SEPARATING BETWEEN-SUBJECT AND WITHIN-SUBJECT NONLINEAR
COVARIATE EFFECTS, WITH APPLICATION TO LONGITUDINAL ROAD
RACING DATA OF COMPETITIVE RUNNERS

by

DANIEL J NELSON

B.A., Stanford University, 1992

A thesis submitted to the
Faculty of the Graduate School of the
University of Colorado in partial fulfillment
of the requirements for the degree of
Master of Science
Biostatistics Program

2014
This thesis for the Master of Science degree by

Daniel J Nelson

has been approved for the

Biostatistics Program

by

Matthew Strand, Chair

Gary Grunwald

Gary Zerbe

Date: 5/6/14
Nelson, Daniel J. (M.S., Biostatistics)

Separating Between-Subject and Within-Subject Nonlinear Covariate Effects, with Application to Longitudinal Road Racing Data of Competitive Runners

Thesis directed by Matthew Strand

ABSTRACT

Twelve years of longitudinal data for competitive runners of a 10k road race can be modeled with a mixed model based on a nonlinear population function that incorporates between-subject and within-subject covariate effects. This offers an improvement on generalized or nonlinear mixed models that ignore these effects. [Neuhaus and Kalbfleisch] The model allows for an examination of how subject-specific changes in performance by age differ from the population average.

The form and content of this abstract are approved. I recommend its publication.

Approved: Matthew Strand
ACKNOWLEDGMENT

I would like to thank Matthew J. Strand for his guidance and knowledge in overseeing the production of this project, and for his inspiring curiosity to return to his original research and elaborate on a fascinating problem. Thanks also to my committee members Gary Grunwald and Gary Zerbe for their expertise in suggesting ways to reframe and re-examine the problem and for providing technical knowledge on how to accomplish difficult programming. Thanks to my family, who patiently supported and nourished me through this process, and to my friends for their encouragement and diversion when my sketches of data curves and notation devolved from looking less like spaghetti plots and more simply akin to spaghetti.
# TABLE OF CONTENTS

## CHAPTER

I. INTRODUCTION ................................................................. 1
   1.1 Setting ................................................................. 1
   1.2 Background .......................................................... 1
   1.3 Questions of Interest ............................................... 2

II. DATA DESCRIPTION .......................................................... 4
   2.1 Data Set ............................................................. 4
   2.2 Software Toolbox .................................................. 4

III. DATA METHODS .............................................................. 7
   3.1 Key Measures ....................................................... 7
   3.2 Basic Visualization ............................................... 8
   3.3 Competitive Runners .............................................. 8
   3.4 Ranking and Inclusion ............................................ 10
   3.5 Refined Visualization ............................................. 11
   3.6 Algorithm for Data Cleaning .................................... 13
   3.7 Identity ............................................................. 15
   3.8 Gender Confusion .................................................. 15
   3.9 Age Correction ..................................................... 17

IV. STATISTICAL METHODS .................................................... 18
   4.1 Nonlinear Model(s) .................................................. 18
   4.2 Mixed Model, Random Effects .................................... 18
   4.3 Subject vs. Population ............................................ 19
   4.4 Multi-Step Approach ............................................... 21
   4.5 Package Inspection ................................................ 22
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.6 Log Transformation and Cluster Covariate Effects</td>
<td>23</td>
</tr>
<tr>
<td>4.7 Interactive Simulation</td>
<td>24</td>
</tr>
<tr>
<td>4.8 Age Grading</td>
<td>25</td>
</tr>
<tr>
<td>V. RESULTS</td>
<td>28</td>
</tr>
<tr>
<td>5.1 Findings</td>
<td>28</td>
</tr>
<tr>
<td>5.2 Collinearity Warning</td>
<td>28</td>
</tr>
<tr>
<td>5.3 Model Comparisons /Progression</td>
<td>28</td>
</tr>
<tr>
<td>5.4 Between versus Within</td>
<td>29</td>
</tr>
<tr>
<td>5.5 Elaboration of Models</td>
<td>31</td>
</tr>
<tr>
<td>5.6 Adding Random Effects</td>
<td>32</td>
</tr>
<tr>
<td>5.7 Simulation Results</td>
<td>35</td>
</tr>
<tr>
<td>5.8 Multi-Steps Taken</td>
<td>38</td>
</tr>
<tr>
<td>5.9 Log-Transformed Final Model</td>
<td>39</td>
</tr>
<tr>
<td>5.10 Comparison of Results with Proc Mixed</td>
<td>41</td>
</tr>
<tr>
<td>5.11 Optimization Issues</td>
<td>42</td>
</tr>
<tr>
<td>VI. DISCUSSION</td>
<td>44</td>
</tr>
<tr>
<td>VII. CONCLUSION</td>
<td>48</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>49</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

## TABLE

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Number of Runners and Standard Deviation of Finishing Time by Age by Sex</td>
<td>5</td>
</tr>
<tr>
<td>3.1</td>
<td>Bolder Boulder Wave Start Times</td>
<td>9</td>
</tr>
<tr>
<td>5.1</td>
<td>Parameter Estimates and their Standard Errors, Final Model</td>
<td>30</td>
</tr>
<tr>
<td>5.2</td>
<td>Estimates and Standard Errors of Random Effects Models</td>
<td>32</td>
</tr>
<tr>
<td>5.3</td>
<td>Estimates and Standard Errors, Log-Transformed Models</td>
<td>41</td>
</tr>
<tr>
<td>5.4</td>
<td>Estimates and Standard Errors, PROC MIXED</td>
<td>42</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

**FIGURE**

- 2.1 Finishing Time by Age, Both Sexes .......................... 6
- 3.1 Raw Data .................................................. 7
- 3.2 Raw Data .................................................. 9
- 3.3 Scatterplot of Men’s Data ................................. 11
- 3.4 Spaghetti Plot, Individual Males ........................... 12
- 3.5 Color Coded Data Cleaning ................................. 14
- 3.6 After Further Processing ................................. 15
- 3.7 Renaming Examples ....................................... 16
- 4.1 3D Visualization .......................................... 25
- 4.2 Age Grading, from Measurement News ..................... 27
- 5.1 Final Model ................................................ 31
- 5.2 Base Model ................................................. 33
- 5.3 Adding a Random Intercept ............................... 34
- 5.4 Adding a Random Intercept and Slope ................. 35
- 5.5 Simulated Range of Left Exponential Parameter ......... 36
- 5.6 Simulated Range of Right Exponential Parameter ..... 37
- 5.7 Overlay of Simulations .................................. 37
- 5.8 Residuals After a Base Model ........................... 38
- 5.9 Residuals Modeled ........................................ 39
- 5.10 Residuals Modeled and Rescaled ......................... 40
- 5.11 AIC Comparison ......................................... 41
- 6.1 Zoomed In Fit ............................................. 45
- 6.2 JMP Smoother .............................................. 46
CHAPTER I
INTRODUCTION

1.1 Setting

The Bolder Boulder 10 kilometer race launches annually each Memorial Day weekend from downtown Boulder, Colorado, tracing a rigorous and challenging course through the streets of the Boulder community and finishing in Folsom Field stadium on the University of Colorado campus. Begun in 1979, the Bolder Boulder has grown to one of the largest road races in North America, with over 54,000 participants in 2013 comprised of athletes of diverse interests and abilities, including world-class elite runners, costumed participants in the citizen’s race, and a wheelchair running division begun in 1990.[1] Noted for its majestic setting and concomitant festivities throughout the race weekend, the Bolder Boulder offers one of the largest cash prizes for competitive road racing, and has innovated in timing measures of road races with its wave start methodology to accommodate the enormous numbers of participants, as well as cutting-edge RFID technology launched in 2007 that tracks every participating runner through tags attached to a competitor’s shoe and surveilled throughout the course of the race, with real-time updates on race split and finish times.

