

# Methods to Quantify Soft-Tissue Based Facial Growth and Treatment Outcomes in Children: A Systematic Review

Sander Brons<sup>1\*</sup>, Machteld E. van Beusichem<sup>1</sup>, Ewald M. Bronkhorst<sup>2</sup>, Jos Draaisma<sup>3</sup>, Stefaan J. Bergé<sup>4</sup>, Thomas J. Maal<sup>4</sup>, Anne Marie Kuijpers-Jagtman<sup>1</sup>

**1** Department of Orthodontics and Craniofacial Biology, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands, **2** Department of Preventive and Curative Dentistry, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands, **3** Department of Pediatric Medicine, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands, **4** Department of Oral and Maxillofacial Surgery, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands

## Abstract

**Context:** Technological advancements have led craniofacial researchers and clinicians into the era of three-dimensional digital imaging for quantitative evaluation of craniofacial growth and treatment outcomes.

**Objective:** To give an overview of soft-tissue based methods for quantitative longitudinal assessment of facial dimensions in children until six years of age and to assess the reliability of these methods in studies with good methodological quality.

**Data Source:** PubMed, EMBASE, Cochrane Library, Web of Science, Scopus and CINAHL were searched. A hand search was performed to check for additional relevant studies.

**Study Selection:** Primary publications on facial growth and treatment outcomes in children younger than six years of age were included.

**Data Extraction:** Independent data extraction by two observers. A quality assessment instrument was used to determine the methodological quality. Methods, used in studies with good methodological quality, were assessed for reliability expressed as the magnitude of the measurement error and the correlation coefficient between repeated measurements.

**Results:** In total, 47 studies were included describing 4 methods: 2D x-ray cephalometry; 2D photography; anthropometry; 3D imaging techniques (surface laser scanning, stereophotogrammetry and cone beam computed tomography). In general the measurement error was below 1 mm and 1° and correlation coefficients range from 0.65 to 1.0.

**Conclusion:** Various methods have shown to be reliable. However, at present stereophotogrammetry seems to be the best 3D method for quantitative longitudinal assessment of facial dimensions in children until six years of age due to its millisecond fast image capture, archival capabilities, high resolution and no exposure to ionizing radiation.

**Citation:** Brons S, van Beusichem ME, Bronkhorst EM, Draaisma J, Bergé SJ, et al. (2012) Methods to Quantify Soft-Tissue Based Facial Growth and Treatment Outcomes in Children: A Systematic Review. PLoS ONE 7(8): e41898. doi:10.1371/journal.pone.0041898

**Editor:** Arin Greene, Children's Hospital Boston, United States of America

**Received:** February 20, 2012; **Accepted:** June 28, 2012; **Published:** August 6, 2012

**Copyright:** © 2012 Brons et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work was supported by a grant from the Dutch Technology Foundation to STW (grant no. 10315). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: orthodontics@dent.umcn.nl

## Introduction

Longitudinal quantitative evaluation of facial dimensions of an individual patient could inform healthcare professionals about growth as well as treatment changes [1,2]. Accurate quantitative evaluation of craniofacial dimensions by comparison of an individual patient to normative values can provide insight into an underlying pathologic process or create a basis for treatment planning [3,4].

Various methods for quantitative evaluation of craniofacial dimensions have been described for a variation of purposes. The standard technique is direct anthropometry which was extensively used for the study of craniofacial dimensions in the past century

[5]. These “direct” measurements are reliable, inexpensive to make and regarded as the gold standard. Limitations include its time-consuming nature, the need for patient compliance and to remain still [6]. Additionally, it is not possible to archive craniofacial surface morphology. Also used for decades are two dimensional (2D) x-ray cephalometry [7–9] and photography [3,10] and even today these are the most commonly used records for dento-skeletal and facial diagnosis. The advantages of these 2D imaging techniques are rapid acquisition, archival capabilities and low cost. Limitations include measurement error due to magnification, parallax and head orientation [11] and exposure to ionizing radiation. Recent technological advancements have led craniofacial researchers and clinicians into the era of three

