Using Mixed Effects Modeling to Quantify Difference Between Patient Groups with Diabetic Foot Ulcers

Rachel French
Western Kentucky University, rachel.french786@topper.wku.edu

Follow this and additional works at: https://digitalcommons.wku.edu/stu_hon_theses

Part of the Biology Commons, Mathematics Commons, and the Statistics and Probability Commons

Recommended Citation
https://digitalcommons.wku.edu/stu_hon_theses/723

This Thesis is brought to you for free and open access by TopSCHOLAR®. It has been accepted for inclusion in Honors College Capstone Experience/Thesis Projects by an authorized administrator of TopSCHOLAR®. For more information, please contact topscholar@wku.edu.
USING MIXED EFFECTS MODELING TO QUANTIFY DIFFERENCES BETWEEN PATIENT GROUPS WITH DIABETIC FOOT ULCERS

A Capstone Project Presented in Partial Fulfillment of the Requirements for the Degree Bachelor of Science with Honors College Graduate Distinction at Western Kentucky University

By
Rachel Noel French
December 2017

*****

CE/T Committee:
Professor Richard Schugart, Chair
Professor Ferhan Atici
Professor Christopher Keller
I dedicate this thesis to my parents, James and Darla French, who are great inspirations to me as lifelong learners. I also dedicate this work to my amazing fiancé James Lundy, who has supported me throughout my educational pursuits.
ACKNOWLEDGEMENTS

I would like to extend my gratitude to Dr. Richard Schugart who has helped to guide my research in this area. Under his leadership our research team has made great strides in many aspects of this wound-healing model.

I would also like to thank Western Kentucky University, the Honors College, and the Mathematics Department for the support contributed during my research through monetary resources, available materials, and valuable time to make this project possible. Throughout my research I have also been fortunate enough to benefit from various grants. The Faculty Undergraduate Student Engagement (FUSE) program at Western Kentucky University supported my project financially which allowed me to travel to and present at numerous conferences across the United States. Western Kentucky University’s Mahurin Honors College Honors Development Grant program also allowed me to travel to conferences as well as cover the charges of computing programs my research required.

I also benefitted greatly from the monetary support of the Kentucky Science and Engineering Foundation (KSEF). An Established Program to Stimulate Competitive Research (EPSCoR) grant from the National Science Foundation allowed me to continue my research throughout the summer months.

Thank you to everyone who has made this research possible.
ABSTRACT

When diabetes progresses, many patients suffer from chronic foot ulcers. In a study described in *Matrix Metalloproteinases and Diabetic Foot Ulcers* (Muller et al., 2008), sixteen patients with diabetic foot ulcers were examined throughout a twelve week healing period. During this period, levels of matrix metalloproteinases (MMP-1), their inhibitors (TIMP-1), and the extracellular matrix in a wound area were measured at distinct time intervals for each patient. The ratios of these healing components are vital in determining whether a wound will heal or become chronic and never properly heal.

*Connecting Local and Global Sensitivities in a Mathematical Model for Wound Healing* (Krishna et al., 2015) mathematically describes the healing interactions between the MMP-1, TIMP-1, the extracellular matrix, and fibroblasts to highlight key differences between those patients categorized as ‘good’ healers or as ‘poor’ healers in the *Matrix Metalloproteinases and Diabetic Foot Ulcers* (Muller et al., 2008) study.

The goal of this research is to utilize the individual patient data obtained from *Matrix Metalloproteinases and Diabetic Foot Ulcers* (Muller et al., 2008) to identify key parameters, through the use of nonlinear mixed effects modeling, a technique that allows for each parameter estimate to be split into a fixed and random effect. The fixed effect is assumed to remain the same across every data collection. Random effects vary from collection to collection and patient to patient. Through this split, information from the population trends, using the fixed effect estimates, can be utilized to help inform the individual patient estimates for patients with fewer data points. The identification of key parameters in the healing process can provide valuable insight about which parameters should be taken into special consideration during the diagnosis and treatment process.
VITA

EDUCATION

Western Kentucky University, Bowling Green, KY December 2017
B.S. in Mathematical Economics with a concentration in Actuarial Sciences and a
minor in Finance
Mahurin Honors College Graduate
Honors Capstone: Using Mixed Effects Modeling to Quantify Differences between
Patient Groups with Diabetic Foot Ulcers

Bowling Green High School, Bowling Green, KY May 2014

PROFESSIONAL EXPERIENCE

Mathematics Department, WKU August 2016-
Research Assistant Present

AWARDS & HONORS

Summa Cum Laude, WKU, December 2017
Omicron Delta Epsilon member, January 2017-present
Pi Mu Epsilon member, August 2017-present
National Society for Collegiate Scholars member, August 2016-present
Recipient of the N.O. Taff Scholarship 2017
Recipient of an Established Program to Stimulate Competitive Research (EPSCoR) grant
Recipient of a grant from the Kentucky Science and Engineering Foundation (KSEF)
Award of Excellence Academic Scholarship, WKU, 2014-2017

INTERNATIONAL EXPERIENCE

Harlaxton College, the British campus for University of Evansville January 2015-
Grantham, England May 2015
Australia and New Zealand study abroad program January 2017
Costa Rica Spanish Institute October 2012

PRESENTATIONS


CONTENTS

Acknowledgements ........................................................................................................ iv

Abstract ......................................................................................................................... v

Vita.................................................................................................................................... vi

List of Equations ............................................................................................................ x

List of Tables .................................................................................................................... xiii

List of Figures .................................................................................................................. xiii

Introduction ..................................................................................................................... 1

The Model ....................................................................................................................... 6

Methods ............................................................................................................................ 9

  *Nonlinear Mixed Effects Modeling* .............................................................................. 9

  *Stochastic Approximation Expectation Maximization Algorithm* ............................. 10

  *MatLab’s nlmefitsa Function* ................................................................................... 12

  *The Issues with nlmefitsa* ....................................................................................... 14

  *SimBiology® Toolbox for MatLab* ......................................................................... 15

Results .............................................................................................................................. 17

  *Preleminary nlmefitsa Results* .............................................................................. 15

  *Preleminary SimBiology® Results* .......................................................................... 25

Conclusions and Future Work ....................................................................................... 30

References ....................................................................................................................... 31
Appendix ........................................................................................................................................33

Appendix 1: Additional graphs from nlmefitsa run for $\tilde{f}$ .................................................. 33

Appendix 2: Additional graphs from nlmefitsa run for $M$ ....................................................... 34

Appendix 3: Additional graphs from nlmefitsa run for $T$ ....................................................... 35

Appendix 4: Additional graphs from the SimBiology® run for $\tilde{f}$ ......................................... 36

Appendix 5: Additional graphs from the SimBiology® run for $M, T, and E$ ......................... 38

Appendix 6: Additional graphs from the SimBiology® run for $M, T, E,$ and $\tilde{f}$ .......... 40
LIST OF EQUATIONS

Equation 1. Rescaled mathematical model describing wound healing................... 7
Equation 2. Original fibroblast differential equation................................. 7
Equation 3. Parameter estimate with fixed and random effects...................... 9
Equation 4. Lavielle’s description of the SAEM algorithm......................... 11
Equation 5. Fixed equations to fit for $\tilde{f}$ and estimate $k_{11}$...................... 18
LIST OF FIGURES

Figure 1. Example of run chains generated by *nlmefitsa* .................................................. 13
Figure 2. Fit for $\bar{f}$ using *nlmefitsa* ............................................................................. 19
Figure 3. Fit for $M$ using *nlmefitsa* .................................................................................. 22
Figure 4. Fit for $T$ using *nlmefitsa* .................................................................................. 25
Figure 5. Fit for $\bar{f}$ using SimBiology\textsuperscript{®} ........................................................... 26
Figure 6. Fit for $M, T,$ and $E$ using SimBiology\textsuperscript{®} .............................................. 27
Figure 7. Fit for $M, T, E,$ and $\bar{f}$ using SimBiology\textsuperscript{®} ..................................... 29
LIST OF TABLES

Table 1. Random effect estimates for $k_{11}$ from the $\tilde{f}$ nlmefitsa run............................... 18
Table 2. Fixed effects for $k_3, k_4, k_6, k_7,$ and $k_8$ from the $M$ nlmefitsa run...................... 20
Table 3. Individual estimates for $k_3, k_4, k_6, k_7,$ and $k_8$ from the $M$ nlmefitsa run .......... 21
Table 4. Fixed effects for $k_3, k_4, k_6, k_7,$ and $k_8$ from the $T$ nlmefitsa run............................ 23
Table 5. Individual estimates for $k_3, k_4, k_6, k_7,$ and $k_8$ from the $T$ nlmefitsa run .......... 24
Table 6. Fixed effects for $k_1, k_2,$ and $k_8$ from the $M, T, E$ SimBiology® run...................... 27
Table 7. Fixed effects for $k_3, k_4, k_6, k_7, k_8$ and $f_i$ from the $M, T, E, \tilde{f}$ SimBiology® run. 28
Table 8. Individual estimates for $k_8$ and $f_i$ from the $M, T, E, \tilde{f}$ SimBiology® run .. 28
INTRODUCTION