The Bolder Boulder race day contains a series of race events within the larger ‘Bolder Boulder’ framework, including the team event and walking waves. In addition, the top 15 male and top 15 female runners in each age category win an award from ages 6 and over. The Bolder Boulder organization has expanded to offer multiple race events throughout the year in their race series, such as a 5k “Flat Out” race and the “Colder Boulder” since 2011, a winter event. Each event provides a unique challenge to competitive runners, and the organization is world-renowned.

1.2 Background

The Bolder Boulder finish times have been published since 1981 to present day on the Bolder Boulder website, and since 2006 have included the additional detail
of split times for every participant. The motivation for this research began with
the particular nature of these data, and with the question of interest in describing
how the finishing times of elite runners change as those runners age. This research
builds on the original analysis that Matt Strand completed for his thesis work, in
which he analyzed the 1995 results of the race using extreme value theory. [10]
Extending beyond that original query, the data for this project represent several years
of results from the race, analyzed longitudinally and with a particular focus on the
hierarchical nature of the data. With the addition of multiple years of results, the data
presents interesting features of between- and within-subject variation, also described
as cluster versus population covariate effects. This research paper will present the
background, methodology and results of the effort to describe a nonlinear mixed
model with random effects for this physiological data. As a chronicle of this research,
the paper will be organized around two topics: 1) the extraction, arrangement and
visualization of the data set, which presents interesting and unique features; and
2) the approach, challenges, and methodology to choosing an appropriate statistical
model for that data set, and those model results.

1.3 Questions of Interest

In his earlier work, Strand identified his question of interest as to determine
“the expected rate of change in performance, loss or gain. . . for a [competitive]
runner of a given age.” [10] He defined a ‘competitive’ runner as those who perform
well in their respective age category, understanding that such runners train hard,
take races seriously but are recreational and not professional competitors. A more
formal discussion of what comprises a competitive runner and the justification for
this perspective will take place in the data methods section of this paper. For now,
let’s accept that the data comprises the top runners of each age category by years
included in a data set separated out by gender. Our primary intent is to investigate
characteristics of finishing times of these top runners, and to investigate patterns that
might occur with the new perspective that we have on this data, notably the addition of a longitudinal dimension for data on individual runners.

Speaking of this longitudinal data, confusion might arise in terminology between the outcome of interest — race finish times — and our predictor variable age, since longitudinal analyses often investigate an outcome as changes over some chronological period. For our purposes, “time” will refer to this primary outcome: race finish times. Our longitudinal dimension is referred to as “age.”

As shall soon be shown, the particular data under study show interesting characteristics when arranged in a useful perspective for the purposes of enlightening this primary question.

These characteristics, namely the longitudinal aspects and subject versus population effects, give rise to secondary questions of interest. The secondary questions of interest are: can we describe the patterns that might occur in individual runners as they age for this particular race with respect to the overall trend for runners as they age, and what models and curves might best fit these subject-specific patterns?

Lastly, as we gain insight on the patterns occurring in the data, tertiary questions arise. These more finely-grained questions can be stated as: investigating trends in subject-specific patterns in subdivided age categories (and inclusively wondering what might be the most reasonable subdivisions of sub-categories of age), and then formally testing these trends to see if statements can be made about the patterns regarding these sub-categories. The investigation into these tertiary questions is least detailed and more exploratory, but worth including in this summary as it suggests directions for future investigation.
CHAPTER II
DATA DESCRIPTION

2.1 Data Set

Having alluded to the data set under study, this section now presents an overview of the summary data and a preliminary visualization of the data set. A more detailed investigation of how the data was assembled, techniques used, and issues raised follows in the section on Data Methods.

I parsed data from the Bolder Boulder website which provides results for 31 years of past race results, from 1981 to 2013 [1]. However, as this site lacks results from the years 1992 and 1993, recent contiguous results data were limited to years 1994 and after. In addition, Bolder Boulder changed the formatting of race results in more recent years. For these reasons, the final data sets comprises 12 years of contiguous data from 1994 to 2005 inclusively. 2007 was also the year that RFID tracking was initiated, so it make sense to limit data to before the measuring technology methodology changed. Furthermore, previous analyses indicate that running time trends differ substantially between men and women [10]. Summary data is therefore presented separately by sex.

A first view of the summary data after it has gone substantial preparation is presented in Table 2.1. These results include the 25th place finish or better for Men and Women from 1995 to 2004. These data are also presented as a scatterplot in Figure 2.1. Note that all subsequent analysis has been performed on a subset of the Males data.

2.2 Software Toolbox

Throughout the data preparation and statistical analysis of this project, a variety of software packages were used. At the time of writing, these were the tools that been employed, with a brief editorial about the value of each tool.
The analysis was begun with SAS 9.3 and continued with the release of SAS 9.4, as well as SAS Enterprise. In later stages, SAS On Demand was used for the analysis. With this fairly large data set, I found that SAS On Demand was an efficient tool that seemed to carry out computational tasks (such as NLMIXED) more quickly than usage on a desktop model. However, SAS On Demand does require an available and reliable internet connection.

JMP 10 and 11 were used substantially for its facile graphing and visualization processes, as well as for quickly analyzing summary data. Nonlinear models were verified through JMP’s Nonlinear Modeling package.

R and RStudio were used for simulation creation and graphing.

RStudio’s Shiny package was used for an interactive visualization of the final model.
Figure 2.1: Finishing Time by Age, Both Sexes
CHAPTER II
DATA METHODS

3.1 Key Measures

Before delving into the history surrounding the data preparation methodology, it’s worth reiterating the setting of the data as it appeared, and the motivation for spending time cleaning it in anticipation of analysis.

The Bolder Boulder website publishes race results by year. These results, however, are quite succinct, and a fair amount of work is needed to tease out the data embedded in the raw output. Although the web site provides some simple searching and filtering functions, downloading results fetches a substantial PDF file containing the names, division and time for all participants. A snapshot of the raw data is shown in Figure 3.1.

![Figure 3.1: Raw Data](image)

These results from 2001 show the bare information: the name of a participant, a categorical division label which concatenates a symbolic reference to sex and a numerical age, and finish time presented as Minutes:Seconds or Hours:Minutes:Seconds.

The primary variables of interest that I needed to extract from this data were the

1. finishing time, in an appropriate and standardized time format;
2. an identifier for participants that would match between years run;
3. the sex label;
4. the age category as recorded;
5. the year of the results.

The eventual model will use all of these factors, albeit using some of them to cross-reference or validate, and some (such as sex) to create a subset of data to analyze.

Extraction of the PDF into data required a pass-through plain text format, after which the plain text file could be parsed into SAS in preparation for labeling and manipulation. In this endeavor the previous work of programmers in the division of Biostatistics at National Jewish Health who provided assistance with the original parsing code from plain text files into SAS data proved invaluable, although I detected that the time format used in the original script was not accurate, possibly having changed since the script was first written.

3.2 Basic Visualization

A first order of business after parsing the raw data was to create a rough visualization eventually much refined through data processing. Looking at the entire data set (all records from 1994–2005) as a scatterplot of age versus time, it’s difficult to detect any features at all, due to the volume of information and density of the data (over 457,000 records) (Figure 3.2)

However, we can note the slightly curved bottom edge of the data at the best finishing times, some potential outliers in the later age ranges (in terms of best finished), and the high proportion of finishing times slower than 1 hour. Overall, the data seems fullest through the young adult and middle ages of 20–55, tapering slightly at the age extremes— though it is worth noting that the Bolder Boulder includes runners from age 6 through age 80 and beyond.

