

Stage line diagram: An age-conditional reference diagram for tracking development

Stef van Buuren^{1,2,*},† and Jeroen C. L. Ooms²

¹*Department of Statistics, TNO Quality of Life, P.O. Box 2215, 2301 CE Leiden, The Netherlands*

²*Department of Methodology and Statistics, FSS, Utrecht, The Netherlands*

SUMMARY

This paper presents a method for calculating stage line diagrams, a novel type of reference diagram useful for tracking developmental processes over time. Potential fields of applications include: dentistry (tooth eruption), oncology (tumor grading, cancer staging), virology (HIV infection and disease staging), psychology (stages of cognitive development), human development (pubertal stages) and chronic diseases (stages of dementia). Transition probabilities between successive stages are modeled as smoothly varying functions of age. Age-conditional references are calculated from the modeled probabilities by the mid- P value. It is possible to eliminate the influence of age by calculating standard deviation scores (SDS). The method is applied to the empirical data to produce reference charts on secondary sexual maturation. The mean of the empirical SDS in the reference population is close to zero, whereas the variance depends on age. The stage line diagram provides quick insight into both status (in SDS) and tempo (in SDS/year) of development of an individual child. Other measures (e.g. height SDS, body mass index SDS) from the same child can be added to the chart. Diagrams for sexual maturation are available as a web application at <http://vps.stefvanbuuren.nl/puberty>. The stage line diagram expresses status and tempo of discrete changes on a continuous scale. Wider application of these measures scores opens up new analytic possibilities. Copyright © 2009 John Wiley & Sons, Ltd.

KEY WORDS: growth; puberty; graphic tool; Tanner stages; quantification

1. INTRODUCTION

Age-conditional reference diagrams of height, weight and body mass index are widely being used to monitor growth and development in children. The methodology to construct diagrams for continuous outcomes has evolved over the last 15 years and its application is now more or less standard [1, 2]. In contrast, relatively little work has been done for ordered measures, such

*Correspondence to: Stef van Buuren, Department of Statistics, TNO Quality of Life, P.O. Box 2215, 2301 CE Leiden, The Netherlands.

†E-mail: stef.vanbuuren@tno.nl

as indicators of motor and mental development, visual acuity, developmental milestones, clinical grades, pubertal stages, and so on.

One approach to ordinal data is to construct an aggregate, such as a sum score, and use this aggregate as an essentially continuous measure. This approach is natural if indicators can be viewed as different manifestations of the same underlying trait, and if the aggregate provides a sensible quantification of the trait. Using this approach, Jacobusse *et al.* fitted a Rasch model to developmental data in infants, thereby deriving an interval measure of development between ages 0–2 years [3]. Calculating age-conditional references is straightforward by applying conventional methods for continuous data to the aggregate.

A second approach is to model the probability of the transitions between successive categories. Aitchison and Silvey extended the binary probit model to ordinal outcomes [4]. They assumed that the observed data are manifestations of an underlying normal distributed variable, which are linked to each other by a series of probit models, one for each category transition. Wade *et al.* applied this type of modeling to developmental data, where the transitions between categories were modeled by an asymmetric logistic models. Their method was applied to derive age-conditional references for the development of visual acuity during childhood and for the recognition of emotions [5–7]. Royston placed this methodology into a general framework, and extended the family of parametric models to include multiple covariates and more liberal forms of dependencies between the outcome and the covariates [8]. He advocated his method as being capable of modeling outcome variables with peculiar distributions (e.g. many zeroes) that are difficult to model in the conventional way.

The present paper provides a third alternative. Our method relies on the assumption of age-continuity (cf. Section 2). Similar to the above approach, age-conditional references of development are expressed as transition probabilities whose value smoothly changes with age. Unlike previous approaches, the reference diagram is calculated from the reference values by the mid- P value. This results in a novel graph that can be used to track development over time.

Section 2 outlines the substantive problem. Section 3 introduces the methodology for calculating reference values and stage lines from discrete data. Section 4 provides an example of the new graphic display and Section 5 contains further quantitative results. Section 6 proposes a way to implement the method by means of a web application. Section 7 contains conclusions and provides suggestions for further research.

