Norsk Regnesentral
NORWEGIAN COMPUTING CENTER

Note

# Assessing agreement between skeletal and dental age estimation methods


**Note no.** SAMBA/2/13

**Authors** Marion Haugen, André Teigland, Sigrid Ingeborg Kvaal

**Date** March 2013

**Norsk Regnesentral**

Norsk Regnesentral (Norwegian Computing Center, NR) is a private, independent, non-profit foundation established in 1952. NR carries out contract research and development projects in information and communication technology and applied statistical-mathematical modeling. The clients include a broad range of industrial, commercial and public service organizations in the national as well as the international market. Our scientific and technical capabilities are further developed in co-operation with The Research Council of Norway and key customers. The results of our projects may take the form of reports, software, prototypes, and short courses. A proof of the confidence and appreciation our clients have in us is given by the fact that most of our new contracts are signed with previous customers.

**Norsk Regnesentral**
Norwegian Computing Center
Postboks 114 Blindern
NO-0314 Oslo, Norway

**Besøksadresse**
*Office address*
Gaustadalleen 23A
NO-0373 Oslo, Norway

**Internett** · internet
www.nr.no
**E-post** · e-mail
nr@nr.no

**Telefon** · telephone
(+47) 22 85 25 00

| | |
|---|---|
| **Title** | **Assessing agreement between skeletal and dental age estimation methods** |
| **Authors** | **Marion Haugen, André Teigland, Sigrid Ingeborg Kvaal** |
| Date | March 2013 |
| Publication number | SAMBA/2/13 |


## Abstract

When two different methods of measuring some quantity are compared, the aim is to find out whether the methods agree well enough for one method to support or replace the other. The product-moment correlation coefficient is often used but this approach is misleading. The correlation coefficient measures association and not agreement. In this paper the limits of agreement approach, also known as the Bland Altman method, is applied to assess the agreement between age estimates from radiographs of the wrist and the teeth.



| | |
|---|---|
| Keywords | Wrist and dental age estimation, difference in estimates, limits of agreement, confidence intervals |
| Target group | Researchers and immigration authorities |
| Availability | Open |
| Project number | 220555 |
| Research field | Image analysis and statistics |
| Number of pages | 18 |

# Contents

# 1 Introduction

When two (or more) methods are compared frequently neither provides the true value of the quantity being measured. In this case we want to know whether the methods agree well enough for one method to support or replace the other, or perhaps for the two methods to be used interchangeably. How far apart measurements can be without causing difficulties is a question of judgment. Methods for assessing agreement in method comparison studies are described in Altman and Bland (1983) and Bland and Altman (1986).

There are several methods for age estimation in youths and young adults. In this paper the agreement between age estimates from radiographs of the wrist and age estimates from radiographs of the teeth is assessed. The two estimates are independent. The true chronological age is unknown so the estimates cannot be compared to the true value.

# 2 Material and methods

From the original data set we have removed five duplicates, eight males with missing observations from radiographs of the wrist, 89 missing observations from radiographs of the teeth (8 females and 81 males) and 390 observations where the dental age assessment was carried out on 2nd molars (30 females and 360 males). With these constraints we are left with a dataset of 1223 observations, 138 females and 1085 males.

The age assessment from radiographs of the wrist is based on the Greulich and Pyle method. Normally, skeletal growth is complete for girls when they reach 18 years of age and for boys aged 19. When skeletal growth has ceased a radiographic examination cannot tell anything more than that a girl is 18 or older or that a boy is 19 or older. It is likewise for dental development. The wisdom tooth, which is the last tooth to complete root development, was assessed from orthopantomogram (OPG) radiographs of the jaw. The root development was assessed from tables published by Liversidge (2008). According to these tables, the median age for completion of the dental development is 20.07 years in females and 19.53 years in males.

113 females (81.9%) and 778 males (71.7%) have a wrist age estimate of 18 and 19 years, respectively, and a higher estimate for dental age. The two estimates are in agreement since we know that the individuals are 18 or older. Assessment of agreement is the aim of the analysis and therefore the dental age estimates for these individuals are censored at 18 and 19 years, respectively.

The age and sex distribution from radiographs of the wrist is given in Table 1. Non-censored age and sex distribution from radiographs of the teeth is given in Table 2 whereas censored age and sex distribution is given in Table 3. In those cases where the ages are not censored at 18 or 19 years there is a difference between the ages estimated from wrist and teeth. The numbers are rounded off, i.e. an age estimate of 13.5 is counted as 14 years of age, 14.5 as 15 years, etc.