3.3 Competitive Runners

Although the raw data presents a great opportunity for research with its huge volume of data, the motivating questions for this research focused on the qualities of
the competitive runner, so some discussion of what will be considered a competitive runner is in order.

The Bolder Boulder itself provides some rough definition with its wave start structure. Although open to any participant, in order to be placed in the waves, a registrant has to provide proof of a qualifying time; proof can be obtained from a previous Bolder Boulder race, another BB race series event (other races held by the same organization throughout the year) or any other timed race between 2 miles and marathon length within the previous year. To understand how finely calibrated the waves are, Table 3.1 provides a snapshot of the first 4 waves for 2014 separated by 1 minute and sometimes 10 second start times:

<table>
<thead>
<tr>
<th>Wave</th>
<th>Group</th>
<th>Start Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Sub 38:00 Qualified</td>
<td>7:00:00 AM</td>
</tr>
<tr>
<td>AA</td>
<td>38:00 - 41:00 Qualified</td>
<td>7:01:00 AM</td>
</tr>
<tr>
<td>AB</td>
<td>41:01 - 43:00 Qualified</td>
<td>7:02:10 AM</td>
</tr>
<tr>
<td>B</td>
<td>43:01 - 44:30 Qualified</td>
<td>7:03:00 AM</td>
</tr>
</tbody>
</table>
Jeff Galloway, an 1972 American Olympic team member and lifelong runner, provides a definition of five categories of a runner in his Book on Running, including the “Competitive Runner” as marked by setting goals for running, including training and timing goals. (15-16)[3]. In fact any of Galloway’s latter 3 of 5 definitions, including the “Competitive Runner,” the “Athlete,” and ultimately the [quintessential] “Runner,” could be considered for inclusion in this project. Likewise Strand identified competitive runners as those who “train hard and take the race seriously... not running for money, but are performing well with respect to others of their age.” (205) [10].

3.4 Ranking and Inclusion

Strand further defines this competitive runner procedurally as “those runners that place in the top k of their age.” (205) [10]. Implicit in this definition is the stratification of the running data previously into sex categories—explicitly thus the top k in an age and sex category defines the top runners.

As the current data extends to the perspective of longitudinal data over several years, of interest was tracking any given year’s top runners as they performed in other years of the Bolder Boulder race (if indeed they returned). Thus a more complex inclusion strategy was required, and carried out algorithmically through iterative SAS processing. The schema used was as follows:

1. Runners were ranked in age and sex category per year results
2. Any runner’s id who appeared in the top k of the results were kept
3. Those runners’ id’s were searched for throughout the entire data set, and kept if found
4. Of those results, finishes better than the rank of j were kept.

Selection was accomplished through a SAS Macro program using the RANK procedure in SAS with default settings, thus the bounds of k and j could be varied until satisfactory inclusion boundaries could be reached via trial and error. The guiding principles for this selection remained the desire to focus on top runners performing at the peak of their abilities and eventually make statements about their performance. Had the secondary j boundary not been included, extreme outliers
representing slow finish times distinct from typical results distorted the analysis.

3.5 Refined Visualization

After refining the data set through the rank selection process, a more restricted view of the data arrived. Examining this visualization highlights the salient characteristics of interest to the investigator, and is presented below.

![Figure 3.3: Scatterplot of Men’s Data](image)

Here we can see more clearly the curve of the lower (faster) threshold of the data, which overall appears roughly U-shaped. Note that ages have also been restricted to between ages 8 and 60. The top 25th or better finishes of the included subjects appear relatively cohesive in this view, representing a band of data around what could be a population curve, also U-shaped although with a steep descent on the left.
(younger) side, and a more gradual ascent on the right (older) side. An outlier at age 10 represents a female runner that indeed did perform very well in her class, and likewise an exception 63-year-old.

Subjects have been marked individually by color, although with the number of subjects the color scale repeats and thus subjects cannot be identified by eye. However, by applying the “cubic spline smoother” model in JMP’s Graph Builder, we can see even at this density interesting within-subject curves in the spaghetti plot below.

![Spaghetti Plot, Individual Males](image)

Figure 3.4: Spaghetti Plot, Individual Males

This visualization paints the way towards the goals of this analysis: within-subject patterns that deviate from the overall population trend, understood as improvement or worsening with respect to the population trend in that particular section of the age results. Furthermore, these subject-specific patterns seem to be influenced by the overall population trend: the left ‘younger’ side looking more streamlined and tighter than the perhaps more variable right and aged runners. These age categorical differences beg then the question of what age categories might represent common development trends in terms of performance.
3.6 Algorithm for Data Cleaning

Some further aspects of the data cleaning process are worth mentioning here, since the arrival at values for age, gender and identity were not as straightforward as they might first appear. The general approach to data preparation used an iterative algorithmic approach in combination with manual verification to efficiently refine the data series in successive steps.

The general approach was to

1. Sort and group records by sex, age, and name, using name as a primary identifier
2. Present in tabular form the records grouped by name
3. Color-code the rows of the table by how likely it appeared that the records might be related to the best place finish
4. Correct for whatever criteria might be the issue.
5. Repeat the process and continue refinement.

Repetition of the process was necessary because any corrective step could have an impact on including or excluding a single observation, which might then impact the rank of any observations within a particular group. An illustrative example might be as follows: a) Jess Smith is identified as 4th place finisher in the year 2000 at age 30 for females. b) A search for Jess Smith reveals 3 additional records. c) Two of those records has Jess Smith recorded as a male runner. d) The records for Jess Smith are corrected and kept or excluded.

More detail about this particular gender issue is included in the section “Gender confusion.” Because of the recursive nature of this data verification and adjustment process, changes such as gender reassignment or age correction could have an impact on ranking, and thus inclusion or exclusion, and thus re-ranking, etc. This issue is particularly critical in this data set, since our primary focus in not on the lumpen middle of the data, but on the extreme value minima that the best ranked finishes represent.
Because time in graduate school is, alas, finite, bounds were placed on the amount of time spent refining the data, and decisions were made as the data set grew more stable not to make major revisions. I attest in good faith that the data I have presented have reasonable integrity and I believe they are quite good for the illustrative purposes of this analysis: with the caveat that more improvement can probably be made through additional investigation.

Below is an illustrative sample of the algorithmic coding as it was practiced for this data set. In this example, ‘M—C—’ has a best place finish in the year 1994 for fourth place in his age category of 23. Several other entries identified as ‘M—C—’ are present in the record— through 2005. Indeed, in 2005 we note two entries for this subject(s?)— one aged 32 and another aged 50. We suspect that this last-name/first-name pair represents at least two intermingled runners’ records, also noticing the wide discrepancy in typical rank placements between the two. Also note the onetime inclusion of a middle name, ‘M—A—C—’ in 2004.

The results are thus color coded from the best finishing time age and year (bold-face red text), with variables calculated relative to the difference between any given record and this marker record of best finish time age and year. For instance, the 2005 aged–50 results occur 11 years later from the best 1994 results, but 27 years more aged than the best 1994 results.

These differences are coded for severity from 1–4 and colored accordingly. After iterations of hand inspection and correction, a more plausible data record for this subject has been refined.
3.7 Identity

Identity was ascribed to the subject records of the data using name information of the original records. This presented several issues.

Names seemed to have been self-assigned on registration. Therefore, there was no restriction on a subject using a legal name or nickname to identify themselves. This lead to searching and attempts to match subjects by likely shortenings of names, if it appeared likely that it was the same subject based on year, age, sex and sometimes rank or finishing time information. It was possible that two or more subjects appeared at first mingled because of identical last- and first-names as in the previous example. Generally the additional information of year, age, and rank could be used to disentangle these records.

In less common cases, subjects appeared to have changed names, perhaps due to marriage or divorce, and adopted new last names or hyphenated last names. In cases where this appeared likely, subject records were grouped together and given a common name.

