POST-INFECTIONOUS HYDROCEPHALUS IN UGANDAN INFANTS: MICROBIAL ORIGINS OF INFECTION AND THE ROLE OF CLIMATE IN DISEASE DISTRIBUTION

SYLVIA L. RANJEVA

Spring 2012

A thesis
submitted in partial fulfillment
of the requirements
for baccalaureate degrees
in Engineering Science and French and Francophone Studies
with honors in Engineering Science

Reviewed and approved* by the following:

Steven J. Schiff
Director, Penn State Center for Neural Engineering
Brush Chair Professor of Engineering
Professor, Department of Engineering Science and Mechanics
Thesis Supervisor

Judith A. Todd
P. B. Breneman Deparment Head Chair
Professor, Department of Engineering Science and Mechanics
Honors Advisor

Christine B. Masters
Associate Professor, Department of Engineering Science and Mechanics
Thesis Reviewer

* Signatures are on file in the Schreyer Honors College and Engineering Science and Mechanics Office.
I. ABSTRACT

The diagnosis and treatment of neonatal infectious disease in Africa and throughout the developing world begs for the collaborative intervention of engineering in biomedical research. The international prevalence of post-infectious hydrocephalus (PIH) in infants serves not only as a fascinating issue in the dynamical consequences of infectious disease but also reveals the socio-economic barrier to treatment of infectious disease in the developing world. Identifying the microbial spectrum that causes PIH could lead to strategies to alleviate the disease from the regions of the world where treatment remains the most elusive. It is further imperative to understand the driving environmental factors, such as climate and rainfall patterns, which lead to the spread of this infectious bacterial syndrome.

My current research with PIH in Uganda addresses these possibilities. Throughout this project I have worked with Dr. Steven Schiff, director of Penn State University's Center for Neural Engineering, Dr. Mary Poss, director of the Penn State Center for Infectious Disease, and Dr. Benjamin Warf, a pediatric neurosurgeon at Harvard University and the former Medical Director of CURE Children’s Hospital of Uganda.

The first phase of the project aimed to determine the infectious agents present in the bacterial diversity of cerebral spinal fluid (CSF) among infants with PIH throughout Uganda. Bacterial DNA was recovered from 94% of patients. The most common potential pathogens were Gram-negative Proteobacteria. Gammaproteobacteria were the most common in patients presenting for infections during the rainy season, and Betaproteobactreia were most common across patients presenting during the dry season. Within the cohort of patients presenting after the rainy season, Acinetobacter species were identified in the majority of samples. This study
suggested a possible bacterial source of infection of PIH, and more notably suggested seasonality in the nature of PIH infections.

In a second study, a spatio-temporal correlation of climate statistics (rainfall averages) with disease distribution aimed to quantify the relationship between climate and the incidence of PIH. Date and location of patients presenting for PIH over a six-year time series were compared with corresponding rainfall values. A time series analysis of cases and rainfall patterns was performed. Infections were shown to occur at intermediate levels of rainfall. Four infection onset peaks overlay twice-yearly rainfall peaks, suggesting a dependence of case onset on phase of the rainfall cycle.

In conjunction with microbial analysis suggesting a pathogenic seasonal difference in PIH infection, this study suggests the importance of climate in preventative measures taken to combat the spread of newborn infections causing PIH.
Table of Contents

I. Introduction

II. Microbial Origins of Post Infectious Hydrocephalus in Ugandan Neonates
   i. Introduction to Techniques Used in Analysis
   ii. Methods
   iii. Results
   iv. Discussion

III. Spatio-Temporal Analysis of Climate and Case Distribution of PIH
   i. Introduction to Techniques Used in Analysis
   ii. Methods
   iii. Results
   v. Discussion

IV. Conclusions and Implications for Future Studies

V. References
I. INTRODUCTION

Post-Infectious Hydrocephalus: Motivation for Study:

Hydrocephalus, literally water on the brain, refers to excess buildup of CSF within the cranial vault due to excessive production or inadequate absorption. One of the most common pediatric developmental disabilities, the leading cause of pediatric brain surgery in the world, hydrocephalus can be heritable or acquired post-natally as the result of hemorrhage or infection (Warf 2011). Fatal if left untreated, the disease severely disrupts cognitive development among children.

Worldwide, the condition affects between one and three infants per thousand, creating a substantial global health burden: in East Africa, treatment and diagnosis of the condition are expensive and technologically complex, exacerbating the debilitating effects of hydrocephalus in developing economies. Treatment remains surgical, relying on the cerebral shunt, which is implanted via neurosurgery into a patient's brain to divert excess CSF. Cranial imaging techniques used to diagnose hydrocephalus are often unrealistic for application throughout East Africa, where poverty and lack of political infrastructure prevent public health interventions. In Uganda, for example, where a majority of the nation’s 28 million people live significantly below the poverty line, most people with hydrocephalus die without treatment (Piquer 2010).

