Oslo
universitetssykehus

BioAlder: A tool for using biological tests to assess the age of unaccompanied minor asylum-seekers

Department of Forensic Sciences| Division of Laboratory Medicine | Oslo University Hospital

Table of contents

2

1   Preface

With effect from 2016, the national scientific responsibility for assessing the age of unaccompanied minor asylum-seekers in Norway has rested with the Department of Forensic Sciences at the Norwegian Institute of Public Health (now the Department of Forensic Sciences at Oslo University Hospital). Since then, a project group has been established, and work to accomplish the assignment has involved making systematic reviews [1-5], international publications [6-9] and searching for new and improved methods.

An external reference group was established in December 2016 and held its first ordinary meeting in February 2017. The group has had the opportunity along the way to provide general input into the work we have done, and the following organisations are represented: The Norwegian Organisation for Asylum Seekers (NOAS), Save the Children (Norway), the Norwegian Psychological Association, the Norwegian Dental Association, the Centre for Medical Ethics at the University of Oslo (UiO), the Norwegian Society of Paediatricians and the Norwegian Society of Paediatric Radiology.

This manual presents the tool BioAlder and describes the work of constructing the tool, which is designed to estimate prediction intervals for age based on radiographs of a wisdom (third molar) tooth and hand skeleton (also referred to as hand-wrist or only hand). The work has been carried out by the Age assessment research group at the Department of Forensic Sciences, Division of Laboratory Medicine, Oslo University Hospital (OUH).

We should like to thank the Norwegian Knowledge Centre for the Health Services for their cooperation on the systematic reviews and Thore Egeland (Norwegian University of Life Sciences/OUH) and Torbjørn Wisløff (UiO/Norwegian Institute of Public Health) for cooperation on the mathematical modelling of data. Thanks also to Jayakumar Jayaraman, Simon Camilleri, Rick R. van Rijn, Eugénia Cunha, Abdul Mueed Zafar, Bernhard Knell, Ivan Galić, Antoine Saadé, Sevcihan Yilmaz, Carolina Medina-Gomez, Sapna Hegde and Miriam Marrero for submitting datasets. We should also like to thank Lil-Sofie Ording Müller for valuable input to improve the updated version of this manual.

2   What is new in BioAlder version 1.3b?

Since the publication of the last version of the manual, we have received comments and suggestions on how to improve BioAlder. Among these are feedback on the way we present and explain the system in this manual, pointing out that it can be too technical and difficult to understand. We have taken these comments into account in this version of the BioAlder-manual. Parts of the text has been simplified and technical parts are moved to the appendix

- Several of the figures are updated and the figure legends have been changed accordingly
- We have tried different a priori age distributions as a part of the modelling and discussed further in the manual and the appendix.
- We have included a more detailed explanation for some of the methodological choices.

3 Main points

- None of the methods currently in use for assessing biological age can determine the exact age of a person, and there is great variation in how the methods are practised and interpreted in different countries.

- BioAlder, the age assessment tool described in this document, makes an automated prediction of chronological age based on results from radiographs of the wisdom tooth and hand skeleton.

- The tool has been developed by OUH to assist the Norwegian Directorate of Immigration in determining the ages of young asylum-seekers. To the best of our knowledge, this tool is the first of its kind worldwide.

- This version of BioAlder is based on research in 34 scientific publications, and includes data from more than 22 000 people. The tool will be updated regularly with new research data.

- All data in the model is truly observed individuals from different studies. Some of these studies have grouped data, meaning that they do not submit data for every individual. Since BioAlder is built with individual data, we have used mathematical modelling to utilize the data from these studies.

- The data include studies conducted in 22 different countries. The significance of regional differences remains unclear.

- The tool is a temporary solution. We are working on development of molecular biological methods (DNA methylation) for future age assessment

6

4 Introduction

Unaccompanied minor asylum-seekers who come to Norway have rights pursuant to Norwegian law and international guidelines and conventions [10]. Their applications must be processed on the best possible basis, amongst other things so that they are accorded the rights that are correct for their age. When there is any doubt about the age of asylum-seekers in Norway, their age is established by the Directorate of Immigration. In most countries, biological methods form an important source of information for determining age [11].

The methods currently in use for determining biological age are unable to provide a precise age [1, 2]. The greatest constraint is the natural biological variation in the development of skeleton and teeth, which are the analytical methods most frequently used in age assessment. Nor are there any scientifically documented systems for the use of psychosocial or cognitive testing to predict chronological age [4].

4.1 About BioAlder

BioAlder has been developed as an aid for determining the age of young, unaccompanied asylum-seekers in cases of doubt. The tool has been constructed as a statistical calculation model on the basis of studies of the development of the hand-wrist skeleton (hand) and lower left wisdom tooth in more than 22000 young persons of known chronological age. BioAlder is used to assess the individual asylum-seeker's developmental stages based on radiographs of the applicant's hand-wrist and teeth, and to compare them with the statistical basis in the model. The model provides an estimate of the applicant's chronological age range. Emphasis is placed on BioAlder being able to present uncertainty in an easily comprehensible manner.

BioAlder has been developed as part of an assignment for Oslo University Hospital (OUH), Department of Forensic Sciences, commissioned by the Norwegian Ministry of Health and Care Services. The assignment is regulated by an agreement between OUH and Directorate of Immigration (UDI). The tool was developed by the Age assessment research group at the Department of Forensic Sciences, Division of Laboratory Medicine, OUH. We should like to thank the Norwegian Knowledge Centre for their cooperation on the systematic review and Thore Egeland (Norwegian University of Life Sciences/OUH) and Torbjørn Wisløff (UiO/Norwegian Institute of Public Health) for cooperation on the mathematical modelling of data.

The tool is based on radiographs of the hand-wrist skeleton (hand) and teeth, which were also components of the system used in Norway before the development of BioAlder. Here, we have

7

selected the best documented methods for staging development, collated all available scientific studies on these stages , and finally constructed a mathematical model has been that makes it possible to combine hand and tooth results. To the best of our knowledge, the system is the first of its kind.

BioAlder has been optimised for assessing the age in asylum cases where there is doubt whether a person is under 18 years of age, and cannot be used indiscriminately in other connections. Some discretionary decisions had to be made, as is the case for any modelling and development of prediction models. In the work on this tool, the primary aim of the discretionary decisions taken was to prevent children being classified as adults, and the secondary aim to prevent adults being classified as children. That is the reason why the tool may not be suitable to determine age in other settings (e.g. in "trafficking" or criminal cases where the posed question might be different).

Four peer reviewed publications have been published during the development: Two are about BioAlder [8, 9] and two are about the systematic reviews on the Greulich & Pyle atlas [6] and the Demirjian grading of the third molar [7], respectively.

BioAlder will be further updated as new scientific publications appear and different versions of the tool may yield somewhat different results for the same developmental stages of hand-wrist skeleton and wisdom teeth.

The tool is the best short-term solution that we have found for the commission assigned to us, but it must be regarded as a temporary solution. In the future, we aim to further develop molecular biological methods of age estimation (DNA methylation). See chapter 10.2 for more information.

5   BioAlder in practice

As the requirements for informed consent is not fulfilled in age assessment, this is not collected from asylum seekers that are eligible for biological age assessment. It is, however, possible for the person to refuse to go through the investigation. In addition, the person must also have had the opportunity to give notification of any chronic diseases, developmental disorders or medication, as those might have implication for the skeletal development and thereby the result from BioAlder.

A report based on gender, estimated Greulich & Pyle skeletal age and/or estimated Demirjian's stage of the lower left wisdom tooth will be delivered to UDI after each new update of the tool.

5.1 Result report generated by the tool

The report that is generated consists of two main parts (Figure 1): an introduction, which explains what the results are based on, and a results section, which shows the various relevant results generated by the tool:

- 75% and 95% prediction intervals for chronological age, given the observed stages.
- Percentages of individuals under the ages of 16, 17 and 18, given the observed stages.

The figures calculated by the tool are presented as follows:

- All prediction intervals for chronological age are given in whole years and months.
    - o Values are rounded off to the nearest whole month.
- All values lower than 5% are reported as "less than 5%".
- All values over 95% are reported as "more than 95%".

9

Figure 1. Example of BioAlder results report.

5.2 Potential and limitations of the tool

The model that generates the results is based on a total of 22941 individuals (11718 boys and 11223 girls). It is important to be aware that the estimates for prediction intervals and percentages under certain age limits were calculated using data based on the individuals included in the tool (see Table 8 and 9). The populations from which many of the unaccompanied, minor asylum-seekers originate from are represented to only a limited extent in the tool's underlying data. Existing research yields no answers regarding the extent to which factors such as regional genetic heterogeneity, nutrition and

health affect the development of skeleton and teeth. However, the possibility that these factors may have substantial effects on certain populations and individuals cannot be excluded (see chapter 7.3). The tool provides a description of probable ages on the basis of the included scientific literature. In other words, the tool does not provide a definite answer for each unaccompanied, minor asylum-seeker who is assessed.

The tool was developed using data from studies on healthy individuals. Disease, medical treatment and nutrition may influence the maturation of the skeleton. How great an influence a disease or disorder may have for the result has not been systematically surveyed in existing studies.

Poor nutrition and a number of diseases will have a negative impact on skeleton maturation, which could lead to a person being assessed as younger than their chronological age in an age determination based on hand radiographs. Medical conditions that may cause precocious skeletal maturation may lead to a person being assessed as older than their chronological age based on hand radiographs. The most common causes of this latter effect in the Western population are overweight/obesity and the use of some medications [12]. A number of rare diseases may also have effects of this nature. These are difficult to detect, even for medical specialists. Some studies suggest that less than 1 per cent of asylum-seekers may have a condition/disease that has a bearing on their biological age assessment [13, 14]. One of BioAlder's strengths is that it is based on two independent observations (one hand stage and one tooth stage) which are influenced by different factors [15].

Should there be any doubt as to whether a medical condition may have influenced the BioAlder results, we recommend that UDI obtain an assessment from a paediatrician on the possibility of disease that might accelerate skeletal maturation. We see the largest discrepancies suggesting accelerated skeletal maturation in the combinations given in Table 1.

Gender	Demirijian's stage	GP skeletal age
Boys	A	18
Boys	A	19
Boys	B	18
Boys	B	19
Boys	C	19
Girls	A	18
Girls	B	18

Table 1. Stage combinations with the largest discrepancies, which may indicate accelerated skeletal maturation.

The reports made by BioAlder for these rare combinations will contain a recommendation that the individual should be examined more closely.

5.3   User training

All those who are to use the tool in case processing must take a training course run by OUH. The course will also be open to others for whom it is of interest. OUH will organise day courses for executive officers as needed. The course will provide insight into the methods used in biological age estimation, research method and understanding of the statistical methods used in BioAlder, including the limitations and eventual discretionary decisions made in the development.

12

6 General information about biological age estimation methods

The biological age assessment systems of various countries are based on different methods [11, 16]. There is also considerable variation in the manner in which the same type of methods are employed. For example, for teeth examinations there are a number of different staging systems [17]. In addition, many operators combine the results of several staging systems into one overall estimate. Therefore, it is difficult to find two European countries that use the same methodology to carry out biological age assessments.

The most commonly used methods are based on studying skeletal and/or dental maturation [11, 16], mostly by taking radiographs of the hand and wrist and a panoramic radiograph (an orthopantomogram, or OPG for short) of all teeth. For most young asylum-seekers who are tested, only the wisdom teeth are not fully developed, and staging of these is therefore most widespread.

The end-stage problem

Hand skeleton and dental maturation ends at a certain age when it is said that development has reached the end stage. In the case of the hand skeleton, this is when all growth zones are closed and signs of maturation end. The dental end-stage is when the root of the wisdom tooth is fully developed and closed. This end-stage will persist for the rest of the individual's life, and commences when the previous stage tapers off. Studies show different ages for these end-stages, and this variation may be a matter of real biological differences, but may also be due to weaknesses in the study design (e.g. age mimicry; see Figure 7). The data for end-stages in boys that we have included in the tool look like this:
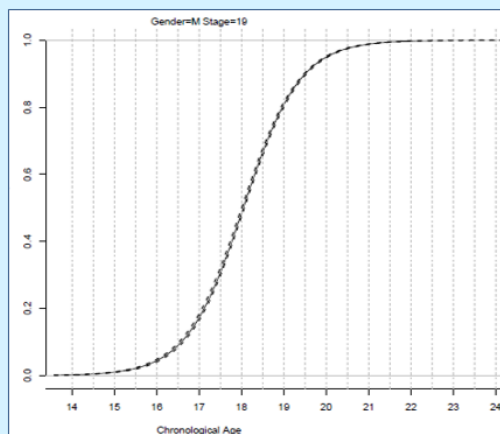
The end-stage for boys in the GP atlas is stage 19 years.
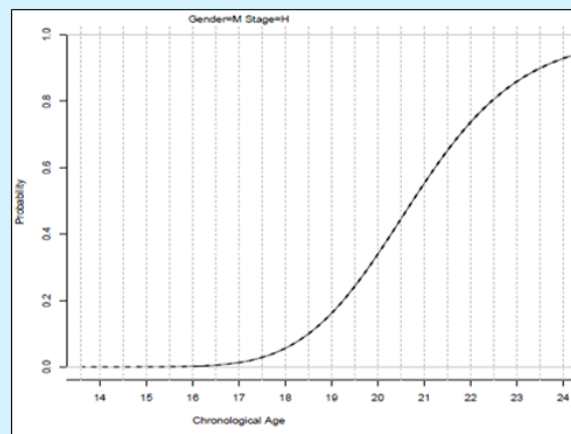
The end stage for Demirjian's staging is H; the curve above is for boys.

If the skeletal age is considered, where the chance of being in the end-stage is 50%, this is about 18 years for boys according to our data, whereas the corresponding age for Demirjian's staging of wisdom teeth is around 21 years. In order to be able to use this stage in the model we have constructed, we have to set an upper age limit. Assumptions of this kind are a necessity for these methods and the tool we have constructed. The reasons for choosing an upper age limit are explained in Chapter 7.2.

Figure 2. The end-stage problem in age estimation based on maturation of hand skeleton and teeth

As both hand-wrist and wisdom teeth are fully developed in the late teens or early twenties, it is difficult to decide whether a person is over or under eighteen years old. This is called *end-stage problem* in this manual, and is explained in detail in Figure 2. Some countries therefore also perform an assessment of bones that mature later [11, 16]. The clavicle, in particular, is frequently imaged. Computed tomography (CT) is usually used for the purpose, but since this involves more radiation than an ordinary X-ray image, and the clavicle is located in an area close to glands and organs, the threshold for conducting this test is higher [13]. There are also limited data on clavicles, particularly with respect to regional differences [3]. Another example is Sweden, where magnetic resonance imaging (MRI) of the knee has been chosen as a basis for determining whether a person is over or under the age of eighteen [18, 19].

A problem common to methods based on the development of skeletons and teeth is that there is substantial variation in natural biological development [1, 2]. This will not vanish even if more research is done on the methods, since the variation is inherent in human biology. Another feature of the development of the hand-wrist skeleton and the third molar teeth is that it stops in the late teens or early twenties (end-stage problem, Figure 2), which presents challenges to making a model for determining whether a person is an adult or a child. Third molar teeth mature later than hand-wrist, meaning it has a later occurring end-stage, and are therefore more suitable as a basis for determining age in the range 17–19 years.

6.1 Age estimation based on radiographs of the hand

When radiographs are taken of the hand and wrist, they can be assessed in relation to a staging system that says something about the development of the skeleton (skeletal age, Figure 3). There are several such systems. In some, like the Greulich & Pyle-atlas (GP-atlas) ,discretionary judgement is used to find the image that is most similar [20], while others are based on scoring predefined bones and their developmental stage and ending up with a total score that gives an age estimate (e.g. the Tanner and Whitehouse methods called TW1, 2 and 3) [21, 22]. The most widely used system, on which there are also most scientific publications, is the GGP-atlas. This is also the staging system for hand/hand wrist that we recommend using for age estimations in Norway.

Figure 3. The Greulich & Pyle-atlas. The grading system used for skeletal development that is recommended for age assessment in Norway, is the Greulich & Pyle-atlas.

The GP atlas was originally developed to determine whether an individual of a known chronological age had skeletal development that was within the normal range. This is described in tables V and VI in the atlas [20]. The tables are structured in such a way that the included individuals are first grouped according to chronological age (for example all boys aged 16) and the results of their estimated skeletal age are reported groupwise as the mean and standard deviation of skeletal age. These results are not directly transferable to a situation where one wants to do the opposite: estimate an unknown chronological age on the basis of skeletal development. In other words, chronological age cannot be read off directly from the GP atlas.

A number of studies have examined the relationship between the developmental stages of the skeleton and chronological age. See in Part 7.1 for further details.

6.2   Age estimation based on dental radiographs

Dental development is estimated by means of various systems on the basis of the development of the crown and roots of teeth. The various staging systems have different numbers of stages, and therefore cannot be compared to each other. Examples of staging systems are Demirjian (Figure 4), Goldstein and Tanner from 1973 (8 stages denoted A to H) [23] and Hunt and Gleiser from 1955 (15 stages) [24]. There are several variations of the latter, such as Moorrees et al., 1963 (14 stages) [25], Haaviko et al., 1970 (12 stages) [26] Kullmann et al., 1992 (7 stages) [27] and Köhler et al., 1994 (10 stages) [28].

15

Demirjian's staging of wisdom tooth formation

Demirjian's staging has the best scientific documentation and has therefore been chosen as the staging system we recommend using in Norway for wisdom teeth. We have chosen to use the left wisdom tooth in the lower jaw (denoted 38), as there is most scientific documentation for this tooth. The figure below shows an outline of Demirjian's tooth formation stages, which are divided into eight stages, A-H, the first four being crown formation stages and the last four root formation stages. The way the method works in practice is that the person assessing the radiographs finds the stage that best describes the wisdom tooth in the radiographs of the individual being assessed.
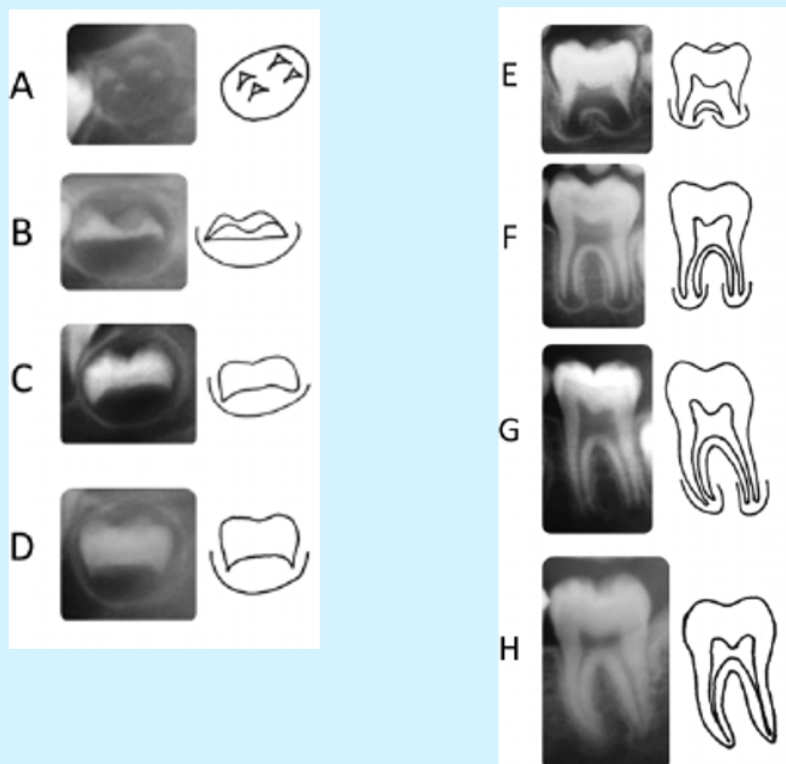
Figure 4. Demirjian's staging of teeth (molars). BioAlder uses Demirjian's grading of teeth in the model.

A number of studies have examined the relationship between the formation stages of wisdom teeth and chronological age. See Part 7.2 for further details.

7   Studies that are used in BioAlder

The studies that are used in the development of BioAlder is based on the GP-atlas for skeletal development of the wrist and Demirjians grading of teeth. The methods and studies were evaluated in each of two systematic reviews in the period from February 2016 to mars 2017, as a collaboration with the Norwegian Institute of Public Health [1, 2]. The systematic reviews resulted in two peer-review publications [6, 7].

16

7.1   Studies on the hand and wrist

March 2017 saw the completion of a systematic review on the use of the GP atlas to estimate age followed by an international publication. Studies in this area normally present their results in one of two ways (Figure 5). Both assume a known chronological age and an observed skeletal age, or GP-stage. The studies can therefore be mainly divided into two categories:

A) **Studies that describe skeletal maturation:** take chronological age as the starting point and present mean and variance of skeletal age (GP-stage) for each age group.

B) **Studies that describe chronological age:** take skeletal age (GP-stage) as the starting point and present mean and variance of chronological age for all individuals in the same skeletal stage collectively.
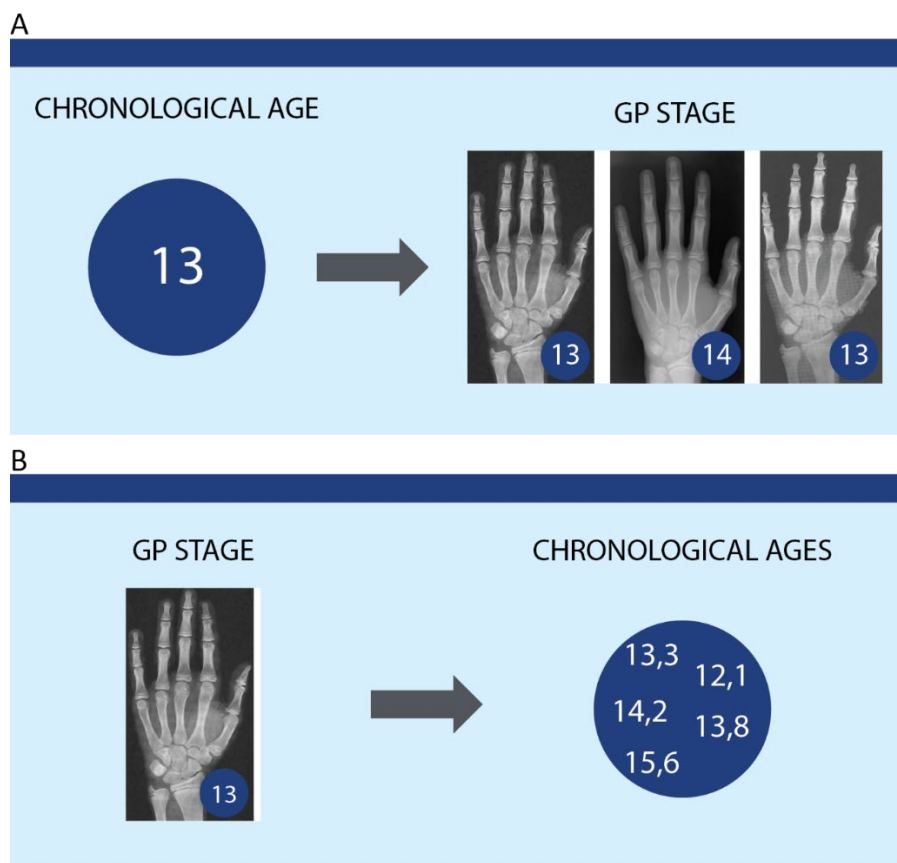
Figure 5. Two different ways of presenting the results in the studies comparing skeletal age and chronological age. A) Studies that group individuals based on chronological age and B) studies that group individuals after the evaluation of skeletal age.

In other words, the two methods of presentation have two different starting points: A groups individuals according to chronological age while B groups them according to the result of estimated skeletal age before the results are presented. As all articles report results groupwise as mean and standard deviation, it is not possible to simply "back-calculate" to a chronological age for each individual in the studies that describe skeletal age. The two approaches are not directly comparable, and two separate analyses were therefore made in the systematic review of the articles that had the two different approaches.

A meta-analysis was performed for the articles with approach A (15 articles). The main finding was that the consistency between skeletal age and chronological age was relatively good for modern populations, where the difference was seldom more than one year at group level.

There were only four articles on approach B. Three of them had an included population that was not evenly distributed age-wise, and when the data is processed, the results reflect this, a phenomenon called age mimicry (Figure 7). We were therefore left with one study using approach B (Chaumoitre 2016, Figure 6) that had more reliable results. This particular study is a relatively large study with an unspecified multi-ethnic population in Marseille. Thus, it is a well-executed modern study of a population of mixed ethnic origin.

Chaumoitre et al. 2016

Chaumoitre et al. (2016) was the only study included in our systematic review whose aim was to show how chronological age was distributed using the pictures in the GP atlas, and had additionally included a population that was relatively uniformly distributed by age. From the illustration below we see the age distribution of the included population of boys. The study included many individuals with non-specified multi-ethnic background. Chronological age in whole years along is given on the x axis and number of individuals on the y axis.
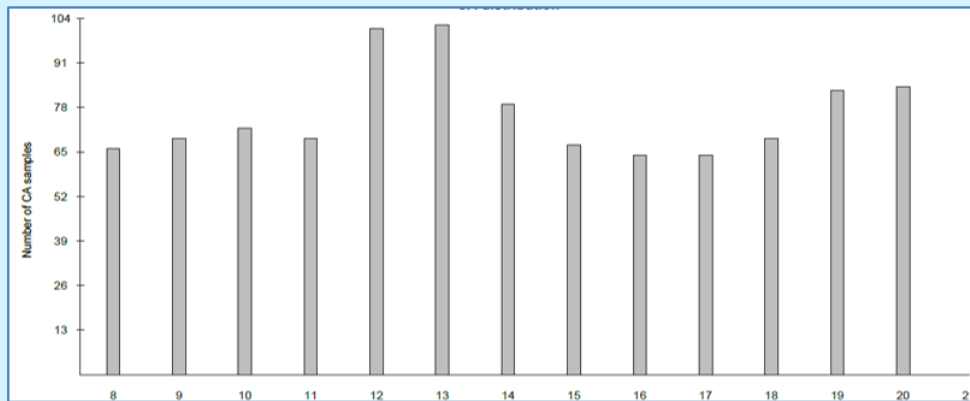
When the included population is large and the age composition even, the results are more reliable for showing directly how chronological age is distributed for each skeletal age. The results of Chaumoitre 2016 are shown in the figure below with the mean (central point), one standard deviation on each side (square brackets) and a 95% confidence interval for the population mean (parentheses). Skeletal ages are given vertically on the left and chronological age horizontally at the top.
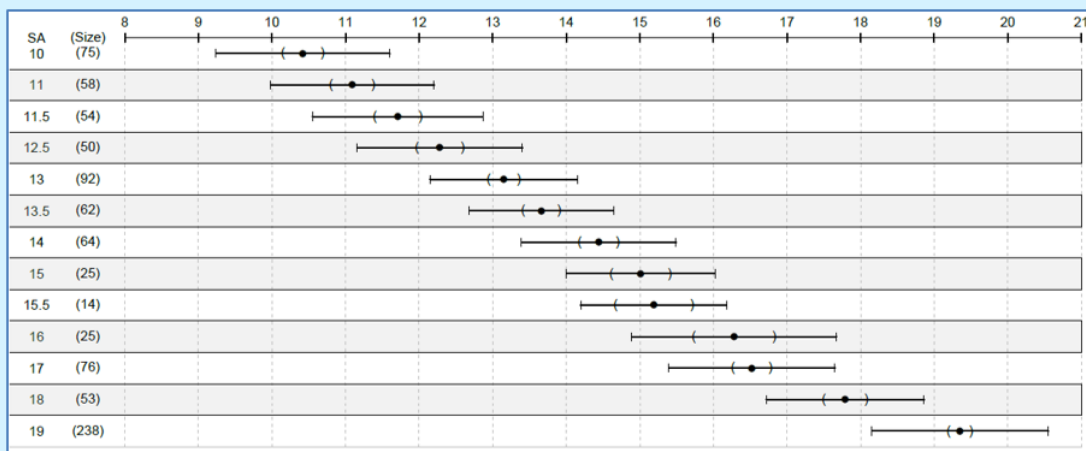
Figure 6. From Chaumoitre et al., 2016 [29]

7.2 Studies on the staging of wisdom teeth formation

In total, we found 21 relevant studies using Demirjian's staging of wisdom teeth, all published after 2005. They were from 15 different countries, spread across all continents except Australia. The studies presented mean age with standard deviation for the participants in each formation stage. The mean chronological age for the different tooth formation stages varied considerably across studies. In the

19

systematic review, we found that the results were strongly affected by the age group selected and the number of individuals in each age group. This bias has previously been described as age mimicry (Figure 7), and the result is that the mean age and standard deviation for each stage strongly reflect the manner in which the participants in each age group were selected, and the age range of the participants. Only a few of the studies were conducted in such a way that they provide an adequate description of the method's ability to estimate age. Because of the differences in the study design, we were unable to combine the studies in a meta-analysis and therefore also unable to reach conclusions as to whether there are differences in the formation of wisdom teeth among populations from different regions.

Age mimicry

On the right is one type of dental study results table, slightly simplified. In this table, four dental stages are labelled with letters at the top, and the ages are given horizontally on the left. The tables themselves show how many individuals from each chronological year were assessed for the four dental studies. We call this presentation a frequency table.

	A	C	E	G
7	1	0	0	0
8	11	2	0	0
9	21	4	0	0
10	22	13	0	0
11	5	37	0	0
12	2	41	1	0
13	1	25	5	0
14	1	11	10	0
15	1	2	13	5
16	0	0	18	9
17	0	0	11	18
18	0	0	7	25
19	0	0	3	17
20	0	0	1	7
21	0	0	0	5
22	0	0	0	4

These results can also be presented as a three-dimensional bar diagram in which the stages are marked with different colours. The height of each bar represent the number of individuals with a certain tooth stage.
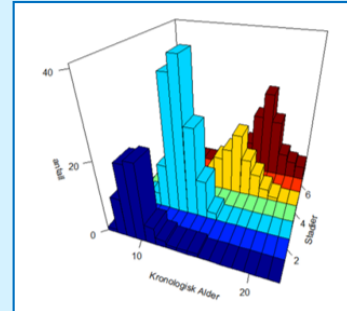
The effect of age mimicry can be demonstrated by changing the number of included individuals of a certain age. If 30 individuals aged eighteen are added, the bar diagram will change for eighteen-year-olds. We see in the diagram that this is strongly reflected in the stages marked yellow and brown. If one publishes these results directly, and gives the mean and standard deviation for each stage, we see how the data 'mimic" the included population. This is why this bias is called "age mimicry".
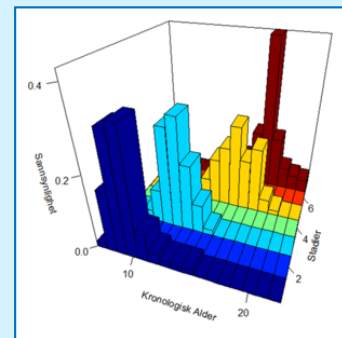
How to solve the problem?

One way to solve age mimicry is to look at the distribution of individuals "the other way round": i.e., not to describe the age for each stage, but rather to describe the distribution of stages on the basis of chronological years. We therefore want to look more closely at the distribution marked in green on the illustration on the right. If we additionally say that the sum of the columns on the right must be one (i.e. we normalise), then this presentation will not be subject to "age mimicry".
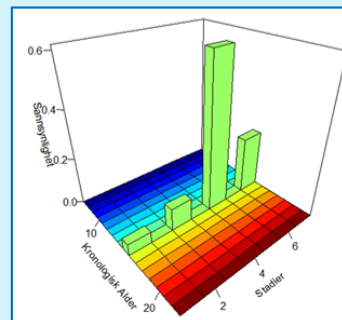
Figure 7. Age mimicry in studies of biological age assessment. The figures explains the concept of age mimicry, where the age distribution of the included study population will affect the mean and standard deviation.

7.3 Regional differences

7.3.1 Development of the hand skeleton

The systematic review of the hand-wrist studies indicates that there may be differences of more than one year between populations from different parts of the world, but that differences that large are rare [1]. A study based on automated measurement of hand-wrist radiographs (BoneXpert software) shows similarly [30], that there may be up to a year's difference on average for studies from different parts of the world. It is not clear, however, whether these variations are due to regional genetic differences, or factors such as dietary and environmental variations. There are also many populations in the world that have not been studied. Mapping of the regional differences would have demanded a very extensive project, and mapping the causes of such differences would be very challenging, as the variation might be the result of multiple and partly unknown factors.

7.3.2 Formation of wisdom teeth

The study material on teeth (Demirjian's staging of wisdom tooth development) is limited, as most of the studies we identified in our systematic review [2] had an included population that was skewed with respect to age, leading to age mimicry (Figure 7) and unreliable results. We therefore ended up with just a few studies that could be used.

The studies Lee 2009 [31], Li 2012 [32] and Johan 2012 [33] have a generally good study design (Table 2). It may be argued that the results of these studies are not representative of other regions and populations. At the same time, we see just as wide a variation among the results of these studies as we find by comparing them with a study from Botswana, which also has a reliable study design [34]. A well-conducted study by Liversidge et al. [35] shows small differences in the timing of third molar development among different populations and suggests that using statistical analysis and datasets avoiding age mimicry are more important than population specific reference data.

STUDIES, Country	Stage F (mean)	SD	Stage G (mean)	SD
Lee 2009, Korea	16.70	1.40	18.60	1.60
Li 2011, South China	18.00	2.50	19.20	2.20
Johan 2014, Malaysia	17.03	1.40	19.03	2.03
Cavric 2016, Botswana	16.60	1.56	18.30	1.57

Table 2. The Table shows mean age and standard deviation (SD) for Demirjian's stages F and G of the third molar in four well-designed studies.

7.3.3   Conclusion

Any regional differences in skeletal and tooth maturation may have a variety of causes, like hereditary factors (regional genetic heterogeneity) or external factors (diet, climate etc.). Many studies indicate that such regional differences exist in the maturation of both skeleton and teeth. These studies often have heterogeneity in the study design or in the way of reporting results, making them difficult to compare in order to create an overall picture. Age mimicry (Figure 7) may explain the inconsistent results that have been attributed to regional differences. Thus, it is not clear how much regional differences affect the results. The effect of eventual systematic differences in staging in the different studies is not fully investigated.

7.4   The situation after the systematic reviews

One of the objectives of the systematic reviews we conducted was to acquire up-to-date data with which to make a new age estimation system. It was therefore disappointing to be left with considerably fewer usable studies than we had expected. On the other hand, we gained insight into the challenges presented by study design, and this gave us a good starting point for working towards a solution. All studies reported their results groupwise, and many of them were biased by age mimicry (Figure 7). In order to use the information in these studies, we initiated a project that uses statistical modelling to produce data in an entirely new way.

8   Statistical modelling of data from included studies

8.1   Purpose

The purpose of the work we have carried out is to obtain the most complete picture possible of what the different stages in the Greulich & Pyle atlas and Demirjian's staging of wisdom teeth tell us about chronological age. It is usual to describe how chronological age is distributed for each stage. In order to understand any regional differences, we have to include studies from different parts of the world. The more observations we can obtain from different geographical regions, the more we can assume that the method will be capable of estimating the chronological age of individuals with different backgrounds. In order to construct a model of this, we primarily wanted individual data: chronological age and stage for every single individual. By using data in this format, we can take account of the effect of age mimicry (Figure 7). We received some datasets with individual data from authors that we contacted directly. In addition we began considering whether mathematical modelling could produce data of this kind from the other studies, where the data are available only at group level. We therefore found eligible studies in the search results from the two systematic reviews. In addition, we have carried out searches in PubMed to identify new publications.

Our aim was to say something about how chronological age is distributed at different stages, calculate a prediction interval for age and a probability that an individual is under a given age limit (for more information about this, see Figure 13).

8.2   Studies that can be used for modelling

The hand and teeth studies consist of empirical data (observations) that have the same basic format: all individuals have a known chronological age and an observed developmental stage. This is recorded for each individual. If, for the sake of simplicity, we say that there are only four stages, numbered with the Roman figures I–IV, a hypothetical dataset consisting of 20 individuals might look like the data in Table 3.