dimensional (3D) digital imaging. Techniques like cone beam computed tomography (CBCT) [12,13], surface laser scanning [14,15] and stereophotogrammetry [16–18] became available to describe and compare 3D facial surfaces, create a diagnosis or virtual treatment planning as well as to evaluate growth and treatment outcomes. These methods allow images to be archived and avoid measurement errors that occur with 2D representations of 3D surfaces. CBCT however, is not an ideal technique for surface measurement because of poor resolution of facial contours, high cost and exposure to ionizing radiation [19]. Laser surface scanning can be reliable and accurate for identifying craniofacial surface landmarks and is relatively inexpensive. Limitations include slow image capture (up to 20 seconds) and potential damage to the eyes [20]. This is particularly difficult for children because they are less able to maintain posture for this period of time and keep their eyes closed. 3D stereophotogrammetry overcomes the limitations of surface laser scanning. It is millisecond fast and has archival capabilities for subsequent morphometric studies, a good-resolution color representation and no exposure to ionizing radiation [19,20]. The disadvantages of stereophotogrammetry are its expense, difficulties in imaging transparent, shiny and shadowed surfaces and inability to measure bony landmarks.

Many studies address validity, accuracy and reliability of craniofacial anthropometric measurements [6,11,21–25]. Differences in types of error, ages of samples and anatomical location of analysis make it difficult to compare reliability. Therefore, the objectives of this systematic review are 1) to give an overview of soft tissue-based methods for quantitative longitudinal assessment of facial dimensions in children until six years of age, 2) to assess the methodological quality of the studies using such a method and 3) to assess reliability of these imaging methods used in studies with good methodological quality.

## Methods

### Protocol and Registration

Inclusion criteria and methods of analysis were specified in advance and documented in a protocol. PROSPERO [26] for prospective registration of systematic reviews was in development at the start of this review. A registration number is therefore not available.

### Eligibility Criteria

Eligible for inclusion were primary publications which report of: 1) soft-tissue based evaluation of head and face; 2) children before 6 years of age at the start of the study; 3) quantitative changes; 4) longitudinal studies.

Excluded were publications which report of: 1) skeletal changes; 2) fetal growth; (2) animal studies, (3) cross-sectional studies, (4) case reports, reviews and letters. No restrictions for language, publication date and publication status were imposed.

### Information Resources

Studies were identified by searching electronic databases. The search was applied to PubMed (from 1948), EMBASE Excerpta Medica (from 1980), Cochrane Library (from 1993), Web of Science (from 1945), Scopus (from 2004) and CINAHL (from 1982). The last search was run on October 1, 2011. In addition, we hand searched the reference lists of included studies for potentially eligible studies. Digital full text publications were retrieved from licensed digital publishers and paper publications were retrieved from the library. In cases where the full text

publication could not be retrieved, authors were requested by e-mail to deliver the publication. Gray literature was not searched.

### Search Strategy

The search strategy was developed and databases were selected with the help of a senior librarian specialized in health sciences. Databases selected were PubMed, EMBASE Excerpta Medica, Cochrane Library, Web of Science, Scopus and CINAHL. Medical Subject Headings and free text words were used for the search strategy of PubMed (Table 1). The search strategies for the other databases are directly derived from the former. The last search was performed on October 1, 2011.

The search strategy focused on four aspects:

1. terms to search for the population of interest (*i.e.*, baby's, infants and pre- school children). A selection of the appropriate terms from the Child search strategy was made to sort out citations not reporting on children between 0 and 6 years of age [27];
2. terms to search for growth and methods for quantitative evaluation (*i.e.*, growth, anthropometrics and imaging techniques);
3. terms to search for the anatomic region of interest (*i.e.*, face and head);
4. terms to search for the longitudinal aspect (*i.e.*, cohort and follow up studies).

### Study Selection

First, studies were independently screened on title and abstract by two reviewers (SB and MB) in a blinded standardized manner. In an additional step, disagreements between reviewers were resolved by discussion and consensus.

Second, full text assessments for eligibility were independently performed by two reviewers in a blinded standardized manner. In an additional step, disagreements were resolved by discussion and consensus.

