Statistical Analyses of Fingerprint Growth

Thomas Hotz*[1], Carsten Gottschlich*[1], Robert Lorenz[2],
Stefanie Bernhardt[2], Michael Hantschel[2], Axel Munk[1]

[1]Institute for Mathematical Stochastics, University of Göttingen,
Goldschmidtstrasse 7, 37077 Göttingen, Germany
{hotz,gottschlich,munk}@math.uni-goettingen.de

[2]Bundeskriminalamt, Wiesbaden, Germany

Abstract: We study the effect of growth on the fingerprints of adolescents, based on which we suggest a simple method to adjust for growth when trying to retrieve an adolescent's fingerprint in a database years later. Here, we focus on the statistical analyses used to determine how fingerprints grow: Procrustes analysis allows us to establish that fingerprints grow isotropically, an appropriate mixed effects model shows that fingerprints essentially grow proportionally to body height. The resulting growth model is validated by showing that it brings points of interest as close as if both fingerprints were taken from an adult. Further details on this study, in particular results when applying our growth model in verification and identification tests, can be found in *C. Gottschlich, T. Hotz, R. Lorenz, S. Bernhardt, M. Hantschel and A. Munk:* Modeling the Growth of Fingerprints Improves Matching for Adolescents*, IEEE Transations on Information Forensics and Security, 2011 (to appear).*

1 Introduction

Consider the following scenario: an adolescent at age 11, say, gives his fingerprints which are entered into an automatic fingerprint identification system (AFIS); later, at age 30, his fingerprints are again taken, and run against the AFIS database. To find the adolescent fingerprint matching the adult one is made difficult by the fact that the adolescent has grown into an adult – as have his fingerprints, compare Figure 1. As these systems are usually engineered for adults, growth effects are not taken into account appropriately, and the AFIS will decide that the adolescent print matches the adult print poorly; indeed, the points of interest (POI), i.e. minutiae and singular points, cannot be brought close by merely rotating and translating the imprints, see Figure 1 (right).

This study aims at determining how fingerprints grow, such that the effects of growth can efficiently be taken into account and corrected for. For more background on the political implications of this question, as well as on the lack of research on fingerprint growth so far, see the complete study report published in [GHL$^+$11]. Here, we focus on the statistical

*T. Hotz acknowledges support by DFG CRC 803, C. Gottschlich by DFG RTS 1023.

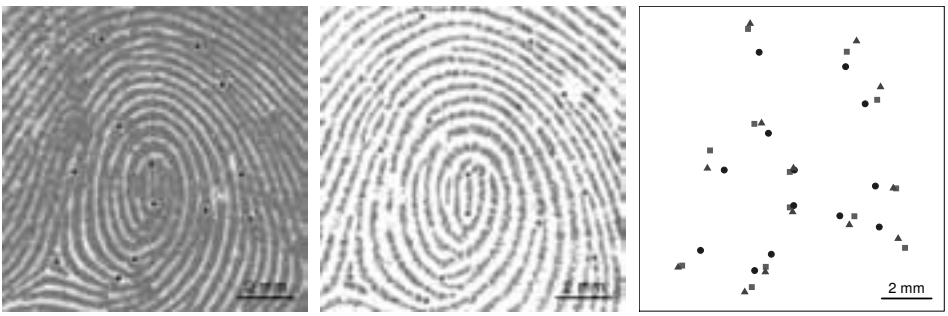

Figure 1: Earliest (left) and latest (middle) rolled imprints of right index finger of person 28 with points of interest (POI) marked by a human expert. Superimposed POI (right) of earliest rolled imprint without rescaling (blue circles), with rescaling (purple triangles), each brought into optimal position w.r.t. the POI of latest rolled imprint (red squares).

analyses that allow us to model fingerprint growth, and to validate the proposed model.

In [GHL+11] the authors also describe the dataset which was used for this study. It consisted of 48 persons (35 male, 13 female) whose fingerprints were taken between 2 and 48 times. At the time when the first fingerprint was taken participants were between 6 and 15 (median 12) years old, when the last fingerprint was taken they were between 17 and 34 (median 25) years old. We thus had longitudinal datasets of 48 persons' fingerprints through their adolescence at our disposal, in order to answer the question of how fingerprints grow.

To be able to measure how growth affects fingerprints, for each person a human expert selected one finger whose imprints were of high quality, showing many POI in all its imprints; the human expert then marked all corresponding POI in this finger's imprints.

