



## Manual

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

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

## Table of contents

1	Preface.....	3
2	Main points.....	3
3	Introduction.....	4
3.1	About BioAlder.....	4
4	BioAlder in practice .....	5
4.1	Result report generated by the tool .....	7
4.2	Potential and limitations of the tool .....	8
4.3	User training .....	9
5	General information about biological age investigation methods .....	10
5.1	Age estimation based on X-rays of the hand .....	12
5.2	Age estimation based on dental X-rays.....	13
6	Systematic reviews .....	13
6.1	Greulich & Pyle age estimation atlas .....	13
6.2	Age estimation based on Demirjian staging of wisdom teeth formation .....	15
6.3	Regional differences .....	18
6.3.1	Development of the hand skeleton.....	18
6.3.2	Formation of wisdom teeth.....	18
6.3.3	Conclusion.....	18
6.4	The situation after the systematic reviews .....	19
7	Statistical modelling of data from included studies.....	19
7.1	Purpose .....	19
7.2	Studies that can be used for modelling.....	20
7.2.1	Type 1.....	21
7.2.2	Type 2.....	21
7.2.3	Type 3.....	22
7.2.4	Type 4.....	22
7.3	Modelling individual data .....	23
7.3.1	Modelling type 2 data .....	23
7.3.2	Modelling type 3 data .....	23
7.3.3	Modelling type 4 data .....	23
7.4	From individual data to probabilities.....	25
7.5	Distribution of chronological age given observed stage .....	27
7.6	Combination of hand and tooth.....	28
8	Results used in BioAlder .....	30
8.1	Overview of studies used in BioAlder .....	30
8.2	Choice of upper age limit .....	31
9	The future of biological age estimation .....	33
9.1	Image-based methods .....	33
9.2	DNA methylation.....	33
10	References.....	34

# 1 Preface

With effect from 2016, national responsibility for assessing the age of unaccompanied minor asylum-seekers in Norway has rested with the forensic toxicology unit at the Norwegian Institute of Public Health (now the Department of Forensic Medicine at Oslo University Hospital). Since then, a project group has been established, and work to accomplish the assignment has involved making systematic reviews (1, 2) 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 describes the work of constructing the BioAlder tool, which is designed to estimate prediction intervals for the unknown true age of an asylum seeker on the basis of X-rays of a wisdom tooth and hand skeleton. The work has been carried out by the research group on age assessment at the Department of Forensic Medicine, Division of Laboratory Medicine, Oslo University Hospital (OUH):

- Liliana Bachs MD PhD (group leader), assistant head of department, OUH
- Øyvind Bleka, PhD, researcher/statistician, OUH
- Pål Skage Dahlberg MSc, researcher, OUH
- Veslemøy Rolseth PhD, researcher, OUH
- Gerd-Jorunn Møller Delaveris MD PhD, head of section, 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 and Ivan Galić for submitting datasets.

In the 1.1 version of BioAlder, new data has been included and some changes in the model have been undertaken in the light of the new information.

Additionally a separate data set from Lebanon with 254 young individuals (132 females and 122 males) has been used for validation (see paper "BioAlder: a tool for assessing chronological age based on two radiological methods", submitted to International Journal of Legal Medicine).

## 2 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 Western countries.

- BioAlder, the age assessment tool described in this document, makes an automated prediction of chronological age on the basis of results from X-rays of the wisdom tooth and hand skeleton.
- The instrument 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 22 scientific publications, and includes data on over 17 000 people. The tool will be updated regularly with new research data.
- The X-ray examinations upon which the tool is based show a wide natural biological variation. BioAlder yields 75% and 95% prediction intervals for chronological age, which clearly shows this variation to executive officers whose job it is to establish/determine an age.
- The data include studies conducted in 16 different countries. The significance of regional differences remains unclear.
- The tool is a temporary solution. In the future we aim to further develop molecular biological methods of age estimation (DNA methylation).

## 3 Introduction

Unaccompanied minor asylum-seekers who come to Norway have rights pursuant to Norwegian law and international guidelines and conventions (1). 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 (2).

The methods currently in use for determining biological age are unable to provide a precise age (3, 4). The greatest constraint is the natural biological variation in the development of skeleton and teeth, which are the analytical methods most frequently used. Nor are there any scientifically documented systems for psychosocial or cognitive testing that can provide a reliable estimate of chronological age.

### 3.1 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 skeleton and lower left wisdom tooth in more than 17000 young persons of known chronological age. BioAlder is used to assess the individual asylum-seeker's developmental stages on the basis of X-ray images of the applicant's hand 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, Department of Forensic Medicine, 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 research group on age assessment at the Department of Forensic Medicine, 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 X-rays of the hand skeleton and teeth, which were also components of the system used in Norway until recently. What is new is that the best documented methods for staging development have been selected, all available scientific studies on these stages collated, and finally a mathematical model has been constructed that makes it possible to combine hand and tooth results. To the best of our knowledge, the system is the first of its kind.

The tool has been optimised for assessing the age of young asylum-seekers, and cannot be used indiscriminately in other connections. The systematic reviews are under the process of being published internationally. The mathematical modelling forming the basis for the tool is published (5).

Some discretionary decisions have to be made in connection with any developments. 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 trafficking.

The tool 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 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 9.2.

## 4 BioAlder in practice

Use of the tool presupposes obtaining informed consent from the person being assessed, in line with current laws and regulations. The person must also have had the opportunity to give notification of any chronic diseases, developmental disorders or medication.

The graphic display of the tool contains a simple and intuitive user interface in which three different items of information are entered: gender, estimated Greulich & Pyle skeletal age and/or estimated Demirjian's stage of the lower left wisdom tooth (see Figure 1). Each combination of these data generates a report, and a collection of reports will be delivered to UDI as pdf files.

The combined use of two independent measures, one from hand and one from tooth, is one of the strengths of our model. The use of the results when only one of the methods is available, either hand or teeth, will yield a weaker evidence and should be used with even more caution.

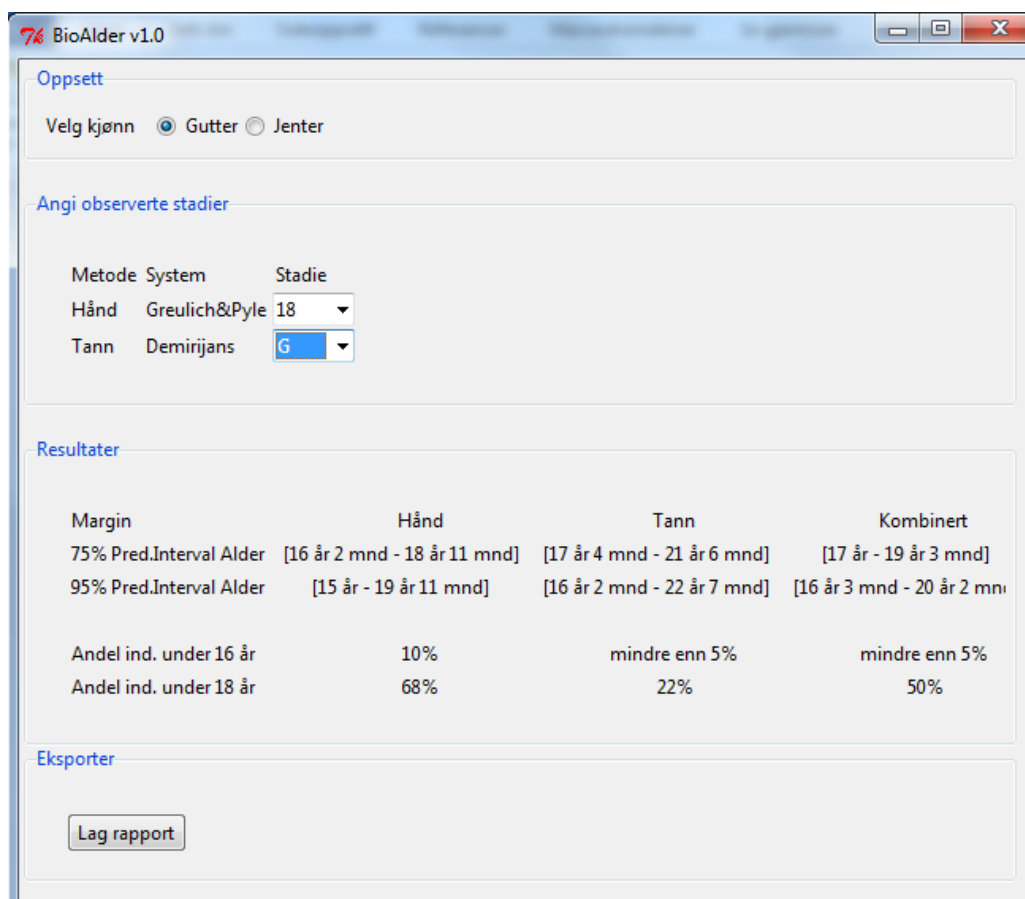


Figure 1. Example of the BioAlder screen display.

The following table gives a translation of all the words contained in Figure 1:

Oppsett	Setup	Resultater	Results
Velg kjønn	Choose gender	Margin	Margin
Gutter	Boys	Hånd	Hand
Jenter	Girls	Tann	Tooth
Angi observerte stadier	Select observed stages	Kombinert	Combined
Metode	Method	75% Pred.Interval Alder	75% Pred.Interval Age
System	System	16 år 2 mnd – 18 år 11 mnd	16 yrs 2 mo. 18 yrs 11 mo.
Stadie	Stage	Andel ind. Under 16 år	Perc. ind. under 16 years
Hånd	Hand	Mindre enn 5%	Less than 5%
Greulich&Pyle	Greulich&Pyle	Eksporter	Export
Tann	Tooth	Lag rapport	Generate report
Demirijans	Demirjian's		

The results of the analysis appear immediately, and are reported with 75% and 95% prediction intervals for chronological age, and the percentages of individuals aged under 16 years, under 17 years and under 18 years (see Figure 17 for more information). As shown in the above illustration, the tool displays estimates for each method and for the methods in combination.

#### 4.1 Result report generated by the tool

The report consists of two main parts: 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 observed stages).
- Percentages of individuals under the ages of 16, 17 and 18 (given 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%”.

### Report on biological age assessment – Boys S 19/G

USED WHEN BOTH HAND AND TOOTH SCORES ARE AVAILABLE. The following results were reached following mathematical modelling of research data. The research compared observed development stages with known chronological age. The total number of observations of development stages and known chronological age used in the mathematical model were 3258 for hands and 4082 for teeth for boys. The modelling calculates prediction intervals and percentages on the basis of the combined observations. It is not clear how representative the observations are for the individual asylum-seeker who is to be assessed, and the results from this tool must be used with caution. Manual v1.0.1 contains more details about the tool.

#### **Combined Greulich & Pyle skeleton age/Demirjian's formation stage: 19/G**

Prediction interval: 75% of the individuals will be between 18 years 3 months and 21 years 3 months.  
Prediction interval: 95% of the individuals will be between 17 years 4 months and 21 years 10 months.  
Percentage of individuals under 16 years: less than 5%.  
Percentage of individuals under 18 years old: 15%

Note: This result presupposes that X-rays of the hand and tooth were performed less than 2 months apart.

-----

#### **Subsidiary assessments are reported for information purposes:**

Observed Greulich & Pyle skeletal ages based on hand X-rays: 19

Prediction interval: 87.5% of the individuals will be over 17 years 10 months.  
Prediction interval: 97.5% of the individuals will be over 16 years 7 months.  
Percentage of individuals under 16 years: less than 5%.  
Percentage of individuals under 18 years old: 15%

Observed Demirjian's formation stage based on X-ray of lower left wisdom tooth: G

Prediction interval: 75% of the individuals will be between 17 years 4 months and 21 years 5 months.  
Prediction interval: 95% of the individuals will be between 16 years 2 months and 22 years 7 months.  
Percentage of individuals under 16 years: less than 5%.  
Percentage of individuals under 18 years old: 22%

*This document was generated using BioAlder v1.0. The tool is an aid to the authorities in establishing the age of young asylum-seekers. It must not be used indiscriminately in other connections.*

Figure 2. Example of BioAlder results report.

## 4.2 Potential and limitations of the tool

The model that generates the results is based on a total of 17151 individuals (8699 boys and 8452 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 7 and Table 8). The populations from which many of the unaccompanied, minor asylum-seekers come 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 6.3). 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 final 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 X-rays. Medical conditions that may cause precocious skeletal maturation may lead to a person being assessed as older than their chronological age on the basis of hand X-rays. The most common causes of this latter effect in the Western population are overweight/obesity and the use of some medications (6). 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 (7, 8). 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 (9).

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 (compared with dental maturation) in the following combinations:

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

**Figure 3.** Stage combinations with the largest discrepancies, which may indicate accelerated skeletal maturation.

The reports for these combinations will contain a recommendation that the individual be investigated in more detail.

### 4.3 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 the tool, and will also take up practical aspects of its use.

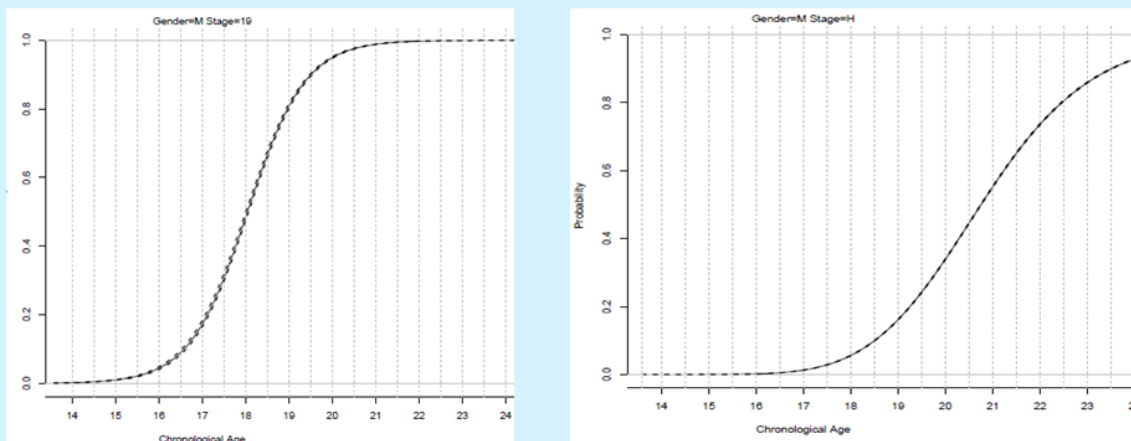
## 5 General information about biological age investigation methods

The biological age assessment systems of various countries are based on different methods (2). There is also considerable variation in the manner in which the same type of method is employed. A number of different staging systems are used for examining teeth, for example (10). In addition, many operators combine the results of several staging systems into one overall estimate. It is therefore 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 (2). In the former, it is most usual to use X-rays of the hand and wrist. When assessing dental development, it is usual to take a panoramic X-ray of all teeth (an orthopantomogram, or OPG). In the case of 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 point. 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 10). 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 4.** The end-stage problem in age estimation based on maturation of hand skeleton and teeth.

As both hand 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. Some countries therefore also perform an assessment of bones that mature later (2). 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 (7). There are also limited data on clavicles, particularly with respect to regional differences. Other bones that mature late are found in the knee (distal femur or proximal tibia). In Sweden, 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 (11). However, the research documentation is very limited, and the maturation stages of the knee extend over a number of years (the method has low resolution) (12). MRI is a technology that does not involve harmful ionising radiation, but the instruments are very expensive, and the test itself is time-consuming and complex to perform. Because of the magnetic field associated with MRI, it may also involve risk for persons with metal in their bodies.

A problem common to methods based on the development of skeletons and teeth is that there is substantial variation in natural biological development (3, 4). 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 skeleton and teeth is that it stops in the late teens or early twenties, which presents challenges to making a model for determining whether a person is an adult or a child. Wisdom teeth mature later than hands, and are therefore most suitable as a basis for determining age in the range 17–19 years.

## 5.1 Age estimation based on X-rays of the hand

When X-rays 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). There are several such systems. In some, discretionary judgement is used to find the image that is most similar (this applies, for example, to the Greulich & Pyle atlas) (13), 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) (14, 15). The most widely used system, on which there are also most scientific publications, is the Greulich & Pyle atlas (GP atlas). This is the staging system for hands that we recommend using for age estimations in Norway.

### The Greulich & Pyle atlas

“The Radiographic Atlas of Skeletal Development of the Hand and Wrist” was first published in 1950, and is still one of the most widely used reference standards for assessing skeletal age. The atlas consists of reference pictures of girls and boys separately, and extends from birth up to 18 years for girls and up to 19 years for boys. Alongside each picture is a skeletal age and a description of important changes that can be observed for this skeletal age. The way the method works in practice is that the person assessing the X-ray image finds the image in the atlas that it most closely resembles.



Figure 5. 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 (13). 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 6.1 for further details.

## 5.2 Age estimation based on dental X-rays

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. Examples of staging systems are Demirjian, Goldstein and Tanner from 1973 (8 stages denoted A to H) (16) and Hunt and Gleiser from 1955 (15 stages) (17). There are several variations of the latter, such as Moorees et al., 1963 (14 stages)(18), Haaviko et al., 1970 (12 stages) (19) Kullmann et al., 1992 (7 stages) (20) and Kohler et al., 1994 (10 stages) (21).

### 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 X-ray image finds the stage that best describes the wisdom tooth in the X-ray image of the individual being assessed.

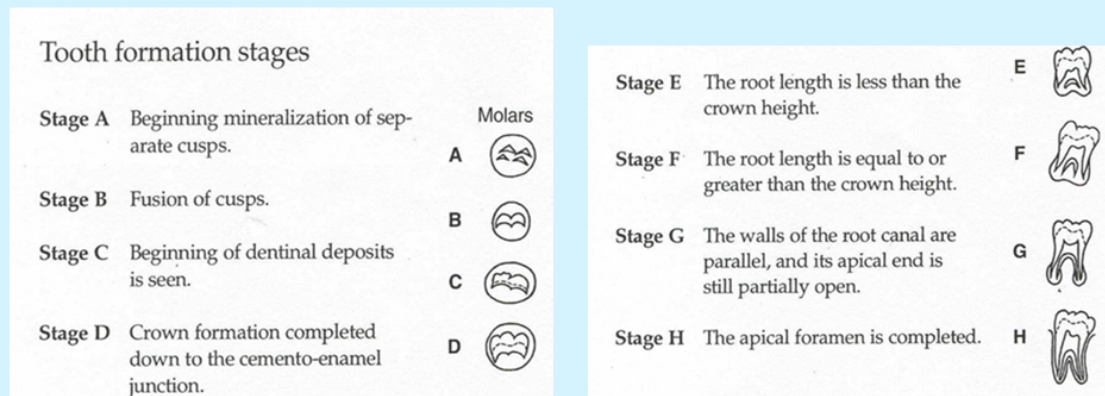


Figure 6. Demirjian's staging of teeth (molars) (*modified from (16)*).

A number of studies have examined the relationship between the formation stages of wisdom teeth and chronological age. See Part 6.2.

## 6 Systematic reviews

In the period February 2016 to March 2017 we collaborated with the Knowledge Centre for the Norwegian Institute of Public Health on two systematic reviews. One dealt with age estimation based on hand X-rays using the Greulich & Pyle atlas, and the other with age estimation based on Demirjian's staging of the formation of wisdom teeth.

### 6.1 Greulich & Pyle age estimation atlas

March 2017 saw the completion of a systematic review on the use of the GP atlas to estimate age (3). Studies in this area normally present their results in one of two ways. Both assume a known chronological age and an observed skeletal age. 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 for each age group.

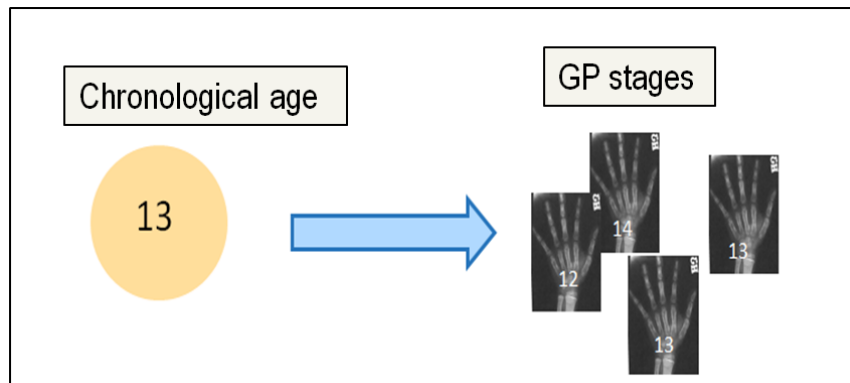


Figure 7. Approach A.

- B) **Studies that describe chronological age:** take skeletal age as the starting point and present mean and variance of chronological age for all individuals in the same skeletal stage collectively.

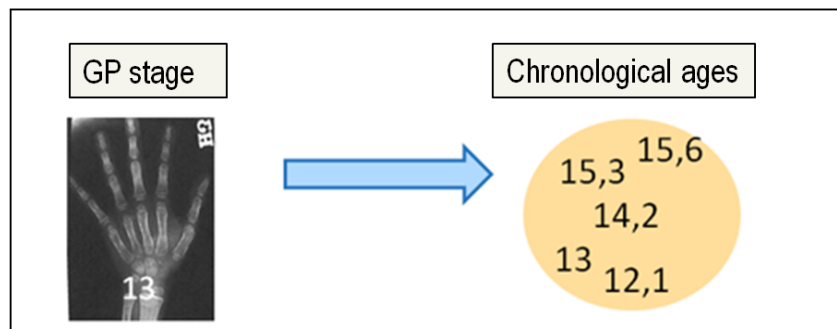


Figure 8. Approach B.

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.

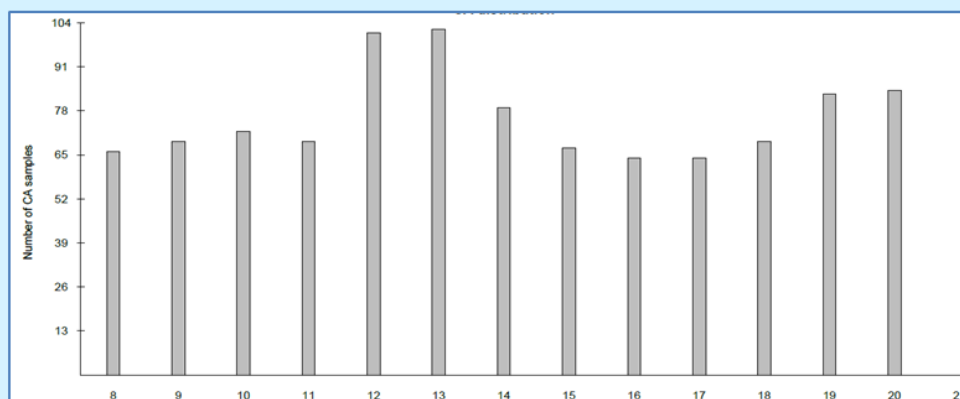
A meta-analysis was performed for the articles with approach A (15 articles). The main finding was that the consistency between skeletal age and chronological age was relatively good for modern populations (the difference was seldom more than one year at group level).

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 are processed the results reflect this (a phenomenon

called age mimicry; see figure 10. We were therefore left with one study using approach B (Chaumoitre 2016) that had more reliable results. Chaumoitre 2016 is a relatively large study with an unspecified multi-ethnic population in Marseille (see Figure 9). Thus it is a well-executed modern study of a population of mixed ethnic origin.

### Chaumoitre et al. 2016

Chaumoitre 2016 was the only study included in our knowledge summary 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 in whole years along 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:

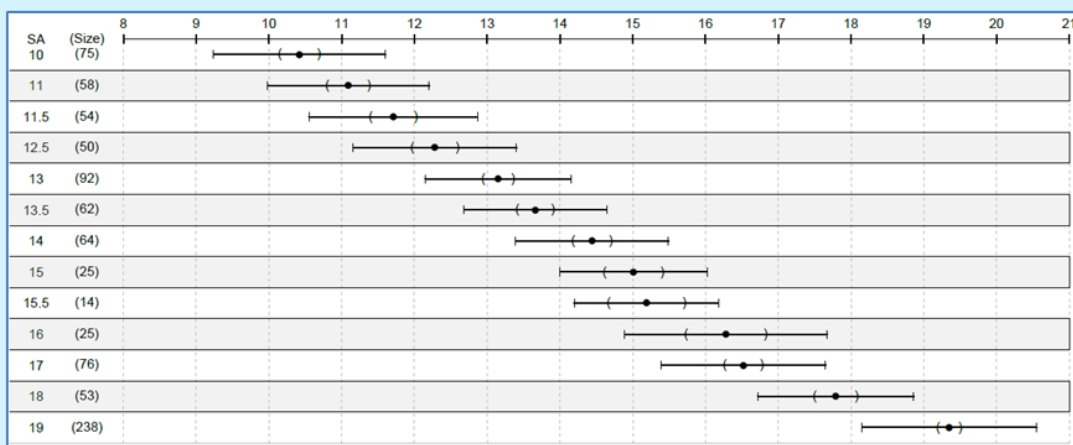


Figure 9. From Chaumoitre et al., 2016.

