

COMBINED CROSS-SECTIONAL AND LONGITUDINAL METHODOLOGY ALLOWS  
ACCURATE MODELING OF GROWTH IN RARE DISEASES

By

DANIEL CHARLES TARQUINIO DO

A THESIS PRESENTED TO THE GRADUATE SCHOOL  
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT  
OF THE REQUIREMENTS FOR THE DEGREE OF  
MASTER OF SCIENCE

UNIVERSITY OF FLORIDA

2008

© 2008 Daniel C. Tarquinio DO

To my mom Marianne, and my wife Keiko

## ACKNOWLEDGMENTS

I am grateful to my mentor at the University of Florida, Dr. Paul Carney, who prompted me to forge new connections in the field of neurology, thus sparking this project and changing the course of my career. Also my mentor at UF, I'd like to thank Dr. Daniel Driscoll, an outstanding clinical researcher and basic scientist who, with his encouraging skepticism, has driven me to prove myself as a researcher while at the same time offering a helping hand whenever I needed one. For teaching me how to write in her brilliant and dynamic style, and then patiently, yet never subtly, correcting my wayward tendencies, I thank Jane Yellowlees Douglas. Dr Marian Limacher and Eve Johnson are the bedrock of the Advanced Postgraduate Program in Clinical Investigation at UF, and I am greatly indebted to them for their continual efforts to train new clinical researchers. In the department of biostatistics I'd like to thank Dr. Cyndi Garvan for teaching me how to think really far outside the box. Her ability to inspire new ideas is unmatched. Also in biostatistics my thanks go out to Dr. Wei Hou who has struggled tirelessly to push the envelope of growth modeling statistics. I sincerely hope we continue on to achieve new breakthroughs in this field together.

Several people in the Department of Pediatrics at UF helped shape and refine my idea of what it means to be a physician. I thank Dr. Maureen Novak, pediatric residency program director, for seeing me through difficult trials with indefatigable optimism and reassurance; and Dr. Robert Lawrence for embodying the archetypal caring and competent physician and teaching me the meaning of professionalism.

I would not have been able to accomplish any of this work without the guidance and support of Dr. Alan Percy, my mentor at the University of Alabama at Birmingham. A master of the mentoring art, Alan advises with timing, moderation, and finesse. While giving prompt support when needed, he continually challenges me to accept new responsibilities and develop

autonomy. A superb clinician and researcher, his dedication and compassion as a child neurologist have set a benchmark for me to aspire to. Also at UAB, I am very grateful to Jane Lane, the rare disease research project coordinator who is always ready to face any problem or find the answer to any question, and who introduced me to database design, and to Susie Geertz, a wonderful nutritionist and anthropometrist who has always been eager to share advice and discuss ideas, and to dig through her library to find references on growth for me.

I cannot thank my mom enough, for her tireless proofreading of this work, and also for raising me to see every obstacle as an opportunity, to never think anything is impossible, and to find fun in everything! Lastly, I'd like to thank my dear wife Keiko for agreeing to share her life with me. Keiko has been a constant source of support, understanding, and tenderness. I couldn't have done it without her.

# TABLE OF CONTENTS

	<u>page</u>
ACKNOWLEDGMENTS .....	4
LIST OF TABLES .....	11
LIST OF FIGURES .....	13
LIST OF EQUATIONS.....	17
LIST OF ABBREVIATIONS.....	18
ABSTRACT .....	22
CHAPTER	
1 INTRODUCTION .....	24
Utility of Growth Reference Charts .....	24
Fundamental Concepts Regarding Growth Reference Charts .....	28
Purpose of Growth Standards .....	28
Interpretation of Growth Standards.....	29
2 HISTORY OF REFERENCE GROWTH CHART CONSTRUCTION.....	32
Early Methodologies for Growth Chart Construction.....	32
Eye Smoothing Techniques .....	32
Data Collection Techniques.....	33
Nationwide and International Surveys .....	35
Mathematical Smoothing Techniques.....	35
International Standards.....	38
Refining Statistical Methodology .....	40
Continuous Versus Discrete Units of Age.....	40
Distributional Assumptions .....	41
Growth References versus Growth Standards .....	44
Persistent Issues Concerning Growth Standards .....	45
Growth Velocity.....	46
Hybrid Size and Velocity Charts .....	48
Confounding Factors .....	49
3 HISTORY OF SPECIALIZED GROWTH CHARTS.....	51
Growth References Based on Racial, Ethnic, and Socioeconomic Disparities .....	51
Early Efforts at Specialized Growth References .....	51
Racial and Ethnic Differences in Growth.....	51
National Growth References versus International Growth References .....	53

Differences in Growth Due to Genetic Conditions.....	55
Down Syndrome .....	55
Turner Syndrome .....	58
Marfan Syndrome.....	60
Williams Syndrome.....	61
Prader-Willi Syndrome.....	62
Cri-Du-Chat Syndrome .....	63
Fragile X Syndrome .....	64
Neurofibromatosis Type I.....	64
Congenital Adrenal Hyperplasia.....	66
Beckwith-Wiedemann Syndrome .....	67
Rubinstein-Taybi Syndrome.....	67
Silver-Russell Syndrome.....	68
Brachmann-de Lange Syndrome.....	68
Duchenne Muscular Dystrophy .....	68
Achondroplasia .....	69
Noonan Syndrome.....	69
Klinefelter Syndrome .....	70
Rett Syndrome .....	70
Non-Genetic Etiologies of Altered Growth.....	74
Cerebral Palsy .....	74
Human Immunodeficiency Virus.....	75
Summary.....	75
4 CASE FOR SPECIALIZED GROWTH CHARTS.....	77
Case Against Specialized Growth Charts.....	77
Counterargument to the CDC Perspective .....	78
Itemized Response to the CDC Critique of Specialized Growth Charts.....	83
Sample Size .....	83
Racial, Ethnic, and Geographic Diversity .....	83
Recency of Data .....	84
Representativeness of Data to Population as a Whole .....	84
Reliability of Data.....	85
Secondary Medical Conditions.....	85
The Role of Disease-Specific Growth Standards.....	85
Counseling .....	85
Detection of Comorbid Diseases.....	86
Understanding the Pathogenesis of the Growth Disorder .....	89
Monitoring Growth-Promoting and Other Treatments .....	90
Should Specialized Growth Charts Be Used for Diagnosis?.....	92
Criteria for a Valid Standard of Disease Specific Growth .....	95
5 RETT SYNDROME: A MODEL FOR SPECIALIZED GROWTH PATTERNS.....	104
Introduction.....	104
Epidemiology.....	104

Clinical Features .....	104
Diagnostic Criteria .....	107
<i>MECP2</i> Mutations.....	107
Genetic Testing .....	110
Neurobiology and Pathophysiology.....	110
Management .....	112
Summary.....	114
6 GROWTH CHART METHODOLOGY SPECIFIC TO RARE DISEASES .....	117
Approach to Study Design.....	117
Study Goals and Design .....	118
Cross-sectional, Longitudinal, and Hybrid Approaches .....	119
Methodology for Data Collection .....	122
Anthropometric Measurement Techniques .....	123
Inter-Observer and Intra-Observer Reliability Assessment .....	123
Statistical Methodology .....	123
Power Analysis .....	124
Data Cleaning.....	124
Techniques for Estimating and Smoothing Percentiles.....	130
Eye Smoothing Versus Mathematical Functions.....	131
Constrained Versus Independent Percentiles.....	131
Choice of Mathematical Function.....	132
Data Distribution.....	133
Transformations .....	134
Choice of Percentiles.....	136
Outliers .....	137
EDF parameters.....	138
Edge Effects.....	138
Age as an Interval or a Continuous Variable.....	139
Summary.....	139
Diagnostic Guides to Balance Smoothing and Fidelity to Data .....	140
Visual Inspection.....	140
Comparison of Observed and Expected Values.....	141
Tests of Normality for the Z-scores .....	141
Quantile-Quantile plot of Z-scores.....	142
Worm Plots .....	142
Techniques for Analyzing and Summarizing Longitudinal Growth Data .....	143
Correlations.....	144
Regression Models .....	144
Techniques for Statistical Comparison of Results .....	145
7 PATTERNS OF GROWTH IN RETT SYNDROME .....	159
Introduction.....	159
Methods .....	161
Participants .....	161

Reliability Testing.....	166
Statistical Analysis.....	166
Results.....	171
Measurements.....	171
Summary Statistics.....	173
Growth Curve Fitting.....	174
Primary Hypothesis: Growth Charts for Classic RTT.....	177
Secondary Hypothesis 1: Height, Weight and FOC are Decreased in RTT Relative to Healthy Children.....	178
Secondary Hypothesis 2: Differences in Growth Exist within RTT Subpopulations...	180
Secondary Hypothesis 3: Growth Velocity in RTT is Decreased Relative to the Normal Population.....	186
Secondary Hypothesis 4: Early Characteristics of RTT Predict Growth Later in Life.	188
Discussion.....	190
 8 FUTURE OF GROWTH STANDARDS FOR SPECIAL POPULATIONS.....	 216
Current Successes in Growth Standard Design.....	216
Applying Growth Standard Principles to Rett Syndrome.....	216
Future of Statistical Modeling of Growth Charts.....	217
Do Disease-Specific Growth Standards Need to be Updated Regularly?.....	218
 APPENDIX	
 A MEASUREMENT TECHNIQUES.....	 220
Basic Guidelines for Length Measurements Including Hands, Feet, Head Circumference and Height.....	220
Weight.....	220
Height.....	221
FOC.....	222
Hand length.....	223
Foot length.....	223
Skin-fold measurements.....	224
Circumference measurements.....	226
Leg length measurement.....	227
Body Mass Index.....	228
 B ANTHROPOMETRIC MEASUREMENT RELIABILITY.....	 229
Abstract.....	229
Introduction.....	230
Methods.....	232
Participants.....	232
Procedure.....	233
Data Analysis.....	234
Results.....	235

Discussion.....	236
C DATA CLEANING TECHNIQUES .....	240
Data preparation.....	240
Error Detection.....	240
Examination of Trends and Detection of Outliers .....	240
D ITEMS FROM STUDY FORMS .....	242
Initial Forms.....	242
Every 6 months (under 12y) or annually (over 12y).....	243
E DIAGNOSTICS FOR CLASSIC RETT SYNDROME GROWTH CHARTS.....	245
F LMS VALUES FOR GROWTH CHARTS .....	257
G GROWTH CHARTS.....	259
LIST OF REFERENCES .....	273
BIOGRAPHICAL SKETCH.....	286

## LIST OF TABLES

<u>Table</u>	<u>page</u>
4-1 Change in height in two individuals with RTT and the corresponding Z-scores.....	97
4-2 Frequency of measurements in the 2000 CDC and 2007 WHO reference chart studies compared to this RTT growth study.....	98
5-1 Stages of RTT.....	115
5-2 Major diagnostic criteria for RTT. ....	115
5-3 Diagnostic criteria for atypical RTT (Must meet 3 main criteria and 5 supportive criteria).....	116
6-1 WHO criteria for an ideal growth standard model.....	148
7-1 Participants by diagnosis (frequency).....	194
7-2 Measurements by diagnosis (frequency). ....	194
7-3 Observations based on specific mutations (frequency). ....	195
7-4 Observations based on mutation type (frequency).....	195
7-5 Observations based on mutation site (frequency). ....	196
7-6 Overall observations at specific age intervals (frequency).....	197
7-7 Classic RTT observations at specific age intervals (frequency). ....	198
7-8 Classic RTT. Birth weight, length FOC compared to CDC reference. ....	199
7-9 Minimum, maximum and mean weight, height, FOC and BMI in classic RTT.....	199
7-10 Differences between calculated Z-scores and ranked percentiles in different age groups, including difference between mean and 50th percentile, maximum and minimum difference across percentiles (expressed as absolute values and percentages of the mean).....	200
7-11 Classic RTT growth. Age at which mean and SD significantly different from three standard references. ....	200
7-12 Comparison of mean and SD among different subgroups, revealing age at which differences in curves appear.....	201
7-13 Type of mutation in classic RTT and adult Weight, Height, and FOC. ....	201

7-14	Site of mutation in classic RTT and adult weight, height, and FOC.....	202
7-15	Specific mutations in classic RTT and adult weight and height. ....	203
7-16	Specific mutations in classic RTT and adult FOC, and BMI.....	204
7-17	Ambulation at most recent visit compared to adult weight, height, and FOC in classic RTT.....	205
7-18	Age at regression compared to adult weight, length, FOC and BMI in classic RTT.....	206
7-19	Purposeful hand use at most recent visit compared to adult weight and height in classic RTT. ....	206
7-20	Ability to speak at most recent visit compared to adult weight, height, FOC, and BMI in Classic RTT.....	207
7-21	Age at which individual with Classic RTT first sat compared to adult weight, height, FOC, and BMI.....	208
7-22	Seizure Severity classifications in Classic RTT compared to adult weight, height, FOC, and BMI.....	209
7-23	Overall Severity Score (CSS and MBA) classifications in Classic RTT compared to adult weight, height, FOC, and BMI. ....	209
B-1	Interpretation of Standardized Reliability Scores .....	239
B-2	Interoperator Reliability Scores based on the ICC, PMCC, and TEM.....	239
B-3	Intraoperator Reliability Scores based on the ICC, PMCC, and TEM.....	239
B-4	Values for Mean Absolute Deviation .....	239

## LIST OF FIGURES

<u>Figure</u>	<u>page</u>
4-1 CDC percentiles for height with classic RTT mean.....	99
4-2 CDC height Z-score chart with classic RTT mean Z-score plotted against age. ....	100
4-3 Examples of how trends can be misinterpreted if non-specialized references are used for genetic syndromes.....	101
4-4 Examples of how using Z-scores from standard references can misinterpret the influence of interventions in genetic syndromes.....	102
4-5 Differences in mean FOC among four commonly used references. ....	103
6-1 Example of Exploratory Data Analysis used to identify outliers and understand differences in dispersion among groups. ....	149
6-2 Boxplots of age groups from birth to adulthood, demonstrating the change in the overall shape of the distribution and the presence of outliers at certain ages.....	150
6-3 Height of 15 individuals plotted longitudinally, demonstrating aberrant measurements which do not follow the typical curve. ....	151
6-4 Scatterplot for data cleaning, demonstrating outliers and clustering at certain ages. ....	152
6-5 Birth weight histogram, demonstrating negative skewness.....	153
6-6 Normal Q-Q plot demonstrating skewness in distribution. ....	154
6-7 Curves undersmoothing data.....	154
6-8 Curves oversmoothing data.....	155
6-9 Curves appropriately smoothing data. ....	155
6-10 Weight transformed using log scale (ln).....	156
6-11 Normal Q-Q plot, demonstrating effect of log transformation (ln).....	156
6-12 Weight-by-age chart demonstrating right edge effect in a sparse data set. ....	157
6-13 Detrended Q-Q plot (worm plot), showing significantly skewed data.....	158
7-1 Participants' years of birth (frequency).....	210
7-2 Ages at enrollment in RDCRN study (frequency). ....	211

7-3	Ages at diagnosis with Rett Syndrome (frequency).....	212
7-4	Simplified mutation types by diagnosis (frequency).....	213
7-5	Specific mutation types by diagnosis (frequency).....	214
7-6	Mutation sites on MECP2 by diagnosis (frequency).....	215
E-1	Weight data plotted on curves.....	245
E-2	Weight power transformation curve.....	245
E-3	Weight mean curve.....	245
E-4	Weight coefficient of variation curve.....	246
E-5	Mean weight velocity curve.....	246
E-6	Mean weight acceleration curve.....	246
E-7	Weight Z-scores plotted against age.....	247
E-8	Weight detrended Q-Q plot.....	247
E-9	Weight Q-statistic data.....	247
E-10	Height data plotted on curves.....	248
E-11	Height power transformation curve.....	248
E-12	Height mean curve.....	248
E-13	Height coefficient of variation curve.....	249
E-14	Mean height velocity curve.....	249
E-15	Mean height acceleration curve.....	249
E-16	Height Z-scores plotted against age.....	250
E-17	Height detrended Q-Q plot.....	250
E-18	Height Q-statistic data.....	250
E-19	FOC data plotted on curves.....	251
E-20	FOC power transformation curve.....	251
E-21	FOC mean curve.....	251

E-22	FOC coefficient of variation curve.....	252
E-23	Mean FOC velocity curve.....	252
E-24	Mean FOC acceleration curve.....	252
E-25	FOC Z-scores plotted against age. ....	253
E-26	FOC detrended Q-Q plot.....	253
E-27	FOC Q-statistic data. ....	253
E-28	BMI data plotted on curves.....	254
E-29	BMI power transformation curve.....	254
E-30	BMI mean curve.....	254
E-31	BMI coefficient of variation curve.....	255
E-32	Mean BMI velocity curve. ....	255
E-33	Mean BMI acceleration curve.....	255
E-34	BMI Z-scores plotted against age.....	256
E-35	BMI detrended Q-Q plot.....	256
E-36	BMI Q-statistic data.....	256
G-1	Weight in classic RTT (blue) compared to normal reference (orange).....	259
G-2	Height in classic RTT (blue) compared to normal reference (orange).....	260
G-3	FOC in classic RTT (blue) compared to normal reference (orange).....	261
G-4	BMI in classic RTT (blue) compared to normal reference (orange).....	262
G-5	Weight in classic (orange) versus atypical RTT (blue). ....	263
G-6	Height in classic versus atypical RTT. ....	264
G-7	Height in classic (orange) versus atypical RTT (blue). ....	265
G-8	Weight in severe RTT (orange) versus mild RTT (blue).....	266
G-9	Height in severe RTT (orange) versus mild RTT (blue). ....	267
G-10	FOC in severe RTT (orange) versus mild RTT (blue). ....	268

G-11	Weight in RTT children born before 1997 (blue) and after 1997 (orange).....	269
G-12	Height in RTT children born before 1997 (blue) and after 1997 (orange).....	270
G-13	FOC in RTT children born before 1997 (blue) and after 1997 (orange).....	271
G-14	Weight in RTT children born before 1997 (blue) and after 1997 (orange).....	272

## LIST OF EQUATIONS

<u>Equation</u>	<u>page</u>
6-1 Formula for calculating Z-scores for a particular measurement using L, M, and S.....	138
B-1 Intraclass Correlation Coefficient (ICC).....	235
B-2 Variable Average Value (VAV).....	235
B-3 Relative Technical Error of Measurement (RelativeTEM) .....	235
B-4 Absolute Technical Error of Measurement (TEM) .....	235
B-5 Pearson Product-Moment Correlation Coefficient (PMCC).....	235

## LIST OF ABBREVIATIONS

3'UTR	3' untranslated region
5'UTR	5' untranslated region
A-1	Preserved speech subtype of atypical Rett syndrome
A-2	Minimal evidence of RTT
A-3	Subtype of atypical Rett syndrome with no hand use regression
A-4	Early onset seizures subtype of atypical Rett syndrome
A-5	Delayed onset subtype of atypical Rett syndrome
AAP	American Academy of Pediatrics
AS	Angelman syndrome
BDLS	Brachman-de Lange syndrome
BDNF	Brain-derived neurotrophic factor
BMI	Body mass index
BWS	Beckwith-Wiedemann syndrome
CAH	Congenital adrenal hyperplasia
CDCS	Cri-du-Chat syndrome
CP	Cerebral palsy
CR	Coefficient of reliability
CRH	Cortisol releasing hormone
CSS	Clinical Severity Score
CV	Coefficient of variation (equivalent to the SD divided by the mean)
CDC	Centers for Disease Control
DMD	Duchenne muscular dystrophy
DOB	Date of birth
DOV	Date of visit

EDA	Exploratory data analysis
EDF	Estimated degrees of freedom
EEG	Electro-encephalogram
EKG	Electrocardiogram
FDR	False detection rate
FTT	Failure to thrive
FOC	Fronto-occipital head-circumference
FXS	Fragile X syndrome
GER	Gastro-esophageal reflux
GH	Growth hormone
GHD	Growth hormone deficiency
$H_0$	Null hypothesis
HIV	Human immunodeficiency virus
HT	Height
HDAC	Histone deacetylase
ICC	Intraclass correlation coefficient
IDR	Interdomain region
IPH	Index of potential height
IQR	Interquartile range
IRB	Institutional Review Board
L	Lambda (Box-Cox power adjustment for skewness)
LMS	Estimated degrees of freedom: L=lambda (Box-Cox power adjustment for skewness), M=mu (median), S=sigma (coefficient of variation)
LMSP	LMS plus Box-Cox power exponential
LMST	LMS plus Box-Cox t-distribution
LLL	Lower leg length

M	Mu (median)
MAD	Mean absolute deviation
MBA	Motor Behavioral Assessment
MBD	Methyl binding domain
<i>MECP2</i>	Methyl CpG Binding Protein 2 gene
MeCP2	Methyl CpG Binding Protein 2
MGRS	Multicentre Growth Reference Study
MR	Mental retardation
NF1	Neurofibromatosis type 1
NHANES	National Health and Nutrition Examination Survey
NHES	National Health Examination Survey
NLS	Nuclear localization signal
non-RTT	Individuals with <i>MECP2</i> mutations who do not fulfill criteria for classic or atypical RTT
NS	Noonan syndrome
NTR	N-terminal region
P-P	Detrended probability-probability
PedNSS	Pediatric Nutrition Surveillance System
PI	Principal investigators
PMCC	Pearson product-moment correlation coefficient
PWS	Prader-Willi syndrome
Q-Q	Detrended quantile-quantile
RDCRN	Rare Diseases Clinical Research Network
RelativeTEM	Relative Technical Error of Measurement
RTS	Rubinstein-Taybi syndrome
RTT	Rett syndrome

S	Sigma (coefficient of variation)
SD	Standard deviation
SGA	Small-for-gestational-age
SNR	Signal to Noise Ratio
SRS	Silver-Russell syndrome
TEM	Technical error of measurement
TRD	Transcription repressor domain
VAV	Variable Average Value
VEEG	Video Electro-encephalogram
WGA	Weeks gestational age
WHO	World Health Organization
WIC	Special Supplemental Nutrition Program for Women, Infants, and Children
WT	Weight
XCI	X-chromosome inactivation

Abstract of Thesis Presented to the Graduate School of the University of Florida in Partial  
Fulfillment of the Requirements for the Degree of Master of Science

A COMBINED CROSS-SECTIONAL AND LONGITUDINAL METHODOLOGY ALLOWS  
ACCURATE MODELING OF GROWTH IN RARE DISEASES

By

Daniel C. Tarquinio DO

May 2008

Chair: Richard Bucciarelli  
Major: Medical Sciences

Many genetic diseases involve abnormal patterns of growth, and few suitable growth references for these diseases exist. Of the specialized growth references in use, most suffer significant methodological issues preventing their use in clinical or research settings as standards of growth. Because of issues with sample size, bias, reliability and secular trends, the Centers for Disease Control (CDC) have suggested that disease-specific charts should not be used. However, adequate growth references addressing all these concerns can be produced using a hybrid study design. In particular, disease specific growth charts would prove useful in Rett syndrome (RTT), a disease with early growth failure and potential for improvement with early recognition and intervention.

Recruiting individuals from a large natural history study of RTT, we collected prospective and retrospective data on growth and other relevant variables. We conducted a reliability study to ensure consistency among our anthropometrists and collected information about nutrition, therapy, and overall disease severity. We analyzed these data with a semiparametric method and compared these results to the reference population using student's t-test.

We reviewed the literature on rare disease growth references, and, using the most suitable techniques, developed the first comprehensive set of growth references for RTT. We found that

growth is significantly different from the normal healthy population at 14 months for weight, 21 months for height, and 15 months for head circumference. Remarkably, BMI was similar in RTT and normal individuals. In secondary longitudinal analysis of growth velocity, we found that individuals with missense mutations in Methyl CpG Binding Protein 2 gene (*MECP2*) grow faster than those with nonsense mutations, and that R133C confers a better growth prognosis than R168X, R255X, and R270X. We also noted higher growth velocity in R306C and C-terminal deletions compared to others.

Our growth charts meet the criteria for reference charts put forth by the CDC, and demonstrate that specialized populations do not pose an insurmountable obstacle to growth reference construction. Researchers will now be able to examine the effects of novel treatments in RTT with a higher degree of precision. Future research on growth will expand the current methodology to include longitudinal correlation analysis within and between subjects. Such an adjustment will refine statistical predictions about individuals and allow researchers to tailor power analyses to detect much more subtle effects of interventions. Moreover, this step will help us to develop prescriptive standards for growth in rare diseases and improve general clinical management.

## CHAPTER 1 INTRODUCTION

### **Utility of Growth Reference Charts**

Growth reference charts are a scientific tool used regularly by physicians to screen children for disease and follow their progress in health. In design, these references are meant to describe the average size of children at particular ages and the dispersion of normal size. However, most physicians consider these references more than a record of how children *do* grow. In actuality, rather than viewing growth charts as descriptive references, physicians see them as prescriptive - as a standard of how children *ought* to grow. Clinical physicians and researchers alike perceive these references as the result of rigorous data collection coupled with state-of-the-art statistical methodology.

Alternatively, the foremost expert in the field of growth reference statistics has described their creation on numerous occasions as “a black art” (Cole 1988; Cole 1990). My goal is to resolve this disconnect between growth charts as a somewhat reliable screening reference and growth standards as a scientific tool - between descriptive and prescriptive. In this study, I will demonstrate that growth reference charts can be created for anthropometric parameters, specific to the general population or to any subpopulation, in a methodical and reproducible manner which any research scientist could employ. Moreover, such charts can be designed with reasonable measurements of tolerance describing their fidelity to the empirical data. Researchers can use these measurements of fidelity to establish the validity of the reference charts in the population of interest. To develop this argument, I will use data on numerous growth parameters extracted from a population of patients with the rare neurodevelopmental disease Rett syndrome (RTT). I will discuss different methodologies and when I have selected an acceptable method, I will examine its performance in evaluating subpopulations within this group. Ultimately, I will

demonstrate a practical approach to developing growth references and standards in rare diseases for both clinical and research use.

Historically, physicians have employed growth reference charts as a screening tool. Admittedly, growth charts are nothing more than a compilation of a random sample of observed values, and as such the creators of these reference values have cautioned physicians against interpreting individual patient values as diagnostic of disease. The primary value of these charts has been to screen – to serve as a sensitive tool for alerting the clinician of potential disease while not specifically implying the presence of disease or suggesting an imminent course of action. If these charts were interpreted as they were intended, a rough screen for extreme values, many clinicians would find them of little use. However, the reality is that most clinicians ignore this admonition, and instead of viewing these charts as how children “do” grow for the most part, physicians interpret these charts as how children “ought” to grow.

The current United States reference charts developed in 2000 for standard parameters of growth – weight, height, fronto-occipital head-circumference (FOC), and body mass index (BMI) – while an improvement over the 1977 reference charts, are based on data which was never evaluated for its validity or reliability. Data on children for charts from birth to age three were collected longitudinally, and data for charts from age 2 to 18 were collected cross-sectionally. Not only the types of data, but also the geographic and ethnic sources were dramatically different for these two segments, yet they were evaluated and merged using the same statistical methodology. Unbeknownst to most clinicians, the issue of childhood obesity was of such concern to the committee designing the 2000 reference charts that weight data collected from 1977 to 2000 was simply discarded. Therefore, the current charts represent the height of children measured up to the year 2000, but the weight of children measured up to 1975.

Despite these shortcomings, physicians regularly employ the 2000 CDC charts not only as a screening tool, but also as a guideline of how children should grow.

The current clinical paradigm with respect to other reference values, such as cholesterol in hypercholesterolemia and hemoglobin A1C in diabetes, is to compel patients to strive for “normal” standard values. This paradigm of treating reference values as standards has proven an effective strategy in these diseases. However, such a strategy with respect to growth is untenable. The principle of the growth standard is not flawed – the references themselves are. Growth reference charts can be as useful as laboratory values in both diagnosing disease and measuring the effects of management. Just as standard laboratory values can vary by age, disease state, family history, risk factors, gender, and even genetic background, growth references must incorporate these variables if they are to be used to their fullest extent. All modern reference curves incorporate standardized values or z-scores which offer a precise numerical representation of the relationship of an individual to the mean within that population. When the variance within the general population is similar to that of the subpopulation of interest (homoschedasticity), z-scores based on the general population can validly represent individuals in the target subpopulation. However, when variances are unequal (heteroschedasticity) or the values of individuals in the subpopulation are at extreme distances from the mean of the general population (Z-score greater than 2.67 or less than -2.67), these values are difficult to interpret. In such cases reference ranges specific to the subpopulation are useful for determining whether values are indicative of the presence of disease or for monitoring response to an intervention.

Historically, data collection for growth charts has been a labor intensive, inconsistent, and somewhat haphazard venture. We are on the cusp of an explosion of data resources available as both hospitals and community clinics transition to electronic medical records. This transition

will address several issues surrounding data collection. Researchers will have the vast numbers of individuals from different geographic, socioeconomic and ethnic regions necessary to conduct population based studies. The issue of reliability has always clouded the validity of growth reference charts, and electronic data collection will facilitate assessment of reliability in individual settings. As data continue to accrue, researchers will have sufficient numbers to assess the validity of charts produced using a data set by confirming their residual error against a second data set. These data will contain variables which can be analyzed using cluster analysis to establish ethnic, racial, geographic, nutritional, socioeconomic or other trends in growth, or simply to weed out confounders. Furthermore, growth charts for specific diseases can be produced and will serve as a valuable research tool to measure the effect of interventions in these diseases.

To establish the case for disease-specific growth references, I will begin with a history of the methodology of reference growth chart construction. I will proceed to discuss several recent attempts at comparing growth in both subpopulations of healthy individuals and populations of diseased individuals. I will develop a case for choosing one flexible methodology to evaluate both general populations and subpopulations. Having chosen RTT as the disease in which to evaluate this methodology, I will then give a summary of characteristics of the disease and previous research on growth in RTT. Reliability assessment is an issue which can be incorporated into this flexible methodology, as I will demonstrate through a study of reliability in RTT I conducted in conjunction with the development of growth charts for RTT. Since a large proportion of the data was collected manually, I will discuss the difference in effort required to produce disease-specific growth charts using a manual chart review versus an electronic medical record accrual of data. I will then present the growth charts produced using this methodology and

compare the growth of a variety of subpopulations within this disease. Ultimately, I will confirm these results on a second set of data which will be collected as part of an ongoing research protocol over the next two years.

## **Fundamental Concepts Regarding Growth Reference Charts**

### **Purpose of Growth Standards**

Growth references can be used to screen for disease and follow normal growth in individuals and can be used to summarize average growth in groups. For individual patients and families, these charts are an educational as well as a screening tool. For example, trends toward obesity or wasting can be used to educate about proper diet and exercise. Additionally, abnormal growth is a frequent reason for clinicians to initiate a diagnostic investigation or refer patients to a specialist in growth. Regarding growth in groups, researchers frequently compare socioeconomic, ethnic, or diseased groups of individuals to the normal healthy population using growth standards. Moreover, growth references serve as an historical control when examining the effect of interventions such as dietary supplementation in failure to thrive (FTT) or growth hormone (GH) in short stature.

Growth references contain a graphical and/or distributional representation of a given anthropometric measurement as it changes with a covariate (Cole 1993). While age is the most common covariate, others use a different covariate, for example weight-for-stature charts. The central tendency is typically represented as the median with “percentiles” above and below the median (50<sup>th</sup> percentile) indicating what proportion of the population lies outside that mark, i.e., if a child’s height is at the 75<sup>th</sup> percentile, then 75 out of 100 children are shorter than that child. If a parametric distribution is assumed, researchers can calculate the mean and standard deviation (SD). Growth monitoring is based on two concepts. First, healthy children are more likely to be closer to the center of the distribution than in the tails. Second, children who are growing well

are likely to maintain a similar position in the distribution relative to the mean (Cole 1993). To screen for health, specific values are chosen as cutoffs for abnormal, either extreme percentiles (3<sup>rd</sup> and 97<sup>th</sup>) or SD scores (-3 and +3). Within the normal range, reference lines drawn within these extreme values describe the child's position relative to the mean. These lines are either convenient ratios of the 50<sup>th</sup> percentile (10<sup>th</sup>, 25<sup>th</sup>, 75<sup>th</sup>, 90<sup>th</sup>) or proportions of the SD score (-2, -1, +1, +2). Using this screening tool, physicians commonly follow an individual's growth by noting both extreme tendencies (above and below the most extreme percentiles) and gross aberration in growth velocity (crossing the "gutters" of several percentile lines). Physicians expect an individual to remain within the percentile range he occupies during early infancy throughout childhood, and any deviation is usually followed with concern. Unfortunately, this common interpretation of growth velocity on size reference charts demonstrates a serious misunderstanding of the design of these charts.

### **Interpretation of Growth Standards**

Although the practice of interpreting percentiles as benchmarks is common, until 2006 prospectively designed growth "standards" did not exist. The growth charts in use over the past century are references, not standards, and were designed not to model "good" growth, but to provide a rough comparison and a screening tool. This deficiency was addressed by the 2006 implementation of the World Health Organization (WHO) growth "standards," the first dataset to offer a prescriptive approach to growth which will be discussed in detail below (WHO 2006).

Since interpretation of a child's position on the growth reference or standard depends on the intent of the chart designer, it is important to understand what the reference is meant to represent. The broadest reference provides an average of large groups of individuals, with no assumption about background or environment. Certain charts are representative of an average geographic population, and are intended for use within that local geographic area. Others are

“elite” charts, intended to represent the variability within a population exposed to ideal circumstances, such as breastfed infants or upper-class individuals. The more general a chart, the more it is considered a “reference”; the more specific, the more a chart is considered a “standard.” The chart’s designer generally restricts “growth standards” to represent individuals who are healthier than the whole based on some modifiable factor (Cole 1996).

Growth rate is a combination of genetic and environmental factors, and so physicians must consider the predominant genetic and environmental factors present in the reference populations before deciding which reference is appropriate for each patient (Ranke 1996). Not only overall growth, but also relative growth of body parts varies according to genetic and environmental influences. For example, Anzo et al., found that Japanese people have larger FOC-to-length ratios than western individuals of the same age and sex (Anzo et al. 2002).

Plotting measurement on growth standard provides a reference as to how extreme a child’s height, weight, FOC or BMI is at that time. The individual point provides no information about how the child has been growing. Once two or more points are plotted, the physician can see the pattern of growth, and at least get a rough sense of growth velocity. Unfortunately, the only possible interpretations of growth velocity based on the current reference standards are “extremely low,” “extremely high,” and somewhere in between.

The methodology proposed in this paper is meant to clarify several of these fundamental issues, not only for rare disease growth charts, but also for growth references in general. I will address the different ways of interpreting growth data and representing it both graphically and mathematically. I will also discuss the conceptual issues of whom to include in the reference and the pros and cons of restricting participants to an elite standard. Incorporating longitudinal and cross-sectional data, I will demonstrate how the paradox of the growth standard, which includes

no information on “growth” rather only on size, can be resolved. To begin this journey, I will describe the history of growth chart construction and how the problems we confront today came to be.

## CHAPTER 2 HISTORY OF REFERENCE GROWTH CHART CONSTRUCTION

### **Early Methodologies for Growth Chart Construction**

The growth reference charts commonly used by physicians today, while far superior to those of a generation ago, are relatively crude in mathematical terms. A recent review of growth chart methodology (Borghini et al. 2006) cites 30 individual techniques for constructing charts, with 19 different strategies for bending the data to fit aesthetically pleasing curves – a process known as “smoothing.” Remarkably, these 30 techniques are only a sample of the techniques used over the past century to create references for plotting the growth of children. In this chapter I will review the theoretical background for the creation of reference charts, describe the introduction of the mathematical paradigm, and conclude with the current standards growth charts are expected to meet.

### **Eye Smoothing Techniques**

The first age-related reference interval charts of any kind were growth references, and the first publication of such references was over a century ago (Wright and Royston 1997). In 1891 H.P. Bowditch, a pathology professor at Harvard Medical School, published a compilation of charts produced by Sir Francis Galton (Bowditch 1891). The source of these data was a survey of 9,337 individuals measured at the 1884 International Health Exhibition (Galton 1885). This endeavor was based on a previous work from 1879 in which Bowditch included charts of average heights and weights by age for children and their parents in Boston (Bowditch 1879). Bowditch divided his charts by nationality and occupation (skilled labor, unskilled labor, and non-laboring). Concurrently, Charles Roberts, a surgeon at Victoria Hospital for Children in London, compiled a work on Anthropometry which included charts and statistical tables as well as a methodology for measuring and recording anthropometrical data (Roberts 1878). Roberts’

earlier work, “The Physical Requirements of Factory Children,” lends insight into the nature of the goals of these investigations into the bulk and fortitude of children. All of these data points were plotted on graph paper, and the distributions were drawn freehand by “eye-fitting” or with the aid of a “spline” – a tool of flexible wood originally used by ship-builders to design the sweeping curves of hulls. While percentiles could only be calculated for individual ages and at great expense of time, Bowditch was able to make observations by plotting data on the same graph paper, and counting the distribution at certain ages. For instance, he noted that in his sample the 7<sup>th</sup> percentile for men was the same as the 93<sup>rd</sup> percentile for women (Cole 1993).

While these early efforts and their tedious methodology may seem rudimentary, they were carried on largely unchanged for almost a century. Researchers compiled a variety of standard charts for weight-for-age, height-for-age, and FOC for-age, specific to geographic and racial populations, as well as specialty charts of anthropometric data on skin-fold thickness, circumferences of other body parts, and limb lengths. As these charts came into common use, pediatricians recognized them as a valuable tool to screen individuals for extreme deviations that could indicate underlying disease. Physicians continued to collect larger samples of data, and began expanding the range of additional information collected along with the growth data.

### **Data Collection Techniques**

Data were collected in two primary fashions: longitudinally and cross-sectionally. An example of cross-sectional data collection was Galton’s survey at the International Health Exhibition – male and female individuals of many ages, and races, and socioeconomic backgrounds were measured at one point in time. This technique has the advantage that data on a predetermined age range can be gathered simultaneously and rapidly. The data are, by definition, independent – no individual is measured twice, and so data points can be considered unrelated. However, the major disadvantage is that to gather enough data for each age group represented

usually requires multiple samples. This frequently involves merging datasets from disparate populations, separated not only by space, but also by time. The growth of two populations that were born twenty years apart is unlikely to be identical based on nutrition and other environmental factors. Therefore, gathering data cross-sectionally and merging it into one reference means that it no longer applies to any one population.

The alternative to cross-sectional collection is to gather data longitudinally. The earliest and most thorough example of this is the Fels Research Institute Longitudinal Study which began in Yellow Springs, Ohio in 1927 and has continued collecting data to the present (2008). The Fels Research Institute was conceived to study the effects of the Great Depression on child development, and in addition to collecting data on growth of children and their families, researchers collected psychological data. In a second example, Tanner used longitudinal data and an eye-fitting technique to develop growth velocity charts for weight velocity and height velocity which are in common use today (Tanner et al. 1966b).

The benefits of longitudinal data collection are that these data are representative of a population subjected to particular conditions (e.g., financial strain in a Midwestern town), and that longitudinal data are by nature dependent. These characteristics allow researchers to use statistical tools like regression and correlation to greater advantage. However, the researchers who constructed reference charts from these data almost universally used techniques identical to those of the generations preceding them – techniques best suited to the concept of cross-sectional data collection. The greatest drawback of longitudinal data, even when analyzed using powerful statistical techniques, is that these data cannot be used to make general statements beyond the research population. In fact, the greatest criticism of the CDC United States growth charts for

children younger than 36 months of age, which were based on the Fels Research Institute data, is that the charts only apply to white mid-western children.

### **Nationwide and International Surveys**

In 1956 the US government recognized the need to study the effects of illness and disability in the country, and the National Health Survey Act was passed. This sparked the creation of several landmark studies of growth including the National Health Examination Survey (NHES), the National Health and Nutrition Examination Survey (NHANES), and the Pediatric Nutrition Surveillance System (PedNSS). The NHES began in 1960, and in three iterations collected data between 1960-1962 (NHES I), 1963-1965 (NHES II), and 1966-1970 (NHES III). The results of these surveys conducted in ten states categorized the nutritional status of children from low-income families as unsatisfactory, specifically noting that low income black and Hispanic children were not consuming sufficient calories, iron, calcium, and vitamins A and C. Subsequently, the NHANES was created in 1971 to study the relationship between nutrition and disease. Since then, NHANES has collected data almost continuously, having been revised three times to produce NHANES II in 1976, NHANES III in 1982, and the present study initiated in 1999. At almost the same time, the CDC created PedNSS in 1973 as a subsidiary of the Special Supplemental Nutrition Program for Women, Infants, and Children (WIC) to study the prevalence and trends of nutrition-related indicators of disease among specific pediatric populations.

### **Mathematical Smoothing Techniques**

Remarkably, throughout this time period the techniques for compiling data and generating charts remained those of Bowditch in the late 19<sup>th</sup> century – plot the individual points on graph paper, count the number of points at various ages to plot the median, and draw in percentile lines by eye. The landmark studies of this sort in the mid 20<sup>th</sup> century were Stuart and Meredith's

reference values for weight, height, chest circumference, hip width, and leg girth published in 1946 (Stuart and Meredith 1946), and Nellhaus' charts of FOC published in 1968 (Nellhaus 1968). Stuart and Meredith's work contains tables of gender specific percentiles for ages four to eighteen as well as "Wetzel Grids," conjoined presentations of height and weight to allow comparison of the two parameters simultaneously. Although the authors note that SD offers a more precise method of categorizing individuals, they opt to present the material in the format of percentiles as conceptually simpler than an SD score. Anthropometric specialists obtained the data for these 1946 charts at University of Iowa experimental schools from 1930 to 1945, and this data served as the standard reference until the late 1970's. Nellhaus' work is a conglomeration of weighted averages from 14 international reports of FOC ranging from birth to 18 years. Nellhaus claims that his charts demonstrate no racial, national, or geographic predilection, and that variance among the sources of these charts was so small as to be subsumed by the noise of measurement error. His sources, although notably diverse for the time period, consisted primarily of English, Irish, and upper-middle class Anglo-Saxon American children measured from 1924 to 1948. Although overshadowed by the current CDC charts of FOC from birth to 36 months of age, Nellhaus' charts are still used commonly for older children in whom head growth is a concern. Other specialized references, such as the Tanner growth curves of height velocity and weight velocity from 1966 (Tanner et al. 1966a) still serve as standard references nearly half a century later.

The introduction of integrated circuits to computers in the 1970's opened new avenues in the application of statistical and mathematical principles to research. However, it would be several years before anthropometric researchers took full advantage of the ability of computers to process vast amounts of demographic data. As recently as 1974, at the University of Colorado,

Duncan et al., used the technique of binning, dividing ages into intervals and averaging the counts of individuals within these intervals, to analyze longitudinal data extracted from McCammon's 1970 reference tables (McCammon 1970). He manually calculated the median, 3<sup>rd</sup> and 97<sup>th</sup> percentiles by hand, and plotted these points on graph paper using a logarithmic scale up to 24 months of age to adjust his curves for the rapid growth that occurs in the first two years of life (Duncan et al. 1974). Duncan was responding to a recent appeal from the American Academy of Pediatrics (AAP) citing the need for new growth reference charts which met four primary objectives: 1) the ability to screen for abnormalities with high sensitivity, 2) the clarity to reassure parents of normal growth, 3) the ability to aid in assessing suspected abnormalities of growth, and 4) the ability to aid in evaluating the effects of treatment (Owen 1973). The AAP hinted at the use of the NHES data and of the ongoing NHANES to fulfill this call. Although Duncan met several of the stipulations of the AAP – a standardized set of intervals for measurement (birth, newborn nursery discharge, 1, 2, 4, 6, 12, 18, 24, 30, and 36 months, and yearly thereafter), and two separate but continuous charts from birth to six years, and six to eighteen years – he included only three percentile curves (3<sup>rd</sup>, 50<sup>th</sup>, and 97<sup>th</sup>), and based his charts on the Denver Child Research Council, a small cohort included in Nellhaus' prior work.

It was not until the CDC growth charts of 1977 that researchers used both mathematical equations and the statistically powerful cohorts of NHES and NHANES to construct and smooth reference charts (Hamill et al. 1977). Extracting longitudinal data from the Fels Research Institute and cross sectional data from the national surveys, Hamill calculated empirical percentiles for small intervals of data across age, again using the process of binning, and smoothed the chosen seven percentile values using mathematical “splines.” Notably, the percentile curves he produced, although achieved through a mathematical process, lacked many

of the characteristics that would make them amenable to further mathematical analysis. For instance, the curves were independent, meaning that not only is it impossible to compare the ratio between certain percentiles at certain ages, but the curves themselves could potentially cross one another, so that the tenth percentile at a certain age could be higher than the twenty-fifth percentile at that age. Clearly, such a paradox would not match the empirical data at that age. Additionally, the nonparametric character of Hamill's splines renders statistical techniques unable to evaluate the fidelity of the percentile curves to the empirical data. Ultimately, the goal of this application of mathematics to growth charts was largely aesthetic. However, this prototypical attempt lit the spark which would lead to the development of dozens of mathematical methodologies over the next three decades.

### **International Standards**

Despite their drawbacks, the CDC charts soon gained worldwide popularity. The WHO recommended in 1978 that the CDC growth reference charts be used as an international standard. Compared to the US population, a significant proportion of the world population fell below the 5<sup>th</sup> percentile of these charts. Since plotting these children below the 5<sup>th</sup> percentile would result in an unacceptably low specificity in screening and would not allow researchers to follow growth over time, the WHO proposed expressing individual values in terms of the number of SDs the measurement falls away from the median, also known as "Z-scores." Percentiles become widely spaced in the tails of the distribution, and plotting children beyond the 5<sup>th</sup> or 95<sup>th</sup> percentiles leads to poor specificity. However, Z-scores elude this issue by taking advantage of parametric assumptions to provide precise values. While this idea is generally attributed to statisticians in the 1970's, (Waterlow et al. 1977) researchers in the 1940's proposed this alternative as more precise than percentile values (Stuart and Meredith 1946). To express measurements in terms of uniform SDs, a researcher is first required to make an assumption about the distribution of the

population. The most common distributional assumption is normality, or the Gaussian distribution, known commonly as the “Bell Curve.” After making an assumption, the researcher tests the distribution against the empirical data using graphical plots and statistical tests of the null hypothesis ( $H_0$ ): the data and assumption follow the same distributional pattern). If  $H_0$  is not rejected, then relative distances from the median can be safely extrapolated out to just over 2.5 SDs.

In the case of the CDC weight-for-age and height-for-age charts, the distribution was non-normal. Hamill recognized that his data, especially those for weight, were skewed with a longer tail above the 50<sup>th</sup> percentile. Using a technique not published until several years later, he divided the data at the median, assigned the two halves completely different SDs, and treated both halves as if they were distributed normally. Hamill derived the SD for the upper percentiles and the lower percentiles separately and used these to calculate and fit curves to new percentile values for the 3<sup>rd</sup> and 97<sup>th</sup> percentiles (equivalent to -1.88 SD and +1.88 SD). Then, to construct new, constrained percentile curves, they simply divided these SD values ( $-1.88/2 = -0.94$  SD or the 17<sup>th</sup> percentile).

This novel application of distributional statistics to growth references allowed children of any size to be plotted at their respective age on the reference charts using SD scores. Although theoretically this technique resulted in discontinuous spacing among the percentiles, the shortcomings of using two separate distributions and the resultant distortion were clinically insignificant. Moreover, in routine use, where extreme anthropometric abnormality is uncommon, these charts met all four of the stipulations proposed by the AAP (Dibley et al. 1987). However, in children with severe malnutrition or, more significantly, children with chronic disorders globally affecting growth, the authors did not fulfill these goals. While children

can easily be plotted at 3 or 4 SDs away from the mean, these relationships are statistically meaningless. Extreme percentiles based on a Gaussian distribution are imprecise, and likewise the variability at extreme SDs is large. At 3 SD from the mean, only 0.14% of the empirical data responsible for creating the distribution lies below that point. Although statistical comparisons are possible using SD scores, any changes beyond 3 SD are impossible to interpret clinically; any assumptions based on these values are simple mathematical extrapolation (Cole 1994). Put more simply, if a sample of 1000 measurements are used to construct the reference for weight in a child of five, and he is determined to be -3 SD based on this reference, he is statistically comparable to only one of the original individuals measured. If an individual is beyond +/- 3 SD, he is unlike any child in the reference population of other five year olds. Consequently, if a patient whose height is initially -4 SD grows to be -3.5 SD from the mean it is impossible for a clinician or researcher to interpret this change in a meaningful way.

### **Refining Statistical Methodology**

Over the following two decades, researchers grappled with the issues of manipulating distributions, incorporating accurate estimation of outliers and extreme values. They also developed techniques to generate curves which reflected physiologic changes in normal growth, and were at the same time aesthetically pleasing. Although statisticians employed several approaches, a handful predominated. The major issues involved the following: 1) continuous versus discrete units of age, 2) distributional assumptions, 3) smoothing of curves, 4) merging longitudinal and cross-sectional data, and 5) handling the effect of extreme ages (birth and completion age of the reference) on curves derived mathematically.

### **Continuous Versus Discrete Units of Age**

The rudimentary techniques used throughout the 19<sup>th</sup> century involved binning data into age groups and smoothing without distributional assumptions. After distributional assumptions

gained favor, researchers continued the practice of treating age discretely, squeezing some older and some younger children's data points into averaged discrete ages. Even the recent CDC growth charts developed in 2000 used the process of binning prior to applying a mathematical model. The fundamental problem with this process is that it introduces bias to the study, as the choice of age cutoffs is arbitrary and must influence what proportion of children are rounded down and what proportion rounded up. The alternate solution is to treat age as a continuous variable as in a regression.

### **Distributional Assumptions**

Given the advantages of calculating Z-scores based on the SD of the normal distribution, most researchers from the 1970s onward opted to use a distributional assumption in constructing their references. Unfortunately, although the Gaussian distribution is a powerful concept, growth data are rarely distributed normally. In particular, weight tends to skew towards the right (with a longer tail of heavier children). Fortunately, mathematical transformations exist to bend the data to a normal distribution. Unfortunately, skewness frequently varies in a population based on age. While birthweights may be skewed to the left due to several low-birthweight individuals in the left tail of the distribution, by adolescence weight invariably skews to the right. Choosing one transformation for the population would improve skewness at one age while worsening it at another.

Data distributions can be summarized by four moments: mean, SD, skewness, and kurtosis. While skewness describes preferential weighting of one tail of the distribution, one type of kurtosis describes clustering of the data either more tightly around the median while another type of kurtosis describes data evenly dispersed away from the median towards the tails. Initially simple transformations such as the log transformation were employed with some success in adjusting for skewness. In 1964, Box and Cox proposed a transformation that resulted in

remarkably consistent and flexible normalization of skewness (Box and Cox 1964). Researchers began employing measures of these variations on the normal distribution to determine how the data needed to be transformed. The detrended Quantile-Quantile (Q-Q) plot displays the data converted to  $Z$ -scores along the x-axis with deviance from the normal distribution represented along the y-axis, and is a useful tool for evaluating how well a particular transformation bends the data to a normal distribution. This technique is discussed in detail in chapter six.

Statisticians who developed techniques for smoothing growth curves regard the technique as highly subjective. They recognize that smoothing a curve implies that the researcher must strike a balance between what *is* and what looks biologically plausible. Even with very large sample sizes, unsmoothed percentile curves are ragged and unwieldy. The researcher performing the smoothing must understand the underlying physiology of growth enough to interpret whether a change in the curve represents sampling noise or a biologically important trend. While some mathematical tools can help judge the goodness-of-fit, or relationship between the empirical values and the curves, ultimately, to avoid introducing bias through smoothing, the statistician must possess a deep understanding of the patterns of the measurements as they change with age.

Over the past 30 years researchers have accomplished mathematical smoothing of growth references with linear functions, cubic splines, and various combinations of polynomial functions. The recent trend has been to develop techniques which both transform the data and estimate smoothed curves simultaneously in a constrained fashion. The Cole and Greene lambda/mu/sigma (LMS) model uses the Box Cox transformation and the principle of penalized likelihood to accomplish simultaneous transformation and smoothing (Cole and Green 1992). Their method generates values the Box-Cox transformation for skewness (L) and the maximum likelihood estimate of the median (M) that change continuously with age while simultaneously

adjusting for the coefficient of variation (S) in the dataset. However, LMS does not account for kurtosis in the distribution. Recently, Rigby and Stasinopoulos expanded on the principles used by Cole and Greene to design a model that also accounts for kurtotic data. The Box-Cox power exponential (LMSP) and Box-Cox t-distribution (LMST) models they propose can be combined with a variety of parametric and nonparametric functions, and can also incorporate other covariates such as gender or parental height (Rigby and Stasinopoulos 2004).

Longitudinal and cross-sectional samples differ in many ways, and merging them is not without issues. Although longitudinal data are dependent, their correlations can be ignored, and the data can be fitted with a cross sectional model. However, by virtue of the collection method, longitudinal data typically exist at discrete, targeted ages, while cross-sectional data are more likely to be randomly dispersed across ages. At points when the datasets of combined studies overlap, researchers have agreed that data from both groups should be compared using the mean and SD of each group, and if they are found to be similar they can be merged.

The distorting effect of extreme ages (close to birth or close to the oldest point in the dataset) on curve smoothing is considerable. This issue raises the question of sample size needed to create a growth reference. While early growth references were created with as few as 30 measurements at particular ages, recent recommendations state that samples should include at least 400 children at each target age. The phenomena that occur at the right and left edges of extreme age, known as right and left edge effect, can be balanced by over-sampling at those age points. One study indicated that oversampling at birth by four times the standard number balances left-edge effect, and extending the age of data collection 20% beyond the desired age limits of the growth reference balances right-edge effect.

## **Growth References versus Growth Standards**

In 1994 the WHO proposed developing growth standards to describe how children should grow in general instead of how they grew at a particular place or time. Essentially, the WHO recognized that physicians use growth references as if they were standards, and that the time had come to produce a legitimate diagnostic tool (de Onis et al. 1997). In 1997 the WHO Multicentre Growth Reference Study (MGRS) began collecting data from children in six countries. The unique aspect of this study was that researchers were very selective about which children were measured. Only healthy, breastfed infants who grew up in a non-smoking household were included in the longitudinal component from birth to 24 months of age. Children were excluded if their family was from a lower socio-economic background, if they had a history of perinatal morbidity, or if their environment was otherwise not conducive to health. The results were somewhat surprising – while large segments of the population were excluded in poorer nations (83% of those initially selected in Brazil were deemed ineligible), the growth data that resulted were remarkably homogeneous. Mean height up to age six was not significantly different among any of the nations, and other measures were only negligibly different. Based on the successful selection of a healthy cohort living in conditions favoring achievement of their full genetic potential, the WHO concluded that the resulting growth standard can be used to make general statements about how children *should* grow across the world.

The concept of creating a reference that serves as a benchmark for the population is very attractive. Unfortunately, this approach presents several problems. The decision of which population to use as “ideal” is somewhat arbitrary. By definition, certain children are excluded from the growth “standards” under the supposition that they have not achieved their genetic potential due to inferior conditions. Therefore these children have no growth reference. Moreover, the children who are excluded are not necessarily less healthy, since the long-term

effects of abnormal height or weight are not uniform. Children in the 85<sup>th</sup> percentile for weight now, still within one SD of the mean, may be less “healthy” than children in the first percentile, beyond two SD from the mean. Many have suggested removing only the most extreme examples from a population to create a standard. However, the elimination of the few children beyond three SD from the mean of a reference population has little effect on the curves that result, and the reference and the standard become almost identical. The WHO approach was to prospectively exclude individuals based on risk factors, but the researchers did not evaluate the individuals they excluded, and so no comparison between the would-be reference and their healthy standard could be made.

### **Persistent Issues Concerning Growth Standards**

While growth charts based on cross-sectional data do an excellent job of comparing how extreme a child’s measurements are with respect to a reference population, they fail to account for the relationship between two measurements on the same individual. The relationship between two measurements of growth, or growth velocity, is perhaps more crucial than the absolute position of the individual on a growth standard, yet the current growth standards offer no means of interpreting growth velocity. Cole phrases it best, “in growth terms the growth standard is no standard at all.” (Cole 1993) Physicians commonly interpret the number of percentile lines an individual crosses over time to indicate pathological growth. For example, an individual who crosses one percentile line in the space of one month may not raise concern, but another individual who crosses two percentile lines in the space of six months may be referred to a specialist or subjected to laboratory testing. In actuality, one, both, or neither of these individuals could be demonstrating abnormal growth, but the common practice of interpreting the crossing of percentiles on size charts is simply not informative. Not only is such practice unfounded, it represents a grave misunderstanding of the growth standards themselves (Cole 1993).

## **Growth Velocity**

Current growth standards, even those compiled from longitudinal data, treat all data cross-sectionally. Therefore, no method exists to determine how individuals typically progress from one point to another. Since all minute individual changes in growth velocity are smoothed on these charts, the reference percentiles give the appearance that growth does, and should, follow these curves. When growth varies from this track these standards offer no information as to how much variation is acceptable. The space between percentile lines varies depending on the reference, but is approximately  $2/3$  of one SD. No foundation exists for accurately interpreting change with reference to percentiles, and healthy individuals frequently cross several percentile lines within the first several years of life. If SD scores are compared, a growth velocity measurement can be calculated and compared with reference values, and such comparisons of true growth are valid. Standards for velocity can only be constructed using longitudinal data, and consequently are able to represent distribution over time, such as the range of normal growth velocity acceleration during puberty. This distribution along the y-axis is another component not present in cross-sectional growth references.

Tanner appropriately referred to the common growth references as “distance” standards, as opposed to velocity standards (Tanner et al. 1966a). Our common reference charts, such as the CDC and WHO weight-for-age and height-for-age charts, measure size. Velocity charts, such as Tanner’s 1966 charts, measure growth.

Few strategies exist to incorporate correlations of longitudinal data. One model proposed by Wade and Ades used an LMS-based approach that incorporated within-subject correlations. They found that although the correlation was highly significant, the resulting median curve and the precision of the fitted percentiles were not affected by incorporating this within-subjects correlation model (Wade and Ades 1998). Other researchers have proposed mixed-effect models

that could incorporate the between and within-subject variation, as well as adjust for other explanatory variables that depend on age. One of the most useful applications of longitudinal data to growth standards involves overlaying growth velocity curves directly on the common reference, thus representing the cutoff for abnormal growth velocity (see thrive lines below). Such an application to growth standards would dramatically improve their utility as a screening tool.

Appropriateness of velocity as a screening tool depends on the degree of average height velocity for that measurement during that time period compared to the average likely error in measurement during that period. Measurement error can be equated to “noise,” and ability to detect meaningful change in growth is equal to “signal.” (Cole 1998) The signal-to-noise ratio (SNR), a simplified measurement of error, is the true variance in a population divided by the error variance. A higher SNR indicates that error plays a smaller role in the distribution, and that the reference is more valid. During periods of rapid growth, the SNR for velocity is very high compared to the SNR for size, and so velocity is a much more sensitive indicator of pathology during these time periods (e.g., infancy, early childhood, and puberty). However, during stable growth the SNR for velocity is very low, and so size is a much more reliable value to interpret. In practice, the variance in discrete measurements (distance) and the variance in velocity can be compared independent of measurements of error, thereby further simplifying the comparison. The SNR ratio, two times the distance variance divided by the velocity variance, is a good estimate of relative utility of distance versus velocity. The growth chart designer should compare SNR ratio in various age groups to determine which measure is best. Therefore, size (growth references) and velocity (growth velocity references) are both important and should be used appropriately depending on the individual’s stage of development.

In general, the measurement of height velocity in normal children is only accurate at times of rapid acceleration or deceleration. During periods of stable growth, measurements such as height or weight provide a more accurate representation of a child's position in the normal distribution. However, when children are ill, growth velocity is a much more sensitive marker of growth failure than standard growth measurements. Therefore, measurement of growth velocity can be justified throughout childhood, both to follow periods of rapid growth and to screen more effectively for disease.

### **Hybrid Size and Velocity Charts**

A third approach involves presenting multiple dimensions of data simultaneously. Emery et al., were the first to do this by developing an infant weight chart on which lines were not based on percentiles or SD scores of weight, but on percentage of weight velocity (Emery et al. 1985). The goal of the charts was to screen for children whose growth velocity was in the 5<sup>th</sup> percentile. Therefore, the lines on the chart are spaced so that if a child's weight moves down the width of the space between lines within two weeks, that child is at risk. Likewise, if a child moves two space-widths in two months, that child is at risk. In this way, the charts simultaneously follow size (weight) and weight velocity.

An improvement on this concept involves not only adjusting for the third dimension of velocity, but also adjusting for the correlation between two points on a longitudinal curve. Charts that incorporate correlations are termed conditional. These charts are automatically adjusted for other predetermined covariates. Many types have been proposed, including those adjusted for tempo (Tanner et al. 1966a), mid-parental height (Tanner et al. 1970), and sibling weight (Tanner et al. 1972). Another approach is to adjust for height one year prior to the date seen in clinic.

Cole proposes a method of representing both size (distance) and conditional velocity simultaneously (Cole 1998). Instead of representing velocity as change-in-measurement over

change-in-time on a separate chart, he proposes using change-in-SD score over change-in-time. This new measure has the benefit of being able to be represented on the same references as size – the standard growth chart. Additionally, Cole incorporates the conditionality of the previous measure, so the velocity curves are not parallel, but vary depending on the correlation of the initial values. He calls these lines “thrive” lines, as they are ideal for identifying FTT, or growth velocity below the 5<sup>th</sup> percentile. He has developed this tool as a clear plastic overlay which can be applied to a standard growth chart to determine, month by month, whether a child is “thriving” or not.

