Curve Matching: A Data-Driven Technique to Improve Individual Prediction of Childhood Growth

Stef van Buuren

Netherlands Organization for Applied Scientific Research TNO, Leiden, and Department of Methodology and Statistics, FSS, University of Utrecht, Utrecht, The Netherlands

Abstract

Longitudinal growth data are valuable for predicting and interpreting future growth of individual children. This note explores the idea of 'curve matching', a new technique to improve prediction of future growth of an individual child. The key idea is to find existing children in existing databases that are similar to the current child. The growth patterns of the matched children suggest how the current child might evolve in the future. This paper describes the various conceptual and practical issues that need to be addressed before the idea can take off. A demo implementation is available at http://vps.stefvanbuuren.nl:3838/frisodemo/.

© 2014 S. Karger AG, Basel

- Given what I know of the child, how will it develop in the future?
- How certain am I of the child's future growth?
- If I do not intervene, will development be normal?
- If I do intervene, will the child's growth be normal or healthy?

Parents are interested in knowing the answers to questions like

- What can be done against my child's inhibited growth?
- What is the prognosis if we do nothing?
- How certain is this prognosis?

Insurance companies might have questions like

- Does this child get the most effective treatment among all options?

In this paper, I discuss *curve matching*, a tool designed to assist in the interpretation and prediction of individual growth curves that addresses questions like these. This note describes the various conceptual and practical issues that need to be addressed before the idea can take off.

Introduction

Longitudinal growth data are valuable for predicting and interpreting future growth of individual children. Historic growth data can be used to improve prediction of future growth for individual patients.

In practice, the health professional would like to be able to have answers to the following questions:

Curve Matching

Curve matching is a tool designed to assist in the interpretation and prediction of individual growth curves. The idea is as follows. Suppose that we have measured length, weight and head circumference of an infant up to half a year and that we have plotted the child's measure-

Stef van Buuren
PO Box 2215
NL–2301 CE Leiden (The Netherlands)
E-Mail stef.vanbuuren@tno.nl

ments onto the growth chart. The idea is to find historic length, weight and head circumference data of, say, 5 other (older) children who are similar to the index child and add their growth curves to the chart of the index child. If the matching is done right, then the historic growth data from the matched children suggest how the index child may develop in the future.

Figure 1 illustrates this idea. The graphs display the growth curves of head circumference, length and weight of a preterm boy aged 5.5 months on the 32-week preterm growth chart [1]. The values of this index child are plotted in red. Also plotted, in grey, are the growth curves of 5 other preterm boys born at 32 weeks who were matched to the index boy. Matching was done here in a rather simplistic way: The average Z-score until the age of 5.5 months was calculated for all infants born at 32 weeks of gestation, and 5 infants were selected that were closest to the index child in terms of the average Z-score. The next section discusses improvements in the approach for obtaining matches. Matching was done separately for each of the three outcomes, so the matched children differ between the panels.

Several observations can be made. The growth of head circumference of this boy slightly lags after 4 months. It would be interesting to know how head circumference would develop further in time. The 5 matched curves suggest that head circumference of the boy is likely to be between –1 and –2 SD of the mean around the age of 1 year. This indeed turned out to be the case for the boy. We know that here because we also collected more recent measurements for this boy.

Observe that the 5 matched curves are relatively close, which indicates that we can predict head circumference at 1 year with reasonable certainty. On the other hand, the matches for body weight (at the bottom panel) diverge somewhat, which indicates that predicting weight is more difficult than predicting head circumference. Thus, the variation between the matched curves tells us something about the certainty of the extrapolation of the child's curve to the future.

Since the matches correspond to real children, it can be useful to drill down electronic health records of the matches for details. For example, some of the matched children might have developed a preventable disease later in life (e.g. diabetes). Finding pathology among the matches may increase awareness of the professional and (the parents of) the child in time of a possible future health problem.

Figure 1 also illustrates a bottleneck. One of the 5 matches for head circumference is located at –2.5 SD

throughout and is actually a bad match. The reason is that the number of candidate matches in the donor database was small. The database contains 121 children born at 32 weeks of gestation. By definition, the number of candidate matches at the extremes of the distribution is low. Thus, curve matching requires relatively large databases.

