \sum x_i z_i \\ \sum z_i & \sum x_i z_i & \sum z_i^{\,2} \end{pmatrix}}_{X} \underbrace{\begin{pmatrix} M \\ \beta \\ \gamma \end{pmatrix}}_{\hat{b}}$$

i.e. $W = X\hat{b}$

$$\hat{b} = X^{-1} W$$

(K is # parameters to estimate)

Least Squares Model

- For rhythm detection, the total sum of squares (TSS) is partitioned into the sum of squares due to the regression model (MSS) and the residual sum of squares (RSS).

TSS = sum of squared differences between the data and the arithmetic mean.

MSS = sum of squared differences between the estimated values based on the fitted model and the arithmetic mean.

$$\mathrm{TSS} = \mathrm{MSS} + \mathrm{RSS} \text{ or } \sum\left(Y_i\text{-}\bar{Y}\right)^2 = \sum\left(\hat{Y}_i\text{-}\bar{Y}\right)^2 + \sum\left(Y_i\text{-}\hat{Y}_i\right)^2$$

Calculate F-statistic to assess whether the model is appropriate

k = # of parameters estimated (3)
n = number of samples

Least Squares Model

For parameter estimation, consider M and (β,γ) separately:

Calculate the confidence interval (based on 1-α=95%)

$$\hat{M} \pm t_{\alpha/2,(N-k)} \hat{\sigma} \sqrt{X_{11}^{-1}}$$

where X_{ij}^{-1} are elements of X^{-1} and:

$$\hat{\sigma} = \sqrt{RSS/(N-k)}$$

Least Squares Model

The covariance matrix (β,γ) is given by:

$$\hat{\sigma}\begin{pmatrix} X_{22}^{-1} & X_{23}^{-1} \\ X_{32}^{-1} & X_{33}^{-1} \end{pmatrix}$$

$$\sum(x_i - \bar{x})^2\left(\beta - \hat{\beta}\right)^2 + 2\sum(x_i - \bar{x})(z_i - \bar{z})\left(\beta - \hat{\beta}\right)(\gamma - \hat{\gamma})$$

$$+ \sum(z_i - \bar{z})^2(\gamma - \hat{\gamma})^2 \leq 2\hat{\sigma}^2 F_{1-\alpha,k-1,n-k}$$

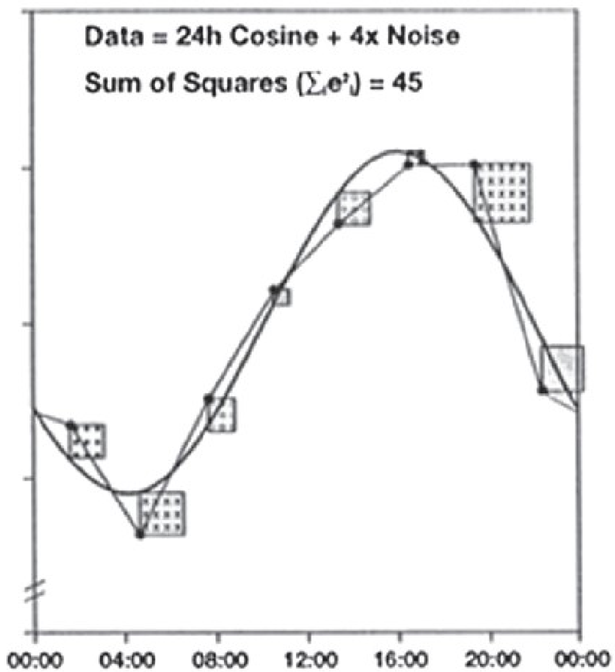
where

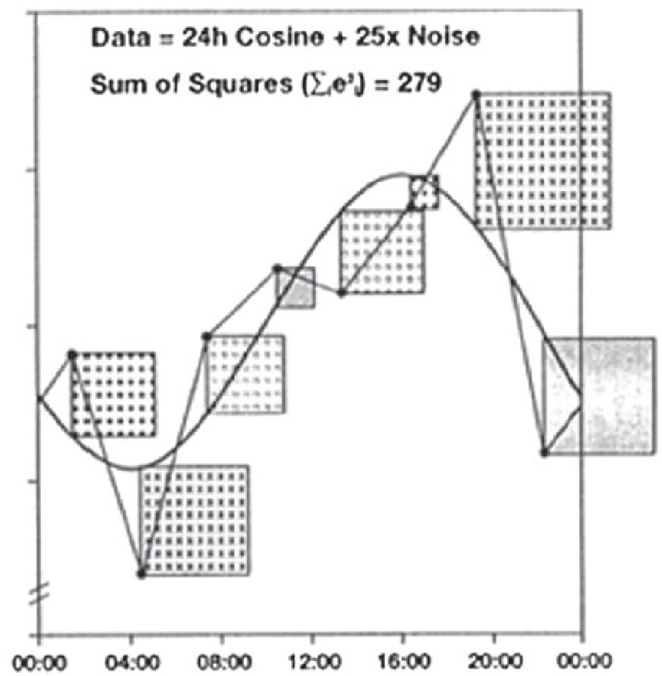
$$\bar{X} = \left(\sum_i X_i\right)/N$$

and

$$\bar{Z} = \left(\sum_i Z_i\right)/N$$

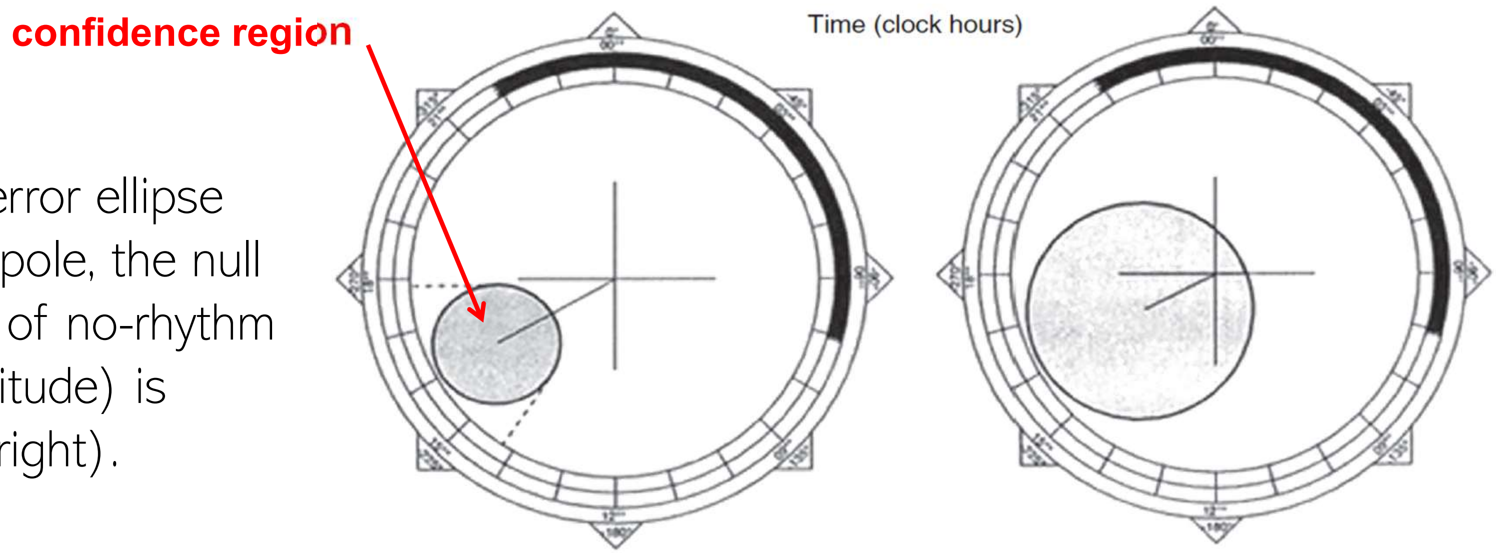
Least Squares Model





Cosinor Representation

When the error ellipse covers the pole, the null hypothesis of no-rhythm (zero amplitude) is accepted (right).



Mesor = 120.00 ± 1.06
Amplitude = 10.98 (5.88 16.09)
Acrophase = -240˚ (-213 -268)

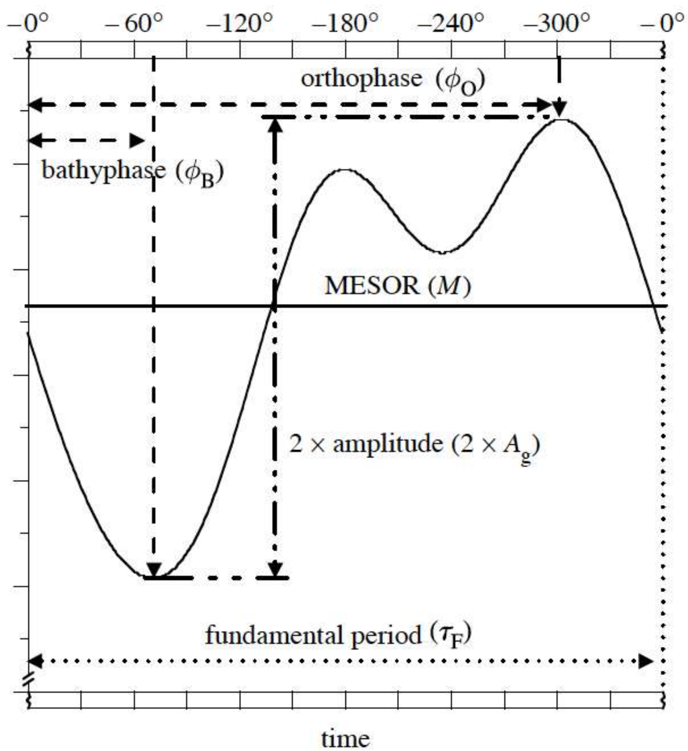
Mesor = 120.00 ± 2.64
Amplitude = 9.47 ( )
Acrophase = -241˚ ( )