Furthermore, in East Africa, the high incidence of hydrocephalus, high infection rates, and obstacles to follow-up care make shunt use less effective than in developed countries. In East Africa, less than 10% of hydrocephalic infants will be operated annually using shunting procedures, with the attendant risks of shunt infection and blockage as high as 25%-50% (Piquer 2010).
Infant hydrocephalus in East Africa has been characterized as 60% post-infectious (Warf 2005), related to preceding neonatal infections, and therefore preventable. Identifying the microbial spectrum present within the CSF of patients presenting with PIH could be a key step towards the development of cost-effective anti-microbial prevention strategies applicable within the constraints of the developing world. Furthermore, the dependence of neonatal infection on climate patterns could shed light onto preventative measures. Climate plays an important role in seasonal infections closely tied to PIH. For example, the dry season across the “meningitis belt” of sub-Saharan Africa brings sweeping dust-born infections of *N. Meningitidis*, causing seasonal meningitis epidemics. The WHO reported 27,000 deaths due to pediatric meningitis across the Meningitis Belt in 2008 (WHO 2008.) In the Sudan, Aziz showed strong seasonality in PIH linked to annual meningitis epidemics (Aziz 1976). This study did not involve bacteriological identification of infection routes, but PIH was presumably caused by *N. meningitidis*. While the exact time of neonatal infection was not characterized, such evidence suggests rainfall patterns (‘wet vs. dry’ seasonality) as a potential key factor in characterizing the nature of infection. There are two rainy seasons in Uganda: generally April–May and October–December. Analyzing the correlation of date of infection with preceding febrile illness and rainfall could establish a vital link between climate and the causative bacterial infections across Uganda.
II. MICROBIAL ORIGINS OF PIH IN UGANDAN INFANTS

i. Introduction to Techniques Used in Analysis

Modern molecular techniques identify the presence of microorganisms by the amplification of their genetic material through PCR. Amplified genes can be used to classify bacteria, often to a species-level specificity, and to establish taxonomic relationships between samples.

a. 16s PCR:

Polymerase Chain Reaction is a molecular assay which amplifies a single sequence of DNA across several orders of magnitude, creating millions of copies. Within a genome, PCR selects specific genes of interest using primers, short DNA fragments containing sequences complementary to the target region, amplifying the region between the forward and reverse primers. Due to the nature of this specificity, traditional molecular assays, unlike bacterial culturing, specifically target one organism. This provides high sensitivity and specificity but implies a preconceived target organism. The present study relies on broad-range characterization of an unknown bacterial spectrum present across patient samples of CSF.

To overcome this limitation, broad-range assays using the 16s ribosomal RNA gene are widely implemented for bacterial identification. Bacterial ribosomal DNA (rDNA) consists of highly-conserved genetic sequences shared by all bacteria, containing variable regions that are genus- or species-specific. Hence, these genes allow for common detection and amplification and provide the basis for phylogenetic classification based on the variable regions (Doolittle, 1999). Among ribosomal DNA, 16s rRNA genes feature highly conserved primer binding sites and hypervariable internal regions, providing an optimal standard for these methods. By using PCR
primers that are targeted at conserved regions of rDNA, it is possible to design broad-range PCRs capable of detecting DNA from almost any bacterial species. The identity of the bacterium captured is revealed by nucleotide sequencing of the PCR product followed by comparison of this sequence with known sequences located in GenBank or other databases (Case 2007).

A major advantage of 16s rRNA PCR as a molecular diagnostic tool is its ability to classify the genetic material of bacteria which cannot be cultured. Under research conditions, this approach has been applied to samples from normally sterile sites to diagnose bacterial infection, including meningitis, pneumonia and bacteremia, when culture results proved negative (Lu et al., 2000). Particularly, in the present study, CSF sampled at the time of shunt insertion was consistently negative in bacterial culture in over 1000 such cases at CURE Children’s Hospital of Uganda since 2000.

b. Basic Local Alignment Search Tool:

Bacterial nucleotide sequences are commonly classified by identification against known reference sequences. Basic Local Alignment Search Tool, or BLAST, is suite of tools for identifying imperfect matches between a given query sequence and a database references. The BLAST algorithm functions as follows:

1.) The query sequence is broken into short subsequences called words. The program identifies the exact matches to the query words first (word hits). A “word” refers to a 11-base nucleotide sequence contained within the query. The reference database is searched for the initial query word that scores above a given threshold value T when compared to the query, according to a scoring matrix. Only those “words” that score higher than the neighborhood word score
threshold (T) are kept, and a table is generated with High-Scoring Pairs for the initial word.