Individual	Chronological age	Stage
1	9.2	I
2	10.7	I
3	10.9	I
4	11.3	II
5	11.5	I
6	12.3	II
7	12.8	I
8	13.1	II
9	13.7	II
10	13.7	II
11	14.5	III
12	15.3	II
13	15.7	III
14	16.2	III
15	16.9	IV
16	17.5	IV
17	17.6	III
18	18.1	IV
19	18.6	IV
20	19.4	IV

Table 3. Example of a hypothetical dataset with four different stages.

Demirjian's staging of teeth contains only eight stages, indicated by the letters A–H [23]. The Greulich & Pale atlas for the hand skeleton contains more stages (one stage for each year, and sometimes also semi-annual images), and each stage is given an age in years [20]. This makes it possible to assign a "skeletal age" to each individual, and thus obtain a numerical system for both chronological and skeletal age. This is also in contrast to Demirjian's staging of teeth, where letters are used to denote stages and thus there is no "tooth age".

As described above, the studies yield the overall results in different ways (Figure 8). The data formats we were able to continue working with are in four formats (called types 1–4).

Data formats included in BioAlder

Type 1	data in individual-based format
Type 1b	individual-based data taken from point-plots in published studies
Type 2	frequency table with number of individuals for each stage within each whole year
Type 3	tables with data on means and standard deviations of chronological age for given stages (skeletal age or tooth formation stage)
Type 4	tables with data on means and standard deviations of chronological age and skeletal age within each whole year

Figure 8. Data formats for the studies included in the development of BioAlder.

8.2.1   Type 1 data

Type 1 data is the optimal data format to be used in BioAlder, in which exact chronological age and stage are given for each individual. This is typically a list, with the data for each individual on the individual rows (Table 4).

Gender	Chronological age	GP stage
1	12	11,5
1	14,1	14
0	11	11
0	13	13
1	14,3	14
1	10,1	11
0	11,9	13
0	12,8	13
0	10	10
1	13,9	13,5
1	12,9	12,5

Table 4. Example on type 1 data.

8.2.2 Type 1b data

Type 1b data is data points for every individual (known chronological age and observed skeletal age) collected from plot points of international publications where we have not been able to access Type 1 data from a table (Figure 9). We have used software in order to locate the points of the plots and extract the data. The challenge with this type of data is that points lying on top of each might not be included because the software does not recognize them.
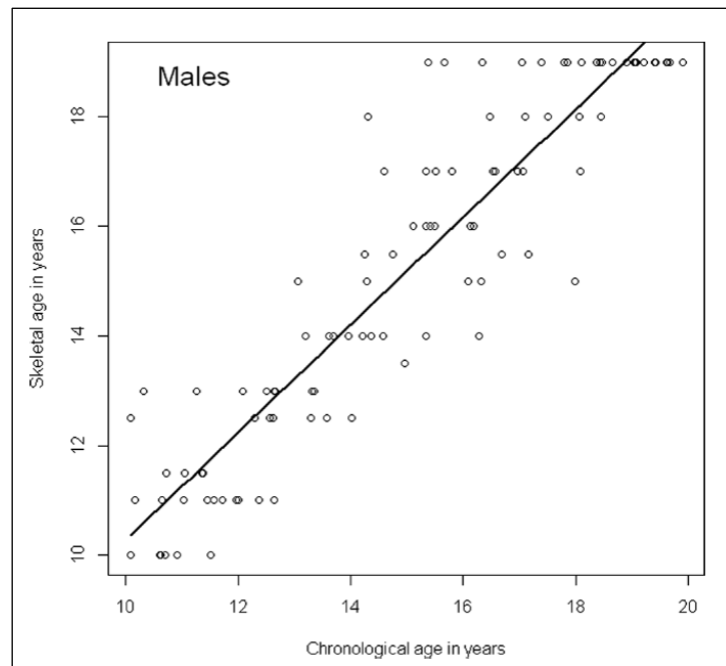
Figure 9. Example from Zabet et al. 2015 (36) on point-plot that Type 1b-data is collected from.

8.2.3 Type 2 data

In the Type 2 data format, the numbers of individuals for each whole chronological year who were assessed for each stage are given (Table 5). These tables show the stages horizontally at the top and chronological age vertically in the left-hand column. The challenge with this type of data is that chronological age is only given in whole years.

Table 4
Frequency of the Demirjian's stage of tooth 38.

Age	Male A	B	C	D	E	F	G	H	Total	Female A	B	C	D	E	F	G	H	Total
7	1								1	5	1							6
8	11	1	2						14	14								14
9	21	8	4						33	22	7	1						30
10	22	20	13	1					56	12	19	16						47
11	5	19	37	2					63	10	11	36	3					60
12	2	5	41	8	1				57	6	10	33	10					59
13	1	4	25	21	5	1			57	2	1	32	21	8				64
14	1		11	20	10	5			47	1	1	10	31	21	4	1		69
15	1		2	13	13	27	5		61	1		4	23	38	17	1		84
16		1		12	18	24	9		64			3	26	33	36	8		106
17				5	11	22	18		56			2	11	16	10	11	1	51
18				1	7	15	25	6	54				3	16	13	17	4	53
19					3	3	17	32	55					6	17	19	20	62
20					1	1	7	47	56					3	9	13	31	56
21						1	5	51	57					1	4	15	30	50
22							4	51	55						1	7	42	50
23															1	3	47	51
24																1	51	52
Total	65	58	135	83	69	99	90	187	786	73	50	137	128	142	112	96	226	964

Table 5. Example of type 2 data, where data is collected from a frequency table.

8.2.4 Type 3 data

Studies with results in this format take a skeletal age or one tooth stage as their starting point and show means and standard deviations (SD) of chronological age for all individuals in the same skeletal (or tooth) stage collectively (Table 6). We thus know the exact skeletal age or tooth stage age of each individual, but the chronological age for each individual is not specified.

SA	Size	CA_mean	CA_SD
10	75	10.42	1.18
11	58	11.09	1.11
11.5	54	11.71	1.16
12.5	50	12.28	1.12

Table 6. Example of type 3 data.

BioAlder Manual Version 1.3b
Valid from 09.03.2023

8.2.5   Type 4 data

Studies with results in this format take chronological age as their starting point and present the mean and standard deviation (SD) of skeletal age for each group of whole chronological years (Table 7). In addition, the correlation (Pearson) between skeletal and chronological ages is given. Here we know neither the skeletal age nor the chronological age of the individual. We have included this type of data for the hand skeleton only.

Size	SA_mean	SA_sd	CA_mean	CA_sd	Corr
15	9.66	1.27	10.23	0.29	0.17
12	11.08	0.65	11.49	0.28	0.13
10	12.01	0.56	12.2	0.15	0.69
11	13.04	0.58	13.37	0.22	0.71
11	13.98	0.3	14.42	0.28	0.75

Table 7. Example of type 4 data.

8.3   Modelling individual data

To be able to use the different types of data into BioAlder, we used well-known mathematical methods to model data that was not type 1 or 1b data, so that we got stage and chronological age for all the individuals used in the tool. Modelling of individual data is described in the appendix, part A.4.

8.4   From individual data to probabilities

With the aid of individual data that we have collected and/or generated from group data as described above, we can construct a distribution of the probability that individuals of a given chronological age will be in the different stages (this is also described in the appendix, part A.2). One way of modelling this is to consider a particular age segment and look at the percentages of the individuals who fall within different stages. We have illustrated modelling of this kind in Figure 10 with the different stages in different colours.
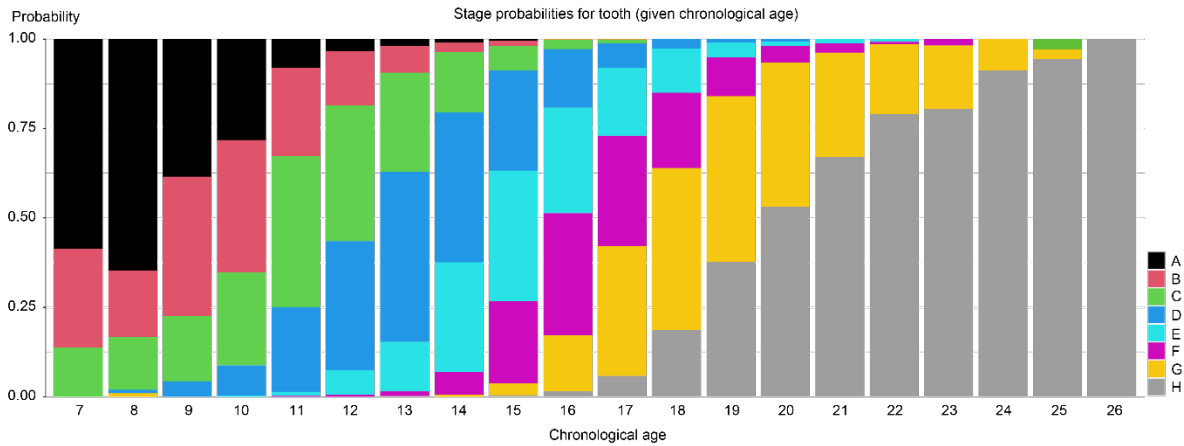
Figure 10. Proportion of individuals (boys) in BioAlder (for given whole chronological ages) who fall into the different tooth stages. The stages are shown in different colours.

With this tool, we want to be able to specify age with greater precision than just whole years. To achieve this, we use a regression model called a "transition analysis" model [36]. This model will give the data a smooth functional fit. An important assumption for this type of model is that the stages are ordered when age increase: First comes the first stage, then the next, etc., and finally an end stage that marks it as the final stage in the method's system. What is special about this model is that it allows for the fact that "the last stage never ends", i.e. there is the same probability of being in the last stage whether you are 25 or 50 years (given that the other stages are completed). In Figure 11, we see an example of such a model. Here we see that the probability of being in a stage is a function of age adapted to the data. In order to check whether this model fits with the data, we need to check that the model is consistent with the observations directly (see Figure 11, for example). For more information on the resulting models and model validation for the tool, see part A.2 of the Appendix.
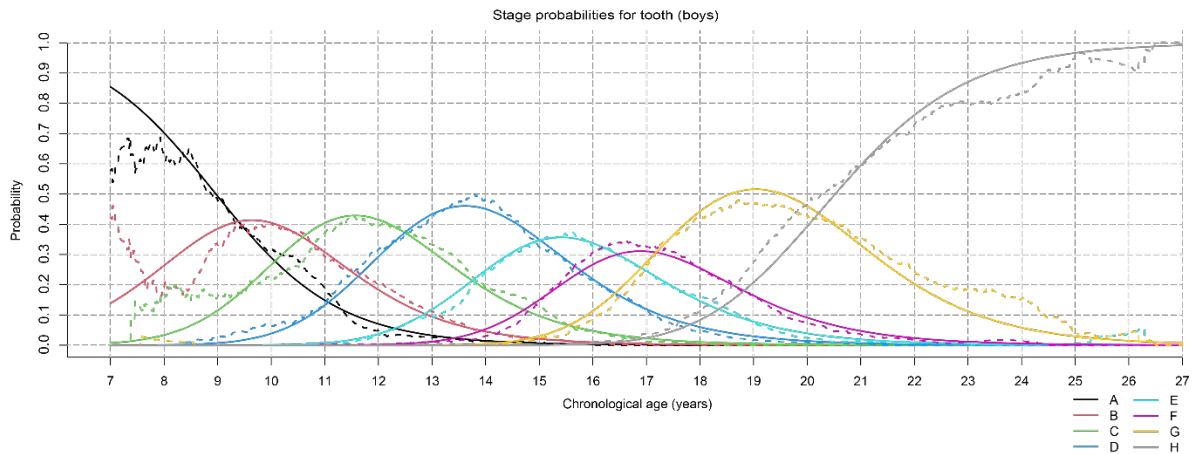
30

Figure 11. Proportions of boys that fall into the different stages for a given chronological age (uneven curves). Each age segment spans one year from 7 to 27 years old.

8.5   Distribution of chronological age given observed stage

Until now, we have only described the stage probability for given chronological ages. Our real objective, however, was to describe how chronological age is distributed for a given stage. In other words: if an individual has a GP stage and/or a Demirjian stage, what is the probable chronological age of this individual? We can describe this with the aid of Bayes' theorem to describe the age distribution for the observed stage.

Age distribution for stage = Probability of stage (age) * assumed age distribution * constant

Bayes' theorem is a commonly used mathematical method that gives us the probability for something to happen, given something else, for example age, given stage.

This enables us to produce the results for the two methods. The "constant" in the above formula is a numerical value such that the area of the age distribution for a given stage is equal to one. In the tool, we assume a uniform age distribution. We have described this in the appendix, part A.7, paragraph 9.2.

8.6   Combination of hand and tooth

Because of the wide biological variation that is reflected by the methods, it is desirable to combine the hand and tooth stages of individuals in order to obtain a more precise estimate of chronological age. Gelbrich et al., 2015 [15] shows that there is no relationship between the age estimation errors by means of the hand-wrist and the third molar tooth of the same individual, and we can therefore

assume that the two methods are independent for a given chronological age. In practice, this means that we can multiply the chronological age distributions for the two methods together to obtain a joint distribution of chronological age based on a tooth formation stage and a hand skeletal age combined (Figure 12). This naturally presupposes that radiographs for these two methods are taken at approximately the same time. This is described in more detailed in the appendix, part A.8.

If the two developments (hand and teeth) were dependent on each other, the situation would be different. In section A.8.1 in the appendix, we illustrate the great significance a moderate dependence on the combined age distribution.
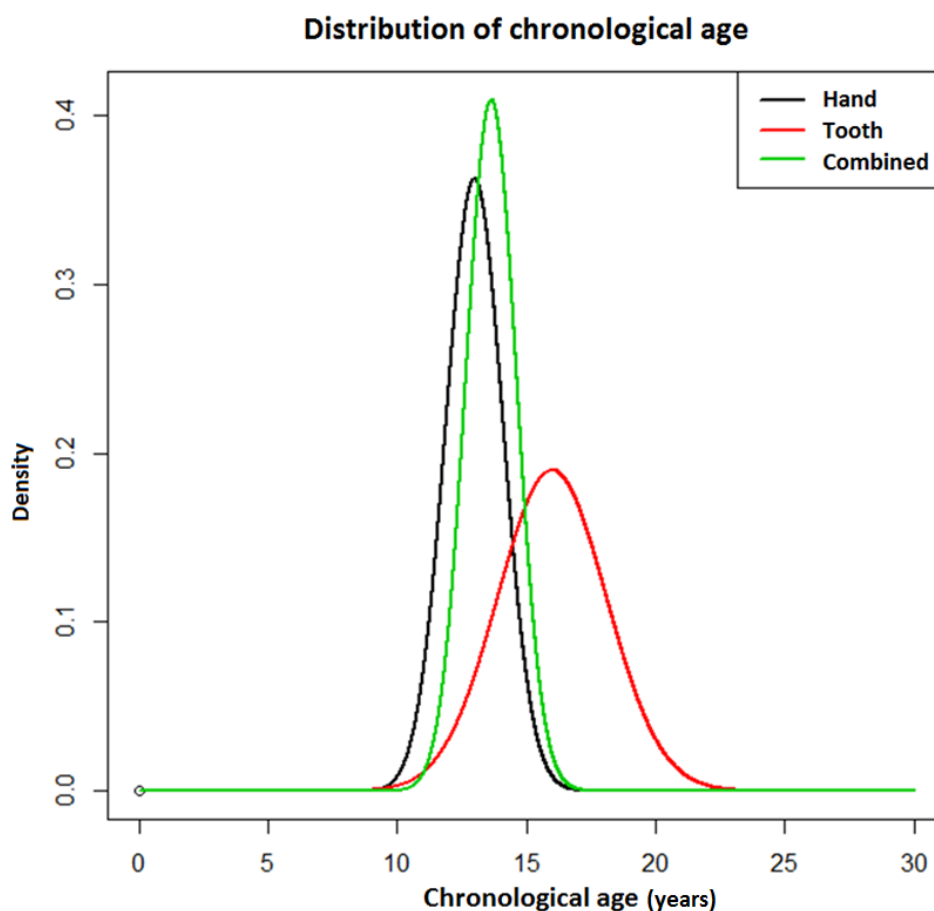

Figure 12. The figure shows how the distributions arrived at using the two methods are combined if they are assumed to be independent of one another for a given chronological age. For each given chronological age, the values of the functions for hand and tooth respectively are multiplied together, and then the multiplied function is normalised to have an area of one.

The statistics underlying the results generated by the tool

BioAlder results are based on distributions that describe the spread of chronological age for a given observed stage. For example, this may be a normal distribution where the most probable age is the mean (average). Other more probable values will be close to this mean, and less common values will be far from the mean.

A 95% prediction interval [x, y] for age means that if a population that is representative of the population upon which the tool was developed were tested, we would expect the age of 95% of the individuals to fall within this interval (i.e. be between x and y years old). This interval is marked black in the illustration on the right. 2.5% of the individuals will fall outside into the area field to the left (i.e. be less than x years old) and the remaining 2.5 per cent outside in the white area to the right (i.e. be more than y years old).

If a lower safety margin is specified, for example 75%, the prediction interval will be narrower. As the distribution graph is curved, and we are interested in the area below the graph, it is clear that the 75% interval is considerably narrower than the 95% interval.

If we want to specify how large a percentage of individuals is expected to lie below a given age limit, e.g. 18 years, we can look at the area under the distribution curve up to this given limit. This is illustrated in the figure on the right.
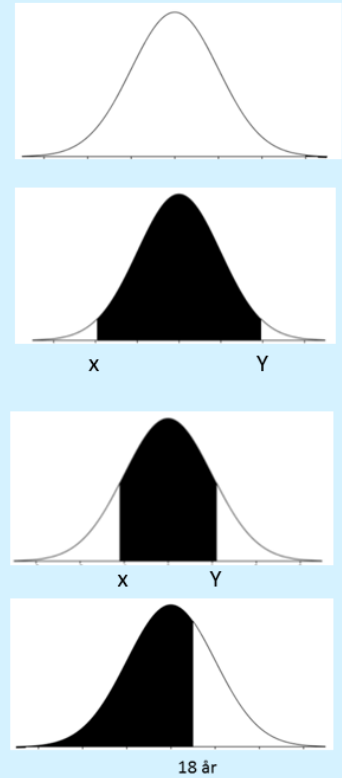
Figure 13. The statistics underlying the results generated by BioAlder.

9 Results used in BioAlder

9.1 Overview of studies used in BioAlder 1.3b

The underlying data for hand are based on the studies in Table 8.

Format	Reference	Boys	Girls	Country	Age span
Type 1	Saadé 2017 [37]	115	129	Lebanon	9-16
Type 1	Santos 2011 [38]	136	94	Portugal	12-20
Type 1	Van Rijn 2001 [39]	178	197	Netherlands	9-20
Type 1	Zafar 2010 [40]	165	64	Pakistan	7-18
Type 1	Tise 2011 [41]	359	126	Italy	11-19
Type 1	Alcina 2018 [42]	312	293	Spain	7-19
Type 1	Yilmaz 2018 [43]	333	379	Turkey	10-15
Type 1b	Buken 2007 [44]	231	219	Turkey	11-19
Type 1b	Hackman 2013 [45]	145	0	Scotland	8-20
Type 1b	Haiter-Neto 2006 [46]	115	105	Brasil	7-15
Type 1b	Kim 2015 [47]	60	40	South Korea	7-12
Type 1b	Paxton 2013 [48]	112	67	Australia	7-18
Type 1b	Schmidt 2007 [49]	172	0	Germany	8-18
Type 1b	Zabet 2015 [50]	98	87	France	10-19
Type 3	Chaumoitre 2016 [29]	886	673	France	7-20
Type 4	Bala 2010 [51]	59	59	India	10-14
Type 4	Cantekin 2012 [52]	259	351	Turkey	10-17
Type 4	Chiang 2005 [53]	141	70	Taiwan	10-17
Type 4	Griffith 2007 [54]	281	105	China	10-18
Type 4	Koc 2001 [55]	185	0	Turkey	10-17
Type 4	Mohammed 2015 [56]	270	270	India	10-18
Type 4	Nahid 2010 [57]	32	45	Iran	10-14
Type 4	Patel 2015 [58]	56	60	India	10-16
	Total	**4700**	**3433**		

Table 8. All hand-wrist studies included in BioAlder. The studies are listed with data format, number of included individuals, country and age range of the included population.

The underlying wisdom tooth data are based the studies in Table 9.

Format	Reference	Boys	Girls	Country	Age span
Type 1	Cavric 2016 [34]	763	907	Botswana	7-23
Type 1	Malta dataset [59, 60]	553	650	Malta	8-24
Type 1	Saadé 2017 [37]	113	119	Lebanon	9-16
Type 1	Jayaraman 2016 [61]	682	617	China	8-24
Type 1	Knell 2009 [62]	591	669	Switzerland	15-22
Type 1	Hegde 2016 [63]	410	267	India	7-16
Type 1	Yilmaz 2018 [43]	70	92	Turkey	10-15
Type 2	Lee 2009 [31]	786	964	South Korea	7-24
Type 2	Johan 2012 [33]	540	539	Malaysia	14-25
Type 2	Duangto 2017 [64]	872	983	Thailand	8-23
Type 2	Li 2012 [32]	648	760	China	7-23
Type 2	Liu 2018 [65]	963	1148	China	8-23
Type 2	Marrero-Ramos 2020	27	75	Spain	13-27
	Total	**7018**	**7790**		

Table 9. All wisdom tooth studies included in BioAlder. The studies are listed with data format, number of included individuals, country and age range of the included population.

All references of data format type 1 are individual data we have received from authors we have contacted. After contacting dozens of authors, we were given access to seven datasets for hand and eight for teeth. The hand datasets correspond to the included population in the published article in question, and we have therefore referred to the article in Table 8. Some of the individual-based datasets for teeth do not represent a single publication. We therefore call them "datasets" with appurtenant geographical area in Table 9.

9.2   Choice of age distribution

When we use Bayes' theorem to predict the age for a given stage, an age range and - distribution of the individuals included in the model must be assumed in advance. This is comparable to the way the age range and - distribution of individuals to be included in a study must be pre-defined. Because we do not know the real chronological age for the individuals that are tested with BioAlder, we assume that the ages are evenly distributed between an upper and lower defined age. This implies that it is equal chance to be any of the defined ages before testing. This assumption is called a "prior".

We have tested how different prior assumptions, for example a normal distribution around 17 or 18 years old, would have affected the prediction, to ensure that the choice of an even prior would not be unfavorable. This is explained in more detail in the appendix, part A.7.

The lowest chronological age of the age prior is set to seven to make sure that all tooth stages were included in the data. The specification of the upper chronological age limit would have a practical impact for the ages around 18 years. Specifically the age 18 years is important since the primary priority for the tool is to reduce the possibility that children are assessed as adults, i.e. the type I error. Table 10 presents an overview of the defined upper ages for the different methods and genders.

	Hand	Tooth	Combined
Boys	20	20.5	20.5
Girls	19	21	21

Table 10. Overview of the upper ages (in years) defined for the assumed age distribution for the different methods and genders

The probability for a certain chronological age in the last stages is partially defined by the upper age limit. See Figure 14 for an example of how the given upper age will affect the prediction intervals. Knell et al. [62] and Olze et al. [67] used the age defining the 50% probability of being in the last stage to bypass this difficulty. Roberts et al. [68] and Lee et al. [31] suggested specifying the upper age limit as the age where the second last stage on tooth ends, so that the complete age distributions, except for the last stage, are described. The choice of the upper age limit for the hand method does not have much impact on the type I error (see also Bleka et al. [9]). Hence, for the hand method we defined the upper age limits as the ages where the second last stage ends: 20 years for males, and 19 years for females. Following a similar strategy for the tooth method gave the age limits 23/25 years for males/females, because the age distributions at stage G have relatively long tails. When we compared this model with the Swiss- and the Botswana dataset, we obtained a high risk of type I error when observing the last stage (i.e., stage H). To reduce this risk we instead decided to define the upper age limit to be the age defining the 50% probability of being in the last stage. We used this definition for both the tooth and the combined method, since the age distributions for the combined method are similar to the tooth method for the last stages. Based on our fitted stage probability models we found that for males, this was 20.5 years for tooth and 20.5 years for the combination (rounded to closest half year), whereas for females this was 21 years for tooth and 21 years for the combination. A natural consequence of this upper age definition is that it removes information about the ages beyond the upper age limit, which mainly affects the distribution of CA for the latest stages. However, the effect

is limited by the fact that the last stages are only described by their lower values in the output of BioAlder, and not as full distributions.
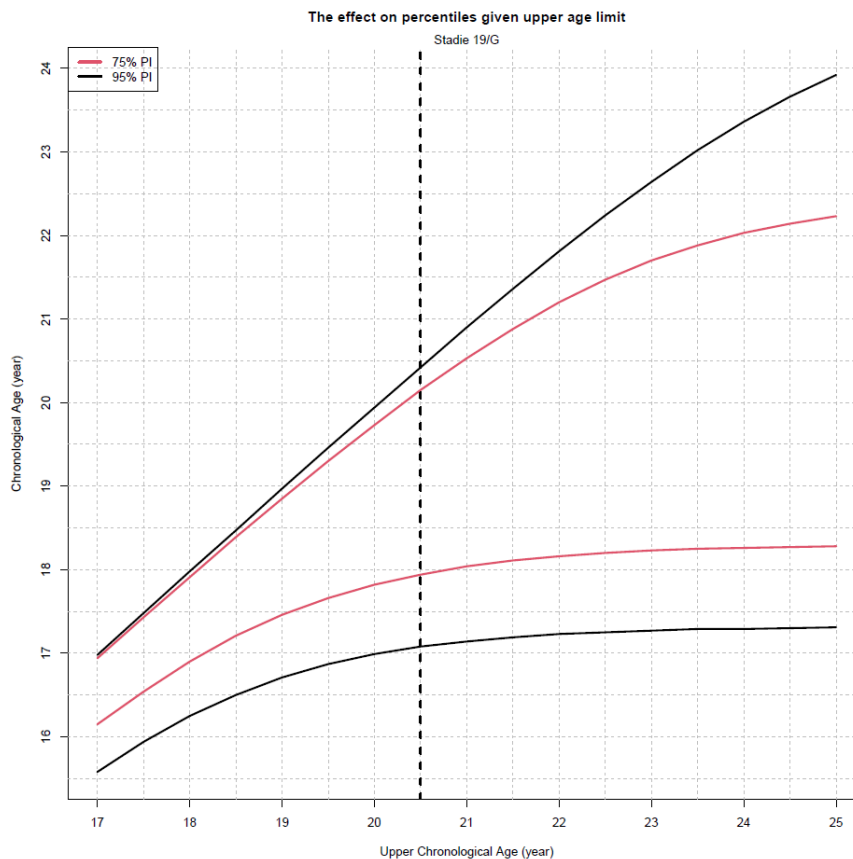

Figure 14. The figure shows how the prediction intervals (PI) vary for skeletal age 19 years combined with tooth stage G for different upper ages in the model. The given upper age is on the X-axis, and the chronological age on the Y-axis. For a given upper age on the X-axis, 75% (red) and 95% (black) intervals for chronological age can be read off with the aid of the lines and values on the Y-axis.

10 The future of biological age estimation

10.1 Image-based methods

Biological variation and uncertainty regarding the significance of regional differences for age estimation by means of hand and tooth radiographs make the methods imprecise. The natural biological variation is an inherent challenge for age estimation based on skeletal and tooth development, and more research or the introduction of other ways of staging development will not change this fact. When it comes to mapping regional differences, we would urge collaboration with researchers in this field, using all the existing data. We hope that the launch of BioAlder, attendance

at conferences and international publications will promote such a collaboration going forward. New studies on unstudied populations are also warranted.

10.2 DNA methylation

We have conducted investigations to find new and better methods of estimating chronological age in children and adolescents. The method that stands out as the most promising is DNA methylation.

DNA methylation changes with increasing age [69]. Several prediction models for estimating chronological age have been developed [69-73], but only a few have been optimised for an adolescent population. DNA methylation appears to have less biological variation and better resolution than skeletal and tooth maturation. Moreover, unlike skeletal and tooth maturation, DNA methylation has no end stage. Another advantage of DNA methylation is that international research activity in many fields is adding rapidly increasing amounts of knowledge and freely available data. Only a small quantity of blood or saliva is required for the analysis, and this makes the method more ethically acceptable in both research and practical use than today's radiological methods. OUH has developed a prediction model based on DNA methylation data for an adolescent population [74]. To validate this model, we have gathered and analysed samples from different geographical regions. This work is still ongoing.

DNA methylation

DNA methylation changes in pace with change in chronological age, and theefore, several prediction models for estimating chronological age have been developed. However, none have been optimised for an adolescent population. We have conducted a pilot study based on datasets from research databases that cover ten different studies and 1000 individuals in relevant age groups (12-25 years old). The method appears to have less variation than the image-based methods that are used in BioAlder, and it has no endpoint. We need more data from different parts of the world, where there are variations in heritage and environment. This would enable validation of the method on different populations.

Figure 15. DNA methylation. DNA methylation is changed in pace with chronological age and might be used as a method for age-assessment in the future.

11 References

1. Dahlberg PS, M.A., Ding KY, Bleka Æ, Straumann GH, Rolseth V, Skjerven-Martinsen M, Delaveris GJM, Vist GE., *Samsvar mellom kronologisk alder og skjelettalder basert på Greulich og Pyle-atlaset for aldersestimering: en systematisk oversikt.* 2017, Folkehelseinstituttet: www.fhi.no.
2. Rolseth V, M.A., Dahlberg PS, Ding KY, Bleka Ø, Skjerven-Martinsen M, Straumann GH, Delaveris GJM, Vist GE, *Demirjians utviklingsstadier på visdomstenner for estimering av kronologisk alder: en systematisk oversikt.* 2017, Folkehelseinstituttet: www.fhi.no
3. Ding KY, R.V., Dahlberg PS, Mosdøl A, Straumann GH, Bleka Ø, Vist GE., *Age estimation by ossification stages of the medial clavicular epiphysis: a systematic review.* 2018, Folkehelseinstituttet: www.fhi.no.
4. Ding KY, M.A., Straumann GH, Vist GE., *Age estimation in adolescents and young adults by psychological assessment of maturity: a systematic review.* 2018, Folkehelseinstituttet: www.fhi.no.
5. Ding KY, D.P., Rolseth V, Mosdøl A, Straumann GH, Bleka Ø, Vist GE., *Development stages of the knee and ankle by computed tomography and magnetic resonance imaging for estimation of chronological age.* 2018, Folkehelseinstituttet: www.fhi.no.
6. Dahlberg, P.S., et al., *A systematic review of the agreement between chronological age and skeletal age based on the Greulich and Pyle atlas.* Eur Radiol, 2019. **29**(6): p. 2936-2948.
7. Rolseth, V., et al., *Age assessment by Demirjian's development stages of the third molar: a systematic review.* Eur Radiol, 2019. **29**(5): p. 2311-2321.
8. Bleka, Ø., et al., *BioAlder: a tool for assessing chronological age based on two radiological methods.* Int J Legal Med, 2019. **133**(4): p. 1177-1189.
9. Bleka, O., et al., *Advancing estimation of chronological age by utilizing available evidence based on two radiographical methods.* Int J Legal Med, 2018.
10. NOAS and Redd Barna, *Over eller under 18? : Aldersvurdering av enslige mindreårige asylsøkere.* 2016.
11. European Asylum Support Office, *EASO Age assessment practice in Europe.* 2013: Luxemburg.
12. Vandewalle, S., et al., *Sex steroids in relation to sexual and skeletal maturation in obese male adolescents.* J Clin Endocrinol Metab, 2014. **99**(8): p. 2977-85.
13. Schmeling, A., et al., *Forensic Age Estimation: Methods, Certainty, and the Law.* Deutsches Ärzteblatt International, 2016. **113**(4): p. 44-50.
14. Rudolf, E., et al., *Standardized medical age assessment of refugees with questionable minority claim-a summary of 591 case studies.* Int J Legal Med, 2015. **129**(3): p. 595-602.
15. Gelbrich, B., et al., *Combining wrist age and third molars in forensic age estimation: how to calculate the joint age estimate and its error rate in age diagnostics.* Ann Hum Biol, 2015. **42**(4): p. 389-96.
16. European Asylum Support Office, *EASO Age assessment practice in Europe.* 2018: Luxemburg.
17. Olze, A., et al., *Validation of common classification systems for assessing the mineralization of third molars.* Int J Legal Med, 2005. **119**(1): p. 22-6.
18. Rättsmedicinalverket. Available from: https://www.rmv.se/verksamheter/medicinska-aldersbedomningar/.
19. Socialstyrelsen. *Om magnetkamera vid bedömning av ålder – En studie av validiteten i radiologisk undersökning.* 2018; Available from: https://www.socialstyrelsen.se/globalassets/sharepoint-dokument/artikelkatalog/ovrigt/2018-5-21.pdf.
20. Greulich, W. and S. Pyle, *Radiograph Atlas of Skeletal Development of the Hand and Wrist.* 2nd ed. 1959, Stanford, California, USA.
21. Tanner, J., et al., *Assessment of Skeletal Maturity and Prediction of Adult Height (TW3) Method.* 2001, London, UK: WB Saunders, Harcourt Publishers Ltd.

22.   Tanner, J., et al., *Assessment of Skeletal Maturity and Prediction of Adult Height (TW2 Method)*. 2nd ed. 1983, London, UK: Academic Press.

23.   Demirjian, A., H. Goldstein, and J.M. Tanner, *A new system of dental age assessment.* Human Biology, 1973. **45**(2): p. 211-27.

24.   Hunt, E.E., Jr. and I. Gleiser, *The estimation of age and sex of preadolescent children from bones and teeth.* Am J Phys Anthropol, 1955. **13**(3): p. 479-87.

25.   Moorrees, C.F., E.A. Fanning, and E.E. Hunt, Jr., *AGE VARIATION OF FORMATION STAGES FOR TEN PERMANENT TEETH.* J Dent Res, 1963. **42**: p. 1490-502.

26.   Haavikko, K., *The formation and the alveolar and clinical eruption of the permanent teeth. An orthopantomographic study.* Suomen Hammaslaakariseuran Toimituksia, 1970. **66**(3): p. 103-170.

27.   Kullman, L., G. Johanson, and L. Akesson, *Root development of the lower third molar and its relation to chronological age.* Swed Dent J, 1992. **16**(4): p. 161-7.

28.   Köhler, S., et al., *Die entwicklung des weisheitszahnes als kriterium der lebensaltersbestimmung.* Annals of Anatomy - Anatomischer Anzeiger, 1994. **176**(4): p. 339-345.

29.   Chaumoitre, K., et al., *Forensic use of the Greulich and Pyle atlas: prediction intervals and relevance.* European Radiology, 2017. **27**(3): p. 1032-1043.

30.   Thodberg, H.H. and L. Savendahl, *Validation and reference values of automated bone age determination for four ethnicities.* Acad Radiol, 2010. **17**(11): p. 1425-32.

31.   Lee, S.H., et al., *Development of third molars in Korean juveniles and adolescents.* Forensic Science International, 2009. **188**(1): p. 107-11.

32.   Li, G., et al., *Dental age estimation from the developmental stage of the third molars in western Chinese population.* Forensic Science International, 2012. **219**(1): p. 158-64.

33.   Johan, N.A., et al., *The variability of lower third molar development in Northeast Malaysian population with application to age estimation.* Journal of Forensic Odonto-Stomatology, 2012. **30**(1): p. 45-54.

34.   Cavric, J., et al., *Time of mineralization of permanent teeth in children and adolescents in Gaborone, Botswana.* Ann Anat, 2016. **203**: p. 24-32.

35.   Liversidge, H.M., et al., *A radiographic study of the mandibular third molar root development in different ethnic groups.* J Forensic Odontostomatol, 2017. **2**(35): p. 97-108.

36.   Boldsen, J.L., et al., *Transition analysis: a new method for estimating age from skeletons.* CAMBRIDGE STUDIES IN BIOLOGICAL AND EVOLUTIONARY ANTHROPOLOGY, 2002: p. 73-106.

37.   Saade, A., et al., *Dental and Skeletal Age Estimations in Lebanese Children: A Retrospective Cross-sectional Study.* J Int Soc Prev Community Dent, 2017. **7**(3): p. 90-97.