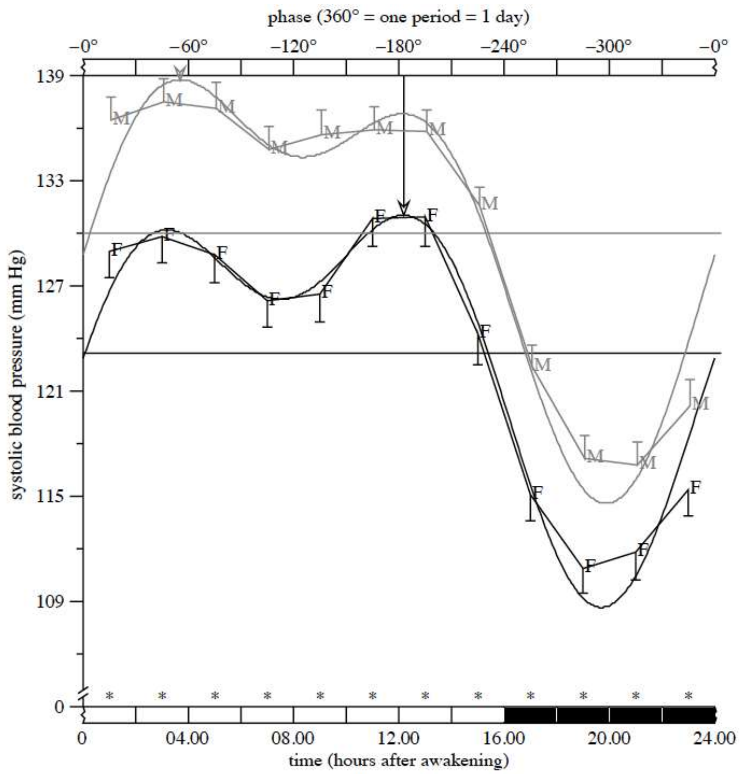
Multiple Rhythms

$$y_n = M + \sum_{c=1}^{C} A_c \cos(\omega_c t_n + \phi_c) + e_n; \quad n = 1, \ldots, N,$$

- where yn is the observed value at time tn (not necessarily equidistant) of the studied variable
- C is the number of sinusoidal components
- ωc are the angular frequencies, i.e. ωc=2π/τc, where τc, with c=1,...,C, are the fitted periods
- N is the number of observed values (sample size)

Multiple Rhythms





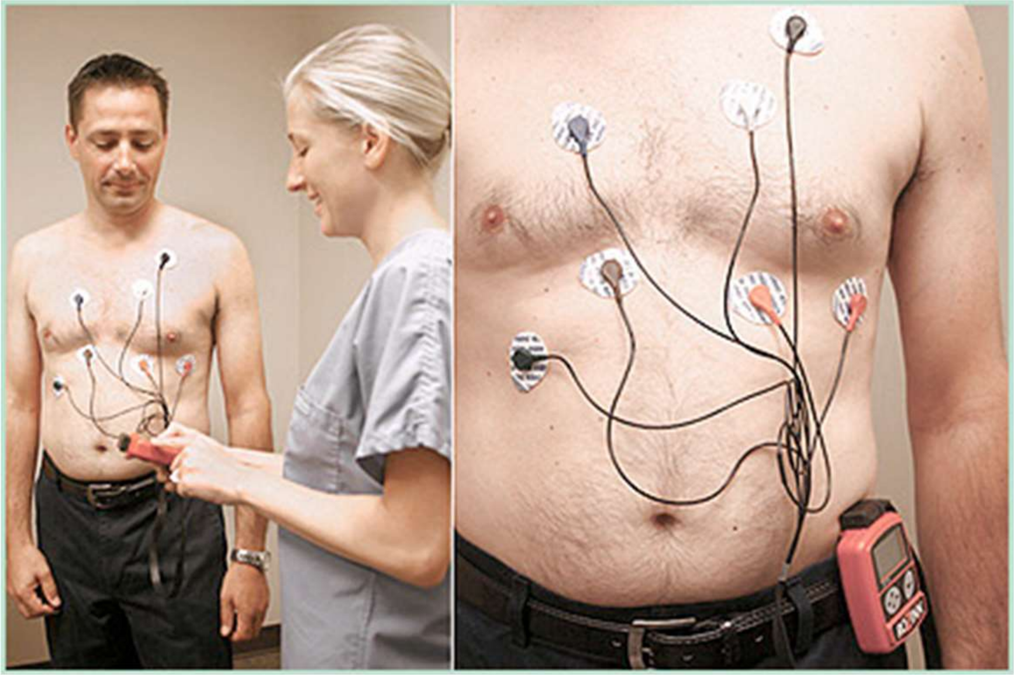
Other Sinusoidally Varying Biological Data:

Heart Rate

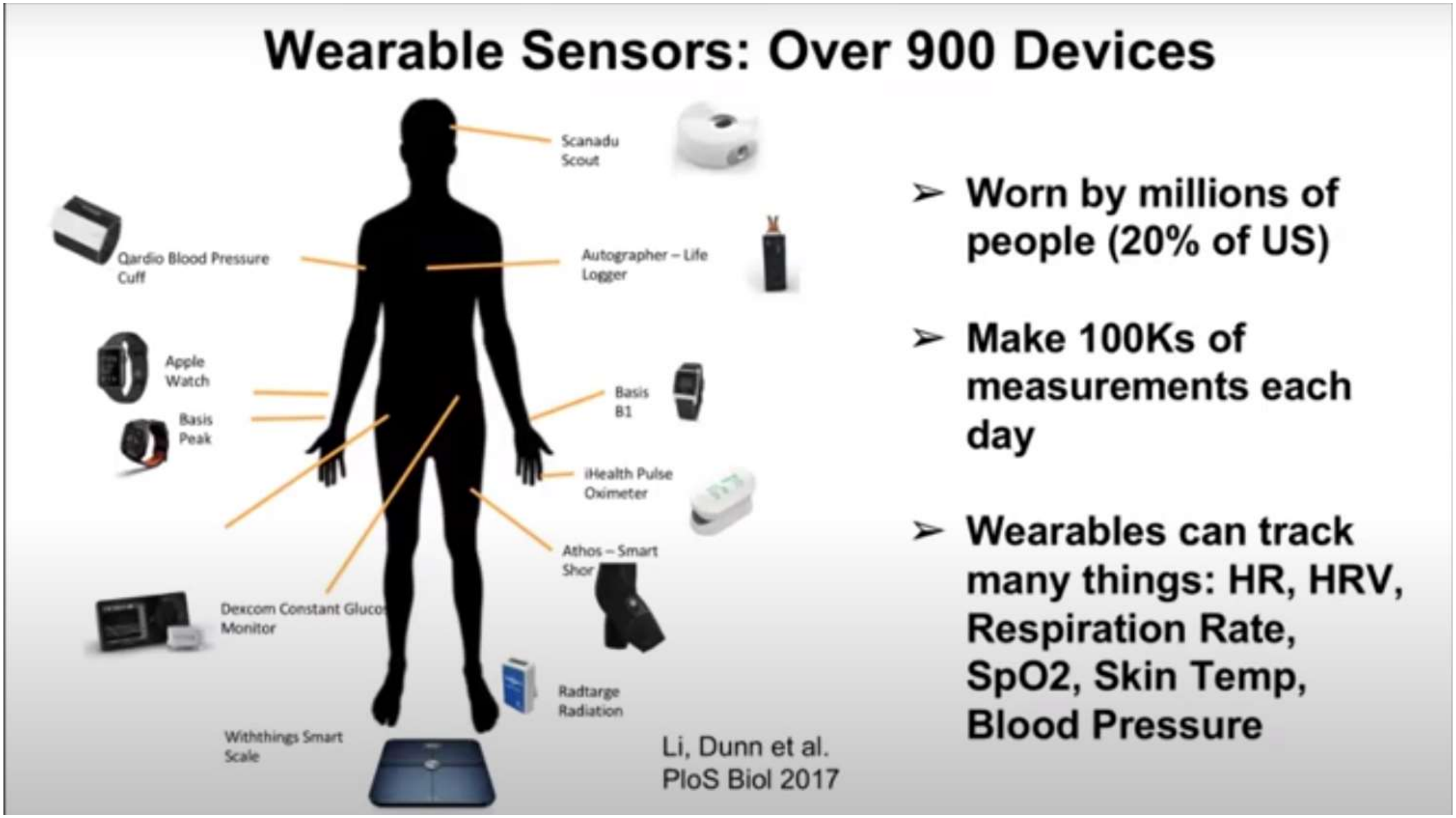
- Heart Rate Variability (HRV)

- Heart Rate Variability (HRV) Lab Analyzes Data from Continuous Electronically-Stored ECGs

Holter device

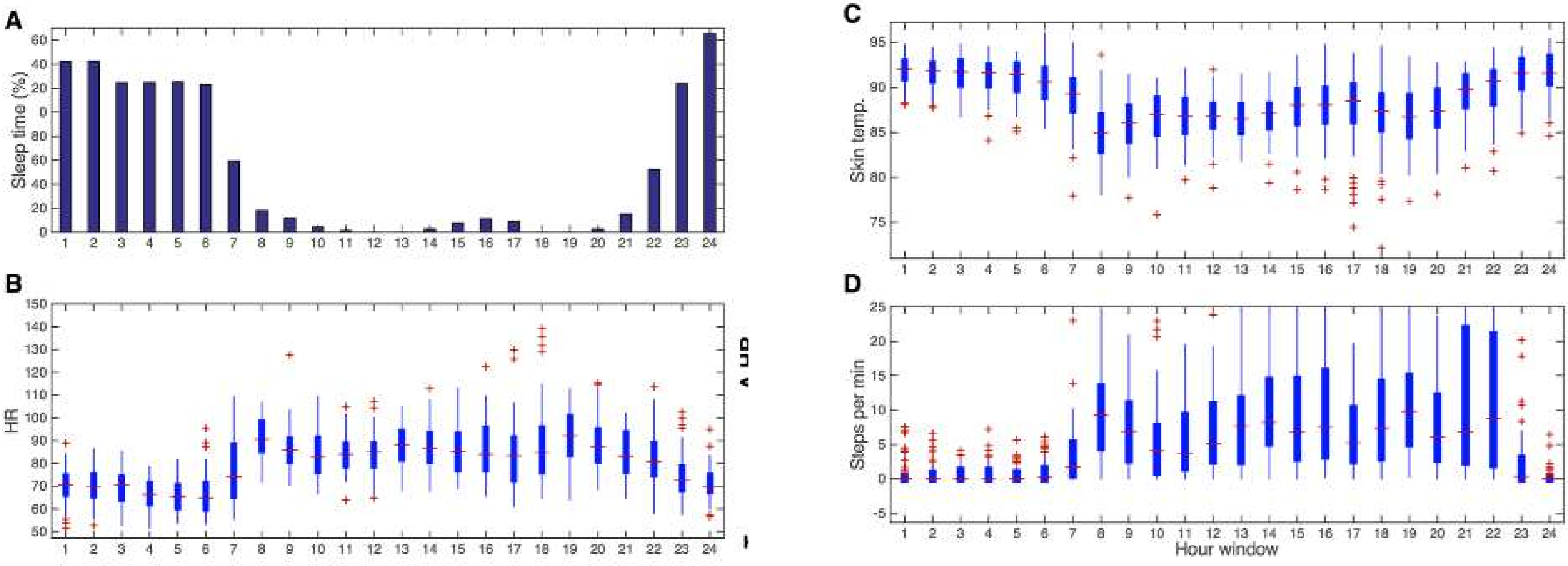



POLAR
Wed 10/11
10:58
A300
Feed
Explore
Diary
Progress
Joanna Hamilton
RUNNING
RELIVE
PRIVATE
00:55:00
10.50
149
MAX 163 / MIN 70
848
TEMPO TRAINING
MORE
HEART RATE ZONES
00:01:33
00:13:12
00:29:15
00:02:14
00:00:00
VERY GOOD
RUNNING INDEX
44
TRAINING BENEFIT
STEADY STATE & TEMPO TRAINING
HEART RATE