Third, a hand search of the reference lists of the included studies was performed by the first author.

Finally, all included studies were categorized as describing facial or cranial evaluation of growth and treatment outcome. The plane connecting glabella with left and right euryon arbitrarily separates the cranium from the face. Measurement on or above this plane are called to be cranial, below this plane are called to be facial. The studies describing facial evaluation of growth and treatment are included in this review for quality assessment. Results of the selection process by two reviewers (SB and MB) were analyzed to assess interrater reliability.

### Quality Assessment

Study quality was assessed by the quality assessment instrument (QAI) for clinical trials used by Gordon et al. (Table 2) [28]. This instrument includes an assessment of study bias. A checkmark was scored when a criterion was fulfilled. Depending on study design quality assessment was performed on a maximum of 15 criteria. In case criteria were not applicable to a certain study design, less than 15 criteria were scored. Study quality is expressed as the percentage of criteria fulfilled in relation to the total number of applicable criteria.

The score per study is calculated as a percentage by dividing the number of checkmarks by the number of applicable criteria and multiplying by 100. Studies were grouped according to similarity of the methods for measurement of facial growth or treatment

**Table 1.** Search strategy PubMed.

Search strategy PubMed
("Face"[Mesh:noexp] OR face[TiAb] OR facial[TiAb] OR craniofacial[TiAb] OR OR OR born* craniomaxillofacial[TiAb] OR maxillofacial[TiAb] OR dentofacial[TiAb] OR "Facies"[Mesh] facies[TiAb] OR "Head"[Mesh:noexp] OR head[TiAb]) AND ("Growth and Development"[Mesh:noexp] OR "Growth"[Mesh:noexp] OR "growth and development"[Sh] growth[TiAb] OR "Anthropometry"[Mesh:noexp] OR anthropometr*[TiAb] OR "cephalometry"[Mesh] OR cephalometr*[TiAb] OR "imaging, three-dimensional"[MeSH Terms] OR "three-dimensional imaging"[TiAb] OR "3d imaging"[TiAb] OR "Photogrammetry"[Mesh] OR photogrammetry[TiAb] OR "Tomography, X Ray Computed"[Mesh] OR "Tomography, X Ray Computed"[TiAb] OR "Lasers"[Mesh:noexp] OR laser[TiAb] OR "Magnetic Resonance Imaging"[Mesh:noexp] OR "magnetic resonance Imaging"[TiAb] OR MRI[TiAb]) AND (infant OR infan* OR newborn OR newborn* OR new OR baby OR baby* OR babies OR neonat* OR perinat* OR postnat* OR toddler* OR kindergar* OR preschool* OR pre school) AND ("Cohort Studies"[Mesh] OR (cohort[TiAb] OR longitudinal[TiAb] OR followup[TiAb] OR follow up*[TiAb]) AND (study[TiAb] OR studies[TiAb]))))

doi:10.1371/journal.pone.0041898.t001

outcome. A mean quality score for each group of methods was calculated. Arbitrarily, a cut-off of 60% or higher is graded as good quality. Below 60% is graded as poor quality. To assess the interrater reliability of the assessment of study quality 19 randomly selected studies were scored by two reviewers (SB and AK).

## Data Extraction

Methods, used in studies with good methodological quality, were assessed for reliability expressed as the magnitude of the measurement error and the correlation coefficient between repeated measurements.

## Statistics

Cohen's kappa statistics were used to assess the interrater agreement for the process of study selection and for each criterion of the quality assessment instrument. According to Landis and Koch the level of interrater agreement is very good if the value of K is 0.81–1.00, good if K is 0.61–0.80, moderate if K is 0.41–0.60, fair if K is 0.21–0.40 and poor if K is <0.20 [29].

Analysis of variance (ANOVA) and non-parametric Kruskal-Wallis test were performed to test differences in mean scores between groups of methods. Fisher's exact test was performed to test for differences between groups of methods with the use of a cut-off of 60%. SPSS version 19.0 was used as statistical software.

## Results

### Study Selection

Interexaminer kappa for screening on title and abstract was 0.76. For full text assessment of eligibility kappa was 0.69. The reliability of both steps in the process of study selection is qualified as good [29].