Table 1. Age and sex distribution from the wrist age estimates.

| Estimated age (years) | Females | Males | Total |
|---|---|---|---|
| **14** | 0 | 1 | 1 |
| **15** | 1 | 7 | 8 |
| **16** | 4 | 16 | 20 |
| **17** | 7 | 47 | 54 |
| **18** | 126 | 64 | 190 |
| **19** | 0 | 950 | 950 |
| **Total** | 138 | 1085 | 1223 |

Table 2. Age and sex distribution from the non-censored dental age estimates.

| Estimated age (years) | Females | Males | Total |
|---|---|---|---|
| **14** | 1 | 0 | 1 |
| **15** | 2 | 4 | 6 |
| **16** | 5 | 24 | 29 |
| **17** | 10 | 73 | 83 |
| **18** | 27 | 114 | 141 |
| **19** | 82 | 398 | 480 |
| **20** | 11 | 472 | 483 |
| **Total** | 138 | 1085 | 1223 |

Table 3. Age and sex distribution from the censored dental age estimates.

| Estimated age (years) | Females | Males | Total |
|---|---|---|---|
| **14** | 1 | 0 | 1 |
| **15** | 2 | 4 | 6 |
| **16** | 5 | 24 | 29 |
| **17** | 10 | 73 | 83 |
| **18** | 118 | 114 | 232 |
| **19** | 1 | 855 | 856 |
| **20** | 1 | 15 | 16 |
| **Total** | 138 | 1085 | 1223 |

In the further analysis we apply the censored dental age estimates.

Cross tables of wrist and censored dental age estimates for females and males are given in Table 4 and Table 5, respectively. Also these numbers are rounded off so an age estimate of 13.5 is counted as 14 years of age, 14.5 as 15 years, etc. There are 113 females (81.9%) and 779 males (71.8%) with concurrent observations, i.e. no difference between the wrist and the dental age estimate.

Table 4. Cross table of wrist and censored dental age estimates for females.

| Females | | Dental age | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | 14 | 15 | 16 | 17 | 18 | 19 | 20 | Total |
| Wrist age | 14 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 15 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| | 16 | 0 | 0 | 0 | 2 | 2 | 0 | 0 | 4 |
| | 17 | 0 | 0 | 1 | 1 | 3 | 1 | 1 | 7 |
| | 18 | 1 | 1 | 4 | 7 | 113 | 0 | 0 | 126 |
| | 19 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | Total | 1 | 2 | 5 | 10 | 118 | 1 | 1 | 138 |

Table 5. Cross table of wrist and censored dental age estimates for males.

| Males | | Dental age | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | 14 | 15 | 16 | 17 | 18 | 19 | 20 | Total |
| Wrist age | 14 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| | 15 | 0 | 1 | 3 | 3 | 0 | 0 | 0 | 7 |
| | 16 | 0 | 0 | 4 | 4 | 5 | 3 | 0 | 16 |
| | 17 | 0 | 1 | 5 | 11 | 16 | 10 | 4 | 47 |
| | 18 | 0 | 0 | 1 | 9 | 20 | 26 | 8 | 64 |
| | 19 | 0 | 1 | 11 | 46 | 73 | 816 | 3 | 950 |
| | Total | 0 | 4 | 24 | 73 | 114 | 855 | 15 | 1085 |

Agreement in method comparison studies are often analysed inappropriately, notably by using correlation coefficients. The use of correlation is misleading (see first paragraph in section 3). We have applied the limits of agreement approach to this, also known as the Bland Altman method (Altman and Bland 1983, Bland and Altman 1986). This method quantifies agreement with the differences between observations made using the two methods on the same individuals. The 95% limits of agreement provide an interval within which 95% of the differences between measurements by the two methods are expected to lie. The interval is estimated by mean difference ± 1.96 x standard deviation of the differences.

# 3 Results

Firstly we give a short remark on the inappropriate use of correlation coefficients (Bland and Altman, 1986). These coefficients depend on the spread of the data and hence on the sampling. The main problem is that they totally ignore bias. For example, for the data in this study, the correlation between the skeletal and dental age estimates for males is r = 0.45. We create systematic bias by adding five years to all dental age estimates but the correlation is still r = 0.45 (Figure 1). It is possible to have a high correlation without agreement between the methods. Perfect agreement appears only if the points in Figure 1 lie along the line of equality, but we will have perfect correlation if the points lie along any straight line. The correlation coefficient is not a measure of agreement; it is a measure of association (Altman and Bland, 1983) and measures the strength of a relation between two variables.