The wound-healing process is often categorized into four stages: coagulation and hemostasis, inflammation, proliferation, and wound remodeling (Velnar et al., 2009). The coagulation and hemostasis phase occurs immediately after an injury is sustained in order to stop blood loss. During this phase, a matrix is created which will allow cells needed later in the healing process access to the wound area. There is a complex balance between endothelial cells, thrombocytes, coagulation, and fibrinolysis that dictates hemostasis and the overall timeline of the healing process. Due to the body’s reflex mechanisms, “injured vessels constrict rapidly due to contraction of vascular smooth muscle cells in the circular muscle layer” (Velnar et al., 2009). Unfortunately, this contraction only lasts for a few minutes before hypoxia and acidosis cause the wound wall and surrounding vessels to relax allowing bleeding to resume. Thankfully, hemostasis deposits enough fibrin at the wound site that an insoluble fibrin plug develops to stop the bleeding. During coagulation, platelet aggregation and clot formation occur to limit blood loss. Clots form due to complex interactions between platelets and exposed collagen or other parts of the extracellular matrix. These clots are composed of fibronectin, fibrin, vitronectin, and thrombospondin. The cytoplasm of the platelets trapped in the blood clots contain various growth factors, such as platelet derived growth factor (PDGF), transforming growth factor-β (TGF-β), and epidermal growth factor, that activate and attract neutrophils, macrophages, endothelial cells, and fibroblasts, important components of the future healing stages.

After the coagulation and hemostasis stage, the humoral and cellular inflammatory phase is next, “with the aim of establishing an immune barrier against invading microorganisms” (Velnar et al., 2009). This phase is often split into an early and late
inflammatory stage. Early inflammatory response begins during the last part of coagulation. During this phase, the neutrophils that were previously attracted to the wound site by the growth factors start phagocytosis, a process that destroys and removes bacteria, foreign particles, and damaged tissue by releasing proteolytic enzymes and oxygen-derived free radical species. “Phagocytic activity is crucial for the subsequent processes, because acute wounds that have a bacterial imbalance will not heal” (Velnar et al., 2009). The attraction of neutrophils occurs within 24-36 hours of injury. However after their job is complete and all the bacteria has been removed, neutrophils are destroyed and their remnants are phagocytosed by macrophages in the next step. Then, the late inflammatory phase occurs between 48-72 hours after injury. This is when blood monocytes go through phenotypic changes to become tissue macrophages. These cells “have a longer lifespan than neutrophils and continue to work at a lower pH” (Velnar et al., 2009). The tissue macrophages regulate the inflammatory stage and provide various tissue growth factors to the wound site allowing healthy tissue to grow and heal the injury. A lack of monocytes or macrophages at the wound causes serious delays in wound healing because of “poor wound debridement, delayed fibroblast proliferation and maturation, as well as delayed angiogenesis” (Velnar et al., 2009). Towards the end of the late inflammatory stage lymphocytes, key factors in the development of a strong immune system, are attracted to the wound area around 72 hours after injury by interleukin-1 (IL-1), which are important components of collagen remodeling and the production and repair of the extracellular matrix.

The next stage of the wound healing-process is the proliferative phase. This is the stage where most of the tissue repair occurs. The proliferation stage generally lasts for
about 2 weeks after injury. This intricate stage encompasses multiple processes: fibroblast migration, collagen synthesis, angiogenesis and granulation tissue formation, protrusion, adhesion, and traction. Fibroblast migration typically occurs within the first 3 days of the proliferative phase. Fibroblasts flood the wound site as they are attracted by TGF-β and PDGF in order to produce the matrix proteins hyaluronan, fibronectin, proteoglycans, and type-1 and type-3 procollagen and deposit them in the local milieu (Velnar et al., 2009). After one week, fibroblasts have generated enough extracellular matrix components to promote cell migration and the repair process. Once there is enough extracellular matrix, fibroblasts change to their myofibroblast phenotype where they attach to fibronectin and collagen in the extracellular matrix to extend the wound edges closer together, towards a closed wound. Afterwards, fibroblasts that have accomplished their job and are no longer useful are flushed out of the wound through apoptosis. Then collagen synthesis occurs in which fibroblast cells synthesize collagen such that the newly synthesized collagen can “impart integrity and strength to all tissues” (Velnar et al., 2009) acting as a foundation for healing. Next, angiogenesis begins to remodel old and generate new blood vessels. Since there is no viable tissue at the wound center, all the new blood vessels are concentrated around the wound margins. After a few days while the collagen grows toward the wound center, new capillary networks extend toward the wound center and connect with capillary sprouts from the edges of the wound clot. The last three parts of the proliferative phase, protrusion, adhesion, and traction, are required to ensure cellular motility as well as movement and migration of cells from one location to another through the combustion of energy.
The final stage of wound healing is the remodeling phase which is “responsible for the development of new epithelium and final scar tissue formation” (Velnar et al., 2009). This phase of the wound healing process can last up to two years. During the remodeling phase of an acute wound, there are many regulatory mechanisms that maintain the strict equilibrium between degradation and synthesis. Though the entirety of the remodeling process may take years, the remodeling of the extracellular matrix around the wound site becomes rather stable about three weeks after injury. Matrix metalloproteinase enzymes breakdown collagen within the extracellular matrix that was damaged when the injury occurred. Tissue inhibitors of metalloproteinases regulated the activity of matrix metalloproteinases. “Gradually, the activity of tissue inhibitors of metalloproteinases increases, culminating in a drop in activity of metalloproteinase enzymes, thereby promoting new matrix accumulation.” (Velnar et al., 2009). In order to close the wound, “underlying connective tissue shrinks in size and brings the wound margins closer together, owing to fibroblast interactions with the extracellular matrix” (Velnar, 2009). As the wound heals, fibroblasts and macrophages exit the wound area through apoptosis, the growth of capillaries stops, and blood flow to the area declines. “The end result is a fully matured scar with a decreased number of cells and blood vessels and a high tensile strength” (Velnar, 2009).

Acute wounds are wounds that follow the healing process described above. Chronic wounds do not follow this typical pattern of healing, often leading to prolonged healing and open wounds. They have been described as “wounds that have failed to return to functional and anatomical integrity in a timely fashion” (Telgenhoff et al., 2005). Most chronic wounds are either pressure sores, diabetic ulcers, or venous ulcers. The fact that
these wounds do not follow the standard healing process typically stems from one of three factors: the cellular and systemic effects of aging, repeated ischemia-reperfusion injury, or bacterial contamination resulting in an inflammatory response (Telgenhoff et al., 2005). Most chronic wounds can be healed, the process simply takes longer and more medical attention than an acute wound. However there is an estimated 15-20% of chronic wound patients that suffer with non-healing wounds. These wounds typically display “decreased growth factors, decreased keratinocyte migration, increased reactive oxygen species, increased tissue proteases, and microbial contamination” (Telgenhoff et al., 2005). Diabetic foot ulcers are the most costly chronic wounds to care for and the most likely to become a non-healing chronic wound, often leading to amputation (Apelquist et al., 1993).

