A Clinical Trial of TrenDx: An Automated Trend-Detection Program

by

Phuc Van Le

B.S., Electrical Engineering and Computer Science
MIT, 1996

Submitted to the Department of Electrical Engineering and Computer Science in Partial Fulfillment of the Requirements for the Degree of

Master of Engineering

in
Electrical Engineering and Computer Science
at the

Massachusetts Institute of Technology

September, 1996

Signature of Author .....
Department of Electrical Engineering and Computer Science
August 26, 1996

Certified By .....
Professor Peter Szolovits
Department of Electrical Engineering and Computer Science
Thesis Supervisor

Accepted by .....
Eric R. Morgenthaler
Chairman, Department Committee on Graduate Students

A Clinical Trial of TrenDx: An Automated Trend-Detection Program

by

Phuc Van Le

Submitted to the Department of Electrical Engineering and Computer Science

September, 1996

in Partial Fulfillment of the Requirements for the Degree of

Master of Engineering

in

Electrical Engineering and Computer Science

Abstract

We carried out a trial designed to assess the performance of a computer program called TrenDx at the task of diagnosing growth disorders in children, given a limited set of data. We compared the performance of TrenDx to that of physicians performing the same task. The task consisted of reviewing a growth chart of a patient and deciding whether the patient should be referred to a growth clinic for a possible growth problem, giving a preliminary diagnosis, and choosing the time which it would have been appropriate to refer the child. The test cases consisted of the height, weight, and bone-age data of 95 children that had been referred to the Boston Children's Hospital. The patient cases were organized into packets of 10 and distributed to physicians. 22 physicians participated. Two gold-standards were used, the medical record diagnosis and the opinion of a pediatric endocrinologist. A scoring algorithm was devised based on the responses of the pediatric endocrinologists. The performance of the experts compared to the medical record indicated that it was possible to accomplish the task that we had designed. When compared to the medical record diagnosis, TrenDx and the physicians performed similarly in terms of referring patients, but the physicians chose the correct diagnosis more often. Compared to the experts, the physicians performed better than TrenDx in terms of referral decision and score.

Thesis Supervisor: Peter Szolovits.

Title: Professor, Electrical Engineering and Computer Science

Acknowledgments

It is 6 in the morning, and I've slept 3 hours in the past 2 days. I'm writing this to remind myself of what I can accomplish, given the right motivation. I think I'm getting carpal tunnel syndrome.

I don't have a wise sayings or quotes. I would just like to thank everyone who has helped me to be what I am today and will become tomorrow.

• My parents and family are the most important people in the world to me. Without you, I would not exist. Thank you.
• My two thesis advisors, "Venerable Grandfather" Peter Szolovits and Isaac "Zak" Kohane, and Ira Joseph Haimowitz, who hired a naive sophomore who had no clue what he was getting himself into. I owe you three more than I can fit onto this page.
• I would also like to thank the people that participated in the trial, both the subjects and the experts.
• To Jane and Yao, who helped me see what being a doctor really meant.

• And to the two groups of people without whom my time at MIT would have been truly miserable :

To my friends on Conner 5 - May you always be able to just sit around and talk
To my volleyball buddies - 1,2,3 "Phil!"

• I would also like to thank the Whitaker Foundation for supporting me for the year and summer that I was working on this .

Table Of Contents

Table of Figures and Tables

1. Introduction

In many domains, experts can judge the state of a process by examining the data produced by the process and then matching these data to stereotypical patterns specific to different states. Haimowitz defines the term **trend** as a clinically significant pattern in a sequence of time-ordered data (Haimowitz). Thus, **trend-detection** is the task of judging the state of a process by matching the data produced by the process to different trends. This trend-detection is applicable in many domains. We are particularly interested in evaluating its application to the domain of medicine, where the task of diagnosis can be viewed as matching a patient's findings to trends that are typical of certain conditions.

The computer program TrenDx was developed by Haimowitz based on this premise - "that a computer program with knowledge of time-varying constraints on measured data can be used for automated trend detection."(Section 1 Haimowitz). TrenDx was tested in a limited manner as part of its original development. This research is a more stringent evaluation of TrenDx at the task of diagnosing growth disorders in children from their height, weight, and bone-age data.

1.1. Pediatric Growth Monitoring

Health care systems have been under intense pressure to become more efficient and cost-effective. This has had many consequences, including forcing physicians to see more patients and spending less time with each individual patient. Often, these time pressures have led to care that is less than optimal. The direct experience of one of the advisors of this thesis has found that children with growth problems are not being diagnosed, and therefore treated, in a timely fashion. Similarly, patients with normal growth are sometimes referred to tertiary care centers for expensive work-ups because their physicians misdiagnose normal patterns of growth.

Being able to diagnose referrals correctly more often would improve health care because patients with abnormal growth would be diagnosed and treated before their conditions become grossly apparent and possibly untreatable. Reducing mortality and morbidity while lowering the cost of care by eliminating needless referrals are the ultimate goals of this research. However, these goals must be accomplished without increasing the physician's workload.

Thus, the task of pediatric growth monitoring is an important one in which the performance of pediatricians can possibly be improved upon by applying automated trend-

detection. In the growth clinic at the Boston Children's Hospital, pediatric endocrinologists routinely quiz one another by asking their peers to make a diagnosis solely from the information contained on a patient's growth chart. Furthermore, Becker notes that "Thus, the monitoring of linear growth is a remarkably cost-effective screening for the documentation of good health or for determining the presence or severity of chronic disease." (Becker)

1.2. Introduction to TrenDx

TrenDx is described in more detail in section 2. It is also described in its entirety in (Haimowitz; Haimowitz and Kohane; Haimowitz and Kohane). Briefly, knowledge-engineers and domain specialists outline stereotypical patterns in temporal data using the modeling language incorporated into TrenDx. These patterns, called trend-templates, consist of partially ordered temporal intervals, each with constraints on all the data that fall into that particular interval. Data within each interval are matched to the constraints associated with the interval using linear regression techniques, producing an error-score which indicates how well the data match to the particular trend-template. Trend-templates are grouped into competing sets called monitor sets, from which the best-scoring trend-template is considered to be the current hypothesis or diagnosis.

1.3. Evaluation

A rigorous evaluation is an essential part of the development of any system (Heathfiled and Wyatt; Waterman). A statement of the exact goals of this evaluation is necessary.

- We wish to assess the performance of a computer program, TrenDx, at the task of recognizing growth disorders in children from a limited data set and then referring the child to a specialist if appropriate.

The immediate goal of the evaluation is to show that TrenDx can perform this task with some expertise. The long-term goal of the development of TrenDx is to create a smart monitoring system that can detect the state of a process by recognizing significant trends in the data produced by the process and cause some type of action to be taken if the data suggest an undesirable state. In addition, we wish to improve the state of expert-systems concerning the incorporation of temporal knowledge into their knowledge base.

Evaluations of decision-support systems can be divided into two categories, laboratory trials and field trials. Laboratory trials are carried out during the earlier phase of

development of a system and are characterized by more controlled conditions such as retrospectively chosen cases, "clean" data, and users who are very familiar with the systems. A handful of systems are carried through to the stage where field trials are appropriate. In field trials, the environment is much less controlled, causing new problems to arise. These generally include a wider range of cases, novice users, a larger setting, different outcome measures, and potential legal and ethical considerations concerning the use of the output generated by these programs.

At this stage in the development of TrenDx, a laboratory trial of the performance of the program is appropriate. Thus, we have designed a retrospective clinical trial of TrenDx in which the performance of TrenDx is compared to human experts - physicians. Both TrenDx and the human subjects, collectively referred to as the test subjects, are given the complete set of height, weight, and bone-age data available for a patient. The test subjects must decide if they would recommend that the patient be referred to the endocrine division to be worked-up for a possible growth problem. The test subjects are also asked to give a preliminary diagnosis and choose the age at which the referral should have been made. All answers are measured against a gold-standard expert panel of pediatric endocrinologists. In addition, the decisions of the gold-standard expert panel are compared to the diagnosis written in the patient's medical record.

This task can be viewed as that of one physician giving a second opinion to a colleague who suspects that one of his or her patients has a growth problem. The task is also analogous to that of the individual in a managed care organization who has to decide whether a referral to a tertiary care center is warranted. In section 3.4, the characteristics of this trial are compared to some of the other types of evaluations of decision-support systems that have been carried out.

1.4. Guide to this Thesis

Section 2 describes TrenDx and the trend-representation language which TrenDx uses. Intervals, value constraints, trend-templates, and monitor-sets are all explained well enough for someone unfamiliar with the program to understand the work presented in this thesis. The section also discusses the most recent evaluations of TrenDx and some of their weaknesses.

In section 3, all the work not directly related to the development of TrenDx and the trend-templates used in this trial are presented. This work includes the collection of test cases, the transcription of the data into electronic format, the creation of test packets for

distribution to the participants, and the task which the participants in the trial must complete. The two gold-standards for this evaluation are also discussed, as well various measures which were used to evaluate the performance of TrenDx and the 22 human participants. Finally, a comparison to other evaluations of expert-systems / decision-support systems is made.

The development of TrenDx comprises the bulk of Section 4. The work that was done can be categorized as programming improvements into TrenDx, creating trend-templates that model the different growth states/disorders involved in the trial, and engineering and refining of the trend-template parameters to achieve a desired level of performance.

Section 5 presents the results of the trial. Some of these include the results of comparisons between the gold-standards and the decisions made by the subjects, comparisons between the gold-standards themselves, and the results of various changes to the mechanism used to interpret the scoring system of TrenDx as actions, such as "refer" or "deny referral."

The discussion of the results can be found in section 6. Conclusions about the trial and the future work / uses of TrenDx are in section 7.

Appendix A - Patient and Subject Result Tables, contains the entire listing of the results of the trial, by patient case and by human subject. Appendix B - Packet Directions and Sample Chart, includes the directions presented to the participants and a sample chart similar to the ones on which they indicated their responses. Appendix C - Trend Template LISP Code, shows the LISP code for all of the trend-templates used the the trial, for those who are interested in such things.

There are several conventions used in this thesis. For example, the word subjects is used to refer to both TrenDx and the physicians that participated in the trial. It does not include the pediatric endocrinologists who provided one of the gold-standards. The terms human subjects, participants, volunteers, and physicians all refer to the human subjects who participated in the trial, even though not all of them are physicians. The gold-standard provided by the pediatric endocrinologists can also be referred to as the expert opinion or the expert decisions. Patient cases, cases, and patients all refer to the patient cases that were reviewed by the subjects and the experts in this trial. Finally, references to tables and figures will usually only contain the caption name and number, such as Table 1. However, if, there is a possibility for ambiguity, then the title of the reference will also be included, such as Table 1:TrenDx matching results on tertiary care patients, from (Haimowitz).

2. TrenDx

The most detailed description of TrenDx can be found in (Haimowitz). The brief description presented here is provided to supplement the discussion of improvements to the program and the engineering choices made in the representation of growth disorders that is presented later in this thesis.

> TrenDx diagnoses trends by matching time-ordered process data to the competing trend templates in each monitor set assigned to that process. TrenDx begins matching by instantiating each trend template for the monitored process. TrenDx then computes all temporal worlds in which the currently interpreted data may be assigned to intervals of the trend template. Each temporal world represents a different hypothesis for the same trend template. For each hypothesis, TrenDx assigns the data to the appropriate trend template intervals and computes the matching scores of the relevant value constraints. The value constraint scores are combined to an overall error score for each hypothesis. Finally, the top hypotheses for each trend template are maintained via a beam search. The output of TrenDx is a list of the top hypotheses for each trend template within a monitor set, with the score of each hypothesis. (Section 4 Haimowitz).

2.1. TUP

TrenDx manipulates temporal assertions and queries using the Temporal Utility Package, or TUP, developed by Kohane (Kohane). TUP is a set of temporal utilities which allow TrenDx to represent time points and intervals, as well as reason about uncertainty in temporal distances. For example, TUP allows the expression of time intervals with uncertain endpoints. Furthermore, TUP provides the ability to deal with alternate temporal worlds. Alternate temporal worlds are contexts in which different temporal assertions apply. For example, say a user specified that Event A occurred sometime between January 1, 1990 and December 31, 1990. Then say that another event, Event B, occurred on July 1, 1990. From the known information, Event B could occur before, at the same time as, of after Event A. TUP allows the formulation of alternate temporal worlds wherein each of the relationships between Event A and Event B is asserted.

The ability to deal with time is an important aspect of any medical decision-support system. This is especially true in the domain of pediatric growth monitoring. One of the conclusions of the INTERNIST-1 program was that the inability to incorporate temporal

information into the program was one of its major weaknesses (Miller, Pople and Myers). The use of TUP enables TrenDx to incorporate this type of knowledge into its models and reasoning.

2.2. Temporal Constraints

The temporal aspects of a trend-template include **landmark points** and **intervals**. Landmark points represent significant events during a process. For example, BIRTH and DEATH would be considered landmark points in the process of a person's life. The temporal distance between landmark points can be specified with a set of lower and upper bounds (MIN MAX), indicating the minimum and maximum difference in time between the two points. For example, for a person who lived exactly 80 years, the temporal distance between their BIRTH and DEATH would be represented by ((years 80) (years 80)). This specifies that both the minimum and maximum distance between the BIRTH and DEATH landmark points is eighty years, meaning that exactly eighty years separated the two points. To represent the fact that a particular person died sometime between the ages of 13 and 20, the (MIN MAX) set would look like ((years 13) (years 20)). Figure 1shows a timeline with the landmark point DEATH occurring 13 to 20 years after BIRTH.

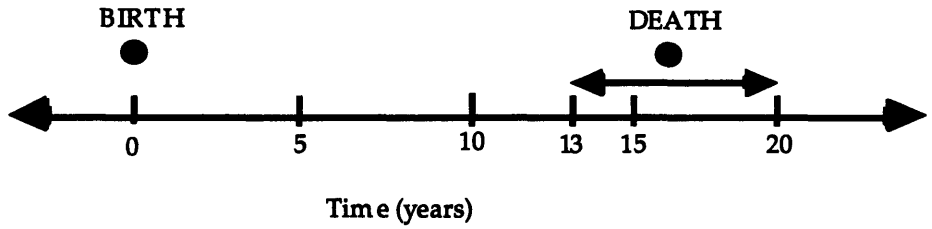


Figure 1: Two landmark points in the life of an individual

Intervals are used to represent different phases of a process. In TrenDx, intervals are represented by a Begin point and an End point. The temporal distance between the two represents the duration of the interval. Begin and End points can also have uncertainty ranges associated with their relation to another time point, either a landmark point or an interval Begin/End point. In Figure 2 we extend the previous example by adding an interval representing the time period over which the person lived.

As suggested before, interval Begin/End points can be defined relative to other interval Begin/End points. One common relationship between two intervals occurs when

one interval directly follows another. In that case, the End point of the first interval is equal in time to the Begin point of the second.

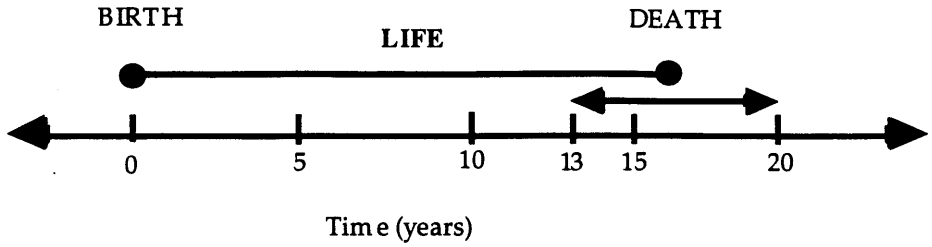


Figure 2: An Interval Representing the Life of an Individual.

In some cases, it is more relevant to view time backwards beginning from a particular time point. Most often, the particular time point is the present. For example, to determine whether a child is obese, the state of the child at the present time is much more important than the child's state 2 years ago. Similarly, to determine whether the child has a fever, it is the current temperature of the child that is relevant. The need to model this view of time led to the development of what is called a Now-Based trend-template.

A Now-Based trend-template has an additional anchor point called 'now' which represents the most recent data point. The 'now' point is updated to be equal in time to each new data point that is processed. This allows a user to design a trend-template and set the interval Begin/End points relative to 'now.' Figure 3 is an example of a Now-Based trend-template that models a child with fever. The duration of the Fever interval is represented relative to 'now' and extending back somewhere between 30 minutes to 2 hours.

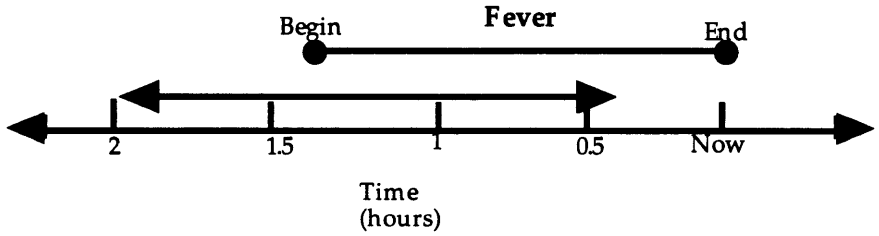


Figure 3:A Now-Based Trend-Template Interval.

2.3. Value Constraints

A value constraint is composed of two main components. The first component is a function that maps the data in the interval to a time-indexed real-valued sequence of

numbers. The second component is a linear regression model describing the pattern of the output of the first component.

The first component can be as simple as a function that simply returns the numerical value of each time-stamped datum. Or it may return some more complicated function of several data points of different types. An example of a simple function may just return a sequence of the height Z-scores, which represent how many standard deviations the child is from the average height at that particular age. A more complicated function is one which returns the ratio of weight to weight for height-age, which we call Build.

The second component, which we call the error function, can be one of a set of up to 2nd order polynomial functions. The constant and 1st order polynomial functions can even specify the value and slope of the function to be matched against the data. In other words, the second component can specify that the sequence returned by the first component should be matched to a constant of known or unknown value, a line of some known or unknown slope, or a 2nd order polynomial curve with first and second derivatives positive or negative. Figure 4 gives some examples of value constraints.

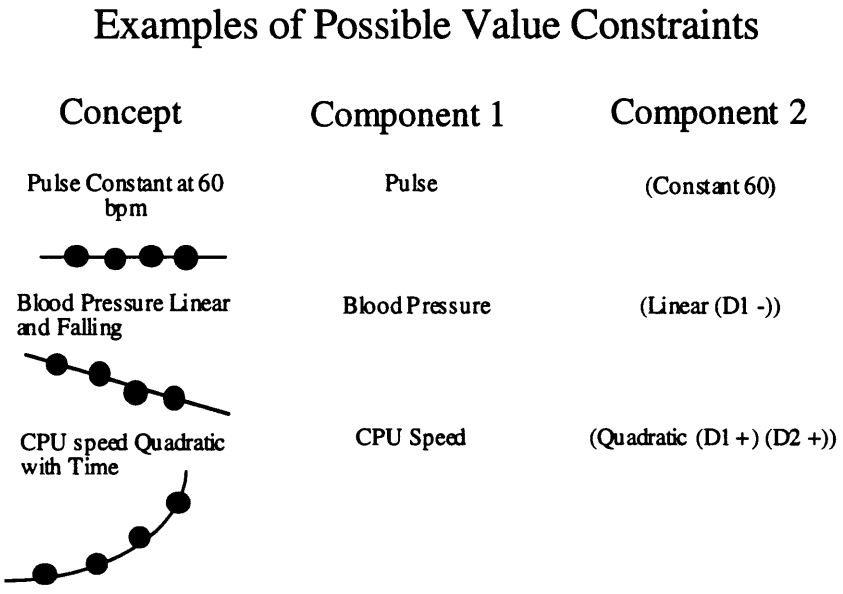


Figure 4: Examples of Possible Value Constraints

The sequence is matched to the error function, producing the residual mean square error. This is repeated for all the value constraints of all intervals of the trend-template. The residuals are combined by using a weighted average of the fits to each value constraint, where the weight is proportional to the number of data points in each value constraint.

Haimowitz suggests using the Mean Absolute Percent Error, or MAPE, because parameters with larger ranges will have a larger variance of residuals (Section 4.4.1 Haimowitz). Using the percent error allows one to combine the errors from different value constraints correctly. However, the use of MAPE is not applicable when the expected value of the parameter is zero. This is discussed in section 4.3.

Originally, TrenDx used simple upper and lower bounds to determine whether a data point matched to an interval well (Haimowitz and Kohane; Haimowitz and Kohane). Only one trend-template was active at any single time and it was considered the current hypothesis. When a data point exceeded the upper or lower thresholds, another trend-template was triggered or some other action, such as an alarm, was taken.

This style of value-constraint matching was known as Constraint-Based TrenDx. It was originally designed to mimic the stream of thought of an expert. For example, an expert would start off with the hypothesis that the child was normal. Then, if the child's height was too low, the expert would then discard the current hypothesis and consider Constitutional Delay of Puberty as the current hypothesis.

Constraint-Based TrenDx suffered from many drawbacks. Similar to other threshold-trigger systems, Constraint-Based TrenDx was brittle. For example, if the lower threshold on a value was -2.0, then a value of -2.1 would cause the current hypothesis to be discarded, while a value of -2.0 would not. In addition, there was no difference between having a data point that was exactly normal and one which fell just within the allowable threshold. Regression-Based TrenDx was developed to remedy some of those problems.

2.4. Trend Templates and Hypotheses

A trend-templates represents an overall state of a process. It is comprised of a partially ordered set of intervals, each with one or more value constraints associated with them. When a trend-template is instantiated for a patient, a temporal context is created in which the BIRTH landmark point of the patient is anchored to the time and date that the patient was born. When alternate temporal worlds are possible, the context branches, producing multiple child contexts that represent each of the possible temporal worlds.

Thus, each hypothesis for a patient consists of a trend-template, a temporal context, and an assignment of the patient data to the intervals of the trend-template that is dependent on the temporal context.

Figure 5, is an example of the complete trend-template for Normal growth in males. Notice that this trend-template only tries to model life up to and a little beyond the point where growth stops.

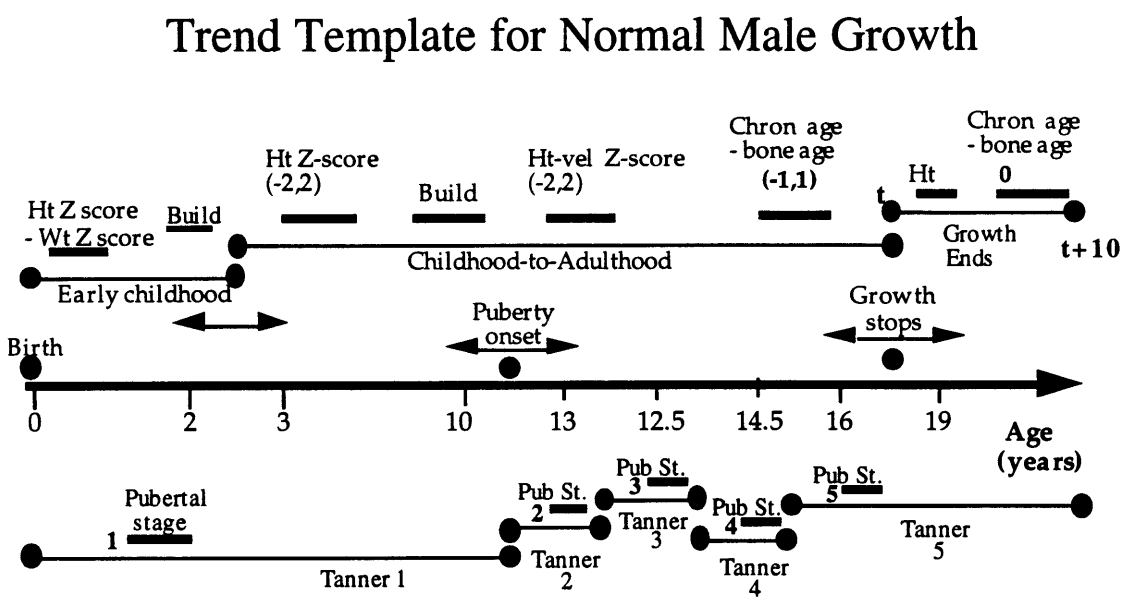


Figure 5: Trend-Template for Normal Male Growth

2.5. Monitor Sets

Trend-templates are grouped into competing sets called monitor sets. Haimowitz defines a monitor set as a set of trend templates forming a clinical context for monitoring. Error scores from each trend-template can be compared with the error scores from the other trend-templates within the monitor set to determine which trend-template has the best fit to the data.

Since a monitor set describes a group of competing trend-templates, it typically consists of a 'normal' trend-template that represents the process in its normal state and one or more 'abnormal' trend-templates that represent fault states.

2.6. Previous Evaluations

In (Haimowitz and Kohane), the performance of constraint-based TrenDx was evaluated, using a panel of three pediatric endocrinologists as a gold-standard. Out of the 20 test cases, 14 were diagnosed correctly by TrenDx.

Haimowitz later performed an evaluation of the regression-based TrenDx as part of his thesis (Section 5.1 Haimowitz). This version of TrenDx will be denoted the thesis-

TrenDx. In the evaluation of thesis-TrenDx, two sets of test cases were used. The first consisted of 30 cases. Of these, there were 4 Normal, 10 had Constitutional Delay of Puberty, 3 were diagnosed with Early Puberty, and 13 suffered from Growth-Hormone Deficiency. The monitor set consisted of trend-templates for Normal growth, Early Puberty, and Constitutional Delay of Puberty. For the Growth-Hormone deficiency cases, a diagnosis of Constitutional Delay of Puberty was considered correct.

