



## Review Article

DOI: 10.36959/763/519

# Is the Demirjian, Goldstein, and Tanner Method of Dental Age Estimation Obsolete? A Critical Review and Re-Assessment

Graham Roberts, MDS\* and Victoria S Lucas, PhD

Department of Orthodontics, King's College London Dental Institute, UK



## Abstract

The Demirjian technique of Dental Age Assessment has been used since 1973 to carry out Dental Age Estimation (DAE). Few authors appear to be aware that the technique comprises two distinct parts. The Description of the Demirjian Tooth Development Stages referred to here as TDS, and the management and integration of the numerical data derived from these TDS to provide a Dental Age (DA). This will be referred to as the Demirjian, Goldstein, and Tanner Method (DGTm). The Tooth Development stages described by Arto Demirjian DTDSs are the most reliable and consistent system of assessing dental development for the purposes of Dental Age Estimation (DAE). The DGTm of DAE has been shown to be inaccurate by up to 3 years, and has demonstrated poor reliability in estimating dental age when compared to the Gold Standard of Chronological Age (CA).

## Keywords

Forensic science, Age estimation, Dental age estimation, Tooth development stages, Mathematical integration, Between rater agreement, Within rater agreement, Forensic odontology, Age at assessment

## Introduction

The role of teeth in forensic age estimation is unequalled in their ability to provide reliable estimates of the age of infants, children, and adolescents [1]. A recent systematic review of the Demirjian 'method' [2] assessed 274 studies [3]. This is a considerable number and is compelling evidence of the negative impact that the 'Demirjian Method' has had on research in Dental Age Estimation (DAE). The main finding of this systematic review paper is that the French-Canadian dataset "... overestimates the age of subjects by more than six months ..." [3]. This is not a suitable degree of accuracy for forensic age estimation. Although the present authors are clear about the '... more than six months ...' being a level of accuracy that is insufficiently accurate, no author has yet defined a level of accuracy that is acceptable for Dental Age Estimation (DAE) in a forensic context. A CA *minus* DA difference of up to 1 year has been described as a '... maximum acceptable difference ...' with a CA *minus* DA difference of up to six months as a '... stringent ...' [4]. This is the first time that a CA *minus* DA difference has been stipulated in a research paper as an acceptable level of accuracy. It is unclear whether this is the result of a forensic legal decision determining plus or minus six months as acceptable or an arbitrary choice by the authors. It is too wide a margin when compared with recent publications which demonstrate a CA *minus* DA difference, on average, of only 0.3 of a year (4 months) for males and 0.12 of a year (1½ months) for females [5]. A further problem with using the DGTm is that it is difficult to provide an expression of the uncertainty of the age estimate other than by visual-

ly reading off percentile values from the graphs presented in the original paper [2].

It is important to understand the precise terminological definitions to aid the comprehension of this review. These are given in Table 1.

A detailed assessment of the DGTm raises the question "can the 'DGTm be relied upon to accurately estimate the age of subjects of unknown date of birth?" Further is it possible to estimate the level of uncertainty [6] using the information provided in the original paper? A detailed analysis of the Age at Assessment (AaA) for the Demirjian tooth formation stages (TDS) shows differences of over a year for individual stages across different studies. The conclusion to this article is "These results suggest no major differences in the timing of tooth formation stages ..." [7]. These seemingly disparate views are difficult to reconcile. Differences of up to a year should be properly accounted for and should not be dismissed as "...groups in this study do not differ greatly..."

**\*Corresponding author:** Graham Roberts, MDS, Department of Orthodontics, King's College London Dental Institute, Floor 25 Guy's Tower Great Maze Pond, London, SE1 9RT, UK

**Accepted:** February 02, 2021

**Published online:** February 04, 2021

**Citation:** Roberts G, Lucas VS (2021) Is the Demirjian, Goldstein, and Tanner Method of Dental Age Estimation Obsolete? A Critical Review and Re-Assessment. *Insights Anthropol* 5(1):325-336

**Table 1:** Definition of terms used in this review. The most important terms to be differentiated are the DTDS - the Demirjian Tooth Development Stages and the DGTm - the mathematical integration of the summary statistics of each of the TDS present in a single individual.

Term(s)	Definition	Explanatory note
Age at Assessment (AaA)		
Demirjian Tooth Development Stages (DTDS)	These are the stages defined by Arto Demirjian in 1973 [2].	There is poor separation of the terminology for the Demirjian TDS and the Mathematical Integration of the data from the TDS to derive the percentile curves.
Demirjian, Goldstein, and Tanner Method (DGTm)	This is the process of Mathematical Integration used to create the percentile curves from which the Dental Age (DA) is derived.	In this review these will now be differentiated by using DTDS when referring to the Tooth Development Stages. The mathematical integration of the data from the TDS is referred to as the DGTm
Demirjian ‘method’	The term used loosely in the dental and forensic literature to refer to the process combining the DTDS and the DGTm.	This term is (or should be) redundant as it leads to confusion about the processes that are involved in estimating the DA of a single subject.

[7]. More recent studies show a large difference from appropriately censored data for Stage H in Southern Chinese males resident in Hong Kong compared to UK Caucasians living in the London area [8]. The important feature of this study is that the methods of assessment used are identical [8].

The population differences shown in the two systematic reviews of the outcomes of the population data using the Demirjian Method [3,9] need to be properly accounted for. A problem which the 1973 authors might take issue with is the widespread criticism of the ‘...inconsistent age estimations on different ethnic and geographic population groups [3]. A re-reading of the original paper [2] reveals that the authors were aware of the potential for separate population groups to give different estimates of dental maturity. It is helpful to quote the exact words from the paper “... *In using the maturity standards ... it should be remembered that the sample on which they are based is entirely of French-Canadian origin. The dental maturity scores for a given chronological age may well be greater or less in other populations, according as to whether they are more or less dentally advanced during growth.*” [2]. This statement clearly demonstrates the original authors’ view that specific ethnic or ancestral derived data are needed to reliably assess dental maturity and conduct DAE in populations other than the French-Canadian children used in the original study. Thus criticisms made about the unsuitability of the Demirjian standards in relation to different ancestral-ethnic groups whilst technically correct are misdirected and unfair to the spirit of the original authors who indicated that specific population data should be used to estimate dental maturity for a subject from an appropriate ancestral population other than French Canadian subjects. This appears to have been overlooked by the numerous authors, who have, as an example of the genre, asked the question “... are Demirjian’s standards applicable?” [10]. This is an inappropriate question given the original author’s clearly expressed view about the need for ‘...other populations ...’[2].