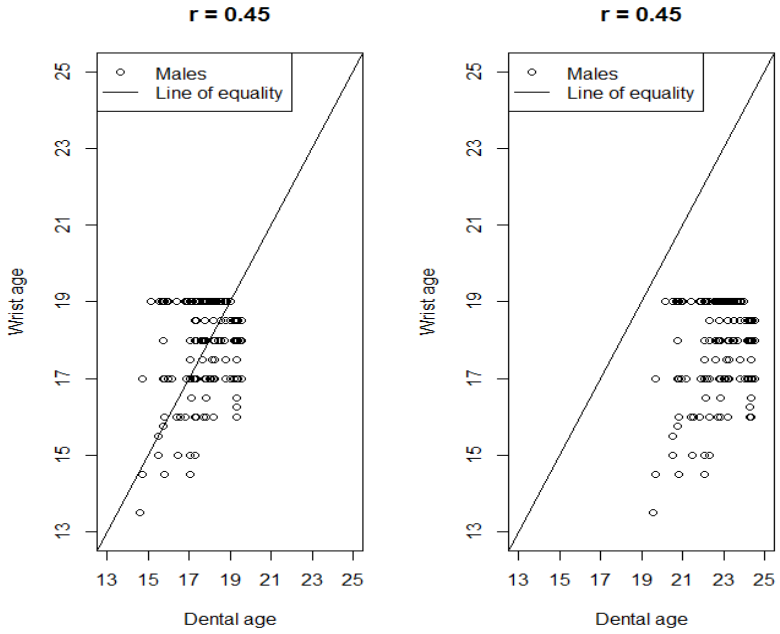


Figure 1. Effect of systematic bias on the correlation coefficient.

Summary statistics for the two age estimation methods, their difference (wrist age minus dental age) and their average are given in Table 6 and Table 7 for females and males, respectively. The location is represented with the mean and median whereas the minimum, maximum and standard deviation (SD) represent the spread.

Table 6. Summary statistics for females.

| | Wrist age | Dental age | Difference (wrist age minus dental age) | Average of wrist age and dental age |
|---|---|---|---|---|
| **Mean** | 17.86 | 17.80 | 0.06 | 17.86 |
| **Median** | 18.00 | 18.00 | 0.00 | 18.00 |
| **Minimum** | 15.00 | 14.20 | -2.50 | 15.00 |
| **Maximum** | 18.00 | 19.50 | 3.80 | 18.00 |
| **SD** | 0.48 | 0.69 | 0.74 | 0.48 |

Table 7. Summary statistics for males.

| | Wrist age | Dental age | Difference (wrist age minus dental age) | Average of wrist age and dental age |
|---|---|---|---|---|
| **Mean** | 18.76 | 18.71 | 0.05 | 18.76 |
| **Median** | 19.00 | 19.00 | 0.00 | 19.00 |
| **Minimum** | 13.50 | 14.57 | -3.31 | 13.50 |
| **Maximum** | 19.00 | 19.53 | 3.85 | 19.00 |
| **SD** | 0.69 | 0.73 | 0.74 | 0.69 |

Figure 2 shows scatter plots of the age estimates from radiographs of the wrist and the teeth for females and males, respectively, together with the line of equality. For females there are 113 observations at the point where both wrist and dental age are 18 years (see Figure 3). For males, 779 out of 1085 observations are concurrent (see Figure 3). The scales for the variables are the same. If the two methods gave exactly the same result for all individuals, all points would lie on the line of equality.