The remainder of this article is organised as follows: in Section 2, Procrustes analysis is employed to determine whether fingerprints grow isotropically in all directions. The affirmative answer to that question allows to model fingerprint growth by rescaling, the factor of which may be obtained from growth charts of body height, as detailed in Section 3. This model allows to effectively predict growth of fingerprints such that adolescent fingerprints can be transformed into fingerprints which can be treated like adult fingerprints as shown in Section 4. Finally, we summarise and discuss the results obtained in Section 5. Further information on this study, including on the model's performance in practice, can be found in [GHL+11].

2 Screening for anisotropic growth effects

It was to be expected that fingerprints grow in size. However, it was unclear whether they grow isotropically, i.e. equally in all directions, which would result in a simple uniform scale-up of the fingerprint. Indeed, it is known that finger bones tend to grow more in length than in width, becoming relatively narrower [Mar70].

To address the question whether there are any anisotropic growth effects, we employed techniques from the field of shape analysis (see e.g. [DM98] for more details) where one considers objects only up to translation, rotation and isotropic rescaling, i.e. up to similarity transformations. In particular, we used *full Procrustes analysis*: for this, let us consider the POI of person i, fixed for the moment, marked in the jth rolled imprint of that person's selected finger. Assuming that n POI have been marked in each of the J imprints, we can represent their coordinates as an n-dimensional complex vector $z_{ij} \in \mathbf{C}^n$.

Next, we aim to define a mean point configuration for that finger, independent of translations, rotations and rescaling. Hence, we look for some $\mu_i \in \mathbf{C}^n$ which minimises the distances to the z_{ij} if the latter are allowed to be translated, rotated and rescaled. Clearly, one should fix μ_i's size and centre of mass, i.e. we require that $\mu_i \in S^n = \{z \in \mathbf{C}^n \,:\, \mathbf{1}_n^* z = 0, \|z\| = 1\}$ where $\mathbf{1}_n$ represents the column vector comprising n ones, and $\|\cdot\|$ denotes the usual Euclidean norm.

Note that rotating and rescaling POI $z \in \mathbf{C}^n$ can be described by multiplication by some $\lambda \in \mathbf{C}^* = \mathbf{C} \setminus \{0\}$, whereas translation is given by adding $\tau \mathbf{1}_n$ for some $\tau \in \mathbf{C}$. Hence, a reasonable choice for μ_i is

$$\mu_i = \operatorname*{argmin}_{\mu \in S^n} \sum_{j=1}^{J} \min_{\lambda \in \mathbf{C}^*, \tau \in \mathbf{C}} \|\lambda z_{ij} + \tau \mathbf{1}_n - \mu\|^2. \tag{1}$$

μ_i is called a *Procrustes mean* of the z_{ij}; it is usually unique up to a rotation which we assume to be chosen arbitrarily. Let us denote the respective minimisers on the right by λ_{ij} and τ_{ij}; they represent the rescaling, rotation and translation bringing z_{ij} into *optimal position* to μ_i. Then,

$$\hat{z}_{ij} = \lambda_{ij} z_{ij} + \tau_{ij} \mathbf{1}_n$$

is rescaled, rotated and translated to optimally match μ_i with respect to (w.r.t.) the Euclidean norm; one might also say that the $\hat{z}_{ij}$ are registered w.r.t. the Procrustes mean. Figure 2 (left) shows the μ_i and the $\hat{z}_{ij}$ for the right index finger of person 28.

Now that rescaling, rotation and translation have been fixed, one can apply methods of classical multivariate data analysis to the $\hat{z}_{ij}$; note however, that the registration – and in fact even the space's dimension – depends on the finger, i.e. on i. We will thus first perform intra-finger analyses which will then get aggregated over the study population; recall that we chose one finger per person.

Assume for a moment that there was an anisotropic growth effect; then, we would expect to see a trend in the $\hat{z}_{ij}$, the POI moving in a preferred direction as time progresses. This growth effect should strongly correlate with the fingerprint's size; note that time is not a good predictor since the effect of one year of growth is expected to be larger for younger persons – for adults there should not be any growth related effects any more.