### **Confounding Factors**

In addition to adjusting measurements for relative time, researchers have adjusted for other conditional covariates. Conditional standards accounting for mid-parental height, birthweight, and sibling weight are particularly important during puberty when changes in growth occur that are not represented on growth standards. Because puberty occurs at different ages in different individuals, growth standards smooth these changes from a dynamic acceleration and deceleration into a linear ascent. Another component of growth not represented by growth standards is the increase in variance with age. A child on the 97<sup>th</sup> percentile needs to grow “faster” than a child on the 3<sup>rd</sup> percentile to maintain his standing. Because of this, a child at the 50<sup>th</sup> percentile whose growth velocity is abnormally low may drift to the 30<sup>th</sup> percentile without calling attention to his physician. However, were this child followed with a standard that incorporated conditional covariates, his abnormal growth would be detected.

The first approach, adjusting for tempo of a physiological phenomenon such as puberty, highlights the variability, not only in growth velocity, but also in time of velocity change. Onset of puberty varies widely, and constructing growth velocity charts based solely on age results in a smoothing of pubertal growth. The increase in velocity is followed very quickly by a decrease,

and so the best way to compare degrees of pubertal growth is to line up the individuals not based on age, but based on age-at-maximum-velocity. Thus a boy who reaches maximum height velocity at age 11 and one who reaches maximum velocity at age 13 would “cancel” each other out if plotted based on chronological age. When they are lined up by peak velocity, the variability in velocity forms a meaningful and uniform distribution. Another option is to adjust for age of onset of puberty using bone-age instead of chronological age.

One concern not addressed by Tanner’s method of simply lining up peak velocities of one event is the common variability in velocity at different ages dependent on other factors. For example, a newborn at the 3<sup>rd</sup> percentile is likely to grow faster than a baby born at the 50<sup>th</sup> percentile. We think of the first baby as “catching up” to the curve, but he is really crossing several percentile lines, growing in a way that may concern us were he born at a normal size. One way to account for such unpredictable and common changes is to incorporate other covariates into the model of growth. Birthweight is an excellent example of a variable that could be incorporated into a conditional velocity standard to make predictions more realistic and interpretations more valid. Mid-parental height is a more accurate adjustment than birthweight in later childhood, and conditional charts adjusted for mid-parental height give a much more precise estimate of whether a child is growing normally. Position on the growth distribution itself must also be considered. Since SD increases with age, and the 3<sup>rd</sup> and 97<sup>th</sup> percentiles move further away from the mean, a child at the 97<sup>th</sup> percentile must grow faster than a child at the mean to maintain his rank. This concept will be discussed in more detail in chapter six under the topic of regression to the mean.

## CHAPTER 3 HISTORY OF SPECIALIZED GROWTH CHARTS

### **Growth References Based on Racial, Ethnic, and Socioeconomic Disparities**

#### **Early Efforts at Specialized Growth References**

While general growth references have existed for over a century, the history of specialized growth charts is just as long. The growth reference by Roberts targeting children likely to engage in factory work is an earlier example of a specialized growth chart (Roberts 1876). While this work is unique in targeting a vocational group rather than an ethnic or socioeconomic one, Roberts' goals were similar to those of later researchers. He documented various anthropometric ranges for children currently participating in factory work and compared these measurements with other children of similar background. Using this information he established "minimum standards of physical capacity for factory work." Later researchers commonly compared anthropometric measurements of individuals from different racial or ethnic backgrounds to determine minimum acceptable rates of growth, below which children would be considered wasting, stunted, or malnourished. The notion that individuals of different backgrounds grow differently was accepted without question, and references suited to these disparate groups were generated accordingly. Remarkably, the question of whether this disparity truly exists has only been brought under severe scrutiny in the late 1990's and early 2000's through the CDC and WHO growth study projects.

#### **Racial and Ethnic Differences in Growth**

Beginning as early as 1925, anthropometrists in the US took an interest in delineating racial and ethnic contributors to growth. In 1935, Manuel writes of plotting Mexican children on American growth references, "one questions whether these tables are well adapted to use with children of a racial group and a geographic location different from that on which they were

based” (Manuel 1934). He based his work on a comparison of skin color as a surrogate marker of race, and documented differences in mean heights. Others shared his concern, and in 1941 Lloyd-Jones compared stature and weight in children of Japanese, Mexican, African American, and European ancestry attending public school in Los Angeles. He reported significant absolute differences among all groups, with whites tallest and heaviest, followed by African Americans, then Mexicans, and then Japanese (Lloyd-Jones 1940). Meredith published numerous works in the 1930’s and 1940’s detailing race-specific growth parameters, including stature in Massachusetts children of Northern European and Italian ancestry (Howard 1939), and the stature of children in Toronto.

In the 1950’s, researchers began to use statistical approaches to compare special groups. Meredith compared groups of children with Mexican ancestry by generating tables of medians and percentiles, using samples that varied dramatically, from nine individuals at four years of age to over 1700 at 10 years (Meredith and Goldstein 1952). This highlights the persistent problem of obtaining consistent samples of special populations across all ages. Meredith identified another confounding issue in his study that would go on to plague researchers developing growth charts for special populations – to obtain sufficient sample sizes researchers must frequently merge data sets with heterogeneous socioeconomic samples, variable statistical reliability, and different data collection techniques. After assimilating data from all these sources, the likelihood that observed differences between special groups are due to chance is much higher. Therefore, the data collection and statistical techniques must be that much more robust to make up for the difficulty in generating sufficient sample sizes.

Long before the WHO endeavored to generate standards of growth sufficient for use throughout the world, researchers were endeavoring to catalogue differences by comparing

partial growth references from different countries. In an effort to reinforce racial disparities, Meredith published data from Okinawa, France, South Africa, and North America. He noted that while previous authors had claimed that racial differences in height did not present until after puberty, he had evidence that these differences were present during the first decade (Meredith 1948). Expanding on this research, in 1968 he published a comparison of 160 individual studies of growth in four-year-old children in Africa, Asia, Australia, Europe, North and South America, the West Indies, and the Malay Archipelago. He found differences among groups of up to 20 cm in average height, 6.2 kg in average weight, and 2.5 cm in average FOC. Meredith noted that these samples compared groups from “the poorest strata” to the most affluent nations.

### **National Growth References versus International Growth References**

By the 1970's and 1980's most developed nations had compiled adequate references of normal growth in the pediatric population. Physicians commonly used these local references as normative values, or standards of growth for that population. The notion of racial and ethnic growth disparities persisted with little opposition, and researchers documented additional comparisons, such as a difference in height between young men in the Netherlands who were reportedly five cm taller than their counterparts in the US. However, when statisticians developed powerful techniques for assimilating data and compiling more accurate and informative charts, some nations initiated new surveys involving thousands of subjects. Rather than reproduce these massive efforts, many nations merely borrowed these more robust reference standards. When the WHO recommended using the NCHS charts in 1978, this issue was again highlighted by the unacceptably high incidence of wasting diagnosed when plotting children in undeveloped countries on these charts based on formula-fed, white, Midwestern children.

Some researchers argued that if the affluent in underdeveloped nations could achieve the same growth as the average individual in industrialized nations, then charts such as the NCHS

charts were suitable worldwide. In 1981, Graitcer et al., produced the first evidence to counter the supposition that race defines growth (Graitcer and Gentry 1981). They measured height and weight in upper-middle class preschool children from Egypt, Togo, and Haiti, and compared them to the NCHS references, finding that percentiles did not differ significantly from the reference population. This argument culminated with the findings of the WHO MGRS study. The WHO found that in comparing length among children from Brazil, Ghana, India, Norway, Oman, and the US, there was only a 3% variability due to differences among sites, compared to a 70% difference in variability in individuals from within each site. They concluded that based on these findings, coupled with recent studies demonstrating remarkable homogeneity in the human genome, reference standards can be applied internationally (de Onis 2007).

Even if certain populations are significantly different in adult height, the difficulty and expense of creating local growth references for every population may not be justified. Since growth patterns in healthy individuals are usually similar, Cole remarks that adjustments of larger growth references usually suffice for smaller populations. For instance, for a child from a non-industrialized country growing up in an industrialized country, the genetic predisposition of the child's background will still play a role in growth. Therefore, as long as data from a small segment of that population can be compared to the larger reference chart, researchers can develop adjustments based on the SD score to apply to that population. Cole cites the example that Asian children growing up in Britain are, on average, 0.5 SD shorter than British children, and the SD of their height is equal to the SD of British height. The adjustment is simple; to plot an Asian individual's height on a British chart, merely calculate the difference of the child's height and the mean of the reference at that age, and divide by the SD of the reference at that age. This calculation produces that child's SD score adjusted for his population of origin (in this

case his race). While this technique works quite well in healthy children, the patterns of growth in children with genetic disorders are quite different, and this technique introduces significant error. Therefore, to accurately interpret the growth of an individual with a genetic condition a researcher must use disease-specific or disease adjusted references.

### **Differences in Growth Due to Genetic Conditions**

Unlike that of racial and ethnic disparities, the issue of patterns of growth specific to genetic conditions was not studied until the mid 1950's, and definitive references for common genetic conditions did not exist until the 1980's. Once again, physicians and anthropometrists assumed that significant differences existed based on disease state and, therefore, to use growth as a marker of health and well-being individuals must be compared to growth references specific to their condition. Researchers have demonstrated that, just as in the normal population, considerable variation exists within specific diseases, and in Turner and Noonan syndromes this variation is on the same magnitude as in the normal population (Lyon et al. 1985; Ranke et al. 1988a; Ranke et al. 1988b). Therefore, merely knowing the mean of a special population is not sufficient (Ranke 1989). As Cole points out, in "genetically unusual" children, such as those with Down syndrome or Turner syndrome, "there is no value in using an international standard, as the differing growth performance is due to the child's genotype, and international comparisons are better made using the syndrome-specific standard." (Cole 1993)

### **Down Syndrome**

Early researchers noted that growth failure in Down syndrome begins *in utero* (Brousseau and Brainerd 1928). In the 1950's, researchers began to quantify the degree of growth failure, and by the 1960's they had conclusively shown that growth begins 0.5 to 1 SD below the normal population mean, and by age five both length and weight are 2 SD below normal. In 1974, Rarick et al., showed that although older children with Down syndrome are typically 2 SD

shorter than control children, their overall growth velocity is similar to controls (Rarick and Seefeldt 1974).

In 1978, just one year after the first CDC growth charts were published, Cronk et al., presented the first complete growth charts specific to Down syndrome based on a longitudinal sample of 90 individuals. The most important finding in this study was that individual growth velocity is dramatically different in Down syndrome, and therefore children are much less likely to cling to a percentile line on standard growth charts. They also found that 30% of children with Down syndrome have excessive weight-for-length by age 3, and that growth velocity after three years is normal relative to their control population. Based on the extreme variations of children with Down syndrome relative to normal children, they concluded that physicians should use disease specific charts to follow growth in these children as these charts “provide a more appropriate reference and set of expectations for the individual child with Down’s syndrome.” (Cronk 1978) The authors speculated that using these charts would lead to more appropriate detection of FTT or overweight children.

Ten years later in 1988, Cronk et al., revised these charts, this time with a population of 730 children and an age range up to 18 years. They were able to demonstrate statistically significant differences in weight and height at all ages compared to normal children. They confirmed the previous findings of overweight tendency in 30% of the population. Notably, Cronk revised his assertive conclusion that the Down syndrome charts should be the primary reference, and appended that the NCHS charts should always be used along with the Down syndrome charts to assure adequate screening for overweight and increased weight-for-length (Cronk et al. 1988).

In 1990 Piro et al., produced growth charts based on 383 Sicilian children with Down syndrome. Unlike Cronk, Piro et al., excluded all children with significant comorbid disease. Piro et al.'s charts were the first specialized growth charts to make the claim that they represented a "normal growth pattern," with which physicians could accurately diagnose other pathologic conditions affecting growth in children with Down syndrome (Piro et al. 1990). In a longitudinal study, Arnell et al., noted an earlier onset of pubertal growth acceleration in Down syndrome but a decreased peak growth velocity compared to normal children (Arnell et al. 1996). Addressing the issue of racial disparity, Ershow demonstrated that race (African American versus Caucasian) had no influence on growth in children with Down syndrome (Ershow 1986). Similarly addressing the question of national differences in growth, Cremers developed references using a mixed cross-sectional and longitudinal model for children in the Netherlands (Cremers et al. 1996). Although referring specifically to Down syndrome, Ranke presents two important points which apply to most rare diseases. First, syndrome specific growth charts remove the ambiguity present when plotting children with aberrant growth on standard growth references and allow the physician to detect additional, overlying disorders such as hypo- or hyperthyroidism, or celiac disease. And second, parents feel much more reassured knowing that their child is in the "normal" range on disease specific growth charts than following them below the third percentile on a standard reference (Ranke 1989).

Recently, researchers have presented an even bolder perspective. Myrelid et al., claim that children with Down syndrome simply should not be plotted on standard growth charts, lest additional diseases be overlooked (Myrelid et al. 2002). These researchers develop Down syndrome charts specific to Sweden, under the premise that since normal Swedish children are

on average taller than American children, the US Down syndrome references are inadequate for their purposes.

Styles et al., point out that Cronk's effort was based on data from five different clinic or research based samples, and therefore not representative of the total population (Styles et al. 2002). Moreover, since Cronk did not exclude any individuals based on presence of severe disease or prematurity, his charts fit more closely to the case of references than standards of growth. Styles et al., incorporated stringent data cleaning techniques similar to those used by the WHO, a component which had rarely been used in specialized growth chart construction.

### **Turner Syndrome**

Individuals with Turner syndrome have markedly abnormal height, and early researchers also noted that this involves several components of growth. In 1983, Ranke compiled a chart based on 150 individuals, dividing growth abnormalities into four segments (Ranke et al. 1983). First, growth retardation begins *in utero* as it does in Down syndrome. Second, despite retarded growth *in utero*, height development (i.e., average growth velocity) is normal up to a bone-age of two years. Third, stunting is most marked from a bone age of two to eleven years. And fourth, after 11 years, although no pubertal growth spurt occurs, the height gain of girls with Turner syndrome is almost equivalent to that of normal girls beyond age 11. Final height is decreased, but this effect is mostly due to the period between age two and eleven years. In 1988 Ranke revisited the issue of growth in Turner syndrome, creating a second set of charts and drawing similar conclusions. Notably, he also concluded that the growth profile in Turner syndrome was not consistent with a primary growth hormone deficiency GHD, making one of the first statements about pathogenesis based on a study of growth in a special population (Ranke et al. 1988b).

Massa et al., studied growth in 100 individuals with Turner syndrome, and found increased growth velocity among those Turner girls who experienced spontaneous puberty. However, because those who did not experience puberty continued to grow beyond age 18, Massa found no difference in final height between groups who did or did not experience puberty. Massa did find a correlation between each Turner syndrome individual's height and their corrected mid-parental height. The findings of Massa et al., suggest that presence of puberty, age at onset of puberty, and mid-parental height could influence all studies of growth (Massa et al. 1990). In 1990 Naeraa et al., retrospectively studied growth in 78 women with Turner syndrome, confirming several previous findings with respect to final height, correlation with mid-parental height, and prenatal and childhood growth retardation. Naeraa et al., also noted the absence of a pubertal growth spurt. However, they did find a change in height velocity, increasing from a negative value to zero at the time when puberty should occur, and then decreasing again after age 12 (Naeraa and Nielsen 1990).

Karlberg et al., studied growth changes in 58 Swedish girls with Turner syndrome. The researchers used a multiple linear regression model to study changes in growth at specific times, and found that height could be predicted in a significant proportion of 12 year olds using height at six months, age at which growth velocity changed appropriate with migration from infancy to childhood, and their change in growth at six years and 12 years of age (Karlberg et al. 1991). Suwa et al., retrospectively studied growth and growth velocity in 704 Japanese girls with Turner syndrome using a mixed cross-sectional and longitudinal model. Suwa et al., did observe a slight and gentle growth spurt during the normal age of puberty regardless of the presence of spontaneous menarche, however the mean growth velocity during this time period was significantly higher in the group who did experience menarche (Suwa 1992). Lyon et al., created

growth references specific to Turner syndrome based on four previously published series of European individuals. This study was one of the first to examine the effect of a treatment, estrogen in this case, on growth, and concluded that, although the treatment resulted in an initial acceleration of height, estrogen produced no change in final height (Lyon et al. 1985).

Wisniewski et al., created growth charts in 2006 using data from 474 girls with Turner syndrome from birth to age six years, arriving at similar conclusions regarding growth trends and birth weight (Wisniewski et al. 2006).

Addressing the issue of racial disparities in growth, Rongen-Westerlaken et al., created growth charts for the Netherlands using 598 girls with Turner syndrome. Rongen-Westerlaken et al., also developed height velocity-for-age curves. These charts confirmed the absence of a pubertal growth spurt that was suggested by the absence of a smoothed increase in growth on standard height-for-age curves. The group also concluded that on average girls with Turner syndrome reached the same average height as those in Germany, but were approximately 2.5cm taller than the average of those in other countries previously studied (Rongen-Westerlaken et al. 1997).

### **Marfan Syndrome**

Erkula et al., stress the importance of understanding growth patterns in Marfan syndrome to predict expected growth, prevent excessive growth using hormone therapy, assess the timing of surgical epiphysiodesis to halt growth, and determine when brace treatment for scoliosis is appropriate (Erkula et al. 2002). They generate growth references of height and weight, including height and weight velocity charts, for Marfan syndrome based on 180 individuals. They found that the increase in growth velocity associated with puberty occurred 2.2 to 2.4 years earlier in children with Marfan's syndrome compared with normal children.

## **Williams Syndrome**

Morris et al., produced the first growth references for height and weight in Williams syndrome in 1988, based on a collection of subjects from three different studies. In this cohort growth disturbance was greatest during infancy and early childhood. Morris et al., corrected the heights of children over four years old based on mid-parental height and compared them to normal children, finding that 70% were below the 3<sup>rd</sup> percentile for normal. Mean FOC was at the 2<sup>nd</sup> percentile of normal for children zero to four years old, and the 25<sup>th</sup> percentile thereafter, and did not correlate with intelligence scores. FOC continues to increase, paralleling linear growth, in a manner uncommon to rare diseases involving microcephaly (Morris et al. 1988).

Pankau et al., conducted two studies of growth in Williams syndrome. The first revealed that a normal pubertal growth spurt occurs (Pankau et al. 1992). The second refutes the findings of Morris et al., regarding FOC. In the Pankau et al., study FOC is indeed smaller in young children with Williams syndrome, but rather than the “catch-up” growth described by Morris et al., they demonstrate head growth paralleling the normal population with a mean adult FOC about 1.5 SD below the normal population (Pankau et al. 1994).

In a prospective study examining growth velocity in 244 children with Williams syndrome, Partsch et al., found both early average onset of the pubertal growth spurt, and lower peak velocity compared to normal children (Partsch et al. 1999). This demonstrates the value of the consistency and reliability offered by a prospective or hybrid study design. Martin et al., conducted a hybrid study, incorporating both prospective and retrospective data, and generated charts specific to British children with Williams syndrome for height, weight, FOC, and BMI (Martin et al. 2007). Martin et al., detected an overall increase in BMI relative to the normal population, a conclusion which had been suggested previously but never confirmed. This finding

suggests that children with Williams syndrome are at higher risk for obesity than the general population as they approach adult life.

### **Prader-Willi Syndrome**

Butler et al., collected data on numerous anthropometric parameters, and in an ambitious effort produced growth charts for Prader-Willi syndrome (PWS) not only of height, weight, and FOC, but also of sitting height, head breadth and length, hand and foot breadth and length, and tricep and subscapular skinfold thicknesses (Butler and Meaney 1991). Although they compared these to normal with rather dramatic differences, the limitation of their study sample to 71 individuals necessitated that they average measurements over large age intervals, sometimes as large as three years, to meet their minimum requirement of five measurements per interval. Therefore, their results are suggestive of trends but not tenable.

Wollman et al., evaluated height and weight and FOC in 315 individuals with PWS (Wollmann et al. 1998). Surprisingly, in a disease in which obesity is considered typical, 50% of the individuals had a BMI in the normal range. Mid-parental height was available in approximately 20% and was consistent with the mean adult height in the reference population. Head circumference was normal in almost all individuals. Wollman et al., also incorporated data on bone age, and found that bone age is retarded in approximately one third of PWS individuals. Notably, the growth pattern assumes an unusual pattern involving normal length at birth, deceleration of linear growth during the first months of life, steady growth during childhood, and decreased growth velocity in adolescence. While growth attenuation in infancy is not currently screened for using growth velocity charts, PWS provides an example to support the use of growth velocity screening for early diagnosis.

In their 2000 study on growth in PWS, Hauffa et al., emphasize the importance of early diagnosis in treatable conditions (Hauffa et al. 2000). They evaluated 100 individuals with PWS,

measuring only height and weight. Despite this limitation in sample size they incorporated a rarely used component which lends credence to their study – reliability testing. Hauffa et al., measured and reported the technical error of measurement (TEM) in their anthropometrists' measurements, a measure of the average SD of the group of anthropometrists. This extra step makes their data much more suitable to be used as a standard than a mere reference. Hauffa et al., processed their data using the sophisticated LMS smoothing technique which allows treatment of age as a continuous variable, and found secular trends in their data. Their German cohort was both taller and heavier on average than the American children with PWS previously reported. They recommended using population specific, diseases specific charts, especially in a case such as PWS where monitoring the effects of GH treatment using growth references or standards is crucial.

Nagai et al., studied the growth of 252 Japanese individuals with PWS from birth to age 24 years (Nagai et al. 2000). Despite the presence of more sophisticated techniques (the Hauffa reference was published in the same year), Nagai et al., used binning with eye-smoothing to plot their reference. They found that although Japanese children with PWS have a relatively similar degree of short stature compared to Caucasian children, Japanese children have a lower incidence of obesity.

### **Cri-Du-Chat Syndrome**

Physicians have long observed low birth weight and slow growth in patients with Cri-du-chat syndrome (CDCS). Due to the extremely rare incidence of the disease (1:27,000), Marinescu et al., combined data collected by clinicians or other trained professionals from North America, Italy, the British Isles, and Australia. The researchers created growth charts for height, weight, FOC, and BMI using data from 374 genetically confirmed individuals (Marinescu et al. 2000). In a later study on growth in 57 individuals with CDCS, Collins et al., found a significant

correlation between FOC and BMI that was previously unrecognized (Collins and Eaton-Evans 2001).

### **Fragile X Syndrome**

After compiling growth references for PWS in 1991, Butler collaborated with others to create growth charts for fragile X syndrome (FXS) (Butler et al. 1992). In addition to references on height, weight, and FOC, Butler et al., reported disease specific ranges for ear length and testicular volume. The researchers' goals were both to provide a reference for medical management and to help identify individuals who should be tested genetically for the syndrome at an earlier age. Although FXS occurs in females as well, Butler et al., only studied growth in 185 male individuals. Using a technique identical to that in the 1991 paper on PWS, they developed curves for the 5<sup>th</sup>, 50<sup>th</sup>, and 95<sup>th</sup> percentiles and compared the results to normal individuals. Butler et al., found that height in FXS parallels that of normal individuals until age 15 when their height velocity decreases, ultimately resulting in an average final height in FXS 5cm shorter than in normal children. Head circumference was consistently higher in FXS, but more markedly so between ages one to five years. Most notably, testicular size was consistently larger than in normal males from an early age, with the 5<sup>th</sup> percentile for FXS paralleling the 95<sup>th</sup> percentile for normal children after age six years. These findings reinforce the concept that specialized growth charts can be useful in confirming clinical suspicion of a genetic condition prior to ordering genetic testing.

### **Neurofibromatosis Type I**

While large FOC is classically associated with Neurofibromatosis type I (NF1), little else was known about growth in the syndrome prior to 1999. Clementi et al., measured height, weight, and FOC in 528 individuals with NF1, and performed repeated measurements in 143 of these to develop height velocity charts based on a longitudinal sample (Clementi et al. 1999).

Clementi et al., used a technique for estimating growth velocity based on a combination of two earlier statistical techniques, those of Healy (Healy et al. 1988) and of Preece and Baines (Preece and Baines 1978). Although more complicated than the techniques used in the studies described above, the Clementi et al., technique resulted in precise estimation of growth velocity taking full advantage of their longitudinal data.

While height in NF1 paralleled normal children until pre-adolescence, height percentiles for NF1 drop relative to normals during adolescence with a significant left skew, unrelated to disease severity. As noted in other genetic syndromes, the pubertal height velocity increase for boys with NF1 was slightly early and somewhat attenuated compared to normal children, however girls had normal pubertal height velocity trends. Neurofibromatosis type 1 represents another condition in which disease specific growth references can guide medical management, as children are frequently treated with GH after radiotherapy for optic glioma.

The following year, Szudek et al., confirmed the findings of Clementi et al., by measuring static parameters on 569 North American patients with NF1 (Szudek et al. 2000). Although they did not obtain longitudinal data nor generate velocity charts, they created slightly more sophisticated reference charts than Clementi et al., using the smoothing techniques that Hamill et al., employed to create the 1977 CDC growth references. Notably, although Szudek et al., reported significant skewness and kurtosis in their data, describing the nature of departure from the normal curve and the effect this has on percentile estimation, they boldly ignored their findings and estimated percentiles as if the data were normally distributed.

The results of the efforts of Szudek et al., are an excellent example of why even more sophisticated smoothing techniques are needed to handle data from populations with rare diseases. The mathematical functions that Szudek et al., used estimate the percentile curves

separately, without considering the position of the adjacent curve or the relationship to the whole distribution. Consequently, the five percentile curves wend their way individually across the chart as if the data followed five separate distributions. The results can be disastrous; in the extreme, percentile lines can touch or cross. In the case of Szudek et al., this method results in dramatically variable distances between percentiles. This example highlights the concept that in special populations, especially those with genetic factors influencing growth, abnormal distributions are more likely to exist. When abnormal distributions are the rule, growth reference designers must pay particular care to the nature, meaning, and graphical interpretation of these trends.

### **Congenital Adrenal Hyperplasia**

Although effective therapy with hormone replacement has been available for congenital adrenal hyperplasia (CAH) for decades, no regimen has been able to address the abnormal growth patterns of this class of diseases. Therefore, specialized growth charts for this condition are of great value not only to ensure that children remain within the “normal” range for their disease, but also so that researchers can strive to improve growth relative to existing standards. To this end, Hargitai et al., in Hungary and Frisch et al., in Austria simultaneously published results from the same international study of 598 children with CAH (Frisch et al. 2002; Hargitai et al. 2001).

Data were available in both cross-sectional and longitudinal format, and so these researchers produced both static references and growth velocity charts. In an unparalleled demonstration of the lack of standardized methods in the field of growth chart construction and the necessity for creativity, these two groups publishing in 2002 resorted to a mathematical model of growth from 1937. The Jenss and Bayley model is best suited to model growth during infancy and early childhood, and these researchers applied it to model a regression curve for age

zero to three since their sample size did not contain sufficient data at those ages to support a more sophisticated approach. For children older than three years, they applied Cole's state-of-the-art approach and generated percentiles for height in two types of CAH. For the longitudinal component, the researchers used the Preece-Baines model mentioned above, and were able to demonstrate an evident, though attenuated, pubertal growth velocity increase which occurs one to two years earlier than in normal children. Once again, understanding the pattern of the disease, in this case the pseudopuberty growth spurt, can help physicians to both care for individuals with the disease and monitor the effect of treatments.

### **Beckwith-Wiedemann Syndrome**

Beckwith-Wiedemann syndrome (BWS) is classically associated with macrosomia at birth and linear growth above the 95<sup>th</sup> percentile through adolescence (Jones and Smith 1997). Sippell et al., conducted a longitudinal study on five girls and two boys with BWS, demonstrating consistently increased height, far beyond predicted genetic potential based on parental height (Sippell et al. 1989). This study took advantage of the longitudinal nature of the data by constructing growth velocity curves of the participants, all of which were above the 90<sup>th</sup> percentile for normal up to 4-6 years of age. However, the lack of summary statistics and measures of dispersion highlights the difficulties in using small samples typically available in rare diseases to generate growth references.

### **Rubinstein-Taybi Syndrome**

Rubinstein-Taybi syndrome (RTS) is a disorder involving postnatal onset growth failure, classically involving decreased height and weight (Jones and Smith 1997). Stevens et al., combined measurements in 95 individuals with RTS to generate growth standards for weight, height, and FOC (Stevens et al. 1990). The sources were almost evenly split between the Netherlands and the US, and observations confirmed the anecdotal observations that children

were significantly smaller soon after birth. However, although height velocity was lower than normal, this longitudinal study found that velocity was still within the normal range with the exception of an absent pubertal growth spurt.

### **Silver-Russell Syndrome**

Another very rare condition, Silver-Russell syndrome (SRS) is associated with prenatal-onset short stature. Two editions of growth references exist for SRS, the first of which was produced by Tanner et al., in 1975 based on 39 children followed longitudinally for 13 years (Tanner et al. 1975). The second study used a mixed longitudinal and cross-sectional methodology to analyze growth in 386 individuals. The longitudinal data allowed the researchers to assess growth velocity, and they found a decrease in the pubertal growth spurt in this condition as well. The researchers concluded that this reference would aid in counseling families, as well as determining the effect of growth promoting therapies in SRS (Wollmann et al. 1995).

### **Brachmann-de Lange Syndrome**

In 1993, Kline et al., created reference standards for Brachman-de Lange syndrome (BDLS) using data obtained from pediatric records in conjunction with study-physician examinations (Kline et al. 1993). This mixed prospective-retrospective design allowed accurate assessment of current growth, and also development of growth velocity references for BDLS. The authors found significantly decreased birthweight and body habitus, but noted that growth velocity in childhood is within the normal range, with the exception of a blunted pubertal growth spurt. Therefore, although children with BDLS are small, they do not fail to thrive on average.

### **Duchenne Muscular Dystrophy**

In 1988 two independent researchers produced growth references for Duchenne muscular dystrophy (DMD). The first addressed weight and length, with a focus on bone maturation related to growth (Eiholzer et al. 1988). Notably, they also accounted for parental height, birth

length, and longitudinal correlation of bone maturation in their model of growth. The second, less comprehensive reference focused on weight, based on the hypothesis that close surveillance of weight in muscle wasting conditions is an important part of continuity care (Griffiths and Edwards 1988).

### **Achondroplasia**

Horton et al., developed the first growth reference for achondroplasia based on measurements of height, growth velocity, upper and lower segment length, and FOC in 400 individuals. The purported goal of this reference was to aid in distinguishing superimposed disorders affecting growth and to assess the effects of growth modifying therapy (Horton et al. 1978). In a more elaborate study, the same authors evaluated growth in three subtypes of achondroplasia, developing references based on 61 to 72 individuals in each group, and established that growth in different types of dwarfism follows different patterns. These references demonstrated that for virtually all individuals, height in achondroplasia is below the lowest reference curve on the normal growth reference. Therefore comparisons of individuals with achondroplasia using normal references are meaningless.

### **Noonan Syndrome**

In 1986, using retrospective data on 112 individuals, Witt et al., developed the first growth references for Noonan syndrome (NS), a disorder associated with short stature (Witt et al. 1986). In doing so, Witt et al., were able to demonstrate that short stature is a uniform feature of the syndrome independent of age. Two years later, Ranke et al., analyzed growth in 144 individuals with NS using longitudinal data (Ranke et al. 1988a). Ranke et al., also calculated growth velocities in subgroups, and found that cardiac defects common in NS do not significantly affect growth within the syndrome. Recently, Limal et al., compared growth in NS individuals of two distinct genetic groups, those with a PTPN11 gene mutation, and those without. They found that

not only do individuals with the mutation grow more slowly, they also respond less vigorously to GH therapy than those without the mutation. This represents one of the first studies to use growth as an outcome to differentiate genetic markers within a syndrome, thus drawing genotype-phenotype correlations.

### **Klinefelter Syndrome**

In 1972, Caldwell et al., produced the first reference for Klinefelter syndrome, one of the most common sex chromosome disorders in males (1:500). They found that not only were males taller on average, but the ratio of upper segment to lower segment was also abnormal, with longer legs on average relative to torso (Caldwell and Smith 1972). Brook et al., found that 27 males with Klinefelter syndrome showed a close correlation in adult height to the adjusted predicted mid-parental height of their parents, as differentiated from Down syndrome, in which height gain is not correlated to parental height.

### **Rett Syndrome**

Despite similar incidence to many of the diseases discussed in this chapter, growth in RTT has received little attention. Rett and Hagberg both recognized the prominent head growth deceleration in the syndrome at its inception. Three years after RTT came to public attention, Holm described poor linear growth in 10 of 21 individuals (48%) with RTT. She noted that in eight of these individuals height fell below the 5<sup>th</sup> percentile between 3 to 42 months of age. Notably, she also found normal sexual development with normal age at menarche in this group (Holm 1986). Another early study involving growth reported FTT as a significant problem in 21 individuals with RTT (Rice and Haas 1988). In 1992, Thommessen et al., examined 10 girls with RTT and found that all but one had height and/or weight below -2 SD of normal healthy children (Thommessen et al. 1992). While these researchers found that energy intake was decreased relative to the normal childhood diet, later researchers noted that despite adequate caloric intake,

girls with RTT still fail to thrive (Motil et al. 1994). This research suggests that low weight and height are features of the syndrome that do not respond to normal nutritional intake, but may respond to increased supplementation.

In the first series of reference charts for RTT, Schultz et al., used a binning technique to summarize longitudinal measurements of weight, height, and FOC in 96 individuals with RTT (Schultz et al. 1993). Schultz et al., then fit a regression line through the summarized data and found that the mean values differed from normal healthy reference values in several respects. First, they found that FOC mean deviated below the normal mean after age three months and persisted at approximately -2 SD from four years onward. Second, they found that mean height deviates from 50<sup>th</sup> percentile at age 16 months, and drops below the 5<sup>th</sup> percentile for normal at age 7.5 years, with a Z-score of -2 SD at age 8.5 and -3 SD at age 12.25 years. Third, they found that mean weight deviated from normal mean weight at four months and crossed below the 5<sup>th</sup> percentile at 4.5 years. Notably, they also found that mean weight at birth was 0.42 SD below the normal mean weight for healthy girls, and fell to -2 SD at eight years, -3 SD at 13 years, and -4 SD at 18 years. Additionally, they compared weight-for-height to normal values, and found that Z-scores for weight-for-height were low at birth (-0.84 SD) and dropped to -1.5 SD at age 10 years. Schultz et al., made the important observation that the pattern of growth failure in RTT is very different from that in chronic diseases in which head growth deceleration is a late finding, and also different from growth in other genetic syndromes. However, this paper suffers several shortcomings. First, binning of age groups introduces significant bias, as described in chapter two. Second, the technique of fitting a regression line only allows meaningful statements about the mean of RTT growth not about its dispersion.

Other studies have focused primarily on specific aspects of growth in RTT. In a later paper, Schultz et al., examined hand and foot growth in RTT, and found that in 28 individuals hand length was shorter than average, but mean length never fell below -3 SD, whereas mean foot length fell below the 5<sup>th</sup> percentile at age 5.5 years and below -3 SD at age 10.5 years (Schultz et al. 1998). Hand length correlated with height, while foot length was distinctly small for height. Schultz et al., also controlled for age and ambulation in their model, and found no significant effect on foot length based on ambulation ability. Hagberg et al., studied head growth in 69 classic and 13 atypical RTT individuals (Hagberg et al. 2000). The study design used by Hagberg et al., included retrospective longitudinal data, however, the authors did not take advantage of the longitudinal component in their statistical analysis. Notably, Hagberg et al., recognized the deficiencies in the standard reference of FOC used in Sweden at the time, and so combined two standard references to use as a comparison. Using cross-sectional analysis techniques, the authors found that FOC was significantly smaller at three months of age in classic RTT compared to the reference population, and dropped to -2 SD by age four years and -3 SD by 8 years. Hagberg et al., also noted that the sooner deceleration occurred, the more extreme the eventual magnitude of deceleration would be.

Although no animal models of RTT mutations have confirmed growth-related findings, McGill et al., recently found that mice with *MECP2* mutations exhibit dysregulation of hypothalamic hormones, specifically cortisol releasing hormone (CRH) (McGill et al. 2006). The high CRH and corticosterone found in these mice is attributed to cause anxiety. This mechanism, or similar hypothalamic dysregulation, could provide a foundation for the altered growth in RTT, specifically the common association of an absent pubertal growth spurt.

In a recent study, Oddy et al., examined the factors affecting growth in RTT, including nutritional support, mutation type, mobility, breath-holding, and hyperventilation (Oddy et al. 2007). The researchers included information on comorbidities such as scoliosis, epilepsy, constipation, breathing disorders, and restricted mobility. The participants were binned into four large age groups, and comparisons were made among groups. They found that mean height, weight, and BMI were decreased relative to healthy normal children, and that low mobility correlated with low BMI. Although this is initially counterintuitive, low mobility could reflect muscle wasting, which would correlate with low BMI. They also found that breathing dysfunction was related to low BMI, although the causality of this relationship is questionable. Notably, they found relationships among mutation type and BMI, specifically that late truncating mutations and milder point mutations (R294X and R306C) had higher average BMI and used less supplemental nutrition than early truncating and more severe point mutations. The authors used boxplots to represent differences in median and dispersion of growth at different ages. This technique is a significant improvement on the regression line fitting used by Schultz et al., however, the clustering into massive age intervals precludes comparison to individual patients. Therefore, although the comparisons are useful, the knowledge is difficult to generalize in a clinical or research setting. This binning into large age groups of up to seven year intervals introduces significant bias into the methodology. This bias could have either blunted or accentuated their results. One additional shortcoming of this study was that all data were cross-sectional, therefore changes in management and survival bias were not accounted for in comparing younger and older participants. All of these issues could have been addressed if the authors had collected longitudinal data in addition to cross-sectional, and had treated age as a continuous variable.

## **Non-Genetic Etiologies of Altered Growth**

Growth charts for non-genetic diseases or scenarios in which growth is influenced by chronic disease are a matter of debate among researchers. Since some researchers view genetic etiologies as a fixed deficit, they believe there is less growth variability within these syndromes compared to chronic diseases, and therefore growth charts are justified in genetic conditions but not in other diseases. Nonetheless, growth charts have been published for cerebral palsy (CP), despite significant heterogeneity within the disorder, and also for certain common infectious diseases, notably human immunodeficiency virus (HIV).

### **Cerebral Palsy**

Growth failure is common in CP and poor nutrition is a component of this. The most frequent intervention is the placement of a gastrostomy tube, however the effect of this intervention on growth is not entirely clear. Rempel et al., studied the effect of gastrostomy feeding on 57 individuals with CP, and found that while gastrostomy feeding improves weight, it has no significant effect on linear growth (Rempel et al. 1988). Not until 1996 did researchers complete growth references specific to CP. Using data on 360 children, Krick et al., generated growth reference charts for weight and height in CP (Krick et al. 1996). They described significant differences in weight and height in CP compared to the CDC reference population, with average measurements below -2 SD. In 2006 Stevenson et al., examined growth in 273 children and developed comprehensive charts for six anthropometric measurements, weight, knee height, upper arm length, midupper arm circumference, triceps and subscapular skinfolds. However, the authors of this reference prudently caution that it is not to be used as a prescriptive guide for growth in CP, merely as a descriptive “snapshot” of how children with CP of varying severity have grown in the past (Stevenson et al. 2006). They propose future research to

delineate growth patterns based on different factors, but remark on the difficulty of such an undertaking.

### **Human Immunodeficiency Virus**

Although perinatally acquired HIV is not a fixed genetic defect, it shares some of the characteristics of genetic syndromes. Newell et al., examined growth in individuals exposed to HIV perinatally, and found that children who contracted the virus grow more slowly and were ultimately shorter and lighter than children who did not (Newell et al. 2003). The authors used 10 years of longitudinal data to create standards of growth in HIV infected children based on Z-scores relative to the normal population. The duration of the longitudinal data makes this study unique, because researchers were able to establish correlations between age at treatment and type of treatment with improvement in growth velocity, stature, and weight. The study also contained a built-in control in the individuals who were exposed to HIV but not infected. The strength and consistency of the correlation help to validate the use of this type of disease-specific reference, both to counsel about growth outcomes and to monitor the effects of treatment.

### **Summary**

This chapter highlights many of the strengths and weaknesses of existing disease specific growth charts. Clearly, researchers have found these references useful, as many are in their third or fourth edition. Refinements in methodology have improved the overall quality of the references and practical experience has validated their use to identify comorbid conditions, counsel families, and evaluate the effects of treatments. However, resistance to the adoption of disease-specific references persists - much of it justified. No standard approach to growth chart construction exists, and many charts produced in the 21<sup>st</sup> century still carry components of bias recognized as problematic in the early 20<sup>th</sup> century. Nonetheless, driven by the need to monitor the effects of emerging new treatments for rare diseases, disease specific growth references and

standards will indeed become a necessary part of medical research and even routine medical practice.

## CHAPTER 4 CASE FOR SPECIALIZED GROWTH CHARTS

### **Case Against Specialized Growth Charts**

“The current CDC recommendation is to use the CDC growth charts in all cases (U.S. Department of Health and Human Services 2008).” Not all researchers agree that specialized growth charts are necessary to interpret changes in individual growth within special populations. The CDC argues strongly against using specialized growth charts in any circumstances.

. . . Recognition of the differing growth pattern in the child with trisomy 21, and the problem posed by comparing that child's growth to that of average children, led investigators to develop and publish alternative growth charts. . . However, it must be emphasized that there are reasons for which these charts should not be used or not used by themselves.

Children with some conditions, such as Rett syndrome or Prader-Willi syndrome, may present no measurement problems with the available anthropometric equipment, but the resulting data may be difficult to interpret because of altered growth potential (U.S. Department of Health and Human Services 2008). – *Center for Disease Control: online tutorial for growth reference use*

The CDC notes the existence of numerous growth charts for specialized populations – diseases with altered growth due to both genetic and non-genetic etiologies. They go so far as to recognize that data from these individuals may be difficult to interpret on the CDC charts due to patterns specific to the disease. Nevertheless, they persevere in their insistence that clinicians continue to use the CDC charts.

The CDC challenges these specialized charts on many levels citing the following concerns as the foundation of their argument:

- Most charts have been developed from very small sample sizes
- Data used for the charts do not reflect racial, ethnic, or geographical diversity
- Old data have been used to construct the charts
- The data may not represent the population as a whole, rather may come only from well-nourished children within that population
- Techniques for measurement are inconsistent, not clearly defined, and varied.

- Most of these charts fail to consider secondary medical conditions which influence growth potential, and which often accompany primary chromosomal disorders (U.S. Department of Health and Human Services 2008).

To support these judgments, the CDC cites the example of Down syndrome, claiming that, in the most recent iteration of Down syndrome-specific growth charts by Cronk et al.:

- The children in the sample were of limited diversity with respect to ethnicity, race, and geographic location
- Nutritional status was not assessed
- Secondary medical conditions such as congenital heart disease and feeding difficulty were not considered

For these reasons, the CDC recommends plotting children with Down syndrome exclusively on the CDC growth charts. They concede that specialized growth charts may have their place to demonstrate to parents that children with chromosomal disorders may have different patterns of growth from normal children, but should not be used as a diagnostic or screening tool, nor to follow the child's growth. Moreover, the CDC insists that specialized charts for non-chromosomal disorders, such as cerebral palsy, should never be used under any circumstances.

### **Counterargument to the CDC Perspective**

Where normal growth patterns differ from the general population, it has been found useful and clinically important to use syndrome specific growth charts (Styles et al. 2002)

The counterargument to the CDC perspective can be broken down into two components, a general argument about the necessity for specialized growth charts, and a specific rebuttal to the faults they find in the current specialized growth charts. In many disorders of growth, measurements fall well outside the normal boundaries of the CDC charts, often beyond 3 SD from the mean. Although values for these measurements can be calculated using Z-scores, the syndromic child cannot be compared using these values. Physicians commonly use Z-score values beyond 3 SD to make relative comparisons of a child's progress. The first problem with

comparisons at extreme SDs is that as SD values move further from the mean, the distance between percentiles becomes much larger, and therefore estimations of where a child lies in reference to other children become very imprecise. The second problem is that following a child's Z-scores relative to a standard population with age assumes that the pattern of growth in the disease follows the mean of the reference population. The following example illustrates both of these issues.

The growth chart in figure 4-1 displays the 2<sup>nd</sup>, 15<sup>th</sup>, 50<sup>th</sup>, 85<sup>th</sup>, and 98<sup>th</sup> percentiles for the CDC height reference (Z-scores indicated in legend), as well as the 50<sup>th</sup> percentile of height in the RTT population. By age 12, the average patient with RTT has dropped below the 2<sup>nd</sup> percentile for height on the CDC charts. Therefore, the majority of individuals plotted in this age range on the CDC charts will be below the lowest cutoff, making interpretation of changes impossible. The Z-score chart of the SD of the average RTT patient with respect to the CDC distribution illustrates the differences in growth pattern between healthy individuals and RTT (figure 4-2). The problem with interpreting RTT values based on CDC Z-scores is best visualized through examining these two hypothetical scenarios.

In the first scenario (example 1), a girl with RTT is seen at 12 years of age, and her height is 131 cm. This value falls below the lowest percentile on the CDC charts, so to aid interpretation the physician calculates her Z-score as -2.7 SD (table 4-1). The physician institutes a program of intensive nutritional therapy, supplemental calories, calcium, vitamin D, as well as physical therapy, regular ambulation, and hippotherapy. Some time goes by and she returns at age 14 for another visit. Her height is 142.4 cm, an absolute gain of 11.4 cm. However, her Z-score is -2.8 SD, slightly lower than her previous Z-score. The physician interprets this intervention as minimally effective. Although she gained over 10 cm in height, relative to the general population

she is shorter than she would have been if she had merely “maintained” her growth in parallel with the CDC percentiles.

In the second scenario (example 2), a female with RTT is seen at 15 years of age and measures 141.4 cm. Again, her height is well below the lowest percentile on the CDC charts, and her Z-score is -3.2, greater than 3 SD from the mean. A similar intervention program is initiated with the goal of maximizing her height before final height is achieved. She returns to clinic at age 18 measuring 143 cm, having gained 1.6 cm in height, and her Z-score has now improved to -3.1 SD. While this is not a dramatic change, it is an improvement, and the physician interprets this intervention as a success.

Examination of the same examples with respect to the average height in RTT yields remarkably different conclusions (figure 4-3). In example 1, the average height in RTT at age 12 is about 135 cm, and so the individual is shorter than average at 131 cm, with a RTT SD of -0.5. However, at age 14 the average height is only 140 cm, and so her height has increased to 0.8 SD above the mean for RTT. While the CDC growth charts revealed a relative loss in height, the conclusion in this case is that the intervention was a success, increasing her height by over 1 SD on the RTT distribution. In example 2 the exact opposite case exists. The individual is at the average height for RTT at age 15 (0 SD), however by age 18 she is -0.1 SD below the mean. The relative improvement on the CDC charts is interpreted as a relative loss in height on the RTT distribution. The Z-score chart with overlying RTT mean emphasizes these relationships (figure 4-4).

This example treats changes on these cross-sectional references as if they represented longitudinal changes, and therefore the comparison is not perfect. However, most physicians use the CDC cross sectional charts as if they dictate how children should grow with age, expecting

children to parallel percentiles or SD lines. The disparity in the conclusions drawn in these examples cannot be ignored – the fact remains that children with many genetic conditions grow “differently” from healthy children.

Some researchers have proposed adjusting the reference curves from a larger population to fit the needs of a special population. The example in chapter three of race and ethnicity illustrated a simple adjustment for population mean as described by Cole. Although this adjustment works well in healthy individuals who follow similar patterns of growth, with similar velocity changes and similar SDs in the population dispersion, the adjustment fails in rare diseases with growth perturbation for the same reasons illustrated above. At an age when the general population height velocity is increasing, the diseased population height velocity may be decreasing, and vice-versa. Complicated adjustments could certainly be contrived which accounted for changes in velocity and SD in the population of interest, but the computations required would be more complicated than those to create stand-alone specialized references, and added benefits do not exist.

Ultimately, four scenarios can occur when data are compared to a standard. These scenarios can be summarized based on the offset (the mean SD score of the data) and the trend (the regression coefficient of the SD scores on age). In the first case, the data and the reference are alike; the mean SD score is zero and this similarity doesn't change with age. In a second case, although the data are different, the relationship doesn't change with age. Therefore, the offset of SD scores would be present, but constant. This could occur, for instance, when comparing Asian children to a Western reference. In the third case, the SD scores are offset, and the offset gradually increases or decreases with age. For example, based on secular trends of overweight and obesity, the CDC weight charts follow this case. Children are, on average,

heavier than the charts depict, and this relative difference increases as they grow older. In a final case, the mean SD score is offset from the reference, and this relationship fluctuates up and down with age, making it impossible to interpret the data using this reference. This final case is the one illustrated in the two examples above, and is the most common scenario in rare diseases affecting growth patterns.

Of the three abnormal scenarios, the offset SD score is the only one that can be corrected relatively easily. If growth references are to be used effectively as a screening tool in rare diseases, physicians need to understand the patterns underlying growth. The only way to thoroughly understand these patterns is to develop a reference built on empirical values. Once this reference has been developed, it makes little sense to then plot these individuals on a reference which does not follow their pattern, e.g., the CDC charts. Moreover, if the goal of monitoring an individual's growth on a reference chart, which is effectively an historical control, is to examine the effect of an intervention, the degree of change must be supported by empirical data. Given the disparity between the current method of using Z-scores to monitor abnormal growth in rare diseases and the actual patterns of growth, researchers will be hard pressed to validate the significance of an intervention using these standards. In actuality, the "standards" themselves are not standardized. The four most popular references for FOC yield values with a variability of up to 1.5 SD for the same measurement (figure 4-5). For example, a two year old female with an FOC of 47cm would have the following Z-scores: British = -1.68 (5<sup>th</sup> %), Nellhaus = -0.84 (20<sup>th</sup> %), CDC = -0.34 (37<sup>th</sup> %), and WHO = -0.13 (45<sup>th</sup> %). Although designers of specialized growth charts should be as rigorous as possible, they can hardly do worse than the current variability in growth standards in use today.

## **Itemized Response to the CDC Critique of Specialized Growth Charts**

### **Sample Size**

Although sample sizes are a persistent challenge in studies of rare diseases, several of the growth charts reviewed in chapter three had adequate sample sizes. Although no definitive formula exists for estimating ideal sample size, tradition dictates that 100-400 participants per age range produce an adequate sample for a general growth reference. The study of growth in RTT in chapter seven had a sample size rivaling that of the 2000 CDC sample size. In fact, at some intervals the RTT database contained more values than the CDC reference. However, the WHO reference contained roughly two to four times as many participants as the RTT reference (table 4-2).

### **Racial, Ethnic, and Geographic Diversity**

Although the CDC point is well taken, and every effort should be made to include a representative population in growth references, the claim that specialized growth charts are less diverse than the current standards is hypocritical. Many examples from chapter two illustrate the fact that the growth references in use today are anything but diverse. The data from the Fels institute, which makes up a large percentage of the current CDC growth reference, is based on formula fed white upper-middle-class children from one town in the Midwest. The current WHO references for adolescents, meant to represent growth across the world, are drawn entirely from the NCHS data in the US. Nellhaus' "international" growth reference for FOC which has been in use for the past 40 years is based predominantly on English, Irish, and white Anglo-Saxon American measurements. Although rare diseases do pose an issue of recruitment, the most important concern is that the sample is representative. Modern multi-center study designs have been able to increase diversity in specialized populations, and future specialized growth

references will likely offer more racial, ethnic, and geographic diversity than the national standards currently in use.

### **Recency of Data**

Again, examples abound of old data resurrected for national growth references. While research has shown that secular trends produce meaningful changes in population growth within as little as 20 years, the US references still include data from over 70 years ago. The WHO growth charts released in 2007, the state-of-the-art prospective childhood standards, were based on data that was already 10 years old. The practice of gathering case reports and small series' of patients from decades past and assembling them into a conglomerate data set has been replaced by modern multicenter or multinational study designs. The goal of these newer designs is to provide the most up-to-date reference data on rare disease natural history. Following the designs outlined in chapters six and seven, researchers will produce rare disease references more current than most national references.

### **Representativeness of Data to Population as a Whole**

Since rare diseases often escape attention if their symptoms are not severe, this point is a very valid concern. Conversely, very ill patients may not be able to participate in natural history studies. The most effective way to address this issue is, of course, to increase the sample size until homogeneity is virtually assured. Statistically, the sample used for the study in chapter seven represents a large enough proportion of the US population as to preclude selection bias. However, for different study designs researchers should seek to interact with advocacy groups, foundations and associations to recruit a broad base of patients. It is mandatory that data also be collected on representative measures of general health and nutrition, so that selection bias can be discovered and addressed.

## **Reliability of Data**

Reliability of data is a major issue in any study of growth, but especially in retrospective studies. While it is unlikely that researchers will be able to assess reliability for all measures and all data types, it is imperative that each study of growth include reliability assessment of a representative sample of data. Many rare disease study designs include both prospective and retrospective data, and data which is assessed for reliability and found to be highly valid and reproducible can be assigned a higher weighting than data from less reliable sources. Although the majority of the CDC data have no measures of reliability associated with them, the recent trend in the literature is to assess reliability of anthropometric measurements early in a study so that errors can be addressed by training operators or refining protocols. The study in chapter seven included a sub-study of reliability which is printed in appendix B.

## **Secondary Medical Conditions**

Secondary medical conditions are common in rare diseases, and physicians are aware of the many specific associations. For example, congenital heart disease in Turner syndrome and Down syndrome, and hypothyroidism in Down syndrome can have major impacts on growth. All studies of growth would be deficient if they did not include this additional information as part of their protocol. This criticism is somewhat counterintuitive, since individuals with rare diseases are likely to come under closer medical scrutiny than the general population. Therefore, physicians would be *more* likely to be aware of and document secondary medical conditions in rare diseases than in studies of growth in the general population.

## **The Role of Disease-Specific Growth Standards**

### **Counseling**

Anticipatory guidance and counseling are challenging issues in rare diseases. When growth abnormalities are prominent, physicians benefit from having discrete, meaningful measures of

progress. In addition to the difficulties posed by interpreting height values well below the lowest curve on a growth chart, many families are dismayed to see their child's growth represented in this grey zone. Disease specific growth charts help to reassure families that their child's growth is "normal" for that disease, in addition to allowing the physician to make informed judgments about treatment.

### **Detection of Comorbid Diseases**

Although certain syndromes exhibit a characteristic growth pattern, considerable overlap exists among these syndromes, and so diagnosis must incorporate other clinical findings, laboratory values, genetic markers, etc. While disease-specific growth charts in isolation are rarely useful in confirming a diagnosis, they can be used to refute or call into question a specific diagnosis when the growth of the individual is not compatible with the normal range of the disease specific chart.

One of the major benefits of using disease specific growth charts in clinical practice is the ability to screen for comorbid or secondary diseases involving growth. Examples abound of syndromes prone to secondary diseases that affect growth. For instance, hypothyroidism occurs frequently in Down syndrome and Turner syndrome (Cutler et al. 1986). Individuals with Turner syndrome are also more likely to develop Crohn's disease or GHD (Price 1979). In other cases, two diseases can occur independently in the same individual, or one disease can result in growth deficiency through two mechanisms. When a second disease occurs randomly in a child with a syndrome, syndrome-specific growth charts are useful for detection. Alternatively, when a single disease incorporates two mechanisms, growth charts can be recalibrated based on the mechanisms involved. For example, in congenital rubella, the virus causes growth failure by directly inhibiting cellular growth and replication. In turn, the effects of this on the hypothalamus and pituitary lead to GHD, a second mechanism for growth failure (Preece et al. 1977). In such a

case, the physician can use the individual's GH levels as a scale to adjust the values of the specialized growth chart to that individual's needs. In this way disease specific growth charts are not only targeted to groups but also flexible enough to accommodate individuals.

Without disease specific growth charts, secondary diseases affecting growth are very difficult to detect. While the incidence of rare diseases affecting growth within rare syndromes is extremely low, the benefits of early detection to the individual child are immeasurable. Once a child is below the lowest percentile on the chart, patterns in growth are very difficult to interpret. For example, the mean for Turner syndrome does not parallel the mean for normal children. Consequently, when clinicians imagine parallel lines below the lowest percentile to track children with a genetic syndrome, the results are nonsensical. In one case from the literature, deceleration in the height of a child with Turner syndrome could have been attributed to the syndrome, in which case another etiology would not have been pursued (Ranke 1989). Because her physician used a disease specific chart revealing that the child fell below the second percentile (-2 SD), he initiated a diagnostic workup which led to her diagnosis with Crohn's disease. Her recovery to a height just below the mean on her disease specific chart after treatment, roughly paralleling her growth at age 2-3, significantly demonstrates the utility of the disease specific chart. Meanwhile, her growth after recovery is still so far below the lowest percentile on the normal growth chart as to be uninterpretable.

Many developers of growth charts have suggested that Z-scores can be used in diseases involving growth abnormalities, as an alternative to disease specific growth charts, to follow children and detect abnormalities. The example above also illustrates the problem with this method. The mean for Turner syndrome decreases with age relative to the mean for normal children. Therefore, every child with Turner syndrome would be suspected of having another

overlying growth abnormality. Consequently, when physicians recognize this pattern in the case of Turner syndrome, they will ignore the decreasing trend of Z-scores, attributing them to the normal course of Turner syndrome. If the individual has Turner syndrome plus a secondary disease affecting growth, physicians will therefore miss the effect the secondary disease. While the advent of routine GH treatment has changed the pattern of growth in Turner syndrome, the pattern still does not mirror the growth pattern of normal children, making the use of CDC growth charts to monitor special populations undergoing treatment irrelevant. Moreover, the majority of children with Turner syndrome in the world live in countries where human GH is still not widely available, therefore their growth follows the pattern of the disease-specific charts for Turner syndrome.

Some researchers have argued that modified Z-scores could be applied to diseases, such that the mean growth for a rare disease could be converted to Z-scores, and individuals could be compared to how closely they parallel the relationship to this mean. Apart from being quite complicated to accomplish in practice, this method has other flaws. As noted in chapter 2, the relationships between measurements calculated beyond 3 SD from the mean are purely mathematical extrapolations. In addition to lacking any empirical foundation, comparisons further from the mean are less specific. To understand this, consider the distribution within 3 SD. The distance between individual percentiles gets larger as percentiles move out from the median, or 50<sup>th</sup> percentile. The distance on a chart from 0 SD (50<sup>th</sup> percentile) to -1 SD (16<sup>th</sup> percentile) is the same as the distance from -1 SD to -2 SD (2<sup>nd</sup> percentile), which is the same as the distance from -2 SD to -3 SD (0.14<sup>th</sup> percentile). Put in another way, the distance from the 50<sup>th</sup> percentile to the 30<sup>th</sup> percentile is 0.5 SD. The distance from the 2<sup>nd</sup> percentile to the 1<sup>st</sup> percentile is *also* 0.5 SD – the same linear distance on the chart. Therefore, as data points move beyond 3 SD from

the mean, comparisons between these points involve hundredths of percentiles. From one perspective, this is the reason why using Z-scores is so much more precise than percentile measurements. From another perspective, it is why judgments and comparisons using Z-scores may be so much *less* accurate than using disease specific references.

### **Understanding the Pathogenesis of the Growth Disorder**

The process involved in developing disease-specific growth references is essentially an in-depth examination of the natural history of the disease. Researchers can learn much about the nature of the disease, and may even receive insight into the causal relationship between the disease and the growth failure. Many mechanisms of disease can cause growth failure, and in genetic syndromes involving growth these causes are usually multiple and overlapping. Understanding the mechanisms behind growth failure can not only help researchers develop targeted treatments for the growth failure, but also for other previously unrecognized components of the disease.

Congenital adrenal hyperplasia serves as an excellent example of this concept. Despite aggressive early treatment for their enzyme deficiency, these patients never achieve predicted height based on mid-parental estimates. Previously, studies of growth found that no pubertal growth spurt occurred in these children. However, in designing a comprehensive longitudinal study of growth in CAH, Hargitai et al., discovered that the longitudinal growth spurt had been smoothed over in previous cross-sectional studies (Hargitai et al. 2001). In studying the growth velocity data, they learned that the growth spurt was spread over a much wider age range than normal. Because early puberty is associated with shorter adult height, this knowledge can lead to strategies to modify treatment to promote growth using the underlying mechanism of the disease.

Turner syndrome provides another example. Contrary to the popular opinion that girls with Turner syndrome are overweight, Rongen-Westerlaken et al., Suwa et al., and Bernasconi et al.,

all found that, based on their reference charts, these girls are relatively underweight for their age (Bernasconi et al. 1994; Rongen-Westerlaken et al. 1997; Suwa 1992). Ranke and Price also discovered that weight-for-height in Turner syndrome is normal until girls reach a height of 120cm at which point this value increases on average (Price and Albertsson-Wikland 1993; Ranke et al. 1983). Rongen-Westerlaken et al., also found that BMI is higher in girls with Turner syndrome, on average, than in normal children, and that the drop in BMI seen on reference charts of normal children between ages 4-7 does not occur in Turner syndrome.