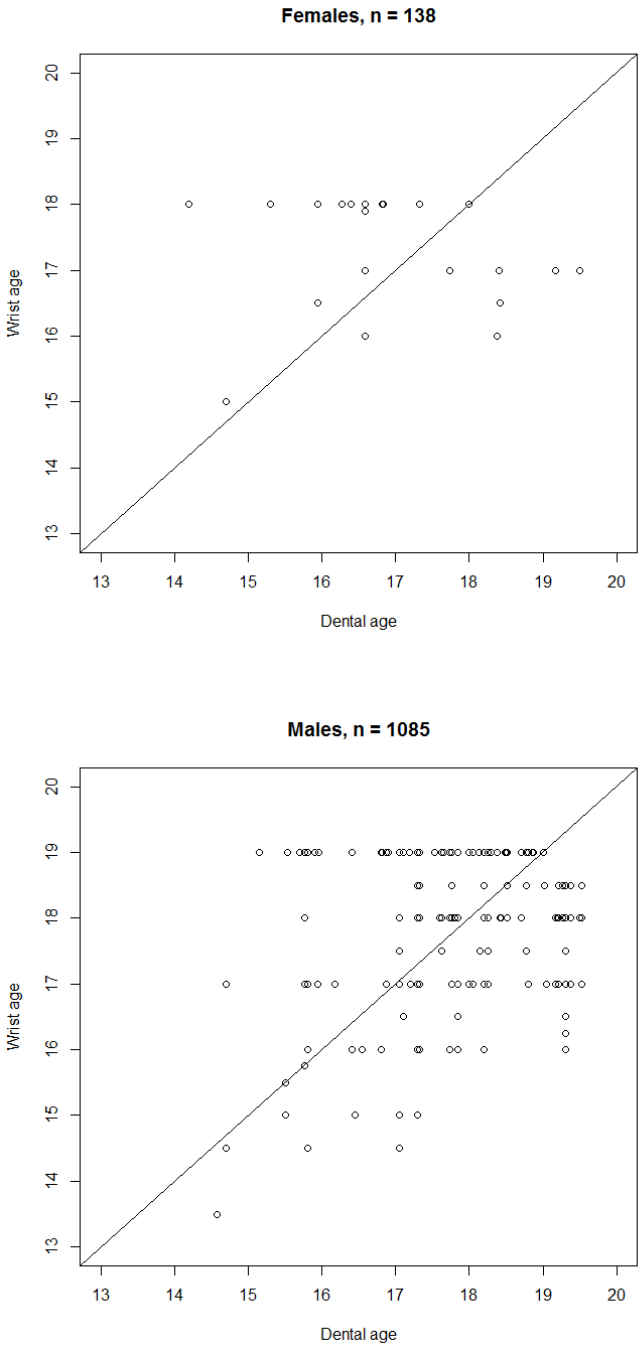


Figure 2. Scatter plot of age estimate from wrist and tooth for females (upper) and males (lower), together with the line of equality.

Figure 3 shows bar plots of the differences between the two age estimates for females and males, respectively. The number of observations corresponding to each bar is given in red above the bar. The number of concurrent observations is 113 females (81.9%) and 779 (71.8%) males. There are 18 females (13.0%) where the difference is more than one year, 7 females (5.1%) where the difference is more than two years and 1 female (0.7%) where the difference is more than three years. There are 153 males (14.1%) where the difference is more than one year, 39 males (3.6%) where the difference is more than two years and 14 males (1.3%) where the difference is more than three years.