A recent systematic review of 274 papers using the ‘Demirjian Method’ reported wide variation in the CA *minus* DA difference using the Demirjian Method [3]. This systematic review showed that most of the CA *minus* DA differences were of the order of 0.65 years for females and 0.60 years for males (females, minus 0.10 years to plus 2.82 years; males, minus 0.23 years to plus 3.04 years). It is surprising to discover, nearly 50 years since the technique was published, that there has not been a systematic study to validate the reliability of the DGTm based on the specific ancestral group of French Canadians.

It is the purpose of this paper to review both the Demirjian Tooth Development Stages (DTDS) and the Demirjian, Goldstein and Tanner Method (DGTm) to determine their suitability for the forensic and archaeological procedure of Dental Age Estimation.

Preliminary Considerations

The ‘Demirjian, Goldstein and Tanner Method’ comprises 5 principle elements that are used to derive data which is then used to estimate the age of a subject of unknown date of birth.

1. Choice of the Reference Population.
2. Selection of a suitable sample to provide a Reference Data Set (RDS) that enables ethnic specific age estimations to be conducted.
3. Assessment and identification of defined Tooth Development Stages (TDS) for the 7 lower permanent teeth excluding the third molar (stages A to H).
4. Integration of the numerical data from the Ages at Assessment (AaA) to produce summary statistics and the potential role of Censoring.
5. Attachment to these weighted scores a value for the uncertainty associated with the age estimates derived based

on the percentile curves published [2].

This process for DAE has been frequently used since the publication of the original paper in the early 1970's [2]. A modification of the '... system of dental maturity ...' was published a few years later by the original authors [11] with the advice that '...the second scoring system should be used...' [Demirjian 1985 personal communication] and '... I would hope is the one people have been using...' [Goldstein 2015 personal communication]. This endorsement of the 1976 scores is difficult to accept as in the 1976 paper there is no evidence to show that the new systems estimate age more accurately than the 1973 systems. For this reason no further consideration will be given to the paper published in 1976 [11]. Indeed, the 1973 weightings using the Canadian data were not validated in the first place and neither were the weightings for the 1976 system validated. Attempts to determine the most accurate DGTM using 7 or 4 teeth show differences of up to half a year [12]. It is clear that the variants of the DGTM do not contribute in a helpful way to the reliability of Dental Age Estimation in modern times.

This issue of validation of Reference Data Sets is a concept that is not applied sufficiently widely. Most assessments of the reliability of DAE methods seem to limit themselves to an assessment of the AaA of individual TDS. This is misguided as the DA is usually determined by the averaging of the AaA of all developing teeth [1]. The validation of a DAE technique is an essential step in the development of an age estimation method as it is the only indicator of the accuracy of any method in terms of the reliability of the estimation of chronological age by using dental age as a proxy. The first clear use of this approach was in the paper published in 2001 that attempted to provide an alternative scoring system [13] but still using the methodological process of the Demirjian, et al. 1973 paper [2].

The purpose of this critique is to investigate the Demirjian TDS and the Demirjian, Goldstein and Tanner Method of Dental Age Estimation with a view to determining their suitability for use as a method for estimating the age of subjects of unknown date of birth.

To do this it is necessary to disassemble or unravel the procedures involved in the DGTM into their component parts and to assess the contribution that each part makes to the process of Dental Age Assessment. Further, to determine the appropriateness of the component procedures and their potential contribution to the reliability of the process of Dental Age Estimation in a logical manner.

## Methods

The data for this study are re-used from papers published over the last 50 years thus ethical approval is not required.

The publications reviewed for this study were obtained by electronic searching from PubMed and Embase and the Cochrane databases. Articles published in English between May 1966 and April 2020 were searched. The information obtained from the papers was data on the Tooth Development Stages and the processes of integrating the data to achieve Dental Age Estimation (DAE) outcomes.

The inclusion criteria were such that paper titles and/or authors that included the word Demirjian were retrieved in pdf format. All papers selected were examined for details as to how the Tooth Development Stages (TDS) were defined, the way in which the numerical data array for each TDS was presented, the process by which the summary data for each TDS was weighted (or managed) to give the tooth score(s) that contributes to the total score for an individual, and the way in which this weighted score was converted to Dental Age.

## The reference data set (RDS)

It is widely recognised that the choice of the RDS underpins the suitability of growth related data for Dental Age Estimation (DAE) [1,14]. The choice of an RDS consists of five elements.

**An ancestral or ethnically homogenous group of human subjects:** This was achieved in the studies on Dental Maturity by the Anglo Canadian research team that developed the Demirjian, Goldstein, and Tanner Method (DGTM) of Dental Age Estimation [2]. It is clearly stated in the Materials and Methods section all children had '...parents and grandparents...' of French-Canadian origin. This is entirely appropriate, and it is difficult to understand why later studies by several authors across the globe have departed from this very clear example of an appropriate specific ancestral composition of the sample. It is alarming that investigators have, over the almost 50 years since the development and publication of the DGTM have ignored the very clear guidance on the need for using specific ancestral RDS [2]. Support for the use of ethnic specific RDS is provided by the test of using the United Kingdom Reference Data Set for estimating the age of Southern Chinese children and adolescents. As was anticipated, it was shown that there is a consistent under-estimation of the age of Southern Chinese subjects [15]. A later study demonstrated that use of an ethnic specific RDS improved the accuracy of DAE [16].

**An even balance between the female and male subjects:** The paper reports the sex distribution as 1,482 girls and 1446 boys from age 2 years to 19 years (see [Table 1](#) of the 1973 paper) [2]. Although there are small differences between the numbers of females and males in each stratum, it is unlikely that these have biased the results.

**Even number of subjects within each age band:** It is clear that although there is a slightly uneven distribution of females and males across the age range from 2 years to 19 years the unevenness is not so great as to materially affect the outcome variables in terms of percentiles which lead to the scores for the individual TDS. Perhaps the most important feature is to have *sufficient* numbers in each age band.

A corollary to this is that there is always an apparent mismatch between the numbers of a given TDS within an age band and the even distribution of chronological age across the age bands.

An underlying difficulty with all DAE studies is that a random sample from a specific ethnic population is not possible. This is the consequence of research ethics committee's un-

willingness to permit studies on dental development based on radiographs taken for the purposes of research. This is to avoid the small risk of radiation induced illness in child and adolescent patients. Thus all studies use extant radiographs from clinical care databases. This is known as a 'convenience sample'. It is widely regarded that such convenience samples from healthy paediatric patients, orthodontic patients and young adults with wisdom tooth problems are essentially normal and exhibit normal growth and development provided clinical or pathological anomalies result in exclusion of such individuals.

**Sufficient numbers of subjects across the age range of development:** This is also well managed although the greatest numbers in clinical databases and research publications are found in the 6.5 years to 14.5 year age range the lower and upper extremes are still well represented. Sample Size Estimates in published work usually recommend a number above 10 for each TDS by gender and ethnicity.