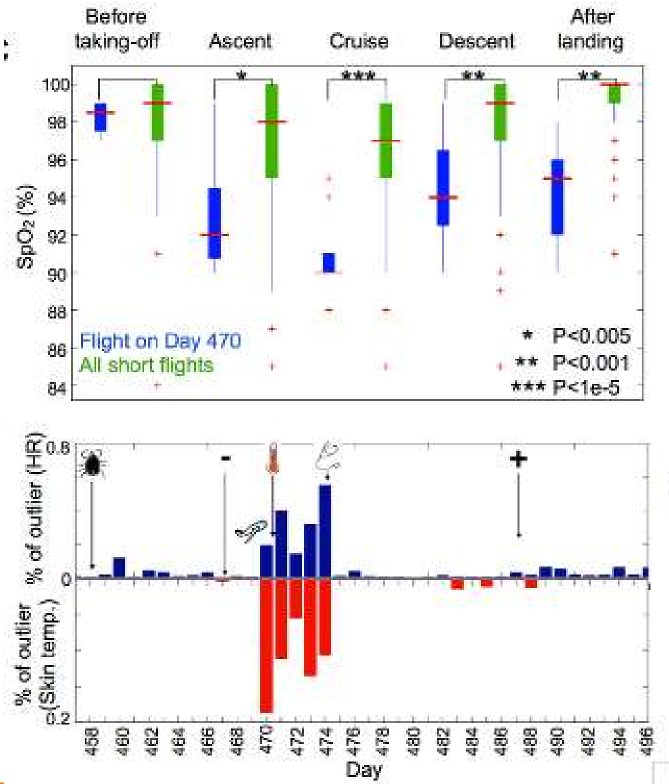
Slides: Dr Michael Snyder MD Stanford

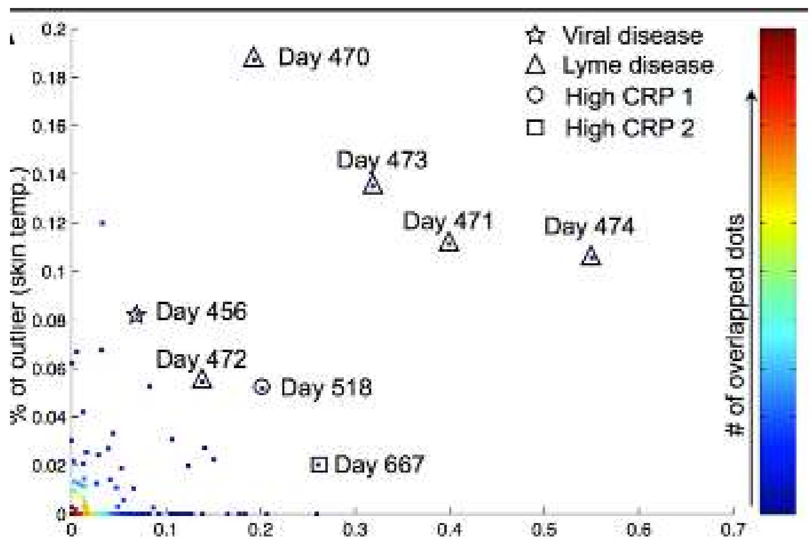
Circadian and Diurnal Rhythms



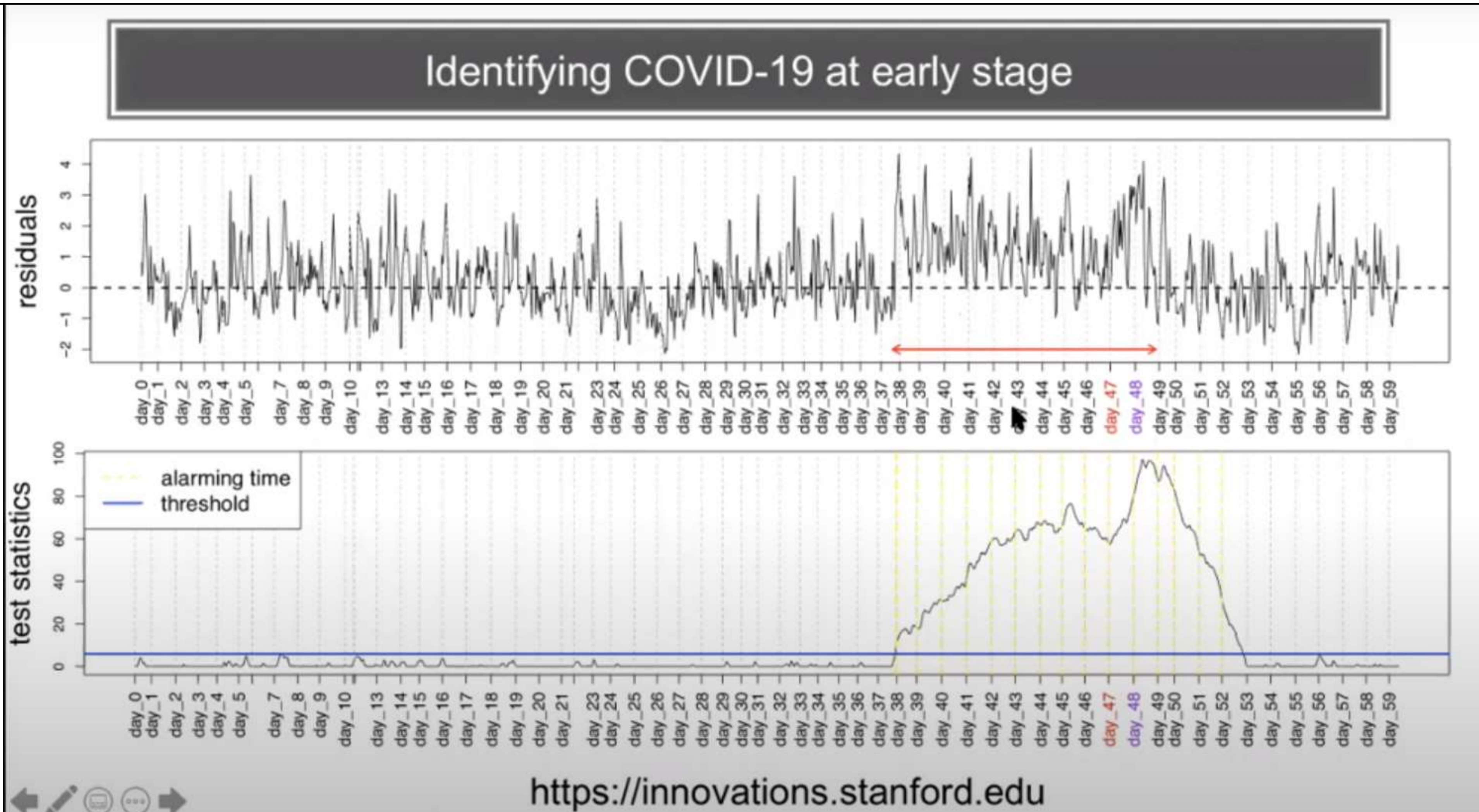
Slides: Dr Michael Snyder MD Stanford

Detection of Lyme Disease

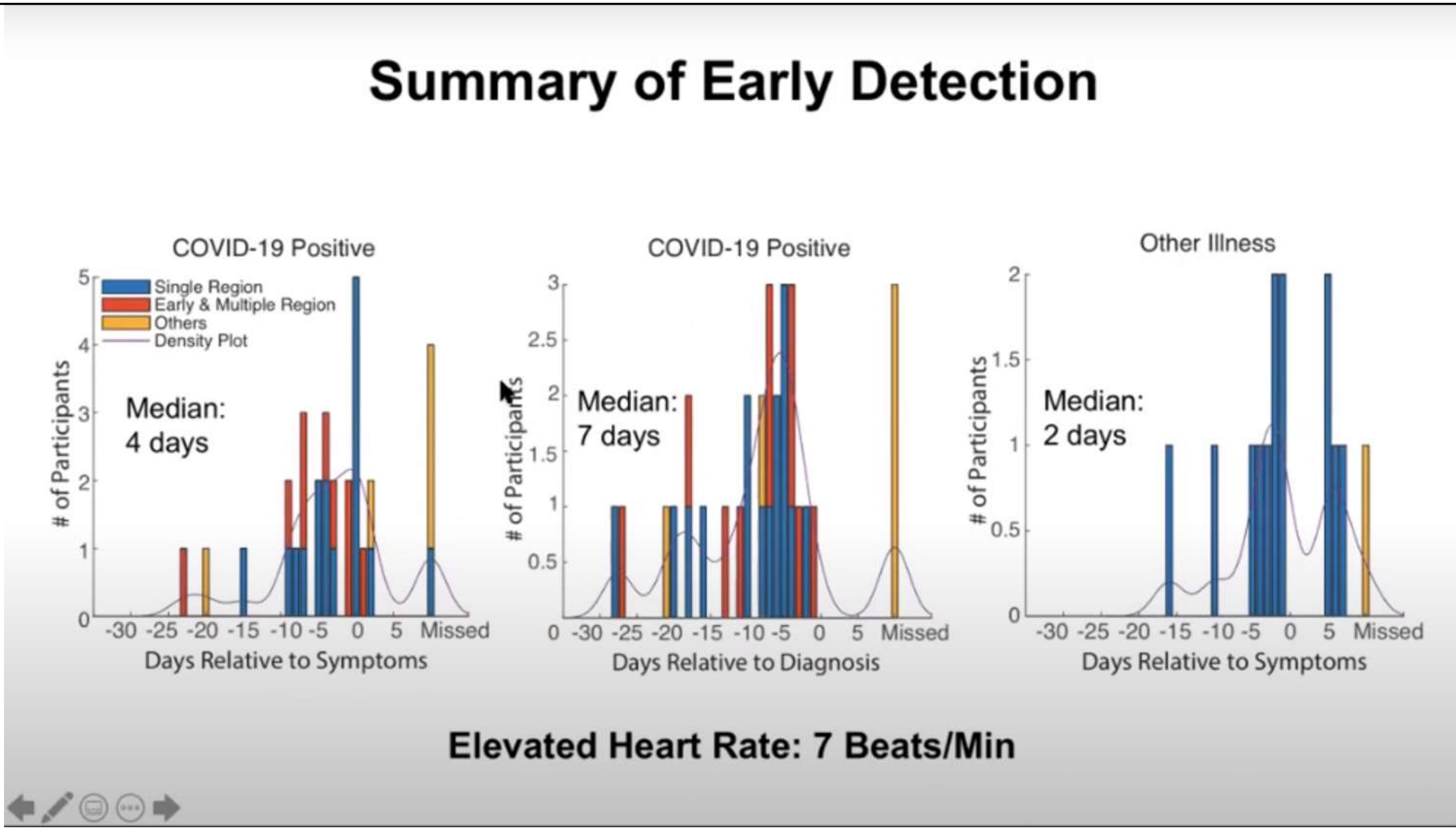




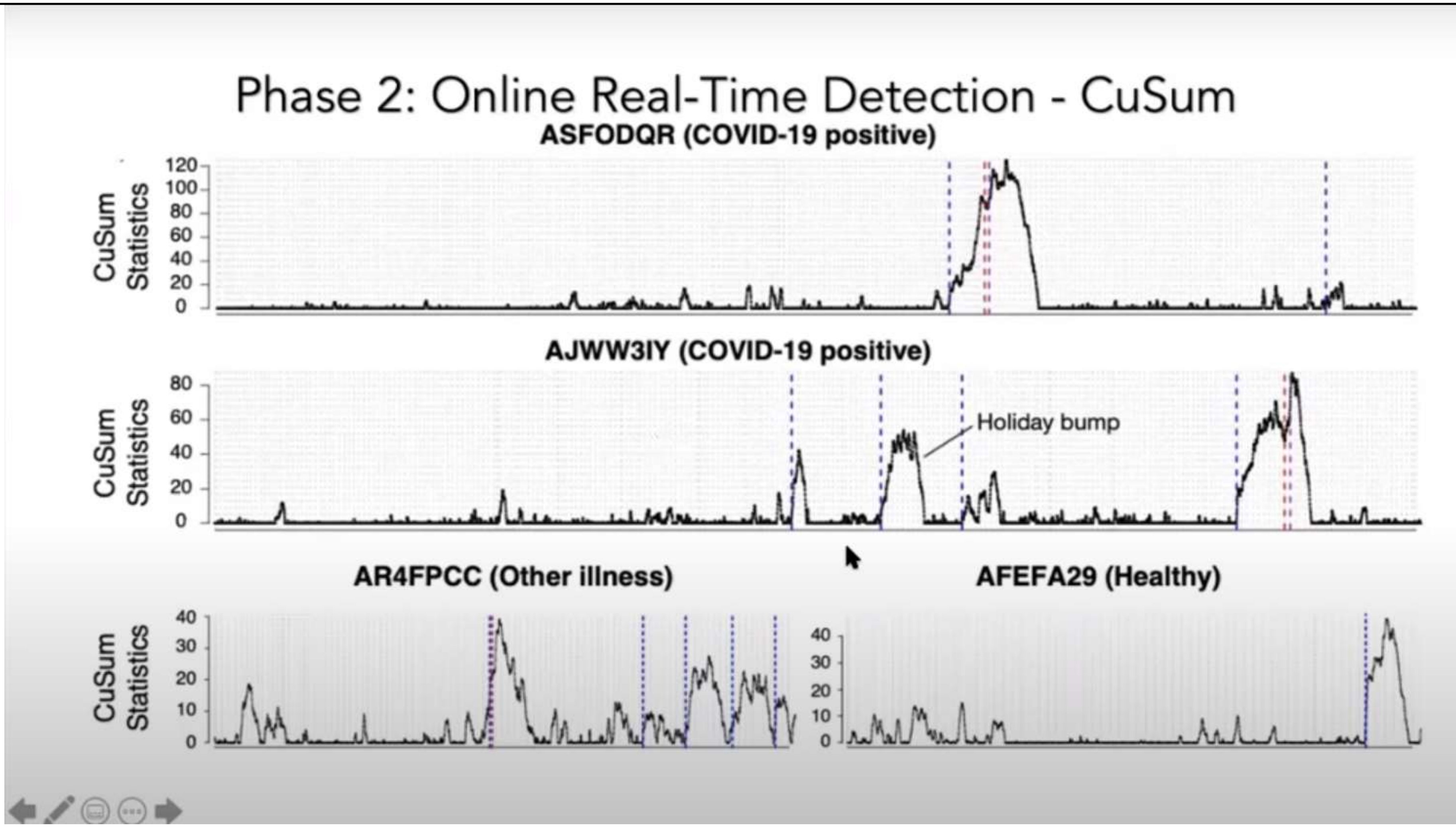
Slides: Dr Michael Snyder MD Stanford



Slides: Dr Michael Snyder MD Stanford



Slides: Dr Michael Snyder MD Stanford



Slides: Dr Michael Snyder MD Stanford

Cohort and Data

- **1015 Fitbit participants**
- **950 Apple Watch participants**

Covid Positives

Apple Watch	33
Fitbit	34

Prospective	28
Retrospective	39

Presymptomatic/Asymptomatic detection rate via NightSignal algorithm

(wrt [-14,+21] relative to symptom onset or test date)