This paper will specifically be examining diabetic foot ulcers. It was found in a study by Lobmann, et al (2002) that when comparing diabetic chronic wounds and nondiabetic acute wounds, the concentrations of matrix metalloproteinases were significantly higher in diabetic foot ulcers. Specifically, the concentration of MMP-1, a type of matrix metalloproteinases, was increased 65-fold in diabetic wounds than in acute, traumatic wounds. Similarly, MMP-2 concentrations were increased threefold, twofold for MMP-8 and 14-fold for MMP-9. It was also found that TIMP-2, the tissue inhibitor for MMP-2, was reduced twofold in the diabetic ulcers. This combination of higher concentrations of MMP’s and lower concentrations of TIMP’s means that the tissue inhibitors are outnumbered and cannot stop the matrix metalloproteinase enzymes from continued destruction of collagen and other parts of the extracellular matrix. Therefore new matrix accumulation never occurs and the wound will not be able to close, leading to a non-healing wound.
THE MODEL

In order to examine the healing process of chronic wounds, a model must be instituted such that variability in the data can be attributed to specific parameters. The model that this paper will be using is a modified version of the model that can be found in the Krishna, et al. (2015) paper *Connecting Local and Global Sensitivities in a Mathematical Model for Wound Healing*. This model uses four state variables and twelve parameters in four differential equations to describe the wound-healing process explained above. Specifically, the model examines concentrations of MMP-1 and TIMP-1 as well as extracellular matrix levels and fibroblast cell counts; these are the four state variables. The model was designed using data collected from chronic diabetic foot ulcers of sixteen type-2 diabetic patients (Muller et al., 2008). MMP-1 and TIMP-1 concentrations were measured using wound fluid collected from absorbent paper strips placed on the edge of the wound for five minutes. Because there was a possibility of variant amounts of fluid collected from each wound, data for MMP-1 and TIMP-1 were recorded as ratios of the number of MMP-1 or TIMP-1 proteins to the total number of proteins collected. Extracellular matrix (ECM) data was measured according to wound area which was examined through numeric photography and analytical software. “ECM levels were taken to be proportional to wound closure” (Krishna et al., 2015) such that the largest ECM value of 1 corresponds to a fully closed wound. Though there is no data in the Muller et al. study for the fibroblast concentrations, fibroblast cells are important to the wound-healing process, since they are the cells that produce the proteins needed for wound healing, and have therefore been included in the model. There were modifications made to the equations because the original model was created using average patient data. Now that de-identified
individual patient data are available some changes were made to combat parameter identifiability issues and to better represent the individual patient data. With all of these changes, the new model is:

\[
\frac{dM}{dt} = \frac{k_1M^3(\tilde{f} + f_i)}{k_2^3 + M^3} - k_3M - k_4MT
\]

\[
\frac{dT}{dt} = \frac{k_5T^3(\tilde{f} + f_i)M}{k_6^3 + T^3} - k_7T - k_4MT
\]

\[
\frac{dE}{dt} = k_8(\tilde{f} + f_i)(1 - E) - k_9ME - k_{10}E
\]

\[
\frac{df}{dt} = k_{11}(\tilde{f} + f_i)[1 - (\tilde{f} + f_i)]
\]

Equation 1: Mathematical model describing wound healing

The four state variables matrix metalloproteinases (MMP-1), tissue inhibitors of matrix metalloproteinases (TIMP-1), extracellular matrix (ECM), and fibroblasts are represented by \(M, T, E,\) and \(\tilde{f},\) respectively. The twelve parameters are \(k_1\) through \(k_{11}\) and \(f_i,\) which is the initial concentration of fibroblasts.

The three main rescalings performed on the data and the model to more adequately describe the individual patient data are updated initial TIMP-1 concentration, changed initial condition of the ECM, and the scaling out of the death term, \(k_{12},\) in the original

\[
\frac{df}{dt} = k_{11}f(1 - f) - k_{12}f
\]

Equation 2: Original fibroblast differential equation
fibroblast differential equation shown in Equation 2. The original MMP-1 and TIMP-1 data was divided by the total number of proteins collected from the wound site. Then, the MMP-1 and TIMP-1 data was rescaled for the Krishna et al. paper by dividing the MMP-1 and TIMP-1 concentrations by the median initial concentration of TIMP-1 for good healers or poor healers, depending upon which group was being modeled. Now through the rescaling with the individual patient data, the MMP-1 and TIMP-1 concentration data is divided by the average initial concentration of TIMP-1 for all patients. The next change to the original model rescales the initial condition of the ECM data. In the Krishna et al. paper, the ECM data was rescaled such that 0 represented the initial wound size and 1 represented a fully closed and healed wound. When the individual patient data was acquired, it was found that some patient’s wound areas increased from the initial measurement before the wound-closing process occurred. To avoid negative ECM levels, the ECM data was rescaled such that the ceiling function of the largest wound area became 0, such that

\[ \text{ceiling(largest wound area)} \overset{\text{rescaled}}{\rightarrow} 0. \]

This method allowed for the various ECM initial points of the individual patients to be graphed showing the variation in initial wound area from patient to patient. Patients with ECM levels reaching 1 by the end of the 12 weeks are still categorized as having closed wounds.

The third main difference between the original model and the updated model is scaling \( k_{12} \) out of the equation due to structural identifiability issues. Structural identifiability analyzes which parameters can be uniquely determined with an idealized, noiseless data set (Eisenberg and Hayashi, 2014). The four differential equations could not be solved for 13 unique parameters, \( k_1 \) through \( k_{12} \) and \( f_i \); the least informative parameter
needed to be eliminated. After individual data was examined, it was determined that a key factor in the individual wound healing process was the initial concentration of fibroblast cells. To acknowledge this variation in the model, $f$ was defined as $\tilde{f} + f_i$. This change also allows the carrying capacity of the $\frac{df}{dt}$ function to vary from patient to patient, such that each individual’s carrying capacity is equal to $1 - f_i$. After these three rescalings were performed, the new model shown in Equation 1 was obtained.
METHODS

Nonlinear Mixed Effects Modeling

This paper utilizes nonlinear mixed effects modeling to estimate parameters and examine variation in the data. Mixed effects modeling is a technique that is well suited for sparse data and is characterized as containing fixed effects and random effects. Fixed effects are population parameters that are assumed to remain constant each time data is collected. Random effects, contrastingly, are sample-dependent variables that vary from patient to patient. The combination of fixed and random effects is such that

$$k_{i,j} = e^{(\beta_i + \phi_{i,j})}, \ i=1,\ldots,12, \ j=1,\ldots,13$$

Equation 3: parameter estimate with fixed and random effects

where $k_{i,j}$ is the estimated value for the $i^{th}$ parameter for the $j^{th}$ patient, $\beta_i$ is the fixed effect for the $i^{th}$ parameter for all patients, and $\phi_{i,j}$ is the random effect for the $i^{th}$ parameter for the $j^{th}$ patient. The exponential equation was chosen to ensure that the parameter values will remain positive for all estimations.

Mixed effects modeling is particularly well suited for situations with sparse data because this technique uses population trends in the data to inform individual curve fits. This idea fits not only this model but many others since “the notion that individuals’ responses all follow a similar functional form with parameters that vary among individuals seems to be appropriate in many situations” (Lindstrom and Bates, 1990). Mixed effects modeling is attractive because of its ability to map high levels of variation within a model, including splitting variation into within- and between-individual components. This
technique is used frequently in the life sciences, specifically in pharmacokinetics, an industry plagued with sparse data due to the difficulty of establishing medical trials.