**Cross-sectional versus longitudinal study sample:** The radiographs used to compile the RDS should be unique in the sense of each subject comprising the sample is represented only once. That is to say the data should be cross-sectional in nature [17]. The reason for this is that extensive studies have shown in somatic growth that individual subjects exhibit what is known as 'canalization' [18]. In summary this means that an individual child grows along a genetically predetermined path or 'percentile'. Put simply, short children usually grow into short adults and tall children into tall adults. It is reasonable to assume teeth grow in a similar manner. Thus, combining longitudinal data with cross-sectional data makes it difficult to interpret the values obtained for the mean and standard deviation of any reference data. It is not stated in the 1973 study [2] whether or not mixed cross-sectional/longitudinal data are used.

Nevertheless, it has been confirmed that data from the 1973 [2] study is *mixed longitudinal/cross-sectional* in nature. A later paper from the Canadian team, drawing on the radiographs from the earlier study, reported that the radiographs were from mixed longitudinal and cross-sectional samples [19]. It thus becomes difficult to interpret the findings in a fully meaningful way. The underlying difficulty is that most subjects are counted once, some are counted twice, and others up to five times. As can be seen in Table 1 of the paper on sexual dimorphism by Levesque and Demirjian there are larger numbers of subjects contributing longitudinal data than cross-sectional data [19].

This makes it impossible to interpret the associated data in terms of the median or mean value for AaA and the true biological variation around that mean AaA. A detailed discussion of this problem is given in the textbook by Professor Jack Tanner [17].

## Sampling from the RDS

To derive appropriate Reference Data Sets it is necessary to use samples comprising a single radiograph per individual, i.e. Cross-Sectional sampling. This is the only effective way to eliminate the possible misleading effect of canalization on the summary statistical values for each of the TDS with-

in the selected Reference Data Set. Examples of an inappropriate mix of Cross-Sectional and Longitudinal data are given in relation to third molar studies [20]. It is difficult to know *how* such mixed data should be interpreted or even if it can be interpreted in a meaningful way. Given the uncertainty associated with such a sampling frame it is strongly recommended in all future DAE studies only an RDS compiled using a verified cross-sectional sampling technique is used. This difficulty of mixed Cross-Sectional data and Longitudinal data occur elsewhere in the DAE literature where Longitudinal and Cross-Sectional mixed data have been used [19].

A further issue is the question of 'age-mimicry' of the sample [21]. This is a term derived from the Archaeological literature. There is some ambiguity about the appropriateness of this term. In essence it appears most archaeological studies on 'age estimation' are based on samples whose demographic data is not known or is unreliable. It is a complex area of age estimation where sophisticated mathematical formulae are applied to data where there is a need to "...avoid age mimicry - the contamination of our age estimates by the age composition of the reference sample?" It is necessary to question this statement. By definition the RDS provides the data from which DAE is conducted.

The underlying problem of a lack of an appropriate Reference Data Set has been intruded into modern DAE inappropriately [21]. The authors have failed to understand that in almost all modern DAE studies the RDS comprises subject of *known* chronological age. This is the Gold Standard for any RDS as data are derived from clinical records where the date of birth and the date of assessment (the date when the radiograph was acquired) are known and verifiable. In this circumstance, the reliable estimation of the AaA, the summary statistics of the sample reflect the ages comprising the sample. It is inappropriate to refer to this Gold Standard of the ages comprising the RDS as 'age mimicry'. It is a major problem for archaeological samples where doubt exists as to age of the individuals comprising the sample. For Dental Age Estimation studies, where the date of birth and date of assessment of each subject in the sample are known, it is now the Gold Standard for Age Estimation Studies. Importantly the validity of DAE studies has been tested by comparing the *known* chronological age with the *estimated* dental age it has been possible to reliably assess the value of the DAE method under consideration. This technique of validating the reliability of the RDS was first reported from The University of Leuven, [13] and later by data on children from London UK [22].

Although the DGTm returns impressive numbers, it fails almost catastrophically with regard to the mixed longitudinal/cross-sectional nature of the sample. To disentangle the data at this time, nearly 50 years after publication of the data, is likely to be impossible.

## The tooth development stages

The Tooth Development Stages utilized for DAE are at the core of the DAE methods which calculate AaA data for each TDS. The number of TDS used by different authors varies from 4 to 21 stages [1]. This is a hugely disparate number. In practice, the most common TDS system are those of the 8 Stage



system of Demirjian, [2] and the 14 (sometimes 15) stage system described by Moorrees and colleagues in 1963 [23]. These two staging systems are the most widely used.

Intuitively it is felt that a large number of TDS would lead to smaller steps between the stages and thus, when the reference data are integrated from the TDSs the resultant age estimation will be more accurate. This has not been explored systematically but a comparison of the age estimation using Demirjian TDS (8 stages) and the Haavikko TDS [24] (12 stages) showed there was no difference in the outcome variable of Dental Age when comparing 80 females and 98 males [1].

The possible explanation for this is that the number of incorrect stage assignments is greater with TDS systems where the number of stages is higher. The issue of potential inaccuracies using the DTDS was investigated in detail, [25] and showed that ‘... nearly all discrepancies ... encountered are of one stage only, and ... closer study of such cases reveals that they can be ... totally attributed to inscription errors [25].’ Although at that date the use of specific statistical methods to assess agreement were not commonly used, it is clear that the DTDS provide very high reliability or ‘within’ and ‘between’ rater agreement. Later studies assess Inter-rater agreement in a formal statistical sense using the Kappa statistic described in 1977 [26]. A study in Germany, [27] and the UK [28] showed that the DTDS gave the best results for Kappa as “Having clearly defined stages and fewer intermediate stages allowed better reproducibility.”

There seems little doubt that the DTDS provide the best approach to assessing and assigning stages. It is, perhaps, the high reliability of this part of the process that has encouraged uncritical investigators to assume that the DGTm is also reliable.

## Numbers in the Reference Data Set and the Numbers for Individual Tooth Development Stages

The RDS should comprise sufficient numbers over the whole range from 4 to 24 years to ensure adequate numbers in each one year age band *and* adequate numbers for each Tooth Development Stage).

These numbers are at two levels:

- i. The number of subjects - referred to as N.
- ii. The number of TDSs referred as n-tds.

This nomenclature assists readers in identifying whether a particular summary statistic refers to the total number of subjects (N), or the total number of a specific TDS (n-tds) [8]. This is important as potentially there are 288 data sets for each tds. The use of the lower case -tds extends to all the summary statistics such as  $\bar{X}$ -tds, sd-tds, and se-tds, and also to the percentile summary statistics.

This is relevant because the usual N is of the order of several hundred subjects. This creates a favourable impression to readers but masks an inherent weakness with most DAE studies. The number of subjects for each TDS by gender usually reduces the n-tds to figures of less than 10.