We therefore need a measure of size of a person's fingerprint. With the data at hand, this can only be an intra-finger measure of size which will not enable us to compare different finger's sizes, but this will suffice for our intra-finger analysis. As a measure of person i's fingerprint j's *size*, we chose

$$S_{ij} = \|z_{ij} - \tfrac{1}{n} \mathbf{1}_n \mathbf{1}_n^* z_{ij}\|, \tag{2}$$

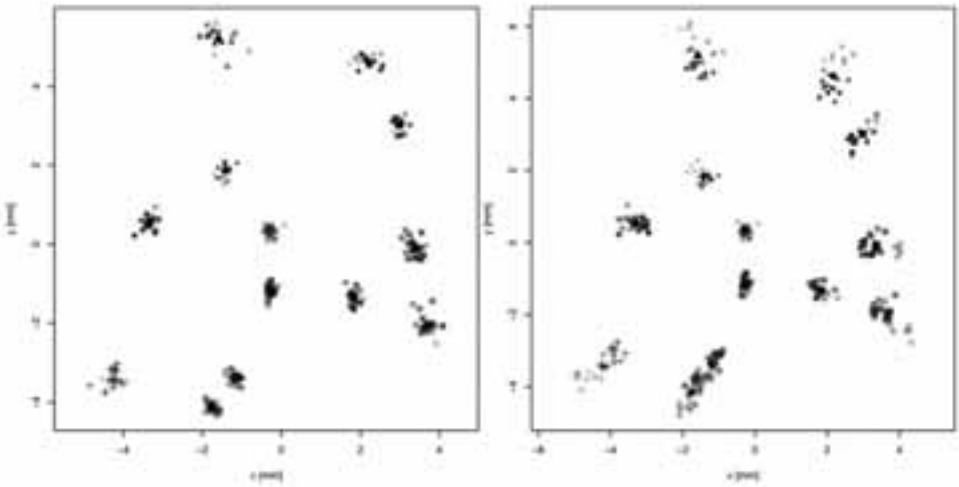

Figure 2: Full (left) and partial (right) generalised Procrustes analysis of POI for the right index finger of person 28: colour codes age (young: dark blue; old: light red), larger black symbols shows the Procrustes mean; different symbols represent different POI.

i.e. S_{ij}^2 is the sum of squared distances of the imprint's POI from their respective centre of mass; one might also call this the spread of the POI.

Using the S_{ij} for fixed i as a linear predictor for the z_{ij} in a multivariate linear model, we expect to be able to explain much of the z_{ij}'s total variance should there be an important anisotropic effect. Indeed, a tendency of change for the POI pattern should give rise to a linear trend that correlates strongly with time and hence with size. For person 28, we were however only able to explain 5% of the total variance, aggregating over all persons, the median was 16% of the total variance explained by size.

This should be compared with the maximum amount a single linear predictor could have explained. The single best linear predictor is given by the first principal component (PC) of the z_{ij}. In the case of person 28, it could explain 45% of the total variance. However, this first PC is not at all correlated with size, see Figure 3 (left). Again averaging over all persons resulted in a median 51% of the total variance explained by the first PC.

Should there be any anisotropic effect, we thus may, for practical purposes, safely assume it to be negligible. To show how large the isotropic growth effect is, we repeated the entire analysis without allowing for rescaling, i.e. restricting λ in equation (1) to fulfil $|\lambda| = 1$, representing a pure rotation, considering

$$\min_{\lambda \in \mathbf{C}^*, |\lambda|=1, \tau \in \mathbf{C}} \|\lambda z_{ij} + \tau \mathbf{1}_n - \mu\|^2. \tag{3}$$

Continuing as before results in what is called *partial Procrustes analysis*.

The isotropic growth effect is then well visible in the partially registered POI of person 28, see Figure 2 (right). Also, the first PC then strongly correlates with size, as is visible from Figure 3 (right). Indeed, averaging over the population, size explains a median 58% of the

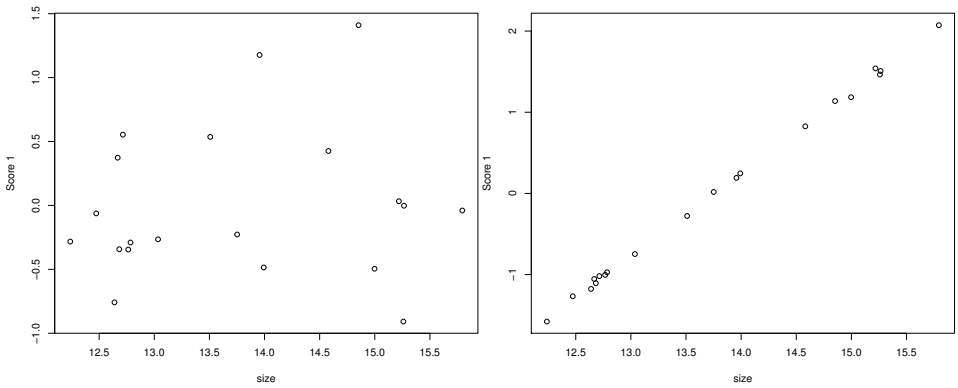

Figure 3: Full (left) and partial (right) generalised Procrustes analysis of POI for the right index finger of person 28: scores (i.e. coordinates in direction) of the respective first PC (vertical axis) vs size (horizontal axis).