## 6.2 Age estimation based on Demirjian staging of wisdom teeth formation

A systematic review on the relationship between age and Demirjian's formation stages for wisdom teeth was also completed in March 2017 (4). We found 18 relevant studies, all published after 2005. They were from 13 different countries, and 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. 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 (see figure 10), 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 bias in the study design, we were unable to combine the studies in a meta-analysis and were therefore unable to reach conclusions as to whether there are differences in the formation of wisdom teeth among populations from different regions.

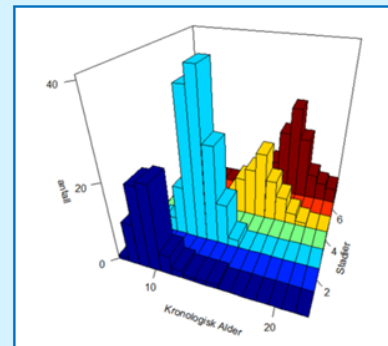


## Age mimicry

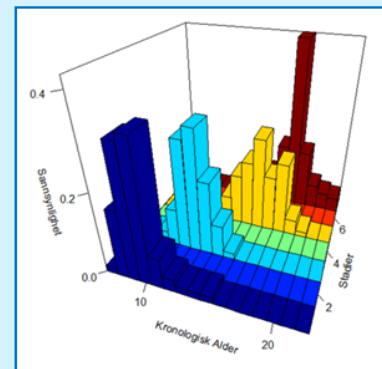
On the right is one type of dental study results table, slightly simplified. In this table, four dental stages are labelled with letters at the top, and the ages are given horizontally on the left. The tables themselves show how many individuals from each chronological year were assessed for the four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

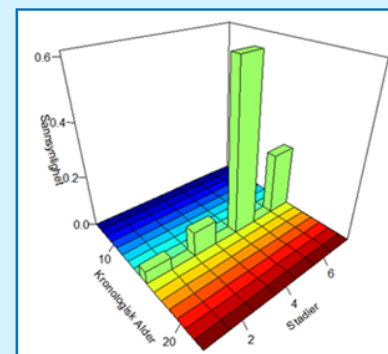


Figure 10. Age mimicry in studies of biological age estimation.

## 6.3 Regional differences

### 6.3.1 Development of the hand skeleton

The systematic review of the hand X-ray studies indicates that there may be differences of more than one year for populations from different parts of the world, but that differences that large are rare (3). A study based on automated measurement off hand X-ray images (BoneXpert software) shows similarly (22) 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 to factors such as dietary variations. There are also many populations in the world that have not been studied. Mapping the regional differences would have demanded a very extensive project, and mapping the causes an even larger project.

### 6.3.2 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 (4) had an included population that was skewed with respect to age, leading to age mimicry (see figure 10) and unreliable results. We therefore ended up with just a few studies that could be used for modelling.

The studies Lee 2009 (23), Li 2012 (24) and Johan 2012 (25) have a generally good study design. 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 (26).

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 1.** The table provides an overview of the mean age for Demirjian's stages F and G for the three well-designed studies in the systematic review in addition to a study from Botswana.

### 6.3.3 Conclusion

Any regional differences in skeletal and tooth maturation may have a variety of causes. These causes may be hereditary factors (regional genetic heterogeneity) or external factors (diet, climate etc.).

Many studies indicate that such differences exist in the maturation of both hands and teeth. However, these studies often have heterogeneity in the study design or method of reporting results, which makes them difficult to compare in order to create an overall picture. Age mimicry (see figure 10) may partially or fully explain the inconsistent results that have been attributed to regional differences. Thus it is not clear how much regional differences affect the results.

## 6.4 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 starting point for working towards a solution. All these studies reported their results groupwise, and many of them were biased by age mimicry (see figure 10). In order to be able to use the information in these studies, we started a project that uses statistical modelling to produce data in an entirely new way.

# 7 Statistical modelling of data from included studies

## 7.1 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 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 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 (see figure 10). 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 looked for studies in our search results from the two systematic reviews. In addition we carried out searches in PubMed to identify completely new publications.

Our aim is to say something about how chronological age is distributed at different stages, to 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 17).

## 7.2 Studies that can be used for modelling

Studies of hands and teeth consist of empirical data (observations) that have the same basic format: All individuals have a known chronological age and an observed developmental stage. These data are 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 this:

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 2.** Example of a hypothetical dataset.

Demirjian’s staging of teeth contains only eight stages, indicated by the letters A–H (16). The Greulich & Pale atlas for the hand skeleton contains more stages (as a rule one stage for each year, and sometimes also semi-annual images), and each stage is given an age in years (13). 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. The data formats we were able to continue working with are in a total of four formats (called types 1–4).

## Data formats included in our tool

**Type 1** consists of data in individual-based format.

**Type 2** is a frequency table with number of individuals for each stage within each whole year.

**Type 3** is tables with data on means and standard deviations of chronological age for given stages (skeletal age or tooth formation stage).

**Type 4** is tables with data on means and standard deviations of chronological age and skeletal age within each whole year.

**Figure 11.** Data formats for the studies included in our tool.

### 7.2.1 Type 1

This is the optimal data format, 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 lines.

TRIALNO	DOB	AGE	SEX	MALE	SKEL_AGE	Date radiogr	Calculated age
1	637	22.des.93	3	0 F	3	09.okt.97	3,8
2	637	22.des.93	3	1 T	3	09.okt.97	3,8
3	638	22.des.93	3	1 T	3	09.okt.97	3,8
4	638	22.des.93	3	1 T	3	09.okt.97	3,8
5	551	15.apr.93	4	0 F	3	09.okt.97	3,8
6	709	07.feb.92	5	1 T	3,3	09.jul.97	4,2
7	709	07.feb.92	5	1 T	3,6	12.nov.97	5,8
8	427	16.mai.92	4	0 F	4,1	07.mai.97	5
9	427	16.mai.92	4	1 T	4,1	07.mai.97	5
10	504	11.mar.92	5	1 T	4,3	18.jun.97	5,3
11	453	21.des.92	4	1 T	4,6	21.mai.97	4,4

**Table 3.** Type 1 data.

### 7.2.2 Type 2

In this data format, the numbers of individuals for each whole chronological year who were assessed for each stage are given. These tables show the stages horizontally at the top and chronological age vertically in the left-hand column.

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

Age	Demirjian's stage																	
	Male								Total Female								Total	
	A	B	C	D	E	F	G	H	A	B	C	D	E	F	G	H		
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 4.** Type 2 data.

The challenge presented by this type of data is that chronological age is only given in whole years (hence not sufficiently exact).

### 7.2.3 Type 3

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.

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 5.** Type 3 data.

We thus know the exact skeletal (or tooth) age of each individual, but chronological ages are not specified individually.

### 7.2.4 Type 4

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. In addition, the correlation (Pearson) between skeletal and chronological ages is given.

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 6.** Type 4 data.

Here we know neither the skeletal age nor the chronological age of the individual.

## 7.3 Modelling individual data

### 7.3.1 Modelling 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.

### 7.3.2 Modelling type 3 data

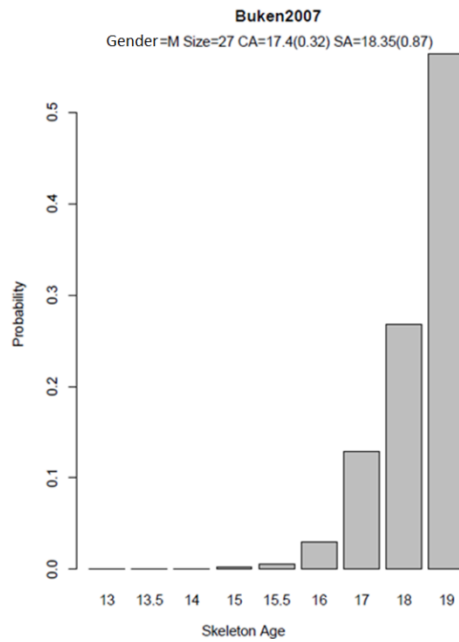
For the studies that report results in this format, we lacked only the chronological age, as all individuals in the same group have identical skeletal age (or tooth stage). Each individual must be assigned a chronological age. These ages are assumed to be normally distributed, and the data give the means and standard deviations (CA\_mean and CA\_SD in Table 5) which are used to assign chronological ages to all individuals in each of the rows in the table.

### 7.3.3 Modelling type 4 data

(See also part A.4.2 of the Appendix)

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 6 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:

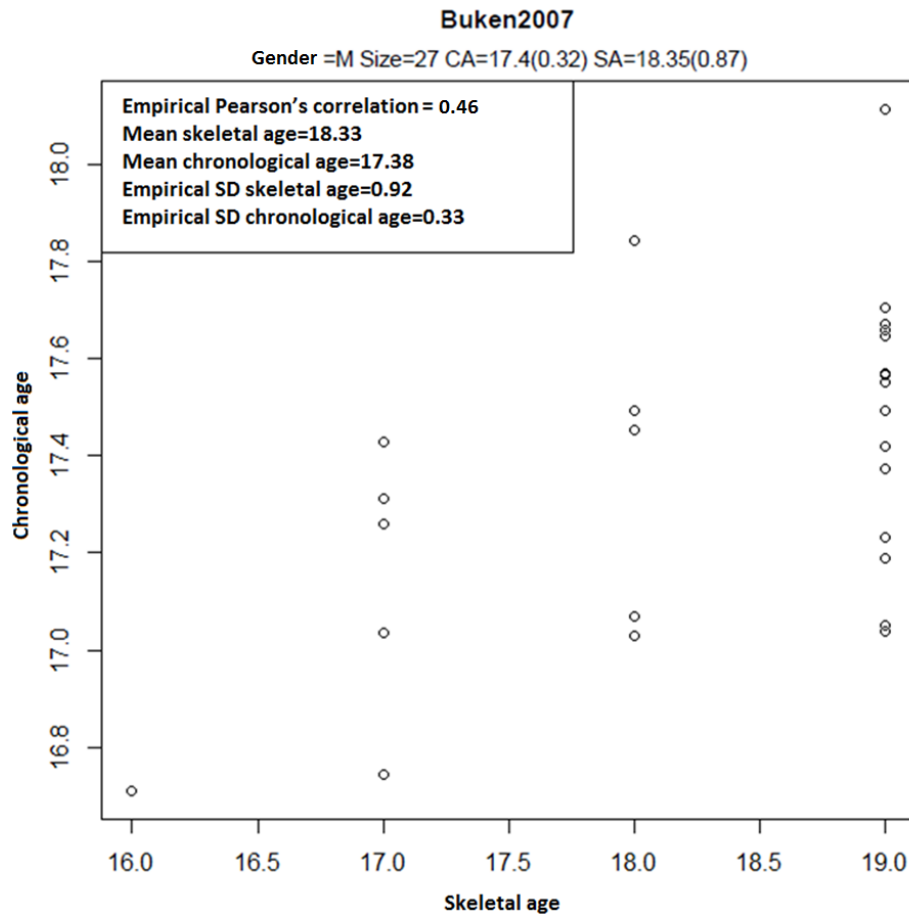


**Figure 12.** Modelling of type 4 data.

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 in Table 6 above as mean (CA mean) and standard deviation (CA sd), and we assume them to be normally distributed. We use the correlation value from Table 6 to 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) (27) are illustrated below in a scatter plot:





**Figure 13.** 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.

#### 7.4 From individual data to probabilities

(See also part A.2 of the Appendix)

With the aid of individual data that we have collected and/or generated as described above, we can construct a distribution of the probability that individuals of a given chronological age will be in the different stages. 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 14, using four stages for the sake of simplicity.

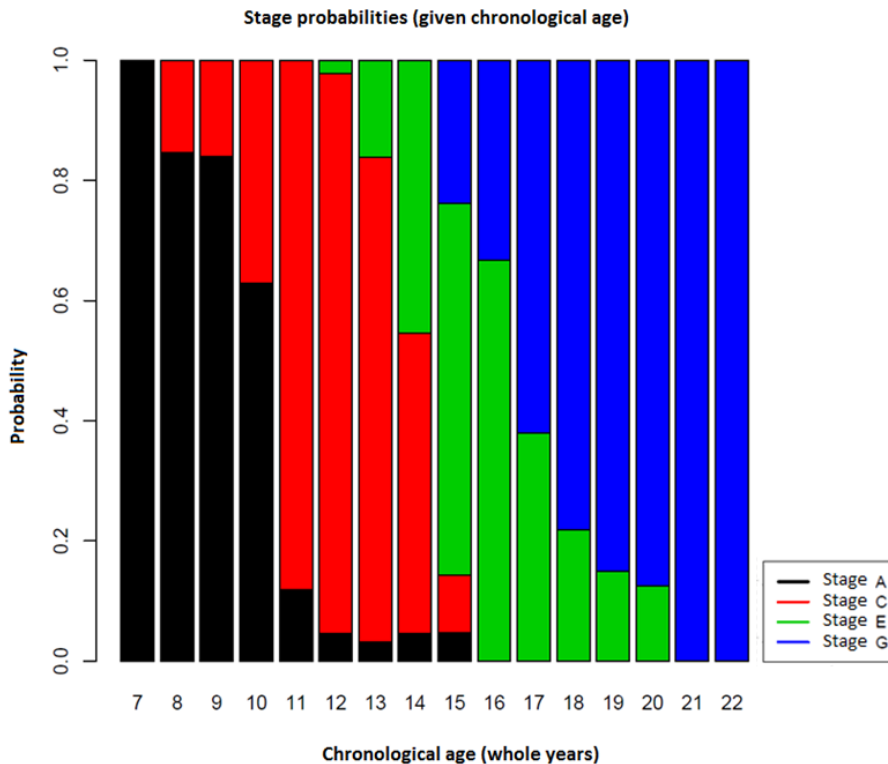
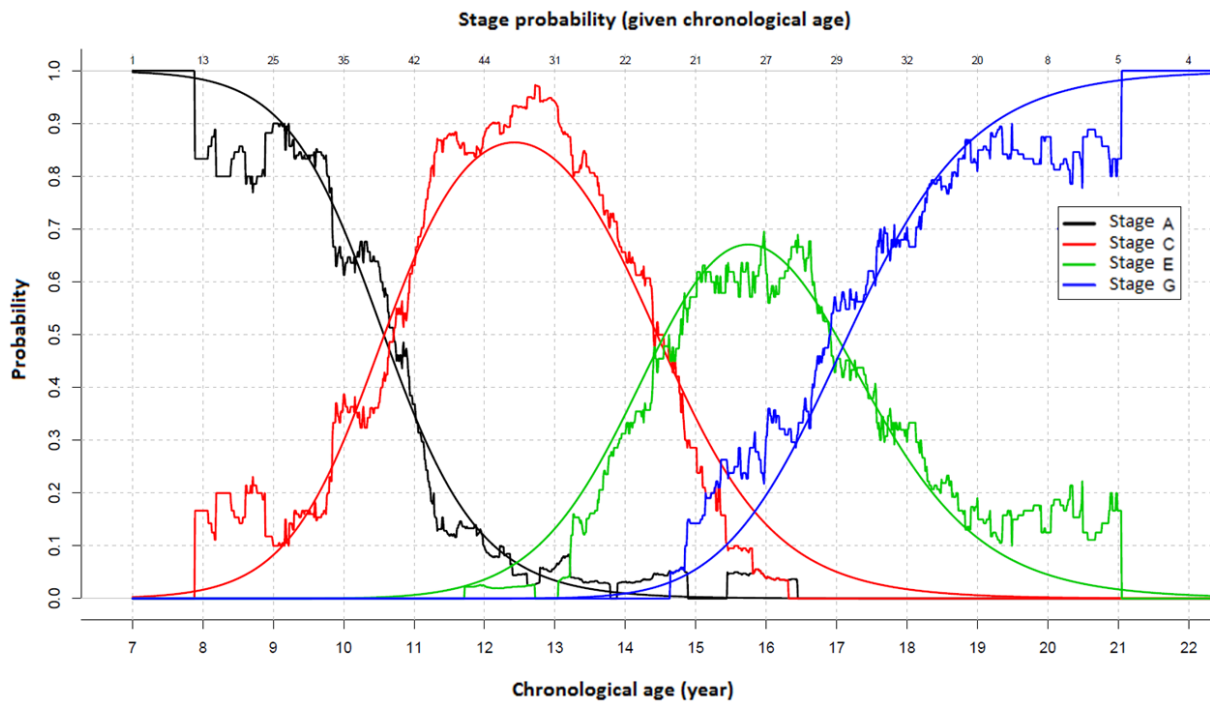


Figure 14. The figure shows the proportions of individuals (for given whole chronological ages) who fall into the different stages. The height of the column (of a given colour) indicates the proportion of individuals that fall into a given stage (black, red, green or blue). We can use these proportions to fit a probability model directly.

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 (28). This model will give the data a smooth functional fit. An important assumption for this type of model is that the stages are ordered: 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 % (given that the other stages are completed). In Figure 15 we see an example of such a model. Here we see that the probability of being in a stage is a “smoothed” function of age. 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 15, for example). For more information on the resulting models and model validation for the tool, see part A.2 of the Appendix.



**Figure 15.** The figure shows the proportions of individuals that fall into the different stages for a given chronological age (uneven curves). 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 give the number of individuals for the closest (rounded off) whole years. For example, for the age 16 on the x-axis, all individuals aged 15.5 to 16.5 years are considered. This applies to a total of 27 individuals. Here, 5% of these individuals fall into stages A and C, while 30% and 60% fall into stages E and G, respectively. The total of these percentages is 100%. The smooth curves are based on a fitted transition analysis model.

## 7.5 Distribution of chronological age given observed stage

(See also part A.7 of the Appendix)

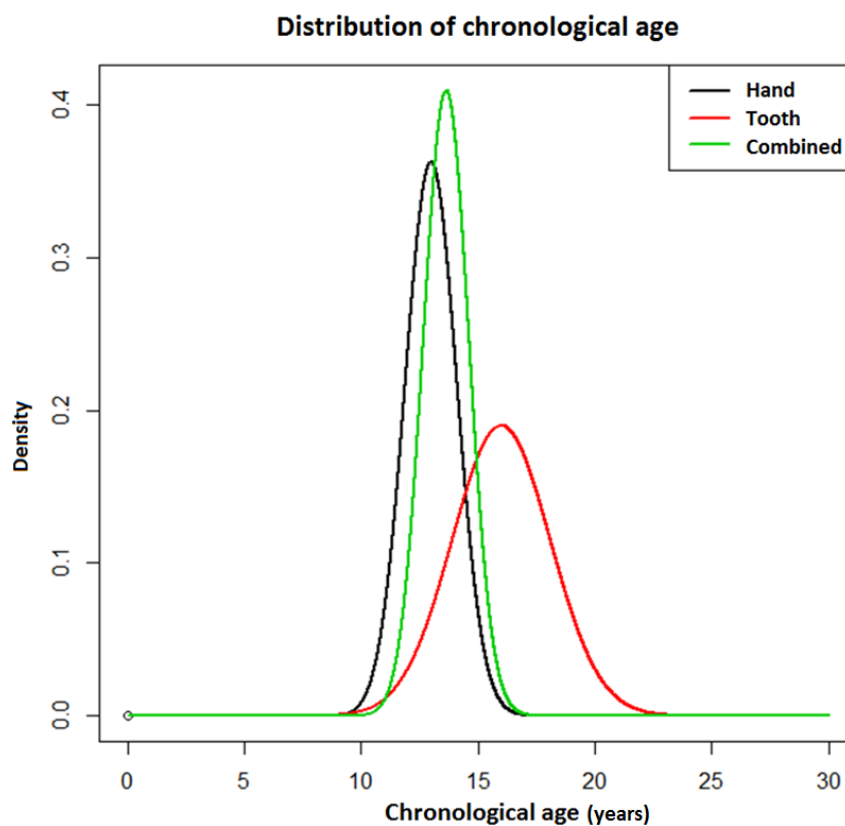
Until now we have only considered the description of the stage probability for given chronological ages. Our real objective, however, is to describe how chronological age is distributed for a given stage. In other words: if an individual has a GP stage hand skeleton and/or a Demirjian stage wisdom tooth,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

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

This enables us to produce the final 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 (prior).

## 7.6 Combination of hand and tooth

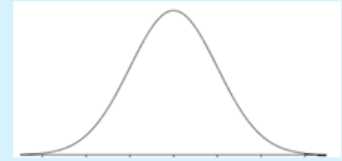
Because of the wide biological variation that is reflected by the methods, it is desirable to combine the hand and tooth stages of individuals in order to obtain a more precise estimate of chronological age. Gelbrich et al., 2015 (9) shows that there is no relationship between the age estimation errors 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. 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 (see Figure 16). This naturally presupposes that X-ray pictures for these two methods are taken at approximately the same time.



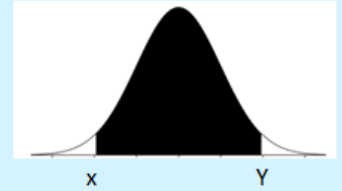
**Figure 16.** The figure shows how the distributions arrived at using the two methods are combined if they are assumed to be independent of one another for a given chronological age: For each given chronological age, the values of the functions for hand and tooth respectively are multiplied together, and then the multiplied function is normalised to have an area of one.

## The statistics underlying the results generated by the tool

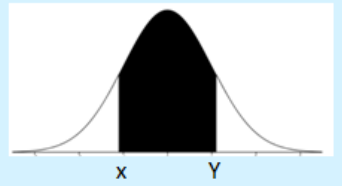
BioAlder results are based on distributions that describe the spread of chronological age for a given observed stage. For example, this may be a normal distribution where the most probable age is the mean (average). Other more probable values will be close to this mean, and less common values will be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



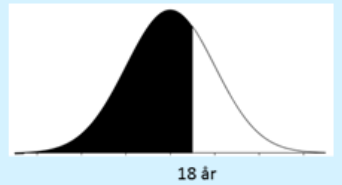
A 95% prediction interval  $[x, y]$  for age means that if a population that is representative of the population upon which the tool was developed were tested, we would expect the age of 95% of the individuals to fall within this interval (i.e. be between  $x$  and  $y$  years old). This interval is 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 17.** The statistics underlying the results generated by the tool.

## 8 Results used in BioAlder

### 8.1 Overview of studies used in BioAlder

The underlying hand data are based on the following studies:

Format	Reference	Boys	Girls	Country	Age range
Type 1	Santos 2011	136	94	Portugal	12-20
Type 1	Van Rijn 2001	178	197	Netherlands	9-20
Type 1	Zafar 2010	165	64	Pakistan	7-18
Type 1	Tise 2011	359	126	Italy	11-19
Type 3	Chaumoitre 2016	886	673	France	7-20
Type 4	Bala 2010	59	59	India	10-14
Type 4	Buken 2007	251	241	Turkey	11-19
Type 4	Cantekin 2012	259	351	Turkey	10-17
Type 4	Chiang 2005	141	70	Taiwan	10-17
Type 4	Griffith 2016	281	105	Kina	10-18
Type 4	Koc 2001	185	0	Turkey	10-17
Type 4	Mohammed 2015	270	270	India	10-18
Type 4	Nahid 2010	32	45	Iran	10-14
Type 4	Patel 2015	56	60	India	10-16
	<b>Total</b>	<b>3258</b>	<b>2355</b>		

**Table 7.** All hand studies included in BioAlder are listed above, with data format, number of included individuals, country and age range of the included population (27, 29-41).

The underlying wisdom tooth data are based on the following studies:

Format	Reference	Boys	Girls	Country	Age range
Type 1	Malta dataset	553	650	Malta	8-24
Type 1	South China dataset	682	617	China	8-24
Type 1	Swiss dataset	591	669	Switzerland	15-22
Type 1	Botswana dataset	768	915	Botswana	6-23
Type 2	Lee 2009	786	964	South Korea	7-24
Type 2	Johan 2012	540	539	Malaysia	14-25
Type 2	Duangto 2017	872	983	Thailand	8-23
Type 2	Li 2012	649	760	China	7-23
	<b>Total</b>	<b>5441</b>	<b>6097</b>		

**Table 8.** All wisdom tooth studies included in BioAlder are listed above, with data format, number of included individuals, country and age range of the included population (23-25, 42).

All references of format type 1 are individual data we have received from authors we have contacted. After contacting dozens of authors, we were given access to three datasets for hands and two 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 7. The individual-based datasets for teeth do not represent a single publication. We therefore call them “datasets” with appurtenant geographical area in Table 8.