Obtaining Appropriate Matches

The scientifically interesting problem centers on finding appropriate matches. This section briefly describes the approach taken in the demo application for forecasting child growth (http://vps.stefvanbuuren.nl:3838/frisodemo/). This application shows a growth chart for head circumference, length and weight. For simplicity, we assume that the clinical interest focuses on predicting the same measurement at a later age. Let Y_0, Y_1, Y_2 and Y_3 denote the measurement at birth (in Z-scale) and at 1, 2 and 3 months, respectively, for the target child (say child 'Laura S' from the application) now aged 3 months. Y_{14} is the yet unknown future measurement of Laura at the age of 14 months (in Z-scale). In addition, we also know the following covariates that influence growth: sex (boy/girl), gestational age (weeks), birth weight (grams), twin (no/yes), the height of the father and mother (cm), ethnicity (Dutch, Turkish or Moroccan), 'in good health' score (yes/no), whether the mother smoked during pregnancy (yes/no), educational level of the father and mother (low/middle/high) and age of the mother when giving birth (≤25, 26–36, ≥37 years) to the target child. We also have available the SMOCK donor database containing individual growth data of 1,933 children aged 0–15 months, as well as the same covariates. For the SMOCK data, a linear regression model is made to predict Y_{14} from all other information, i.e.

$$Y_{14} = [Y_0, Y_1, Y_2, Y_3, X] \, \beta + \text{error},$$

where X denotes the matrix of collective covariates. After fitting the model, we calculate the predicted value for each person in the donor dataset. In addition, we calculate the predicted value for the target using the same model. The idea is now to select those children from the donor set that have (almost) the same predicted value as the target child. Taking the 5 closest donors (in the scale of predicted values) provides the 5 'best matches'. The measured curves from the matches are then drawn onto the chart as backdrop information. If desired, we may calculate point estimates by averaging the measurements of the matched
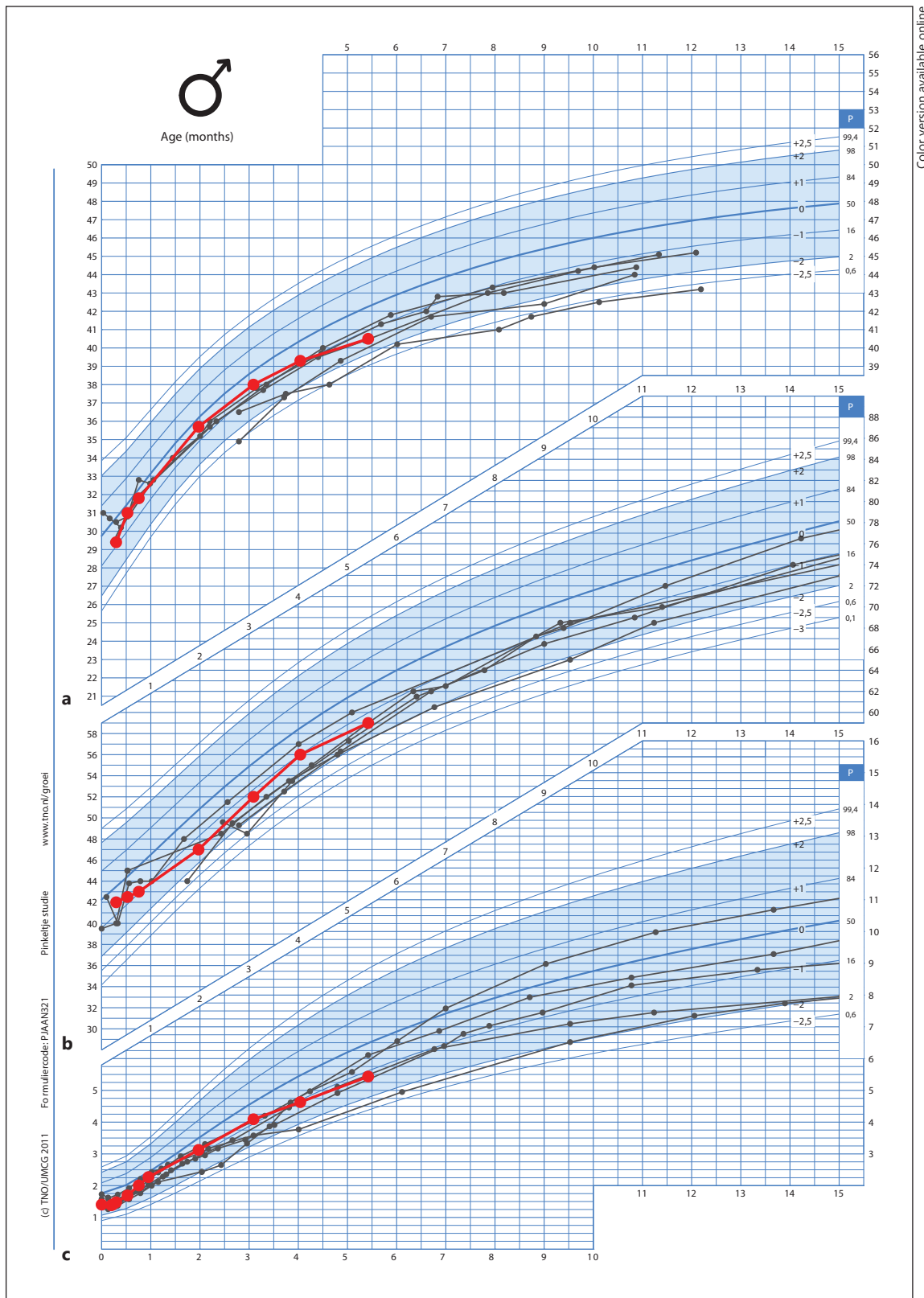
Fig. 1. Growth charts for preterm boys born at 32 weeks of gestation plotted as the growth curves of 1 index child (red in the online version) and 5 matched children (in grey). **a** Head circumference. **b** Length. **c** Weight.