Total = 49/67 = 73.1%

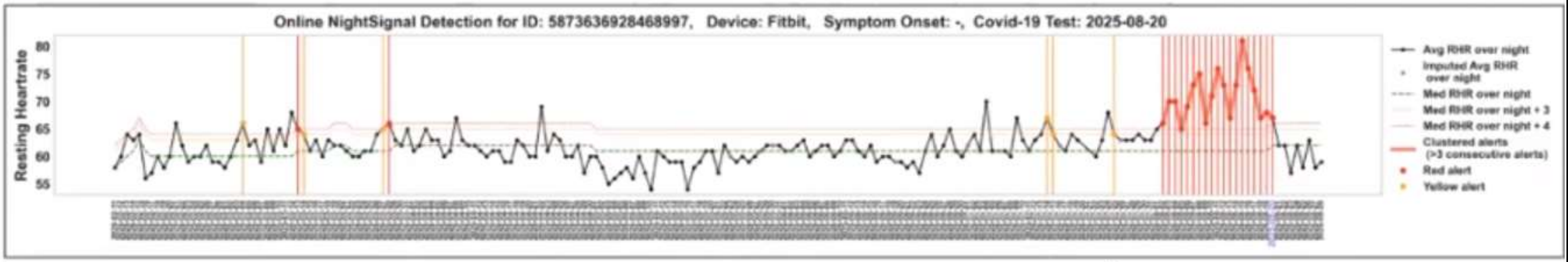
Prospectives = 24/28

Slides: Dr Michael Snyder MD Stanford

Asymptomatic Detection Examples

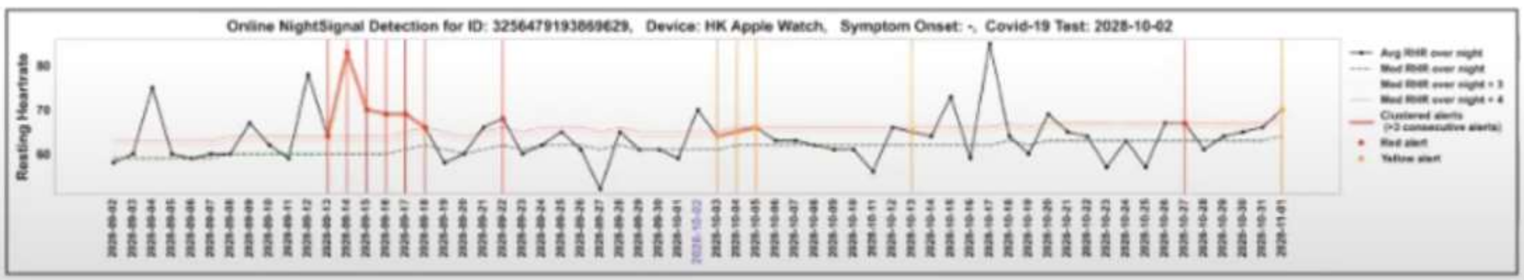
FitBit

NightSignal



Apple Watch

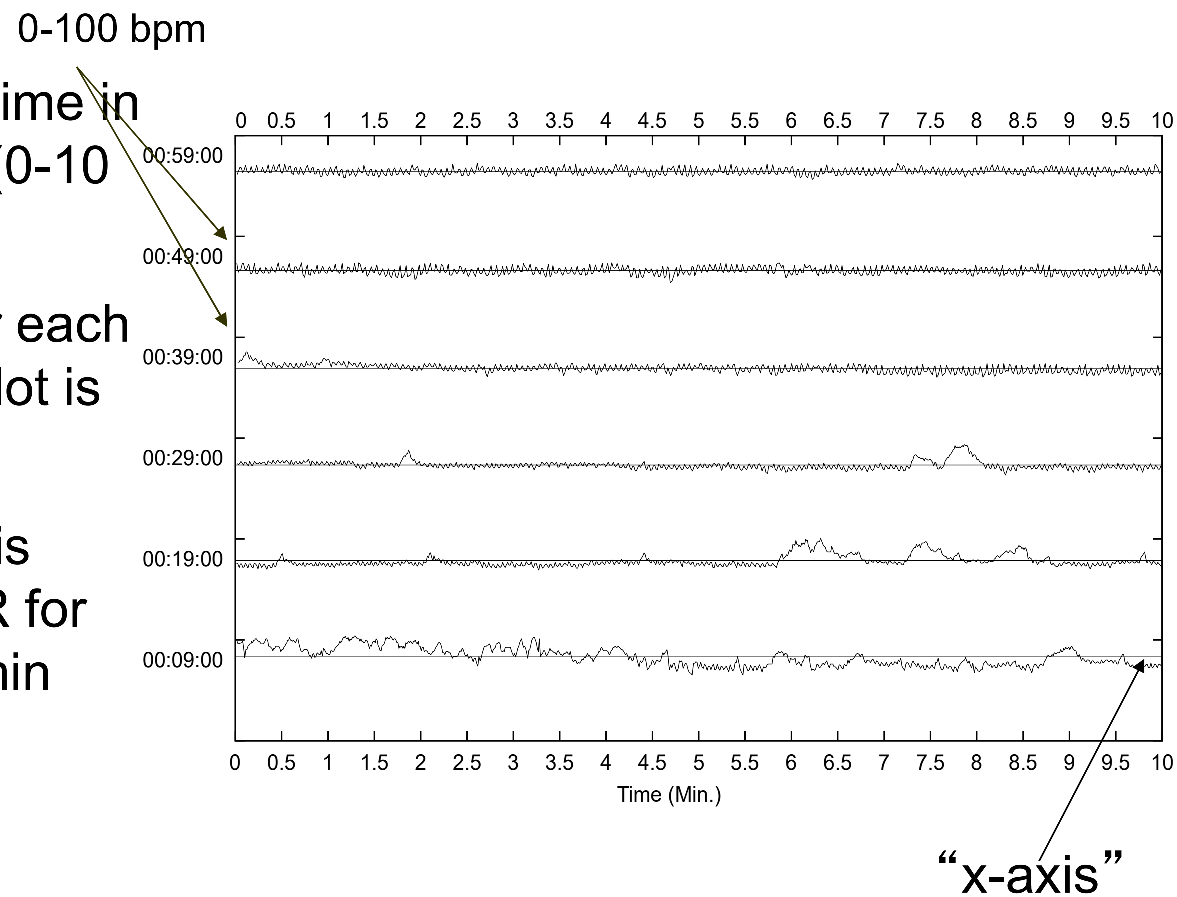
NightSignal

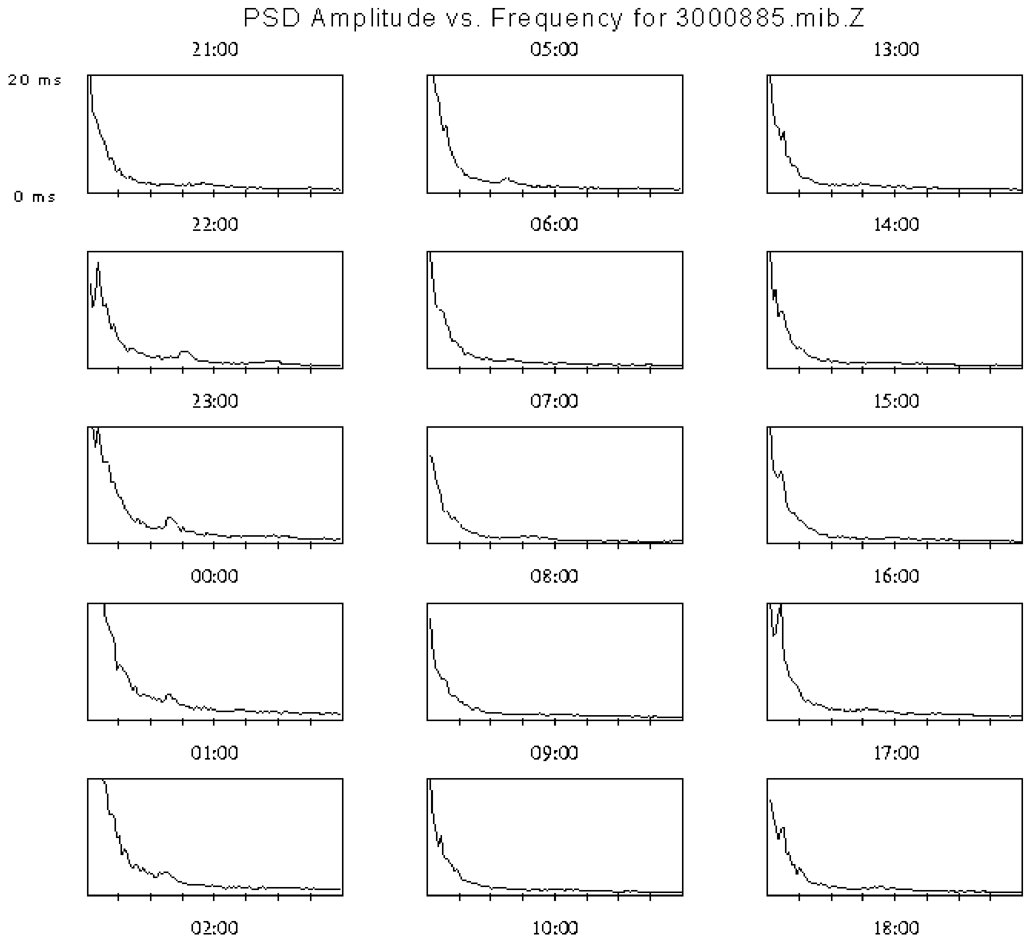


Slides: Dr Michael Snyder MD Stanford

Heart Rate Tachogram

- x-axis = time in minutes (0-10 minutes)
- y-axis for each 10-min plot is HR
- "x-axis" is mean HR for that 10-min segment





Hourly HRV Power Spectral Plots

By the way... What’s wrong with these spectra?

Assessment of HRV

Approach 1

Physiologist's Paradigm

HR data collected over short period of time (~5-20 min), with or without interventions, under carefully controlled laboratory conditions.

https://physionet.org/tutorials/hrv-toolkit/

Assessment of HRV

Approach 2

Clinician's/Epidemiologists's Paradigm

Ambulatory Holter Recordings usually collected over 24-hours or less, usually on outpatients.

Approaches 1 and 2 can be combined

HRV Perspectives

Longer-term HRV

- quantifies changes in HR over periods of >5min.

Intermediate-term HRV

- quantifies changes in HR over periods of <5 min.

Short-term HRV

- quantifies changes in HR from one beat to the next

Ratio HRV

- quantifies relationship between two HRV indices.

Sources of Heart Rate Variability

Intrinsic Periodic Rhythms

- Respiratory sinus arrhythmia
- Baroreceptor reflex regulation
- Thermoregulation
- Neuroendocrine secretion
- Circadian rhythms
- Other, unknown rhythms

Extrinsic

- Activity
- Mental Stress
- Physical Stress

- Sleep Apnea
- Smoking

Ways to Quantify HRV

Approach 1: How much variability is there?

Time Domain and Geometric Analyses

Approach 2: What are the underlying rhythms? What physiologic process do they represent? How much power does each underlying rhythm have?

Frequency Domain Analysis

Approach 3: How much complexity or self-similarity is there?

Non-Linear Analyses

Commonly used time-domain measures

AVNN Average of all NN intervals (i.e. normal to normal)

SDNN Standard deviation of all NN intervals

SDANN Standard deviation of the averages of NN intervals in all 5-minute segments of a 24-hour recording

SDNNIDX Mean of the standard deviations of NN intervals in all 5-minute segments of a 24-hour recording

rMSSD Square root of the mean of the squares of differences between adjacent NN intervals

pNN50 Percentage of differences between adjacent NN intervals that are greater than 50 ms

Time Domain HRV

Longer-term HRV

- SDNN

Standard deviation of NN intervals (in msec) (Total HRV)

- SDANN
- Standard deviation of mean values of QRS-QRSs for each 5 minute interval in msec
- (Reflects circadian, neuroendocrine and other rhythms + sustained activity)

Time Domain HRV

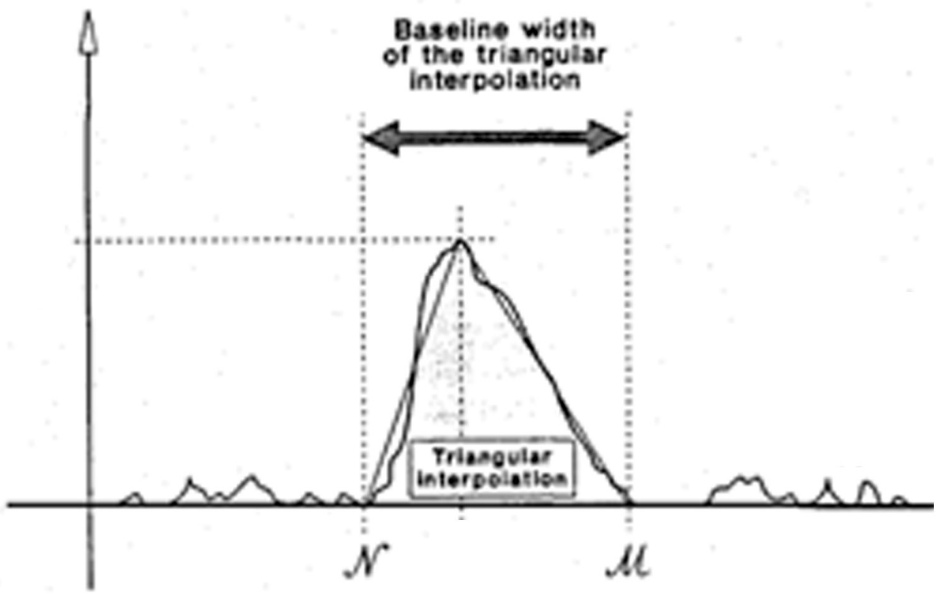
Intermediate-term HRV

- SDNNIDX
- Average of standard deviations of QRS to QRS for each 5 min interval in ms (Combined SNS and PNS HRV)

Coefficient of variance (CV)

- = SDNNIDX / AVNN.
- Heart rate normalized SDNNIDX.