2. AGE-CONTINUITY OF PUBERTAL DEVELOPMENT

Puberty is a phase of life with important medical and psychosocial consequences. Pubertal maturation is generally classified into distinct stages of secondary sexual characteristics. The system of ‘Tanner stages’ is widely accepted and applied [9]. For boys, there are three types of measures: genital development (5 stages G1–G5), pubic hair (6 stages PH1–PH6) and testis size (12 stages T1–T12 corresponding to volumes of 1, 2, 3, 4, 5, 6, 8, 10, 12, 15, 20, 25 ml). For girls, we have breast development (5 stages B1–B5), pubic hair (6 stages PH1–PH6) and menarche (2 stages no/yes). Tanner stages form ‘reasonably reliable indicators’ of pubertal development that provide the most accurate results if scored by a trained health professional [10].

Standards of pubertal maturation are typically published as the ages t_{10} , t_{50} and t_{90} at which respectively, 10, 50 and 90 per cent of the reference population achieves a certain stage. In clinical practice, the physician examines the child, determines the stage appropriate for the child, and

compares the child's age to the 'normal' age range $t_{10}-t_{90}$ for that stage. This procedure outlined will answer the question 'Does this child mature 'early', 'normal' or 'late'?'

The above procedure works well if only a classification into 'early' vs 'normal' vs 'late' is needed, but lacks any sense of continuity between 'early' and 'late'. It is not difficult to imagine pubertal development as continuous though. Even if the phenomenon under study is a rigid succession of distinct stages, the timing to go from one stage to the next differs between individuals. This induces a form of continuity that we call age-continuity. Suppose that two children, one young and one old, are in the same developmental stage. We can say that, on average, the younger child matures earlier than the older child. To see why this is the case, consider the fact that the younger child still has the opportunity to move into the next stage before his/her age reaches that of the older child, whereas the older child does not have this opportunity anymore. Thus, depending on age, the same stage is associated with different degrees of maturation. Moreover, the difference in maturation grows with the age gap.

Reference standards describe the natural variation of a measure in a reference population. For a discrete measure like the Tanner stages that use five categories, the references take on only five distinct maturation scores at a given age. These values, however, gradually change with age, thereby producing a measure that is continuous in age. The new methodology exploits age-continuity to quantify and compare development of children measured at the different ages.

3. METHOD

Let Y be an ordered stochastic variable whose values $Y \in \{1, 2, \dots, m\}$ correspond to stages 1 to m , and let X be represent the decimal age. We denote $P(Y \geq c|X)$ as the probability of achieving stage c ($c=2, \dots, m$) at age X , and define $P(Y < c|X) = 1 - P(Y \geq c|X)$. For each stage transition, we model a reference curve conditional on age c by the additive probit model

$$\Phi^{-1}(P(Y \geq c|X)) = \alpha_c + f_c(X), \quad c=2, \dots, m \quad (1)$$

where Φ is the standard normal cumulative distribution function and where $f_c(X)$ is a smooth univariate function of age. The use of $f_c(X)$, instead of the linear term $X\beta_c$, allows for a flexible and smooth change of the transition probability with age. Models of this type overcome limitations of probit and logistic regression [11].

The probability of observing stage c at age X is given by

$$P(Y = c|X) = P(Y < c+1|X) - P(Y < c|X), \quad c=1, \dots, m \quad (2)$$

where $P(Y < 1|X) \equiv 0$ and $P(Y < m+1|X) \equiv 1$, so that $P(Y = 1|X) + \dots + P(Y = m|X) = 1$ for all X .

In order to assess the relative position of a score Y at age X , we assume that the observable stages $Y = c$ are a coarse version of a continuous latent variable \tilde{Y} . More specifically, we assume that $\tilde{Y} \sim U(0, 1)$ has a uniform distribution between 0 and 1 that is linked to the observed variable Y as

$$Y = c \quad \text{if} \quad P(Y < c|X) \leq \tilde{Y} < P(Y < c+1|X), \quad c=1, \dots, m \quad (3)$$