Table 1 is adapted from (Haimowitz). It shows the performance of thesis-TrenDx on the 30 cases. The column labeled 'Persistent Gap,' 'Single Gap,' and 'Union' describe the criteria used to determine whether the Constitutional Delay or Early Puberty trend-templates overtake the Normal Growth trend-template. In essence, 'Persistent Gap' requires that the Normal Growth trend-template scores somewhat worse than either of the other two trend-templates for two consecutive time points. 'Single Gap' requires that the Normal Growth trend-template scores *significantly* worse than either of the other two trend-templates. 'Union' is the union of the both Persistent and Single gap triggering mechanisms. The use of Persistent Gap is to try to reduce some of the brittleness inherent in using threshold-triggering, as discussed in Section 2.3. Formally:

- Persistent gap: For two consecutive visits, the best hypothesis for P scored 0.8 or less times the best hypothesis for Normal Growth.
- Single gap: For one visit, the best hypothesis for P scored 0.6 or less times the best hypothesis for Normal Growth.
- Union: Either Persistent Gap or Single Gap.

Where P is the trend-template for either Constitutional Delay of Puberty or Early Puberty. The sensitivity and specificity of thesis-TrenDx was calculated for the diagnosis of both Constitutional Delay of Puberty and Early Puberty as follows:

sensitivity for P =(number with P and TrenDx triggers P) /

(Number with P)

specificity for P = (number without P and TrenDx does not trigger P) /

(Number without P)

Notice that the sensitivities are low, ranging from 0.0 to 0.66, while the specificities are good, ranging from 0.77 to 1.0. Clearly, the results could have been changed by changing the triggering criteria. Lowering the threshold value used to trigger

the alternate trend-templates would most likely have increased the sensitivity at the cost of decreasing specificity. This is a common tradeoff that is described by a Receiver Operator Characteristic Curve(Pagano and Gauvreau).

Table of Results from the Previous Evaluation of TrenDx

		Persistent Gap			Single Gap			Union		
Disorder	#	Norm	Cons Delay	Early Pub	Norm	Cons Delay	Early Pub	Norm	Cons Delay	Early Pub
Normal	4	4	0	0	3	0	1	3	0	1
Cons. Delay	10	7	3	0	5	5	1	4	6	1
Early Puberty	3	3	0	0	1	0	2	1	0	2
GH Deficiency	13	4	6	4	5	6	3	4	6	4
Cum Sensitivity	30		0.39	0.00		0.48	0.66		0.52	0.66
Cum. Specificity	30		1.00	0.85		1.00	0.81		1.00	0.77

Table 1:TrenDx matching results on tertiary care patients, from (Haimowitz)

The second set of test cases used consisted of 20 cases taken from the files of a general pediatrician. These cases were considered 'normal' by the pediatrician, but were not reviewed by any pediatric endocrinologists. The table presenting the results of those 20 additional cases (Table 2 Section 5.1 Haimowitz) has several small inconsistencies and will not be included here. The results of the second set of test cases has very little effect on the Cumulative Specificity of the program.

There are several characteristics of the trial that incorporate bias or weaken the strength of the results. One of the most obvious problems with the trial is the small number of trend-templates. In fact, combining the diagnosis of Growth-Hormone Deficiency with Constitutional Delay of Puberty is very suspect. While both conditions exhibit short stature as a result of delayed growth, Constitutional Delay of growth is generally considered a benign condition and not worthy of a growth clinic referral unless the patient has a very extreme case. On the other hand, Growth Hormone Deficiency is a true, secondary

disturbance of growth that can be related to even more serious problems such as a craniopharyngioma (brain tumor). A second problem with the trial is that some of the Growth-Hormone Deficiency cases had been used previously as test cases in a previous trial (Haimowitz and Kohane) of Constraint-Based TrenDx. Thus, the test was biased because those cases had influenced the previous development of the Regression-Based TrenDx. However, the design of the test was appropriate to the level of development of the program. Formal evaluations, such as double-blind, comparative studies are less appropriate at early levels of development because the experts should probe the inference engine and knowledge base of the program, not just be worried about final program results.

In summary, the results of the evaluation were promising. However, they indicated that further work would be necessary to improve the performance of the program and that a more complete evaluation of the program would be appropriate at that time.

3. Methods

We have conducted an experimental trial of a revised version of TrenDx using 95 newly-collected patient cases taken sequentially from the patients referred to the Endocrine Division at the Boston Children's Hospital. The cases were screened for inclusion into the trial, then the data in the cases were transcribed and placed into packets for each human subject. Each packet contained growth charts for 10 test cases, distributed to try to achieve an even distribution of test cases. The packets were distributed as participants were found. The participants consisted of physicians, medical students, and a registered nurse. Over 80 packets were created and distributed, but only 22 were returned. The medical record diagnosis for each of the cases was obtained and used as one gold-standard. A pediatric endocrinologist not involved with the development of the program provided a second gold-standard for the trial.

TrenDx was updated and improved, independent of the test cases that were going to be used for the trial. New trend-templates were designed to try to take into account a new variation of triggering as well as improve over the performance of thesis-TrenDx. In fact, this trial is the first formal test of any trend-templates other than those for Normal Growth, Early Puberty, or Constitutional Delay of Puberty. Several problems with the use of 'now-based trend-templates were uncovered and solved. The development of TrenDx is discussed in Section 4.

3.1. Packet Creation and Distribution

3.1.1. Test Case Criteria

The occurrence of growth abnormalities in the general population is too low to use it as a test population - for example, Congenital Growth Hormone Deficiency occurs in approximately 1 out of 16,000 people (Kaplan). However, arbitrarily picking a certain number of cases of different pathologies is difficult to justify because the numbers would not reflect the relative frequency of the different pathologies that are referred to the growth clinic. Consequently, we decided to take cases randomly chosen from patients that had been referred to the Division of Endocrinology at the Boston Children's Hospital. This population has both a high proportion of abnormal patients, as well as normal patients that had some characteristics of abnormal patients. Some of the cases were not referred to the growth clinic because of a suspicion of growth abnormality, but were referred for some other reason. We consider these cases to be "normal" as well.

Recall that the motivation of the program was to improve the performance of physicians in the domain of pediatric growth monitoring by helping to diagnose children with growth disorders and by reassuring the physician that a normal child is truly normal. One might argue that using a referral population only tests the ability of the test subjects to do the latter task and not the former - i.e. that using this test population only allows us to catch patients with normal growth who were referred incorrectly and that it does not allow us to catch patients with abnormal growth who were not referred. However, a child who does have a true growth disorder will become more symptomatic as time progresses. In fact, most of the disorders cause the children to fall 2 or 3 standard deviations below the mean height for that age and to keep falling away from their peers. Thus, it is a somewhat simple task to recognize a child with a growth disorder if the child has been suffering from it long enough. Our aim is to improve the timing of the diagnosis and referral to minimize morbidity in these children.

Other evaluations have used referral cases in a similar fashion. Heckerman uses referral cases in his evaluation of the Pathfinder program (Heckerman and Nathwani). In an evaluation of four decision-support systems by (Berner et al.), the test cases consisted of referrals to a "...group of 10 nationally recognized consultants in the fields of general internal medicine, eight subspecialties of internal medicine, and neurology..." They chose to use the referral cases to ensure that the cases were diagnostically challenging.

3.1.2. Patient Record Collection

To decide whether to accept a patient case into the trial, the physical record was scanned and the age at which the child was first referred to the clinic was noted. If the record contained data for at least three time points before the referral date, we tentatively accepted the case. A second criterion was that only patients that were referred more than one year ago were accepted. This was done to allow the true clinical outcome of the patient to be used as one of the standards. We then screened out previously diagnosed cancer patients. There were two reasons for this decision. First, both cancer and its treatments have complex effects on growth. Second, patients receiving cancer treatment were assumed to be under close clinical observation and the original motivation of the trial was to catch cases which were diagnosed late because of time pressures on the pediatrician. Of the patient cases that were screened, approximately 70% had enough data and the right background for us to accept the case. We collected approximately 120 patient records and numbered them consecutively starting at 2000.

3.1.3. Data-Entry

All height, weight, and bone-age data of the child were entered into a spreadsheet. Recall that only the data available before the date the child was seen at the clinic were used in this trial. The growth clinic usually acquires the information from the referring doctor by calling the referring doctor's office and verbally transcribing the data or by receiving a faxed copy of the patient's growth chart. Therefore we took photocopies of the patient's growth charts and the verbal transcriptions.

Once the information was entered into a spreadsheet, it was transformed into TrenDx-readable LISP code by a series of programs. Here is some example code:

```
(make-patient 'BOY-PATIENT :id 3 :dob "1/1/80"
    :name "Fake patient ID# 3")
(add-patient-datum 'height 3 84 :age 2)
(add-patient-datum 'weight 3 12.5 :age 2)
(add-patient-datum 'height 3 93 :age 3)
(add-patient-datum 'weight 3 14.5 :age 3)
```

The above code creates a male-patient whose date of birth is January 1, 1980 and assigns the patient the id number 3. It then adds 4 data elements to the patient - 2 height and 2 weight data, taken at ages 2 and 3. Data that have the same time-stamp, such as the height and weight pair taken at age 2, are considered a data-cluster and are processed together.

At this stage, approximately 8 of the 120 cases had to be removed from the trial because some portion of the record was unreadable. From the remaining 112 cases, the first 100 were chosen to be included in the trial and distributed. The others were not used.

3.1.4. Packet Creation, Distribution, and Return

Each patient's data were displayed on a growth chart (Appendix B - Packet Directions and Sample). The charting process was automated by writing a Hypercard application that automatically plotted the data on either an infant chart (age 0-3) or a childhood chart (age 2-18), or both if appropriate. The test cases were then distributed among the packets in a way to try to equalize the number of test subjects that saw each case, while preventing any two packets from containing the same 10 individual patient cases. However, since packets were not returned frequently, the number of responses per patient case varies significantly.

The human subjects were recruited in several different manners. Many of the subjects were physicians at the Boston Children's Hospital. They were asked to participate and those who agreed were handed packets with return envelopes. Other subjects were found by placing a message on the usenet newsgroup sci.med.informatics asking for participants. This resulted in a wide range of participants, from the United States, Canada, and even a physician from France. There were no criteria for participation in the trial except that the individual had to be a medical doctor or in medical school. One respondent was a registered nurse. She was allowed to participate, but to help interpret results, all participants were grouped according to the amount of clinical training that they had received. Overall, over 80 packets, each containing 10 cases, were created and distributed. Of these, 22 packets were completed and returned. The large number of unreturned packets is due to several factors. One participant asked for 30 additional packets to be distributed to interns at the teaching hospital where he worked. Repeated queries were able to effect the return of the individual's packet, but the 30 additional ones were never returned. Similarly, a total of 14 packets were sent to a medical school where a colleague of Dr. Kohane was attending. Only 1 of those packets came back. Another participant distributed 10 packets to his colleagues, of which only 1 was returned. Of a total of 13 packets distributed to medical students, only 3 made it back.

After the packet distribution had begun, it was discovered that 3 of the 100 cases had typographical mistakes that were not caught earlier. They were removed from the trial, leaving 97 cases. Later, 2 more cases were removed from the trial because their medical records could not be located to obtain the medical record diagnoses. Thus, the final number of cases used in the trial was 95.

3.1.5. The Task

Each of the participants was told that the cases which they were reviewing came from files from the endocrine clinic at the Boston Children's Hospital. They were also told that the data that they were presented with consisted of all the height, weight, and bone-age data available to the physician at the time that the child was referred. In addition, they were reminded that not all of the patients had growth disorders. They were asked not to discuss the case with others or "study" in preparation for participation, and they were told to spend the same amount of time that they would normally spend if asked in a clinical setting to give an opinion.

Each growth chart presented the data graphically and in tabular form. At the bottom of each chart is a response area in which the subject was asked to do three things (See

Figure 6). First, they had to decide whether to refer the child to the growth clinic. Then they were asked to give a preliminary diagnosis. Finally, if they felt that a referral was warranted, they were asked to choose a time point at which it would have been appropriate to refer the child, only having seen the patient's data up to that point. For example, if the subject felt that the data suggested that the child had a Short Bone Syndrome, the subject should then choose to refer the child to the clinic and place a check next to the Short Bone Syndrome / Turner's Syndrome / Hypochondroplasia diagnosis. Then, if the subject felt that the child's clinical measurements clearly showed a growth abnormality that should have been noted by the data point at age 6, then he/she would circle the 6 in the tabular listing of the clinical measurements adjacent to the graphical picture of the patient's growth chart.

Sample Human Subject Response

1. Based on the data presented, would you recommend:

X Approve referral to endocrine clinic — Deny referral to endocrine clinic

2. Please place a checkmark next to exactly one congenital condition and any number of acquired conditions that you feel best describe the patient.

Congenital Conditions		Acquired Conditions
— Normal Growth	— Precocious Puberty	— Acquired Growth Hormone Deficiency
— Early Puberty	X Short Bone Syndrome / Turner's Syndrome / Hypochondroplasia	— Hypothyroidism
— Constitutional Delay	— Not Enough Information	— Obesity
— Congenital Growth Hormone Deficiency		

3.Please circle the age at which you feel the patient should be referred

Figure 6: Sample Human Subject Response

3.2. Gold-Standard - Medical Record and Experts

There were two gold-standards in this evaluation. The first gold-standard was the diagnosis written in the medical record of the patient. A second gold-standard was the evaluation of the patient by a pediatric endocrinologist.

The medical record diagnoses for the cases were first obtained from the on-line problem list of the patient. In approximately 75% of the cases, the problem list was empty so the most recent referral letter was scanned and any diagnoses made by the endocrinologist who saw the patient were accepted. A referral letter is the letter sent back to the pediatrician who referred the patient to the growth clinic. It contains the patient's

history, findings, diagnoses, and other clinical information. Because the on-line problem list appeared so incomplete, the most recent referral letter was consulted for each and every patient, even if the problem list was not empty. The union of the diagnoses from the problem list and the referral letter was accepted as the correct diagnoses. There were no cases where the two sources were incompatible. Some of the patients had no data in the on-line medical record. In those cases, the physical medical record was used. In the end, 2 of the cases had incomplete medical records and were removed from the trial.

To obtain the answers for the second gold-standard, a pediatric endocrinologist was given the same sheet that was given to the human subjects. Four endocrinologists each saw one quarter of the approximately 100 cases. The endocrinologist that helped develop TrenDx was **not** one of the four endocrinologists who provided the gold-standard. In a similar fashion to the human subjects, the experts were asked to either recommend or deny a referral to the growth clinic and to circle the appropriate time of referral. However, instead of choosing one diagnosis, the experts were asked to rank up to three acceptable preliminary diagnoses. Figure 7 shows a sample gold-standard response for a patient.

Sample Gold-Standard Response

1. Based on the data presented, would you recommend:

X Approve referral to endocrine clinic ___ Deny referral to endocrine clinic

2. Please place a checkmark next to exactly one congenital condition and any number of acquired conditions that you feel best describe the patient.

Congenital Conditions		Acquired Conditions
___ Normal Growth	___ Precocious Puberty	___ Acquired Growth Hormone Deficiency
___ Early Puberty	2 Short Bone Syndrome / Turner's Syndrome / Hypochondroplasia	___ Hypothyroidism
___ Constitutional Delay	___ Not Enough Information	X Obesity
1 Congenital Growth Hormone Deficiency		

3.Please circle the age at which you feel the patient should be referred

Figure 7:Sample Gold-Standard Response

While the performance of TrenDx will be compared to both gold-standards, we focused the results on the comparison of all the test-subjects, both TrenDx and human subjects, to the second gold-standard. There were several reasons for this choice.

As discussed in Section 3.4, expert opinion is an accepted gold-standard for trials of medical expert-systems because true gold-standards often do not exist.. This applied to our trial as well. Since some of the patients in our test set were referred for problems that were not related to growth, it would not be possible to achieve the medical record diagnosis

from only the height, weight, and bone-age data. Instead, we consider these cases "normal," even though the medical record does not explicitly state that the child is normal. Moreover, the doctor in the growth clinic who saw the patient had much more information available to him or her. This included a complete physical exam as well as the ability to take more measurements and labs. In fact, the medical record diagnosis might not have been made until several visits after the first referral visit. This was too high a standard for any individual to be held against, especially considering the limited amount of data available. A final reason to use an expert opinion was the poor quality of the information available in the medical record. As noted, the problem lists were incomplete and the referral letters often mentioned the *possible* presence of other disorders that were never confirmed or denied. One possible reason for the incompleteness of the medical records is that the patient may have moved or changed physicians.

We also compared the gold-standard panel's diagnosis to the medical record diagnosis. The results gave an indication as to whether it is possible to make a diagnosis using only the data that we used, or if the task itself was impossible.

3.3. Performance Measures

A scoring mechanism which uses a numeric score to evaluate the performance of TrenDx and the human subjects is useful to quantitatively express differences in performance. Numeric scores allow the performance to be quickly summarized as well as grouped over many patients. However, any scoring mechanism has biases and weaknesses. In addition, (Hayes-Roth, Waterman and Lenat) note that:

> Principle 1. Complex objects or processes cannot be evaluated by a single criterion or number.
>
> Principle 2. The larger the number of distinct criteria evaluated or measurements taken, the more information will be available on which to base an overall evaluation.

For these reasons, the performance of TrenDx and the other test subjects were evaluated in several different categories.

3.3.1. Comparison To Medical Record Diagnosis

First, the expert opinions, the diagnosis of TrenDx, and the diagnosis by the human physicians were all compared to the medical record diagnosis. In one analysis, just the decision to refer the child was compared to the clinical outcome of the child. A child with a clinical diagnosis of normal growth, early puberty, constitutional delay of puberty, or

familial short stature was considered normal and not worthy of referral. In addition, any patient that was diagnosed with a disorder that would not affect the child's growth, as determined by the expert who helped develop TrenDx, was also categorized as "normal." Then, for all of the abnormal patients, the preliminary diagnosis made by TrenDx and the human subjects was compared to the medical record diagnosis. All of these comparisons were performed for the human subjects as a group, and for the sub-populations divided by the amount of training that they had received.

3.3.2. Comparison to Expert.

Then, the decisions of the test subjects were compared to the second gold-standard set by four pediatric endocrinologists. Again, the decisions that the test subjects needed to make were:

1 - Refer or not to Refer

2 - Make a Preliminary Diagnosis

3 - Choose a Time to Refer.

Clearly, the decision to refer or not to refer was the most important. This is because if the patient did have a growth problem and was referred to a specialist, the specialist should be able to make the correct diagnosis. Giving the decision to refer a value of 5, the preliminary diagnosis a possible value of 3, and the timing of the referral a possible value of 2 gave each case a total possible score of 10.

The possible number of points for the preliminary diagnosis score was 3. Recall that the gold-standard expert ranked up to three preliminary congenital diagnoses, as well as checking off any number of the acquired conditions (See Figure 7). Since the number of acquired conditions that were chosen by the expert varied, the 3 points were divided as follows, depending on whether the expert felt that any acquired conditions were appropriate to note.

- **No Acquired Conditions Appropriate**: 3 points for choosing the top-ranked congenital diagnosis as the diagnosis, 2 points for choosing the second-ranked diagnosis, and 1 point for the third.

- **One or More Acquired Conditions Appropriate**: 1 point for choosing any of the acquired conditions. 2 points remaining for the congenital conditions: 2 points for the top-ranked congenital diagnosis, 1 point for the second, 0 points for the third.

To make the scoring system more concrete, let's match the sample human subject response in Figure 6 to the gold-standard response in Figure 7. First, note that both the human subject and the gold-standard felt that the patient should be referred. That gives the human subject 5 points out of 5 possible points. For the preliminary diagnosis, the gold-standard ranked 2 congenital disorders, Congenital Growth Hormone Deficiency and the Short Bone Syndrome / Turner's Syndrome / Hypochondroplasia combination. The gold-standard expert also noted that the patient was obese. This causes the scoring possibility to fall into the second of the two scoring categories listed above - "One or More Acquired Conditions Appropriate." The human subject did not note the acquired condition of obesity, and his/her choice of congenital conditions matches the 2nd choice of the gold-standard expert. According to the scoring scheme this scores 1 point out of 3 possible points.

The 2 remaining points, based on the timing of the referral, were also split - 2 points for choosing the referral at the same time as the gold-standard, and 1 point for coming within 1 data point, if it was within 2 years of the correct time. This scoring mechanism allowed subjects to score if the time at which they felt that the referral should have been made was "close enough" to that of the expert. To continue our example, assume that both subjects felt that the referral should have been made at the time that the patient was 6 years old. Thus, the subject scores 2 out of 2 possible points, for a total score of 8 out of 10 for this case. Note that the timing of the referral is not shown on the small portion of the response sheet that we have presented in Figure 6 or Figure 7. It can be seen in on a complete response sheet as shown in Appendix B - Packet Directions and Sample Chart.

As a side-effect of the scoring scheme, the number of possible points became fewer than 10 if the Gold-Standard expert decided that the patient should not be referred, because the 2 points for the timing of the referral could not be scored. In those cases, the referral decision became worth 7 points to keep the total possible points the same for all cases.

Table of Results for a Sample Patient

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obes	Score
MR	N			X								X	1
Exp	Y	4.25	2				1					X	N/A
TrDx	N		X									X	2
Sub1	Y	4.25		X								X	9
Sub21	N			X								X	1
Sub60	Y	5.5					X					X	9

Table 2: Table of Results for a Sample Patient

Table 2 is an example of a table of results for a sample patient. The table lists the complete set of answers for a patient, including the medical record diagnosis, the opinions of the gold-standard expert, the decisions of TrenDx, and all the responses of the human subjects. The complete set of results is listed in Appendix A - Patient and Subject Result Tables. The columns are:

- Sub: The subject giving the answers.
- Refer (y/n): Whether the subject decides to refer the patient
- Ref Age: If referral was recommended, the age of the child that the referral should have been made
- Early Pub: Early Puberty
- Normal: Normal Growth
- Cons Del.: Constitutional Delay of Growth
- Cong GH: Congenital Growth Hormone Deficiency
- Prec Pub: Precocious Puberty
- Short Bone: Short Bone Syndrome/Turner's Syndrome
- No. Info: Not enough information
- Acq. GH: Acquired Growth Hormone Deficiency
- Hyp Thyr: Hypothyroidism
- Obes: Obesity
- Score 1: Score relative to answers provided by Gold-Standard Expert.

Within the Sub column, the rows are:
- MR: Medical Record Gold-Standard
- Exp: Gold-Standard Expert
- TrDx: TrenDx answers
- Sub##: Human Subject number ## answers

To illustrate the scoring mechanism again, the score for TrenDx is calculated as follows. The decision to refer is Exp- Y, TrDx- N, so TrenDx scores 0 points for the decision not to refer the child. That automatically prevents TrenDx from scoring in regards to the timing of the referral, since TrenDx did not refer the child. In the area of the

preliminary diagnosis, TrenDx gets 1 point for choosing the acquired condition of Obesity and 1 point for choosing the second-ranked condition of Early Puberty. Thus TrenDx scores 2 points in comparison to the Gold-Standard Expert.

3.3.3. Multiple Comparisons

In the evaluation of four diagnostic decision-support systems, Berner uses several measures of performance based on a consensus of the programs being tested (Berner et al.). We made similar measurements by noting the number of cases with more than one participant review and in which all participant referral decisions were in agreement. Out of these cases, we looked at several statistics such as the number of cases in which both gold-standards agreed and the number of cases in which any group made a singular decision (e.g. chose to refer while all other groups chose not to).

3.4. Comparison to Other Evaluations

Since evaluations are an integral part of the development of decision-support systems, or expert-systems, many different types of evaluations have been performed(Berner et al.; Feldman and Barnett; Heckerman and Nathwani; Miller, Pople and Myers). Most of these evaluations attempt to measure the performance of the system on some number of cases. Forsythe argues that performance should not be the only aspect of a system that is measured(Forsythe and Buchanan). This is especially true of field trials of systems that are very advanced along the development cycle.

In evaluations of expert systems, several ways of obtaining a gold-standard or evaluating the answers produced by the system have been devised. The gold-standard is usually either the "real" answer, such as the correct diagnosis as confirmed by laboratory studies, or the opinion of one or more experts in the domain. In some cases, the expert that helped develop the system is also involved in the evaluation (Heckerman and Nathwani) introducing bias into the evaluation. Often, the system is the only thing being tested and its answers are evaluated by an expert and given either a subjective rating or a quantitative rating. Sometimes both the system and other physicians, who are not considered experts in the particular domain, are both evaluated and their performance is compared.

Several systems which try to cover a wide domain, such as the entire field of internal medicine, try to produce a ranked differential diagnosis list (Bankowitz, Lave and McNeil; Feldman and Barnett; Miller, Pople and Myers). In these cases, evaluation is more complicated because the goal of the program may not be to just suggest the most likely

answer, but to also stimulate the user by suggesting rare conditions. In these cases, more complex measures are devised to represent favorable and unfavorable traits. These generally involve counting disagreements and agreements between all of the participants. Specifically, not suggesting a diagnosis that every other participant suggested and being the only participant to suggest a particular diagnosis are two unfavorable characteristics.

In terms of the patient cases, our evaluation of TrenDx differs from many other evaluations because most of our patients are normal. Again, the cases were chosen serially with some screening criteria and then "cleaned" by collecting all the measurement data and by removing patients whose medical record contained unreadable measurements or had similar problems. While this introduces bias into the evaluation by limiting the scope of the trial, it also simplifies the evaluation by making it easy to categorize patients and avoid problems with missing data or cases in which the child somehow loses 10 centimeters in height between one visit and the next. As noted before, these cases were simply removed from the trial. In a field trial of the program, TrenDx would have to be programmed to deal with poor data.