The number of subjects per age group is important to give the dental age assessor confidence in utilising the RDS. In the original paper there are 2,928 subjects and in general terms each single year span for age groups has over 100 subjects. This is an excellent number [although the use of serial radiographs mixed in with this sample diminishes its value]. As time has passed there has been a trend to use smaller samples but intuitively these are seen as less reliable than adequate numbers for n-tds in each RDSs. This feature of large numbers in the original paper is an excellent yard stick and where-ever possible investigators should attempt to approach this number [1]. A further difficulty with small numbers for n-tds is the use of Student's t test to demonstrate that there are no significant differences therefore justifying a conclusion of misleading comparability leading to combination of disparate ethnic groups. This is an inappropriate test for this purpose and is especially misleading with small n-tds. The very nature of the calculation with small n-tds will lead to the conclusion of no -significant difference [10]. This is inappropriate as “... worse agreement decreases the chance of finding a significant difference and so increase the chance that methods will appear to agree!” [29].

## The radiographs should provide a clear and detailed view of all tooth morphology types (TMTs)

The radiographic technology available at the time of the original paper (1973) was less technologically advanced compared to modern machines. The consequence of this is that upper teeth were less easily visualised than lower teeth. For this reason the DGTm technique was limited to lower teeth only. As improvements have been made it has become possible to obtain good quality images of all the TMTs. This has given greater opportunities to use up to 16 teeth in the process of DAE [5]. The DGTm approach has been to limit the number of teeth to seven, viz all those in the lower arch excluding the third permanent molar. It is ironic that since 1993, there have been numerous publications to determine the applicability of the third molar for age estimation at the 18 year threshold.

Modern Rotational Tomographic Techniques provide satisfactory images of all TMTs and have increased the applicability of the Demirjian Tooth Development Stages to DAE by enabling inclusion of all 16 TMT's on the left side.

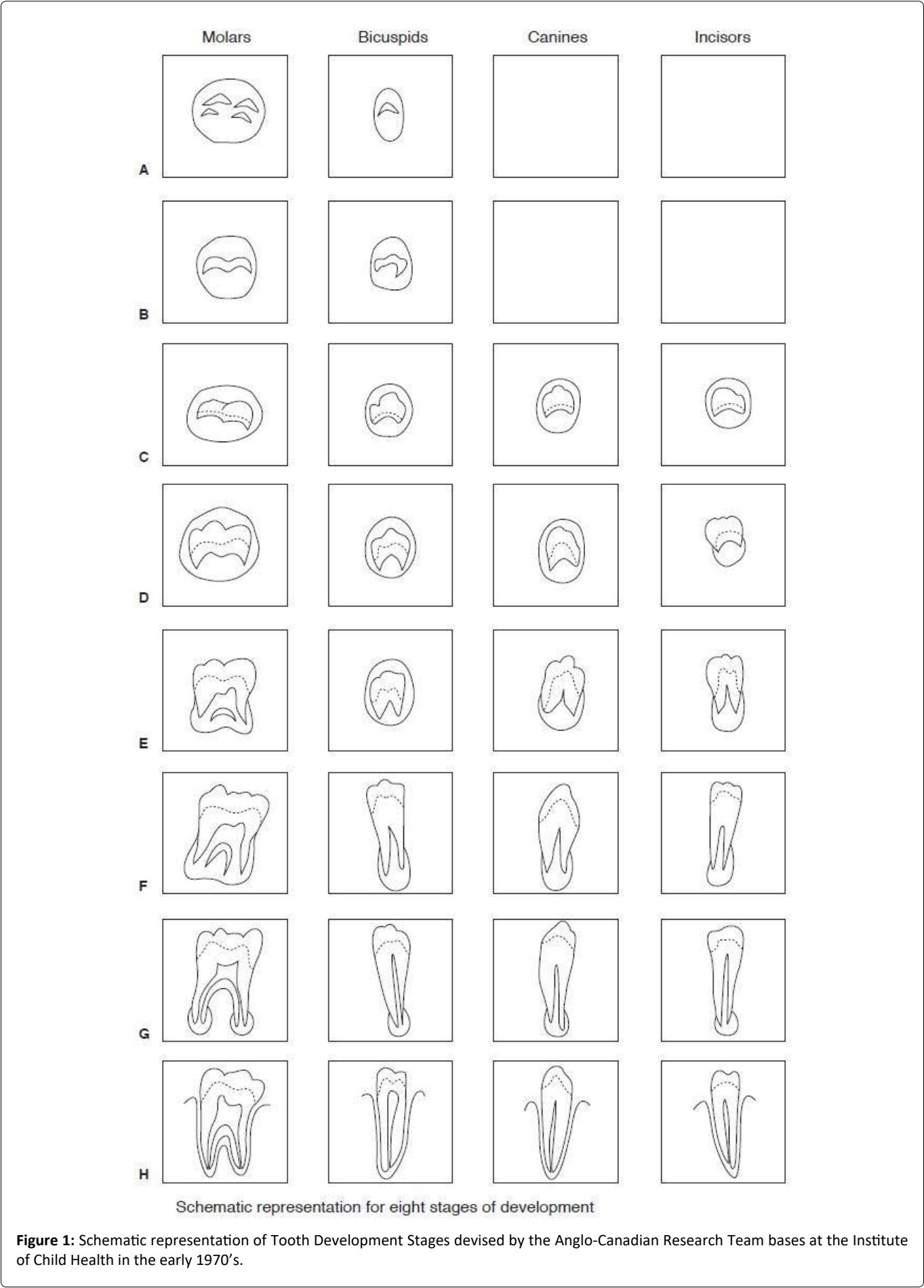
It is pertinent to draw attention to the development of other imaging techniques particularly the use of Cone Beam Computed Tomography (CBCT, and Magnetic Resonance Imaging (MRI).

The former creates greater exposure to ionising radiation and the latter has insufficient Reference Data Sets for general applicability.

The TDS used should be clearly defined. The DARLInG team at King's College London has selected the 8 stages defined by the team in Montreal, Canada (Demirjian, Goldstein and Tanner 1973).

## Application of different tooth development stage systems

The use of schematic diagrams with accompanying de-



tailed written descriptions of the Tooth Development Stages were devised by Arto Demirjian during a sabbatical from Montreal at the Institute of Child Health, University of London in the early 1970's. The principal then adopted was to emulate the approach used in Skeletal Maturity Assessment using a system of anatomically defined developmental stages published in book form in 1975, and now in its 3<sup>rd</sup> edition [30]. This system relates skeletal development to data arrays for each anatomically defined bone stage. It has been a standard approach for many years.