The assumption is $\tilde{Y} \sim U(0, 1)$ will be reasonable if the reference model adequately fits the reference data as evaluated through worm plots or Q -statistics [12, 13]. The subset of values of \tilde{Y} that

correspond to stage c has a uniform distribution $\tilde{Y}_c \sim U[P(Y < c|X), P(Y < c+1|X)]$. The mean and variance of \tilde{Y}_c are given by

$$\pi_c|X = P(Y < c|X) + P(Y = c|X)/2, \quad c = 1, \dots, m \quad (4)$$

$$\sigma_c^2|X = P(Y = c|X)^2/12, \quad c = 1, \dots, m \quad (5)$$

The quantity $\pi_c|X$ is known as the ‘mid- P value’, and was introduced as an alternative to the conventional P -value to correct for continuity [14]. Until now, the mid- P value is primarily used in tests for discrete data such as Fisher’s exact test [15]. Note that a proportion of $P(Y = c|X)$ subjects will have \tilde{Y} -values between the category boundaries $P(Y < c|X)$ and $P(Y < c+1|x)$. It is known that the mean of these values is a summary, that is, optimal in the sense that it has the minimum squared error in a wide variety of situations [16]. Under the uniform distribution, the mean and mid- P values coincide. A further convenient property is $P(\tilde{Y} < \pi_c|X) = \pi_c|X$, i.e. for all X the probability that \tilde{Y} falls below π_c is identical to the mid- P value itself, π_c .

We interpret the mid- P value $\pi_c|X$ as the location on the \tilde{Y} -scale that optimally represents the maturity of children in stage c at age X . Reference diagrams for pubertal stages can be drawn as m *stage lines*, each of which portrays the development of a stage over age as the mid- P value $\pi_c|X$. The vertical axis can be presented in the probability scale, the logistic scale, or the probit scale. In the sequel, we choose the probit scale because its values correspond to standard deviation scores (SDS) or Z -scores. The SDS measure is popular in growth and development because it provides high resolution at the extremes, the areas of most clinical significance. Probabilities are transformed into SDS by the probit transformation $Z_c = \Phi^{-1}(\pi_c|X)$. In the Z -scale, the location Z_c corresponds to a weighted average of Z scores within the interval $[\Phi^{-1}(P(Y < c|X)), \Phi^{-1}(P(Y < c+1|X))]$, where the weights correspond to the standard normal density.

Equation (4) can be adapted to calculate Z_c . Suppose that reference values $P(Y \geq c|X = t)$ are available for $c = 2, \dots, m$ at tabulated ages $X = t$. For a given age X and stage c , the SDS of an observation is calculated as

$$Z = \Phi^{-1} \left(1 - \frac{P(Y \geq c|X) + P(Y \geq c+1|X)}{2} \right) \quad (6)$$

where $P(Y \geq c|X)$ is interpolated linearly from the surrounding tabulated ages t_1 and t_2 ($t_1 \leq X \leq t_2$) by $P(Y \geq c|X) = hP(Y \geq c|X = t_1) + (1-h)P(Y \geq c|X = t_2)$ and $h = (t_2 - X)/(t_2 - t_1)$. Equation (6) allows us to draw a diagram that relates the concept of early/late maturation to the proportion of children in the reference sample that achieves some stage.

For completeness, we also provide the inverse relations of (4) and (5).

$$P(Y = c|X) = 2\sqrt{3}\sigma_c|X \quad (7a)$$

$$P(Y < c|X) = \pi_c|X - \sqrt{3}\sigma_c|X \quad (7b)$$

$$P(Y < c+1|X) = \pi_c|X + \sqrt{3}\sigma_c|X \quad (7c)$$

These equations allow us to switch back and forth between the Z -score and probability scales.

Stage lines are estimated from the reference data, and so are subject to sampling variability. We can study the influence of sample size on the stage lines by the bootstrap [17]. The procedure involves drawing a random sample with replacement from the reference data, fitting the reference model (1) to the sample, calculating the set of stage lines by (4), and transforming these to the

Z-scale by (6). We replicate these steps 1000 times, and calculate confidence bands that contain 80 per cent of the fitted stage lines. The widths of these bands portray the influence of sample size on the stage lines.

4. A REFERENCE DIAGRAM OF BREAST DEVELOPMENT

Estimates $P(Y \geq c|X)$ for Tanner stages were available from data of the Fourth Dutch Growth Study, which measured pubertal development in a group of 2524 boys and 3028 girls [18, 19]. The probabilities were maximum likelihood estimates calculated by local scoring using the *S-Plus* `gam()` function, and where smoothing was done by splines using the *s()* function. The estimated probabilities $P(Y \geq c|X)$ for $c=2, \dots, 5$ are given as a four numbers at a grid of *X*-values, which together form a growth references for breast development, pubic hair and menarche.

Since we cannot directly observe the age at which a transition to a next stage occurs, tabled references are not directly suited for drawing reference diagrams. The available data measure the stage in which the child currently resides, so reference diagrams must be based on the probability of being in a stage given age.