Several additional examples exist. While creating a growth reference for Marfan syndrome, Erkula et al., recognized that puberty occurs 2.2-2.4 years earlier in Marfan syndrome than in healthy individuals (Erkula et al. 2002). Researching growth in Williams syndrome, Martin et al., detected an overall increase in BMI relative to the normal population, a conclusion which had been suggested previously but never confirmed. This finding suggests that children with Williams syndrome are at higher risk for obesity than the general population (Martin et al. 2007). In a study on growth in 57 individuals with CDCS, Collins et al., found a significant correlation between FOC and BMI that was previously unrecognized (Collins and Eaton-Evans 2001). Surprisingly, in a study of 315 individuals with PWS, a disease in which obesity is considered typical, 50% of the individuals had a BMI in the normal range (Wollmann et al. 1998).

### **Monitoring Growth-Promoting and Other Treatments**

Ranke describes the variability in growth relative to potential outcome based on ideal genetic and environmental conditions. He discusses achievement of growth potential in terms of the “somatostat,” a measure of how well the individual conforms to parental growth cues (Ranke 1996). As certain syndromes have an altered correlation to parental height, the extent of the influence of mid-parental height on the child’s height in specific diseases should be established.

While girls with Turner syndrome will most likely never achieve a calculated mid-parental height, girls with taller parents will, on average, be taller than girls with shorter parents, despite their genetic defect (Brook et al. 1977; Ranke and Grauer 1994). However, in other syndromes, such as Down syndrome, researchers have demonstrated the opposite – that height of the affected individual does not correlate with mid-parental height. This difference implies that only certain syndromes require researchers to adjust for race or incorporate individuals with diverse genetic backgrounds in their references. Unfortunately, for the majority of syndromes the presence or degree of correlation between growth of the affected individual and parental measurements is unknown. According to the somatostat principle, in Down syndrome and all other syndromes with altered growth potential the somatostat has an altered set point. The hormonal mechanisms behind the altered growth potential in these syndromes likely involve both central (e.g., abnormal GH secretion), and peripheral (e.g., altered hormone receptors) components.

In syndromes with altered growth, both the degree and the timing of growth are altered. Since both regulatory mechanisms and tissue receptors can be involved, growth does not necessarily parallel normal growth. Prader-Willi syndrome serves as an excellent example of this rule. During infancy, when normal children have a much higher growth velocity, PWS children experience significant FTT with very low growth velocity. When normal children move from infancy to childhood their growth velocity decreases precipitously, while in PWS weight velocity increases dramatically during this same time period (Hargitai et al. 2001).

The standard approach to evaluating drugs designed to affect growth directly is to measure height velocity before and after treatment, or to compare height velocity to accepted standards (Ranke 1989). Disease specific standards for growth velocity exist for very few disorders, as discussed in chapter 3. However, the utility of disease specific height velocity charts in

measuring response to GH is dramatic, as the following example illustrates. If an individual's growth velocity were plotted on the standard growth velocity charts pre- and post-treatment with GH, the values would likely fall within the normal range for velocity, while on the Turner-specific charts, the values would likely migrate out of the normal range, above the 98<sup>th</sup> percentile. This example illustrates why it is crucial to consider how to best monitor the effects of treatment when growth is an outcome. When measuring response to treatment, the most important issue to consider is the ultimate goal of treatment. In the case of height, individuals will be more concerned if they have achieved a "normal" height relative to the normal population, than if they have increased 0.5 SD relative to the disease specific growth reference (Ranke 1989). In this case, the disease specific reference is useful to set a benchmark for the population, from which the physician can derive a goal of treatment. If the patient's mid-parental-adjusted target adult height is 150 cm based on the prognosis from the disease-specific charts, and this value is at the 2<sup>nd</sup> percentile on the standard charts, then the physician should chose a value in the normal range, say 155 cm (the 10<sup>th</sup> percentile) as a goal height for treatment.

Specialized growth charts serve as an excellent tool to track changes in diseases with abnormal growth patterns. The effect of GH in many disorders has become much clearer through the use of specialized growth charts. Moreover, as new treatments are developed to target the underlying pathophysiology of rare diseases, growth will be an important outcome measure when studying these treatments. In fact, without specialized growth references, research on future medical and genetic interventions will be hampered by inadequate estimation of the significance of these interventions.

### **Should Specialized Growth Charts Be Used for Diagnosis?**

Many clinicians believe that growth parameters of some sort, growth velocity, growth acceleration, or even absolute measurements, can be used to screen for or even diagnose specific

diseases. One of the most tenable examples is short stature. Typically, a cutoff value is chosen on standard growth references, and if a child is below that cutoff on the height-for-age chart, that child has short stature. This definition is based on a comparison between the individual and the general population. The cutoff changes with one dimension – age. Therefore, it is a useful, but perhaps not a sensitive or specific measure of disease state. The simplest extension of this principle that qualifies as a disease-specific growth chart is to incorporate mid-parental height as a target height, and to establish a range 2 SD above and below this target height that will be accepted as normal variation for this individual. Thus a child whose parents were both at the 5<sup>th</sup> percentile adult height would not be considered abnormal at just below the 2<sup>nd</sup> percentile, while a child whose parents were both in the 95<sup>th</sup> percentile would be considered markedly abnormal if he were at the 5<sup>th</sup> percentile. Although such a method does not satisfy some of the conditions for a valid reference standard, it is a more specific technique than the standard cutoff and can result in fewer false positives.

In another example, change in a measurement as it changes with age may be the concerning value. For instance, although absolute measures of FOC in infancy can be plotted within the normal range, the dimension changes so quickly that the change in size over the change in time is a more sensitive indicator of abnormal growth. For example, in RTT changes in head growth velocity are noticeably different from the normal population by three months of age. Screening for this type of parameter would increase the sensitivity for such conditions. In this case the specialized growth reference would be a growth velocity reference, not specific to a particular disease, but specific to a different interpretation of data. For instance, while growth attenuation in infancy is not currently screened for using growth velocity charts, PWS provides

an example to support the use of growth velocity screening for early diagnosis (Wollmann et al. 1998).

In a third example, a child is prenatally diagnosed with Down syndrome through chorionic villous sampling. As demonstrated in the previous chapter, in Down syndrome abnormal growth is the rule, not the exception. This child's physician may be interested in either the one-dimensional measurement value of the first example or the growth velocity measurement of the second example. However, if the physician's goal is to use these values to screen for abnormalities, they must be compared to what is normal *for that child* given the child's genetic constitution. The most sensitive and specific method to achieve this goal is by using a disease-specific growth chart.

Pursuing this example further, children with Down syndrome are highly prone to develop hypothyroidism, and one of the most important screens for hypothyroidism is growth. However, if an infant with Down syndrome is assumed to be at the 2<sup>nd</sup> percentile for height solely on the basis of his genetic condition, then the growth chart loses its utility as a screening tool. Those who oppose the use of disease specific growth charts suggest that in addition to plotting these children at a location which may be below the lowest available percentile line, the physician should calculate a Z-score to determine exactly how many SDs the individual is from the mean. This method does provide an exact proportionate distance from the mean of a normal reference population, but the reference population does not contain information about how this child with this genetic condition should progress from that point. In the case of Down syndrome, where many individuals move closer to a normal weight after infancy, if the individual in our example was calculated to be -2.5 SD at the first visit, and was -2.5 SD at a follow-up visit, the physician might view this as normal growth. If, however, this growth were compared to the standard

growth for children with Down syndrome, the physician would see that the child has moved down a percentile line. In fact, to have “maintained” his percentile rank on the Down syndrome growth chart, the individual would have had to progress from -2.5 SD up to -2.0 SD on the normal growth reference.

In all the previous examples of disease-specific growth references, the concern for missing comorbid diagnoses affecting growth when plotting special populations on references of the normal population concerned the absence of reference points below the lowest percentile line on the chart. Alternately, in PWS the issues are both having no reference lines *above* the highest percentile line on the chart, and having a 3<sup>rd</sup> percentile, or “cut-off” line for abnormal growth, that is too low. A child with PWS with a weight at the 3<sup>rd</sup> percentile on a normal growth reference may have been demonstrating evidence of a comorbid disease for some time, as this value would be significantly below the 3<sup>rd</sup> percentile for weight on the PWS-specific chart (Hauffa et al. 2000). Rather, the lowest cut-off for PWS on a standard chart may be the 10<sup>th</sup> or 25<sup>th</sup> percentile for weight.

### **Criteria for a Valid Standard of Disease Specific Growth**

Although researchers have produced many disease-specific growth references over the past five decades, most suffer from significant flaws. The majority are descriptive, and are not amenable to the calculation of SD scores. Stevenson et al., comment on the importance of understanding the difference between prescriptive and descriptive growth references (Stevenson et al. 2006). While representative samples from any diseased population will include individuals of variable ultimate potential and a mixture of comorbid disorders, standards of growth should include clinically useful associations between growth and functional outcome. Prescriptive growth references should take into account the individual’s genetic potential based on conditional factors. If studies demonstrate that mid-parental height correlates with final height,

this covariate should be included. If genetic variability can be linked to growth potential, this should be taken into account. If growth outcomes in a disease can be linked to interventions earlier in life, growth under the influence of these interventions should be considered the standard to strive for in management. This approach is quite different, in fact superior in terms of both screening and prediction, to the approach generally taken in the construction of references of growth for the normal healthy population. However, it is an arduous approach that requires validation of all prescriptive factors to be incorporated in the standard. The first step is to collect high quality data, both longitudinally and cross-sectionally, which can account for changes both within and among various age groups, genetic backgrounds, and levels of overall care.

Such a standard must incorporate precise, high quality data, on the level of the reliability achieved by the WHO study. The timeframe of data collection should be relatively narrow to avoid secular influences, or if the time span is wide, the data should be examined for secular differences. The highest quality statistical data processing should be used, including data cleaning techniques (chapter six), summary statistics and goodness-of-fit testing. Ideally, longitudinal data would be collected to allow researchers to calculate height velocity, and cross-sectional data would be used to fill in the “gaps” left by the bias of longitudinal data for specific well-child check-up age ranges. The resulting charts must be both graphically and numerically useful. They should either incorporate formulas from which Z-scores can be calculated directly, or include tabulated records for calculating Z-scores using a computer program.

The important question remains; should specialized growth charts be references or standards? If the ideal of the population is known and achievable, then the standard “provides a reminder of what might be possible in better conditions” (Cole 1993). Growth standards should only be norms representing optimal growth if they account for velocity differences, since obesity

and malnutrition can be within the “standard,” but still abnormal. Despite the recent WHO findings of extremely low height variability among populations, racial and ethnic differences most likely exist, and so individual standards may be necessary. In the absence of valid standards, Cole recommended against using references as norms (Cole 1993). Therefore, if specialized growth charts are adopted for counseling and research purposes, prudence dictates that they should be standards, not references.

Table 4-1. Change in height in two individuals with RTT and the corresponding Z-scores.

	Age in Years	Height in cm	CDC Z- score in SD	RTT Z- score in SD
Example 1	12	131.0	-2.7	-0.5
	14	142.4	-2.8	0.8
Difference		11.4	-0.1	1.3
Example 2	15	141.4	-3.2	0.0
	18	143.0	-3.1	-0.1
Difference		1.6	0.1	-0.1

Table 4-2. Frequency of measurements in the 2000 CDC and 2007 WHO reference chart studies compared to this RTT growth study.

Age in Month	CDC-Wt	WHO- Wt	RTT- Wt	CDC- Ht	WHO- Ht	RTT- Ht	CDC- FOC	WHO- FOC	RTT- FOC
0	40162934	838	212	7287	842	209	174	841	122
1		448	143	4231	448	136		449	117
2	34	447	119	2474	447	119	33	447	94
3	118	446	72	1562	449	67	117	448	49
4	93	444	104	2167	447	108	92	447	98
5	98	447	31	99	450	28	98	450	21
6	92	445	104	92	448	101	91	448	84
7	122	444	57	122	448	43	122	448	27
8	104	440	24	104	445	26	103	445	11
9	137	446	84	137	449	85	139	449	60
10	126	444	41	126	446	25	125	446	21
11	135	444	18	135	445	20	133	446	10
12	304	449	104	300	451	100	299	452	80
15	312	450	146	310	451	127	306	452	98
18	331	447	160	327	449	126	327	449	98
21	310	445	109	306	444	78	307	445	72
24	636	449	165	453	449	135	612	449	102
30	572	229	213	445	238	144	551	237	126
36		224	231		229	185		226	134

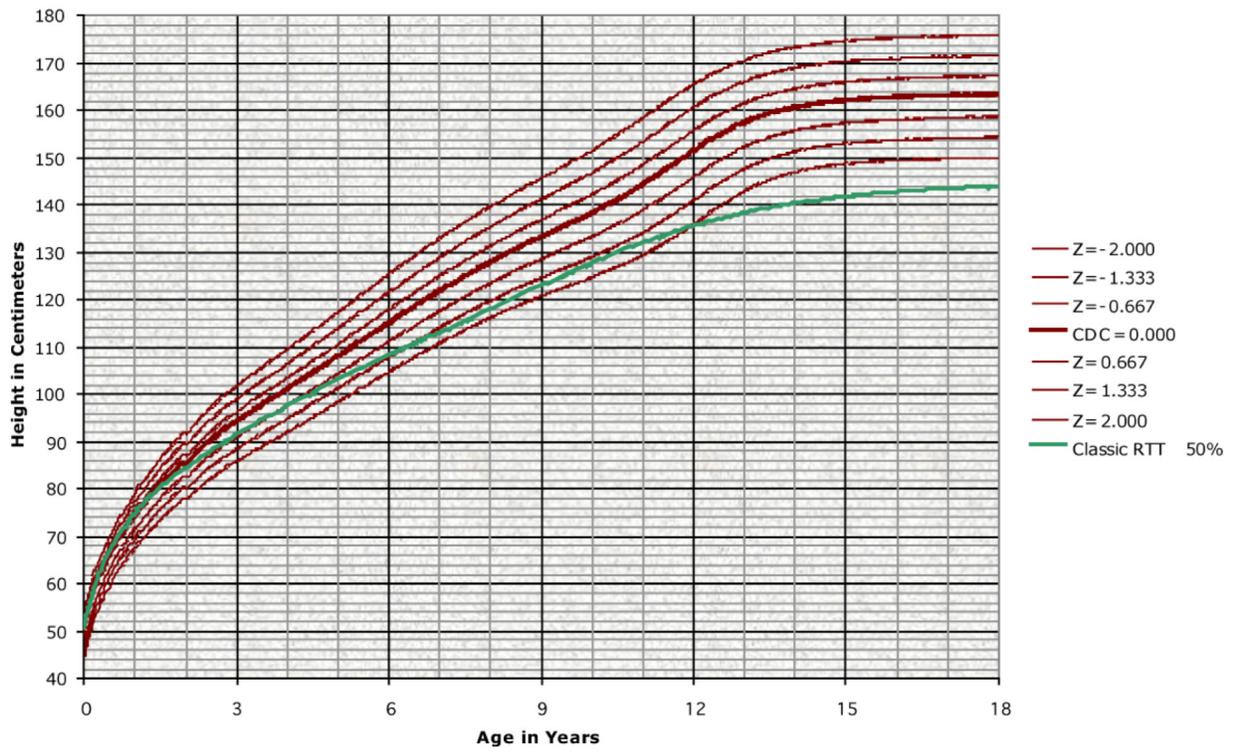


Figure 4-1. CDC percentiles for height with classic RTT mean.

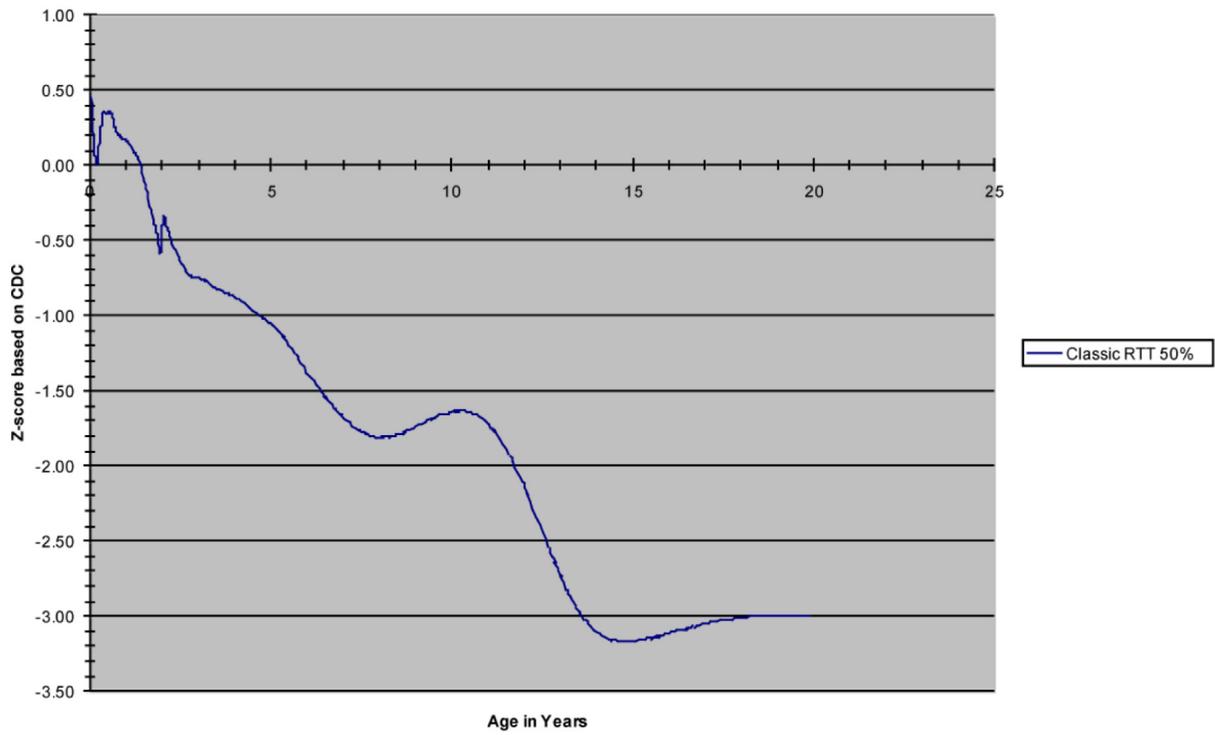


Figure 4-2. CDC height Z-score chart with classic RTT mean Z-score plotted against age.

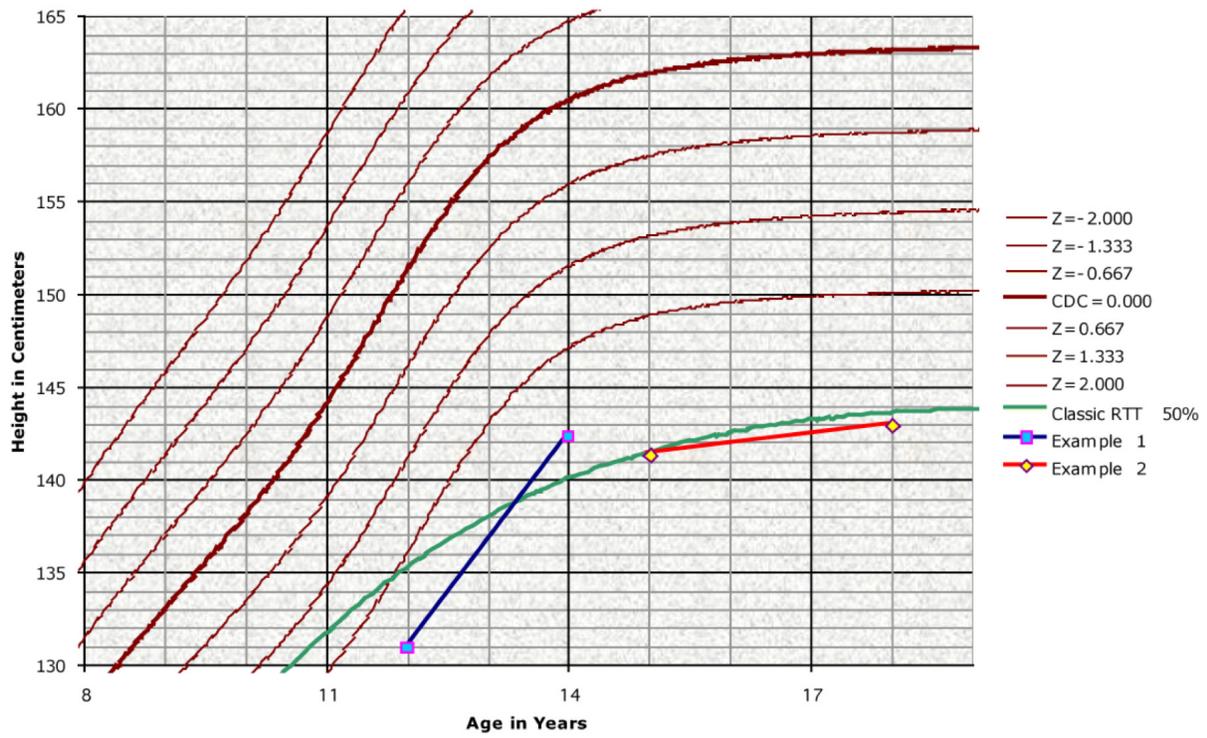


Figure 4-3. Examples of how trends can be misinterpreted if non-specialized references are used for genetic syndromes.

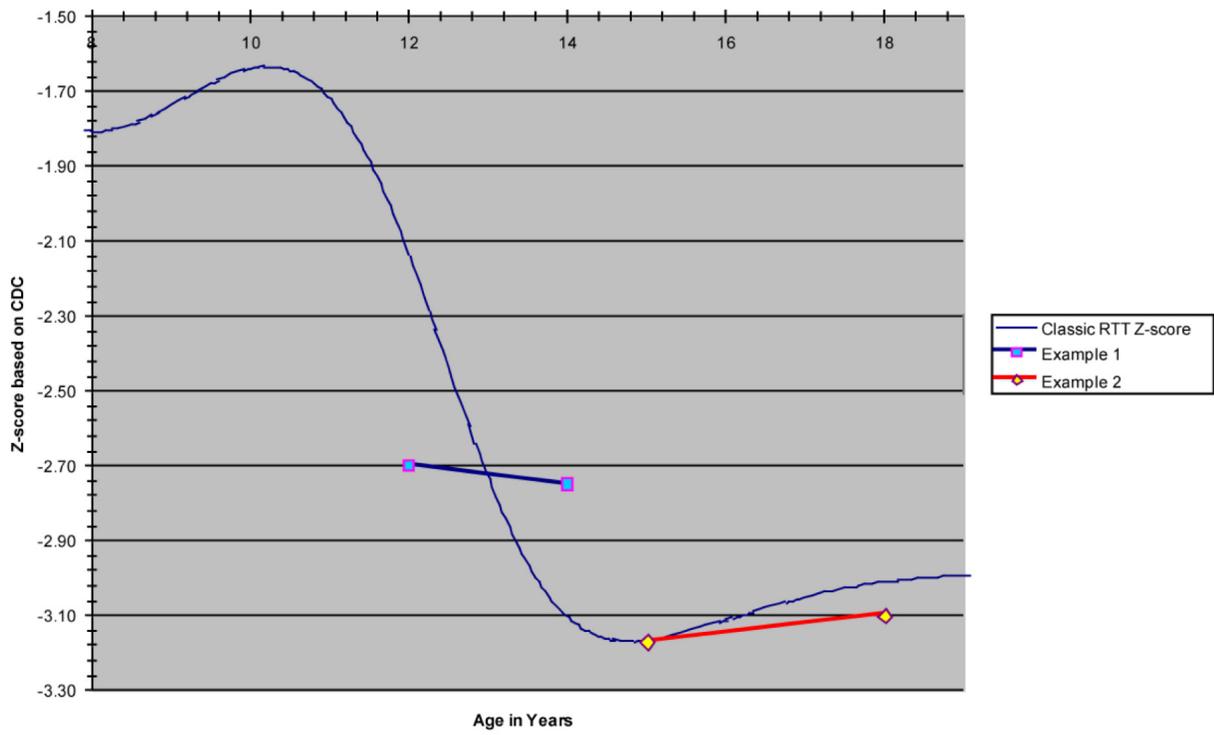


Figure 4-4. Examples of how using Z-scores from standard references can misinterpret the influence of interventions in genetic syndromes.

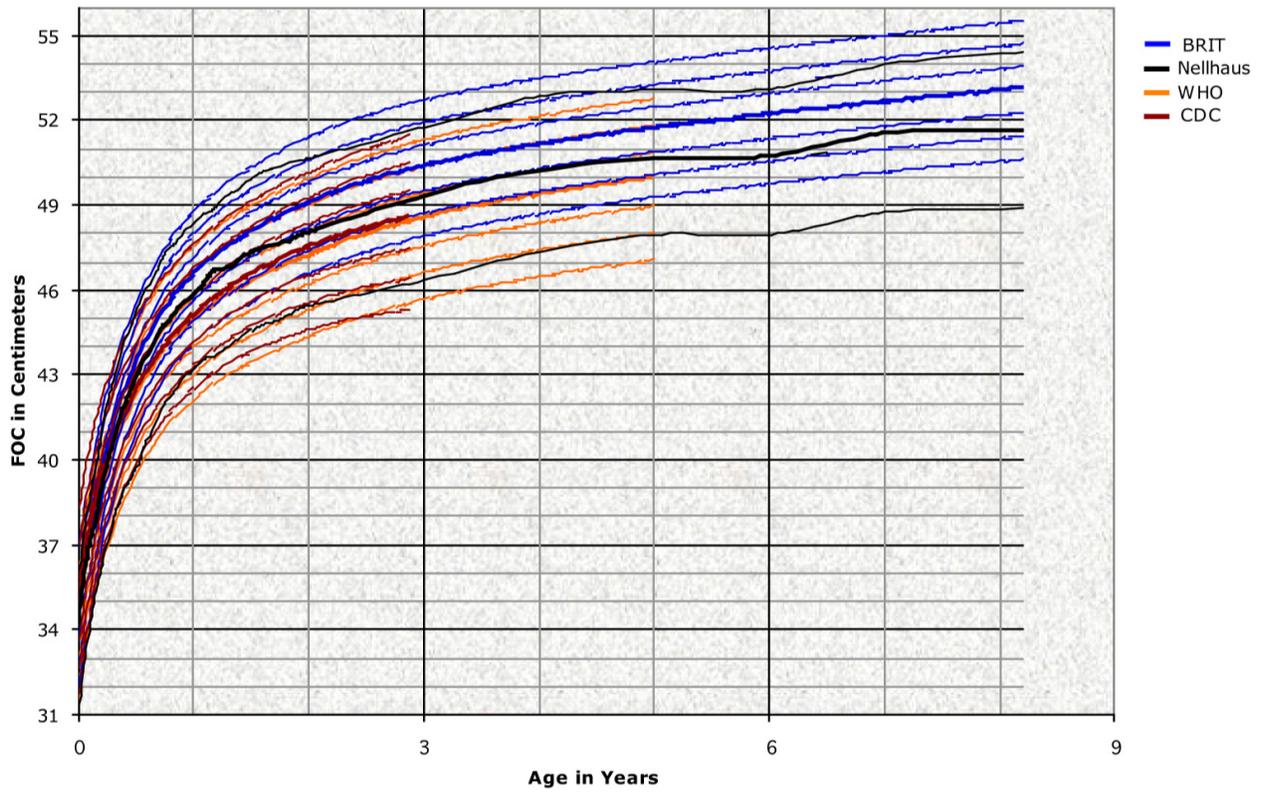


Figure 4-5. Differences in mean FOC among four commonly used references.

## CHAPTER 5 RETT SYNDROME: A MODEL FOR SPECIALIZED GROWTH PATTERNS

### **Introduction**

This chapter introduces a rare disease which is an ideal example of genetically abnormal growth patterns. Rett syndrome is a rare, X-linked dominant neurodevelopmental disorder of females characterized by apparently normal early development, psychomotor regression involving loss of purposeful hand skills, language and social interaction, and the emergence of stereotypic hand movements. Although Andreas Rett described this pattern in 1966 (Rett 1966), the medical community did not recognize RTT until Bengt Hagberg and colleagues reported thirty-five cases in 1983 (Hagberg et al. 1983). Despite this initial gap of almost two decades, in the 25 years since then, researchers have elucidated many clinical, neurobiological and genetic aspects of RTT, including the identification of mutations in the Methyl CpG Binding Protein 2 gene (*MECP2*) in the majority of RTT individuals (Amir et al. 1999).

### **Epidemiology**

Rett syndrome is the leading cause of profound cognitive impairment in females. Incidence estimates vary from 0.43 to 0.71 per 10,000 females in France (Bienvenu et al. 2006) to 1.09 per 10,000 females in Australia (Laurvick et al. 2006). Rett syndrome transmission is sporadic, the risk of recurrence being less than 0.5%. No racial or ethnic predilection exists.

### **Clinical Features**

The pattern of development in RTT is unique. Girls appear normal in early infancy, achieving appropriate developmental milestones including sitting, walking, single words or phrases, hand transfer and pincer grasp. Although parents and physicians usually do not recognize abnormalities until six to eighteen months of age, parents later remark that the child was “too good,” and relatively hypotonic from birth. A review of videos from early infancy

revealed abnormalities including tongue protrusion, stiffness, asymmetric eye opening, unusual finger movements, abnormal facies and a “bizarre” smile (Einspieler et al. 2005).

Signs and symptoms develop in a characteristic order. One of the earliest signs, growth failure, begins with deceleration of fronto-occipital FOC as early as two to three months of age. However, microcephaly is only present in 50% at age five (Schultz et al. 1993), and deceleration of head growth is no longer a necessary criterion for classic RTT. Weight velocity decelerates at one year, and linear growth follows at 15 months. The median for weight and height both fall below the 5<sup>th</sup> percentile of normal children by age seven. Early symptoms include increasing irritability and a plateau in the acquisition of motor and language skills. Regression begins between 6 months and 2.5 years with the loss of fine motor skills (interest in playing with toys or manipulation of objects). Communication skills can deteriorate abruptly or insidiously, and speech disappears between 9 months to 2.5 years. Decreased socialization gives the impression of autism. As purposeful hand use deteriorates, midline stereotypic hand movements emerge. Common hand stereotypies include wringing, washing, tapping, mouthing, picking, claspings, squeezing and finger rubbing and can incorporate foot and oromotor activity. Over time girls build a repertoire of evolving stereotypies and additional behaviors that include bruxism, air-swallowing, hyperventilation and breath-holding (often alternating), and may include self-mutilation. While stress exacerbates these incessant behaviors, stereotypies and breathing disorders cease during sleep (Percy 2007b). This diversity results in a spectrum of clinical phenotypes that are dissimilar and confound diagnosis.

After regression, development stabilizes and then gradually improves. Eighty percent learn to walk with a dyspraxic, wide-based gait, and although many lose the ability to ambulate during regression, 60% remain ambulatory in adolescence. Social interaction and decision-making

abilities improve throughout childhood, while stereotypies and breathing abnormalities may intensify. Communication is primarily nonverbal, and girls indicate desire through eye gaze, resulting in a characteristically intense, piercing gaze. Interaction improves with age, and autistic features are replaced by a tendency to socialize and seek attention. As girls develop into women, hand stereotypies and breathing dysregulation decrease in intensity or disappear altogether.

Hagberg originally described four stages of progression (Hagberg and Witt-Engerstrom 1986). The transition between stages is difficult to discern, and the length of each stage and age of transition are difficult to predict. However, staging offers a temporal profile that can be used for anticipatory guidance (table 5-1).

Additional features present at different ages and with variable incidence. While EEG abnormalities are ubiquitous after age 2, seizure prevalence estimates range from 30-80%, and seizure types vary. However, many paroxysmal events are non-epileptic on video EEG, even in those with an abnormal baseline EEG (Glaze 2005). Other neurologic abnormalities are common including dystonia and autonomic dysfunction. Girls typically have small, cold feet as a result of autonomic dysfunction, a phenomenon that reverses after lumbar sympathectomy during scoliosis surgery. Growth failure is frequently accompanied by osteopenia. Gastro-esophageal reflux (GER) is common, and constipation is almost universal. Scoliosis is present in 80% and can be severe, requiring surgery in up to 10%. More recently recognized features include increased incidence of cholecystitis and prolonged QT syndrome (Sekul et al. 1994). Oddy et al., have suggested that nearly all of these features can have an impact on growth, and that growth failure is most likely due to a combination of all or most of them (Oddy et al. 2007).

Although Andreas Rett considered the disorder progressive, research on the neurobiology and pathophysiology indicates that RTT is a stable neurodevelopmental condition, not a

neurodegenerative disorder. This notion is borne out by the longevity in RTT. Survival is normal through age 10, and women are likely to outlive their parents. Survival to age 35 is 70% of the normal population; so long-term care plans are crucial. Notably, RTT survival is superior to other disorders of profound cognitive impairment, averaging 27% at age 35 (Percy 2002). While sudden death is more common in RTT, the etiology is unclear, and may be related to seizures, autonomic dysfunction, or cardiac conduction (Julu et al. 2001; Kerr et al. 1997).

### **Diagnostic Criteria**

Despite the discovery of a genetic etiology, the diagnosis of RTT remains clinical (Erlandson and Hagberg 2005). With wide variability in phenotype, criteria aid in diagnosing classic RTT and distinguishing and categorizing atypical variants (Hagberg et al. 2002). Based on the most recent revision of these criteria, individuals with classic RTT fulfill all 8 necessary criteria (table 5-2) and meet none of the exclusion criteria (organomegaly, retinopathy, metabolic disorder, acquired neurologic disease or trauma). Individuals have atypical RTT if they meet three of six atypical criteria and five of eleven supportive criteria (table 5-3). Within the category of atypical RTT are four variants: 1) early onset seizures, 2) preserved speech, 3) delayed onset, or *forme fruste*, and 4) congenital onset with absence of normal early development. Of all individuals with RTT characteristics, >80% have classic RTT and 15-20% have atypical RTT (Percy 2007a). A third category, provisional RTT, exists for those who may not have manifested all the components of classic RTT. Although diagnosis of RTT does not require *MECP2* testing, the role of *MECP2* mutations in RTT and other disorders is under investigation.

### ***MECP2* Mutations**

Given that RTT is classically X-linked dominant, early efforts focused on mutations of the X-chromosome. Although familial cases were scarce, linkage studies helped narrow the search to Xq28 and ultimately identify mutations in *MECP2* as the etiology of RTT (Amir et al. 1999).

Researchers soon found several *MECP2* mutations causing RTT. Although over 200 specific mutations are now associated with the RTT phenotype, 66% of affected individuals have one of eight common mutations.

Over 95% with classic RTT and over 50% with atypical RTT have a mutation in *MECP2*. However, not all with a *MECP2* mutation have RTT, and those with pathological mutations range from asymptomatic, to mildly learning disabled, to autistic; some may fulfill diagnostic criteria for Angelman syndrome (AS) (Jedele 2007; Watson et al. 2001). Phenotype in X-linked dominant disease depends on the process of lyonization, or X-chromosome inactivation (XCI), as well as other epigenetic factors (Amir et al. 2000). X-chromosome inactivation occurs early in embryonic life, and randomly silences one of the two X chromosomes in each cell for the rest of the female's life. Although XCI is random, unbalanced silencing can either mask or expose a mutation on one of the X chromosomes. Family studies have shown that identical mutations with differing degrees of XCI can result in dramatically different phenotypes, including "silent" carriers or monozygotic twins in which one girl with balanced XCI developed RTT and the other, with skewed XCI, was asymptomatic (Hoffbuhr et al. 2002).

The association between genotype and phenotype is complicated and not thoroughly elucidated. Although XCI can affect the phenotypic penetration of a *MECP2* mutation, the majority of individuals have balanced XCI. Other factors, such as the functional site and the type of mutation can account to some extent for the variable loss of function of the final protein product. In fact, recent studies have shown that in most cases XCI fails to explain the degree of severity of females with X-linked diseases (Takahashi et al. 2007; Xinhua et al. 2008). Due to small sample sizes and inconsistent rating scales, most comparisons of mutations with clinical phenotype have found only weak trends. However, two recent studies using consistent criteria to

assess larger samples found significant associations. Neul et al., found that ambulation, hand use, and language ability are more severely affected in those with the R168X mutation than in those with the R294X and late carboxy-terminal mutations. Those with the R133C mutation are less severely affected than those with R168X or large deletions. Surprisingly, while those with the R306C mutation (thought to confer a less severe phenotype) showed better scores for ambulation and hand use, their language was severely affected (Neul et al. 2008). Comparing similar criteria, Bebbington et al., found that R270X and R255X are the most severe, while R133C and R294X are the mildest (Bebbington et al. 2008). The authors of this chapter have found similar associations regarding hand use, ambulation, language, and overall severity. In general, nonsense mutations confer a more severe phenotype than missense. Notably, these findings did not change when controlling for age, arguing against age as a mediator of phenotypic heterogeneity. In comparing two milder mutations, R133C and R306C, seizure severity is significantly lower in R306C (Tarquinio, DC et al., Phenotypic Differences in RTT Are Associated with Specific *MECP2* Mutations, unpublished).

Several males have been identified with *MECP2* mutations, and a recent publication suggests that as many as 1.3-1.7% of males with mental retardation (MR) have mutations in *MECP2* (Villard 2007). Males with *MECP2* mutations fall into three general categories. First, mutations found in females with RTT cause a severe neonatal encephalopathy in males resulting in death during infancy. However, males with a *MECP2* mutation and either Klinefelter syndrome (47 XXY) or somatic mosaicism of the X chromosome have a RTT phenotype “diluted” by the presence of a normal *MECP2* gene. Second, males with mutations that do not cause RTT in females can present with a variety of phenotypes ranging from mild MR to severe cognitive impairment with or without motor abnormalities. Third, males with duplication of the

entire *MECP2* gene, as well as other genes at the Xq28 locus, present with infantile hypotonia, recurrent respiratory infections, seizures, spasticity, absent speech, and severe MR.

### **Genetic Testing**

Testing for *MECP2* is expensive and should be approached in a logical fashion. As the phenotypic range is so broad, physicians are tempted to test all children with MR or autism. However, testing should be reserved for specific scenarios. Children with clinical criteria of classic or atypical RTT should be tested. Females, age six to twenty-four months with features of RTT, should be tested if they also display the following: low muscle tone, deceleration of FOC, or unexplained developmental delay. Three other situations warrant *MECP2* testing: 1) females who fulfill clinical criteria for AS but have negative methylation or mutation studies at the AS locus, 2) males with X-linked MR and normal FXS testing, and 3) unexplained neonatal or infantile encephalopathy (Percy 2007b). All four exons of *MECP2* should be sequenced first. If sequencing is unrevealing, then testing for large deletions is appropriate. In familial X-linked MR, or the third male phenotype above, duplications should be pursued. Genes other than *MECP2* may be responsible for a minority of those with RTT, but screening for mutations in related genes is currently not recommended (Kerr and Ravine 2003).

### **Neurobiology and Pathophysiology**

Methyl CpG Binding Protein 2 (MeCP2) is involved in repression of gene transcription. Although expressed in all body tissues, it is most prominent in the central nervous system. Abnormal MeCP2 protein results in immature neurons (small with abnormal dendritic morphology) via faulty gene repression and subsequent increase in transcription of yet to be defined proteins. Autopsy studies reveal that brain weight in RTT is reduced in all age groups to 60-70% of expected. Frontal cortex and deep nuclei have reduced volume, and melanin pigmentation is decreased in the substantia nigra. The neuropil is denser in RTT; small, tightly

packed neurons possess fewer processes than normal. Dendrites are short and primitive with reduced arborization, leading to fewer synaptic connections. Cortex in RTT shares remarkable similarities with that in AS, FXS, and Down syndrome. However, no evidence of neurodegeneration exists – instead, the arrest of normal neuronal maturation in the third trimester or early in infancy represents a neurodevelopmental disorder.

Clinically, RTT suggests global neuropathology; however neurophysiology studies have elucidated specific deficits. The brainstem exhibits inappropriate serotonin transporter binding in vagal nuclei, which could explain poor autonomic control over gastrointestinal and cardiac functions (Paterson et al. 2005). Moretti et al., demonstrate abnormal hippocampal synaptic connections in a mouse model with abnormal socialization and poor target focus reminiscent of the motor apraxia seen in RTT (Moretti et al. 2006). Other models show abnormal secretion of brain-derived neurotrophic factor (BDNF), norepinephrine, and serotonin in brainstem nuclei and adrenal chromaffin cells (Chang et al. 2006; Sun and Wu 2006; Wang et al. 2006).

Although the targets of MeCP2-mediated gene silencing are not completely known, the process of silencing is at least partially understood. Methyl CpG Binding Protein 2 binds to CpG dinucleotides on methylated chromatin in gene promoter regions. Methyl CpG Binding Protein 2, along with the corepressor protein Sin3A attracts Histone Deacetylase (HDAC) to methylated DNA effecting gene silencing. Loss of HDAC activity results in increased transcriptional noise, overtranscription of certain genes, and downstream effects on other processes (Kerr and Ravine 2003). The role of transcription repression in neurodevelopment is unclear, but decreased transcriptional noise presumably promotes efficient cellular function.

In vitro studies suggest a mechanism for phenotypic variability. An R106W mutation results in a 100-fold reduced affinity for methylated DNA (Ballestar et al. 2000), while T158M

causes a moderate reduction in binding (Kudo et al. 2001). Recent experiments suggest that R133C affinity is similar to that of the wild-type protein (Galvao and Thomas 2005).

While much remains to be learned about the role of MeCP2, one of the most encouraging experiments asked – could mature individuals with defective MeCP2 benefit from presence of the normal protein? Researchers silenced *MECP2* in mice with a genetic “switch” and activated the gene after the classic phenotype was evident. This proof of concept experiment showed that neurological defects can be reversed by the presence of a normal *MECP2* gene (Guy et al. 2007). A similar experiment demonstrated partial recovery of function (Giacometti et al. 2007).

### **Management**

No treatment targets defective MeCP2 production in RTT. Strategies for modifying protein production or inserting exogenous DNA have not reached the stage of clinical application. Therefore, management strategies are supportive, symptomatic, and preventive. One of the primary concerns and principle targets of management strategy is growth failure.

The degree of cognitive impairment in RTT is very difficult to assess. Classical methods reveal a mental age of 8-10 months and a gross motor age of 12-18 months, and tests using visual response also show severe impairment. Except in rare instances, girls never acquire adaptive skills such as dressing, and toileting. However, aggressive physical, occupational, and speech therapy, and augmentative communications are crucial to improve functional ability and prevent deterioration. Girls with RTT demonstrate universal appreciation of music, and music therapy should cultivate personal interactions and choice-making (Kerr and Ravine 2003).

Breathing irregularities can be dramatic and involve breath-holding in excess of one minute or dramatic abdominal distention from air swallowing. However, all of these behaviors should subside during sleep. Any irregular breathing or apnea during sleep should prompt an investigation for obstructive sleep apnea. Interrupted sleep, playing or laughing in bed, are

common and may disrupt caregivers' sleep. Melatonin can help induce sleep, but will not prevent arousals. Antihistamines may restore sleeping patterns, but tachyphylaxis is common. Zolpidem and trazodone are effective sleep aids, and have been successful in RTT (Prater and Zylstra 2006).

Seizures in RTT can be difficult to differentiate from non-epileptic events. If diagnosis is unclear, video electro-encephalography (VEEG) should be pursued. The EEG typically shows background slowing with recurrent spike or slow spike-and-wave activity. In the absence of clinical seizures, this pattern alone does not warrant anti-epileptic medication. Clinical seizures frequently respond to carbamazepine, sodium valproate, or lamotrigine. However, sodium valproate inhibits HDAC and could exacerbate the effects of a *MECP2* mutation (Phiel et al. 2001). With potential for anorexia, topiramate could exacerbate growth failure. Levetiracetam may provoke adverse behaviors or exacerbate existing behaviors. Many non-epileptic events are related to anxiety, and can be treated effectively with selective serotonin reuptake inhibitors.

As the majority of girls with RTT develop scoliosis, screening should begin at an early age. Scoliosis occurs in 8% before school age and 80% by age 16. Bracing is used for curves over 25 degrees, although efficacy is unclear. Surgery is effective for curves over 40 degrees. Girls should receive adequate calcium and vitamin D, and should bear weight often, in a standing frame if necessary. Dual-energy X-ray absorptiometry scans should be used to follow osteopenia.

Routine care should include bi-annual consultation with an experienced nutritionist. Girls with RTT have increased protein and calorie requirements. Feeding difficulties are common, resembling those of Parkinson's disease, and occupational therapists can advise on positioning and augmentative devices. Constipation is pervasive in RTT, and either magnesium hydroxide suspension or polyethylene glycol powder can be titrated into daily fluids to achieve normal

bowel movements without the complications of enemas or mineral oil. Reflux is a common source of irritability and should be diagnosed and treated appropriately. In severe growth failure a gastrostomy tube allows caregivers to provide adequate calories and nutrition.

### **Summary**

While a rare disease, RTT is a common cause of profound cognitive impairment in females. The unique and characteristic features of RTT change with age, and diagnosis depends on understanding the course of the disease. While informative, genetic testing has not replaced clinical criteria as the means for diagnosing RTT. Testing for *MECP2* should be pursued in specific instances, and when in doubt about what testing to pursue, physicians should consult a geneticist or pediatric neurologist. Although there is no treatment to address the genetic defect in RTT, many management strategies exist to address the diverse clinical manifestations of the syndrome. Additionally, trials are in planning or currently underway to study the efficacy of interventions to manage osteopenia, anxiety, and breathing disorders, and to assess quality of life. The RDCRN study on the natural history of RTT discussed in chapter seven is laying the foundations for studies that will test the efficacy of novel treatments. Much has been learned about the role of MeCP2 in neurodevelopmental disorders over the past decade. With RTT serving as a model for other neurodevelopmental disorders, future research on both the natural history of RTT and the functional aspects of RTT neurobiology and neurogenetics will help to unravel the mysteries behind this enigmatic class of disorders. In the near future animal trials of medications to improve translation of DNA into functional MeCP2 protein will take place, as a precursor to clinical trials. Additionally, the hope remains that gene therapy will provide an effective treatment for the severe developmental disability in RTT. In order to measure the outcomes of these future research endeavors, accurate growth references must be available to researchers.

Table 5-1. Stages of RTT.

Stage	Age at Onset	Duration	Characteristics
I – Early Onset Stagnation	5 mo – 1.5 y	Weeks – Months	Halted developmental progress Early postural delay Dissociated development Stereotypical hand movements “Bottom-shufflers”
II – Developmental Regression	1 y – 4y	Weeks – Months	Loss of acquired skills: fine finger, babble/words, active playing Eye contact preserved Seizures in 15% Breathing problems modest Described as “In another world”
III – Pseudostationary Period	At end of Stage II	Years – Decades	Prominent hand apraxia/dyspraxia Slow neuromotor regression Preserved ambulation Some communicative restitution “Wake up” period
IV – Late Motor Degeneration (IVa – ambulated previously, IVb – never walked)	When Stage III Ambulation Ceases	Decades	Complete wheelchair dependency Severe disability and wasting Rigidity, dystonia, hypomimia, bradykinesia Stable cognition, improved communication and fewer seizures

Table 5-2. Major diagnostic criteria for RTT.

All of the following:
1. apparently normal prenatal/perinatal development
2. apparently normal psychomotor development through age 5 months
3. normal head circumference at birth
Onset of all of the following after the period of normal development:
1. deceleration of head growth between ages 5 and 48 months (in most)
2. loss of previously acquired purposeful hand skills between 5 and 30 months followed by stereotypical hand movements
3. loss of social engagement early in the course (although often social interaction develops later)
4. appearance of poorly coordinated gait or trunk movements
5. severely impaired expressive and receptive language development with severe psychomotor retardation

Table 5-3. Diagnostic criteria for atypical RTT (Must meet 3 main criteria and 5 supportive criteria).

Atypical Criteria	Atypical Supportive Criteria	
1) Absent/reduced hand skills	1) Breathing irregularities	7) Cold, purple feet
2) Reduced/lost speech	2) Air swallowing	8) Sleep disturbance
3) Monotonous hand stereotypies	3) Teeth grinding	9) Screaming/laughing spells
4) Absent/reduced communication	4) Unusual gait	10) Decreased pain response
5) Head growth deceleration	5) Scoliosis	11) Intense eye contact
6) Regression followed by recovery	6) Lower extremity wasting	

## CHAPTER 6 GROWTH CHART METHODOLOGY SPECIFIC TO RARE DISEASES

### **Approach to Study Design**

Designing and constructing a growth chart involves several steps. First, the researcher must identify the reference population. Second, he must organize individuals to make the actual measurements. Third, he must analyze and reduce the data to graphical and mathematical summaries.

Cole proposed that one way to address the differences in special populations would be to recalibrate a larger reference using data from the special population. He uses ethnicity as an example, citing that it would be impractical to create large databases of anthropometrics for every ethnicity in every region. Instead, he proposes using a pre-existing larger reference as a model to be amended for creating smaller references (Cole et al. 1998). For example, if all diseases with head growth deceleration can be considered as a group, then a large database of growth trends in disease involving head growth deceleration could be created. Smaller surveys in specific diseases could then be used to summarize growth in the specific disease relative to the mean and SD of the larger population. Growth charts could be created for the disease by recalibrating the larger references based on the mean and SD of the disease. While this approach is frequently successful in describing racial and ethnic differences within normal populations, the patterns of growth in genetic syndromes are often entirely unrelated to normal growth. When disease alters the genetic and hormonal cues for growth, direct comparisons to normal individuals are unhelpful.

Waterlow et al., recommended specific criteria be followed when constructing growth references (Waterlow et al. 1977). Individuals should be well nourished, measurement techniques should be well defined and reliable, the sample should be large and cross-sectional,

and techniques for smoothing should be reproducible. Remarkably, very few growth references fulfill these criteria, and none of those in common use in the US do so.

### **Study Goals and Design**

Guidelines for selecting the study population were discussed in chapter four – the population must be racially and geographically diverse, and ancillary information about the population must be available. In the case of rare diseases, the varying severities and comorbidities of the disease must be accounted for and, if possible, represented to a degree similar to their natural incidence. If resources are scarce but a standard reference is truly inadequate for the disease of interest, an alternative to appending references of normal individuals for diseased populations is to survey diseased populations at specific ages and create summary statistics for those discrete ages of interest. For example, in Down syndrome, birth weight is significantly different from the general population, and early FTT is an issue, primarily due to feeding aversion. Therefore, early researchers focused primarily on detailed summary statistics for birth weight and interval summaries during the first three years of life. In doing so, they provided clinicians with a reliable tool for evaluating the small-for-gestational-age (SGA) neonate suspected of Down syndrome on a syndrome-specific reference, and a rough estimate of expected growth in infancy and early childhood. Once children with Down syndrome progress beyond three years, they continue to grow abnormally, but the issues in older children are primarily overweight-for-height. While obesity is a significant source of morbidity in children, it is less of an immediate problem than FTT.

A more comprehensive approach involves collecting data on children throughout childhood and adolescence. While comprehensive growth charts are more useful to clinicians following children from birth through adolescence, these charts present numerous design issues. Most studies confined to a limited age range (e.g., birth to three years) collect data longitudinally

on patients over the entire reference range. In this case, the study participants, the measurement equipment, and the anthropometrists are usually all identical, therefore variability is minimized. In studies of a comprehensive age range, data must be collected simultaneously on individuals of different ages, frequently involving multiple researchers at different study sites using different equipment. The resulting cross-sectional references, although clinically useful, must be interpreted with caution if researchers did not consider this variability in the study design.

### **Cross-sectional, Longitudinal, and Hybrid Approaches**

Overall, the cross-sectional approach to growth chart design is more common than the longitudinal approach, especially in large survey-based studies such as the NCHS study on which the CDC charts were primarily based. These studies are able to collect massive amounts of data over a wide age range in a relatively short period of time. However, they do not account for issues such as secular trends in growth. Researchers cannot adjust for the environmental influences on individuals 18 years prior to the study, therefore these charts ultimately represent growth of many different “populations” stratified by age differences, represented simultaneously within one sample. Recent evidence has shown that environmental factors such as the prevalence of breastfeeding versus formula feeding, average exposure to cigarette smoke, and rates of physical activity have changed dramatically over the past two decades, and contribute significantly to patterns of normal growth. Other factors are difficult to evaluate and incorporate, such as ethnic and racial diversity among groups of different ages. Nevertheless, this biased method remains the predominant approach to growth chart design, because cross-sectional data analyze dispersion very well.

Alternately, longitudinal studies, in addition to offering decreased variability, present more options for interpretation than cross-sectional studies. The relationships of individuals in a cross sectional chart are only meaningful when compared as a distribution at the same age. Growth

velocity cannot be calculated using cross-sectional data. However, longitudinal data from individuals can be compared throughout the duration of the study. Longitudinal data offers benefits in data cleaning and detection of measurement error not available through purely cross-sectional studies, therefore overall data quality is higher. In addition to growth velocity, researchers can calculate correlations between growth rates and other variables, such as measures of nutrition, general health, or laboratory markers of disease severity. Moreover, if the cohort is large enough, summaries of longitudinal relationships between patients allow researchers to draw conclusions about the effects of environmental modifications at different ages. In diseased subpopulations, this concept ultimately extends to the effects of treatments on growth. Investigators have used this technique to great effect while examining the influence of GH at different degrees of skeletal maturation in Turner's syndrome and GH deficiency. Especially in rare diseases, following a large enough cohort longitudinally is impractical.

Since the cross-sectional and longitudinal approaches both have benefits and drawbacks, a powerful alternative is the hybrid approach. According to Arnell, "the most accurate method for studying growth is to follow it longitudinally. The cross-sectional method is reliable, if the patient group is very large and covers every age period." (Arnell et al. 1996) The essence of the hybrid design, which takes advantage of the benefits offered by each approach, is to obtain longitudinal data for periods of rapid change when the signal-to-noise ratio will be highest. These ages are typically infancy, early childhood, and puberty. Cross-sectional data is then obtained to fill in the remaining time periods, and to fill any gaps within the longitudinal time period due to bias in interval selection for longitudinal visits. Since collection of data for growth charts is a time consuming and arduous process, longitudinal data should only be collected for periods of rapid change when they will be most useful. If longitudinal data are collected on

individuals during periods of stable growth, these data can be interpreted as if they were cross-sectional. Although these data are technically “dependent” since they come from the same individual, this dependency has little effect on the finished product. In fact an entirely longitudinal sample can be analyzed cross-sectionally with very little bias resulting from data dependency.

Another benefit of the hybrid approach involves how the researcher treats the data after it has been collected. Rare diseases present the inherent problem of insufficient numbers to produce robust references. Therefore, it is useful to collect retrospective data on each individual in addition to prospective data. However, retrospective data are longitudinal with irregular intervals, and occasionally individuals will have dozens of measurements one year and only one or two the next. This approach is problematic from the perspective of a cross-sectional reference. To incorporate these data without filtering into a cross-sectional reference would introduce bias based on higher weighting of individuals who had more measurements. One solution is to choose an interval of one month and average the measurements and the age-of-measurement within that month, thus reducing the weighting of that one individual. The bias introduced by this binning procedure is small, since individuals with very frequent measurements are rare. Alternatively, what is problematic in cross-sectional analysis is actually a boon in longitudinal analysis. The more frequent data points will produce a better longitudinal model, actually reducing bias in the model and averaging out error from poor measurement reliability.

Therefore, a hybrid approach offers the best solution to reference chart construction in rare disease. The combination of retrospective and prospective data collection allows the researcher to collect sufficient data in a short period of time. This approach also allows reliability assessment of the prospective component, and reliability attenuation through model-fitting of the

retrospective component. Some have criticized retrospective growth data as being highly unreliable due to numerous and uncontrolled sources, but these criticisms are largely unfounded. In fact, school nurses have been shown to be as reliable as trained anthropometrists at measuring height (Cotterill et al. 1993; Majrowski et al. 1994; Voss et al. 1990)

The combination of longitudinal and cross-sectional analysis allows the researcher to generate velocity references in addition to measurement references. This combination also allows the researcher to control the amount of bias introduced by weighting individuals with more frequent measurements. Many researchers have acknowledged that the hybrid approach effectively addresses the two most important issues in growth chart design, sufficient numbers and sufficient reliability (Martin et al. 2007; Partsch et al. 1999).

### **Methodology for Data Collection**

Many approaches exist for collecting cross-sectional and longitudinal data. The most reliable and most time-consuming approach is to collect all data longitudinally and prospectively as the WHO did in the MGRS. Since rare diseases do not offer large enough cohorts to follow for the duration of childhood, children of different ages must be recruited. Therefore, to collect sufficient longitudinal data on these individuals of different ages, retrospective data from variable sources must be used. All retrospective data should be tagged as to its source so that, if enough points exist and weighting can be applied, the less reliable data can be weighted lower than the more reliable. The result of this combined technique is to accrue patients enrolled at different ages into a prospective arm in which reliability can be assessed and longitudinal data can be collected, and to gather retrospective data from these individuals from all possible sources.

## **Anthropometric Measurement Techniques**

Growth chart references are an excellent example of the maxim, “garbage in, garbage out.” Researchers should make every effort to ensure that data are of the highest quality, prior to any data cleaning techniques. This includes obtaining suitable equipment, training operators in specialized techniques such as skinfold measurements and limb circumferences, and testing the reliability of the operators’ measurements periodically. The first step in developing a methodology for measuring a population is to decide on a standard for measurement technique. This guide should be as objective as possible, and no doubt should exist about reproducibility. The techniques employed in the study in chapter seven serve as a good model and are reported in detail in appendix A.

### **Inter-Observer and Intra-Observer Reliability Assessment**

Children with a genetic predisposition to altered growth are likely to be more difficult to measure accurately than normal children. Some potential sources of difficulty include scoliosis, contractures, lack of head and trunk stability, and the inability to bear weight for standing height. In such cases, anthropometrists and physicians commonly use alternative measurements to assess linear growth, including crown-rump length or sitting height, arm span, upper arm length, and lower leg length. Therefore, remeasurement and reliability testing are especially important in special populations with altered growth. The prudent researcher should assess inter-observer and intra-observer reliability on a small representative subpopulation at the outset of the study to ensure high quality data. Reliability study design is discussed in appendix B, the reliability study accompanying the study of growth in chapter seven.

### **Statistical Methodology**

Tanner states that the three reasons for smoothing empirical data are to remove measurement error, adjust for seasonal variation, and attenuate “other factors” that contribute to

variability in growth (Tanner et al. 1966a). Achieving a functional and aesthetically attractive reference requires sufficient participants, variety in racial and geographic background, and powerful statistical tools to attenuate these “other factors,” that confound distributional assumptions.

### **Power Analysis**

Unfortunately, no good statistical techniques exist to estimate the sample size needed for a particular power to detect changes in growth, or to estimate a valid curve. Since no precise tests of goodness-of-fit exist for the complicated semi-parametric estimation of growth curves, no re-sampling technique or traditional power calculation will suffice. The traditional wisdom is that in excess of 100 individuals for a longitudinal study to be measured at pre-specified time intervals. The intervals are based on the longitudinal calculations that will be performed. If growth velocity will be calculated in infancy, measurements should occur monthly; if in adolescence, annual measurements suffice. The “magic number” 100 is derived from the fact that with 100 individuals, rank-order statistics can be used to create a “percentile,” an ordering of 100 ranks from lowest to highest, without any gaps.

### **Data Cleaning**

The purpose of data cleaning is first to prepare the data to be analyzed efficiently, second to identify errors in data collection or data entry, and third to describe trends in data distribution. This process, expanded into 13 steps, is included in appendix B as it was applied in the study in chapter seven. The first component should ideally be completed prior to study implementation, but not all difficulties can be foreseen. For example, rare diseases research frequently involves collecting data on type of genetic mutation. If mutation types are many and in various categories, including all possible types as bubble entries on a scannable form is cumbersome. Therefore, many researchers opt to include blanks on study forms for copying mutation details from lab

reports, and character recognition software to interpret the records. This technique introduces two sources of potential variability. First, different labs frequently report results in different formats, and second, character recognition errors can lead to multiple expressions of the same mutation. In our study, this type of error led to redundant expression of mutations which could not be analyzed without significant recoding. The error in design was detected and remedied by including a bubble entry on a scanable form for mutation category in addition to specific type.

The second component of data cleaning involves identifying errors. Errors in data collection include measurement error due to poor technique or participants who are difficult to measure. For example, children with RTT tend to be anxious, may be resistant to temporary restraint, and may have contractures or spasticity. These difficulties can lead to poor accuracy and/or precision. Such difficulties are common in genetic syndromes, both when landmarks are displaced due to structural defects, and when individuals are anxious and unwilling to cooperate with measurements due to developmental disability. The frequency and degree of these errors can be assessed using reliability testing as described above, and appropriate corrective measures can be initiated.

Errors in data entry are common, and occur both with manual and computer-scanned entry methods. Data cleaning techniques can screen for all of these errors with high sensitivity. One of the simplest screening procedures involves searching for discrepancies in data that is represented more than once in the database. Date of birth (DOB) is frequently recorded twice, and non-zero differences in DOB can be investigated. Calculating age by subtracting DOB from date of visit (DOV) and comparing to recorded age or querying for negative age values can also reveal entry errors. Duplicate entries can be identified with most statistical software, and several permutations should be tested to search for partial duplicates. Careful identification of duplicates is

particularly important in rare diseases, because data merged from various sources may include identical patients. Determining if less obvious errors are simple errors of entry involves returning to the source documentation. Therefore, it is beneficial to have source documentation which cannot be modified, whether a paper hard copy, or a read-only tablet computer file recording of digital pen markings. One option would be to record to a digitally signed file which could not be altered. Alternatively, determining if unusual values are errors of measurement or features of the data involves the third component of data cleaning.