The same principle was utilized for the development of Tooth Development Stages. There were already available a set of TDS 'descriptions' comprising 13 stages for single rooted teeth and 14 stages for two rooted teeth [23] which were based on the descriptions derived to describe the development of the permanent first molar [31]. These descriptions were limited to ten permanent teeth i.e. all the mandibular teeth and maxillary central and lateral incisors. A cursory examination of these fourteen stages immediately raises the question as to how easy it is to discriminate between the stages. The descriptions of the stages given in the 1963 paper are limited and clearly rely on the observer's guesstimate of the length of the root e.g. 'Root Length  $\frac{1}{4}$  {R $\frac{1}{4}$ }' compared to Root Length  $\frac{1}{2}$  {R $\frac{1}{2}$ } [23]. A few moments attempting to apply these simplistic criteria when assessing a clinical radiograph indicate the difficulty of discriminating between the two stages. It is of note that these earlier papers did not provide any information on the within rater agreement (WRA) and between rater agreement (BRA). The team at the Institute of Child Health, led by Dr. Arto Demirjian, resolved to overcome this problem by providing detailed descriptions of the appearance of the teeth at a radiographically discernible level (Figure 1 and Table 1), and by systematic testing of the percentage reliability of assessments [25].

It is the detailed descriptions accompanied by schematic drawings that provide the basis for the easy application of the 8 Stage system of tooth development when assessing the maturity of developing teeth. When performing assessments of TDS for age estimation it is our experience that the combined use of the schematic drawings (Figure 1), and the written de-

scriptions (Table 1), enable a prompt decision to be made. For example it is our experience when hovering between a Stage F or G that referral to the drawings and descriptions usually makes a choice easy.

The underlying difficulty is that the 8 stage system is considered by some to insufficiently reliable for estimating dental age, and advocates have proposed a 10 stage system [32], a 12 stage system [24], continued use of the 14 stage system [23], and development of a new 21 stage system (Malekniazi 2010 personal communication). Each of these systems has its merits. The underlying issue with all of them is that there has been no reliable validation test using the outcome of Dental Age (DA) to see which is the most accurate method of assessment using the Gold Standard of Chronological Age for comparison. In essence the value of the CA *minus* DA should be expressed as decimal years and also in words as weeks, months or years as appropriate.

**Data Management and Integration of the Numerical Data for the AaA of each Tooth Development Stage**

A crucial part of DAE is the management of the data derived from each TDS. The approach used by the DARLInG study group is to summarize the Normal distribution statistics and the Percentile distribution statistics for each TDS. It is helpful to use the data from the UK Caucasian Database [33]. Namely the Lower (L) Left (L) Second Permanent Molar (7) stages B to H for males (m) (Table 2).

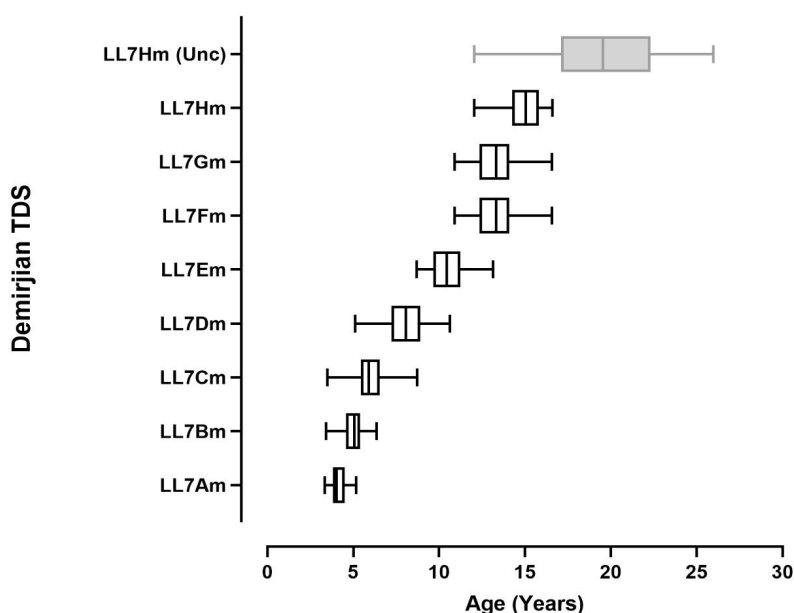
The impact of this results in mean value for uncensored data 2½ years greater than for the appropriately censored data. [see LL7H-cens and LL7H-uncens above]This large difference is, perhaps, better illustrated with a comparison of the 5 value summary using box and whisker graphs (Figure 2).

**How is the censor point identified?**

This is well illustrated using the Stacked Bar graph of the LL8 (Figure 3).

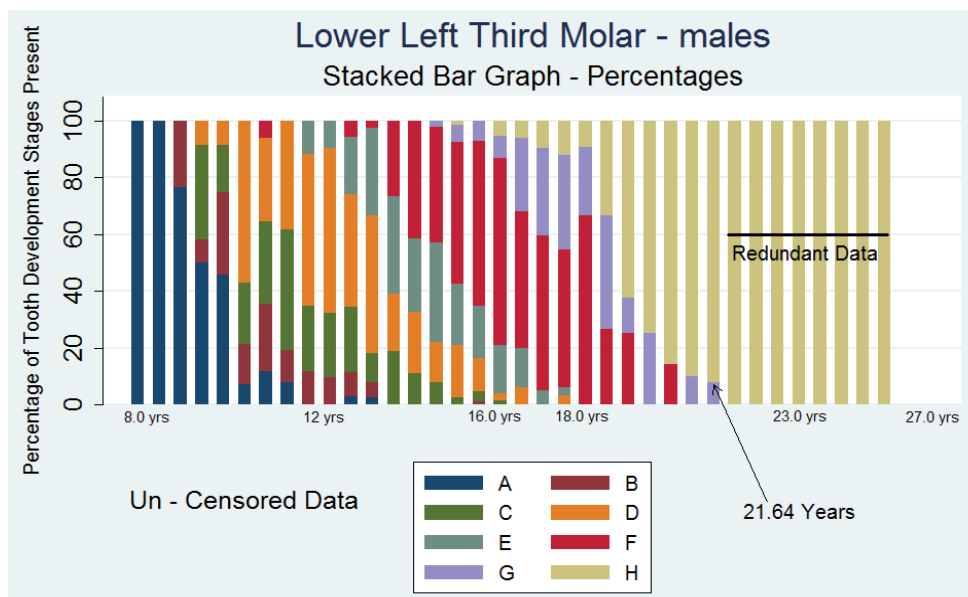
**Table 2:** Key - the names of the individual tooth Development Stages is given in the first column. The values for the individual TDS's are given in all the adjacent columns. The bottom row is the uncensored data for Stage H. The values are high because there is redundant data in the upper part of the distribution of values for Age at Assessment.