A code excerpt shows the typical name reassignments that were made through IF-THEN statements after the kind of color-coding and manual inspection that occurred in the previous example.

3.8 Gender Confusion

Although the Bolder Boulder appears to subscribe to a binary gender classification system (current online registration provides radio buttons for males or females only), the published records contained some gender confusion. Generally divisions were marked with an uppercase ‘M’ or ‘F’ to denote the sex category that was paired with the age. In less common circumstances, the gender was marked with a lowercase
variant of these, or with other letters (‘G’, ‘D’, ‘N’). These letters represent contiguous keys on the standard Western-layout keyboard to F and M and could be reasonably assumed to be mistyped sexes. Still other records recorded the gender as a U category, which might be presumed to be ‘Unknown,’ perhaps because of hand-written registration forms or other data integrity issues.

Generally, observations with genders that did not subscribe to the binary classification had to be reassigned or excluded, since the process of re-assigning a gender impacted the ranking system as ranks were calculated within sex and age categories.

In rare cases sex might have been recorded as an F or M, but all other records of that particular subject were assigned the opposite sex. In these cases, it was possible that misclassification gave them a inclusive finish time which they would otherwise not have achieved in their opposite gender class. Generally, this benefitted males who might have been miscast as a female, thereby giving them a better-ranked female finishing time than would have happened in their male age category (females on average having slower finish times than men.) In these cases, if the subject had other place finishes that would keep them qualified for inclusion, the records were kept. Otherwise, all records of that subject were excluded.

We have no method to address an actual gender change on the part of a runner during the course of the race. In that case, however, we would group the results with
the sex and age class results that were recorded during the time of the race, in effect splitting the records between the two gender classes.

3.9 Age Correction

As the variable age would become the primary predictive measure, it required some care to refine the age values per subject in order to complete reasonable analysis.

The Bolder Boulder race is held on Memorial Day, which is defined as the “the final Monday of May.” Because this is not a fixed but a calendar-relative date, it was possible that participants with birthdays at or near the race date could display identical ages for successive yearly races. Therefore, two different approaches were used to correct for ages. In one approach, the age of the best place finish of a particular subject was used as the referent age, and age and year data compared to that key value were used to recalculate a derived age for that participant, as presented in the worked example earlier.

Sometimes, however, it was apparent that the best place finishing age was an anomaly: in the sense that most records were off by one year from the best race result finish. In this case, a ‘modal age’ was used as the best guess for the age of that participant, and a corrected age was assigned based on the year differences of the ‘modal age.’ Since participants could not run twice in one year (assuming the record does not represent mingled individuals), this recalibration is a reasonable solution.

Of important note, however, is that consequently one years’ age difference in the model does not necessarily represent one full year of aging. Therefore statements that try to calculated predictive or fractional age results should be aware of possible error and potential bias.
CHAPTER IV
STATISTICAL METHODS

4.1 Nonlinear Model(s)

The basis for this current work is nonlinear modeling methods, or models of the general form \( y = f(X, \beta) + e \) with errors normally distributed, \( e \sim N(0, \sigma^2) \). In this case, the primary model as shown by Strand that best fits the population data is a non-linear equation of the form

\[
f(x) = \alpha_0 x^{\alpha_1} e^{x\alpha_2}
\]

— a fixed effect model. Neatly, this model has the advantage that the peak (fastest performance) age is found when the derivative is set to 0, which turns out to be \(-\alpha_1/\alpha_2\).

This main population model will be used as the basis for all subsequent models, and may be abbreviated \( f(x) \).

4.2 Mixed Model, Random Effects

Elaboration of this model began with extension into random effects. First, a random intercept was added.

\[
f(x) = b_{0i} + \alpha_0 x^{\alpha_1} e^{x\alpha_2}
\]

Then, a random slope and subsequently random quadratic term were added as well.

\[
f(x) = b_{0i} + b_{1i} x + \alpha_0 x^{\alpha_1} e^{x\alpha_2}
\]

and

\[
f(x) = b_{0i} + b_{1i} x + b_{2i} x^2 + \alpha_0 x^{\alpha_1} e^{x\alpha_2}
\]
Conversely, it was proposed to explore random terms not added linearly but in the exponents of the main model. So, equations in the variations below were explored.

\[ f(x) = \alpha_0 x^{(b_1 + \alpha_1)} e^{(\alpha_2) x} \]
\[ f(x) = \alpha_0 x^{(\alpha_1)} e^{(b_2 + \alpha_2) x} \]
\[ f(x) = \alpha_0 x^{(b_1 + \alpha_1)} e^{(b_2 + \alpha_2) x} \]

The random effect errors are also assumed to be normally distributed around zero.

4.3 Subject vs. Population

At issue with the model building is the main focus on achieving distinction between the between- and within-subject effects. Conceptually, our ideal model would account for individual runners both displaying a trend in concert with the main population model but distinct to their own performance at the age range during which they ran—within-subject effects. How they differ from other runners of the same age would be modeled by between-subject effects.

The separation of effects into within- and between-subject is discussed and Hedeker and Gibbons [5], and in their example (74) they note that the estimated coefficients are very different and even of opposite sign, though neither is statistically significant to warrant rejecting the assumption of equality for within- and between-effects. Nevertheless they provide the first steps of a methodology similar to Neuhaus and Kalbfleisch.

Following the work of Neuhaus and Kalbfleisch (638) [8], we can summarize the between-subject (or between-“cluster”) component with the mean age of the observations for that subject, \( \bar{x}_i \). The within-subject component can be summarize by \( x_{ij} - \bar{x}_i \). Neuhaus and Kalbfleisch provide illumination for the application of this structure within longitudinal studies, perfectly apt in this scenario:
For example, in longitudinal studies, we can distinguish between (i) an overall effect of age on response, as measured by the association of the mean age $X_i$ with the response, and (ii) the effects of deviations from the average age $X_{ij} - \bar{X}_i$ on the series of responses within the cluster. Typically, authors do not distinguish between- and within-cluster covariate effects in generalized linear mixed models and so implicitly assume that these effects are the same... as a consequence, models that incorrectly assume common effects can lead to very misleading assessments of the association of covariates with response. (639) [ibid]

In our particular model, we wonder if this status of the main covariate age can likewise be decomposed into a mean age per subject and a deviation of age of that subject from his mean for subject or random effects distinct from the population function, and if that articulation of the covariate will provide a meaningful addition to our model. We anticipate that it will, given the patterns noted in the early visualization of the spaghetti plot earlier.

Olofsen et al. [9] speak further to the benefit of developing models that embrace the complexity of subject versus population effects and intra- and inter-subject variability, both within the nonlinear mixed-effects model and more generally while studying biomathematical models of fatigue and performance. The authors note that if approaches use an average of observations across subjects or fitting a model of subjects first and subsequently averaging parameter estimates creates inaccurate standard errors of those parameters. “The reason,” they explain, “is that intra- and inter-individual variabilities are intertwined.” A good approach is the one we’ve taken here, to use mixed-effects modeling that “separates fixed effects (usually constant parameters or functions of time [in our case, age]) from random effects, describing the sampling of subject-specific parameter values from probability distributions.” This has the benefit, they conclude, of describing the experimental observation involving multiple subjects “properly” and “parsimoniously.”
4.4 Multi-Step Approach

As an approach to investigate the complex structure of the proposed model, a “multi-step” modeling attempt was assayed. This stepwise approach sought to disassemble the knotty problem into more manageable and well-established techniques, to establish if the piecewise analysis of nonlinear, mixed, longitudinal models could perhaps produce a satisfactory result.