## 8.2 Choice of upper age limit

(See also part B.5 of the Appendix)

When we use Bayes’ theorem to model age composition for a given stage, an age range/distribution of the individual points included in the model must be assumed in advance, in the same way as the age range/distribution of individuals to be included in a study must be pre-defined. In order to avoid age mimicry we let the prior age distribution to be uniform. This implies that, a priori, it is equal chance to be any of the ages as defined. The lowest chronological age of the age prior is set to seven, whereas the table below provides/presents an overview of the defined upper ages for the different methods and genders:

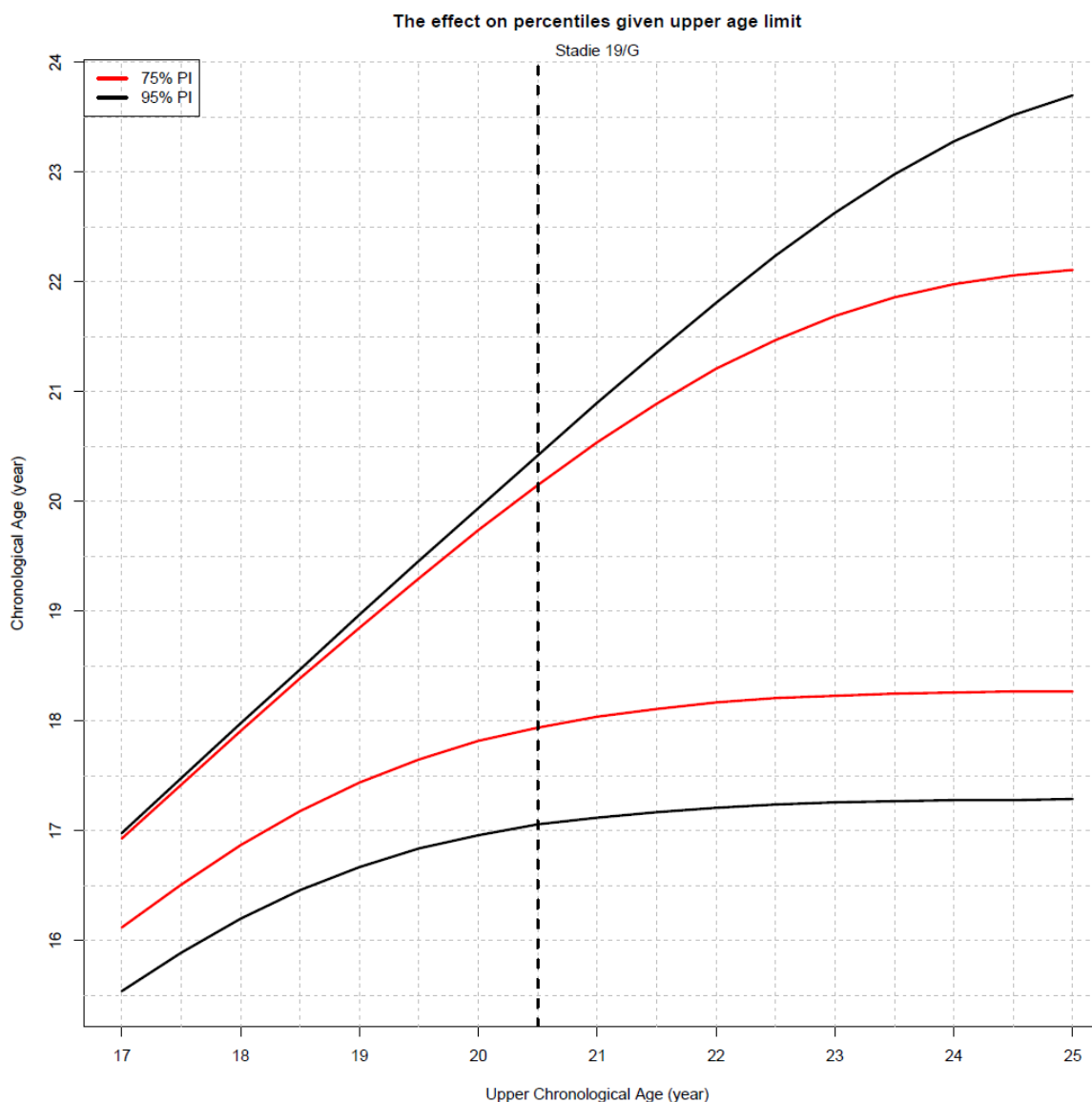
	Hand	Tooth	Combined
Boys	20 years	20.5 years	20.5 years
Girls	19 years	21 years	21 years

**Table 9.** The table provides an overview of the upper ages defined for the assumed age distribution for the different methods and genders.

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 important priority for the tool is to reduce the possibility that children are assessed as adults, i.e. the type I error.

The posterior age distribution for the last stage is partially defined by the upper age limit. See Figure 18 for an example of how the given upper age will affect the prediction intervals. Knell et al. (43). and Olze et al. (44) used the age defining the 50% probability of being in the last stage to bypass this difficulty, without describing any age variation. Roberts et al. (45) and Lee et al. (23) 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. (5)). 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 18.** 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.



## 9 The future of biological age estimation

### 9.1 Image-based methods

Biological variation and uncertainty regarding the significance of regional differences for age estimation by means of hand and tooth X-rays 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, rather than the initiation of new studies. We hope that the launch of BioAlder, attendance at conferences and international publications will promote such a collaboration going forward.

### 9.2 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 (46). Several prediction models for estimating chronological age have been developed (46-50), but none 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 also makes the method more ethically acceptable in both research and practical use than today's radiological methods. OUH is therefore working on a prediction model based on DNA methylation data for an adolescent population.

#### DNA methylation

- DNA methylation changes in pace with change in chronological age.
- Several prediction models for estimating chronological age have been developed, but 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.
- The method appears to have less variation than the image-based methods in use now, 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 19. DNA methylation.

## 10 References

1. NOAS, Redd Barna. Over eller under 18? : Aldersvurdering av enslige mindreårige asylsøkere. 2016.
2. European Asylum Support Office. EASO Age assessment practice in Europe. Luxemburg; 2013.
3. Dahlberg PS MA, 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-atlasen for aldersestimering: en systematisk oversikt. [www.fhi.no](http://www.fhi.no): Folkehelseinstituttet; 2017 03.2017. Contract No.: ISBN (elektronisk): 978-82-8082-822-4.
4. Rolseth V MA, 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. [www.fhi.no](http://www.fhi.no) Folkehelseinstituttet; 2017 03.2017.
5. Bleka Ø, Wisløff T, Dahlberg PS, Rolseth V, Egeland T. Advancing estimation of chronological age by utilizing available evidence based on two radiographical methods. . International Journal of Legal Medicine (in press).
6. Vandewalle S, Taes Y, Fiers T, Van Helvoirt M, Debode P, Herregods N, et al. Sex steroids in relation to sexual and skeletal maturation in obese male adolescents. J Clin Endocrinol Metab. 2014;99(8):2977-85.
7. Schmeling A, Dettmeyer R, Rudolf E, Vieth V, Geserick G. Forensic Age Estimation: Methods, Certainty, and the Law. Deutsches Ärzteblatt International. 2016;113(4):44-50.
8. Rudolf E, Kramer J, Gebauer A, Bednar A, Recsey Z, Zehetmayr J, 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):595-602.
9. 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. Ann Hum Biol. 2015;42(4):389-96.
10. Olze A, Bilang D, Schmidt S, Wernecke KD, Geserick G, Schmeling A. Validation of common classification systems for assessing the mineralization of third molars. Int J Legal Med. 2005;119(1):22-6.
11. Rättsmedicinalverket. [Available from: <https://www.rmv.se/verksamheter/medicinska-aldersbedomningar/>].
12. Saint-Martin P, Rerolle C, Pucheux J, Dedouit F, Telmon N. Contribution of distal femur MRI to the determination of the 18-year limit in forensic age estimation. Int J Legal Med. 2015;129(3):619-20.
13. Greulich W, Pyle S. Radiograph Atlas of Skeletal Development of the Hand and Wrist. 2nd ed. Stanford, California, USA1959.
14. Tanner J, Healy M, Goldstein H, Cameron N. Assessment of Skeletal Maturity and Prediction of Adult Height (TW3) Method. London, UK: WB Saunders, Harcourt Publishers Ltd; 2001.
15. Tanner J, Whitehouse R, Cameron N, Marshall W, Healy M, Goldstein H. Assessment of Skeletal Maturity and Prediction of Adult Height (TW2 Method). 2nd ed. London, UK: Academic Press; 1983.
16. Demirjian A, Goldstein H, Tanner JM. A new system of dental age assessment. Human Biology. 1973;45(2):211-27.
17. Hunt EE, Jr., Gleiser I. The estimation of age and sex of preadolescent children from bones and teeth. Am J Phys Anthropol. 1955;13(3):479-87.
18. Moorrees CF, Fanning EA, Hunt EE, Jr. AGE VARIATION OF FORMATION STAGES FOR TEN PERMANENT TEETH. J Dent Res. 1963;42:1490-502.
19. Haavikko K. The formation and the alveolar and clinical eruption of the permanent teeth. An orthopantomographic study. Suomen Hammaslaakariseuran Toimituksia. 1970;66(3):103-70.

20. Kullman L, Johanson G, Akesson L. Root development of the lower third molar and its relation to chronological age. *Swed Dent J*. 1992;16(4):161-7.
21. Köhler S, Schmelzke R, Loitz C, Püschel K. Die entwicklung des weisheitszahnes als kriterium der lebensaltersbestimmung. *Annals of Anatomy - Anatomischer Anzeiger*. 1994;176(4):339-45.
22. Thodberg HH, Savendahl L. Validation and reference values of automated bone age determination for four ethnicities. *Acad Radiol*. 2010;17(11):1425-32.
23. Lee SH, Lee JY, Park HK, Kim YK. Development of third molars in Korean juveniles and adolescents. *Forensic Science International*. 2009;188(1):107-11.
24. Li G, Ren J, Zhao S, Liu Y, Li N, Wu W, et al. Dental age estimation from the developmental stage of the third molars in western Chinese population. *Forensic Science International*. 2012;219(1):158-64.
25. Johan NA, Khamis MF, Abdul Jamal NS, Ahmad B, Mahanani ES. The variability of lower third molar development in Northeast Malaysian population with application to age estimation. *Journal of Forensic Odonto-Stomatology*. 2012;30(1):45-54.
26. Cavric J, Vodanovic M, Marusic A, Galic I. Time of mineralization of permanent teeth in children and adolescents in Gaborone, Botswana. *Annals of anatomy = Anatomischer Anzeiger : official organ of the Anatomische Gesellschaft*. 2016;203:24-32.
27. 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.
28. 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.
29. Santos C, Ferreira M, Alves FC, Cunha E. Comparative study of Greulich and Pyle Atlas and Maturus 4.0 program for age estimation in a Portuguese sample. *Forensic Sci Int*. 2011;212(1-3):276.e1-7.
30. van Rijn RR, Lequin MH, Robben SG, Hop WC, van Kuijk C. Is the Greulich and Pyle atlas still valid for Dutch Caucasian children today? *Pediatric Radiology*. 2001;31(10):748-52.
31. Zafar AM, Nadeem N, Husen Y, Ahmad MN. An appraisal of Greulich-Pyle Atlas for skeletal age assessment in Pakistan. *JPMA - Journal of the Pakistan Medical Association*. 2010;60(7):552-5.
32. Chaumoitre K, Saliba-Serre B, Adalian P, Signoli M, Leonetti G, Panuel M. Forensic use of the Greulich and Pyle atlas: prediction intervals and relevance. *European Radiology*. 2017;27(3):1032-43.
33. Tise M, Mazzarini L, Fabrizzi G, Ferrante L, Giorgetti R, Tagliabracci A. Applicability of Greulich and Pyle method for age assessment in forensic practice on an Italian sample. *International Journal of Legal Medicine*. 2011;125(3):411-6.
34. Bala M, Pathak A, Jain RL. 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):95-9.
35. Cantekin K, Celikoglu M, Miloglu O, Dane A, Erdem A. Bone age assessment: the applicability of the Greulich-Pyle method in eastern Turkish children. *Journal of forensic sciences*. 2012;57(3):679-82.
36. Chiang KH, Chou ASB, Yen PS, Ling CM, Lin CC, Lee CC, et al. The reliability of using Greulich-Pyle method to determine children's bone age in Taiwan. *Tzu Chi Medical Journal*. 2005;17(6):417-20+53.
37. Griffith JF, Cheng JCY, Wong E. 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):28-32.
38. Koc A, Karaoglanoglu M, Erdogan M, Kosecik M, Cesur Y. Assessment of bone ages: is the Greulich-Pyle method sufficient for Turkish boys? *Pediatrics International*. 2001;43(6):662-5.

39. Mohammed RB, Rao DS, Goud AS, Sailaja S, Thetay AA, Gopalakrishnan M. 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):218-25.
40. Nahid G, Abdorrahim A, Gharib SM, Anvar E. Assessment of bone age in Kurdish children in IRAN. *Pakistan Journal of Medical Sciences*. 2010;26(1):36-9.
41. Patel PS, Chaudhary AR, Dudhia BB, Bhatia PV, Soni NC, Jani YV. 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):18-27.
42. Duangto P, Iamaroon A, Prasitwattanaseree S, Mahakkanukrauh P, Janhom A. 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):559-68.
43. Knell B, Ruhstaller P, Prieels F, Schmeling A. Dental age diagnostics by means of radiographical evaluation of the growth stages of lower wisdom teeth. *Int J Legal Med*. 2009;123(6):465-9.
44. Olze A, Pynn BR, Kraul V, Schulz R, Heinecke A, Pfeiffer H, et al. Studies on the chronology of third molar mineralization in First Nations people of Canada. *International Journal of Legal Medicine*. 2010;124(5):433-7.
45. Roberts GJ, McDonald F, Andiappan M, Lucas VS. 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;36:177-84.
46. Horvath S. DNA methylation age of human tissues and cell types. *Genome Biol*. 2013;14(10):R115.
47. Bekaert B, Kamalandua A, Zapico SC, Van de Voorde W, Decorte R. Improved age determination of blood and teeth samples using a selected set of DNA methylation markers. *Epigenetics*. 2015;10(10):922-30.
48. Huang Y, Yan J, Hou J, Fu X, Li L, Hou Y. Developing a DNA methylation assay for human age prediction in blood and bloodstain. *Forensic Sci Int Genet*. 2015;17:129-36.
49. Yi SH, Xu LC, Mei K, Yang RZ, Huang DX. Isolation and identification of age-related DNA methylation markers for forensic age-prediction. *Forensic Sci Int Genet*. 2014;11:117-25.
50. Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sada S, et al. Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol Cell*. 2013;49(2):359-67.



# Appendix to BioAlder Manual Version 1.1

## Appendix to Manual

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

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

## Contents

A.	Statistical modelling .....	3
A.1	Purpose.....	3
A.2	Modelling stage probabilities .....	4
A.3	Combining studies .....	7
A.4	Modelling of individual data.....	8
A.4.1	Modelling of tooth data (Type 2) .....	9
A.4.2	Modelling of hand data (Types 3 and 4).....	9
A.5	Model chosen for the stage probabilities in the tool.....	12
A.5.1	Overview of candidate models.....	13
A.5.2	The likelihood function.....	13
A.5.3	Model fitting.....	14
A.5.4	Details of each individual transition model.....	14
A.6	Model validation.....	15
A.7	Distribution of chronological age given observed stage .....	15
A.7.1	How to work backward to the age distribution .....	15
A.7.2	Calculated results based on age distribution .....	16
A.7.3	Effect of assumed age distribution.....	17
A.7.4	Choice of results as a consequence of the fact that the data are generated .....	18
A.7.5	Details of calculating the results used by the tool .....	20
A.8	Modelling of stage probability for hand and tooth combined.....	20
B.	Results used in the tool .....	22
B.1	Overview of studies used in the tool.....	22
B.2	Overview of models used in the tool .....	23
B.3	Randomly generated variation for stage probabilities.....	23
B.3.1	Overview figures.....	23
B.3.2	Model validation.....	31
B.4	Credibility interval for stage probabilities.....	32
B.4.1	Credibility interval for non-parametric model .....	32
B.4.2	Credibility interval for parametric model.....	32
B.4.3	Overview figures.....	32
B.5	Choice of upper age limits in the tool (defining the prior age distribution) .....	34
B.5.1	The effect of different upper age limits .....	34
B.5.2	Final choices for upper age limits in the tool .....	35
B.5.3	Overview figures of the effect of assumed upper age .....	36
C.	References.....	44

## A. Statistical modelling

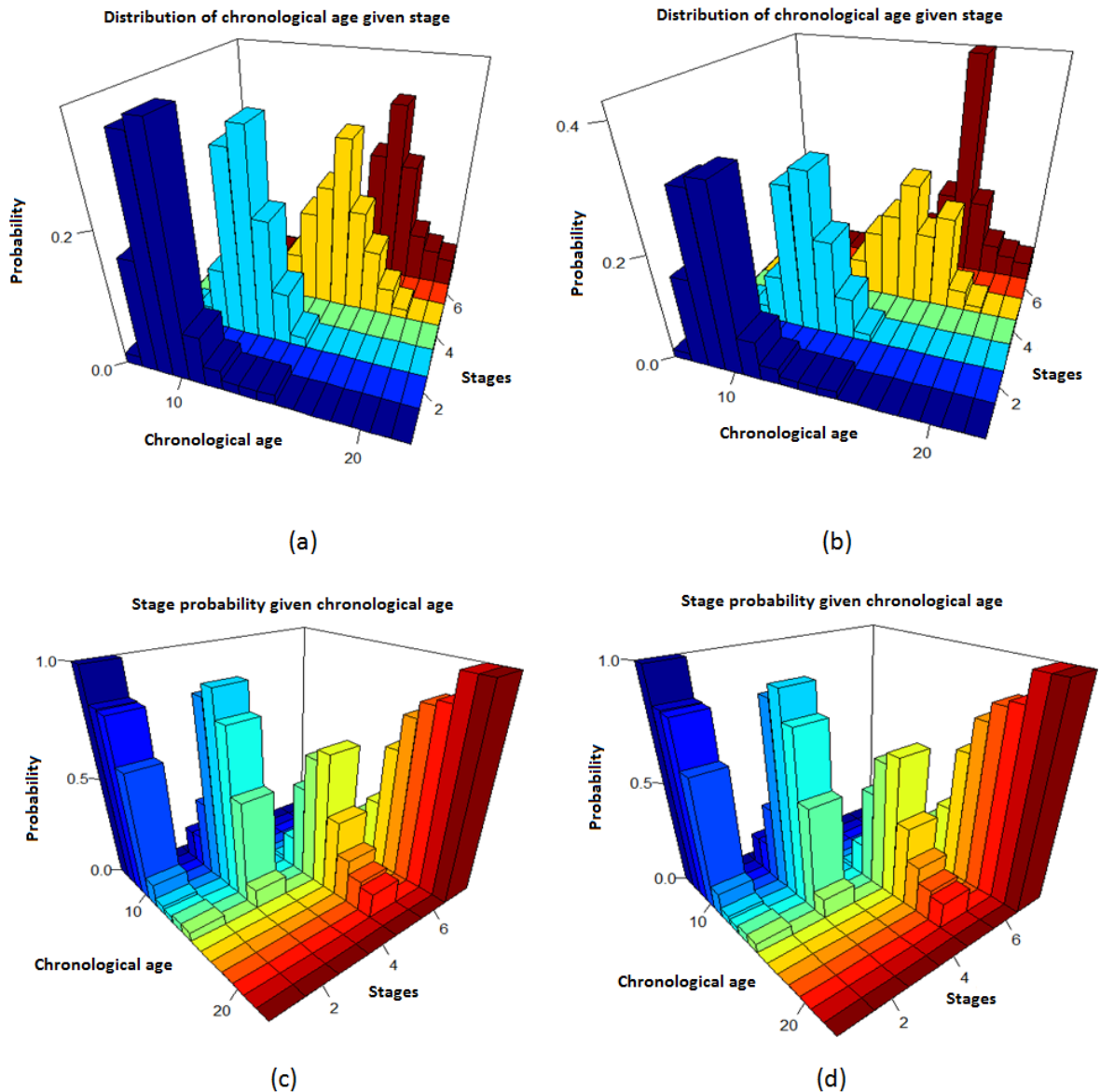
### A.1 Purpose

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 X-ray picture 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 A1** 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.





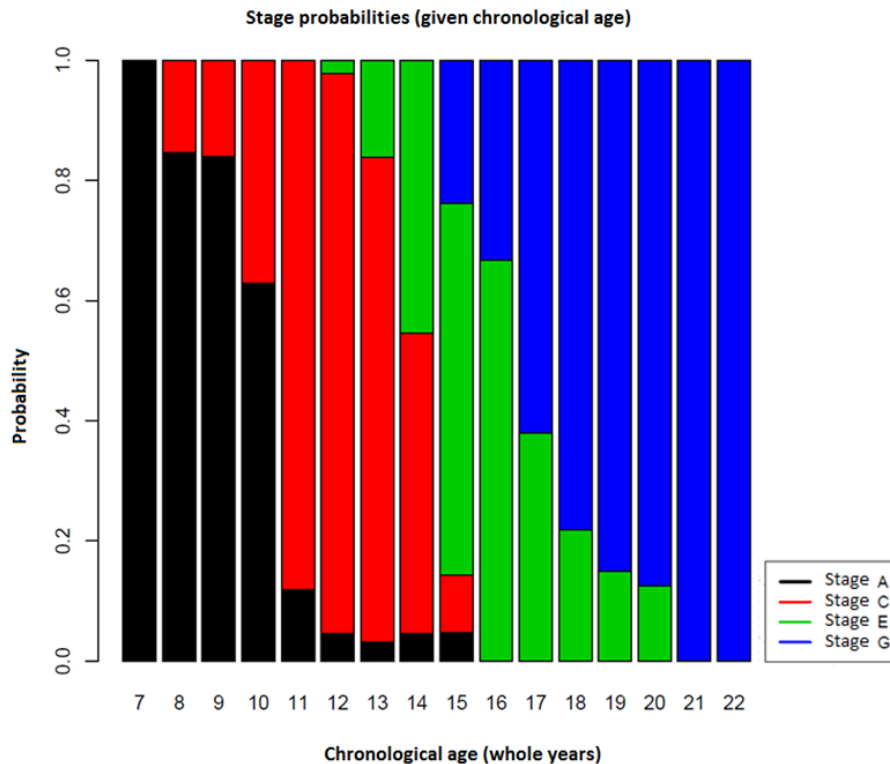
**Figure A1:** The figure shows a 3D histogram that illustrates what happens when we include thirty extra 18 year-olds for two types of approach. The one horizontal axis represents the variable 'Chronological age' and the other the category variable 'Stage'. Figures (a) and (b) show the distribution 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 A2** shows an

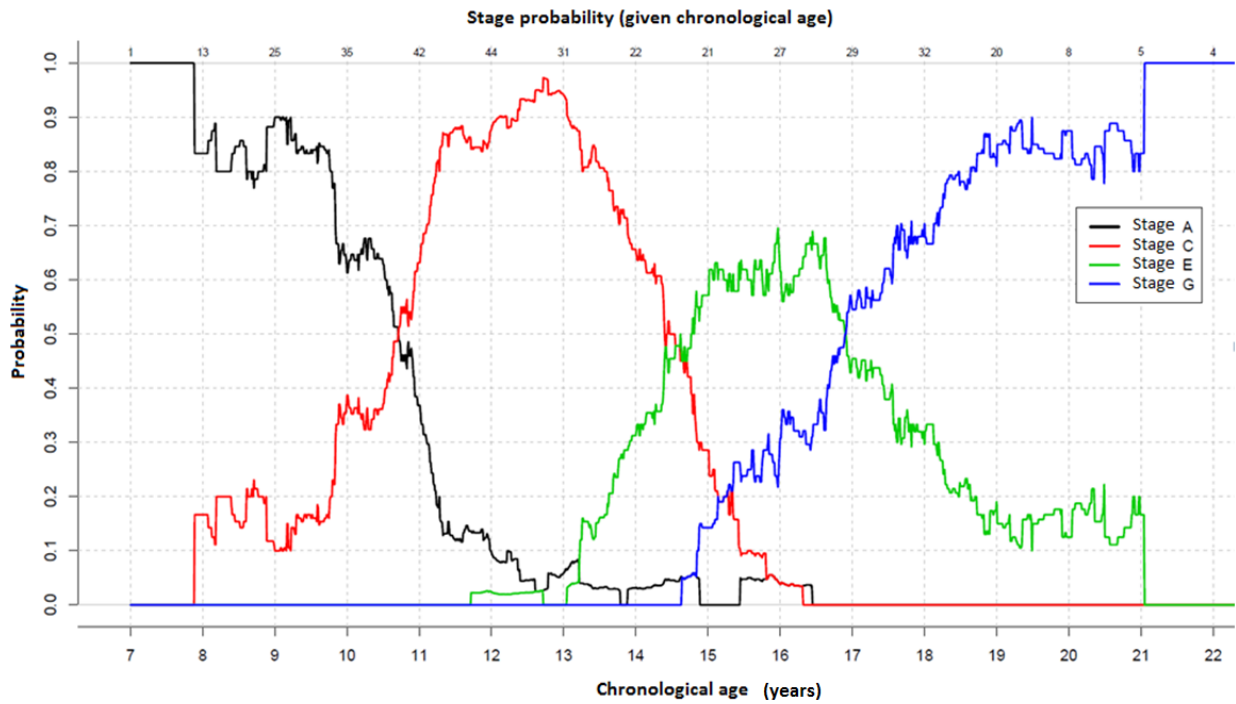
example where stage probabilities are estimated as the proportions of individuals that fall into four different stages. We call this type of model a non-parametric model.