Summary Statistics for LL7-m UK Caucasian Reference Data Set								
Males	n-tds	x-tds	sd-tds	0 <sup>th</sup> %ile	25 <sup>th</sup> %ile-	50 <sup>th</sup> %ile-	75 <sup>th</sup> %ile	100 <sup>th</sup> %ile-
LL7A	-	-	-	-	-	-	-	-
LL7B	8	4.75	0.78	3.76	4.00	4.71	5.58	5.69
LL7C	13	5.80	0.80	4.17	5.33	5.79	6.26	7.45
LL7D	61	7.98	1.33	4.19	6.94	7.99	9.29	10.22
LL7E	64	9.58	0.97	6.98	8.99	9.58	9.90	12.96
LL7F	82	11.46	1.48	7.86	10.36	11.48	12.46	15.31
LL7G	181	13.50	1.34	10.86	12.50	13.41	14.35	16.90
LL7H-cens	435	15.08	1.35	9.83	14.32	15.38	16.09	16.90
LL7H-uncens	853	17.57	3.19	9.83	15.35	16.84	19.51	24.98



**Figure 2:** Key.

The box and whisker plots are typical of the standard output from most computer software packages. The mid-line of the box represents the median (50<sup>th</sup> %ile) of each TDS data set. These show a steady increase with age from 4.75 years at Stage B to 15.08 years in the appropriately censored data set. The uncensored data is the uppermost of the graphs and illustrates the effect of having redundant data above the censor point.



**Figure 3:** The Stacked Bar graph above shows the proportion of each TDS present at each half year interval. At 8.0 years 100% of the cases are Stage A. As one moves to the right viewing increasing age, the older stages, for example 14 years, has a small proportion of stage C (about 5%), a slightly larger proportion of Stage D (about 12%), a similar proportion of stage E (25%) and about 42% of Stage F. Toward the top end of the graph Stage H appears at age 16.0 years. By the age of 21.64 years all of Stage are completed and only Stage H are left. After this point all the stages are H Once this point has been reached there are no immature 3<sup>rd</sup> molars so dental development has complete. The data cases collected in the older age groups is redundant because they do not represent ongoing development.

The issue of censoring is important as it clearly identifies the point at which growth and development of the dentition has stopped. The effect of this has been explored in detail where the gradual reduction to the censor point reduces the mean value of Stage H from 24.99 years to 21.64 years [34].

The importance of this is that this phenomenon of censoring stage age should apply to each of the Stage H for each of the teeth used in the DGTm. Without censoring the box and whisker plots show the large amount of redundant data at the upper end of the data span for each Tooth Development



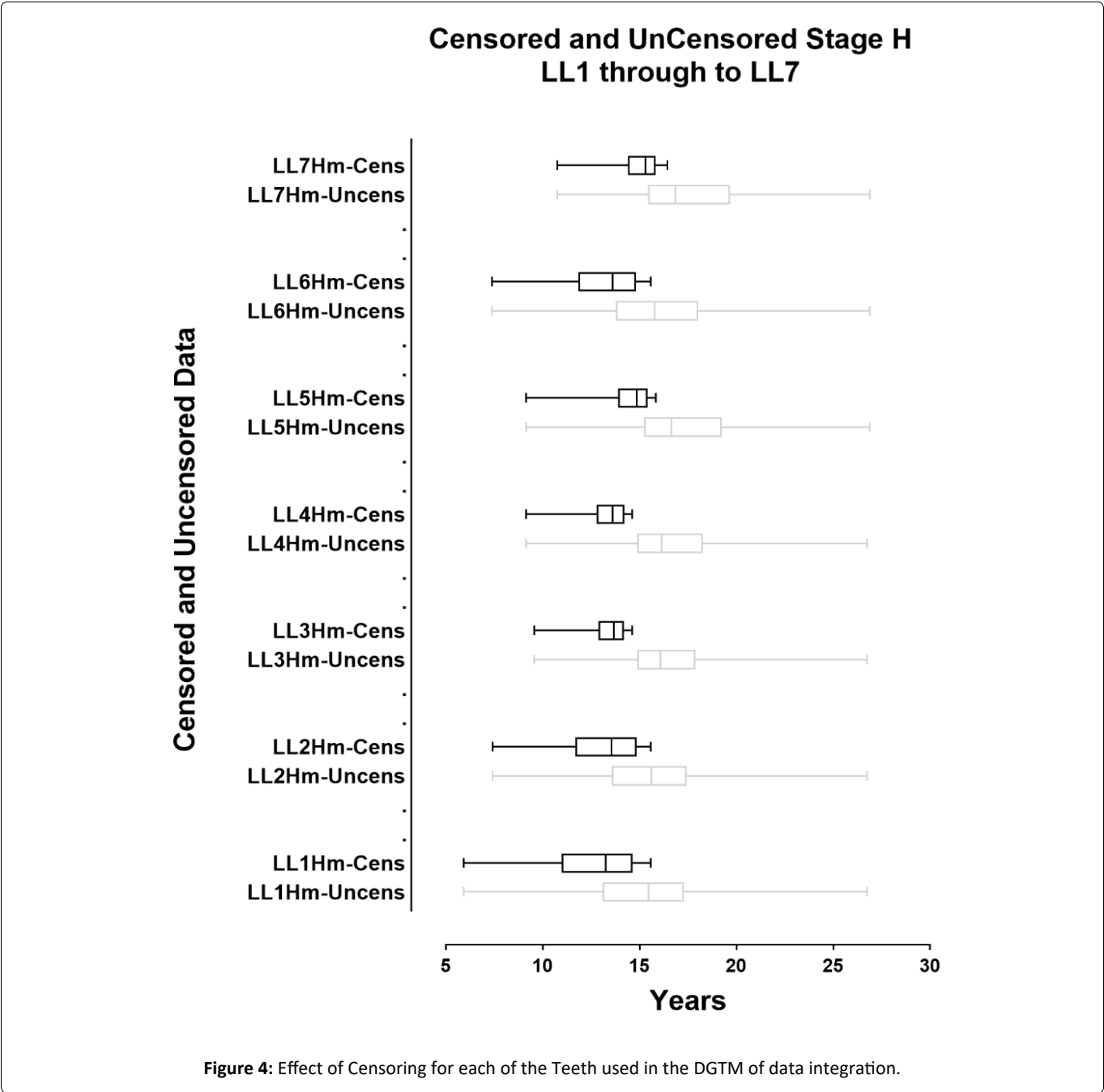
Stage (Figure 4).

Figure 4 shows the effect of Censoring on Stage H of each of the 7 teeth used in the DGTm. The above graphs are created from the UK Caucasian RDS. The age range of the radiographs is from 5.50 years to 25.99 years. The LL7Hm provides an exemplar of the issue of censoring. For the whole data set there are several hundred cases making up the TDS sample for LL7Hm. This is illustrated with the faint print box and whisker graph. The use of censoring brings down the mean. The box and whisker plots are based on the mean value with the maximum and minimum as the extreme values (the whiskers). The censored data, with black print box and whisker provides a plot that is greatly constrained [31] [the impact of censoring on the spread of the box an whisker graph will be explored in the discussion].