The methodology of this approach was laid out in a series of steps, summarized as:

Let:

\[ Y_{ij} = f(x_{ij}) + h_i(x_{ij}) + \epsilon_{ij} \]  \hspace{1cm} (4.1a)

where the population function is:

\[ f(x) = \alpha_0 x^{(a_1)} e^{(a_2)x} \]  \hspace{1cm} (4.1b)

the subject-specific deviation is:

\[ h_i(x) = b_{0i} + b_{1i} x + b_{2i} x^2 \]  \hspace{1cm} (4.1c)

with errors normally distributed:

\[ \epsilon_{ij} \sim N(0, \sigma_\epsilon^2) \]  \hspace{1cm} (4.1d)
Assume:
\[ h_i(x) = g_i(x_{ij}^*) \]  
where
\[ x_{ij}^* = x_{ij} - \bar{x}_i \]

1. Fit \( Y_{ij} \) versus \( f(x_{ij}) \) to obtain \( \hat{f}(x_{ij}) \).

2. Fit \( Y_{ij} - \hat{f}(x_{ij}) \) (residuals from Step 1) versus \( x_{ij}^* \) to obtain \( \hat{g}_i(x_{ij}^*) \).

3. Combine \( \hat{f}(x_{ij}) \) and \( \hat{g}_i(x_{ij}^*) \) from the separate fits to get \( \hat{Y}_{ij} \).

### 4.5 Package Inspection

The main statistical packaged used for this analysis is SAS 9.4’s NLMIXED package. Although the final log-transformed model essentially presents a linear function, NLMIXED has the advantage of allowing a syntactically explicit model formulation, as well as more flexibility to calculate transformations of covariates without having to return to a data step to create a quadratic or logged transformation of a covariate. NLMIXED also allows us to return to the non-log-transformed, original model if, after working through so many variations, it might be fit.

However, PROC MIXED has been used for confirmation and comparison, although the particular optimization routines might differ slightly, as well as consideration of PROC GLIMMIX and PROC NONLIN. One main comparison will be presented in the conclusion, otherwise fuller specifications and variations can be found in the Appendices.

NLMIXED offers the additional advantage of allowing distributions other than normal to be specified for the errors of the main function, although it restricts random effects errors to be normally distributed at the time of writing. Although we have assumed normal distributions for the errors, other structures could be explored.
4.6 Log Transformation and Cluster Covariate Effects

After examination of the proposed models, Gary Grunwald suggest a log transformation of the outcome [4], simplifying the population main equation into a linear equation, to which the subject specific random effects could be added. The log transformation thus modified the population equation into:

\[ f'(x) = \alpha'_0 + \alpha_1 \ln(x) + \alpha_2 x \]

With the addition of the Neuhaus-Kalbfleisch covariate effects

\[ Y_{ij} = f'(\bar{x}_i) + g_i(x_{ij} - \bar{x}_i) + \epsilon_{ij} \]

Some discussion was had as to whether the model needed to use \( \bar{x}_i \) or \( x_{ij} \) as the main covariate for the population portion of the overall model, \( f(x) \). Both approaches were tried, centered around \( \bar{x}_i \) and around \( x_{ij} \). In using \( \bar{x}_i \), the model more closely aligns with the notation and model specification used by Neuhaus and Kalbfleisch, who consider this covariate the “cluster-constant” or “cluster-level covariate.” This becomes the between-cluster, or in our case the between-subject component, while the within-cluster components, those which are allowed to vary with identical distribution across the units within each cluster, is noted as \( x_{ij} - \bar{x}_i \). In shorthand, we denote the quantity \( x_{ij} - \bar{x}_i \) as \( x_{ij}^\ast \).

Therefore the fully specified model can be written as

\[ Y_{ij} = f'(\bar{x}_i) + g_i(x_{ij}^\ast) + \epsilon_{ij} \text{ where } x_{ij}^\ast = x_{ij} - \bar{x}_i \]

Or decomposed into main and random effects:

\[
\begin{align*}
    f'(\bar{x}_i) + g_i(x_{ij}^\ast; b_{ki}) &= \alpha'_0 + \alpha_1 \ln(\bar{x}_i) + \alpha_2 \bar{x}_i + \underbrace{\beta_0 + \beta_1 x_{ij}^\ast + \beta_2 (x_{ij}^\ast)^2}_{\text{subject effects}} + \underbrace{b_{0i} + b_{1i} x_{ij}^\ast + b_{2i} (x_{ij}^\ast)^2}_{\text{within-subject effects}} \\
\end{align*}
\]
Similarly to the way in which Neuhaus and Kalbfleisch articulate the simplest
between/within model as an extension of a mixed effects linear model \( a_i + \beta X_{ij} \) into
\( a_i + \beta_B \bar{X}_i + \beta_W (X_{ij} - \bar{X}_i) \); our model uses the nonlinear log-transformed base model
to inform the population terms, with a quadratic function shaping the between- and
within-subject components. Our \( \beta_{ki} \) terms correspond to the \( \beta_B \) components of the
N-K model, with our \( b_{ki} \) terms similar to the \( \beta_W \) components.

We could formulate this model in more classical terms for the subject predictive
values, that is:

\[
\tilde{Y}_{ij} = \hat{\alpha'}_0 + \hat{\alpha}_1 \log(\bar{x}_i) + \hat{\alpha}_2 \bar{x}_i + \hat{b}_0i + \hat{b}_1i \bar{x}_{ij}^\star + \hat{b}_2i (x_{ij}^\star)^2
\]

where \( \hat{\alpha'}_0 \), \( \hat{\alpha}_1 \), and \( \hat{\alpha}_2 \) are maximum likelihood estimates of the fixed effects and \( \hat{b}_0i \),
\( \hat{b}_1i \), and \( \hat{b}_2i \) are predictions of the random effects [ref needed] ; if we consider that the
random effects have non-zero means. This begs the question, and goes to the heart
of the analysis, if the average of the subjects is the same as the population average
at any age \( x \), which the analysis will show is not a correct assumption. Furthermore,
our population average curve at age \( x \) is not \( \bar{Y}(X) = \frac{1}{N} \sum_{i=1}^{N} \tilde{Y}_i(x) \) for the reasons
just mentioned, and that not all subjects will have relevant estimates at age \( x \). For
instance, if we have a young subject that only has observations for ages 13 through
17, it may not make good sense to average an extrapolated value for age 35 for that
subject into the model.

4.7 Interactive Simulation

Lastly, some simulation was attempted with R in order to understand the effects
of parameter variation in the main model and ranges that random effects might
theoretically have on the model. These simulations were carried out in R, with a final
interactive simulation created in RStudio’s Shiny software, allowing parameter ranges
to be varied by slider. This approach was helpful to understand how patterns might
be described in the data, the contribution of individual parameters to those patterns
and what the final model results could be if model convergence was achieved.

In the main model, three parameters, $\alpha_0$, $\alpha_1$ and $\alpha_2$, contribute to the upward U-shaped curve of the population curve. This model and shape as formulated by Strand, looks something like a quadratic form but the particular Gamma-like function allows for some subtle but important distinctions that adhere nicely to the data. In particular, $\alpha_1$ contributes to the steepness of the left-most portion of the curve as it descends through the younger ages (they improve running times). As the curve turns around and lengthens into the older age ranges, $\alpha_2$ dictates its curvature and steepness. This will be illustrated in the simulation results in the next section.

Figure 4.1: 3D Visualization

4.8 Age Grading

Alternative methods of mapping age to running time have been proposed, in particular a knot-and spline function that allows for the population curve to flex at
different knots, or age cutoffs. Although this presents a more complexly specified model in some respects, it does allow for the discussion of what critical ages or age ranges might be tied to physical changes in terms of running performance. We will return to this discussion briefly in the future directions section of conclusions. We would like to point out, however, that age correction formulae and such critical age ranges remain an important issue in road racing literature, in the interests of making age-specific performative evaluations and also in equating performance across gender. In particular, it becomes crucial in the aspect of placement and ranking.