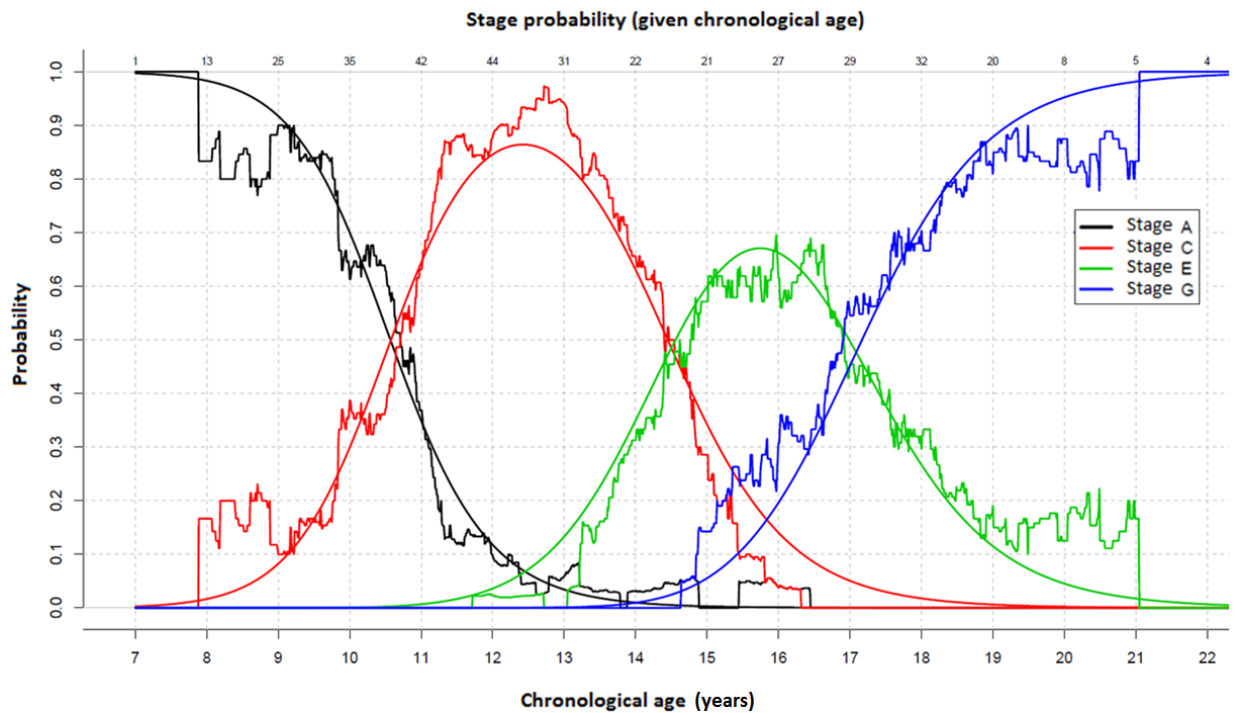
**Figure A2:** The figure shows the proportions of individuals (for given whole chronological ages) who fall into the different stages. The height of the column (of a given colour) indicates the proportion of individuals that fall into a given stage (black, red, green or blue). We can use these proportions 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 A3**. 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 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 (given that the other stages are completed). In **Figure A4** 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.



**Figure A3:** 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 percentages is 100%.



**Figure A4:** The figure shows the directly observed proportions (uneven curves) together with the stage probabilities from a fitted transition model (smooth curves).

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 which of these two model variants is best suited to 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

### Data formats included in our tool

**Type 1** consists of data in individual-based format.

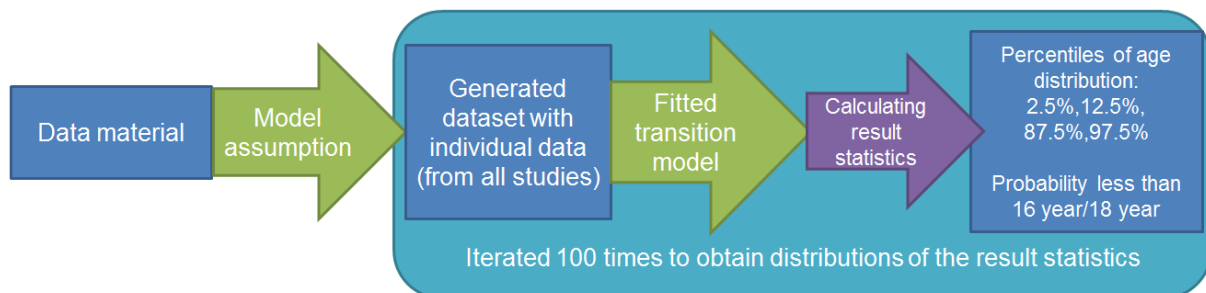
**Type 2** is a frequency table with number of individuals for each stage within each whole year.

**Type 3** is tables with data on means and standard deviations of chronological age for given stages (skeletal age or tooth formation stage).

**Type 4** is tables with data on means and standard deviations of chronological age and skeletal age within each whole year.

**Table A1:** The figure is from the manual, and provides an overview of the different types of data format used by various relevant studies.

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 the great majority 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 A5:** 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 A5**). 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 tooth data (Type 2)

As described in the manual, there are only two datasets that provide complete tooth data for individuals (Type 1). We received these data from other research communities, and have consent to use them. Four of the studies (Johan 2002 (2), Lee 2009 (3), Li 2012 (4) and Duangto 2017 (5)) 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.

### A.4.2 Modelling of hand data (Types 3 and 4)

As stated in the manual, the formats for hand data are of three different types: Individual-based (Type 1), 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, we have not observed what the actual skeletal and chronological ages per individual are. 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 A2: 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 A2**, 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).

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	17.46	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

**Table A2:** 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).

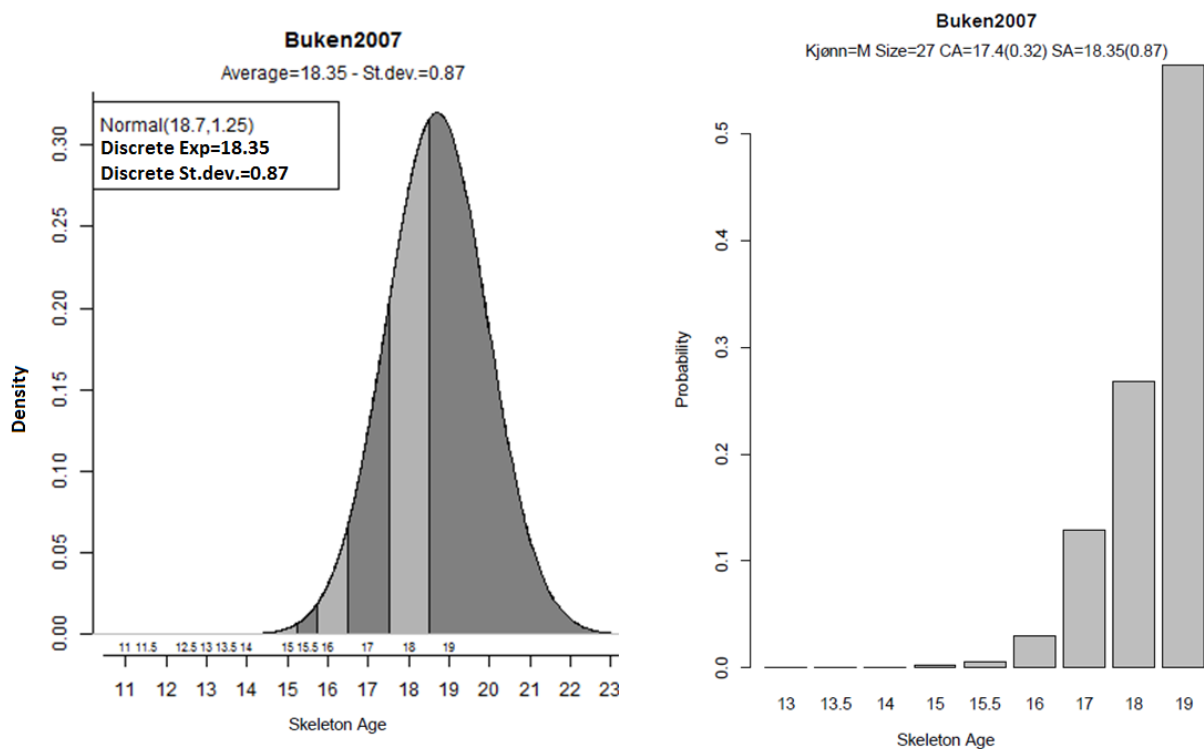
#### *Step 1: Model assumption for skeletal age:*

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 A3 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 A6** for an illustration in which row 3 of **Table A2** is considered).

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	

**Table A3:** The table shows the defined skeletal ages (from 5 years) from the Greulich & Pyle Atlas for both genders.



**Figure A6:** 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 (6) (age 17 years).

#### A.4.2.1 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  $\mu$  and standard deviation  $\sigma$ .

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

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 A3**). 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  $\mu$  and  $\sigma$  such that

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

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

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 A2**).

In technical terms, the choice of  $\mu$  and  $\sigma$  is made by minimising the function

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

### 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 A2**).  $Corr$  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 here 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 A7**.



**Steps 1–2** are carried out for each row in the table (see **Table A2**) 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.

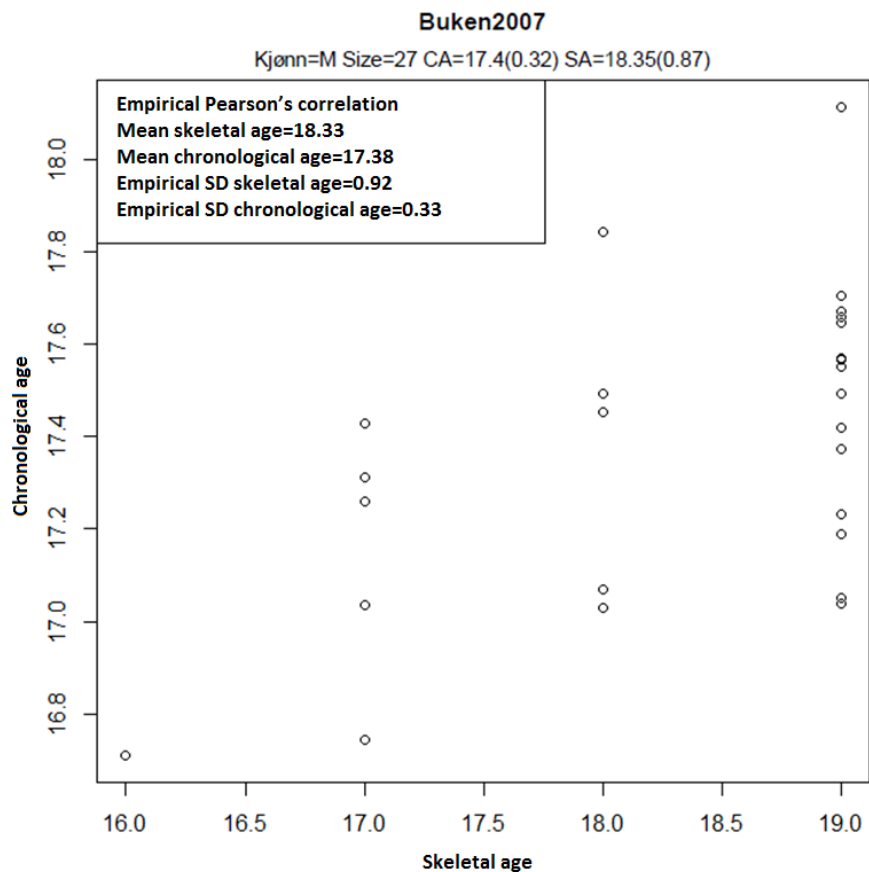


Figure A7: Example of generating 27 random individuals for the age group boys 17 years old for Buken 2007.

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

Boldsen et al. (2001) (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 (7), i.e. that we have predefined an age transformation. In our approach we allow the data decide the transformation of chronological age,  $g(\text{age}) = \text{age}^\lambda$  for  $\lambda = 0.1, 0.2, \dots, 1.0$ . 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  $j = 1, \dots, J$ , chronological age  $x$  and the parameters  $\theta = (\alpha_1, \dots, \alpha_{j-1}, \beta, \lambda)$ . We assume that  $\lambda$  takes the values 0.1, 0.2, ..., 1.0. By letting  $Y$  be a discrete stochastic variable with the stage outcomes  $1, \dots, J$  that an individual of age  $x$  may be in, we can describe the candidate models (1–4) as follows for the stages  $j = 1, \dots, J - 1$ :

- 1) Proportional-odds cumulative model with logit link
  - a.  $\text{logit}(P(Y \leq j|\theta, x)) = \alpha_j + \beta * x^\lambda$
- 2) Proportional-odds cumulative model with probit link
  - a.  $\text{probit}(P(Y \leq j|\theta, x)) = \alpha_j + \beta * x^\lambda$
- 3) Continuation-ratio model with logit link
  - a.  $\text{logit}(P(Y = j|Y \geq j, \theta, x)) = \alpha_j + \beta * x^\lambda$
- 4) Continuation-ratio model with probit link
  - a.  $\text{probit}(P(Y = j|Y \geq j, \theta, x)) = \alpha_j + \beta * x^\lambda$

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

The link function  $\text{logit}(x) = \log\left(\frac{x}{1-x}\right)$ , 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  $\theta$ . 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)

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

where  $n_j$  is the number of individuals in stage  $j$  and  $x_{i,j}$  is the chronological age of the individual  $i$  observed with 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 (8) 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  $\theta$  (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  $f(P(Y \leq j|\theta, x))$  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

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

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

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

#### A.5.4.2 Continuation-ratio

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

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

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

## 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 A4**). 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 is carried out for all stages of the hand and tooth methods for both genders.

## A.7 Distribution of chronological age given observed stage

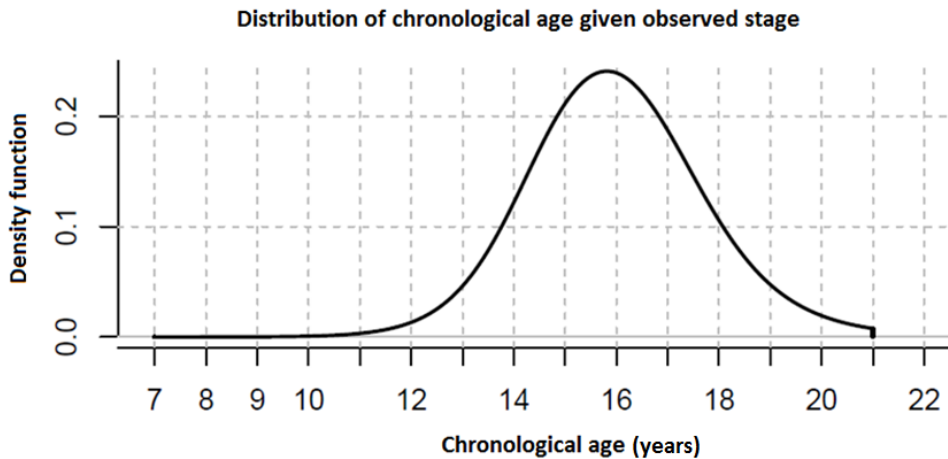
### A.7.1 How to work backward to the age distribution

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 A8** 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 A8:** The figure shows age distribution for a given stage based on Bayes' theorem, with the age distribution defined as from 7 to 21 years.

#### A.7.1.1 Bayes' theorem explained mathematically

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.

### A.7.2.1 Formulas for calculating results statistics

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

$$P(\text{Age} = a | \text{Stage} = j) = P(\text{Stage} = j | \text{Age} = a) * C(u)$$

Is calculated for  $a = 7.00, 7.01, \dots, u - 0.01, u$  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:

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

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:

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

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

Overview of calculated result statistics:

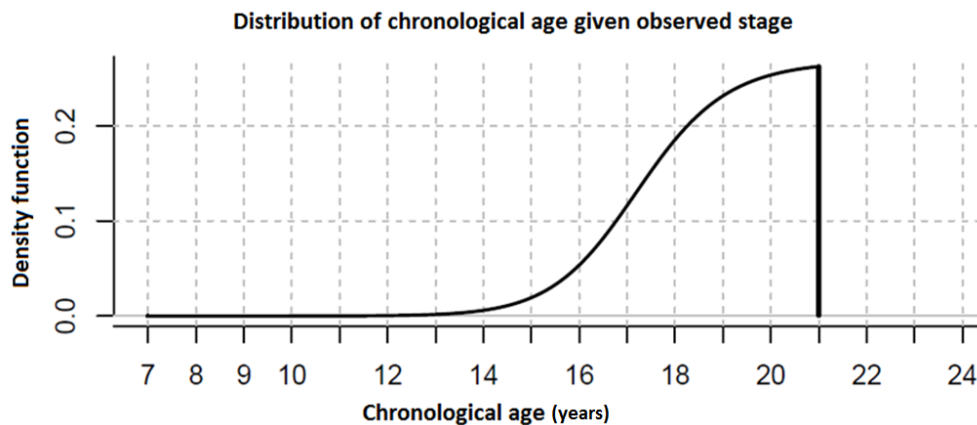
- 1) The probability of age less than  $T$  years will then be  $P(\text{Age} \leq T | \text{Stage} = j)$
- 2)  $q$ -percentile =  $\text{argmax}_a P(\text{Age} \leq a | \text{Stage} = j) \leq q$

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

### 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 A9** 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

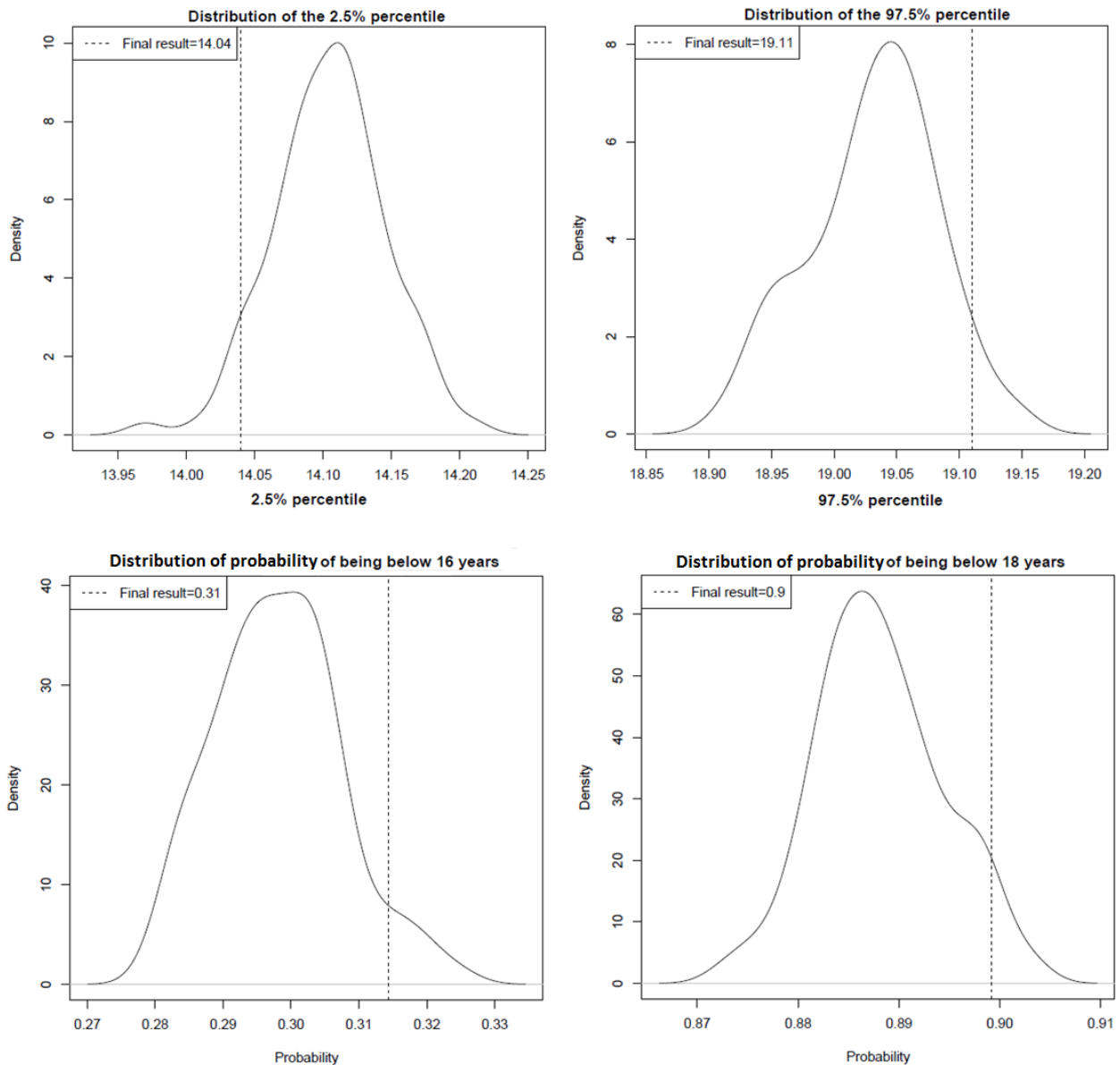
the age distribution must always be one, both the percentiles and the probabilities of being less than 16/17/18 years will be influenced by the upper defined age: The lower the upper age limit that is defined, the lower the age distribution percentiles will be, and the higher the probabilities of being 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.



**Figure A9:** 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.

#### 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 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 A10** for an illustration of these distributions.



**Figure A10:** 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 A10**). For all the other results statistics (87.5% and 97.5% and the probabilities for under 16 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 A10**). 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.

### 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

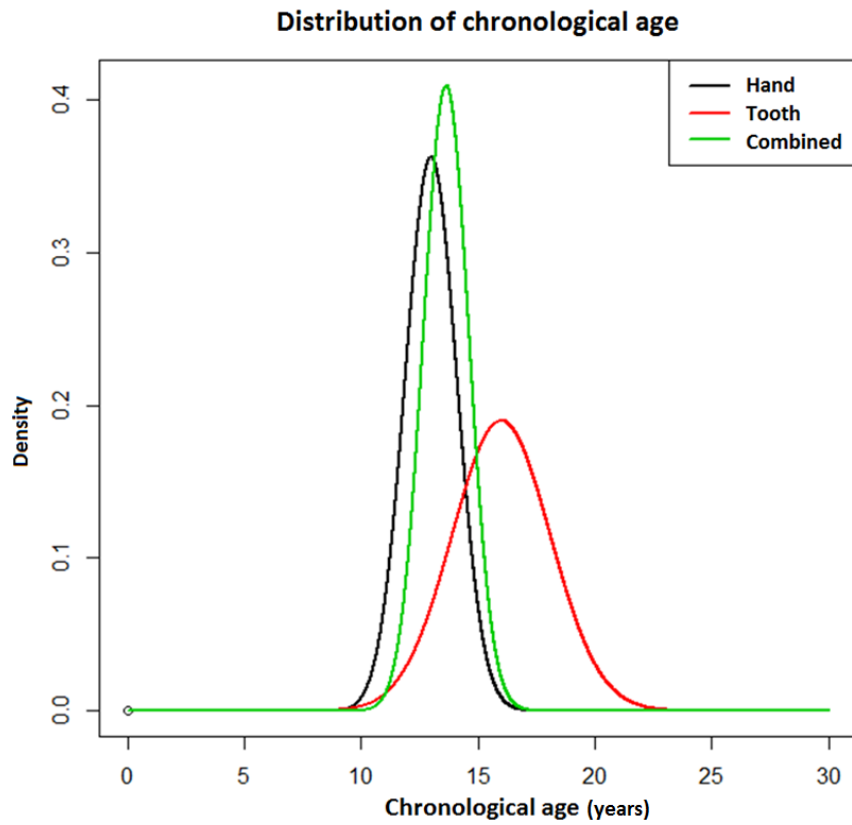
Because of the great biological variability affecting the methods, it is desirable to combine several methods in order to obtain a more precise estimate. Gelbrich et al. (2015) (9) point out that there is no relationship between the age estimation errors arrived at by means of the hand and tooth of the same individual, and we can therefore assume that the two methods are independent for a given chronological age. This naturally presupposes that the X-ray pictures 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

$$\begin{aligned} P(\text{skeletal age} = s, \text{tooth stage} = t | \text{Age} = a) \\ = P(\text{skeletal age} = s | \text{Age} = a) * P(\text{tooth stage} = t | \text{Age} = a) \end{aligned}$$

Applying Bayes’ theorem with defined a priori age  $P(\text{Age} = a)$  we get

$$\begin{aligned} P(\text{Age} = a | \text{skeletal age} = s, \text{tooth stage} = t) \propto \\ P(\text{skeletal age} = s | \text{Age} = a) * P(\text{tooth stage} = t | \text{Age} = a) * P(\text{Age} = a). \end{aligned}$$

Figure A11 illustrates the distribution of chronological age, given the data for an observed skeletal age and tooth stage (combined).