children. Note that the procedure uses no future information from the target child.

Matching of the predicted value is most informative in the sense that it represents the (linear) single-number aggregate of the predictors. The technique has been developed for missing data imputation and is known as predictive mean matching [2]. In reality, the calculations are slightly more complex because in practice the measurements do not occur exactly at the ages of the planned visits. The broken-stick model, a simple multilevel model, is applied to the Z-scores to address this problem (see Hauspie et al. [6] for more details). Predictive mean matching is known to be extremely robust to violations of linearity and normality, very fast and the state of the art in missing data imputation.

It is an open research question what the number of matches should be. A small number of matches (say 5 or 10) already suffices to get a good grasp of inherent variability in the matched curves and still allows for a good view on the individual curves. However, 5 or 10 may be too small to estimate the prediction interval, e.g. the interval in which 80% of the real values lie. Given a sufficiently large donor dataset, we could easily enlarge the number of matches, but this is not without dangers as the increase in precision also comes with an increase in bias. In extremis, taking all donors as matches will predict the mean of the donor population irrespective of the person's growth, which is biased and useless. Using 5 or 10 matches is a reasonable compromise in practice for graphing purposes, whereas at least 50 matches are needed to get an estimate of the full posterior predictive density.

Architecture

It will be clear that curve matching is only feasible in a computerized environment. Curve matching is most naturally implemented as a client-server framework. Today's health professional typically already has access to a database containing medical records of his or her patients. In the client-server model, the functionality of the local database needs to be extended with client software that is able to communicate with a server that can calculate the matches.

The task of the online server is to create matches for the given patient data. For this purpose, the client needs to have access to a central donor database including patients that can act as potential matches. The donor database can be external to the local patient databases, but it could also be a distributed database of local databases. Such a setup is flexible, but is also slower as it requires more online data communication. The server also needs to implement a statistical matching algorithm. The task of the algorithm is to search the donor database for the best matches to the index patient given the patient data and, optionally, given preferences and settings specified by the end user.

From the perspective of the end user, drawing charts with matched curves should be immediate. It is thus important to have fast online communication and curve matching algorithms so that overall response time per chart stays, say, below 1 s. This requires fast network communication, small graphic files, an indexed and fast donor database with relevant patient data, robust matching algorithms and established communication protocols that bind everything together using thin software layers. As all of these tools are available nowadays, there seem to be no real technological barriers that complicate developing a client-server system.

Special attention is needed for data problems that are likely to be encountered. In particular, the registration of diseases, dates, background factors and anthropometric measurements can differ widely in different electronic health record systems. This may range from trivial scale differences (e.g. centimeters or meters) to more complicated problems that affect comparability of information or that lead to missing data. To prevent such problems, the matching algorithm needs to be relatively robust against data issues.

All technological components for building a client-server model are available, many of which are free. What lacks is the integration of the different components into one system. In addition, as the system is novel, there is no user experience yet that may be helpful to steer development. In addition, user guidelines and instructions on the appropriate interpretation of matched curves need to be developed.

Implementation

Curve matching is an add-on to aid interpretation of growth charts. The 'MARK I' system could start off as a simple version of the server, a database with matches and a few early users who will be using the same electronic health record system. Additional facilities can be built on top to create 'MARK II', where hopefully only minimal alterations to the client are required to use new functionality.

One technical extension is the ability for on-the-fly calculation of referral guidelines. Another technical extension is to build live databases where index patients are automatically added to the donor database as a potential match for other patients. In this way, the matching pool will grow with its intensity of use.