38.   Santos, C., et al., *Comparative study of Greulich and Pyle Atlas and Maturos 4.0 program for age estimation in a Portuguese sample.* Forensic Sci Int, 2011. **212**(1-3): p. 276.e1-7.

39.   van Rijn, R.R., et al., *Is the Greulich and Pyle atlas still valid for Dutch Caucasian children today?* Pediatric Radiology, 2001. **31**(10): p. 748-52.

40.   Zafar, A.M., et al., *An appraisal of Greulich-Pyle Atlas for skeletal age assessment in Pakistan.* JPMA - Journal of the Pakistan Medical Association, 2010. **60**(7): p. 552-5.

41.   Tise, M., et al., *Applicability of Greulich and Pyle method for age assessment in forensic practice on an Italian sample.* International Journal of Legal Medicine, 2011. **125**(3): p. 411-6.

42.   Alcina, M., et al., *Reliability of the Greulich and Pyle method for chronological age estimation and age majority prediction in a Spanish sample.* Int J Legal Med, 2018. **132**(4): p. 1139-1149.

43.   Günen Yılmaz, S., et al., *Evaluation of the relationship between the Demirjian and Nolla methods and the pubertal growth spurt stage predicted by skeletal maturation indicators in Turkish children aged 10-15: investigation study.* Acta Odontol Scand, 2019. **77**(2): p. 107-113.

44. Buken, B., et al., *Is the assessment of bone age by the Greulich-Pyle method reliable at forensic age estimation for Turkish children?* Forensic Science International, 2007. **173**(2): p. 146-53.

45. Hackman, L. and S. Black, *The reliability of the Greulich and Pyle atlas when applied to a modern Scottish population.* J Forensic Sci, 2013. **58**(1): p. 114-9.

46. Haiter-Neto, F., et al., *Skeletal age assessment: a comparison of 3 methods.* Am J Orthod Dentofacial Orthop, 2006. **130**(4): p. 435.e15-20.

47. Kim, J.R., Y.S. Lee, and J. Yu, *Assessment of bone age in prepubertal healthy Korean children: comparison among the Korean standard bone age chart, Greulich-Pyle method, and Tanner-Whitehouse method.* Korean J Radiol, 2015. **16**(1): p. 201-5.

48. Paxton, M.L., A.C. Lamont, and A.P. Stillwell, *The reliability of the Greulich-Pyle method in bone age determination among Australian children.* J Med Imaging Radiat Oncol, 2013. **57**(1): p. 21-4.

49. Schmidt, S., et al., *Comparative analysis of the applicability of the skeletal age determination methods of Greulich-Pyle and Thiemann-Nitz for forensic age estimation in living subjects.* Int J Legal Med, 2007. **121**(4): p. 293-6.

50. Zabet, D., et al., *Can the Greulich and Pyle method be used on French contemporary individuals?* Int J Legal Med, 2015. **129**(1): p. 171-7.

51. Bala, M., A. Pathak, and R.L. Jain, *Assessment of skeletal age using MP3 and hand-wrist radiographs and its correlation with dental and chronological ages in children.* Journal of the Indian Society of Pedodontics and Preventive Dentistry, 2010. **28**(2): p. 95-99.

52. Cantekin, K., et al., *Bone age assessment: the applicability of the Greulich-Pyle method in eastern Turkish children.* Journal of Forensic Sciences, 2012. **57**(3): p. 679-82.

53. Chiang, K.H., et al., *The reliability of using Greulich-Pyle method to determine children's bone age in Taiwan.* Tzu Chi Medical Journal, 2005. **17**(6): p. 417-420+453.

54. Griffith, J.F., J.C.Y. Cheng, and E. Wong, *Are western skeletal age standards applicable to the Hong Kong Chinese population? A comparison of the Greulich and Pyle method and the tanner and whitehouse method.* Hong Kong Medical Journal, 2007. **13**(3 Supplement 3): p. 28-32.

55. Koc, A., et al., *Assessment of bone ages: is the Greulich-Pyle method sufficient for Turkish boys?* Pediatrics International, 2001. **43**(6): p. 662-5.

56. Mohammed, R.B., et al., *Is Greulich and Pyle standards of skeletal maturation applicable for age estimation in South Indian Andhra children?* Journal of pharmacy and bioallied sciences., 2015. **7**(3): p. 218-25.

57. Nahid, G., et al., *Assessment of bone age in Kurdish children in IRAN.* Pakistan Journal of Medical Sciences, 2010. **26**(1): p. 36-39.

58. Patel, P.S., et al., *Accuracy of two dental and one skeletal age estimation methods in 6-16 year old Gujarati children.* Journal of forensic dental sciences : JFDS, 2015. **7**(1): p. 18-27.

59. Alsaffar, H., et al., *Dental age estimation of children and adolescents: Validation of the Maltese Reference Data Set.* J Forensic Leg Med, 2017. **45**: p. 29-31.

60. Elshehawi, W., et al., *Dental age assessment of Maltese children and adolescents. Development of a reference dataset and comparison with a United Kingdom Caucasian reference dataset.* J Forensic Leg Med, 2016. **39**: p. 27-33.

61. Jayaraman, J., et al., *Development of a Reference Data Set (RDS) for dental age estimation (DAE) and testing of this with a separate Validation Set (VS) in a southern Chinese population.* J Forensic Leg Med, 2016. **43**: p. 26-33.

62. Knell, B., et al., *Dental age diagnostics by means of radiographical evaluation of the growth stages of lower wisdom teeth.* Int J Legal Med, 2009. **123**(6): p. 465-9.

63. Hegde, S., A. Patodia, and U. Dixit, *Staging of third molar development in relation to chronological age of 5-16 year old Indian children.* Forensic Sci Int, 2016. **269**: p. 63-69.

64. Duangto, P., et al., *New models for age estimation and assessment of their accuracy using developing mandibular third molar teeth in a Thai population.* Int J Legal Med, 2017. **131**(2): p. 559-568.

65.   Liu, Y., et al., *Third molar mineralization in relation to chronologic age estimation of the Han in central southern China.* Int J Legal Med, 2018. **132**(5): p. 1427-1435.

66.   Marrero-Ramos, M.D., et al., *Estimation of the age of majority through radiographic evaluation of the third molar maturation degree.* Med Oral Patol Oral Cir Bucal, 2020. **25**(3): p. e359-e363.

67.   Olze, A., et al., *Studies on the chronology of third molar mineralization in First Nations people of Canada.* International Journal of Legal Medicine, 2010. **124**(5): p. 433-7.

68.   Roberts, G.J., et al., *Dental Age Estimation (DAE): Data management for tooth development stages including the third molar. Appropriate censoring of Stage H, the final stage of tooth development.* Journal of Forensic and Legal Medicine, 2015. **36**: p. 177-184.

69.   Horvath, S., *DNA methylation age of human tissues and cell types.* Genome Biol, 2013. **14**(10): p. R115.

70.   Bekaert, B., et al., *Improved age determination of blood and teeth samples using a selected set of DNA methylation markers.* Epigenetics, 2015. **10**(10): p. 922-30.

71.   Huang, Y., et al., *Developing a DNA methylation assay for human age prediction in blood and bloodstain.* Forensic Sci Int Genet, 2015. **17**: p. 129-36.

72.   Yi, S.H., et al., *Isolation and identification of age-related DNA methylation markers for forensic age-prediction.* Forensic Sci Int Genet, 2014. **11**: p. 117-25.

73.   Hannum, G., et al., *Genome-wide methylation profiles reveal quantitative views of human aging rates.* Mol Cell, 2013. **49**(2): p. 359-67.

74.   Aanes, H., et al., *A new blood based epigenetic age predictor for adolescents and young adults.* Scientific Reports, 2023. **13**(1): p. 2303.

Appendix to BioAlder
Manual Version 1.3b

Appendix to Manual

BioAlder: A tool for using biological tests to assess the age of unaccompanied minor asylum-seekers

Department of Forensic Sciences | Division of Laboratory Medicine | Oslo University Hospital

Contents

A. Statistical modelling

A.1 Purpose

The purpose of the work we have carried out is to produce the most complete picture possible of what the different stages in the Greulich & Pyle (GP) atlas and Demirjian's staging of the (lower left) wisdom tooth tell us about chronological age. It is usual to describe how chronological age is distributed for each stage. By chronological age, we mean the time from birth until the radiograph was taken, expressed as number of days, for example. We call the stages in the GP atlas skeletal age (for the hand), and the stages of the tooth, tooth stages. For a further description of the stages, please see the manual.

In order to understand any regional differences, we have to include studies from different parts of the world. The more observations we can obtain from different geographical regions, the more justified we are in assuming that the method will be capable of estimating chronological age for individuals with different backgrounds.

Our aim with the BioAlder tool is to describe how chronological age is distributed at given stages in one of the methods hand, tooth, or hand and tooth combined. We want to use this distribution to calculate the 2.5% and 97.5% percentiles of the distribution, in order to define a 95% prediction interval for chronological age, or the probability that chronological age is under a given age limit (for example 18 years). Ideally, all studies found in the literature would have presented the distribution of chronological age at given stages in the form of a histogram, or the like. As mentioned previously, this might result in the selection bias known as age mimicry, which may strongly influence the results (see the manual for an explanation of this effect). To take account of this effect, it is therefore necessary to approach the problem from a different angle: to describe the distribution of the different stages for a given chronological age. This prevents the age mimicry effect as described in the manual. Figure 1 shows why this may be a sensible approach. This figure illustrates the fact that the approach that indicates the probability of being in different stages (for a given age) is not influenced by adding extra individuals of a given age. This is an important argument for why we want to build a probability distribution for the various stages for an individual's given chronological age.
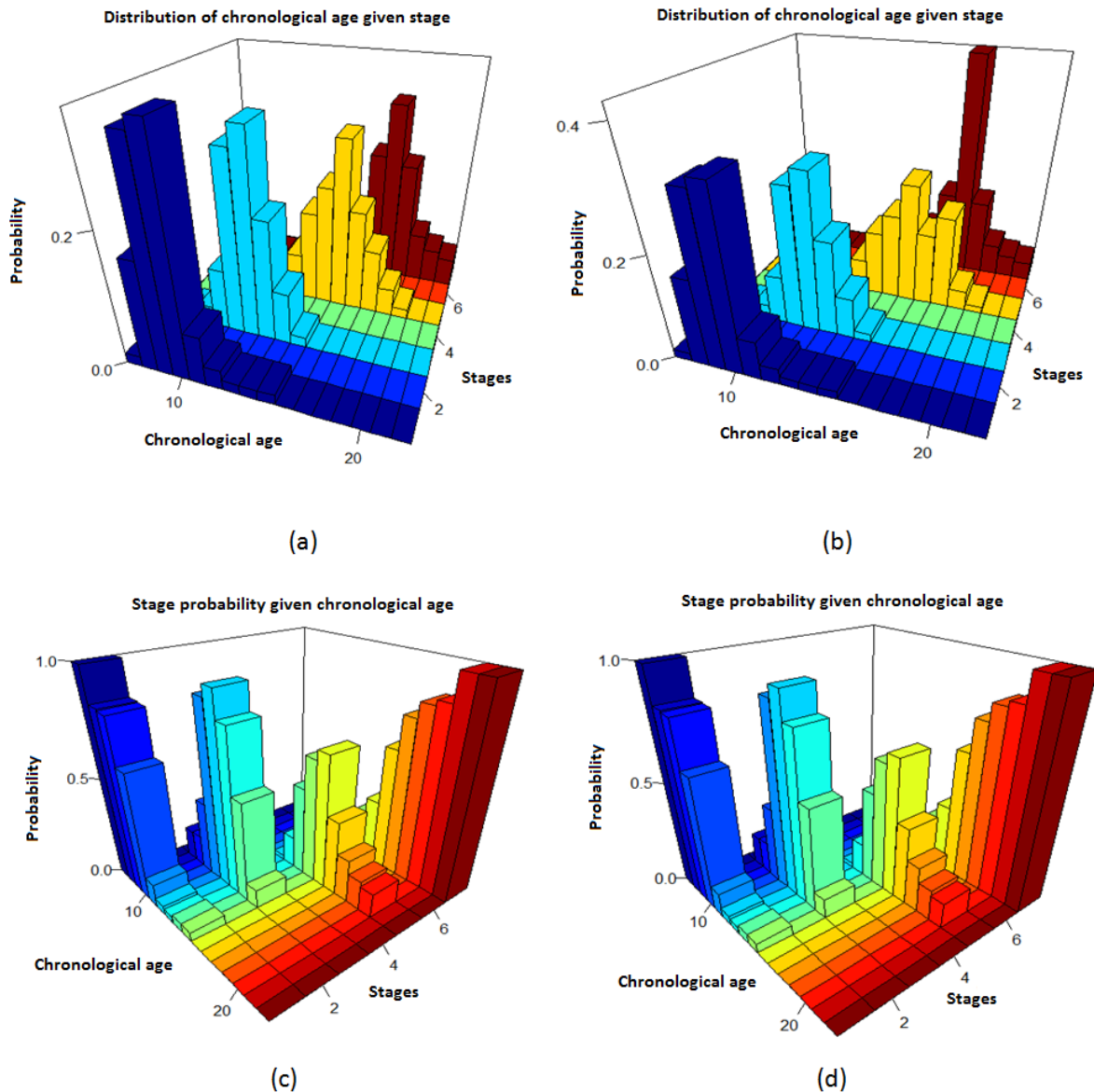
3

Figure 1. The figure shows a 3D histogram that illustrates what happens when we include thirty extra 18 year-olds for two types of approach. The one horizontal axis represents the variable 'Chronological age' and the other the category variable 'Stage'. Figures (a) and (b) show the distribution of chronological age for given stages, while figures (c) and (d) show the probability of different stages, given chronological age. Figures (a) and (c) show the distributions before thirty 18 year-olds are included, while figures (b) and (d) show the effect after thirty 18 year-olds are included.

A.2   Modelling stage probabilities

In order to be able to construct a probability distribution for the various stages for a given chronological age, we assume a regression model of the form **Stage probability = function of chronological age**. The point of this model is to be able to say something about the possibility of being in one of the defined stages at a given age.

One way of using data to model this probability is to consider an age segment (for example all those between 10 and 11 years old) and see how many fall into different stages. Figure 2 shows an example

4

where stage probabilities are estimated as the proportions of individuals that fall into four different stages. We call this type of model a non-parametric model.
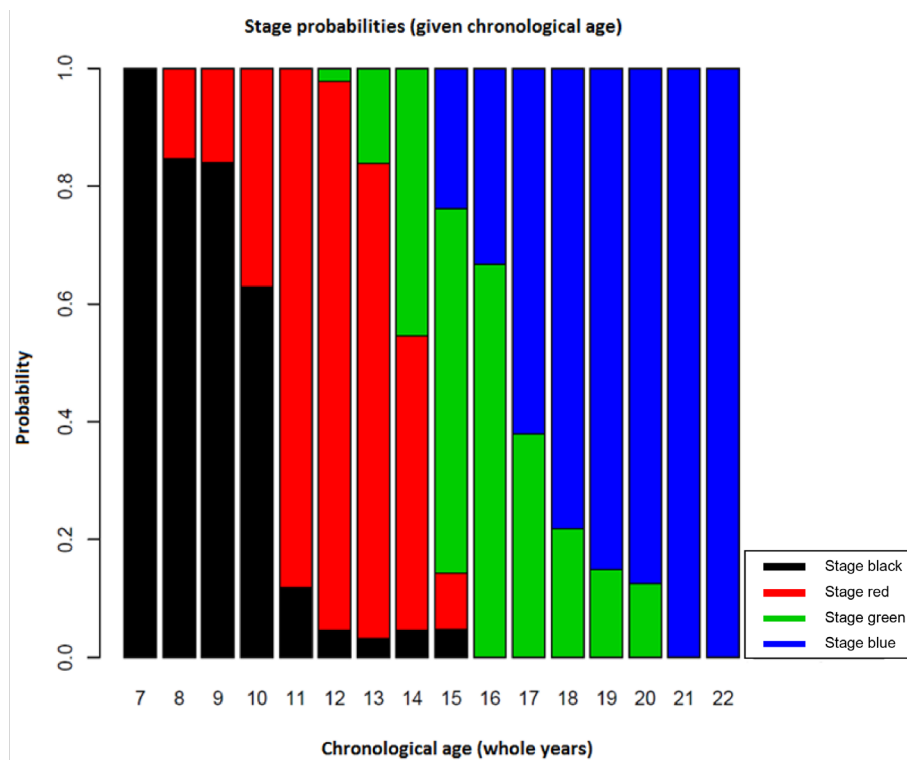

Figure 2. The figure shows the proportions of individuals (for given whole chronological ages) who fall into the different stages. The height of the column (of a given colour) indicates the proportion of individuals that fall into a given stage (black, red, green or blue). We can use these proportions to fit a probability model directly.

One challenge is that we want chronological age to be continuous, not discrete. We therefore consider a regression model that shows the probability of being in one of the defined stages at a given numerical age (not just whole years). An example of such a model is shown in Figure 3. The value on the y axis indicates the proportion of all individuals aged within -0.5 and +0.5 years of the year given on the x axis who fall into the different stages. We see here that the curves are very irregular and variable, a consequence of the variation in the proportions of individuals who fall into the different stages (for the whole-year age segments).

An alternative to looking at the direct, non-parametric approach of considering proportions is to assume a parametric transition model (1). This model will give the data a smooth function fit. An important assumption for this type of model is that the stages come in order: First comes the first stage, then the next, etc., and finally an end stage that marks it as the final stage in the method's system. What is special about this model based on skeletal and tooth-development is that it allows for the fact that "the last stage never ends", i.e. there is the same probability of being in the last stage whether you are 25 years or 50 years (given that the other stages are completed). In Figure 4, we see an example of such a transition model. Here we see that the probability of being in a stage is a "smoothed" function of age.
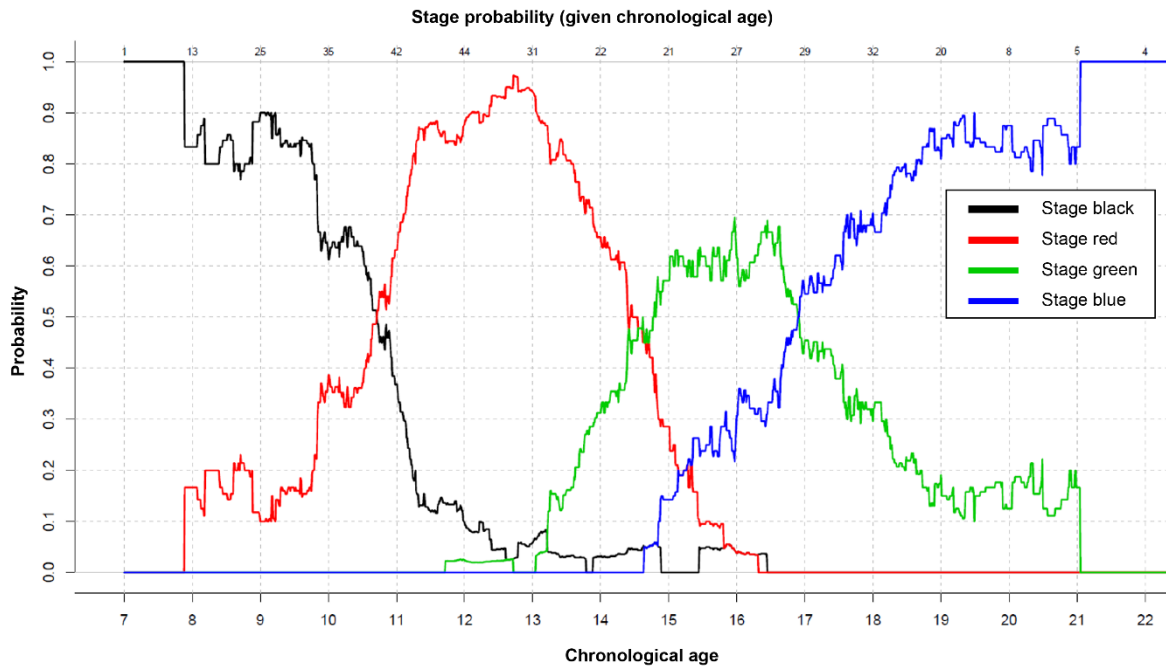
5

Figure 3. The figure shows the proportions of individuals of a given chronological age that fall into the different stages. An age segment spanning one year is considered, with mid-points considered at 7.00, 7.01, 7.02 years etc. up to 22.00 years. The figures above the curves show the number of individuals for the closest (rounded off) whole year. For example, for age 16 on the x-axis, all individuals aged 15.5 to 16.5 years are considered. This is a total of 27 individuals. Here, 5% of these individuals fall into stages A and C, while 30% and 60%, respectively, fall into stages E and G. The total of these proportions is 1.
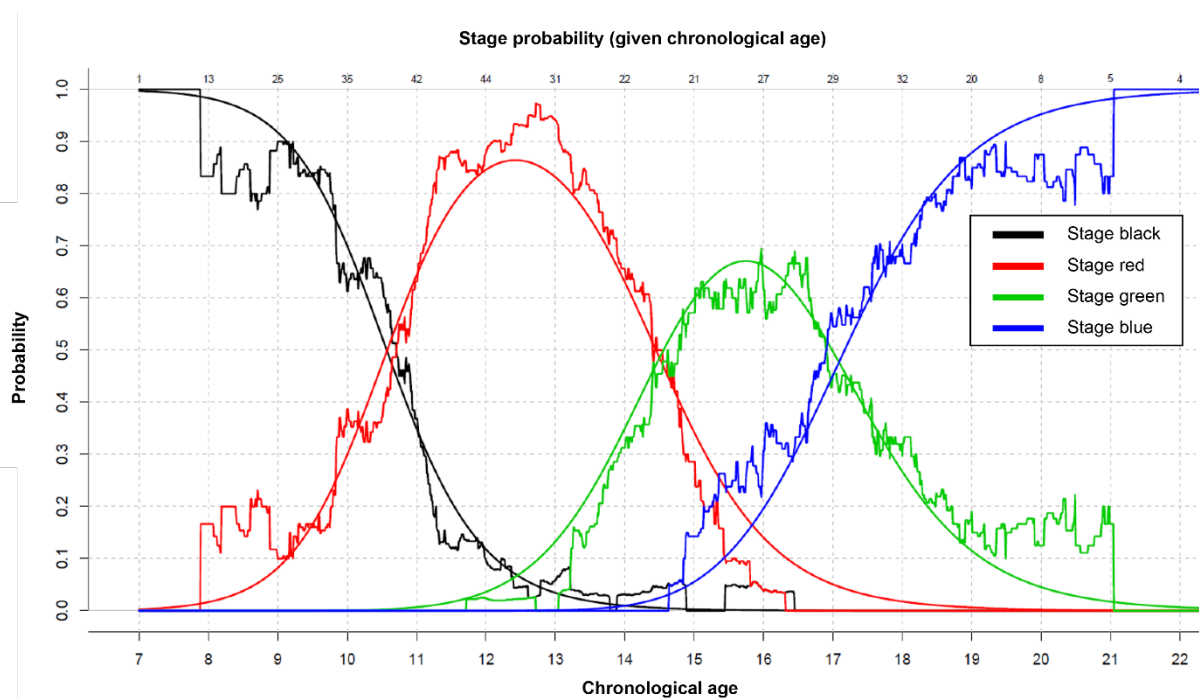
Figure 4. The figure shows the directly observed proportions (uneven curves) together with the stage probabilities from a fitted transition model (smooth curves).

6

In order to check that this parametric model is consistent with the data, it should be checked that the transition model tallies with respect to the percentages of individuals that fall into different stages (check that these curves are consistent with one another).

It should be noted that it is not self-evident that the transition model is better than the empirical model in the prediction of new individuals, but as a rule a "simple" described model is most appropriate. That is to say, a model that is adapted with as few parameters as possible but that can still explain the data. Using the percentages of individuals in the various stages as a non-parametric model for stage probabilities can be regarded as a model with very many parameters, since a probability is assigned to each age segment that is moved. This is not the case for the transition model, which is a great advantage when it comes to describing the mechanism underlying the data.

In order to be able to use a transition model, we need information on chronological age (preferably on a date scale) and observed stage for each individual. Unfortunately, the studies do not publish these data. They are typically provided in summarised form: for example, the mean and standard deviation of skeletal age for groups of individuals in various age segments, mean and standard deviation of chronological age for given stages/skeletal ages. A major challenge in the development of this method has thus been to recreate the data for each individual, by means of an extra layer of modelling. This is essential to enable individuals from the different studies to be combined into a single model.

A.3   Combining studies

In this tool we assume that the individuals from all studies follow a common parametric transition model, and that the parameters for this model are the same across all studies. This means that we assume that the individuals in these studies come from the same population and have the same distribution. This enables any differences between the studies to be "smoothed over".

One objection to the defined model is that it does not make allowance for study heterogeneity, with the result that "outsider" studies can influence the final model to a greater extent than is the case for the underlying effect that is common to all studies. What is of importance to the common model is the information on the stages that the majority of individuals (of a given age) across the studies as a whole are in. A study with a large number of individuals with a certain type of development will thus have more weight than a study with a small number with a different type of development. See section **A.5.1** for an overview of different transition models.

A.4 Modelling of individual data

Table 1. The table is from the manual, and provides an overview of the different types of data format used by various relevant studies

Data formats included in BioAlder

Type 1	data in individual-based format
Type 1b	individual-based data taken from point-plots in published studies
Type 2	frequency table with number of individuals for each stage within each whole year
Type 3	tables with data on means and standard deviations of chronological age for given stages (skeletal age or tooth formation stage)
Type 4	tables with data on means and standard deviations of chronological age and skeletal age within each whole year

In order to fit transition models, we need individual data with information about the stage and chronological age of all individuals in the studies included in the model. As mentioned previously, we do not have this information for many of the studies. In order to recreate this information, we therefore carry out an extra step of modelling of individual data. This is done slightly differently for hand and tooth, since they yield information about stage development (for groups of chronological ages) in different ways. Common to both is that we generate a dataset consisting of individuals with the same number of observations as in the tables (applies to Types 2–4). This dataset generation is iterated multiple times, to take account of the inherent uncertainty of not knowing the actual chronological age and development stage (tooth stage or skeletal age) of each individual.
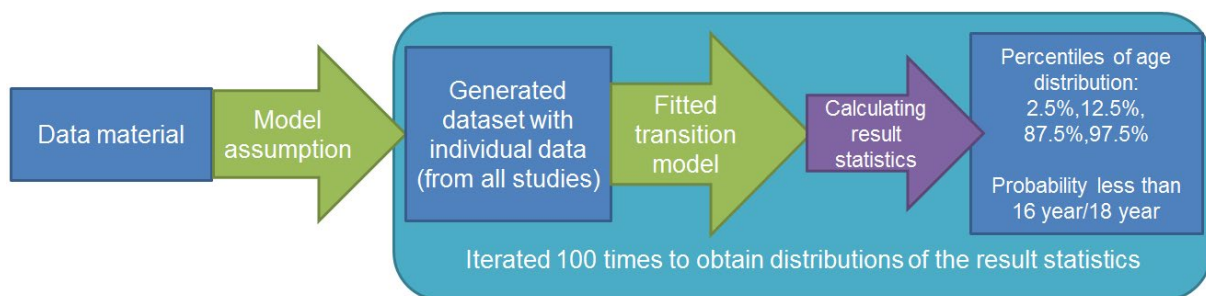

Figure 5. The figure shows the steps from the data material whereby we can use model assumptions to generate a dataset with only individual data (from all the studies combined). A stage probability model (transition model) is first fitted for each dataset generated, and then results are calculated on the basis of this model. This is iterated 100 times in order to obtain a distribution for all the result statistics.

The uncertainty of not knowing the actual chronological age with appurtenant development stage is taken account of by iterating the generation of the complete dataset 100 times. Each time a transition model is fitted to form a basis for the result statistics, e.g. the probability of the chronological age being under 18 years for a given observed stage (see Figure 5). These 100 iterations generate a distribution of the various result statistics we are interested in. In section **A.7.4:** "Choice of results as a consequence of the fact that the data are generated", we explain further what we do to arrive at the results calculated by the tool. We will now provide a description of the modelling of individual data.

A.4.1 Modelling of type 2 data

These data lacked only the exact chronological age of each individual. In order to recreate individual data for these studies, we assume that the individuals within a given age segment (e.g. 12 and 13 years) are uniformly distributed in this segment. In practice this means that we generate a chronological age that may have any value within this age segment with equal probability.

A.4.2 Modelling type 3 data

For the studies that report results in this format, we lacked only the chronological age, as all individuals in the same group have identical skeletal age (or tooth stage). Each individual must be assigned a chronological age. These ages are assumed to be normally distributed, and the data give the means and standard deviations (CA_mean and CA_SD in table 6) which are used to assign chronological ages to all individuals in each of the rows in the table.

A.4.3 Modelling type 4 data

Here we have neither the chronological nor the skeletal age of the individual, only grouped data. Skeletal age is defined in what we call discrete stages, which means that there are no values between the different stages (for example, an individual is either in the 17-year stage or in the 18-year stage). However, chronological age is a continuous scale where an individual can, for example, be 17.36 years old.

The data in Table 2 give only the mean (SA mean) and standard deviation (SA sd) for skeletal age, and these are used to define how the discrete stages are distributed:
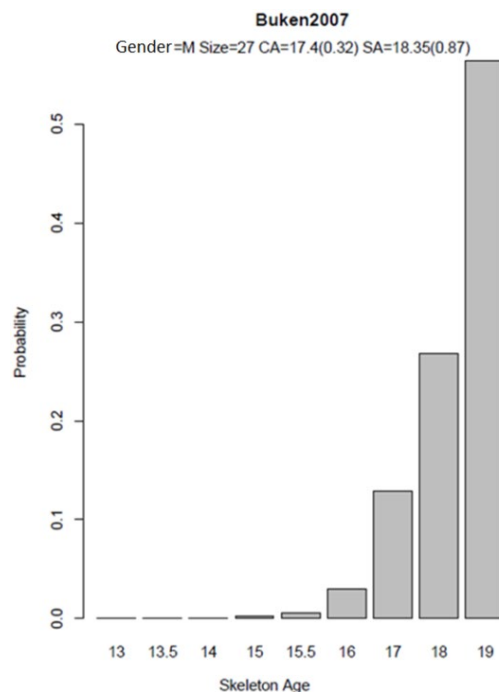

Figure 6. Example of modelling of type 4 data.

This distribution of skeleton age is used in its turn to assign a skeletal age to each individual. These individuals must also be assigned a chronological age. These data are given as mean (CA mean) and standard deviation (CA sd) in Table 2, and we assume them to be normally distributed. We use the

9

correlation value from the tableto assign a chronological age to each individual with the value of the individual's specified skeletal age as the starting point (see Appendix part A.4.2 for more information).

The specified chronological and skeletal ages for a row in a table of the study (Buken 2007) (2) are illustrated in Figure 7.
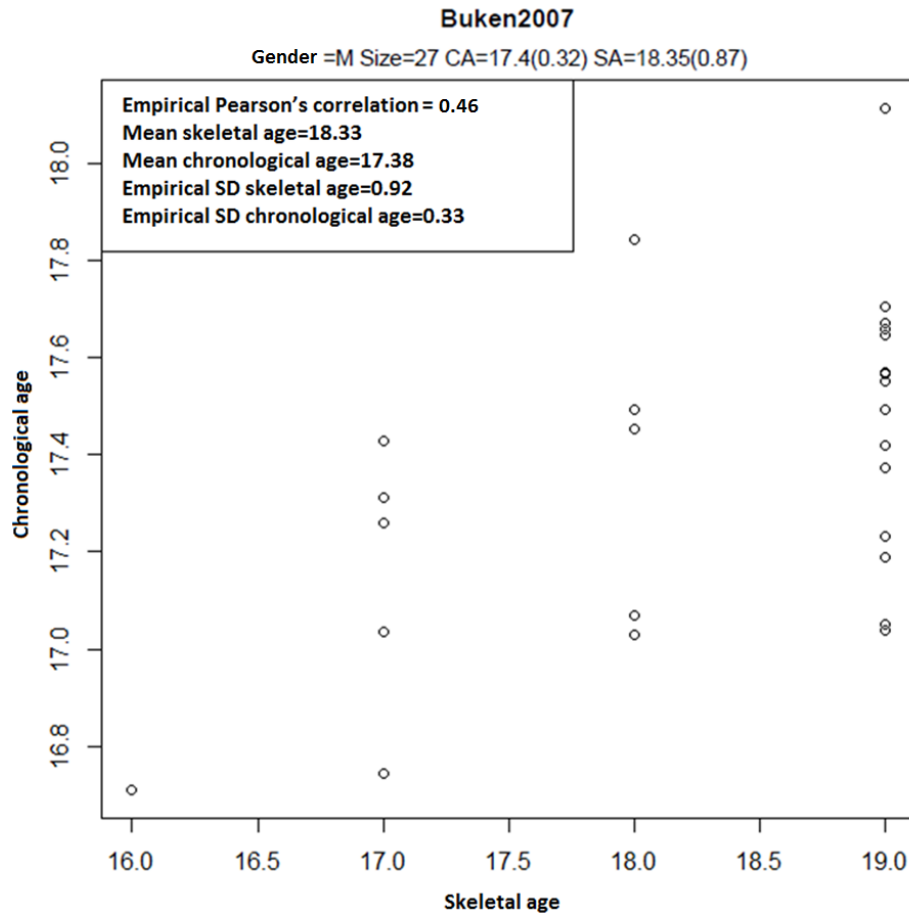

Figure 7. Example of generating 27 random individuals for the age group boys 17 years old for Buken 2017.

This method is applied to all the rows for all the study tables generating a dataset containing information of the skeletal and chronological age for each individual, whom the tables are based on.

A.4.4   Modelling of tooth data (type 2)

As described in the manual, there are only seven datasets that provide complete tooth data for individuals (Type 1). We received these data from other research communities, and have consent to use them. Six studies contain tables (Type 2) showing how many individuals in a whole-year age segment fall into the different tooth stages. In order to recreate individual data for these studies, we assume that the individuals within a given age segment (e.g. 12 and 13 years) are uniformly distributed in this age segment. In practice this means that we generate a chronological age that may have any value within this age segment with equal probability.

10

A.4.5 Modelling of hand data (Types 3 and 4)

As stated in the manual, the formats for hand data are of four different types: Individual-based (Type 1), extracted plot points (Type 1b), age distribution for observed skeletal ages (Type 3) and distribution of skeletal age for segments of whole chronological years (Type 4). For Type 3 we assume that the chronological ages of the individuals in the study are normally distributed for the given skeleton stages, with expectation and standard deviations as given directly by the tables in the articles.

In the case of Type 4, no information on the actual skeletal and chronological ages per individual are given. We therefore have to make extra model assumptions in order to recreate these data. As stated in the manual, Type 4 data are specified as shown in Table 2. The individuals for a study are first grouped according to chronological age, and then the whole group's skeletal age is presented as mean and standard deviation. In order to recreate the skeletal and chronological ages for a given row in Table 2, we first fit a model for the skeletal ages and then assume a model for chronological age given skeletal age. We make use of the fact that we know the correlation coefficient (Pearson).

Table 2. The table shows a section of the information specified for Type 4 data (for the study Buken 2007). The data provided are the mean and standard deviation of both skeletal age (SA_mean and SA_sd) and chronological age (CA_mean and CA_sd) with Pearson's correlation coefficient (Corr) for groups of individuals (of number Size) segmented on whole chronological years (Age).

Study	Age	Size	SA_mean	SA_sd	CA_mean	CA_sd	Corr
Buken2007	15	28	16.39	1.56	15.5	0.3	0.027
Buken2007	16	32	17.45	1.19	16.49	0.28	0.438
Buken2007	17	27	18.35	0.87	17.4	0.32	0.459
Buken2007	18	18	18.44	1.19	18.47	0.29	0.238
Buken2007	19	23	18.95	0.2	19.43	0.29	0.296

Step 1: **Model assumption for skeletal age:**

Note that we know only the mean and standard deviation of the discrete variable skeletal age. Because the latter is discrete, we fit a discrete model to it, which results in a probability for each of the skeletal ages. We calculate these probabilities as follows:

We assume a continuous normal distribution for the "underlying" distribution of skeletal age (prior to discretisation), assumed to be defined from skeletal age zero years. We then consider all the discrete skeletal ages that are defined in the GP atlas (see Table 3 for a subset of these). All the studies grade skeletal age on the basis of this atlas. To arrive at the probability of, for example, the discrete stage "17", the area under the normal distribution from 16.5 to 17.5 years is calculated. This is done for all the discrete skeletal ages such that a discrete model of skeletal age is constructed. We fit the discrete model so that the expectation and standard deviation of the model are equal to the empirical mean and standard deviation of skeletal age for a given row in the study table (see Figure 6. Example of modelling of type 4 data. See Figure 8 for an illustration in which row 3 of Table 2 is considered).