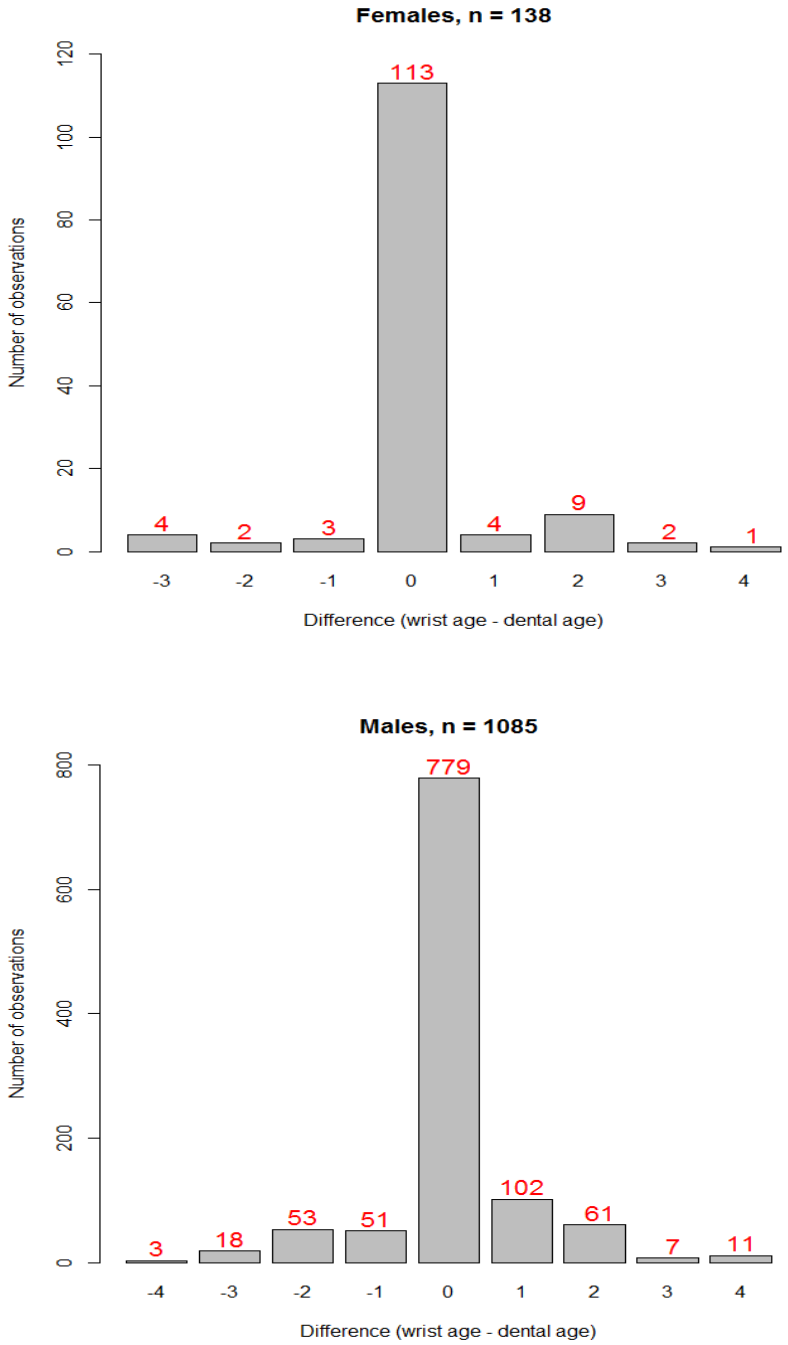


Figure 3. Bar plot of the difference between the wrist and dental age estimate for females (upper) and males (lower).

A plot of the difference between the methods (wrist age minus dental age) against their mean may be more informative than the plot in Figure 2. We do not know the true chronological age so the mean of the two measurements is the best estimate. Figure 4 and Figure 5 show the data from Figure 2 replotted in this way, for females and males, respectively. In the plots lines showing the mean difference and a 95% range calculated from the standard deviation are given. From these plots it is much easier to assess the magnitude of disagreement (both error and bias), find outliers, and see if there is any trend, e.g. an increase in the difference between the methods for high values.

From the points in Figure 4 and Figure 5 we see an association between the differences and the size of the measurements. The scatter of the differences increases as the age estimate increases, both for females and for males. The correlation of the individual differences with the average value is 0.42 (p-value < 0.001) and 0.49 (p-value < 0.001) for females and males, respectively. A log transformation of the data does not change the relation between the difference and the mean. With independence the methods may be compared by analysing the individual wrist age minus dental age differences. The mean of these differences will be the relative bias, and their standard deviation is the estimate of error. With a relation between the difference and the mean this method will tend to give limits of agreement which are too far apart rather than too close, and so should not lead to the acceptance of poor methods of measurement (Bland and Altman, 1986).

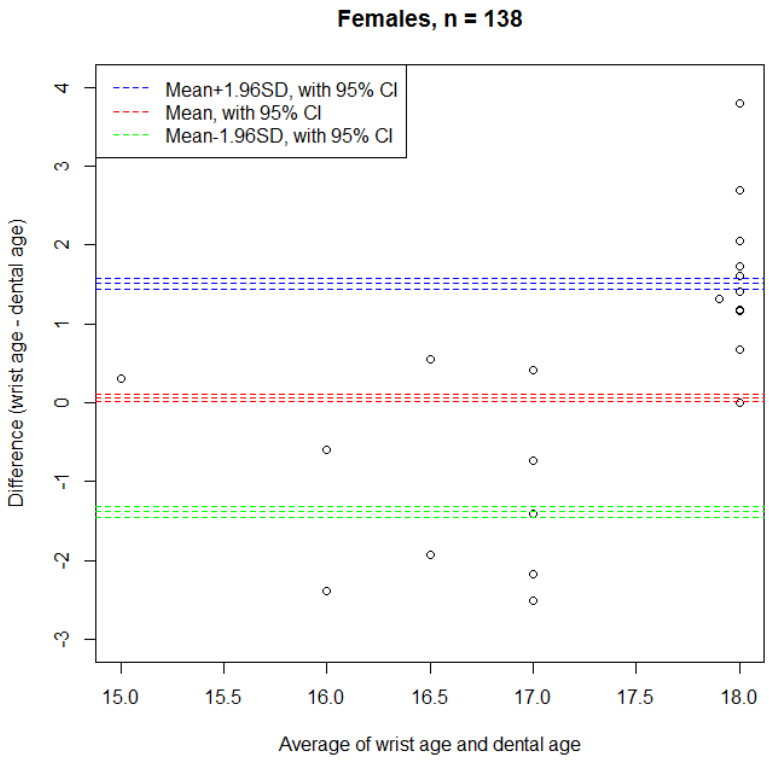


Figure 4. Difference (wrist age minus dental age) against the average of the two age estimates. 113 observations for females are concurrent.