total variation. Also, the first PC captures much of this trend, explaining a median 64% of the variation; thus size is close to explaining as much of the total variation as any linear predictor, if we do not correct for isotropic growth effects.

3 Modeling fingerprint growth

From the last section, we may assume that, essentially, fingerprints grow isotropically. What remains to be determined are the factors by which fingerprints get enlarged during growth. Since it is well known that the lengths of upper limbs and body height exhibit a strong correlation during growth [McC70, HYM77], we decided to utilise growth charts of boys' and girls' body heights, in particular those given in [KOG$^+$02].

If for two fingerprints to be compared we only know the person's age and sex at the time the imprints were taken, we cannot hope to be able to predict that person's individual growth – e.g. we do not know whether some imprint was taken immediately before or after her pubertal growth spurt. Instead we must resort to using the amount a "typical" person would have grown over the same period. We therefore use growth charts of boys' and girls' *median* body height, given in [KOG$^+$02].

This gives a simple model for fingerprint growth: for person i and imprint j taken at age x_j, look up the population's median body height at age x_j of the corresponding sex; let us denote that by G_{ij}. If later imprint j' is taken, we propose to predict how fingerprint j has grown in the meantime by simply enlarging it by a factor of $G_{ij'}/G_{ij}$, i.e. by the factor the "median person" would have grown in body height over the same period.

To check whether growth charts for body height can in fact be used for growth of fingerprints, we compared these to the measure of fingerprint size defined in equation (2). Recall that this is an intra-finger measure of size, lacking an absolute scale. For person 28,

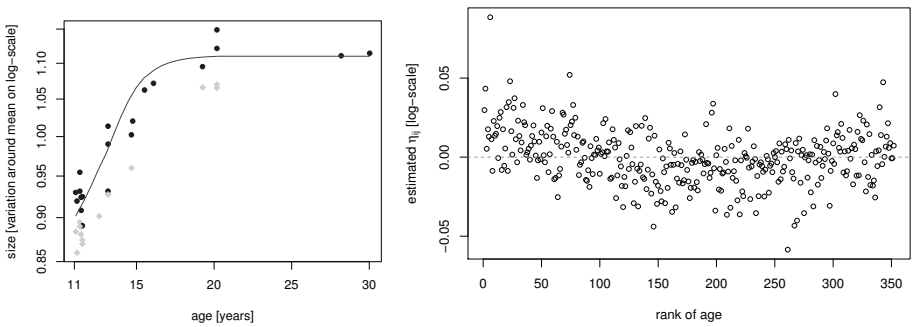

Figure 4: Left: imprints' sizes (blue circles: rolled, green diamonds: plain control) of person 28's right index finger, as well as fitted growth according to the growth chart (purple line); right: estimated η_{ij} of the mixed effects model on a log-scale vs rank of age.

we show its fingerprints' sizes S_{ij} divided by their geometric mean (the arithmetic mean on the log-scale) in Figure 4 (left). Plotted are the rolled imprints' sizes as blue circles; at times, additional plain imprints had been taken as controls at the same occasion, their sizes are plotted as green diamonds for comparison. Similarly, the G_{ij}, the corresponding median body heights, were divided by their geometric mean and plotted in violet – in fact, the entire normalised growth curve for that period is shown.

The growth curve in Figure 4 (left) is seen to fit person 28's fingerprint growth very well. To check how good growth charts for body height can predict fingerprint growth in the entire study population, we used the following linear mixed effects model:

$$\log S_{ijk} - \log G_{ij} = \rho_k + \nu_i + \eta_{ij} + \varepsilon_{ijk},$$

where the additional index k serves to distinguish between rolled imprints and plain controls. Since we are only interested in relative increases in size, the analysis works on a log-scale. The outcome on the left is given by the log-ratio of the imprint's observed size S_{ijk} and the corresponding median body height G_{ij}. Since size is an intra-finger measure, there cannot be a universal factor of proportionality between our fingerpint size and body height, thus ν_i on the right allows for an individual factor of proportionality, fixed for that finger. Even if two imprints are taken at the same point in time, like the rolled and the plain control imprint, there may be variations in the locations of the POI, due to distortions arising from the skin being pressed on a flat surface, and also due to variations in the process of marking the POI. These random effects are captured by the ε_{ijk} which we assume to be independent and to follow a Gaussian distribution with mean 0 and variance σ_ε^2. Systematic differences between rolled imprints and plain controls are taken into account by ρ_k; note how in Figure 4 (left) the plain controls in green appear systematically smaller than the rolled imprints in blue. Finally, η_{ij} models deviations of person i at the time when imprint j was taken from the growth model; the η_{ij} are again assumed to be random – since persons and times were random as far as the analysis is concerned – and independently Gaussian distributed with mean zero and variance σ_η^2.

The standard deviation σ_ε of the ε_{ijk} represents how much a fingerprint's size appears

to vary between two imprints taken at the same age; this is the measurement noise any AFIS has to cope with. On the other hand, the model misfit is given by the η_{ij}s' standard deviation σ_η. Comparing these two enables us to determine how well the proposed growth model fits the data in relation to the noise level. By maximum likelihood, see e.g. [PB00], we estimated them to be about $\hat{\sigma}_\varepsilon = 0.0225$ and $\hat{\sigma}_\eta = 0.0223$ on the log-scale, i.e. both are of similar magnitude; the difference between the fixed effects ρ_k for rolled imprints and plain controls was estimated as 0.0242. Bringing them back to the original scale, we obtain variations in size due to noise of about $\pm(\exp(\hat{\sigma}_\varepsilon)-1) = \pm 2.28\%$, similar to the variations due to model misfit of about $\pm(\exp(\hat{\sigma}_\eta)-1) = \pm 2.26\%$; for the difference between rolled ink and plain control we get 2.44% larger rolled imprints than plain controls. This analysis used a total of 578 imprints (combinations of i, j, and k), taken on 352 occasions (pairs of i and j) from 48 persons (i). For a visual inspection, we show the estimated η_{ij} sorted by age in Figure 4 (right); systematic deviations from the growth model would be expected to result in regions where the η_{ij} are driven away from the horizontal line; however, no large systematic deviations in comparison to the random noise are discernible.

4 Validating the growth model

The two previous sections established that fingerprints grow isotropically, and proportionally to body height. This led to the proposed strategy of enlarging the younger imprint prior to comparing it with an older imprint, by a factor to be read off from growth charts of boys' and girls' body height.

To validate this model for use in fingerprint matching, we determine how closely POI can be registered when the younger imprint is rescaled, in comparison to when it is left unscaled. To do so, we chose to compare the earliest ($j = 1$) and the latest ($j = J$) rolled imprint for each person i, and determined the square root of the POI's average squared Euclidean distance after bringing the last rolled imprint into partially optimal position by rotation and translation,

$$\delta_i(\mu) = \sqrt{\min_{\lambda\in\mathbf{C}^*,|\lambda|=1,\tau\in\mathbf{C}} \tfrac{1}{n}\|\lambda z_{iJ} + \tau\mathbf{1}_n - \mu\|^2}.$$

cf. equation (3); we determine the distance of the last rolled imprint to the first one with $\mu = z_{i1}$ without rescaling; with rescaling, we use $\mu = \alpha z_{i1}$ where the scale factor is given by $\alpha = G_{iJ}/G_{i1}$; in order to see typical variations when the same imprint is taken twice on the same occasion, we also use the corresponding last plain control for μ (if available). Note that this distance does not depend on which of the two imprints was fixed and which was allowed to be rotated and translated, since λ is required to fulfil $|\lambda| = 1$.

The mean distances thus obtained per person are shown in Figure 5 where they are represented across persons by the respective cumulative distribution functions. The median over the mean distances without rescaling is found to be about 0.62 mm, which is approximately halved when growth is taken into account by rescaling, resulting in a median of 0.33 mm. In fact, this is the same order of magnitude as the median mean distance of 0.31 mm when comparing with the plain control; see [GHL$^+$11] for box-plots conveying