**Figure A11:** 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.

## 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	Santos 2011	136	94
Type 1	Van Rijn 2001	178	197
Type 1	Zafar 2010	165	64
Type 1	Tise 2011	359	126
Type 3	Chaumoitre 2016	886	673
Type 4	Bala 2010	59	59
Type 4	Buken 2007	251	241
Type 4	Cantekin 2012	259	351
Type 4	Chiang 2005	141	70
Type 4	Griffith 2016	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	3258	2355

#### Tooth studies

Format	Study	Boys	Girls
Type 1	Malta dataset	553	650
Type 1	South China dataset	682	617
Type 1	Swiss dataset	591	669
Type 1	Botswana dataset	768	915
Type 2	Lee 2009	786	964
Type 2	Johan 2012	540	539
Type 2	Duangto 2017	872	983
Type 2	Li 2012	649	760
All	Total	5441	6097

**Table B1:** The tables provide an overview of the numbers of individuals in different studies upon which the results in the tool are based. 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 B1** 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 3258 and 2355 for hands, and 5441 and 6097 for teeth. We see from the table that the Chaumoitre 2015 study contributes most for hands, and some studies contribute very little (Bala 2010, Nahid 2010 and Patel 2015), while all the studies contribute quite similarly for teeth.

## B.2 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	$Age^{0.7}$	Continuation-ratio	probit
Tooth	Girls	$Age^{0.7}$	Continuation-ratio	probit

**Table B2:** 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 B2** provides an overview of the selected parametric models upon which the results in the tool are based. We found here 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.3 Randomly generated variation for stage probabilities

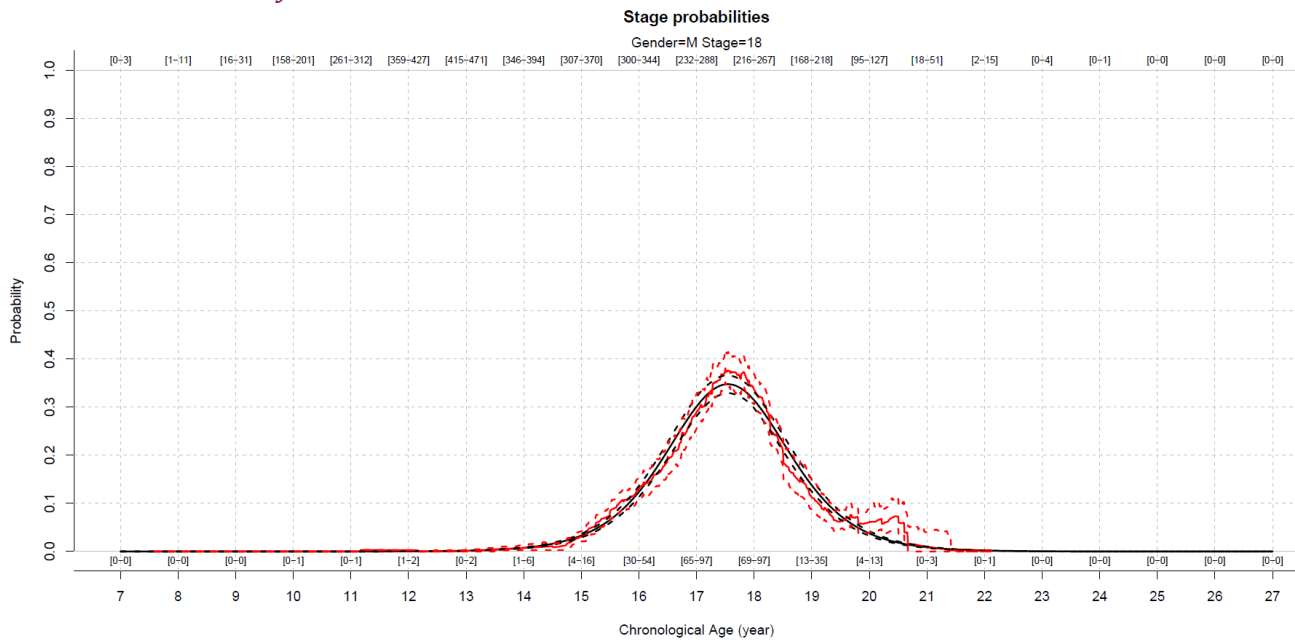
In this section we show the effect of randomly generated variation for only the second last and last stages for all methods and both genders. This is useful for describing the variability of the fitted stage probabilities in view of the fact that we do not know the individual-based data.

### B.3.1 Overview figures

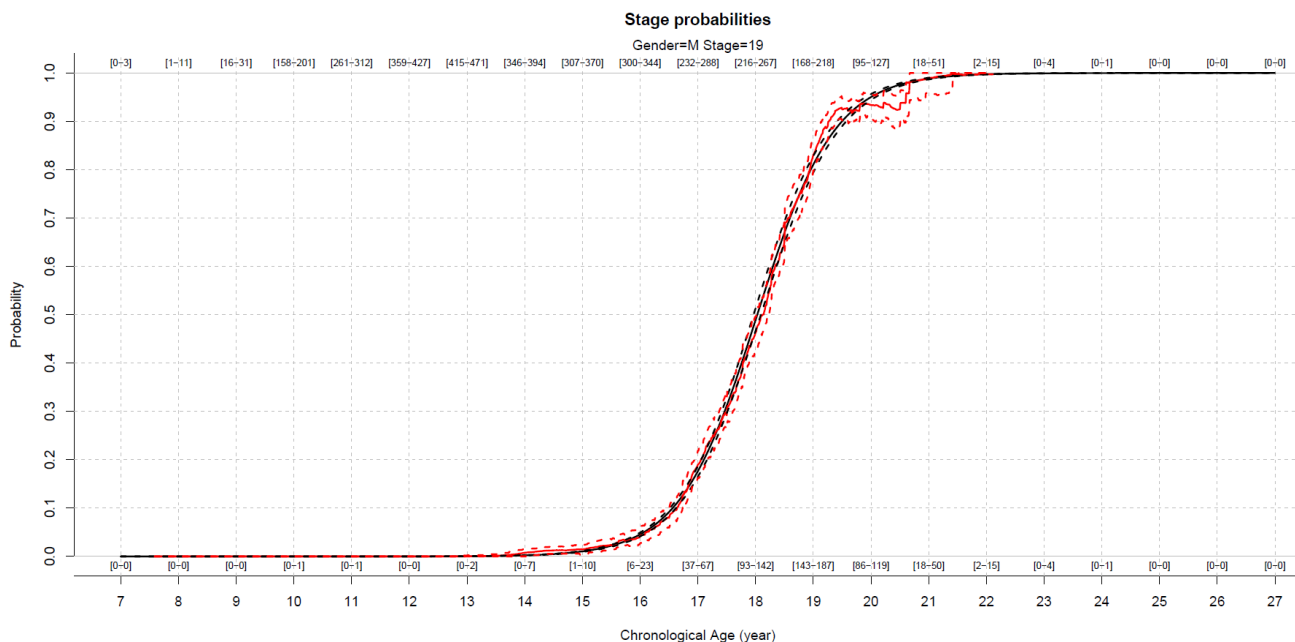
The following figures (**Figures B1–B16**) in sections **B.3.1.1** and **B.3.1.2** (for boys) and **B.3.1.4** and **B.3.1.5** (for girls) show the mean (solid line), 5% (lower stippled) and 95% (upper stippled) quantiles for the fitted parametric regression models of stage probabilities (black) and the non-parametric probabilities (red) across the 100 generated datasets. The fitted regression models are as listed in **Table B2**. 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 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.3.1.1 For boys hand

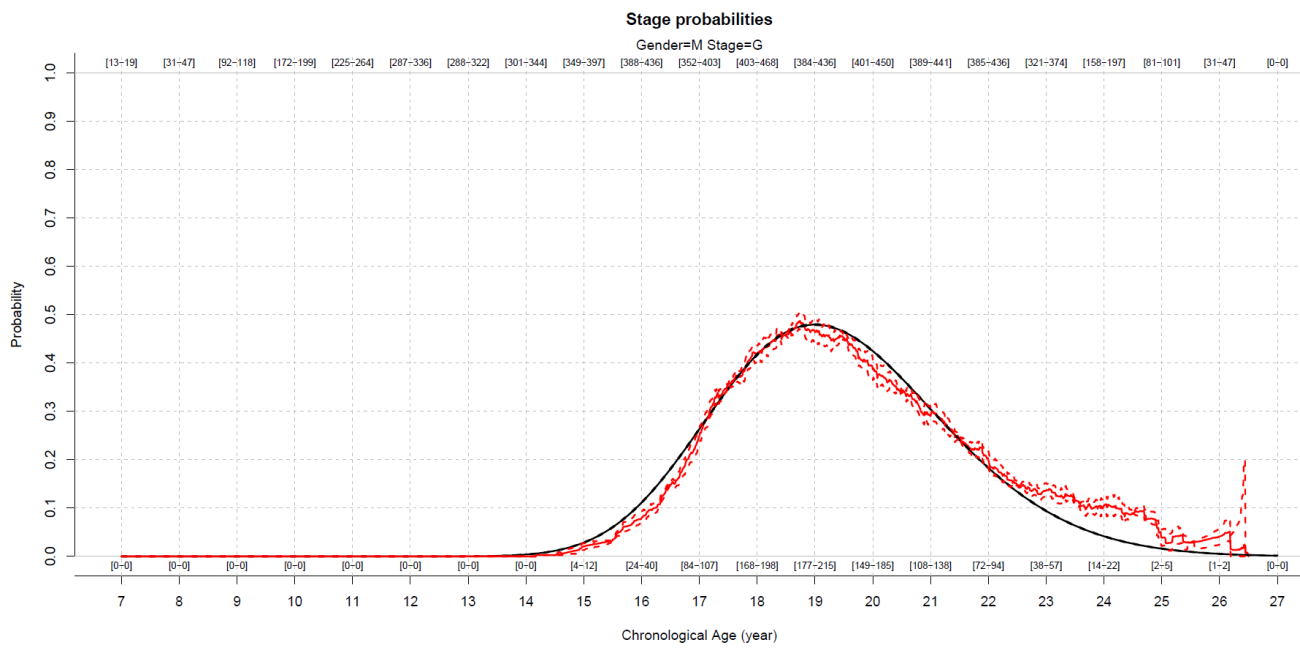


**Figure B1:** The figure shows the stage probability for boys with a skeletal age of 18 years for given chronological ages.

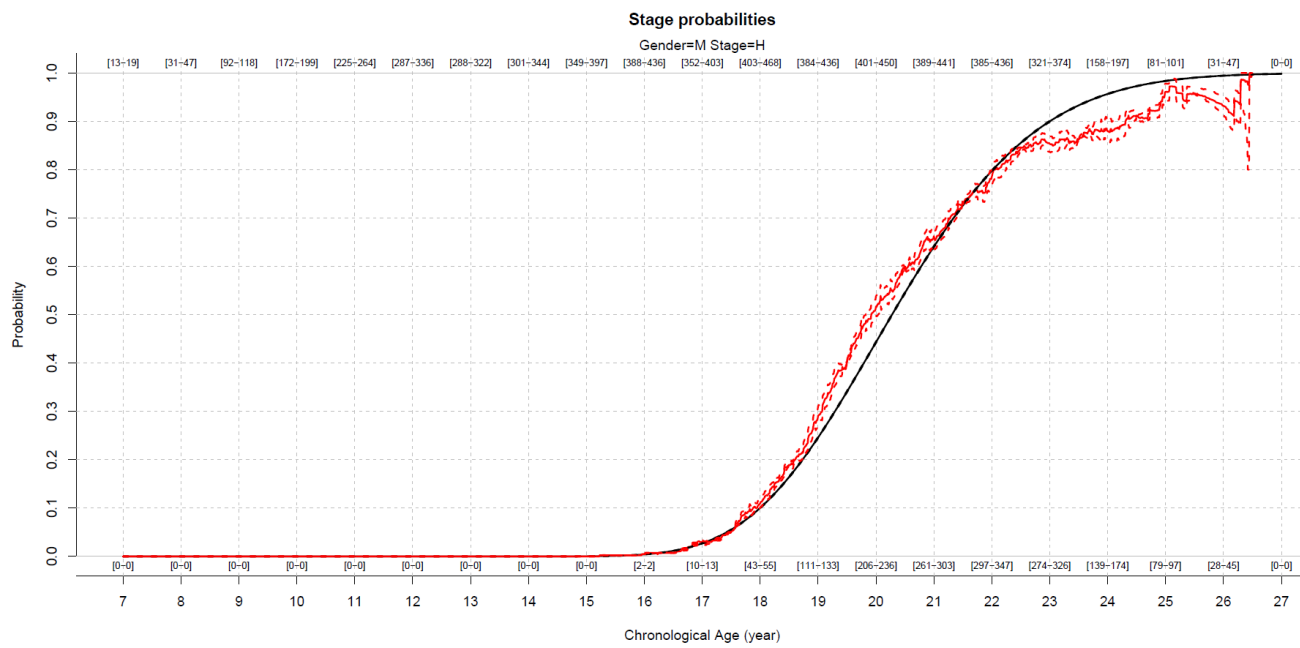


**Figure B2:** The figure shows the stage probability for boys with a skeletal age of 19 years for given chronological ages.

### B.3.1.2 For boys tooth

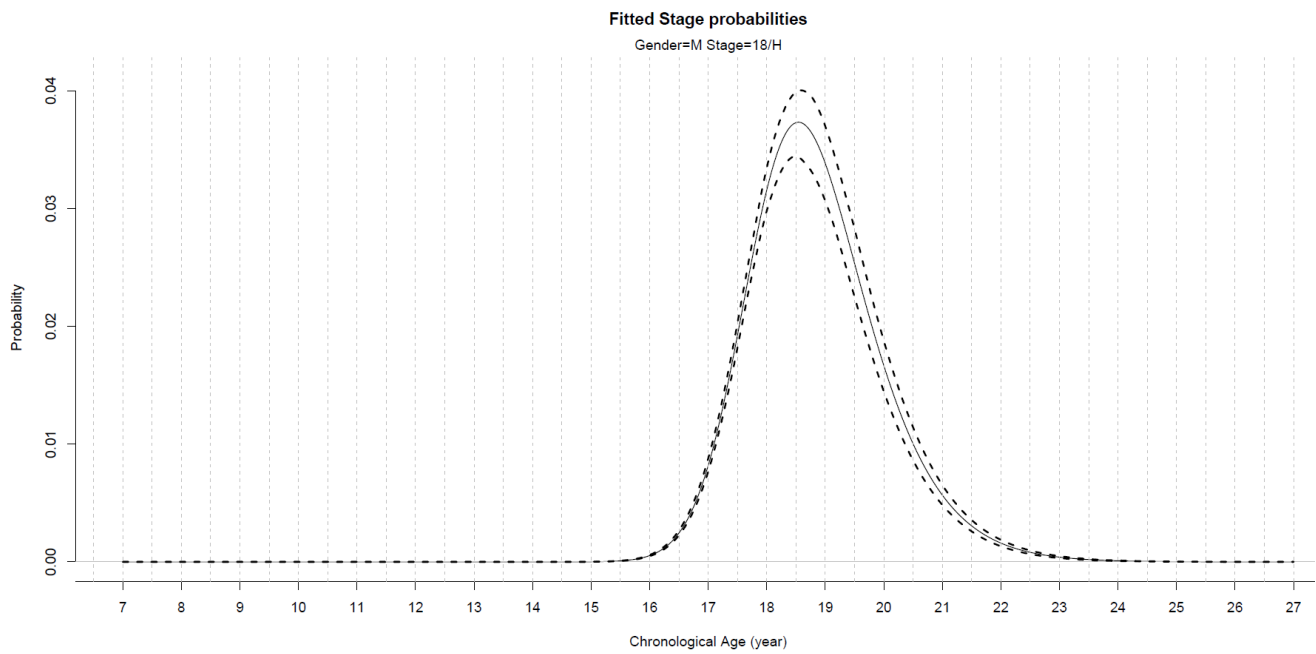


**Figure B3:** The figure shows the stage probability for boys with tooth stage G for given chronological ages.

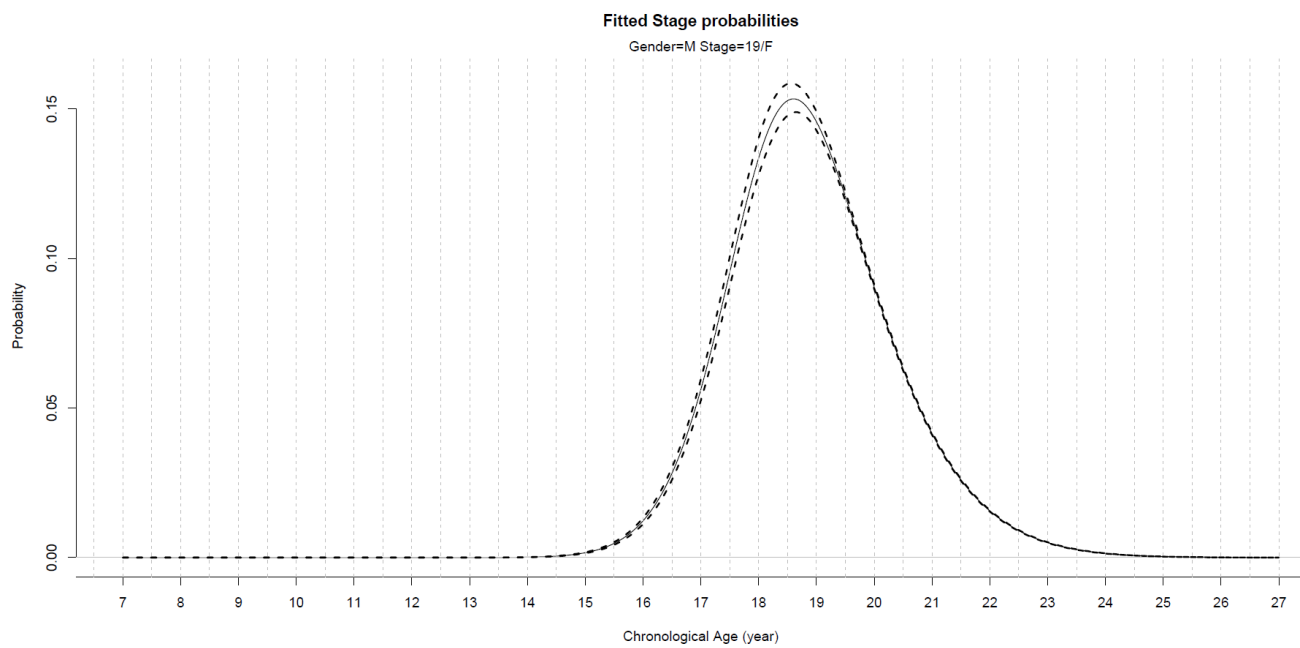


**Figure B4:** The figure shows the stage probability for boys with tooth stage H for given chronological ages.

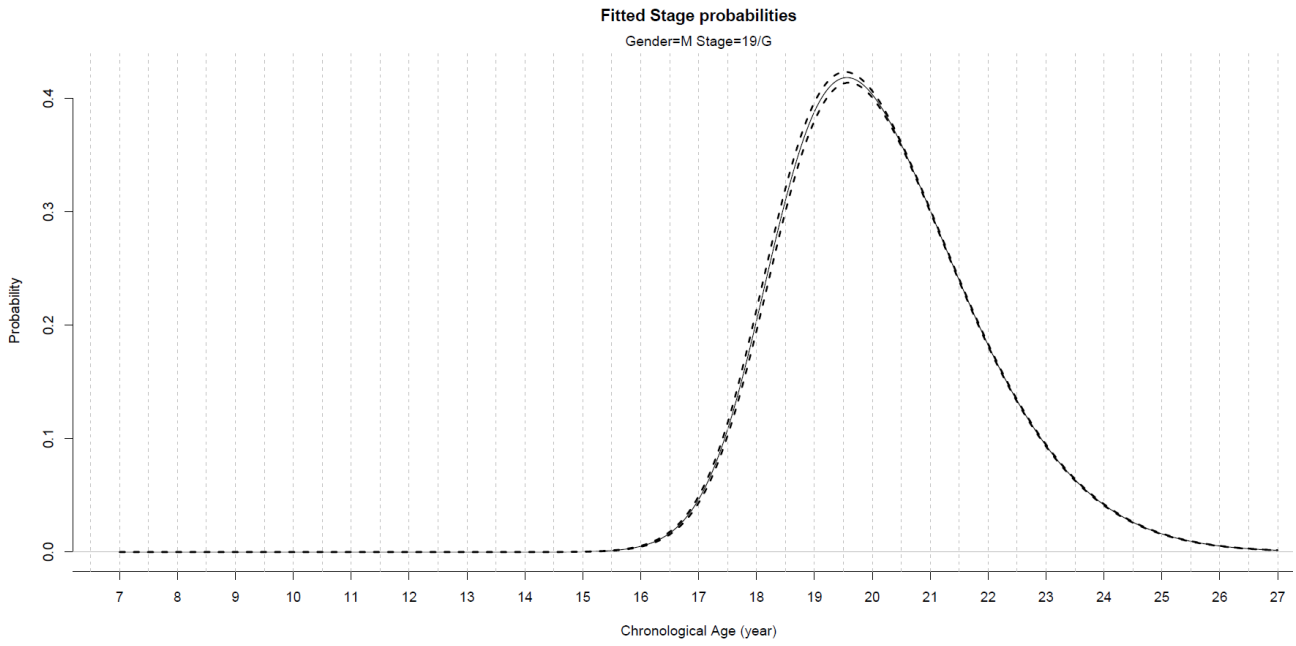
*B.3.1.3 For boys combined*



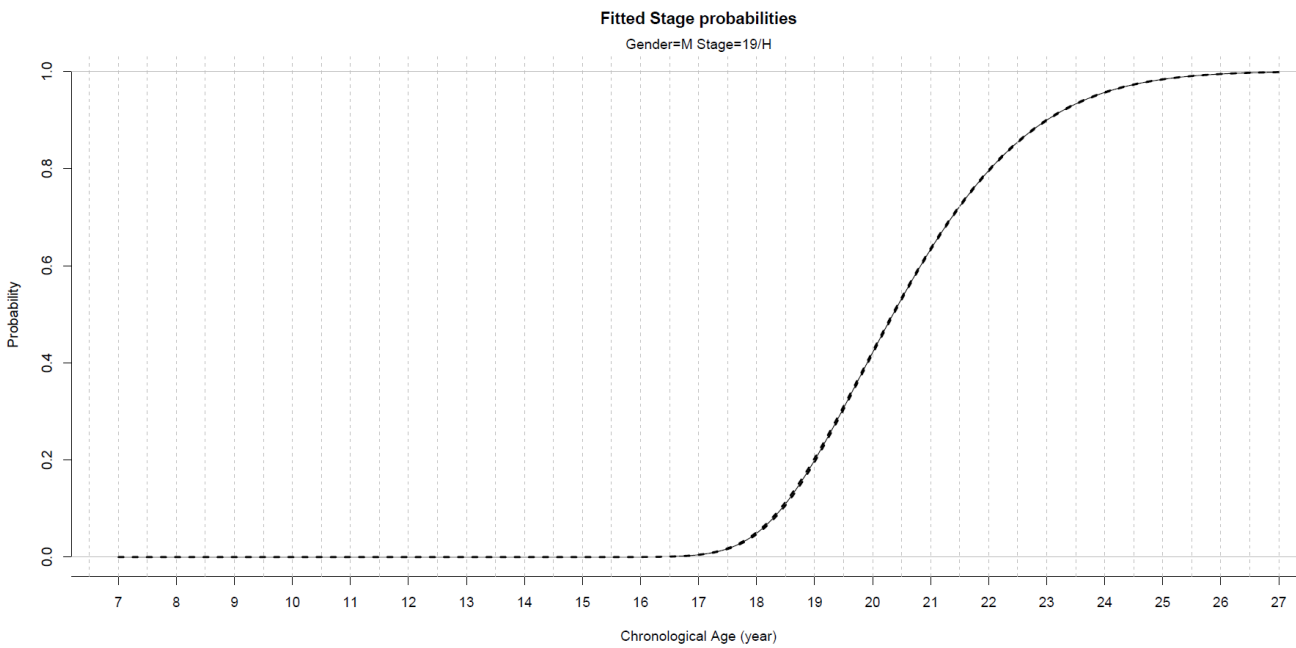
**Figure B5:** The figure shows the stage probability for boys with a skeletal age of 18 years and tooth stage H for given chronological ages.



**Figure B6:** The figure shows the stage probability for boys with a skeletal age of 19 years and tooth stage F for given chronological ages.