Figure 1 is a *stage line diagram* of breast development, an age-conditional reference diagram. The horizontal axis represents age between 8 and 21 years. The vertical axis indicates maturation status as SDS correcting for age. Lower values indicate delayed development, and higher values signal early maturation. The diagram contains five stage lines. Each stage line corresponds to one of five Tanner's stages of breast development.

The use of the diagram is as follows. A trained observer determines the child's stage B1–B5 in the usual way, places a mark on the stage line corresponding to the child's stage at the child's age, and connects the mark to the previous measurement. The curve gradually tails off as long as

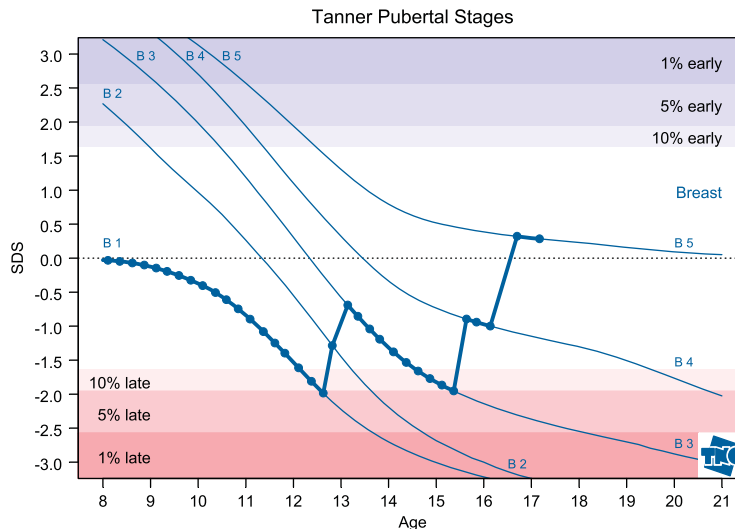


Figure 1. Stage line diagram of breast development (B-SDS) based on Dutch growth references 1997. The plotted curve corresponds to longitudinal observations on one patient.

the child remains in the same stage. A move to the next stage produces a jump in the curve. The age at which the child reaches the next stage is unknown, and can be anywhere between the two ages surrounding the jump. Steeper jumps occur for measurements that are closer in time. Jumps can span two or more stages.

Growth curves of normally developing children are located in the middle region of the diagram, roughly between -2 SDS and $+2$ SDS. Early maturing children are placed near the top of the diagram, while children with developmental delay appear near the bottom. Regions signifying 10, 5 and 1 per cent extreme children are marked at both sides for easy reference. These regions can be used to set action levels and monitor treatment. The slope of the curve starting from the last B1-mark is a measure of developmental tempo in SDS/year.

The stage line diagram fully supports the conventional procedure that compares the age of the child to a stage-specific reference interval. In clinical use, the physician determines the developmental stage of the child, and compares the child's age to the 'normal' age range t_{10} – t_{90} for that stage. The age interval t_{10} – t_{90} of stage B2 can be read off from the diagram as the age t_{10} at which line B2 crosses the 10 per cent early region, i.e. $t_{10(B2)} = 9.01$. Age t_{90} is the age at which line B1 crosses the 10 per cent late region, i.e. $t_{90(B1)} = 12.16$ years. Thus, the new diagram supports the conventional procedure for classifying children into 'early', 'normal', 'late' by t_{10} and t_{90} .

In addition, the diagram can be used to compare maturation across different ages. For individual use, we recommend plotting all three Tanner measures in the same diagram. This display visualizes differences in status and tempo of different aspects of pubertal development within the same individual. Figure 2 shows pubertal development of a girl with Turner Syndrome. Pubic hair development is normal, and clearly ahead of breast development and menarche. Menarche has only two stages, so its curve has only one jump. For this girl, menarche occurred very late somewhere between ages 16.5 and 17. If desired, other measures on the SDS scale, such as height SDS or body mass index SDS, can be added.