4. TrenDx Development

The development of TrenDx in preparation for the trial can be broken down into three areas: programming new features and fixing bugs in the inference engine, modeling the processes that we wanted to monitor in order to get trend-templates, and knowledge-engineering the trend-templates to achieve improved performance.

4.1. Programming

The state of TrenDx at the conclusion of Haimowitz's thesis was poor. In fact, in running the trial discussed in his thesis (Section 5.1 Haimowitz), he was hampered by the fact that it took several hours to run a single patient. This was a direct result of the speed of the TUP functions and the large number of unconstrained intervals in his trend-templates. There were also several bugs in the inference engine that caused Now-Based trend-templates to crash.

Together, Haimowitz and I were able to fix the problems with Now-Based trend-templates. This improvement was essential in order to use several of the new trend-templates such as the trend-template for Obesity and Acquired Hypothyroidism. I also worked with Kohane to improve the speed of TUP calculations, since they constituted the majority of the time it took to process a patient.

The modeling language was improved by adding the ability to specify ranges on constant and linear constraints. For example, a user could specify that a parameter should be constant and that the parameter's value should be within the interval -5 to 5.

4.2. Modeling Growth

4.2.1. Standards

Several standards are routinely used to measure the growth of a child. The National Center for Health Statistics, or NCHS, has produced cross-sectional population curves for growth of both boys and girls up to age 18. Tanner and Davies produced longitudinal standards for height and height velocities that are more appropriate to use when following the growth of a child once puberty has begun (Tanner and Davies). These standards include curves for children in whom puberty occurs at the average time, as well as 2 standard deviations early and late. To compare the patient to a standard, we calculate the Z-

score for that measurement. Again, the Z-score represents the number of standard deviations from the mean.

During the previous development of thesis-TrenDx, it was noted that a clinically relevant piece of information was a measure of the weight of the child relative to his or her height. Using the weight and height curves developed by the NCHS, we created a measurement known as Build to represent the stockiness of the child. This Build measurement was chosen over another measurement known as the Body-Mass Index, or BMI, because the BMI was developed as a measurement of adult obesity. Several studies have shown that BMI is not a good indicator of obesity, or body build, in children (Roche et al.; Roland-Cachera et al.). During the current trial of TrenDx, the use of the Build measurement over the BMI was reconfirmed by showing that the BMI for fictitious patients who follow different percentile curves is <u>not</u> consistent over the longitudinal growth of a child. By definition, the Build measurement is consistent over the longitudinal growth of a child who follows percentile curves.

The pubertal development of children is measured on what is known as a Tanner Scale(Tanner). The scale ranges between 1 and 5, where 1 signifies no pubertal development and 5 signifies full pubertal development. A score of 2 signifies the beginning of puberty, and progression to each successive stage generally occurs within 6 months to 1.5 years. There are three Tanner scores - for pubic hair, breast, and penis development. Generally, only pubic hair and penis development scores are appropriate for boys, while pubic hair and breast development scores are appropriate for girls. In this trial, since the pubertal development information was not available, it was not used in the trend-templates actually used in the trial. However, pubertal development constraints should certainly be included in any model of pediatric growth.

The skeletal development of a child is measured by taking an X-ray of the left hand and wrist and comparing to published standards. The skeletal age, which we term bone-age, is very valuable in determining the growth state of a child. In fact, Kaplan uses bone age to differentiate between primary and secondary disturbances of growth(Kaplan). In essence, a short child who has a bone-age that is younger than his or her chronological age has the potential to continue growing when other children have stopped. This means that the child may "catch up" to his or her peers. The disorders which result in this type of growth delay are termed secondary disturbances of growth. Constitutional Delay of Puberty and Congenital Growth Hormone Deficiency are both secondary disturbances of growth. In comparison, a short child that does not have a delay in his or her skeletal growth does not have this same potential to "catch up." This characterizes a primary

disturbance of growth, such as a skeletal dysplasia - which we group under short-bone syndrome in our set of trend-templates.

4.2.2. Description of Growth States

The following is a characterization of all the process states, or growth conditions, that we chose to model. Collectively they are referred to as the set of disorders, even though they include normal growth. These characterizations were taken from consulting the expert as well as from descriptions listed in medical texts (Becker; Kaplan; Roche et al.; Roland-Cachera et al.; Tanner; Tanner and Davies). They are presented to aid the reader in understanding the trial and are only grossly correct from a medical standpoint. All of the trend-templates that we developed are listed in Appendix C - Trend Template LISP Code.

NORMAL GROWTH

Normal infants generally develop with consistent height and weight Z-scores up until sometime between ages 2 and 3. Sometime between 2 and 3, they 'establish centiles,' meaning that they naturally fall into a certain percentile channel that they stay within until they stop growing. The exact time of the onset of puberty is generally accepted to be around the age of 9 to 12 years in females and 10 to 13 years of age in males. Females continue growing slowly almost to the age of 17 whereas males often continue growing beyond age 18.

CONSTITUTIONAL DELAY OF PUBERTY

In children with Constitutional Delay, the skeletal development of the child is generally delayed in proportion to the delay in development of their height. This delay can be up to two years or more in extreme cases. Consequently, puberty occurs somewhat later, and they continue to grow for a longer period of time. Final adult height is equivalent to children with Normal Growth.

EARLY PUBERTY

Children with Early Puberty are characterized as having an advanced skeletal age. Puberty occurs earlier, as does cessation of growth. Final adult height is still equivalent to children with Normal Growth.

Being described as early or delayed signifies that the child is early in relation to the normal population. There is no agreed-upon definition of what is early and what is not. While both Constitutional Delay and Early Puberty are considered benign, the extent of the delay or acceleration may cause an expert to consider it differently.

CONGENITAL GROWTH HORMONE DEFICIENCY

Children suffering from Congenital GH Deficiency have several traits that we can model. They are generally significantly short, directly due to an inability to produce or respond to growth hormone. In addition, the extent of their skeletal and sexual delay is much greater than in Constitutional Delay of Puberty. Often these children have sexual infantilism.

SHORT BONE SYNDROME / TURNER'S SYNDROME

We use the category Short Bone Syndrome/Turner's Syndrome to describe a large class of disorders which we feel can not be differentiated by only height, weight, pubertal, and bone-age data. We characterize the group as being even shorter than Congenital GH Deficient children, with very little delay in skeletal age.

PRECOCIOUS PUBERTY

True precocious puberty describes a condition in which children develop sexual characteristics at an extremely early age. As one would expect, skeletal age is more advanced than chronological age, and the children are often taller than their peers because of their advanced development.

ACQUIRED GROWTH HORMONE DEFICIENCY

Acquired Growth Hormone Deficiency is a condition that is acquired at some time during the patient's life. Prior to this time, he or she may be perfectly normal. This condition is often secondary to some other problem, such as a craniopharyngioma. The effects of this disease are a marked deceleration of growth at the onset of the deficiency. Experts often talk about a child "falling off" the growth curve. However, the rate of deceleration is often variable since it is affected by the source and extent of the problem. This makes it difficult to model.

CONGENITAL AND ACQUIRED HYPOTHYROIDISM

There are many clinical manifestations of Acquired Hypothyroidism such as lethargy, decreased appetite, and other findings which are not available to us. However, Kaplan notes that "...the most important sign of acquired hypothyroidism in childhood is growth failure."(Kaplan) Height moves in a progressive downward deviation and weight tends to increase modestly. Skeletal development is delayed in proportion to height age. Kaplan also notes that these are characteristic of hypopituitarism (Growth Hormone

Deficiency), and consequently it is difficult to differentiate between the two from the data available to us.

OBESITY and MALNUTRITION

These conditions were included because the growth clinic is often sent referrals for children with extreme weight problems. As noted, weight, or more precisely build, is clinically significant in several of the disorders we are trying to diagnose. We cannot actually capture obesity information from the data that we have, since we cannot differentiate between a muscular individual and an obese individual from only height and weight data.

It is very difficult to differentiate between Growth-Hormone Deficiency and Hypothyroidism from the available data. Moreover, it is also difficult to differentiate between Growth-Hormone Deficiency and Hypothyroidism, and the Short Bone Syndrome except by knowing Bone Age and, to a lesser extent, Pubertal Stage data. However, both of these data types are not routinely collected and thus are not present in very many of the test cases. Consequently, Growth Hormone Deficiency and Hypothyroidism were considered the same diagnosis in both the creation of our trend-templates and in scoring TrenDx and the human subjects.

4.2.3. Modeling

The trend-templates used to model each of the disorders presented in the previous section were designed based on consultations with the expert, reading medical texts, and adapting the trend-templates used in thesis-TrenDx. The entire set of trend-templates is listed in Appendix C - Trend Template LISP Code.

The new (current) and old (thesis-TrenDx) trend-templates for normal growth in

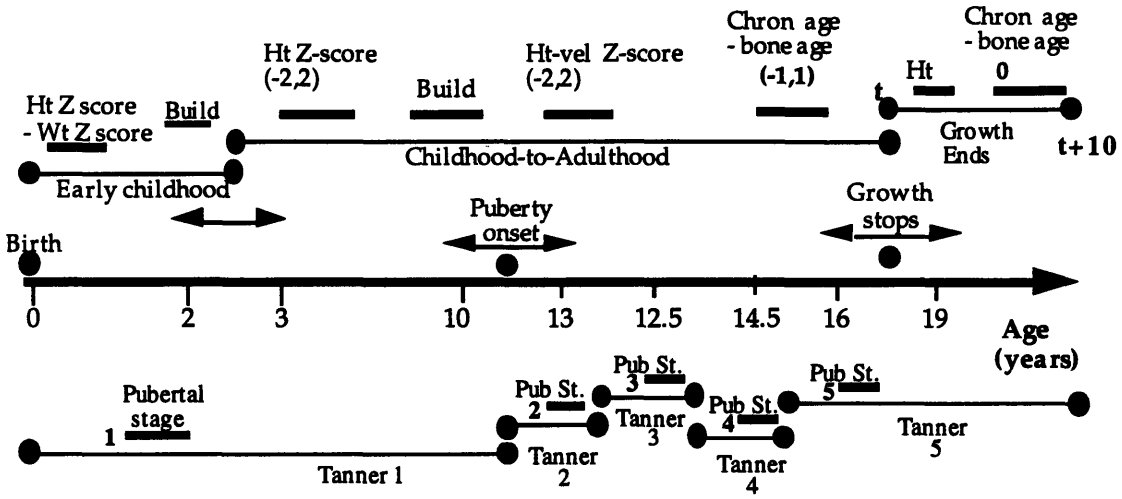


Figure 8:Current Trend-Template for Normal Male Growth

males are shown in Figure 8 and Figure 9 for easy comparison. The new trend-template differs from the previous trend-template in several ways. One of the most notable is the reduction of Intervals 2 through 4 in Figure 9 to a single interval in Figure 8. Another is the absence of the second order value constraints, and a third is the addition of numeric values as part of the constraints.

Haimowitz notes that:

> First, in order to insure reliable value constraint matches, TrenDx should assign at least three or four data points per trend template interval. When modeling trends in sparse data sets, a knowledge engineer should minimize the number of disjoint intervals constraining a particular parameter. If trend templates contain too many such intervals, TrenDx may find a low-scoring or trivial match by assigning only one or two data points to each interval. (Section 5.1.5 Haimowitz)

This is especially true for the domain of pediatric growth monitoring, when measurements are not necessarily taken with any regularity. In fact, because data in growth monitoring is often sparse, each individual data point takes on much more significance. This is in contrast to some domains where data collection occurs at a high frequency, such as ICU data monitoring. In those domains, individual data points may have little significance compared to the overall trend in the data.

Previous Trend-Template for Normal Male Growth

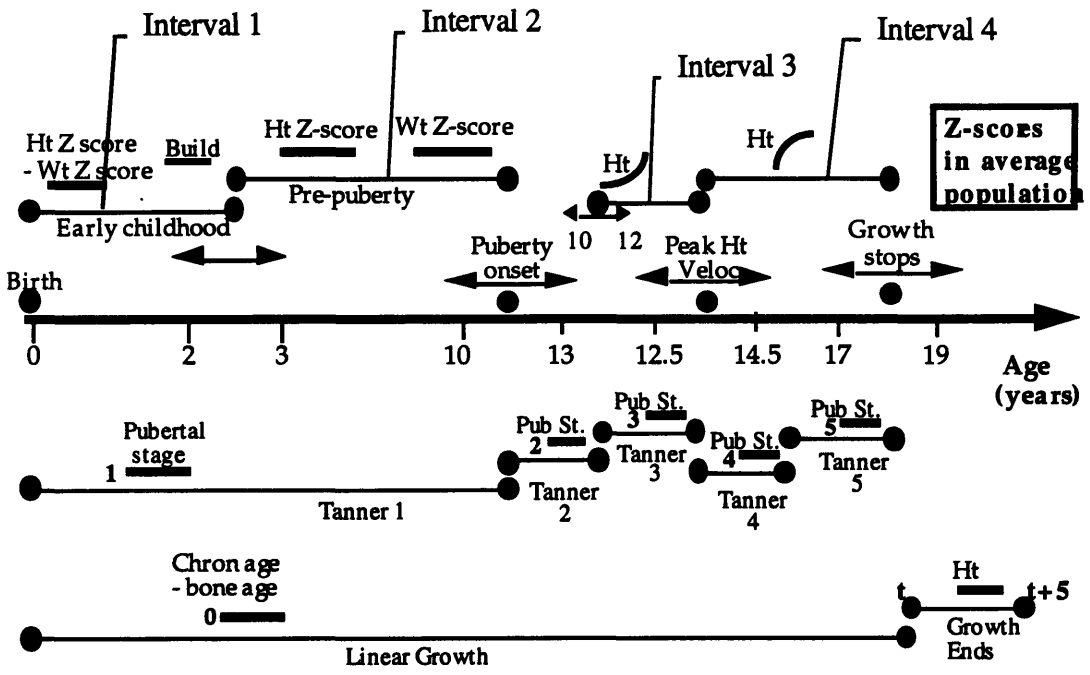


Figure 9: Previous Trend-Template for Normal Male Growth(Haimowitz)

Note that in the previous trend-template (Figure 9), there are four different intervals that constrain the patient's height or height Z-score (Intervals 1-4). Depending on the value constraint, each interval needed two or three data points before a non-trivial score could be assigned. This problem is compounded by the ability to specify variable value constraints. For example, a user could specify that a particular parameter should be constant at some unknown value. In that case, a single value in the interval would match trivially, since the value constraint would match a single value to a wildcard and get no error. Moreover, the presence of multiple intervals with uncertain endpoints increases the computational complexity of matching data to trend-templates. Each data point that falls into the uncertainty range of an interval endpoint causes TrenDx to branch and create two new contexts - one in which the data belongs to the interval and one in which it doesn't. For both of the new hypotheses, all the data must be matched to the trend-template and both of the hypotheses kept until one or both are pruned by the beam-search.

There is also a discontinuity between Interval 2 and Interval 3. Since the temporal constraint allowed the Begin point of Interval 3 to vary, TrenDx would branch when there was a data point between the ages of 10-12. One branch would include the data point in the interval and the other would not. The beam search would then choose the second

hypothesis over the first because the data point contributes no error-score to the score of the second hypothesis. In essence, TrenDx was 'throwing away' data.

Using the lessons learned from these mistakes, new models of the growth states were created and trend-templates were written from these models. These trend-templates were further refined during iterative sessions of knowledge-engineering, discussed in Section 4.3.

Haimowitz suggests that "A trend template should only include information that will distinguish it from competing trend templates."(Section 6.5.1 Haimowitz) This suggestion is based on the fact that having the same information in each trend-template will contribute the same amount of error to each trend-template. While this is true, it ignores the fact that knowing that a process matches to all the trend-templates poorly is important information in itself. In fact, in a domain where all of the possible states are not known (most real-world domains), using this information suggests that the process is in some unmodelled state, assuming that the models are correct. The principle is amended as follows:

> A trend-template should include as much information as practical, even information that does not distinguish it from competing trend-templates.

It is important to examine exactly what knowledge is expressed while modeling with TrenDx. Unlike many other expert-systems, such as INTERNIST-1 (Miller, Pople and Myers) or PATHFINDER (Heckerman and Nathwani), there is no implicit or explicit mention of the probability of the occurrence of each growth disorder or some type of evoking strength of some particular symptom. There are some implicit probabilities embedded in the performance of the program, as well as assumptions about the distributions in height and weight that are used to make Z-score calculations; however, there are no statements equivalent to "Short Bone Syndrome occurs in 1 out of every 35,000 individuals," or "A Bone Age 1 year behind chronological age indicates the presence of Congenital Growth Hormone Deficiency with a probability of 0.35." While it is true that knowledge of prior probabilities of diseases and findings is important and useful for diagnosis, the absence of these probabilities allows TrenDx to model processes in which these probabilities are not known.

In modeling the Acquired Growth-Hormone Deficiency trend-template, another lesson was learned. The first attempt to model the disorder only included an interval that extended back 2 to 3 years from the most recent data point. TrenDx created alternate temporal worlds for each data point that might have been included in the interval. Since each hypothesis received an error-score and the hypotheses were pruned using a beam

search, it was most likely that hypothesis in which the sole interval contained the fewest number of points would score the best. In essence, TrenDx would "throw away" data. Thus, the trend-template was extended by adding an interval that represents the time before the onset of the growth-hormone deficiency. By defining the intervals to be consecutive, no data is thrown away. This is the same problem that was encountered with the non-consecutive intervals in the previous trend-templates used in Haimowitz's thesis trial.

Overall, it is difficult to characterize abnormal growth. Medicine does not understand every disorder that afflicts man. Even when the disorder is known and has been characterized, its effects on each individual can vary in presence or absence of symptoms and the severity of those symptoms. The possibility of multiple disorders being present and interacting further complicates matters. Because of all these factors, it is imperative to model the normal state as well as possible.

4.3. Trend Template Refinement

Once the broad outline of the trend-templates were generated for each of the disorders, they were refined to achieve the desired performance. Many of the age ranges from the previous trend-templates were used, and the improvements mentioned in Section 4.2.3, such as reducing the number of intervals, were made. In addition, the values for parameter error-function models and thresholds had to be established.

4.3.1. Establishing Performance Goals and Training Sets

The development and refinement of trend-templates is an iterative process. A user must define a trend-template and then process the data for a particular patient and see how the trend-template performs. If the performance does not match the desired performance, then the trend-template must be refined and the process must be repeated.

A set of performance goals were defined to give a concrete definition of when the trend-templates were acceptable. The performance goals consisted mainly of specifying that the trend-templates should score well or poorly on a set of fake patients with stereotypical patterns of growth, which was called the training set of fake patients. The training set also consisted of some actual patients such as the cases used in earlier trials of TrenDx. The existence of the training set was important to ensure that new additions to the knowledge base did not interact with the previous information in unexpected ways.

4.3.2. Error-Function Models

At the conclusion of Haimowitz's thesis, TrenDx was somewhat limited in the expressive capabilities of the error-function of a value constraint. It allowed the user to specify an expected value of a parameter, such as (constant 5), but the ability to specify a normal range was absent. After this ability was incorporated, it was discovered that it still did not capture the knowledge that the expert wanted to express. The expert wanted to say, "the value should be in the range of A to B. If the value is somewhere near the middle, that's good. If the value is near the endpoints, then the value does not match the trend-template very well. If it is outside the range, the trend-template should score very badly." Constraint-Based TrenDx, discussed in Section 2.3, was only able to distinguish between the first and last conditions while Regression-Based TrenDx gave the user the ability to express the first and second conditions. Looking at the definition of the value constraint as a composition of functions, *the limitations of the second component can be overcome by making the first component more complex.* As a simple example, the performance of Constraint-Based TrenDx could be imitated by the Regression-Based TrenDx as shown in Figure 10.

Comparison of Simple and Compound Value Constraints to Constraint-Based TrenDx

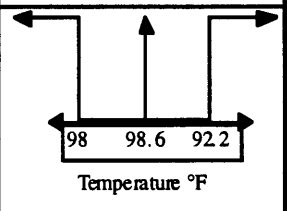
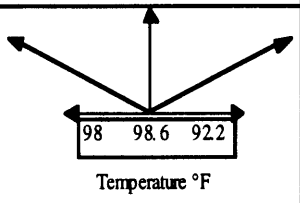
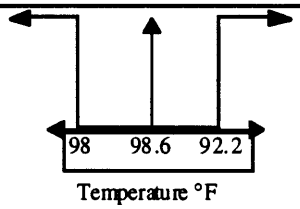
	Constraint-Based TrenDx	Regression-Based TrenDx	
	Normal Value Constraint	Simple Value Constraint	Compound Value Constraint
Parameter	Temperature	Temperature	Function (Temperature) if Temperature < 98.0°F return 1 if Temperature > 99.2°F return 1 else return 0
Model	98.0° to 99.2°F	(Constant 98.6°F)	(Constant 0)
Resulting Error-Function			

Figure 10: Comparison of Simple and Compound Value Constraints to Constraint-Based TrenDx

In fact, almost any error-function can be created with the use of various parameter functions. A few possible ones are shown in Figure 11.

By using a function in the parameter component of the value constraint, many different composite error-functions could be created. For many of the value constraints used in the final trend-templates, the composite error-function IV from Figure 11 was used.

It is arguable that these permutations on the first component of the value constraint are unnecessary and that instead, programming changes should be made within TrenDx to allow the expression of the more commonly used composite error-functions. The same results could have been achieved by programming new functions into the second component of the value constraint. However, time did not permit this and the original programmer of TrenDx was unavailable to help complete the task. Furthermore, each user of TrenDx should not be expected to go into the internal code of TrenDx to program new error-functions.

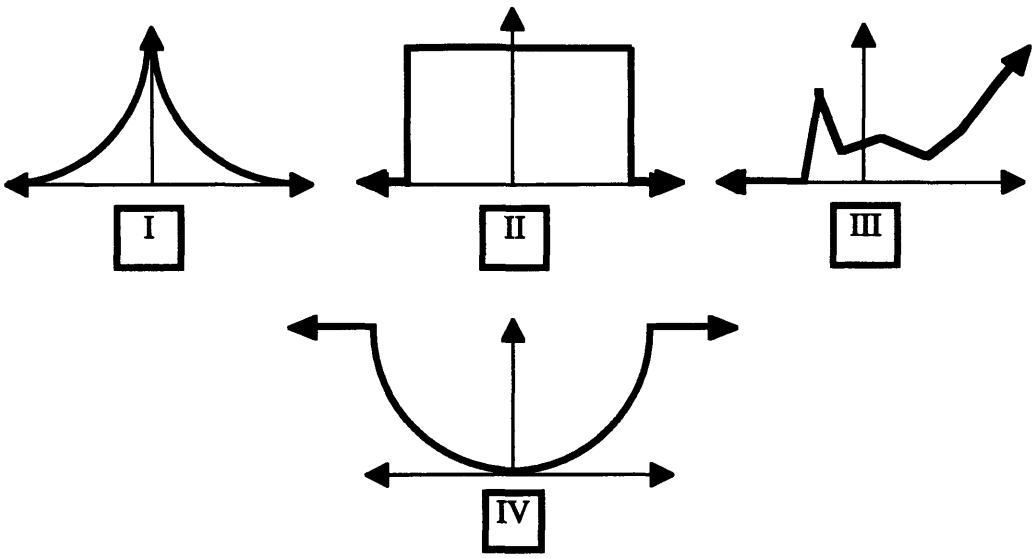


Figure 11: Possible Composite Error-Functions

We chose to use an error-function that looks like Error Function IV in Figure 11. This error-function represents the way the expert compares a value to its expected range. If the value is somewhere near the middle, the error is small. However, as the value moves away from the middle, the error increases gradually. Near the boundaries, the error is large. Since it is continuous and fairly smooth, this error-function avoids the brittle behavior that the Constraint-Based TrenDx suffered from. A value just below the boundary scores almost as poorly as a value beyond the boundary.

4.3.3. Residual Mean Square Error vs. MAPE

TrenDx allows the option of using either the Residual Mean Square Error or the Mean Absolute Percent Error when matching data to value constraints. The Residual Mean Square Error is defined in Equation 1.

$$\frac{\sum_t (\text{Expected}_t - \text{Actual}_t)^2}{\text{Degrees of Freedom}}$$

Equation 1: Residual Mean Square Error

Haimowitz used the Mean Absolute Percent Error, or MAPE, to calculate the match of data to a value constraint. MAPE is defined in Equation 2.

$$\frac{\sum_t \left| \frac{\text{Expected}_t - \text{Actual}_t}{\text{Expected}_t} \right|}{\text{Degrees of Freedom}}$$

Equation 2: Mean Absolute Percent Error

The ability to use the MAPE was added because it is useful for comparing the goodness of fit between models of variables with different scales. For example, a constraint with an expected value of 100 and an actual value of 110 would be off by 10, but really only deviates from the expected value by 10%. In comparison, a constraint with an expected value of 10 and an actual value of 20 is still off by 10, but deviates from the expected value by 100%.

In using the MAPE, an unexpected problem arose because the expected value of many of the constraints is 0, and division by 0 is undefined. There were two possible solutions to the problem. One solution would be to add some offset to both the expected and actual values and then perform the calculation. However, that solution causes the error to be affected by the size of the offset. The second solution was to use the Residual Mean Square Error. The weakness of the residual error, matching to variables of different scales, was minimized by keeping the range of values of all variables approximately equal and by using the error-function IV shown in Figure 11.