*Stochastic Approximation Expectation Maximization algorithm*

A new development in mixed effects modeling incorporates the stochastic approximation expectation maximization (SAEM) algorithm which has “proven very efficient, quickly converging to the maximum likelihood estimators and performing better than linearization-based algorithms” (Comets et al., 2017). A common issue found with linearization-based algorithms is the increase in type I errors of likelihood tests that often result due to the lack of knowledge of the transition probability density. To combat this issue, SAEM algorithm codes have various ways of approximating the transition density. This paper specifically uses Markov Chain Monte Carlo (MCMC) algorithms to approximate the transition probability. The traditional method for mixed effects modeling involves solving for the expected value of the loglikelihood function and then maximizing that function for each iteration. The SAEM algorithm proposes a process that eliminates the requirement of solving for the expected value of the loglikelihood function by approximating this function through stochastic approximation of the loglikelihood function informed by MCMC estimates of the transition probability density. This approximated expected value of the loglikelihood function is then maximized and the estimate for the parameter is updated. A generic example of this process, as explained by Marc Lavielle

```
and his collaborators in their 2010 presentation *SAEM algorithm: a powerful stochastic algorithm for population pharmacology modeling*, is displayed below.

Let $\pi_\theta$ be the transition probability of an ergodic markov chain with the limiting distribution $p_{\psi|y}(\cdot|y;\theta)$. Steps one through three are performed for each iteration $k$.

**Step 1 – simulation**: draw $\psi^{(k)}$ according to the transition probability $\pi_{\theta_{k-1}}(\psi^{(k-1)}, \cdot)$.

**Step 2 – stochastic approximation**:

$$Q_k(\theta) = Q_{k-1}(\theta) + \gamma_k[\log p(y, \psi^{(k)}; \theta) - Q_{k-1}(\theta)]$$

where $\gamma_k$ is a decreasing sequence: $\sum \gamma_k = +\infty$, $\sum (\gamma_k)^2 < +\infty$

**Step 3 – maximization**:

$$\theta_k = \text{Argmax } Q_k(\theta)$$

Equation 4: Lavielle’s description of SAEM algorithm

where $\psi$ is a vector of random effect estimates, $\theta$ is a vector of estimated population parameters, $p(y; \theta) = \prod_{i=1}^N p(y_i; \theta)$ is the likelihood function, and $n$ is the number of patients (Lavielle, et al., 2010). The simulation step generates simulated values of the random effects given the current parameter estimates according to the transition probability function and the posterior density. Then, the stochastic approximation “update[s] the expected value of the loglikelihood function by taking its value from the previous step, and moving part way toward the average value of the logliklihood calculated from the simulated random effects” (MathWorks, 2017). The last step in the process, the
maximization step, chooses the parameter estimates for the next iteration based on values that maximize the $Q_k(\theta)$ function.

**MATLAB’s nlmefitsa function**

This paper uses the MATLAB function *nlmefitsa*, nonlinear mixed effects fit of the data using the stochastic algorithm. This built-in function requires inputs of an $n$-by-$h$ matrix of $n$ observations on $h$ predictor variables, an $n$-by-$1$ vector of responses, a group variable indicating from which patient the observation was taken, and a model function. For this model, we also utilize a group-specific predictor variable matrix to assign patient specific initial values to be used in the ordinary differential equation (ODE) solver. The *nlmefitsa* function starts the process with user-selected initial guesses ($\Phi_0$) of the fixed effects for the estimated parameters. For this paper, the original $\Phi_0$ guesses were taken from average, curve-fitted values for the estimated parameters provided through work completed by other members of the research group using the process described in the Krishna et al. 2015 paper applied to individual patient values rather than average values. After passing the $\Phi_0$ values through to the model function and solving the ODE system with these parameter estimates, *nlmefitsa* records the differences between the solved-for values and the recorded data for a specified state variable. Following this initial run, *nlmefitsa* utilizes the SAEM algorithm to identify new, more accurate parameter values to be used in the next iteration.

In the default programming of this function, each iteration utilizes six MCMC algorithms split such that the first two MCMC iterations provide full multivariate updates, the next two provide single coordinate updates, and the last two provide multiple
coordinate updates. As the mixed effects modeling iterations are being completed, \( \text{nlmefitsa} \) displays run chains such as those shown in Figure 1. The \( \beta_1 \) seen in Figure 1 represents the estimated fixed effects for one parameter for iterations 1 through 450. The estimated fixed effect in the first iteration is simply the user-entered initial guess \( \Phi_0 \).

Figure 1: Example of run chains generated by \( \text{nlmefitsa} \)

However by the 450\textsuperscript{th} iteration, a new, more accurate fixed effects value has been converged upon. The \( \Psi_{11} \) variable graphed in Figure 1 is the variance of the random effects for the estimated parameter. Preferably, the variance would be large enough to allow the parameter estimate to adequately explain the variation in the data but also small enough to indicate an accurate fixed effect was chosen. If a covariance matrix was designed as an input for \( \text{nlmefitsa} \), covariances between two parameters’ random effects will also be
shown in the output run chains. Figure 1 shows that an adequate parameter estimate has been calculated by the function because the estimated values converge upon a single value. This pattern often occurs at the end of the iteration cycle because the last third of the iterations uses a reduced step-size in the algorithm when picking a new fixed value estimate. The first third of the iterations combats poor initial estimates with a simulated annealing algorithm, a probabilistic technique for approximating a global optimum of a given function over a large space. In between the simulated annealing and reduced step size iterations is the middle third of iterations which simply uses a full step-size process.

The other outputs of *nlmefitsa* are estimates of the fixed effects, variance and covariance estimations for the random effects, the Akaike information criterion, the Bayesian information criterion, the standard errors for the fixed effects estimates, the error degrees of freedom, and the population, conditional, and individual weighted residuals.

**Issues with nlmefitsa**

While *nlmefitsa* is a useful function and readily available to implement through MATLAB, there are certain drawbacks to this generic, built-in function. First, only local, unconstrained optimizers are available to be selected as the optimization method, *fminsearch* or *fminunc*. Since the model describes a biological process, there are certain bounds in which a parameter value is biologically feasible. The limitation of a local optimizer is that a constrained parameter space cannot be specified. Constrained optimizer are available in MATLAB, such as *fmincon*. However their implementation is not compatible with *nlmefitsa* at this time. Because of this issue, *nlmefitsa* may return parameters values that are outside the realm of biological feasibility; this constrained space
has been set as [0,200] for parameters $k_1$ though $k_{11}$ and [0,1] for $f_i$. The value of 200 has been set due to a combination of biological reasoning and practical parameter value estimations. Due to biological circumstances, it is unreasonable for $k$ values to become arbitrarily large. The specific value of 200 was selected through a Latin Hypercube Sampling and Partial Rank Correlation Coefficient process which showed that even when the upper bound was increased to 500 or 1,000 the large majority of parameter estimates occurred in the [0,200] range. Using 200 as an upper limit rather than 500 or 1,000 enables the algorithms to sample a smaller space, saving computing time, while maintaining the same confidence in the results. In attempt to resolve the local search function issue, a minimum and a maximum function was inserted into the ODE system around the estimated parameters such that $max(0, min(k_1,200))$ would replace $k_1$ in the model function. Essentially this function returns 200 if the parameter estimate computed by *nlmefitsa* is greater than 200 and 0 if the estimate is less than 0. While this does not manipulate the optimization function to generate more accurate estimates, it does allow us to take the biologically feasible parameter space into account without modifying the complex *nlmefitsa* built-in coding. However, issues do arise with parameter estimates outside of this range since all numbers outside of the [0, 200] range enter either 0 or 200 into the ODE system. This does not allow the mixed effects modeling system to witness the variations in solutions.