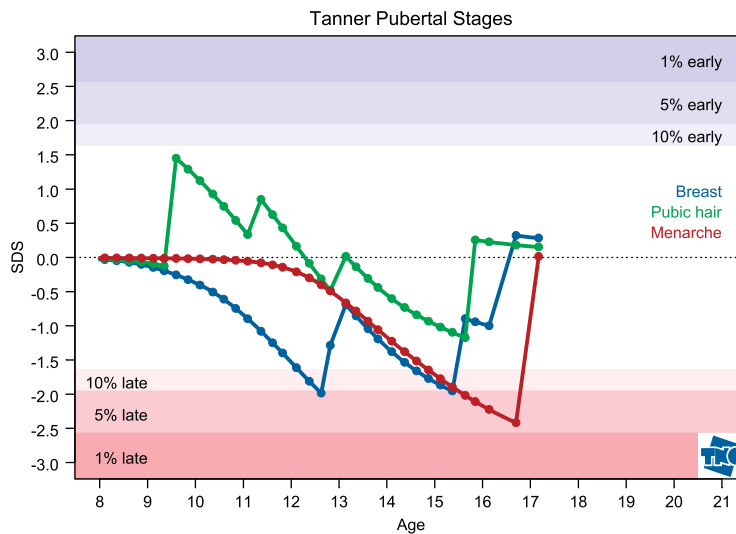


Figure 2. Stage line diagram where B-SDS, PH-SDS and M-SDS are combined on one display.

Figure 3 plots breast development from 200 girls with TS on the top of each other. Since breast development is severely delayed in this group, most of the curves occur in the lower region. The graph also allows us to spot some peculiarities in the data. Nearly all jump lines have positive slopes, signaling an advance in developmental stage. However, some negative slopes appear which can only occur if a higher state is followed by a lower state. Whether this represents a genuine developmental phenomenon, a measurement error, or a data error is not known, but the diagram allows us to quickly spot anomalies.

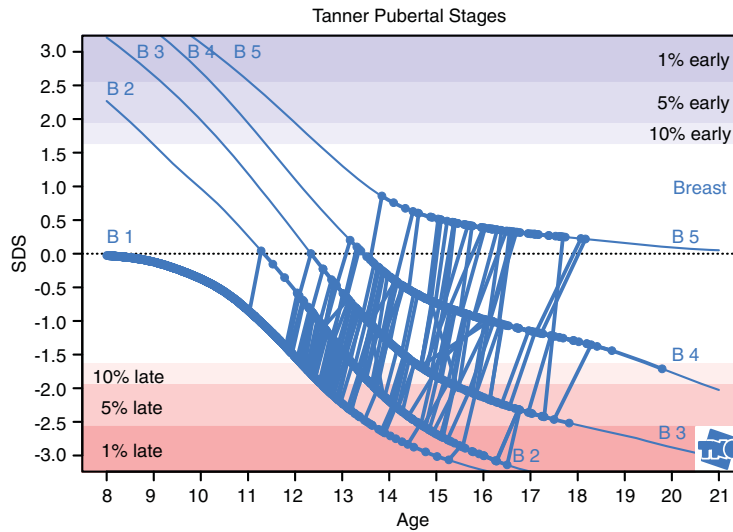


Figure 3. Stage line diagram that superposes longitudinal B-SDS data of 200 girls with Turner Syndrome.

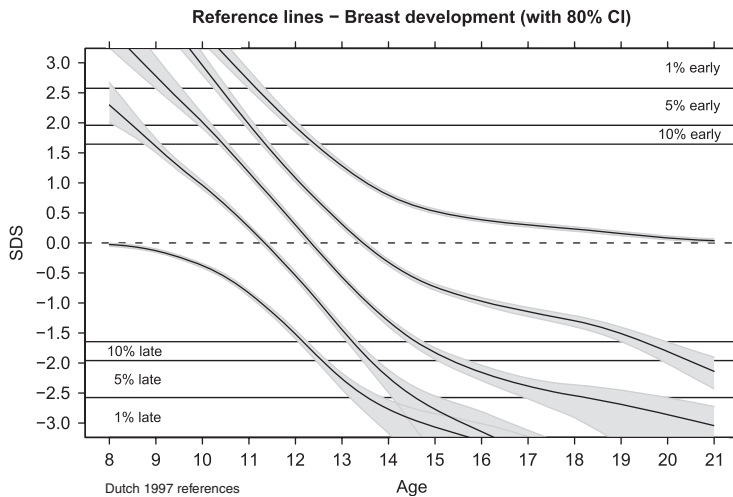


Figure 4. Stage line diagram of breast development with 80 per cent confidence intervals.