4.3.4. Thresholds for Triggering

Recall that TrenDx calculates an error score for each hypothesis of each trend-template. Thus, the hypothesis that matches the data the best is the one with the lowest error score. The decision to refer is based on the assumption that the trend-templates for Normal Growth, Constitutional Delay of Puberty, and Early Puberty are considered "normal" and do not require a referral to the growth clinic. Often, all error-scores are trivially small because only a few data points have been processed or because data points match trivially to certain value constraints. To avoid triggering a referral at these times, a threshold score was used. It was similar to the one described in the previous trial of TrenDx. When the error-scores for all the "normal" trend-template hypotheses scored extremely high for a single data point, or somewhat high for two consecutive points, the patient was considered worthy of referral. Then, the best-scoring "abnormal" trend-template was taken as the diagnosis.

The triggering mechanism used to evaluate thesis-TrenDx (Haimowitz) was revised for this trial. Instead of triggering when the abnormal hypotheses scored a certain amount

better than the normal hypotheses, we chose to trigger whenever all the normal hypotheses score worse than a certain threshold. We kept the use of the union of the two triggering styles: high single-point thresholds and lower consecutive-point thresholds (Section 2.6). Reasons for the revision of the triggering mechanism are given in section ----.

The threshold values were obtained from an analysis of 20 normal patient cases processed by TrenDx. The cases consisted of 10 male and 10 female patients, each with at least three data points. Of these cases, at least 3 of the patients of each gender had data points from infancy. The lowest error-score of the three "normal" trend-templates at each data point was analyzed. The mean and standard deviation were 0.19 and 0.08, respectively. We chose to set the threshold for the single-point triggering at 2 standard deviations from the mean, and the threshold for the consecutive-point triggering at 1.5 standard deviations from the mean. Thus, if a patient's lowest error-score for all of the three "normal" trend-template hypotheses was above 0.35 at any time point, or above 0.31 for two consecutive time points, the patient was considered abnormal and referred. The lowest non-trivial trend-template was then taken as the preliminary diagnosis. In some cases, most notably in infants, none of the trend-templates scored better than any of the others. In that case, we considered TrenDx unable to make a diagnosis, representing the "Not enough information" diagnosis. In addition, the Obesity acquired condition was considered "checked off" if the trend-template for Obesity scored better than the trend-template for normal-build at any time.

The results obtained by changing the triggering thresholds were also calculated to help justify the triggering mechanism. Recall that one of the reasons for the use of the union of the two triggering mechanisms was to reduce the brittle nature of thresholds, where a value just below the threshold would not trigger, while a value a little bit higher would trigger. To help understand whether this goal was achieved and whether the triggering threshold values that were chosen were appropriate, the effects of raising and lowering the threshold values were examined.

5. Results

The decisions made by TrenDx and the human subjects were compared to the medical record diagnosis and then to the recommendation of a pediatric endocrinologist. Then the decisions of the pediatric endocrinologists were compared to the diagnosis recorded in the medical record. Inter-group comparisons were performed and then the results of changing the triggering thresholds were examined.

5.1. Comparison to Medical Record Diagnosis

The referral decisions made by TrenDx were compared to those made by the human subjects. Next, the abnormal cases were analyzed by looking at the number of correct referrals and correct diagnoses within each sub-population.

The 95 patients in the trial were categorized according to their medical record diagnosis. They were grouped as shown in Table 3:

Medical Record Diagnoses of Trial Cases

Category Name	Description / Diagnosis	Number
Normal -	Normal Growth, Early Puberty, Constitutional Delay of Puberty, Familial Short Stature	50
Normal - Other	Referred for non-growth problem	18
Precocious Puberty	Precocious Puberty	6
GH-Def and Hypothyroidism	Congenital Growth Hormone Deficient, Acquired Growth Hormone Deficient, Hypothyroidism	11
Complex Cases -	Multi-congenital abnormalities/Cancer	8
SB/ Turner's	Short Bone Syndrome, Turner's Syndrome	2

Table 3: Medical Record Diagnoses of Trial Cases

Combining the two normal groups into one category and all the abnormal groups into another category resulted in a **total of 68 normal patients and 27 abnormal patients.**

The 22 human subjects were categorized by the amount of training that they had received. One group, called the Post-Residency group, was composed of human subjects who had completed a residency program and were either in a fellowship program or were attending. Of the 17 members of this group, 6 were fellows, 7 were attending in an

teaching hospital, and 4 were general practitioners. The Pre-Residency group was composed of all subjects who had not completed a residency program. This group comprised the remaining 5 of the 22 human subjects in the experiment - 1 resident, 3 medical students, and 1 registered nurse.

5.1.1. Referral Decision of TrenDx vs. Human Subjects

Table 4 through Table 7 list the results of the comparing the referral decisions made by TrenDx and the human subjects to the medical record diagnosis. The sensitivity and specificity for the referral decisions were calculated as follows:

- Sensitivity: (Patient is abnormal and referral approved) / (Total # of abnormal patients)
- Specificity: (Patient is normal and referral denied) / (Total # of normal patients)

Fisher's Exact Test was performed to obtain an exact Chi-Square Test value.

Comparing the performance of TrenDx to the set of all the physicians, the sensitivities were the same (0.59), but the human subjects had a higher specificity value (0.53 vs. 0.47). When the participants were separated by training group, the Pre-Residency group had the highest sensitivity value (0.765) and a low specificity (0.47). The Post-Residency group had the lowest sensitivity value (0.52), and the highest specificity (0.54). None of the results were statistically significant at α=0.05.

TrenDx Decision vs. Medical Record Diagnosis

Decision	Abnormal	Normal	Total
Refer	16	36	52
No Referral	11	32	43
Total	27	68	95[1]

Table 4: TrenDx Decision vs. Medical Record Diagnosis

Sensitivity: 16/27 = 0.59

Specificity: 32/68 = 0.47

Fisher's Exact Test - P=0.6512

[1] TrenDx crashed while processing 3 of the patients. 1 was abnormal and 2 were normal. We assume that it got the wrong answer (referred normal patients and did not refer abnormal patients).

All Physicians vs. Medical Record Diagnosis

Decision	Abnormal	Normal	Total
Refer	35	75	110
No Referral	24	83	107
Total	59	158	217

Table 5: All Physicians vs. Medical Record Diagnosis

Sensitivity: 35/59 = 0.59

Specificity: 83/158 = 0.53

Fisher's Exact Test - P=0.1296

Pre-Residency Subjects vs. Medical Record Diagnosis

Decision	Abnormal	Normal	Total
Refer	13	17	30
No Referral	4	15	19
Total	17	32	49

Table 6: Pre-Residency Subjects vs. Medical Record Diagnosis

Sensitivity: 13/17 = 0.76

Specificity: 15/32 = 0.47

Fisher's Exact Test - P=0.1347

Post-Residency Subjects vs. Medical Record Diagnosis

Decision	Abnormal	Normal	Total
Refer	22	58	80
No Referral	20	68	88
Total	42	126	168

Table 7: Post-Residency Subjects vs. Medical Record Diagnosis

Sensitivity: 22/42 = 0.52

Specificity: 68/126 = 0.54

Fisher's Exact Test - P=0.4825

5.1.2. Decision-Breakdown by Abnormal Sub-Populations

Table 8 through Table 11 show results of the referral decision and diagnosis, categorized by the different sub-populations of the abnormal patients. Recall that all of these patients should have been referred, according to the medical record gold-standard. Most of the columns are self-explanatory. There are two exceptions. The column labeled '#' represents the total number of patients with that disorder for whom a decision was made. This column total is not consistent between tables. This is because TrenDx saw each patient exactly once, but both the Pre-Residency subjects and the Post-Residency subjects saw different patients, causing some patients to be evaluated more than others. The column labeled 'Correct Dx% (of Refs)' is the percentage of cases that had the correct diagnosis, out of the population of patients that were referred. It gives an indication of how often the diagnosis was correct, given that the subject felt that the patient had some kind of abnormality. For the 'Correct Dx%(of Refs)' Total calculation, the Cancer/Complex cases are not included.

Each of the groups referred patients fairly consistently across different disorder sub-populations. Again, one can see that the referral decisions of TrenDx were comparable to that of the aggregate group of human subjects. However, the physicians chose the correct preliminary diagnosis more often (33.3% vs. 21.0% and 56.0% vs. 36.4%). Table 10 and Table 11 show why. While the Total Correct Dx% and Correct Dx%(of Refs) for TrenDx and the Pre-Residency groups are comparable (~22% and ~35%), the performance of the Post-Residency group was almost twice as good (~38% and 69%).

Disorder Population Referral and Diagnosis Results
TrenDx vs. Medical Record

Disorder Sub Population	#	No Refer	Refer	Correct Ref %	Correct Dx	Correct Dx %	Correct Dx % (of Refs)[2]
Precocious Pub	6	2	4	66.7%	2	33.3%	50.0%
GH-Def / Hypothyroid	11	5	6	54.5%	1	9.1%	16.7%
Cancer / Complex	8	3	5	62.5%	N/A	N/A	N/A
Short Bone / Turner's	2	1	1	50%	1	50.0%	100%
Total	27	11	16	59.2%	4/19	21.0%	36.4%

Table 8: Disorder Population Referral and Diagnosis Results, TrenDx vs. Medical Record

Disorder Population Referral and Diagnosis Results
All Physicians vs. Medical Record

Disorder Sub Population	#	No Refer	Refer	Correct Ref %	Correct Dx	Correct Dx %	Correct Dx % (of Refs)
Precocious Puberty	20	9	11	55.0%	6	30.0%	54.5%
GH-Def / Hypothyroid	19	6	13	68.4%	8	42.1%	61.5%
Cancer / Complex	17	7	10	58.8%	N/A	N/A	N/A
Short Bone / Turner's	3	2	1	33.3%	0	0%	0%
Total	59	24	35	59.3%	14/42	33.3%	56.0%

Table 9: Disorder Population Referral and Diagnosis Results,

All Human Subjects vs. Medical Record

[2] Correct Dx% of refs does not include Cancer/Complex cases. Neither does Total.

Disorder Population Referral and Diagnosis Results
Pre-Residency vs. Medical Record

Disorder Sub Population	#	No Refer	Refer	Correct Ref %	Correct Dx	Correct Dx %	Correct Dx % (of Refs)
Precocious Pub	6	2	4	66.7%	1	16.7%	25.0%
GH-Def / Hypothyroid	6	1	5	83.3%	2	33.3%	40.0%
Cancer / Complex	4	0	4	100%	N/A	N/A	N/A
Short Bone / Turner's	1	1	0	0%	0	0%	0%
Total	17	4	13	76.5%	3/13	23.1%	33.3%

Table 10: Disorder Population Referral and Diagnosis Results,

Pre-Residency vs. Medical Record

Disorder Population Diagnosis Results
Post residency vs. Medical Record

Disorder Sub Population	#	No Refer	Refer	Correct Ref %	Correct Dx	Correct Dx %	Correct Dx % (of Refs)
Precocious Puberty	14	7	7	50.0%	5	35.7%	71.4%
GH-Def / Hypothyroid	13	5	8	61.5%	6	46.2%	75.0%
Cancer / Complex	13	7	6	46.2%	N/A	N/A	N/A
Short Bone / Turner's	2	1	1	50%	0	0%	0%
Total	42	20	22	52.4%	11/29	37.9%	68.8%

Table 11: Disorder Population Referral and Diagnosis Results,

Post-Residency vs. Medical Record

5.1.3. Summary

The performance of TrenDx and the human subjects was comparable, though the specificity of TrenDx was lower. The Pre-Residency subjects referred the highest percentage of the abnormal cases that they were presented with, but were only able to choose the correct diagnosis about as often as TrenDx. The Post-Residency subjects had the lowest sensitivity and highest specificity, but the highest percent of correct preliminary diagnoses. None of the Chi-Square Tests resulted in statistically significant results.

5.2. Comparison to Experts

5.2.1. Referral Decision

Table 16 compares the referral decisions of TrenDx to the Expert Gold-Standard. The sensitivity and specificity are improved, compared to those in Table 4: TrenDx Decision vs. Medical Record Diagnosis. A more detailed breakdown is presented in Table 13. It subdivides the patients by whether the medical record diagnosis was normal or abnormal. The middle columns represent the cases in which the expert agreed with the medical record diagnosis, either by referring an abnormal case or not referring a normal case. The consensus sensitivity (sensitivity over patients in whom both gold-standards agreed should be referred) was even higher, at 0.68. The consensus specificity was lower than the non-consensus specificity (Table 12), but still higher than the specificity compared to the medical record diagnosis (Table 4). Again, the P-values were not statistically significant.

TrenDx Decision vs. Expert Gold-Standard

Decision	Expert Refer	Expert Not Refer	Total
Refer	36	17	53
No Referral	23	19	42
Total	59	36	95

Table 12: TrenDx Decision vs. Expert Gold-Standard

Sensitivity: 36/59 = 0.61

Specificity: 19/36 = 0.52

Fisher's Exact Test - P=0.2080

TrenDx Decision vs. Expert and Medical Record Consensus

Decision	Expert Refer		Expert Not Refer		Total
	Normal	**Abnorm**	**Normal**	Abnorm	
Refer	22	**13**	**14**	3	53
No Referral	18	**6**	**14**	5	42
Total	40	**19**	**28**	8	95

Table 13: TrenDx Decision vs. Expert and Medical Record Consensus

Consensus Sensitivity: 13/19 = 0.68[3]

Consensus Specificity: 14/28 = 0.50

Fisher's Exact Test - P=0.2438

Table 14 and Table 15 show the results of the same analysis performed for the physicians. Note that their performance is also improved and the specificity values are consistently higher than those of TrenDx. The results of Fisher's Exact Test indicate a statistically significant correlation between the responses of the physicians and those of the experts at α=0.05.

Physicians vs. Expert Gold-Standard

Decision	Expert Refer	Expert Not Refer	Total
Refer	91	19	110
No Referral	51	56	107
Total	142	75	217

Table 14: Physicians vs. Expert Gold-Standard

Sensitivity: 91/142 = 0.64

Specificity: 56/75 = 0.75

Fisher's Exact Test - P < 0.0001

[3] Sensitivity and Specificity calculated out of the two middle columns representing the consensus of Expert and Medical Record

Physicians vs. Expert and Medical Record Consensus

Decision	Expert Refer		Expert Not Refer		Total
	Normal	**Abnorm**	**Normal**	Abnorm	
Refer	59	**32**	**16**	3	110
No Referral	40	**11**	**43**	13	107
Total	99	**43**	**59**	16	217

Table 15: Physicians vs. Expert and Medical Record Consensus

Consensus Sensitivity: 32/43 = 0.74[4]

Consensus Specificity: 43/59 = 0.73

Fisher's Exact Test - P< 0.0001

5.2.2. Preliminary Diagnosis

The number of times that the preliminary diagnosis of the test subjects matched to those of the experts was counted and is shown in Table 16.

Test Subject Preliminary Dx Matches to Expert Dx

Group	# Decisions	# Match	% Match
TrenDx	95	39	41.0%
All Human Subjects	217	123	56.7%

Table 16: Test Subject Preliminary Diagnosis Matches to Expert Diagnosis

5.2.3. Scores

The responses of all the subjects were scored according to the algorithm described in section 3.3. The entire set of scores for all the participants is listed in Appendix A - Patient and Subject Result Tables. Table 17 shows the average score received by each subject group - TrenDx, all participants, the Pre-Residency group, and the Post-Residency group. The average score of TrenDx was lower than the average scores of any of the participant groups. The two Pre-Residency and Post-Residency groups performed comparably. The table also shows the 95% confidence interval calculations and the results of the Wilcoxon Signed-Rank Test comparing the human subject scores to the scores

[4] Sensitivity and Specificity calculated out of the two middle columns representing the consensus of Expert and Medical Record

received by TrenDx. The differences are statistically significant for the entire group of participants and for the Post-Residency participants at α=0.05, indicating a high probability that TrenDx does not score as well as the other groups at this task.

Test Subject Scores

Test Subject Group	# Decisions	Avg±S.D.	95% C.I.	t-test P Value
TrenDx	95	4.7±0.8	3.7-5.6	N/A
All Humans	217	5.4±1.2	4.9-5.9	0.0033[5]
Pre-Residency	49	5.5±0.6	4.8-6.2	0.0625
Post-Residency	168	5.4±1.3	4.7-6.0	0.0135[6]

Table 17: Test Subject Scores

The scores of TrenDx and the human subjects over the different disorder sub-populations are presented in Table 18. The two groups performed comparably on the Precocious Puberty cases, but the physicians scored considerably better on the GH-Def/Hypothyroid cases and the Complex / Cancer cases. TrenDx scored better on the Short Bone/Turner's cases, but there were very few of those cases. At the bottom of the table, the weighted average of the scores is shown. Note that the weighted average score of TrenDx on these abnormal cases is the same as the average score of TrenDx over all cases (Table 17). This is in contrast to the physicians; their weighted average score on the abnormal cases is much higher than their score over all cases. The weighted average score of the physicians on the normal population is 5.18.

Average Score by Disorder Sub-Population

Disorder	# Pats	TrenDx Avg Score	# Dec	Physician Avg Score
Precocious Puberty	6	5.8	20	5.6
GH-Def / Hypothyroid	11	4.4	19	6.0
Complex / Cancer	8	4.2	17	6.9
SB/Turner's	2	5	3	3.3
Total / Weighted Avg	27	4.7	59	6.0

Table 18: Average Score by Disorder Sub-Population

[5] Statistically Significant

[6] Statistically Significant

5.2.4. Summary

Overall, both test subject groups performed better when compared to the Expert Gold-Standard than when compared to the medical record diagnosis. The specificity of the referral decision of the human subjects improved the most, going from 0.53 to above 0.7. This contributed to the resulting statistically significant correlation between the expert's referral decisions and those of the human subjects.

Using the scoring algorithm, the human subjects performed better than TrenDx, earning higher average scores over all the patient cases and when the patient cases were grouped as normal/abnormal. The difference in scoring was statistically significant at $\alpha=0.05$.

5.3. Comparisons Without Gold-Standards

This section details the results of several comparisons in which the two gold-standards are compared to each other and comparisons between the different groups, in which there is no clear gold-standard.

5.3.1. Expert vs. Medical Record

Table 19 shows the results of the comparison of referral decisions between the expert pediatric endocrinologist and the medical record diagnosis. The expert referral decisions are similar to those in Table 6: Pre-Residency Subjects vs. Medical Record Diagnosis; they have a high sensitivity and low specificity.

Expert Decision vs. Medical Record Diagnosis

Decision	Abnormal	Normal	Total
Refer	19	40	59
No Referral	8	28	36
Total	27	68	95

Table 19: Expert Decision to Refer vs. Medical Record Diagnosis

Sensitivity: 19/27 = 0.70

Specificity: 28/68 = 0.41

Fisher's Exact Test - P=0.3534

Table 20 is in the same format as Table 8 through Table 11. It lists the referral and preliminary diagnosis performance of the experts, using the medical record as the gold-

standard. In this performance measure, the experts performed better than all other groups, having both a high Correct Referral % and the highest Correct Dx percentages.

Disorder Population Referral and Diagnosis Results
Expert vs. Medical Record

Disorder Sub Population	#	No Refer	Refer	Correct Ref %	Correct Dx	Correct Dx %	Correct Dx % (of Refs)[7]
Precocious Pub	6	1	5	83.3%	3	50.0%	60.0%
GH-Def / Hypothyroid	11	3	8	72.7%	7	63.6%	87.5%
Cancer / Complex	8	2	6	75.0%	N/A	N/A	N/A
Short Bone / Turner's	2	2	0	0%	0	0%	0%
Total	27	8	19	70.4%	10/19	52.6%	76.9%

Table 20: Disorder Population Referral and Diagnosis, Expert vs. Medical Record

5.3.2. Multiple Comparisons of Consensus Cases

Of the 95 patient cases in this trial, 59 of them had more than one human subject/reviewer. Out of those 59 patient cases, there was unanimous consensus among 29 of the human subjects in terms of the referral decision, with 18 referrals and 11 patients not referred. For each of those 29 cases, we looked at the combination of the medical record diagnosis, the expert's decision, TrenDx's decision, and the consensus decision of the physicians. Among the 29 cases, there were 4 cases in which every group did not refer the patient. There were 4 cases in which every group did refer the patient. Then we looked at cases in which there were singular decisions (one group made one decision and all other groups made the opposite decision). There were 9 cases in which the medical record stated that the patient was normal but all of the other groups referred the patient. There were three cases in which the experts had a singular decision; they were all decisions to refer. There were 2 cases in which the decision of TrenDx differed from those of all the other groups. In one, TrenDx referred and in the other, TrenDx did not refer. The human subject group had no singular decisions. There were 7 remaining decisions in which two of the groups referred and two didn't. These results are summarized in Table 21.

---

[7] Correct Dx% of refs does not include Cancer/Complex cases. Neither does Total.

Consensus and Singular Referral Decisions

Description	Number
All groups agree - Refer	4
All groups agree - No referral	4
Normal (MR), all others referred	9
Expert refer, no one else referred	3
TrenDx refer, no one else referred	1
TrenDx no referral, all others referred	1
Split decision	7
Human subject singular decisions	0

Table 21: Consensus and Singular Referral Decisions

Looking at all of the patient cases, there were only 2 cases in which the medical record indicated that a referral was necessary, but none of the other groups referred the patient. Only 1 human subject reviewed those patient cases. There were 3 cases in which the patient was abnormal and only TrenDx triggered a referral (including experts). And there were 3 cases in which the patient was abnormal and only one of the human subjects referred the case. There were no cases in which the patient was abnormal and the expert was the only one to refer the case.

5.3.3. Summary

In the comparison of the expert decisions to the medical record diagnoses, the experts performed fairly well, having a high sensitivity score and the highest percentage of correct preliminary diagnoses. However, their specificity was lower than any of the subject groups. The consensus analysis had several interesting results which will be discussed later.

5.4. Variations in Threshold Triggering

As described in Section 4.3.4, the threshold triggering values were obtained by processing twenty 'normal' cases and using the lowest error-score of the trend-templates for Normal Growth, Constitutional Delay of Puberty, and Early Puberty. The threshold for single point triggering was set at 0.35, which was 2 standard deviations from the mean. The threshold for consecutive triggering was set at 0.31, approximately 1.5 standard deviations from the mean.

This section describes the results obtained from raising and lowering the threshold triggering values by one half of the standard deviation, and the effects of ignoring certain data points.

5.4.1. Results of Raising Threshold Triggering Values

In the first test, both thresholds were raised by one half a standard deviation. The single-point triggering threshold became 0.39 and the consecutive-point triggering threshold became 0.35. Table 22 shows the results of raising the triggering thresholds. Recall that TrenDx referred 50 of the 95 cases. Raising the triggering thresholds prevented 7 of the 50 cases from triggering a referral. Of those 7 patients, 6 were normal and 1 was abnormal.

Results of Raising Triggering Thresholds

Decision	Abnormal	Normal
Refer	15	28
No Referral	11	38

Table 22: Results of Raising Triggering Thresholds

Sensitivity: 0.56[8]

Specificity: 0.56

Chi-Square Test - P=0.2468

5.4.2. Results of Lowering Threshold Triggering Values

Then, the effects of lowering the triggering thresholds by 1/2 a standard deviation was examined. The single-point threshold was lowered to 0.31 and the consecutive-point threshold was lowered to 0.27. These values represent 1.5 and 1 standard deviation from the mean error-value, respectively. Using the new, lowered triggering thresholds, 7 new

[8] Sensitivity and Specificity calculated out of 27 abnormals and 68 normals.

referrals occurred. Of these, 2 were abnormal patients that deserved to be referred and 5 were normal patients. The results are shown in Table 23.

Results of Lowering Triggering Thresholds

Decision	Abnormal	Normal
Refer	18	39
No Referral	8	27

Table 23: Results of Lowering Triggering Thresholds

Sensitivity: 0.67

Specificity: 0.40

Chi-Square Test - P= 0.997

Neither raising nor lowering the triggering thresholds had an immense or unexpected impact on the sensitivity and specificity of TrenDx referrals. Raising the threshold caused TrenDx to deny more referrals and 'miss' one abnormal child but prevent 6 unnecessary referrals. Lowering the threshold caused TrenDx to recommend more referrals, 'catching' two abnormal children while causing 5 normal children to be referred.

5.4.3. Ignoring Certain Error-Scores

During the processing of the patients by TrenDx, it was noticed that all hypotheses frequently scored very poorly on the first non-trivial (<> 0) error-score. Specifically, in many cases the error-scores for a patient's first data point would be very high, then drop to below the triggering level. The reason for this was because some constraints require a certain number of data points before a meaningful match can be made. For example, if an interval had 2 constraints, one constraining a parameter to be equal to 0 and a second which constrained the same parameter to be linear with a slope of 0, then any single data point would match to the first constraint, but would match trivially to the second. Then, when the next data point was examined, TrenDx would be able to generate an error-score for the second constraint. Since the overall error-score for the hypothesis is a weighted average of all the error-scores, the overall error-score could change dramatically between the first few data points. Therefore, we examined the results of "ignoring" the referral if it was made on the first, non-trivially scoring data point. In that case, there were 7 fewer referrals. However, 2 of those 7 were abnormal patients who should have been referred. Table 24 lists the resulting sensitivity and specificity.