The second issue with *nlmefitsa* is the limitation in recognizing only a single vector of responses. This means that only one of the model’s four state variables can be fit at a time. This issue also means that the other three state variables not being fit during the estimation process must be fixed to some central value for all patients. Fixing such crucial
components of the model reduces the amount of information that can be extrapolated from the results. The fixed state values were calculated as the average $M$, $T$, $E$, and $\hat{f}$ values of all patients over all time points. There was no modification available for the built-in \textit{nlmefitsa} function found that would provide the opportunity to fit for multiple response variables at the same time.

\textit{SimBiology$^\text{®}$ Toolbox for MATLAB}

A solution found for the one state variable limitation in \textit{nlmefitsa} was the nonlinear mixed effects modeling option in the SimBiology$^\text{®}$ Toolbox addition available for MATLAB. SimBiology$^\text{®}$ “provides an app and programmatic tools to model, simulate, and analyze dynamic systems, focusing on pharmacokinetic/pharmacodynamics (PK/PD) and systems biology applications” (MathWorks, 2017). Since SimBiology$^\text{®}$ is specifically designed to handle biological systems, this process offers more promising and accurate results for the data fitting. SimBiology$^\text{®}$ also handles complex systems, necessary when describing biological systems, more efficiently than \textit{nlmefitsa} through built-in simulation accelerators, ODE solvers, and stochastic algorithms. The nonlinear mixed effects modeling function in SimBiology$^\text{®}$ offers a better suited system to fit the model as well as the ability to fit multiple state variables at the same time. In addition, SimBiology$^\text{®}$ offers a simplified, user-friendly interface which allows for easy changes and adjustments to the model. This system uses the same nonlinear mixed effects modeling process described above with the addition of a modification that enables rapid simulation such that each iteration of the mixed effects modeling process proceeds quickly. Because of this addition, the SimBiology$^\text{®}$, even when fitting for all four state variables, delivers results in an eighth
of the time that the *nlme*fitsa* function does. Because of the significant time difference and more robust results, SimBiology® results will be focused on more heavily in this paper than the results achieved through *nlme*fitsa.
RESULTS

Preliminary nlmefitsa Results

Due to the time expensive results from nlmefitsa, only three fits have been included in this paper. While there may be some inconsistencies regarding which and how many parameters were estimated, the time cost of rerunning these simulations would take away from more informative results. From searching the latest articles, it was found that nonlinear mixed effects modeling, specifically using MATLAB’s nlmefitsa, is an underutilized tool for ODE systems with multiple response variables. Because of this lack of information, developing the nlmefitsa code was challenging. Therefore, the simplest equation in the system, \( \frac{df}{dt} \), was fitted first. Figure 2 is the fit obtained for the \( f \) variable when all parameters were fixed except for \( k_{11} \). Even \( f_i \) was fixed to previously curve fitted values for each patient. Each color represents a different patient according to the given key. The points are simulated \( f \) data using previously curve fitted \( k_{11} \) and \( f_i \) values substituted into the \( f(t) \) equation and then numerically solved at the corresponding time points. Noise was also added to the simulated data using a Gaussian curve with a mean of 0 and a variance of 0.1. The corresponding colored lines are the curve fits for individual patients. The black line represents a population fit which is the solution for the system when no random effects are added to the equation, such that \( k_{11} = e^{(\beta_{11})} \). This is a fit of Equation 5 where parameter represents a parameter that has been fixed to an average of the curve fitted patient data, \( \beta_{11} \) is the fixed effects of \( k_{11} \), and \( \phi_{11,j} \) is the random effect for \( k_{11} \) for patient \( j \). Because eleven of the twelve parameters and three of the four state variables were fixed,
the parameter estimate and state variable fit returned from this run leaves much of the variation of the data unexplained.

\[
\frac{dM}{dt} = \frac{k_1 M^3 (\tilde{f} + \tilde{f}_i)}{k_3^3 + M^3} - \frac{k_3 M}{k_3} - \frac{k_4 M T}{k_3}
\]

\[
\frac{dT}{dt} = \frac{k_5 T^3 (\tilde{f} + \tilde{f}_i) M}{k_6^3 + T^3} - \frac{k_7 T}{k_6} - \frac{k_4 M T}{k_6}
\]

\[
\frac{dE}{dt} = \frac{k_8 (\tilde{f} + \tilde{f}_i)(1 - \tilde{E})}{k_9 M E - k_{10} E} - \frac{k_9 M E}{k_9}
\]

\[
\frac{d\tilde{f}}{dt} = (e^{(\beta_{11} + \phi_{11,j})})(\tilde{f} + \tilde{f}_i)[1 - (\tilde{f} + \tilde{f}_i)]
\]

Equation 5: Fixed equations to fit for \(\tilde{f}\) and estimate \(k_{11}\)

<table>
<thead>
<tr>
<th>Patient</th>
<th>(\phi_{11})</th>
<th>(e^{\phi_{11}})</th>
<th>(k_{11} = e^{(\beta_{11} + \phi_{11,j})} = e^{(\beta_{11})} \times e^{(\phi_{11})})</th>
<th>Previously curvefitted</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0829</td>
<td>1.0864</td>
<td>8.0420</td>
<td>5.2866</td>
</tr>
<tr>
<td>2</td>
<td>6.2337\times10^{-4}</td>
<td>1.0006</td>
<td>7.4070</td>
<td>81.4774</td>
</tr>
<tr>
<td>3</td>
<td>-0.0023</td>
<td>0.9977</td>
<td>7.3850</td>
<td>49.2129</td>
</tr>
<tr>
<td>4</td>
<td>7.1706\times10^{-5}</td>
<td>1.0001</td>
<td>7.4029</td>
<td>10.1484</td>
</tr>
<tr>
<td>5</td>
<td>0.0757</td>
<td>1.0787</td>
<td>7.9847</td>
<td>13.5138</td>
</tr>
<tr>
<td>6</td>
<td>-1.8095</td>
<td>0.1637</td>
<td>1.2121</td>
<td>10.9497</td>
</tr>
<tr>
<td>7</td>
<td>-2.5644</td>
<td>0.0770</td>
<td>0.5698</td>
<td>0.6609</td>
</tr>
<tr>
<td>8</td>
<td>-2.9677</td>
<td>0.0514</td>
<td>0.3806</td>
<td>0.2660</td>
</tr>
<tr>
<td>9</td>
<td>3.5590\times10^{-4}</td>
<td>1.0004</td>
<td>7.4050</td>
<td>37.7009</td>
</tr>
<tr>
<td>11</td>
<td>-2.7371</td>
<td>0.0648</td>
<td>0.4794</td>
<td>0.5006</td>
</tr>
<tr>
<td>12</td>
<td>8.7074\times10^{-4}</td>
<td>1.0009</td>
<td>7.4088</td>
<td>74.8162</td>
</tr>
<tr>
<td>14</td>
<td>-0.9372</td>
<td>0.3917</td>
<td>2.8997</td>
<td>5.8088</td>
</tr>
<tr>
<td>15</td>
<td>0.1501</td>
<td>1.1619</td>
<td>8.6011</td>
<td>49.5219</td>
</tr>
</tbody>
</table>

Table 1: Random effect estimates for \(k_{11}\) from the \(\tilde{f} nlmefitsa\) run
The fixed effect of $k_{11}, e^{(\beta_{11})}$ was found to be 7.4024 through this run of `nlmefitsa`. The random effects and $k_{11}$ estimates for each patient can be found in Table 1. The $k_{11}$ estimates for patients 7, 8, and 11 are very close to previously curve fitted values. Due to the other fixed state variables, the $k_{11}$ estimates were prevented from reaching large enough values to match previous curve fitted values for the rest of the patients. The run chains for this fit were used as the example run chains in Figure 1 above. This figure also shows the variance of the random effects, $\phi_{11}$, which was calculated as 7.2437. This is a fairly large variance considering the fixed effect is 7.4024. The root mean squared error of this nonlinear fit is 0.0654, suggesting that the fit is relatively helpful for predictions of Figure 2: Fit for $\tilde{f}$ using `nlmefitsa`