Table 3. The table shows the defined skeletal ages (from 5 years) from the Greulich & Pyle Atlas for both genders.

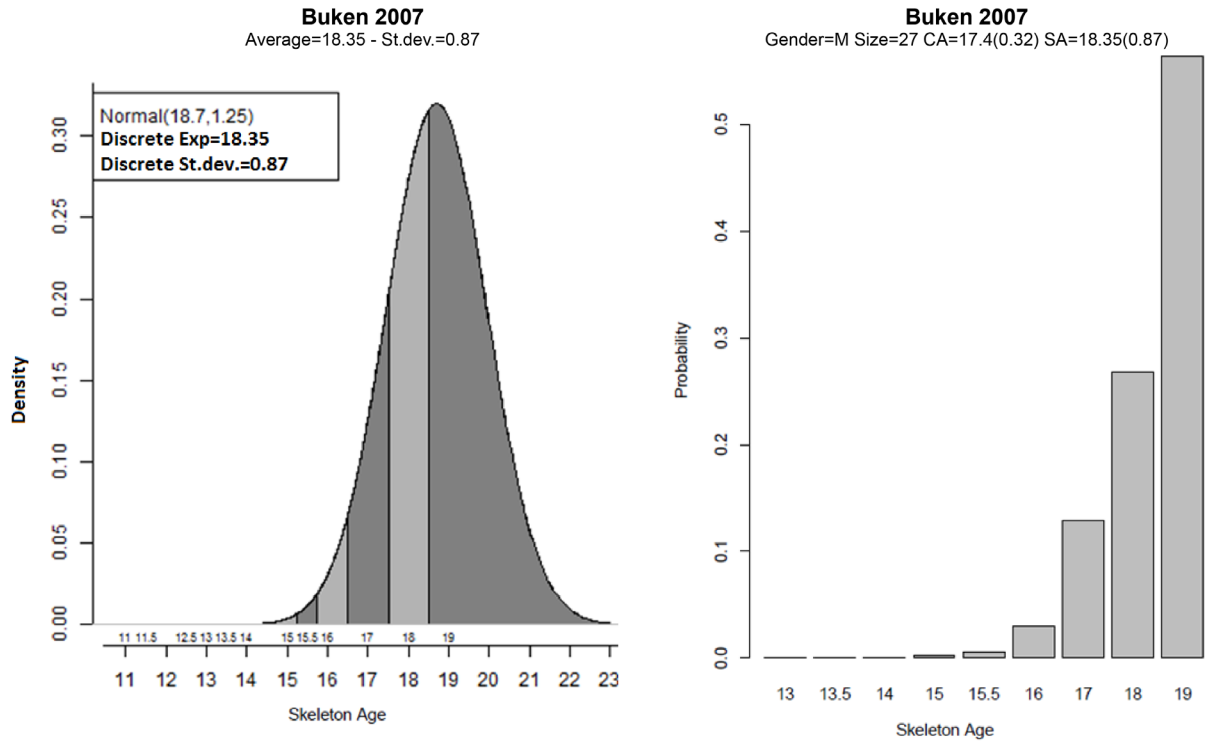| Boys	5	6	7	8	9	10	11	11.5	12.5	13	13.5	14	15	15.5	16	17	18	19
Girls	5	5.75	6+5/6	7+5/6	8+5/6	10	11	12		13	13.5	14	15		16	17	18	

Figure 8. The left-hand plot shows how a continuous "underlying" normal distribution of skeletal age is adapted such that the discrete distribution of skeletal age has the same expectation and standard deviation as given in the table. The right-hand plot shows an adapted probability model of skeletal ages. The example is taken from Buken 2007 (2) (age 17 years).

A.4.5.1 Mathematical description of Type 4 model

The model we have described is a "continuous latent response variable": Here we assume that the observed discrete skeletal ages are actually a categorised version of an unobserved (latent) continuous variable X. We assume this variable to be normally distributed with the unknown parameters expectation μ and standard deviation σ.

$$X \sim N(\mu, \sigma^2)$$

The probability of observing skeletal age s is then defined by using the cumulative distribution of X, F_X, where s^+ is the defined skeletal age after s, while s^- is the defined skeletal age before s (see **Table 3**). The interval we consider for a given skeletal age s is $[\frac{s+s^-}{2}, \frac{s+s^+}{2}]$, which gives the probability:

$$P(S = s|\mu, \sigma) = F_X\left(\frac{s + s^+}{2}\right) - F_X\left(\frac{s + s^-}{2}\right)$$

Note that $F_X\left(\frac{s+s^+}{2}\right) = 1$ for the last defined skeletal age and $F_X\left(\frac{s+s^-}{2}\right) = 0$ for the first defined skeletal age. We fit the probability model for the discrete skeletal ages ($P(S = s)$) by choosing the parameters μ and σ such that

$$E[S = s|\mu, \sigma] = \sum_s s * P(S = s|\mu, \sigma) = SA_{mean}$$

12

$$Var[S = s|\mu, \sigma] = \sum_s (s - SA_{mean})^2 * P(S = s|\mu, \sigma) = SA_{sd}^2$$

where SA_{mean} and SA_{sd} are the empirical mean and standard deviation of skeletal age (based on a given number of individuals "Size") which is given in the row from a study table (see **Table 2**).

In technical terms, the choice of μ and σ is made by minimising the function

$$f(\mu, \sigma) = (E[S = s|\mu, \sigma] - SA_{mean})^2 + \left(Var[S = s|\mu, \sigma] - SA_{sd}^2\right)^2.$$

Step 2: **Model assumption for chronological age given skeletal age**

Let "s" be a generated skeletal age for an individual from Step 1. Given an observed skeletal age "s", we assume that chronological age is normally distributed with expectation and variance as

$$E[A|S = s] = CA_{mean} + \frac{CA_{sd}}{SA_{sd}} * Corr * (s - SA_{mean})$$

$$Var[A|S = s] = (1 - Corr^2) * CA_{sd}^2$$

Where CA_{mean} and CA_{sd} are the empirical mean and standard deviation of chronological age (based on a given number of individuals "Size") which is entered in the row in a study table (see **Table 2**). $Corr$ is Pearson's correlation coefficient between chronological age and skeletal age (based on a given number of individuals, "Size", for similar rows).

The underlying assumption is that skeletal age and chronological age are bivariate normally distributed. This means that the expected chronological age increases if the generated skeletal age was higher than its expectation (for positive $Corr$), and reduces the variation of the chronological age if $Corr$ is not zero. An illustration of generation of individual data based on **Stage 1-2** is given in Figure 9.

Steps 1–2 are carried out for each row in the table (see **Table 2**) using the numbers in the table. This yields a complete dataset with information on skeletal age and chronological age for each individual. A model to describe the probability of being at a specific stage at a given age is fitted on the basis of such a complete dataset.
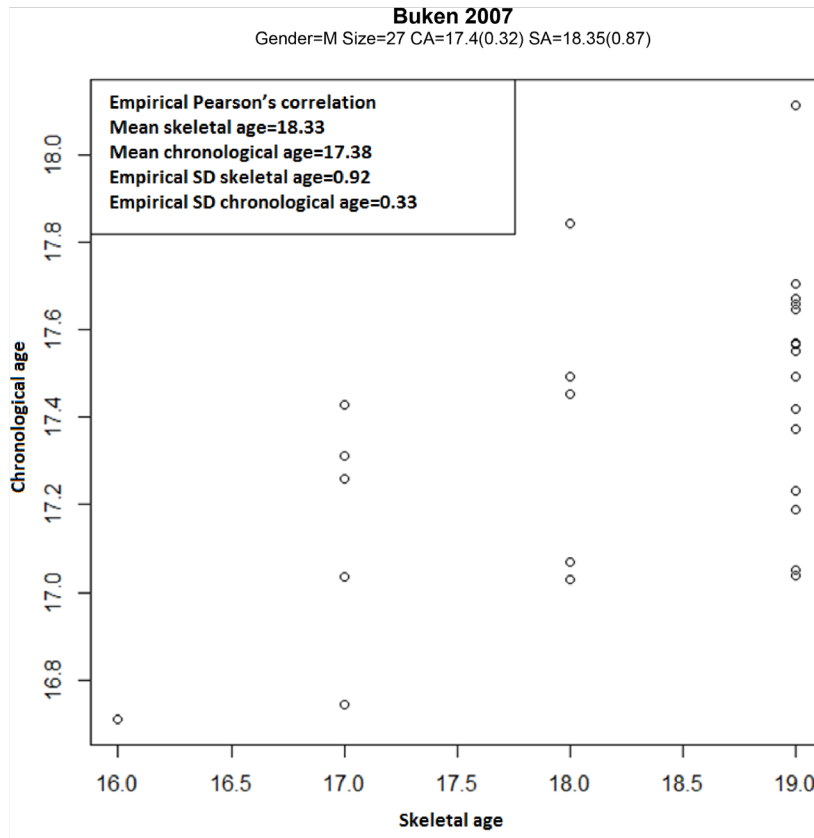
13

Buken 2007
Gender=M Size=27 CA=17.4(0.32) SA=18.35(0.87)

Figure 9. Example of generating 27 random individuals for the age group boys 17 years old for Buken 2007.

A.5 Model chosen for the stage probabilities in the tool

Boldsen et al. (2002) (1) describe two different types of transition models for modelling ordinal discrete variables as responses in a regression model. In the BioAlder tool, we consider several candidate models of this type in order to model the stage probabilities as a function of chronological age. Although these are very similar in form, we still want the data to tell us which variant of the models is most suitable for the different genders (boys or girls) and methods (hand or tooth). For example, it can be assumed that the chronological age variable is on a log scale (3), i.e. that we have predefined an age transformation. In our approach we allow the data decide the transformation of chronological age, $g(age) = age^{\lambda}$ for $\lambda = 0.1, 0.2, ..., 1.0$. This makes it possible for the stage probabilities to be asymmetrical about chronological age (as for a log transformation).

A.5.1 Overview of candidate models

Consider the ordered stages $j = 1, ..., J$, chronological age x and the parameters $\theta = (\alpha_1, ..., \alpha_{J-1}, \beta, \lambda)$. We assume that λ takes the values $0.1, 0.2, ..., 1.0$. By letting Y be a discrete stochastic variable with the stage outcomes $1, ..., J$ that an individual of age x may be in, we can describe the candidate models (1–4) as follows for the stages $j = 1, ..., J - 1$:

1) Proportional-odds cumulative model with logit link
   a. $logit\big(P(Y \le j|\theta, x)\big) = \alpha_j + \beta * x^{\lambda}$

14

2) Proportional-odds cumulative model with probit link

    a. $probit(P(Y \leq j|\theta, x)) = \alpha_j + \beta * x^\lambda$

3) Continuation-ratio model with logit link

    a. $logit\big(P(Y = j|Y \geq j, \theta, x)\big) = \alpha_j + \beta * x^\lambda$

4) Continuation-ratio model with probit link

    a. $probit\big(P(Y = j|Y \geq j, \theta, x)\big) = \alpha_j + \beta * x^\lambda$

For $j = J$ (last stage) we have $P(Y \leq J|\theta, x) = 1$ and $P(Y = J|Y \geq J, \theta, x) = 1$.

The link function $logit(x) = \log(\frac{x}{1-x})$, while the link function $probit$ is cumulative standard normal distribution.

A.5.2 The likelihood function

In order to fit these models to the data, maximum likelihood estimates are chosen for the unknown parameters θ. In other words, they are chosen such that the data are those most likely to be observed. These are found by maximising the likelihood function (on the log scale)

$$L(\theta) = \prod_{j=1}^{J} \prod_{i=1}^{n_j} P\big(Y = j|\theta, x_{i,j}\big)$$

where n_j is the number of individuals in stage j and $x_{i,j}$ is the chronological age of the individual i observed in stage j. Thus the likelihood function is a product of all the probabilities of the individuals (of a given age) being in their given stages.

Although the variants of the transitions models (1–4) are fairly similar, they have different features that it will be useful to take into account: The logit function appears to function better than the probit function in the great majority of cases – a concave maximum of the likelihood function (on the log scale) cannot always be achieved with the probit function.

Since all variants of the transition models (1–4) have the same number of parameters, the final model chosen is the one that gives the best fit with the observed data. The value of the maximised likelihood function is used to measure this. Since the observed data are generated, the candidate model that gives the best fit over the 100 generated complete datasets is chosen. This model is also chosen to calculate the stage probabilities (for a given age) for each of the 100 generated complete datasets that the tool results are based on.

A.5.3 Model fitting

We use the VGAM R package (4) with the **vglm** function to fit the models and the **predict** function to calculate the stage probabilities for a given age. We have also made our own functions in R which

calculate the likelihood function and stage probabilities for a given age, for all the candidate models numerically. We can use the optimising function **nlm** to produce a Hessian matrix in order to be able to describe the asymptotic variation in the estimators of θ (the covariance is the inverse of the negative Hessian matrix). This is very useful for calculating the confidence interval for the model parameters, or the confidence interval for model-based stage probabilities for a given age.

A.5.4 Details of each individual transition model

In section **A.5.1** we defined the models in the form $f\big(P(Y \leq j|\theta, x)\big)$ and $f(P(Y = j|Y \geq j, \theta, x))$ where the link function f was either $logit$ or $probit$. But to calculate the likelihood function we need an expression for the stage probabilities $P(Y = j|\theta, x)$. We now describe this mathematically for each type of model:

A.5.4.1 Proportional-odds cumulative

$$P(Y = 1|\theta, x) = f^{-1}(\alpha_1 + \beta * x^\lambda)$$

$$P(Y = j|\theta, x) = f^{-1}\big(\alpha_j + \beta * x^\lambda\big) - f^{-1}\big(\alpha_{j-1} + \beta * x^\lambda\big) \quad \text{for } j = 2, \dots, J - 1$$

$$P(Y = J|\theta, x) = 1 - f^{-1}\big(\alpha_{j-1} + \beta * x^\lambda\big) = 1 - \sum_{j=1}^{J-1} P(Y = 1|\theta, x)$$

A.5.4.2 Continuation-ratio

$$P(Y = 1|\theta, x) = f^{-1}(\alpha_1 + \beta * x^\lambda)$$

$$P(Y = j|\theta, x) = f^{-1}\big(\alpha_j + \beta * x^\lambda\big) \prod_{l=1}^{j-1}[1 - f^{-1}\big(\alpha_{l-1} + \beta * x^\lambda\big)] \quad \text{for } j = 2, \dots, J - 1$$

$$P(Y = J|\theta, x) = 1 - \sum_{j=1}^{J-1} P(Y = 1|\theta, x)$$

A.6 Model validation

As mentioned above in this Appendix, it is important to check the fitted parametric model against the actual observed data in order to see that it is consistent with the model assumptions (see Figure 4**)**. This is done by comparing the non-parametric model (with one-year age segments) with the adapted transition model. This will be useful documentation of the fact that the underlying model that is used to indicate the uncertainty of chronological age (given observed stage) is consistent with the underlying data. This comparison was carried out for all stages of the hand and tooth methods for both genders.

16

A.7   Distribution of chronological age given observed stage

A.7.1   How to work backward to the age distribution

So far we have only considered the description of the stage probabilities for given chronological ages. Our reason for this is to take account of the effect of age mimicry. Our purpose with the tool is to describe how chronological age is distributed for a given observed stage. In order to find this distribution, we use Bayes' theorem (see section **A.7.1.1** for a mathematical description) as follows:

Age distribution given stage = 'stage probability (age)' ∗ 'assumed age distribution' ∗ constant

where "constant" is a numerical value such that the area of the posterior distribution "Age distribution given stage" adds up to one. Thus the age distribution for a given stage consists of two main parts that are multiplied together: The probability for stage (a function of age) that we fit on the basis of a transition model, and an assumed age distribution (a priori distribution). This corresponds to the definition of which chronological ages (or rather, their distribution) we choose that the individuals included in a study should have (this presupposes, of course, that we want to describe chronological age for a given observed stage).

In the tool, we assume that the assumed age distribution (a priori distribution) is uniformly distributed and defined, for example from 7 to 21 years, with a view to take account of the effect of age mimicry. In Figure 10 we see an example of how age distribution (posterior distribution) for a given stage is affected by how one defines assumed age distribution. We see here that the age distribution for the given stage is cut off at 21 years since this is the assumed upper age.
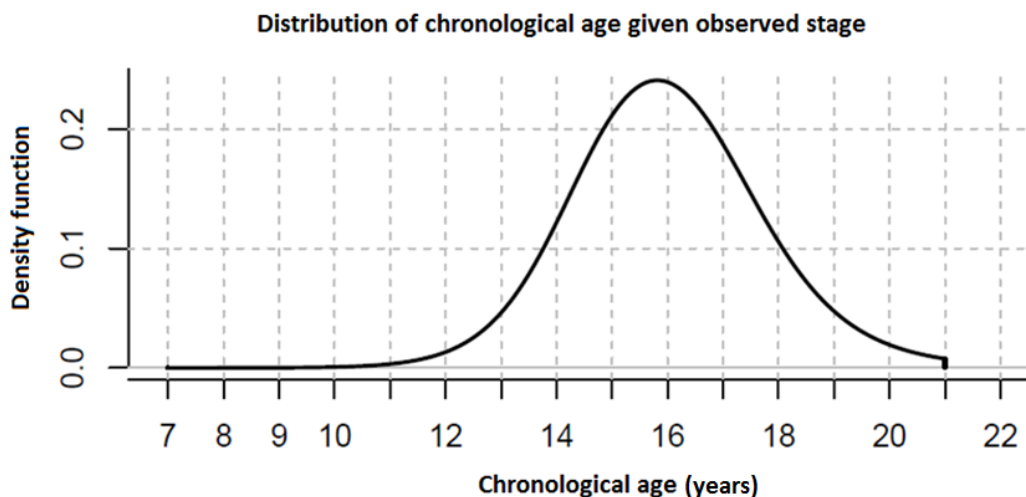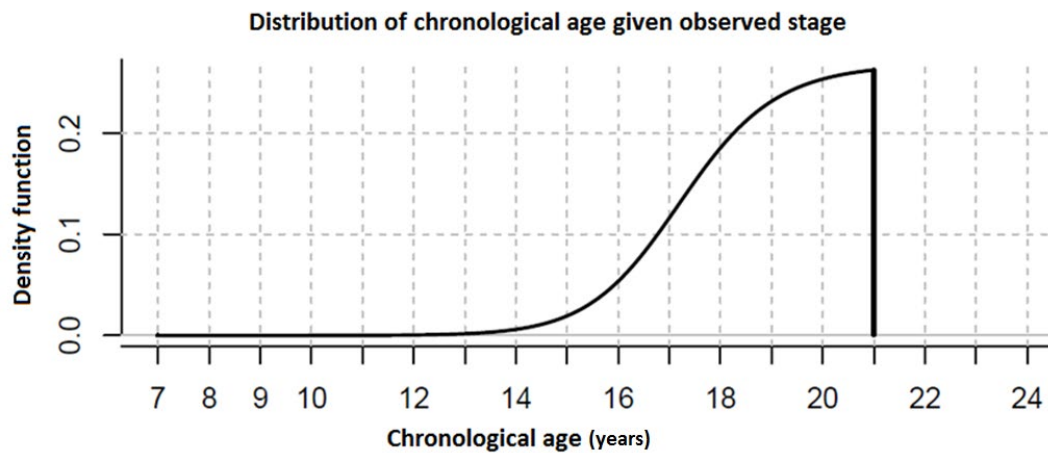

Figure 10. The figure shows age distribution for a given stage based on Bayes' theorem, with the age distribution defined as from 7 to 21 years.

Consider the stochastic variables X and Y and that the outcome of these $(x$ and $y)$ is given with probability $p(X = x, Y = y)$. Assume that we have the observed outcome x. The rule for calculating the conditioned probability for Y given observed x is given as

$$P(Y = y | X = x) \; = \; p(X = x, Y = y)/p(X = x)$$

It is worth noting that $p(X = x)$ is a constant such that $P(Y = y | X = x)$ adds up to one. The only thing that varies is the variable y. Using this rule, we can also show that

$$p(X = x, Y = y) \; = \; p(X = x) * P(Y = y | X = x) \; = \; P(Y = y) * P(X = x | Y = y).$$

Thus we also have Bayes' theory, which is a reformulation of this:

$$P(Y = y | X = x) \; = \; P(X = x | Y = y) * P(Y = y) / P(X = x).$$

This gives us $f(y) \; = \; P(Y = y | X = x) \; = \; P(X = x | Y = y) * P(Y = y) * constant$

Thus by defining $P(X = x | Y = y)$ and $P(Y = y)$, we can calculate $P(Y = y | X = x)$.

A.7.2   Calculated results based on age distribution

Our aim is to use the model-fitted age distribution (see previous section) to yield the age variation for an observed stage. As mentioned in the manual, we choose to define this age variation as 75% and 95% age prediction intervals. In order to estimate these, we estimate the 2.5%, 12.5%, 87.5% and 97.5% percentiles of the age distribution (for a given stage). We are also interested in finding the probabilities that the age of an individual is less than 16, 17 or 18 years. These probabilities are found by calculating the areas under the age distribution up to 16, 17 and 18 years, respectively.

The following is an overview of the results statistics produced by the tool:

1) The estimated 2.5, 12.5, 87.5 and 97.5 percentiles of the age distribution.
2) The areas under the age distribution curve up to 16, 17 and 18 years. These give the estimated probabilities of individuals being under 16, 17 and 18 years, respectively.

In the next section, we describe how these are calculated by the tool.

We calculate $P(Stage = j | Age = a)$ for age $a = 7.00, 7.01, 7.02, \dots, 26.99, 27.00$, i.e. with a grid size of 0.01. Then

$$P(Age = a | Stage = j) = P(Stage = j | Age = a) * C(u)$$

Is calculated for $a = 7.00, 7.01, \ldots, u - 0.01, u$ where u is the upper defined age limit in whole years (e.g. 21.00 or 23.00 years) and $C(u)$ is a constant that depends on this upper defined age limit:

$$C(u) = 0.01 * \sum_{i=7.00}^{u} P(Stage = j | Age = i)$$

which is calculated with a simple rectangular approximation to the integral.

The cumulative distribution of the age distribution given stage is calculated by simple summation:

$$P(Age \leq a | Stage = j) = \sum_{i=7.00}^{a} P(Age = i | Stage = j)$$

for $a = 7.00, 7.01, 7.02, \ldots, 26.99, 27.00$.

Overview of calculated result statistics:

1) The probability of age less than T years will then be $P(Age \leq T | Stage = j)$
2) q-percentile = $argmax_a \, P(Age \leq a | Stage = j) \leq q$

i.e. the highest age of $a = 7.00, 7.01, \ldots, u$ where $P(Age \leq a | Stage = j) \leq q$.

A.7.3   Effect of assumed age distribution

As described in the previous section, the approach for describing the age distribution for a given stage is as follows: First define probabilities for the given stage for the outcome of age values, then assume a prior age distribution in order to "work backwards" to the posterior age distribution for the given stage. An effect that cannot be avoided with this approach, is that the definition of assumed prior age distribution may influence the results that the tool generates (to different extents for different stages). As an example, we consider the last stage of one of the methods. Figure 11 shows how the age distribution for the last stage suddenly stops at age 21 years. This is because we have defined 21 years as an upper age in the assumed age distribution. Since the sum of the area under the age distribution must always be one, both the percentiles and the probabilities of being less than 16/17/18 years will be influenced by the upper defined age: The lower the upper age limit that is defined, the lower the age distribution percentiles will be, and the higher the probabilities of being less than 16/17/18 years old. An important part of the manual has been to describe the effect on the results of assuming different upper age limits, and to argue for the choice of the upper age limit we define.

Distribution of chronological age given observed stage

Figure 11. The figure shows the age distribution for a given stage based on Bayes' theorem, with the age distribution defined as from 7 to 21 years. The x-axis is chronological age in years.

In the study of any phenomenon, one often has some knowledge in advance, and in Bayesian statistics, this knowledge is called "prior information". Such information says something about what the result may become before the data is collected, and this could for example be data from previous studies on tooth- and hand investigations.

In BioAlder, a uniform prior was chosen (Figure 12). This means that the probability of being 15 is equal to the probability of being 18, for example. The relevant ages is from 7 to 20.5 years for men, and 7 to 21 years for women, when both methods are included. One could, of course, argue that other prior distributions would be better, and this would affect the predicted age interval for each stage, or combination of such. One distribution that we tested, was a normal distributed prior with expectation 18 years of age and with a standard deviation of 2 years (as seen in Figure 12). Some have also argued that it would be more appropriate to adapt the prior distribution in each individual case, based on other available information (5). E.g., if other circumstances point towards a chronological age of 21 years, expectation of the prior could manually be set to 21 years with a standard deviation that reflects the uncertainty. If the prior distribution were set to an age above 18 years, this would push the prediction interval towards higher ages, making it more likely for someone to be over 18 years old.

The prior probability distribution can also be used to increase the positive predictive value (PPV). This would be the case if the majority of the test population was under 18 years, as recently problematized (6, 7). This criticism is, however, impossible to fully accommodate, as the chronological age of the test population is, and will always be, unknown. The best way to avoid a low PPV will be to only test individuals where there is doubt whether the person is under 18 years.

20
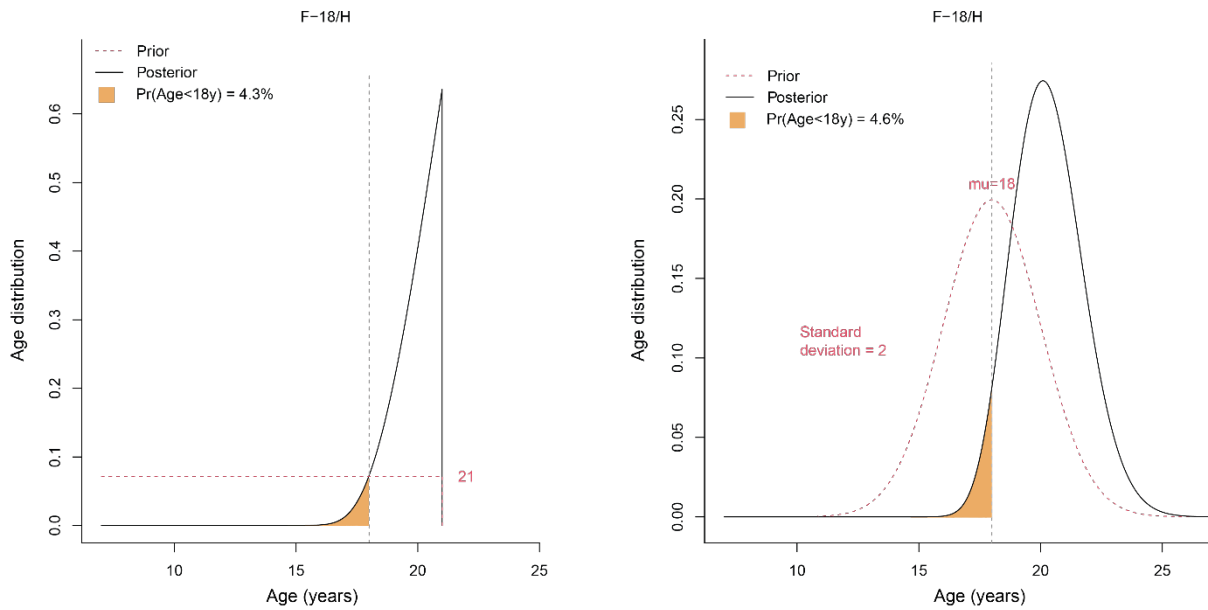
Figure 12. The figure shows two different prior age distributions. The first one is uniform, as it is in BioAlder, and the second with a normal distribution around 18 years with 2 years standard deviation. The number of individuals under 18 years will be slightly different between the two; 4.3% and 4.6%.

A.7.4 Choice of results as a consequence of the fact that the data are generated

As previously explained in section **A.4**, for most of the studies we do not know the actual chronological age with appurtenant development stage for each individual (individual-based data). The approach we have chosen for using the information in the studies (in the form of tables) is to generate complete datasets with extra model assumptions in order to recreate the individual-based data that these studies are based on (as described in section **A.4**). By going through all the rows in the tables in the studies, a complete dataset is generated (which also includes individual-based data). The "result statistics" (percentiles for age distribution and areas under the age distribution up to 16, 17 and 18 years) are calculated for this (partially) generated dataset. We iterate this 100 times, so that each of these results statistics gets a distribution. See Figure 13 for an illustration of these distributions.

21
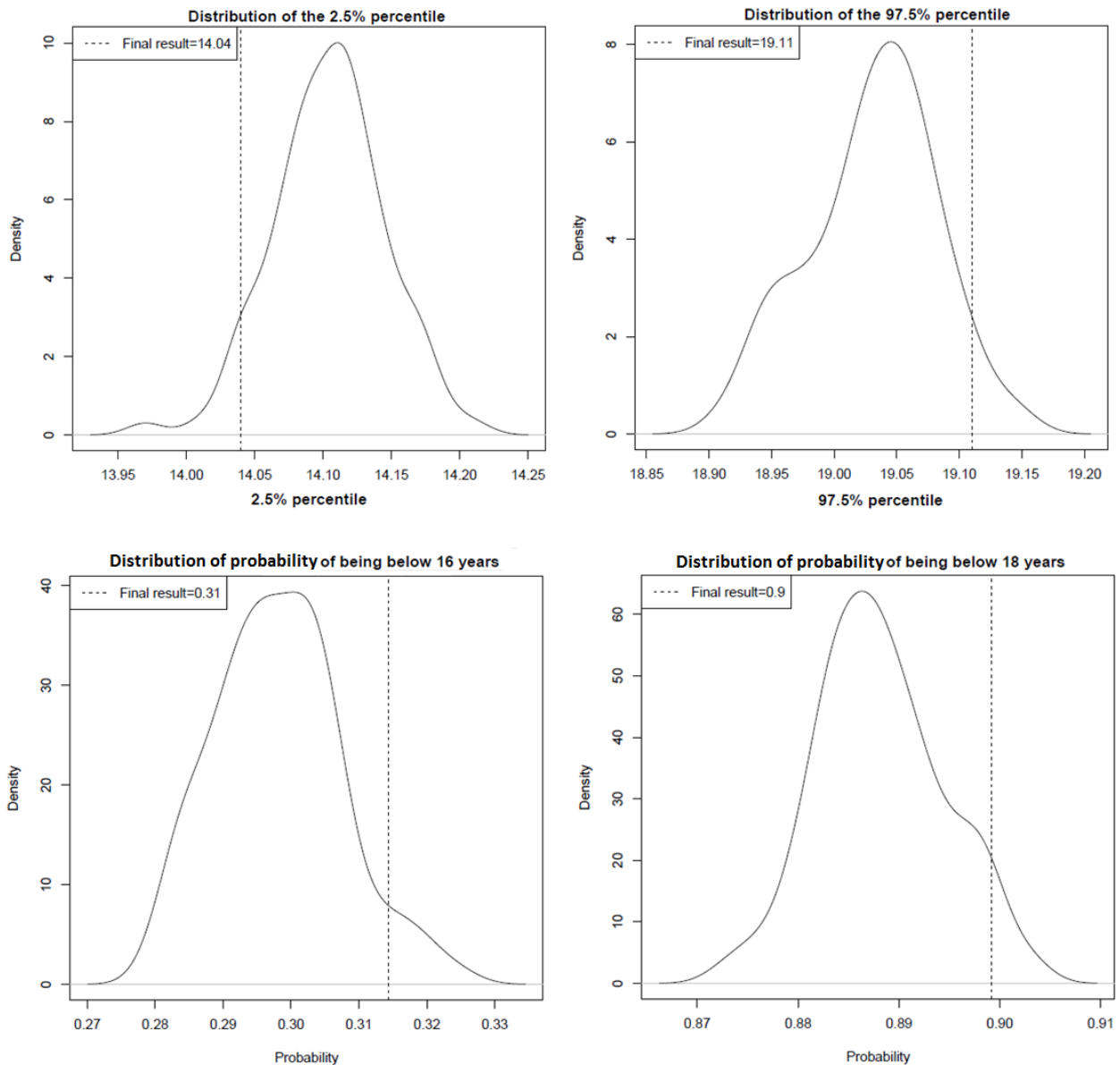
Figure 13. The figure shows the distribution of the 2.5 percentile (upper left plot) and the 97.5 percentile (upper right plot) for chronological age, and the probabilities under 16 years (lower left plot) and 18 years (lower right plot) over 100 generated datasets for an observed stage. The final results are chosen as either 5% or 95% quantiles of the 100 generated results.

For the distributions of the result percentiles that are under 50% (these are defined as 2.5% and 12.5%), we use the 5% quantile of the 100 results as the final result for the tool (see upper left plot in Figure 13). For all the other results statistics (87.5% and 97.5% and the probabilities for under 16, 17 and under 18 years) we use the 95% quantile of the 100 results as the final result (see upper right and two lower plots in Figure 13). The basis for these choices is that the prediction interval (e.g. the 2.5 and 97.5 percentiles) should be broader because we ought to allow for the fact that we do not know the actual individual-based data. For the probabilities of being under 16, 17 and under 18 years, we use the 95 percentiles so that the probabilities are increased in favour of not classifying a minor as an adult.

It is worth noting that the "resolution" of the data formats (whether they are Type 2, 3 or 4) that are included and the number of observations influence the variation of the distribution of the result statistics.

22

A.7.5 Details of calculating the results used by the tool

From the 100 complete datasets, we get 100 values for each of the types of result statistics defined under 1) and 2) in section **A.7.2.** We take either the 5% or 95% quantile for these as described in the previous section by using the function **quantile** in statistics program R, where the type of quantile is specified as "Type 7", which is the default.

A.8 Modelling of stage probability for hand and tooth combined

Because of the great biological variability affecting the methods, it is desirable to combine several methods in order to obtain a more precise estimate. Gelbrich et al. (2015) (8) point out that there is no relationship between the age estimation errors arrived at by means of the hand and tooth of the same individual, and we can therefore assume that the two methods are independent for a given chronological age. This naturally presupposes that the radiographs of hand and tooth are taken at approximately the same time for the same individual. In practice, this means that we can multiply the distributions of chronological age for the two methods together in order to obtain a joint distribution of chronological age based on a combination of development stage for tooth and skeletal age for hand. For a given observed skeletal age s and tooth stage t, the model for the combination can be written

$$P(skeletal\ age = s, tooth\ stage = t | Age = a)$$
$$= P(skeletal\ age = s | Age = a) * P(tooth\ stage = t | Age = a)$$

Applying Bayes' theorem with defined a priori age $P(Age = a)$ we get $P(Age = a | skeletal\ age = s, tooth\ stage = t) \propto P(skeletal\ age = s | Age = a) * P(tooth\ stage = t | Age = a) * P(Age = a)$.

Figure 14 illustrates the distribution of chronological age, given the data for an observed skeletal age and tooth stage (combined).