Results of Ignoring First Non-Trivial Point

Decision	Abnormal Patients	Normal Patients
Refer	14	29
No Referral	12	37

Table 24: Results of Ignoring First Non-Trivial Point

Sensitivity: 14/27 = 0.52

Specificity: 37/68 = 0.54

Fisher's Exact Test - P=0.4877

Results of Ignoring Infant Scores

Decision	Abnormal Patients	Normal Patients
Refer	15	28
No Referral	11	38

Table 25: Results of Ignoring Infant Scores

Sensitivity: 15/27 = 0.56

Specificity: 38/68 = 0.56

Fisher's Exact Test - P=0.2468

In addition, it was noted that the error-scores for infants were somewhat higher than for adults. This suggested that the trend-templates did not model infancy well enough. We considered the results of ignoring infant scores and only looking at childhood data points (> age 3) for those patients that had childhood data. Out of the patients whose error-scores triggered a referral, ignoring the infant data point error-scores prevented 7 of them from being referred. Of those 7, only 1 was abnormal and deserved to be referred. The consequences of ignoring infant data points (< age 3) are presented in Table 25.

5.5. Referral Decision Timing

One goal of TrenDx was to try to "catch" children with growth disorders early in order to minimize morbidity. Thus, we compared the timing of the referral decision of TrenDx to the time at which the expert felt that the patient should have been referred. This comparison only makes sense for cases in which the child was abnormal and the expert and TrenDx both referred the child. There were 13 cases which met that criteria (Table 13: TrenDx Decision vs. Expert and Medical Record Consensus).

Of those 13 cases, 4 were in the Precocious Puberty disorder subpopulation, 5 were GH-Def/Hypothyroidism, and 4 were Complex/Cancer cases. The results are shown in Table 26.

Timing of Referrals for TrenDx vs. Expert

Disorder Subpopulation	# Ref	> 1 yr. before Expert	Within 1 yr. of Expert	> 1 yr. after Expert
Precocious Puberty	4	1	1	2
GH-Def / Hypothyroid	5	3	2	0
Complex / Cancer	4	2	0	2
Total	13	6	3	4

Table 26: Timing of Referrals for TrenDx vs. Expert

Table 27 shows the results of the same analysis for the human subjects. Again, for the cases which were abnormal and where the expert and subject both referred the case, we looked at the timing of the referral decision by the subject. Sometimes, the subjects did not circle a referral date on their response sheet.

Timing of Referrals for Human Subjects vs. Expert

Disorder Subpopulation	# Ref[9]	> 1 yr. before Expert	Within 1 yr. of Expert	> 1 yr. after Expert
Precocious Puberty	9 (8)	2	4	2
GH-Def / Hypothyroid	8 (5)	2	3	0
Complex / Cancer	8 (7)	0	5	2
Total	25 (20)	4	12	4

Table 27: Timing of Referrals for Human Subjects vs. Expert

[9] Numbers in parentheses indicate the number of cases in which the subject did circle a referral date. These were the only responses that could be evaluated.

6. Discussion

The performance of TrenDx and issues such as changing the triggering mechanisms are discussed separately from the interpretation of the patient cases.

6.1. Analysis of Results

6.1.1. Performance of Test Subjects vs. Medical Record Diagnosis

Overall, the human subjects referred abnormal patients about as often as TrenDx but with a higher percentage of correct preliminary diagnoses. The Pre-Residency subjects referred a higher percentage of the abnormal patients than TrenDx or the Post-Residency subjects, while the Post-Residency subjects chose the correct preliminary diagnosis almost twice as often as the other two groups.

It is conceivable that the Pre-Residency subjects had a higher sensitivity score because they did not have as much training as the other subjects and conservatively referred any patient whose data looked even slightly suspicious. This simple explanation is supported by the fact that they also had the lowest specificity score, actually referring over half of the normal patients that they were presented with. However, it is difficult to make any conclusion because this population only consisted of 5 subjects, for a total of 49 cases (one of the cases was removed from the trial). In comparison, the 17 Post-Residency subjects were presented with over 150 cases. In fact, if the performance of the Pre-Residency subjects was copied two more times, to get a total of 147 decisions, the P value from the Fisher's Exact Test would be 0.0074. Of course, this analysis has no statistical basis, but is interesting nonetheless.

The difference in training is also a possible explanation for the fact that the Post-Residency subjects were able to choose the correct preliminary diagnosis twice as often as the Pre-Residency subjects (and TrenDx), even though they had the lowest referral sensitivity.

Another explanation is that the Pre-Residency subjects had a higher sensitivity score because the material (pediatric endocrinology) was more recently learned. For example, some of the Post-Residency subjects were attending in other divisions of pediatrics such as Emergency Medicine or Neonatology. They may not have remembered the subtle differences between the disorders that this trial involved.

6.1.2. Performance of Test Subjects vs. Expert Gold-Standard

All test subjects (TrenDx and human subjects) performed better when compared to the Expert Gold-Standard as opposed to the Medical Record Gold-Standard. The physicians performed significantly better than TrenDx. The P-value of the Fisher's Exact Test indicated a high probability of correlation ($P<0.0001$) between the referral decisions of the physicians and those of the experts. Furthermore, the scores of the Post-Residency subjects and the human subjects as a whole were statistically likely to be higher than the scores received by TrenDx ($P < 0.05$).

The use of "Consensus Sensitivity and Specificity" in the comparison represented an arguably more concrete gold-standard than either the medical record diagnosis or the expert response individually. Consensus sensitivity was used because it represented the sensitivity for the set of cases in which there was a disorder and the limited data that was available was enough to suggest a referral. Similarly, the consensus specificity represented the specificity for those cases in which there was no underlying abnormality and the data did not suggest any problems.

There are also some interesting results from Table 12: TrenDx Decision vs. Expert Gold-Standard, Table 14: Physicians vs. Expert Gold-Standard, and Table 17: Test Subject Scores. The sensitivity and specificity of TrenDx referral decisions are fairly close, 0.610 and 0.528, respectively. TrenDx scores an average of 4.7 points over both normal and abnormal patients. In contrast, for the participants' decisions, the specificity is quite a bit higher than the sensitivity, at 0.747 vs. 0.640, respectively. From this, one might reasonably expect the average score for the normal patients to be higher, since they referred a higher percentage of them. However, the average score received for the set of abnormal patients was 6.0 whereas the average score for the normal patients was 5.18.. The confounder in this relationship is the fact that the score is relative to the decisions of the Expert Gold-Standard, while the disorder subpopulations are based on the medical record diagnosis.

6.1.3. Comparisons Without Gold-Standard

The comparison of the two gold-standards was necessary to help prove whether the task that we had set for the participants was possible to accomplish. Comparing the expert referral decisions to the medical record diagnoses resulted in a sensitivity value of 0.70, a specificity value of 0.41, and a P-value of 0.3534 (Table 19: Expert Decision to Refer vs. Medical Record Diagnosis). These results do not strongly support the conclusion; the experts appeared to have over-referred the cases to help achieve such a high sensitivity score. However, in Table 20, one can see that the experts performed much better than any

of the other groups. Not only was their referral sensitivity high, they had the highest percentage of correct preliminary diagnoses. Thus, in terms of the patients with disorders, the experts were able to refer them correctly 70% of the time and get the correct diagnosis more than 50% of the time.

The results shown in Table 21: Consensus and Singular Referral Decisions, also helps support the conjecture that the task can be accomplished. First, note that out of the 59 cases with more than 1 participant reviewer, there was consensus on 29 of the cases, or 49.2%. Having almost half of the cases have participants be unanimous in their referral decisions, when anywhere from 2 to 5 physicians reviewed each case is unlikely unless there is some correlation between the data and the referral decision. Furthermore, of all 95 cases, there were only 2 cases in which the patient was abnormal and was not referred by either TrenDx, an expert, or a human subject. In each of those cases, only 1 participant reviewed the case.

Admittedly, the task is very difficult. This is especially true because of the limited data set that was available. However, the data listed above suggests that the task can be accomplished, perhaps with the availability of a few more types of information such as pubertal stage information or maybe even just increasing the frequency and precision of the collection of the measurements that we are already using.

6.1.4. Referral Timing

The timing of the referral decision compared to that of the experts does not lead to any strong conclusions. TrenDx referred many cases earlier than the expert (6 of 13), while the physicians referred most of the cases at about the same time as the expert (12 of 20). One could consider this evidence that the referral decisions of TrenDx were because it is not likely that TrenDx could catch the abnormality before the expert, especially since the expert was also seeing all the data at once.

6.2. Performance of TrenDx

6.2.1. Worse than Physicians (?)

Some of the results have indicated very strongly that the performance of TrenDx at the task of referring and diagnosing children with possible growth disorders is not as good as that of physicians, when compared to an expert. At this stage of development of TrenDx, this is likely to be true.

One might also wish to consider the statement that TrenDx is not yet able to refer and diagnose children with possible growth disorders in the same manner as experts and

physicians. The following information may be considered evidence for this statement. TrenDx and the participants had exactly the same sensitivity value (0.59), when compared to the medical record diagnoses. Their specificity values were similar, but not identical (0.47 vs. 0.53). When compared to the expert, their sensitivities were still similar, 0.61 vs. 0.64 for TrenDx and the physicians, respectively, but their specificities were very different, 0.52 vs 0.75 for TrenDx and the physicians, respectively. This indicates that the physicians chose not to refer many of the same cases that the experts chose not to refer. Now, comparing Table 13 and Table 15, out of the 8 abnormal patients that the experts did not refer, TrenDx referred 3 of them, or 37.5%. For those same 8 cases, the human participants had to make the referral decision 16 times and only referred 3 of them, or 18.8% (half as often). In fact, none of the physicians referred those 3 cases either. As another comparison, out of the 16 abnormal cases that TrenDx referred, the experts referred 13 of them, or 81.3%. Out of the 35 abnormal cases that the physicians referred, the experts referred 32 of them, or 91.4%. In summary,

1. TrenDx and physician sensitivity / specificity is similar when compared to the medical record but dissimilar when compared to the expert. In fact, there is a statistically significant correlation between the decisions of the physicians and the decisions of the experts, while the correlation between TrenDx and the experts is <u>not</u> statistically significant.

2. Of the abnormal patients that the experts did not refer, TrenDx referred twice as high a percentage as the physicians did. Remember that TrenDx and the physicians have similar sensitivities.

3. Of the 16 abnormal cases that TrenDx referred, the experts only referred 81.3% of them, while they referred 91.4% of the 36 abnormal cases that the physicians referred.

This "evidence" may have no relevance, but was interesting.

6.2.2. Reasons for Poor Performance

The reasons for the poor performance of TrenDx can be categorized under problems with the engineering and problems with TrenDx itself. The engineering performance issues will be discussed first.

The engineering issues can also be divided into two categories - preprocessing and postprocessing. Preprocessing refers to the development of trend-templates and the design of value constraints to obtain the desired performance. In this trial, the models of infancy were insufficient. The models were based on those taken from the previous trials of

TrenDx and were not developed to the same degree as the for childhood and puberty. It was assumed that early puberty and constitutional delay of puberty would have little effect in infant growth. The modeling with the expert for the disorder trend-templates such as Precocious Puberty and Short Bone Syndrome focused on the effects during childhood and afterwards, so very little time was spent on modeling the effects of the different disorders during infancy. It is likely that more modeling and engineering of the trend-templates, such as changing value constraint parameters and weights, could produce better performance.

We have already demonstrated that variations in the Postprocessing of TrenDx can lead to changes in performance (Table 22: Results of Raising Triggering Thresholds and Table 23: Results of Lowering Triggering Thresholds). Table 24: Results of Ignoring First Non-Trivial Point and Table 25: Results of Ignoring Infant Scores present different ways to process the output of TrenDx. The results of Table 25 are actually diminished because the infant scores contributed a lot of error to the overall error-score, even into childhood / puberty. It is likely that the high-error-scoring infant portion of the trend-templates caused many normal patients to be referred that otherwise would not have been referred.

TrenDx also crashed while processing 3 of the patients. These were due to the fact that the patients had either height or weight measurements that were extreme (either high or low). TrenDx does not gracefully handle functions which may not return an answer. This is a developmental problem due partly to the fact that TrenDx is an experimental program and not a commercial software.

Aside from the engineering issues, there are some problems inherent in applying TrenDx to this domain. One of the most obvious problems is the fact that regression-based TrenDx uses linear regression techniques to match data to value constraints. These techniques often require several data points to even have a meaningful value. In a domain such as pediatric growth monitoring, data points come at irregular intervals and a single measurement may often be enough evidence in itself to diagnose the condition or refer the child. As another side effect of the linear regression techniques, TrenDx is very sensitive to the first few initial data points. For example, the trend-templates for most growth states constrained the Z-score of the ht to be within some certain range, and for the Z-score to have a slope of 0 from childhood to adulthood. For the normal growth template, the range was set to be ± 2 standard deviations from the mean. Since the second constraint constrained the velocity of the Z-score, it would not contribute to the error score if there were only 1 data point and so the overall error-score for the trend-template hypothesis at the first time point would be composed of the error-score for the constraint that the Z-score be within ± 2. However, later data points would allow the second constraint to be applied and

could change the overall error-score significantly. Thus, TrenDx frequently triggered a referral on the first non-zero error-score that it produced. In other words, TrenDx didn't know that the strength of its conclusions should be based on the number of points it had matched.

There are numerous other weaknesses of TrenDx. For example, TrenDx has no way to explicitely state the probability that the process will be in a certain state. This is actually both a strength and a weakness, because it allows TrenDx to model processes in which these probabilities are unknown or do not exist. Another weakness is the fact that TrenDx does not model the thought processes of an expert. Currently, TrenDx matches the data to each trend-template hypothesis - in effect, TrenDx considers every possibility. Human experts do not reason in this manner because they lack the computational power and because this style of reasoning lacks focus.

Another problem inherent in the domain is the fact that the system that is being dealt with is the human body, which varies from person to person. This was briefly discussed in section 4.2. Since the body varies, the effects of each disorder vary as well. In some cases, a disorder may affect height and weight to a great extent, in which case it would be easier to diagnose. If a correct model were available, then the creation of trend-templates would be much simpler. However, this is also a strength of TrenDx, that it is able to perform without a detailed model of human physiology or endocrinology.

Adding to the difficulty of the task is the fact that there is a lot of noise in the data. There were several instances in the patient cases where the height of the patient dropped a small amount. Since it is very unusual for children to shrink, one can assume that some type of measurement error took place. TrenDx is unable to reason in this fashion - like most other expert systems it takes its data at face value. Moreover, even if the measurements were incredibly accurate, the growth of a child does not follow the model exactly. There are even differences in growth rate between the different seasons (Tanner and Davies).

One valuable by-product often associated with expert systems is the ability to explain the reasoning process, such as in MYCIN {Shortliffe, 1976 #56}. The knowledge embedded in the trend-templates and the tracing output of TrenDx as it processes data provides a similar feature. It explains the results as it matches each new data point to the appropriate value-constraint of a trend-template. From this output, a user can see which data point matched to which value constraint poorly.

6.3. Limitations of This Trial

In the evaluation of the program DxPlain, Feldman notes that "There are often inherent difficulties in designing elegant tests: these may not have been appreciated and dealt with adequately in the study design." (Feldman and Barnett). As with any other trial, there were limitations and biases inherent in the design and the execution of this trial.

The Hawthorne Effect, which describes the tendency of subjects in a trial to perform at a higher level than normal because they know that they are being examined, must always be considered a source of bias. However, in this trial we hope that the subjects always try to perform at their highest level, since they are physicians. One manner in which their performance may differ from normal practice is the amount of time that they spent on the cases in this trial, since a doctor in a busy clinic may not have the same amount of time.

Another effect, the so-called Checklist Effect, describes the effect of having clean data presented to the subject in a uniform, coherent manner. This effect would contribute to the improved performance of the test subjects over the original pediatrician who referred the child to the growth clinic. However, we do not make that comparison explicitly. A second source of bias comes from the fact that the participants see all the data at once. That prevents them from looking at the data serially, which both TrenDx and the original pediatrician of the patient had to do. For example, if the patient had a height measurement that was relatively high or relatively low but was later corrected, the human subject could easily give less weight to the outlying point. This is another case in which "hindsight is 20/20".

This trial only evaluated the performance of the test subjects on a set of 95 referral cases. Not only were the cases difficult, but the data that we presented to the subjects was limited to the patient's height, weight, and bone-age information available up to the time of referral to the endocrine clinic. This made diagnosis a difficult task, but it was necessary to simulate the condition in which we wished to evaluate TrenDx.

Another problem with the trial was that there was geographical bias. This bias comes from the fact that the cases were obtained from a single institution and that the gold-standard experts were taken from the same institution. Most of the participants were also from the same geographical area (Boston Longwood Medical Area), but the recruitment of participants from all over the country and internationally by the use of internet postings help remove this bias. A breakdown of performance by geographical location would be

interesting but we are constrained by time. In addition, the small number of participants would preclude making any kind of conclusion. Another, more severe bias comes from the fact that the cases come from the patients of the same clinic as the gold-standard experts. This bias is limited by the use of non-active files (> 1 yr. old), but is present nonetheless.

One must also question the gold-standards that were used in this trial. As noted in section 3.2, there are many problems with the use of the medical record as the gold-standard. There were some problems with incomplete problem lists and uncertain diagnoses; some medical records became lost from the hospital and their cases had to be removed from the trial; and the labeling of the patients that were not referred for suspicion of growth abnormalities as "normal" is questionable. Furthermore, even the medical record diagnosis may not be applicable. For example, many of the "normal" patients in the trial have height measurements below 2 standard deviations from the mean for their age could easily argue that a child that short should not be categorized as "normal." Some of these biases were unavoidable. Once we decided that we wanted to process all the cases that were referred to the clinic so that the true proportions of each disorder would be represented, we had to deal with cases that were not referred for suspicion of growth abnormalities.

The second gold-standard introduced several problems as well. The original design of the trial had 2 pediatric endocrinologists review each case individually. Each test subject would receive 2 scores, based on both of the responses of the experts. In addition, this allowed their consensus, or lack thereof, to give an indication of the difficulty of diagnosing a particular patient. If both experts agreed, then that suggested that there was little room for questioning the decision and that the same diagnosis was clearly available from the data. However, due to time constraints the use of 2 experts was not possible. There were other problems as well. Often, the expert did not fill out the entire response sheet. For those patients, it was not possible to get a total score of 10 because the subject's response could not be matched against the expert's.

The weaknesses of the scoring algorithm also need to be discussed. First, the assignment of equal value to normal and abnormal cases must be questioned. Depending on the goals of the evaluation, it may be desirable to weigh one group more heavily. In addition, our scoring mechanism gives no weight to the difficulty of the case. Another weakness of the scoring system can be described in the following scenario: If the patient's data suggests that the most likely diagnosis is Constitutional Delay, but that there is a possibility of Growth Hormone Deficiency, then the expert might choose to refer the patient. However, if TrenDx or the human subject correctly diagnoses the patient as having

Constitutional Delay, they might choose to deny the referral. Furthermore, if the subject does choose to refer the patient and chooses Growth Hormone Deficiency as his or her preliminary diagnosis, then should they get only 2 points for getting the second-ranked disorder? Or should the expert rank Growth Hormone Deficiency first? These are just some of the issues which face any scoring mechanism and are one reason why we evaluated the performance of the subjects in multiple ways.

Another weakness is that we have no way to evaluate the real value of getting the referral decision and preliminary diagnosis correct. How much money would it save? Would it help children reach their growth potential? Would it help diagnose cancer patients early? This evaluation has no way of addressing those questions.

The participation of subjects in this trial was another major source of problems. As described in Section 3.1, over 80 packets were distributed but only 22 were returned. This caused the number of times a patient was reviewed by a participant to vary between 1 and 5. Since it was not possible to predict which participant would return their packet, some type of solution for this problem should have been conceived and implemented.

7. Conclusion

7.1. Summary of Results

The comparison between the medical record diagnosis and the decisions of the experts indicates that this task can be accomplished, to some extent. One might suggest the use of more data to improve the possible performance at this task.

Compared to the medical record diagnosis, TrenDx and the physicians performed similarly in terms of referral decision sensitivity and specificity. However, the physicians were able to choose the correct diagnosis more often (33% vs 21%). There was a statistically significant correlation between the referral decisions of the experts and the physicians. The correlation between the experts and TrenDx was not significant. The physicians also scored significantly higher than TrenDx, according to our scoring algorithm.

7.2. Lessons Learned

It is important to characterize what we have learned about TrenDx from this experiment. We have presented results which suggest that the computer program TrenDx is able to diagnose and refer children with possible growth disorders near the level of proficiency of physicians. In addition, we have shown that the task which we defined can be accomplished to a certain extent by only looking at the data that we have chosen to use. Yet the difference between the performance of TrenDx and the physicians leads one to ask, "Is TrenDx the right tool to use for this task?" and "Is TrenDx suitable for use as a screen for the general population?"

The answer to the first question is uncertain. TrenDx has many attributes which make it useful for monitoring growth. The trend-representation language is fairly simple and intuitive. The ability to manipulate time points and express uncertainty in time is useful, especially for monitoring processes in which events can occur at different times. In truth, the task is a difficult one, and whichever tool one chooses to use will be able to do some parts of the task well, but other parts poorly.

The answer to the second question is also complicated. The performance of TrenDx on this set of patients cannot be extrapolated to its performance as a growth screening program in the general population. In general, a predictive test will have to make a tradeoff

between sensitivity and specificity. To be used as a screening tool in the general population, specificity is often considered more important, especially if the disorder is very rare, because of the monetary and psychological cost of dealing with the more numerous false-positives. However, since the test population in our trial has a high percentage of abnormal cases, one may argue that our sensitivity and specificity values have somewhat different interpretations than they usually do. To test the performance of TrenDx as a growth-screening program, an entirely new study would have to be devised because the test population in this study does not represent the general population. However, the results that we have obtained so far suggest that TrenDx is not ready for use on the general population.

7.3. Future Work

: Improvement on TrenDx can take the form of more engineering and improvement of the trend-templates, intervals, value-constraints, and triggering mechanisms. Changing these parameters may allow TrenDx to perform as well as physicians. To go beyond that, however, will probably require much more work to improve TrenDx.

While modelling with the trend-templates, I considered the usefulness of adding some type of simple boolean logic operators to the trend-templates. For example, it is conceivable that a parameter will decrease at some velocity or stay constant. TrenDx currently has no mechanisms to express this.

Currently, TrenDx does not actually "trigger" any action when the error-scores pass a certain threshold. This feature may be useful to represent the reasoning of experts, who may consider one monitor set, but if a certain trigger goes off, then a new monitor set is loaded.

To conclude, TrenDx did not perform better than the physicians at the task of diagnosing growth disorders in children from their height, weight, and bone-age data. However, TrenDx did perform in a somewhat intelligent manner and appears to have the potential to be improved upon.

8. Appendix A - Patient and Subject Result Tables

Y* = complex

N* = Normal, ref for other reasons

- means no answer written down.