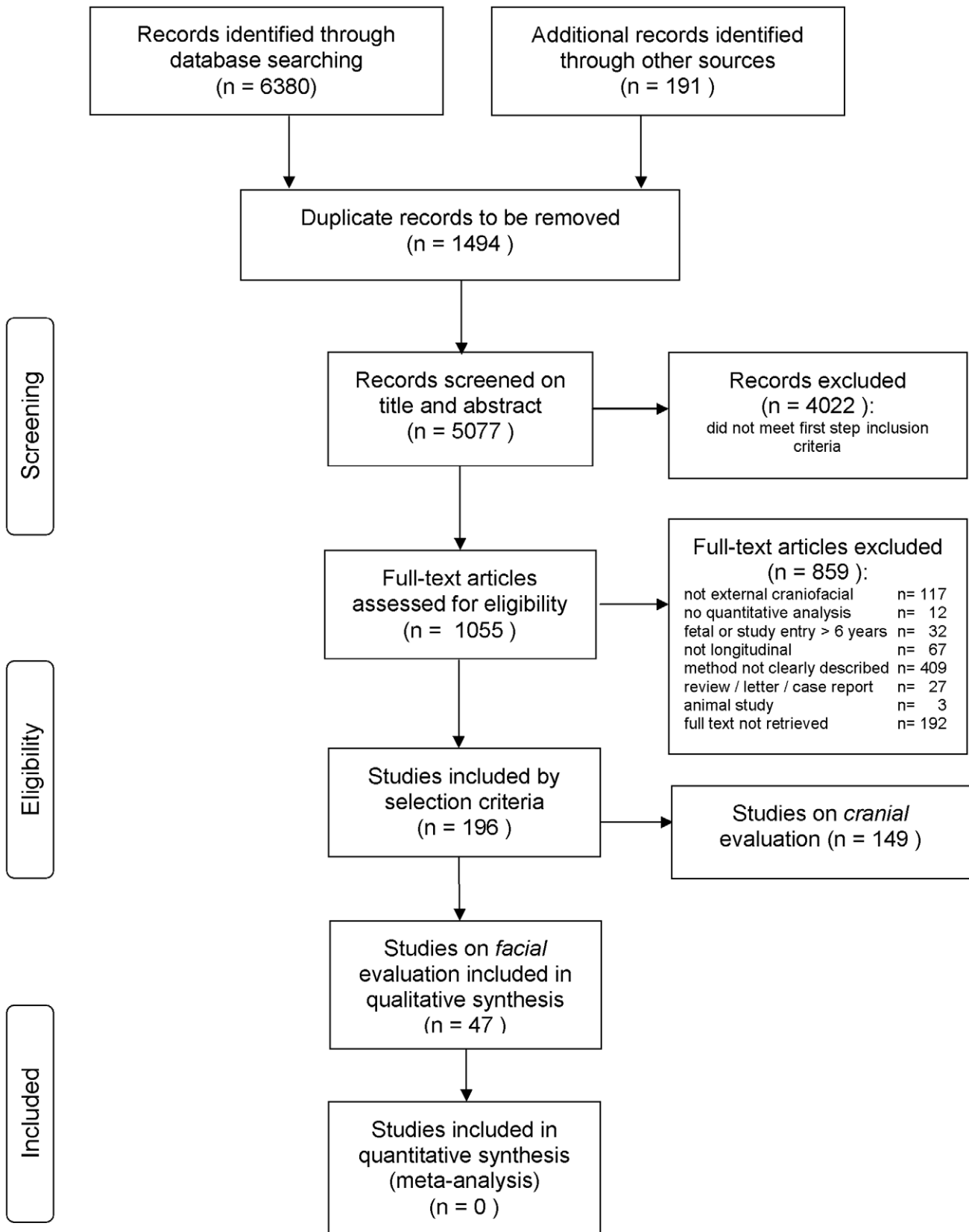
The search of PubMed, EMBASE, Cochrane Library, Web of Science, Scopus and CINAHL provided a total of 6380 citations and the hand search provided 191 citations. After adjusting for duplicates 5077 remained for screening of title and abstract. Of these, 4022 studies were discarded because these did not meet the eligibility criteria. A total of 1055 studies remained for full text assessment of eligibility. Of these, 859 studies were excluded with reasons. Of these excluded studies, 192 were discarded because the full text publication could not be retrieved. The last step in the inclusion process divided the studies into facial evaluation (n = 47) and studies on cranial evaluation (n = 149). A total of 196 studies was identified meeting the inclusion criteria; 175 studies originated