because the mean squared error is close to 0. Another way to evaluate the fit is graphical evidence shown in Figure 2. The individual curve fits generally match the data trend of the individual simulated data. The population curve, the black line, shows an initial population value around 0.45. The fact that the population curve follows the trend of the majority of individuals indicates that the fit ran properly and that the population curve is an adequate representation of the population. It is important to note the three patients whose concentrations of fibroblast cells do not increase as quickly as the others; those patients are 7, 8, and 11. The fibroblasts may have encountered obstacles when migrating to these patients’ wound site due to a slow healing response, perhaps due to high levels of MMPs or low levels of TIMPs. The average residual for data points for patient 7 is 0.0134. The average residual for patient 8 is 0.0061. It is 0.0047 for patient 11. These rather low average residuals for patient 7, 8, and 11 suggest that the model decently captures the variations in the data points for these three patients. Since the model offers low residuals even for the three patients varying the most for the general population trend, it can be inferred that decent fits can be obtained for most patients.

The next run accomplished with *nlmefitsa* was the fit for $M$, MMP-1. For this run, all parameters, except for $k_3, k_4, k_6, k_7$, and $k_8$, were fixed to the curve fitted values for patient 7. The estimated fixed effects can be found in Table 2. Since the minimum-maximum function has been put in place, the fixed values for $k_7$ and $k_8$ which are clearly

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$k_3$</th>
<th>$k_4$</th>
<th>$k_6$</th>
<th>$k_7$</th>
<th>$k_8$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed effect</td>
<td>$1.991 \times 10^{-33}$</td>
<td>0.0024</td>
<td>$3.548 \times 10^{-5}$</td>
<td>$2.902 \times 10^{4}$</td>
<td>$7.511 \times 10^{5}$</td>
</tr>
</tbody>
</table>

Table 2: Fixed effects for $k_3, k_4, k_6, k_7$, and $k_8$ from the $M \ nlmefitsa$ run
above the 200 threshold are actually entered into the ODE as 200. Therefore, the fixed effects for $k_7$ and $k_8$ are actually 200. The fixed effect for $k_3$ is extremely close to 0. This is concerning as it means that the population trend has no decay rate for the matrix metalloproteinases, represented by $k_3$ in the model. However, this does follow the Lobmann et al. study which shows that the levels of MMP-1 in diabetic chronic wounds is 65-fold larger than the levels in acute wounds. Since the death rate is very small, matrix metalloproteinases stay longer at the wound site and break apart more sections of the ECM, preventing the wound from healing. The decay rate of the TIMP-1 is $k_7$. Thus a large value for $k_7$ also follows Lobmann et al.’s findings that the concentration of TIMP-1 is decreased twofold in diabetic chronic wounds compared to acute wounds. A large $k_8$ value indicates a large ECM production rate and promotes wound healing. This value may represent an overproduction of ECM to combat the increased concentrations of MMP proteins.

The estimated parameter values for each patient are shown in Table 3. Beside each estimate from the nlmefitsa code is the previously curve fitted values to indicate whether or not the estimates are similar. The nlmefitsa values do not seem to capture the variability of the model. Almost none of the nlmefitsa values are of the same order of magnitude as the previously curve fitted values. This is most likely because the fminsearch function has allowed the system to converge to values that are not in the biologically feasible range. The implementation of a constrained optimizer would deliver better results. The run chains for this nlmefitsa run are shown in Appendix 2.1.

Another reason that the parameter estimates may be different from the previously curve fitted values is that one patient’s curve fit skews the population trend and therefore
<table>
<thead>
<tr>
<th></th>
<th>$k_3$</th>
<th>Curve fitted $k_3$</th>
<th>$k_4$</th>
<th>Curve fitted $k_4$</th>
<th>$k_6$</th>
<th>Curve fitted $k_6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.99×10^{-33}</td>
<td>15.3514</td>
<td>0.0021</td>
<td>0.5337</td>
<td>3.55×10^{-5}</td>
<td>0.3520</td>
</tr>
<tr>
<td>2</td>
<td>2.18×10^{-33}</td>
<td>0.2758</td>
<td>1.8526</td>
<td>1.0186</td>
<td>3.56×10^{-5}</td>
<td>3.2047</td>
</tr>
<tr>
<td>3</td>
<td>1.52×10^{-33}</td>
<td>2.5809</td>
<td>0.3104</td>
<td>0.8865</td>
<td>3.57×10^{-5}</td>
<td>0.6041</td>
</tr>
<tr>
<td>4</td>
<td>2.10×10^{-33}</td>
<td>1.9271</td>
<td>3.3484</td>
<td>11.1258</td>
<td>3.92×10^{-5}</td>
<td>44.7713</td>
</tr>
<tr>
<td>5</td>
<td>2.35×10^{-33}</td>
<td>0.0095</td>
<td>0.0014</td>
<td>0.4050</td>
<td>3.59×10^{-5}</td>
<td>2.1726</td>
</tr>
<tr>
<td>6</td>
<td>1.99×10^{-33}</td>
<td>0.2118</td>
<td>0.0023</td>
<td>0.0932</td>
<td>3.55×10^{-5}</td>
<td>0.3929</td>
</tr>
<tr>
<td>7</td>
<td>1.95×10^{-33}</td>
<td>0.0020</td>
<td>0.0005</td>
<td>0.0036</td>
<td>3.57×10^{-5}</td>
<td>4.6668</td>
</tr>
<tr>
<td>8</td>
<td>1.99×10^{-33}</td>
<td>8.89×10^{-6}</td>
<td>0.0022</td>
<td>1.14×10^{-5}</td>
<td>3.55×10^{-5}</td>
<td>33.6742</td>
</tr>
<tr>
<td>9</td>
<td>1.82×10^{-33}</td>
<td>0.0028</td>
<td>0.0011</td>
<td>0.0360</td>
<td>3.45×10^{-5}</td>
<td>0.9942</td>
</tr>
<tr>
<td>11</td>
<td>1.99×10^{-33}</td>
<td>36.4</td>
<td>0.0022</td>
<td>0.0289</td>
<td>3.55×10^{-5}</td>
<td>2.1492</td>
</tr>
<tr>
<td>12</td>
<td>1.95×10^{-33}</td>
<td>0.0530</td>
<td>0.1102</td>
<td>0.0681</td>
<td>3.55×10^{-5}</td>
<td>2.2605</td>
</tr>
<tr>
<td>14</td>
<td>1.93×10^{-33}</td>
<td>0.0147</td>
<td>0.0025</td>
<td>0.0840</td>
<td>3.53×10^{-5}</td>
<td>1.0110</td>
</tr>
<tr>
<td>15</td>
<td>2.93×10^{-33}</td>
<td>0.0039</td>
<td>0.0006</td>
<td>0.8940</td>
<td>3.65×10^{-5}</td>
<td>0.1195</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$k_7$</th>
<th>Curve fitted $k_7$</th>
<th>$k_8$</th>
<th>Curve fitted $k_8$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>2.90×10^{4}</td>
<td>4.5641</td>
<td>7.51×10^{7}</td>
<td>32.9586</td>
</tr>
<tr>
<td>12</td>
<td>2.74×10^{4}</td>
<td>0.1016</td>
<td>8.84×10^{7}</td>
<td>26.7042</td>
</tr>
<tr>
<td>11</td>
<td>2.91×10^{4}</td>
<td>5.4668</td>
<td>4.45×10^{7}</td>
<td>42.5586</td>
</tr>
<tr>
<td>14</td>
<td>3.51×10^{4}</td>
<td>7.9222</td>
<td>6.56×10^{7}</td>
<td>5.7413</td>
</tr>
<tr>
<td>12</td>
<td>2.95×10^{4}</td>
<td>0.1060</td>
<td>8.23×10^{7}</td>
<td>35.5018</td>
</tr>
<tr>
<td>10</td>
<td>2.90×10^{4}</td>
<td>2.8441</td>
<td>7.50×10^{7}</td>
<td>36.6025</td>
</tr>
<tr>
<td>12</td>
<td>2.93×10^{4}</td>
<td>0.0672</td>
<td>1.04×10^{6}</td>
<td>68.9492</td>
</tr>
<tr>
<td>10</td>
<td>2.90×10^{4}</td>
<td>1.13×10^{-6}</td>
<td>7.51×10^{7}</td>
<td>39.8318</td>
</tr>
<tr>
<td>14</td>
<td>3.11×10^{4}</td>
<td>11.3855</td>
<td>9.95×10^{7}</td>
<td>160.9344</td>
</tr>
<tr>
<td>10</td>
<td>2.90×10^{4}</td>
<td>0.0644</td>
<td>7.51×10^{7}</td>
<td>61.9300</td>
</tr>
<tr>
<td>14</td>
<td>2.91×10^{4}</td>
<td>0.0699</td>
<td>7.80×10^{7}</td>
<td>32.6444</td>
</tr>
<tr>
<td>10</td>
<td>2.93×10^{4}</td>
<td>0.8100</td>
<td>7.62×10^{7}</td>
<td>11.3146</td>
</tr>
<tr>
<td>10</td>
<td>2.96×10^{4}</td>
<td>0.0149</td>
<td>1.22×10^{6}</td>
<td>16.8799</td>
</tr>
</tbody>
</table>

Table 3: Individual patient estimates for $k_3, k_4, k_6, k_7,$ and $k_8$ from the $M nlmefitsa$ run.
the estimated fixed effects corresponding random effects. The concentration of MMP’s should increase during the first few weeks of healing, but once the MMP proteins begin to interact with the TIMP’s, the concentration should decrease as the wound heals. Patient 9’s concentration of MMP’s steadily increases during the first four weeks and then healed before week 8. Because of this, the only data available for patient 9 is at time points [0,1,2,4]. The curve fit produced by nlmefitsa observed this increase and then interpreted the remaining eight weeks, without data points, as a near exponential increase in MMP’s shown in Figure 3. The parameter estimates for $k_3, k_4, k_6, k_7,$ and $k_8$ for patient 9 are more similar to the previously fitted estimates, taking into consideration that the $k_7$ and $k_8$ values were returned to the ODE solver as 200. The estimate for $k_4$ for patient 9 was more similar
to the previously fitted value than most of the patients’ estimates. The curve fit for patient 9 has skewed the population fit and led to poor parameter estimates. If this run were to be reexamined, better results may occur if patient 9 is excluded from the data set.

The last fit run with the \textit{nlmefitsa} algorithm was for state variable $T$ while estimating for parameters $k_3, k_4, k_6, k_7,$ and $k_8$, as in the fit for $M$ above. The fixed effects estimates for these parameters in Table 4 are very different than the estimates in the $M$ fit. These estimates are within the [0, 200] range. However, these estimates are smaller than