1* means no answer, assume No Info

Patient #2000

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obes	Score
MR	N			X								X	10
Exp	N			1								X	
TrDx	N		X	.								X	8
Sub1	Y	4.25		X								X	3
Sub21	N			X								X	10
Sub60	Y	5.5					X					X	1

Patient #2001

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obes	Score
MR	N				X								0
Exp	Y	8.75							1	X	x		
TrDx	Y	11				X				X	x	X	6
Sub1	N				X								0
Sub13	N				X								0
Sub21	Y	9.75			X					X	x		6
Sub58	Y	11			X	x				x	X		6
Sub60	Y	8.75			X								7

Patient #2002

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	N			X									10
Exp	N			1									
TrDx	Y	1.25							X			X	0
Sub13	N				X								7
Sub21	N					x			X	x	X		7
Sub58	N					x			X	X	X		7
Sub60	Y	1.25				X					x		0

Patient #2003

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	N			X									10
Exp	N			1									
TrDx	Y	8						X					0
Sub13	N				X								7
Sub22	N				X	x				X	X		7
Sub58	Y	10			X								0
Sub60	Y	11				X							0

Patient #2004

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	N				X								0
Exp	Y	6.75							1				
TrDx	Y	3.92						X					5
Sub14	Y	2							X				8
Sub22	Y	3.92			X					X	x		5
Sub59	Y	3.92							X				8
Sub60	Y	4.17						X					5

Patient #2005

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	N			X									0
Exp	Y	-							1	X	X		
TrDx	N				X								0
Sub14	N				X								0
Sub22	N			X		x				x	X		1
Sub59	N			X									0
Sub60	N								X				0
Sub81	Y	11							X				5

Patient #2006

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	N				X								7
Exp	N			1									
TrDx	N				X								7
Sub14	N								X				7
Sub23	N								X				7
Sub59	N								X				7
Sub60	N								X				7
Sub81	N				X								7

Patient #2008

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	N*			X									7
Exp	N			-								X	
TrDx	N		X									X	8
Sub15	N			-								X	8
Sub23	N			X								X	8
Sub60	N			-								X	8
Sub81	Y	9							X			X	1

Patient #2009

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	Y						X						6
Exp	Y	7	1										
TrDx	Y	7.33					X					X	6
Sub15	Y	7					X						7
Sub24	N			X									0
Sub60	Y	7	X										10
Sub81	Y	5.75							X			X	5

Patient #2010

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	Y*				X				X				8
Exp	Y	8.25			1								
TrDx	Y	7				X					x		5
Sub16	Y	8.25				x		X		X	X		7
Sub24	Y	8.25							X				7
Sub61	Y	10.75				X			X	X	X		5

Patient #2011

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	N*			X									10
Exp	N			1									
TrDx	Y	0.25							X				0
Sub16	N			X								X	10
Sub24	N			X									10
Sub61	N			X									10

Patient #2012

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	Y*												5
Exp	Y	13						1					
TrDx	Y	14.8				X					x	X	5
Sub16	Y	13				x		X		X	X		10
Sub61	Y	13.8					X						6

Patient #2013

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	N				X								0
Exp	Y	13						1					
TrDx	Y	7.7						X					8
Sub17	N				X								0
Sub61	N			X									0

Patient #2014

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	Y			X							X	X	3
Exp	N			1									
TrDx	Y	15.17					X			X	x	X	0
Sub17	N			X								X	10
Sub61	N			-								X	5

Patient #2015

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	N*			X									0
Exp	Y	.75					1						
TrDx	Y	0.25							X				5
Sub17	Y	?							X				5
Sub61	Y	1.33							X				5

Patient #2016

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	N			X									0
Exp	Y	1.25		-						X	x		
TrDx	Y	0.75							X				5
Sub18	Y	1.29		?		x				x	X		7
Sub61	Y	1.29						X					6

Patient #2017

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	Y			X							X		8
Exp	Y	1		-						1	x		
TrDx	N			X									0
Sub18	Y	1				X					x		8
Sub61	Y	1				x		X		x	X		8

Patient #2018

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	Y						X					X	6
Exp	Y	2		-								X	
TrDx	Y	3.5							X			X	7
Sub18	N								X				0
Sub27	Y	3.5										X	7
Sub61	Y	4.17							X			X	6

Patient #2019

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	N*			X									0
Exp	Y	0.17					1					X	
TrDx	N			X								X	1
Sub27	N			X									0
Sub61	N			X									0

Patient #2020

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	N			X									0
Exp	Y	9						1					
TrDx	Y	6.25				X						X	5
Sub27	Y	9				x			X	x	X		7

Patient #2021

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	N				X								10
Exp	N				1								
TrDx	N				X							X	10
Sub90	Y	2.17				X							0

Patient #2022

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	N*			X									10
Exp	N			1									
TrDx	-			-									0
Sub20	Y	.67				x			X	x	X	X	0

Patient #2023

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	N				X								0
Exp	Y	3							1*				
TrDx	Y	.33							X				8
Sub20	Y	1.5				X					x		5

Patient #2024

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	N				X								7
Exp	N			1									
TrDx	Y	4.5				X					x		0
Sub20	N			-									7

Patient #2025

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obes	Score
MR	N			X									0
Exp	Y	4.58			1								
TrDx	Y	2							X				5
Sub1	Y	2				X				X	X		5
Sub22	N				X					X	x		3
Sub59	Y	2				X					x		5
Sub75	N			X									0

Patient #2027

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cons GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hypo Thyr	Obesity	Score
MR	N*			X									10
Exp	N			1									
TrDx	Y	3							X				0
Sub1	N			X								X	10
Sub23	N			X									10
Sub75	Y	6?			X								0

Patient #2028

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	Y*									X	X		6
Exp	Y	8				x				X	X		
TrDx	Y	10.5				X					x	X	8
Sub23	Y	10.5				X				X	x		8
Sub75	Y	8.33						X					6

Patient #2029

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	Y							X					0
Exp	N			1									
TrDx	N			X									10
Sub24	Y	5							X	X	x		0
Sub75	N			-									0

Patient #2030

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	Y					X					X		5
Exp	Y	6.33			1								
TrDx	Y	3						X					5
Sub24	Y	6.5				X					x		6

Patient #2031

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	N*			X								X	10
Exp	N			1								X	
TrDx	Y	0.58							X				0
Sub90	N			X								X	10

Patient #2032

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	N				X								2
Exp	Y	4.5			1						X		
TrDx	Y	4.75				X					x		7
Sub90	Y	1.25				X					x		6

Patient #2033

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	Y							X					0
Exp	N				1								
TrDx	Y	5.1						X					0
Sub13	N				X								10

Patient #2034

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	N			X									2
Exp	Y	0.79		1		x				x	X		
TrDx	N			X									2
Sub13	Y	0.79							X				7

Patient #2035

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	N*			X									0
Exp	Y	9.25							1				
TrDx	Y	7.75						X					5
Sub14	Y	7.75				X					x		5
Sub27	N								X				3
Sub87	N			X									0

Patient #2036

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	N			X									0
Exp	Y	9.67							1				
TrDx	N				X								0
Sub14	N			X									0
Sub27	N			X									0
Sub87	Y	1.17							X				5

Patient #2037

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	N*			X									0
Exp	Y	1.75			1								
TrDx	Y	3.5						X					5
Sub15	Y	1.75						X		X	x		7
Sub87	Y	2.5			X								8

Patient #2039

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	N*				X								0
Exp	Y	11.58	1										
TrDx	Y	5.5						X					5
Sub16	Y	?			X					X	X		5
Sub87	Y	5.5							X	X	x		5
Sub90	N							X					0

Patient #2040

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	N				X								10
Exp	N				1								
TrDx	Y	0.12							X			X	0
Sub16	Y	1.58				X				X	X		0

Patient #2041

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	N			X									3
Exp	Y	2.25		1									
TrDx	N				X								0
Sub17	N								X				0

Patient #2042

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	N				X								7
Exp	N								1				
TrDx	N				X							X	7
Sub17	N			X									7

Patient #2043

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	N				X								10
Exp	N				1								
TrDx	Y	4.2						X					0
Sub18	N			X									7

Patient #2044

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	N			X								X	0
Exp	Y	1.5					1						
TrDx	Y	1.25							X			X	6
Sub18	Y			-								X	5

Patient #2046

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	N			X									0
Exp	Y	7.75					1						10
TrDx	Y	7.75					X						10
Sub89	Y	7.75					X						10
Sub90	Y	8.3					X						8

Patient #2047

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	N			X									10
Exp	N			1	2								
TrDx	N				X								9
Sub58	N			X									10
Sub89	N				X								9

Patient #2049

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	Y*												0
Exp	N			1									
TrDx	Y	1							X				0
Sub20	N			X									10
Sub58	N			X									10
Sub89	N			X									10

Patient #2050

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	N				X								0
Exp	Y	6.5		-									
TrDx	N				X								0
Sub20	N			X									0
Sub58	Y	9.5			X	x				x	X		5
Sub89	N				X								0

Patient #2051

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	Y						X						10
Exp	Y	6.75					1						
TrDx	N			X								X	0
Sub21	N								X				0
Sub59	N			X									0
Sub89	Y	6.75					X						10

Patient #2053

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	N*			X									10
Exp	N			1			3		2				
TrDx	N			X								X	10
Sub1	Y	5.75					X					X	1
Sub21	N			X								X	10
Sub65	N								X			X	9
Sub89	N								X				9

Patient #2054

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	N				X								0
Exp	Y	5		2			1			X	x		
TrDx	N			X									1
Sub1	N			X									1
Sub13	N			X									1
Sub20	N			X									1
Sub65	N			X									1
Sub89	N			X									1

Patient #2056

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	Y*					X				X	X	X	7
Exp	Y	5		2			1					X	
TrDx	-			-									0
Sub13	N											X	1
Sub21	N						X		X			X	3
Sub65	Y	2.17		X		x				x	X	X	8
Sub89	N			X								X	2

Patient #2057

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	N				X								0
Exp	Y	11.5		1					2	X	x		
TrDx	N			X									2
Sub13	N			X									2
Sub22	N			X									2
Sub65	Y	7.25				X					x		6
Sub89	Y	9.5			X					X	x		6

Patient #2058

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	N				X								2
Exp	Y	11.75			1			2	3	X	x		
TrDx	Y	0.50							X				5
Sub14	Y	10.75							X				5
Sub22	Y	?						X					6
Sub65	Y	1.5		X						X	x		6
Sub58	Y	9.25				X					x		6

Patient #2059

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	Y										X		6
Exp	Y	5.21		1	2	3				X	X		
TrDx	Y	0.12		-								X	5
Sub14	Y	0.5						X					5
Sub22	N			X									2
Sub65	Y	0.12						X					5

Patient #2060

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	Y										X		6
Exp	Y	0.85			3	1			2	X	X		
TrDx	Y	0.25							X			X	6
Sub14	Y	0.25							X				6
Sub23	N			X									0
Sub65	Y	0.12			X								5

Patient #2061

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	Y						X					X	8
Exp	Y	7.25	2	3			1						
TrDx	Y	8.83					X						8
Sub15	Y	8	X				X					X	9
Sub23	N		X			x				x	X	X	0
Sub65	Y	6					X						8

Patient #2062

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	N			X									0
Exp	Y	10			2	1			3	X	x		
TrDx	Y	10				X					x		10
Sub15	Y	12.5		?		x				X	X		8
Sub23	Y					X				X	X		8
Sub65	Y	10							X	X	x		8

Patient #2063

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	N				X								0
Exp	Y	11				1		2	3	X	x		
TrDx	N				X								0
Sub15	Y	13		?		x				X	X		8
Sub24	Y								X				5
Sub58	Y					x		X		X	X		8
Sub65	Y	10.33			X					X	x		6

Patient #2065

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	N*			X									10
Exp	N			1									
TrDx	N			X									10
Sub16	N			X									10
Sub24	N			-									7
Sub58	N			X									10
Sub87	N			X									10

Patient #2066

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	Y						X						0
Exp	N			1								X	
TrDx	N			X									9
Sub16	N			X								X	10
Sub24	N			X								X	10
Sub59	N			X								X	10
Sub87	Y	3		X								X	3

Patient #2067

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	N		X	X									10
Exp	N			1									
TrDx	Y	3						X				X	0
Sub16	N				X								7
Sub59	Y								X			X	0
Sub87	Y	5						X					0

Patient #2068

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	Y										X		0
Exp	N			1									
TrDx	N				X								7
Sub17	N				X								7
Sub87	Y	2.25						X		X	x		0

Patient #2069

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	N				X								2
Exp	Y	1.5		1	2	3							
TrDx	N			X									3
Sub17	N								X				0
Sub87	Y	1.25			X								8

Patient #2070

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	Y										X		0
Exp	N			1									
TrDx	N			X									10
Sub17	N			X									10

Patient #2071

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	N			X								X	7
Exp	N								1				
TrDx	-			-									0
Sub18	N								X				10

Patient #2072

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	N				X								7
Exp	N			1									
TrDx	N				X								7
Sub18	N								X				7

Patient #2073

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	N		X										0
Exp	Y	4.25		3			1		2			X	
TrDx	Y	4.25					X					X	10
Sub18	Y	4.25					X						9
Sub27	Y	4.25					X						9

Patient #2074

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	Y					X							7
Exp	Y	2.83			2	1			3	X	X		
TrDx	N		X										0
Sub27	Y	2.83								X	x		8

Patient #2075

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	N*			X									10
Exp	N			1									
TrDx	N			X									10
Sub27	N			X									10

Patient #2077

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	Y*												6
Exp	Y	2.33		3	2	1							
TrDx	N			X									1
Sub20	Y					X						X	8

Patient #2078

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	N				X								7
Exp	N			1									
TrDx	Y	2							X				0
Sub20	Y	2							X	X	x		0

Patient #2080

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	N			X									2
Exp	Y	4.75		1		3			2	X	X		
TrDx	N			X								X	2
Sub20	Y	3.75			X								6

Patient #2081

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	N			X								X	2
Exp	Y	7.83	1	2			3					X	
TrDx	N		X									X	3
Sub75	Y	7.83	X										9

Patient #2082

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	N*			X									7
Exp	N								1				
TrDx	Y	3.6					X						0
Sub1	N			X									7
Sub20	Y						X						0
Sub75	Y	3.6					X						0

Patient #2083

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	N				X								0
Exp	Y	0.75				1		3	2	x	X		
TrDx	N			X									0
Sub1	Y	1.25				X				X	X		9
Sub20	Y	1.5				X							7
Sub75	Y	1				X							8

Patient #2084

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	N				X								0
Exp	Y	3				3		2	1				
TrDx	Y	3						X					9
Sub1	N			X									0
Sub21	Y	3				x			X	X	X		10
Sub75	N				X								0

Patient #2085

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	Y										X	X	8
Exp	Y	11.9			1	2			3	X	X		
TrDx	Y	11.9				X				X	x	X	9
Sub21	Y					x			X	X	X		7
Sub75	Y	?		?		x				x	X		7

Patient #2086

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	Y					X			X				9
Exp	Y	0.88				1		2		x	X		
TrDx	N			X								X	0
Sub22	Y	0.88				x			X	x	X		10

Patient #2087

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	Y*								X				0
Exp	N			1		2		3					
TrDx	N			X									10
Sub22	N								X				7

Patient #2088

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	N			X									1
Exp	Y	0.67		2		3			1		X		
TrDx	Y	1.5							X				7
Sub23	N				X	x				x	X		0

Patient #2089

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	Y					X			X				6
Exp	Y	8.75		2	1	3							
TrDx	Y	7.75				X					x	X	7
Sub23	Y					x		X		X	X		6

Patient #2090

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	N				X								7
Exp	N			1									
TrDx	N				X								7
Sub24	N			X									10

Patient #2091

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	N				X								0
Exp	Y	1	2				1						
TrDx	Y	0.12							X				5
Sub24	Y								X	X	x		5

Patient #2093

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	N				X								7
Exp	N		2	1			3						
TrDx	Y	14								X	x		0
Sub13	N			-								X	7

Patient #2094

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	N		X										9
Exp	N		2	1									
TrDx	N		X										9
Sub14	N			X									10
Sub58	N								X			X	7

Patient #2095

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	N*			X									2
Exp	Y	8.25		1								X	
TrDx	N			X								X	3
Sub14	Y	2.42							X			X	6
Sub58	Y	7.33							X			X	7

Patient #2096

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	N		X									X	1
Exp	Y	6.33		1								X	
TrDx	Y	3.5							X			X	6
Sub15	Y	7.5		?								X	6
Sub27	Y								X			X	6
Sub59	N			X								X	3
Sub81	Y	5.5							X			X	6

Patient #2097

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	N				X								2
Exp	Y	5.42		3	1	2				X	X		
TrDx	Y	0.27							X				5
Sub15	Y	5.42			X			X					9
Sub27	Y	0.67							X				5
Sub59	Y	5.75				X					x		8
Sub81	Y	0.42				X					x		7

Patient #2098

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	N*			X									1
Exp	Y	4	2	3			1						
TrDx	N			X								X	1
Sub16	N			X								X	1
Sub81	N								X			X	0
Sub90	N			X									1

Patient #2099

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	Y						X						10
Exp	Y	8	2	3			1						
TrDx	Y	5.7								X	x	X	5
Sub16	Y	?					X						8
Sub81	Y	8					X						10
Sub90	N								X				0

Patient #2100

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	N				X								0
Exp	Y	2.67		1		2				X	X		
TrDx	N				X								0
Sub17	N								X				0
Sub81	N				X								0

Patient #2101

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	N				X								0
Exp	Y	2		1		2				X	X		
TrDx	N				X								0
Sub17	N								X				0

Patient #2102

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	N				X								7
Exp	N			1									
TrDx	N				X								7
Sub18	N			X									10

Patient #2103

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	Y*				X	X		X	X				7
Exp	Y	2				1		2		x	X		
TrDx	Y	0.50							X				5
Sub18	Y					X				X	X		10

Patient #2104

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	N*			X								X	3
Exp	Y	2.83		1									
TrDx	N		X										0
Sub90	N			X									3

Patient #2105

Sub:	Refer (y/n)	Ref Age	Early Pub	Norm	Cons Del.	Cong GH	Prec Pub	Short Bone	No. Info	Acq. GH	Hyp Thyr	Obesity	Score
MR	N*			X									10
Exp	N			1									
TrDx	N			X									10
Sub90	N			X									10

* means case was removed from trial

Sub 001 - Fellow

Pat #	Score	TrenDx Score
2000	3	8
2001	0	6
2025	5	5
2026	*	*
2027	10	0
2053	1	10
2054	1	1
2082	7	0
2083	9	0
2084	0	9
Avg	4.0	4.3

Sub 013 Attending Academic

Pat #	Score	TrenDx Score
2001	0	6
2002	7	0
2003	7	0
2033	10	0
2034	7	2
2054	1	1
2056	1	0
2057	2	2
2092	*	*
2093	7	0
Avg	4.7	1.2

Sub 014 Attending Academic

Pat #	Score	TrenDx Score
2004	8	5
2005	0	0
2006	7	7
2035	5	5
2036	0	0
2058	5	5
2059	5	5
2060	6	6
2094	10	9
2095	6	3
Avg	5.2	4.5

Sub 015 Attending GP

Pat #	Score	TrenDx Score
2007	*	*
2008	8	8
2009	7	6
2037	7	5
2038	*	*
2061	9	8
2062	8	10
2063	8	0
2096	6	6
2097	9	5
Avg	7.8	6.0

Sub 016 Attending GP

Pat #	Score	TrenDx Score
2010	7	5
2011	10	0
2012	10	5
2039	5	5
2040	0	0
2065	10	10
2066	10	9
2067	7	0
2098	1	1
2099	8	5
Avg	6.8	4.0

Sub 017 Attending GP

Pat #	Score	TrenDx Score
2013	0	8
2014	10	0
2015	5	5
2041	0	0
2042	7	7
2068	7	7
2069	0	3
2070	10	10
2100	0	0
2101	0	0
Avg	3.9	4.0

Sub 018 Attending, Academic

Pat #	Score	TrenDx Score
2016	7	5
2017	8	0
2018	0	7
2043	7	0
2044	5	6
2071	10	0
2072	7	7
2073	9	10
2102	10	7
2103	10	5
Avg	7.3	4.7

Sub 020 Fellow

Pat #	Score	TrenDx Score
2022	0	0
2023	5	8
2024	7	0
2049	10	0
2050	0	0
2077	8	1
2078	0	0
2080	6	2
2082	0	0
2083	7	0
Avg	4.3	1.1

Sub 021 Attending, Academic

Pat #	Score	TrenDx Score
2000	10	8
2001	6	6
2002	7	0
2051	0	0
2052	*	*
2053	10	10
2054	1	1
2056	3	0
2084	10	9
2085	7	9
Avg	6.0	4.8

Sub 022 Fellow

Pat #	Score	TrenDx Score
2003	7	0
2004	5	5
2005	1	0
2025	3	5
2026	*	*
2057	2	2
2058	6	5
2059	2	5
2086	10	0
2087	7	10
Avg	4.8	3.6

Sub 023 Fellow

Pat #	Score	TrenDx Score
2006	7	7
2007	*	*
2008	8	8
2027	10	0
2028	8	6
2060	0	6
2061	0	8
2062	8	10
2088	0	7
2089	6	7
Avg	5.2	6.6

Sub 024 RN

Pat #	Score	TrenDx Score
2009	0	6
2010	7	5
2011	10	0
2029	0	10
2030	6	5
2063	5	0
2065	7	10
2066	10	9
2090	10	7
2091	5	5
Avg	6.0	5.7

Sub 027 Resident

Pat #	Score	TrenDx Score
2018	7	7
2019	0	1
2020	7	5
2035	3	5
2036	0	0
2073	9	10
2074	8	0
2075	10	10
2096	6	6
2097	5	5
Avg	5.5	4.9

Sub 058 Attending Academic

Pat #	Score	TrenDx Score
2001	6	6
2002	7	0
2003	0	0
2047	10	9
2049	10	0
2050	5	0
2063	8	0
2065	10	10
2094	7	9
2095	7	3
Avg	7.0	3.7

Sub 059 Attending,GP

Pat #	Score	TrenDx Score
2004	8	5
2005	0	0
2006	7	7
2025	5	5
2051	0	0
2052	*	*
2066	10	9
2067	0	0
2096	3	6
2097	8	5
Avg	4.6	4.1

Sub 060 Med Student

Pat #	Score	TrenDx Score
2000	1	8
2001	7	6
2002	0	0
2003	0	0
2004	5	5
2005	5	0
2006	7	7
2007	*	*
2008	8	8
2009	10	6
Avg	4.8	4.4

Sub 061 Med Student

Pat #	Score	TrenDx Score
2010	5	5
2011	10	0
2012	6	5
2013	0	8
2014	5	0
2015	5	5
2016	6	5
2017	8	0
2018	6	7
2019	0	1
Avg	5.1	3.6

Sub 065 Med Student

Pat #	Score	TrenDx Score
2053	9	10
2054	1	1
2056	8	0
2057	6	2
2058	6	5
2059	5	5
2060	5	6
2061	8	8
2062	8	10
2063	6	0
Avg	6.2	4.7

Sub 075 Fellow

Pat #	Score	TrenDx Score
2025	0	5
2026	*	*
2027	0	0
2028	6	8
2029	0	10
2081	9	3
2082	0	0
2083	8	0
2084	0	9
2085	7	9
Avg	3.3	4.9

Sub 081 Attending, Academic

Pat #	Score	TrenDx Score
2005	5	0
2006	7	7
2007	*	*
2008	1	8
2009	5	6
2096	6	6
2097	7	5
2098	1	1
2099	10	5
2100	0	0
Avg	4.7	4.2

Sub 087 Attending, Academic

Pat #	Score	TrenDx Score
2035	0	5
2036	5	0
2037	8	5
2038	*	*
2039	5	5
2065	10	10
2066	3	9
2067	0	0
2068	0	7
2069	8	3
Avg	4.3	4.9

Sub 089 Attending, Academic

Pat #	Score	TrenDx Score
2046	10	10
2047	9	9
2049	10	0
2050	0	0
2051	10	0
2053	9	10
2054	1	1
2056	2	0
2057	6	2
2058	6	5
Avg	6.3	3.7

Sub 090, Fellow

Pat #	Score	TrenDx Score
2021	0	10
2031	10	0
2032	6	7
2039	0	5
2046	8	10
2076	*	*
2098	1	1
2099	0	5
2104	3	0
2105	10	10
Avg	4.2	4.8

9. Appendix B - Packet Directions and Sample Chart

Direction Sheet

Within this packet, you will find patient growth charts for several patients. On each patient chart, you are asked to determine whether the patient should be referred to an endocrine clinic, give a preliminary diagnosis, and determine the time at which the patient should have been referred, if at all.

1. If you feel that the growth data suggests a growth abnormality, please indicate that you approve a referral to the endocrine clinic by marking an 'X' in the space next to that choice.

2. Then, whether or not you chose to refer the child, choose **exactly one congenital condition** and **any number of acquired conditions** that you feel best describe the child, from the data that is available.

3. Finally, if you chose to refer the child, **circle the age in the data table** after which you believe that it would have been appropriate to refer the child, only knowing the data that was recorded up to that age.

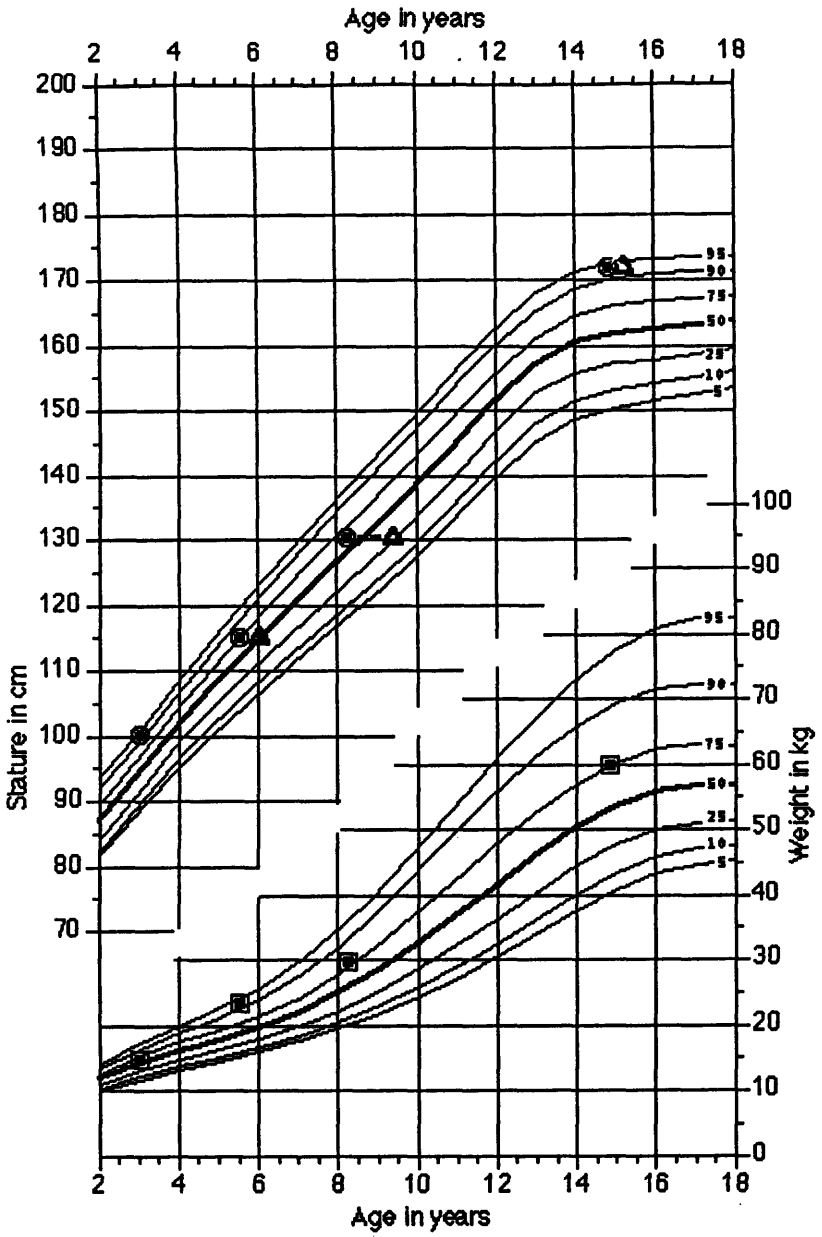
Each growth chart has the data plotted in graphical form, as well as listed in a table alongside the graph. The charts are all identified by the patient ID#, listed in the upper right hand corner of the chart. Patients that have both infant and adult data points have both infant and adult growth charts stapled together. For these cases, please mark your answers on the adult chart, except for step #3 above, which can be done on either chart.