Geometric HRV



HRV Index-Measure of longer-term HRV

Examples of Normal and Abnormal Geometric HRV

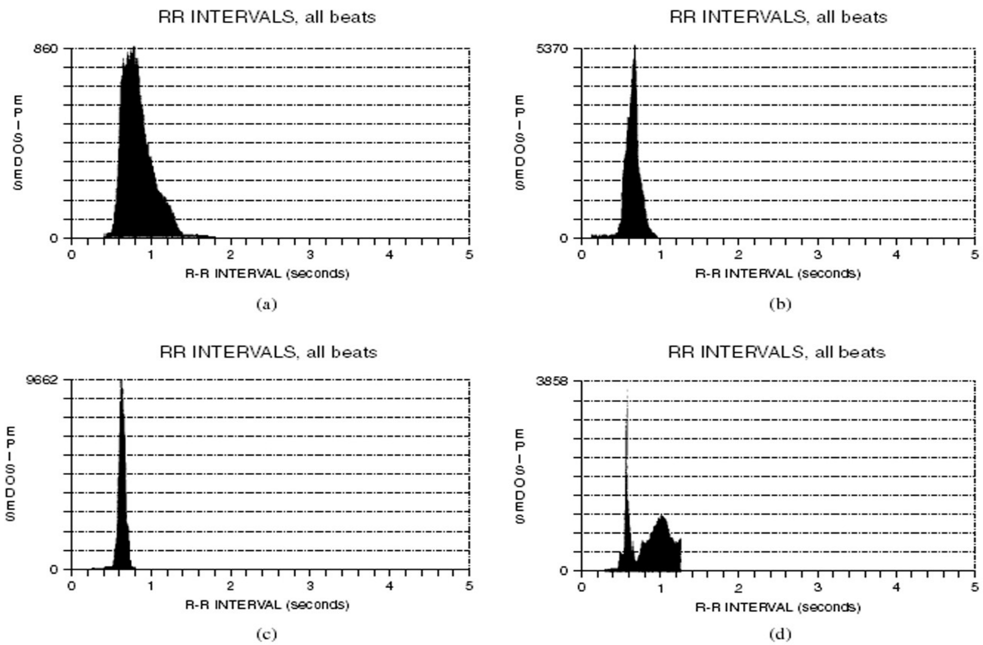


Fig. 1. *Examples of R-R histograms from (a) a normal subject, (b) a cardiac patient with decreased HRV, (c) a cardiac patient with very low HRV and (d) a patient with an extremely abnormal R-R interval distribution and a large number of ventricular ectopic beats.*

Frequency Domain HRV

Based on autoregressive techniques or fast Fourier transform (FFT).

Partitions the total variance in heart rate into underlying rhythms that occur at different frequencies.

These frequencies can be associated with different intrinsic, autonomically-modulated periodic rhythms.

Commonly used frequency-domain measures

TOTPWR
- Total spectral power of all NN intervals up to 0.04Hz

ULF
- Total spectral power of all NN intervals up to 0.003Hz

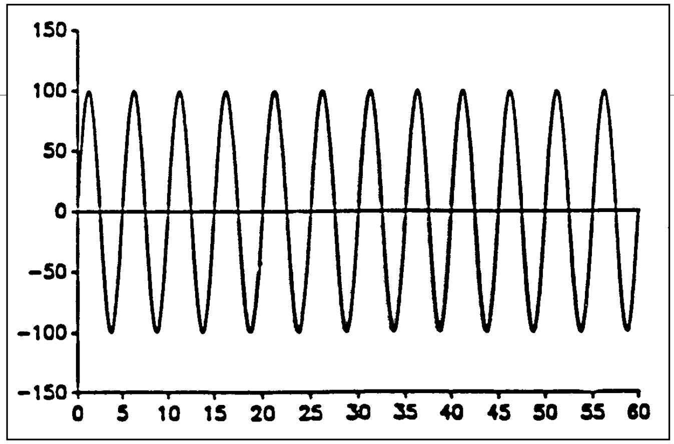
VLF
- Total spectral power of all NN intervals between 0.003 and 0.04Hz

LF
- Total spectral power of all NN intervals between 0.04 and 0.15Hz

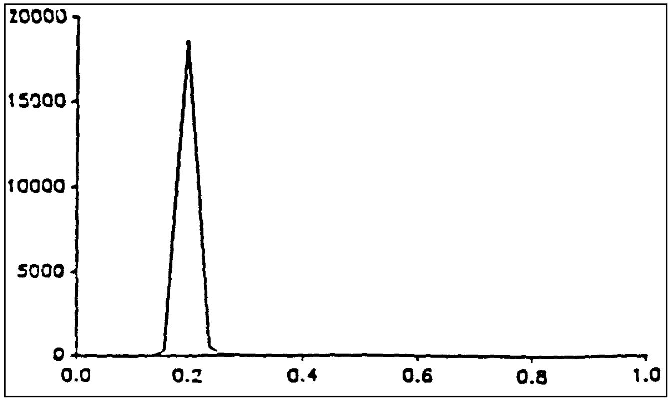
HF
- Total spectral power of all NN intervals between 0.15 and 0.4Hz

LF/HF
- Ratio of low to high frequency power

What are the Underlying Rhythms?



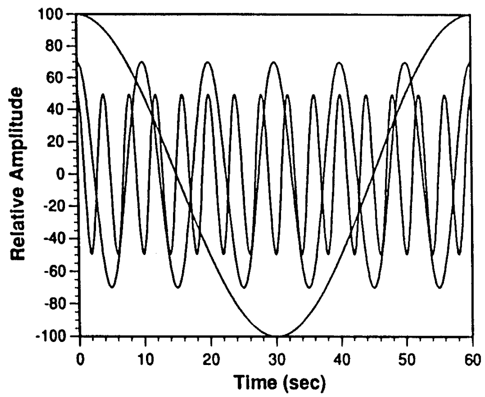
One rhythm
5 seconds/cycle or
12 times/min

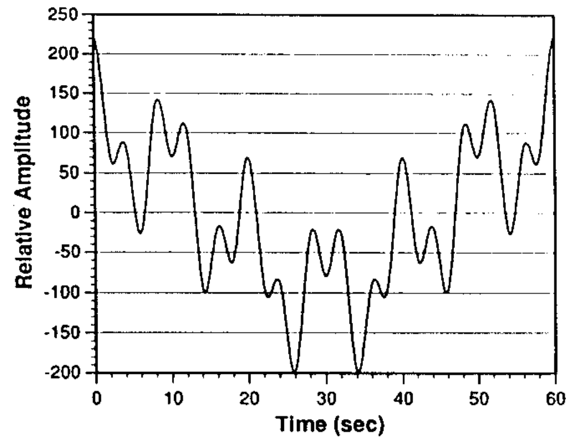


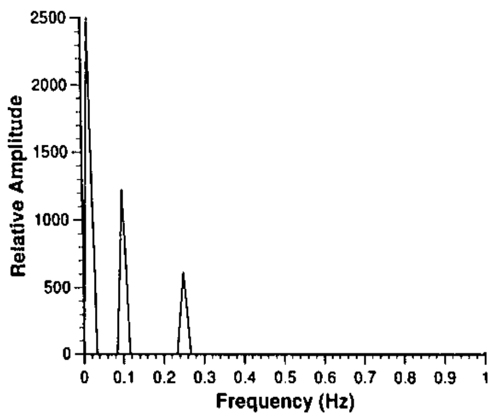
5 seconds/cycle=
1/5 cycle/second

1/5 cycle/second=
0.2 Hz

What are the Underlying Rhythms?







Three Different Rhythms

High Frequency = 0.25 Hz (15 cycles/min
Low Frequency = 0.1 Hz (6 cycles/min)
Very Low Frequency = 0.016 Hz
(1 cycle/min)

Ground Rules for Measuring Frequency Domain HRV

Only normal-to-normal (NN) intervals included

At least one normal beat before and one normal beat after each ectopic beat is excluded

Cannot reliably compute HRV with >20% ectopic beats

With the exception of ULF, HRV in a 24-hour recording is calculated on shorter segments (5 min) and averaged.

Frequency Domain HRV

Longer-Term HRV

Total Power (TP)

Sum of all frequency domain components.

Ultra low frequency power (ULF)

At >every 5 min to once in 24 hours. Reflects circadian, neuroendocrine, sustained activity of subject, and other unknown rhythms.

Frequency Domain HRV

Very low frequency power (VLF)

At ~20 sec-5 min frequency

Reflects activity of renin-angiotensin system, vagal activity, activity of subject.

Exaggerated by sleep apnea. Abolished by atropine

Low frequency power (LF)

At 3-9 cycles/min Baroreceptor influences on HR, mediated by SNS and vagal influences. Abolished by atropine.

Frequency Domain HRV

Short-term HRV

High frequency power (HF)

- At respiratory frequencies (9-24 cycles/minute, respiratory sinus arrhythmia but may also include non-respiratory sinus arrhythmia).
- Normally abolished by atropine.
- Vagal influences on HR with normal patterns.

Frequency Domain HRV

Ratio HRV

LF/HF ratio may reflect SNS:PNS balance under some conditions.

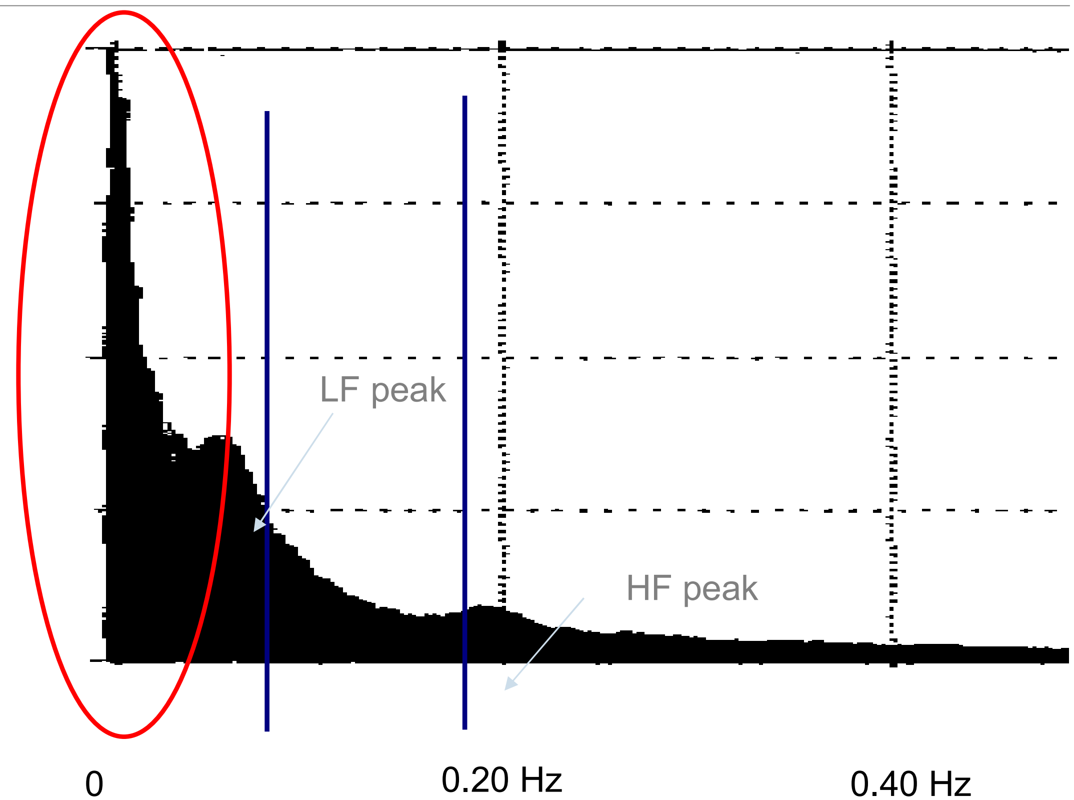
Normalized LF power= LF/(TP-VLF) correlates with SNS activity under some conditions.