Attachment of a ‘weight’ to each of the TDS identified in the radiograph to give a score for each TDS

This is the most difficult part of the DGTm to comprehend. A sustained effort over the last 5 years has failed to reveal a formal process by which the numerical value can be converted to a weighted value suitable for integration into the overall assessment of the matrix for age estimation (Table 2 of the 1973 paper [2]).

Presumably an upper limit was determined when preparing the data for analysis. This was probably by inappropriate censoring [35]. The method used in the original paper could not be applied to the data in the UK Caucasian RDS (Goldstein H personal communication). An attempt to use a modern ver-



sion of dual scaling was not successful [36]. This is a crucial issue as the mean age at attainment for all stage Hs in the teeth involved in the DGTm is of critical importance. As is shown in [Figure 4](#) appropriate censoring materially alters the values of the summary statistics [34].

### **Summation of the weighted scores for the TDS identified to provide a maturity score - this is between 0 to 100**

Summation of the weighted scores for the TDS identified to provide a maturity score - this is between 0 to 100. This is an understandable but unnecessary way to present the data for age estimation. In essence the AaA of each TDS are acquired using summary statistics, these ages are then converted to maturity scores (by a method that is incomprehensible). The maturity scores for each TDS are the summed to give an overall score. This score is then converted back to a Chronological Age. This was always considered to be a cumbersome method especially with the unnecessary conversion and reversion. It has been replaced with the straightforward method of collecting together the summary statistics in a child and then carrying simple mathematical procedures to provide an estimate of the Dental Age with conventional measures of uncertainty [37].

The conversion of the sum of the weighted scores to chronological age (CA) for that individual by looking up a table of maturity/chronological scores and thus estimating the equivalent value for the CA is fraught with difficulty. It relies upon the investigator measuring by eye where the intersection of the curve for maturity matches the Chronological Age. This is a similar process to the reverse engineering of the Moorrees method [38] and is intrinsically a poor approach to the method of estimating age from a graphical presentation.

## **Discussion**

The majority of the weaknesses in the DGTm have been highlighted above. The main reason for drawing attention to these weaknesses is the need to fully understand the methods involved in DAE. This is important because the methods used by assessors need to be properly understood by the individual conducting the DAE. This has important implications as the DAE report may be scrutinized by lawyers who have access to expert advice and may take exception to the details and main thrust of the DAE report.

Sections 1 and 2 are, in essence very straightforward and follow standard clinical research practice [29].

It is a matter of great concern that in the published literature on DAE these fundamental and orthodox processes for clinical research are often violated.

The issue of Tooth Development Stages is crucial to the process of DAE. It is important that the most suitable TDS system is used. An important element is – How reliable are the assessments of the TDS? This currently requires that investigators conduct Intra and Between Agreement (IRA and BRA) studies using Cohen's Kappa statistic as the outcome [39]. The assessments by the Canadian team using %age agreement were detailed and thorough and provide more detailed in-

formation on the discrepancies occurring in TDS assessments [25]. The use of the Kappa Statistic has become widely used but reliance on a single value is questionable. The Canadian research team provided data on the number of differences observed. This information is helpful when interpreting the outcome of IRA and BRA studies. The additional information from the %age agreement given in the study by Levesque and Demirjian engenders confidence in the outcomes.

The careful assessment of individual TDS schemes has shown by two independent research groups that the Demirjian TDS are the most reliable as regards IRA and BRA [27,28]. It is difficult to see how any alternative system of TDS can be justified especially given the outcome DA is the same for the 8 stage system of Demirjian and the 12 Stage system of Haavikko [1].

The issue of managing the data before integrating the values from individual TDS is crucial and is poorly reported in the several papers published by the group. In an attempt to understand this issue the authors of the present review contacted the original authors of the 1973 paper. Professor Arto Demirjian promptly and deftly passed this issue back indicating that the person responsible for this is Professor Harvey Goldstein. Subsequently communication with Professor Goldstein over a period of a year or more fizzled out when it became impossible to receive guidance on how the data were managed. This was disappointing as intrinsically the simplicity of the DGTm is attractive. This simplicity has been utilised in an attempt to use estimated AaA for the TDS [13]. It is of interest that this method based on a Belgian population usually provides more accurate results than the DGTm. However, the final process of integrating the summary statistics and the way that this leads in to the age values presented for the TDS is unclear and despite several enquires to the author it has not been possible to identify the method used.

This is discouraging but the varied results shown in systematic reviews for both the DGTm [3] and the Willems method [40] show considerable variations the extremes of which are difficult to accept as reasonably accurate estimates of CA. It is, perhaps, the simplicity of these two techniques that has led investigators to use them without considering the effects of ancestry. This issue of ancestry or ethnicity is at the heart of all biological phenomena and failure to recognise and act on this fundamental aspect of human biology has led innumerable clinical investigators concluding 'good' or 'not so good' result on the basis of methods which, thus far, have been impossible to unravel.

## **Conclusion**

The DTDS (Demirjian Tooth Development Stages) are the most suitable for use in DAE systems. They have the additional value of being applicable to different ethnic or ancestral groups. Their importance and the universality of the highly reliable stages descriptions is clear from the many papers published on this important aspect of human dental development.

The other aspects of the DGTm are problematic. Perhaps the most worrying of these is the difficulty of comprehending exactly how the data processing is managed. This is a problem that is fatal to the DGTm.

## Recommendations

For DAE studies it is important that the following conditions are fulfilled:

1. A clearly defined ethnic or ancestral group is used to create the Reference Data Set.
2. The sample must comprise Cross-Sectional Data only
3. The Age range should be from 4 years to 25 years to ensure that the full range of variability for each Tooth Development Stage is represented.
4. The final stage of development - Stage H, must be appropriately censored to ensure redundant data beyond the time of completion of each tooth is not included.
5. The summary statistics for each of the Tooth Development Stages should be presented using n-tds for each Stage the mean age at attainment and the standard deviation. It would be helpful to present the five number summary comprising the Minimum, 25th %ile, Median, 75Th %ile, and the Maximum.
6. The method of integrating these data to produce an estimated age with the associated level of uncertainty is required to be presented in a simple and comprehensible way.
7. Consideration should be given to including all Tooth Morphology Types - all 16 teeth on the left side.