Distribution of chronological age
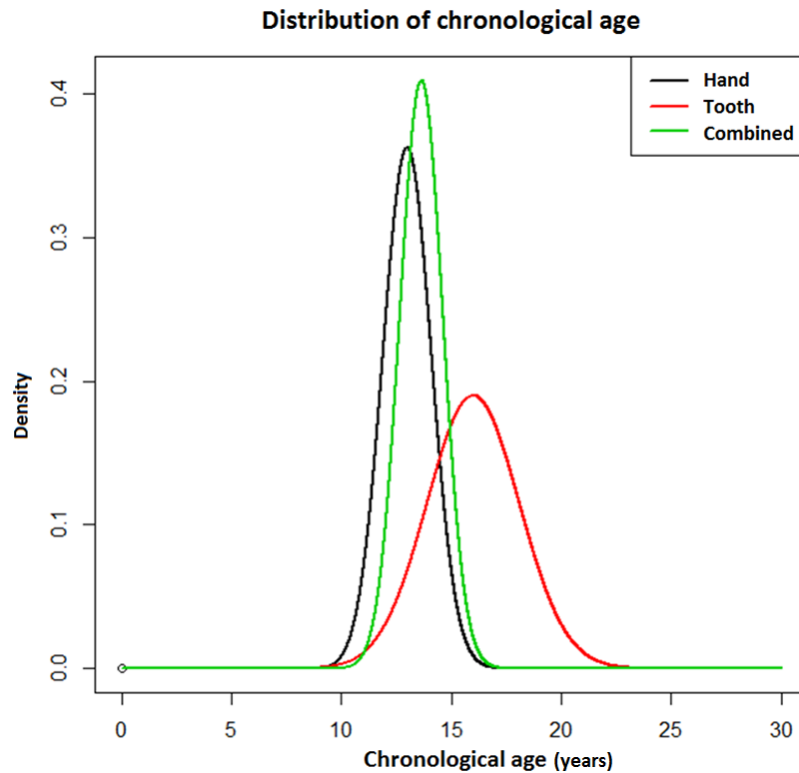
Figure 14. The figure shows how the distribution based one the two methods are combined if they are assumed to be independent of one another for a given chronological age: For each given chronological age, the values of the functions for hand and tooth respectively are multiplied together, and then the multiplied function is normalised to have an area of one.

The combined age distribution under moderate dependency

In this section we carry out a simulation experiment to see how the combined age distribution changes if there is a dependency of 0.3 in correlation between the two methods (given chronological age). We considered a proportional-odds cumulative model (with probit link function) with the parameters $\alpha_1^H, \alpha_2^H, .., \alpha_{I-1}^H, \beta^H$ for hand and parameters $\alpha_1^T, \alpha_2^T, .., \alpha_{J-1}^T, \beta^T$ for tooth, and $\lambda = 1$ for both methods. The parameters were fitted based on the data from all studies, for the 1st data generation:

Parameter	α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8	α_9	α_{10}	α_{11}	α_{12}	β
Hand (H)	8.23	9.03	9.66	10.31	10.77	11.21	11.9	12.28	12.52	13.07	13.8	14.56	-0.81
Tooth (T)	5.35	6.38	7.59	8.58	9.48	10.33	11.85						-0.57

We found that there is not any particular big change for the combined age distribution when there is a correlation of 0.3 compared to zero (independence). The largest difference is when the age distribution for the hand method does not overlap much with the age distribution for the tooth method – here the combined age distribution should slightly more shifted towards the hand method. See Figure 15 for some example variants.

24
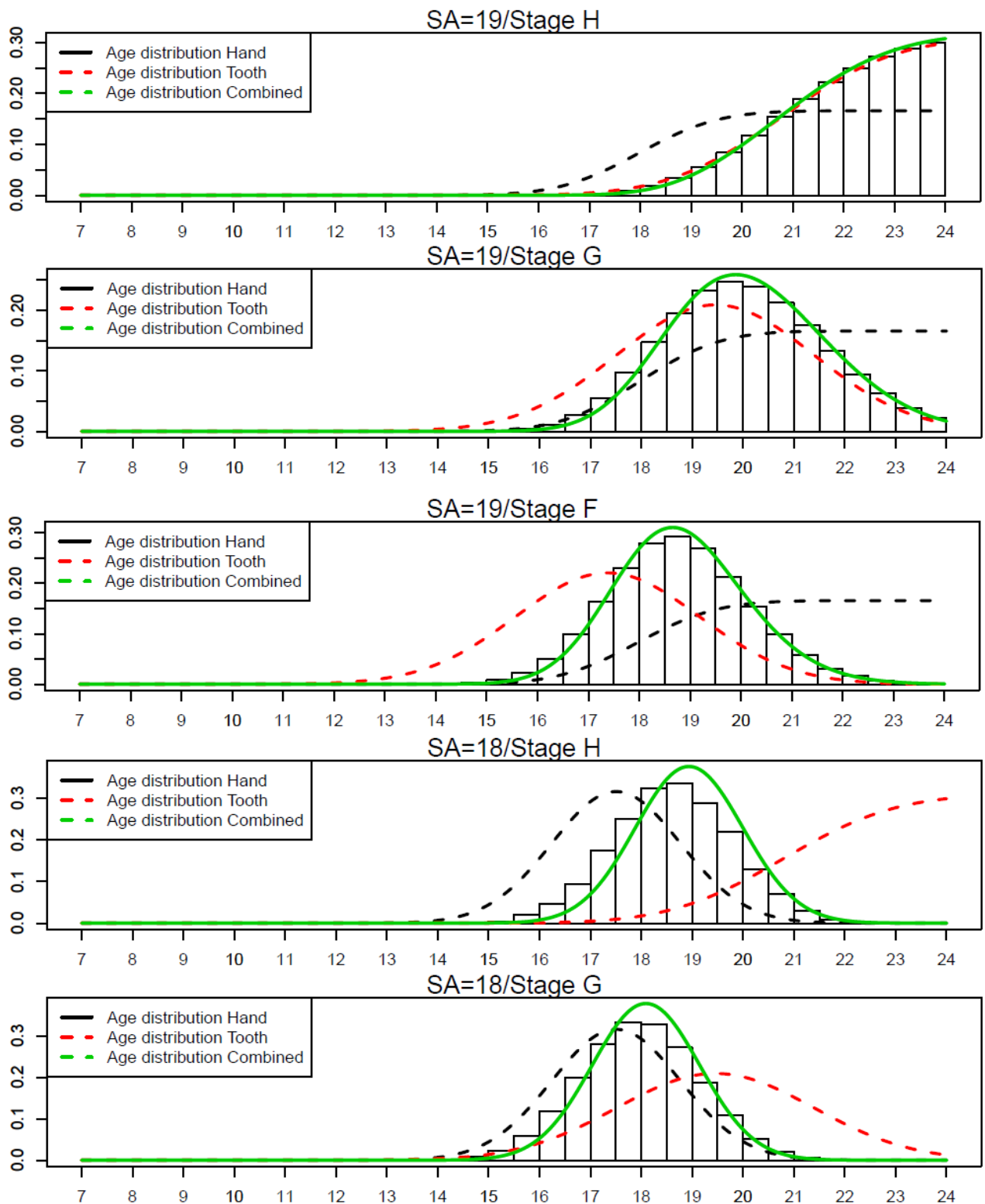
Figure 15. The figure shows the age distribution for some variants of the combined methods. The x-axis gives chronological age (year), whereas the y-axis is the density. The histogram was created based on simulating 10 mill. individuals from a uniform distribution between 7-24 years and applied on the proportional-odds cumulative models. The conditional dependency between the two methods was set to have correlation 0.3.

25

B. Results used in the tool

This chapter gives the user of the tool a thorough overview of all the elements on which the final results shown by the tool are based. The chapter is intended to be documentation to support the assumptions made along the way in order to arrive at these results. A large portion of this chapter will also show the effect of making different assumptions: for example, how to allow for the fact that we do not know the individual-based information for many of the studies, or the choice we make to define an upper age limit for the prior age when using Bayes' theorem.

In the first section, **B.1**, we present the data upon which the results are based. In the next section, **B.2**, we state which model type we found to give the best fit for the different methods for different genders. Sections **B.3–B.5** are "effect sections", in which we demonstrate the model uncertainty and the effect of making different assumptions. The chapter then concludes with sections that describe our final assumptions, which form the basis for the results used in the tool.

B.1 Overview of studies used in the tool

Hand studies

Format	Study	Boys	Girls
Type 1	Saadé 2017	115	129
Type 1	Santos 2011	136	94
Type 1	Van Rijn 2001	178	197
Type 1	Zafar 2010	165	64
Type 1	Tise 2011	359	126
Type 1	Alcina 2018	312	293
Type 1	Yilmaz 2018	333	379
Type 1b	Buken 2007	231	219
Type 1b	Hackman 2013	145	0
Type 1b	Haiter-Neto 2006	115	105
Type 1b	Kim2015	60	40
Type 1b	Paxton2013	112	67
Type 1b	Schmidt2007	172	0
Type 1b	Zabet2015	98	87
Type 3	Chaumoitre 2016	886	673
Type 4	Bala 2010	59	59
Type 4	Cantekin 2012	259	351
Type 4	Chiang 2005	141	70
Type 4	Griffith 2007	281	105
Type 4	Koc 2001	185	0
Type 4	Mohammed 2015	270	270
Type 4	Nahid 2010	32	45
Type 4	Patel 2015	56	60
All	**Total**	**4700**	**3433**

Tooth studies

Format	Study	Boys	Girls
Type 1	Cavric 2016	763	907
Type 1	Malta dataset	553	650
Type 1	Saadé 2017	113	119
Type 1	Jayaraman 2016	682	617
Type 1	Knell 2009	591	669
Type 1	Hedge 2016	410	267
Type 1	Yilmaz 2018	70	92
Type 2	Lee 2009	786	964
Type 2	Johan 2012	540	539
Type 2	Duangto 2017	872	983
Type 2	Li 2012	648	760
Type 2	Liu 2018	963	1148
Type 2	Marrero-Ramos 2020	27	75
All	**Total**	**7018**	**7790**

Table 4. The tables provide an overview of the numbers of individuals in different studies upon which the results in the tool are based (see the reference list of the manual for details). The data format for the appurtenant study is given in the column "Format". Type 1 studies have data in an individual-based format, Type 2 have a frequency table with the number of individuals for each stage within each whole year, Type 3 are tables with information on the mean and standard deviation of chronological age for given stages (skeletal age or tooth stage), while Type 4 are tables with information on means and standard deviations of chronological age and skeletal age within each whole year.

Table 4 provides an overview of all studies upon which the results in the tool are based. The total numbers here for boys and girls, respectively, are **4700** and **3433** for hand, and **7018** and **7790** for teeth.

Overview of models used in the tool

Method	Gender	Transformation age	Model type	Link function
Hand	Boys	Age^1	Proportional-odds cumulative	logit
Hand	Girls	Age^1	Proportional-odds cumulative	logit
Tooth	Boys	$Age^{0.8}$	Continuation-ratio	logit
Tooth	Girls	$Age^{0.6}$	Continuation-ratio	probit

Table 5. The table shows the model selected for each method and each gender. The selection criteria for the models are based on a model search as described in section A.5.2. "Transformation age" indicates which transformation was carried out on the variable chronological age in the regression model.

Table 5 provides an overview of the selected parametric models upon which the results in the tool are based. We found that the stage probabilities given chronological age were approximately symmetrical about chronological age for hands, while for teeth there were slightly longer tails for higher chronological ages. This is important information that will be taken into account in the final results, since it is these distributions that form the basis when we work backwards to describe the distribution of chronological age for a given stage (using Bayes' theorem).

B.2   Randomly generated variation for stage probabilities

In this section we show the effect of randomly generated variation for only the second last and last stages for all methods and both genders. This is useful for describing the variability of the fitted stage probabilities in view of the fact that we do not know the individual-based data.

B.2.1   Overview figures

The following figures (**Figures 16–31**) in sections **B.3.1.1** and **B.3.1.2** (for boys) and **B.3.1.4** and **B.3.1.5** (for girls) show the mean (solid line), 5% (lower stippled) and 95% (upper stippled) quantiles for the fitted parametric regression models of stage probabilities (black) and the non-parametric probabilities (red) across the 100 generated datasets. The fitted regression models are as listed in Table 5. The non-parametric probabilities are percentages of individuals in the data (within one year centred round a given chronological year) that fall into the various stages. The numbers at the top of the figure give the

27

total number of individuals that fall into a whole year (centred around whole given years), while the numbers at the bottom of the figures indicate the number of individuals who fall into the stage in question. These figures are given as the lowest and highest observed numbers of individuals across all the 100 generated datasets. The percentages of individuals for the combined datasets in sections **B.3.1.3** (boys) and **B.3.1.6** (girls) are not shown, since we do not have these numbers.

B.2.1.1   For boys hand
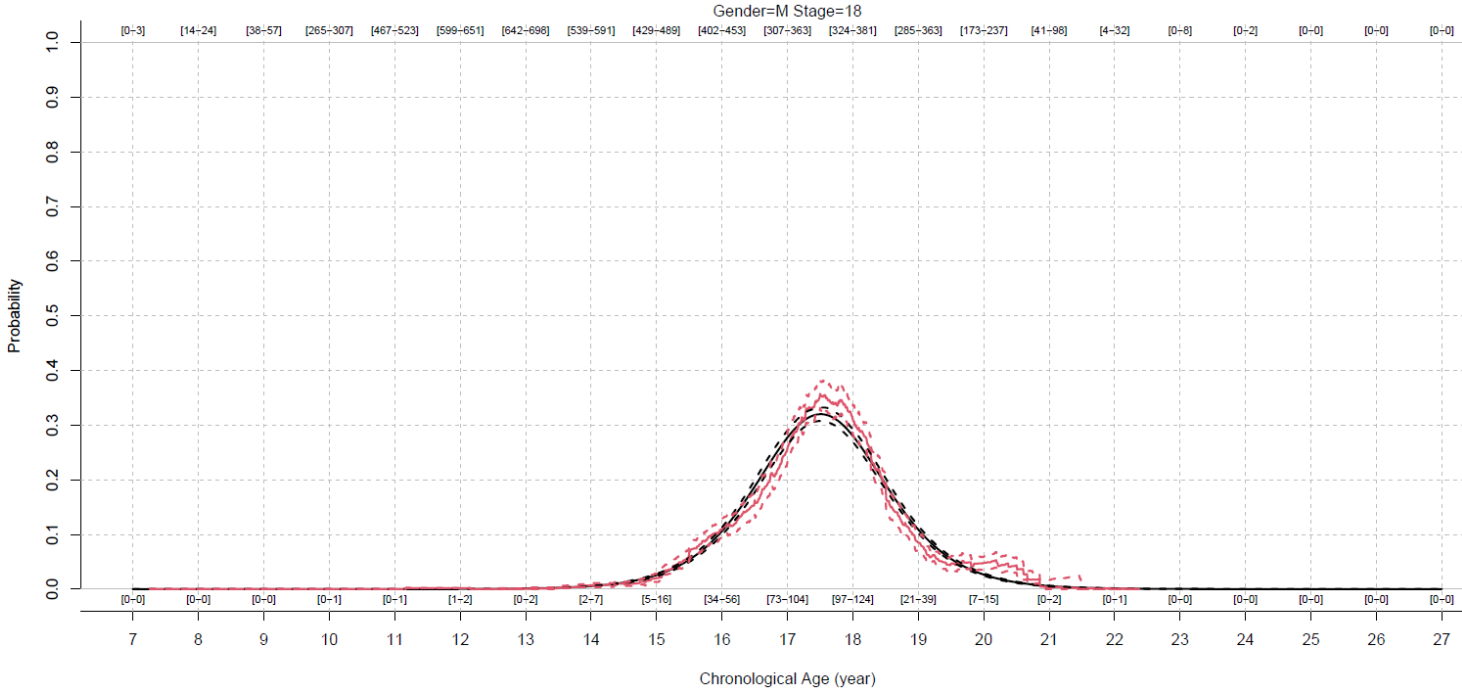
Figure 16. The figure shows the stage probability for boys with a skeletal age of 18 years for given chronological ages.
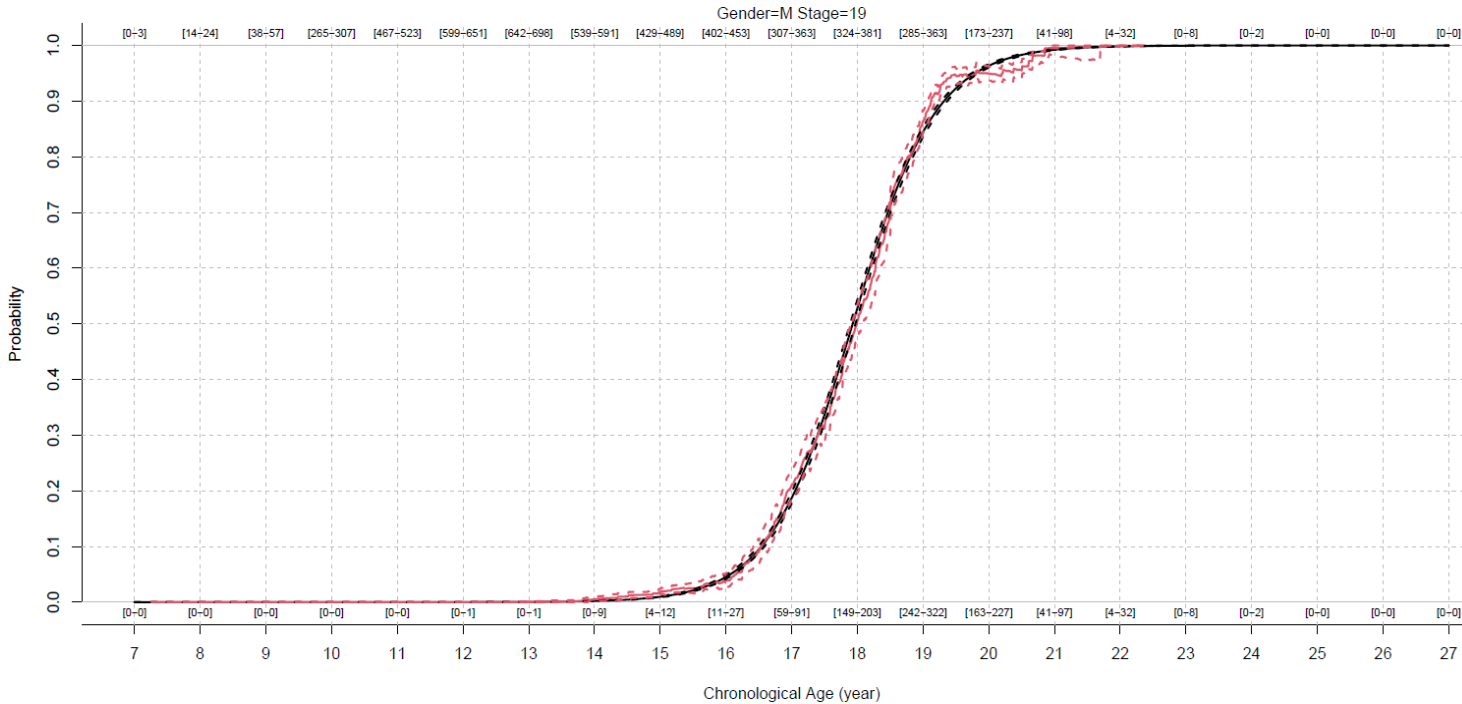
Figure 17. The figure shows the stage probability for boys with a skeletal age of 19 years for given chronological ages.

28

B.2.1.2   For boys tooth
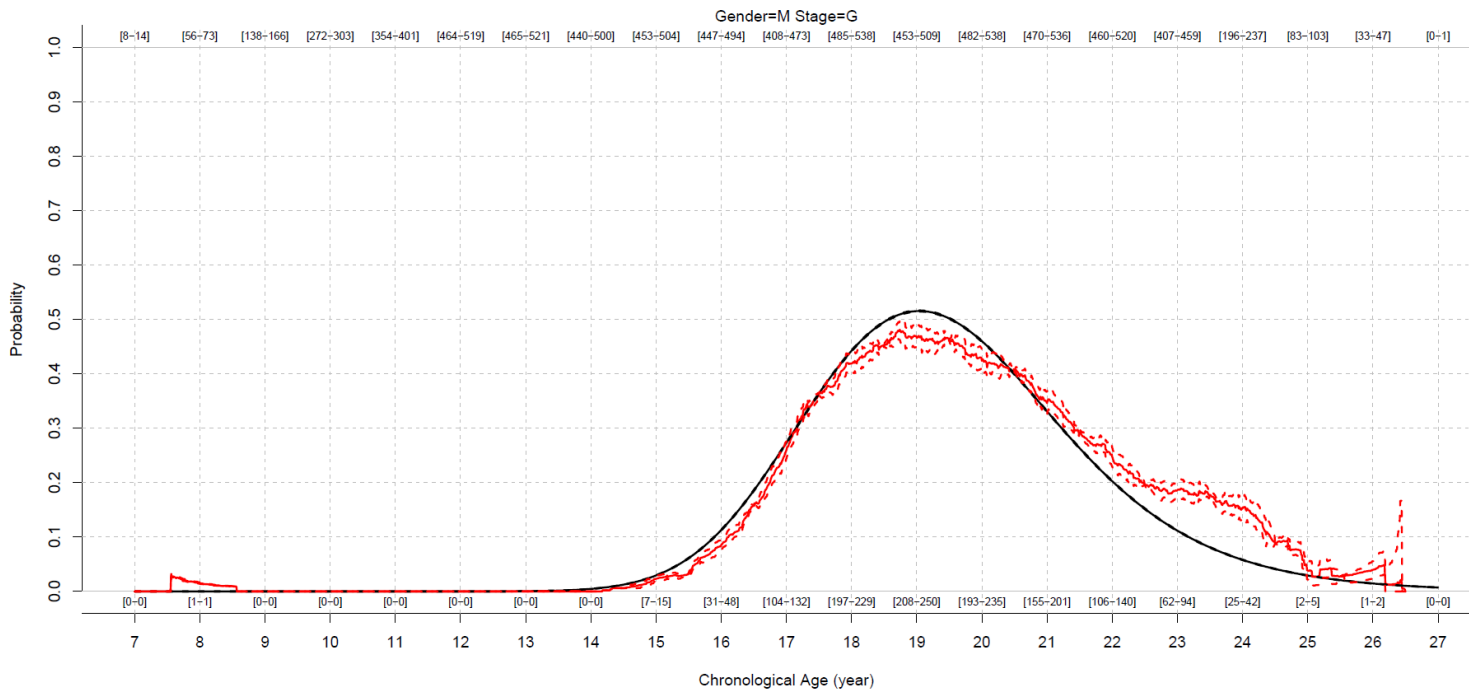
Figure 18. The figure shows the stage probability for boys with tooth stage G for given chronological ages.
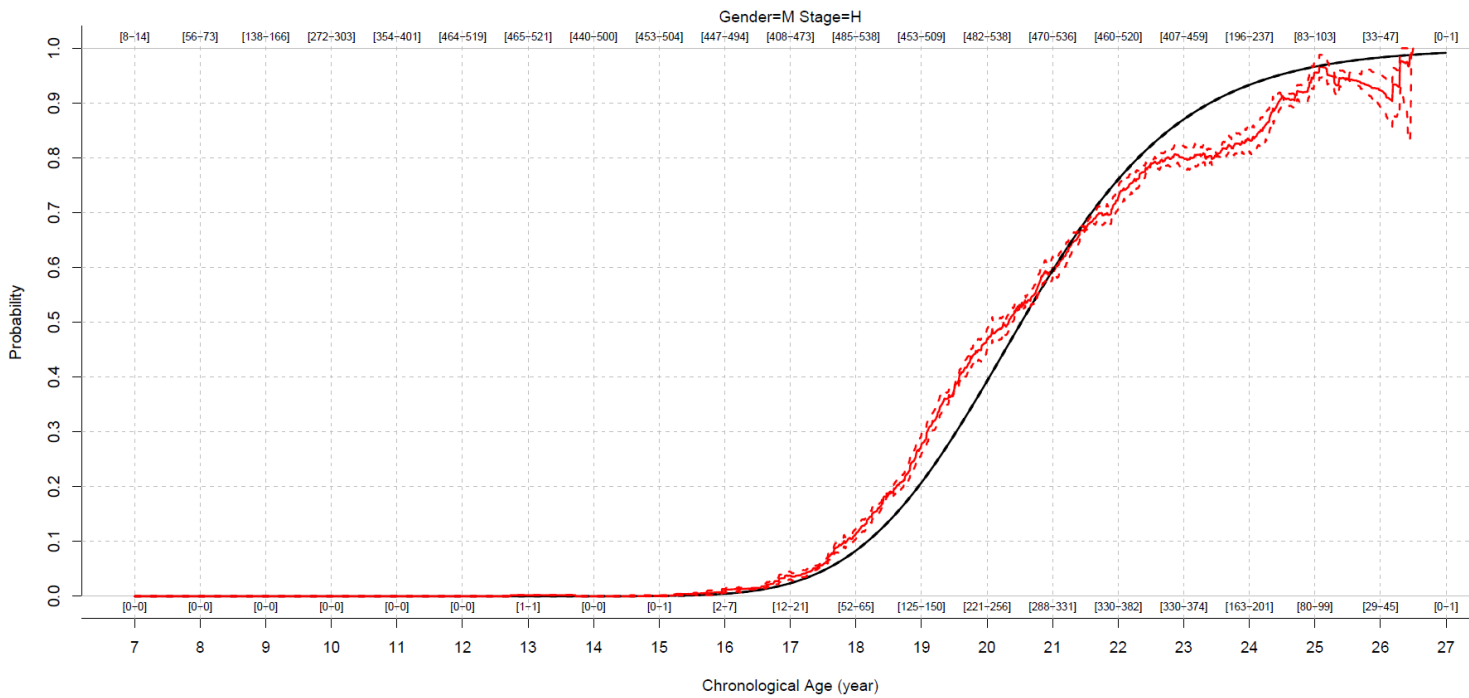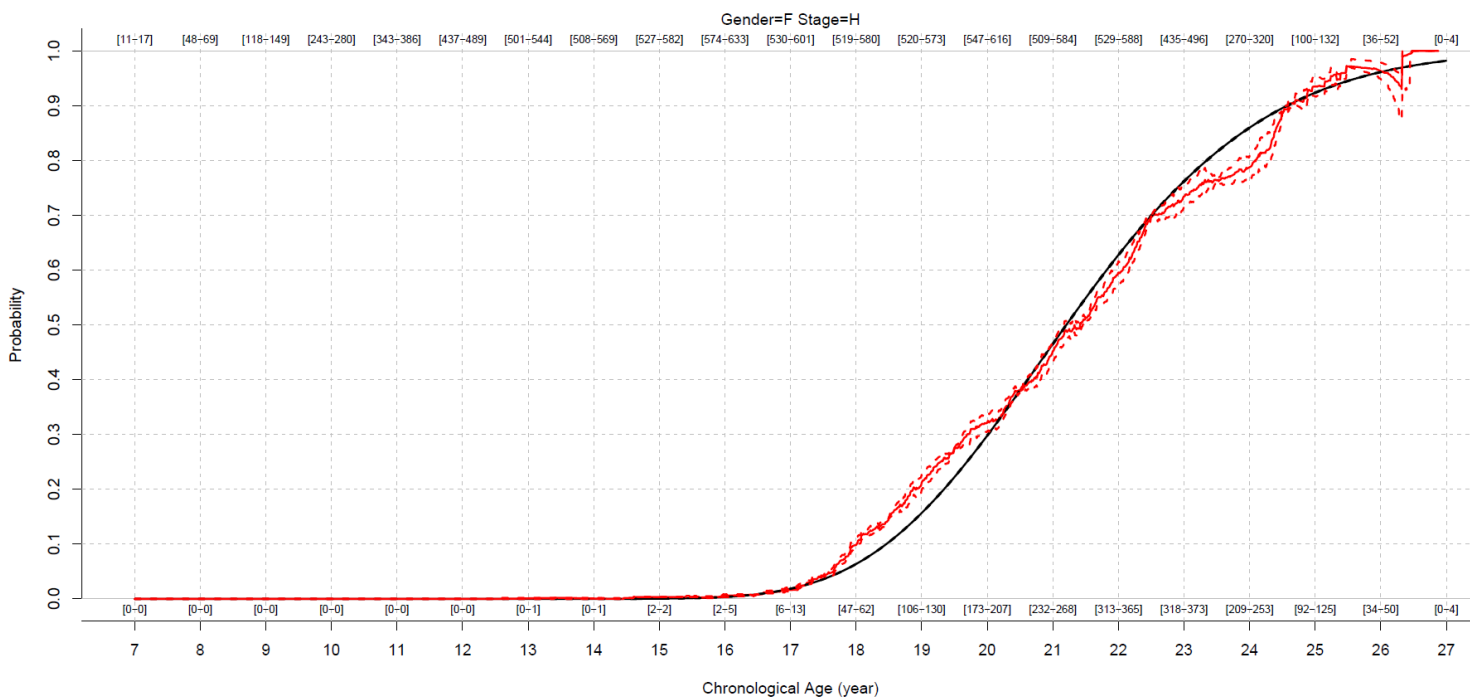

Figure 19. The figure shows the stage probability for boys with tooth stage H for given chronological ages.
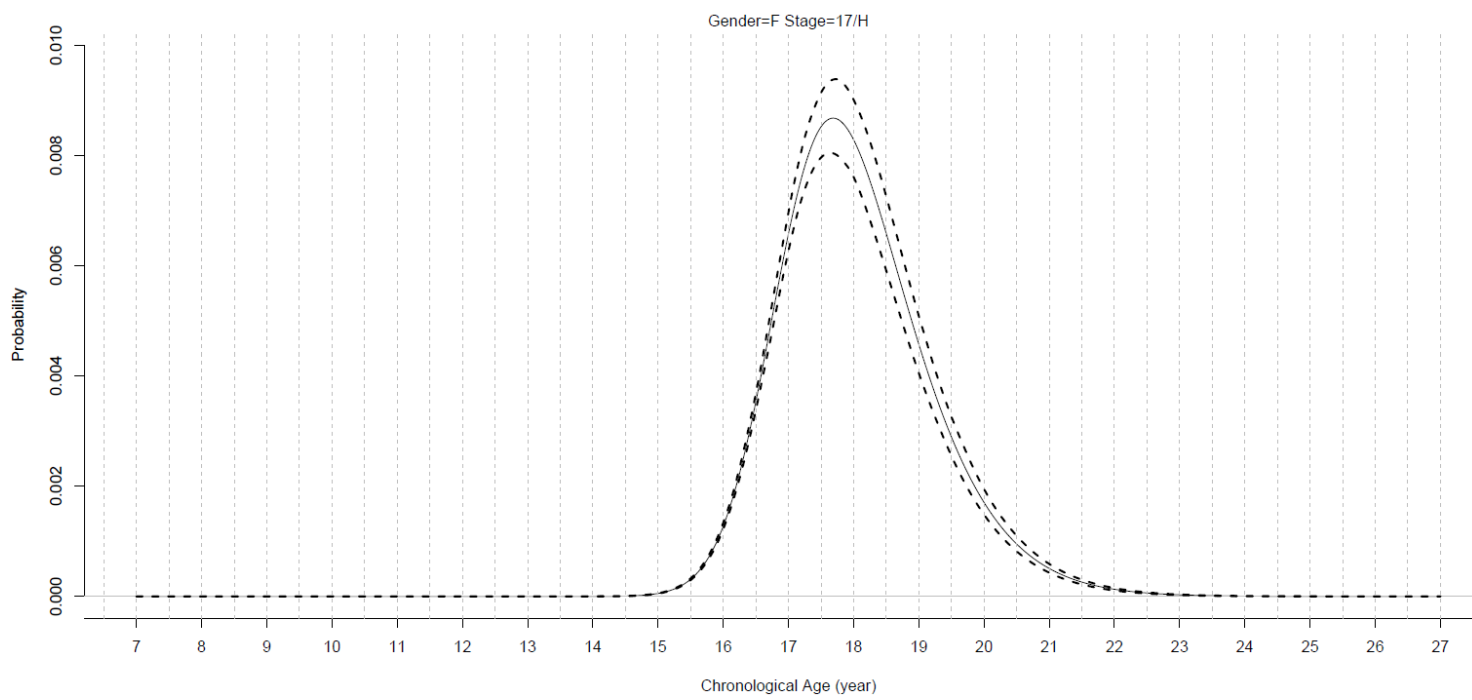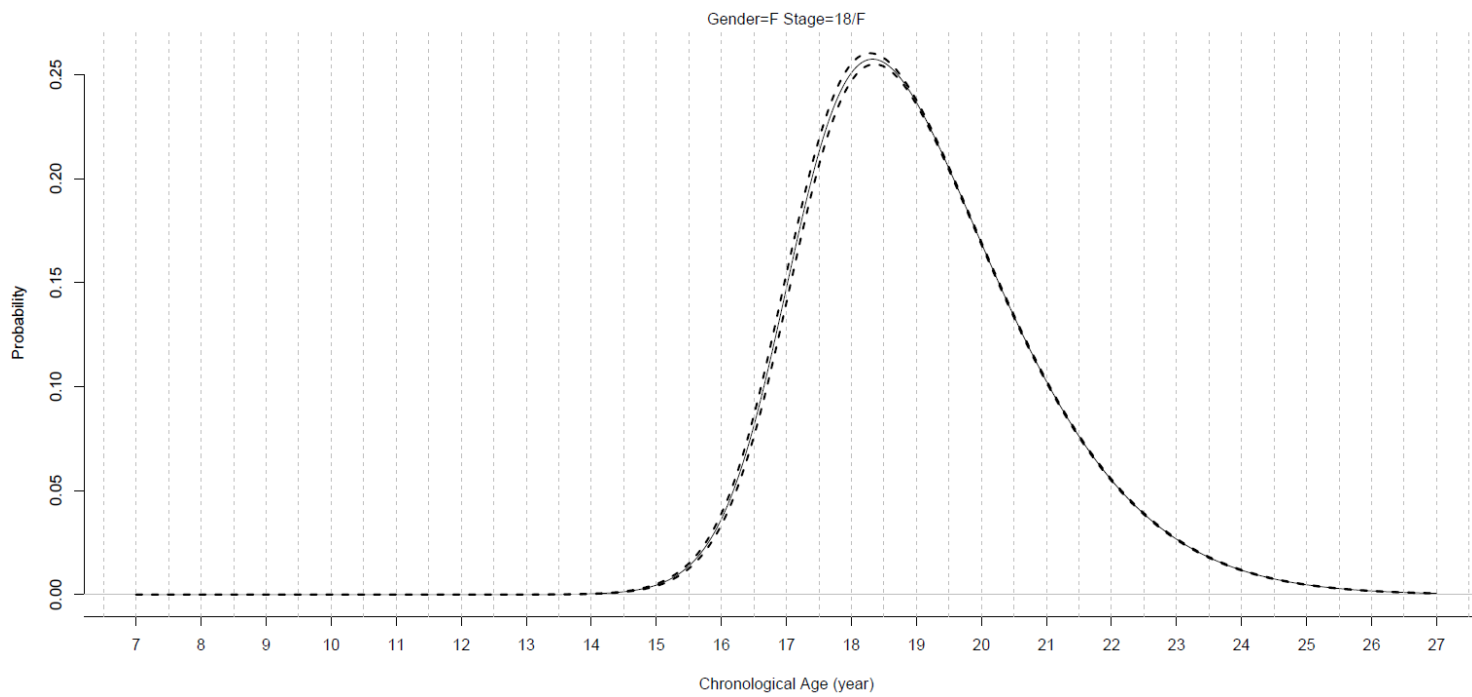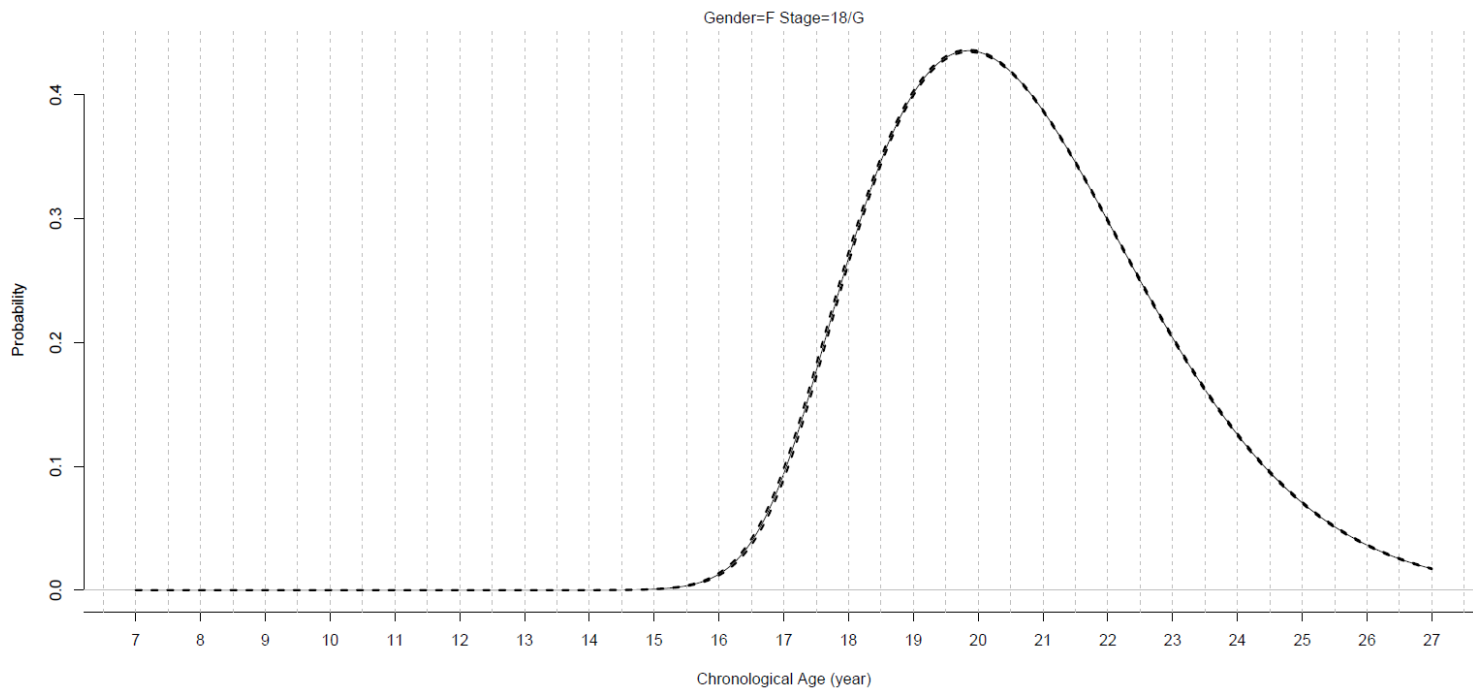
29

Gender=M Stage=18/H
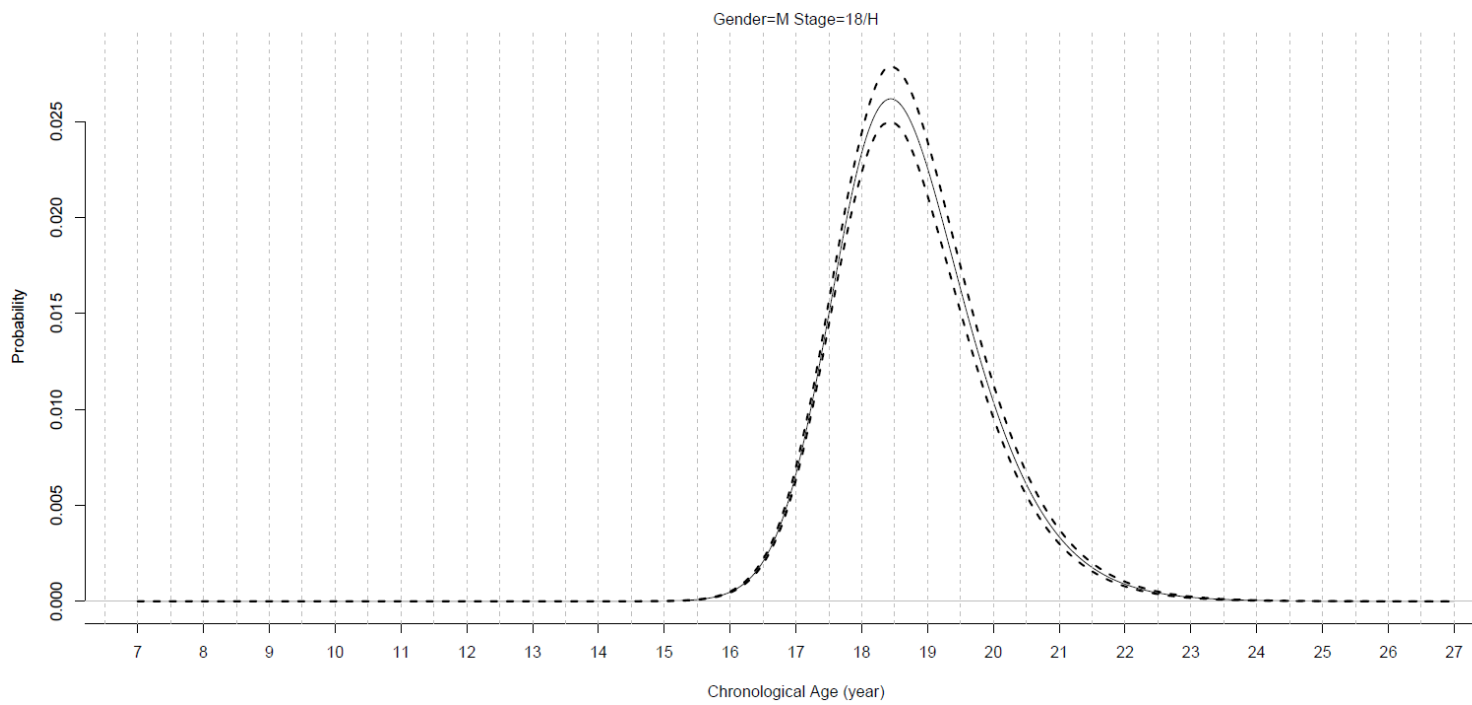
Figure 20. The figure shows the stage probability for boys with a skeletal age of 18 years and tooth stage H for given chronological ages.