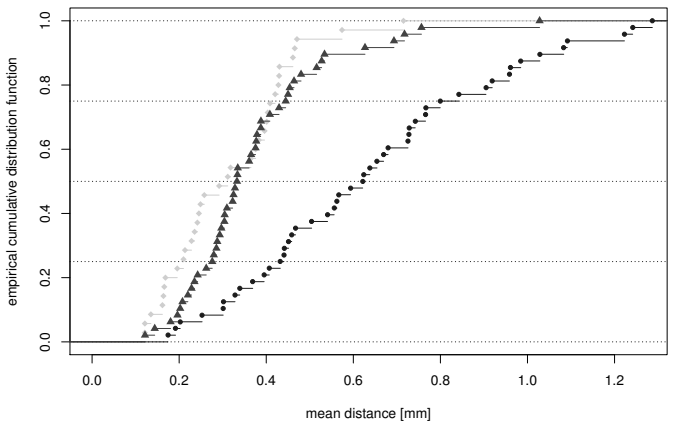

Figure 5: Empirical cumulative distribution functions of mean distance from POI of the latest rolled imprint in partially optimal position to the POI of the first rolled imprint without (blue circles) and with (violet triangles) rescaling, or to latest plain control (green diamonds); horizontal dashed lines correspond to quartiles.

the same information. There, we also plot the relative improvement against the predicted factor, showing that they correlate strongly with Spearman's rank correlation being 0.63. This clearly demonstrates that one gains the more by rescaling the more growth is predicted, as was to be expected.

We conclude that the proposed model transforms the younger fingerprint to match the size of the older one, allowing to compare the two as if they were taken at the same age.

5 Summary and Outlook

Using extensive statistical analyses, we have demonstrated that fingerprints essentially grow isotropically by an amount proportional to the growth in body height. Furthermore, the latter can be read off population-based growth charts of boys and girls, whence the growth model could be successfully constructed without using any training data. We employed `R` [R D09] for all statistical analyses; in particular, the Procrustes analysis was carried out using package `shapes` [Dry09], the mixed effects model using package `nlme` [PBD+09]. The findings thus obtained have been corroborated by a series of verification and identification tests which are described in detail in [GHL+11].

For the future, we plan to repeat the above statistical analyses on an independent data set, testing the hypothesis that rescaling as proposed leads to a reduction in minutiae distances and improves upon error rates obtained without rescaling.

Since most of the participants in this study were more than 11 years old, the results obtained only apply to adolescents from that age onwards. It would be most interesting to

perform similar analyses using children's fingerprints; indeed, the different way in which younger children develop may also lead to anisotropic effects. Another point worth investigating is whether the growth predictions become more accurate when measurements of the person's true body height can be used instead of the population's median body height at that age. Otherwise, noting that the growth charts [KOG+02] we used were based on United States National Health Examination Surveys, it is conceivable that one might benefit from growth charts which are specific for the population under consideration.

References

[DM98] I. L. Dryden and K. V. Mardia. *Statistical Shape Analysis*. Wiley, Chichester, 1998.

[Dry09] I. L. Dryden. shapes: statistical shape analysis, 2009. R package version 1.1-3, `http://CRAN.R-project.org/package=shapes`.

[GHL+11] C. Gottschlich, T. Hotz, R. Lorenz, S. Bernhardt, M. Hantschel, and A. Munk. Modeling the Growth of Fingerprints Improves Matching for Adolescents. *IEEE Transations on Information Forensics and Security*, 2011. To appear, early access available at `http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5751684`.

[HYM77] J. H. Himes, C. Yarbrough, and R. Martorell. Estimation of stature in children from radiographically determined metacarpal length. *J. Forensic Sci.*, 22(2):452–455, 1977.

[KOG+02] R. J. Kuczmarski, C. L. Ogden, S.S. Guo, et al. *2000 CDC Growth Charts for the United States: Methods and Development*, volume 11 of *Vital and Health Statistics*. National Center for Health Statistics, May 2002.

[Mar70] M. M. Maresh. Measurements from roentgenograms, heart size, long bone lengths, bone, muscle and fat widths, skeletal maturation. In McCammon [McC70], pages 155–200.

[McC70] R. W. McCammon, editor. *Human growth and development*. Charles C. Thomas, Springfield, IL, 1970.

[PB00] J. C. Pinheiro and D. M. Bates. *Mixed-effects models in S and S-PLUS*. Springer, New York, NY, 2000.

[PBD+09] J. C. Pinheiro, D. M. Bates, S. DebRoy, D. Sarkar, and R Development Core Team. nlme: linear and nonlinear mixed effects models, 2009. R package version 3.1-96, `http://CRAN.R-project.org/package=nlme`.

[R D09] R Development Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing, 2009. `http://www.R-project.org`.