Figure A below demonstrates this process:

Figure A: BLAST Algorithm: Generation of HSPs for Initial Word

2.) Once a complete word identification is found between the query sequence and a database sequences, the BLAST program extends the algorithm in both directions multiple times to generate the final gapped alignments.

Scoring is determined using an identity matrix: an identity between the reference sequences receives a score of 1, a mismatch receives a score of -3, and a gap receives a score of -1. Further penalties are assigned for gap extension. Figure B below shows the generation of a total (raw) score between a word on a query sequence and a reference sequence.

Figure B: BLAST Scoring
A key parameter of the BLAST search is the Expectation (E) value, which provides a measure of confidence for a given alignment. The Expectation value indicates how probably a given alignment could have arisen by chance. By default, BLAST produces alignments which fall within an E-value of 10, meaning that the chance of the alignment occurring randomly is $1\times10^{-10}$ (Tao 2011). The E-value is determined as follows:

$$ E = K \cdot m \cdot n \cdot \lambda^{-S} $$

Where:

- $E$ = Expectation Value
- $K$ = value dependent upon the substitution matrix used and adjusted for database size
- $m$ = number of nucleotides in query
- $n$ = number of nucleotides in database
- $\lambda$ = natural scale for the scoring system
- $S$ = raw score

**c.) Maximum Likelihood Model for Phylogenetic Analysis:**

Bacterial phylogenies present evolutionary models of the relationships between bacterial specimens based on the similarity of their genetic sequences. The topology of a rooted phylogenetic tree depicts the branching history of common ancestry relating the taxa, while an unrooted tree depicts branching as the result of degrees of genetic similarity, without...
assumptions about ancestry. Maximum Likelihood is a method for the inference of phylogeny. It evaluates a hypothesis about evolutionary history in terms of the probability that the proposed model and the hypothesized history would give rise to the observed data set. The supposition is that a history with a higher probability of reaching the observed state is preferred to a history with a lower probability. The Maximum Likelihood method operates under several key assumptions. First of all, changes at different sites along the genome occur independently. Under the assumption that nucleotide sites evolve independently, we can calculate the likelihood for each site separately and combine the likelihood into a total value towards the end. After speciation, lineages are assumed to evolve independently and the same process can be applied at each node along the tree. To calculate the likelihood for any site, we have to consider all the possible scenarios by which the nucleotides present at the tips of the tree could have evolved. So the likelihood for a particular site is the summation of the probabilities of every possible reconstruction of ancestral states. For the entire tree, the Likelihood based on all N sites is the product of the likelihood of each site, or:

\[ L = L(1) \times L(2) \ldots \times L(N) \]

Individual Likelihoods are often very small numbers, so Likelihood is often reported by summing the natural log of individual likelihoods;

\[ \ln L = \ln L(1) + \ln L(2) \ldots + \ln L(N) \]
The Maximum Likelihood method searches for the tree with the highest probability or likelihood (Guindon 2003).

ii. Methods:

Samples of CSF were collected from infants presenting with hydrocephalus (in a majority of cases after the resolution of preceding febrile illness) in three cohorts. Cohort 1 contained specimens from 25 consecutive patients collected in January 2008 for DNA analysis; Cohort 2, specimens from 25 consecutive patients obtained in July 2008 for culture; and Cohort 3, specimens from 25 consecutive patients obtained in October 2008 for DNA analysis. Each specimen was obtained from an individual patient. The presenting infants were aged 1-6 months, and the mean time from neonatal infection to onset of PIH falls at .8 months (Warf 2005); therefore, patients presenting in January were presumptively infected during the rainy season, while those presenting in October were presumably infected during the dry season. Environmental samples were also obtained. We collected samples for culture from hut floors, nearby animal dung, and water supplies from the villages of 8 of 17 Cohort 1 patients with evidence of Acinetobacter species in their CSF. Our culture efforts targeted gram-negative bacteria, and a wide variety of species were recovered. Out of 39 surveillance cultures, Acinetobacter was recovered from 2 hut floors and 1 cow dung sample. The 16S rRNA genes were amplified from these 3 positive cultures to identify the bacterial species.

DNA was recovered using a standard extraction technique from Whatman FTA cards and Biomatrica CrudE tubes spotted with 100µL CSF. Both methods permit the stabilization of nucleic acids for long-distance transport at room temperature. No bacteria from Cohort 2, samples collected for culture, were culture positive either at Mbale, where culture was attempted
on chocolate agar and blood agar, or at Penn State University, where extensive culture techniques were employed.