Please give the same consideration to these cases as you would give to real cases in which you were asked for a second opinion. However, we must ask you not to ask colleagues for their opinions and not to use any reference materials. In addition, please do not spend more time thinking about a particular case than you think that you would be able to spend in a clinical setting.

Finally, although these cases have no identifying information that relates them back to the actual medical records at the Children's Hospital, we must ask you **not to discuss the cases with others** and to minimize exposure of the charts to other people.

Within this packet, there is a sample patient case with patient ID - SAMPLECASE. It has already been marked to demonstrate how to correctly fill out a chart. NOTE: The diagnosis marked on the sample chart is not necessarily correct.

SAMPLE CASE



Female Adult NCHS Chart

Patient ID: SAMPLECASE

Age	Height in cm	Weight in kg	Bone Age
3	100.3	15	
5.5	115	23.5	6
8.2	130.28	30	9.4
14.82	172	60	15.25

Legend
- Height
- Weight
- Bone Age

1. Based on the data presented, would you recommend:

✓ Approve referral to endocrine clinic ___ Deny referral to endocrine clinic

2. Please place a checkmark next to exactly one congenital condition and any number of acquired conditions that you feel best describe the patient.

Congenital Conditions		Acquired Conditions
___ Normal Growth	✓ Precocious Puberty	___ Acquired Growth Hormone Deficiency
___ Early Puberty	___ Short Bone Syndrome / Turner's Syndrome / Hypochondroplasia	___ Hypothyroidism
___ Constitutional Delay		✓ Obesity
___ Congenital Growth Hormone Deficiency		

3. Please circle the age at which you feel the patient should be referred

10. Appendix C - Trend Template LISP Code

```
(in-package :tt)


(deftt "boy-average-growth"
  :landmarks '(birth growth-stops)
  :intervals (list (definterval :name "early-childhood"
                    :constraints
                    (list (defconstraint :name "constant build"
                           :func #'bd-infant-normal
                           :parameters '(weight height)
                           :model '(constant *)
                           :weight 1)
                          (defconstraint :name "height weight co-vary"
                            :func #'wt-minus-ht-infant-zscores
                            :parameters '(weight height)
                            :model '(constant *)
                            :weight 1)))
                   (definterval :name "childhood-to-adulthood"
                    :constraints
                    (list (defconstraint :name "constant build"
                           :func #'bd-child-normal
                           :parameters '(weight height)
                           :model '(constant 1)
                           :weight 1)
                          (defconstraint :name "Avg SD Constant"
                           :func #'avg-sd
                           :parameters '(height)
                           :model '(constant *)
                           :weight 3)
                          (defconstraint :name "Avg SD between +-2"
                           :func #'(lambda (height-datum)
                                    (let ((ht-sd (avg-sd height-datum)))
                                      (error-fourth-then-constant ht-sd '(-2 2))))
                           :parameters '(height)
                           :model '(constant 0)
                           :weight 1)
                          (defconstraint :name "Avg ht-vel SD between +-2"
                           :func #'(lambda (height-vel-datum)
                                    (let ((ht-vel-sd (avg-sd height-vel-datum)))
                                      (error-fourth-then-constant ht-vel-sd '(-2 2))))
                           :parameters '(ht-vel)
                           :model '(constant 0)
                           :weight 1)
                          (defconstraint :name "Chron Age - Bone Age = +- 1"
                           :func #'(lambda (bone-age-datum)
                                    (let ((bone-age-delta
                                           (chron-age-minus-bone-age bone-age-datum)))
                                      (error-fourth-then-constant bone-age-delta '(-1 1))))
```

```
                    :parameters '(bone-age)
                    :model '(constant 0)
                    :weight 5)
                  ))
            (definterval :name "linear-growth-ends"
              ;; height is constant for the next 10 years.
              :constraints
              (list (defconstraint :name "height-constant"
                      :parameters '(height)
                      :model '(constant *)
                      :weight 5)
                    (defconstraint :name "Bone Age = Chron Age"
                      :parameters '(bone-age)
                      :func #'chron-age-minus-bone-age
                      :model '(constant 0)
                      :weight 5)))
            ) ;; close off list of intervals

 :relations  '(
          ((begin early-childhood) birth 0 0)
          ((begin early-childhood) (end early-childhood) (years 2.05) (years 2.95))
          (consecutive-phase early-childhood childhood-to-adulthood)
          ((end childhood-to-adulthood) growth-stops 0 0)
          (birth growth-stops (years 16.5) (years 18.5))
          (consecutive-phase childhood-to-adulthood linear-growth-ends)
          ((begin linear-growth-ends) (end linear-growth-ends) (years 10) (years 10))
          ))



(deftt "boy-constitutional-delay"
 :landmarks '(birth growth-stops)
 :intervals (list (definterval :name "early-childhood"
              :constraints
              (list (defconstraint :name "constant build"
                    :func #'bd-infant-normal
                    :parameters '(weight height)
                    :model '(constant *)
                    :weight 1)
              (defconstraint :name "height weight co-vary"
                      :func #'wt-minus-ht-infant-zscores
                      :parameters '(weight height)
                      :model '(constant *)
                      :weight 1)))
            (definterval :name "childhood-to-adulthood"
              :constraints
              (list (defconstraint :name "constant build"
                    :func #'bd-child-normal
                    :parameters '(weight height)
                    :model '(constant 1)
                    :weight 1)
                  (defconstraint :name "Late SD Constant"
                    :func #'late-sd
                    :parameters '(height)
```

```
                :model '(constant *)
                :weight 3)
               (defconstraint :name "Late SD between -2,2"
                :func #'(lambda (height-datum)
                        (let ((ht-sd (late-sd height-datum)))
                          (error-fourth-then-constant ht-sd '(-2 2))))
                :parameters '(height)
                :model '(constant 0)
                :weight 1)
               (defconstraint :name "Late ht-vel SD between -2,2"
                :func #'(lambda (height-vel-datum)
                        (let ((ht-vel-sd (late-sd height-vel-datum)))
                          (error-fourth-then-constant ht-vel-sd '(-2 2))))
                :parameters '(ht-vel)
                :model '(constant 0)
                :weight 1)
               (defconstraint :name "Chron Age - BA = 0.5 to 2.5"
                :func #'(lambda (bone-age-datum)
                        (let ((bone-age-delta
                               (chron-age-minus-bone-age bone-age-datum)))
                          (error-fourth-then-constant bone-age-delta '(0.5 2.5))))
                :parameters '(bone-age)
                :model '(constant 0)
                :weight 5)
               ))

          (definterval :name "linear-growth-ends"
            ;; height is constant for the next 10 years.
            :constraints
            (list (defconstraint :name "height-constant"
                   :parameters '(height)
                   :model '(constant *)
                   :weight 5)
                  (defconstraint :name "Bone Age = Chron Age"
                   :parameters '(bone-age)
                   :func #'chron-age-minus-bone-age
                   :model '(constant 0)
                   :weight 5)))

          ) ;; close off list of intervals

 :relations '(
        ((begin early-childhood) birth 0 0)
        ((begin early-childhood) (end early-childhood) (years 2.05) (years 2.95))
        (consecutive-phase early-childhood childhood-to-adulthood)
        ((end childhood-to-adulthood) growth-stops 0 0)
        (birth growth-stops (years 18) (years 20))
        (consecutive-phase childhood-to-adulthood linear-growth-ends)
        ((begin linear-growth-ends) (end linear-growth-ends) (years 10) (years 10))
))


(deftt "boy-early-puberty"
 :landmarks '(birth growth-stops)
```

```
:intervals (list (definterval :name "early-childhood"
                  :constraints
                  (list (defconstraint :name "Constant build"
                         :func #'bd-infant-normal
                         :parameters '(weight height)
                         :model '(constant *)
                         :weight 1)
                   (defconstraint :name "height weight co-vary"
                          :func #'wt-minus-ht-infant-zscores
                          :parameters '(weight height)
                          :model '(constant *)
                          :weight 1)))
                 (definterval :name "childhood-to-adulthood"
                  :constraints
                  (list (defconstraint :name "Constant build"
                         :func #'bd-child-normal
                         :parameters '(weight height)
                         :model '(constant 1)
                         :weight 1)
                        (defconstraint :name "Early sd constant"
                         :func #'early-sd
                         :parameters '(height)
                         :model '(constant *)
                         :weight 3)
                        (defconstraint :name "Early SD between -2,2"
                         :func #'(lambda (height-datum)
                                  (let* ((ht-sd (early-sd height-datum)))
                                    (error-fourth-then-constant ht-sd '(-2 2))))
                         :parameters '(height)
                         :model '(constant 0)
                         :weight 1)
                        (defconstraint :name "Early ht-vel SD between -2,2"
                         :func #'(lambda (height-vel-datum)
                                  (let ((ht-vel-sd (early-sd height-vel-datum)))
                                    (error-fourth-then-constant ht-vel-sd '(-2 2))))
                         :parameters '(ht-vel)
                         :model '(constant 0)
                         :weight 1)
                        (defconstraint :name "Chron Age - BA = -2.5 to -0.5"
                         :func #'(lambda (bone-age-datum)
                                  (let* ((bone-age-delta
                                          (chron-age-minus-bone-age bone-age-datum)))
                                    (error-fourth-then-constant bone-age-delta '(-2.5 -0.5))))
                         :parameters '(bone-age)
                         :model '(constant 0)
                         :weight 5)))

                 (definterval :name "linear-growth-ends"
                  ;; height is constant for the next 10 years.
                  :constraints
                  (list (defconstraint :name "height-constant"
                         :parameters '(height)
                         :model '(constant *)
                         :weight 5)
```

```
                (defconstraint :name "Bone Age = Chron Age"
                  :parameters '(bone-age)
                  :func #'chron-age-minus-bone-age
                  :model '(constant 0)
                  :weight 5)))

            ) ;; close off list of intervals

 :relations  '(
          ((begin early-childhood) birth 0 0)
          ((begin early-childhood) (end early-childhood) (years 2.05) (years 2.95))
          (consecutive-phase early-childhood childhood-to-adulthood)
          ((end childhood-to-adulthood) growth-stops 0 0)
          (birth growth-stops (years 15) (years 17))
          (consecutive-phase childhood-to-adulthood linear-growth-ends)
          ((begin linear-growth-ends) (end linear-growth-ends) (years 10) (years 10))
          ))


(deftt "boy-congenital-gh-def"
 :landmarks '(birth growth-stops)
 :intervals (list (definterval :name "early-childhood"
              :constraints
              (list (defconstraint :name "constant build"
                    :func #'bd-infant-normal
                    :parameters '(weight height)
                    :model '(constant *)
                    :weight 1)
             (defconstraint :name "height weight co-vary"
                     :func #'wt-minus-ht-infant-zscores
                     :parameters '(weight height)
                     :model '(constant *)
                     :weight 1)))
            ;; sd should be falling...
            (definterval :name "childhood-to-adulthood"
              :constraints
              (list (defconstraint :name "increasing build"
                    :func #'bd-child-normal
                    :parameters '(weight height)
                    :model '(linear (D1 +))
                    :weight 1)
                  (defconstraint :name "Late SD Falling"
                    :func #'late-sd
                    :parameters '(height)
                    :model '(linear (D1 -))
                    :weight 3)
                  (defconstraint :name "Late SD between -4 and 0"
                    :func #'(lambda (height-datum)
                          (let ((ht-sd (late-sd height-datum)))
                            (error-fourth-then-constant ht-sd '(-4 0))))
                    :parameters '(height)
                    :model '(constant 0)
                    :weight 1)
                  (defconstraint :name "Late ht-vel Zscore between -4 to 0"
```

```
                :func #'(lambda (height-vel-datum)
                        (let ((ht-vel-sd (late-sd height-vel-datum)))
                          (error-fourth-then-constant ht-vel-sd '(-4 0))))
                :parameters '(ht-vel)
                :model '(constant 0)
                :weight 1)
               (defconstraint :name "Chron Age - BA = 2 to 4"
                :func #'(lambda (bone-age-datum)
                        (let* ((bone-age-delta
                                (chron-age-minus-bone-age bone-age-datum)))
                          (error-fourth-then-constant bone-age-delta '(2 4))))
                :parameters '(bone-age)
                :model '(constant 0)
                :weight 5)
               (defconstraint :name "Chron Age - BA Increasing"
                :func #'chron-age-minus-bone-age
                :parameters '(bone-age)
                :model '(linear (D1 +))
                :weight 5)
               ))
          (definterval :name "linear-growth-ends"
           ;; height is constant for the next 10 years.
            :constraints
           (list (defconstraint :name "height-constant"
                  :parameters '(height)
                  :model '(constant *)
                  :weight 5)
                 (defconstraint :name "Bone Age = Chron Age"
                  :parameters '(bone-age)
                  :func #'chron-age-minus-bone-age
                  :model '(constant 0)
                  :weight 5)))

          ) ;; close off list of intervals

:relations '(
        ((begin early-childhood) birth 0 0)
        ((begin early-childhood) (end early-childhood) (years 2.05) (years 2.95))
        (consecutive-phase early-childhood childhood-to-adulthood)
```

```
        ((end childhood-to-adulthood) growth-stops 0 0)
        (birth growth-stops (years 18) (years 22))
        (consecutive-phase childhood-to-adulthood linear-growth-ends)
        ((begin linear-growth-ends) (end linear-growth-ends) (years 10) (years 10))
        ))


(deftt "boy-short-bone-syndrome"
 :landmarks '(birth growth-stops)
 :intervals (list (definterval :name "early-childhood"
             :constraints
             (list (defconstraint :name "constant build"
                   :func #'bd-infant-normal
                   :parameters '(weight height)
                   :model '(constant *)
                   :weight 1)
              (defconstraint :name "height weight co-vary"
                    :func #'wt-minus-ht-infant-zscores
                    :parameters '(weight height)
                    :model '(constant *)
                    :weight 1)))
            (definterval :name "childhood-to-adulthood"
             :constraints
             (list (defconstraint :name "increasing build"
                   :func #'bd-child-normal
                   :parameters '(weight height)
                   :model '(linear (D1 +))
                   :weight 1)
                (defconstraint :name "Late SD Falling"
                   :func #'late-sd
                   :parameters '(height)
                   :model '(linear (D1 -))
                   :weight 3)
                (defconstraint :name "Late SD between -5 and -1"
                   :func #'(lambda (height-datum)
                         (let ((ht-sd (late-sd height-datum)))
```

```
            (error-fourth-then-constant ht-sd '(-5 -1))))
      :parameters '(height)
      :model '(constant 0)
      :weight 1)
     (defconstraint :name "Late ht-vel Zscore between -5 to -1"
      :func #'(lambda (height-vel-datum)
            (let ((ht-vel-sd (late-sd height-vel-datum)))
              (error-fourth-then-constant ht-vel-sd '(-5 -1))))
      :parameters '(ht-vel)
      :model '(constant 0)
      :weight 1)
     (defconstraint :name "Chron Age - BA +-1"
      :func #'(lambda (bone-age-datum)
            (let* ((bone-age-delta
                 (chron-age-minus-bone-age bone-age-datum)))
              (error-fourth-then-constant bone-age-delta '(-1 1))))
      :parameters '(bone-age)
      :model '(constant 0)
      :weight 5)
     ;; this is primary disturbance of growth, no BA delay.
     ))
(definterval :name "linear-growth-ends"
 ;; height is constant for the next 10 years.
 :constraints
 (list (defconstraint :name "height-constant"
        :parameters '(height)
        :model '(constant *)
        :weight 5)
       (defconstraint :name "Bone Age = Chron Age"
        :parameters '(bone-age)
        :func #'chron-age-minus-bone-age
        :model '(constant 0)
        :weight 5)))

) ;; close off list of intervals
```

```
 :relations  '(
          ((begin early-childhood) birth 0 0)
          ((begin early-childhood) (end early-childhood) (years 2.05) (years 2.95))
          (consecutive-phase early-childhood childhood-to-adulthood)
          ((end childhood-to-adulthood) growth-stops 0 0)
          (birth growth-stops (years 15.5) (years 18.5))
          (consecutive-phase childhood-to-adulthood linear-growth-ends)
          ((begin linear-growth-ends) (end linear-growth-ends) (years 10) (years 10))
          ))


(deftt "boy-acquired-gh-def"
  :landmarks '(birth now growth-stops)
  :intervals (list (definterval :name "early-childhood"
                :constraints
                (list (defconstraint :name "constant build"
                     :func #'bd-infant-normal
                     :parameters '(weight height)
                     :model '(constant *)
                     :weight 1)
                (defconstraint :name "height weight co-vary"
                       :func #'wt-minus-ht-infant-zscores
                       :parameters '(weight height)
                       :model '(constant *)
                       :weight 1)))
               (definterval :name "childhood-to-onset"
                :constraints
                (list (defconstraint :name "constant build"
                      :func #'bd-child-normal
                      :parameters '(weight height)
                      :model '(constant 1)
                      :weight 1)
                    (defconstraint :name "Avg SD Constant"
                      :func #'avg-sd
                      :parameters '(height)
                      :model '(constant *)
```

```
    :weight 1)
   (defconstraint :name "Avg SD between +-2"
    :func #'(lambda (height-datum)
          (let ((ht-sd (avg-sd height-datum)))
           (error-fourth-then-constant ht-sd '(-2 2))))
    :parameters '(height)
    :model '(constant 0)
    :weight 1)
   (defconstraint :name "Avg ht-vel SD between +-2"
    :func #'(lambda (height-vel-datum)
          (let ((ht-vel-sd (avg-sd height-vel-datum)))
           (error-fourth-then-constant ht-vel-sd '(-2 2))))
    :parameters '(ht-vel)
    :model '(constant 0)
    :weight 1)
   (defconstraint :name "Chron Age - Bone Age = +- 1"
    :func #'(lambda (bone-age-datum)
          (let ((bone-age-delta
              (chron-age-minus-bone-age bone-age-datum)))
           (error-fourth-then-constant bone-age-delta '(-1 1))))
    :parameters '(bone-age)
    :model '(constant 0)
    :weight 5)
   ))
(definterval :name "onset-to-now"
 :constraints
 (list (defconstraint :name "Build increasing"
    :func #'bd-child-normal
    :parameters '(weight height)
    :model '(linear (D1 +.05 per-year))
    :weight 1)
   (defconstraint :name "Late SD dropping"
    :func #'late-sd
    :parameters '(height)
    :model '(linear (D1 -.6 per-year))
    :weight 3)
```

```
            (defconstraint :name "Late SD between -4 0, dropping (constant)"
              :func #'(lambda (height-datum)
                        (let ((ht-sd (late-sd height-datum)))
                          (error-fourth-then-constant ht-sd '(-4 0))))
              :parameters '(height)
              :model '(constant 0)
              :weight 1)
            (defconstraint :name "Late ht-vel SD between -4 0, dropping (constant) "
              :func #'(lambda (height-vel-datum)
                        (let ((ht-vel-sd (late-sd height-vel-datum)))
                          (error-fourth-then-constant ht-vel-sd '(-4 0))))
              :parameters '(ht-vel)
              :model '(constant 0)
              :weight 1)
            (defconstraint :name "BA-CA increasing"
              :func #'chron-age-minus-bone-age
              :parameters '(bone-age)
              :model '(linear (D1 +))
              :weight 1)))
        (definterval :name "linear-growth-ends"
          ;; height is constant for the next 10 years.
          :constraints
          (list (defconstraint :name "height-constant"
                  :parameters '(height)
                  :model '(constant *)
                  :weight 5)
                (defconstraint :name "Bone Age = Chron Age"
                  :parameters '(bone-age)
                  :func #'chron-age-minus-bone-age
                  :model '(constant 0)
                  :weight 5)))
        ) ;; close off list of intervals

:relations '(
        ((begin early-childhood) birth 0 0)
        ((begin early-childhood) (end early-childhood) (years 2.05) (years 2.95))
```

```
(consecutive-phase early-childhood childhood-to-onset)
(consecutive-phase childhood-to-onset onset-to-now)
((begin onset-to-now) now (years 2) (years 3))
((end onset-to-now) now 0 0)
(growth-stops (begin linear-growth-ends) 0 0)
((begin linear-growth-ends) (end linear-growth-ends) (years 10) (years 10))
(birth growth-stops (years 17) (years 21))
))


(deftt "boy-precocious-pub"
 :landmarks '(birth growth-stops)
 :intervals (list (definterval :name "early-childhood"
                  :constraints
                  (list (defconstraint :name "constant build"
                        :func #'bd-infant-normal
                        :parameters '(weight height)
                        :model '(constant *)
                        :weight 1)
                  (defconstraint :name "height weight co-vary"
                         :func #'wt-minus-ht-infant-zscores
                         :parameters '(weight height)
                         :model '(constant *)
                         :weight 1)))
                 (definterval :name "childhood-to-adulthood"
                  :constraints
                  (list (defconstraint :name "increasing build"
                        :func #'bd-child-normal
                        :parameters '(weight height)
                        :model '(linear (D1 +))
                        :weight 1)
                       (defconstraint :name "Ht SD Rising .2"
                        :func #'child-nchs-sd
                        :parameters '(height)
                        :model '(linear (D1 0.2 per-year))
                        :weight 3)
```

```
            (defconstraint :name "Early SD between (0 4)"
              :func #'(lambda (height-datum)
                    (let* ((ht-sd (early-sd height-datum)))
                         (error-fourth-then-constant ht-sd '(0 4))))
              :parameters '(height)
              :model '(constant *)
              :weight 1)
            (defconstraint :name "Chron Age - BA = -4 to -2"
              :func #'(lambda (bone-age-datum)
                    (let* ((bone-age-delta
                         (chron-age-minus-bone-age bone-age-datum)))
                        (error-fourth-then-constant bone-age-delta '(-4 -2))))
              :parameters '(bone-age)
              :model '(constant 0)
              :weight 5)))
        (definterval :name "linear-growth-ends"
          ;; height is constant for the next 10 years.
          :constraints
          (list (defconstraint :name "height-constant"
                  :parameters '(height)
                  :model '(constant *)
                  :weight 5)
                (defconstraint :name "Bone Age = Chron Age"
                  :parameters '(bone-age)
                  :func #'chron-age-minus-bone-age
                  :model '(constant 0)
                  :weight 5)))

        ) ;; close off list of intervals

:relations  '(
        ((begin early-childhood) birth 0 0)
        ((begin early-childhood) (end early-childhood) (years 2.05) (years 2.95))
        (consecutive-phase early-childhood childhood-to-adulthood)
        ((end childhood-to-adulthood) growth-stops 0 0)
        (birth growth-stops (years 13) (years 17))
```

```
        (consecutive-phase childhood-to-adulthood linear-growth-ends)
        ((begin linear-growth-ends) (end linear-growth-ends) (years 10) (years 10))
        ))

(deftt "boy-normal-build"
 :landmarks '(birth now)
 :intervals (list (definterval :name "Normal"
              :constraints
              (list (defconstraint :name "constant build"
                   :func #'build-normal
                   :parameters '(weight height)
                   :model '(constant 1)
                   :weight 1))))
 :relations '(
        (birth (begin Normal) 0 (years 18))
        ((begin Normal) (end Normal) (months 1) (months 1))
        ((end Normal) now 0 0)))



(deftt "boy-obese"
 :landmarks '(birth now)
 :intervals (list
        (definterval :name "Obese"
         :constraints
         (list (defconstraint :name "high build"
              :func #'build-normal
              :parameters '(weight height)
              :model '(constant 1.3)
              :weight 1))))
 :relations '(
        (birth (begin Obese) 0 (years 18))
        ((begin Obese) now (months 1) (months 1))
        ((end Obese) now 0 0)))
```

```
(deftt "boy-malnurished"
 :landmarks '(birth now)
 :intervals (list
          (definterval :name "Malnurished"
           :constraints
           (list (defconstraint :name "low build"
                 :func #'build-normal
                 :parameters '(weight height)
                 :model '(constant .8)
                 :weight 1))))
 :relations '(
          (birth (begin Malnurished) 0 (years 18))
          ((begin Malnurished) now (months 1) (months 1))
          ((end Malnurished) now 0 0)))



(make-instance 'monitor-set
 :name "boy-growth"
 :tts (list
     "boy-average-growth"
     "boy-constitutional-delay"
     "boy-early-puberty"
     "boy-congenital-gh-def"
     "boy-short-bone-syndrome"
     "boy-acquired-gh-def"
     "boy-precocious-pub"
     "boy-normal-build"
     "boy-obese"
     "boy-malnurished"
     )
 :grain 0
 :error-calc 'mean-res
 :weights 'explicit)
```

```
;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;
;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;
;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;
;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;
;; Females
;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;
;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;
;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;

(deftt "girl-average-growth"
  ;; Uses error-fourth-then-constant methods
  :landmarks '(birth growth-stops)
  :intervals (list (definterval :name "early-childhood"
                 :constraints
                 (list (defconstraint :name "constant build"
                        :func #'bd-infant-normal
                        :parameters '(weight height)
                        :model '(constant *)
                        :weight 1)
                 (defconstraint :name "height weight co-vary"
                          :func #'wt-minus-ht-infant-zscores
                          :parameters '(weight height)
                          :model '(constant *)
                          :weight 1)))
               (definterval :name "childhood-to-adulthood"
                 :constraints
                 (list (defconstraint :name "constant build"
                         :func #'bd-child-normal
                         :parameters '(weight height)
                         :model '(constant 1)
                         :weight 1)
                       (defconstraint :name "Avg SD Constant"
                         :func #'avg-sd
                         :parameters '(height)
```

```
      :model '(constant *)
      :weight 3)
     (defconstraint :name "Avg SD between +-2"
      :func #'(lambda (height-datum)
            (let ((ht-sd (avg-sd height-datum)))
              (error-fourth-then-constant ht-sd '(-2 2))))
      :parameters '(height)
      :model '(constant 0)
      :weight 1)
     (defconstraint :name "Avg ht-vel SD between +-2"
      :func #'(lambda (height-vel-datum)
            (let ((ht-vel-sd (avg-sd height-vel-datum)))
              (error-fourth-then-constant ht-vel-sd '(-2 2))))
      :parameters '(ht-vel)
      :model '(constant 0)
      :weight 1)
     (defconstraint :name "Chron Age - Bone Age = +- 1"
      :func #'(lambda (bone-age-datum)
            (let ((bone-age-delta
                (chron-age-minus-bone-age bone-age-datum)))
              (error-fourth-then-constant bone-age-delta '(-1 1))))
      :parameters '(bone-age)
      :model '(constant 0)
      :weight 5)
     ))
(definterval :name "linear-growth-ends"
  ;; height is constant for the next 10 years.
  :constraints
  (list (defconstraint :name "height-constant"
        :parameters '(height)
        :model '(constant *)
        :weight 5)
       (defconstraint :name "Bone Age = Chron Age"
        :parameters '(bone-age)
        :func #'chron-age-minus-bone-age
        :model '(constant 0)
```

```
                :weight 5)))


            ) ;; close off list of intervals

 :relations  '(
          ((begin early-childhood) birth 0 0)
          ((begin early-childhood) (end early-childhood) (years 2.05) (years 2.95))
          (consecutive-phase early-childhood childhood-to-adulthood)
          ((end childhood-to-adulthood) growth-stops 0 0)
          (birth growth-stops (years 14) (years 16))
          (consecutive-phase childhood-to-adulthood linear-growth-ends)
          ((begin linear-growth-ends) (end linear-growth-ends) (years 10) (years 10))
          ))


(deftt "girl-constitutional-delay"
 :landmarks '(birth growth-stops)
 :intervals (list (definterval :name "early-childhood"
              :constraints
              (list (defconstraint :name "constant build"
                   :func #'bd-infant-normal
                   :parameters '(weight height)
                   :model '(constant *)
                   :weight 1)
              (defconstraint :name "height weight co-vary"
                    :func #'wt-minus-ht-infant-zscores
                    :parameters '(weight height)
                    :model '(constant *)
                    :weight 1)))
             (definterval :name "childhood-to-adulthood"
              :constraints
              (list (defconstraint :name "constant build"
                   :func #'bd-child-normal
                   :parameters '(weight height)
                   :model '(constant 1)
```

```
        :weight 1)
      (defconstraint :name "Late SD Constant"
        :func #'late-sd
        :parameters '(height)
        :model '(constant *)
        :weight 3)
      (defconstraint :name "Late SD between -2,2"
        :func #'(lambda (height-datum)
                (let ((ht-sd (late-sd height-datum)))
                  (error-fourth-then-constant ht-sd '(-2 2))))
        :parameters '(height)
        :model '(constant 0)
        :weight 1)
      (defconstraint :name "Late ht-vel SD between -2,2"
        :func #'(lambda (height-vel-datum)
                (let ((ht-vel-sd (late-sd height-vel-datum)))
                  (error-fourth-then-constant ht-vel-sd '(-2 2))))
        :parameters '(ht-vel)
        :model '(constant 0)
        :weight 1)
      (defconstraint :name "Chron Age - BA = 0.5 to 2.5"
        :func #'(lambda (bone-age-datum)
                (let ((bone-age-delta
                      (chron-age-minus-bone-age bone-age-datum)))
                  (error-fourth-then-constant bone-age-delta '(0.5 2.5))))
        :parameters '(bone-age)
        :model '(constant 0)
        :weight 5)
        ))
(definterval :name "linear-growth-ends"
  ;; height is constant for the next 10 years.
  :constraints
  (list (defconstraint :name "height-constant"
          :parameters '(height)
          :model '(constant *)
          :weight 5)
```

```
                (defconstraint :name "Bone Age = Chron Age"
                  :parameters '(bone-age)
                  :func #'chron-age-minus-bone-age
                  :model '(constant 0)
                  :weight 5)))

            ) ;; close off list of intervals

 :relations  '(
          ((begin early-childhood) birth 0 0)
          ((begin early-childhood) (end early-childhood) (years 2.05) (years 2.95))
          (consecutive-phase early-childhood childhood-to-adulthood)
          ((end childhood-to-adulthood) growth-stops 0 0)
          (birth growth-stops (years 15) (years 17))
          (consecutive-phase childhood-to-adulthood linear-growth-ends)
          ((begin linear-growth-ends) (end linear-growth-ends) (years 10) (years 10))
))


(deftt "girl-early-puberty"
 :landmarks '(birth growth-stops)
 :intervals (list (definterval :name "early-childhood"
                :constraints
                (list (defconstraint :name "constant build"
                     :func #'bd-infant-normal
                     :parameters '(weight height)
                     :model '(constant *)
                     :weight 1)
                  (defconstraint :name "height weight co-vary"
                       :func #'wt-minus-ht-infant-zscores
                       :parameters '(weight height)
                       :model '(constant *)
                       :weight 1)))
               (definterval :name "childhood-to-adulthood"
                 :constraints
                 (list (defconstraint :name "constant build"
```

```
    :func #'bd-child-normal
    :parameters '(weight height)
    :model '(constant 1)
    :weight 1)
   (defconstraint :name "early sd constant"
    :func #'early-sd
    :parameters '(height)
    :model '(constant *)
    :weight 3)
   (defconstraint :name "Early SD between -2,2"
    :func #'(lambda (height-datum)
          (let* ((ht-sd (early-sd height-datum)))
            (error-fourth-then-constant ht-sd '(-2 2))))
    :parameters '(height)
    :model '(constant 0)
    :weight 1)
   (defconstraint :name "Early ht-vel SD between -2,2"
    :func #'(lambda (height-vel-datum)
          (let ((ht-vel-sd (early-sd height-vel-datum)))
            (error-fourth-then-constant ht-vel-sd '(-2 2))))
    :parameters '(ht-vel)
    :model '(constant 0)
    :weight 1)
   (defconstraint :name "Chron Age - BA = -2.5 to -0.5"
    :func #'(lambda (bone-age-datum)
          (let* ((bone-age-delta
               (chron-age-minus-bone-age bone-age-datum)))
            (error-fourth-then-constant bone-age-delta '(-2.5 -0.5))))
    :parameters '(bone-age)
    :model '(constant 0)
    :weight 5)))
(definterval :name "linear-growth-ends"
  ;; height is constant for the next 10 years.
  :constraints
  (list (defconstraint :name "height-constant"
       :parameters '(height)
```

```
                    :model '(constant *)
                    :weight 5)
                  (defconstraint :name "Bone Age = Chron Age"
                    :parameters '(bone-age)
                    :func #'chron-age-minus-bone-age
                    :model '(constant 0)
                    :weight 5)))

            ) ;; close off list of intervals

 :relations  '(
          ((begin early-childhood) birth 0 0)
          ((begin early-childhood) (end early-childhood) (years 2.05) (years 2.95))
          (consecutive-phase early-childhood childhood-to-adulthood)
          ((end childhood-to-adulthood) growth-stops 0 0)
          (birth growth-stops (years 13) (years 15))
          (consecutive-phase childhood-to-adulthood linear-growth-ends)
          ((begin linear-growth-ends) (end linear-growth-ends) (years 10) (years 10))
          ))


(deftt "girl-congenital-gh-def"
  :landmarks '(birth growth-stops)
  :intervals (list (definterval :name "early-childhood"
                :constraints
                (list (defconstraint :name "constant build"
                      :func #'bd-infant-normal
                      :parameters '(weight height)
                      :model '(constant *)
                      :weight 1)
                (defconstraint :name "height weight co-vary"
                      :func #'wt-minus-ht-infant-zscores
                      :parameters '(weight height)
                      :model '(constant *)
                      :weight 1)))
              (definterval :name "childhood-to-adulthood"
```

```
:constraints
(list (defconstraint :name "increasing build"
        :func #'bd-child-normal
        :parameters '(weight height)
        :model '(linear (D1 +))
        :weight 1)
      (defconstraint :name "Late SD Falling"
        :func #'late-sd
        :parameters '(height)
        :model '(linear (D1 -))
        :weight 3)
      (defconstraint :name "Late SD between -4 and 0"
        :func #'(lambda (height-datum)
                  (let ((ht-sd (late-sd height-datum)))
                    (error-fourth-then-constant ht-sd '(-4 0))))
        :parameters '(height)
        :model '(constant 0)
        :weight 1)
      (defconstraint :name "Late ht-vel Zscore between -4 to 0"
        :func #'(lambda (height-vel-datum)
                  (let ((ht-vel-sd (late-sd height-vel-datum)))
                    (error-fourth-then-constant ht-vel-sd '(-4 0))))
        :parameters '(ht-vel)
        :model '(constant 0)
        :weight 1)
      (defconstraint :name "Chron Age - BA = 2 to 4"
        :func #'(lambda (bone-age-datum)
                  (let* ((bone-age-delta
                          (chron-age-minus-bone-age bone-age-datum)))
                    (error-fourth-then-constant bone-age-delta '(1 4))))
        :parameters '(bone-age)
        :model '(constant 0)
        :weight 5)
      (defconstraint :name "Chron Age - BA Increasing"
        :func #'chron-age-minus-bone-age
        :parameters '(bone-age)
```

```
                :model '(linear (D1 +))
                :weight 5)
               ))

          (definterval :name "linear-growth-ends"
            ;; height is constant for the next 10 years.
            :constraints
            (list (defconstraint :name "height-constant"
                    :parameters '(height)
                    :model '(constant *)
                    :weight 5)
                  (defconstraint :name "Bone Age = Chron Age"
                    :parameters '(bone-age)
                    :func #'chron-age-minus-bone-age
                    :model '(constant 0)
                    :weight 5)))

          ) ;; close off list of intervals

 :relations  '(
          ((begin early-childhood) birth 0 0)
          ((begin early-childhood) (end early-childhood) (years 2.05) (years 2.95))
         (consecutive-phase early-childhood childhood-to-adulthood)
          ((end childhood-to-adulthood) growth-stops 0 0)
          (birth growth-stops (years 16) (years 20))
          (consecutive-phase childhood-to-adulthood linear-growth-ends)
          ((begin linear-growth-ends) (end linear-growth-ends) (years 10) (years 10))
          ))




(deftt "girl-short-bone-syndrome"
  :landmarks '(birth growth-stops)
  :intervals (list (definterval :name "early-childhood"
                :constraints
```

```
 (list (defconstraint :name "constant build"
       :func #'bd-infant-normal
       :parameters '(weight height)
       :model '(constant *)
       :weight 1)
 (defconstraint :name "height weight co-vary"
         :func #'wt-minus-ht-infant-zscores
         :parameters '(weight height)
         :model '(constant *)
         :weight 1)))
(definterval :name "childhood-to-adulthood"
 :constraints
 (list (defconstraint :name "increasing build"
       :func #'bd-child-normal
       :parameters '(weight height)
       :model '(linear (D1 +))
       :weight 1)
      (defconstraint :name "Late SD Falling"
       :func #'late-sd
       :parameters '(height)
       :model '(linear (D1 -))
       :weight 3)
      (defconstraint :name "Late SD between -5 and -1"
       :func #'(lambda (height-datum)
             (let ((ht-sd (late-sd height-datum)))
               (error-fourth-then-constant ht-sd '(-5 -1))))
       :parameters '(height)
       :model '(constant 0)
       :weight 1)
      (defconstraint :name "Late ht-vel Zscore between -5 to -1"
       :func #'(lambda (height-vel-datum)
             (let ((ht-vel-sd (late-sd height-vel-datum)))
               (error-fourth-then-constant ht-vel-sd '(-5 -1))))
       :parameters '(ht-vel)
       :model '(constant 0)
       :weight 1)
```

```
                (defconstraint :name "Chron Age - BA +-1"
                  :func #'(lambda (bone-age-datum)
                            (let* ((bone-age-delta
                                    (chron-age-minus-bone-age bone-age-datum)))
                              (error-fourth-then-constant bone-age-delta '(-1 1))))
                  :parameters '(bone-age)
                  :model '(constant 0)
                  :weight 5)
                ;; this is primary disturbance of growth, no BA delay.
                ))
              (definterval :name "linear-growth-ends"
                ;; height is constant for the next 10 years.
                :constraints
                (list (defconstraint :name "height-constant"
                        :parameters '(height)
                        :model '(constant *)
                        :weight 5)
                      (defconstraint :name "Bone Age = Chron Age"
                        :parameters '(bone-age)
                        :func #'chron-age-minus-bone-age
                        :model '(constant 0)
                        :weight 5)))

              ) ;; close off list of intervals

 :relations  '(
          ((begin early-childhood) birth 0 0)
          ((begin early-childhood) (end early-childhood) (years 2.05) (years 2.95))
          (consecutive-phase early-childhood childhood-to-adulthood)
          ((end childhood-to-adulthood) growth-stops 0 0)
          (birth growth-stops (years 13) (years 16))
          (consecutive-phase childhood-to-adulthood linear-growth-ends)
          ((begin linear-growth-ends) (end linear-growth-ends) (years 10) (years 10))
          ))

(deftt "girl-acquired-gh-def"
```

```
:landmarks '(birth now growth-stops)
:intervals (list (definterval :name "early-childhood"
                  :constraints
                  (list (defconstraint :name "constant build"
                         :func #'bd-infant-normal
                         :parameters '(weight height)
                         :model '(constant *)
                         :weight 1)
                        (defconstraint :name "height weight co-vary"
                          :func #'wt-minus-ht-infant-zscores
                          :parameters '(weight height)
                          :model '(constant *)
                          :weight 1)))
                 (definterval :name "childhood-to-onset"
                  :constraints
                  (list (defconstraint :name "constant build"
                         :func #'bd-child-normal
                         :parameters '(weight height)
                         :model '(constant 1)
                         :weight 1)
                        (defconstraint :name "Avg SD Constant"
                         :func #'avg-sd
                         :parameters '(height)
                         :model '(constant *)
                         :weight 1)
                        (defconstraint :name "Avg SD between +-2"
                         :func #'(lambda (height-datum)
                                   (let ((ht-sd (avg-sd height-datum)))
                                     (error-fourth-then-constant ht-sd '(-2 2))))
                         :parameters '(height)
                         :model '(constant 0)
                         :weight 1)
                        (defconstraint :name "Avg ht-vel SD between +-2"
                         :func #'(lambda (height-vel-datum)
                                   (let ((ht-vel-sd (avg-sd height-vel-datum)))
                                     (error-fourth-then-constant ht-vel-sd '(-2 2))))
```

```
    :parameters '(ht-vel)
    :model '(constant 0)
    :weight 1)
   (defconstraint :name "Chron Age - Bone Age = +- 1"
    :func #'(lambda (bone-age-datum)
          (let ((bone-age-delta
              (chron-age-minus-bone-age bone-age-datum)))
           (error-fourth-then-constant bone-age-delta '(-1 1))))
    :parameters '(bone-age)
    :model '(constant 0)
    :weight 5)
   ))
(definterval :name "onset-to-now"
 :constraints
 (list (defconstraint :name "Build increasing"
    :func #'bd-child-normal
    :parameters '(weight height)
    :model '(linear (D1 +.05 per-year))
    :weight 1)
   (defconstraint :name "Late SD dropping"
    :func #'late-sd
    :parameters '(height)
    :model '(linear (D1 -.6 per-year))
    :weight 3)
   (defconstraint :name "Late SD between -4 0, dropping (constant)"
    :func #'(lambda (height-datum)
          (let ((ht-sd (late-sd height-datum)))
           (error-fourth-then-constant ht-sd '(-4 0))))
    :parameters '(height)
    :model '(constant 0)
    :weight 1)
   (defconstraint :name "Late ht-vel SD between -4 0, dropping (constant)"
    :func #'(lambda (height-vel-datum)
          (let ((ht-vel-sd (late-sd height-vel-datum)))
           (error-fourth-then-constant ht-vel-sd '(-4 0))))
    :parameters '(ht-vel)
```

```
                :model '(constant 0)
                :weight 1)
              (defconstraint :name "BA-CA increasing"
                :func #'chron-age-minus-bone-age
                :parameters '(bone-age)
                :model '(linear (D1 +))
                :weight 1)))
          (definterval :name "linear-growth-ends"
            ;; height is constant for the next 10 years.
            :constraints
            (list (defconstraint :name "height-constant"
                :parameters '(height)
                :model '(constant *)
                :weight 5)
              (defconstraint :name "Bone Age = Chron Age"
                :parameters '(bone-age)
                :func #'chron-age-minus-bone-age
                :model '(constant 0)
                :weight 5)))


          ) ;; close off list of intervals

:relations  '(
        ((begin early-childhood) birth 0 0)
        ((begin early-childhood) (end early-childhood) (years 2.05) (years 2.95))
        (consecutive-phase early-childhood childhood-to-onset)
        (consecutive-phase childhood-to-onset onset-to-now)
        ((begin onset-to-now) now (years 2) (years 3))
        ((end onset-to-now) now 0 0)
        (growth-stops (begin linear-growth-ends) 0 0)
        ((begin linear-growth-ends) (end linear-growth-ends) (years 10) (years 10))
        (birth growth-stops (years 14) (years 18))
        ))
```

```
(deftt "girl-precocious-pub"
  :landmarks '(birth growth-stops)
  :intervals (list (definterval :name "early-childhood"
                    :constraints
                    (list (defconstraint :name "constant build"
                           :func #'bd-infant-normal
                           :parameters '(weight height)
                           :model '(constant *)
                           :weight 1)
                          (defconstraint :name "height weight co-vary"
                           :func #'wt-minus-ht-infant-zscores
                           :parameters '(weight height)
                           :model '(constant *)
                           :weight 1)))
                   (definterval :name "childhood-to-adulthood"
                    :constraints
                    (list (defconstraint :name "increasing build"
                           :func #'bd-child-normal
                           :parameters '(weight height)
                           :model '(linear (D1 +))
                           :weight 1)
                          (defconstraint :name "Ht SD Rising .2"
                           :func #'child-nchs-sd
                           :parameters '(height)
                           :model '(linear (D1 0.2 per-year))
                           :weight 3)
                          (defconstraint :name "Early SD between (0 4)"
                           :func #'(lambda (height-datum)
                                    (let* ((ht-sd (early-sd height-datum)))
                                     (error-fourth-then-constant ht-sd '(0 4))))
                           :parameters '(height)
                           :model '(constant *)
                           :weight 1)
                          (defconstraint :name "Chron Age - BA = -4 to -2"
```

```
                :func #'(lambda (bone-age-datum)
                        (let* ((bone-age-delta
                                (chron-age-minus-bone-age bone-age-datum)))
                               (error-fourth-then-constant bone-age-delta '(-4 -2))))
                :parameters '(bone-age)
                :model '(constant 0)
                :weight 5)))
          (definterval :name "linear-growth-ends"
            ;; height is constant for the next 10 years.
            :constraints
            (list (defconstraint :name "height-constant"
                    :parameters '(height)
                    :model '(constant *)
                    :weight 5)
                  (defconstraint :name "Bone Age = Chron Age"
                    :parameters '(bone-age)
                    :func #'chron-age-minus-bone-age
                    :model '(constant 0)
                    :weight 5)))

          ) ;; close off list of intervals

 :relations  '(
        ((begin early-childhood) birth 0 0)
        ((begin early-childhood) (end early-childhood) (years 2.05) (years 2.95))
        (consecutive-phase early-childhood childhood-to-adulthood)
        ((end childhood-to-adulthood) growth-stops 0 0)
        (birth growth-stops (years 11) (years 15))
        (consecutive-phase childhood-to-adulthood linear-growth-ends)
        ((begin linear-growth-ends) (end linear-growth-ends) (years 10) (years 10))
        ))



(deftt "girl-normal-build"
  :landmarks '(birth now)
```

```
 :intervals (list (definterval :name "Normal"
                  :constraints
                  (list (defconstraint :name "constant build"
                        :func #'build-normal
                        :parameters '(weight height)
                        :model '(constant 1)
                        :weight 1))))
 :relations '(
          (birth (begin Normal) 0 (years 18))
          ((begin Normal) (end Normal) (months 1) (months 1))
          ((end Normal) now 0 0)))


(deftt "girl-obese"
 :landmarks '(birth now)
 :intervals (list
          (definterval :name "Obese"
           :constraints
           (list (defconstraint :name "High build"
                 :func #'build-normal
                 :parameters '(weight height)
                 :model '(constant 1.3)
                 :weight 1))))
 :relations '(
          (birth (begin Obese) 0 (years 18))
          ((begin Obese) now (months 1) (months 1))
          ((end Obese) now 0 0)))



(deftt "girl-malnurished"
 :landmarks '(birth now)
 :intervals (list
          (definterval :name "Malnurished"
           :constraints
           (list (defconstraint :name "Low build"
```

```
                :func #'build-normal
                :parameters '(weight height)
                :model '(constant .8)
                :weight 1))))
  :relations '(
          (birth (begin Malnurished) 0 (years 18))
          ((begin Malnurished) now (months 1) (months 1))
          ((end Malnurished) now 0 0)))



(make-instance 'monitor-set
  :name "girl-growth"
  :tts (list
      "girl-average-growth"
      "girl-constitutional-delay"
      "girl-early-puberty"
      "girl-congenital-gh-def"
      "girl-short-bone-syndrome"
      "girl-precocious-pub"
      "girl-acquired-gh-def"
      "girl-normal-build"
      "girl-obese"
      "girl-malnurished"
      )
  :grain 0
  :error-calc 'mean-res
  :weights 'explicit)
```