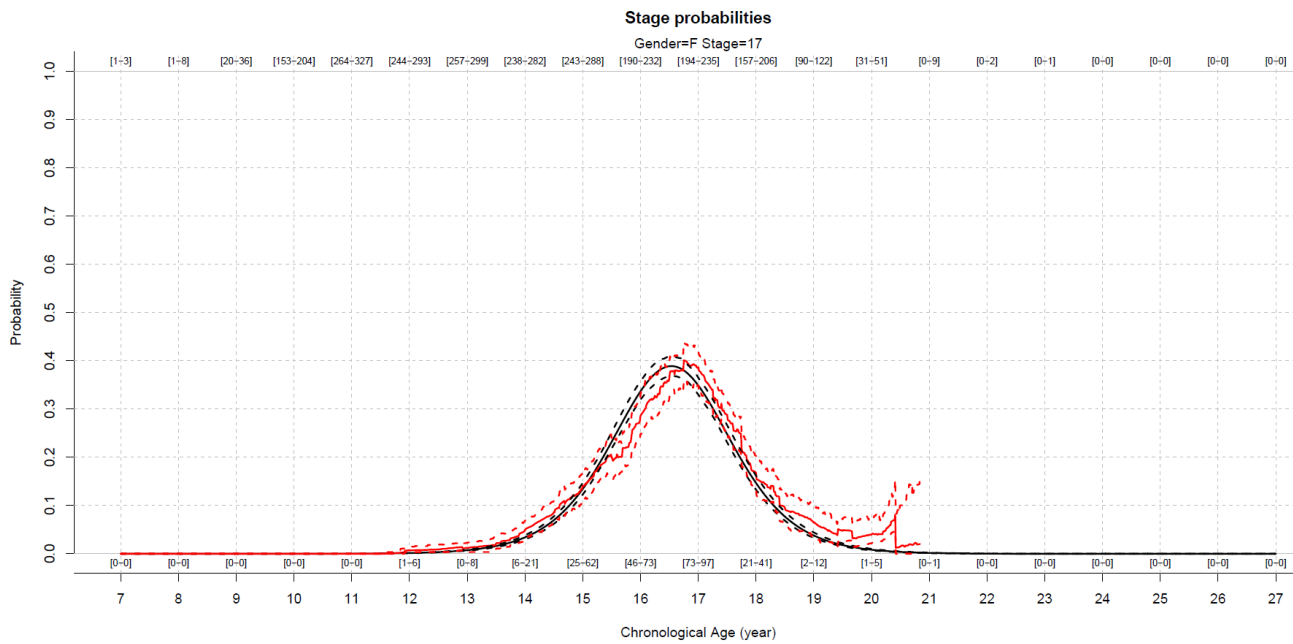
**Figure B7:** The figure shows the stage probability for boys with a skeletal age of 19 years and tooth stage G for given chronological ages.



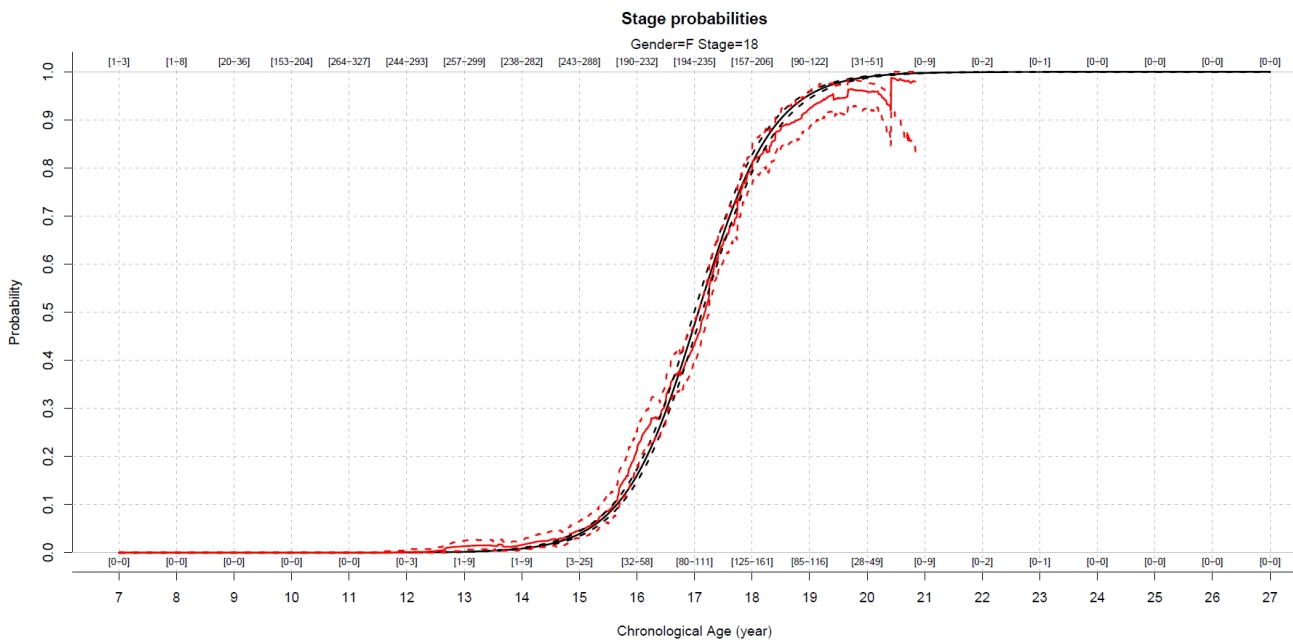
**Figure B8:** The figure shows the stage probability for boys with a skeletal age of 19 years and tooth stage H for given chronological ages.



### B.3.1.4 For girls hand



**Figure B9:** The figure shows the stage probability for girls with a skeletal age of 17 years for given chronological ages.



**Figure B10:** The figure shows the stage probability for girls with a skeletal age of 18 years for given chronological ages.

### B.3.1.5 For girls tooth

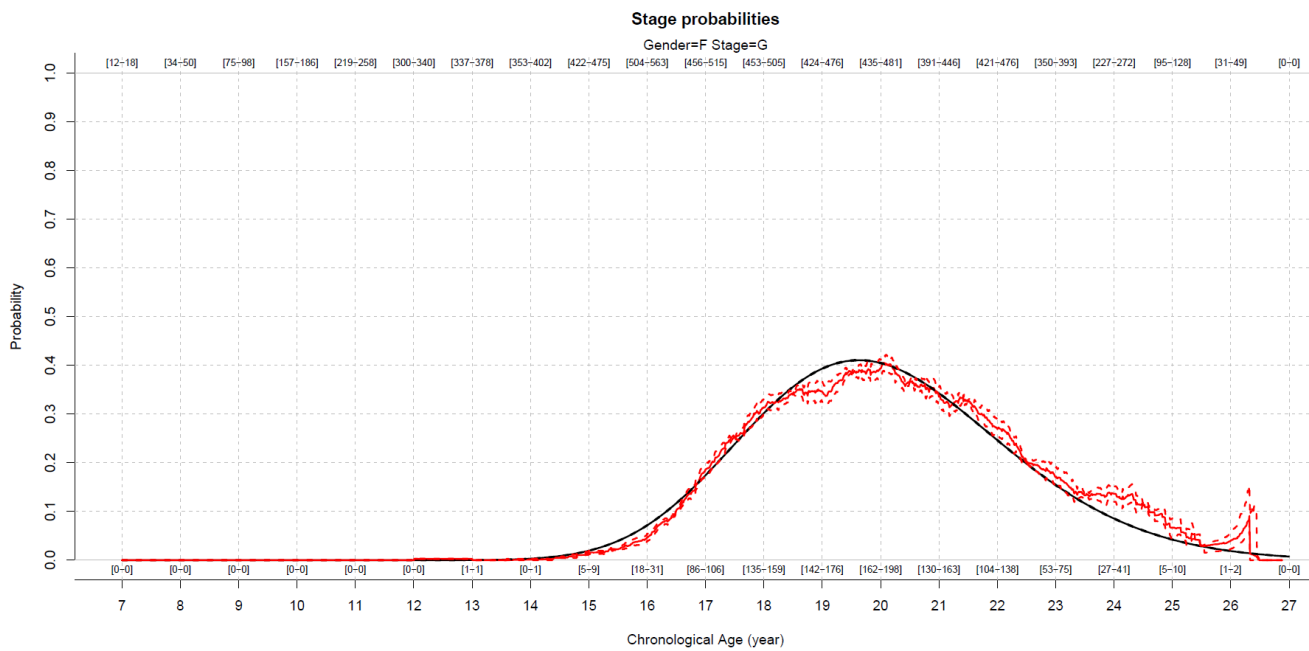


Figure B11: The figure shows the stage probability for girls with tooth stage G for given chronological ages.

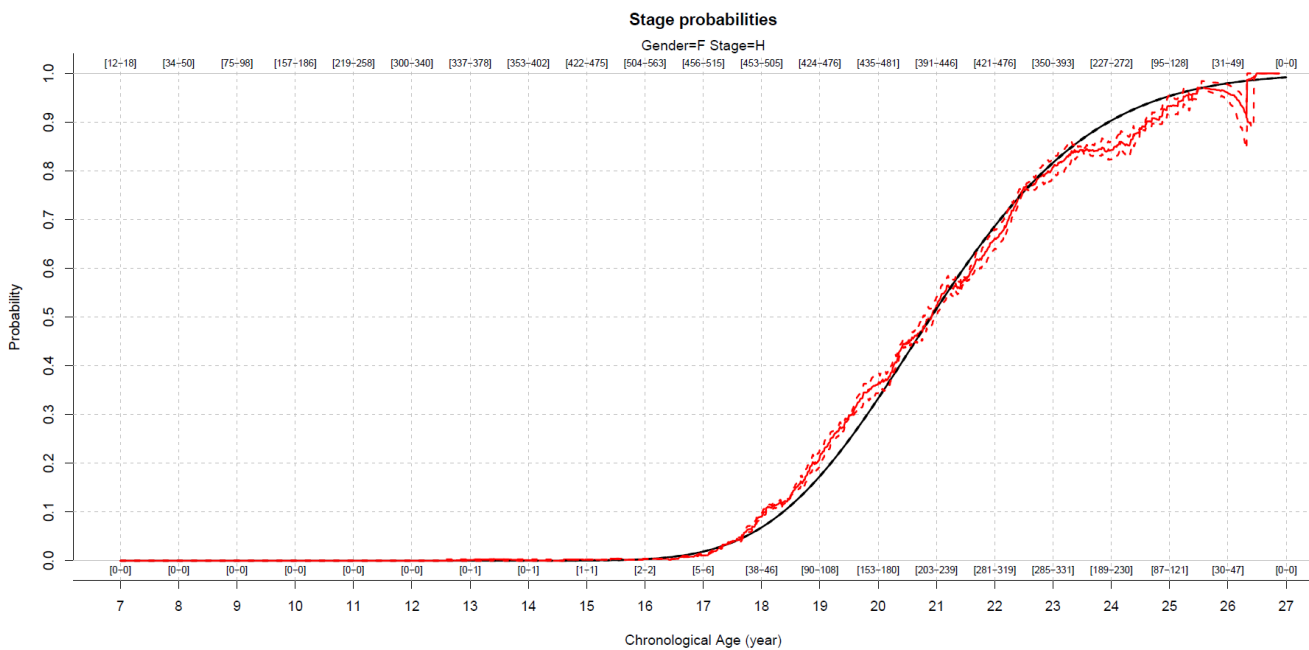
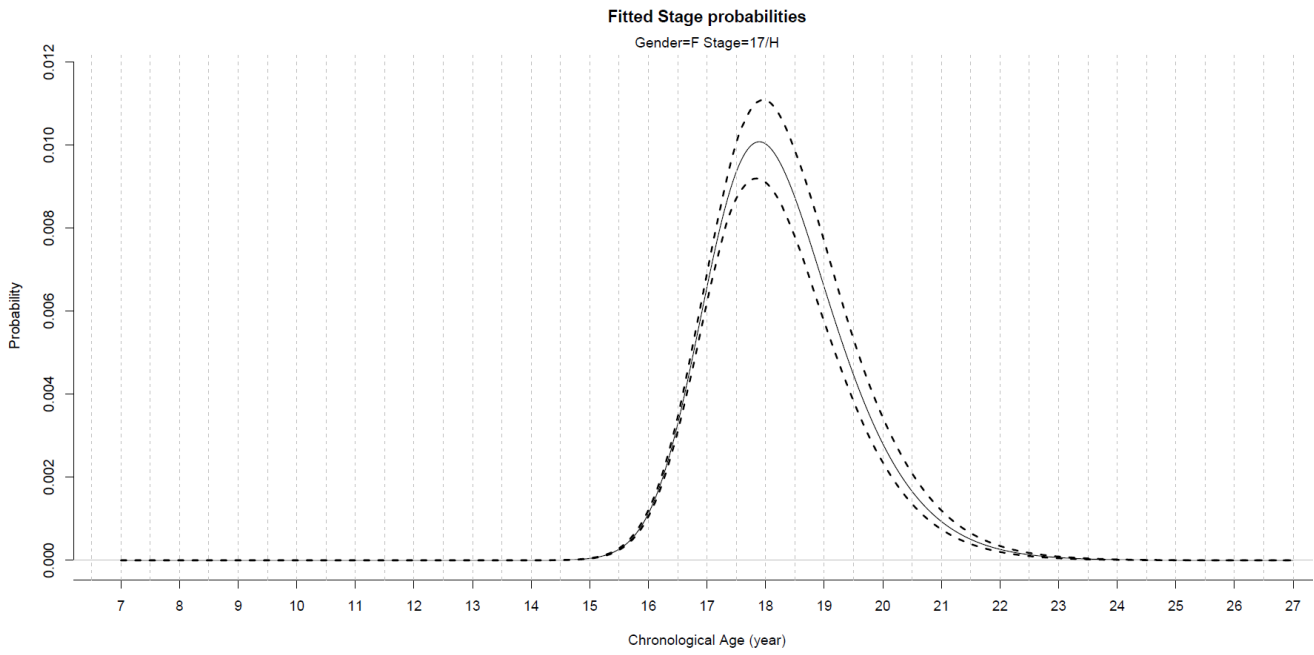
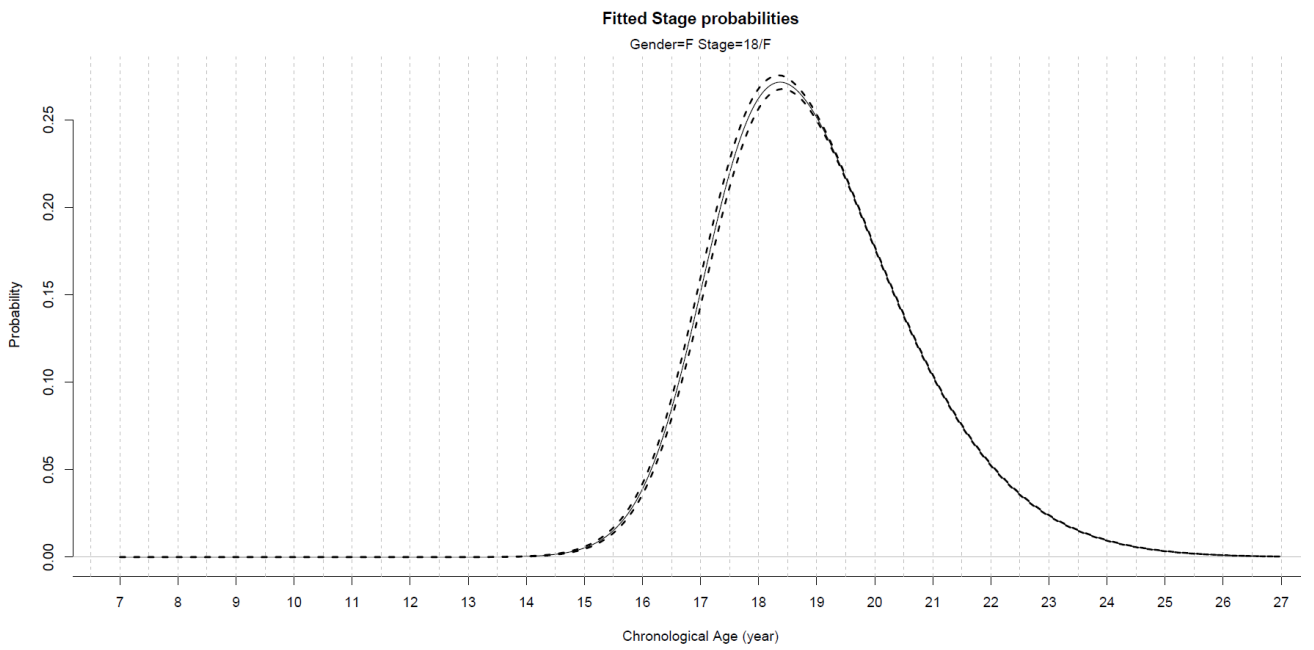


Figure B12: The figure shows the stage probability for girls with tooth stage H for given chronological ages.

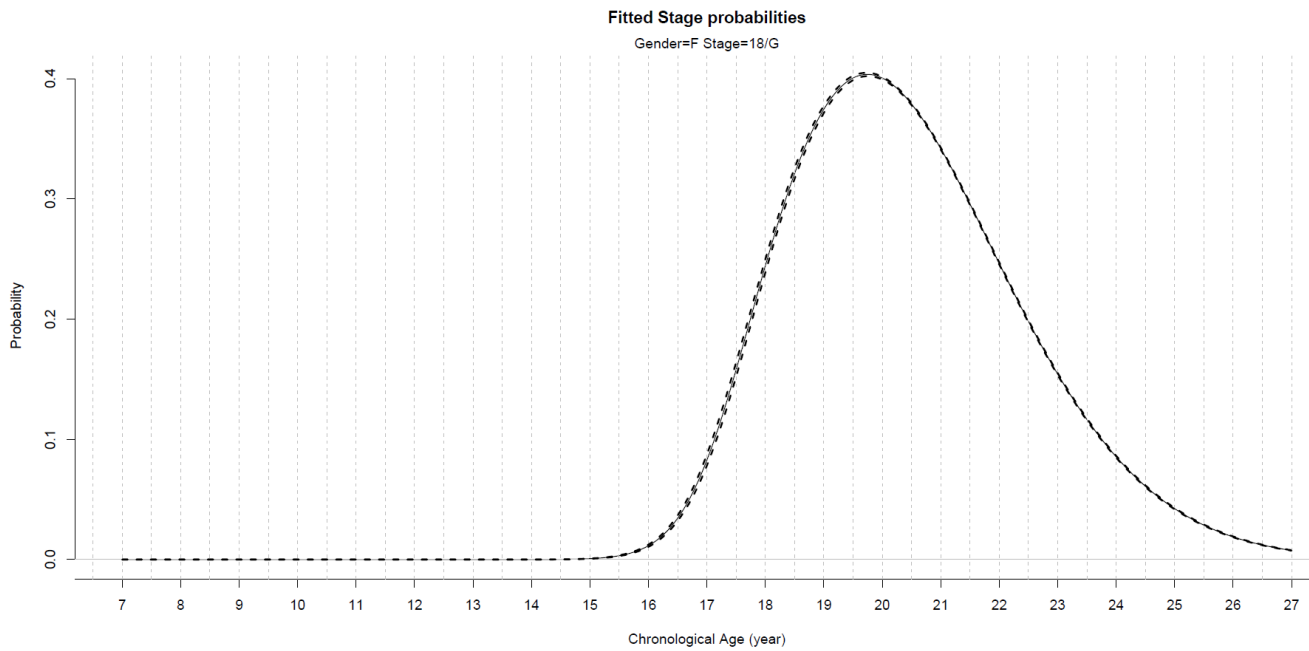
### B.3.1.6 For girls combined



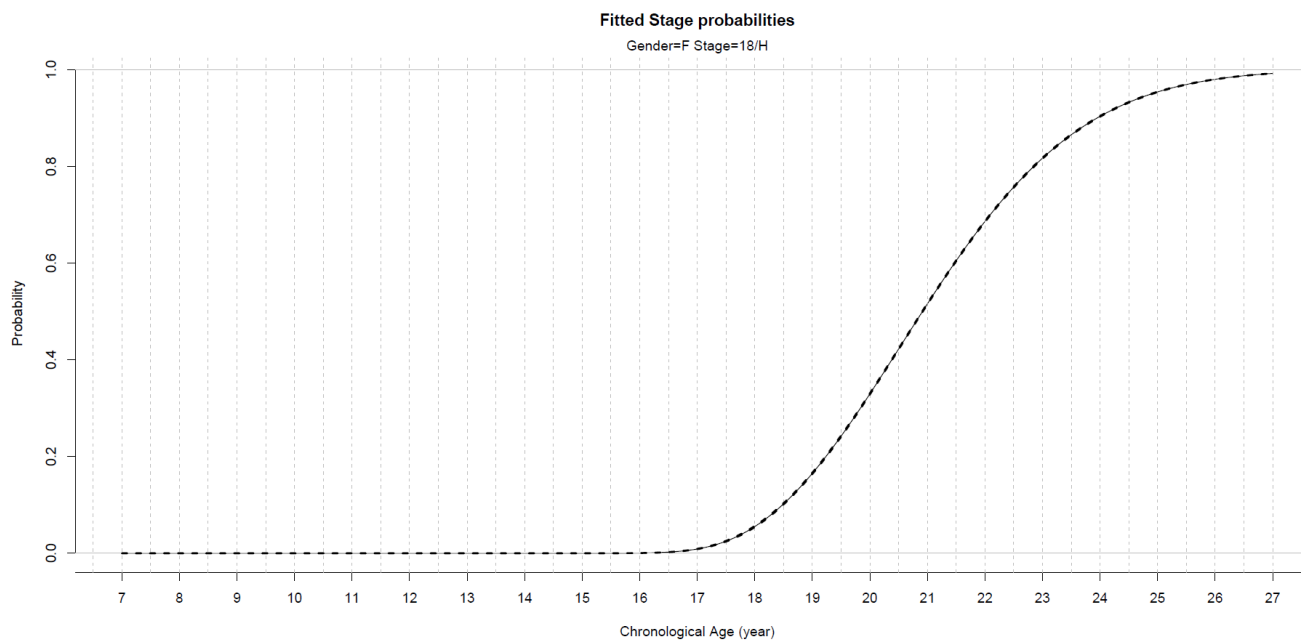
**Figure B13:** The figure shows the stage probability for girls with a skeletal age of 17 years and tooth stage H for given chronological ages.



**Figure B14:** The figure shows the stage probability for girls with a skeletal age of 18 years and tooth stage F for given chronological ages.



**Figure B15:** The figure shows the stage probability for girls with a skeletal age of 18 years and tooth stage G for given chronological ages.



**Figure B16:** The figure shows the stage probability for girls with a skeletal age of 18 years and tooth stage H for given chronological ages.