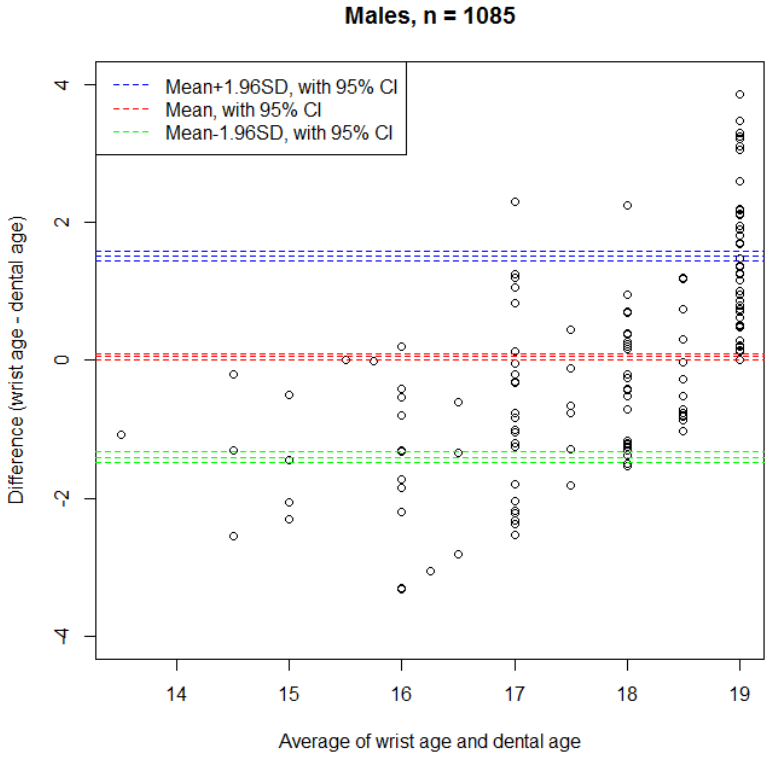


Figure 5. Difference (wrist age minus dental age) against the average of the two estimates. 779 observations for males are concurrent.

From Figure 4 and Figure 5 we discover discrepancies between wrist age and dental age of -2.5 years and 4 years for females, and between - 3.5 years and 4 years for males. These differences are not obvious from Figure 2. The mean difference between wrist age and dental age for females is 0.06 with 95% confidence interval -0.06 to 0.19 and the mean difference is 0.05 for males with 95% confidence interval (0.01, 0.10). With a 5% significance level we cannot reject the null hypothesis that the true difference in means for the two age estimates for females is equal to zero (since zero is included in the 95% confidence interval). The hypothesis is rejected for males but zero is barely outside the confidence interval and the differences given in this interval are very small. For most of the individuals the two age estimates are in agreement and large discrepancies are uncommon. There are no significant differences between these estimates, neither for females nor for males.

Provided that differences within mean difference ± 1.96 x standard deviation of the differences would not be clinically important, skeletal and dental age estimation methods may be used interchangeably for differences within the interval (-1.4, 1.5) both for females and for males. These are the limits of agreement for the differences between skeletal and dental age estimates. They are highlighted in green and blue, respectively, in Figure 4 and Figure 5.

The precision of estimated limits of agreement can be illustrated with standard errors and confidence intervals, provided the differences follow a distribution which is approximately Gaussian. Figure 6 shows histograms of the differences between the wrist and dental age estimates together with the estimated normal distribution curve, females at left and males at right. There are some deviations from the normal distribution but we assume that the

approximation is good enough. 95% confidence intervals for the mean difference and for the limits of agreement are given in Figure 4 and Figure 5. We see that these intervals are very narrow, reflecting the small variation in the differences both for females and for males.