Extraction of DNA from the FTA cards was conducted using the procedure described by Biek et al. Briefly, the entire area containing sample was washed, minced, and eluted with 300–400 ml water at 95°C for 60 minutes by using sterile reagents and instruments. The paper and eluate were transferred to a spin column to remove solid material, the filtrate was adjusted with 1 M Tris to 10 mM Tris pH 8, and DNA recovery was achieved using a Qiagen DNeasy blood and tissue kit.

The PCR reaction was conducted using primers 16S-8F (5’-AGAGTTTGATCCTGGCTCAG) and 16S-534R (5’-ATTACCGCGGCTGCTGGC). Polymerase chain reaction products were sequenced directly or cloned into the pGEM-T vector (Promega), and cloned products were sequenced. Sequences were aligned using the DNASTAR package, version 7, and MEGA, version 4. For accurate reconstruction of the relationships of bacteria based on the 16S rRNA gene, a model of how the sequences evolve must be developed. Over 50 models of sequence evolution were evaluated, and the best model for each dataset was selected using the Akaike Information Criterion as implemented in the Modeltest, version 3.7. The phylogenetic trees, which display the genetic relationships among the individual sequences of Acinetobacter species, were produced using a maximum likelihood tree method implemented in the program PhyML, version 2.4. Using the same program, the strength of support for each node in the tree was estimated with 100 bootstrap replicates (a statistical resampling method) in the same program. To infer the relationships among individual bacterial sequences within each Acinetobacter species, parsimony networks were reconstructed using TCS, version 1.2. Gaps
were treated as missing data, and all the sequences were set to connect at the 95% confidence limit (Li et al 2011).

**iii. Results:**

Bacterial DNA fragments were recovered by 16sPCR for 94% of samples from patients in Cohorts 1 and 3. Cohort 2, samples collected for culture, failed to yield viable organisms.

**Figure 1:** Phylogenetic relationship of all unique CSF samples, classified by comparison to known bacterial sequences obtained from GenBank.

The DNA from patient CSF samples was classified based on comparison to the sequences of known bacterial specimens. Samples were BLAST-ed against the GenBank database, an open-access collection of annotated nucleotide sequences maintained by the National Center for
Biotechnology Information (Benson 2008). As demonstrated above, bacterial diversity among CSF samples was high, with many sequences failing to match classified reference strains available on the GenBank. Proteobacteria dominated across samples.

**Figure 2:** Expanded phylogeny showing relationships among Gammaproteobacteria genera detected among samples. Bootstrap values at the base of each node indicate statistical confidence.
In Cohort 1, of 21 patients for whom 16s PCR succeeded in amplifying bacterial DNA, 19 reflected the presence of Gammaproteobacteria, a Gram-negative enteric bacteria. Additionally, 95% of these samples reflected infection by *Acinetobacter*, as compared to reference strains.

**Figure 3:** Expanded phylogeny showing relationships among Betaproteobacteria genera detected among samples of CSF, primarily from Cohort 3.
While Proteobacteria was again the most popular phylum represented among Cohort 3 patients, 14 of 22 patients analyzed identified with references from the Betaproteobacteria family rather than the dominant Gammaproteobacteria of Cohort 1. The remaining 8 were classified as Gammaproteobacteria. The remaining patients in this cohort presented sequences that could not be classified.

**Figure 4:** Species grouping of *Acinetobacter* detected in samples from Cohort 1, Cohort 3 and the environment. The sequences highlighted in red represent patients from Cohort 1 presenting samples which did not cluster with the rest of the Cohort.

**Figure 5:** Parsimony Network: Each circle represents one unique sequence, with the size of circles proportional to the number of identical sequences across patients. The color of each circle
iv. Discussion:

The recovery of bacterial DNA in 94% of patient CSF samples supports previous assertions that the majority of neonatal hydrocephalus in East Africa is post-infectious in nature. Figure 1 demonstrates a striking bacterial diversity of the bacteria sampled, with Proteobacteria emerging as the dominant phylum. This study characterized degree of similarity between sequences acquired across patients. Acinetobacter infections appeared common, particularly among patients from Cohort 1. Within samples from Cohort 1, 95% of the 19 samples which identified with Gammaproteobacteria were identified within the genus Acinetobacter. Figure 5 displays the tightly-related cluster among these samples which identify with A. junii. Within the A. junii cluster, only 1 sample was derived from a patient in Cohort 3. A second smaller cluster
of Cohort 1 sequences matched more closely with A. parvus, which is a close relative of A. junii. The two other patients in Cohort 3 from whom Acinetobacter was amplified clustered with environmental samples from hut floors or from dung collected during the dry season.