In the March 2003 issue of Measurement News, a newsletter dedicated to the measurement, validation and standardization of distance running, an article by Alan Jones lays out a more complex formula for an age correction approach to standardizing race results. His formula looks similar to a knot-and-spline approach:

\[
f = 1 - B(b - x) - A(a - x)^2 \text{ for } x < a
\]

\[
f = 1 - B(b - x) \text{ for } x \geq a \text{ and } x < b
\]

\[
f = 1 \text{ for } x \geq b \text{ and } x < c
\]

\[
f = 1 - C(x - c) \text{ for } x \geq c \text{ and } x < d
\]

\[
f = 1 - C(x - c) - D(x - d)^2 \text{ for } x \geq d
\]

Jones’ Age Grading Factor graph, while it presents a similar curve to our own population model, identifies 4 ages that are analogous to knots: visually appearing to be approximate ages 16, 21, 33 and 69. These ages, or similar break points, could be used to define age-class categories for categorical comparison.
Figure 4.2: Age Grading, from Measurement News
CHAPTER V
RESULTS

5.1 Findings
In this section we present the findings of the methodology of searching for a robust model that can describe between-subject and within-subjects effects in the road racing data of competitive runners from the Bolder Boulder data set. Most of this analysis has, for simplicity, been carried out on the data from the male runners only, although the female data set was fit with final models. We mention but are not overly concerned with the numerical estimates that the model fitted, and focus more on the methodology and its implications.

5.2 Collinearity Warning
The main model as implemented uses the measures log(age) and age, which are highly collinear. In fact, by investigation this portion of the model through a collinear analysis, we note the high collinearity between the terms and a high Variance Inflation Factor (VIF). Though of concern because of the effect on potentially destabilizing the standard errors of the parameters, we note that this particular model needs both these terms to effectively model the population curve, and dropping one or the other does not adequately model the main trend. We also believe that this is a relatively economical and elegant population model, and so are willing to live with the collinearity as present and noted.

The VIF is quite high at 16.46 for the parameters log(age) and age, when analyzed by PROC REG above a guideline cutoff of 10. Again, we note these concerns but retain the base model as specified for the most reasonable description of the population trend within these competitive running results.)

5.3 Model Comparisons / Progression
The final model for the Bolder Boulder data is a log-transformed nonlinear model that decomposes into population and subject functions. While this model was fit using PROC NLMIXED in SAS for continuities’ sake, the essentially
The linear form of the logged model can also be fit using PROC MIXED or any other mixed linear model software. However, once transformed back to the original scale, NLMIXED becomes necessary. One further feature of the applied model is scaling of the log-transformed outcome, necessary to obtain convergence of the extreme range of parameters and their standard errors that the model encompasses.

The final applied model obtains stable estimates for the population parameters of interest, as well as estimates for the parameters and standard errors of the random effects. These estimates are presented in the table below. In addition, a plot of the predicted within-subject and between-subject effects compared to the population trend is presented in the following figure.

Reprising our model enumeration, the between-subject covariates are centered around the subject means, $x^{-i}$, while the within-cluster components are formulated around $x^{*ij} = x_{ij} - x^{-i}$. We will eventually compare these results with models that use only $x_{ij}$ components. The model reprised:

$$f'(ar{x}_i) + g_i(x^{*ij}; b_{ki}) = \alpha'_0 + \alpha_1 \log(\bar{x}_i) + \alpha_2 \bar{x}_i + \beta_0 + \beta_1 x^{*ij} + \beta_2 (x^{*ij})^2 + b_{0i} + b_{1i} x^{*ij} + b_{2i} (x^{*ij})^2$$

5.4 **Between versus Within**

The advantage and interesting feature of the final model is the separation of between-versus within-effects, particularly within the more complex setup of the nonlinear function (although it has been transformed to a linear form.) In contrast to the Neuhaus and Kalbfleish model, where the parameter $\beta$ which “measures the change in expectation within the $i$th cluster corresponding to a unit increase in the covariate” [8] is separated, this model allows for a more complex specification of the parts of each sub-model, allowing for greater flexibility in potentially contrasting between-and within-behaviours. The separated sections can have show distinct curvatures, as in our case, not adhering to the overall trend.
Table 5.1: Parameter Estimates and their Standard Errors, Final Model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha'_0$</td>
<td>561.48</td>
<td>2.4545</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>-88.6461</td>
<td>1.0441</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>3.1147</td>
<td>0.03437</td>
</tr>
<tr>
<td>res</td>
<td>12.1987</td>
<td>0.3499</td>
</tr>
<tr>
<td>$\beta_{1i}$</td>
<td>-0.4318</td>
<td>0.0846</td>
</tr>
<tr>
<td>$\beta_{2i}$</td>
<td>0.1576</td>
<td>0.01475</td>
</tr>
<tr>
<td>g11</td>
<td>28.8398</td>
<td>1.3881</td>
</tr>
<tr>
<td>g21</td>
<td>-0.6511</td>
<td>0.6167</td>
</tr>
<tr>
<td>g22</td>
<td>5.2303</td>
<td>0.3437</td>
</tr>
<tr>
<td>g31</td>
<td>-0.2846</td>
<td>0.07978</td>
</tr>
<tr>
<td>g32</td>
<td>-0.3313</td>
<td>0.046</td>
</tr>
<tr>
<td>g33</td>
<td>0.04259</td>
<td>0.007537</td>
</tr>
</tbody>
</table>

Without similar formulae for $\beta_B$ and $\beta_W$, this model provides more a more complex relation-ship between the decomposed covariate parts, allowing very different responses in the distinct components of the covariates. The linearized gamma-like population function in this case elegantly and parsimoniously models the main curve of the response for sex across age, while the quadratic form of the subject effects allow a distinct curvature to the responses happening within subject trajectories. In this particular model, modeling subject effects with a similar gamma-like function would not make sense, since the within-subject deviation from the subject’s mean age could have a negative value, and a log of this potential value is not possible.

This disaggregation of the within- and between-subject effects follow as well the methodology explored by Curran and Bauer [2], who investigate the possibility as well as using the grand mean pooling over all individuals (in their notation, $\tilde{z}_{ti}$,) however noting that the person-mean centered predictor is the currently regarded as the best practice for disaggregation. In our case this makes sense too, since a grand mean across all individuals is not informative, subjects having been measured for
only a maximum of twelve years within the time span of the age data that represents adolescence to advanced age.

With this analysis, we can make a statement about the average subject response with respect to the population, although the expression may seem challenging to comprehend. We can make a statement about the average 27-year-old male, for instance, and his response 3 years before he ran his race at 27 (corresponding to a deviation from his mean-centered age). This is not the same statement as for a 24-year-old male.

5.5 Elaboration of Models

The final model was arrived at through exploration of several alternative model formulations and methodologies, each representing evolutionarily more complex elaborations of a previous model or alternative side-branches on the path from the original model to what would end up being the destination. In this section I present a brief
narrative of this exploration, with summaries of the results. The stages can be divided into roughly four sections:

1) adding random effects linearly or exponentially to the original model; 2) the multi-stage approach of modeling the main function and subsequently its residuals; 3) the log-transformed model; and 4) back-transforming to the original scale.

5.6 Adding Random Effects

The original nonlinear model was fit to the new longitudinal data simply with the NLMIXED package. This provided base parameter estimates and a fitted population function as shown in tables and figures below. Fit statistics in the form of Akaike information criterion (AIC) is provided, to compare models’ goodness of fit.

Subsequently, random effects were added linearly to this base model, progressively as a random intercept; a random intercept and slope; and a random intercept and slope and quadratic term. Not all models in this sequence converged, but the results are summarized in the table.