The third purpose of data cleaning, identifying trends, overlaps with the first component of actual data analysis – exploratory data analysis (EDA). Many abnormal values will be due to errors in collection or entry, but other apparent “outliers” will represent unusual features of the data that need to be accounted for in analysis. If these outliers are valid, then they must either be accounted for in the final distribution of the data or their removal from the dataset must be explained.

In growth chart construction the most useful initial screen is producing a scatterplot of time on the x-axis and measurement on the y-axis (figure 6-1). Using this procedure, researchers can investigate any variables clearly outside the cluster on the plot for error. Frequently, the dataset must be divided and scatterplots for age segments must be examined separately. In this way gross outliers can be differentiated from the cluster of main data. However, scatterplots do not describe the distribution of the data within individual age ranges.

Another graphical technique to focus on the general distribution is the box-plot. Box-plots are a non-parametric technique for sorting data and identifying the relationship of the distribution of data relative to the median. The boxplot identifies the median, 25<sup>th</sup> percentile, and 75<sup>th</sup> percentile (boundaries of the box), and uses the distance between 25<sup>th</sup> and 75<sup>th</sup> percentile to

define the distribution of the data (the interquartile range, or IQR). Any point further than 1.5 times the IQR below the 25<sup>th</sup> quartile or above the 75<sup>th</sup> quartile is marked as an outlier, and points further than three times the IQR outside the quartiles are marked as extreme outliers. The points that are furthest from the median without qualifying as outliers are marked by a “whisker,” or thin line with a serif. Since boxplots examine the distribution within an entire population, the population must be divided into age groups with smaller intervals for rapidly changing growth. This technique of binning is a good screening tool for identifying suspected outliers, but introduces substantial bias when used to analyze the distribution. Once age intervals are created, boxplots can be generated for the entire age range of the study simultaneously. For example, weight plotted against age groupings labeled by month of life demonstrate that relatively few outliers exist in this dataset below one year of age (figure 6-2). After one year of age, many outliers exist above the distribution, but almost none below it. The only possible conclusions are that many errors of increased weight occurred, or the distribution is positively skewed, with many extremely heavy individuals relative to the rest of the population. To determine which, the researcher must take advantage of the longitudinal component of the data by examining individual plots. In purely cross-sectional studies this step cannot be performed, and significant errors can go unnoticed.

Individual plots can be created for every participant, or only for those whose distribution falls outside the boxplot distribution. Since significant errors can occur within the distribution, the conservative approach is to examine scatterplots of each participant’s data. These can be created in groups of 10-20 for screening purposes, and clearly demonstrate if values are abnormal. The following example represents 15 simultaneous scatterplots of height in RTT, and several individuals show improbable trends of decreasing and increasing height (figure 6-3).

Criteria can be established that are within the acceptable limits for measurement precision for the study population, and any value falling outside an otherwise smoothly changing curve can be examined for errors. Notably, because of population variability only one of these values would have been detected as an outlier by the boxplot technique.

After completing this initial screening of the dataset, the data need to be examined in more detail to identify trends and distributions within various anthropometric measurements. These trends both inform the researcher about possible hypotheses that can be tested and determine what types of adjustments will need to be made if a standard distribution is to be imposed on the data. One example of hypothesis generation includes changes in growth velocity at unexpected times which might suggest a change in the underlying pathophysiology of the disease in the population. Another example is a gradually increasing skewness or even a bimodal distribution, suggesting that certain clusters with differing growth exist within the population.

The first components of EDA were completed in the initial data cleaning process. Boxplots revealed important summary information about the population distribution, including how tightly measurements cluster about the median, and the extremity of variation away from the median in both directions. Scatterplots demonstrated less clearly what percentages of data cluster around the center, but gave a much more explicit sense of how the data are distributed in multiple directions. One example clearly shows that data cluster around specific ages with decreased density in between (figure 6-4). The explanation in this case is obvious – more children visit their physician at standard age intervals than other times. Moreover, this technique can reveal previously unsuspected trends, as seen in figure 6-1, which demonstrates subpopulations and relative height.

Another graphical technique that indicates how closely the data approximates the normal distribution is the histogram. Because this technique allows the researcher to compare empirical data to a distributional assumption, it has statistical testing implications. The arrangement of frequencies along a continuous x-axis gives an immediate impression of how well the data suggest a set distribution. This example of birth weight includes an overlying Gaussian distribution to graphically test how normal the data are (figure 6-5). In this case, the data exhibit noticeable negative skewness, explained in this dataset by the inclusion of six premature babies.

The distribution seen in the histogram can be described based on a distributional summary or tested statistically. The most common descriptives are the mean, median, variance, SD, skewness, and kurtosis. If the two measures of central tendency (mean and median) are noticeably different, this suggests the data are not normally distributed. The variance and its square root, the SD, are measures of the average difference from the mean in the population. Skewness is a measure of which side of the distribution (right/positive or left/negative) stretches further from the mean, and kurtosis is a measure of how much the data cluster about the mean (leptokurtosis) or cluster in the tails of the distribution (platykurtosis). An absolute value of skewness greater than one is likely to be significantly non-normal, and an absolute kurtosis greater than three is also likely to be non-normal. All of these “moments,” mean, SD, skewness and kurtosis are important components of the population. Because estimating each of these at each age manually is impractical, summary statistics can be generated to do so, and are described in the section on goodness-of-fit testing.

Further testing at discrete ages can be performed using the Shapiro-Wilk test for samples of 3 to 2000, and the Kolmogorov-Smirnov test for samples greater than 2000. These tests use a significance of  $p < .05$  to compare the data distribution to the normal distribution, and are

“positive” if the data is significantly different. Plots comparing observed values to expected values (were the data normally distributed) can reveal subtle differences in the distribution not detected by these two tests of normality. Q-Q plots are useful to detect differences in the tails of the distribution, and probability-probability (P-P) plots are more sensitive to differences close to the mean of the distribution. These are discussed below as diagnostic tests for the continuous distribution over all ages, but can be used for discrete ages, such as birth measurements. In this example, the birth weights from the histogram above were not significantly different from normal using the Shapiro-Wilk test ( $p=.10$ ), however the skewness seen in the histogram (due to the premature participants) was clearly present in the Q-Q plot as several points in the lower left of the chart rising above the line of normal (figure 6-6). This disparity in test results again highlights the importance of examining the data from many perspectives.

### **Techniques for Estimating and Smoothing Percentiles**

To generate a smoothed percentile, a mathematical function must be able to smooth the reference distribution in two directions simultaneously – between ages, and within ages (van Buuren and Fredriks 2001). The act of smoothing percentiles, removing the inherent noise present in any data set, by its very nature introduces a degree of artifice to the data. The researcher’s goal is to balance the aesthetic and harmonious ends of smoothing against fidelity to the empirical values.

The growth chart designer has many choices to confront. Should the data be plotted out and curves fit by eye, or should a mathematical function be used to fit the curves? If a function is used, which type of function is best? Should the functions for all percentile curves change simultaneously, or should they be independent? Should the data be molded to suit a preconceived distribution, or should the mathematical function bend to accommodate the idiosyncrasies of the data? If a distribution is assumed, should the data be transformed, or does the researcher assume

that the data are naturally similar to the distribution? Should outliers beyond a certain degree be discarded from the dataset as “physiologically unlikely” individuals? Which percentiles should be used for extreme curves and subdivisions? If data are sparse at the extremes of age (birth and the oldest age of the reference), should the dataset be extended beyond the desired age, and should additional individuals be sought to fill in those sparse areas? If data are missing for certain age groups, should the data be lumped into large bins, discrete ages where enough individuals are present, and smoothed based on those fixed intervals?

### **Eye Smoothing Versus Mathematical Functions**

Until relatively recently, the most common technique for smoothing percentile curves was the highly subjective process of visually approximating the maximum number of data points within the curve of a metal spline. Cubic splines and kernel smoothing functions have replaced mechanical eye-smoothing, however these mathematical functions are not without subjectivity. All smoothing procedures are a balancing act, removing the noise of the actual values in favor of aesthetic shape. The benefit of mathematical functions is that it is somewhat easier to quantify the degree of fidelity to the empirical data, or goodness-of-fit.

### **Constrained Versus Independent Percentiles**

“Growth charts are like fishing nets,” is the clearest analogy for the complicated issue of constraint (Cole 1993). Few would dispute the concept that the horizontal distribution of age should be evenly spaced on a reference chart, although some unusual distributions do exist in which months of infancy are spaced out wider than years of childhood. However, the concept that percentiles should be evenly spaced throughout age is difficult to justify, and mathematically even more difficult to achieve. Most of the functions above perform well when the task is generating a moving average constrained by a continuously changing age. However, when each of seven or nine percentile curves are expected to change relative to one another, most

approaches only succeed if the data are perfectly normally distributed throughout. As discussed above, weight and height do not follow the normal distribution in most populations, and in many rare diseases skewness is even more prominent. Notably, if this constraint is not enforced and the distribution is not normal, percentiles will weave independently, sometimes touching or crossing, and rarely achieving the desired symmetry. When a function successfully constrains the curves in both directions, tugging down on the 2<sup>nd</sup> percentile of the fishing net will proportionally drag the 10<sup>th</sup> down, not only at that point in age, but in younger and older children as well.

### **Choice of Mathematical Function**

Smoothing is a balance – an exchange of fidelity to empirical data for aesthetic and practical utility. Each potential smoothing formula presents a limited number of choices, known as smoothing parameters, for the researcher to manipulate. One combination of smoothing parameters, “under-smoothing,” will result in jagged peaks and valleys with improbable distributions, unlikely to represent physiologic changes in growth (figure 6-7). Another choice, “over-smoothing,” will present smooth, gradually changing curves which are just as likely to ignore important physiologically relevant shifts in the data (figure 6-8). The examples illustrate these extremes by presenting the data as they would plot out on these curves. The two curves are generated with identical data and different smoothing parameters. The first set of curves shift with every perturbation in the data distribution. Since much of the data are cross-sectional, no dependent distribution exists along the time axis. The fluctuations are not representative of any physiologic trend common to any one individual. The second example presents a graceful set of curves within which many gaps exist. While the first example “fits” the data better and the second fits very poorly, the third set of curves strikes a balance between fit and logical change in growth (figure 6-9). Therefore, to fit curves well the researcher must not only understand the general distribution of growth in the population, but must also be able to extrapolate based on the

underlying physiology of the population to determine whether shifts represent true changes in growth velocity or merely sampling artifacts.

Each type of function comes with benefits as well as concessions. Parametric models provide many options for goodness-of-fit testing, allow easy calculation of SD scores, and can be used in regressions with other covariates. However, parametric models rarely fit the nuances of growth data well, as they tend to create asymptotes at the extremes of age that don't truly exist, and may ignore physiologically relevant trends. Nonparametric and semiparametric models tend to follow the data much more closely, but do not result in a formula that can be used to estimate percentiles or Z-scores of individuals; instead percentiles and Z-scores must be presented in tabular form.

Polynomials are attractive in their simplicity, but they tend to be suited to only small segments of growth data, beyond which their curves tend to be wild and unpredictable. Exponential curves, especially those incorporating log or natural log components, tend to fit the smoothly changing characteristics of growth well. Fractional polynomials are an extension of simple exponential curves, in which the model incorporates several powers of age, such as log, square root, etc. Although some models seem to offer tremendous flexibility, no model is perfect for all situations. The measurement being modeled, the age range, and the patterns of growth in the disease all affect what model will work best to fit the data. Therefore, growth chart designers searching for models to fit rare diseases may not be satisfied with models used for healthy children.

### **Data Distribution**

If the researcher does not wish to assume that the data fit a known distribution, each percentile must be estimated individually and curves must be fitted to these independently.

Although there are ways to constrain the multiple equations so that curves follow similar paths,

this method lacks a distribution to estimate extreme percentiles. Z-scores cannot be used since they require the normal distribution to be present, and so precision in the extreme ranges of the chart suffers.

Prior to the development of robust statistical techniques for simultaneously analyzing measurements and continuous ages, researchers applied a minimum sample size requirement of five individuals per age interval to the construction of growth charts (Butler and Meaney 1991). Another common criterion was that SD should be less than 25% of the interval range. Since these guidelines were developed a radical change has occurred in the ability to estimate curves using non-parametric or semi-parametric models to fit age continuously. These requirements were developed prior to statistical techniques which simultaneously transform data along a continuous variable and generate estimates of the distribution. Researchers at that time relied on these rank order systems to estimate percentiles at fixed ages. However, with only five individuals per age interval, the distribution will never approximate normal, and the resulting curves cannot be relied upon. Ultimately, because the number of individuals in the “tails” of any sample is so small, the only way to accurately estimate the outer percentiles in a growth reference is to rely on an underlying distribution.

### **Transformations**

In cases where non-normal data appear to be clearly skewed, but not bimodal or haphazard, transforming the data by performing a mathematical operation on all variables may bend the distribution into a normal shape. Several mathematical operations exist to reframe the distribution, and the choice of which one to use is somewhat random. Taking the natural logarithm of each datapoint is often successful in strongly skewed data, and taking the square root of each is useful in less skewed datasets. If the data are skewed in an exponential manner, the reciprocal can be used to adjust the distribution.

The histogram in figure 13A shows weight of 10 year old females with RTT, and the one in figure 13B is the natural log of weight in the same individuals (figure 6-10). Empirical observation and all estimates of normality show that the transformed weight is much closer to the normal distribution. The Shapiro-Wilk test showed a significant difference from normal ( $p < .001$ ) in the original weight, and no significant difference ( $p = .53$ ) after transformation. This difference is also evident in the Q-Q plots as a decrease in the S-shaped variation from the normal line in figure 6-10B compared to figure 6-10A (figure 6-11).

With data cleaning and transformation researchers can adjust data to accurately fit preconceived distributions. In 1968, Meredith claimed that there is no difference between mean and median standing height, essentially stating that height follows a normal distribution, at least with respect to skewness (Meredith 1968). Numerous examples exist in the literature to refute this claim, however many growth chart designers have accepted his perspective as dogma and have dismissed concerns about variation from the normal distribution in their data. Rare diseases are more likely to be subject to unusual distributions, both based on unusual patterns of growth within the disease and on sampling bias. Therefore, a strong methodology for transforming data is necessary.

However, since growth data for height are right-skewed after puberty, a log transformation such as in the example should be able to approximate the normal distribution. Unfortunately, a simple transformation of the measurements fails to account for the fact that data at other ages may not be skewed or may even be skewed in the opposite direction. Nor would the log-distribution necessarily be the best choice to adjust the distribution. Instead of applying one rigid transformation to the entire dataset, Cole suggested applying the Box-Cox power transformation,

a flexible power transformation that accounts for different degrees of skewness at different ages, and is able to adjust for a wide variety of skew distributions (Cole and Green 1992).

The LMS method assumes that the reference data are distributed based on the Gaussian distribution after a Box-Cox transformation. The model is flexible enough to estimate the mean and SD while simultaneously transforming the distribution at that age using a Box-Cox power transformation. Moreover, using penalized likelihood, a maximum log-likelihood approximation technique, the model can simultaneously estimate curves for three parameters along a continuous age. The three parameters it estimates are the median (M), the coefficient of variation (S), and the power for the Box-Cox transformation of the measurement values (L). The user input for smoothness parameters determine the flexibility of each portion of the model, and are represented as estimated degrees of freedom, or EDFs. Therefore, not only do the mean and SD change continuously with age in this model, but the L parameter also allows the transformation for skewness to change with age. Moreover, the method of log-likelihood estimation that Cole chose (maximum penalized likelihood) constrains all three curves (mean, coefficient of variation, and skewness) to change simultaneously and smoothly. Although this method does not adjust for kurtosis in the transformation, kurtosis is generally mild in typical measurements such as height, weight, and FOC, but can be a significant issue in skinfold measurements. Additional models, such as the Box-Cox power exponential used in the LMSP method, do adjust for kurtosis, but at a cost of fitting the empirical data. In rare diseases with marginally sized datasets, the tradeoff in fidelity is probably not worth the extra fitting parameter of kurtosis.

### **Choice of Percentiles**

Historically, the choice of percentiles has been somewhat arbitrary. Many early researchers suggested that the small sizes of their datasets precluded estimation of extreme percentiles. Therefore, many used the 5<sup>th</sup> and 95<sup>th</sup> or even the 10<sup>th</sup> and 90<sup>th</sup> as extreme values.

When better techniques for estimation were developed and larger sample sizes were collected, practical decisions were made regarding the meaning of the percentiles. The extreme values are considered screening cutoffs – anyone above or below the extremes is at risk for disease. Therefore, these values can vary based on the measure being expressed. Since a percentile reflects the percentage of the population found below that value, the CDC chose the relatively low 85% cutoff for BMI as indicating risk for obesity, since at least 15% of the population in the US are obese. Alternately, organic causes of short stature are rare, and so the CDC chose 5% in 1977 and 3% in 2000 as cutoffs for risk of short stature. Based on the nature of percentiles, 3% of the population will fall below this curve. Of this 3%, only a small number have a medical cause for their short stature. Ultimately, the purpose of the screening cutoffs is to alert a physician to perform a diagnostic workup or to refer a child to a specialist. Therefore, the cutoffs dictate the percentage of the population who will receive a diagnostic evaluation, and therefore are an important factor in distributing medical resources. In rare diseases, the percentage of comorbid conditions and the risk for associated findings varies with the condition. For example, in Down syndrome 15-20 percent have hypothyroidism, a cause of short stature, and therefore the lower cutoff in this syndrome should probably be higher. In RTT, FTT is much more common an issue than overweight, and so the lower BMI cutoff should be higher than the third percentile. Conversely, in diseases where a disorder of growth in a particular parameter is rare, a higher percentile can be chosen; the 99.6 percentile selects only 4 per 1000 as at risk, and is conveniently two and two-thirds SD from the mean.

## **Outliers**

Outliers can be assessed in a number of ways. The serial boxplot technique illustrated above estimates outliers in large age intervals. Once individual outliers in these groups have been addressed prior to model fitting, outliers in the transformed model can be identified using a Z-

score plot of the distribution over age. One of the benefits of the LMS technique is that it allows calculation of Z-scores directly from the following formula if the L, M, and S values are known at that age.

$$Z_{score} = \frac{(Measurement / M)^L - 1}{LS} \quad (6-1)$$

Equation 6-1. Formula for calculating Z-scores for a particular measurement using L, M, and S.

### **EDF parameters**

The LMSChartMaker program provides the ability to select datapoints on the Z-score plot and identify their age, measurement, and ID number so that the researcher can investigate the nature of the extreme value.

### **Edge Effects**

Edge effects occur when the model has too few empirical values at the edge of the distribution relative to the preceding ages, and so extrapolates to non-physiological asymptotes. Right edge effect generally results in extreme percentiles deviating vertically rather than assuming the expected horizontal asymptote of adulthood. This example comes from a sparse set of data, in which the curve predicts that 98<sup>th</sup> percentile for weight will increase from 50 kg at 10 years to 150 kg at 25 years (figure 6-12). Left edge effect is less common but may result in a flaring at birth rather than a condensation to the tight distribution normally seen at birth. Each edge effect is best dealt with in two different manners. For right edge effect, expanding the sample size to an age 20 to 25% greater than the desired age usually allows the desired curves to follow a physiologic trend. For example, the WHO reference collected data on children up to age six, but only published charts up to age five. Left edge effect requires a different strategy, since no individuals can be measured younger than birth. Therefore, oversampling is the best way to approximate a better left edge, by improving the strength of the distribution at birth. In general,

two to four times the number of individuals seen at other times during infancy should be measured at birth. The WHO dealt with this by using birth data on individuals who they excluded based on post-natal risk factors (non-breast feeding, maternal smoking, etc.). The CDC, on the other hand, resorted to a random sampling of birth records, collecting 40 million birth-weight records and 400,000 birth-length records to supplement an average sample of 100-400 measurements at other intervals.

### **Age as an Interval or a Continuous Variable**

With easy availability of the statistical techniques described above, few researchers opt to bin measurements into predefined ages for analysis. Functions which estimate distributions over age as a continuous variable are much more elegant, and eliminate the operator bias of binning the ages into user-determined groups. The WHO has stipulated a number of prerequisites for any high-quality study of growth, and this is one of the strongest recommendations for growth models to conform to.

### **Summary**

The WHO reviewed over thirty methodologies for creating growth references, and developed eight criteria they feel every growth reference should meet. The first five are absolute criteria, and the last three are optional criteria (table 6-1). The origins of these criteria can be traced back through the history of growth chart design. Failure to observe one or more of these criteria has marked significant flaws in most of the models of growth used over the past century. The WHO identified three of the thirty models as potential candidates to use in their study, and they ultimately chose the LMST model (Borghini et al. 2006). The LMST model, which does account for kurtosis, reduces to the LMS model if no kurtosis adjustment is necessary. Notably, in all of the growth measures the CDC examined, the adjustment for kurtosis was omitted and the model reduced to Cole's LMS approach.

## **Diagnostic Guides to Balance Smoothing and Fidelity to Data**

No matter how aesthetically pleasing a set of reference percentiles, if the mathematical functions do not fit the empirical data, then the curves they produce are useless. Unfortunately, at this point in considering the science behind growth chart methodology it becomes clear why Cole designated the process of percentile curve construction as a “black art.” Surprisingly few techniques exist for evaluating the goodness-of-fit of the complex techniques for curve fitting to the empirical data. Two overall categories of quantitative techniques are available: inferential tests and graphical analysis (Royston and Wright 2000). This section will begin with descriptive or qualitative assessments and progress to quantitative measures.

### **Visual Inspection**

Visual inspection has been described above, and should be one of the first steps in assessing the fit of curves to the data. Clearly, in the examples of over- and under-smoothing, visual inspection is adequate to determine that the model has failed to fit the data.

### **Overall Shape of Curves**

Likewise, curves that appear too simplistic or too convoluted are likely to be poor fits to the data. Although it is possible to smooth out a physiologically significant bump in the curves in the interests of appearance, this can be avoided by considering the nature of the underlying distribution. If longitudinal data are available, the shape of the longitudinal curves can be examined independently to determine if the shift represents a trend in multiple individuals or is simply an artifact of sampling bias.

### **Data Points With Percentile Curves Superimposed**

Plotting the data points directly over the smoothed percentiles allows the researcher to detect outliers, and to see gaps in the data where few empirical values support the curves. This

technique, along with the Z-score plots with superimposed data, serve to evaluate the density of the data distribution.

### **Empirical Percentiles and Fitted Percentiles Superimposed**

Once the curves have been generated, researchers can artificially choose age intervals and percentage estimates to compare with the transformed, normalized values. If the transformation has perfectly normalized the data, then -2 SD will line up with the 2<sup>nd</sup> percentile, -1.33 with the 10<sup>th</sup>, -0.66 with the 25<sup>th</sup>, etc. An easy comparison to begin this investigation is to compare transformed means with the median at a certain age. If the two do not match, then skewness is still present in the distribution.

### **Comparison of Observed and Expected Values**

Extrapolating on the concept of percentile-to-Z-score comparisons, the researcher can test the distribution using a chi-square test of the expected and observed values which fall within certain percentile ranges. However, the estimation of percentiles at various ages introduces selection bias, and the choice of ranges for the chi-square test increases this bias and is not mathematically justified. Other statistical tests offer less biased approaches.

### **Tests of Normality for the Z-scores**

When Z-scores are available (in cases where the methodology assumes a fixed distribution), the researcher can test the empirical data against the null hypothesis that these data fall inside the normal distribution within a set confidence interval (90 or 95%). Several statistical tests exist to compare the values of the smoothed reference curves to the normal distribution. The researcher can use either the Shapiro-Wilk test or the Kolmogorov-Smirnov test, depending on the sample size and the moment of the distribution under scrutiny (i.e., skewness or kurtosis). However, failure to reject the null hypothesis in these tests does not adequately validate the model. In that respect, these tests are difficult to interpret.

The Q-statistic incorporates modified D'Agostino and Shapiro-Wilk tests to produce a composite test of normality (Royston and Wright 2000). This test gives overall information about the normality of the fitted model with respect to four moments of the normal distribution, mean, coefficient of variation, skewness, and kurtosis. However, just as with any test of significance, the Q-statistic does not indicate the magnitude of the discrepancy between the empirical data and the fitted curves. When tests of significance on very large sets of data are positive, this statistical significance may not equate to clinical or practical significance. Consequently, Q-statistics should be used along with other diagnostics that interpret magnitude.

### **Quantile-Quantile plot of Z-scores**

Quantile-Quantile plots compare the data to a normal distribution. The x-axis represents the normal distribution in terms of SDs. The y-axis represents the Z-scores of the empirical data. Q-Q plots follow a diagonal line from the bottom left to the top right of the graph. If the data are perfectly normally distributed, they will track along this diagonal line flawlessly. If skewness or kurtosis are present they will wind around the line like a snake or deviate above or below it.

### **Worm Plots**

The detrended version of the Q-Q plot, in which the y-axis represents the *deviation* of the empirical data from the normal distribution, is a more sensitive indicator of skewness and kurtosis. The 95% confidence interval of normal is represented as two opposing parabolic curves above and below the horizon. In the detrended Q-Q plot perfectly normal data cling to the horizontal axis. A collection of detrended Q-Q plots for various ages is called a worm plot. Since the morphology of the worm formed by the data represents specific deviations from the normal distribution, the worm plot is an excellent diagnostic tool for modeling growth charts.

Each morphologic interpretation of the “worm” on the plot refers to how well the data conform to one of the four statistical moments in the fitted model – mean, variance, skewness,

and kurtosis. If the fitted mean is smaller than the mean of the empirical data, the worm raises above the horizon (origin); if the fitted mean is too large, the worm sinks below. If the fitted variance is too small, the worm slopes up to the top right; if the variance is too large the worm points down to the bottom right. If the fitted distribution is skewed to the left, the worm curves up into a U-shaped parabola; if the distribution skews to the right, the worm curls down into an inverted U. This example of highly skewed data is taken from the “oversmoothing” data set (figure 6-13). Lastly, if the tails of the fitted distribution are too light (platykurtosis), the worm forms an S-curve with left side down; if the tails are too heavy (leptokurtosis), the left side of the S-curve points up. The strength of the worm plot is that it is able to give a graphical representation of age-related fidelity, whereas the Q-statistic provides only a global test of appropriateness with no indication of magnitude of fit. Pan and Cole recommend using the worm plot as an initial tool for building the model, and then using the Q-test to refine the adjustments of the estimated degrees of freedom (Pan and Cole 2004).

### **Techniques for Analyzing and Summarizing Longitudinal Growth Data**

One of the primary purposes for creating specialized growth charts is to determine the effect of interventions. Although disease-modifying interventions are not available for many of the diseases discussed in chapter 3, treatments will become available in the near future. To monitor the effects of these treatments, syndrome-specific, perhaps genotype-specific longitudinal data must be available. The number of patients included in a placebo-controlled trial of a new treatment agent is usually too small to allow meaningful comparisons of growth parameters. Nevertheless, disease-specific growth references can act as a historical control for both the treatment and the placebo arms of efficacy trials. However, to produce meaningful results, the best historical control is a longitudinal cohort.

## **Correlations**

The simplest type of correlation is the autoregression, where data are compared to the previous value and adjusted accordingly. This concept was discussed above under the concept of the “conditional” growth standard. Researchers have begun to examine more complicated models of correlation that account for within-subject correlations using various correlation structures and various mathematical models for the mean and SD. In 1998, Wade et al., examined CD-4 counts in infants exposed to HIV perinatally, and although they found a very strong within-subject correlation structure, the incorporation of correlation in the model did not have a significant effect on the curves (Wade and Ades 1998). However, when longitudinal data are collected correlation could prove a significant tool in examining differences among groups and modeling unusual growth patterns.

## **Regression Models**

Researchers have developed techniques for summarizing longitudinal data that are much more powerful than the summary statistics for cross-sectional data, especially when applied to a longitudinal comparison of outcomes. Regression formulas for predicted adult height can be derived from longitudinal growth data and other variables, such as mid-parental height, birth length, bone age and genotype. For example, rather than comparing the height of patients with Turner syndrome to a standard reference range, Lenko et al., developed and refined a formula to predict adult height based on longitudinal reference values called the index of potential height (IPH) (Lenko et al. 1979). Using this formula, Joss et al., were able to assess the effect of anabolic steroids on final height in Turner syndrome much more effectively than if they had compared Z-scores pre- and post-treatment to a disease-specific growth standard. This same technique can be applied to all longitudinal data sets, so that researchers will be able to judge not

only size, but also true growth response. Moreover, these judgments will be based on the height potential of the individual under study, rather than the height potential of the entire population.

Berkey et al., developed an approach which can be used to reliably predict the future height of a child using a regression equation. This tool could be used to conduct research in circumstances where a control population is not practical or ethical (Berkey et al. 1983).

Ultimately, the treatments themselves can be included in these mathematical models, allowing researchers to design individualized treatment regimens, adjust these regimens based on the patient's degree of responsiveness and predict individual outcomes.

Notably, growth velocity does not typically follow the percentiles on a size reference chart. For example, children who tend to follow the 3<sup>rd</sup> percentile on a height-for-age chart are actually growing slower than average. The average child plotted at the 3<sup>rd</sup> percentile on one visit would increase towards the mean by the next visit, in a phenomenon known as regression to the mean. Therefore, the child who parallels the 3<sup>rd</sup> percentile is growing in the 25<sup>th</sup> percentile based on height velocity (i.e., slower than average). Likewise, the child who parallels the 97<sup>th</sup> percentile is growing faster than average. Most children at the 97<sup>th</sup> percentile for height on one visit will be below it by the next. This phenomenon occurs because height charts are based on cross sectional data with no accounting for correlation among ages. However, even height charts based on longitudinal samples (the Fels chart from 0-3 years for example) do not incorporate correlation, and regression to the mean still holds true. To express velocity on a height chart requires a hybrid technique such as thrive lines discussed above.

### **Techniques for Statistical Comparison of Results**

Once a disease specific growth chart has been created, it may become useful to compare this reference to other growth. While visual inspection of the references is usually sufficient to determine where the patterns of growth differ, occasionally researchers will wish to know how

early or to what degree anthropometric values differ in two populations. Additionally, the creators of the reference may wish to compare growth among subpopulations within the reference, and statistical techniques for this type of comparison exist.

The most accessible method of comparison is to compare the mean values of the references at certain ages. If the researcher is interested in differences at discrete ages, for example standard well-child check-up intervals, he can use the model to compute the mean and SD. Using these values and the number of participants in each group he can perform student's t-test to determine if the reference mean and distribution are significantly different at those ages. Several researchers have used the t-test to compare growth in a special population with other special populations or a control population (Cronk et al. 1988; Cronk 1978; Kimura et al. 2003). Alternately, if a standard reference is used for comparison and the number of participants used to create it is not known, the researcher can use a one-sample t-test. The researcher will need to correct the test statistic based on the number of comparisons to control type one error, but in general the number can be minimized by using visual comparison and EDA to help generate directed research hypotheses. One useful adjustment is the false detection rate (FDR), a computation which reveals the percentage of incorrect assumptions likely to be made within a group of data based on a statistical test.

In addition to comparing mean values among references, the coefficient of variation (CV) offers a stable measure of dispersion in the population as it varies with age. The CV is simply the SD divided by the mean, and since the SD increases as the mean increases with age, the CV is much more stable across ages. The CV is also a unitless number, and so allows comparison of references in different units.

Comparisons of rate of change, or velocity, can be made in several ways. First, if a polynomial or other linear function was used to calculate the curves, then the first derivative of this function represents the growth velocity as it changes with age. The absolute values of this curve can also be compared using student's t-test. This technique is particularly useful if individual growth curves and growth velocity curves are estimated for each participant. Another technique is to compare the change in velocity, or acceleration by taking the second derivative of the curve and comparing values at various ages. Velocity is a more sensitive indicator of growth than size, and acceleration is more sensitive than velocity. Therefore the choice of which curve to compare depends on the trend of the curve at the ages of interest. Puberty is an excellent example. Several studies of populations with genetic factors affecting growth have noted that the pubertal increase in growth velocity is less profound than in normal children (Cronk et al. 1988; Roche 1965). Since most length-for-age charts only show a very gentle increase in the slope of the size curve at puberty, comparison using size is very difficult. Velocity curves at the same age, however, show a dramatic increase and decrease in velocity. If a researcher is examining changes related to puberty, velocity curves are an excellent tool.

Notably, certain adjustments may need to be made to models of velocity to be used for comparison, depending on the phenomenon under study. For example, in the case of pubertal growth velocity the effects can be unacceptably smoothed over if velocity is compared directly to age. This smoothing occurs because the pubertal growth spurt begins at a wide range of ages in the normal population. An adjustment based on peak velocity (lining up peak velocity irrespective of age), or by plotting growth-at-bone age (a marker of skeletal maturation that correlates with onset of puberty) can address the problem of over-smoothing important trends.

The next chapter will demonstrate the development of growth charts for RTT. Not all of the techniques described above were applicable to the goals of the RTT growth study, and so only a selection were used. However, they will likely prove useful in other studies of growth in rare diseases.

Table 6-1. WHO criteria for an ideal growth standard model.

Necessary Criteria	Desirable Criteria
Must estimate 3rd and 97th percentiles precisely	Should be able to assess goodness-of-fit to empirical data
Must constrain percentiles so that they cannot cross	Should be well documented and easy to apply
Must be able to estimate Z-scores using a formula	Should be flexible to different anthropometric measures (e.g., skinfolds)
Must treat age as a continuous variable	
Must account for skewness at all ages (should be able to adjust for kurtosis when present)	

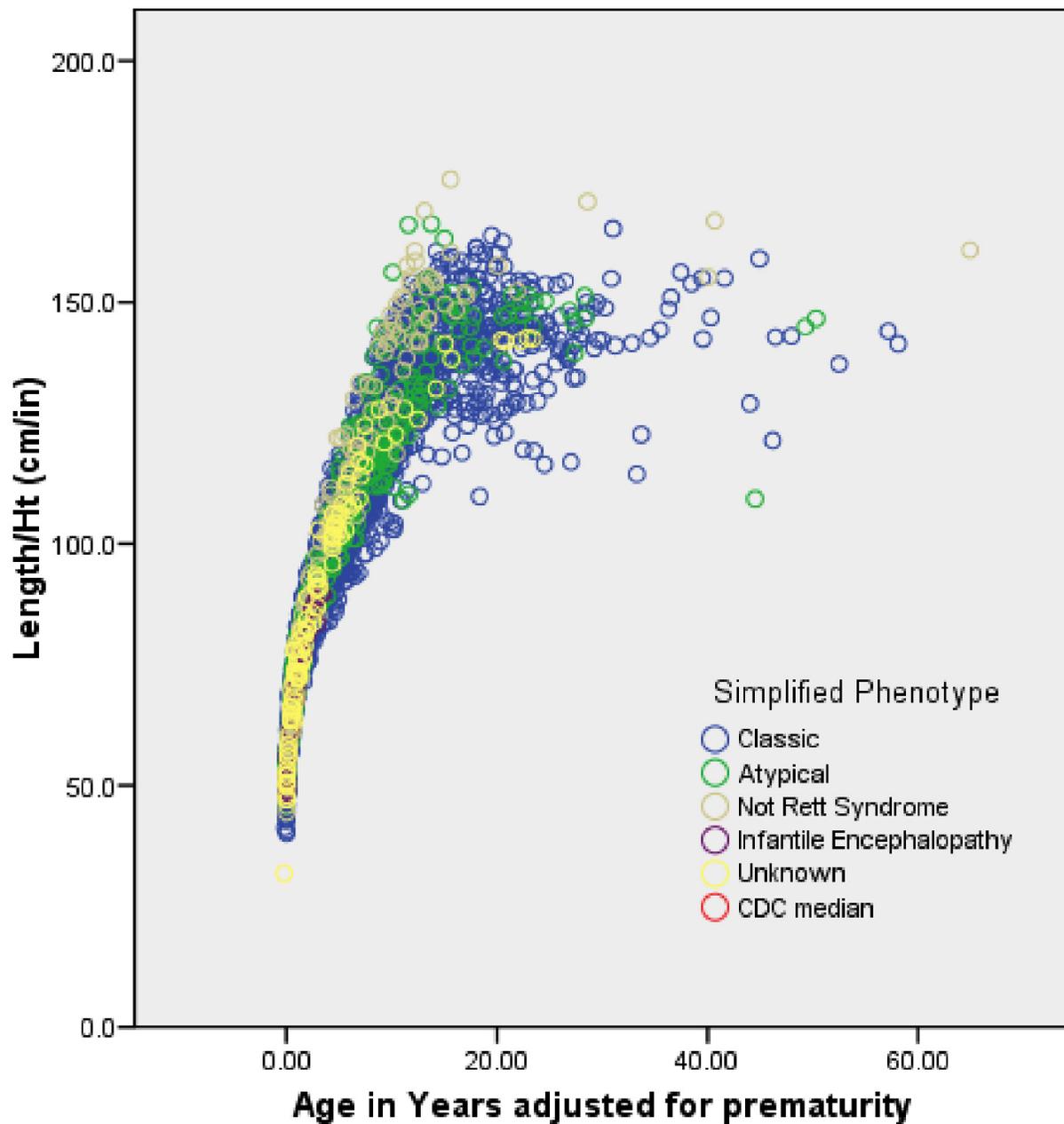


Figure 6-1. Example of Exploratory Data Analysis used to identify outliers and understand differences in dispersion among groups.

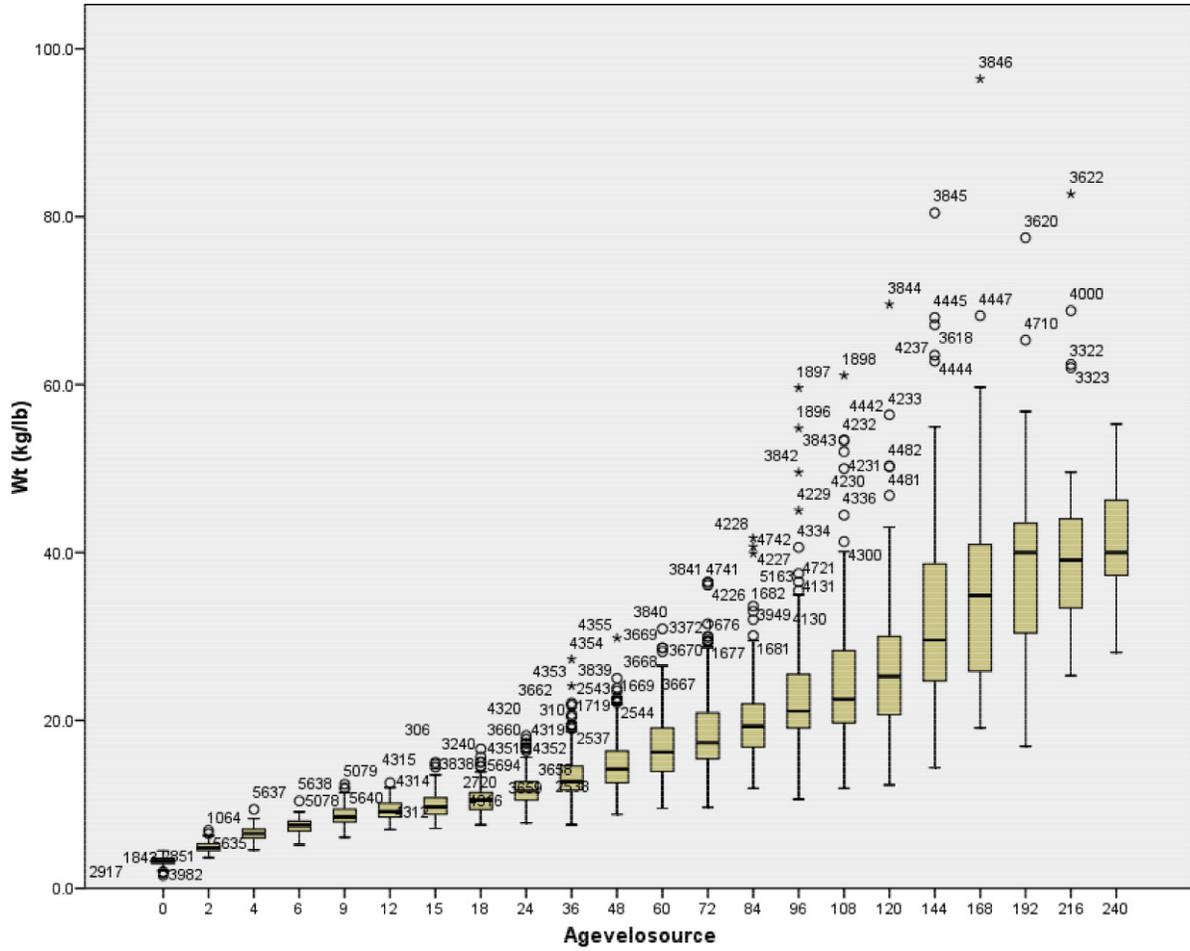


Figure 6-2. Boxplots of age groups from birth to adulthood, demonstrating the change in the overall shape of the distribution and the presence of outliers at certain ages.

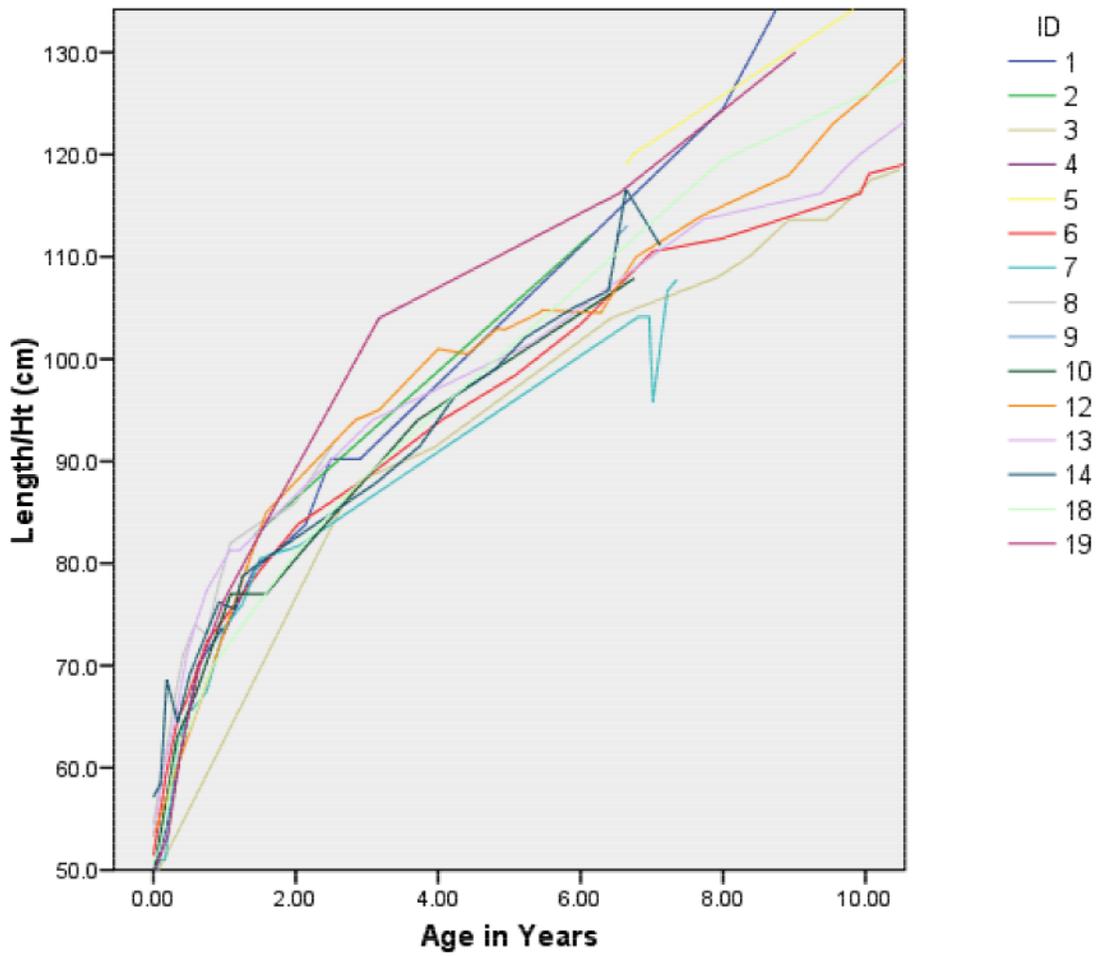


Figure 6-3. Height of 15 individuals plotted longitudinally, demonstrating aberrant measurements which do not follow the typical curve.



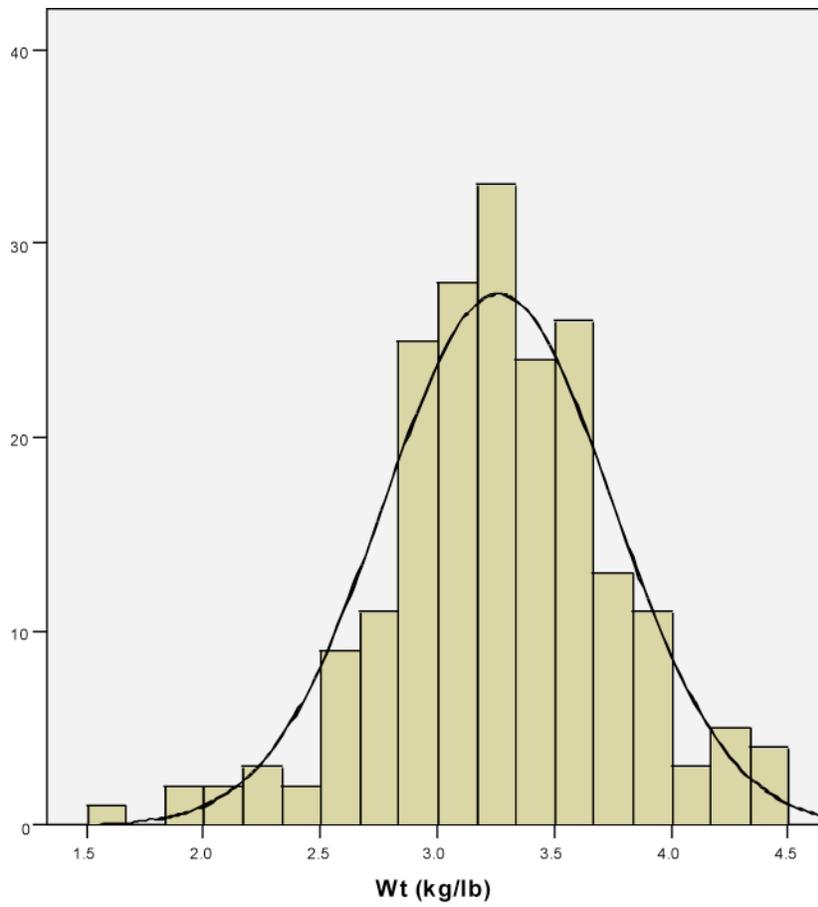


Figure 6-5. Birth weight histogram, demonstrating negative skewness.

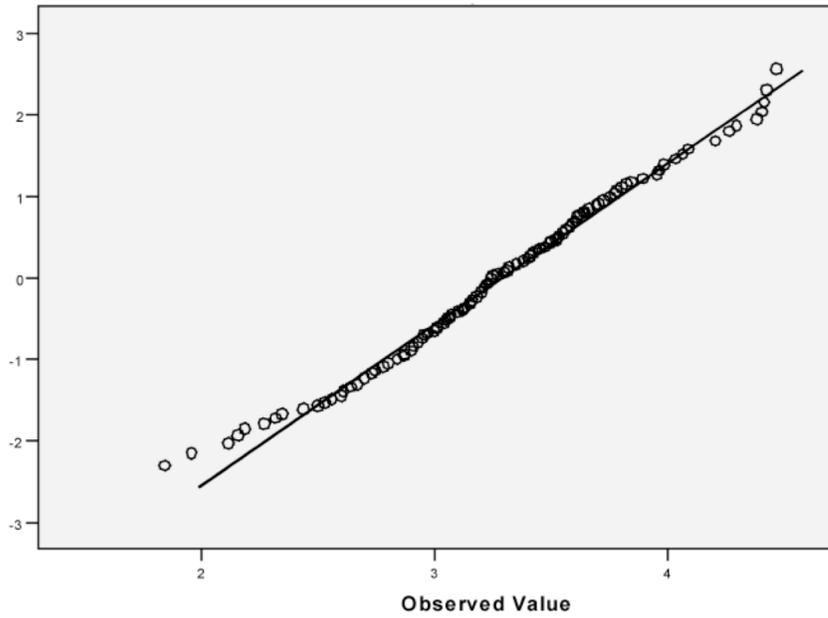


Figure 6-6. Normal Q-Q plot demonstrating skewness in distribution.

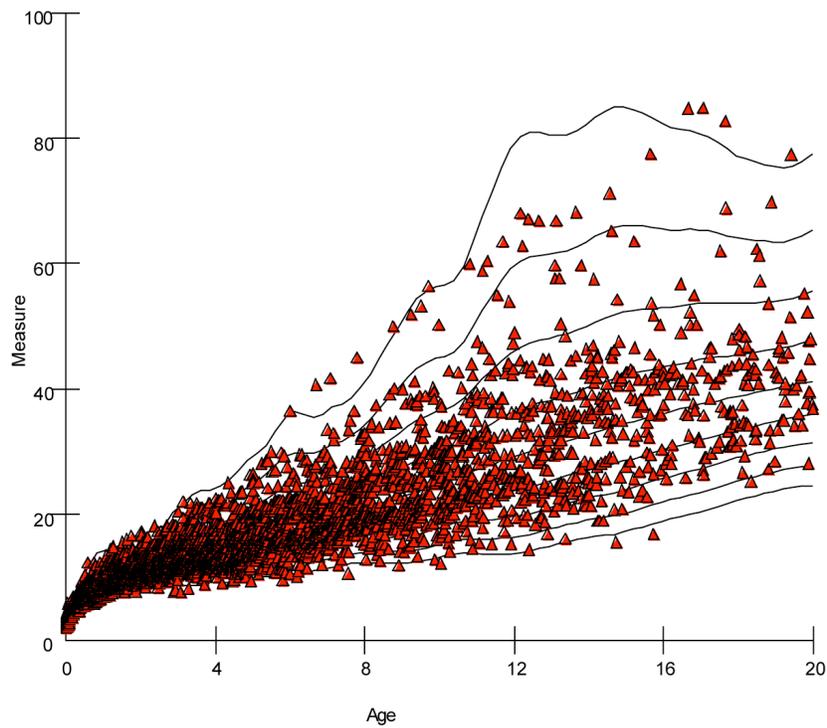


Figure 6-7. Curves undersmoothing data.

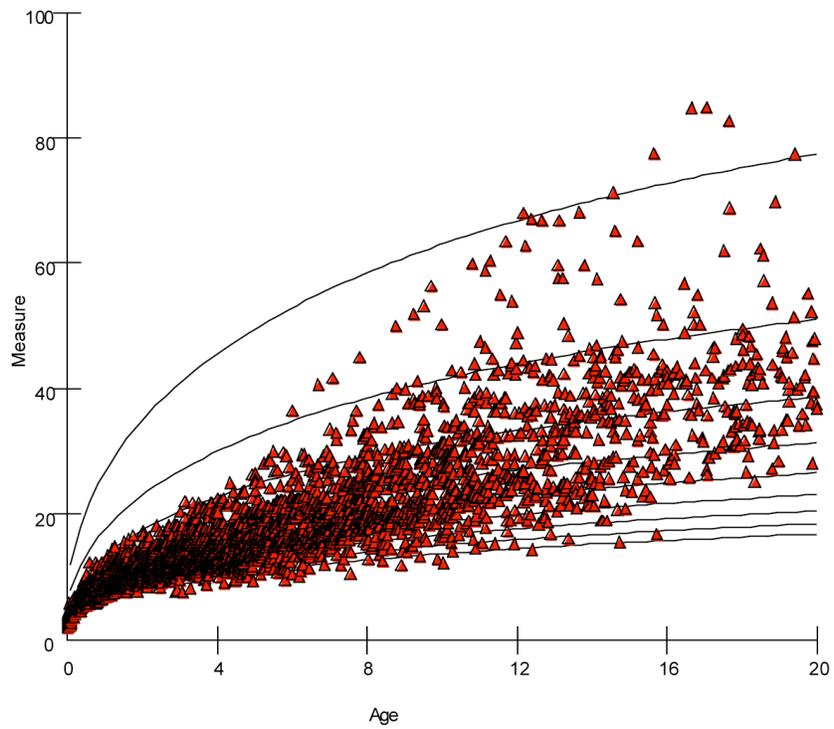


Figure 6-8. Curves oversmoothing data.

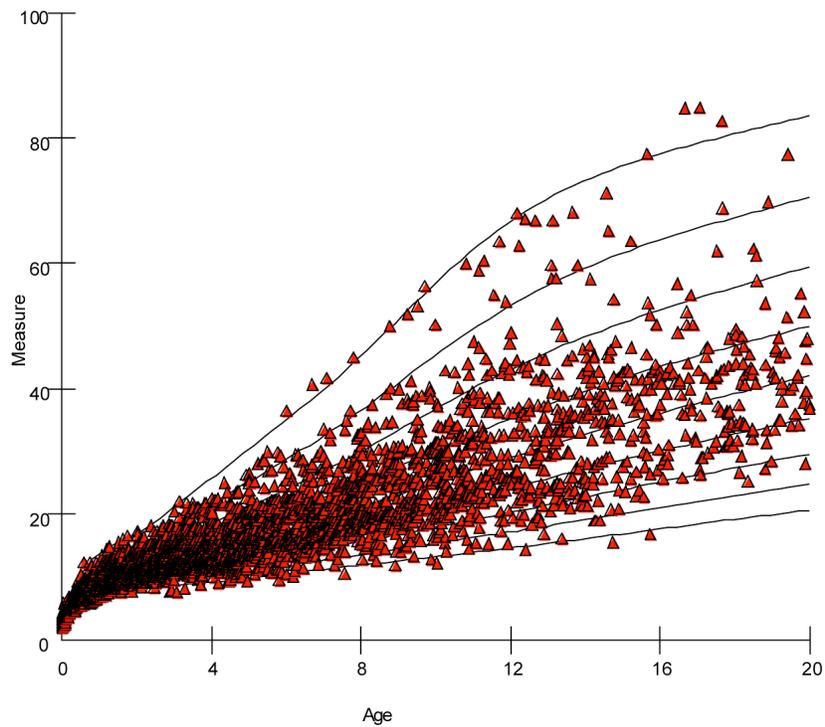


Figure 6-9. Curves appropriately smoothing data.

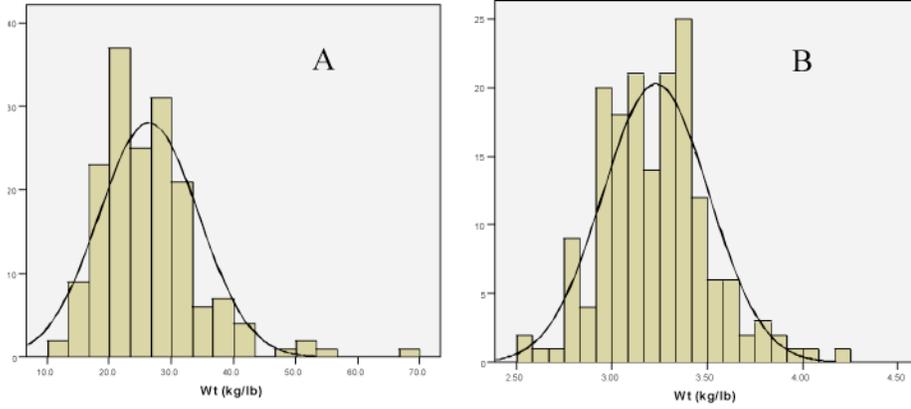


Figure 6-10. Weight transformed using log scale (ln).

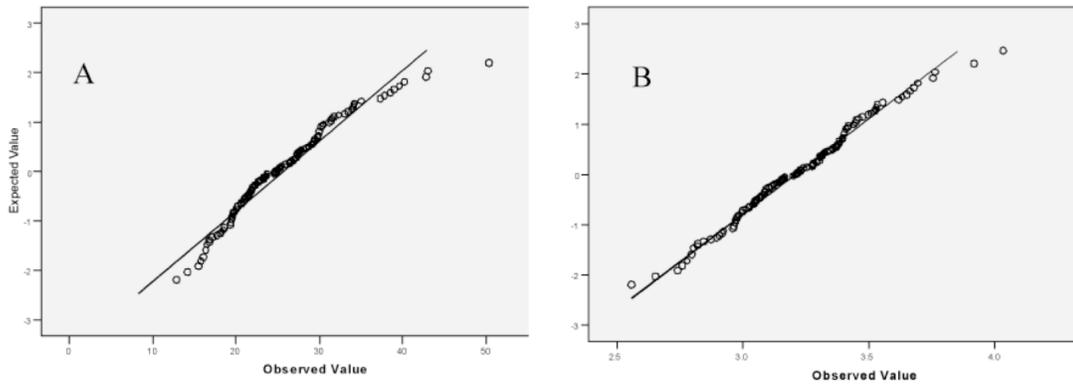


Figure 6-11. Normal Q-Q plot, demonstrating effect of log transformation (ln).

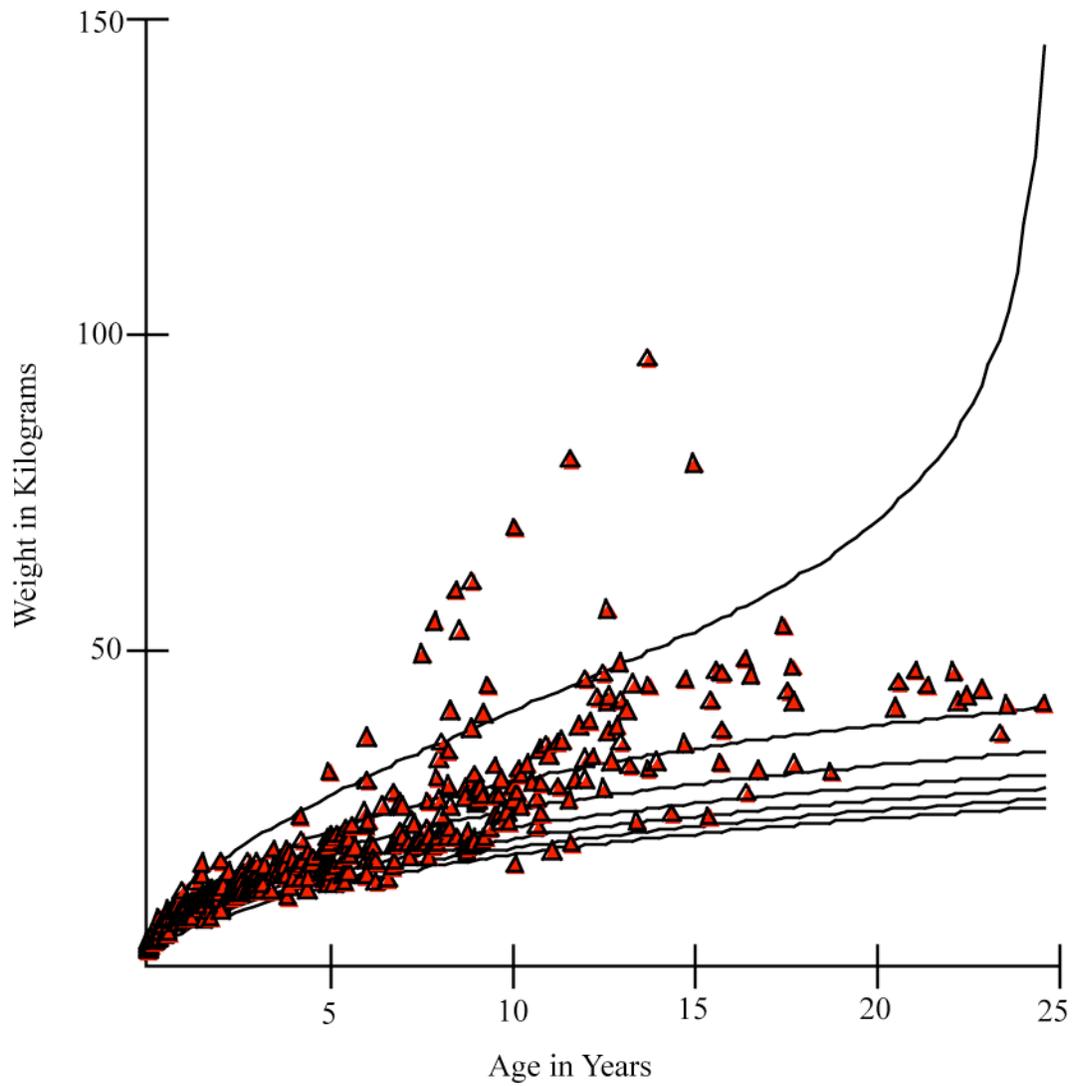


Figure 6-12. Weight-by-age chart demonstrating right edge effect in a sparse data set.