Perhaps the most notable finding of this study is the suggestion of seasonality with regards to the bacterial routes of infection for children presenting with PIH. The species of Acinetobacter identified in patients with PIH varies between patient cohorts, and the prevalence of A. junii is particularly high in the rainy season. In general, the presence of enteric Gram-negative bacteria suggests environmental sources of infection, and the shifting nature of bacterial genera with season of infection indicates that climate could be an important driving force in the nature of infection. Figure 5 demonstrates the relationship between samples from Cohort 3, patients infected during the dry season, and samples drawn from environmental sources. Note that within the parsimony network, these samples present an isolated cluster that cannot be related with reasonable statistical confidence to the samples from Cohort 1. The association of bacteria found in samples from patients infected during the dry season with environmental bacteria calls to reference the strong presence of climate as a driving factor in “dry epidemics” of meningitis just north of Uganda which spread bacterial infections seasonally (WHO 2008).
III. SPATIO-TEMPORAL ANALYSIS OF CLIMATE AND CASE DISTRIBUTION OF PIH

A database detailing the treatment of 696 infants presenting for PIH over 6 years was parsed by location (District of origin) and time (month of onset of febrile illness, birth, and date of presentation for surgery). The data was compared to satellite-derived rainfall averages for the same geographical distribution and time period. The dependency of case distribution on rainfall patterns was examined by frequency filtering and statistical methods.

i. Introduction to Techniques used in Analysis

a.) Frequency filtering:

In signal processing and electronics, filters can be used to select for a desired range of frequencies within a “noisy” input signal. Filters are described according to their behavior. For example, a low-pass filter will pass all signals below a specific frequency, but will attenuate or block signals of a higher frequency. High pass filters perform the opposite functions. Combining these characteristics, filter circuits can further be designed and modeled mathematically to extract a specific band of frequencies from a wider range of input signals. The result is called a band-pass filter. The pass band is the range of frequencies over which it will pass an incoming signal. Signal frequencies lying outside the pass band are attenuated. The following block diagram describes a narrow band filter, demonstrating the combination of transformation by a high pass filter g(f) and a low-pass signal h(f). The graphs above each block represent the effect on the response of the signal at given frequencies.

Figure C: Block Diagram Describing Narrow Band Filter
Many types of analog filters are available to suit the needs of specific filter designs. A Chebychev filter provides maximally steep attenuation in the stop band, while accepting ripple in the passband. Generally, Chebychev filters provide the steepest slope in the transition band, or the frequency range just following the cutoff frequency, or maximum desired frequency of the passband, before the stop band. A schematic of these concepts is shown below for the higher frequency limit of the passband:

**Figure D: Characteristics of Band Regions in an Analog Filter**

![Image of filter characteristics]

Image Courtesy of *Practical Analog and Digital Filter Design* (Thede 2004 p.4)

The Chebyshev response is based on a Chebyshev polynomial that has equal ripple about the desired pass-band response, producing a maximally accurate approximation to the desired pass-band response. The Chebychev polynomial is a representation of the filter order. The order of a filter determines how steep the response falls off in the stop-band. The voltage response of a Chebychev filter is given by the transfer function:
\[ \frac{V_{out}}{V_{in}} = G/(a_0 + a_1s + a_2s^2 + \ldots + a_{n-1}s^{n-1} + s^n) \]

where \( V_{out} \) and \( V_{in} \) denote the voltage responses of the output and input signals of the filter, respectively, \( G \) denotes the filter gain, \( a \) denotes the coefficient which moderates the filter response over ranging frequencies, and \( n \) denotes the filter order (Thede 2004). The higher the order, the steeper the slope of the response versus frequency. In the case of the Chebychev filter, a higher order provides maximum separation of desired and attenuated frequencies by providing the steepest transition band. A pole exists for a filter wherever its transfer function becomes infinite. Therefore, given the Chebychev polynomial, the number of poles is equal to the filter order. The graph below compares the responses of different Chebychev filter orders \( (n=4, n=3, \text{ and } n=2) \) at the upper limit of the pass-band transitioning to the stop-band.

Figure E: Attenuation Response Based on Filter Order
This investigation used a narrow-band Chebychev filter to extract a desired frequency from raw data due to the optimal precision of the frequency range in the pass-band. A high filter order of five was used to ensure maximally sharp attenuation of frequencies outside the pass-band.

**b.) Statistical Characterization of Non-Uniformity:**

Statistical measures provide confidence estimates to observable peaks in the distribution of histograms. The present study investigates the distribution of cases, a total-sum point process, as a function of phase of continuous rainfall cycles.