As the models add a random intercept, slope, and attempted quadratic term to the main model, the parameter estimates seem to be converging and standard errors appear to be diminishing. However, the fourth model failed to converge, and we cannot be conclusive in these results.

Table 5.2: Estimates and Standard Errors of Random Effects Models

<table>
<thead>
<tr>
<th>Model</th>
<th>( \alpha_0 )</th>
<th>( \alpha_1 )</th>
<th>( \alpha_2 )</th>
<th>Res</th>
<th>g11</th>
<th>g21</th>
<th>g22</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter Estimates</td>
<td>286.79</td>
<td>-0.9007</td>
<td>0.03161</td>
<td>7.3897</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>02.Random Intercept</td>
<td>272.18</td>
<td>-0.8773</td>
<td>0.03066</td>
<td>3.4303</td>
<td>3.6978</td>
<td></td>
<td></td>
</tr>
<tr>
<td>03.Random Intcpt+Slope</td>
<td>261.95</td>
<td>-0.852</td>
<td>0.02965</td>
<td>3.4104</td>
<td>10.1199</td>
<td>-0.2467</td>
<td>0.007648</td>
</tr>
</tbody>
</table>

In figures, the addition of these terms can be seen in the sequence below, with apparent progress being made toward the hypothesized shape of the data. AIC may not be the most accurate measure for model comparison within the framework of adding random effects to a base model, but for this provisional study will suffice to indicate the trend of improvement.
Figure 5.2: Base Model
Figure 5.3: Adding a Random Intercept
An alternative approach suggested adding the random components to the exponents of the model, either in the ‘first’ part of the model (the component involving $\alpha_1$) or in the ‘second’ part of the base model (the component involving $\alpha_2$). This provided results summarized in the tables and figures following.

Although this approach proved interesting, simulation suggested that there would be restricted success with the exponential approach of this form. The model could not be made flexible enough to describe the between- and within-subject patterns with an approach that only varied the values of the exponential parameters, even with those parameters working in concert.

5.7 Simulation Results

To test the potential and deepen understanding of the approach of adding random effects to the exponents, I used R to simulate data sets through a range of values for either the ‘first,’ ‘second,’ or ‘both’ exponents. These simulations allowed under-
standing of the potential range of parameter values for either component that could be useful, as well as providing visualization of how each component contributed to the final model, and how they interacted together.

![Simulated Range of Left Exponential Parameter](image.png)

Figure 5.5: Simulated Range of Left Exponential Parameter

In the figure above, it’s apparent that adding terms to the first component of the population model affects the curvature of the population trend on the left side, where the steepness of the younger ages approach the minimal peak performance age for men. In contrast, varying the second component contributes to the right side of the model where the population trend more gradually slows down from the peak performance age.

Clearly from Figure 16, the combination of the these two influences has a limited potential to displace and deform the population trend, yet these results do not immediately point toward a formulation that could encompass the within-subject patterns that motivated this exploration.
Figure 5.6: Simulated Range of Right Exponential Parameter

Figure 5.7: Overlay of Simulations
5.7 Multi-Steps Taken
As described in the methods section, the “Multi-Step” approach was devised to tackle a potentially intractable problem through piecewise approach.

The ‘first step’ involved modeling the population trend to achieve estimates for the population parameters, in our model $\alpha_0$, $\alpha_1$, and $\alpha_2$. This model also serves as a check to see if the nonlinear formula of previous research still appears to fit the longitudinal data well. Abbreviated output below shows the results.

The base model yields parameter estimates for the population parameters. Using these results, residuals were found by subtracting model predictions from the observations. This residual data set was then modeled using PROC MIXED with a quadratic formulation.

![Residuals with random effects up to quadratic (VC) Males](image)

Figure 5.8: Residuals After a Base Model

This multi-stage approach gives us results that look similar to what turned out to be the final model, with the kind of patterning between and within-subjects that was sought after. I do not include any estimates for either model, as these are not strictly of interest in this exploratory case, and as we have not striven to combine the estimates mathematically. Furthermore, this base model fitting was carried out on
the original scale (race finish time) and not on the log-transformed outcome.

5.9 Log-Transformed Final Model

It seemed that the attempt to combine some kind of within-subject component involving random effects additively to the nonlinear base model could not provide the sought-after results, nor could random effects added to the exponential terms yield a well-behaved model.

However, it can be noted that a log transformation of the base model neatly alters the problem. The log transformed base model becomes $f'(x) = \alpha'_0 + \alpha_1 \ln(x) + \alpha_2 x$. From this perspective, adding random intercepts, slopes, and a quadratic effect become more tractable. Effects up to cubic were considered, but the resulting G matrix demands optimization of (it seems) too many parameters. In this approach, scaling of the outcome (log finishing time) was needed in order to overcome convergence issues with the standard errors of the G matrix. Although I tried also different convergence
methods, convergence criteria, and a Cholesky root reparameterization, scaling the outcome worked most successfully to provide results.

Technically, the method used for optimization of this problem and this package was first-order optimization and not the default adaptive gauss-hermite quadrature. This FIRO method as implemented in SAS demands that the model be specified as a normal distribution and not an alternate distribution as could be proposed; furthermore it demands a RANDOM statement in the model specification, which suits well our particular problem. Note however, in the model comparisons, that the simplest model specification made for comparison without random terms necessitated a switch to the the default gauss-hermite quadrature (a method for approximating the integral necessary for optimization.)

In the comparison of the table below, we can see the impact of adding random intercept, slope and quadratic terms to the base model. In addition to the changes
in parameter estimates and standard errors, note the decreasing AIC. A comparison across the four models is made in the figure below.

Table 5.3: Estimates and Standard Errors, Log-Transformed Models

<table>
<thead>
<tr>
<th>model</th>
<th>Label</th>
<th>$\alpha_0$</th>
<th>$\alpha_1$</th>
<th>$\alpha_2$</th>
<th>res</th>
<th>B1i</th>
<th>B2i</th>
<th>g11</th>
<th>g21</th>
<th>g22</th>
<th>g31</th>
<th>g32</th>
<th>g33</th>
</tr>
</thead>
<tbody>
<tr>
<td>10a.Final Mixed Estimate</td>
<td>561.48</td>
<td>-88.65</td>
<td>3.11</td>
<td>12.20</td>
<td>-0.43</td>
<td>0.16</td>
<td>28.84</td>
<td>-0.65</td>
<td>5.23</td>
<td>-0.28</td>
<td>-0.33</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td>10b.Final+ri+sl Estimate</td>
<td>564.56</td>
<td>-89.66</td>
<td>3.14</td>
<td>14.22</td>
<td>-0.40</td>
<td>23.89</td>
<td>0.10</td>
<td>4.40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10c.Final+rnd i Estimate</td>
<td>563.74</td>
<td>-89.14</td>
<td>3.11</td>
<td>34.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10d.Final NO rn Estimate</td>
<td>564.06</td>
<td>-88.96</td>
<td>3.10</td>
<td>54.15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10a.Final Mixed SE</td>
<td>2.45</td>
<td>1.04</td>
<td>0.03</td>
<td>0.35</td>
<td>0.08</td>
<td>0.04</td>
<td>1.39</td>
<td>0.62</td>
<td>0.34</td>
<td>0.08</td>
<td>0.05</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>10b.Final+ri+sl SE</td>
<td>2.32</td>
<td>0.99</td>
<td>0.03</td>
<td>0.37</td>
<td>0.08</td>
<td>1.04</td>
<td>0.48</td>
<td>0.28</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10c.Final+rnd i SE</td>
<td>2.27</td>
<td>0.99</td>
<td>0.03</td>
<td>0.74</td>
<td></td>
<td></td>
<td>1.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10d.Final NO rn SE</td>
<td>1.71</td>
<td>0.74</td>
<td>0.03</td>
<td></td>
<td></td>
<td></td>
<td>1.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 5.11: AIC Comparison

5.10 Comparison of Results with Proc Mixed

Because the final applied model is essentially linear, it can also be fitted with PROC MIXED or other linear modeling package. In this case the results in terms of parameter estimates are very similar if not identical to the estimates achieved by NLMIXED. The results below are included for comparison: Maximum likelihood
optimization (instead of the default REML, restricted maximum likelihood) was used in PROC MIXED for comparison with NLMIXED.