Normalized HF power= HF/(TP-VLF) proposed as a measure of relative vagal control of HR. Increased for abnormal HRV.

Frequency Domain HRV

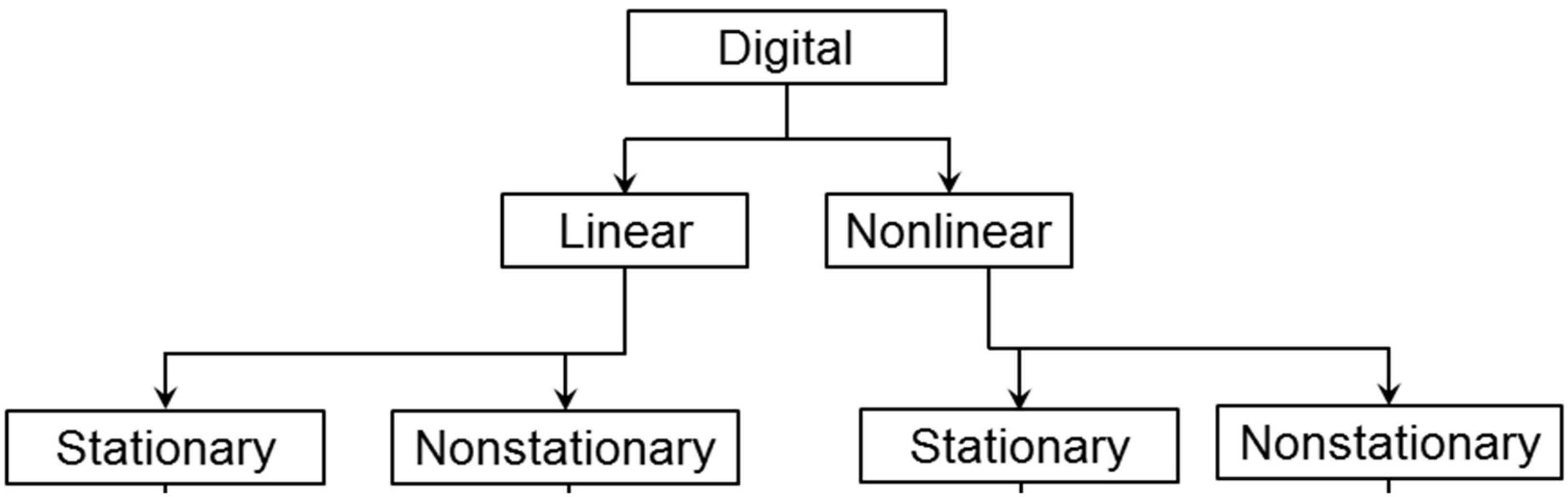
24-hour average of 2-min power spectral plots in a healthy adult

FYI- what's wrong about this spectrum?



How else can we assess variability?

Digital Signal Types



Signal Encoding

All signals involve some type of encoding scheme.

Most encoding strategies can be divided into two broad categories or domains:

- continuous
- discrete.

Continuous signals usually encode information in terms of signal amplitude (the intensity of the signal, voltage, or current values) as a function of time.

Example

The temperature in a room can be encoded so that 0 volts represents 0.0 ^oC, 5 volts represents 10 ^oC, 10 volts represents 20 ^oC, and so on. If *linear*, the encoding equation would be:

Temp	Voltage
0	0
5	10
10	20
15	30
20	40

Linear Signals

input (temperature), output (voltage) following the classic linear relationship:

$$y = mx + b$$

where m is the slope of the input-output relationship and b is the offset which in this case is 0.0.

The temperature can be found from the voltage output of the transducer as:

$$temperature = 2^* \ voltage \ ^\circ C$$

When the information is encoded in terms of signal amplitude, it is known as an analog signal.

Linearity

If you double the input into a linear system, you will double the output.

basic concept is proportionality.

- if the independent variables of linear function are multiplied by a constant, k, the output of the function is simply multiplied by k.

If $y = f(x)$ where f is a linear function:

Properties of Linear Signals

If f is a linear function:

$$f\left(x_1(t)\right) + f\left(x_2(t)\right) = f\left(x_1(t) + x_2(t)\right)$$

- If $z = \frac{df(x)}{dx}$ then $\frac{df(kx)}{dx} = k\left(\frac{df(x)}{df}\right) = kz$
- If $z = \int f dx$ then $\int f(kx)dx = k\int f(x)dx = kz$
- Derivation and integration are linear operations.
- Systems that contain derivative and integral operators and other linear operators produce linear signals.

Time Invariance

If a system's response characteristics do not change over time, it is said to be *time-invariant*.

Time invariance is a stricter version of stationarity since a time-invariant system would also be stationary.

Mathematically: if f is a linear function, then for time invariance:

$$y(t-T) = f\left(x(t-T)\right)$$

LTI Systems

A system that is both linear and time-invariant is referred to as a *linear time-invariant* (*LTI*) system.

The LTI assumptions allow us to apply a powerful array of mathematical tools known collectively as linear systems analysis or linear signal analysis.

Most living systems change over time, they are adaptive, and they are often nonlinear, but the power of linear systems analysis is simplifies everything so much that simplifying assumptions or approximations are made so that these tools can be used.

Causality

A system that responds only to current and past inputs is termed *causal.*

Systems that exist in the real-world (e.g., analog electronic filters) must be causal.

Computer programs can operate using values that appear to be in the future with respect to a given operation.

Such systems are *noncausal.*

Superposition

Linearity is required for the application of an important concept known as *superposition*.

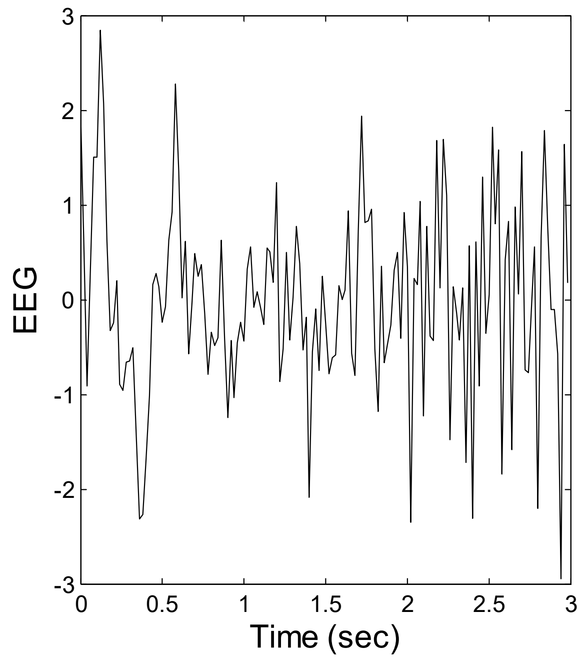
Superposition states that if there are two (or more) inputs acting on a linear system, the system responds to each as if it were the only input

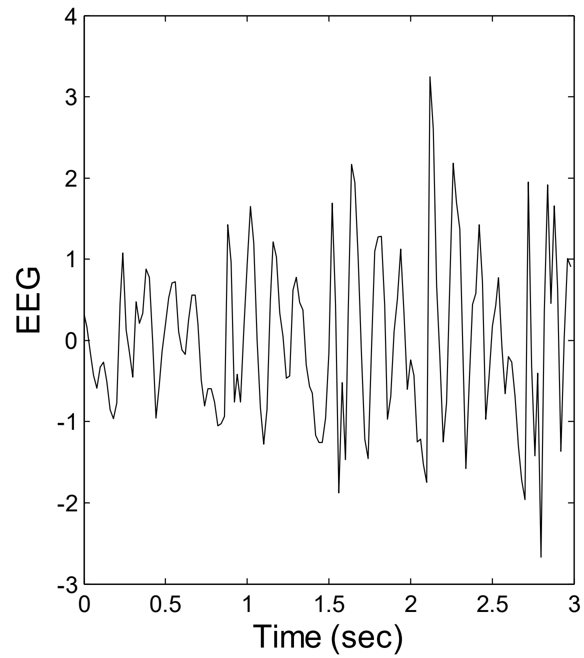
This allows a "divide and conquer" approach

Data Functions and Transforms

Basic measurements do not definitively describe signals.

For example, these two EEG segments have the same mean, RMS, and variance, but are clearly different.





Describing Signals

We would like some method to capture the differences between these two (and other) signals, and preferably to be able to quantify these differences.

Other functions (or waveforms) can be used to describe signals and their differences.

In signal processing, functions fall into two categories:

1) Data, including waveforms and images;

2) Functions that operate on data.

Transformations: Functions that Operate on Signals

Transformations are operators that modify data.

Transformations are used to:

- Improve data quality
- make the data easier to interpret
- Reduce the size of the data by removing unnecessary components

Transformations: depend on data type

Data Type	Single Observation Data		Multiple Observation Data	
Examples	Heart Rate, Respiration Cardiac pressure (*dp/dt*) pCO_2, pO_2		EEG, ECG, VER fMRI	
Transforms	Correlation-based	Convolution-based	Averaging	PCA
Analyses	Fourier transform Continuous Wavelet transform	Filters Discrete Wavelet transform	Ensemble averaging, noise reduction	ICA Data reduction

Comparing Waveforms: Correlation

Correlation seeks to quantify how much one function (i.e., signal) is like another.