Figure 4 portrays the influence of sample size on stage lines for breast development. The gray bands around the stage lines are 80 per cent confidence intervals obtained by bootstrapping. The sample sizes in the Fourth Dutch Growth Study were large, so the confidence intervals are narrow most of the time. Though sampling variability grows near the extremes, the differences between the stage lines remain 'statistically significant' almost everywhere. We may therefore safely ignore sampling variability and adopt the usual view that takes the references as fixed.

5. QUANTITATIVE MEASURE OF MATURATION

Let B-SDS, PH-SDS and M-SDS denote maturation scores for breast, pubic hair and menarche, respectively, calculated by equation (6). Table I lists the mean and variance of these measures, split according to age, in the reference girls. Observe that mean B-SDS, PH-SDS and M-SDS are close to zero, which is expected in the reference population.

In contrast to SDS of continuous data, the variance depends on age and is less than 1. The variance starts off low around the age of 8 years, reaches its peak around the ages 11–13 years and gradually tails off at older ages. The explanation of this pattern is as follows. Around the age of 8 years, the distribution of Tanner scores is highly asymmetrical since most girls are in stage 1. Likewise, for ages around 20 years, the distribution of Tanner score is skewed in the opposite way. These distributions represent the floor and ceiling of the scoring system. As age progresses, the distribution gradually changes from the floor to the ceiling.

The maximum variance found in Table I is 0.82. The variance of categorical data depends on two factors: the number of categories and the distribution of the measurements over the categories. The maximal efficiency of a five-category system relative to a truly continuous measure is 92 per cent [20]. Thus, about 8 per cent of the variance is lost because of the inherent discreteness of the measurement. The maximal efficiency in a five-category system will be achieved only if the percentages of cases per category are equal to 10.9, 23.7, 30.7, 23.7 and 10.9 per cent. In fact, at the age of 12.25 years, the empirical distribution for breast development is remarkably

Table I. Summary measures of SDS maturation scores in the reference population.

Age	N	Breast		Pubic hair		Menarche	
		Mean	Variance	Mean	Variance	Mean	Variance
8–9	63	–0.07	0.05	–0.03	0.07	0.03	0.11
9–10	255	0.09	0.44	0.10	0.40	0.04	0.13
10–11	282	0.01	0.59	0.05	0.51	–0.01	0.06
11–12	273	0.04	0.82	0.01	0.78	0.00	0.18
12–13	243	0.01	0.81	0.00	0.81	0.04	0.41
13–14	243	0.01	0.71	0.00	0.70	0.02	0.42
14–15	227	–0.02	0.58	–0.07	0.57	–0.06	0.34
15–16	225	–0.07	0.56	–0.04	0.44	–0.02	0.16
16–17	106	–0.06	0.47	–0.08	0.35	0.00	0.05
17–18	82	0.00	0.39	–0.04	0.29	0.01	0.00
18–19	105	–0.09	0.45	–0.07	0.33	0.00	0.00
19–20	45	–0.01	0.19	–0.12	0.34	–0.06	0.02
20–21	52	–0.03	0.20	0.05	0.00	0.00	0.00

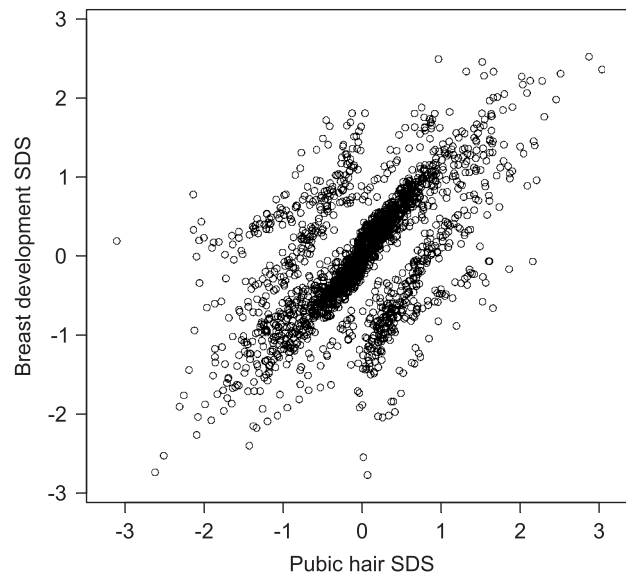


Figure 5. Scatter diagram of B-SDS and PH-SDS of Dutch reference data. Points are slightly jittered.

close: 9.7, 23.5, 36.3, 22.5 and 7.9 per cent. At this age, the relative efficiency is almost 92 per cent. The variance is lower at other ages because the asymmetry in the distribution increases as we move away from the optimum age. For binary data, the maximal efficiency is 64 per cent. This optimum for menarche is achieved around the P50 in the reference population, i.e. at the age of 13.25 years. The fact that the variance is lower than 1 is not an error, but it reflects the quality of the measurement at different ages.