Gender=M Stage=19/F
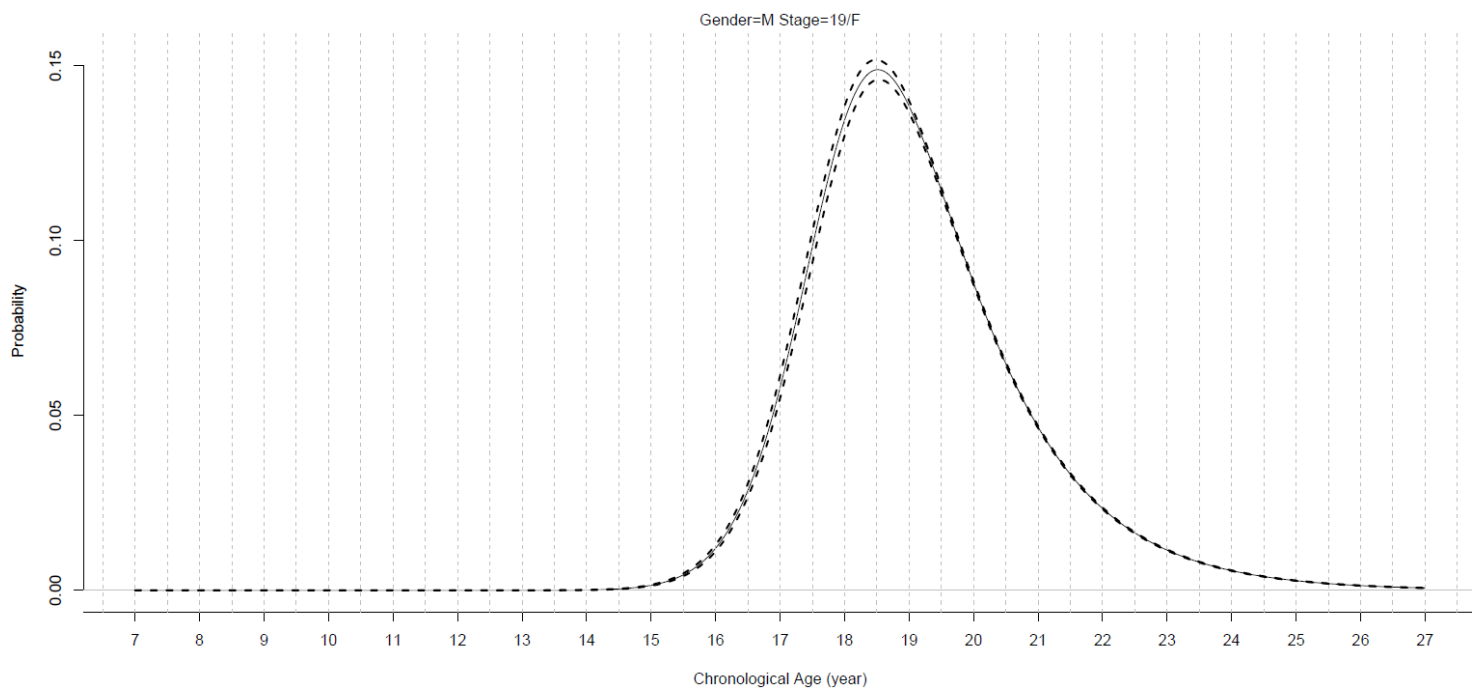
Figure 21. The figure shows the stage probability for boys with a skeletal age of 19 years and tooth stage F for given chronological ages.

30
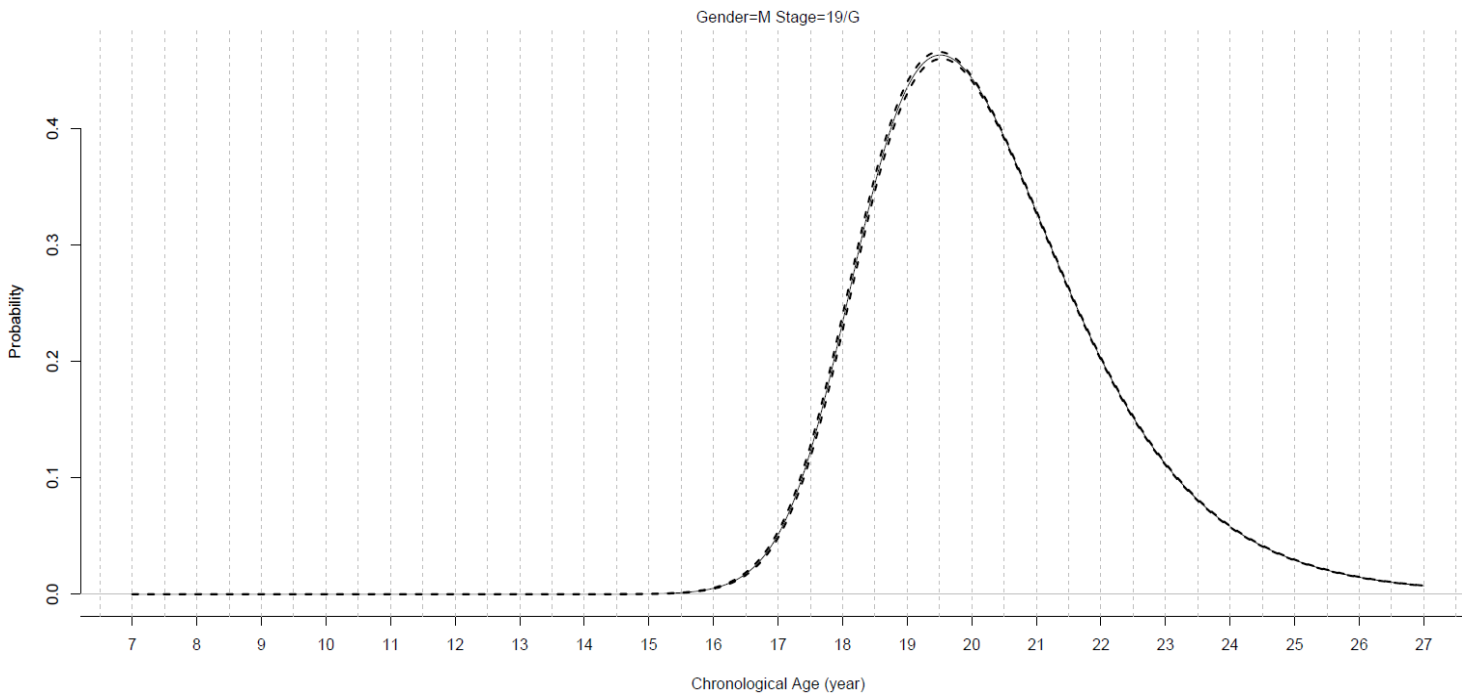
Figure 22. The figure shows the stage probability for boys with a skeletal age of 19 years and tooth stage G for given chronological ages.
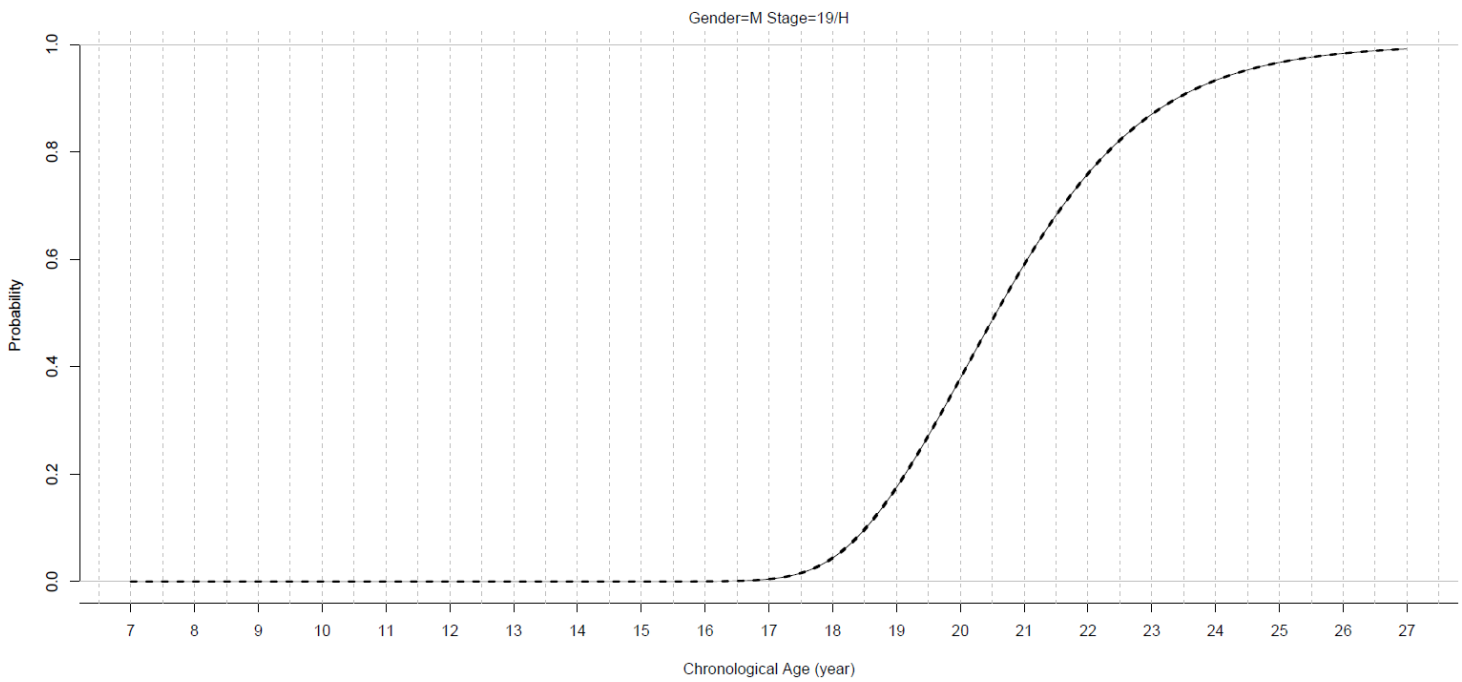
Figure 23. The figure shows the stage probability for boys with a skeletal age of 19 years and tooth stage H for given chronological ages.
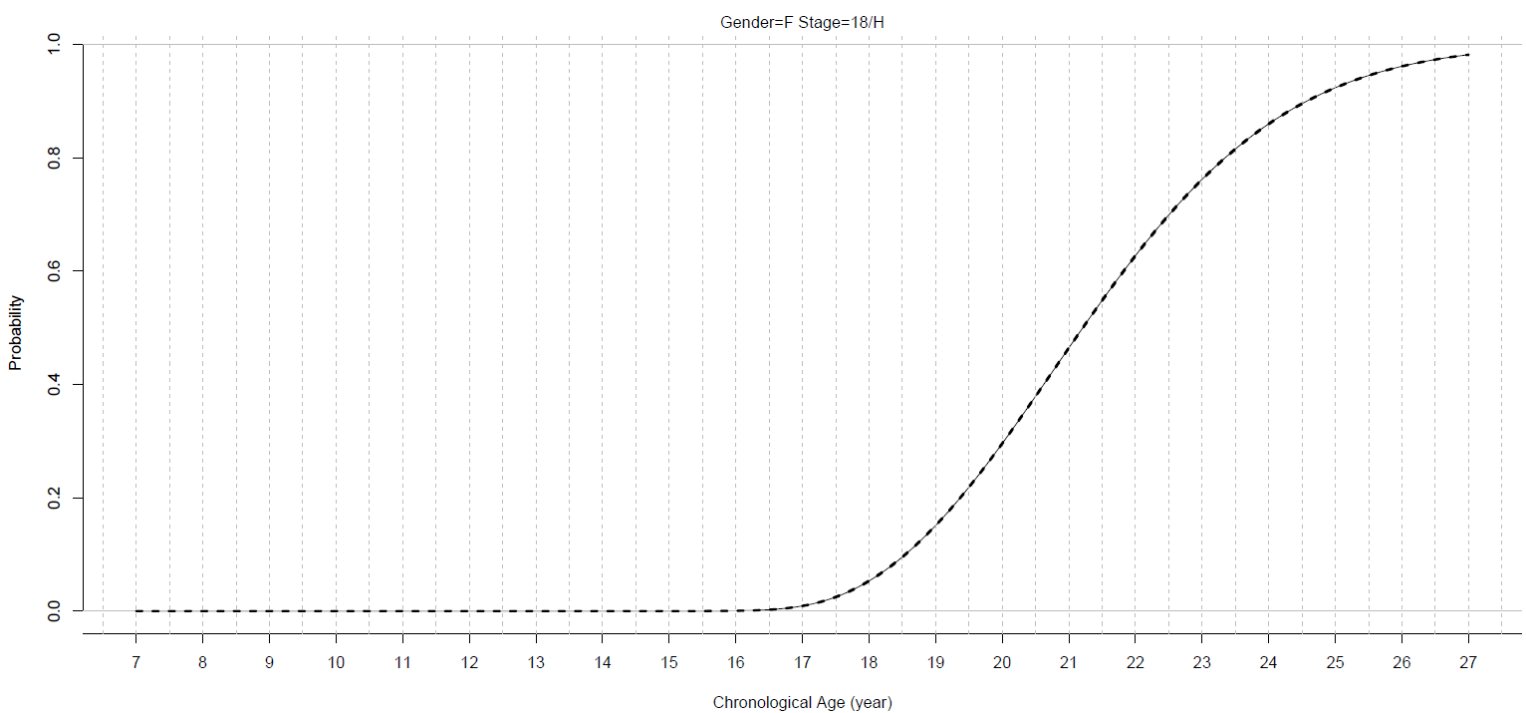
31

B.2.1.4   For girls hand
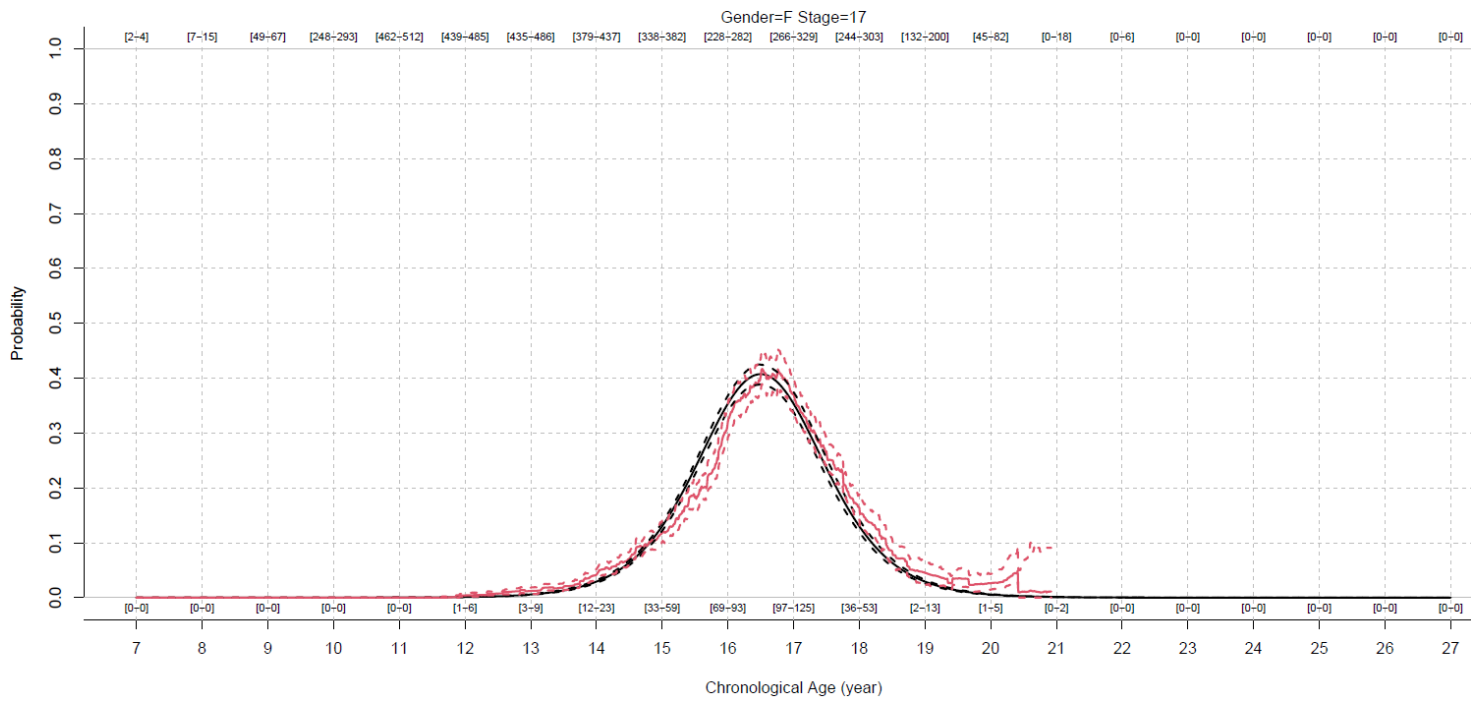
Figure 24. The figure shows the stage probability for girls with a skeletal age of 17 years for given chronological ages.
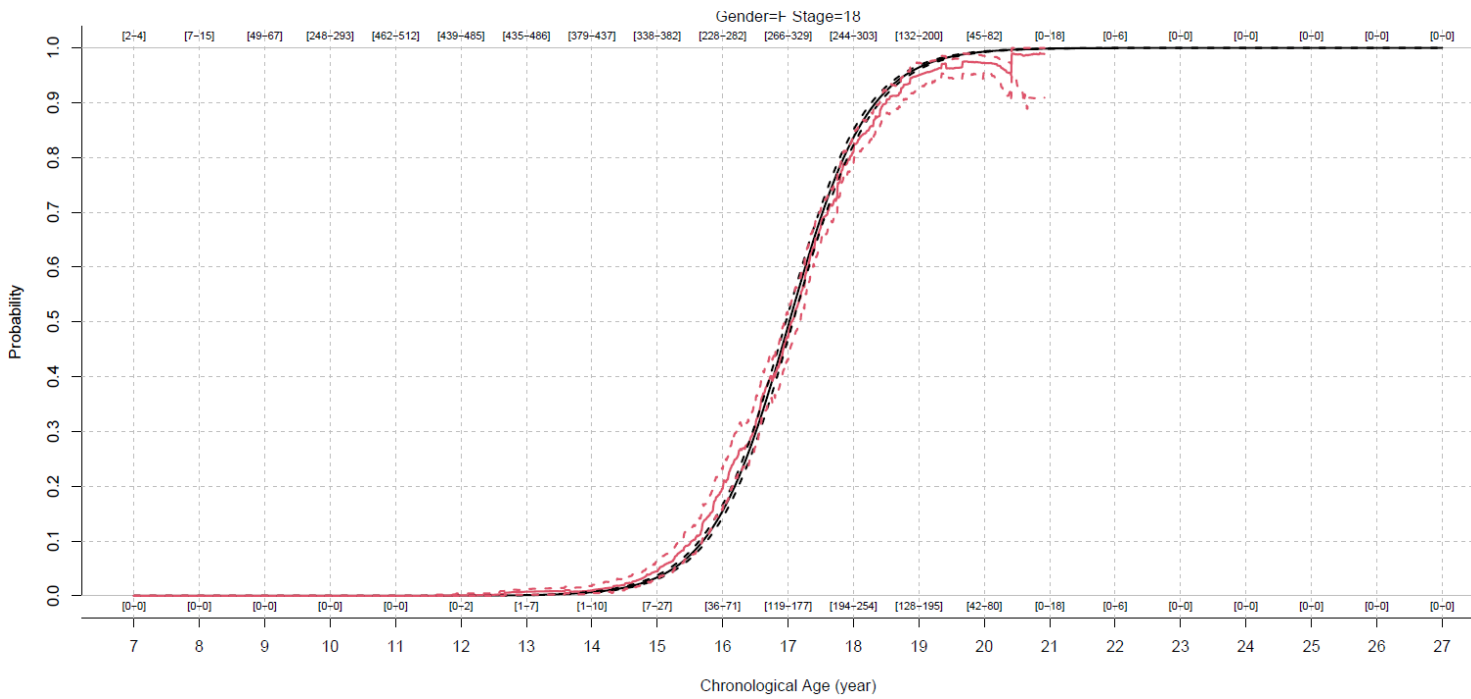
Figure 25. The figure shows the stage probability for girls with a skeletal age of 18 years for given chronological ages.

32
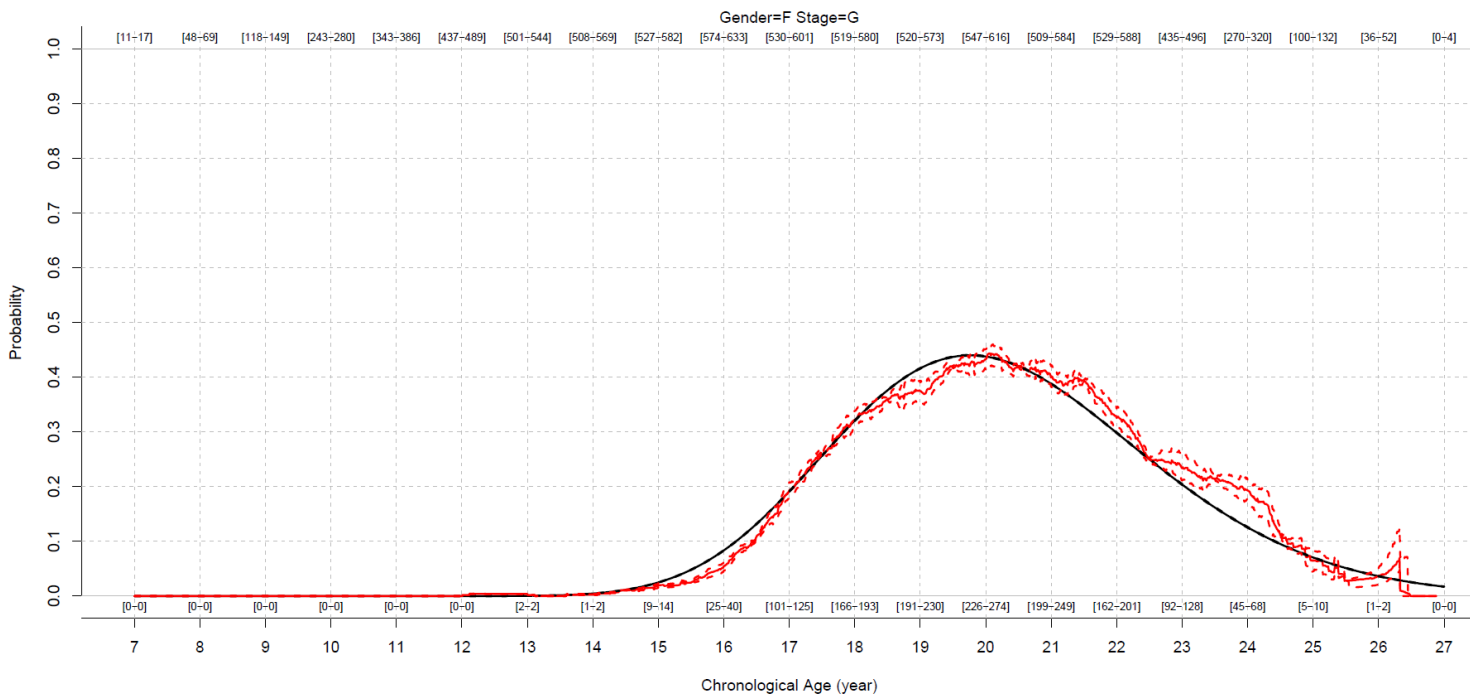
Figure 26. The figure shows the stage probability for girls with tooth stage G for given chronological ages.

Figure 27. The figure shows the stage probability for girls with tooth stage H for given chronological ages**.**

33

B.2.1.6   For girls combined

Figure 28. The figure shows the stage probability for girls with a skeletal age of 17 years and tooth stage H for given chronological ages.

Figure 29. The figure shows the stage probability for girls with a skeletal age of 18 years and tooth stage F for given chronological ages.

34

Figure 30. The figure shows the stage probability for girls with a skeletal age of 18 years and tooth stage G for given chronological ages.

Figure 31. The figure shows the stage probability for girls with a skeletal age of 18 years and tooth stage H for given chronological ages.

B.2.2   Model validation

The figures in section **B.3.1** for hand and tooth are useful for model validation since they show the fitted model compared with the actual data. We see here that the stage probabilities G and H for boys'

teeth, and the stage probabilities for skeletal age 17 years for girls, deviate somewhat from the observed data. Since the estimation of stage probabilities is based on a limited number of individuals, we ought also to take this into account by estimating confidence intervals for the probabilities. This is done in section **B.4.**

B.3   Credibility interval for stage probabilities

In this section we consider 95% credibility intervals (Bayesian analogue of confidence intervals) for stage probabilities (for each given age) to allow for the fact that the estimation of these probabilities is based on a limited number of individuals.

B.3.1   Credibility interval for non-parametric model

For the non-parametric model, we use a "Jeffreys Interval", where we assume a beta prior with shape parameters equal to a half. The 95% credibility interval $[L, U]$ for probability given age a will then be equal to

$$[Beta_{2.5\%}\left(n_j(a) + \frac{1}{2}, n(a) - n_j(a) + \frac{1}{2}\right), Beta_{97.5\%}\left(n_j(a) + \frac{1}{2}, n(a) - n_j(a) + \frac{1}{2}\right)]$$

where $n(a)$ and $n_j(a)$ are the total numbers of individuals across all stages and within stage j, respectively, within the age segment $[a - 0.5, a + 0.5)$. Special cases: For $n_j(a) = 0, L = 0$. For $n_j(a) = n(a), U = 1$. For $n(a) = 0, L = 0, U = 1$.

B.3.2   Credibility interval for parametric model
For the appurtenant selected parametric models (see section **B.2**), we calculate the 95% credibility intervals for probability given age a as

$$[P(Stage = j|a, \theta^*)_{2.5\%}, P(Stage = j|a, \theta^*)_{97.5\%}]$$

$$\theta^* \sim MVN(\theta^{hat}, -Hessian^{-1}(\theta^{hat}))$$

where MVN is multivariate normal distribution (with expectation and covariance matrix as arguments) and $Hessian$ is the second order derivation matrix of the likelihood function (on a log scale). We generate 1000 random samples from the multivariate normal distribution in order to calculate the credibility interval.

B.3.3   Overview figures

We will now show 95% credibility intervals based on non-parametric and parametric models for the cases where we were in doubt as to whether the model assumption for the parametric models is adequate. Since we get one credibility interval $[L, U]$ for each generated dataset, we choose the 5%

36

quantile of the lower thresholds L, and the 95% quantile of the upper thresholds U, as the final values of the credibility intervals.

The figures below (**Figures 32-34)** show the mean of the fitted parametric regression models for the stage probabilities (black), and 95% credibility intervals for the stage probabilities for both the parametric transition model (black stippled) and the non-parametric probabilities (red stippled) across the 100 generated datasets.

Note 1: The variability is less for the probabilities based on the parametric transition model than the non-parametric model.