### B.3.2 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.4 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.4.1 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

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

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.4.2 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

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

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

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.4.3 Overview figures

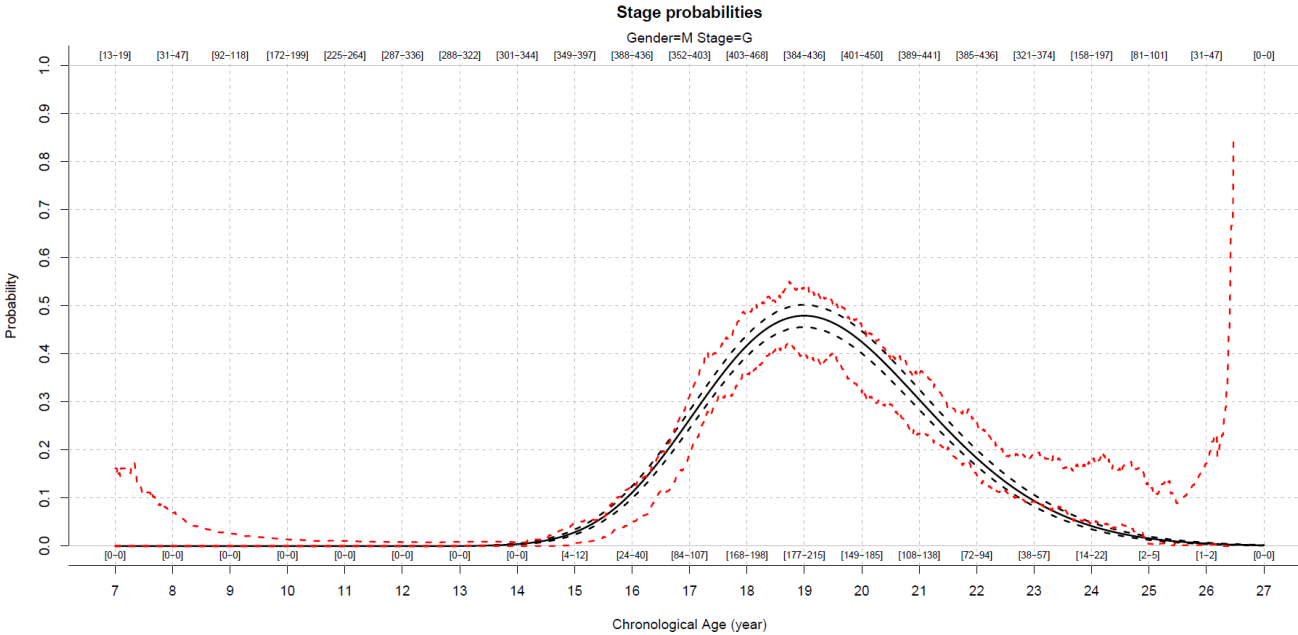
We will now show 95% credibility intervals based on non-parametric and parametric models for the cases where we were in doubt as to whether the model assumption for the parametric models is adequate. Since we get one credibility interval  $[L, U]$  for each generated dataset, we choose the 5%

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

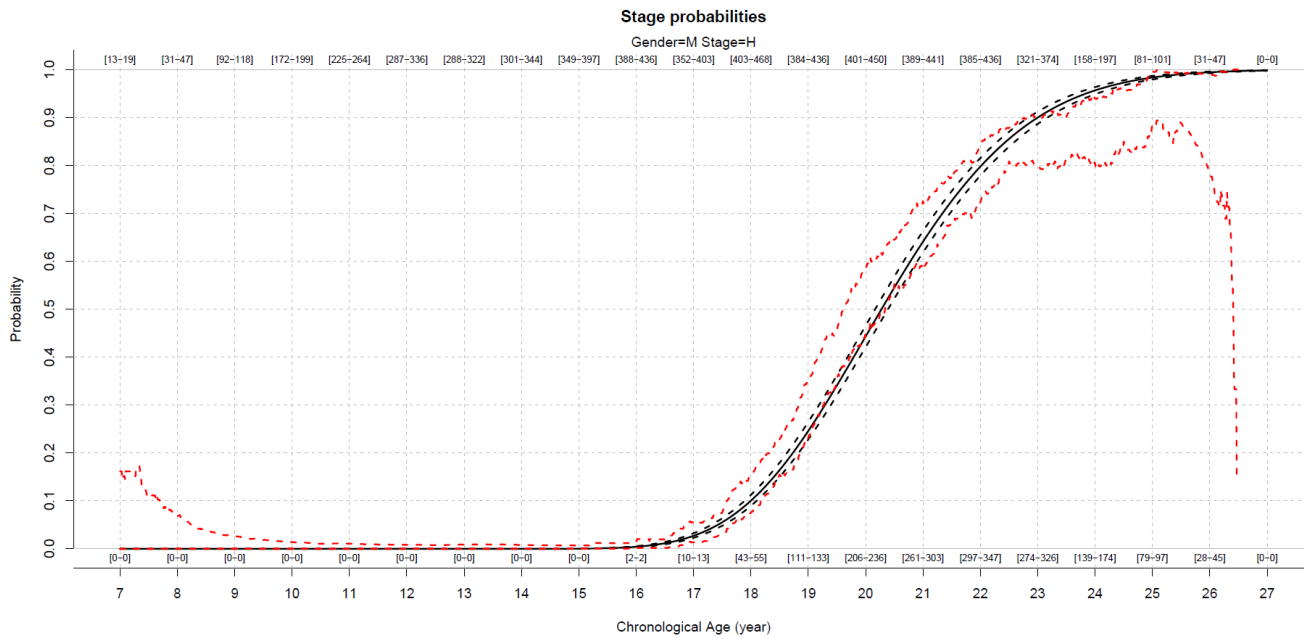
The figures below (**Figures B17–B19**) show the mean of the fitted parametric regression models for the stage probabilities (black), and 95% credibility intervals for the stage probabilities for both the parametric transition model (black stippled) and the non-parametric probabilities (red stippled) 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.

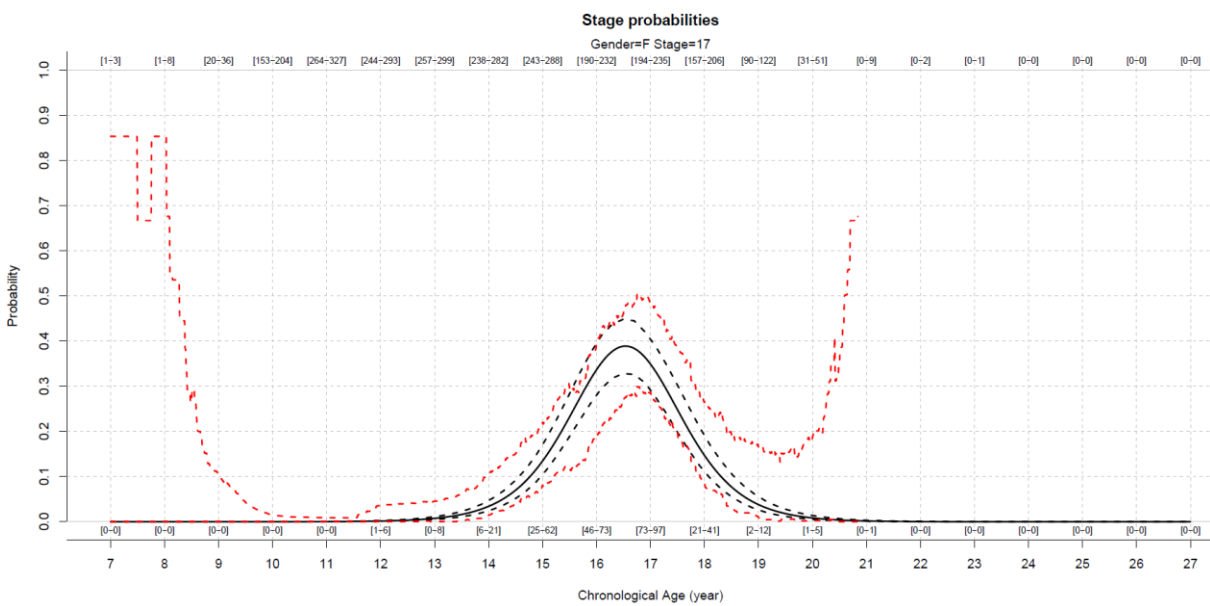
Note 2: The average parametric model (almost) always lies within the 95% credibility intervals for the probabilities based on the non-parametric model.



**Figure B17:** The figure shows the stage probability for boys with tooth stage G for given chronological ages.



**Figure B18:** The figure shows the stage probability for boys with tooth stage H for given chronological ages.



**Figure B19:** The figure shows the stage probability for girls with a skeletal age of 17 years for given chronological ages.

## B.5 Choice of upper age limits in the tool (defining the prior age distribution)

### B.5.1 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 B3** 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	19	20	20.5	21	23	25
Boys	Hand	19	18	0.41	0.23	0.19	0.16	0.09	0.07
Boys	Tooth	G	18	0.57	0.4	0.35	0.32	0.27	0.26
Boys	Tooth	H	18	0.31	0.13	0.09	0.07	0.03	0.02
Boys	Comb	19/F	18	0.48	0.34	0.31	0.29	0.27	0.27
Boys	Comb	19/G	18	0.31	0.16	0.13	0.12	0.09	0.08
Boys	Comb	18/H	18	0.4	0.29	0.27	0.26	0.25	0.25
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	0.54	0.35	0.3	0.27	0.2	0.19
Girls	Tooth	H	18	0.31	0.12	0.08	0.06	0.02	0.01
Girls	Comb	18/F	18	0.56	0.41	0.38	0.35	0.32	0.31
Girls	Comb	18/G	18	0.4	0.23	0.19	0.16	0.12	0.11
Girls	Comb	18/H	18	0.23	0.08	0.06	0.04	0.01	0.01
Boys	Hand	19	17	0.12	0.07	0.06	0.05	0.03	0.02
Boys	Tooth	G	17	0.25	0.17	0.15	0.14	0.12	0.11
Boys	Tooth	H	17	0.06	0.03	0.02	0.01	0.01	0
Boys	Comb	19/F	17	0.14	0.1	0.09	0.08	0.08	0.08
Boys	Comb	19/G	17	0.05	0.03	0.02	0.02	0.01	0.01
Boys	Comb	18/H	17	0.06	0.04	0.04	0.04	0.04	0.04
Girls	Hand	18	17	0.22	0.15	0.13	0.11	0.07	0.06
Girls	Tooth	G	17	0.23	0.15	0.13	0.11	0.09	0.08
Girls	Tooth	H	17	0.07	0.03	0.02	0.01	0.01	0
Girls	Comb	18/F	17	0.19	0.14	0.13	0.12	0.11	0.11
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	0.05	0.04	0.04	0.03
Girls	Hand	18	16	0.06	0.04	0.04	0.03	0.02	0.02
Girls	Tooth	G	16	0.07	0.05	0.04	0.03	0.03	0.02

**Table B3:** 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.5.2 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. The table below provides/presents 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

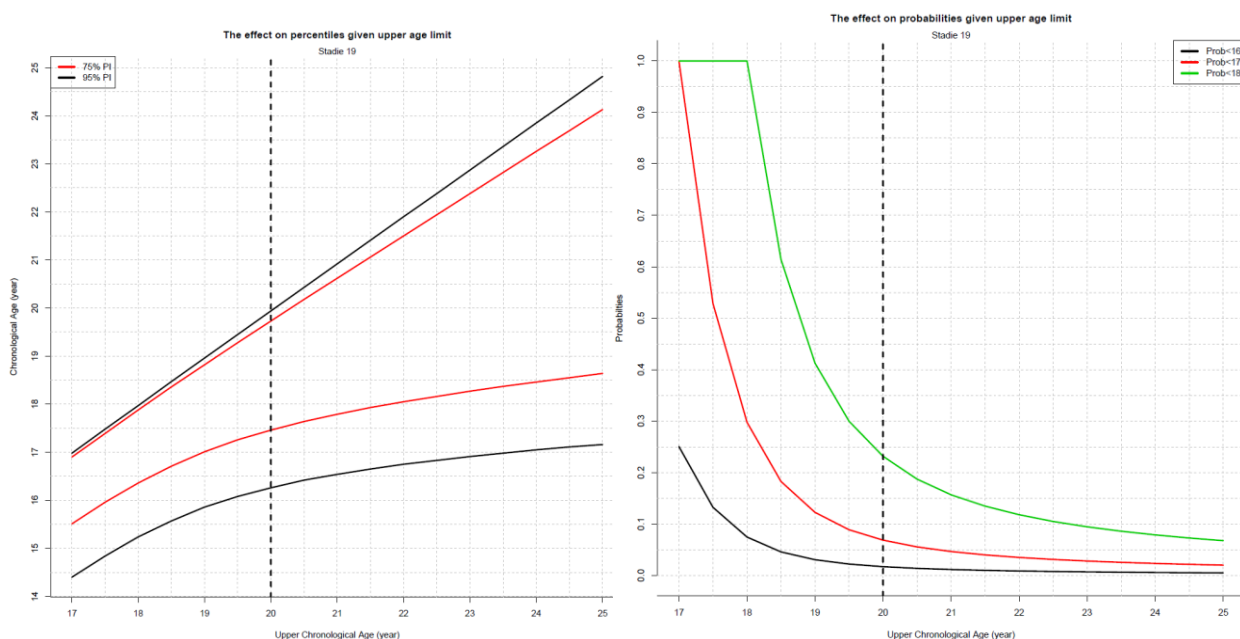
**Table B4:** The table provides an overview of defined upper age of the assumed age distribution for the different methods and genders used in the tool.

### B.5.3 Overview figures of the effect of assumed upper age

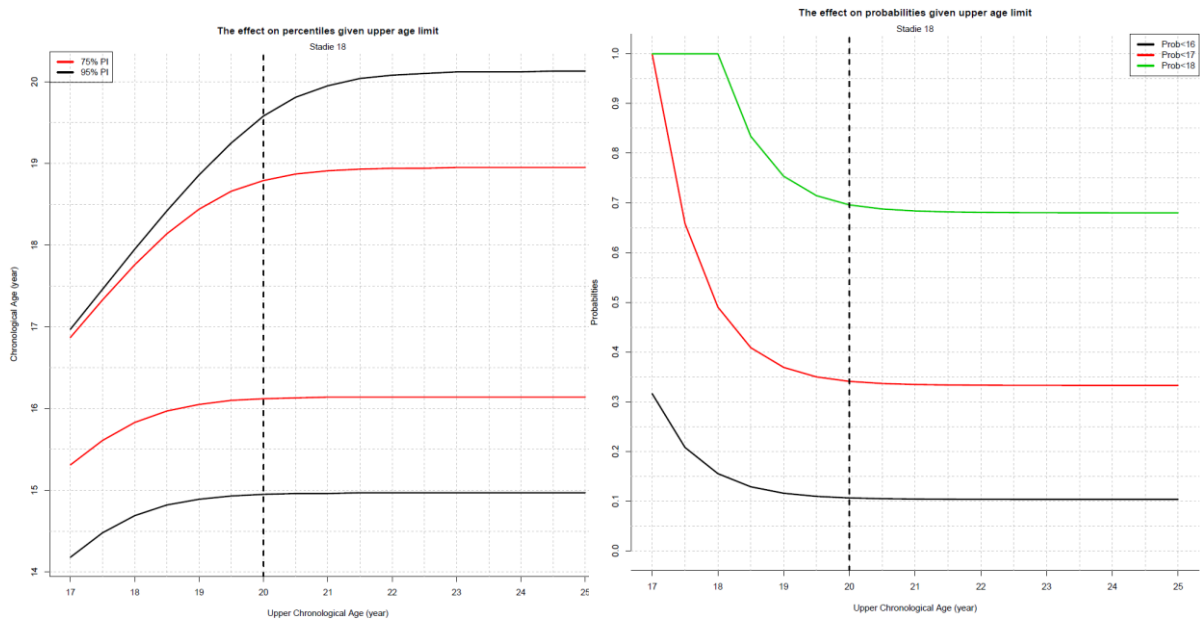
The following figures (**Figures B20–B35**) in sections **B.5.3.1–B.5.3.6** show the effect of assuming different upper ages in order to work backwards (i.e. applying Bayes theorem) to describe the distribution of chronological age for given stages.

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.5.3.1 Effect for boys: hand skeletal ages 18 and 19 years

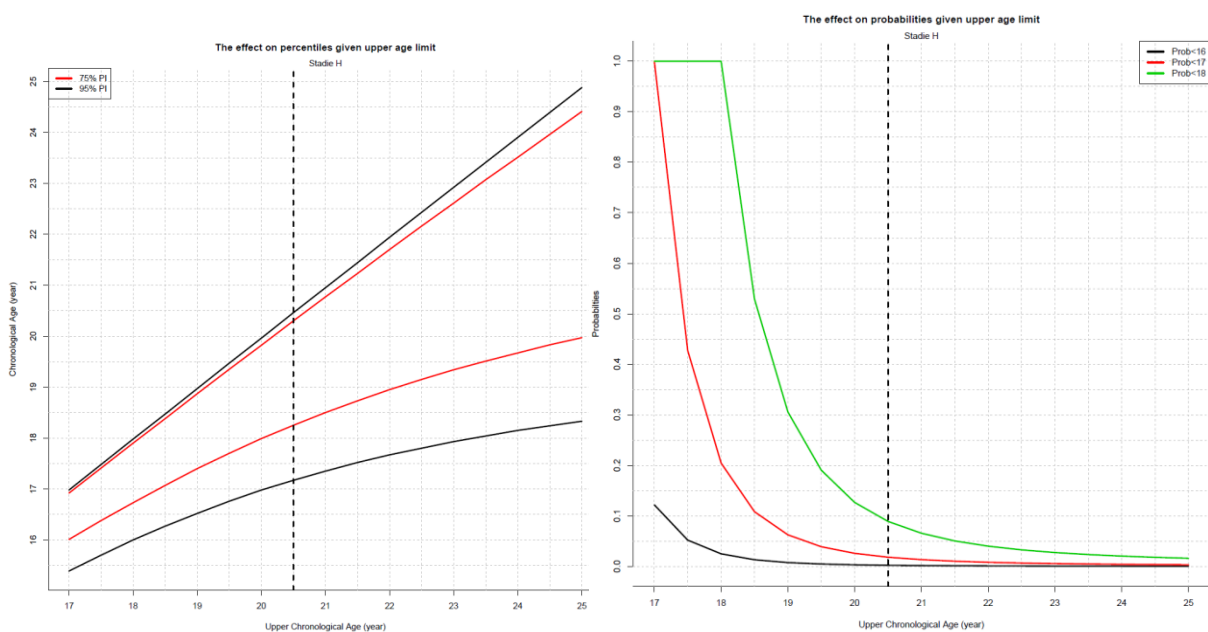


**Figure B20:** The figure shows prediction intervals and probabilities for boys with a skeletal age of 19 years.

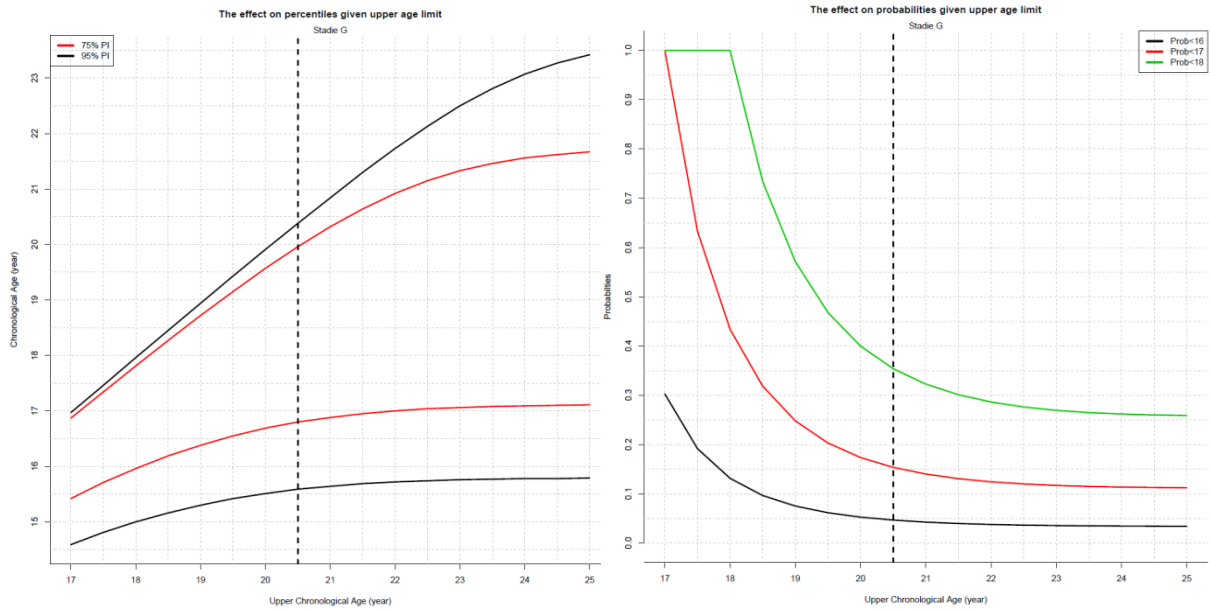


**Figure B21:** The figure shows prediction intervals and probabilities for boys with a skeletal age of 18 years.

### B.5.3.2 Effect for boys: tooth stages G and H

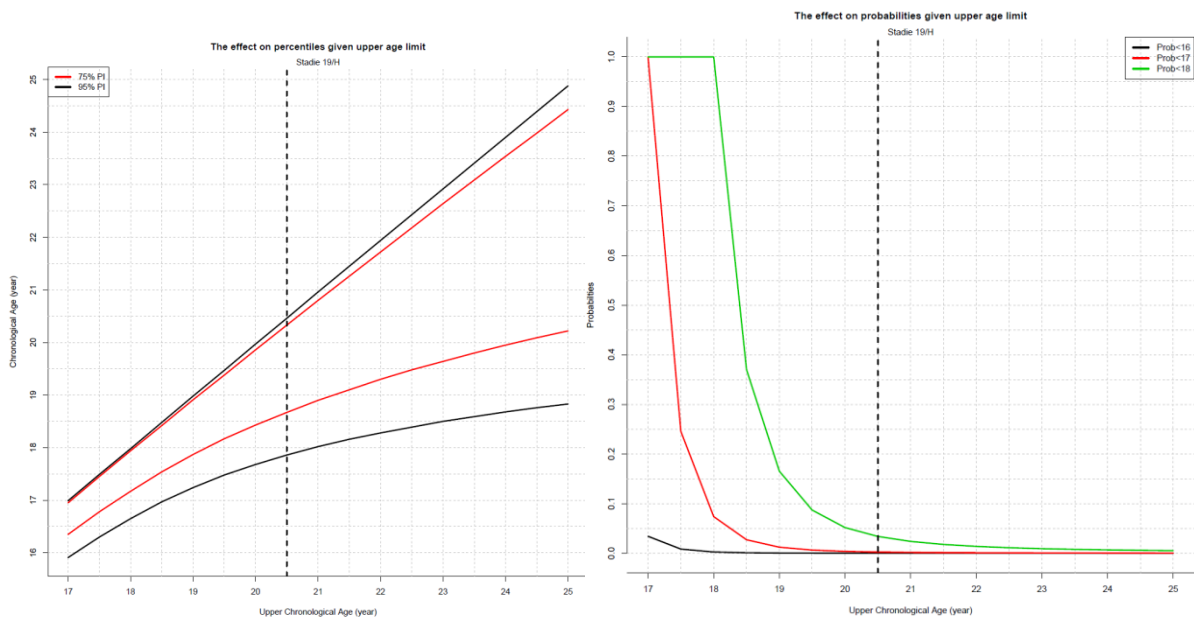


**Figure B22:** The figure shows prediction intervals and probabilities for boys with tooth stage H.

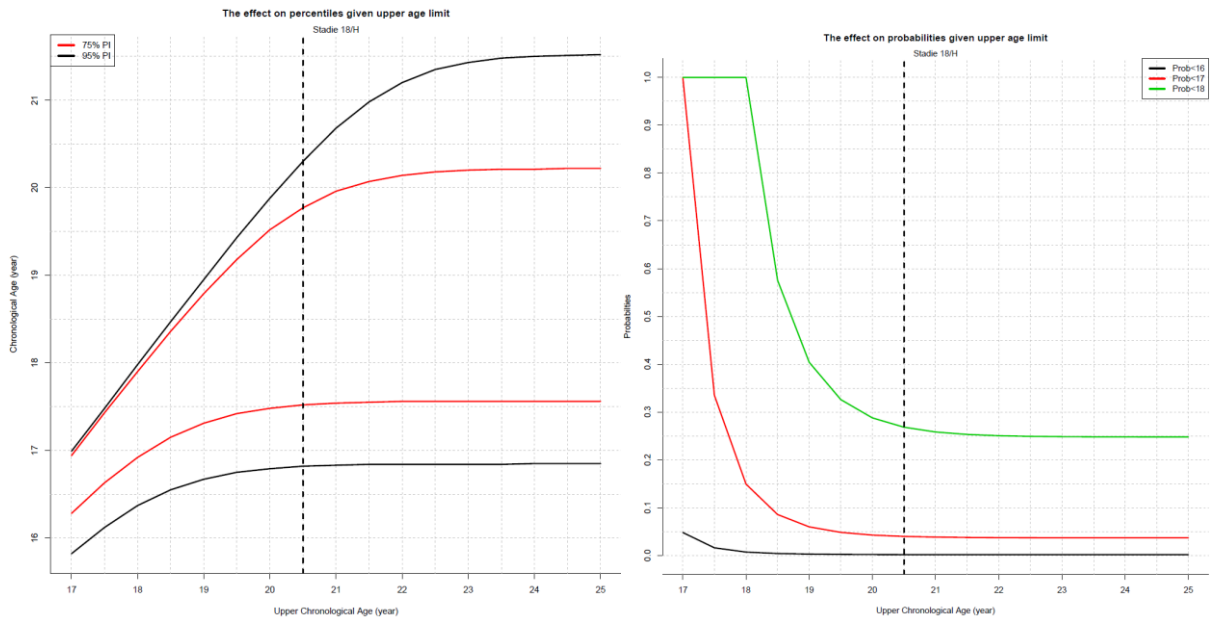


**Figure B23:** The figure shows prediction intervals and probabilities for boys with tooth stage G.

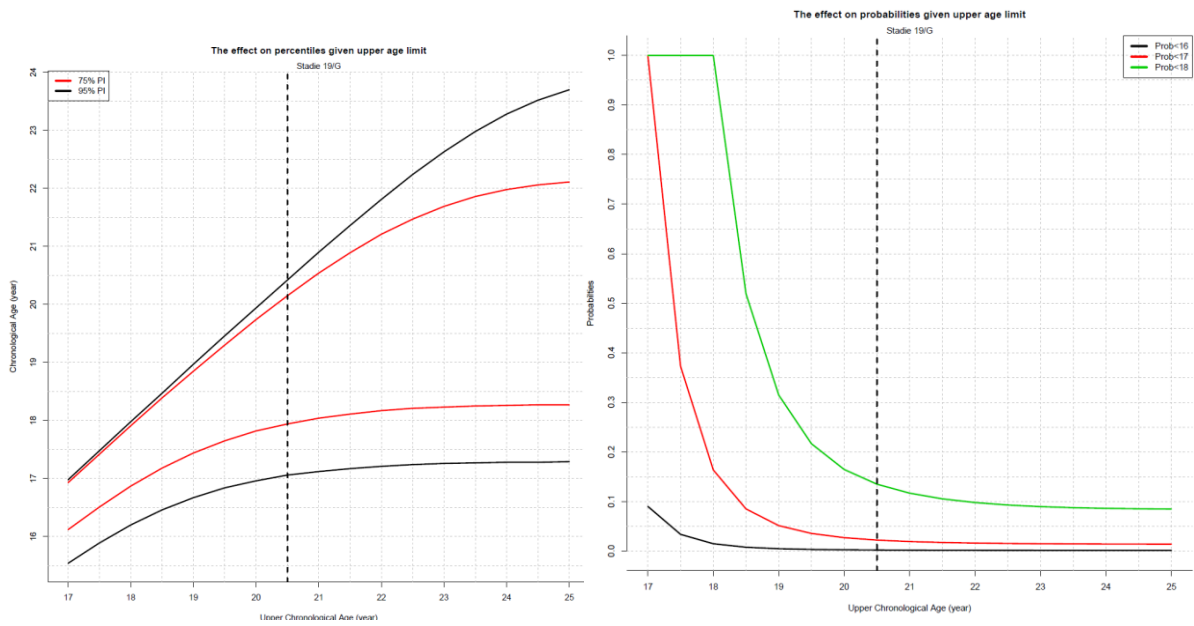
*B.5.3.3 Effect for boys: combined stages 19/F, 19/G, 18/H and 19/H*