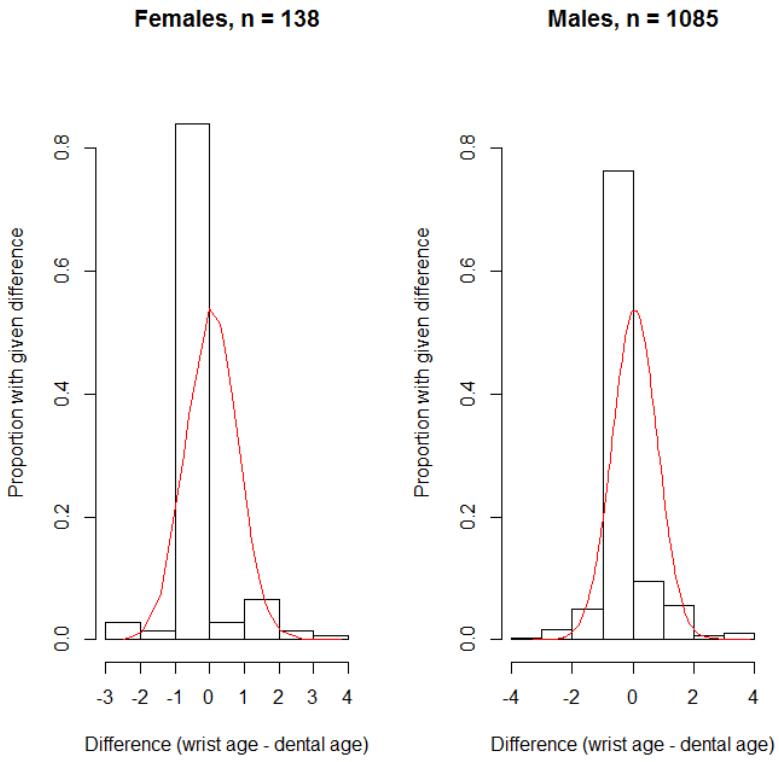


Figure 6. Histogram of differences (wrist age minus dental age) together with estimated normal distribution curve.

There are 25 females (18.1%) where the difference between the wrist and dental age estimate is different from zero. The mean difference for these females is 0.35 with 95% confidence interval -0.36 to 1.07. The mean difference for the 306 males (28.2%) where the difference is different from zero is 0.19 with 95% confidence interval 0.03 to 0.35. Again, zero is included in the confidence interval for females and barely excluded in the interval for males. With 5% significance level, there are no significant differences between the two age estimates for females. There are small significant differences between the age estimates for males.

Figure 7 shows the age estimates for the 25 females with a difference between the wrist age (given in blue circles) and the dental age (given in red triangle point-up). The value of the age estimate is given on the leftmost y-axis. The differences between the wrist and dental age estimate are given as the grey vertical bars and the value of these differences are specified on the rightmost y-axis. A positive value means that the wrist age is higher than the dental age and vice versa. The number of positive grey bars is almost twofold the number of negative grey bars. The wrist age estimate is higher than the dental age estimate for 16/25 = 64.0% of the females. The maximum difference in absolute value is 3.8 years, belonging to a female with a wrist age estimate of 18 years and a dental age estimate of 14.2 years (observation number one in Figure 7).

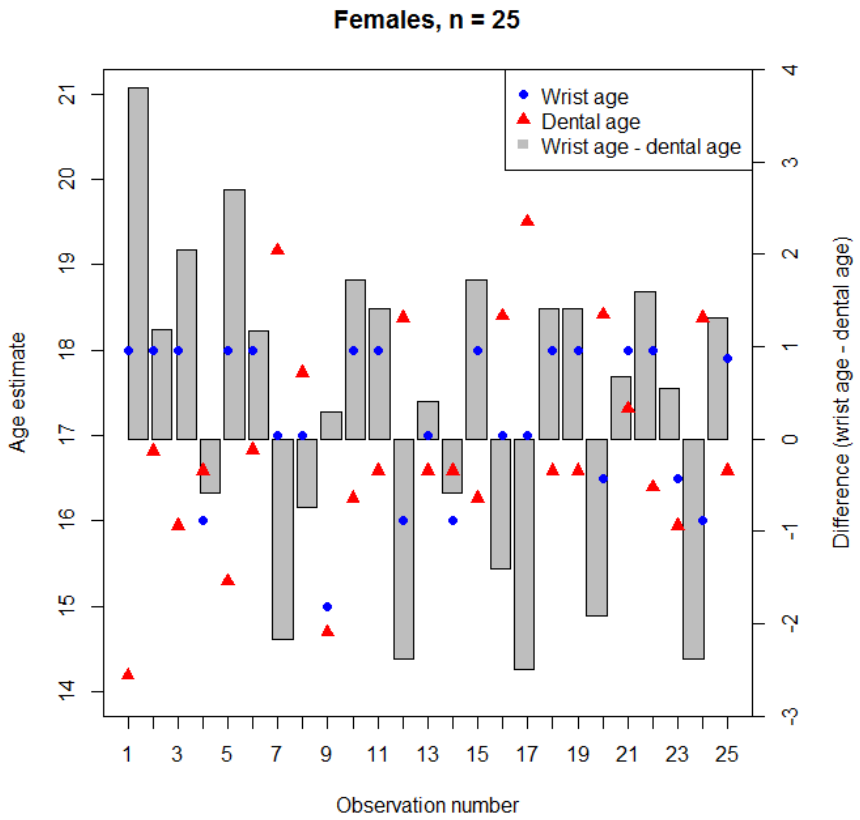


Figure 7. Two age estimates together with their difference (wrist age minus dental age) for females.