Note 2: The average parametric model (almost) always lies within the 95% credibility intervals for the probabilities based on the non-parametric model.
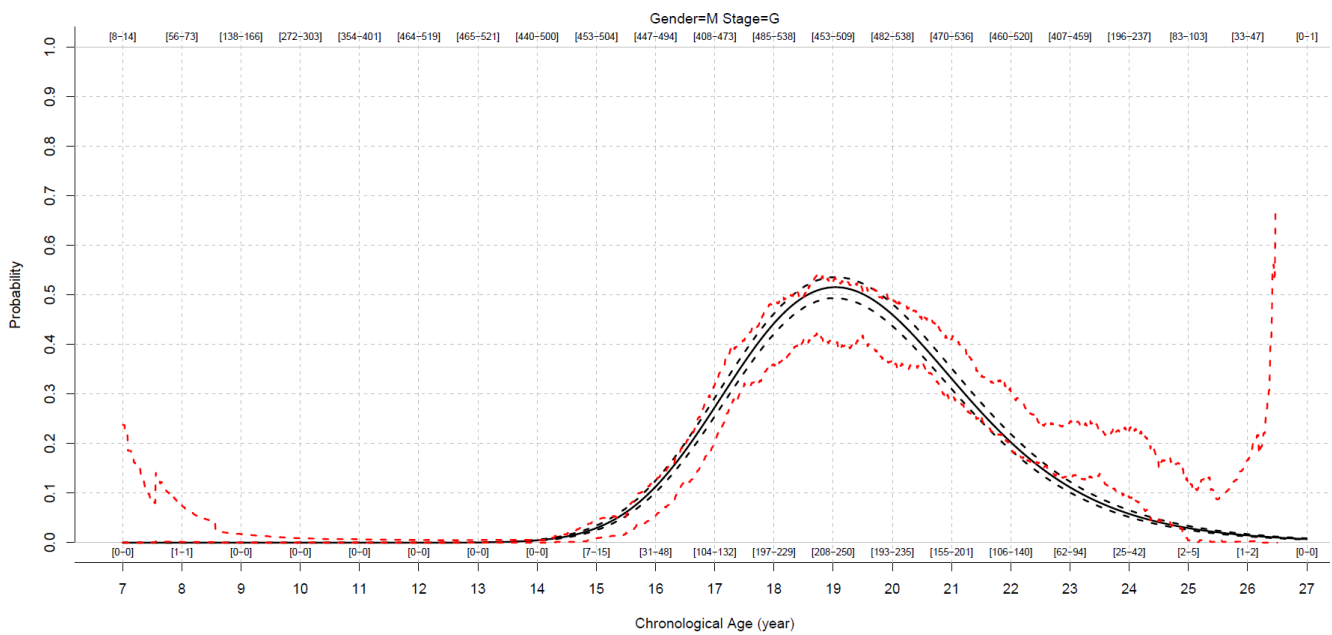
Figure 32. The figure shows the stage probability for boys with tooth stage G for given chronological ages.
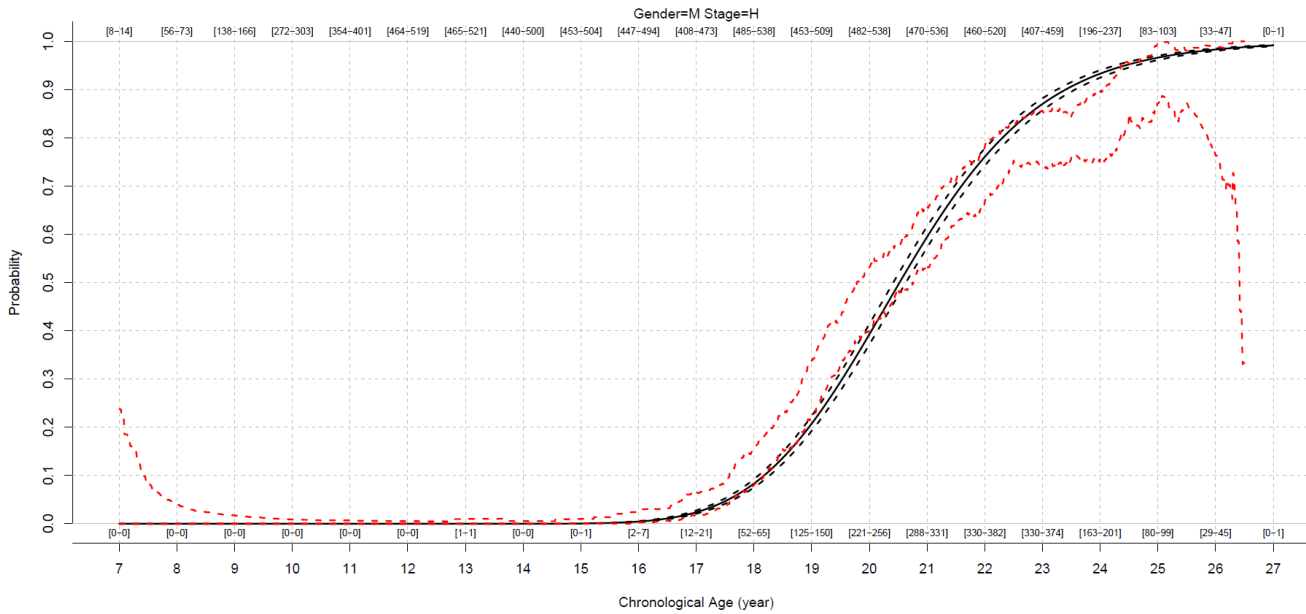
Figure 33. The figure shows the stage probability for boys with tooth stage H for given chronological ages.
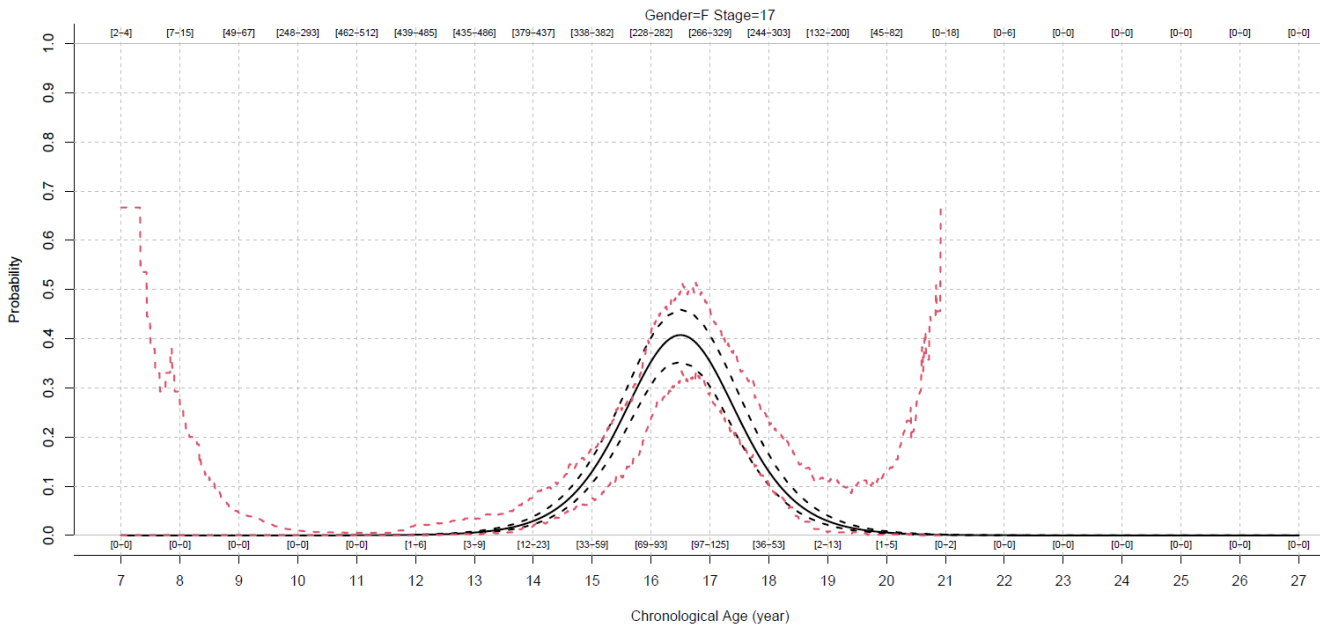
Figure 34. The figure shows the stage probability for girls with a skeletal age of 17 years for given chronological ages

B.4   Choice of upper age limits in the tool (defining the prior age distribution)

B.4.1   The effect of different upper age limits

Both teeth and hand skeleton have end-stages that last for the rest of the individual's life. This means, for example, that there will be the same probability of a 50 year-old having the end stage as a 25 year-old, assuming that no 25 year-olds can have the second-last stage. This makes it very challenging to describe the distribution of chronological ages for those stages that do not "end", since this description will depend on which ages are included in the study (see section **A.7.3**). It is essential to know how strong the effect of the defined age is on the results. In this section we therefore investigate how the choice of various defined age ranges affects the results. Table 6 presents an overview of the results

that are most strongly affected by a change in the upper defined age. See section **B.5.3** for extensive illustrations of how defined upper age limits affect the results.

Gender	Method	Stage	Threshold	19y	20y	20.5y	21y	23y	25y
Boys	Hand	19	18	0.41	0.23	0.19	0.16	0.10	0.07
Boys	Tooth	G	18	0.56	0.39	0.35	0.32	0.26	0.25
Boys	Tooth	H	18	0.31	0.13	0.09	0.06	0.03	0.01
Boys	Comb.	19/E	18	0.59	0.46	0.43	0.41	0.38	0.38
Boys	Comb.	19/F	18	0.49	0.35	0.32	0.30	0.27	0.27
Boys	Comb.	19/G	18	0.31	0.16	0.14	0.12	0.09	0.08
Boys	Comb.	18/H	18	0.43	0.31	0.30	0.29	0.28	0.28
Boys	Comb.	19/H	18	0.17	0.05	0.03	0.02	0.01	<0.01
Girls	Hand	18	18	0.55	0.37	0.32	0.28	0.19	0.14
Girls	Tooth	G	18	0.55	0.36	0.31	0.27	0.2	0.18
Girls	Tooth	H	18	0.32	0.13	0.09	0.07	0.02	0.01
Girls	Comb.	18/F	18	0.56	0.41	0.38	0.35	0.31	0.31
Girls	Comb.	18/G	18	0.41	0.23	0.19	0.17	0.12	0.10
Girls	Comb.	18/H	18	0.24	0.09	0.06	0.04	0.02	0.01
Boys	Hand	19	17	0.12	0.07	0.05	0.05	0.03	0.02
Boys	Tooth	G	17	0.24	0.17	0.15	0.13	0.11	0.11
Boys	Tooth	H	17	0.07	0.03	0.02	0.01	0.01	<0.01
Boys	Comb.	19/E	17	0.23	0.18	0.17	0.16	0.15	0.15
Boys	Comb.	19/F	17	0.14	0.1	0.09	0.08	0.08	0.07
Boys	Comb.	18/H	17	0.07	0.05	0.05	0.04	0.04	0.04
Girls	Hand	18	17	0.21	0.14	0.12	0.11	0.07	0.05
Girls	Tooth	G	17	0.24	0.16	0.14	0.12	0.09	0.08
Girls	Tooth	H	17	0.07	0.05	0.03	0.02	0.01	<0.01
Girls	Comb.	18/F	17	0.19	0.14	0.13	0.12	0.11	0.10
Girls	Comb.	18/G	17	0.09	0.05	0.04	0.04	0.03	0.02
Boys	Tooth	G	16	0.07	0.05	0.04	0.04	0.03	0.03
Boys	Comb.	19/E	16	0.06	0.05	0.04	0.04	0.04	0.04
Girls	Hand	18	16	0.06	0.04	0.03	0.03	0.02	0.01
Girls	Tooth	G	16	0.08	0.05	0.04	0.04	0.03	0.02

Table 6. The table shows an overview of the cases where the probability of being below a "threshold" (16, 17 or 18 years) is at least 1.5 times as high if the upper age limit is set at 19 years as opposed to an upper age limit of 25 years (at least one of the probabilities must also be at least 5%). The figures with dark grey background indicate the values that are used as final results in the tool.

B.4.2   Final choices for upper age limits in the tool

In the previous section, we saw that the assumed upper age in the age distribution affects the results, in particular the last stage (see also section **B.5.3** for a broader overview). This is a major challenge, since the value of the upper age should not be too low, as this would mean excluding information about the chronological ages a stage can have, but not too high either, as this could reduce the probability of an individual being, for example, under 18 years (which could increase the possibility of

children erroneously being classified as adults). Section 8.2 describes the motivation of our final choices. Table 7 provides an overview of the defined upper ages for the different methods and genders.

	Hand	Tooth	Combined
Boys	20 years	20.5 years	20.5 years
Girls	19 years	21 years	21 years

Table 7. The table provides an overview of defined upper age of the assumed age distribution for the different methods and genders used in the tool.

B.4.3   Overview figures of the effect of assumed upper age

The following figures (**Figures 35–50**) in sections **B.5.3.1–B.5.3.6** show the effect of assuming different upper ages in order to work backwards (i.e. applying Bayes theorem) to describe the distribution of chronological age for given stages.

The figures on the left show the 75% (red lines) and 95% (black lines) prediction intervals for chronological age for given observed stages, while the figures on the right show the probabilities of being under 16 (red lines) and under 18 years (black lines) in chronological age for given observed stages.

B.4.3.1   Effect for boys: hand skeletal ages 18 and 19 years
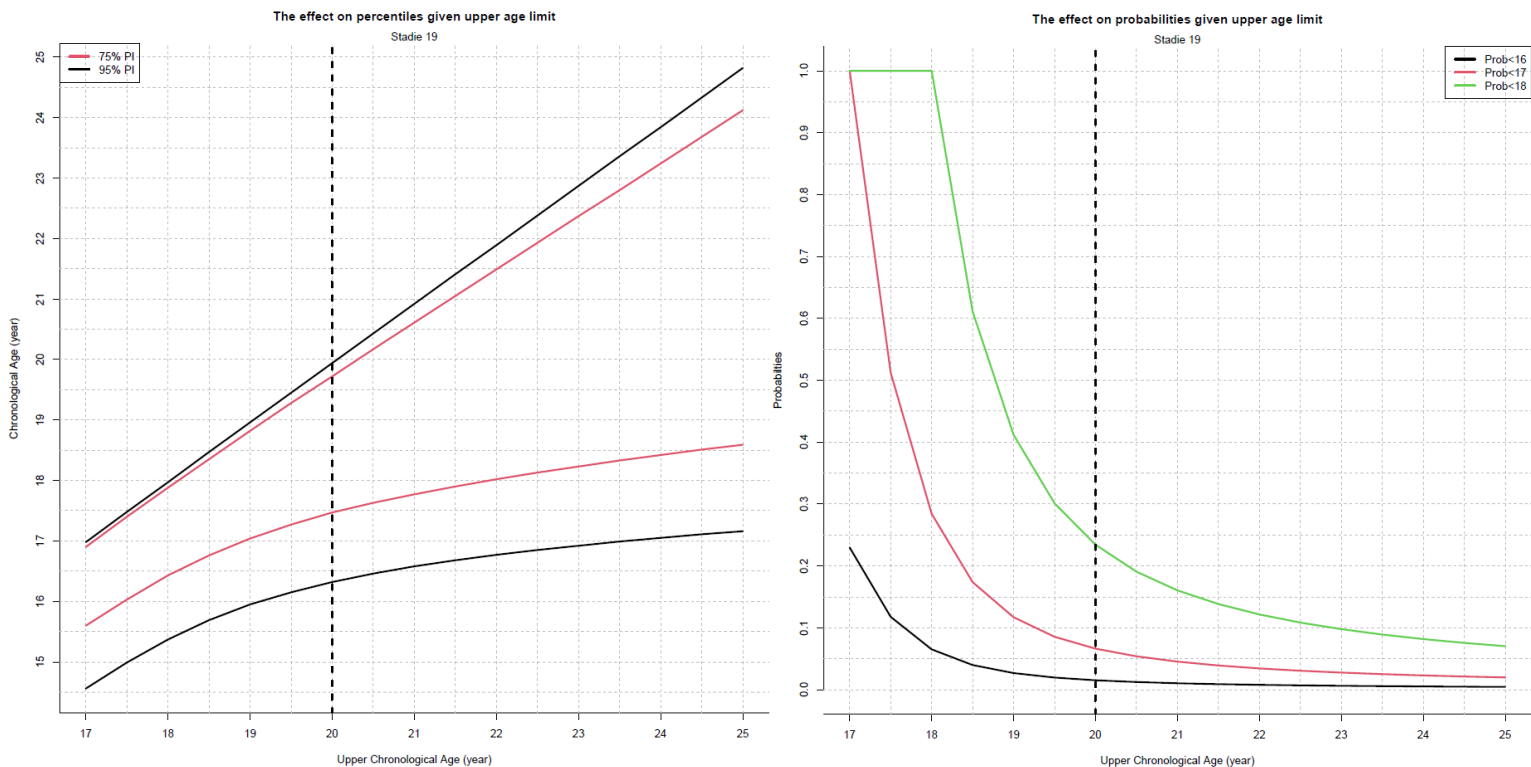
Figure 35: The figure shows prediction intervals and probabilities for boys with a skeletal age of 19 years.

41
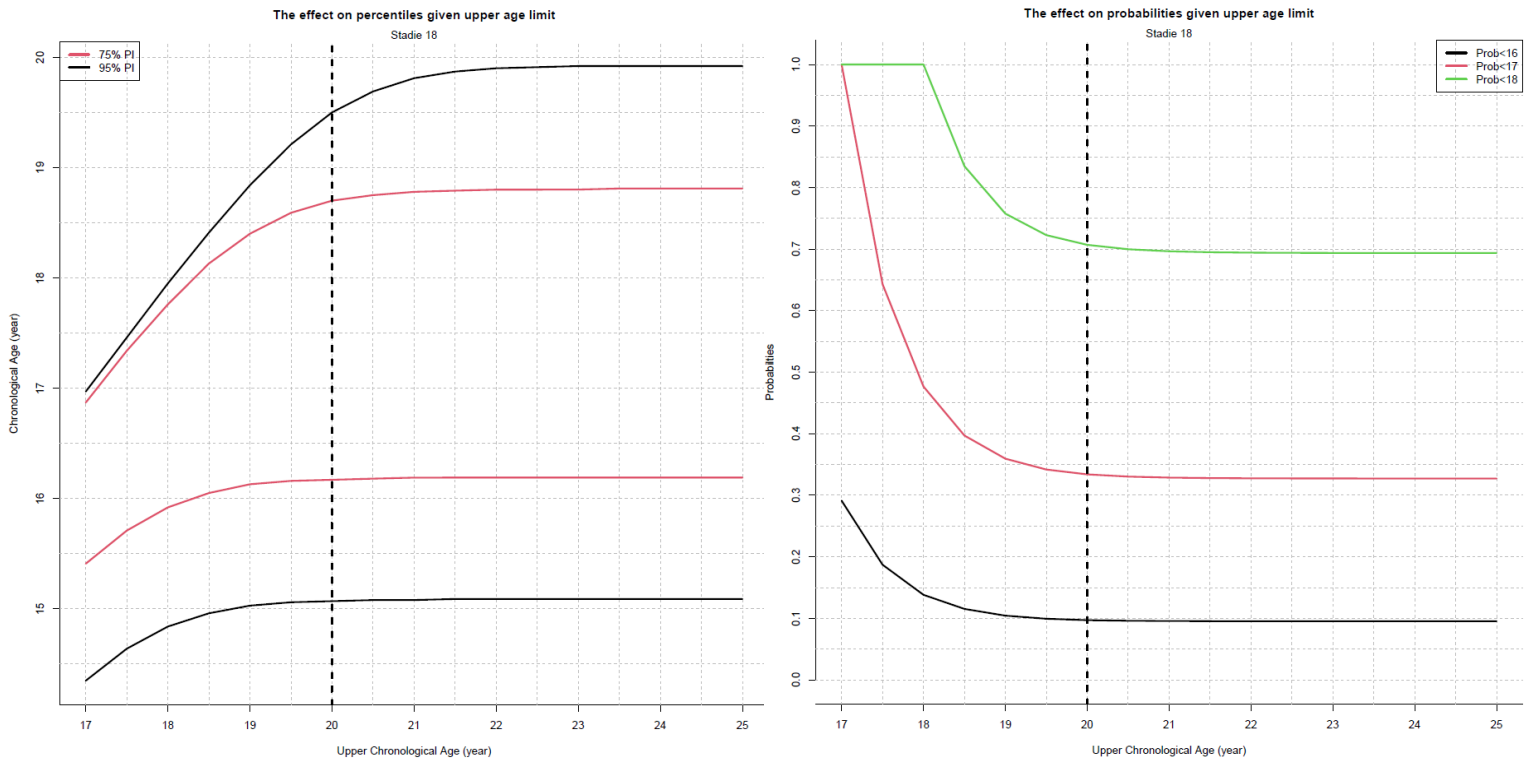
Figure 36: The figure shows prediction intervals and probabilities for boys with a skeletal age of 18 years.

B.4.3.2   Effect for boys: tooth stages G and H
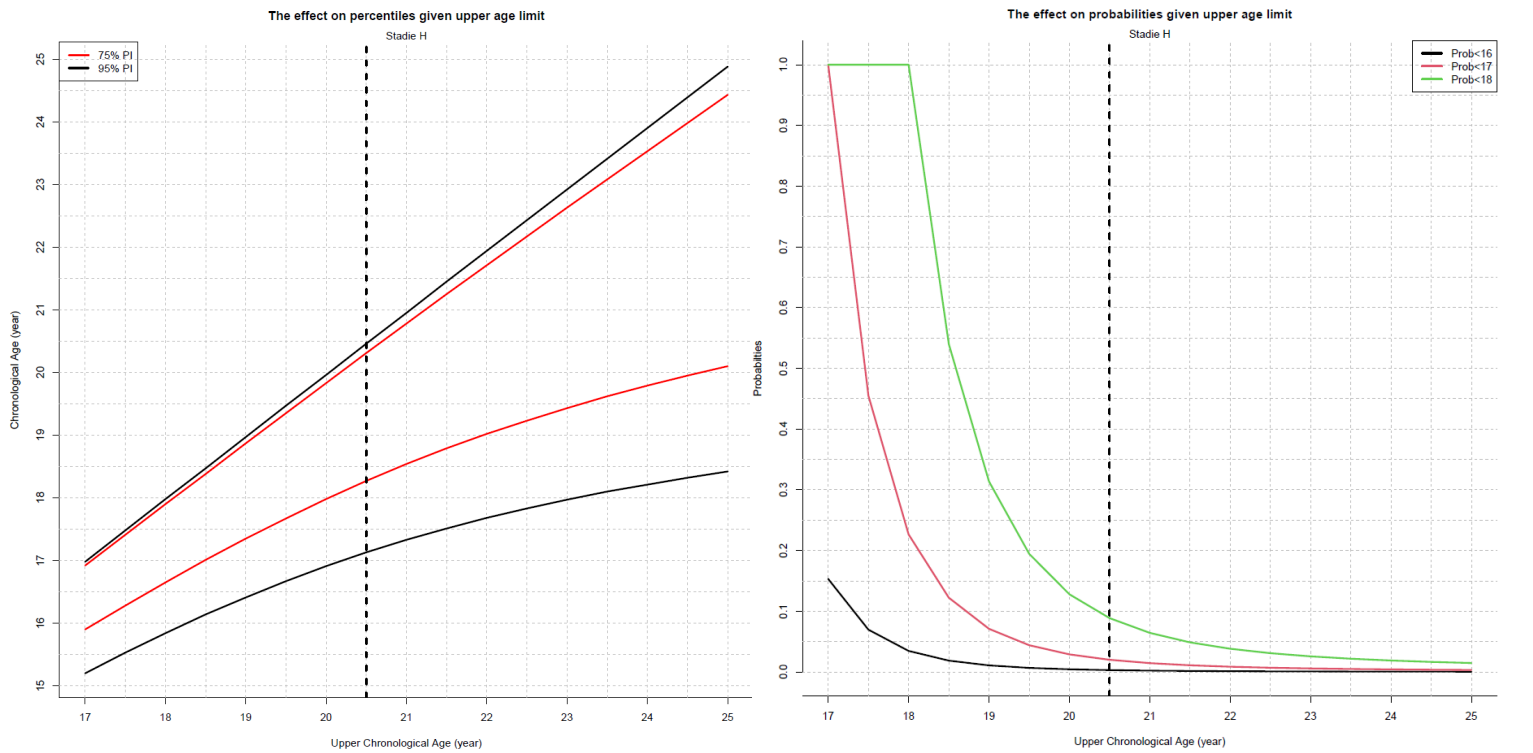
Figure 37: The figure shows prediction intervals and probabilities for boys with tooth stage H.
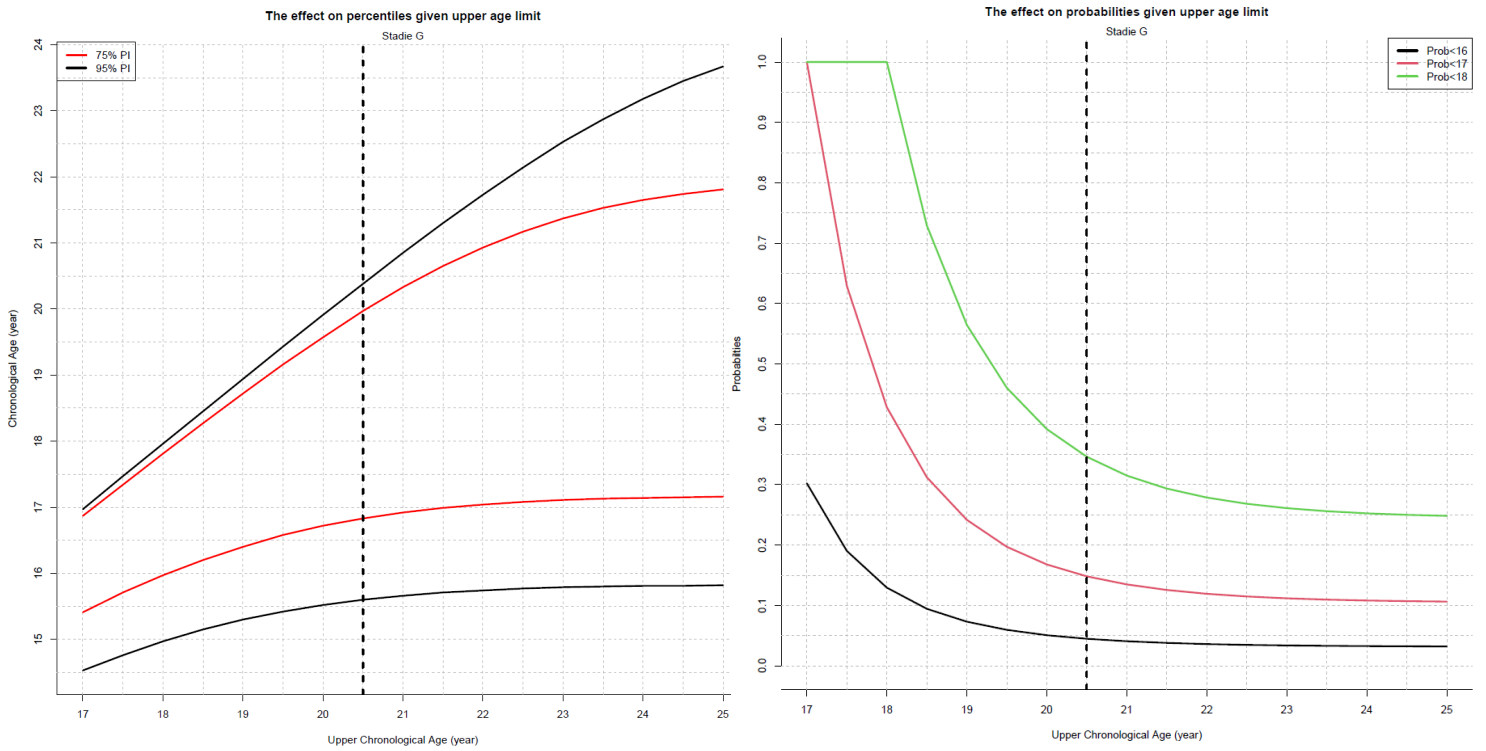
Figure 38: The figure shows prediction intervals and probabilities for boys with tooth stage G.

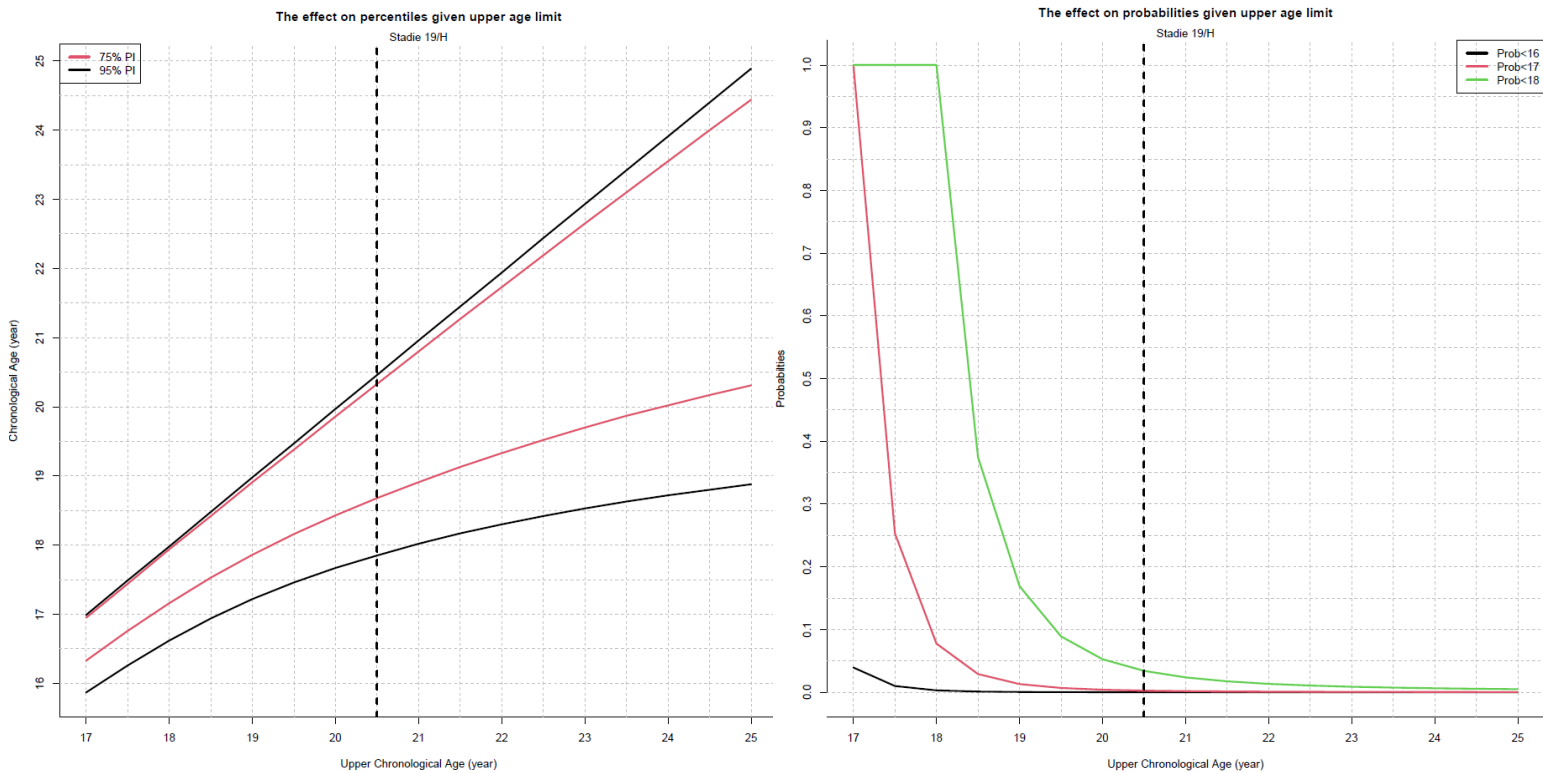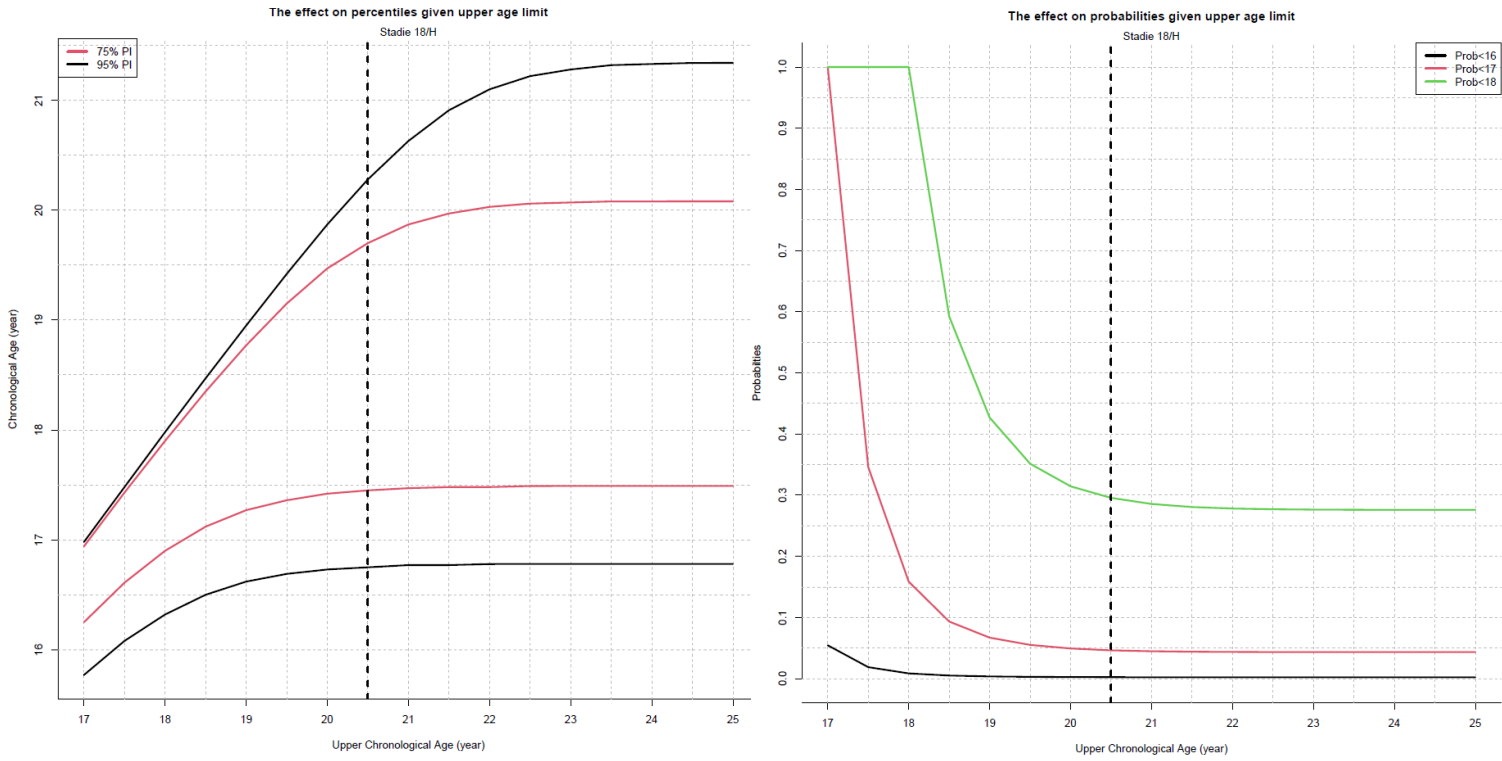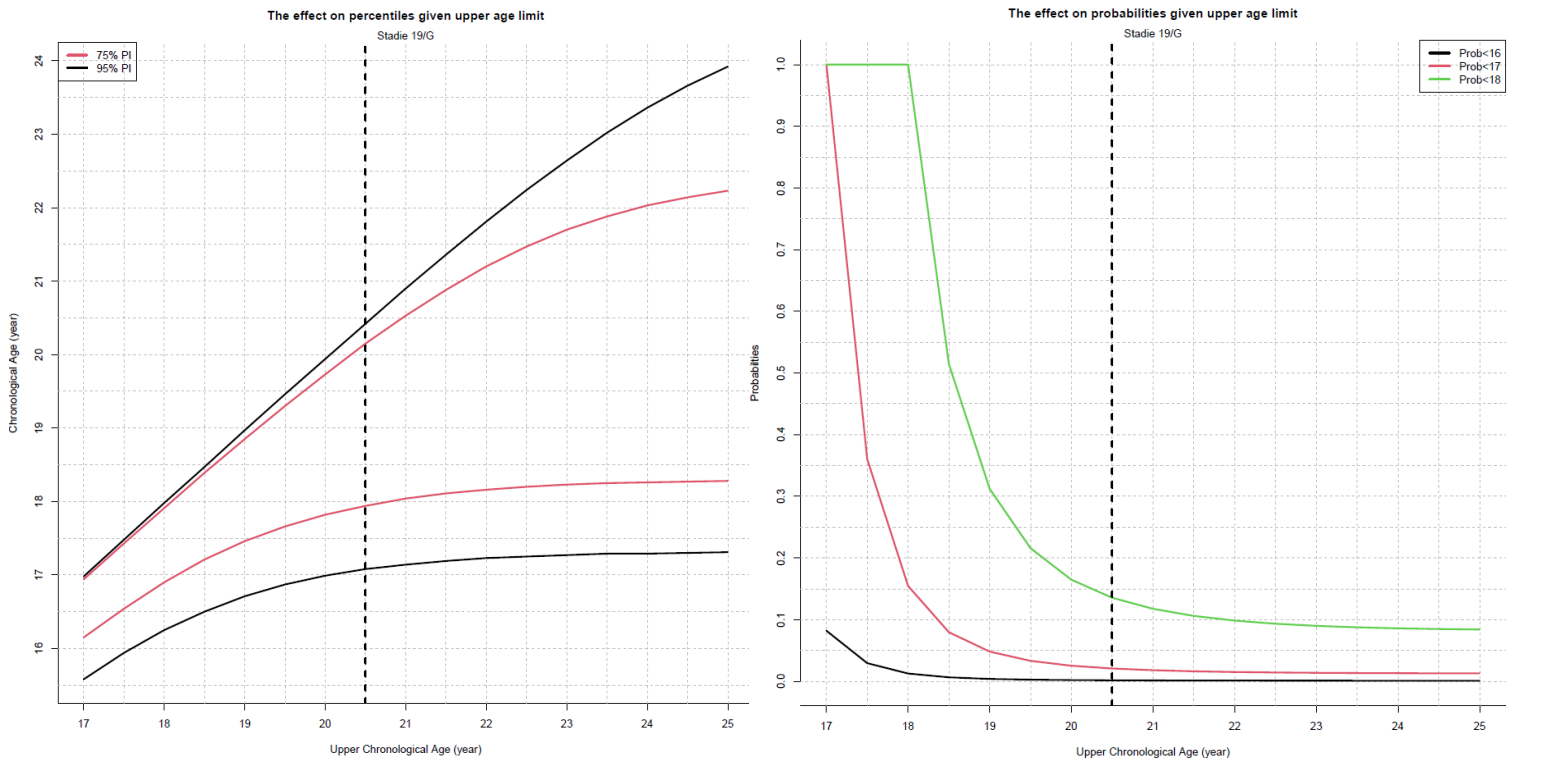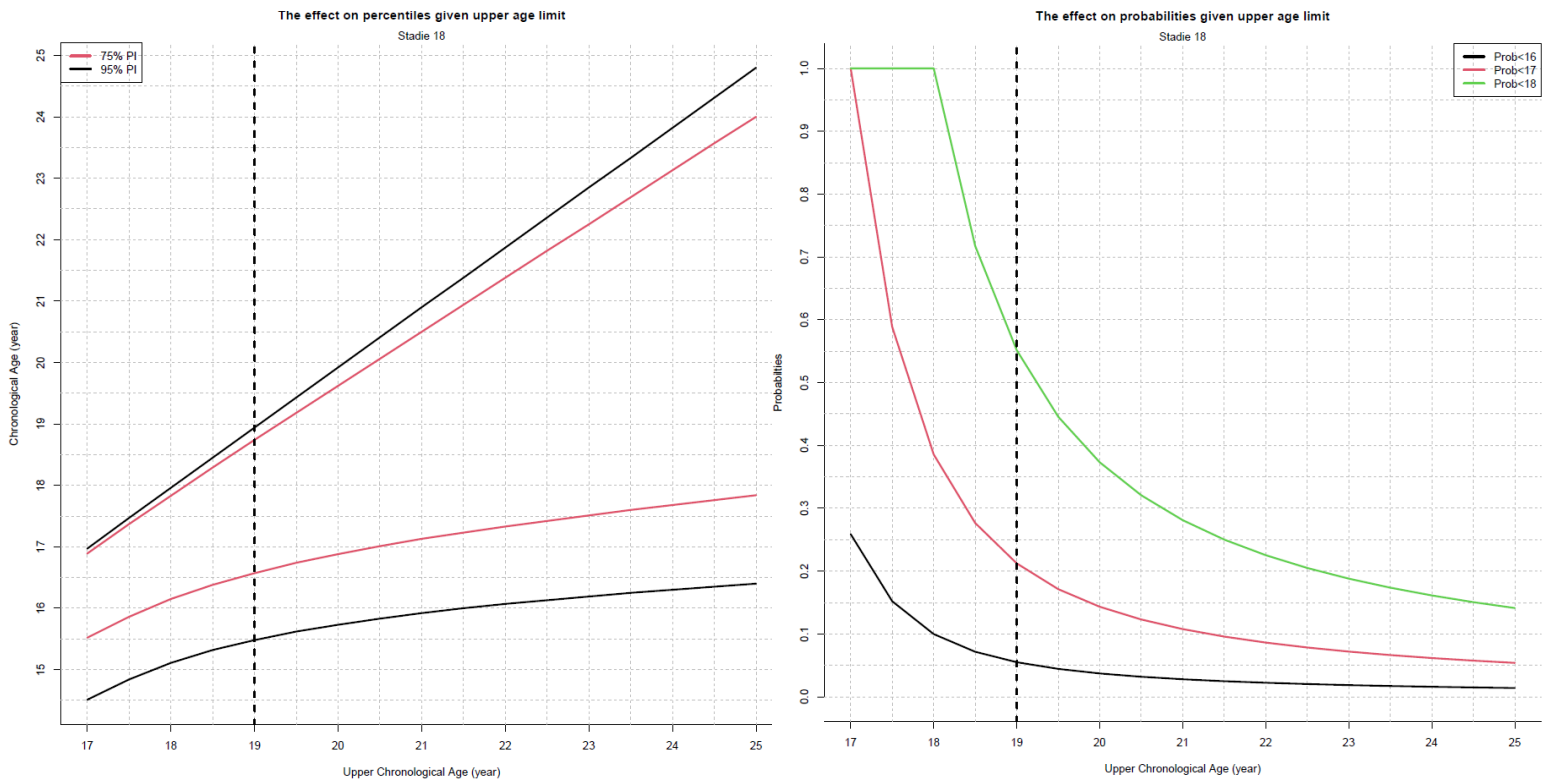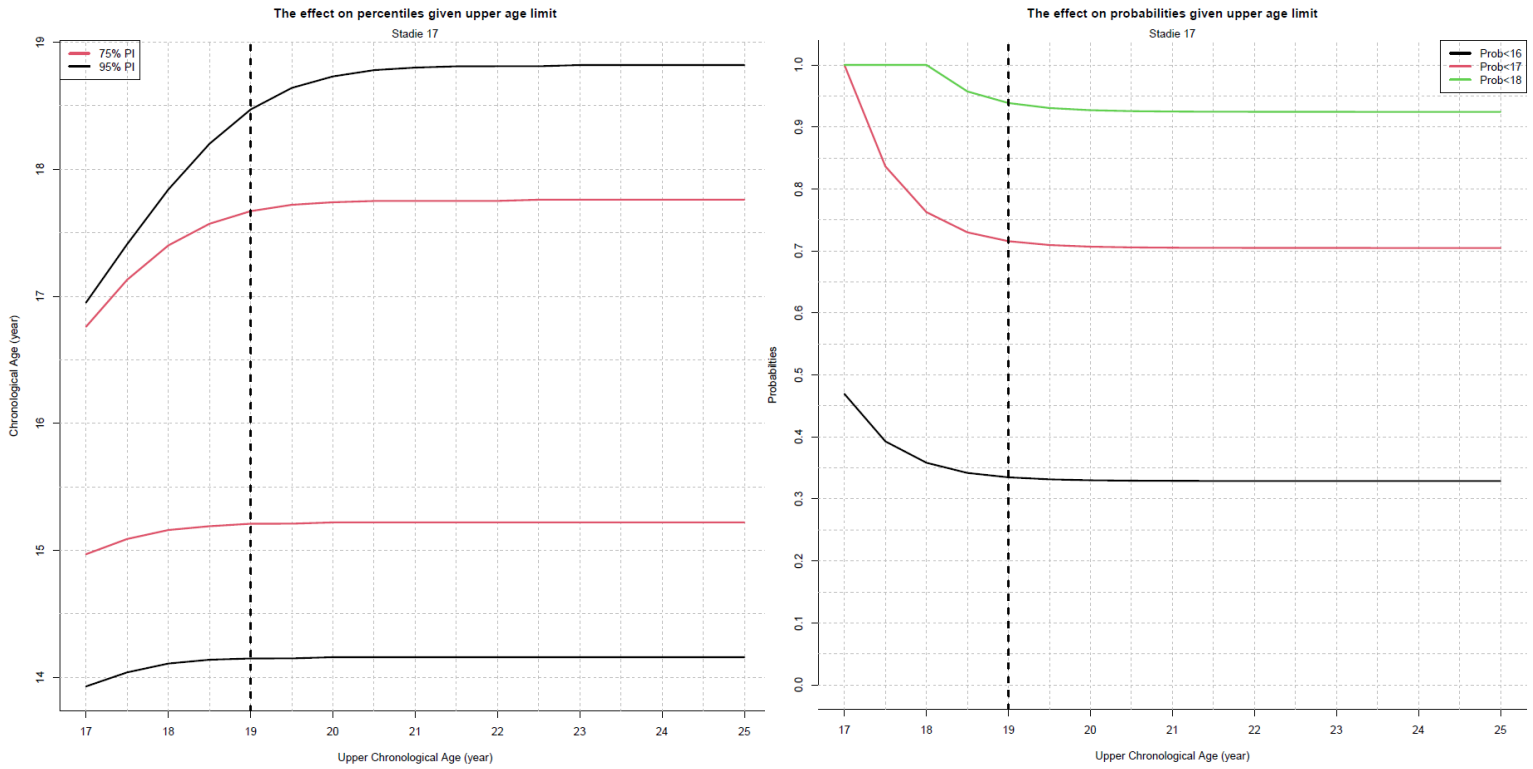### B.4.3.3   Effect for boys: combined stages 19/F, 19/G, 18/H and 19/H