**Table 2.** Quality assessment instrument [28].

<b>I. Study design (7 Y)</b>
A. Objective—objective clearly formulated (Y)
B. Sample size—considered adequate (Y)
C. Sample size—estimated before collection of data (?)
D. Selection criteria—clearly described (Y)
E. Baseline characteristics—similar baseline characteristics (Y)
F. Timing—prospective (Y)
G. Randomization—stated (Y)
<b>II. Study measurements (3 Y)</b>
H. Measurement method—appropriate to the objective (Y)
I. Blind measurement—blinding (Y)
J. Reliability—adequate level of agreement (Y)
<b>III. Statistical analysis (5 Y)</b>
K. Dropouts—dropouts included in data analysis (Y)
L. Statistical analysis—appropriate for data (Y)
M. Confounders—confounders included in analysis (Y)
N. Statistical significance level—P value stated (Y)
O. Confidence intervals provided (Y)

Maximum number of Ys = 15.

doi:10.1371/journal.pone.0041898.t002



**Figure 1. PRISMA flow diagram of study selection process.**

doi:10.1371/journal.pone.0041898.g001

from the electronic databases; the remaining 21 studies originated from the additional handsearch of the references of the included studies. Figure 1 shows the PRISMA flow diagram and figure S1 shows the PRISMA checklist [30]. This study is restricted to studies on facial evaluation of growth and treatment outcome in children.

Of the 47 included studies, 24 studies used 2D x-ray cephalometry [31–54], 9 studies used 2D photography [55–63], 7 studies used anthropometry [64–70] and 7 studies used 3D imaging (2 surface laser scanning [71,72], 4 stereophotogrammetry [73–76] and 1 CBCT [77]).

### Study Quality Assessment

Interrater reliability for all 15 criteria of the quality assessment instrument were between 0.19 and 1 (interexaminer kappa), 11 out of 15 criteria had a kappa of 0.50 or higher. Interrater agreement on criteria E (similar baseline characteristics), I (blind measurement) and K (dropouts included in data analysis) were below 0.20.

All included studies could be categorized into one of following methods for quantitative evaluation of soft-tissue based growth or treatment changes: 2D X-ray cephalometry, 2D photography, direct and indirect anthropometry, and 3D imaging techniques (surface laser scanning, stereophotogrammetry, cone beam computed tomography). Assessment of methodological quality of all reviewed studies resulted in scores ranging from 30% to 100%. 24 studies qualified as good according to a methodological quality score equal to or above 60%. Score summaries are shown in Table 3.

Analysis of variance ( $p=0.41$ ) and Kruskal-Wallis test ( $p=0.15$ ) showed no statistical significant difference for quality expressed as a percentage between groups of methods. Also Fisher's exact test ( $p=0.07$ ) showed no statistical significant difference in the amount of studies with good methodological quality between groups of methods.

### Reliability

Scores for reliability of methods for soft-tissue based quantitative longitudinal assessment are shown in Table 4.

All good quality studies using 2D x-ray cephalometry report a measurement error below 1 mm and 1° except for the studies of Hermann et al. [41,42] where the range is up to 2 mm for linear and 3° for angular measurements. Correlation coefficients between repeated measurements range from 0.665 to 0.989 and are qualified as good to very good. Two studies report on reliability as “no significant” error and three studies do not report on reliability at all.

Studies with good methodological quality using 2D photography report a measurement error of 0.01 in case of ratios [56], “no significant” error [61] and a correlation coefficient of 0.9956 [60] which can be qualified as very good.