\begin{table}[h]
\centering
\begin{tabular}{l|ccccc}
\textbf{Parameter} & $k_3$ & $k_4$ & $k_6$ & $k_7$ & $k_8$ \\
\hline
\textbf{Fixed effect} & 1.8307 & 0.9764 & 2.1555 & 0.7864 & 1.7453 \\
\end{tabular}
\caption{Table 4: Fixed effects for $k_3, k_4, k_6, k_7,$ and $k_8$ from the $T$ \textit{nlmefitsa} run}
\end{table}

the average of the curve fitted values for all patients which are 7.5167, 14.9013, 25.5439, 9.9722, and 53.7389 for $k_3, k_4, k_6, k_7,$ and $k_8$, respectively. The cause of this may be that too many of the parameters were fixed to accurately convey the variability in the data or the fixed values were inadequate. The individual parameter estimates in Table 5 are very tightly clustered with little variation from patient to patient. These values can be compared to the previously curve fitted values displayed in Table 3. When compared to the previously curve fitted values, $k_8$ stands out as the most different. While the previous estimates for this parameter boast a range of 155.1931, the \textit{nlmefitsa} estimates have a range of 0.021. Clearly, the \textit{nlmefitsa} estimates are missing key variation in the data. The root mean squared error of this fit is 0.5018 which indicates that this is not a particularly helpful model to predict $T$ values, since the mean squared error is not significantly close to 0.
<table>
<thead>
<tr>
<th></th>
<th>$k_3$</th>
<th>$k_4$</th>
<th>$k_6$</th>
<th>$k_7$</th>
<th>$k_8$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.8899</td>
<td>0.9657</td>
<td>2.1553</td>
<td>0.7767</td>
<td>1.7453</td>
</tr>
<tr>
<td>2</td>
<td>1.8621</td>
<td>0.9447</td>
<td>1.6287</td>
<td>0.7566</td>
<td>1.7451</td>
</tr>
<tr>
<td>3</td>
<td>1.2243</td>
<td>0.8433</td>
<td>0.6289</td>
<td>0.6956</td>
<td>1.7491</td>
</tr>
<tr>
<td>4</td>
<td>2.1919</td>
<td>1.0476</td>
<td>2.0664</td>
<td>0.7657</td>
<td>1.7420</td>
</tr>
<tr>
<td>5</td>
<td>1.5540</td>
<td>1.0222</td>
<td>2.2157</td>
<td>0.8494</td>
<td>1.7454</td>
</tr>
<tr>
<td>6</td>
<td>1.6715</td>
<td>1.0051</td>
<td>2.1616</td>
<td>0.8166</td>
<td>1.7453</td>
</tr>
<tr>
<td>7</td>
<td>1.7806</td>
<td>0.9845</td>
<td>2.1668</td>
<td>0.7965</td>
<td>1.7453</td>
</tr>
<tr>
<td>8</td>
<td>1.8448</td>
<td>0.9738</td>
<td>2.1554</td>
<td>0.7841</td>
<td>1.7453</td>
</tr>
<tr>
<td>9</td>
<td>1.4990</td>
<td>0.9337</td>
<td>0.5824</td>
<td>0.8873</td>
<td>1.7613</td>
</tr>
<tr>
<td>11</td>
<td>1.8290</td>
<td>0.9767</td>
<td>2.1554</td>
<td>0.7867</td>
<td>1.7453</td>
</tr>
<tr>
<td>12</td>
<td>2.3396</td>
<td>0.8957</td>
<td>1.8660</td>
<td>0.6991</td>
<td>1.7453</td>
</tr>
<tr>
<td>14</td>
<td>1.9187</td>
<td>0.9622</td>
<td>2.1419</td>
<td>0.7703</td>
<td>1.7453</td>
</tr>
<tr>
<td>15</td>
<td>1.2707</td>
<td>0.8564</td>
<td>0.4341</td>
<td>0.7455</td>
<td>1.7403</td>
</tr>
</tbody>
</table>

Table 5: Individual patient estimates for $k_3, k_4, k_6, k_7, \text{ and } k_8$ from the $T$ nlmefitsa run

The estimated fits for $T$ from this run of nlmefitsa in Figure 5 seem to be stuck in an area bounded by 0.2 and 1.5 concentration of TIMPs, even though there are data points outside of this band. Most patients tend to follow the biological pattern of increased concentrations of TIMP’s that then are reduced by their interaction with the MMP’s and then reach a steady state. The three patients that do not follow this trend are patients 1, 8, and 11. These patients have constant concentrations of TIMP curve fits. This is particularly concerning because the scaled data for these three patients show the same observed trend as in the rest of the patients. The fixed values, local search optimizer, and working in a 5-dimensional system, with the other seven parameters fixed, has created issues with using nlmefitsa and curve fits in general. Using the SimBiology® toolbox will allow for multiple state variables to be estimated during a single run but it still suffers from the three above listed problems that plague this technique for this model. However, the benefit of being
able to take all four data sets, for $M, T, E,$ and $\tilde{f}$, will decrease the parameter identifiability issues since the program will then be working with a more informed model.

**Figure 4: Fit for $T$ using nlmefitsa**

![Fit for T using nlmefitsa](image)

**Preliminary SimBiology® Results**

A couple SimBiology® fit results have been included to show how this toolbox fits all four state variables for all patients. The first successful run using SimBiology® was similar to the first nlmefitsa run. All parameters were fixed except for $k_{11}$ and only $\tilde{f}$ was fitted, the other state variables were fixed to the average value from the data. The estimated fixed effect for $k_{11}$ is 15.8115. This value is more than double the fixed effect estimated from the nlmefitsa run using the same fixed parameter values. The individual patient estimates are all equal to the fixed effect estimate. Because of this, the $\tilde{f}$ curve fits are the same for all patients as seen in Figure 5. All curve fits are basically just a straight line at 1.
SimBiology® currently does not have an option to supply different initial conditions for each patient without adding each initial condition as a new parameter to be estimated. This may be one of the reasons that all the curve fits are the same, as the model is using the same initial condition for all patients. This fit returned a root mean square value of 0.5133 indicating that this fit is significantly less predictive than the _nlmefitsa_ run for the same parameter-state variable combination. The variance of $k_{11}$ was found to be 2.2941. With a larger variance and a less predictive fit, _nlmefitsa_ seems to have outperformed SimBiology® on this fit of the model. Additional graphs generated from this run can be found in Appendix 4.

The next fit run through SimBiology® estimates $k_1, k_2,$ and $k_8$ while fitting for $M$, $T$, and $E$ state variables. Since these three parameters were determined to be the most informative subset of the parameter for patient 1, patient 1 previously curve fitted values were used for the other parameters. The fixed effect estimates for $k_1, k_2,$ and $k_8$ are listed
in Table 6. However, the individual parameter estimates do not vary enough to explain the variation in the data. The variances for \( k_1, k_2, \) and \( k_8 \) were estimated to be 0.0059, 0.1999, and 0.0065, respectively.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>( k_1 )</th>
<th>( k_2 )</th>
<th>( k_8 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed effect</td>
<td>6.8159</td>
<td>0.0155</td>
<td>5.0916</td>
</tr>
</tbody>
</table>

Table 6: Fixed effects for \( k_1, k_2, \) and \( k_8 \) from the \( M, T, E \) SimBiology® run

Figure 6: Fit for \( M, T, E \) using SimBiology®

What this run lacks in predictive capabilities, it makes up for in the ability to view estimates for three of the state variables at the same time. While the individual fits for the state variables are not significantly predictive curve fits, this was the first successful run that allowed multiple state variables to be fit, as shown in Figure 6. The blue data points and curve fits correspond to state variable \( M \), red for \( T \), and yellow for \( E \). The curve fits for
each individual are extremely similar, a reflection of all patients having the same parameter estimates.

The last SimBiology® run analyzed in this paper offers a more varied fit of all four state variables and estimates for $k_3, k_4, k_6, k_7, k_8$ and $f_i$. The fixed effects are listed in Table 7. Once again, since the minimum-maximum function has been executed, the fixed effect estimate for $k_6$ is entered into the ODE system as 200. The random effects for all patients for $k_3, k_4, k_6$ and $k_7$ are very similar, resulting in basically equal parameter estimates for each patient. However, the estimates for $k_8$ and $f_i$ reflect the data’s variability more accurately. The individual patient estimates for $k_8$ and $f_i$ can be found in Table 8. The entire covariance matrix was calculated for this run which can be found in Appendix 6.1. The root mean square value for this fit of the model is 0.4600, meaning that this is not the most predictive included in this paper however it is the most predictive model for the most number of state variables. The individual curve fits

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$k_3$</th>
<th>$k_4$</th>
<th>$k_6$</th>
<th>$k_7$</th>
<th>$k_8$</th>
<th>$f_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed effect</td>