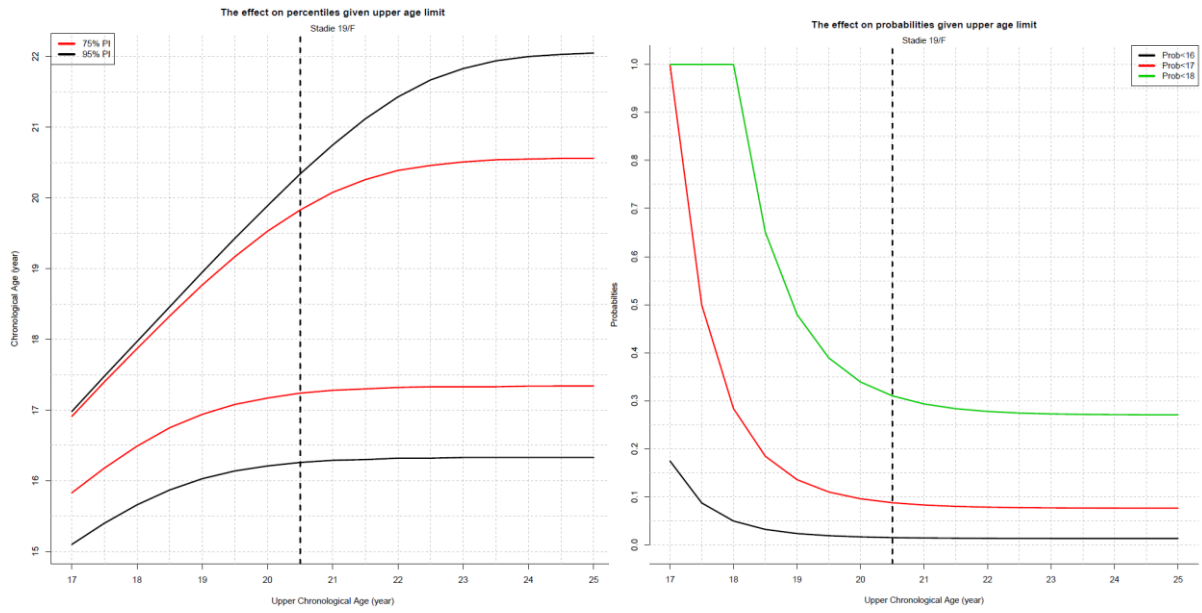
**Figure B24:** The figure shows prediction intervals and probabilities for boys with skeletal age 19 years and tooth stage H.



**Figure B25:** The figure shows prediction intervals and probabilities for boys with skeletal age 18 years and tooth stage H.

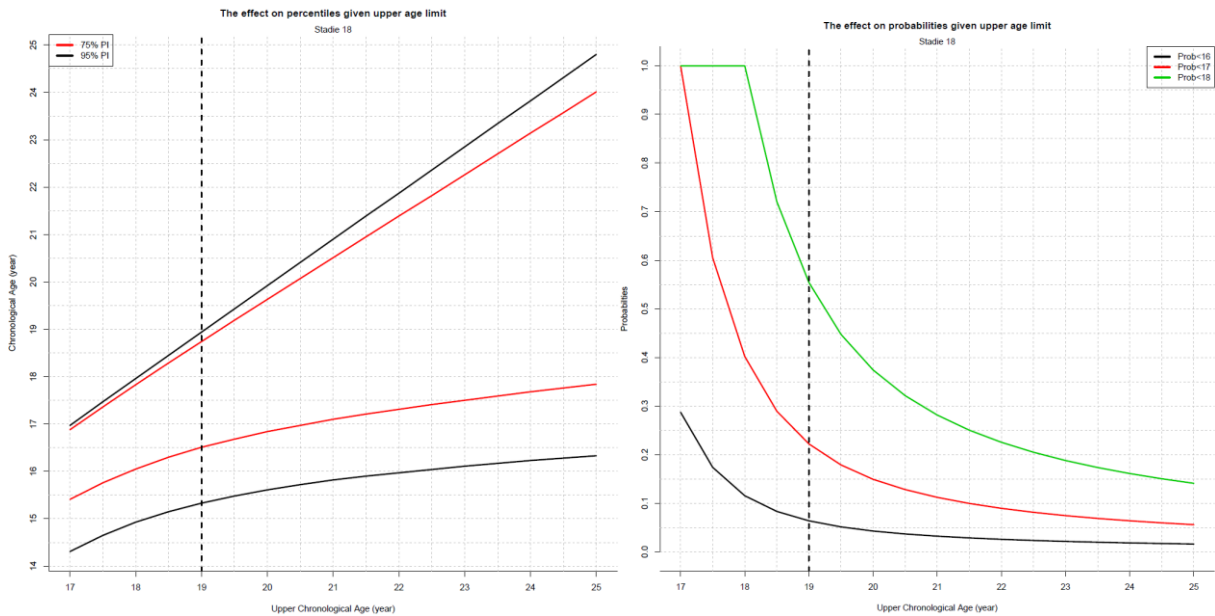


**Figure B26:** The figure shows prediction intervals and probabilities for boys with skeletal age 19 years and tooth stage G.

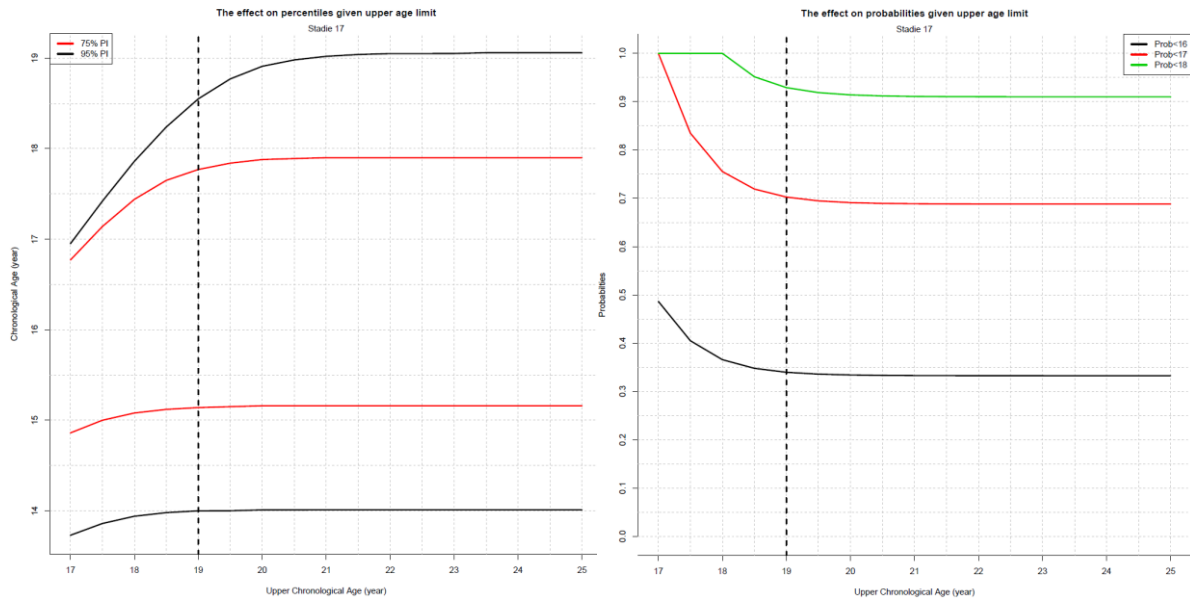


**Figure B27:** The figure shows prediction intervals and probabilities for boys with skeletal age 19 years and tooth stage F.

*B.5.3.4 Effect for girls: hand skeletal ages 17 and 18 years*

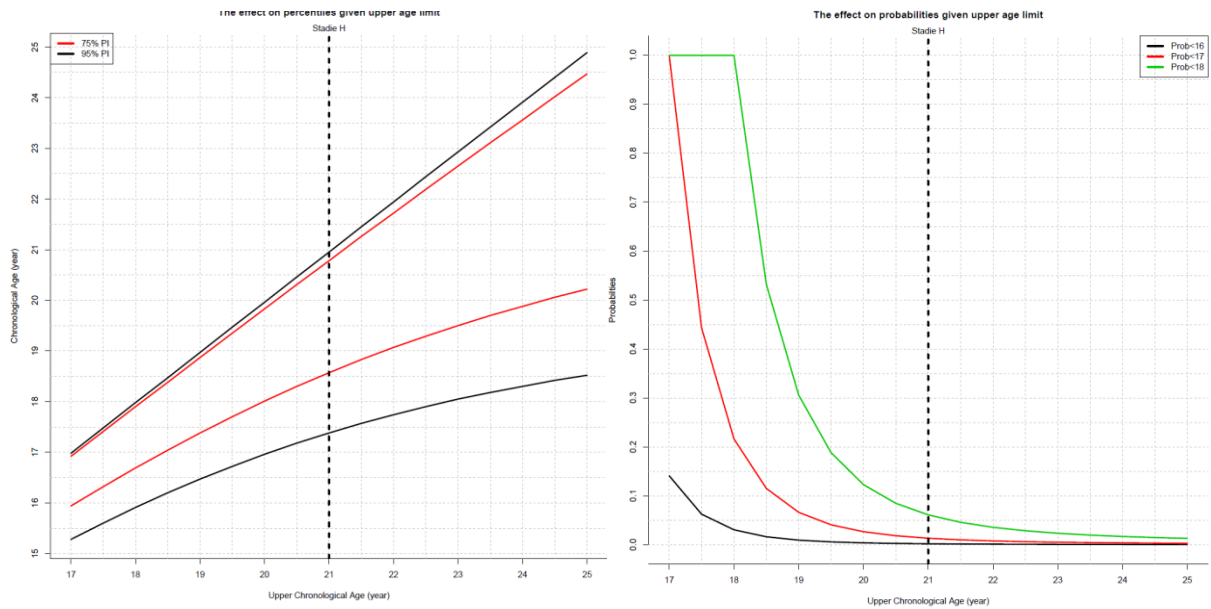


**Figure B28:** The figure shows prediction intervals and probabilities for girls with a skeletal age of 18 years.

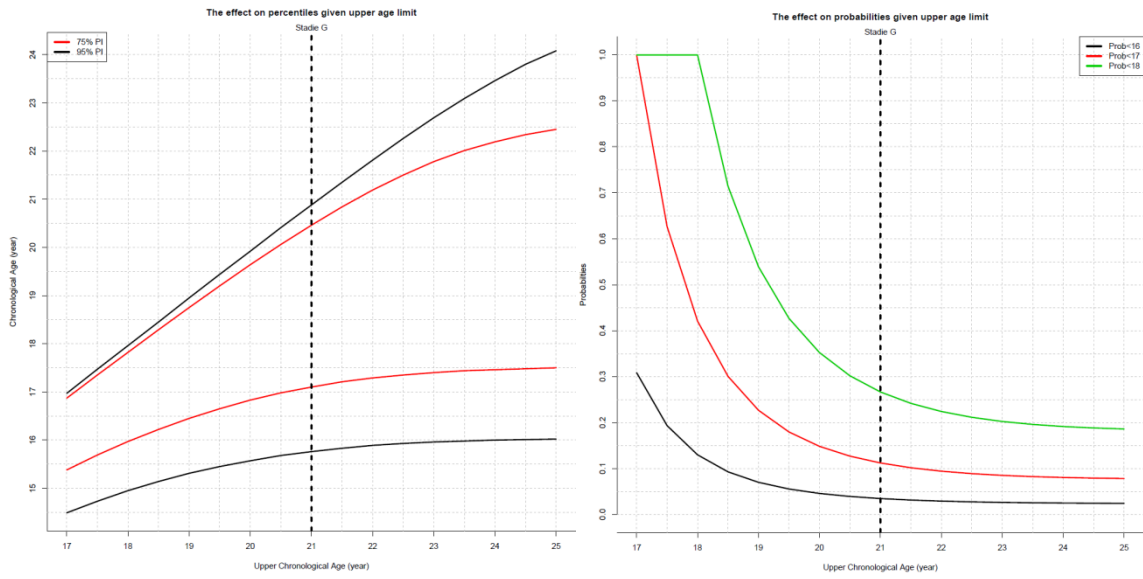


**Figure B29:** The figure shows prediction intervals and probabilities for girls with a skeletal age of 17 years.

### B.5.3.5 Effect for girls: tooth stages G and H

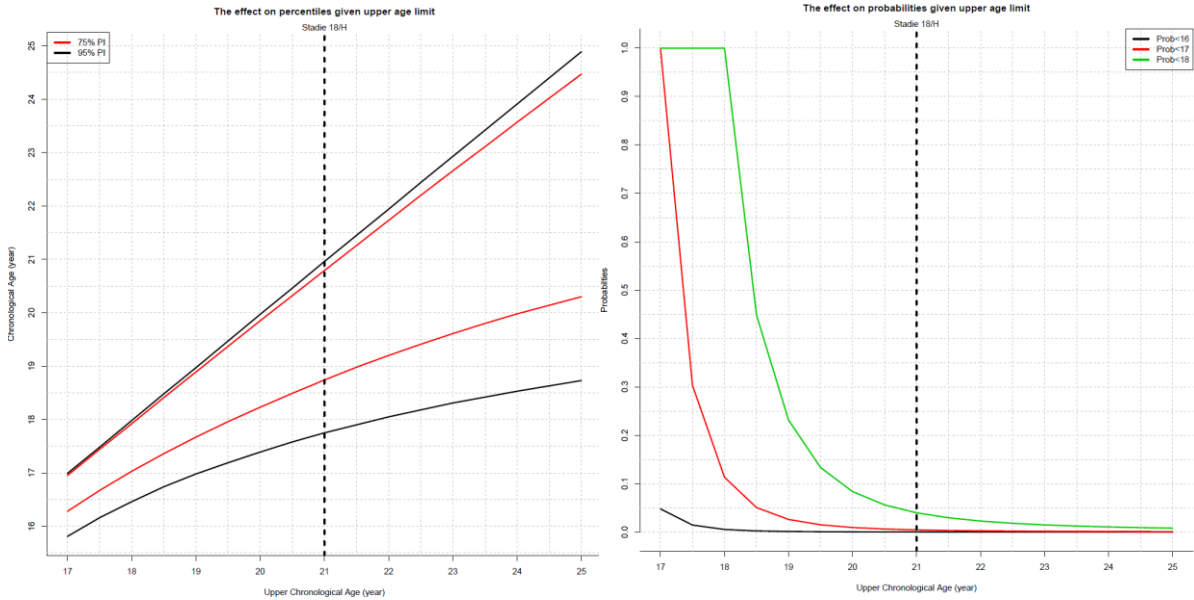


**Figure B30:** The figure shows prediction intervals and probabilities for girls with tooth stage H.

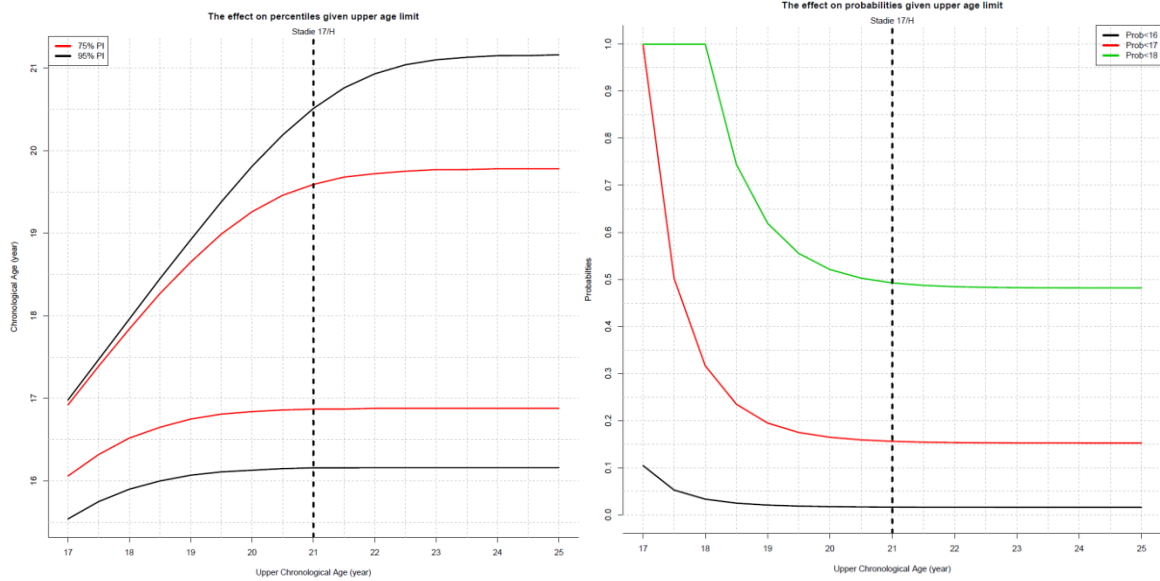


**Figure B31:** The figure shows prediction intervals and probabilities for girls with tooth stage G.

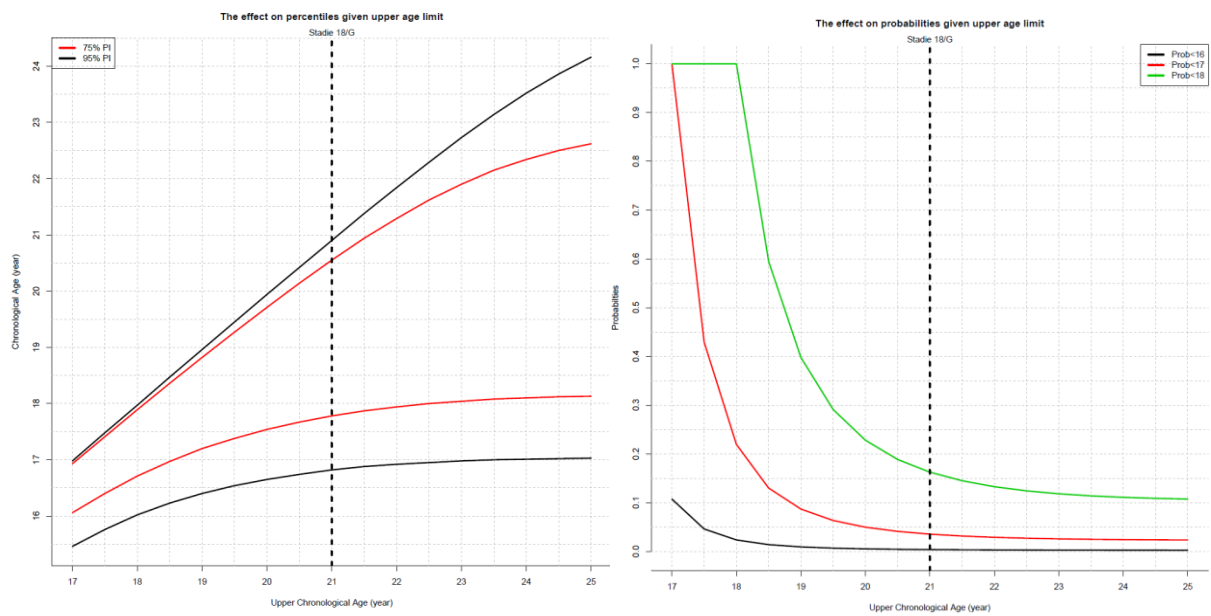
*B.5.3.6 Effect for girls: combined stages 18/F, 18/G, 17/H and 18/H*



**Figure B32:** The figure shows prediction intervals and probabilities for girls with skeletal age 18 years and tooth stage H.

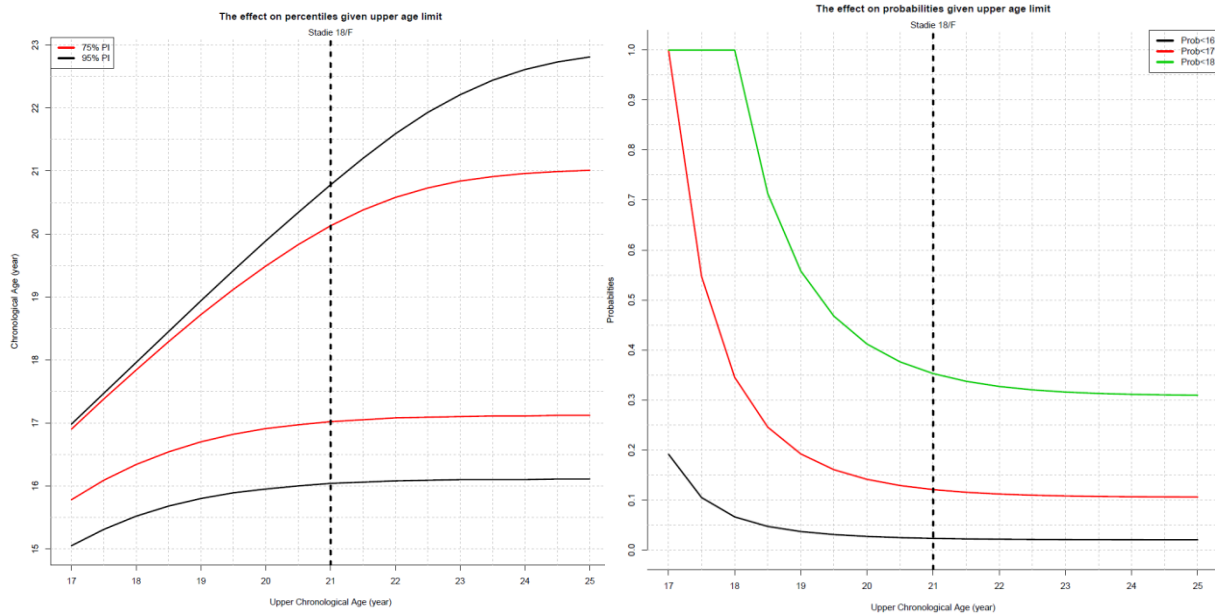


**Figure B33:** The figure shows prediction intervals and probabilities for girls with skeletal age 17 years and tooth stage H.



**Figure B34:** The figure shows prediction intervals and probabilities for girls with skeletal age 18 years and tooth stage G.





**Figure B35:** 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. Johan NA, Khamis MF, Abdul Jamal NS, Ahmad B, Mahanani ES. The variability of lower third molar development in Northeast Malaysian population with application to age estimation. *Journal of Forensic Odonto-Stomatology*. 2012;30(1):45-54.
3. Lee SH, Lee JY, Park HK, Kim YK. Development of third molars in Korean juveniles and adolescents. *Forensic Science International*. 2009;188(1):107-11.
4. Li G, Ren J, Zhao S, Liu Y, Li N, Wu W, et al. Dental age estimation from the developmental stage of the third molars in western Chinese population. *Forensic Science International*. 2012;219(1):158-64.
5. Duangto P, Iamaroon A, Prasitwattanaseree S, Mahakkanukrauh P, Janhom A. New models for age estimation and assessment of their accuracy using developing mandibular third molar teeth in a Thai population. *International journal of legal medicine*. 2017;131(2):559-68.
6. 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.
7. 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.
8. Yee TW. VGAM: Vector Generalized Linear and Additive Models 2017 [Available from: <https://CRAN.R-project.org/package=VGAM>].

9. 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. *Ann Hum Biol.* 2015;42(4):389-96.
10. Chaumoitre K, Saliba-Serre B, Adalian P, Signoli M, Leonetti G, Panuel M. Forensic use of the Greulich and Pyle atlas: prediction intervals and relevance. *European Radiology.* 2017;27(3):1032-43.