There are 306 males (28.2%) with a difference between the wrist age and the dental age estimates. To illustrate the differences in the same manner as with females, 30 of these observations are drawn at random and plotted in Figure 8. Again, the number of positive grey bars is almost twofold the number of negative grey bars meaning that the wrist age estimate is higher than the dental age estimate for these males. If we plot Figure 8 for all males and count the positive and negative grey bars, the wrist age estimate is higher than the dental age estimate for approximately 59.2% of the males. The maximum difference in absolute value for the plotted data is 3.24 years, belonging to three males with a wrist age estimate of 19 years and a dental age estimate of 15.76 years (observation number 8, 10 and 22 in Figure 8).

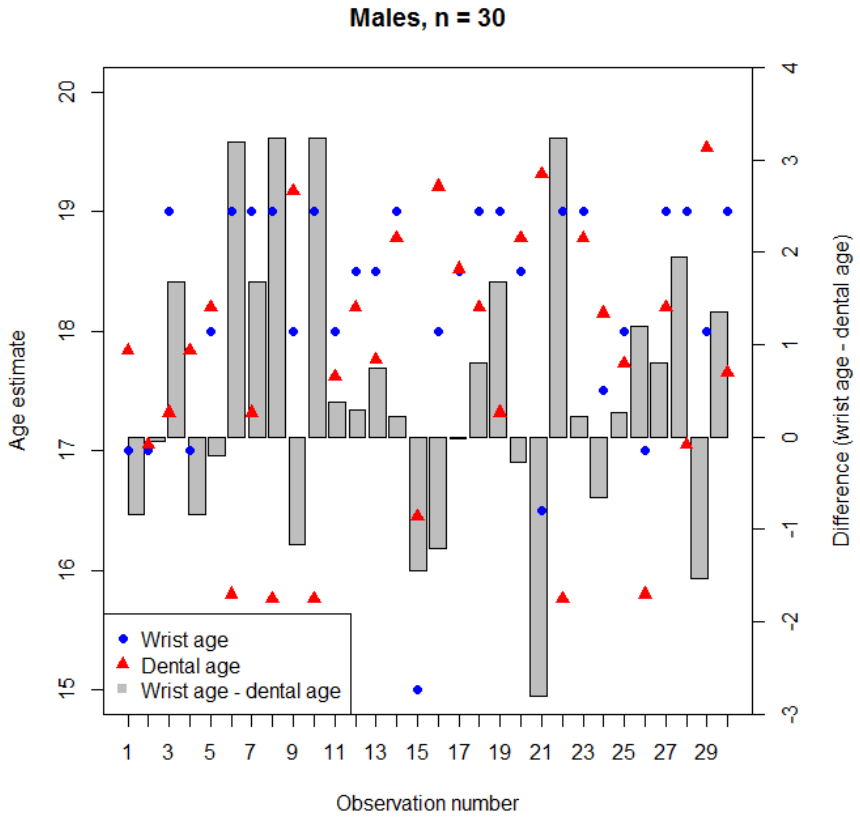


Figure 8. Two age estimates together with their difference (wrist age minus dental age) for males.

# 4 Discussion

There are no significant differences between the age estimates from radiographs of the wrist and age estimates from radiographs of the teeth for females. The differences found for males are very small and barely significant. There are many concurrent observations (113/138 ≈ 81.9% for females and 779/1085 ≈ 71.8% for males). We conclude that these two age estimates are in agreement.

From the limits of agreement approach, skeletal and dental age estimation methods may be used interchangeably for differences within approximately ± 1.5 years, both for females and for males.

The assessment of repeatability is an important aspect of studying method comparison because the repeatability of the methods limits the amount of agreement which is possible. The agreement between two methods is bound to be poor if one method has poor repeatability (Bland and Altman, 1986). There were no repeated observations in our data so this aspect could not be analysed.

As mentioned in Section 2, when skeletal growth has ceased a radiograph examination cannot tell anything more than that a girl is 18 or older or that a boy is 19 or older. For individuals with this maximum value and a higher estimate for dental age (81.9% of all females and 71.7% of all

males), the dental age estimate is censored at 18 and 19 years, respectively. This is the best way to handle these data since they are in agreement, i.e. both estimates say that the individual is 18 or older. But for individuals with this maximum value for wrist age and a lower estimate for dental age (8.7% of all females and 13.6% of all males), the difference between the two estimates may be underestimated since a radiographic examination cannot tell anything more than that a girl is 18 or older or that a boy is 19 or older.

# 5 References

Altman D.G., Bland J.M. (1983) Measurement in medicine: the analysis of method comparison studies. The Statistician; 32, 307-317

Bland J.M., Altman D.G. (1986) Statistical methods for assessing agreement between two methods of clinical measurement. The Lancet; 327, 307-310

Liversidge HM. (2008) Timing of human mandibular third molar formation. Annals of Human Biology; 35(3), 294-321