Initially, circular statistical tests of non-uniformity were considered to determine the deviation of angular histograms (viewing one year, with two rainy seasons, as a cycle of phases from zero to $4\pi$) from uniform distribution. A common such test is Rayleigh’s test of non-uniformity. For cases binned according to phase $\theta$, Rayleigh’s test treats each count (or case of PIH) in a bin as a discrete count with direction $\theta_i$. First, the Rayleigh test establishes a mean resultant direction over the entire histogram:

For each count, or angle, $\Theta_i$ in a histogram of $N$ counts, let $c_i = \cos(\Theta_i)$ and $s_i = \sin(\Theta_i)$.

Under the Rayleigh test:

\[ C = \sum c_i \text{ for } i \in [1,N] \]
\[ S = \sum s_i \text{ for } i \in [1,N] \]
\[ R^2 = C^2 + S^2 \]
\[ R = \sqrt{R^2} \]

The significance of the distribution is assessed as follows:

\[ \hat{R} = R/n, \] where $n$ is the total number of counts over the histogram.

\[ z = n \ast (\hat{R}^2); \]
\[ p = e^{-z} \]

The value \( p \) ranging from \([0,1]\) determines the degree of presence of a strong mode, or preferential direction in the distribution. The null hypothesis \( H_0 \) of uniform distribution \((p=1)\) implies that no significant fluctuation of PIH cases occurs with variations in the rainfall cycles. Thus, small values of \( p \) indicate strong deviation from uniformity. Generally, significance of non-uniformity is accepted for \( p < .05 \) (Zar 1999).

However, the nature of Rayleigh’s test renders it unsuitable for application in the present study. Because the Rayleigh test, applied to a histogram, treats every count as a unique sample with assigned direction \( \theta \), the method is heavily impacted by the number of bins used in the analysis. For data binned into only twelve phase values (as determined by the resolution of monthly values provided by the time series analysis), this approach is statistically weak.

Instead, the Shannon Entropy can be used to characterize the non-uniformity of histograms with any given number of bins. Let \( S_{\text{max}} = \ln(n) \), where \( n \) is the number of bins in the histogram, and \( N \) is the total number of counts in the histogram. Then \( S = \sum [p(i) \ln p(i)] \) for \( i \in [1,N] \), where \( i \) reflects an individual bin and \( p(i) = \frac{\text{(number of counts per bin)}}{N} \). Shannon’s entropy, or the probability of deviation from uniform distribution is then \( p = \frac{S_{\text{max}} - S}{S_{\text{max}}} \). (Tass 1998)

c.) Hilbert Transform: The Hilbert transform can be used in signal processing to assign an instantaneous phase to a time-domain signal. The Hilbert transform is achieved for a time-domain signal \( s(t) \) via the transformation:
The real signal and its Hilbert transform are combined to form a new analytic signal, \( z(t) \) where:

\[
z(t) = s(t) + i\hat{s}(t) = A(t)e^{i\theta(t)}
\]

It follows that:

\[
\theta(t) = \arctan\left(\frac{\hat{s}(t)}{s(t)}\right) = \arctan\left(\frac{\text{Im}(z(t))}{\text{Re}(z(t))}\right).
\]

Note that the Fourier transform of \( z(t) \) is \( Z(f) \) where:

\[
Z(f) = S(f) + i\hat{S}(f) = S(f) + S(f)(\text{sgn}(f))
\]

\[
= \begin{cases} 
  2S(f) & \text{for } f > 0 \\
  0 & \text{for } f \leq 0
\end{cases}.
\]

(King 2009).

Therefore, in order to obtain a Hilbert transform for a signal \( s(t) \), we calculate the Fourier transform \( S(f) \), double positive frequencies and eliminate negative frequencies, and take the inverse Fourier transform to obtain \( z(t) \). By convention, phase assignment with a Hilbert transform yields a phase on the interval \([-\pi, \pi]\).

i. Methods:

a. Generation of Data Sets: Geographical maps of the districts of Uganda were generated in the Penn State University Maps library. We employed Digital Charts of the World, a product of the Environmental Systems Research Institute, to produce a digital Global Positioning System grid overlay as shown in Figure 1Aa. Localized rainfall estimates were derived from a mixture of ground and satellite data, assimilated using the African Rainfall Estimation Algorithm Version 2.0 (RFE2.0) developed at the NOAA Climate Prediction Center. The political map of Uganda
containing 77 political districts (Figure 3Aa) was condensed into a 10x10 spatial matrix (Figure 3Ab) for mathematical modeling, conserving adjacencies between districts. Black squares represent void spaces on the 10x10 matrix. A political map of Uganda, marking political districts, was overlain with a GPS grid in .1x.1 degree grid spacing. Each gridpoint-rainfall value was assigned to a political district, and an average district-rainfall value was calculated from the composite values. Numbers of cases by month and location were generated from a database of patients presenting for PIH. “Location of cases” was recorded as the political district of origin as listed in the database. Three different measures of “date of cases” were compared: date by month of birth, month of onset of febrile illness, and month of surgery. This served to validate perceived causational relationship between rainfall and case values: if indeed climate motivated infections preceding hydrocephalus, then the strongest mathematical correlation should be between rainfall pattern and date of onset of febrile illness. Birth month, which often accompanies infection (the mean time from birth to onset of febrile illness is .8 months) should be closely correlated, with date of presentation for surgery serving as the most statistically uncertain measure.