Adoption, implementation and continuation can be done in a gradual fashion. There is no need to develop or implement the entire system at one time. Adoption can be facultative. The introduction of the system can be done primarily by word of mouth.

Prospective users may need to develop new skills for adequate interpretation. Current decision-making guidelines may need revision so as to incorporate expected growth. For example, referral guidelines could be generalized to include future growth in the decision-making process (e.g. refer the child if 4 out of 5 matches end up below the future threshold). The technology to assess and optimize the validity of evidence-based referral guidelines has matured over the last decade and can be readily applied [4].

Effectiveness

As the system has yet to be built, it is not known whether the system will actually be effective at reducing morbidity. We may investigate effectiveness by performing embedded randomized experiments where various aspects of the system (e.g. type of matching or type of the graph) are systematically varied, and where the outcome (morbidity) is recorded. It is also possible to build in automatic checkups of old predictions based on more recent data, so that the system can learn and improve its matches and predictions.

My hypothesis is the curve matching will enable better predictions for individual patients and thus enhance the possibilities for tailored medicine. The reason it works is that it gives matched curves of *real growth* of *real children*. Health professionals are trained to think what is best for individual patients. Their gradually growing experience enables them to make implicit matches of the index child to similar children they may have seen in the past. The statistical matching algorithm is a way to enhance, enrich and assist the memory of the health professional by a comparable strategy, so that better-informed decisions based on more 'experience' are possible. Of course, we need evidence to document this claim.

Added Value

Curve matching may be attractive to different users for different aspects. Curve matching

- objectifies the effects of an intervention of a child before treatment starts. We could ask the matching algorithm to find matches from the group that got the intervention in the past. Also, we could narrow our search to matches from the group that did not get the intervention. The difference between these two groups of matches may show whether the intervention has a relevant effect for this child;
- allows to see the uncertainty associated with the future. So it comes with a built-in warning for overprediction;
- attempts to enhance, enrich and assists the memory of the health professional, and may help to identify and question false beliefs;
- renews itself as the data of new children can be added to the donor database and those of older children can be discounted;
- creates a visual representation of 'what lies ahead'. This makes it easier for the child or parent to make better-informed preferences about possible treatments, and hence contribute to patient emancipation and empowerment;
- can be applied without adaptation to endpoints other than anthropometric measures;
- may stimulate fellow-sufferer contact by forming real-life groups of matches, and
- puts existing information to a new use.

Different users may emphasize different points, but, together, each point adds to the attractiveness of curve matching.

User Acceptance

The primary user group consists of professionals working in the youth health care sector: youth health care professionals, nurses, assistants, pediatricians and endocrinologists. Each of these groups deals with individual children. Curve matching enables comparison with similar children. Plotting the curves of the matched children onto the same graph will look familiar. We may therefore expect that adoption of matched curves can be quick once the initial step has been set of plotting multiple children on one chart. This presentation format might connect more easily to the end user than presenting results of statistical models in an unfamiliar display.

It is likely that the health professional may adopt the tool more quickly if the donor database consists of children from his or her own practice. It makes sense to provide the user with the option of matching to the own local database. The risk of doing so is that the donor database may be too small, resulting in the same matches for the extreme children (which are generally of most clinical interest). Some users may wish to influence the set variables for which matches are made (e.g. restrict matches to an ethnic group). Options like this may contribute to the confidence users have in the matching methodology.

Risks and Barriers

Curve matching will only work if it can be integrated into the local patient database. Thus, an important requirement is that there is a software plug to communicate with a curve matching server. Some of the current systems are closed, which represents a potential barrier against successful implementation.

Privacy concerns form another potential risk. Curve matching can work from anonymous data. However, some unique combination of personal data may lead to disclosure of the identity of the matched person data, which is undesirable. The nondisclosure risk may put a boundary on the set of characteristics that can be used to find matches (e.g. no ethnicity and age categories instead of exact ages, for example).

Another potential limitation is what I call 'dislike of inherent uncertainty of predictions'. In reality, predictions are often highly uncertain, and some people may find it difficult to deal with that fact. The inherent uncertainty may be difficult to explain to the parents and could invoke reactions like 'even the doctor does not know'. The best available evidence often adds to our confusion. A proper interpretation of the results for a given child may go against accepted practice and recommended guidelines. It takes some courage to deviate from mainstream advice for this child, so curve matching may not be for everybody. There could even be adverse legal consequences in some settings. Experience is needed to determine how such influences operate in practice.