Figure 5 is a scattergram of B-SDS vs PH-SDS of Dutch reference girls, a cross-sectional study. The values are slightly jittered. The correlation is equal to 0.59. Different quadrants correspond to different maturation patterns. Girls located in the upper right corner are early in both *B* and PH development, while those located in the lower left corner are late on both aspects. Girls in the upper left corner are late on PH and early on *B*. Girls in the lower right corner are early on PH, but late on *B*.

6. WEB APPLICATION

To illustrate how stage line diagrams can be made directly available we implemented the puberty diagram as a web application. The application is publicly available at <http://vps.stefvanbuuren.nl/puberty>, and requires only a standard internet browser on the client side.

The web site allows the user to enter a series of observed Tanner stages with the corresponding ages for a patient. These data are sent to a server that dynamically generates the stage plot, and instantly returns the plot to the client by means of an Ajax request [21]. The server side process of the application is handled using rapache, a module for the Apache web server that allows adding *R* scripts to web pages [22]. The *R* script calls a function that calculates the SDS scores for the

observed Tanner data and that draws the appropriate stage line diagram. After the diagram has been shown to the user, reference lines can be added or removed, and the plot can easily be exported to high-quality PDF format with a single click.

The concept of dynamically offering server-based statistical tools by use of web applications can easily be extended to any other method that has a clear input–output form. Moving toward web-based applications is a clear trend in many fields, both commercial and scientific. The main advantage above classical client-based software is that new or improved applications can be made directly available to a wide audience. Furthermore, web applications can easily be connected to existing services and databases, which makes integration of new statistical tools into for example diagnostics a small step for the user. Server-based systems make by design more efficient use of resources and are easier to maintain than client-based software, which makes them less expensive.

7. CONCLUSION

The stage line diagram is a highly useful instrument for tracking development processes over time. The diagram itself can be used to monitor the development in both individuals and groups. The SDS can be used as a continuous early/late score that corrects for age. Both the diagram and the SDS ease the quantitative analysis of developmental processes.

The process to create the stage line diagram is straightforward. First, we model the probabilities of successive category transitions in the reference data as functions that are smooth in age. Next, we calculate the mid- P value for each category, and transform the mid- P values into the Z -scale by a probit transformation. The resulting scores can be plotted against age to produce the stage line diagram.

We calculated references by a series of generalized additive models on the transition probabilities. Our methodology is not limited to this class. In principle, one could apply any method for ordered data that produces probability estimates that vary smoothly in age. An overview of such models can be found in Clogg and Shihadeh [23]. The data requirements needed to produce a stage line diagram are relatively low.

The interpretation of the mid- P value as an optimal quantitative value is new. The key to the new diagram is the use of the mid- P value to assign scores to categories. The argument for using the mid- P relies on the property that P -values are uniformly distributed if the model fits the reference population. Consequently, the middle of the interval, i.e. the mid- P value, is the best one-number summary. It would be interesting to study the relationships between the mid- P value and other methods for assigning optimal scores to categories [24, 25].

The stage line diagram expresses status and tempo of discrete changes on a continuous scale. The diagram provides quick insight into both status (in SDS) and tempo (in SDS/year) of maturation. Reference standards for tempo do not yet exist, but could play a role in discriminating between a rapid and slow succession of states. Such standards are likely to depend on the age of the first transition.

Maturation scores are of interest in their own right. One common approach to compare groups is to calculate the difference between the average stage numbers in each group. Another common approach is to compare the average time spent in each stage between groups. Both approaches are highly problematic if the study groups are observed at different ages, which is often the case. The maturation score takes age into account and is straightforward to calculate by equation (6).

We expect that maturation scores will replace stage numbers, ages and time spent in a stage as the primary outcome of developmental delay. Wider application of maturation measures opens up new analytic possibilities.