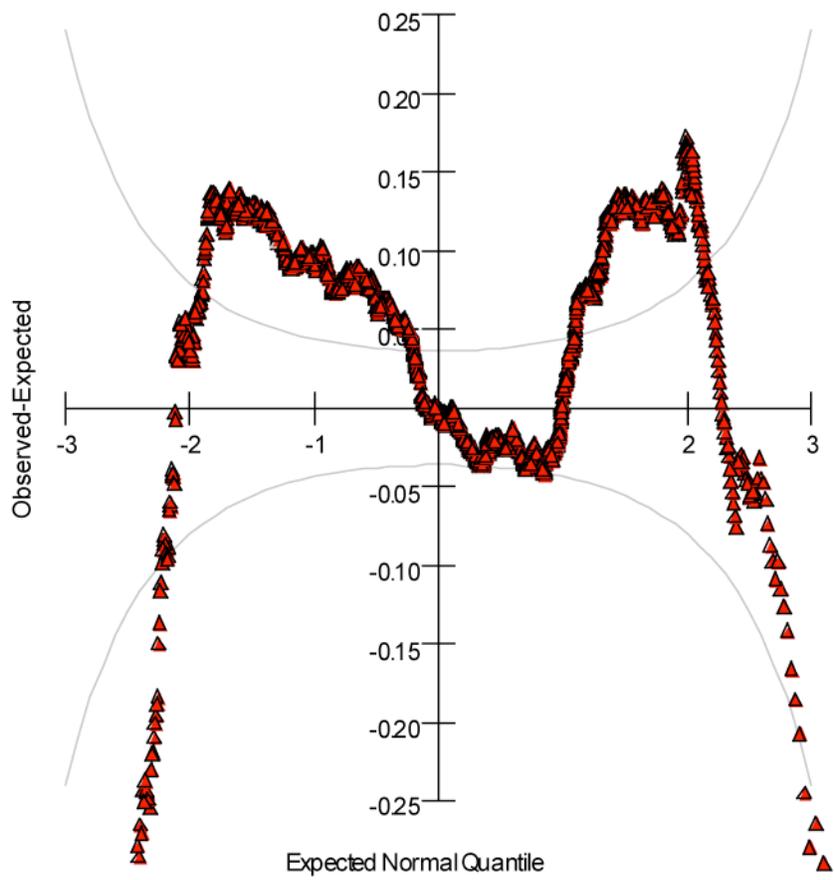


Figure 6-13. Detrended Q-Q plot (worm plot), showing significantly skewed data.

## CHAPTER 7 PATTERNS OF GROWTH IN RETT SYNDROME

### **Introduction**

Compared to other rare diseases, reference charts for growth in patients with RTT are in their infancy. Many of the disorders discussed in chapter three have an incidence similar to RTT. While numerous editions of growth charts exist for Down syndrome (1:1000 live births), Turner syndrome (1:2500 female births), and NF1 (1:3500 live births), a complete set of growth references for RTT has yet to be produced. Although the incidence of RTT is only 1:10,000 female births, growth charts exist for equally rare syndromes such as CDCS (1:50,000 live births), RTS (1:100,000 live births), and three editions of charts exist for PWS (1:15,000 live births). While many attempts at comprehensive growth charts have been made in these other syndromes, the quality of syndrome specific growth charts varies significantly in areas such as data assimilation, measurement reliability, statistical smoothing techniques, and accessibility. However, the means exist to produce growth references for RTT that meet the standards put forth by the CDC.

Previous research on growth in RTT has either addressed only one aspect of growth, such as FOC, or has analyzed data using binning and regression techniques which do not allow detailed comparison with other references. Moreover, no study has examined charts of BMI in RTT, despite the prominent concern for wasting in this population. Oddy et al., examined BMI in large age groups in RTT, but this only provided a sense of general trends. Additionally, many factors may play a role in the growth failure in RTT, and the contribution of each of these has not been examined in detail. Moreover, to compare groups within RTT requires RTT-specific references. Several studies have demonstrated that the majority of individuals with RTT after age 7-8 years have measurements below -3 SD compared to normal individuals (Hagberg et al. 2001;

Hagberg et al. 2000; Oddy et al. 2007; Schultz et al. 1993). As demonstrated in chapter two, comparisons beyond -3 SD are unjustified, as they are based on too few empirical data, and the further from the mean the less precise the comparisons are (Cole 1994). Thus, to compare groups within RTT in a statistically valid manner, RTT-specific references and Z-scores are needed.

This study's principal aim is to produce comprehensive specialized growth reference charts for RTT which are both robust to the CDC criticisms of specialized growth charts discussed in chapter four and amenable to use in future research. The four secondary aims are as follows: 1) to compare growth between individuals with RTT and the normal healthy population, 2) to compare growth among various subgroups within the Rett research population, 3) to provide reference values for growth velocity in RTT and compare these to normal growth velocity, and 4) to examine the influence of various RTT characteristics on growth.

Our primary hypothesis was that useful growth references could be created for girls RTT from birth to 18 years using combined longitudinal and cross-sectional data. The four secondary hypotheses are summarized below.

- Height, weight and FOC are decreased in RTT relative to the normal healthy reference
- Differences in growth exist among individuals within the following categories:
  - Diagnosis: classic, atypical, and non-RTT
  - Secular trend: individuals born/after before aggressive nutritional supplementation in RTT
  - Disease severity: mild vs severe
  - Mutation types: 11 specific mutations, 3 mutation types, and 5 mutation sites on *MECP2*
- Growth velocity in RTT is decreased relative to the normal healthy reference
- Early characteristics of RTT predict growth later in life, such as:
  - Ambulation
  - Scoliosis
  - Age at regression
  - Hand use ability
  - Language ability
  - Age at sitting
  - Seizures

To address these hypotheses, we used a mixed study design, including prospective and retrospective data, and analyzing these data with longitudinal and cross-sectional techniques. Prospective measurements were recorded from individuals at the time of study encounter, while retrospective data were collected on the same individuals to complete the longitudinal component.

## **Methods**

### **Participants**

The majority of participants were recruited as part of the Rare Diseases Clinical Research Network (RDCRN) study on the natural history of RTT. The RDCRN study is a multicenter longitudinal observational study of the natural history of classic RTT, atypical RTT, and individuals with *MECP2* mutations who do not fulfill criteria for classic or atypical RTT (non-RTT). Participants were recruited at one of three main sites (Baylor College of Medicine, Houston, TX; University of Alabama at Birmingham, Birmingham, AL; Greenwood Genetic Center, Greenwood, SC), or four remote sites (New Brunswick, NJ; Chicago, IL; Oakland, CA; Boca Raton, FL). Institutional Review Board approval was obtained at each institution-based site, and informed assent was obtained from participants' families. A detailed developmental history was recorded at the first encounter and updated at each subsequent visit. After the first encounter, individuals were seen at six month intervals up to age 12 and annually thereafter. Participants were examined by an RDCRN physician who completed a full neurological exam and recorded FOC, and hand and foot measurements, and then by a nutritionist/anthropometrist who recorded height, weight, skinfold measurements, lower leg length and limb circumferences. All measurement techniques are outlined in the manual of operations for the study (appendix A). Data were marked on forms using several techniques including bubble-entry, and character entry.

Forms were faxed to a data collection center where they were scanned into the database, or entered manually via online access to the database.

The primary aims of the continuing RDCRN study are to establish phenotype-genotype correlations, and advance the understanding of RTT clinical features, quality of life, and longevity. The target population of 1000 participants represents nearly one-third of the members of the International Rett Syndrome Foundation, and approximately half of this target had been accrued when the growth study was initiated. The growth study reports on the anthropometric findings of the parent study. Data were supplemented by additional non-RDCRN patients from the practice of one of the principal investigators (PI).

### **Inclusion and exclusion criteria**

The 495 participants enrolled in the RDCRN study were screened for inclusion in the growth charts. 13 males were excluded, as they were too few to produce meaningful gender-specific growth curves. Of the 482 remaining females, 451 ranged in age from 8 months to 25 years and were selected to create the growth charts. An additional 59 individuals were included who had been followed by one of the RDCRN PIs, but who were either deceased or lost to follow-up prior to implementation of the current study. Although the goal was to create charts to age 18, the age range of 25 years was chosen specifically to deal with right-edge effect, discussed in chapter six.

### **Participant attributes**

These supplementary data were extracted through chart review, incorporating additional variables when available, such as standard measures of disease severity, scoliosis, nutritional supplementation, and comorbid diseases. Thus a total of 510 females were included in further growth analysis (table 7-1). Of these, information on seven deceased individuals was included, one from the RDCRN database and six from the supplemental population. Of these, none was

born prior to 1988. The range of birth year for individuals was from 1945 to 2006 (figure 7-1). Earlier data were included specifically so that secular trends could be examined in detail. Ninety-five percent of the population were born after 1982 (i.e., within the 25 year age range of the reference chart). The age-at-enrollment in the RDCRN study ranged from one year to forty-eight years, however, 95% enrolled prior to 21 years of age (figure 7-2). Although some were diagnosed later in life due to the recent discovery of the syndrome, ninety percent were diagnosed before age 8 (figure 7-3).

Five participants had secondary medical conditions not associated with RTT. One was born with neonatal arthrogryposis. One had type 1 Diabetes Mellitus and hypothyroidism, treated successfully with levothyroxine. One had a clinically insignificant ventricular septal defect. Another had evidence of precocious puberty with pubic hair at age 2y. Another had precocious puberty, with pubic hair at age 20 months, as well as gelastic seizures. Additionally, two participants were identical twins, both with the same *MECP2* mutation (R168X). Race and ethnicity data were available on all but six (1%) of the RDCRN participants, and 88% were white, 3.5% were black, 4% were Asian, 0.5% were American Indian or Alaskan, and 3% reported more than one race. Regarding ethnicity, 12.5% were Hispanic or Latino, 1% was unknown or not reported, and the remainders were not Hispanic or Latino. Of the 59 supplemental individuals, race and ethnicity data were available on 56, and 83.9% were white, 8.9% were black, 1.8% were Asian, and 1.8% were from the Indian subcontinent. Data on residence were sparsely recorded, and of 18 records, 16 lived at home and 2 lived in group home facilities. Only four of 520 individuals were born prior to 36 weeks gestational age (WGA), and the range of these was from 32 WGA to 34 WGA.

## Diagnosis and genetic information

Of the three possible diagnoses described in chapter five, 440 (86%) had classic RTT, 54 (11%) had atypical RTT, and 16 (3%) had *MECP2* mutations but neither classic nor atypical RTT (non-RTT). All diagnoses were confirmed by one of four RDCRN neurologists or geneticists. Of the atypical individuals, only 13 had subclass diagnoses: two had preserved speech (A-1), three had minimal evidence of RTT (A-2), one had no hand use regression (A-3), two had early onset seizures (A-4), and five had delayed onset RTT (A-5). Therefore, the atypical RTT individuals were all considered as one group. A total of 453 individuals had mutations in *MECP2*. Of the 139 individuals with one of the four common missense point mutations, 60 had T158M (11.8%), 36 had R306C (6.9%), 30 had R133C (5.9%), and 14 had R106W (2.7%). Additionally, 140 individuals had one of the four common nonsense point mutations divided as follows: 43 had R255X (8.4%), 41 had R168X (8.0%), 28 had R294X (5.5%), and 28 had R270X (5.5%). Only six individuals were affected with the next most common point mutation, P152R, and only one or two individuals had each of the remaining 27 point mutations. Deletions were divided into the following groups: carboxy-terminal deletions after base-pair (BP) 930 (C-terminal del) affected 44 (8.6%), 26 (5.1%) had deletions before BP 930, 14 (2.7%) had unspecified deletions. Insertions were also divided into those before BP 930 affecting eight individuals (1.6%), and after BP 930 affecting five (1.0%). Four individuals had mutations with unknown details. Other miscellaneous mutations included polymorphisms in three (0.6%), intervening sequences in two (0.4%), a small duplication in one (0.2%), and four unspecified mutations: one insertion, two point mutations, and one splice site mutation. No mutations were found in 33 individuals (17 classic and 16 atypical), however 23 of these had not had expanded testing as described in chapter five. Mutation testing was pending on 9 individuals (1.8%) at the time of the study, and results were unknown in 15 (2.9%).

In addition to anthropometric variables, historical and clinical information was collected at each RDCRN visit. Historical information included the following: a detailed birth and developmental history; comorbid diseases, surgeries, and hospitalizations; ethnicity and race; residence type; seizure type, frequency, and treatment; RTT supportive criteria, behaviors, and stereotypies; scoliosis history; electrocardiogram (EKG) reports; menstruation and pubertal history; and genetic mutation details. Current management information included diet history and nutritional supplementation, GER, constipation, sleep dysfunction, recent hospitalizations, fractures, and types of organized therapy (physical, occupational, speech, music, etc.).

Clinical assessment included RTT subtype diagnosis based on RDCRN physician assessment, a complete physical exam with general and RTT-specific characteristics, Tanner staging, ambulation and scoliosis assessment. Two commonly used severity scales, the Motor Behavioral Assessment (MBA) and Clinical Severity Score (CSS), were employed at each visit to measure both historical development and recent changes. The MBA includes current information on verbal ability, social interaction, apraxia, ability to eat, toilet training, seizures, bruxism, breathing dysfunction, dystonia, spasticity, and scoliosis. The CSS includes a mixture of information including the following: current information on growth failure, seizures, scoliosis, language, hyperventilation, and dysautonomia; mixed past/present information on hand use; and past information on regression onset, and age at sitting and walking. Additional information was collected on quality of life of the individual and the family. These variables were used to compare growth among various degrees of severity with regard to scoliosis, seizures, communication, hand use, ambulation, and age at regression. A detailed list of these variables is included in appendix D.

## **Reliability Testing**

As part of the growth chart study we performed an interim assessment of reliability to determine if adjustments needed to be made to study design or measurement technique. The methodology, selection of statistical tools, and results of this additional study are included in appendix B. In summary, the TEM, a measurement of average absolute error in measurement, was well within acceptable limits for the types of measurements conducted in this study.

## **Statistical Analysis**

To simplify analysis, three separate categories were created by categorizing common mutation types. The first category (I) included a reduction of the above 52 types, including the eight common point mutations, the four deletion subsets, and the two insertion subsets (figure 7-4). The second category (II) included nine groups redefined based on mutation type as follows: nonsense mutations, missense mutations, frameshift mutations, inframe variations, splice site mutations, polymorphisms, silent variants, duplications, and no-mutation-present (figure 7-5). The third category (III) was divided based on 11 mutation sites on the *MECP2* gene as follows: 5' untranslated region (5'UTR), N-terminal region (NTR), methyl binding domain (MBD), interdomain region (IDR), transcription repressor domain (TRD), nuclear localization signal (NLS), C-terminal region, 3' untranslated region (3'UTR), intron mutation, exon 1, exon 4, or unknown (figure 7-6).

## **Data cleaning**

Data were initially examined for errors using multiple approaches (appendix A). First, frequency tables were generated for all text variables. This technique identified a significant problem in the database involving redundant coding of mutation type. Since identical mutations can be expressed as amino acid change or as DNA sequence change, this variable needed to be

examined closely for redundancy and reduced to common units. One additional preparation step involved sorting the mutations into categories (described above) to facilitate analysis.

Mathematical impossibilities were addressed next. Date of birth was recorded in two locations in the data, and non-zero differences were corrected. Age was calculated by subtracting DOB from DOV and negative ages were examined. Additional mathematical impossibilities were identified, such as discrepancies in hand measurements (palm and middle finger must add to total hand length) and composite severity scores. Duplicates were eliminated by searching first for duplicates of ID number, age, and all three measurements, and second for partial duplicates in which age, ID and one measurement were identical, but other measurements were either absent or different. For variables which were only collected through the RDCRN, any observations which were not coded as RDCRN and contained spurious data were deleted.

To begin searching for outliers, data were arranged in ascending age and measurement values were scanned to look for obvious anomalies. Measurements themselves were then individually arranged in ascending order, and adjacent values were scanned for unlikely values. Next, general scatterplots were generated to search for outliers using age-by-measurement at four age intervals, birth to one year, one to three years, three to twelve years, and twelve to twenty-five years. After abnormal values in the general plots were addressed, individual scatterplots for each participant were then generated in batches of 10-20. Lines were fitted between the data points, and spurious values outside of that individual's continuous trend were identified. If datapoints for an individual described a smooth curve with a single exception within two months proximity to another point, that exception was examined. The point was deleted if it deviated from the expected value of the curve by greater than 2kg for weight, 2cm for height, or 1cm for

FOC. No adjustments were made if only three or fewer observations were present for the patient, as no trends could be extrapolated.

To check for outliers and unusual distributions in certain age groups, data were binned into the standard check-up age groups described above, and boxplots were created for each measurement parameter at each age group. The standard definition of 1.5 times the IQR was used to identify outliers. Any outliers identified in this manner were checked for consistency against the original records, but none were deleted based solely on this technique.

Common errors included keystroke entry, and pound/kilogram or centimeter/inch transposition. Many errors resulted from scanning numbers directly into the database, and after this was discovered manual entry was used exclusively. After these errors were remedied, discrete narrow age ranges were chosen (birth, 0.45-0.55y, 0.95-1.05y, and 1.95-2.05y), and measurement distributions were examined for normality using histograms, Q-Q plots, and the Shapiro-Wilk test. In each data cleaning exercise, discrepancies were recorded and investigated through source documentation. If no resolution could be achieved the data were discarded. Less than 0.1% of the data needed to be discarded after this cleaning process.

### **Statistical comparisons**

All comparisons between categorical or ordinal variables, such as mutation type or disease severity, and continuous variables, such as height or weight, were done using t-test or ANOVA as appropriate. Levene's test for homoscedasticity was performed for all ANOVA tests and is only mentioned if results were significant. All tests were considered significant if  $p < .05$ . Corrections for multiple t-tests were made using the FDR, which is interpreted as the expected percent of false predictions in a set of predictions, and an  $FDR < .10$  concurrent with a  $p < .05$  was considered significant.

## **Growth curve modeling**

To address the primary hypothesis, both cross-sectional and longitudinal data were combined to create references for all classic RTT participants. These references were compared to references of normal healthy children to address secondary hypothesis 1. Both the CDC and WHO references were used to compare height, weight, FOC and BMI. The normal reference used for FOC in older children and on the growth charts themselves was the British 1990 growth reference. This reference was chosen for two primary reasons. First, it was created using the same state-of-the-art statistical techniques as the current RTT reference, whereas the CDC reference used a hybrid approach involving approximation of polynomials and *a priori* curves fit using LMS. Second, because our study includes participants to age 25, and data in this reference are available up to age 17 for female FOC. To examine secondary hypotheses 2a, 2b, and 2c, additional groups were created to compare growth within classic RTT. The first included charts for atypical RTT and non-RTT. The second divided participants who were born before January 1<sup>st</sup>, 1997 and those born on or after that date. The third examined the distributions of the global severity scales, the CSS and MBA. Both of these scales appeared to have a bimodal histogram, suggesting that the data clustered into “mild” and “severe” groups. Therefore, to test the hypothesis that growth is better among individuals with overall lower severity, two further sets of growth charts were created by dividing the severity score data between the bimodal peaks. Since the CSS contains two questions of growth of individuals, a modified CSS was created by removing these questions, and this was used to divide the individuals.

Statistical modeling of the growth curves was done using the LMS method described in detail in chapter six. The parameters for the EDF of L, M, and S were chosen based on examination of the GOF testing described in that chapter. The growth charts are displayed in

appendix G, and the values for L, M, and S, and results of the GOF testing along with the individual L, M, and S charts are displayed in Appendices E and F.

### **Growth velocity**

To examine secondary hypothesis three regarding growth velocity, two velocity modeling curves were used. For longitudinal data available from birth to adulthood, the Preece-Baines model 1 was used (Preece and Baines 1978). This model approximated velocity which trended to zero (an asymptote for height and FOC) in adulthood. To examine early growth in isolation, the Jenss model, a model using nonlinear least squares methods, was used to fit curves for weight velocity, length velocity, and FOC velocity. This model fit early growth velocity data better than high degree polynomials.

### **Group comparisons**

Early characteristics described in secondary hypothesis four were compared using t-test or ANOVA depending on the number of categorical or ordinal variables. Initial comparisons were based on absolute adult values of height, weight and FOC. Since RTT is associated with normal measurements at birth, comparisons were also made based on change in height, weight and FOC (relative to birth), and change in Z-score for these measurements from birth to adulthood.

### **Growth standard selection**

To determine if growth standards would be significantly different from the references produced without removing unhealthy individuals, modifiable risk factors for growth were evaluated. T-tests were used to compare change in growth from birth to adulthood in individuals who had received certain therapies, had nutritional intervention, and lived at home versus group home.

All statistical analyses were performed on SAS and SPSS with the exception of the curve fitting, smoothing, and Z-score calculations which were performed on LMSchartmaker (Cole and

Pan). The WHO iGrowup macro (2005) for SPSS was used for comparing Z-scores in RTT against the WHO and CDC data.

## **Results**

### **Measurements**

A total of 5723 observations were recorded, averaging 11.2 observations per individual. These total observations included 5312 weight measurements, 3745 length measurements, 3338 FOC measurements, and 3709 BMI calculations. Type of height measurement was also recorded in 50% (including all RDCRN participants) and distributed as 1623 recumbent, 1177 standing, and 82 calculated using segmented measurements. All individuals under two years were measured in the recumbent position, and from age three to eighteen, 30-50% were measured recumbent. This percentage is consistent with the percentage in the study population able to ambulate (61.6%). Source of measurement was also recorded, and 1428 measurements (25.0%) were performed at RDCRN sites, 2788 (48.7%) were obtained from physician notes, 190 (3.3%) were from hospitalization records, 20 (0.3%) were from emergency department notes, 1042 (18.2%) were from physician growth charts, and 255 (4.5%) came from lists provided by families. In addition to these standard measurements, several additional anthropometric variables were measured, including 1205 hand measurements, 1163 foot measurements, 890 arm circumferences, 890 tricep skinfolds, 886 bicep skinfolds, 876 subscapular skinfolds, 884 suprascapular skinfolds, 884 thigh circumferences, 851 thigh skinfolds, 885 shin circumferences, 858 shin skinfolds, 882 and lower leg lengths.

The number of measurements varied according to diagnosis and mutation category (table 7-2). Regarding diagnosis, the majority of measurements were on classic RTT individuals (88-89%), followed by atypical RTT (9-10%) and then non-RTT (3-4%). Certain groups within the simplified categories did not contain sufficient observations for meaningful analysis. For the

mutation reduction category (I), the observation frequency ranged from a minimum of 29 observations in C-terminal insertions, to a maximum of 543 observations for R168X (table 7-3). However, the mutation type category (II) yielded groups large enough for comparison, with up to 2162 measurements in the nonsense group and 1622 in the missense group (table 7-4). Mutation site groups (III) also yielded larger groups with 1612 measurements in the TRD group and 1134 in the MBD group (table 7-5).

The number of observations also varied dramatically with age. Many observations were recorded in the first year of life, and the number of observations decreased almost logarithmically with age. Birth measurements ranged from 139 observations for FOC to 234 for weight. While measurements in the first two months of life range from 287 for FOC to 430 for weight, measurements at other ages were significantly lower. However, frequencies at standard well-child check-up ages (2, 4, 6, 9, 12, 15, 18 and 24 months, and yearly thereafter) were above 100 until age 10 with no exceptions. The number of observations per age plateaued after age 10 to a range of 26 to 136 per year (table 7-6). Concerning only classic RTT individuals, the frequency at standard check-up intervals was above 100 until age 10 with only seven exceptions (table 7-7).

Although the RDCRN study is a longitudinal natural history study, at the time of analysis no participants had greater than three measurements recorded by RDCRN staff. However, many of these participants contributed historical data from their pediatric records, and as a result these historical data were treated as longitudinal. To qualify as longitudinal, data were defined as five or more data points from the same individual. Therefore, for purposes of analysis all data recorded by RDCRN staff were treated as cross-sectional. Data were coded as to source (RDCRN, pediatrician, hospital records) so that they could be weighted according to relative

significance. Although the reliability of data from pediatric offices is unknown, the reliability of the RDCRN data could be tested and was therefore considered of higher quality. Individuals were almost evenly divided between those contributing longitudinal data (N=224) and those providing only cross-sectional data (N=286). The majority of early measurements (under two years) were longitudinal, while from age 8 to 25 most were cross-sectional.

### **Summary Statistics**

The distributions of weight, height and FOC at birth in RTT were very similar to those in normal individuals (table 7-8). Of the birth data available in 238 individuals with classic (N=207), atypical (N=21), and non-RTT (N=10), no significant differences existed among groups for weight, length, or FOC (p-value ranged .31-.79). Weight in study individuals was not significantly different from normal WHO individuals (p = .29). Although study individuals were longer at birth compared to both CDC and WHO references (mean CDC=49.29cm, WHO=49.15cm, study=50.4cm, p<.001), this discrepancy disappeared by two months of age (p=.65). Mean birth FOC ranges somewhat in standard references (CDC=34.71cm, WHO=33.88cm, British=34.54cm), and mean study FOC of 34.24 was between these values, significantly larger than WHO mean (p=.001) and significantly smaller than both the CDC mean (p=.004) and the British mean (p=.009).

The Shapiro-Wilk test of normality at birth was significant in classic RTT for length (p<.001), weight (p=.04) and FOC (p=.02). Quantile-Quantile plots and histograms revealed that each of these measurements was slightly negatively skewed. These non-normal distributions became normal by two months of age both by graphical examination and statistical tests. Notably, the Shapiro-Wilk test became non-significant for weight and FOC after removal of six premature birth values from the 238 total values. The distributions were sporadically non-normal for all measurements after 11 months, and adjustment was made for these distributions through

transformations for the growth charts. Final adult height, weight, and FOC showed trends to larger measurements in non-RTT individuals, but since few adult height measurements were available and the SD was large, these results were not statistically significant (p-value range .07-.25).

Adult values were available on 86 classic RTT participants, including 85 weights, 74 heights, 69 FOCs and 74 BMIs. Based on the CDC and British criteria, the adult values ranged dramatically among classic RTT women (table 7-9). The heaviest woman (R294X mutation), 82.7 kg at age 17.7 also had the largest FOC and highest BMI. All of the minimum adult values are well below -3 SD. In fact, the minimum weight value is so low that the estimate of that individual's percentile using the normal distribution would be on the order of  $10^{-28}$ . The mean values for height and FOC are below -3 SD and the mean for weight is below -2.5 SD, yet the mean BMI balances out to -0.37, very near the CDC mean.

## **Growth Curve Fitting**

### **Choice of statistical methodology for estimating and smoothing percentiles**

Based on the ideals for growth chart methodology set forth by the WHO (see chapter 6), I chose to use Cole's LMS method for curve fitting. Although the model does not adjust for kurtosis, the models that do include the extra parameter needed for kurtosis adjustment require a significantly larger sample size than most rare disease populations can supply. However, the diagnostic tests used to evaluate model fit assessed the degree of fit with respect to all four moments, M, SD, skewness, and kurtosis.

I adjusted the ages for prematurity by subtracting the amount of prematurity in years (weeks/52) from the ages of all premature children. This resulted in negative ages which had to be offset when using a log transformation, but otherwise did not pose a computational problem.

## **Diagnostic techniques for curve fitting and adjustment of the model**

These diagnostic tests available in the LMS ChartMaker program by Pan and Cole include both graphical and statistical aids in fitting curves to empirical data. As the user changes the model input variables – the estimated degrees of freedom (EDF) – the diagnostics provide information about changes in the fit of the model. First, the penalized log likelihood deviance, a measure of the degree to which the iterative approach is able to fit the empirical data, will decrease with a better fit. Therefore, in fitting the curves I monitored the decreases in deviance with changes in EDF to see if the change produced a better fit. A change in deviance of 8 or more is statistically significant with a moderate sample such as our group had collected, however with larger sample sizes over 10,000 changes in deviance of over 20 can correspond to trivial changes in the shape of the fitted curves. The next test, the Q-statistic or studentised residual of the Chi-square goodness of fit test, measures the Z-scores of the data against the normal distribution. To indicate that the data follow a normal distribution, the Q-statistic should be between -1.96 and 1.96 (the 95% confidence interval). If the Q-statistic was outside this range for any of the moments, I increased the EDF associated with that moment. Although the Q-statistic is informative about Kurtosis in the model, the LMS model does not adjust for kurtosis. The final diagnostic test is the detrended Q-Q plot. Although not as sensitive as the “worm plot,” or collection of detrended Q-Q plots based on age group, the individual detrended Q-Q plot is a good summary diagnostic test of normality. If the worm did not pass through the origin, I increased the EDF of M. If the slope was not zero, the EDF of S was increased. Lastly, if the worm was U-shaped, the EDF of L was increased. I also plotted the empirical data on the curves and examined for areas of low density, suggesting that few data supported those portions of the curves. If this were true I adjusted the EDF of M to attempt to fit the curve to an area with greater density. However, in smaller data sets it is to be expected that some portions of the curve

will have sparse empirical data, theoretically compensated for by the continuous nature of the curve estimation.

### **The LMS technique in practice**

The M curve represents the most important variation, followed by the S curve, and finally the relatively small influence of the L curve. Therefore, I first optimized the EDF for M, by beginning with an EDF of 5 and increasing by intervals of 1 while monitoring the decrease in the penalized likelihood deviance until the change in deviance became less than 8. If the change in EDF resulted in a change in deviance of less than 4, I changed the EDF back to the previous value. I simultaneously observed the curves for the appearance of physiologically unlikely trends (peaks and valleys or wiggles at ages when growth should be continuous). I examined the data points plotted within the curves for evidence of outliers or large gaps in empirical values to support the fitted curves. Pan and Cole's guidelines specify that for small datasets (<500) the EDF for M will be smaller (approximately 5) than larger datasets, and for over 10,000 values the EDF may be over 15 (Pan and Cole 2005). If the curve appeared appropriate, I turned to the Q-statistic, and if the M value was outside the desired range (-2 to 2) I increased the EDF further. In some cases, especially when the data in older age groups was sparse relative to younger ages, I also performed an age transformation, beginning with a log transformation (power = 0, offset = absolute value of negative ages), and changing the power on intervals of 0.1 while noting the change in deviance, shape of the curves, and Q-statistic. Finally, I examined the detrended Q-Q plot for confirmation that the data were normally distributed across all Z-scores.

I then optimized S by starting at an EDF of 1, increasing by intervals of 1, and following the deviance, curve shape, and Q-statistic. Cole recommends skipping EDF values of 2 for L and S to evade "silly values at the extremes" (van Buuren and Fredriks 2001). Finally, I examined the S curve itself for informative trends in the coefficient of variation, both to note potential errors in

the dataset, and also to search for unexpected trends in the variance of the population at different ages.

Two additional curves are available in the LMS Chartmaker program, the M curve velocity and the M curve acceleration. Median curve velocity is an even more sensitive indicator of smoothness than the M curve, since the first derivative changes more dramatically with changes in the direction of the M curve. Acceleration, the second derivative, changes with the slope of the velocity, and is therefore even more sensitive to change. In practice, although the velocity curve can reveal trends in the data such as growth spurts or dips, the acceleration curve is difficult to interpret.

At periodic intervals in fitting the model, I checked the Z-score plot to see if any values were beyond -3 SD of +3 SD (outliers). I attempted to fit these back within the distribution by changing the model, and if this altered the original results significantly, I identified the specific values of these outliers to examine the original data for errors.

### **Primary Hypothesis: Growth Charts for Classic RTT**

This fitting method resulted in excellent fit for the dataset of classic RTT (see Growth Charts, appendix 177). The diagnostic tests (Q-statistic, detrended Q-Q plot) revealed excellent fit, and the empirical data were very evenly distributed at younger ages and moderately dense at older ages (appendix E). The data became significantly sparser, especially for FOC, after age 20. To test the degree to which the curves coincide with non-parametric estimates of the distribution, we compared percentiles at various ages (0, 2, 3, 5, 10, 21 years) against the Z-score generated values coinciding with these percentiles based on the normalized distribution derived through LMS. Percentile-to-Z-score differences are reported as ranges of absolute difference and percentage difference (table 7-10). The differences at the mean (between mean and median) were negligible for all values, with the exception of a 9.8% difference in weight mean and median at

age 21 (the curves over estimate weight by 3.25kg). The percentiles were calculated out to the 2<sup>nd</sup> percentile, and several significant differences occurred at these extreme percentiles, notably an overestimation of weight at the 2<sup>nd</sup> percentile by 0.25kg (11.6%), an overestimation of weight at the 95<sup>th</sup> percentile by 2.4kg (11.6%) at five years of age, and an overestimation of height at the 2<sup>nd</sup> percentile by 1.87cm (5.2%) at 10 years of age. Overall the fit for height was good, and for FOC the fit was remarkably consistent through all age groups.

### **Secondary Hypothesis 1: Height, Weight and FOC are Decreased in RTT Relative to Healthy Children**

References for healthy children included the CDC 2000 growth charts, the WHO 2007 growth charts, and the 1990 British growth charts. Significant differences existed between normal children and classic RTT for weight, height, and FOC, but not for BMI (table 7-11).

#### **Weight**

Weight mean for classic RTT begins to drop below that of both the WHO and CDC at six months of age. Using the t-test for significance and controlling for familywise error using the FDR method, the median for RTT was significantly below that of the WHO beginning at 25 months ( $p=.029$ ,  $FDR=.047$ ), and after 32 months the difference between groups remained strongly significant throughout growth ( $p<.0001$ ,  $FDR<.0001$ ). Notably, RTT weight was significantly different based on the CDC beginning as early as 14 months ( $p=.04$ ,  $FDR=.05$ ), and remained strongly significant after 22 months ( $p<.0001$ ,  $FDR<.0001$ ). By age 12.5 years the median weight for RTT was -2 SD (2<sup>nd</sup> percentile) compared to that of the British charts. Notably, the SD of weight in RTT was significantly wider than in normal children. For example, at six years, the RTT mean is -1 SD on the British reference, yet the RTT 98<sup>th</sup> percentile is greater than the British 98<sup>th</sup> percentile. Although the RTT mean was well below -2 SD on the normal curves after age 12.5 years, the 98<sup>th</sup> percentile persisted at about +1 SD on the British

reference. The average weight in RTT at 18 years was 40kg, compared to 58kg for normal 18 year-old females.

## **Height**

The slope of the mean for RTT begins to drop below the WHO mean at 17 months. Using the t-test and FDR method, the mean for RTT was also significantly below the WHO beginning at 18 months ( $p=.004$ ,  $FDR=.006$ ), and after 20 months the difference between groups remained strongly significant throughout growth ( $p<.0001$ ,  $FDR<.0001$ ). Alternatively, based on the CDC the mean for RTT became lower slightly later at 21 months ( $p=.02$ ,  $FDR=.03$ ), and after 24 months the difference between groups remained strongly significant ( $p<.0001$ ,  $FDR<.0001$ ). By age 12 years the mean height for RTT was -2 SD (2<sup>nd</sup> percentile) that of normal children. The SD for height was also wider than in the normal population, though not as dramatically as for weight. After age 12 years the mean RTT height was well below -2 SD of normal, but the 98<sup>th</sup> percentile for RTT was almost at the mean for normal children. The average height in RTT at 18 years was 144cm, compared to 164cm for normal 18 year old females.

## **FOC**

The slope of the mean for RTT begins to drop below that for normal children based on the British reference as early as 1.5 months, based on Nellhaus by three months, and based on the CDC and WHO at about 12-15 months. Using the t-test and FDR method, the mean FOC for RTT was significantly below the WHO at 15 months ( $p=.029$ ,  $FDR=.038$ ), and after 22 months the difference between groups remained strongly significant throughout growth ( $p<.0001$ ,  $FDR<.0001$ ). Similarly, mean FOC for RTT was significantly below the CDC at 15 months ( $p=.017$ ,  $FDR=.026$ ), and after 20 months the difference between groups remained strongly significant throughout growth ( $p<.0001$ ,  $FDR<.0001$ ). Remarkably, the classic RTT FOC was significantly different from the British mean at birth ( $p=.03$ ,  $FDR=.03$ ), and became strongly

significant after one month ( $p < .0001$ ,  $FDR < .0001$ ). By age two years the mean FOC for RTT was -2 SD (2<sup>nd</sup> percentile) compared to that of normal children. The average FOC in RTT at 18 years was 51cm, compared to 55.5cm for normal 18 year old females.

## **BMI**

Mean BMI slope for RTT begins to drop below that for normal children at 4-5 months, but BMI velocity is roughly parallel that of normals from nine months on. Using the t-test and FDR method, the mean BMI for RTT was not consistently below that for normal children based on either the WHO or CDC references. Although a significant difference emerged at 13 months with RTT BMI lower than BMI in WHO children, it did not reach significance based on the adjustment for familywise error, and the p-value significance disappeared at 23 months (p range .006-.033, FDR range .065-.140). The average BMI at 18 years for RTT was 20, compared to 21 for normal 18 year old females.

## **Secondary Hypothesis 2: Differences in Growth Exist within RTT Subpopulations**

### **2a: Based on diagnosis of classic, atypical, or non-RTT**

#### **Weight**

Comparing classic and atypical, the slope of the mean weight for Classic RTT decreases slightly relative to atypical RTT beginning at 21 months (table 7-12). At age nine years, classic growth velocity increases relative to atypical, and the mean growth crosses at age 14 with classic weight higher than atypical. Average weight for classic at 18 years is 39.5 kg compared to 38 kg for atypical. No significant differences exist between the two groups based on t-test (p range .06-.98).

Comparing classic and non-RTT, the slope of the mean weight for classic decreases relative to non-Rett beginning at 18mo (table 7-12). This trend continues with average height for classic at 18 years equal to 40 kg compared to 58 kg for non-Rett (equal to the normal reference

population). Using repeated t-tests and controlling for familywise error using the FDR method, the mean weight for classic RTT was significantly below that for non-Rett children beginning at 43 months (p range <.001-.034, FDR range .005-.050), and this difference persisted throughout childhood.

## **Height**

Comparing classic and atypical height, the slope of the mean height for classic RTT decreases slightly relative to atypical RTT beginning at 18 months (table 7-12). At age 8 years, classic growth increases relative to atypical, with the median growth crossing higher than that for atypical and then decreasing in velocity from ages 13-18. By age 18 years the median growths are equal at 144 cm. No significant difference exists between the two groups based on t-test (p range .13-.99).

The slope of mean height for classic RTT decreases relative to non-Rett beginning at 18 months (table 7-12). This trend continues with average height for classic at 18 years is equal to 144cm compared to 159cm for non-Rett. Using repeated t-tests and the FDR method, the mean height for Classic RTT was significantly below that for Non-Rett children beginning at 31 months (p range .001-.033, FDR range .001-.090).

## **FOC**

Mean FOC for classic RTT decreases slightly relative to atypical RTT beginning at two years (table 7-12). This trend stabilizes and the two subtypes have parallel growth in FOC with average FOC for classic at 18 years equal to 51cm compared to 52cm for atypical. No significant difference exists between the two groups based on t-test (p range .08-.99).

The slope of mean FOC for classic RTT decreases relative to non-Rett beginning at 15 months (table 7-12). This trend continues with average FOC for classic at 18 years equal to 51cm compared to 56cm for non-Rett (just above the normal population). Using repeated t-tests and

the FDR method, the mean for classic RTT was significantly below that for non-Rett children beginning at 33 months ( $p=.047$ ,  $FDR=.060$ ), but this difference disappeared by 34 months. Classic RTT FOC was again significantly lower at 46 months and this time the difference persisted throughout childhood ( $p$  range .001-.023,  $FDR$  range .007-.026).

## **BMI**

Although classic RTT demonstrates a sharper rise in BMI after birth compared to atypical RTT, classic also experiences a sharper drop in BMI beginning at 8 months (table 7-12). Consequently, the average BMI for classic RTT is lower than that for atypical from age 1.5y to 16y. At 16y, the average BMI for classic RTT catches up to that of atypical. Although there is one statistically significant comparison at six months of age (classic BMI higher than atypical,  $p=.044$ ), the familywise correction did not reach significance ( $FDR = .741$ ). The average BMI at 18y was 20 for classic RTT and 19 for Atypical RTT.

Classic and non-Rett BMI are parallel from birth to age 2y, at which point classic BMI decreases relative to non-Rett BMI (table 7-12). There is a persistent gap in BMI with final average BMI at 18y for classic equal to 20 compared to 23 for Non-Rett. Although the graphical differences appear dramatic, the wide SD and small number of patients in the non-Rett group led to no significant differences on statistical testing. While  $p$ -values reached significant differences with Classic BMI lower than Non-Rett BMI at 49 to 60 months ( $p$  range .018-.044), and 73 to 114 months ( $p$  range .017-.036), the familywise correction did not reach significance ( $FDR$  range .142-.172).

## **2b: Based on secular trends**

### **Weight**

Comparing individuals born prior to 1997 when vigilant nutritional management was not common and those born after 1997 when supplementation was very common revealed no

significant differences in weight between classic RTT individuals born before 1997 and after 1997 (p range .06 to 1.00) (table 7-12). Although the mean after age 7.5 years for girls born before 1997 was consistently lower than those born after 1997, the SD in both groups was wide, and the differences after age 7.5 years did not approach significance (p range .13-.64). Final average weight at 10 years was 25kg for those born before 1997 and 27 kg for those born after 1997 (not significantly different).

### **Height**

Similarly, no discernable differences exist in height between classic RTT individuals born before 1997 and after 1997 (p range .35 to 1.00). Final average height at 10 years was 124 cm for before 1997 and 127 cm for after 1997.

### **FOC**

No discernable differences exist in FOC between classic RTT individuals born before 1997 and after 1997 (p range .15 to 1.00). Final average FOC at 10 years was 49.5 cm for before 1997 and 49.75 cm for after 1997.

### **BMI**

Individuals born before 1997 had a slightly lower BMI relative to those born after 1997, although this never achieved significance (p range .14 to .87). Final BMI at 10 years was 15 for before 1997 and 16 for after 1997.

## **2c: Based on disease severity at most recent examination**

### **Weight**

Comparing classic RTT individuals with milder CSS scores (<20, N=177) to those with more severe CSS scores ( $\geq$ 20, N=191), the slope of the mean weight decreased for severe relative to mild at age 15 months, but weight difference between groups was not statistically significant after correction (p range .04-.99, FDR range .34-.99) (table 7-12). However, final

mean weights in the two groups were dramatically different at 33kg in the severe group and 45 kg in the mild group.

Likewise, comparing classic RTT individuals with milder (<52, N=206) versus more severe MBA scores (>=52, N=195), the slope of the mean weight decreased for severe relative to mild at age six years, but weight difference was not significant after correction (p range .05-.98, FDR range .14-.98). Mean final weights in the two groups were also different, at 34kg in the severe group and 42.5 kg in the mild group.

### **Height**

The mean height was similar in both CSS groups until age seven years, when the slope of the severe group decreased, resulting in a mean adult height of 138 in the severe group and 148 in the mild group. This difference was only weakly significant beginning at age 7.25 years (p=.05, FDR = .15), and fluctuated throughout the rest of growth (p range .01-.18, FDR range .15-.28).

The MBA groups also showed a difference in height beginning at five years when the slope of severe MBA height dropped below that of mild MBA. The significance was more prominent in this group reaching cutoff values at 68 months (p=.04, FDR=.08), and strong significance at 77 months (p<.01, FDR<.01). Final adult height was 139 in the severe group and 147 in the mild group.

### **FOC**

Mean FOC in the severe CSS group dropped precipitously away from that in the mild group at about six months of age. This difference became statistically significant at one year (p=.03, FDR = .11), however this significance fluctuated throughout growth (p range .005-.278, FDR range 0.11-0.61). Adult mean FOC at 18 years in the severe group was 50 cm, compared to 52.5 cm in the mild group.

Similarly, the MBA groups showed a difference in FOC beginning at 8 months when the slope of severe MBA mean FOC dropped below mild FOC. The difference reached significance at 24 months ( $p < .01$ , FDR=.02), and persisted through age 18 years. Final adult FOC was 50 cm in the severe group and 52.5 cm in the mild group.

## **BMI**

Mean BMI in the severe CSS group paralleled that in the mild group until 8 years of age when the mean slope began to decrease. This difference never achieved significance (p range .32-.98, FDR range .95-.98). However, mean BMI at 18 years was 17 in the severe group and 20 in the mild group.

Likewise, mean MBA BMI in the severe and mild groups was similar to age 11 years when severe BMI slope decreased. This difference did not achieve significance after correction (p range .03-.99, FDR range .81-.99), although final BMI at 18 years was 17.5 in the severe group and 20 in the mild group.

## **2d: Based on mutation, mutation type, and mutation site**

Several categories of mutation type were compared with growth variables. The simplest categories were based on type of mutation, including missense, nonsense, frameshift, inframe variations, duplications, splice-site mutations, silent mutations and polymorphisms. Of those who achieved adulthood, there were no significant differences among measurements except with the exception of classic RTT (table 7-13). In the case of classic RTT, nonsense mutations were significantly taller than missense, who were taller than those with frameshift mutations ( $p = .036$ ), however significant heteroschedasticity existed ( $p = .005$ ).

Several mutation sites are defined by functional or biochemical positions on the gene. Although trends existed with better growth for those with mutations in the TRD and C-terminal region, these differences were not statistically significant (table 7-14). An even more detailed

comparison of genotype and growth involves specific mutations. Individuals with eight common point mutations, four types of deletions, one type of insertion, and those who do not have any mutation were compared with respect to growth. Significant differences existed in adult height ( $p=.003$ ) (table 7-15) and a trend was present in adult FOC ( $p=.075$ ) (table 7-16). Individuals with deletions before amino acid 930 were shortest (with the exception of one individual with an R106W mutation), and those with insertions after amino acid 930 had the smallest average FOC. Those with R294X were the tallest, and those with C-terminal deletions had the largest average FOC.

### **Secondary Hypothesis 3: Growth Velocity in RTT is Decreased Relative to the Normal Population**

#### **Weight**

While in normal children the average weight velocity increases between age 11y and 13y (reflecting smoothing of pubertal increase in weight velocity), the weight velocity in RTT is flat during this time period. This absence of velocity change results in an average weight at 18y of 40kg, compared to 58kg for normal 18yo females.

Comparing mutation types using the Jenss curves, early weight velocity was significantly higher at nine months in individuals with missense mutations ( $M_{\text{missense}}=4.14$  kg/y) compared to those with nonsense mutations ( $M_{\text{nonsense}}=3.50$  kg/y,  $p=.011$ ), as well as at 12 months ( $M_{\text{missense}}=3.38$  kg/y,  $M_{\text{nonsense}}=2.60$  kg/y,  $p<.001$ ). At 18 months missense mutations ( $M_{\text{missense}}=2.50$  kg/y) grew faster than both nonsense ( $M_{\text{nonsense}}=1.78$  kg/y,  $p<.001$ ) and frameshift mutations ( $M_{\text{frameshift}}=1.82$ ,  $p=.018$ ); this difference persisted at 24 months ( $M_{\text{missense}}=2.04$  kg/y,  $M_{\text{nonsense}}=1.41$  kg/y ( $p<.001$ ),  $M_{\text{frameshift}}=1.36$  kg/y ( $p=.023$ )). Regarding specific mutation types, weight velocity was significantly higher in R133C at 18 months ( $M_{\text{R133C}}=3.50$  kg/y) compared to the following three mutations: R168X ( $M_{\text{R168X}}=1.73$  kg/y,  $p=.002$ ), R255X ( $M_{\text{R255X}}=1.42$  kg/y,

$p < .001$ ), and R270X ( $M_{R270X} = 1.72$  kg/y,  $p = .010$ ). This superior weight velocity in R133C persisted at age 24 months ( $M_{R133C} = 3.20$ ,  $M_{R168X} = 1.36$  ( $p = .001$ ),  $M_{R255X} = 1.03$  ( $p < .001$ ),  $M_{R270X} = 1.31$  ( $p = .007$ )), and age 36 months ( $M_{R133C} = 2.89$ ,  $M_{R168X} = 1.13$  ( $p = .013$ ),  $M_{R255X} = 0.77$  ( $p = .002$ ),  $M_{R270X} = 1.00$  ( $p = .026$ )).

## Height

While in normal children the average height velocity increases between age 11.5y and 13y (reflecting smoothing of pubertal increase in height velocity), the height velocity in RTT decreases noticeably during this same time period. This decrease in velocity results in an average height at 18y of 144cm, compared to 164cm for normal 18yo females.

Comparing early height velocity in specific mutation types using the Jenss curves, height velocity was significantly higher in R306C at nine months ( $M = 17.30$  cm/y) compared to those with large deletions ( $M = 13.52$  cm/y,  $p < .05$ ).

## FOC

Head circumference growth velocity was zero for RTT after age 2.5 and was relatively stable (growing with slightly decreasing velocity) for normals after age 4. Therefore, after age 4, growth was roughly parallel with  $+0.66SD$  for RTT overlapping  $-2.0SD$  for normals. This translates as 75% of the RTT population over 4y were consistently below 2<sup>nd</sup> percentile of normal (microcephalic).

Comparing early growth velocity using the Jenss curves revealed that mean FOC velocity was significantly higher at nine months in individuals born before 1997 ( $M_{pre1997} = 5.11$  cm/y) compared to those born after 1997 ( $M_{post1997} = 4.63$  cm/y,  $p = .036$ ), at 12 months ( $M_{pre1997} = 3.35$  cm/y,  $M_{post1997} = 2.87$  cm/y,  $p = .006$ ), at 18 months ( $M_{pre1997} = 1.79$  cm/y,  $M_{post1997} = 1.41$  cm/y,  $p = .002$ ), and at 24 months ( $M_{pre1997} = 1.23$  cm/y,  $M_{post1997} = 0.94$  cm/y,  $p = .02$ ). Regarding mutation site, those with C-terminal deletions had a greater FOC velocity ( $M_{C-term} = 1.62$  cm/y) at

age 24 months compared to those with MBD mutations ( $M_{\text{MBD}}=0.84$  cm/y,  $p=.03$ ), and at 36 months ( $M_{\text{C-term}}=1.47$  cm/y,  $M_{\text{MBD}}=0.57$ ,  $p=.026$ ).

#### **Secondary Hypothesis 4: Early Characteristics of RTT Predict Growth Later in Life**

While dozens of historical variables were collected on each RDCRN subject, the following are pertinent to growth. Ambulation was variable, and a total of 61.6% could ambulate to some degree during RDCRN examination, although 12% of these needed assistance. The majority of those who could not ambulate had never acquired the ability (29.4%), while 8.9% had lost the ability to ambulate. Ambulation in adulthood correlated strongly with final height, weight, and FOC (all  $p<.001$ ). This association was present among groups based on age at which ambulation developed, and those who walked earlier were heavier, taller, and had larger average FOC (table 7-17).

Scoliosis prevalence (58.6%) was similar to other studies of RTT, and 12.1% had undergone scoliosis surgery. Final height was significantly different among scoliosis groups ranging from an average of 149.2cm in those with no scoliosis to 138.3cm in those who had scoliosis surgery ( $p=.007$ ). Remarkably, those with no scoliosis and those with a 1-20 degree ( $M=147.1$ cm) and a 20-40 degree scoliotic curve ( $M=144.8$ cm) were similar in height ( $p=.43$ ). Surprisingly, scoliosis was not only associated with height, but also with weight ( $p=.035$ ) and FOC ( $p=.002$ ), however the association with weight disappeared when only classic RTT was examined ( $p=.075$ ). BMI was not associated with degree of curve ( $p=.34$ ).

One variable which could affect both disease severity and growth is the age at onset of regression. Age at regression varies widely in RTT, and was recorded in five intervals. The first interval included 26 individuals who regressed between birth and six months (6.1%), followed by 25 who regressed between 6-12 months (5.9%), then 135 from 12-18 months (31.7%), then the largest group of 179 from 18-30 months (42.0%), and finally 61 who regressed at an age

greater than 30 months (14.3%). Although ratios were similar among diagnoses, regression generally occurred earlier in atypical RTT and later in non-RTT. Although not a strong association, those with earlier regression had smaller average FOC ( $p=.05$ ); length, weight, and BMI were not associated with developmental regression (table 7-18).

Hand use, another marker of disease severity, also varied significantly in RTT (table 7-19). Hand use was significantly correlated with both weight ( $p<.015$ ) and FOC ( $p=.023$ ). On average, those with better hand use were heavier and had larger FOC than those who had lost hand use, however, the test of variance showed that significant heteroschedasticity existed in the weight groups ( $p=.01$ ). This association disappeared when controlling for disease type ( $p=.10$ ). Language ability was generally poor in RTT, but some atypical and non-RTT individuals had preserved contextual language. Although significant trends of increasing height and FOC existed with improvements in language, none was statistically significant ( $p$ -value range .17-.58) (table 7-20).

Sitting ability was measured based on age at which the skill was acquired. Most RTT patients only had mild to moderate delay in sitting. The ability to sit earlier in life correlated with all three measurements – weight, height, and FOC ( $p<.01$  for all). In adulthood, individuals who sat before nine months were significantly heavier, taller, and longer than those who acquired the ability later or never acquired it (table 7-21). This association was also unaffected by disease type.

Seizures, historically associated with a prevalence of greater than 50% in RTT, were less common in this study. The overall prevalence was 36% with little variability between classic (36%) and atypical RTT (35%). No significant associations existed among seizure severity and measurements (table 7-22).

Summary severity scores, the MBA and CSS, correlated strongly with weight, height, and FOC ( $p < .001$  for all). All these correlations were inverse, indicating that lower severity scores resulted in taller, heavier individuals with larger FOC. These correlations persisted with equal significance when controlling for age at final height measurement and RTT diagnosis subtype (table 7-23).

## **Discussion**

Differences between individuals in the study and the reference population were insignificant at birth with the exception of FOC, and this value was higher than the WHO reference and lower than the CDC and British references, so the significance of this is difficult to interpret. The differences are likely due to sampling bias accentuated by the small SD at birth which increases rapidly in the first 1-2 months. Likewise, differences in FOC among diagnoses in the study population were likely due to the small sample sizes of the atypical and non-RTT groups and therefore clinically insignificant.

The accuracy of the LMS growth curve estimates of the distribution vary according to age, based on the number of individuals in the underlying population. Although no perfect test exists to judge the fit of the percentiles, a reasonable assessment can be made by calculating empirical percentiles for an age range (1 year) and comparing the resulting percentiles to the Z-score correlates based on the normal distribution. The comparison of mean and median effectively judges how well the normalization transformation performed at that age. For ages birth to 10 years, the transformation successfully normalized the data in all measurement cases, since the discrepancy was less than 3%. The exception occurred in the mean weight curve for older individuals which overestimated the median. Regarding the 2<sup>nd</sup> to 98<sup>th</sup> percentiles, the curves performed very well using this technique in the age range of 2-3 years, likely due to the high volume of empirical values in this age range. The Z-scores in this age range are within 0-6% of

the percentiles for weight, 0-2% for height, 0-1.5% for FOC, and 0-5% for BMI. The curves perform less well for the adult estimation of weight and height. Considering all the adult individuals from age 17 to 25 (N=58), the mean age is 21.4 and the percentile estimates varying among the 2<sup>nd</sup> and 98<sup>th</sup> percentiles at this age vary anywhere from underestimating weight by 10% in the 98<sup>th</sup> percentile, to overestimating height in the 2<sup>nd</sup> percentile. Yet, this method of estimating percentiles is, in itself, misleading, since for adults it sums individuals over an 8 year time interval when measurement values are changing. However, too few individuals exist within any smaller adult time interval to generate meaningful percentile ranks.

Notably, while height in the normal population is invariably positively skewed, in this study height in the adult RTT population was noticeably negatively skewed. This suggests that very few women with RTT are tall, while some are extremely short.

The choice of the combination of references for normal data comparison was a compromise. The recent WHO reference employed the most rigorous data collection techniques to date for their 2007 reference. Unfortunately, the WHO reference only contains data to five years of age, and our data set extends to age 25. Moreover, the statistical techniques they used were effectively identical to the LMS method, and therefore provided no added incentive to use their reference. The CDC data set contains height and weight information to age 19, but only contains FOC information to age three, and does not include BMI reference information prior to age two. Few other references contain FOC data beyond early childhood. A Dutch reference includes height, weight, and FOC data to age 18 constructed with LMS, but no evidence suggested that this reference was superior to the British reference, and less detailed information was publicly available for it. The most popular reference of FOC in the US is the Nellhaus reference from 1968. Although the Nellhaus reference is considered an “international” reference,

the majority of the sources used to create this conglomerated work were from English, Irish, or white Anglo-Saxon American communities. Therefore it presented no added benefit based on diversity issues. Discrepancies do exist among all of these charts, and therefore the choice of which one to use is somewhat arbitrary. Both the creators of the WHO charts and the British charts have commented that no good explanation exists for these differences, and there is no objective measure of which is the “right” chart to use. Objectively, the Nellhaus chart is about 0.5 SD smaller but otherwise parallels the British chart.

Recently, Oddy et al., examined growth in RTT and suggested that other factors such as breath-holding and hyperventilation may play a role in poor growth (Oddy et al. 2007). Reilly et al., have also suggested that the growth failure in RTT is multi-factorial, involving oromotor function, motor control (self-feeding), communication ability, anxiety, social interaction, adverse behaviors, respiratory dysfunction (hyperventilation and breath-holding), seizure disorders, GER, constipation, developmental and cognitive level, posture (scoliosis and stereotypies), and other orthopedic issues such as hip dislocation (Reilly and Cass 2001). Since we have also collected data on these variables, future analyses will examine other behavioral and neurological factors that may play a role in growth failure. Oddy et al., also remark on the effects of interventions such as gastrostomy-tube supplementation. They note that in CP this intervention effects increases in height and weight thought to be catch-up growth, however this phenomenon has not been adequately demonstrated in RTT. As the RDCRN study continues and data on girls pre- and post-gastrostomy placement accrue, further longitudinal analyses will be able to answer this question with respect to height velocity.

Remarkably, although women with RTT are consistently well below -2 SD for height and weight, their small size appears to be proportionate, as their BMI is generally within the normal

limits as determined by the CDC. It is possible that BMI is not a good indicator of wasting in the RTT population, since BMI is more sensitive to changes in height, which is raised to the second power, than to changes in weight. Future efforts examining differences between RTT and the healthy population and differences within the RTT population could focus on weight-for-height as a measure of wasting instead of BMI-for-age.

Several authors have commented on the difficulty in measuring children with special needs (Oddy et al. 2007; Stevenson et al. 2006). This difficulty leads to frequent errors in measurement, both due to habits of individual operators consistently measuring differently from each other (inter-rater error), and variability based on cooperation at the time of measurement (intra-rater error). The operators in this study were all highly trained, and reliability measures bore out their skill. Nonetheless, all studies on growth and growth references, for clinical or research purposes, should incorporate measures of reliability. Ideally, studies should report the TEM, the coefficient of variation, the intraclass correlation coefficient (ICC), and the minimum and maximum absolute error.

One drawback of our study design is the absence of parental height, weight, and FOC data. These data are especially useful in longitudinal analyses as they allow correction formulae to be derived specific to the specialized population. Additionally, researchers can calculate the degree of correlation between height gain in the affected individual with mid-parental height to determine if significant correlation exists. Future studies incorporating longitudinal measurements should incorporate this variable, as well as sibling height, weight, and FOC if applicable.

In summary, we successfully generated growth reference curves for height, weight, FOC and BMI for girls and women with RTT from birth to 18 years. In doing so, we discovered

several significant trends in RTT growth, including the absence of a pubertal growth spurt. We also found many significant differences among different subpopulations of girls with RTT syndrome, including confirmation that individuals with lower overall severity growth much better than those who are severely affected. We also noted significant trends in growth among different mutation types which occur as early as nine months of age based on growth velocity. These differences warrant further investigation, as they will likely necessitate interpretation of the effect of interventions on growth using a mutation-adjusted growth reference.

Table 7-1. Participants by diagnosis (frequency).

	Frequency	Percent
Classic	440	86.3
Atypical	54	10.6
Not Rett Syndrome	16	3.1
Total	510	100.0

Table 7-2. Measurements by diagnosis (frequency).

Simplified Phenotype	Weight	Length	FOC	BMI
Classic	4,716	3,289	2,913	3,257
Atypical	449	355	307	351
Not Rett Syndrome	147	101	118	101
Total	5,312	3,745	3,338	3,709

Table 7-3. Observations based on specific mutations (frequency).

	Weight	Length	FOC	BMI
R168X	543	416	386	414
R255X	494	335	284	331
R306C	498	338	325	337
R294X	419	312	241	311
T158M	474	395	397	391
R133C	325	189	180	187
R270X	324	169	152	167
R106W	123	78	67	77
None	519	358	320	355
Large Deletion	267	159	133	159
Deletion After aa 930	374	266	221	262
Deletion Before aa 930	149	124	122	121
Insertion After aa 930	52	41	29	40
Insertion Before aa 930	92	70	63	68
Deletion Unspecified	63	47	57	47
Total	4716	3297	2977	3267

aa = amino acid number

Table 7-4. Observations based on mutation type (frequency).

	Weight	Length	FOC	BMI
Missense	1622	1135	1075	1126
Nonsense	2162	1490	1290	1480
Frameshift	659	490	451	481
Silent	2	2	2	2
Inframe Variation	82	68	46	68
Splice Site	19	20	9	18
None	549	387	352	383
Duplication	16	13	9	13
Polymorphism	54	33	19	33
Unknown	147	107	85	105
Total	5312	3745	3338	3709

Table 7-5. Observations based on mutation site (frequency).