Figure 39: The figure shows prediction intervals and probabilities for boys with skeletal age 19 years and tooth stage H.

Figure 40: The figure shows prediction intervals and probabilities for boys with skeletal age 18 years and tooth stage H.

Figure 41: The figure shows prediction intervals and probabilities for boys with skeletal age 19 years and tooth stage G.

44
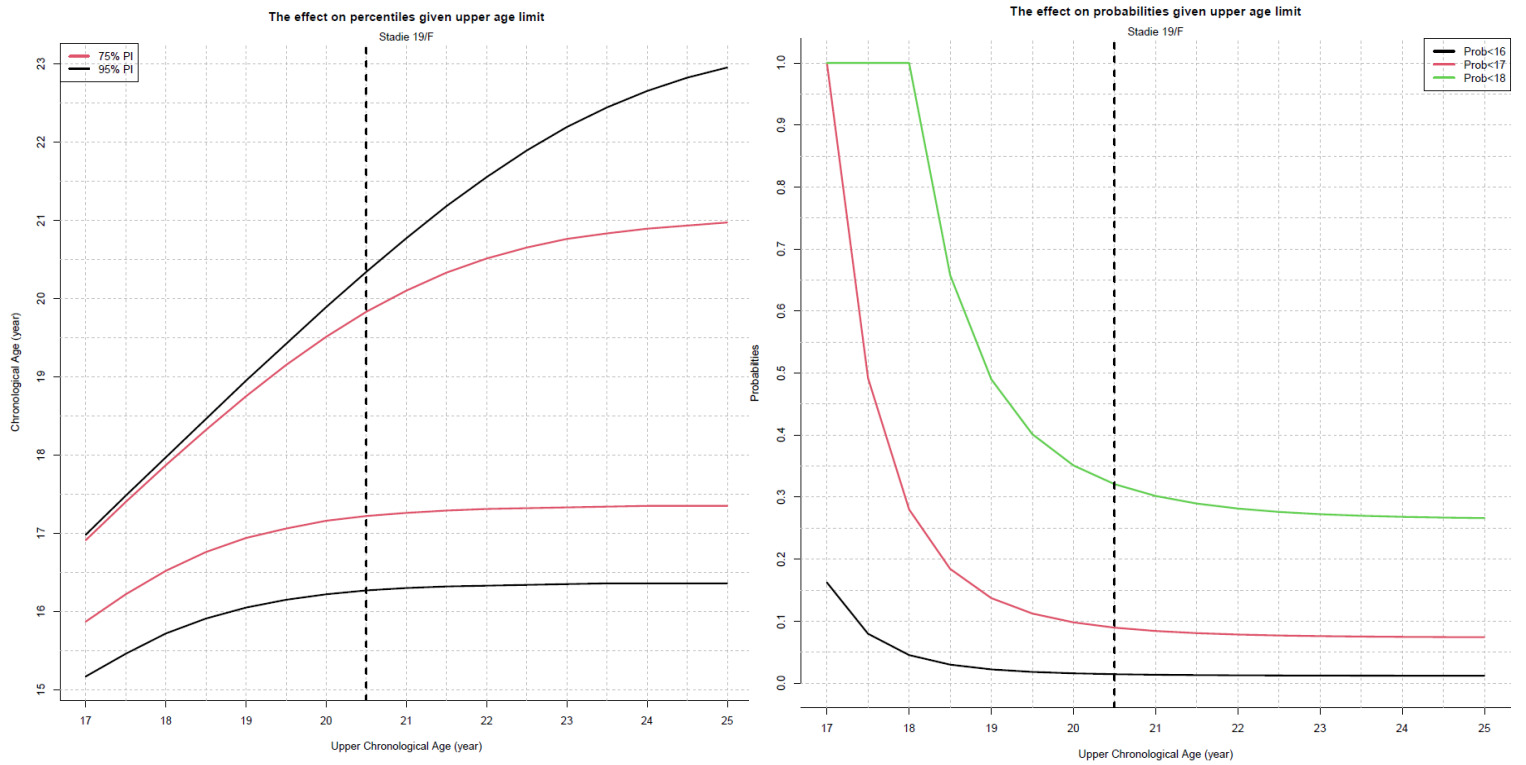
Figure 42: The figure shows prediction intervals and probabilities for boys with skeletal age 19 years and tooth stage F.

B.4.3.4 Effect for girls: hand skeletal ages 17 and 18 years

Figure 43: The figure shows prediction intervals and probabilities for girls with a skeletal age of 18 years.

45

Figure 44: The figure shows prediction intervals and probabilities for girls with a skeletal age of 17 years.
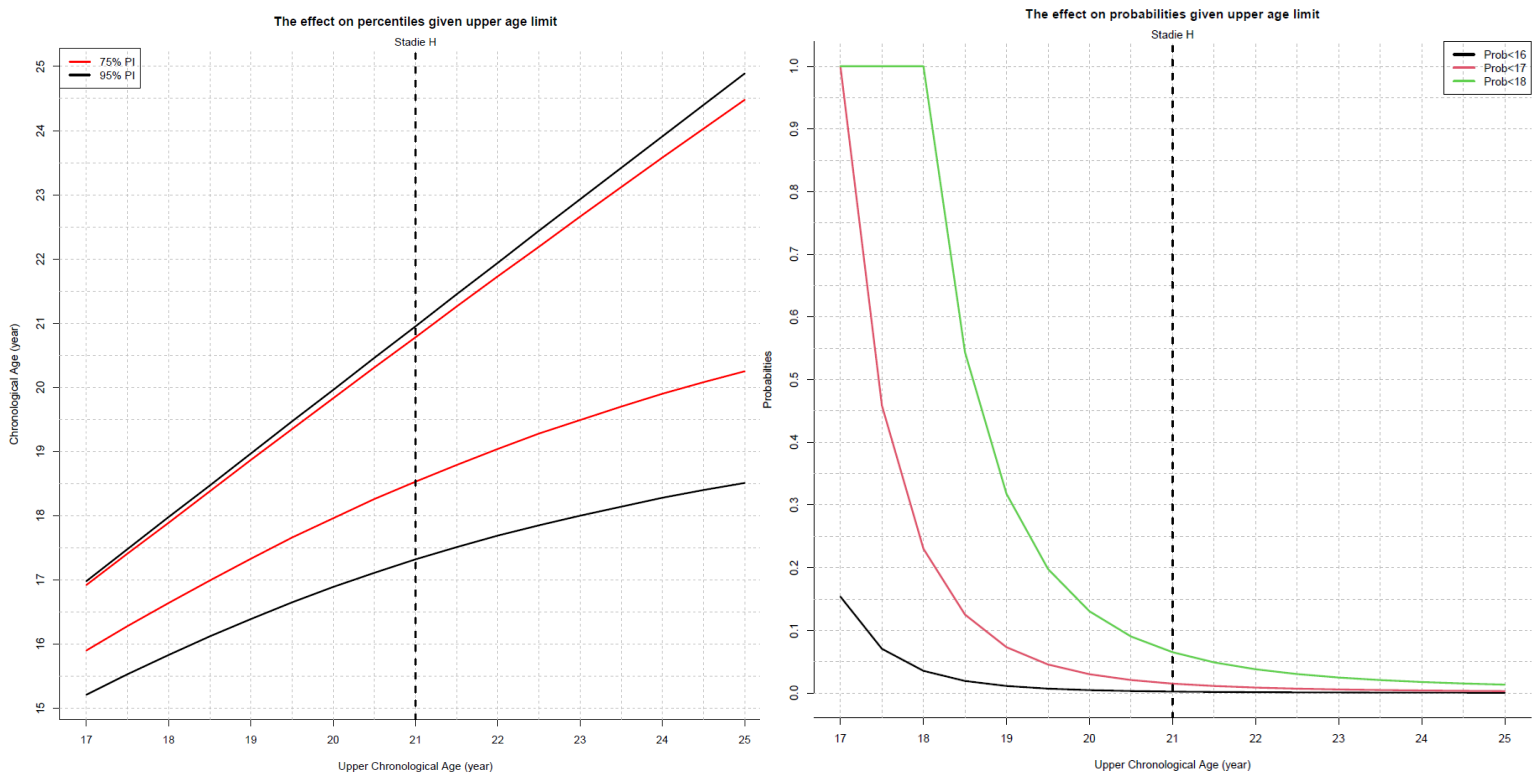
B.4.3.5   *Effect for girls: tooth stages G and H*

Figure 45: The figure shows prediction intervals and probabilities for girls with tooth stage H.
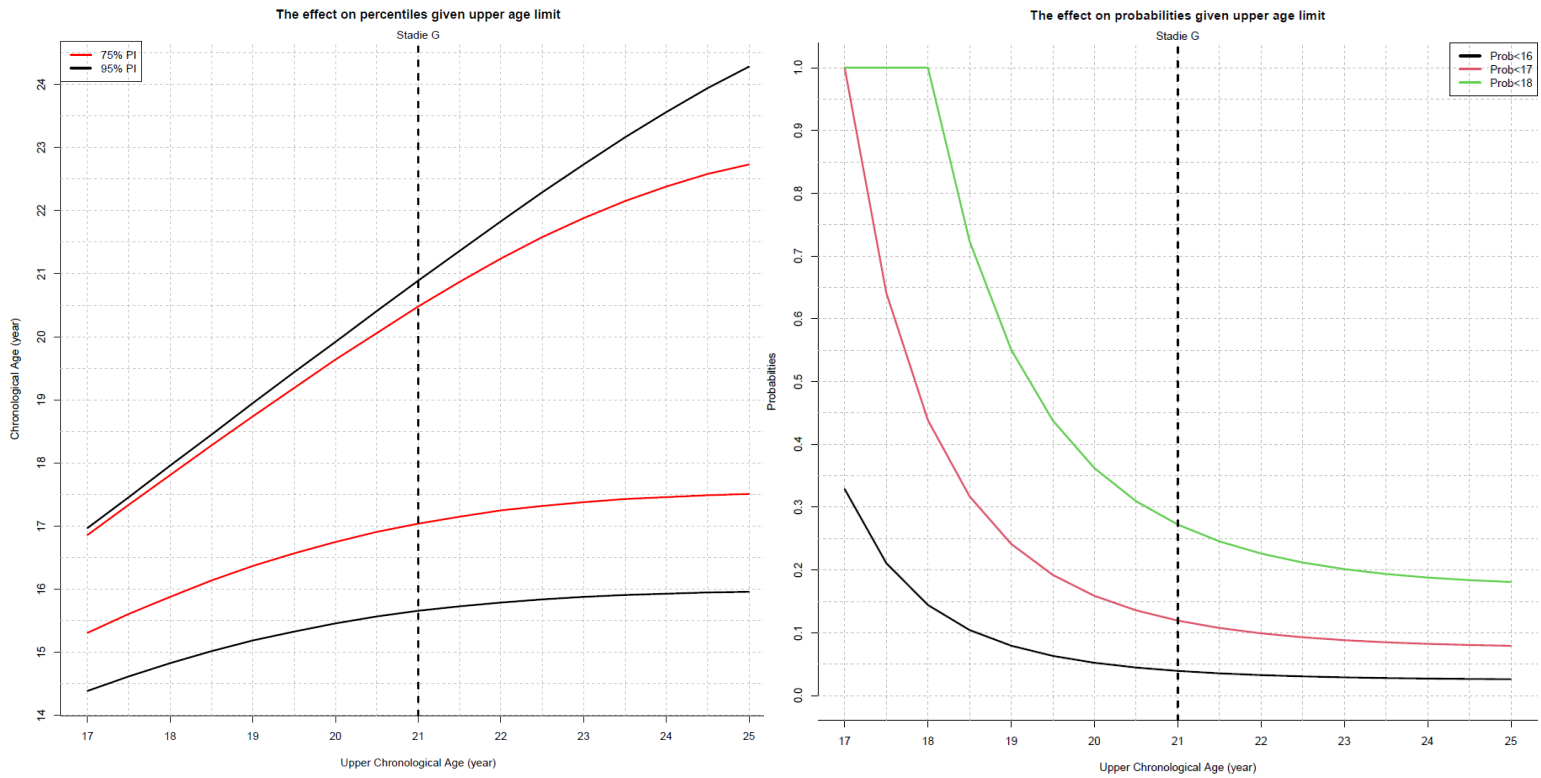
46

Figure 46: The figure shows prediction intervals and probabilities for girls with tooth stage G.

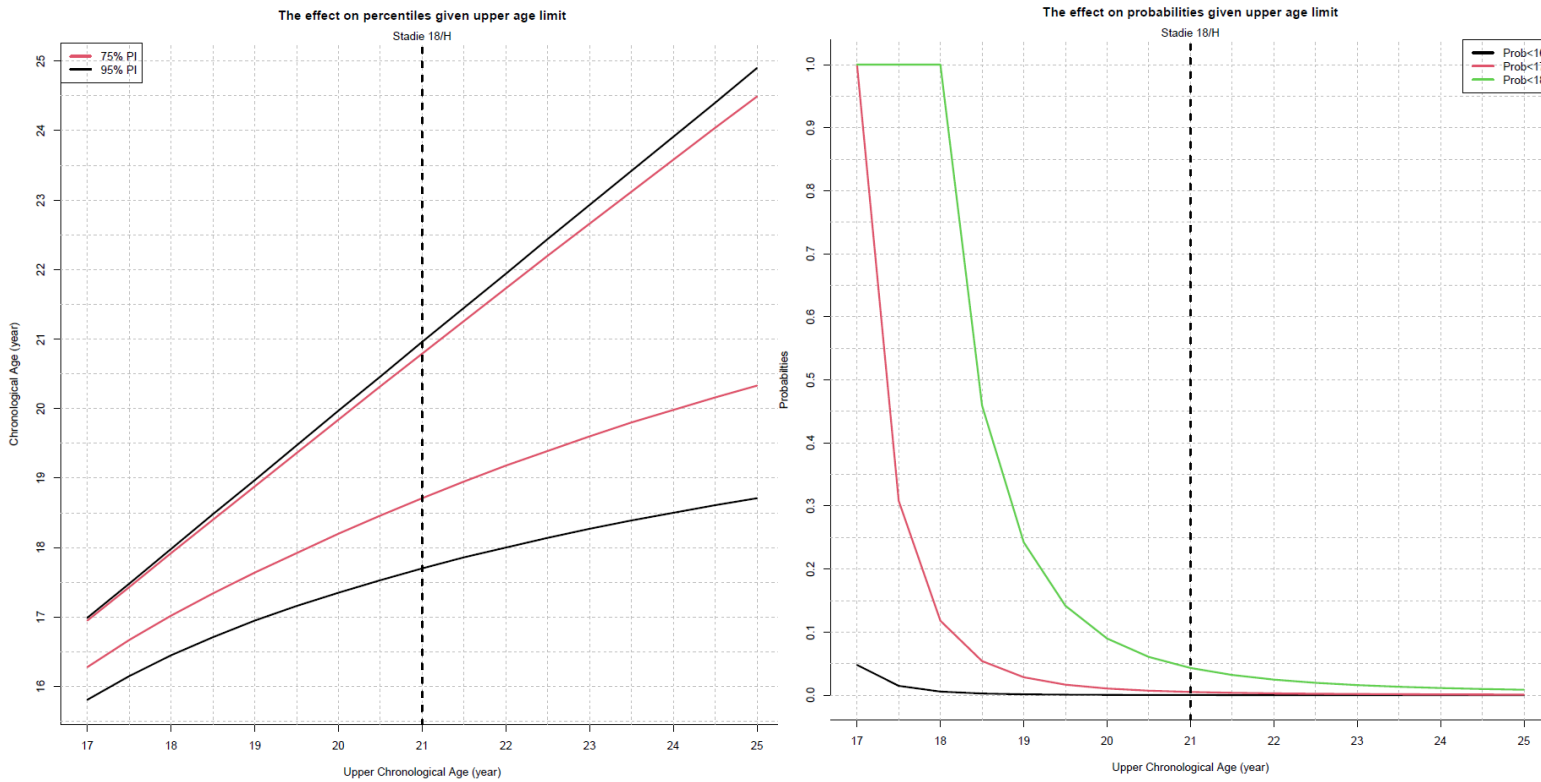### B.4.3.6 Effect for girls: combined stages 18/F, 18/G, 17/H and 18/H

Figure 47: The figure shows prediction intervals and probabilities for girls with skeletal age 18 years and tooth stage H
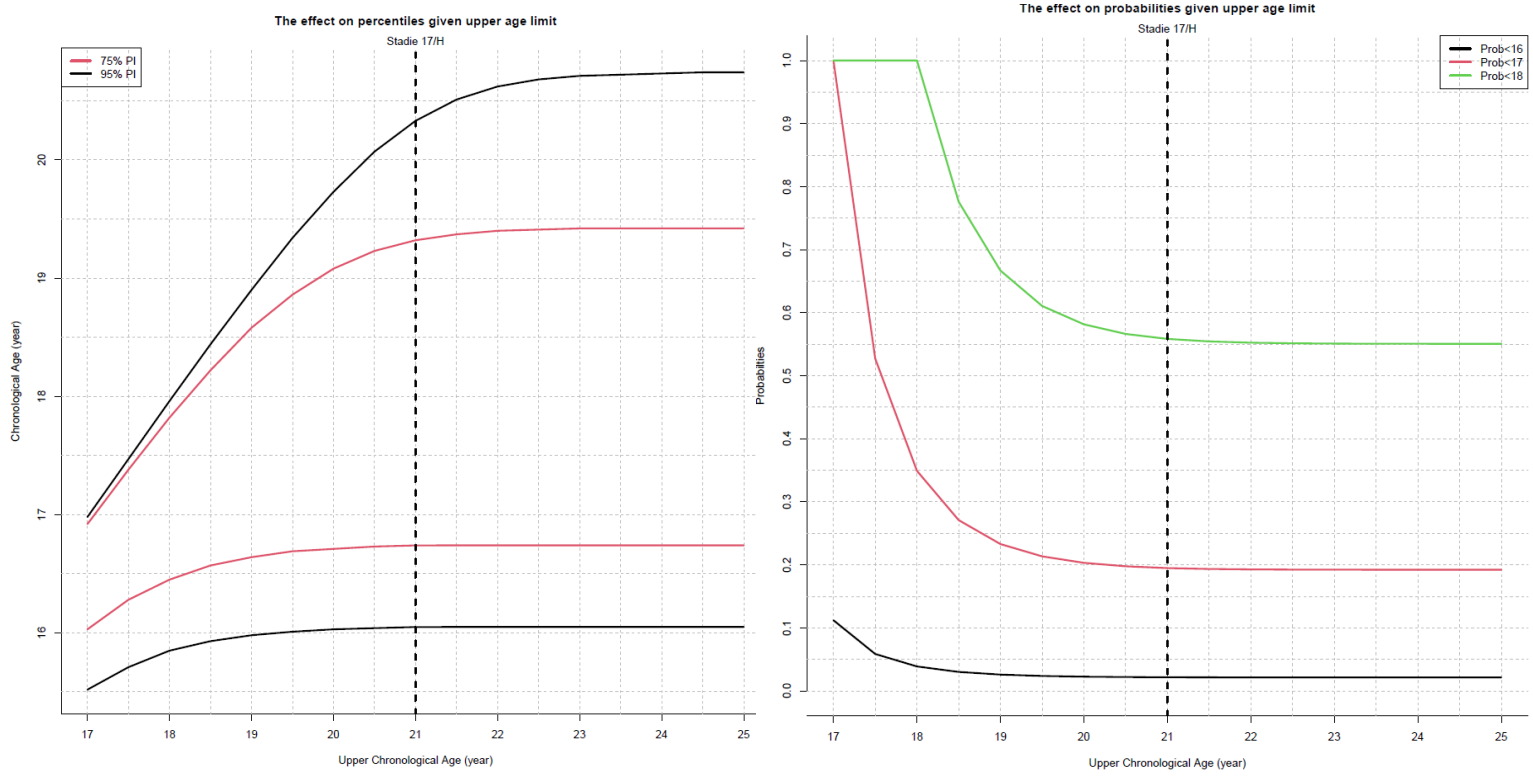
47

Figure 48: The figure shows prediction intervals and probabilities for girls with skeletal age 17 years and tooth stage H.
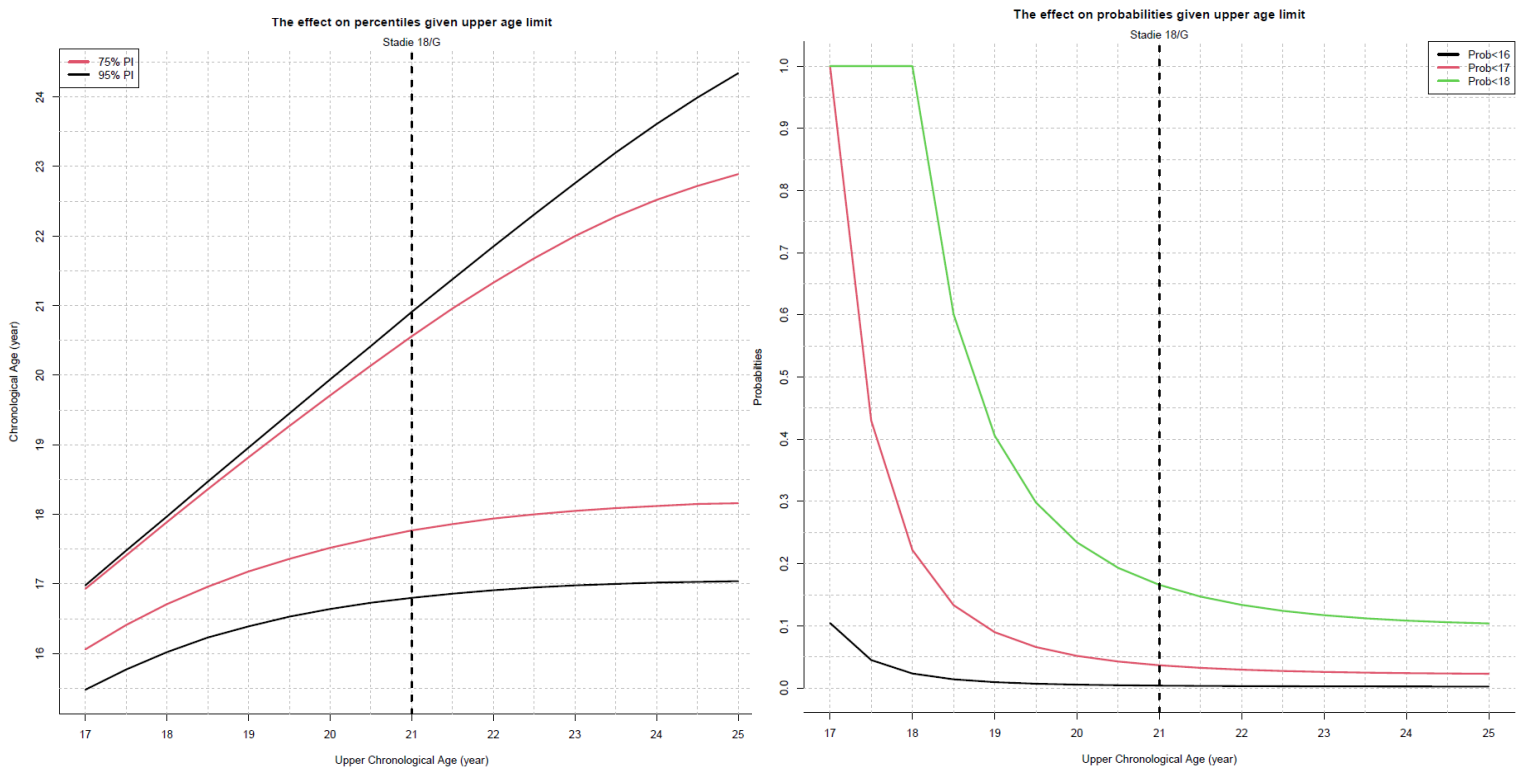

Figure 49: The figure shows prediction intervals and probabilities for girls with skeletal age 18 years and tooth stage G.
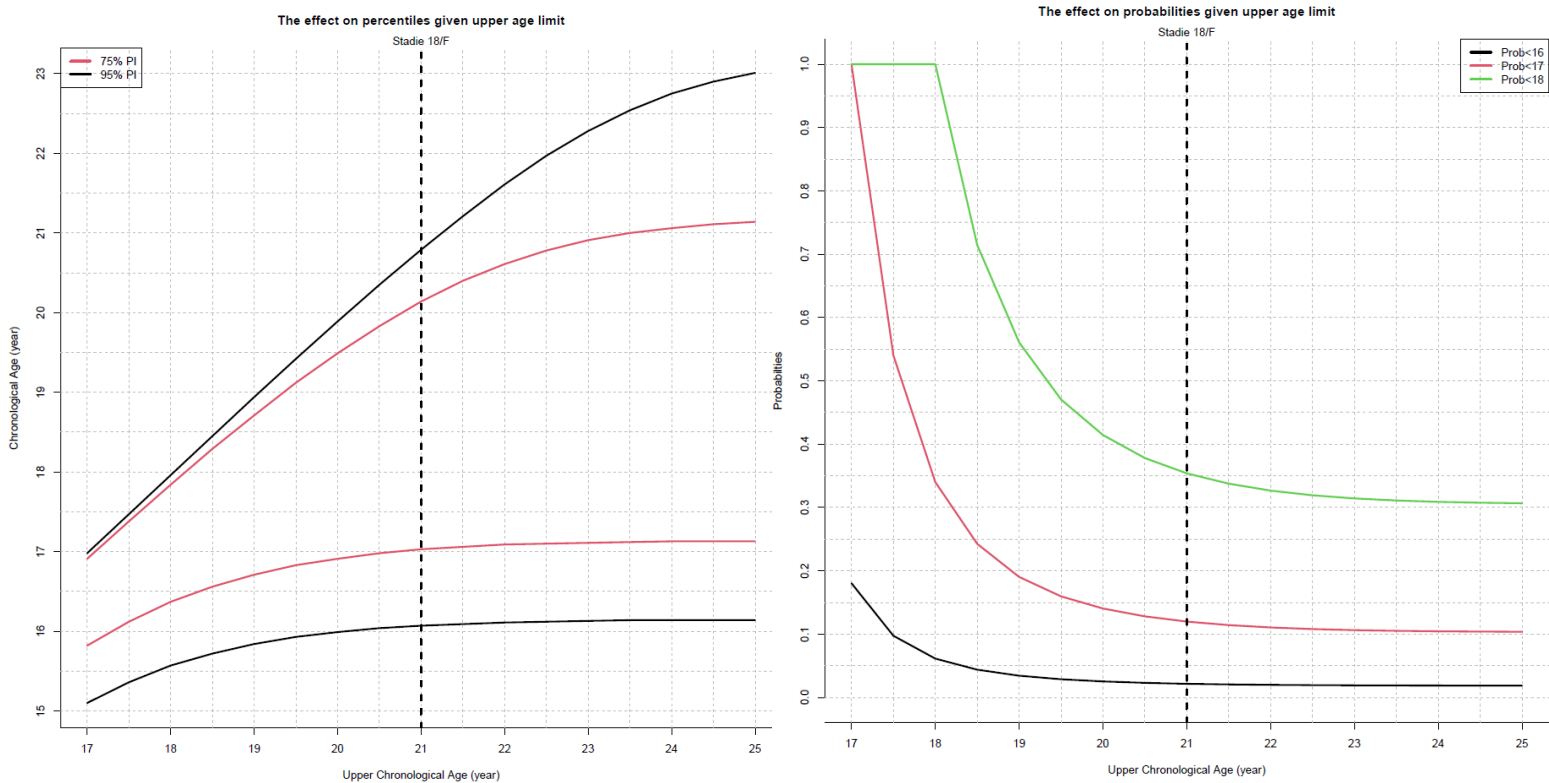
48

Figure 50: The figure shows prediction intervals and probabilities for girls with skeletal age 18 years and tooth stage F.

C. References

1.   Boldsen JL, Milner GR, Konigsberg LW, Wood JW. Transition analysis: a new method for estimating age from skeletons. CAMBRIDGE STUDIES IN BIOLOGICAL AND EVOLUTIONARY ANTHROPOLOGY. 2002:73-106.

2.   Buken B, Safak AA, Yazici B, Buken E, Mayda AS. Is the assessment of bone age by the Greulich-Pyle method reliable at forensic age estimation for Turkish children? Forensic Science International. 2007;173(2):146-53.

3.   Konigsberg LW, Herrmann NP, Wescott DJ, Kimmerle EH. Estimation and evidence in forensic anthropology: age-at-death. Journal of forensic sciences. 2008;53(3):541-57.

4.   Yee TW. VGAM: Vector Generalized Linear and Additive Models 2017 [Available from: https://CRAN.R-project.org/package=VGAM.

5.   Mostad P, Schmeling A, Tamsen F. Mathematically optimal decisions in forensic age assessment. International journal of legal medicine. 2022;136(3):765-76.

6.   Aarseth S, Sund T, Müller LSO, Bring J. Feil i aldersvurderingen av unge asylsøkere. Tidsskrift for den Norske Laegeforening. 2022;142(11).

7.   Bachs L, Bleka Ø, Aanes H, Rolseth V. Vi tar usikkerheten rundt biologisk aldersvurdering på alvor. Tidsskrift for den Norske Laegeforening. 2022;142(13).

8.   Gelbrich B, Frerking C, Weiss S, Schwerdt S, Stellzig-Eisenhauer A, Tausche E, et al. Combining wrist age and third molars in forensic age estimation: how to calculate the joint age estimate and its error rate in age diagnostics. Annals of human biology. 2015;42(4):389-96.