<td>$5.15 \times 10^{-5}$</td>
<td>0.0019</td>
<td>$4.07 \times 10^3$</td>
<td>$3.34 \times 10^{-4}$</td>
<td>15.1373</td>
<td>0.0321</td>
</tr>
</tbody>
</table>

Table 7: Fixed effects for $k_3, k_4, k_6, k_7, k_8$ and $f_i$ from the $M, T, E, \tilde{f}$ SimBiology® run

<table>
<thead>
<tr>
<th>Patient</th>
<th>$k_8$</th>
<th>$f_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>16.3774</td>
<td>0.0261</td>
</tr>
<tr>
<td>2</td>
<td>8.4179</td>
<td>0.1461</td>
</tr>
<tr>
<td>3</td>
<td>19.2386</td>
<td>0.0169</td>
</tr>
<tr>
<td>4</td>
<td>17.2850</td>
<td>0.0225</td>
</tr>
<tr>
<td>5</td>
<td>21.8686</td>
<td>0.0120</td>
</tr>
<tr>
<td>6</td>
<td>17.6867</td>
<td>0.0212</td>
</tr>
<tr>
<td>7</td>
<td>6.5000</td>
<td>0.3592</td>
</tr>
<tr>
<td>8</td>
<td>4.4883</td>
<td>0.9614</td>
</tr>
<tr>
<td>9</td>
<td>16.3581</td>
<td>0.0261</td>
</tr>
<tr>
<td>11</td>
<td>3.9870</td>
<td>1.0000</td>
</tr>
<tr>
<td>12</td>
<td>18.2929</td>
<td>0.0193</td>
</tr>
<tr>
<td>14</td>
<td>21.8635</td>
<td>0.0273</td>
</tr>
<tr>
<td>15</td>
<td>16.1096</td>
<td>0.0321</td>
</tr>
</tbody>
</table>

Table 8: Individual patient estimates for $k_8$ and $f_i$ from the $M, T, E, \tilde{f}$ SimBiology® run
are shown in Figure 7. In the graphs, blue data points and lines represent $M$, red is for $T$, yellow is $E$, and purple is $\tilde{f}$.

Figure 6: Fit for $M, T, E, \tilde{f}$ using SimBiology®
CONCLUSIONS AND FUTURE WORK

While *nlmefts*a offers more versatility by way of modifying and adapting the code in the regular MATLAB interface, it also comes with significant limitations such as only fitting for one state variable and only using local search algorithms. On the other hand, SimBiology® offers an easy-to-navigate user interface while also allowing for multiple state variables to be fit. SimBiology® also boasts a faster simulation time due to the modifications built into the PK/PD models for rapid simulation. However, SimBiology® is still simply an extension of the MATLAB *nlmefts*a function and suffers from the lack of a constrained optimizer.

After trying both of these methods of implementing mixed effects modeling, it has been decided that SimBiology® offers more opportunities for a reasonable convergence to be obtained and more advanced fits to be generated. However, there are a few issues that should be addressed in later simulations and fit of this model. One of this issues is the crucial initial guesses for estimated parameters and fixed values for the non-estimated parameters. Incorrect fixed values may encourage the ODE system to converge to unreasonable estimates. Dozens of runs should be simulated using a variety of previously curve fitted data and previously calculated estimates. The fixed effect estimates from these runs should be compared and inform the choosing of the most appropriate fixed values for this model.

Another effort in furthering this work would be to review other programming software that may be better equipped for dealing with such complex nonlinear mixed effects modeling. There built-in functions for nonlinear mixed effect modeling in both R
and SAS that should be examined. However, there is a very promising program MONLIX that was specifically designed for nonlinear mixed effects modeling with the SAEM algorithm addition. MONOLIX was designed and still involves a multi-disciplinary group of academic statisticians from several universities in Paris, researchers in the pharmacology industry, researchers in the agronomy, animal genetics, industry, and scientists in the oncology department of the medical institute at the Lyon-Sud University in Villeurbanne, France (Mentré et al., 2008). MONOLIX may provide the robust computing ability and the intricate fitting algorithms this technique needs.

In the future, there may be adaptations or additions that could be made to the model to improve the predictive and informative capacity. Since it has been found that a lack of monocytes or macrophages can cause serious delays in wound healing (Velnar et al., 2009), it may be beneficial to collect data on these cells in future data collection as well as expand the model. This may prove to be difficult since the blood monocytes transition into tissue macrophages during the early part of the inflammatory phase of healing. The transition may be challenging to capture in both a modeling and data collection sense. In general, additional data collection would encourage more accurate results from \textit{nlmefitsa}. A temporary solution to this would be to include all sixteen patients in future runs of nonlinear mixed effects modeling. In the paper only thirteen patients were included because these were the patients with the most data points. Since nonlinear mixed effects modeling is specifically designed to handle sparse data, the additional data points for any patients should help inform results.
REFERENCES


APPENDIX

Appendix 1.1: Individual patient residual for *nlme*fitsa run for $\tilde{f}$

![Individual Residuals for nlme*fitsa run for F](image1)

Appendix 1.2: Conditional weighted patient residual for *nlme*fitsa run for $\tilde{f}$

![Conditional weighted patient residual for nlme*fitsa run for F](image2)
Appendix 2.1: Run chains for the \textit{nlmeSs} run for $M$
Appendix 2.2: Individual patient residuals from the nlmefitsa run for $M$

Appendix 2.3: Conditional weighted residuals from the nlmefitsa run for $M$
Appendix 3.1: Individual patient residuals from the *nlmefts* run for $T$

Appendix 3.2: Conditional weighted residuals from the *nlmefts* run for $T$
Appendix 4.1: Observation versus prediction from the SimBiology® run for $\tilde{f}$

![Observation versus Prediction](image1)

Appendix 4.2: Box plot of random effects from the SimBiology® run for $\tilde{f}$

![Random Effects Box Plot](image2)
Appendix 5.1: Observation versus prediction for the SimBiology® run for M, T, and E

*Top left:* Observation versus prediction graph for M

*Top right:* Observation versus prediction graph for T

*Bottom left:* Observation versus prediction graph for E
Appendix 5.2: Random effects box plots for the SimBiology® run for $M$, $T$, and $E$
Appendix 5.3: Residual distribution plots for the SimBiology® run for $M$, $T$, and $E$

CWRES: Conditional Weighted Residuals
IWRES: Individual Weighted Residuals
Appendix 6.1: Covariance of random effects for the SimBiology® run for $M$, $T$, $E$, and $f$

<table>
<thead>
<tr>
<th></th>
<th>$k_3$</th>
<th>$k_4$</th>
<th>$k_6$</th>
<th>$k_7$</th>
<th>$k_8$</th>
<th>$f_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_3$</td>
<td>1.1431</td>
<td>0.9028</td>
<td>0.0217</td>
<td>0.8422</td>
<td>-0.9865</td>
<td>2.4835</td>
</tr>
<tr>
<td>$k_4$</td>
<td>0.9028</td>
<td>0.7991</td>
<td>0.0390</td>
<td>0.8112</td>
<td>-0.8883</td>
<td>2.3676</td>
</tr>
<tr>
<td>$k_6$</td>
<td>0.0217</td>
<td>0.0390</td>
<td>0.0222</td>
<td>-0.0245</td>
<td>-0.0194</td>
<td>0.0711</td>
</tr>
<tr>
<td>$k_7$</td>
<td>0.8422</td>
<td>0.8112</td>
<td>-0.0245</td>
<td>1.3038</td>
<td>-1.0520</td>
<td>2.8991</td>
</tr>
<tr>
<td>$k_8$</td>
<td>-0.9865</td>
<td>-0.8883</td>
<td>-0.0194</td>
<td>-1.0520</td>
<td>1.0369</td>
<td>-2.7931</td>
</tr>
<tr>
<td>$f_i$</td>
<td>2.4835</td>
<td>2.3676</td>
<td>0.0711</td>
<td>2.8991</td>
<td>-2.7931</td>
<td>7.7208</td>
</tr>
</tbody>
</table>

Appendix 6.2: Population fit for the SimBiology® run for $M$, $T$, $E$, and $f$
Appendix 6.3: Observation versus prediction for the SimBiology® run for $M$, $T$, $E$, and $\tilde{f}$

*Top left:* Observation versus prediction for state variable $M$
*Top right:* Observation versus prediction for state variable $T$
*Bottom left:* Observation versus prediction for state variable $E$
*Bottom right:* Observation versus prediction for state variable $\tilde{f}$
Appendix 6.4: Box plot of random effects for the SimBiology® run for $M$, $T$, $E$, and $\tilde{f}$
Appendix 6.5: Residual distributions for the SimBiology® run for $M$, $T$, $E$, and $f$

<table>
<thead>
<tr>
<th></th>
<th>CWRES</th>
<th>IWRES</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td><img src="image1.png" alt="Graph" /></td>
<td><img src="image2.png" alt="Graph" /></td>
</tr>
<tr>
<td>T</td>
<td><img src="image3.png" alt="Graph" /></td>
<td><img src="image4.png" alt="Graph" /></td>
</tr>
<tr>
<td>E</td>
<td><img src="image5.png" alt="Graph" /></td>
<td><img src="image6.png" alt="Graph" /></td>
</tr>
<tr>
<td>$f$</td>
<td><img src="image7.png" alt="Graph" /></td>
<td><img src="image8.png" alt="Graph" /></td>
</tr>
</tbody>
</table>

CWRES: Conditional Weighted Residuals
IWRES: Individual Weighted Residuals