b. Data Transformation and Phase Assignment:

Phases were assigned for rainfall data using a Hilbert transform. Prior to phase assignment, data for each district were zero-phase filtered (both forward and then backwards) using a 5th order Chebyshev filter centered on 2 cycles per year (Figure 3Ca). Rainfall data was de-trended (the mean was set to zero) prior to analysis.

c. Histograms:
Histograms were generated in Matlab to determine corresponding case and rainfall distribution. Histograms were binned both by month of year and by phase of the rainfall cycle over a year-long period. Shannon entropy was used to characterize the nonuniformity of the phase distributions.

ii. Results:

Figure 1:

Figures 1Aa and 1Ab detail the spatial arrangement used to model the geography of Uganda and to assign location to GPS rainfall values. Figure 1Ba shows the overlay of raw rainfall average for each district over the 72-month time series. Within this raw data, a striking twice yearly cycle is visible. Figures 1Bb, 1Bc, and 1Bd display total cases (summed across spatial coordinates) across the time series for date of birth, date of preceding febrile illness, and
date of presentation for surgery. In addition, these figures display coherograms (generated by Dr. Steven Schiff using techniques programmed in Chronux, Version 2) comparing the correlation in frequencies between cases and rainfall data across the 72-month time series. These spectrograms are complex, suggesting that the total amount of rainfall may not be the determining factor of spread of PIH cases, but rather the underlying cycle of the climate pattern itself. Figure 1Ca presents total rainfall (mm) for the six-year data set distributed spatially, reflecting high overall amounts of rain. Figure 1Cb presents a complex distribution of cases (based on date of onset of febrile illness), presented by a log scale to facilitate visibility. Figure 1Cc presents a plot of cases as a function of total rainfall and a subsequent plot of residuals from the applied linear fit with increasing rainfall amounts. The large residual (R=.44 and R increasing with increased rainfall) suggests that number of cases, i.e. incidence of infection preceding PIH, does not increase linearly with amount of rainfall.
Figures 2Aa – 2Ac present histograms reflecting cases as a function of average monthly rainfall (mm) for cases based on date of onset of febrile illness, date of birth, and date of surgery. All three situations show heavy peaks in cases towards intermediate values of rainfall. Each histogram was fit with a normal distribution, reflecting identical mean values of rainfall for cases based on birth and onset of febrile illness, which are offset from the mean of cases based on surgery date. Figure 2Ba displays a histogram of average national rainfall values per month as a function of month of the year (i.e. January corresponds to month number one) summed over the six-year data set. Two yearly peaks in April and October are noticeable. Figures 2Ba, 2Bb, and 2Bc present histograms of total cases binned per month of year over the 6-year data set for cases...
based on date of onset of febrile illness, date of birth, and date of surgery, respectively. Shannon’s entropy values suggest significant non-uniformity (p<.05) in all three case-representations, with date of onset of febrile illness proving the most statistically-peaked or non-uniform (p=.0066). From a visual perspective, four peaks in febrile illness cases surround each rainfall peak, suggesting a relationship between phase of the rainfall cycle and case incidence.

In order to extract and label phase independently of month of year, rainfall for each district was narrow band filtered using a 5th-order Chebychev filter with permissible frequencies spanning 1.5Hz-2.5Hz. Phase was preserved by filtering forwards and backwards. The resulting curve was assigned a yearly phase from 0 to 4pi (with each 6-month rainy cycle representing a complete 2pi) as displayed in Figure Cb. Figures Da through Dd display rainfall values and case values for each febrile illness dates, birth dates, and date of surgery, based on phase of the rainfall cycle rather than based on month of year. Figure Da fits a sine wave to the average rainfall per phase value to reinforce the extracted filtered signal. Notably, febrile illness cases show a marked four peaks per year, at characteristic phase values in the rainfall cycle.