	Weight	Length	FOC	BMI
None	549	387	352	383
N-terminal Region	39	37	39	35
Methyl Binding Domain	1134	822	789	813
Interdomain Region	567	435	401	433
Transcription Repressor Domain	1612	1114	968	1106
Nuclear Localization Signal	332	210	200	208
C-terminal Region	489	342	266	337
3' Untranslated Region	16	13	9	13
Other	5	5	7	5
Exon 1	50	38	31	37
Exon 4	60	53	50	53
Unknown	459	289	226	286
Total	5312	3745	3338	3709

Table 7-6. Overall observations at specific age intervals (frequency).

Age	Weight	Length	FOC	BMI
0-1 mo	416	331	278	327
2-3 mo	220	180	170	180
4-5 mo	158	132	139	128
6-8 mo	221	166	152	163
9-11 mo	175	125	115	125
12-14 mo	119	95	98	94
15-17 mo	172	121	112	120
18-23 mo	315	201	198	199
2 y	389	252	217	247
3 y	482	311	286	306
4 y	403	260	210	258
5 y	332	250	208	248
6 y	254	197	175	196
7 y	230	157	134	155
8 y	215	149	120	148
9 y	182	123	102	121
10 y	170	130	111	129
11 y	136	100	86	100
12 y	97	67	67	67
13 y	107	72	71	73
14 y	88	50	53	50
15 y	57	36	38	36
16 y	53	40	37	40
17 y	40	29	25	29
18 y	54	35	29	35
19 y	31	24	21	24
20 y	30	23	19	22
21 y	32	25	18	25
22 y	30	24	16	24
23 y	56	22	16	22
24 y	32	12	12	12
25 y	16	6	5	6
Total	5312	3745	3338	3709

Table 7-7. Classic RTT observations at specific age intervals (frequency).

Age	Weight	Length	FOC	BMI
0-1 mo	369	293	246	290
2-3 mo	194	157	146	157
4-5 mo	140	117	123	113
6-8 mo	191	139	125	136
9-11 mo	147	105	96	105
12-14 mo	107	83	82	82
15-17 mo	152	104	102	103
18-23 mo	277	179	174	177
2 y	352	235	203	230
3 y	447	287	261	282
4 y	364	234	192	233
5 y	292	223	187	221
6 y	226	171	152	170
7 y	207	142	122	140
8 y	187	128	103	128
9 y	156	100	81	99
10 y	145	110	91	109
11 y	118	86	72	86
12 y	82	57	53	57
13 y	87	55	56	56
14 y	80	43	43	43
15 y	51	30	32	30
16 y	45	32	29	32
17 y	36	25	21	25
18 y	50	31	25	31
19 y	30	23	20	23
20 y	29	22	18	21
21 y	28	21	14	21
22 y	27	21	15	21
23 y	54	20	14	20
24 y	31	11	11	11
25 y	15	5	4	5
Total	4716	3289	2913	3257

Table 7-8. Classic RTT. Birth weight, length FOC compared to CDC reference.

	N	Mean	SD
Classic Wt (kg)	200	3.27	0.51
Atypical Wt (kg)	18	3.36	0.43
Not Rett Syndrome Wt (kg)	7	3.28	0.21
CDC Wt (kg)		3.40	0.48
Classic Ht (cm)	174	50.31	3.04
Atypical Ht (cm)	12	51.67	2.59
Not Rett Syndrome Ht (cm)	7	50.62	2.00
CDC Ht (cm)		49.29	2.47
Classic FOC (cm)	120	34.25	1.64
Atypical FOC (cm)	8	34.43	1.71
Not Rett Syndrome (cm)	5	33.78	1.10
CDC FOC (cm)		34.71	1.63

Table 7-9. Minimum, maximum and mean weight, height, FOC and BMI in classic RTT.

	Age (years)	Mutation	Reference		
			Value	Z-score	RTT Z-score
Maximum Wt (kg)*	17.65	R294X	82.70	1.74	2.78
Mean Wt (kg)	19.00		41.50	-2.65	0.07
Minimum Wt (kg)	21.62	R270X	24.00	-11.34	-2.29
Maximum Ht (cm)	20.51	R294X	162.50	-0.13	2.16
Mean Ht (cm)	19.00		142.60	-3.17	-0.12
Minimum Ht (cm)	18.34	Early Deletion	109.80	-8.15	-2.79
Maximum FOC (cm)*	17.65	R294X	56.00	0.35	2.17
Mean FOC (cm)	19.00		51.20	-3.12	-0.01
Minimum FOC (cm)	23.58	None	47.50	-5.78	-1.87
Maximum BMI (kg/m <sup>2</sup> )*	17.65	R294X	32.96	1.97	1.90
Mean BMI (kg/m <sup>2</sup> )	19.00		20.45	-0.37	0.14
Minimum BMI (kg/m <sup>2</sup> )	24.59	Unknown	12.56	-7.31	-1.88

References used: CDC (Wt, Ht, BMI), British (FOC). \*Same individual

Table 7-10. Differences between calculated Z-scores and ranked percentiles in different age groups, including difference between mean and 50th percentile, maximum and minimum difference across percentiles (expressed as absolute values and percentages of the mean).

	Age (y)	Discrepancy at Mean	Discrepancy % at Mean	Abs Min Discrepancy	Abs Max Discrepancy	% Min Discrep.	% Max Discrep.
Weight in kg or %kg	0	0.03	0.93	0.01	0.25	0.20	11.61
	2	0.29	2.89	0.05	0.85	0.45	6.29
	3	0.01	0.08	0.01	0.93	0.08	5.21
	5	0.15	1.08	0.15	2.41	1.08	11.62
	10	0.51	2.34	0.04	4.05	0.23	8.18
	21	3.25	9.80	1.21	8.59	4.12	10.66
Height in cm or %cm	0	0.31	0.64	0.08	1.47	0.17	3.44
	2	0.56	0.69	0.16	1.79	0.18	1.93
	3	0.57	0.65	0.01	1.69	0.01	1.70
	5	0.53	0.54	0.02	2.10	0.02	1.89
	10	0.33	0.27	0.25	5.39	0.18	5.16
	21	0.78	0.57	0.18	10.57	0.14	9.63
FOC in cm or %cm	0	0.31	0.93	0.00	1.14	0.01	3.79
	2	0.01	0.02	0.01	0.52	0.02	1.09
	3	0.12	0.27	0.02	0.75	0.04	1.65
	5	0.25	0.52	0.04	0.38	0.09	0.86
	10	0.22	0.44	0.08	0.86	0.15	1.92
	21	0.09	0.18	0.09	0.43	0.18	0.91
BMI in kg/m <sup>2</sup> or %kg/m <sup>2</sup>	0	0.06	0.54	0.06	0.77	0.53	4.64
	2	0.08	0.55	0.08	0.89	0.45	4.90
	3	0.05	0.32	0.00	0.72	0.00	3.93
	5	0.20	1.51	0.11	1.07	0.76	5.77
	10	0.26	1.91	0.07	2.23	0.44	10.30
	21	0.56	3.21	0.02	0.82	0.14	3.42

Table 7-11. Classic RTT growth. Age at which mean and SD significantly different from three standard references.

	Based on Slope	CDC	WHO	British
Weight	6 mo	14 mo	25 mo	
Height	17 mo	21 mo	18 mo	
FOC	1.5 - 12 mo	15 mo	15 mo	Birth
BMI	5 mo	Never	Never	

Table 7-12. Comparison of mean and SD among different subgroups, revealing age at which differences in curves appear.

Comparison	Measure	Based on Slope	Based on t-test
Classic vs Atypical	Weight	21 mo	Never
	Height	18 mo	Never
	FOC	2 y	Never
	BMI	Birth	Never
	Weight	18 mo	43 mo
Classic vs Non-RTT	Height	18 mo	31 mo
	FOC	24 mo	46 mo
	BMI	24 mo	49 mo
Before 1997 vs After 1997	Weight	Never	Never
	Height	Never	Never
	FOC	Never	Never
	BMI	Never	Never
Mild CSS vs Severe CSS	Weight	15 mo	Never
	Height	5 y	7.25 y
	FOC	6 mo	1 y
	BMI	8 y	Never
	Weight	6 y	Never
Mild MBA vs Severe MBA	Height	5 y	5.5 y
	FOC	8 mo	2y
	BMI	11 y	Never

Table 7-13. Type of mutation in classic RTT and adult Weight, Height, and FOC.

	Mutation Type	N	Mean	SD
Weight (kg)	Missense	25	41.54	11.71
	Nonsense	35	43.89	11.46
	Frameshift	15	43.59	15.90
Height (cm)	Missense	22	141.26	6.49
	Nonsense	35	145.52	10.35
	Frameshift	15	137.52	13.76
FOC (cm)	Missense	21	51.41	1.78
	Nonsense	34	51.58	1.77
	Frameshift	15	52.26	2.79

Table 7-14. Site of mutation in classic RTT and adult weight, height, and FOC.

		N	Mean	SD
Wt (kg)	Methyl Binding Domain	23	41.8	11.6
	Interdomain Region	8	39.4	5.9
	Transcription Repressor Domain	20	47.2	14.1
	Nuclear Localization Signal	6	38.3	12.0
	C-terminal Region	10	49.8	15.2
Length/Ht (cm)	Methyl Binding Domain	23	141.8	6.8
	Interdomain Region	8	143.1	9.9
	Transcription Repressor Domain	18	147.9	10.0
	Nuclear Localization Signal	6	136.3	14.8
	C-terminal Region	9	146.0	11.2
FOC (cm)	Methyl Binding Domain	21	51.5	1.8
	Interdomain Region	7	51.3	1.7
	Transcription Repressor Domain	19	52.1	2.1
	Nuclear Localization Signal	6	51.6	1.5
	C-terminal Region	9	53.2	3.4

Table 7-15. Specific mutations in classic RTT and adult weight and height.

	Specific Mutation	N	Mean	SD
Wt (kg)	R168X	7	40.6	5.2
	R255X	5	43.7	11.8
	R306C	3	38.9	10.5
	R294X	8	50.9	16.4
	T158M	12	37.4	8.4
	R133C	6	51.7	15.6
	R270X	6	43.1	11.4
	R106W	1	41.8	.
	None	12	40.7	9.6
	Large Deletion	8	43.3	8.4
	Deletion After aa 930	6	52.7	18.5
	Deletion Before aa 930	4	43.1	19.2
	Insertion after aa 930	3	43.1	8.9
	Deletion Unspecified	3	33.4	9.4
	Length/Ht (cm)	R168X	7	142.1
R255X		5	145.2	5.6
R306C		1	149.1	.
R294X		8	152.9	10.6
T158M		12	141.2	4.3
R133C		6	143.7	7.1
R270X		6	142.1	7.3
R106W		1	124.5	.
None		11	146.0	11.4
Large Deletion		8	143.5	13.0
Deletion After aa 930		6	152.0	5.1
Deletion Before aa 930		4	128.9	17.6
Insertion after aa 930		3	134.1	11.0
Deletion Unspecified		3	130.3	2.6

aa = amino acid #

Table 7-16. Specific mutations in classic RTT and adult FOC, and BMI.

	Specific Mutation	N	Mean	SD
FOC (cm)	R168X	6	51.6	1.5
	R255X	5	50.9	2.0
	R306C	2	53.5	0.0
	R294X	8	52.6	2.2
	T158M	11	51.4	1.7
	R133C	6	52.0	1.8
	R270X	6	51.9	1.3
	R106W	0		
	None	10	51.5	2.4
	Large Deletion	8	51.0	1.4
	Deletion After aa 930	6	54.7	3.0
	Deletion Before aa 930	4	51.8	2.6
	Insertion after aa 930	3	50.1	1.7
	Deletion Unspecified	3	51.5	0.8
	BMI	R168X	7	20.2
R255X		5	20.5	4.1
R306C		1	22.8	
R294X		8	21.6	5.8
T158M		12	18.8	4.1
R133C		6	25.1	7.8
R270X		6	21.1	4.8
R106W		1	27.0	
None		11	19.1	3.4
Large Deletion		8	21.1	3.8
Deletion After aa 930		6	22.7	7.1
Deletion Before aa 930		4	24.8	4.4
Insertion after aa 930		3	23.9	3.8
Deletion Unspecified		3	19.8	6.3

aa = amino acid #

Table 7-17. Ambulation at most recent visit compared to adult weight, height, and FOC in classic RTT.

		N	Mean	SD
Weight (kg)	Acquired < 18 months / Apraxic gait	28	49.57	13.32
	18 months < walks alone <30 months	10	47.81	15.45
	>30 months walks alone	6	39.10	9.60
	>50 months walks with help	8	40.12	7.15
	Lost	11	36.36	4.73
	Never acquired	19	35.63	7.80
Height (cm)	Acquired < 18 months / Apraxic gait	27	148.85	8.28
	18 months < walks alone <30 months	10	147.76	10.20
	>30 months walks alone	6	141.99	8.52
	>50 months walks with help	8	146.58	5.32
	Lost	10	137.37	9.54
	Never acquired	18	134.46	11.18
FOC (cm)	Acquired < 18 months / Apraxic gait	26	52.74	1.94
	18 months < walks alone <30 months	9	52.27	1.96
	>30 months walks alone	6	51.32	2.02
	>50 months walks with help	8	51.26	2.02
	Lost	10	50.75	1.49
	Never acquired	18	50.16	1.46

Table 7-18. Age at regression compared to adult weight, length, FOC and BMI in classic RTT

		N	Mean	SD
Weight (kg)	>30 months	20	45.2	9.7
	18 - 30 months	39	42.0	13.4
	12 - 18 months	17	41.8	13.7
	6 - 12 months	1	33.9	.
	< 6 months	5	42.3	8.2
Height (cm)	>30 months	19	148.2	8.6
	18 - 30 months	38	141.6	11.9
	12 - 18 months	16	142.8	10.0
	6 - 12 months	1	145.9	.
	< 6 months	5	137.1	8.7
FOC (cm)	>30 months	19	52.5	1.9
	18 - 30 months	36	51.6	2.0
	12 - 18 months	16	50.6	2.0
	6 - 12 months	1	51.3	.
	< 6 months	5	50.5	0.8
BMI	>30 months	19	20.7	3.5
	18 - 30 months	38	20.6	5.2
	12 - 18 months	16	20.6	5.9
	6 - 12 months	1	15.9	.
	< 6 months	5	22.4	2.9

Table 7-19. Purposeful hand use at most recent visit compared to adult weight and height in classic RTT.

		N	Mean	SD
Wt (kg)	Acquired and conserved	3	47.3	3.8
	Holding acquired on time: 6 to 8mo (partially conserved)	27	48.6	15.5
	Holding acquired late: >10mo (partially conserved)	11	42.3	13.0
	Holding acquired and lost	40	38.3	7.9
	Never acquired hand use	1	43.7	.
HC (cm)	Acquired and conserved	3	53.4	0.6
	Holding acquired on time: 6 to 8mo (partially conserved)	25	52.4	2.1
	Holding acquired late: >10mo (partially conserved)	9	51.4	2.0
	Holding acquired and lost	39	50.9	1.9
	Never acquired hand use	1	50.5	.

Table 7-20. Ability to speak at most recent visit compared to adult weight, height, FOC, and BMI in Classic RTT.

		N	Mean	SD
Weight (kg)	Preserved, Contextual	1	48.0	.
	Short phrases only	2	53.3	15.7
	Single Words	6	45.3	21.1
	Vocalization, babbling	51	42.9	10.9
	Screaming, no utterance	22	40.1	12.7
	Total	82	42.7	12.3
Height (cm)	Preserved, Contextual	1	151.4	.
	Short phrases only	2	150.6	10.0
	Single Words	6	143.7	12.2
	Vocalization, babbling	48	144.6	10.6
	Screaming, no utterance	22	139.1	10.5
	Total	79	143.2	10.8
FOC (cm)	Preserved, Contextual	1	54.0	.
	Short phrases only	2	54.2	2.6
	Single Words	6	52.2	2.7
	Vocalization, babbling	47	51.5	2.0
	Screaming, no utterance	21	51.1	1.7
	Total	77	51.6	2.0
BMI	Preserved, Contextual	1	20.9	.
	Short phrases only	2	23.2	3.8
	Single Words	6	21.4	7.3
	Vocalization, babbling	48	20.5	4.4
	Screaming, no utterance	22	20.6	5.4
	Total	79	20.7	4.8

Table 7-21. Age at which individual with Classic RTT first sat compared to adult weight, height, FOC, and BMI.

		N	Mean	SD
Weight (kg)	Sits alone, acquired <=8mo	43	45.75	12.87
	Sits with delayed acquisition >8mo	11	48.16	13.89
	Sits with delayed acquisition >18mo	3	33.27	2.48
	Sits with delayed acquisition >30mo	2	44.15	1.34
	Lost sitting	13	35.24	6.66
	Never acquired sitting	10	35.53	7.59
Height (cm)	Sits alone, acquired <=8mo	42	147.29	7.52
	Sits with delayed acquisition >8mo	11	147.80	9.82
	Sits with delayed acquisition >18mo	3	140.17	10.56
	Sits with delayed acquisition >30mo	2	142.90	7.21
	Lost sitting	12	132.33	9.44
	Never acquired sitting	9	134.33	13.61
FOC (cm)	Sits alone, acquired <=8mo	39	52.22	1.93
	Sits with delayed acquisition >8mo	11	52.19	2.20
	Sits with delayed acquisition >18mo	3	51.80	1.73
	Sits with delayed acquisition >30mo	2	49.30	1.98
	Lost sitting	13	50.11	1.55
	Never acquired sitting	9	50.44	1.34
BMI	Sits alone, acquired <=8mo	42	20.95	5.44
	Sits with delayed acquisition >8mo	11	21.73	4.32
	Sits with delayed acquisition >18mo	3	17.01	1.57
	Sits with delayed acquisition >30mo	2	21.74	2.85
	Lost sitting	12	20.23	4.24
	Never acquired sitting	9	19.80	4.34

Table 7-22. Seizure Severity classifications in Classic RTT compared to adult weight, height, FOC, and BMI.

		N	Mean	SD
Weight (kg)	Absent	43	44.03	12.33
	< monthly	14	44.35	10.41
	< weekly to monthly	7	31.17	3.26
	Weekly	7	46.44	20.04
	More than weekly	11	40.08	8.31
Height (cm)	Absent	43	144.67	10.81
	< monthly	13	144.61	7.95
	< weekly to monthly	7	134.60	13.34
	Weekly	6	144.87	13.36
	More than weekly	10	140.28	9.10

Table 7-23. Overall Severity Score (CSS and MBA) classifications in Classic RTT compared to adult weight, height, FOC, and BMI.

		Weight (kg)	Height (cm)	FOC (cm)	BMI
CSS severity	Correlation	-.470	-.613	-.599	.858
	Sig (2-tailed)	.000	.000	.000	.000
	N	57	57	57	57
MBA severity	Correlation	-.455	-.627	-.511	-.151
	Sig (2-tailed)	.000	.000	.000	.230
	N	67	65	63	65

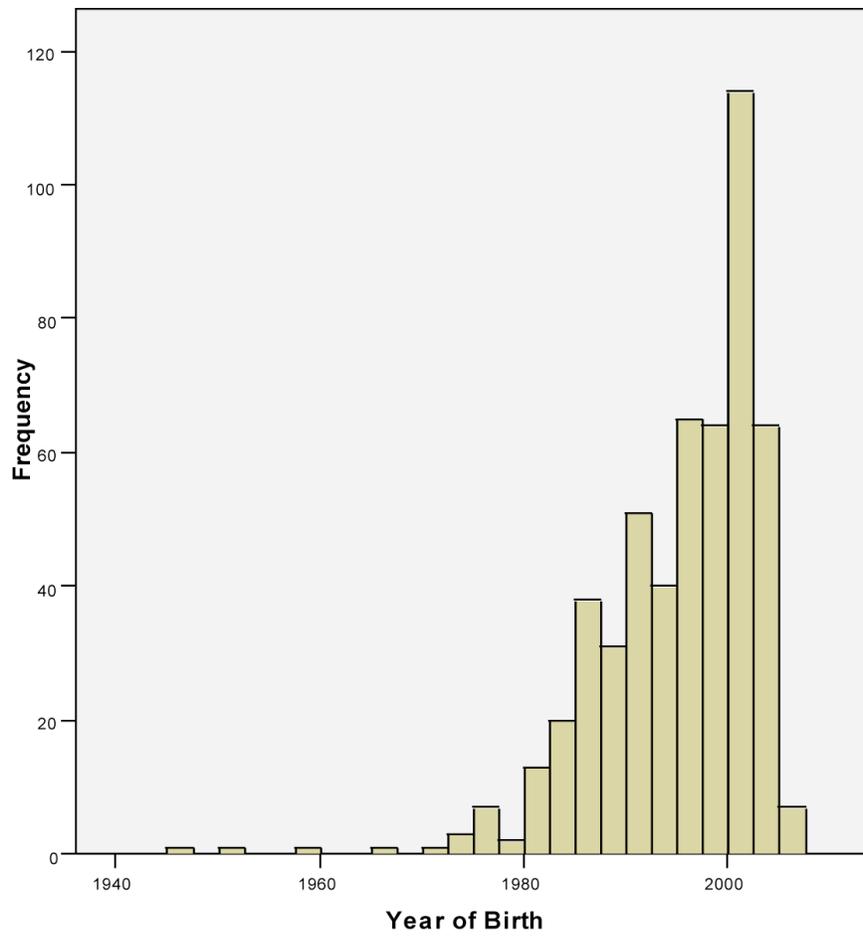


Figure 7-1. Participants' years of birth (frequency).

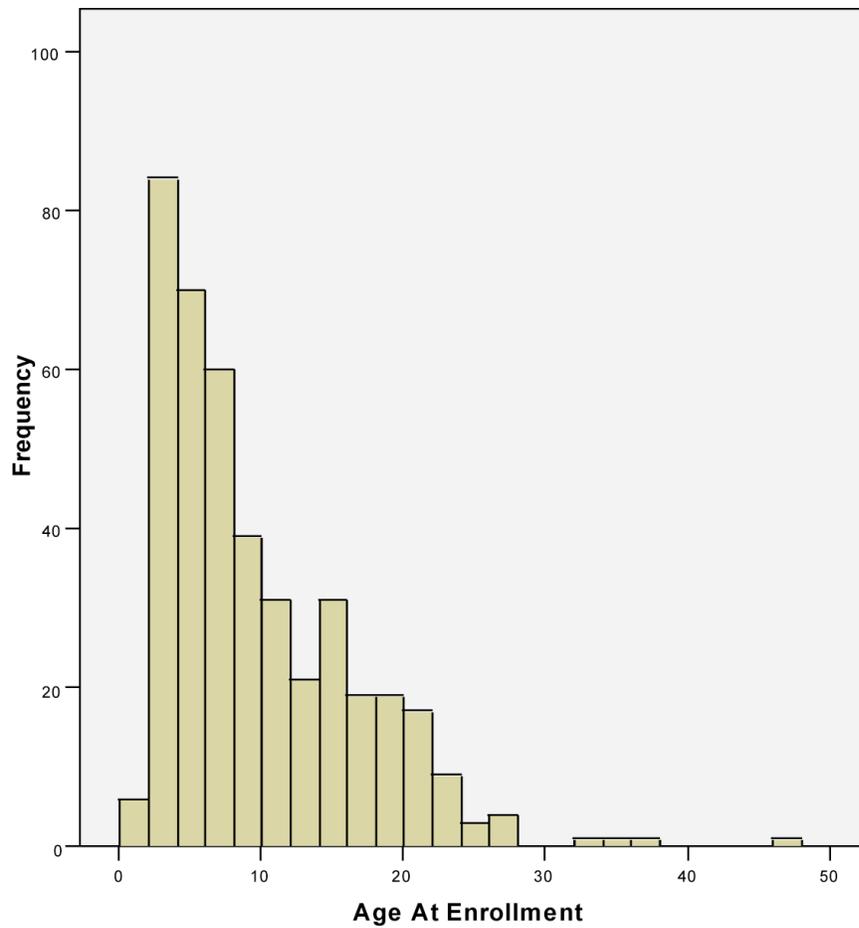


Figure 7-2. Ages at enrollment in RDCRN study (frequency).

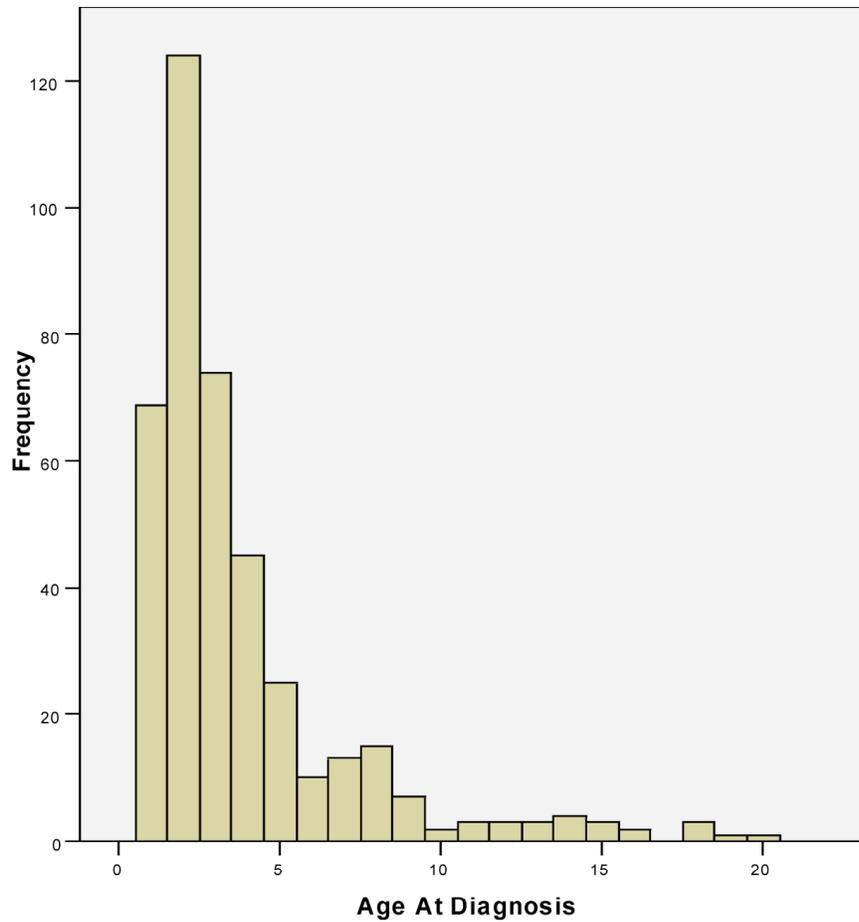


Figure 7-3. Ages at diagnosis with Rett Syndrome (frequency).

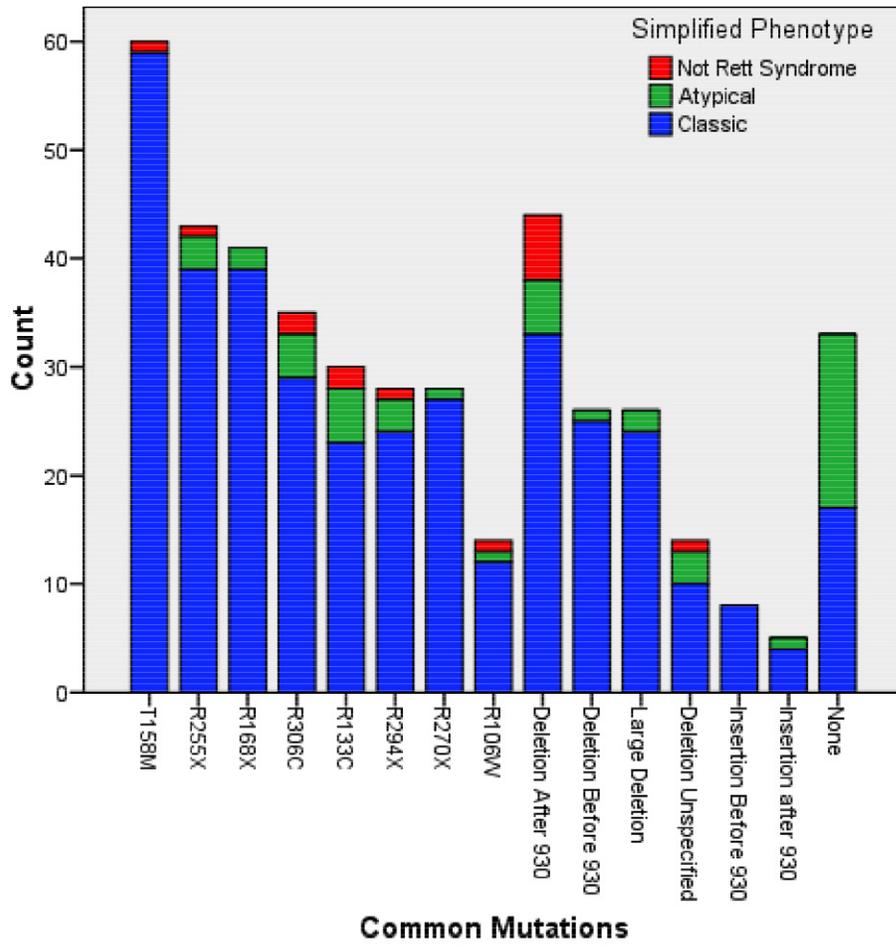


Figure 7-4. Simplified mutation types by diagnosis (frequency).

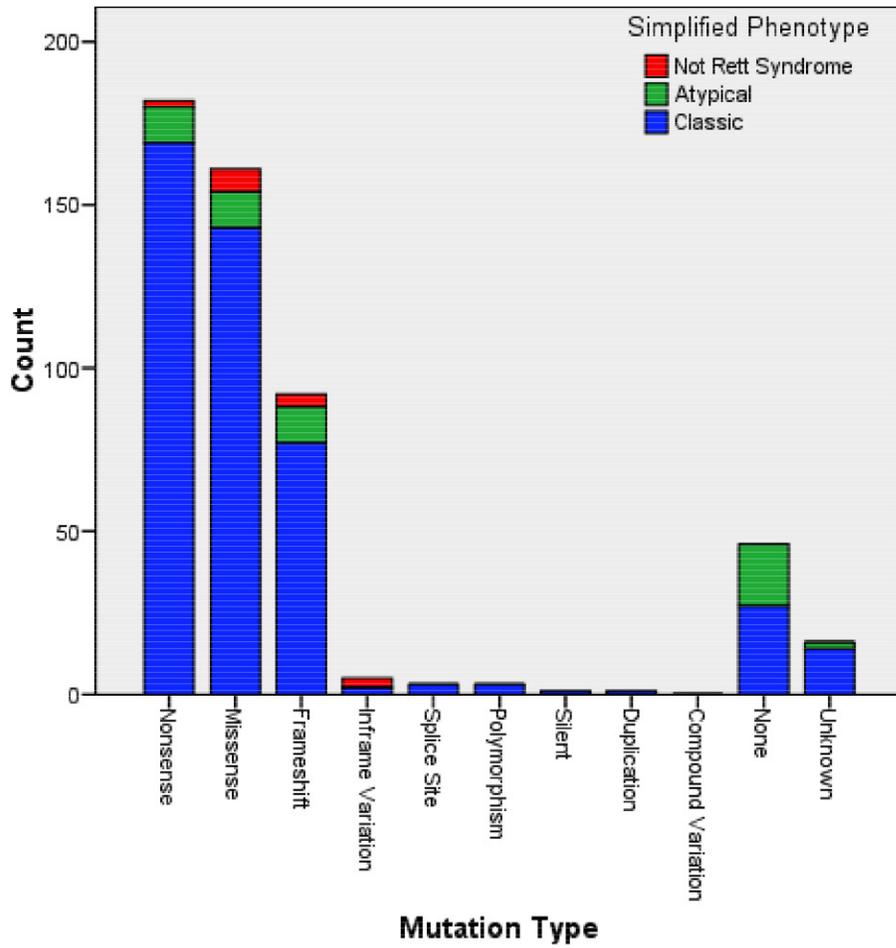


Figure 7-5. Specific mutation types by diagnosis (frequency).

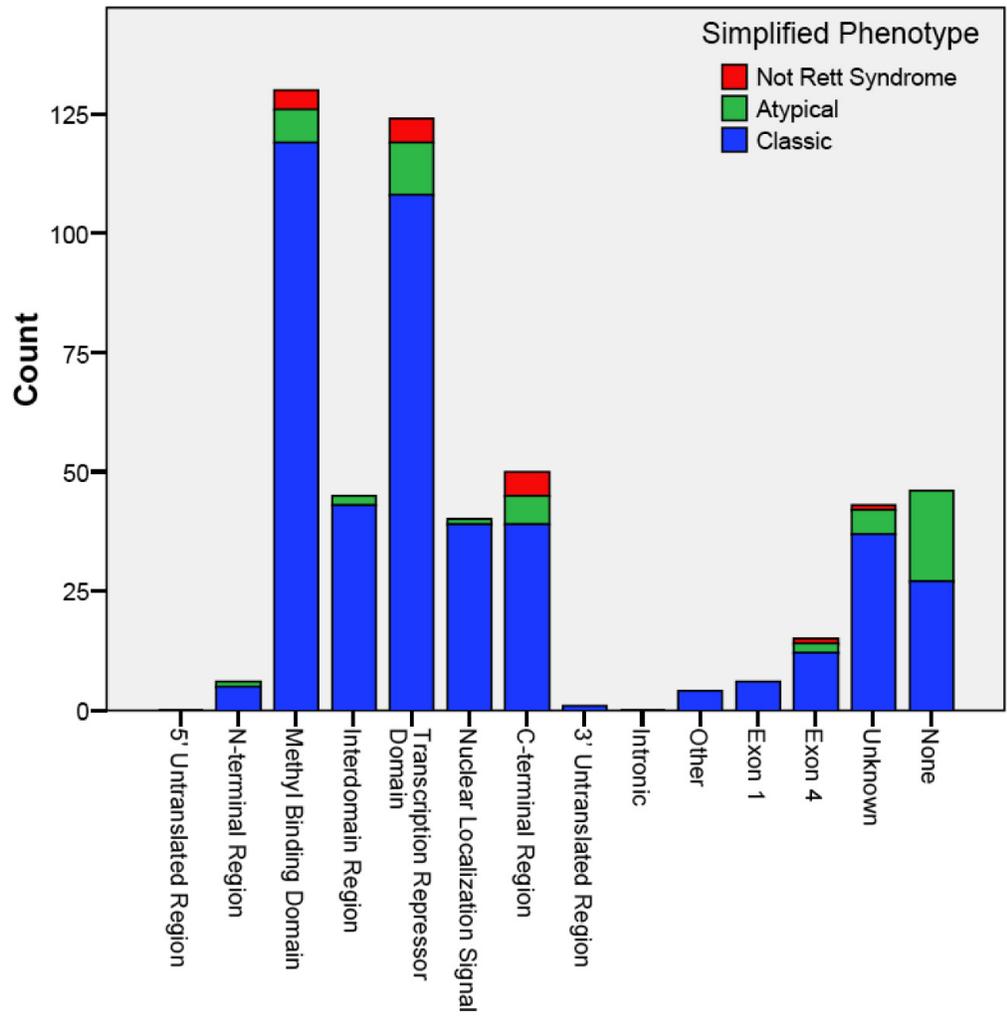


Figure 7-6. Mutation sites on MECP2 by diagnosis (frequency).

## CHAPTER 8 FUTURE OF GROWTH STANDARDS FOR SPECIAL POPULATIONS

### **Current Successes in Growth Standard Design**

Although growth chart design has developed rapidly over the past two decades, many excellent designs have yet to be fully realized. The concept of applying one or more covariates has been effective in certain scenarios. Height corrected for mid-parental height using a regression equation has met some success (Lenko et al. 1988; Lyon et al. 1985). Age can be adjusted for degree of skeletal maturity by measuring bone age, a measurement made by examining radiographic evidence of epiphyseal formation and comparing to standard normal references. In many cases bone age is a much more accurate indicator of skeletal age and progression to final height, and can be used not only for prognosis of final height, but also to detect abnormalities of growth beyond normal skeletal maturation. In one example, two individuals, one with Turner syndrome and one with idiopathic GHD, begin with identical height-for-bone-age measurements. However, as the girl with Turner syndrome's bone age advances faster relative to the individual with GHD, her height-for-bone-age falls below normal on the height-for-bone age chart. Alternatively, on the standard CDC height-for-age chart, the girl with GHD is consistently shorter than the girl with Turner syndrome. The GHD patient's bone age is markedly delayed, but significantly, her height-for-bone-age is relatively normal. This concept has important implications in achieving a desired final height, since early skeletal maturation is associated with short adult height.

### **Applying Growth Standard Principles to Rett Syndrome**

The concept of a growth standard in rare diseases is not only attractive from a counseling standpoint, but may also be necessary for growth charts to be used effectively in research. Researchers have hypothesized that many factors contribute to the growth failure in RTT. Many

of these factors, such as nutrition, seizure disorders, anxiety, constipation, GER, and orthopedic issues, can all be modified. Therefore it is possible to compare RTT individuals who have had more aggressive treatment with respect to these factors to those who have not. Should significant differences exist, growth standards will represent both a benchmark for clinicians to strive to, and a hurdle for researchers to overcome.

Many questions regarding growth in RTT remain unanswered. The longitudinal component of this study was preliminary, and future efforts will focus on assessing correlations within RTT. Since adjustments for growth, including mid-parental height, birth measurements and sibling measurements are only valid for normal healthy individuals, conditional growth charts for rare diseases cannot be produced using the current formulas for estimation (Zachmann et al. 1978). These formulas have proven accurate in certain diseases and grossly inaccurate in others. However, longitudinal data from birth to adulthood will allow researchers to create new models for conditional growth standards that incorporate these covariates. Since cognitive level is so difficult to assess in RTT, growth remains one of the most important outcome measures of intervention, and targeted therapies will certainly require a conditional growth standard for comparison.

### **Future of Statistical Modeling of Growth Charts**

Unfortunately, as the statistical methodology for smoothing and fitting data to models progresses, the task of evaluating the results against the empirical data becomes more complicated. Buuren et al., comment that Cole has replaced the “black art” of growth chart construction with a “black box” of statistical smoothing (van Buuren and Fredriks 2001). The process of generating curves for size or growth is still very subjective. Although the methodology proposed in this study uses the best diagnostic tools to assess fidelity to empirical

data, these tools are not truly goodness-of-fit tests. Hopefully the following 20 years will be as rich in statistical improvements on growth modeling as the past 20 have been.

### **Do Disease-Specific Growth Standards Need to be Updated Regularly?**

Just as standards for growth in the normal population must be updated, disease specific growth charts must be reviewed regularly and changes must be incorporated. The population-wide differences seen between the 1977 CDC charts to the 2000 CDC charts, and the disparities among the values found in disease specific charts all deserve attention. Moreover, the discrepancy in standards for FOC is concerning to all who are aware of it. Many opportunities exist for improving the references of the healthy population, in addition to developing disease specific references.

Some of these differences represent secular trends. For example, a dramatic change in the mean and SD of weight occurred between 1977 and 2000. In this case the mean and SD increased and the data became skewed further to the right relative to the 1977 CDC weight values. The researchers chose to exclude the more recent data in favor of the data from the lighter children of the 1960's and 1970's. Another example involves the FOC of infants. Physicians in England used the growth references of Gairdner and Pearson (Gairdner and Pearson 1971) from 1971 until the early 2000s, despite the development of other references in the late 1990s. To emphasize the need to change to the newer references, Savage et al., demonstrated significant secular increases in FOC in the 20 years since the earlier charts were published (Savage et al. 1999). A third example in the Japanese literature involved a comparison of data from over 50 years. Ishikawa et al., found that, although Japanese FOC is smaller than British FOC on average, Japanese FOC was considerably smaller in the 1940 compared to 1980 (Ishikawa et al. 1987). However, a study of growth in Japan from 1990-1994 found only a small secular trend (1 cm maximum) of increasing FOC when compared to studies from 1978 to 1981

(Anzo et al. 2002). Anzo et al., postulate that secular trends in Japanese culture are diminishing, and that Japanese have nearly achieved their maximum genetic growth potential. Therefore, although secular trends must be accounted for, the degree and time interval over which they occur are uncertain.

Just as all of these changes - secular trends, seasonal variation, and conditional variation - occur in the general population, shifts occur in special populations as well. As the RDCRN study continues to collect new data on the current generation of RTT individuals, we will improve this growth reference in size and statistical refinement. As we gather more data on modifiable risk factors we hope to develop the reference into a growth standard. Additionally, we intend to develop models which can simultaneously adjust for multiple conditional variables, including mutation type, racial and ethnic background, mid-parental size, and previous growth velocity, using regression techniques incorporating between-subject and within-subject variation. Hopefully as we refine these techniques, the black box will turn clear, and we will gain further insight into the logic behind the modeling of human growth.

## APPENDIX A MEASUREMENT TECHNIQUES

### **Basic Guidelines for Length Measurements Including Hands, Feet, Head Circumference and Height**

- Always use flexible, nonstretchable tape measure
- Read all measurements at eye level and in good lighting
- Optimally it is best to use two people to position and hold participant and tape measure in correct position and record readings for each measure.
- Parents may also have to help hold, distract or calm the participant
- If exceptions are noted during exam (i.e. presence contractions, a cast, severe scoliosis, etc), try to follow guidelines but note exceptions when recording results.

### **Weight**

Weight will be obtained at each visit in either a standing or sitting position and will be reported in kilograms (kg) to the nearest 0.1 kg. Weight measurements will be plotted on age appropriate standardized forms (National Center for Health Statistics in collaboration with the National Center for Chronic Disease Prevention and Health Promotion).

Equipment: A beam scale with non-detachable weights or an equally accurate electronic scale can be used for participants who can stand. A chair or wheelchair (platform) scale may be used for participants who cannot stand. Prior to measuring, the scale should be calibrated. A standard weight should be used to check scale performance. If a platform or chair scale is not available or if the participant moves excessively, and if the participant is small enough to be held, the participant can be weighed while held in the arms of an adult. The weight of the adult alone would be subtracted from the total weight, thus providing the weight of the participant.

Technique: Regardless of type of scale used, scale should be adjusted to zero. If a pad or sheet is used, zero the scale with these in place. If the participant is in a wheelchair, a tare weight of the chair is obtained after all extra bags and attachments have been removed and subtracted accordingly from the total weight of the participant and the wheelchair. The chair tare weight can be recorded and used at the next visit if the wheelchair equipment does not change. Regardless of

the type of scale, the participant should remove shoes, excess clothing and wet diaper and the weight is read either from the beam or the electronic readout. If there is excessive movement or if two separate measurements do not agree within 5 grams, take a third measurement and record an average of the three. Record the numerical value on data forms and plot on age-appropriate standardized form for weight.

## **Height**

Height will be obtained at each visit and will be reported in centimeters (cm) to the nearest 0.1 cm. Height measurements will be plotted on standardized forms (National Center for Health Statistics in collaboration with the National Center for Chronic Disease Prevention and Health Promotion). Participants who are able to stand should be measured in a standing position. Young participants or ones who are unable to stand will be measured in recumbent position. Participants with conditions such as scoliosis, kyphosis or contractures of legs or feet that prevent measurement by conventional recumbent methods will be measured with modified methods. Position of measurement (either standing, recumbent or calculated) should be noted on the data collection form

**Standing Height Equipment:** A vertical measuring rod, which is at least 175 cm high and capable of measuring to an accuracy of 0.1 cm should be used. A digital readout device is optimal.

**Standing Height Technique:** After removing the shoes, the participant should stand on a flat surface by the scale with feet parallel and with heels, buttocks, shoulders and back of head touching the upright. The head should be comfortably erect. The arms should be hanging loosely at the sides. The headpiece of the measuring device is gently lowered, crushing the hair and making contact with the top of the head. One observer should position the feet and straighten the knees, while a second observer positions the shoulders and holds the head erect, if necessary.

Height is measured at the crown of the head, excluding hair accessories, pony tails or braids.

Repeat the measurement until two measurements agree within 0.5 cm, average these and record the numerical value on data collection forms and plot on standardized form for height.

Recumbent Height Equipment: A firm, flat surface or a recumbent board, and a soft, non stretchable measuring tape will be used.

Recumbent Height Technique: Two people are required to measure length accurately. The participant should be in recumbent position. Person "A" should hold the head with crown against the headboard or a flat surface so that the child is looking straight upward. Make sure the trunk and pelvis are properly aligned. Person "B" should straighten the child's legs and hold the ankles together with the toes pointed directly upward. With a flat surface pressed against the soles of the feet, mark the bottom of the foot and measure from the head to the mark at the foot. Repeat the procedure and measurement until two measurements agree within 0.2 cm (1/4 in).

Technique for calculation of height in the case of severe scoliosis or contractures: When an accurate height or length is not obtainable because of severe scoliosis or contractures, a conversion formula that considers regression between height (HT) and lower leg length (LLL) is used to calculate length. The LLL is equal to - 6.39 plus 0.354 times HT.

## **FOC**

Head circumference will be obtained while the participant is in a seated position and will be reported in centimeters (cm) to the nearest 0.1 cm.

Equipment: A flexible, non-stretching measuring tape will be used.

Technique: Objects such as pins or rubber bands will be removed from the hair. The measurer will stand to the side of the participant and will position the tape just superior to the eyebrows anteriorly and over the external occipital protuberance (inion) posteriorly. The tape should be pulled tightly to compress hair and braids should be removed or avoided.

Measurements should be performed twice or until two measurements agree to within 0.1cm.

These measurements will be plotted on standardized forms for head circumference until age 19 years (Nellhaus, Pediatrics, 41:106, 1968).

### **Hand length**

Hand length of each hand will be measured at each visit until age 15 years. Measurements of each hand will be reported in centimeters to the nearest 0.1 cm.

Equipment: A flexible, non-stretching measuring tape will be used.

Technique: Measurement should begin with the measurement of the palm beginning at the wrist-palm crease and ending at the palm-proximal middle phalanx crease; measurement of the total hand length will begin from wrist-palm crease and ending at the tip of the middle finger. The middle finger length will be calculated by subtracting the palm length from the total hand length. These measurements will be plotted on standardized forms for hand measurement (Feingold and Bossert, Birth Defects, 10[suppl.13], 1974).

### **Foot length**

Foot length measurements of each foot will be obtained at each visit until age 19 years. Measurements of each foot will be reported in centimeters to the nearest 0.1cm.

Equipment: A flexible, non-stretching measuring tape and an appropriate sized clean piece of paper will be used.

Technique: Place the foot on a firm surface while standing, if possible, and marking the end of the heel and of the great toe on a clean, appropriate sized piece of paper. If standing is not possible, the foot will be placed against a firm, flat surface while the participant is seated and the points marked as above. After marking, the distance will be measured in centimeters to the nearest 0.1 cm and plotted on standardized forms for foot measurement (Feingold and Bossert, Birth Defects, 10[suppl.13], 1974).

## **Skin-fold measurements**

Skin-fold measurements of six sites including subscapular, suprailiac, triceps, biceps, thigh and calf will be completed at each visit. Measurements will be reported in millimeters to the nearest 0.1 mm. Specific information on skinfold methods was derived from Heyward and Wagner [in Applied Body Composition Assessment. 2<sup>nd</sup> Edition. Human Kinetics. Champaign, IL. 2004. Chapter 4, pp.49-66].

Basic guidelines for skinfold measurements: Each measurement should be made in duplicate; if readings do not agree within 3 mm for skinfolds, then take a third measurement and record the average of the two closest measurements.

To ensure that measurements are made at the correct location, identify the point first by making a small mark on the skin. Marks should be removed with an alcohol wipe when finished.

Make all measurements on the right side of the body.

All anthropometric measurements should be made consistently by one individual.

Equipment: Lange Calipers adjusted to perfect zero will be used for all skinfold measurements. The calibration/accuracy of the calipers should be checked with calibration blocks on a regular basis. The same calipers should be used for each subsequent measurement of a participant. A washable marker is also needed.

Technique – General guidelines: For all skinfold measurements, the skinfold should be pinched between the thumb and forefinger one cm above (towards the participant's head) the point at which you are going to place the calipers. The width of the skinfold that is enclosed between the fingers is an important factor. The amount that is pinched cannot be standardized for all sites, but in general for “thicker” skinfolds a wider segment of skin must be pinched than for a “thin” skinfold. For a given site, the width of the skin should be minimal but still yielding a well-defined fold. The skinfold must be free of muscle tissue and this necessitates that the underlying

muscles be fully relaxed. Hold the fold firmly between the fingers while the measurement is being made. The depth of the skinfold, at which the jaws of the calipers are placed, should be approximately equal to the thickness of the fold. Most importantly, it must be at a point where the two surfaces of the skinfolds and the contact surfaces of the caliper jaws are approximately parallel to each other. The reading on the dial should be taken when the gross movement of the needle has stopped. This will vary from 5 to 45 seconds. When making replicate measurements at the same site, the skinfold should be completely released between measurements.

**Triceps Skinfold:** Direction of fold – Vertical. Anatomical reference - Acromial process of scapula and olecranon process of ulna. Technique - A vertical fold should be raised midway between the right olecranon and acromion process on the posterior of the upper arm. Position the subject with the upper arm flexed at a 90 degree angle to the forearm. Using a tape measure, determine the distance between the acromion and the tip of the olecranon. Mark this point. The skinfold is picked up one cm above this mark and the measurement is taken on the midpoint.

**Biceps Skinfold:** Direction of fold – Vertical (midline). Anatomical reference – Biceps brachii. Technique - The biceps skinfold is taken at the anterior aspect of the arm over the Biceps brachii and opposite the triceps on a vertical line joining the anterior border of the acromion and the center of the antecubital fossa.

**Subscapular Skinfold:** Direction of fold – Diagonal. Anatomical reference – Inferior angle of scapula. Technique - A diagonal fold, inclined approximately 45 degrees from horizontal, in the natural cleavage of the skin should be picked up at the inferior angle of the scapula. The participant should be sitting or lying comfortably on their left side with the arms relaxed at the sides of the body.

Suprailiac Skinfold: Direction of fold – Oblique. Anatomical reference – Iliac crest.

Technique - The suprailiac skinfold is a diagonal fold measured immediately superior to the iliac crest. The skinfold is taken one cm posteriorly to the midaxillary line and the caliper applied on the midaxillary line. The fold should follow the natural cleavage of the skin.

Thigh Skinfold: Direction of fold – Vertical. Anatomical reference – Inguinal crease and patella. Technique - The thigh skinfold is measured at the midpoint of the thigh. The midpoint is determined by measuring the length from the inguinal crease to mid-patella while the participant lies flat with the leg extended. Grasp the skin one cm from the midpoint of the thigh. The measurement is made approximately one cm distal to the fingers holding the skinfold.

Lateral Calf (Shin) Skinfold: Direction of fold – Vertical. Anatomical reference – Maximal calf circumference. Technique - The right foot is placed on the bed or exam table with the knee flexed at a 45 degree angle. The calf skinfold is measured at the midpoint of the calf. The midpoint is determined by measuring the length of the lower leg from the mid-patella to the ankle crease of the right foot. Grasp the skin one cm from the midpoint of the lower leg. The measurement is made approximately one cm distal to the fingers hold the skinfold.

### **Circumference measurements**

Arm (bicep), thigh and calf circumference measurements will be obtained at each visit and will be reported in centimeters to the nearest 0.1cm. Specific information on circumference measurement methods was derived from Heyward and Wagner [in Applied Body Composition Assessment. 2<sup>nd</sup> Edition. Human Kinetics. Champaign, IL. 2004. Chapter 5, pp.67-85].

Basic guidelines for circumference measurements: Take all circumference measurements on the right side of the body. Carefully identify and measure the anthropometric site. Be meticulous about locating anatomical landmarks used to identify the measurement site. Take a

minimum of three measurements at each site in a rotational order. For body segments with small girths such as the calf and arm, take three measurements within 0.2 cm.

Equipment: A flexible, non-stretching measuring tape and a washable marker will be used.

Technique: Arm circumference: Anatomical reference - Acromion process of scapula and olecranon process of ulna. Position - Perpendicular to the long axis of arm. Technique - With participant's arms hanging freely at sides and palms facing thighs, apply tape snugly around the arm at the level midway between the acromion process and olecranon process of the ulna.

Thigh circumference: Anatomical reference - Inguinal crease and proximal border of patella. Position - With the participant's knee flexed 90 degrees (right foot on the exam table), apply tape at level midway between inguinal crease and proximal border of the patella.

Calf (shin) circumference: Anatomical reference - maximum girth of calf muscle. Position - Perpendicular to long axis of leg. Technique - With the participant sitting on the end of the table and legs hanging freely, apply tape horizontally around the maximum girth of the calf.

### **Leg length measurement**

Leg length measurement of the lower leg length will be obtained at each visit and will be reported in centimeters to the nearest 0.1 cm.

Equipment: A flexible, non-stretching measuring tape and a washable marker will be used. If available, a Ross™ knee height caliper can be used for the lower leg measurement.

Technique: Lower Leg length - If Ross™ calipers are not available, place participant in recumbent position on the exam table and locate the top of the patella. Place a removable mark on the table at the point adjacent to the patella. Flex foot to a 90 degree angle. Place a removable mark on the table adjacent to the bottom of the heel. Measure the distance between marks for each leg and record. If Ross™ calipers are available, place the participant in the recumbent position on the exam table. Bend both the knee and the ankle to a 90 degree angle. Open the

caliper and place the fixed blade under the heel. Press the sliding blade down against the thigh about 2 inches behind the patella. The shaft of the caliper should be in line with the tibia in the lower leg and over the lateral malleolus. To hold the measurement, push the locking lever away from the blades. Read the measurement through the viewing window. Release the locking lever by pushing it towards the caliper blades. Repeat the process twice until measurements within 0.5 cm are obtained. Repeat the technique for the other leg.

### **Body Mass Index**

Body Mass Index will be calculated at each visit and reported in  $\text{kg}/\text{m}^2$ .

## APPENDIX B ANTHROPOMETRIC MEASUREMENT RELIABILITY

### Abstract

**Introduction:** Studies of growth seldom report the reliability of measurements, and those studies that report reliability vary widely in analytical technique. While reference charts lacking estimates of reliability may be adequate for clinical practice, such charts are not satisfactory as research tools. This study investigates two topics: (1) inter- and intraobserver measurement error in an anthropometric study of children with Rett syndrome (RTT), and (2) the utility of specific estimators of measurement precision in anthropometry. **Methods:** A sample of 20 female participants with RTT were selected from a larger population participating in a study on growth. Participants included children from three to fourteen years of age from three study sites with a diagnosis of classic RTT confirmed by genetic sequencing who ranged from three to fourteen years of age. Children with severe scoliosis, kyphosis, or contractures were excluded. Repeated measurements of height (HT), weight (WT), and fronto-occipital head circumference (FOC) were performed on all participants by three trained observers who were blinded to each other's measurements. However, due to time constraints, intraobserver repeated measurements were only performed for FOC. Measurements were compared using three common estimators of measurement precision: the intraclass correlation coefficient (ICC), the technical error of measurement (TEM) with corresponding coefficient of reliability, and the Pearson product-moment correlation coefficient (PMCC). **Results:** Interobserver measurement analysis using the three techniques revealed that the TEM (HT = 0.65 cm, WT = 0.10 kg, FOC = 0.09 cm) and its CR (HT = 0.999, WT = 1.000, FOC = 0.999) yielded the highest values, followed by the PMCC (HT = .999, WT = 1.000, FOC = .998), and then the ICC (HT = .998, WT = 1.000, FOC = .997). Intraobserver analysis using the TEM (FOC = 0.13 cm) and CR (FOC = 0.998) also yielded the

highest mean precision, followed by the PMCC (FOC = .997), and then the ICC (FOC = .992). The TEM, PMCC and ICC results all fell into the “excellent” range based on standard interpretation, however they each have unique attributes: the TEM can be converted back to the original units of measurement, and the ICC accounts for all degrees of variation within and among the measurements. **Conclusions:** This study is one of the first to assess the reliability of anthropometric data in a rare disease. The ICC and the TEM provide the most meaningful measures of reliability. Moreover, they should be used in conjunction to provide a comprehensive picture of both the magnitude (TEM) and source (ICC) of error. This study does not support the use of the PMCC, as this test does not account for the magnitude of error. This methodology combining the ICC and TEM, when employed at the conception of anthropometric studies, will help validate reference charts as suitable research tools in rare diseases.

### **Introduction**

Growth reference charts are a crucial component of health maintenance in children. However, anyone who has examined growth charts in detail has noted that they are peppered with values that simply don't fit – values which lie outside expected percentiles, and are usually labeled as spurious. This observation provokes two questions concerning validity. First, are data recorded in these charts accurate or reproducible? Second, are the conclusions that physicians draw from these charts valid? To answer these questions we must first consider the potential sources of error. Once we have delineated the sources, we can choose a methodology for measuring the degree of error. Finally, we can determine if the degree of error is significant enough to invalidate any conclusions based on the data.

This concept of validating measurements applies to both individual patient records and the reference charts themselves. Reference charts are composed of data from individuals, and that data is collected with a degree of precision which is typically not measured. Reliability testing

addresses the issue of “common practice” to which the unknown errors in accuracy and precision are usually ascribed in other studies (Hauffa et al. 2000). Therefore, a methodology for measuring precision would not only be useful to determine error within a specific patient’s data or an individual clinical setting, but also to measure the degree of error present within reference charts. The process of testing the reproducibility of measurements, (i.e., reliability) can be divided into two categories. Interobserver reliability is defined as the precision of measurements taken by many observers on the same individual, while intraobserver reliability is the consistency of repeated measurements taken by the same observer on one individual.

In this age of electronic medical records researchers have access to an abundance of anthropometric information, including large samples of participants with rare diseases. Growth charts for individual diseases are in demand, both for health maintenance and as a research tool to quantify the effects of novel treatments. As vast amounts of data accumulate, researchers should seek to establish the reliability of the source measurements prior to developing reference charts from these data. Fortunately, statisticians have developed several techniques over the past 60 years suited to assessing reliability of observations. Some examples are the Technical Error of Measurement (TEM) of Dahlberg (1944), Cohen’s Kappa (1960), the Intraclass Correlation Coefficient (ICC) of Shrout and Fleiss (1977), and the Product-Moment Correlation Coefficient (PMCC) of Pearson. Surprisingly, very few growth reference chart developers employ these techniques to assess reliability. Most growth studies consist of retrospective analyses, but to evaluate reliability a study must be designed prospectively. The 2000 CDC growth charts, considered the definitive reference for growth, were obtained from seven sources of varying reliability. However, only one of these sources has been assessed for internal reliability. Several

smaller studies of growth have incorporated prospective tests of reliability, but their analytic techniques and interpretations of the results vary widely.

The two aims of this study are to examine measurement reliability in a rare disease and to compare techniques for assessing the reliability of anthropometric measurements to determine which is best suited for rare diseases. To accomplish these goals, we first performed redundant measurements on a cross-section of patients with RTT, a rare neurodevelopmental disorder. We used three common statistical techniques (the TEM, the ICC and the PMCC) to test both our first hypothesis, that all observers' measurements would be consistent with each other, and our second hypothesis, that each observer's repeated measurements would be internally consistent. To address our second aim, we compared and contrasted these three techniques for computing reliability scores against the minimum and maximum error, the coefficient of reliability (derived from the TEM), and the mean absolute deviation, and have suggested a standard method for assessing the reliability of anthropometric measurements in rare diseases.

## **Methods**

### **Participants**

We selected a convenience sample of twenty participants already involved in an ongoing study of growth in RTT. The parent study is a multicenter observational study of the natural history of RTT which includes anthropometric measurements such as height, weight, and FOC. The population was enrolled during the period of July to September 2007 from five sites in five different states: Oakland, CA; Chicago, IL; New Brunswick, NJ; Boca Raton, FL; and Birmingham, AL. All participants who met clinical criteria for classic RTT were included. Although not a requirement for inclusion, all participants in this sample had a genetic mutation consistent with RTT. Children whose height could not be measured due to severe scoliosis, kyphosis, or contractures were excluded. Growth in such children is typically followed by

calculated measurements, the reliability of which is beyond the scope of this study. Informed consent and institutional review board approval were obtained at each research site through the natural history study.

## **Procedure**

Participants traveled to a local study site to participate. Three observers recorded height, weight, and FOC using standardized techniques as described in the natural history study protocol. Observers were blinded to each other's measurements. In most cases two observers worked together, one acting as an assistant to stabilize the participant, but unable to see the measurements taken by the first observer. The observer and assistant then switched roles for the second measurement. Each observer measured each patient's FOC twice, prior to and following the physical exam.

Height was measured in a standing position for all participants over two years of age who were able to stand, and in a recumbent position for all participants less than two years of age and those unable to stand. In both cases the participant's hair clips, shoes and ankle-foot orthoses (AFOs) were removed before measuring. Standing height was measured using a vertical measuring rod with an accuracy of 0.1 centimeters (cm), or a calibrated digital device when available. The participant stood on a flat surface while an assistant assured that the feet were parallel, knees were locked, and heels buttocks, shoulders, and back of head were touching the upright rod. The observer, ensuring that the head was erect, lowered the perpendicular headpiece of the device to crush the hair and make contact with the crown of the head, and recorded height to the nearest 0.1 cm. Recumbent height was measured with the participant lying supine on a firm, flat table. An assistant standing to the side of the participant assured that the trunk and pelvis were in neutral position, and then held the head with the crown against a headboard perpendicular to the table. The observer straightened the participant's legs, brought the ankles

together and slid an adjustable footboard flat against the soles of the feet. The observer then made a mark on the table at the perpendicular intersection of the footboard, and measured the distance between the headboard and mark to the nearest 0.1 cm using a flexible, non-stretching measuring tape.