All correlation-based approaches involve taking the sum of the sample-by-sample product of the two functions:

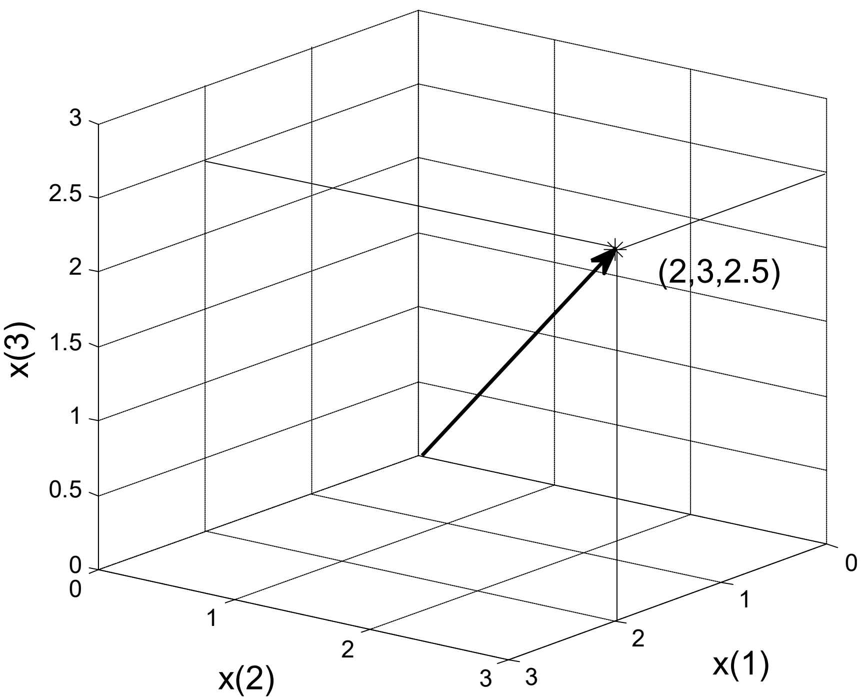
$$r_{xy} = x[1]y[1] + x[2]y[2] + x[3]y[3] + \ldots + x[N]y[N]$$

$$== \sum_{n=1}^{N} x_n y_n$$

where rxy is used to indicate correlation and the subscripts x and y indicate what is being correlated.

- Different normalizations ($\frac{1}{N}\sum_{n=1}^{N} x_n y_n$) can be used.

Vector Representation



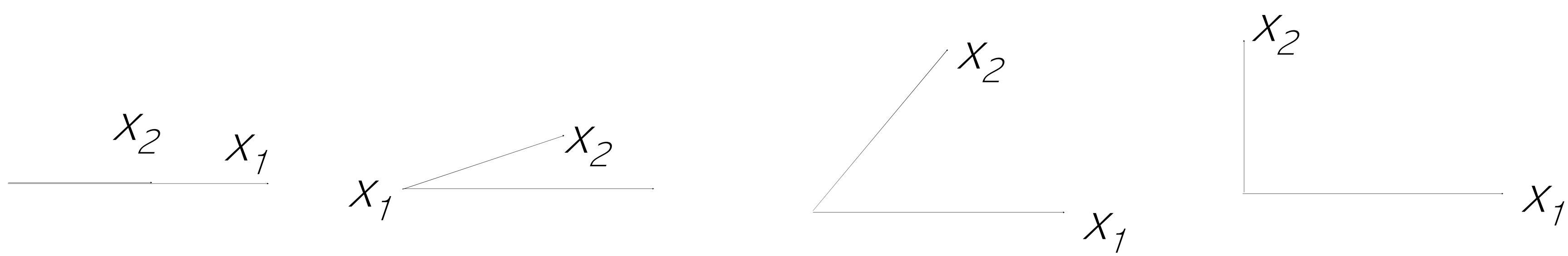
A string of numbers can be thought of as a vector in N-dimensional space: $x[n] = [x_1, x_2, x_3, \ldots x_n]$

Data sequence:
$x[n] = 2, 3, 2.5$
represented as a vector in 3-dimensional space.

This curious way of thinking about a data string does have its uses.

Signal Comparison using Vector Representation

If two strings are mathematically similar, their vector representations will project on one another:



Completely similar
ϑ = 0 deg.

Highly similar
ϑ = small

Moderately similar
ϑ = larger

Completely different
ϑ = 90 deg

Correlation and the Scalar Product

The projection of one vector on another is found by taking the scalar product of the two vectors.

This shows the relationship between vector projection and correlation.

$$Scalar\ product\ of\ x\ \&\ y \equiv \langle x, y \rangle = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}$$

$$= x_1 y_1 + x_2 y_2 + \ldots + x_N y_N$$

The scalar product is defined as:

$$= \sum_{n=1}^{N} x_n y_n$$

Scalar Product – Correlation (cont)

Note that the scalar product results in a single number (i.e., a scalar), not a vector.

The scalar product can also be defined in terms of the magnitude of the two vectors and the angle between them:

where θ is the angle between the two vectors.

Projection (correlation) is an excellent way to compare two signals or to compare a signal with a 'probing' or 'test' waveform.

In MATLAB, the scalar product is just: sum(x.*y)

Orthogonality

In common usage, "orthogonal" means perpendicular: if two lines are orthogonal they are perpendicular.

In vector representation, orthogonal signals would have orthogonal vectors.

The formal definition for orthogonal signals is that their correlation (or scalar product) is zero:

Orthogonality (cont)

An important characteristic of signals that are orthogonal (i.e., uncorrelated) is that when they are combined or added together they do not interact with one another.

Orthogonality simplifies many calculations and some analyses could not be done, at least not practically, without orthogonal signals.

Orthogonality is not limited to two signals. Whole families of signals can be orthogonal (or orthonormal*) and are called orthogonal or orthonormal sets.

Coherence

- The coherence C between two signals x and y is defined as the cross-spectrum Sxy normalized by the power spectra Sxx and Syy

- However, to make the coherence a dimensionless number between 0 and 1, Sxy is squared:

NOTE: Sxy will usually be a complex function, whereas Sxx and Syy are both real functions

Coherence (cont)

If we calculate the normalized cross-spectrum as a complex number for a single frequency and a single trial, the outcome always has magnitude 1 and phase angle φ.

E.g. define:

$$X(\omega) = a + bj \qquad Y(\omega) = c + dj$$

Can write expression for each of the components of the coherence equation as:

$$S_{xx}(\omega) = X(\omega)X^*(\omega) = (a + bj)(a - bj) = a^2 + b^2\,,$$

$$S_{yy}(\omega) = Y(\omega)Y^*(\omega) = (c + dj)(c - dj) = c^2 + d^2\,, \quad \text{and}$$

$$|S_{xy}(\omega)|^2 = |X(\omega)Y^*(\omega)|^2 = |(a + bj)(c - dj)|^2 = |(ac + bd) - j(ad - bc)|^2$$

Coherence (cont)

Because S_{xy} is a complex number, the magnitude squared is the sum of the squares of the real and imaginary parts:

$$|S_{xy}(\omega)|^2 = (ac + bd)^2 + (ad - bc)^2 = a^2c^2 + b^2d^2 + a^2d^2 + b^2c^2$$

Looking at the original coherence equation it is easy to see why it equals 1:

$$C(\omega) = \frac{a^2c^2 + b^2d^2 + a^2d^2 + b^2c^2}{(a^2 + b^2)(c^2 + d^2)} = 1$$

Coherence (cont)

- coherence is typically estimated by averaging over several epochs or frequency bands
- Thus Sxy is determined by averaging over n epochs:

$$C(\omega) = \frac{\left| \left\langle S_{xy}(\omega) \right\rangle_n \right|^2}{\left\langle S_{xx}(\omega) \right\rangle_n \left\langle S_{yy}(\omega) \right\rangle_n}$$

VERY Important Note: The averaging of cross-spectrum Sxy occurs before the absolute value is taken. A common beginner's mistake is to average the absolute which will always come out to one!

Coherence(cont)

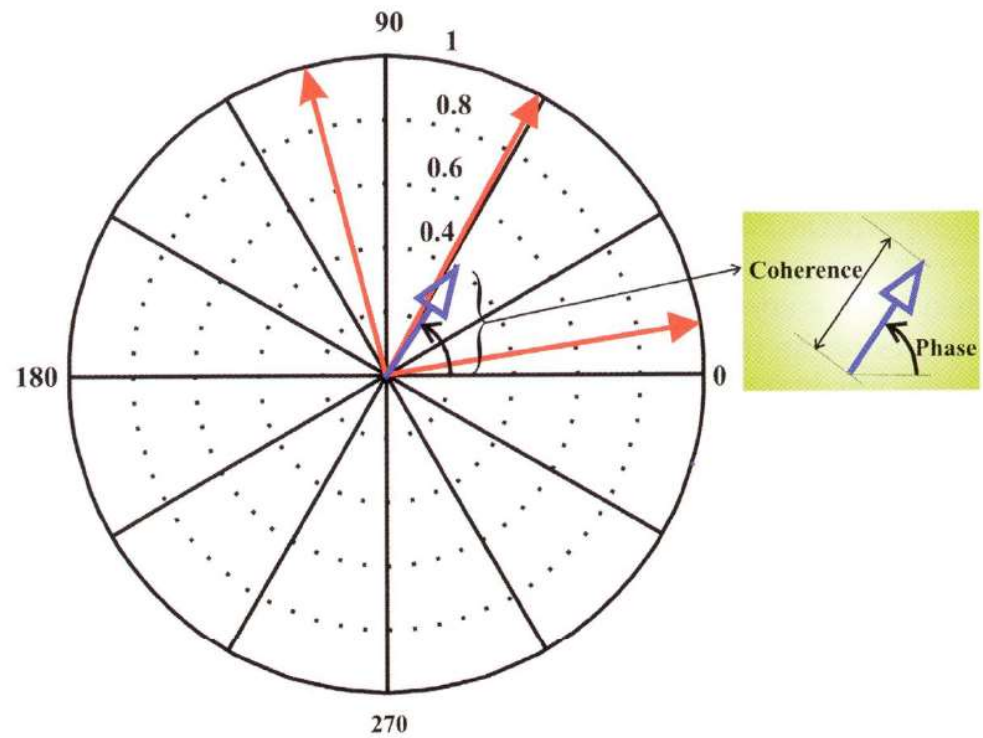
- When determining $C(\omega)$ for a single frequency, ω over different samples (from an ensemble) one obtains several vectors on the unit circle, typically with different phase angles.

- The magnitude of sum of individual vectors indicates the degree of coherence, and the resulting phase angle is the phase coherence.

Coherence (cont)

Complex numbers (red vectors) in complex plane are values for normalized cross-spectrum obtained from different samples out of an ensemble.

The blue arrow is average of the 3 (i.e. amplitude coherence, or just coherence), and the phase is the phase coherence.



Coherence (cont)

The magnitude of a single coherence estimate is always 1. The use of the coherence metric therefore only makes sense if the value is determined repeatedly and subsequently averaged.

Usually the coherence values are:

(1) averaged over different frequencies in a frequency band

(2) averaged for a given frequency band for different epochs, or

(3) averaged over both frequencies and epochs of the signal

Example

An example of coherence calculations associated with subdural electrode arrays implanted over the frontal cortex (red and blue, 1cm spacing) and temporal cortex (green-yellow, 5-mm spacing) of a patient with medically intractable epilepsy.

- colored pipes indicate pairs of electrodes with high coherence between them.

White pipes are not associated with a phase shift.

Green pipes indicate a phase delay at the blue end of the pipe.