iii. Discussion: These findings provide evidence for a previously unknown link between rainfall and PIH. On a broader level, the implication of environmental conditions as key factors in the spread of neonatal infections opens the door to potential environmental public health initiatives to prevent preceding febrile infection and avoid PIH. Infection appears to occur at intermediate values of rainfall in four strong peaks over the yearly phase of the four-season rainy cycle. Four peaks in infection cluster around two strong peaks in rainfall. The detection of infection peaks at points flanking peaks in the rainfall cycle, presents a compelling and novel dependence of infection incidence on intermediate amounts rather than on extreme amounts of rainfall. The
most common seasonally-associated patterns of infection, (i.e. epidemics of meningococcal meningitis across the meningitis belt of sub-Saharan Africa) are closely tied to extremes in the seasonal climate spectrum (WHO 2008).
IV. CONCLUSIONS AND IMPLICATIONS FOR FUTURE STUDIES:

The present study characterized the bacterial spectrum present in the CSF of infants presenting for PIH in Uganda, and presented novel evidence for the correlation of seasonal rainfall patterns with incidence of PIH. From a bacteriological perspective, we found evidence of fragments of bacterial DNA in almost all patients, with a predominance of gram-negative enteric bacteria. Furthermore, the study suggests an aspect of seasonality in the species of bacteria linked to infection. On a clinical level, these results imply that PIH may be preventable among Ugandan infants. While these results associate PIH with environmental microbes, they do not establish a causal relationship between the identified bacteria and the onset of PIH. Identification of the exact spectrum of pathogens causing neonatal sepsis could pave the way for the development of antimicrobial treatments that could prevent the onset of hydrocephalus and eliminate the need for costly neurosurgical intervention.

The correlation of satellite-derived rainfall averages across Uganda with case data revealed a novel link between infantile PIH and environmental conditions. Cases of onset of febrile illness were analyzed as a function of phase assigned to the twice-yearly rainfall cycle: four strong peaks occurred in the febrile illness data, flanking each of the two peaks in rainfall. This is consistent with a comparison of febrile illness cases with average amount of rainfall, which suggested that cases peak at intermediate values of rainfall. Broadly, these results imply that PIH could be preventable even before the onset of febrile illness by interrupting the environmental mechanisms that allow infection to spread. Future studies could better characterize the properties of environmental media such as soil which likely aid in the spread of pathogens. Basic public health measures, aimed at the better separation of infants from such media, could provide life-saving protection from infection leading to PIH in impoverished
settings. Furthermore, the model of spatio-temporal correlation of climate data with disease distribution can be readily applied to studies of other infectious diseases for which the role of environmental interaction is poorly characterized.
V. REFERENCES


OBJECTIVE: To pursue collaborative research and clinical practice research to innovate in Infectious Diseases and Disease diagnostics as a physician scientist.

EDUCATION

The Pennsylvania State University

Graduation Anticipated May 2012
Schreyer Honors College
Primary Degree: Bachelor of Science in Engineering Science (Honors Program of the College of Engineering)
Secondary Degree: Bachelor of Science in French and Francophone Studies

PUBLICATIONS

• Article Title: "Association of Bacteria with Hydrocephalus in Ugandan Infants."
  Position: First Author (shared)
  Field: Microbiology and Infectious Disease
• Article Title: "Climate Drives Hydrocephalus in East Africa"
  Position: First Author (shared)
  Field: Neural Engineering
  Citation: Schiff SJ, Ranjeva SL, Sauer T, Warf BC, Climate Drives Hydrocephalus in East Africa, Journal of Nature Medicine, submitted

RESEARCH EXPERIENCE

Penn State Centers for Neural Engineering and Infectious Diseases
Infectious Disease Research: Project Head and Researcher

• Alongside Dr. Steven Schiff (Head of the PSU Center for Neural Engineering), conducted a spatio-temporal analysis of hydrocephalus across Uganda in correspondence with seasonal patterns.
• Between fall 2008 and fall 2009, analyzed the bacterial diversity across neonates affected by Post Infectious Hydrocephalus in Uganda. Extracted DNA from Cerebral Spinal Fluid across patients and analyzed the clustering of the resulting sequences to determine a prevalent bacterial agent

Institut Pasteur: Paris, France
Research Fellow

• Six-month laboratory fellowship exploring the host-pathogen interactions of Neisseria meningitidis bacteria

RECENT ACADEMIC AWARDS AND HONORS

2011:
-Barry M. Goldwater Scholarship (national prestigious scholarship for excellence in research)

2010:
-Penn State Finalist for Barry M. Goldwater Prestigious Scholarship
- Penn State Engineering Science Academic Excellence Scholarship
- Featured Presenter for PSU Engineering Science Global Initiatives Panel

2009:
-President's Freshman Award
-Engineering Science Fall 2009 Award for Academic Excellence
-Distinguished Presenter at the GlobeMed National Leadership Institute

2008:
-Schreyer Honors College George J. Coleman Alumni Memorial Award

-Dean’s List (All five semesters of academic study)