Weight was obtained with the participant sitting or standing, and reported to the nearest 0.1 kilograms (kg). A variety of automatically averaging electronic scales were used for measuring weight, depending on what was available at the site. All were accurate to within 0.1 kg. Excessive clothing, shoes and AFOs were removed before measuring, and the scale was zeroed prior to each measurement with a chair or wheelchair on the platform when indicated.

Head circumference was measured using a flexible, non-stretching measuring tape while the participant was seated. Braids and any objects, such as pins or clips, were removed. The observer sat next to the patient and positioned the tape just superior to the eyebrows anteriorly, and over the external occipital protuberance posteriorly. The tape was pulled tightly to compress the hair, and was read to within 0.1 cm.

### **Data Analysis**

Data analysis was performed using SAS 9.1 software. The sample size of 20 was based on an 80% power calculation ( $\beta=.2$ ). Statistical significance was defined as  $p < .05$ . The data were analyzed using the three most common methods for assessing reliability in anthropometric studies: the ICC, the TEM (expressed as relative TEM), and the PMCC. The values of these results were then compared qualitatively to each other as well as to the mean absolute, minimum, and maximum differences in each case.

$$ICC = \frac{s_b^2}{(s_b^2 + s_w^2)} \quad (B-1)$$

B-1. Intraclass Correlation Coefficient (ICC),  $s_b^2$  = pooled variance between subjects,  $s_w^2$  = pooled variance within subjects

$$VAV = \sum \frac{m_1 + m_2}{2} \quad (B-2)$$

Equation B-2. Variable Average Value (VAV), m = measurement (e.g. cm, kg)

$$RelativeTEM = TEM / VAV \times 100 \quad (B-3)$$

Equation B-3. Relative Technical Error of Measurement (RelativeTEM)

$$TEM = \sqrt{\sum \frac{d_i^2}{2n}} \quad (B-4)$$

Equation B-4. Absolute Technical Error of Measurement (TEM),  $d$ =deviation in measurement,  $i$ =number of deviations,  $n$ =number of participants measured

$$PMCC = \sum \frac{Z_x Z_y}{(n-1)} \quad (B-5)$$

Equation B-5. Pearson Product-Moment Correlation Coefficient (PMCC),  $z_x$  = standard score of first observer,  $z_y$  = standard score of second observer

## Results

All patients were similar in their degree of cooperation and level of disability based on physical and epidemiological attributes relevant to anthropometric measurements. Standards for acceptable ranges of reliability (clustered as unacceptable, acceptable, good, and excellent) were used to interpret and categorize results (table B-1).

Results for degrees of reliability were very similar for both inter- and intraobserver reliability measurements regardless of which statistical technique was employed. The ICC, PMCC and TEM scores were in the “excellent” range for interobserver reliability (table B-2), and all three observers also showed reliability in the “excellent” range on intraobserver analysis as well (table B-3).

The results of the different analytic techniques can be compared based on the ranges of the values that they produced. To aid in interpretation of results, the minimum, maximum, and mean absolute deviation (MAD) are provided for each measure (table B-4). For interobserver reliability, each technique produced an identical hierarchy with weight reliability higher than height, and height reliability higher than FOC. The range between variables was smallest for TEM, and the ICC technique had the largest range among measurements.

Interobserver measurement analysis using the three techniques revealed TEM for height of 0.65 cm, weight of 0.10 kg, and FOC of 0.09 cm. The TEM values overestimate error in height and weight when compared to the MAD values of 0.33 cm and 0.03 kg, and underestimate FOC compared to the MAD FOC of 0.18 cm. Intraobserver analysis using the TEM revealed error just over one mm for Observer 1 for FOC, and just under one mm for Observer 2.

### **Discussion**

Both interobserver and intraobserver reliability of anthropometric measurements on children with RTT were excellent. Although several reliability studies of anthropometric measurements exist in normal children, this study represents one of the few documenting reliability in a rare disease. Rare diseases pose unique challenges to the anthropometric researcher, including unusual body habitus, uncooperative participants, and limited sample sizes. Since several of the measures of reliability rely on the variance within a target population,

researchers must validate reliability within each of these groups prior to generating reference charts for that population.

We were able to integrate the process seamlessly within the structure of the parent study on growth by measuring participants first on entering the exam area, and again on leaving it. Based on the large number of participants in the parent study, we were able to easily recruit a sufficient number to meet our power analysis requirements. The process of recording and interpreting the results would have been simplified if we had initiated the project prior to design of the parent database and forms. One significant drawback was the limitation of intra-observer measurements to the parameter of FOC. We excluded height and weight based on time constraints of the study; in future prospective studies researchers could address this drawback during study design.

The interpretation of reliability depends strongly on the type of analysis performed. In our review of 44 recent studies of the reliability of anthropometric observations, researchers employed nine separate techniques for measuring reliability. Four of these methods are now considered obsolete, and the other five techniques can be summarized as follows. Ten studies used the PMCC, although this measure has recently fallen out of favor as it is not sensitive to degrees of magnitude. Four used the kappa correlation, another test which has received criticism. Although kappa is excellent at determining significant differences between groups, researchers have used the test inappropriately as a rating system. In addition, Kappa carries the drawback of only allowing comparison between two groups. Three studies employed the mean absolute difference, which tends to overestimate differences in reliability and is not sensitive to differences among specific individuals within a group. Regarding the two techniques which we found to be most useful, twenty of these studies used the ICC, and ten used the TEM.

In addition to measuring error, the ICC accounts for the variation within the population, as well as variation among raters and trials. Nevertheless the ICC has drawbacks; the correlation coefficient is dimensionless and is only generalizable to populations with similar variation. Alternately, the TEM is more stable across different population samples and is also expressed in the same units as the original measurement. Therefore, the TEM is a more intuitive measure and can be converted to exact margins of error in the original units. For example, in our study the ICC coefficient for height of .998 is difficult to interpret, and the range of “excellent” in which it falls may seem large. Alternatively, after converting back to the original units, the TEM indicates with 95% confidence that the difference in height measurements among the observers was no greater than 0.65 cm. In addition to these unique and beneficial attributes, disagreements between the two tests can call attention to idiosyncrasies of the design. The ICC and TEM may produce opposite results, since one depends on variation, and the other reports in units of original measurement. In such cases interpretation demands a thorough understanding of the study design and measurement procedure. The minimum, and maximum errors can be useful to identify specific individuals who are more difficult to measure, and the MAD provides a rough average error to compare to the TEM results.

With so many approaches to examining reliability and agreement within and among observers, ideally anthropometric studies should begin with a systematic array of thoughtfully chosen tests based on the variables being measured. Database designers can incorporate macros which continuously monitor measurements and perform several of these tests as data accrue, so that, when a critical number of participants are recruited, the program will automatically generate a reliability assessment. Investigators can then determine if the study design requires adjustment or if individual observers would benefit from more training. This approach was originally

proposed by Eliasziw, *et al.* in 1994 (Eliasziw et al. 1994), and first implemented on a large scale by the WHO in their 1999 MGRS (de Onis 2007).

This study is one of the first to measure and assess the reliability of anthropometric data in a rare disease. We demonstrated that combining the techniques of the ICC and TEM, researchers can generate meaningful reliability data based on a small sample size. This process should begin at the conception stage of any study of anthropometrics in rare diseases. The results not only lend power to the conclusions drawn from the resulting reference charts but can help to modify the study design during its execution.

Table B-1. Interpretation of Standardized Reliability Scores

Test	Unacceptable	Acceptable	Good	Excellent
TEM (% error)	>7.5%	5-7.5%	1-5%	<1%
ICC (cos (%var*100))	<0.5	0.5-0.75	0.75-0.9	0.9-1
PMCC (cos (%var*100))	<0.5	0.5-0.75	0.75-0.9	0.9-1

Table B-2. Interoperator Reliability Scores based on the ICC, PMCC, and TEM

Test	Height Variance	Weight Variance	FOC Variance
TEM/CV	0.999	1.000	0.999
ICC	0.998	1.000	0.997
PMCC	0.999	1.000	0.998

Table B-3. Intraoperator Reliability Scores based on the ICC, PMCC, and TEM

Test	FOC Variance	
	1	2
Observer Number		
TEM/CV	0.998	0.999
ICC	0.992	0.997
PMCC	0.998	0.998

Table B-4. Values for Mean Absolute Deviation

	Minimum	Maximum	Mean	Std. Deviation
FOC (cm)	0.025	0.475	0.175	0.140
HT (cm)	0.000	1.450	0.333	0.396
WT (kg)	0.000	0.422	0.034	0.087

## APPENDIX C DATA CLEANING TECHNIQUES

### Data preparation

- 1 – Run frequency check on all text variables and correct abnormal entries and redundancy within variables
- 2 – Count days from DOB to DOV, Convert to age in years variable, combine with months variable (i.e., growth chart data without DOV)
- 3 – Eliminate Scoliosis if Source=1
- 4 – Sort Source ascending, and if RDCRN data present code as 1 (e.g., Scoliosis, anthropometrics)
- 5 – Change Source=5 to 1 if RDCRN and DOV present (this was accidentally left as default=5)

### Error Detection

- 1 – Search for error in DOB:
  - Calculate days between 2 values for DOB and sort descending
  - Correct non-zero values
- 2 – Search for duplicate or blank age, wt, height
  - Examine as duplicate of all 3 (Ht+Wt+HC)first
  - Delete true duplicates (all columns identical: DOV or age, ID, and all measurements)
  - Re-examine for individual duplicates (Ht or Wt or HC) and delete true duplicates
- 3 – Mathematical errors
  - difference between hand length/palm+middle finger length
  - difference between MBA total and sum of 1,2,3

### Examination of Trends and Detection of Outliers

- 1 – Arrange in ascending order of Age
  - Screen for negative ages
  - Correct or explain
  - Search for date “0” = 1/1/1904
  - Search for obvious discrepancies in age/measurement
- 2 – Arrange all ht/wt/hc in ascending order and look for outliers
- 3 – Scatterplot of Age vs: Height, Wt, HC
  - Scatter for entire group, then Sorted for age 0-1, 1-3, 3-12, 12-21
  - AgeYear <= 1
  - AgeYear >= 1 & AgeYear <= 3
  - AgeYear >= 3 & AgeYear <= 12
  - AgeYear >= 12 & AgeYear <= 25
  - Mark outliers and explain – potential explanations
    - Wrong DOV/DOB
    - Wrong units used (inches vs cm)
    - Wrong decimal place (g instead of kg)
    - Numeral keyed in wrong – go back to source data to verify
  - Connected by ID to examine inappropriate trends (head or height shrinking)

- 4 – Scatterplot of each variable for each participant
- 5 – Boxplot of DOB for Ht, Wt, HC
  - AgeYear <= .1
  - AgeYear = 0
- 6 – Boxplots of months (0-12, 12-36, etc in 3y intervals) vs Ht, Wt, HC
  - AgeRoundMonth <= 12
  - AgeRoundMonth >= 12 & AgeRoundMonth <= 36
  - Repeat for 6 month intervals (to be used for groups over 3y)
    - Agesixmonth >= 3 & Agesixmonth < 9
    - Agesixmonth >= 9 & Agesixmonth < 15
    - Agesixmonth >= 15 & Agesixmonth < 25

APPENDIX D  
ITEMS FROM STUDY FORMS

**Initial Forms**

- Clinical Criteria
  - 15 Primary Items
  - 11 Supportive Items
- Diagnostic and MECP2 Status
  - Diagnosis
  - Age at diagnosis
  - MECP2 results and Lab
- Demographic Information
  - Gender
  - Adopted
  - Ethnicity
  - Race
  - Age at enrollment
  - Residence
- Initial History
  - Perinatal history
  - Early infancy
  - Development
    - Gross motor
    - Fine motor
    - Language and affect
    - Receptive language
    - Adaptive behaviors
  - Rett-related events
    - Seizure information
    - Breathing disorders
    - Oromotor disorders
    - Constipation
    - Cholecystitis
    - Gastro-esophageal reflux
    - Fractures
    - Stereotypies
    - Self-abuse
    - Scoliosis
    - Abnormal EKG
    - Age at menstuation
- Current History
  - General health
  - Rett-related events (as above)
  - Hand use
  - Ambulation
  - Seizures

- Diet and supplementation
- Dysautonomia
- Sleep dysfunction
- Hospitalization/Surgery since last visit
- Clinical Assessment
  - Alertness
  - Activity level
  - Affect
  - Self-abuse
  - Eye contact
  - Language
  - Breathing pattern
  - Bruxism
  - Hand preference
  - Stereotypies
  - Hand/finger position
  - Habitus (thorax, abdomen)
  - Tanner staging
  - Spine
  - Motor: Tone, gegenhalten, contractures, muscle mass, strength, reaches/holds, sitting, ambulation, foot position, reflexes (stretch and primitive), dyskinesia
  - Autonomic: sensation, skin temperature, autonomic function
- Motor-Behavioral Assessment (37 items, 0-148)
  - Behavioral/Social Assessment (16 items, 0-4)
  - Orofacial/Respiratory Assessment (7 items, 0-4)
  - Motor Assessment/Physical signs (14 items, 0-4)
- Clinical Severity Scale (13 items, 0-58)
  - Onset (1 item, 0-5)
  - Growth (2 items, 0-4)
  - Motor (4 items, 0-4/0-5)
  - Communication (2 items, 0-4/0-5)
  - Rett Behavior/Neurologic (4 items, 0-4/0-5)
- Measurements
  - Height, Weight, FOC
  - Hand/Foot
  - Anthropometrics:
    - Skinfolds: Triceps, Biceps, Subscapular, Suprailiac, Thigh, Shin
    - Circumferences: Upper mid-arm, Thigh, Shin
    - Lower leg length
- SF-36 (14+) – 11 general health items

**Every 6 months (under 12y) or annually (over 12y)**

- Current History
- Clinical Assessment
- Motor-Behavioral Assessment
- Clinical Severity Scale

- Measurements
- SF-36 (14+) – (only annually)

APPENDIX E  
DIAGNOSTICS FOR CLASSIC RETT SYNDROME GROWTH CHARTS

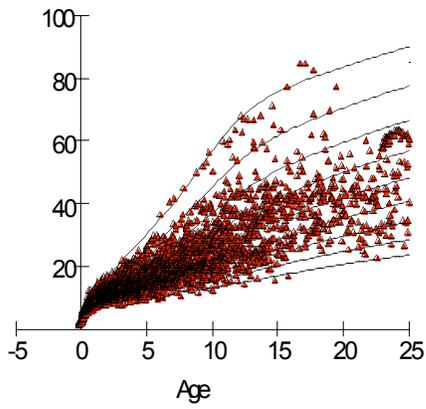


Figure E-1. Weight data plotted on curves.

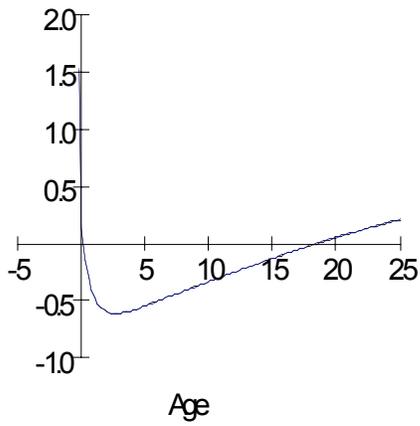


Figure E-2. Weight power transformation curve.

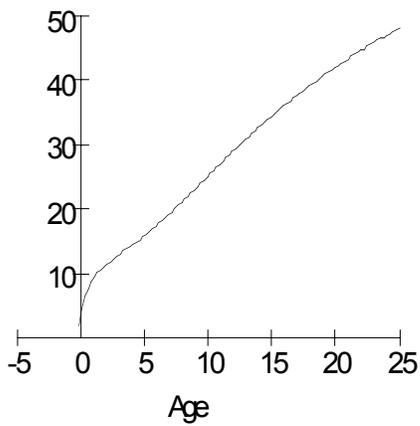


Figure E-3. Weight mean curve.

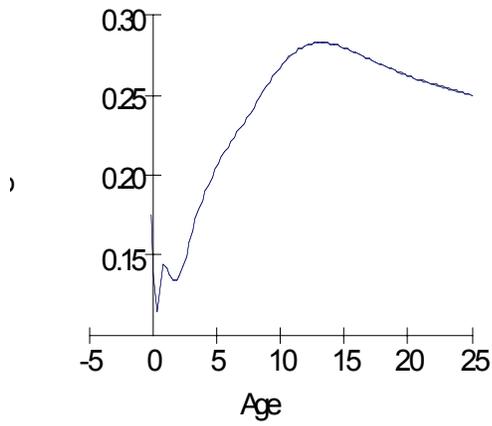


Figure E-4. Weight coefficient of variation curve.

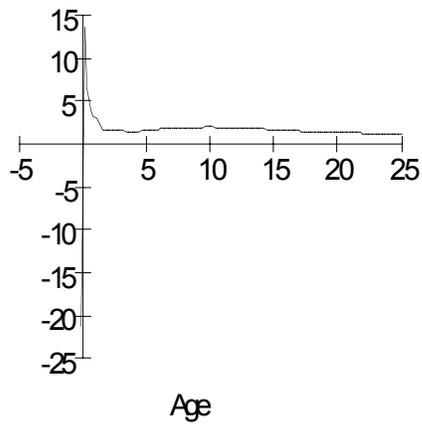


Figure E-5. Mean weight velocity curve.

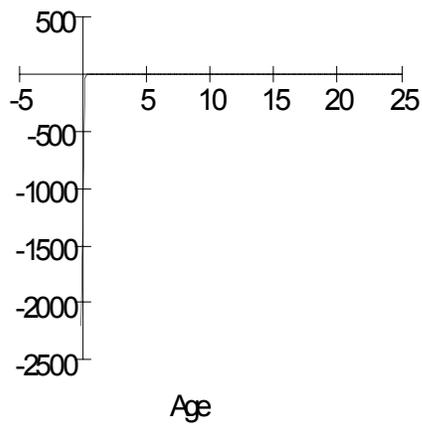


Figure E-6. Mean weight acceleration curve.

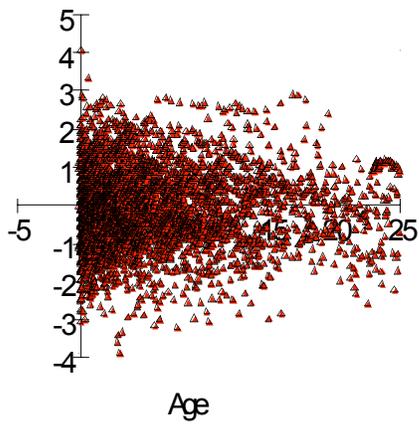


Figure E-7. Weight Z-scores plotted against age.

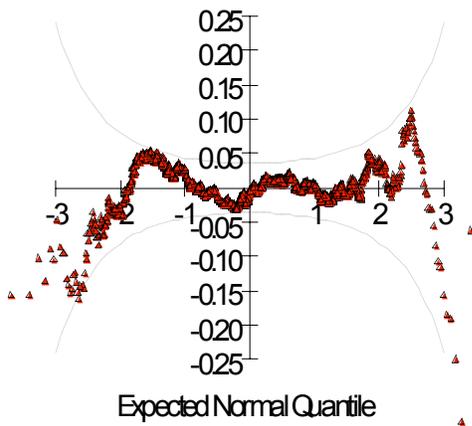


Figure E-8. Weight detrended Q-Q plot.

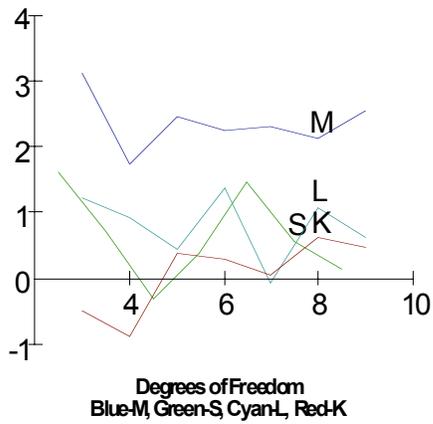


Figure E-9. Weight Q-statistic data.

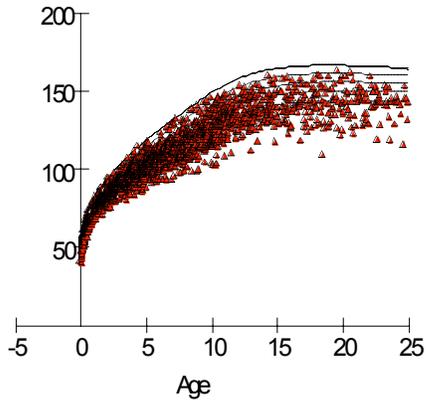


Figure E-10. Height data plotted on curves.

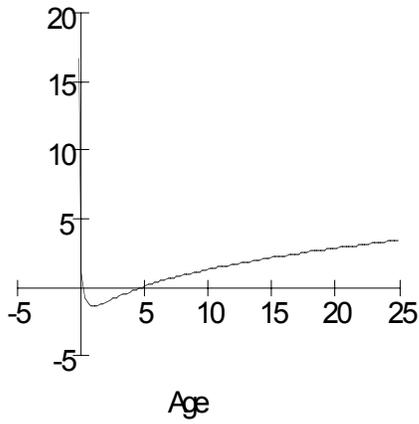


Figure E-11. Height power transformation curve.

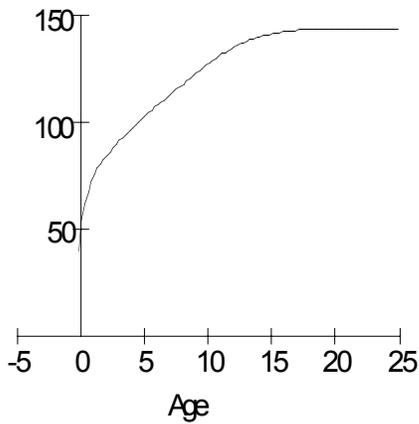


Figure E-12. Height mean curve.

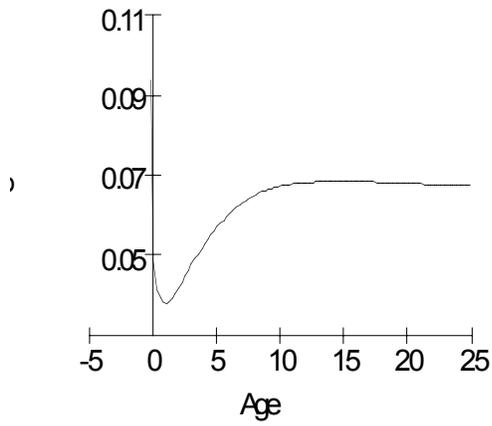


Figure E-13. Height coefficient of variation curve.

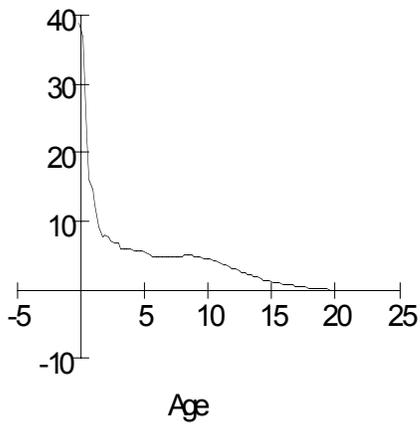


Figure E-14. Mean height velocity curve.



Figure E-15. Mean height acceleration curve.

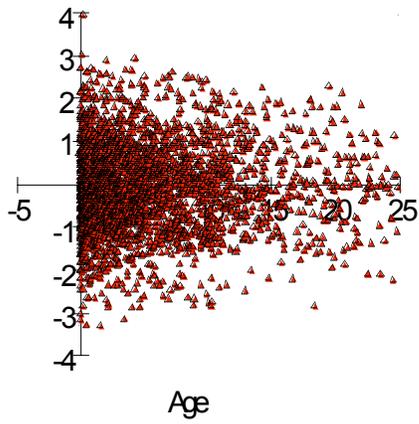


Figure E-16. Height Z-scores plotted against age.

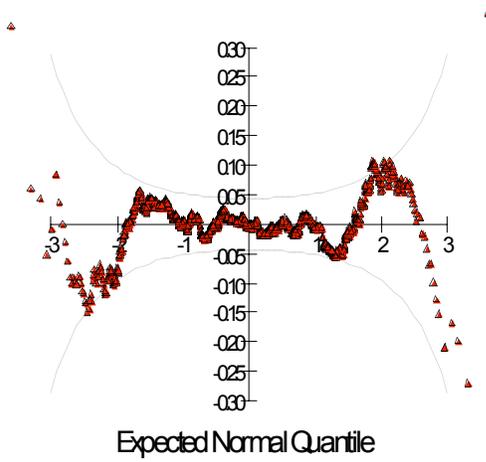


Figure E-17. Height detrended Q-Q plot.

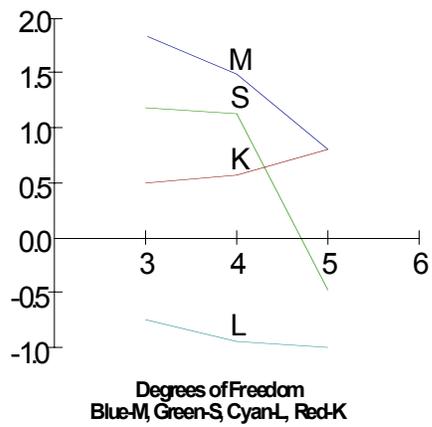


Figure E-18. Height Q-statistic data.

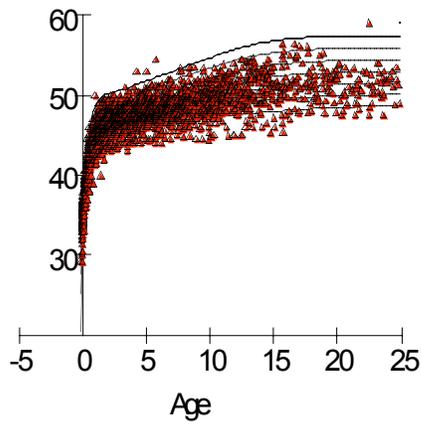


Figure E-19. FOC data plotted on curves.

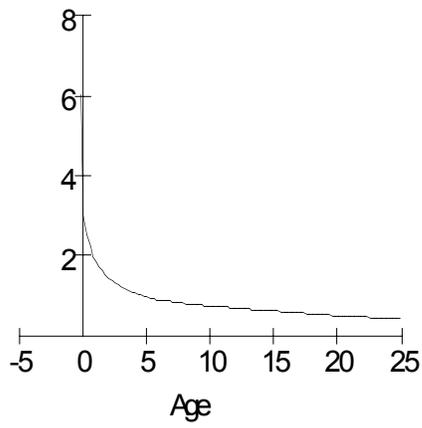


Figure E-20. FOC power transformation curve.

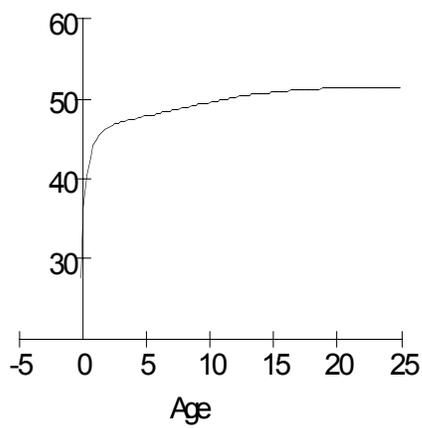


Figure E-21. FOC mean curve.

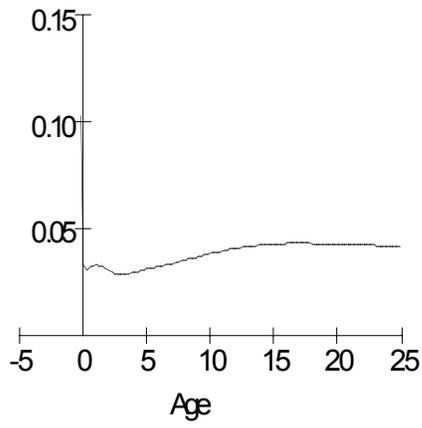


Figure E-22. FOC coefficient of variation curve.

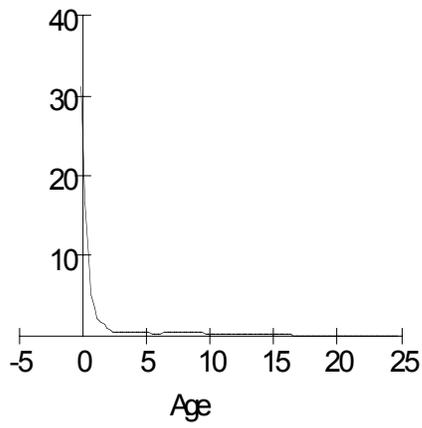


Figure E-23. Mean FOC velocity curve.

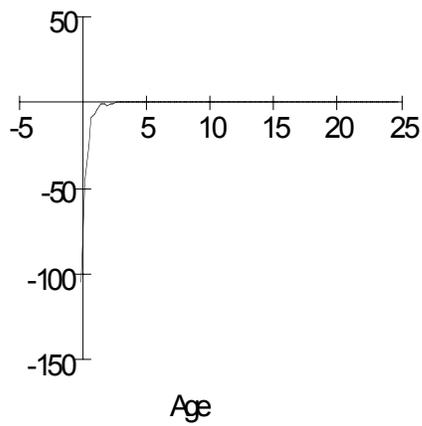


Figure E-24. Mean FOC acceleration curve.

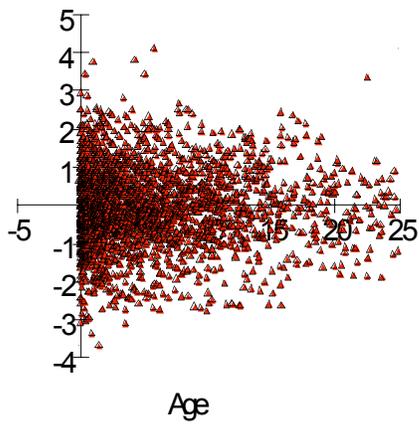


Figure E-25. FOC Z-scores plotted against age.

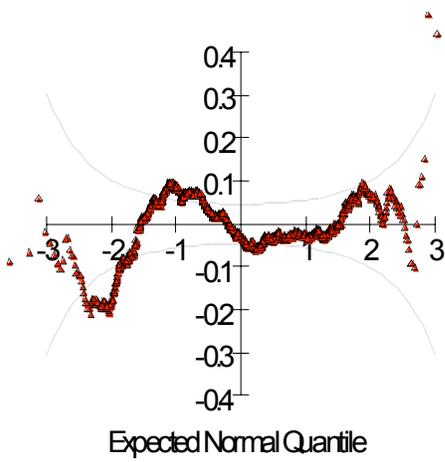


Figure E-26. FOC detrended Q-Q plot.

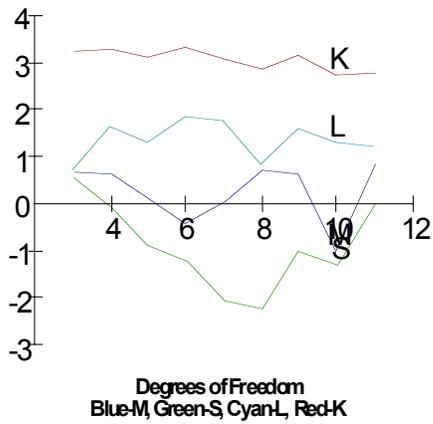


Figure E-27. FOC Q-statistic data.

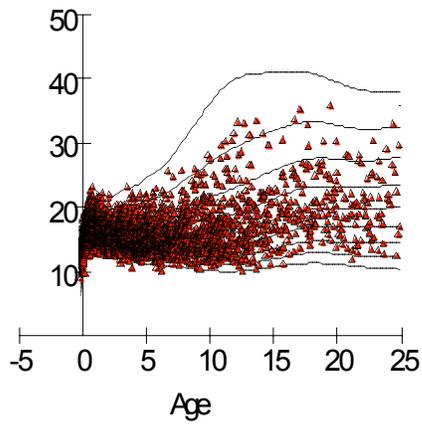


Figure E-28. BMI data plotted on curves.

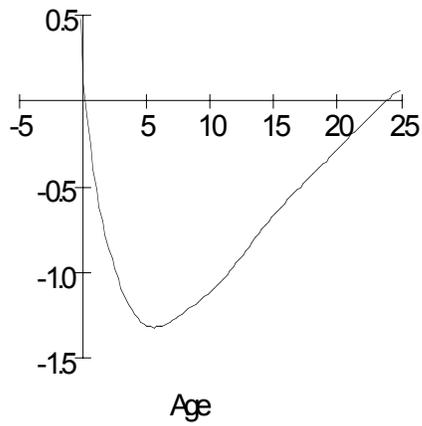


Figure E-29. BMI power transformation curve.

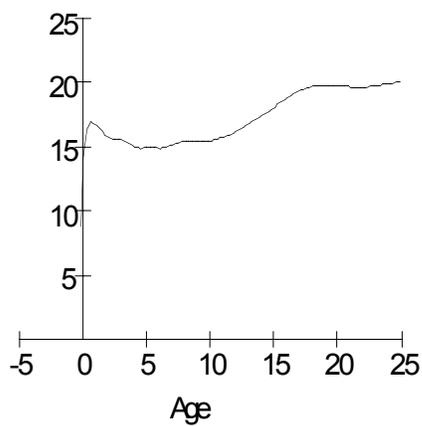


Figure E-30. BMI mean curve.

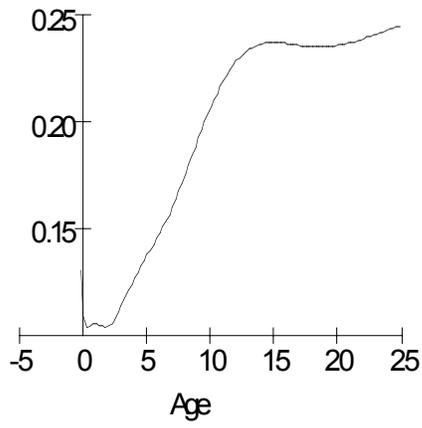


Figure E-31. BMI coefficient of variation curve.

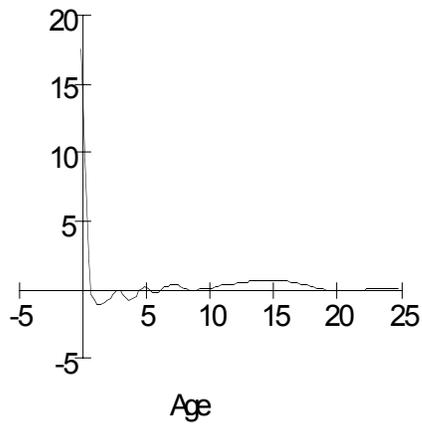


Figure E-32. Mean BMI velocity curve.

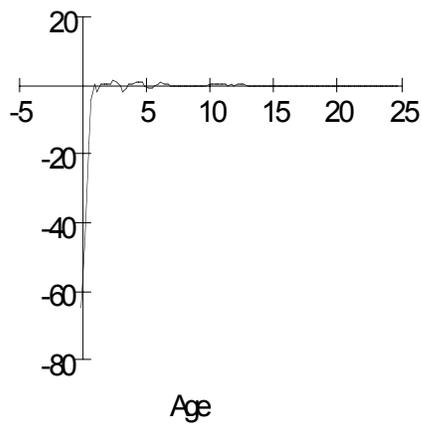


Figure E-33. Mean BMI acceleration curve.

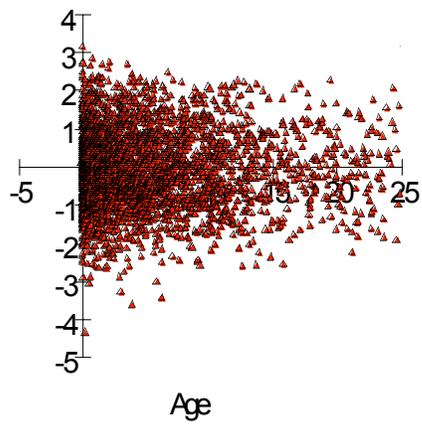


Figure E-34. BMI Z-scores plotted against age.

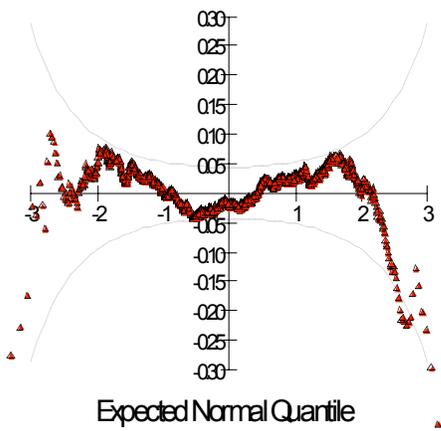


Figure E-35. BMI detrended Q-Q plot.

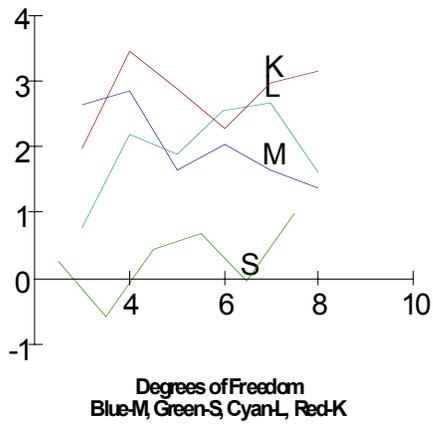


Figure E-36. BMI Q-statistic data.

APPENDIX F  
LMS VALUES FOR GROWTH CHARTS

Key:

- L= Power transformation curve EDF
- M= Mean curve EDF
- S= Coefficient of variation curve EDF
- Transf = Y/N – was a power transformation also performed on the age variable
- Power = Power transformation of age variable
- Offset = Number added to age prior to computation and subtracted afterward (to prevent taking log of zero)

Models for:	L	M	S	Transf	Power	Offset
0-25y						
Classic:						
Weight:	4	20	12	Y	0	0.18
Height	2.5	20	7	Y	0	0.18
FOC	3	13	9	Y	0	0.18
BMI	5	17	8	Y	0.5	0.18
Atypical:						
Weight:	3	7	6	Y	0	0.18
Height:	3	6	6	Y	0	0.18
FOC:	3	7	3	Y	0	0.18
BMI:	3	6	3	Y	0.5	0.18
Non-Rett:						
Weight:	3	12	4	Y	0	0.18
Height:	3	8	4	Y	0	0.18
FOC:	3	5	3	Y	0	0.18
BMI:	3	7	4	Y	0	0.18
Pre-1997						
Weight:	5	16	4	Y	0.5	0.18
Height:	2.4	11	3	Y	0	0.18
FOC:	5	10	5	Y	0	0.18
BMI:	5	12	7	Y	0.4	0.18
Post-1997						
Weight:	4	17	6	Y	0.4	0.18
Height:	3	8	4	Y	0	0.18
FOC:	3	8	3	Y	0	0.18
BMI:	4	9	4	Y	0	0.18

MildCSS						
Weight:	3	14	5	Y	0.3	0.18
Height:	3	12	5	Y	0.4	0.18
FOC:	4	11	5	Y	0.4	0.18
BMI:	3	11	4	Y	0.4	0.18
MildCSS2						
Weight:	4	14	5	Y	0.3	0.18
Height:	3	11	5	Y	0.5	0.18
FOC:	3	10	4	Y	0.2	0.18
BMI:	3	11	4	Y	0.3	0.18
SevereCSS						
Weight:	3	12	5	Y	0.3	0.18
Height:	3	11	3	Y	0.3	0.18
FOC:	3	11	3	Y	0	0.18
BMI:	3	9	6	Y	0	0.18
SevereCSS2						
Weight:	3	12	4	Y	0	0.18
Height:	2	10	6	Y	0.4	0.18
FOC:	3	8	4	Y	0	0.18
BMI:	3	9	4	Y	0.1	0.18
MildMBA						
Weight:	4.0	13.0	4.0		0.4	0.18
Height:	3.0	14.0	5.0	T	0.2	0.18
FOC:	4.0	13.0	4.0	T	0.2	0.18
BMI:	3.0	12.0	5.0	T	0.3	0.18
SevereMBA						
Weight:	3.0	10.0	4.0	Y	0.5	0.18
Height:	3.0	12.1	5.0	Y	0	0.18
FOC:	3.0	8.0	3.0	Y	0	0.18
BMI:	3	8	6	Y	0	0.18

APPENDIX G  
GROWTH CHARTS

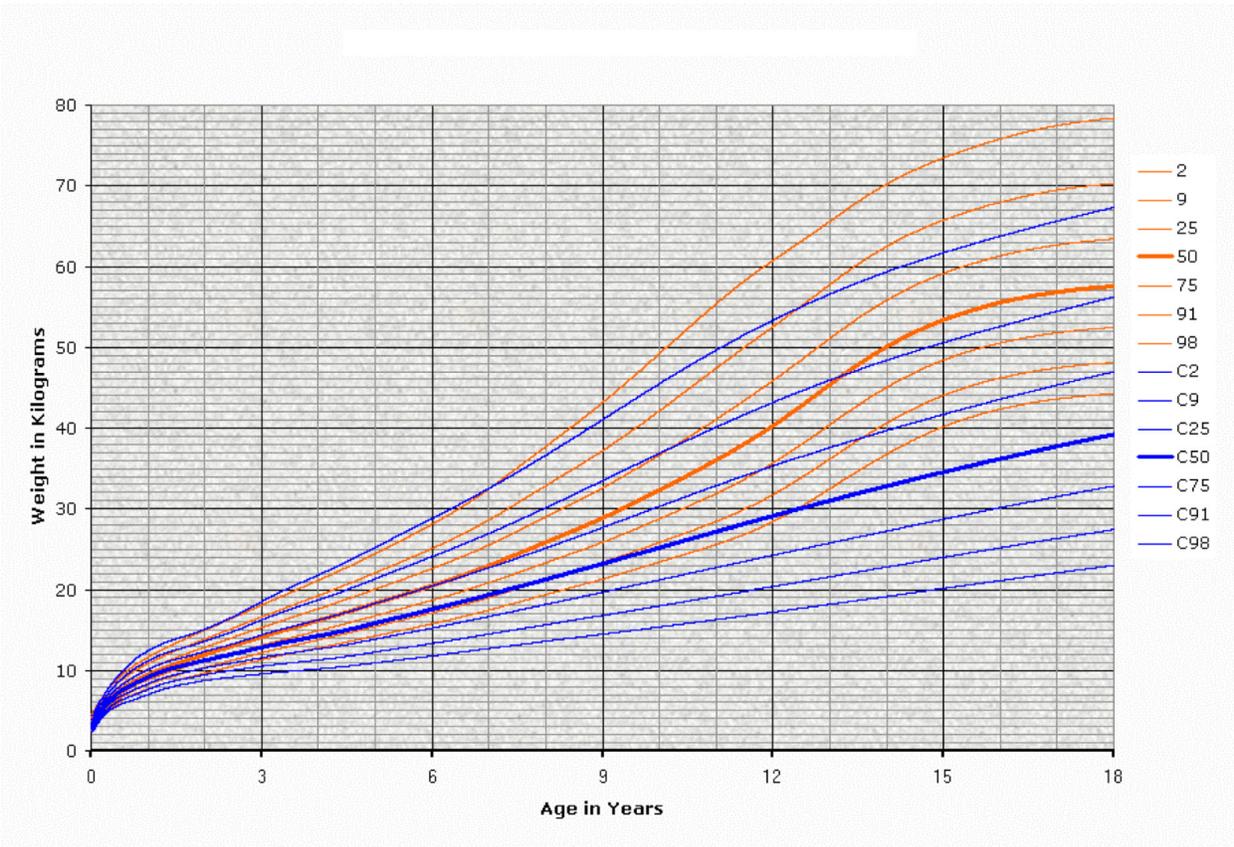


Figure G-1. Weight in classic RTT (blue) compared to normal reference (orange).

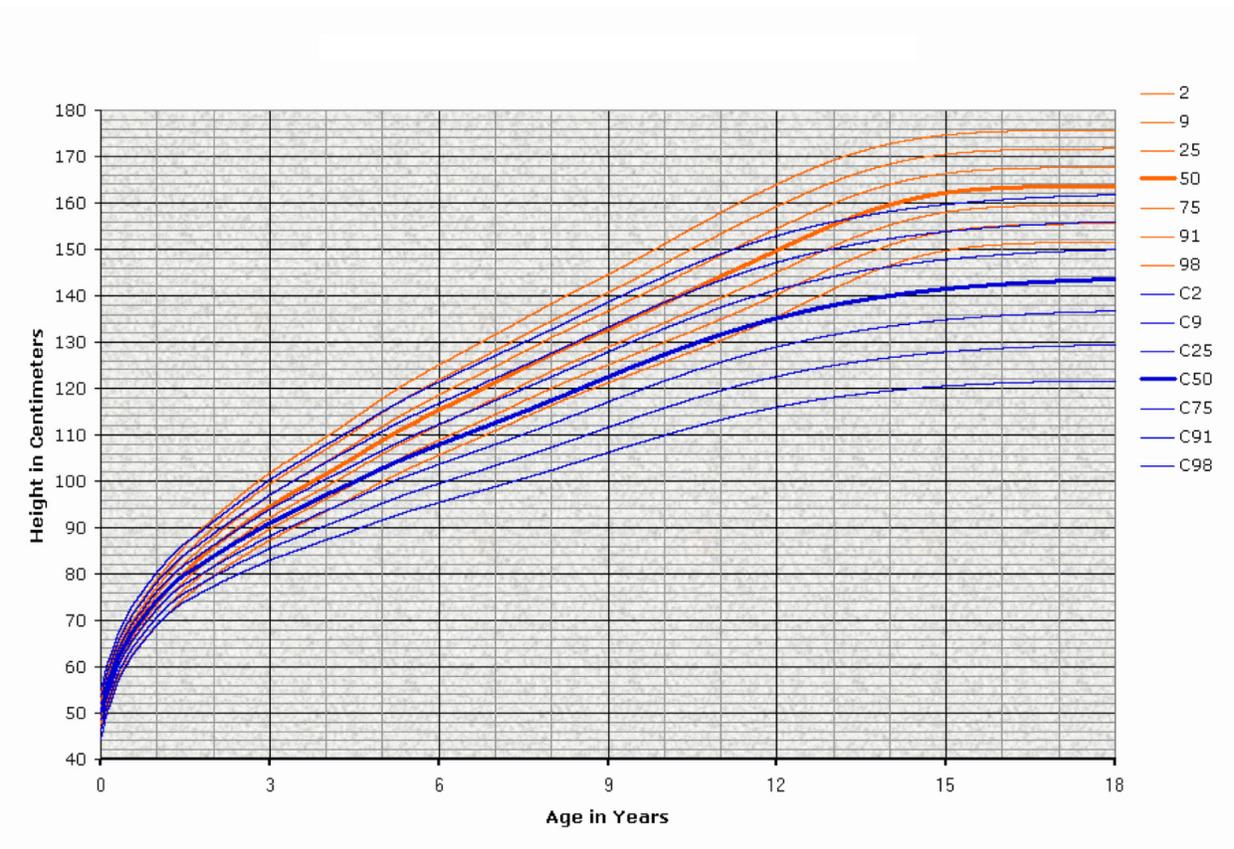


Figure G-2. Height in classic RTT (blue) compared to normal reference (orange).

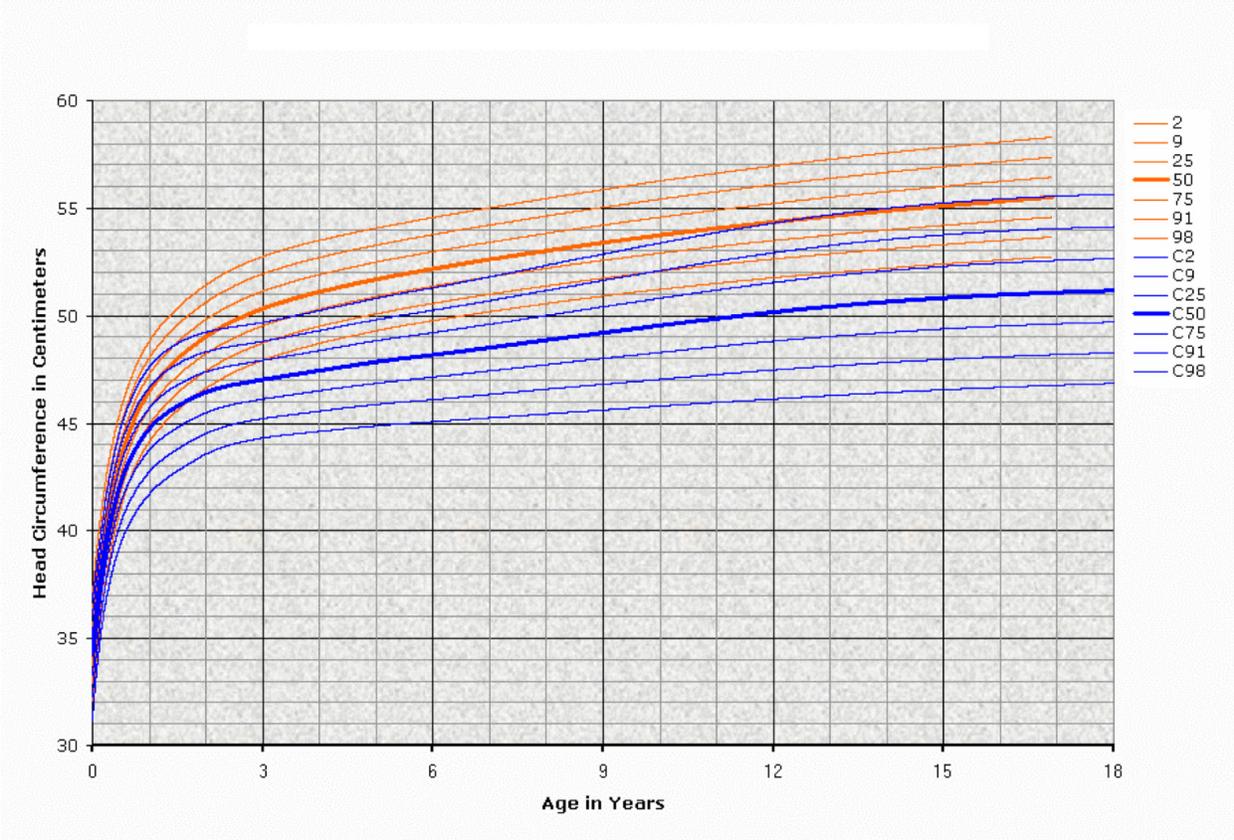


Figure G-3. FOC in classic RTT (blue) compared to normal reference (orange).

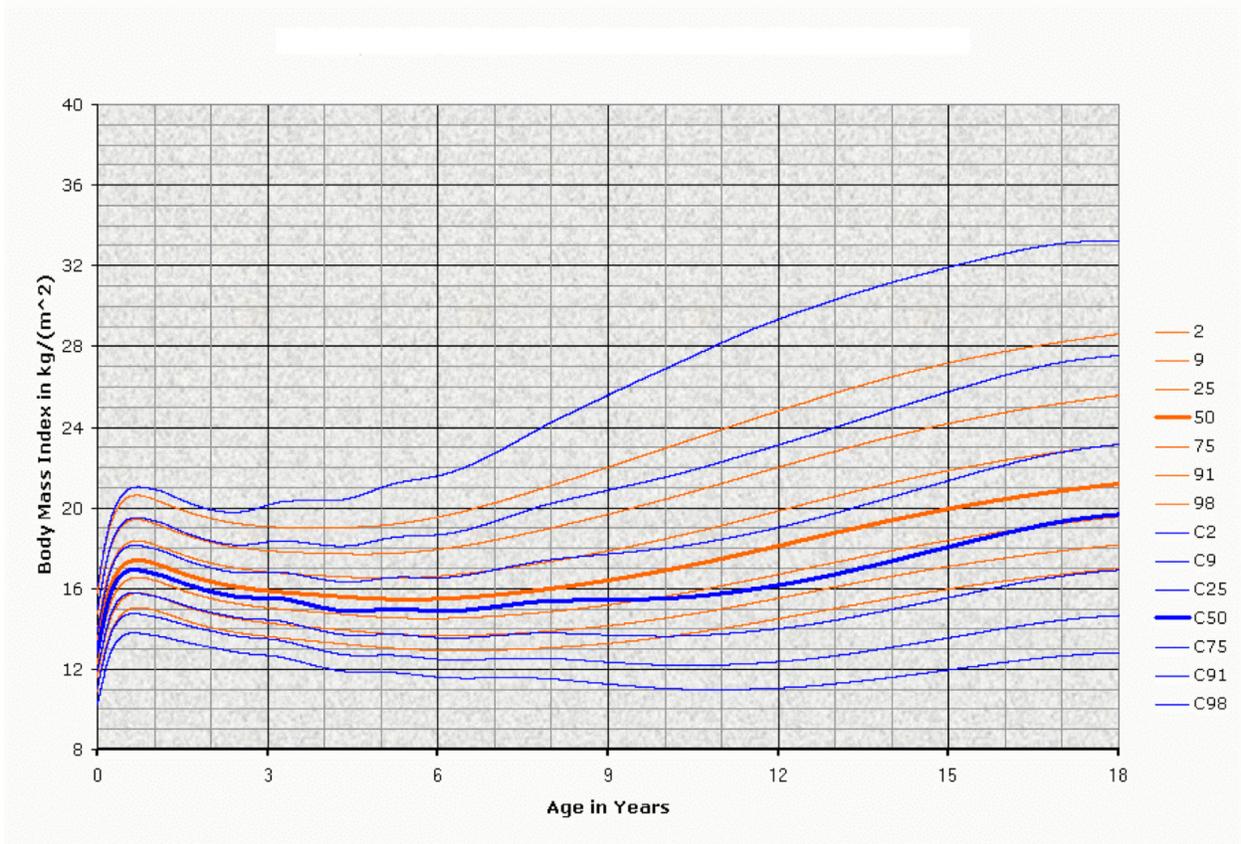


Figure G-4. BMI in classic RTT (blue) compared to normal reference (orange).

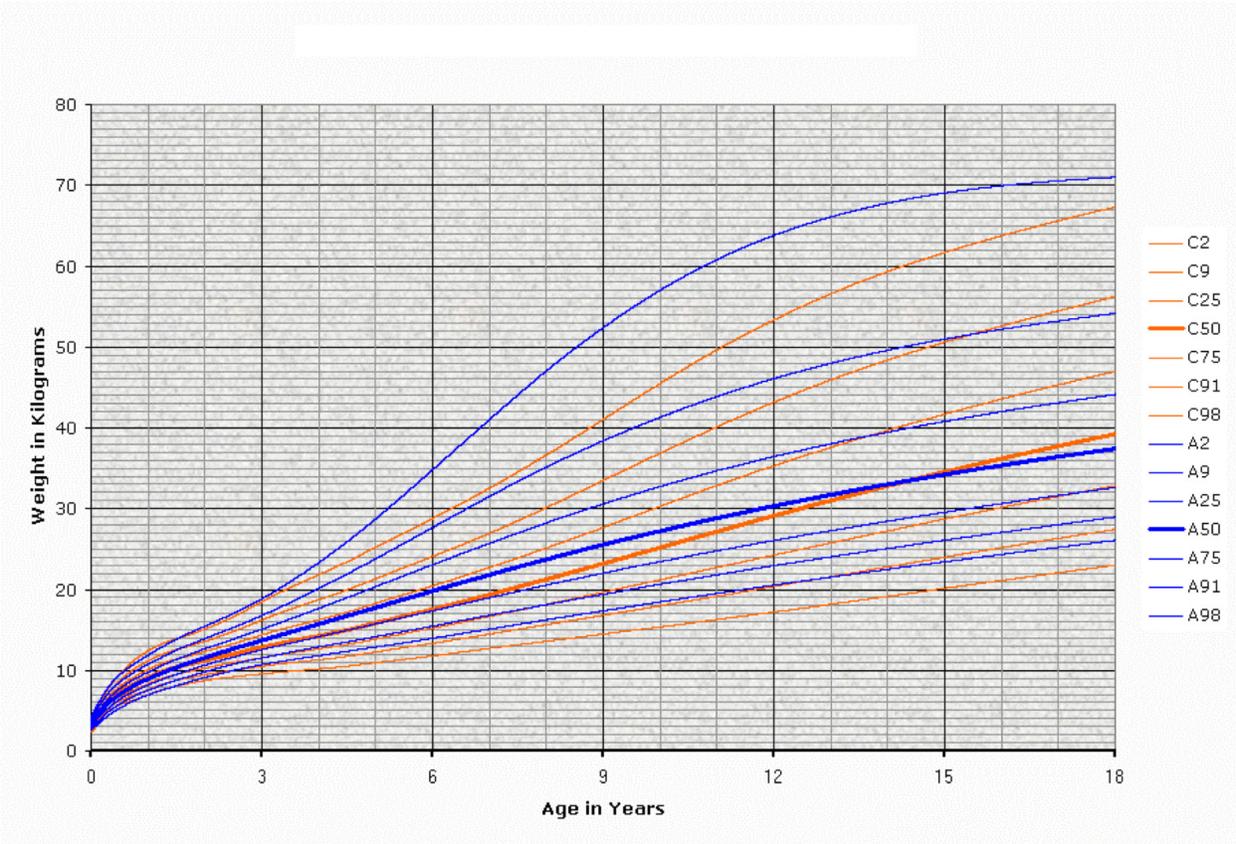


Figure G-5. Weight in classic (orange) versus atypical RTT (blue).

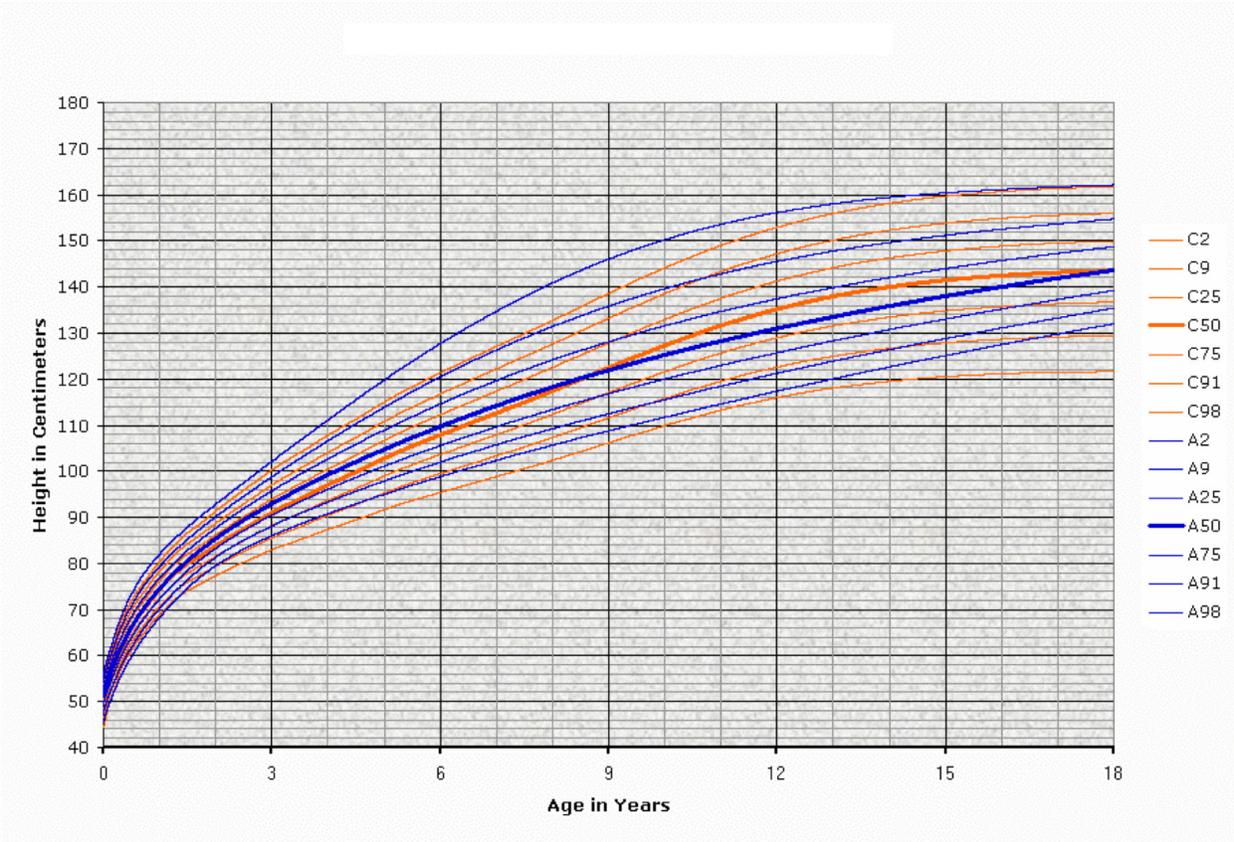


Figure G-6. Height in classic versus atypical RTT.