Curve matching will only be useful if the user has confidence in the appropriateness of the matches found. Building trust takes time, and the matches have to tap into the experience of the user. There is a high penalty of inappropriate matches, so the matching algorithm should be carefully user tested before going live.

Related to this is the danger of too small or inadequate donor databases. Almost by definition, extremes in growth are the areas of most clinical significance. As the open population contains only very few children located in the extremes, the danger is that curve matching may produce the same matches repeatedly for the most relevant cases. To counter this effect, the donor database should also include clinical cases.

Despite these dangers, I believe that – when properly done – curve matching represents a major improvement over the current practice. In the ideal scenario, the health professional will be able to form an adequate image of what will happen to this particular child some time from now. When in fact that does happen, the system gains credibility and may eventually replace accepted, but misguided, wisdom. Nevertheless, as always, the health professional should stay alert to exceptions.

Costs

Curve matching is an alternative to one-size-fits-all medicine. The promise of the technique is that care can be tailored to the individual. Ideally, this results in less overtreatment and hence in more efficient health care. Whether that actually will happen in reality will of course depend on more factors than curve matching alone.

Short-term costs of curve matching include system development, the setup of a donor database, data harmonization and education of users. The primary long-term costs include system maintenance and fine-tuning of the algorithms used.

The budget for information technology within an organization often contains substantial posts for data storage, backup and security. Online storage has become very cheap with the arrival of cloud-based computing and free Linux-based operating systems, so it is only natural to transfer some of the funds for 'data storage' to 'data use'.

Conclusion

There is a great demand for sound techniques that can predict individual health. Curve matching is a simple yet powerful approach that can put huge collections of medical data to a new use. The key idea is to find and present individual historic data of patients who are similar to the patient sitting in the consultation room. A convincing presentation of relevant historic data assists in evaluating what might happen to this patient.

Creating robust matches from longitudinal data requires some care. Ideally, the matches should (1) be close to the index child; (2) on average predict the correct outcome, and (3) have the lowest possible prediction error. Figure 1 displays 5 matches, but it is an open research question what the optimal number of matches should be. Also, different matching metrics can be used. It is not yet clear which metric is best. We also develop statistical methodology for solving problems related to comparability and harmonization resulting from the combination of different data sources. All these issues can be usually framed as missing data problems for which solutions are now increasingly becoming available [3]. It would be useful to compare curve matching to existing methods for individual prediction methods that use parametric descriptions of future growth [5–7].

This paper has concentrated on the setting of child health, but the idea is certainly not restricted to that. We can easily extend to other settings. For example, curve matching may assist in answering questions like:

- Can this patient walk to the shops 2 months after the operation? And can he do it without the operation?
- If I keep smoking, how long will I live? And for how long if I quit?
- How many people like this patient will get a cerebrovascular accident within 10 years? And how does that number change if this brand of medication is given to reduce blood pressure?

- I can choose among 5 different treatment options for this patient. Which of these improves health most? And which of these if health is divided by costs?

Successful adoption and implementation of curve matching requires the cooperation of many different disciplines. We need to connect to consortia that have adequate donor data. We need to tap into the relevant clinical knowledge and tailor matching algorithms to that. We need to set up a technical infrastructure so that advances in matching techniques are easily available to a wide group. And we need the interest of commercial partners who are able to apply our matching technology as part of their products.

Acknowledgment

Disclosure Statement

The author has no relevant financial or nonfinancial relationships to disclose.

References

1 Bocca-Tjeertes IFA, van Buuren S, Bos AF, Kerstens JM, ten Vergert EM, Reijneveld SA: Growth of preterm and fullterm children aged 0–4 years: integrating median growth and variability in growth charts. J Pediatr 2012; 161:460–465.
2 Little RJA: Missing-data adjustments in large surveys (with discussion). J Business Econ Statist 1988;6:287–301.
3 van Buuren S: Flexible Imputation of Missing Data. Boca Raton, Chapman & Hall/CRC, 2012.
4 van Dommelen P, van Buuren S: Methods to obtain referral criteria in growth monitoring. Stat Methods Med Res 2013;23:369–389.
5 Berkey CS, Kent RL: Longitudinal principal components and non-linear regression models of early childhood growth. Ann Hum Biol 1983;10:523–536.
6 Hauspie R, Cameron N, Molinari L (eds): Methods in Human Growth Research. Cambridge, Cambridge University Press, 2004.
7 Meigen C, Hermanussen M: Automatic analysis of longitudinal growth data on the website willi-will-wachsen.de. Homo 2003;54: 157–161.