### 5.11 Optimization Issues

We briefly note some practical considerations and solutions to optimization, convergence and model-fitting issues.

The approach of log transformation provided several benefits to model-fitting, including stabilizing the tail-end variance of the data set. This issue was also tamed by truncating the age range of research to runners below 65 (the Bolder Boulder results provide sparser data up to ages in the 90s for both sexes). However, before the log transformation a Cholesky root reparameterization proved useful for mathematically simplifying the underlying optimization.

Cholesky root reparameterization breaks down the covariance G matrix into mathematically simpler but related quantities. Since the covariance matrix cannot in NLMIXED be specified as it can within the MIXED procedure, Cholesky provides something like a similar approach to specification. Though perhaps obscure, instructions for Cholesky can be found within the SAS literature and helpful web sites.

Since a requirement for the likelihood to be maximized is a positive definite G matrix, and neither NLMIXED nor MIXED may necessarily produce this positive definite estimate, the Cholesky root reparameterization can overcome this issue. As it estimates G by the product of a lower triangular matrix T and its transpose $T'$, thus $G = TT'$, the estimate of G will be positive definite. This can be accomplished by other programming methods in NLMIXED and MIXED [13]

---

<table>
<thead>
<tr>
<th>model</th>
<th>Label</th>
<th>$a_0^*$</th>
<th>$a_1$</th>
<th>$a_2$</th>
<th>res</th>
<th>B1i</th>
<th>B2i</th>
<th>g11</th>
<th>g21</th>
<th>g22</th>
<th>g31</th>
<th>g32</th>
<th>g33</th>
</tr>
</thead>
<tbody>
<tr>
<td>10a.Final Mixed</td>
<td>Estimate</td>
<td>561.48</td>
<td>-88.65</td>
<td>3.11</td>
<td>12</td>
<td>-0.43</td>
<td>0.16</td>
<td>28.84</td>
<td>-0.65</td>
<td>5.23</td>
<td>-0.28</td>
<td>-0.33</td>
<td>0.04</td>
</tr>
<tr>
<td>PROC MIXED</td>
<td>Estimate</td>
<td>560.83</td>
<td>-88.35</td>
<td>3.11</td>
<td>13</td>
<td>-0.21</td>
<td>0.065</td>
<td>24.72</td>
<td>0.31</td>
<td>0.92</td>
<td>-0.21</td>
<td>0.027</td>
<td>0.015</td>
</tr>
<tr>
<td>10a.Final Mixed</td>
<td>SE</td>
<td>2.45</td>
<td>1.04</td>
<td>0.35</td>
<td>0.08</td>
<td>0.01</td>
<td>1.39</td>
<td>0.62</td>
<td>0.34</td>
<td>0.08</td>
<td>0.05</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>PROC MIXED</td>
<td>SE</td>
<td>1.92</td>
<td>0.84</td>
<td>0.03</td>
<td>0.45</td>
<td>0.01</td>
<td>1.39</td>
<td>0.62</td>
<td>0.34</td>
<td>0.08</td>
<td>0.05</td>
<td>0.01</td>
<td></td>
</tr>
</tbody>
</table>
In our example, three random effects necessitated a 3x3 covariance matrix to be estimated.

\[
\text{random } b0i \ b1i \ b2i \ \text{normal([0, } \mu1i, \mu2i],

g11, \ g21, \ g22, \ g31, \ g32, \ g33
\) subject=id;
\]

Asking NLMIXED to directly estimate these matrix terms in some cases failed, or thwarted convergence of the model. Although this approach will not be applicable to every problem, a Cholesky reparameterization solved the issue:

\[
g11 = a*a;
g21 = a*b;
g22 = b*b + c*c;
g31 = a*d;
g32 = b*d + c*e;
g33 = d*d + e*e + f*f;
\]

These terms are then asked for as estimates from the procedure:

\[
\text{estimate 'g11' } a*a; \\
\text{estimate 'g21' } a*b; \\
\text{estimate 'g22' } b*b + c*c; \\
\text{estimate 'g31' } a*d; \\
\text{estimate 'g32' } b*d + c*e; \\
\text{estimate 'g33' } d*d + e*e + f*f;
\]

By decomposing the positive definite matrix into a lower triangular matrix and its conjugate transpose, the Cholesky reparameterization aids in achieving efficient numerical solutions. [7]
CHAPTER VI
DISCUSSION

As we have shown, the articulation of between-subject and within-subject effects within a complex (nonlinear) mixed model can be achieved, and benefits the model specification. In this case of competitive running data, the articulation of the effects enriches our understanding of the performance of runners as they age. Runners do not simply follow the population trend by improving until a peak performance age and then declining in performance. Within that broad trend, individual subjects articulate specific trajectories that differ from others in their own category, and show improvement and decline relative to their own capabilities.

If we ‘zoom in’ on a section of the data and randomly filter records out for clarity, we can see more clearly the model predicting quadratic trajectories modified by information from the population trend (Figure 22). Actual data points are represented by color-coded symbols, the population trend in a heavy black line, and individual modeled curves in colored and patterned lines.

Compared to a zoomed in spaghetti plot (created in JMP, which uses a knotted spline smoother to describe a curve between data points), our model takes information from the population trend.

It seems from visual inspection as we approach the data limits (either the youth or the aged) that subject curves flatten out and adhere more to the population trend. This potential effect could be an distortion of the vertical scale of the data, or could be tested formally through further work.

In addition, we wonder if particular age groups show more of a ‘improvement,’ ‘steady,’ or ‘declining’ trend compared to other groups. If we return to an early question of age grading, we could divide the population into age factors, and analyze and estimate the linear and quadratic trends of the random effects of these groups. If most aged runners show a positive intercept (slower than the population) and positive
Figure 6.1: Zoomed In Fit
Figure 6.2: JMP Smoother
quadratic coefficient, it might seem that a certain portion of these aged runners start slower, improve, and then decline with respect to the population. Perhaps other age categories show similar or distinct patterns.
CHAPTER VII
CONCLUSION

The Bolder Boulder race results present a fascinating case study in the application of between-subject and within-subject modeling for a complex mixed model. The data presents itself unconventionally, in that only twelve years of race data are (in this study) incorporated, but when reorganized provide information on a full spectrum of runners’ ages as they relate to finishing time outcomes. Although the base nonlinear population function well models this data that represents a class of competitive runners, it can be enriched through the incorporation of between- and within-subject effects constructed around components that use the mean subjects’ ages and deviations from these ages. The incorporation of these effects improve the model, providing information that can’t be captured by ignoring these effects.

Although this data set presented problematic issues adherent to its complexity, with time and creative thinking these challenges can be overcome, and perhaps the methods attempted can prove applicable to similar data and modeling issues.

We take inspiration from Galloway’s ethos in running, and find it pertinent to the problem at hand:

Real growth in running occurs when you pull yourself out of the motivational dumps, learn from the problem at hand, try a few new things, and suddenly find yourself looking at running in a different way. (13)[3]

Swap out the word ‘running’ for ‘biostatistics’ or another discipline, and we think you’ll find a winning philosophy.
REFERENCES