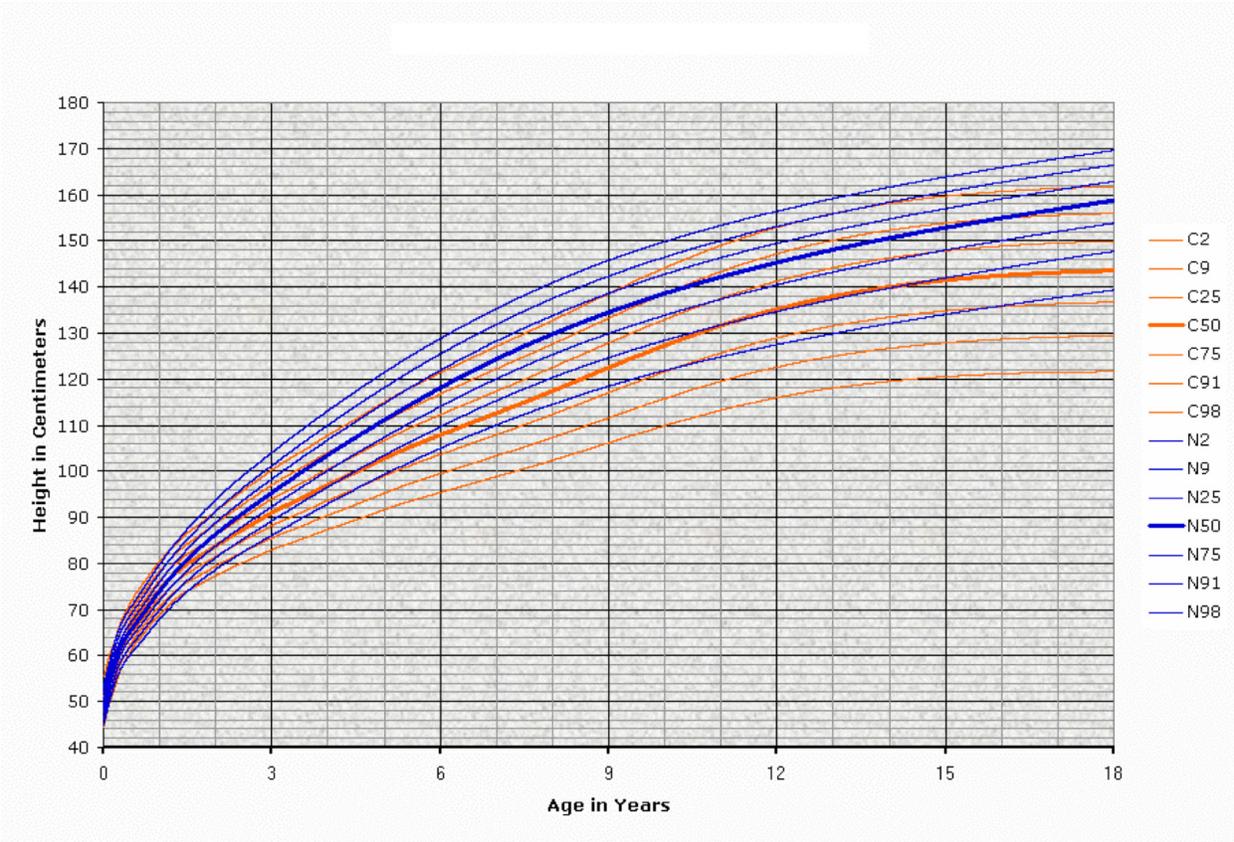


Figure G-7. Height in classic (orange) versus atypical RTT (blue).

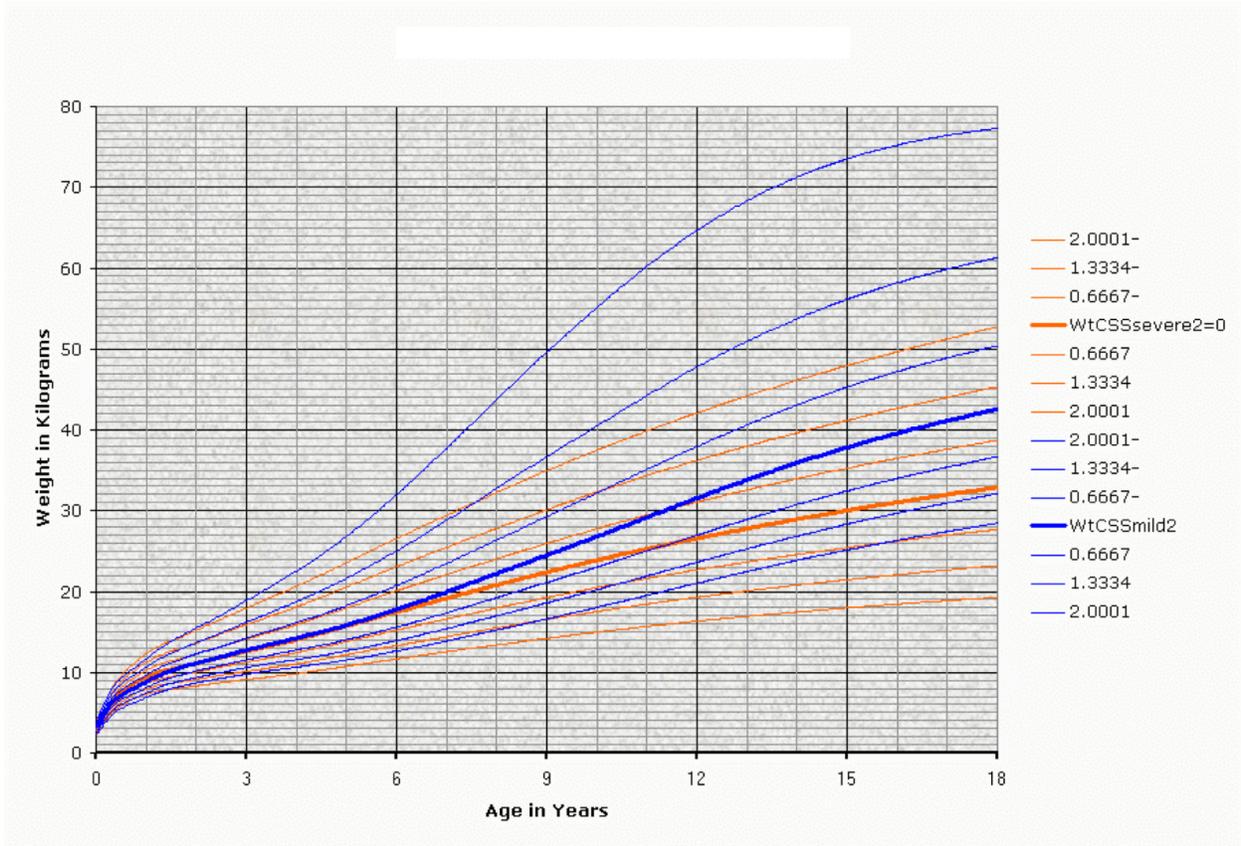


Figure G-8. Weight in severe RTT (orange) versus mild RTT (blue).

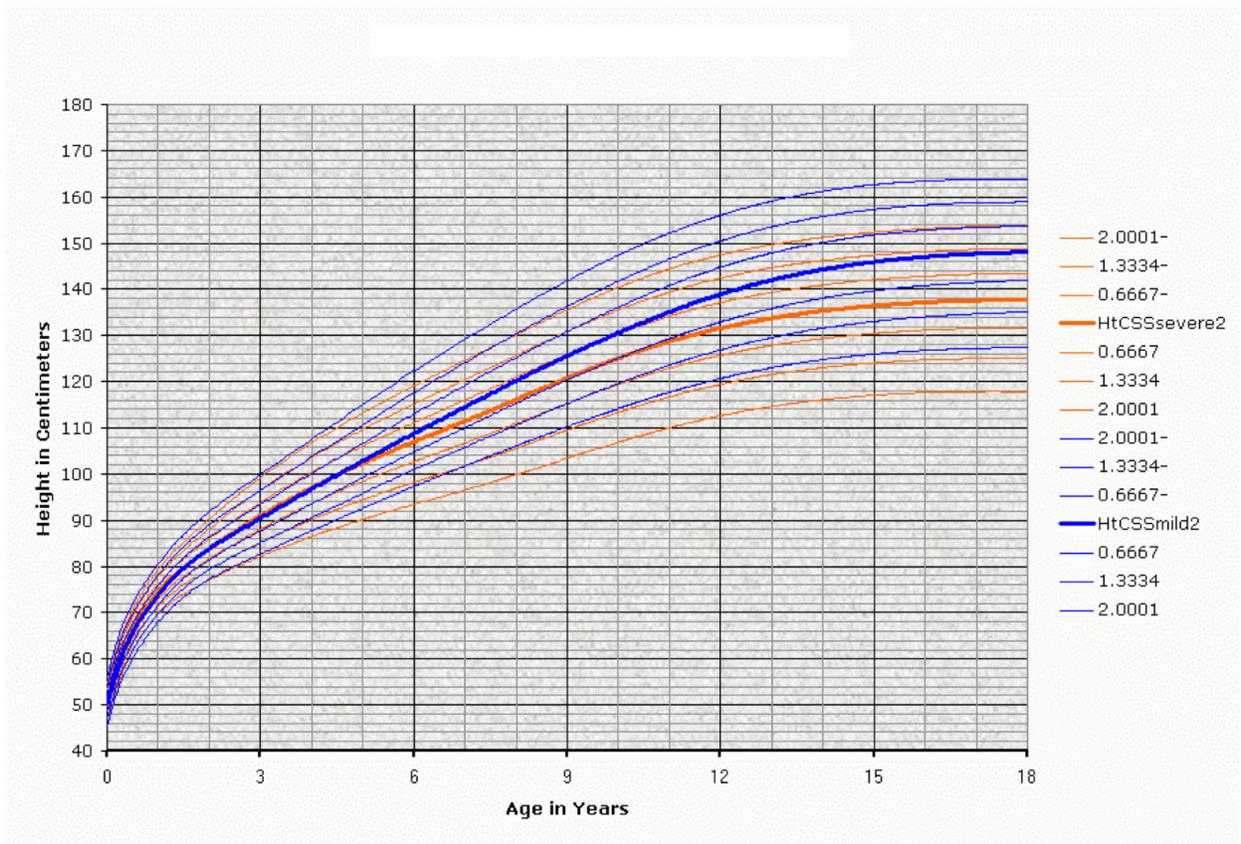


Figure G-9. Height in severe RTT (orange) versus mild RTT (blue).

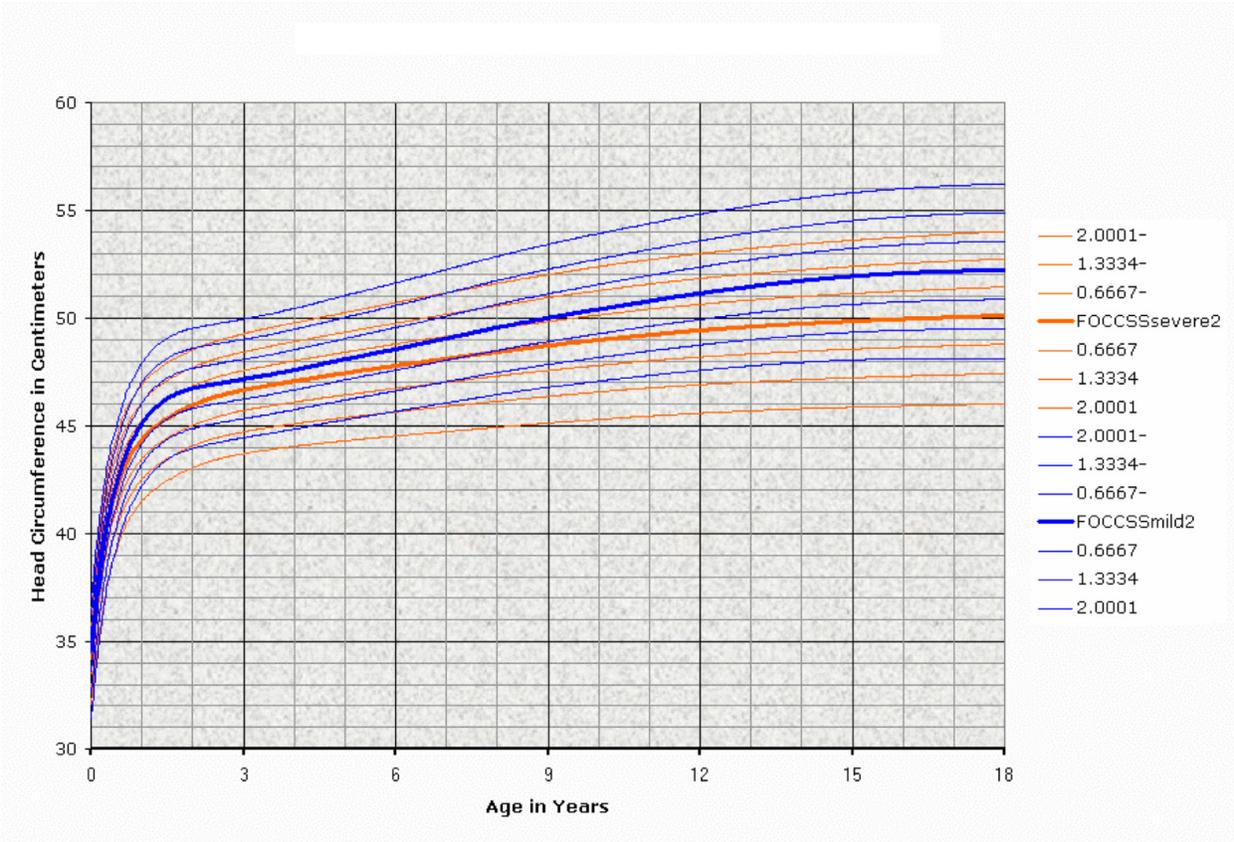


Figure G-10. FOC in severe RTT (orange) versus mild RTT (blue).

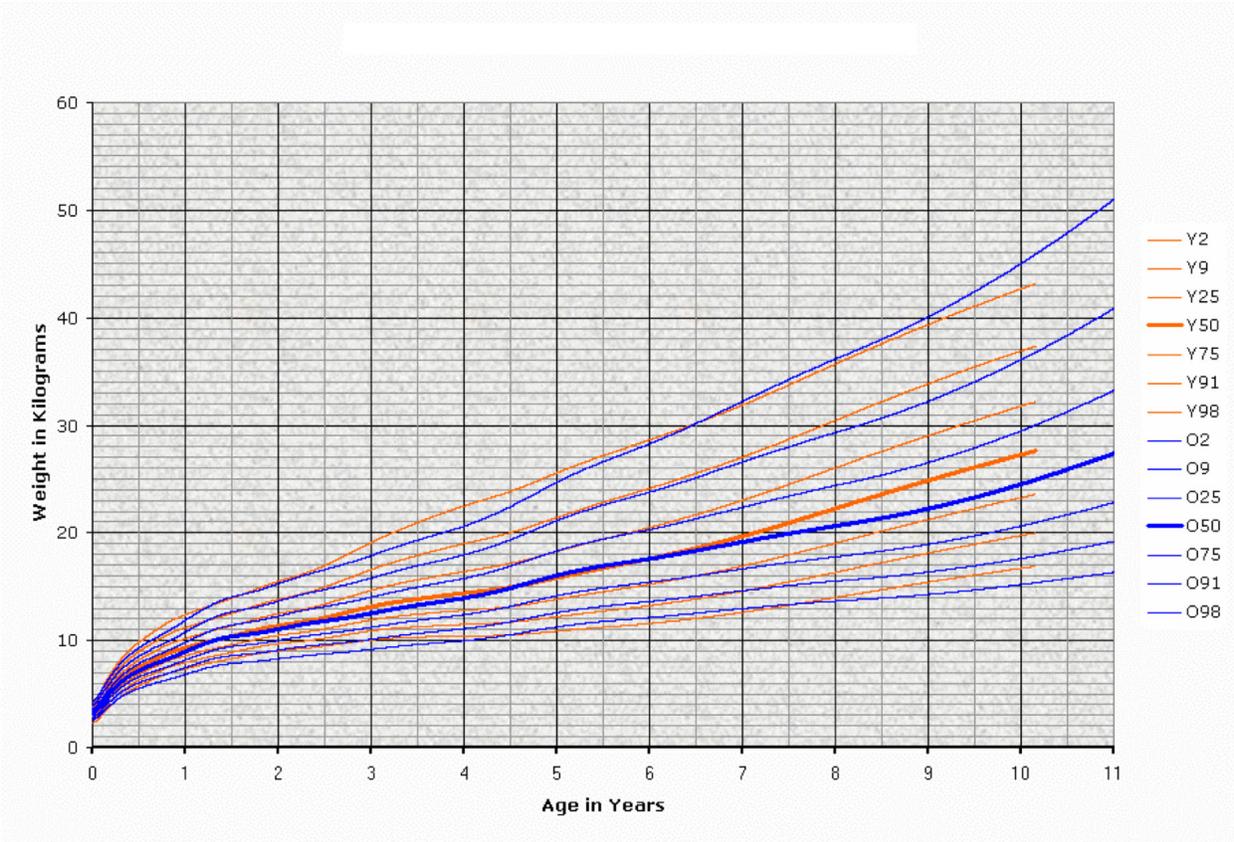


Figure G-11. Weight in RTT children born before 1997 (blue) and after 1997 (orange).

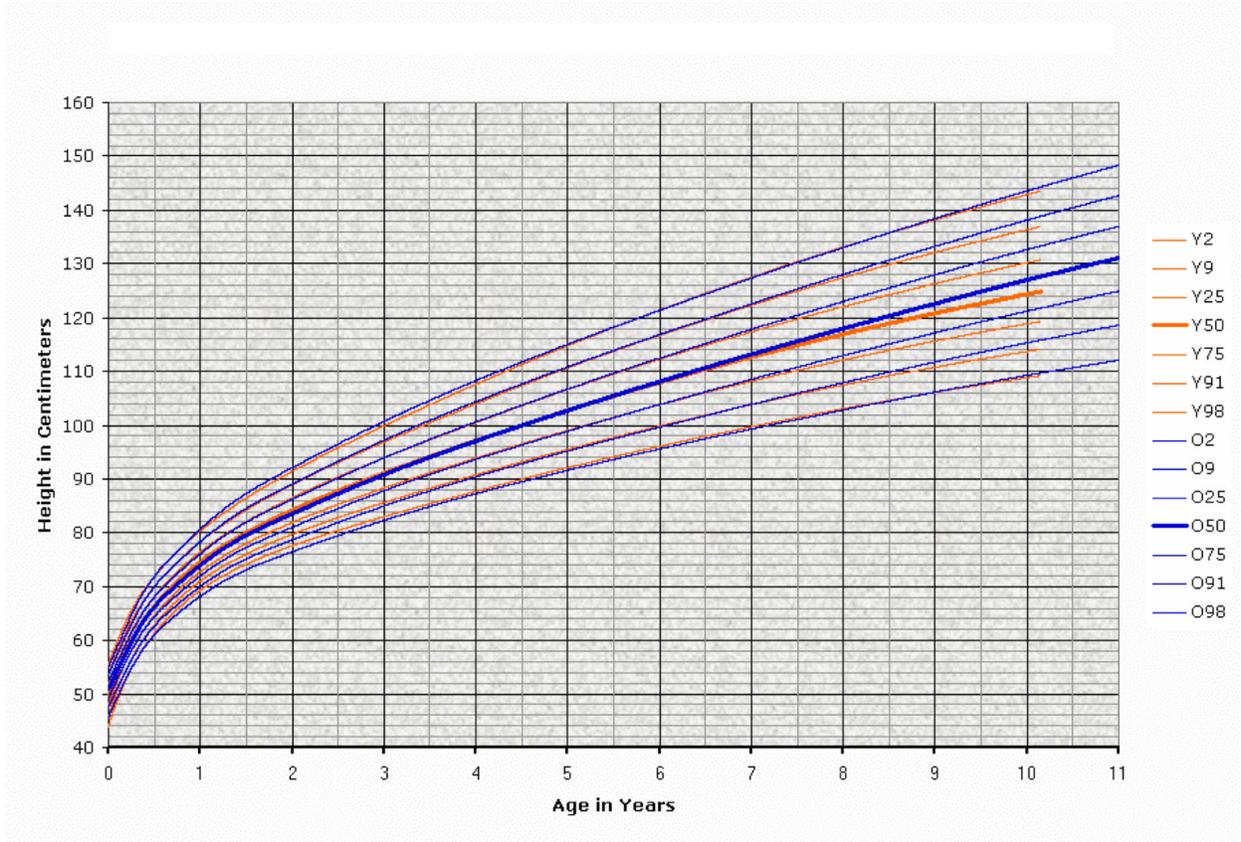


Figure G-12. Height in RTT children born before 1997 (blue) and after 1997 (orange).

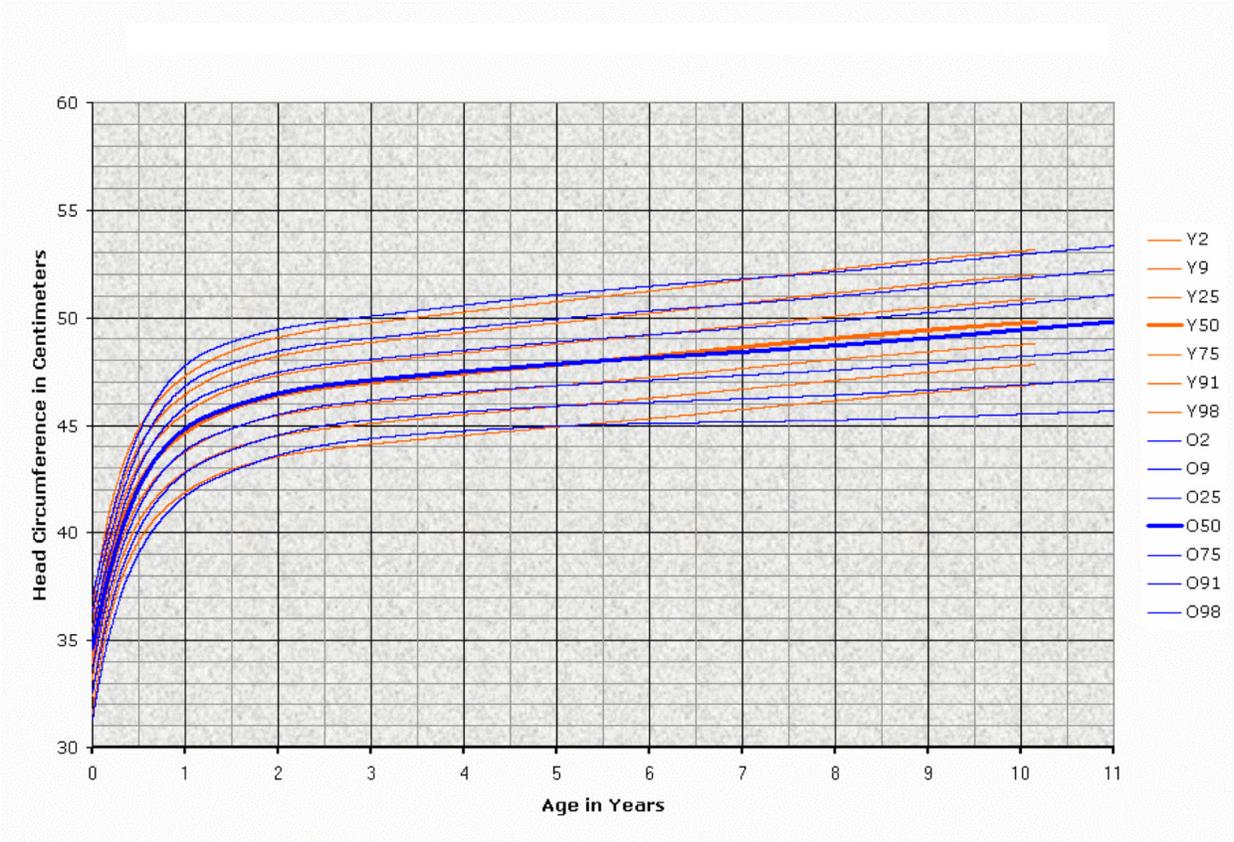


Figure G-13. FOC in RTT children born before 1997 (blue) and after 1997 (orange).

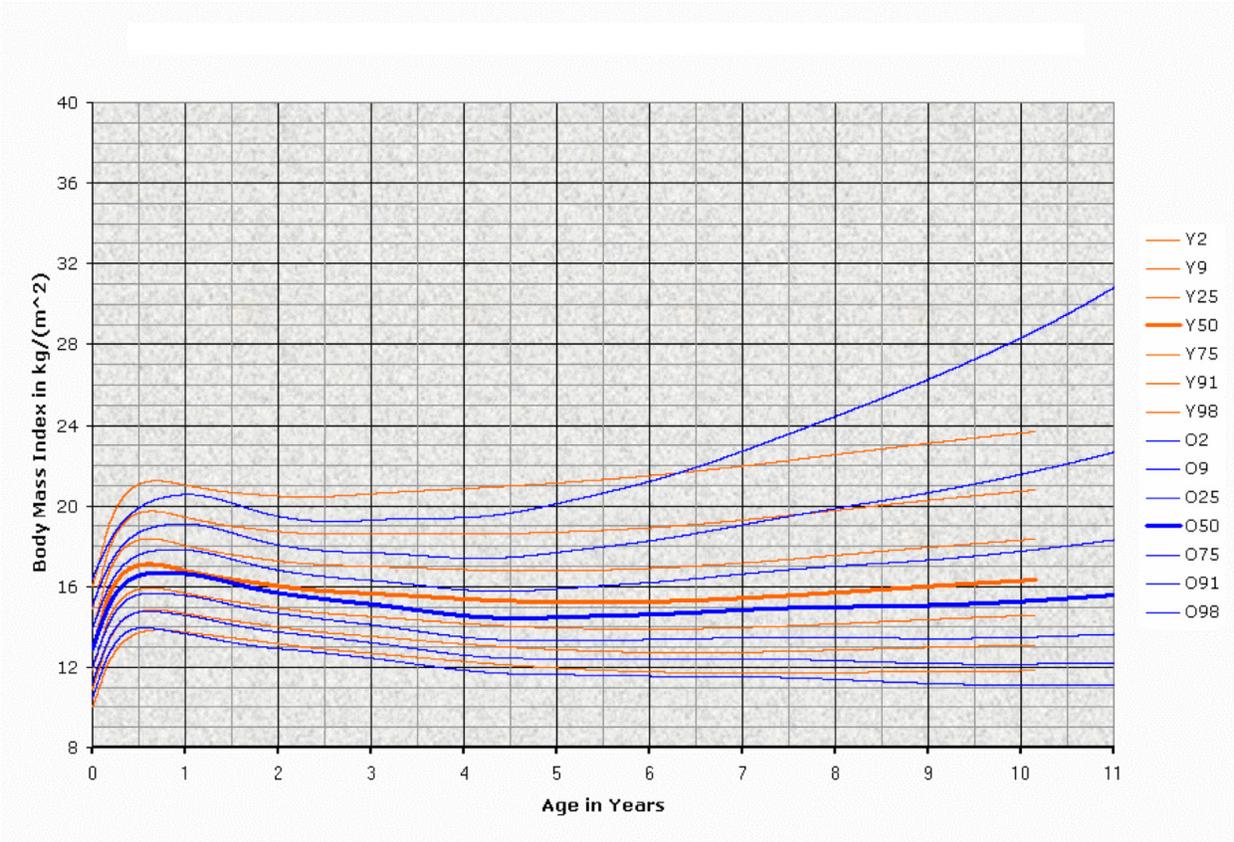


Figure G-14. Weight in RTT children born before 1997 (blue) and after 1997 (orange).

## LIST OF REFERENCES

- Amir RE, Van den Veyver IB, Schultz R, Malicki DM, Tran CQ, Dahle EJ, Philippi A, Timar L, Percy AK, Motil KJ and others. 2000. Influence of mutation type and X chromosome inactivation on Rett syndrome phenotypes. *Ann Neurol* 47 (5):670-679.
- Amir RE, Van den Veyver IB, Wan M, Tran CQ, Francke U, and Zoghbi HY. 1999. Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nat Genet* 23 (2):185-188.
- Anzo M, Takahashi T, Sato S, and Matsuo N. 2002. The cross-sectional head circumference growth curves for Japanese from birth to 18 years of age: the 1990 and 1992-1994 national survey data. *Ann Hum Biol* 29 (4):373-388.
- Arnell H, Gustafsson J, Ivarsson SA, and Anneren G. 1996. Growth and pubertal development in Down syndrome. *Acta Paediatrica* 85 (9):1102-1106.
- Ballestar E, Yusufzai TM, and Wolffe AP. 2000. Effects of Rett syndrome mutations of the methyl-CpG binding domain of the transcriptional repressor MeCP2 on selectivity for association with methylated DNA. *Biochemistry* 39 (24):7100-7106.
- Bebbington A, Anderson A, Ravine D, Fyfe S, Pineda M, de Klerk N, Ben Zeev Ghidoni B, Yatawara N, Percy A, Kaufmann W and others. 2008. Investigating genotype-phenotype relationships in Rett syndrome using an international dataset. *Neurology*:in press.
- Berkey CS, Reed RB, and Valadian I. 1983. Longitudinal growth standards for preschool children. *Ann Hum Biol* 10 (1):57-67.
- Bernasconi S, Larizza D, Benso L, Volta C, Vannelli S, Milani S, Aicardi G, Berardi R, Borrelli P, Boscherini B and others. 1994. Turner's syndrome in Italy: familial characteristics, neonatal data, standards for birth weight and for height and weight from infancy to adulthood. *Acta Paediatr* 83 (3):292-298.
- Bienvenu T, Philippe C, De Roux N, Raynaud M, Bonnefond JP, Pasquier L, Lesca G, Mancini J, Jonveaux P, Moncla A and others. 2006. The incidence of Rett syndrome in France. *Pediatr Neurol* 34 (5):372-375.
- Borghi E, de Onis M, Garza C, Van den Broeck J, Frongillo EA, Grummer-Strawn L, Van Buuren S, Pan H, Molinari L, Martorell R and others. 2006. Construction of the World Health Organization child growth standards: selection of methods for attained growth curves. *Stat Med* 25 (2):247-265.
- Bowditch H. 1879. *The Growth of Children - A Supplementary Investigation - With Suggestions in Regard to Methods of Research*. Boston, MA: Band, Avery & Co, Printers to the Commonwealth, 117 Franklin St.

- Bowditch H. 1891. *The Growth of Children, Studied by Galton's Method of Percentile Grades*. Boston, MA: Wright and Potter.
- Box G, and Cox D. 1964. An Analysis of Transformations. *Journal of the Royal Statistical Society Series B* 26:211-252.
- Brook CG, Gasser T, Werder EA, Prader A, and Vanderschueren-Lodewyckx MA. 1977. Height correlations between parents and mature offspring in normal subjects and in subjects with Turner's and Klinefelter's and other syndromes. *Ann Hum Biol* 4 (1):17-22.
- Brousseau K, and Brainerd H. 1928. *Mongolism : a study of the physical and mental characteristics of mongolian imbeciles*. Baltimore: The Williams & Wilkins Company.
- Butler MG, Brunshwig A, Miller LK, and Hagerman RJ. 1992. Standards for selected anthropometric measurements in males with the fragile X syndrome. *Pediatrics* 89 (6 Pt 1):1059-1062.
- Butler MG, and Meaney FJ. 1991. Standards for selected anthropometric measurements in Prader-Willi syndrome. *Pediatrics* 88 (4):853-860.
- Caldwell PD, and Smith DW. 1972. The XXY (Klinefelter's) syndrome in childhood: detection and treatment. *J Pediatr* 80 (2):250-258.
- Chang Q, Khare G, Dani V, Nelson S, and Jaenisch R. 2006. The disease progression of *Mecp2* mutant mice is affected by the level of BDNF expression. *Neuron* 49 (3):341-348.
- Clementi M, Milani S, Mammi I, Boni S, Monciotti C, and Tenconi R. 1999. Neurofibromatosis type 1 growth charts. *Am J Med Genet* 87 (4):317-323.
- Cole T. 1988. Fitting Smoothed Centile Curves to Reference Data. *Journal of the Royal Statistical Society Series A (Statistics in Society)* 151 (3):385-418.
- Cole T. 1994. Do Growth Charts Need a Facelift? *British Medical Journal* 308.
- Cole T, and Pan H. *LMSChartMaker*. UK: Medical Research Council.
- Cole TJ. 1990. The LMS method for constructing normalized growth standards. *Eur J Clin Nutr* 44 (1):45-60.
- Cole TJ. 1993. The Use and Construction of Anthropometric Growth Reference Standards. *Nutrition Research Reviews* 6 (01):19-50.
- Cole TJ. 1996. Some questions about how growth standards are used. *Horm Res* 45 Suppl 2:18-23.

- Cole TJ. 1998. Presenting information on growth distance and conditional velocity in one chart: practical issues of chart design. *Statistics in Medicine* 17 (23):2697-2707.
- Cole TJ, Freeman JV, and Preece MA. 1998. British 1990 growth reference centiles for weight, height, body mass index and head circumference fitted by maximum penalized likelihood. *Stat Med* 17 (4):407-429.
- Cole TJ, and Green PJ. 1992. Smoothing reference centile curves: the LMS method and penalized likelihood. *Stat Med* 11 (10):1305-1319.
- Collins MS, and Eaton-Evans J. 2001. Growth study of cri du chat syndrome. *Arch Dis Child* 85 (4):337-338.
- Cotterill AM, Majrowski WH, Hearn SJ, Jenkins S, and Savage MO. 1993. Assessment of the reliability of school nurse height measurements in an inner-city population. (The Hackney Growth Initiative). *Child Care Health Dev* 19 (3):159-165.
- Cremers MJG, Tweel I, Boersma B, Wit JM, and Zonderland M. 1996. Growth curves of Dutch children with Down's syndrome. *Journal of Intellectual Disability Research* 40 (5):412-420.
- Cronk C, Crocker AC, Pueschel SM, Shea AM, Zackai E, Pickens G, and Reed RB. 1988. Growth charts for children with Down syndrome: 1 month to 18 years of age. *Pediatrics* 81 (1):102-110.
- Cronk CE. 1978. Growth of children with Down's syndrome: birth to age 3 years. *Pediatrics* 61 (4):564-568.
- Cutler AT, Benezra-Obeiter R, and Brink SJ. 1986. Thyroid function in young children with Down syndrome. *Am J Dis Child* 140 (5):479-483.
- de Onis M. 2007. WHO Child Growth Standards: length/height-for-age, weight-for-age, weight-for-length, weight for height and body mass index-for-age: methods and development. Geneva, Switzerland: World Health Organization.
- de Onis M, Garza C, and Habicht JP. 1997. Time for a new growth reference. *Pediatrics* 100 (5):E8.
- Dibley MJ, Goldsby JB, Staehling NW, and Trowbridge FL. 1987. Development of normalized curves for the international growth reference: historical and technical considerations. *Am J Clin Nutr* 46 (5):736-748.
- Ducan B, Lubchenco LO, and Hansman C. 1974. Growth charts for children 0 to 18 years of age. *Pediatrics* 54 (4):497-502.

- Eiholzer U, Boltshauser E, Frey D, Molinari L, and Zachmann M. 1988. Short stature: a common feature in Duchenne muscular dystrophy. *Eur J Pediatr* 147 (6):602-605.
- Einspieler C, Kerr AM, and Prechtel HF. 2005. Is the early development of girls with Rett disorder really normal? *Pediatr Res* 57 (5 Pt 1):696-700.
- Eliasziw M, Young SL, Woodbury MG, and Fryday-Field K. 1994. Statistical methodology for the concurrent assessment of interrater and intrarater reliability: using goniometric measurements as an example. *Phys Ther* 74 (8):777-788.
- Emery JL, Waite AJ, Carpenter RG, Limerick SR, and Blake D. 1985. Apnoea monitors compared with weighing scales for siblings after cot death. *Arch Dis Child* 60 (11):1055-1060.
- Erkula G, Jones KB, Sponseller PD, Dietz HC, and Pyeritz RE. 2002. Growth and maturation in Marfan syndrome. *Am J Med Genet* 109 (2):100-115.
- Erlanson A, and Hagberg B. 2005. MECP2 abnormality phenotypes: clinicopathologic area with broad variability. *J Child Neurol* 20 (9):727-732.
- Ershow AG. 1986. Growth in black and white children with Down syndrome. *Am J Ment Defic* 90 (5):507-512.
- Frisch H, Waldhauser F, Lebl J, Solyom J, Hargitai G, Kovacs J, Pribilincova Z, Krzisnik C, and Battelino T. 2002. Congenital adrenal hyperplasia: lessons from a multinational study. *Horm Res* 57 Suppl 2:95-101.
- Gairdner D, and Pearson J. 1971. A growth chart for premature and other infants. *Arch Dis Child* 46 (250):783-787.
- Galton SF. 1885. Anthropometric Per-centiles. *Nature* 31:223-225.
- Galvao TC, and Thomas JO. 2005. Structure-specific binding of MeCP2 to four-way junction DNA through its methyl CpG-binding domain. *Nucleic Acids Res* 33 (20):6603-6609.
- Giacometti E, Luikenhuis S, Beard C, and Jaenisch R. 2007. Partial rescue of MeCP2 deficiency by postnatal activation of MeCP2. *Proc Natl Acad Sci U S A* 104 (6):1931-1936.
- Glaze DG. 2005. Neurophysiology of Rett syndrome. *J Child Neurol* 20 (9):740-746.
- Graitcer PL, and Gentry EM. 1981. Measuring children: one reference for all. *Lancet* 2 (8241):297-299.
- Griffiths RD, and Edwards RH. 1988. A new chart for weight control in Duchenne muscular dystrophy. *Arch Dis Child* 63 (10):1256-1258.

- Guy J, Gan J, Selfridge J, Cobb S, and Bird A. 2007. Reversal of neurological defects in a mouse model of Rett syndrome. *Science* 315 (5815):1143-1147.
- Hagberg B, Aicardi J, Dias K, and Ramos O. 1983. A progressive syndrome of autism, dementia, ataxia, and loss of purposeful hand use in girls: Rett's syndrome: report of 35 cases. *Ann Neurol* 14 (4):471-479.
- Hagberg B, Hanefeld F, Percy A, and Skjeldal O. 2002. An update on clinically applicable diagnostic criteria in Rett syndrome. Comments to Rett Syndrome Clinical Criteria Consensus Panel Satellite to European Paediatric Neurology Society Meeting, Baden Baden, Germany, 11 September 2001. *Eur J Paediatr Neurol* 6 (5):293-297.
- Hagberg B, and Witt-Engerstrom I. 1986. Rett syndrome: a suggested staging system for describing impairment profile with increasing age towards adolescence. *Am J Med Genet Suppl* 1:47-59.
- Hagberg G, Stenbom Y, and Engerstrom IW. 2001. Head growth in Rett syndrome. *Brain Dev* 23 Suppl 1:S227-229.
- Hagberg G, Stenbom Y, and Witt Engerstrom I. 2000. Head growth in Rett syndrome. *Acta Paediatr* 89 (2):198-202.
- Hamill PV, Drizd TA, Johnson CL, Reed RB, and Roche AF. 1977. NCHS growth curves for children birth-18 years. United States. *Vital Health Stat* 11 (165):i-iv, 1-74.
- Hargitai G, Solyom J, Battelino T, Lebl J, Pribilincova Z, Hauspie R, Kovacs J, Waldhauser F, and Frisch H. 2001. Growth patterns and final height in congenital adrenal hyperplasia due to classical 21-hydroxylase deficiency. Results of a multicenter study. *Horm Res* 55 (4):161-171.
- Hauffa BP, Schlippe G, Roos M, Gillessen-Kaesbach G, and Gasser T. 2000. Spontaneous growth in German children and adolescents with genetically confirmed Prader-Willi syndrome. *Acta Paediatr* 89 (11):1302-1311.
- Healy MJ, Rasbash J, and Yang M. 1988. Distribution-free estimation of age-related centiles. *Ann Hum Biol* 15 (1):17-22.
- Hoffbuhr KC, Moses LM, Jerdonek MA, Naidu S, and Hoffman EP. 2002. Associations between MeCP2 mutations, X-chromosome inactivation, and phenotype. *Ment Retard Dev Disabil Res Rev* 8 (2):99-105.
- Holm VA. 1986. Physical growth and development in patients with Rett syndrome. *Am J Med Genet Suppl* 1:119-126.
- Horton WA, Rotter JI, Rimoin DL, Scott CI, and Hall JG. 1978. Standard growth curves for achondroplasia. *J Pediatr* 93 (3):435-438.

- Howard VM. 1939. Stature of Massachusetts children of North European and Italian ancestry. *American Journal of Physical Anthropology* 24 (3):301-346.
- Ishikawa T, Furuyama M, Ishikawa M, Ogawa J, and Wada Y. 1987. Growth in head circumference from birth to fifteen years of age in Japan. *Acta Paediatr Scand* 76 (5):824-828.
- Jedele KB. 2007. The overlapping spectrum of rett and angelman syndromes: a clinical review. *Semin Pediatr Neurol* 14 (3):108-117.
- Jones KL, and Smith DW. 1997. *Smith's recognizable patterns of human malformation*. Philadelphia: Saunders. xviii, 857 p.
- Julu PO, Kerr AM, Apartopoulos F, Al-Rawas S, Engerstrom IW, Engerstrom L, Jamal GA, and Hansen S. 2001. Characterisation of breathing and associated central autonomic dysfunction in the Rett disorder. *Arch Dis Child* 85 (1):29-37.
- Karlberg J, Albertsson-Wikland K, Nilsson KO, Ritzen EM, and Westphal O. 1991. Growth in infancy and childhood in girls with Turner's syndrome. *Acta Paediatr Scand* 80 (12):1158-1165.
- Kerr AM, Armstrong DD, Prescott RJ, Doyle D, and Kearney DL. 1997. Rett syndrome: analysis of deaths in the British survey. *Eur Child Adolesc Psychiatry* 6 Suppl 1:71-74.
- Kerr AM, and Ravine D. 2003. Review article: breaking new ground with Rett syndrome. *J Intellect Disabil Res* 47 (Pt 8):580-587.
- Kimura J, Tachibana K, Imaizumi K, Kurosawa K, and Kuroki Y. 2003. Longitudinal growth and height velocity of Japanese children with Down's syndrome. *Acta Paediatr* 92 (9):1039-1042.
- Kline AD, Barr M, and Jackson LG. 1993. Growth manifestations in the Brachmann-de Lange syndrome. *Am J Med Genet* 47 (7):1042-1049.
- Krick J, Murphy-Miller P, Zeger S, and Wright E. 1996. Pattern of growth in children with cerebral palsy. *J Am Diet Assoc* 96 (7):680-685.
- Kudo S, Nomura Y, Segawa M, Fujita N, Nakao M, Dragich J, Schanen C, and Tamura M. 2001. Functional analyses of MeCP2 mutations associated with Rett syndrome using transient expression systems. *Brain Dev* 23 Suppl 1:S165-173.
- Laurvick CL, de Klerk N, Bower C, Christodoulou J, Ravine D, Ellaway C, Williamson S, and Leonard H. 2006. Rett syndrome in Australia: a review of the epidemiology. *J Pediatr* 148 (3):347-352.

- Lenko HL, Perheentupa J, and Soderholm A. 1979. Growth in Turner's syndrome: spontaneous and fluoxymesterone stimulated. *Acta Paediatr Scand Suppl* 277:57-63.
- Lenko HL, Soderholm A, and Perheentupa J. 1988. Turner syndrome: effect of hormone therapies on height velocity and adult height. *Acta Paediatr Scand* 77 (5):699-704.
- Lloyd-Jones O. 1940. Race and Stature: A study of Los Angeles school children. *Am J Dis Child* 60:11-21.
- Lyon AJ, Preece MA, and Grant DB. 1985. Growth curve for girls with Turner syndrome. *Arch Dis Child* 60 (10):932-935.
- Majrowski WH, Hearn S, Rohan C, Jenkins S, Cotterill AM, and Savage MO. 1994. Comparison of school nurse and auxologist height velocity measurements in school children with short stature. (The Hackney Growth Initiative). *Child Care Health Dev* 20 (3):179-188.
- Manuel H. 1934. Physical Measurements of Mexican Children in American Schools. *Child Dev* 5 (3):237-252.
- Marinescu RC, Mainardi PC, Collins MR, Kouahou M, Coucourde G, Pastore G, Eaton-Evans J, and Overhauser J. 2000. Growth charts for cri-du-chat syndrome: an international collaborative study. *Am J Med Genet* 94 (2):153-162.
- Martin ND, Smith WR, Cole TJ, and Preece MA. 2007. New height, weight and head circumference charts for British children with Williams syndrome. *Arch Dis Child* 92 (7):598-601.
- Massa G, Vanderschueren-Lodeweyckx M, and Malvaux P. 1990. Linear growth in patients with Turner syndrome: influence of spontaneous puberty and parental height. *Eur J Pediatr* 149 (4):246-250.
- McCammon RW. 1970. *Human Growth and Development*. Springfield, Illinois: Charles C Thomas.
- McGill BE, Bundle SF, Yaylaoglu MB, Carson JP, Thaller C, and Zoghbi HY. 2006. Enhanced anxiety and stress-induced corticosterone release are associated with increased Crh expression in a mouse model of Rett syndrome. *Proc Natl Acad Sci U S A* 103 (48):18267-18272.
- Meredith HV. 1948. *Body Size in Infancy and Childhood: A Comparative Study of Data from Okinawa, France, South Africa, and North America*. *Child Dev* 19 (4):179-195.
- Meredith HV. 1968. Body size of contemporary groups of preschool children studied in different parts of the world. *Child Dev* 39 (2):335-377.

- Meredith HV, and Goldstein MS. 1952. Studies on the body size of North American children of Mexican ancestry. *Child Dev* 23 (2):91-110.
- Moretti P, Levenson JM, Battaglia F, Atkinson R, Teague R, Antalffy B, Armstrong D, Arancio O, Sweatt JD, and Zoghbi HY. 2006. Learning and memory and synaptic plasticity are impaired in a mouse model of Rett syndrome. *J Neurosci* 26 (1):319-327.
- Morris CA, Demsey SA, Leonard CO, Dilts C, and Blackburn BL. 1988. Natural history of Williams syndrome: physical characteristics. *J Pediatr* 113 (2):318-326.
- Motil KJ, Schultz R, Brown B, Glaze DG, and Percy AK. 1994. Altered energy balance may account for growth failure in Rett syndrome. *J Child Neurol* 9 (3):315-319.
- Myreliid A, Gustafsson J, Ollars B, and Anneren G. 2002. Growth charts for Down's syndrome from birth to 18 years of age. *Arch Dis Child* 87 (2):97-103.
- Naeraa RW, and Nielsen J. 1990. Standards for growth and final height in Turner's syndrome. *Acta Paediatr Scand* 79 (2):182-190.
- Nagai T, Matsuo N, Kayanuma Y, Tonoki H, Fukushima Y, Ohashi H, Murai T, Hasegawa T, Kuroki Y, and Niikawa N. 2000. Standard growth curves for Japanese patients with Prader-Willi syndrome. *Am J Med Genet* 95 (2):130-134.
- Nellhaus G. 1968. Head circumference from birth to eighteen years. Practical composite international and interracial graphs. *Pediatrics* 41 (1):106-114.
- Neul J, Fang P, Barrish J, Lane J, Caeg E, Smith E, Zoghbi H, Percy A, and Glaze D. 2008. Specific Mutations in Methyl-CpG-Binding Protein 2 Confer Different Severity in Rett Syndrome. *Neurology*:in press.
- Newell ML, Borja MC, and Peckham C. 2003. Height, weight, and growth in children born to mothers with HIV-1 infection in Europe. *Pediatrics* 111 (1):e52-60.
- Oddy WH, Webb KG, Baikie G, Thompson SM, Reilly S, Fyfe SD, Young D, Anderson AM, and Leonard H. 2007. Feeding experiences and growth status in a Rett syndrome population. *J Pediatr Gastroenterol Nutr* 45 (5):582-590.
- Owen GM. 1973. The assessment and recording of measurements of growth of children: report of a small conference. *Pediatrics* 51 (3):461-466.
- Pan H, and Cole T. 2005. *User's Guide to lmsChartMaker*. London, UK: Medical Research Council.
- Pan H, and Cole TJ. 2004. A comparison of goodness of fit tests for age-related reference ranges. *Stat Med* 23 (11):1749-1765.

- Pankau R, Partsch CJ, Gosch A, Oppermann HC, and Wessel A. 1992. Statural growth in Williams-Beuren syndrome. *Eur J Pediatr* 151 (10):751-755.
- Pankau R, Partsch CJ, Neblung A, Gosch A, and Wessel A. 1994. Head circumference of children with Williams-Beuren syndrome. *Am J Med Genet* 52 (3):285-290.
- Partsch CJ, Dreyer G, Gosch A, Winter M, Schneppenheim R, Wessel A, and Pankau R. 1999. Longitudinal evaluation of growth, puberty, and bone maturation in children with Williams syndrome. *J Pediatr* 134 (1):82-89.
- Paterson DS, Thompson EG, Belliveau RA, Antalffy BA, Trachtenberg FL, Armstrong DD, and Kinney HC. 2005. Serotonin transporter abnormality in the dorsal motor nucleus of the vagus in Rett syndrome: potential implications for clinical autonomic dysfunction. *J Neuropathol Exp Neurol* 64 (11):1018-1027.
- Percy A. 2007a. Rett Syndrome: North American Database. Under Review.
- Percy AK. 2002. Rett syndrome. Current status and new vistas. *Neurol Clin* 20 (4):1125-1141.
- Percy AK. 2007b. Rett Syndrome Needs Your Attention. Clinton, MD: IRSA.
- Phiel CJ, Zhang F, Huang EY, Guenther MG, Lazar MA, and Klein PS. 2001. Histone deacetylase is a direct target of valproic acid, a potent anticonvulsant, mood stabilizer, and teratogen. *J Biol Chem* 276 (39):36734-36741.
- Piro E, Pennino C, Cammarata M, Corsello G, Greci A, Lo Giudice C, Morabito M, Piccione M, and Giuffre L. 1990. Growth charts of Down syndrome in Sicily: evaluation of 382 children 0-14 years of age. *Am J Med Genet Suppl* 7:66-70.
- Prater CD, and Zylstra RG. 2006. Medical care of adults with mental retardation. *Am Fam Physician* 73 (12):2175-2183.
- Preece MA, and Baines MJ. 1978. A new family of mathematical models describing the human growth curve. *Ann Hum Biol* 5 (1):1-24.
- Preece MA, Kearney PJ, and Marshall WC. 1977. Growth-hormone deficiency in congenital rubella. *Lancet* 2 (8043):842-844.
- Price DA, and Albertsson-Wikland K. 1993. Demography, auxology and response to recombinant human growth hormone treatment in girls with Turner's syndrome in the Kabi Pharmacia International Growth Study. International Board of the Kabi Pharmacia International Growth Study. *Acta Paediatr Suppl* 82 Suppl 391:69-74.
- Price WH. 1979. A high incidence of chronic inflammatory bowel disease in patients with Turner's syndrome. *J Med Genet* 16 (4):263-266.

- Ranke MB. 1989. Disease-Specific Growth Charts, Do We Need Them? *Acta Paediatrica* 78 (s356):17-25.
- Ranke MB. 1996. Disease-specific standards in congenital syndromes. *Horm Res* 45 Suppl 2:35-41.
- Ranke MB, and Grauer ML. 1994. Adult height in Turner syndrome: results of a multinational survey 1993. *Horm Res* 42 (3):90-94.
- Ranke MB, Heidemann P, Knupfer C, Enders H, Schmaltz AA, and Bierich JR. 1988a. Noonan syndrome: growth and clinical manifestations in 144 cases. *Eur J Pediatr* 148 (3):220-227.
- Ranke MB, Pfluger H, Rosendahl W, Stubbe P, Enders H, Bierich JR, and Majewski F. 1983. Turner syndrome: spontaneous growth in 150 cases and review of the literature. *Eur J Pediatr* 141 (2):81-88.
- Ranke MB, Stubbe P, Majewski F, and Bierich JR. 1988b. Spontaneous growth in Turner's syndrome. *Acta Paediatr Scand Suppl* 343:22-30.
- Rarick GL, and Seefeldt V. 1974. Observations from longitudinal data on growth in stature and sitting height of children with Down's syndrome. *J Ment Defic Res* 18 (0):63-78.
- Reilly S, and Cass H. 2001. Growth and nutrition in Rett syndrome. *Disabil Rehabil* 23 (3-4):118-128.
- Rempel GR, Colwell SO, and Nelson RP. 1988. Growth in children with cerebral palsy fed via gastrostomy. *Pediatrics* 82 (6):857-862.
- Rett A. 1966. [On a unusual brain atrophy syndrome in hyperammonemia in childhood]. *Wien Med Wochenschr* 116 (37):723-726.
- Rice MA, and Haas RH. 1988. The nutritional aspects of Rett syndrome. *J Child Neurol* 3 Suppl:S35-42.
- Rigby RA, and Stasinopoulos DM. 2004. Smooth centile curves for skew and kurtotic data modelled using the Box-Cox power exponential distribution. *Stat Med* 23 (19):3053-3076.
- Roberts C. 1876. The Physical Requirements of Factory Children. *Journal of the Statistical Society of London* 39 (4):681-733.
- Roberts C. 1878. *Manual of Anthropometry or A Guide to the Physical Examination and Measurement of the Human Body: Containing a Systematic Table of Measurements, and Anthropometrical Chart or Register, and Instructions for Making Measurements on a Uniform Plan.* London: J. & A. Churchill, New Burlington Street.

- Roche AF. 1965. The Stature of Mongols. *J Ment Defic Res* 9:131-145.
- Rongen-Westerlaken C, Corel L, van den Broeck J, Massa G, Karlberg J, Albertsson-Wikland K, Naeraa RW, and Wit JM. 1997. Reference values for height, height velocity and weight in Turner's syndrome. Swedish Study Group for GH treatment. *Acta Paediatr* 86 (9):937-942.
- Royston P, and Wright EM. 2000. Goodness-of-fit statistics for age-specific reference intervals. *Statistics in Medicine* 19 (21):2943-2962.
- Savage SA, Reilly JJ, Edwards CA, and Durnin JV. 1999. Adequacy of standards for assessment of growth and nutritional status in infancy and early childhood. *Arch Dis Child* 80 (2):121-124.
- Schultz R, Glaze D, Motil K, Hebert D, and Percy A. 1998. Hand and foot growth failure in Rett syndrome. *J Child Neurol* 13 (2):71-74.
- Schultz RJ, Glaze DG, Motil KJ, Armstrong DD, del Junco DJ, Hubbard CR, and Percy AK. 1993. The pattern of growth failure in Rett syndrome. *Am J Dis Child* 147 (6):633-637.
- Sekul EA, Moak JP, Schultz RJ, Glaze DG, Dunn JK, and Percy AK. 1994. Electrocardiographic findings in Rett syndrome: an explanation for sudden death? *J Pediatr* 125 (1):80-82.
- Sippell WG, Partsch CJ, and Wiedemann HR. 1989. Growth, bone maturation and pubertal development in children with the EMG-syndrome. *Clin Genet* 35 (1):20-28.
- Stevens CA, Hennekam RC, and Blackburn BL. 1990. Growth in the Rubinstein-Taybi syndrome. *Am J Med Genet Suppl* 6:51-55.
- Stevenson RD, Conaway M, Chumlea WC, Rosenbaum P, Fung EB, Henderson RC, Worley G, Liptak G, O'Donnell M, Samson-Fang L and others. 2006. Growth and health in children with moderate-to-severe cerebral palsy. *Pediatrics* 118 (3):1010-1018.
- Stuart HC, and Meredith HV. 1946. Use of Body Measurements in the School Health Program. *Am J Public Health Nations Health* 36 (12):1365-1386.
- Styles ME, Cole TJ, Dennis J, and Preece MA. 2002. New cross sectional stature, weight, and head circumference references for Down's syndrome in the UK and Republic of Ireland. *Arch Dis Child* 87 (2):104-108.
- Sun YE, and Wu H. 2006. The ups and downs of BDNF in Rett syndrome. *Neuron* 49 (3):321-323.
- Suwa S. 1992. Standards for growth and growth velocity in Turner's syndrome. *Acta Paediatr Jpn* 34 (2):206-220; discussion 221.

- Szudek J, Birch P, and Friedman JM. 2000. Growth in North American white children with neurofibromatosis 1 (NF1). *J Med Genet* 37 (12):933-938.
- Takahashi S, Ohinata J, Makita Y, Suzuki N, Araki A, Sasaki A, Murono K, Tanaka H, and Fujieda K. 2007. Skewed X chromosome inactivation failed to explain the normal phenotype of a carrier female with MECP2 mutation resulting in Rett syndrome. *Clin Genet*.
- Tanner JM, Goldstein H, and Whitehouse RH. 1970. Standards for children's height at ages 2-9 years allowing for heights of parents. *Arch Dis Child* 45 (244):755-762.
- Tanner JM, Lejarraga H, and Cameron N. 1975. The natural history of the Silver-Russell syndrome: a longitudinal study of thirty-nine cases. *Pediatr Res* 9 (8):611-623.
- Tanner JM, Lejarraga H, and Healy MJ. 1972. Within-family standards for birth-weight. *Lancet* 2 (7790):1314-1315.
- Tanner JM, Whitehouse RH, and Takaishi M. 1966a. Standards from birth to maturity for height, weight, height velocity, and weight velocity: British children, 1965. I. *Arch Dis Child* 41 (219):454-471.
- Tanner JM, Whitehouse RH, and Takaishi M. 1966b. Standards from birth to maturity for height, weight, height velocity, and weight velocity: British children, 1965. II. *Arch Dis Child* 41 (220):613-635.
- Thommessen M, Kase BF, and Heiberg A. 1992. Growth and nutrition in 10 girls with Rett syndrome. *Acta Paediatr* 81 (9):686-690.
- U.S. Department of Health and Human Services. 2008. The CDC Growth Charts for Children with Special Health Care Needs. <http://depts.washington.edu/growth/cshcn/text/page2a.htm>. [Accessed April 2008]
- van Buuren S, and Fredriks M. 2001. Worm plot: a simple diagnostic device for modelling growth reference curves. *Stat Med* 20 (8):1259-1277.
- Villard L. 2007. MECP2 mutations in males. *J Med Genet* 44 (7):417-423.
- Voss LD, Bailey BJ, Cumming K, Wilkin TJ, and Betts PR. 1990. The reliability of height measurement (the Wessex Growth Study). *Arch Dis Child* 65 (12):1340-1344.
- Wade AM, and Ades AE. 1998. Incorporating correlations between measurements into the estimation of age-related reference ranges. *Statistics in Medicine* 17 (17):1989-2002.

- Wang H, Chan SA, Ogier M, Hellard D, Wang Q, Smith C, and Katz DM. 2006. Dysregulation of brain-derived neurotrophic factor expression and neurosecretory function in *Mecp2* null mice. *J Neurosci* 26 (42):10911-10915.
- Waterlow JC, Buzina R, Keller W, Lane JM, Nichaman MZ, and Tanner JM. 1977. The presentation and use of height and weight data for comparing the nutritional status of groups of children under the age of 10 years. *Bull World Health Organ* 55 (4):489-498.
- Watson P, Black G, Ramsden S, Barrow M, Super M, Kerr B, and Clayton-Smith J. 2001. Angelman syndrome phenotype associated with mutations in *MECP2*, a gene encoding a methyl CpG binding protein. *J Med Genet* 38 (4):224-228.
- WHO. 2006. WHO Child Growth Standards based on length/height, weight and age. *Acta Paediatr Suppl* 450:76-85.
- Wisniewski A, Milde K, and Stupnicki R. 2006. [Spontaneous growth of girls with Turner's syndrome until 6 years of age]. *Endokrynol Diabetol Chor Przemiany Materii Wieku Rozw* 12 (1):7-11.
- Witt DR, Keena BA, Hall JG, and Allanson JE. 1986. Growth curves for height in Noonan syndrome. *Clin Genet* 30 (3):150-153.
- Wollmann HA, Kirchner T, Enders H, Preece MA, and Ranke MB. 1995. Growth and symptoms in Silver-Russell syndrome: review on the basis of 386 patients. *Eur J Pediatr* 154 (12):958-968.
- Wollmann HA, Schultz U, Grauer ML, and Ranke MB. 1998. Reference values for height and weight in Prader-Willi syndrome based on 315 patients. *Eur J Pediatr* 157 (8):634-642.
- Wright E, and Royston P. 1997. A comparison of statistical methods for age-related reference intervals. *Journal of the Royal Statistical Society, Series A* 160:47-69.
- Xinhua B, Shengling J, Fuying S, Hong P, Meirong L, and Wu XR. 2008. X Chromosome Inactivation in Rett Syndrome and Its Correlations With *MeCP2* Mutations and Phenotype. *J Child Neurol* 23 (1):22-25.
- Zachmann M, Sobradillo B, Frank M, Frisch H, and Prader A. 1978. Bayley-Pinneau, Roche-Wainer-Thissen, and Tanner height predictions in normal children and in patients with various pathologic conditions. *J Pediatr* 93 (5):749-755.

## BIOGRAPHICAL SKETCH

Daniel C. Tarquinio D.O., is currently a postdoctoral research fellow in the Department of Child Neurology at the University of Florida, and research postdoc at the Civitan International Research Center at the University of Alabama, Birmingham. Dr. Tarquinio received his degree in osteopathic medicine from the University of New England College of Osteopathic Medicine in Biddeford, Maine, in 2004. He completed his pediatrics residency at the University of Florida in June of 2007. He joined the Advanced Postgraduate Program for Clinical Investigation in July of 2007 and is currently developing growth charts for patients with Rett syndrome and evaluating genotype-phenotype correlations in this syndrome. He will go on to a child neurology fellowship at Boston Medical Center in July 2008. His research interests include Rett syndrome, drug induced osteopenia, and diagnosis and management of epilepsy.