11. References

Bankowitz, R.A., J.R. Lave, and M.A. McNeil. "A Method for Assessing the Impact of a Computer-Based Decision Support System on Health Care Outcomes." Meth Inform Med 31 (1992): 3-11.

Becker, Kenneth, ed. Principles and Practices of Endocrinology and Metabolism. NY: J.B. Lippincott Co., 1990.

Berner, Eta S., et al. "Performance of Four Computer-Based Diagnostic Systems." NEJM 330.179 (1994): 218-225.

Feldman, M.J., and G.O. Barnett. An Approach to Evaluating the Accuracy of DXplain, 1990. 38-43.

Paul Clayton, ed. Forsythe, Diana E., and Bruce G. Buchanan. Broadening Our Approach to Evaluating Medical Information Systems. Washington, DC: AMIA, 1992. 8-12.

Haimowitz, Ira J. "Knowledge-Based Trend Detection and Diagnosis." Ph.D. Massachusetts Institute of Technology, 1994.

Haimowitz, Ira J., and Isaac S. Kohane. Automated Trend Detection with Multiple Temporal Hypotheses. Chambery, France, 1993. 146-151.

Haimowitz, Ira J., and Isaac S. Kohane. An Epistemology for Clinically Significant Trends. Washington, DC: AAAI Press, 1993. 176-181.

Hayes-Roth, Frederick, Donald A. Waterman, and Douglas B. Lenat, eds. Building Expert Systems. 2 vols. London: Addison-Wesley, 1983.

Heathfiled, H.A., and J. Wyatt. "Philosophies for the Design and Development of Clinical Decision-Support Systems." Meth Inform Med 32 (1993): 1-8.

Heckerman, David E., and Bharat N. Nathwani. "An Evaluation of the Diagnostic Accuracy of Pathfinder." Comp & Biomed Research 25.1 (Feb) (1992): 56-74.

Kaplan, Solomon A. Clinical Pediatric Endocrinology. Philadelphia: W.B. Saunders Co., 1982.

Kohane, Isaac S. "Temporal Reasoning in Medical Expert Systems." : MIT Laboratory for Computer Science, 1987.

Miller, Randolph A., Harry E. Jr Pople, and Jack D. Myers. "INTERNIST-1, An Experimental Computer-Based Diagnostic Consultant for General Internal Medicine." NEJM 307 (1982): 468-476.

Pagano, Marcello, and Kimberlee Gauvreau. Principles of Biostatistics. Belmont: Wadsworth Publishing Company, 1993.

Roche, Alex F., et al. "Grading body fatness from limited anthropometric data." Am. J. Clin. Nutr. 34 (1981): 2831-2838.

Roland-Cachera, Marie Francoise, et al. "Adiposity indices in children." Am. J. Clin. Nutr. 36 (1982): 178-184.

Tanner, J.M. Foetus into Man: Physical Growth from Conception to Maturity. Cambridge, MA: Harvard University Press, 1990.

Tanner, J.M., and P.S.W. Davies. "Clinical Longitudinal Standards for Height and Height Velocity for North American Children." Journal of Pediatrics 107.3 (1985): 317-329.

Waterman. A Guide to Expert Systems. Ed Frederick Hayes-Roth: Addison-Wesley, 1985.