REFERENCES

1. Borghi E, de Onis M, Garza C, Van den Broeck J, Frongillo EA, Grummer-Strawn L, Van Buuren S, Pan H, Molinari L, Martorell R, Onyango AW, Martines JC for the WHO Multicentre Growth Reference Study Group. Construction of the World Health Organization child growth standards: selection of methods for attained growth curves. *Statistics in Medicine* 2006; **25**(2):247–265.
2. Van Buuren S. Growth references. In *Growth Disorders*, Kelnar C, Savage M, Saenger P, Cowell C (eds). Hodder Arnold: London, 2007; 165–181.
3. Jacobusse G, Van Buuren S, Verkerk PH. An interval scale for development of children aged 0–2 years. *Statistics in Medicine* 2006; **25**(13):2272–2283.
4. Aitchison J, Silvey SD. The generalization of probit analysis to the case of multiple responses. *Biometrika* 1957; **44**:131–140.
5. Wade AM, Lawrence K, Mandy W, Skuse D. Charting the development of emotion recognition from 6 years of age. *Journal of Applied Statistics* 2006; **33**(3):297–315.
6. Wade AM, Salt AT, Proffitt RV, Heavens SJ, Sonksen PM. Likelihood-based modelling of age-related normal ranges for ordinal measurements: changes in visual acuity through early childhood. *Statistics in Medicine* 2004; **23**(23):3623–3640.
7. Wade AM, Ades AE, Salt AT, Jayatun R, Sonksen PM. Age-related standards for ordinal data: modelling the changes in visual acuity from 2 to 9 years of age. *Statistics in Medicine* 1995; **14**(3):257–266.
8. Royston P. A parametric model for ordinal response data, with application to estimating age-specific reference intervals. *Biostatistics* 2000; **1**(3):263–277.
9. Marshall WA, Tanner JM. Variations in pattern of pubertal changes in girls. *Archives of Diseases in Childhood* 1969; **44**:291–303.
10. Coleman L, Coleman J. The measurement of puberty: a review. *Journal of Adolescence* 2002; **25**(2):535–550.
11. Hastie TJ, Tibshirani RJ. *Generalized Additive Models*. Chapman & Hall: London, 1990.
12. Van Buuren S, Fredriks M. Worm plot: a simple diagnostic device for modelling growth reference curves. *Statistics in Medicine* 2001; **20**(8):1259–1277.
13. Royston P, Wright EM. Goodness of fit statistics for age specific reference intervals. *Statistics in Medicine* 2000; **19**:2943–2962.
14. Lancaster HO. The combination of probabilities arising from data in discrete distributions. *Biometrika* 1949; **36**:370–382.
15. Berger VW. Mid-*P* values. In *Encyclopedia of Statistics in Behavioral Science*, Everitt B, Howell D (eds). Wiley: Chichester, 2005; 1221–1223.
16. Curry RE. *Estimation and Control with Quantized Measurements*. MIT Press: Cambridge, MA, 1970.
17. Efron B, Tibshirani RJ. *An Introduction to the Bootstrap*. Chapman & Hall/CRC: London, 1994.
18. Fredriks AM, Van Buuren S, Burgmeijer RJF, Meulmeester JF, Beuker RJ, Brugman E, Roede MJ, Verloove-Vanhorick SP, Wit JM. Continuing positive secular growth change in the Netherlands 1955–1997. *Pediatric Research* 2000; **47**(3):316–323.
19. Mul D, Fredriks AM, Van Buuren S, Oostdijk W, Verloove-Vanhorick SP, Wit JM. Pubertal development in the Netherlands 1965–1997. *Pediatric Research* 2001; **50**(4):479–486.
20. Cox DR. Note on grouping. *Journal of the American Statistical Association* 1957; **52**:543–547.
21. Wikipedia TFE. *Ajax (programming)*, Online resource: [http://en.wikipedia.org/w/index.php?title=Ajax_\(programming\)&oldid=263271982](http://en.wikipedia.org/w/index.php?title=Ajax_(programming)&oldid=263271982), 2009.
22. Horner J. *rapache: Web application development with R and Apache*, Online resource: <http://biostat.mc.vanderbilt.edu/rapache/>, 2009.
23. Clogg CC, Shihadeh ES. *Statistical Models for Ordinal Variables*. Sage: Thousand Oaks, 1994.
24. Healy MJR, Goldstein H. An approach to the scaling of categorized attributes. *Biometrika* 1976; **63**(2):219–229.
25. Gifi A. *Nonlinear Multivariate Analysis*. Wiley: New York, 1990.