## References

1. Roberts G, McDonald F, Lucas VS (2016) Age estimation in the living: Dental age estimation - theory and practice. In: Payne-James J, Byard RW, Encyclopedia of forensic and legal medicine. (2<sup>nd</sup> edn), 46-69.
2. Demirjian A, Goldstein H, Tanner JM (1973) A new system of dental age assessment. *Hum Biol* 45: 211-227.
3. Jayaraman J, Ming Wong H, King NM, et al. (2013) The French-Canadian data set of Demirjian for dental age estimation: A systematic review and meta-analysis. *J Forensic Leg Med* 20: 373-381.
4. Ambarkova V, Galic I, Vodanovic M, et al. (2014) Dental age estimation using Demirjian Willems methods: Cross sectional study on children from former Yugoslav Republic of Macedonia. *Forensic Sci Int* 234: 187.
5. Yadava M, Roberts GJ, Lucas VS (2011) Dental age assessment (DAA): Reference data for British children at the 10 year threshold. *Int J Legal Med* 125: 651-657.
6. Vosk E, Emery AF (2015) Forensic metrology. Scientific measurement and inference for lawyers, judges, and criminalists. International Forensic Science and Investigation Series. CRC Press, Boca Raton, Florida.
7. Liversidge HM, Chaillet N, Mornstad N, et al. (2006) Timing of Demirjian's tooth formation stages. *Annals of Human Biology* 33: 454-470.
8. Jayaraman J, Roberts GJ (2018) Comparison of dental maturation between Hong Kong Chinese and United Kingdom Caucasian populations. *Forensic Science International* 292: 61-70.
9. Yan J, Lou X, Xie L, et al. (2013) Assessment of dental age of children aged 3.5 to 16.9 years using Demirjian's method: a meta-analysis based on 26 studies. *Public Library of Science One* 8.
10. Liversidge HM, Speechly T, Hector MP (1999) Dental maturation in British Children: Are Demirjian's standards applicable? *Int J Paediatr Dent* 9: 263-269.
11. Demirjian A, Goldstein H (1976) New systems for dental maturity based on seven and four teeth. *Ann Hum Biol* 3: 411-421.
12. Kelmendi J, Vodanovic M, Kocani F, et al. (2018) Dental age estimation using four Demirjian's, Chaillet's and Willems methods in Kosovar children. *Leg Med (Tokyo)* 33: 23-31.
13. Willems G, Van Olmen A, Spiessens B, et al. (2001) Dental age estimation in Belgian children: Demirjian's technique revisited. *J Forensic Sci* 46: 893-895.
14. Usher BM (2002) Reference Samples: The first step in linking biology and age in the human skeleton. In: Hoppa RD, Vaupeij JW, Paleodemography. Cambridge University Press, NY, USA.
15. Jayaraman J, Roberts GJ, King NM, et al. (2012) Dental age assessment of Southern Chinese using the United Kingdom Caucasian reference data set. *Forensic Sci Int* 216: 68-72.
16. Jayaraman J, Wong HM, King NM, et al. (2016) Development of a Reference Data Set (RDS) for dental age estimation and testing of this with a separate Validation Set (VS) in a Southern Chinese population. *J Forensic Leg Med* 43: 26-33.
17. Tanner JM (1962) Growth at adolescence. (2<sup>nd</sup> edn), Blackwell Scientific Publications, Oxford, 4-10.
18. Cameron N (2002) Human growth and development. Academic Press, London, 15-17.
19. Levesque GY, Demirjian A, Tanguay R (1981) Sexual dimorphism in the development, emergence, and agenesis of the Mandibular third molar. *J Dent Res* 60: 1735-1741.
20. Liversidge HM (2008) Timing of human mandibular third molar formation. *Ann Hum Biol* 35: 294-321.
21. Rolseth V, Mosdol A, Dahlberg S, et al. (2019) Age assessment by Demirjian's development stages of the third molar: A systematic review. *Eur Radiol* 29: 2311-2321.
22. Roberts GJ, McDonald F, Neil M, et al. (2014) The weighted average method 'WAM' for Dental Age Estimation: A simpler method for children at the 10 year threshold. *Journal of Forensic and Legal Medicine* 26: 56-60.
23. Moorrees CF, Fanning LA, Hunt LE (1963) Age variation of formation stages for ten permanent teeth. *J Dent Res* 42: 1490-1502.
24. Haavikko K (1970) The formation and the alveolar and clinical eruption of the permanent teeth. An orthopantomographic study. *Suom Hammaslaak Toim* 66: 103-170.
25. Levesque GY, Demirjian A (1980) The inter examiner variation in rating dental formation from radiographs. *Journal of Dental Research* 59: 1123-1126.
26. Landis JR, Koch GG (1977) The measurement of Observer Agreement for categorical data. *Biometrics* 33: 159-174.
27. Olze A, Bilang D, Schmidt S, et al. (2005) Validation of common classification system for assessing the mineralization of third molars. *Int J Legal Med* 119: 22-26.
28. Dhanjal KS, Bhardwaj MK, Liversidge HM (2006) Reproducibility of radiographic stage assessment of third molars. *Forensic Science International* 159: S74-S77.

29. Altman DG (1991) *Practical statistics for medical research*. Chapman & Hall, London.
30. Tanner JM, Healy MJR, Goldstein H, et al. (2001) *Assessment of skeletal maturity and prediction of adult height*. WB Saunders. London.
31. Gleiser I, Hunt EE (1955) The permanent mandibular first molar: its calcification, eruption and decay. *Am J Phys Anthropol* 13: 253-284.
32. Kohler S, Schmelzle RL, Puschel K (1994) Development of wisdom teeth as a criterion of age determination. *Ann Anat* 176: 339-345.
33. Roberts G (2014) UK reference data set.
34. Roberts GJ, McDonald F, Andiappan M, et al. (2015) 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. *J Forensic Leg Med* 36: 177-184.
35. Boonpitakthasit T, Hunt N, Roberts G, et al. (2011) Dental Age Assessment of adolescents and emerging adults in United Kingdom Caucasians using censored data for stage H of third molar roots. *European Journal of Orthodontics* 33: 503-508.
36. Ueda T (2006) Dual Scaling using mathematical programming and its application. *The Sixth International Symposium on Operations Research and Its Applications*.
37. Draft D, Lucas VS, McDonald F, et al. (2019) Expressing uncertainty in Dental Age Estimation: A comparison between two methods of calculating the 'average standard deviation'. *Journal of Forensic Sciences* 64: 1506-1509.
38. Harris EF, Buck AL (2002) Tooth mineralization: A technical note on the Moorrees-Fanning-Hunt standards. *Dental Anthropology* 16: 15-19.
39. Petrie A, Sabin C (2009) *Medical statistics at a glance*. (3<sup>rd</sup> edn), Wiley-Blackwell, Oxford.
40. Sehwat JS, Singh M (2017) Willems method of dental age estimation in children: A systematic review and meta-analysis. *J Forensic Leg Med* 52: 122-129.

**DOI: 10.36959/763/519**