No studies with good methodological quality using direct or indirect anthropometry in children below 6 years of age report on measurement error. One study reports a correlation coefficient of 0.96 to 1.0 which can be qualified as very good [66].

One study with good methodological quality using 3D stereophotogrammetry reports a measurement error of 0.5 mm [74].

There are no good quality studies using 3D surface laser scanning or CBCT in children below 6 years of age.

## Discussion

### Summary of Evidence

The objectives of this systematic review were 1) to give an overview of soft tissue-based methods for quantitative longitudinal assessment of facial dimensions in children until six years of age and 2) to assess the methodological quality of the studies using such a method and 3) to assess reliability of these quantitative measurement methods used in studies with good methodological quality. 2D X-ray cephalometry is the method used most often and has demonstrated its potential to be used in studies with a good methodological quality. Also 2D photography and anthropometry are used in studies with good methodological quality. However, only one study using 3D imaging has shown its use with a good methodological quality despite its potential benefits. A possible explanation might be that researchers pioneering these relatively new methods are more focused on application of these methods than on development of the best possible study design. Future studies using 3D imaging for quantitative evaluation of facial growth and treatment outcome should focus on proper design to demonstrate its potential to be used in studies with good methodological quality in order to take advantage of their benefits.

In literature various terms to describe the measurement error exist. Some studies use accuracy to describe landmark identification error which in turn may consist of operator error, capture error and registration error [78]. More often in literature reliability is used to describe landmark identification error of a method. Reliability can be expressed by the measurement error or correlation coefficient between repeated measurements [11,25,79]. Reliability represents the ability of observers to make a consistent analysis. In this systematic review reliability in studies with good methodological quality is assessed and expressed by duplicate measurement errors and correlation coefficients between repeated measurements.

Reliability in included studies using 2D x-ray cephalometry report a measurement error below 1 mm and 1°. Correlation coefficients range from 0.665 to 0.989 and are qualified as good to very good. This is in concordance with the reported reliability of digital 2D x-ray cephalometry in older children (from 9.2–11.0 years) [79]. Reliability in one of the included studies using 2D photography is qualified as very good. This is in partial agreement with Farkas et al. [11] who found only 20 out of 62 measurements to be reliable in adolescents with a measurement error equal to or below 1 mm and 2°. It is key to select reliable measurements when using 2D photography. Reliability in one of the included studies using anthropometry is qualified as very good. Well-trained anthropologist are indeed able to reliably measure craniofacial dimensions, as was shown for older individuals [5]. Finally, reliability of one included study using 3D stereophotogrammetry is good with a measurement error of 0.5 mm. This is in agreement with literature with reported measurement errors in adults between 0,20 mm and below 1 mm and a correlation coefficient of 0.91 [25,78].

When comparing the accuracy of a technique to the standard technique or the gold standard, anthropometry -direct anthropometric measurements- correlated highly with digital 3D stereophotogrammetry (mean  $r=0.88$ ) [25]. Furthermore, millisecond fast image capture, archival capabilities for subsequent morphometric studies, a good-resolution color representation and no exposure to ionizing radiation make stereophotogrammetry the best 3D method for quantitative longitudinal assessment of facial dimensions in children until six years of age.

**Table 3.** Methodological quality scores of studies reporting on soft tissue-based quantitative longitudinal assessment of facial dimensions in children until six years of age.

First author	Year	Design							Measure			Statistics					Score
		A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	
<i>2D x-ray cephalometry</i>																	
Bishara	1998	Y	Y	o	Y	.	o	.	Y	.	Y	.	Y	.	Y	o	70%
Bishara	1997	Y	Y	o	Y	Y	o	.	Y	o	Y	.	Y	Y	Y	o	69%
Bishara	1985	Y	Y	o	Y	.	o	.	Y	.	Y	.	Y	.	Y	o	70%
Bishara	1985	Y	Y	o	Y	Y	o	.	Y	o	Y	.	Y	o	Y	o	62%
Bishara	1984	Y	Y	o	Y	.	o	.	Y	.	Y	.	Y	.	Y	o	70%
Bongaarts	2009	Y	Y	o	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	93%
Coccaro	1965	Y	Y	.	o	o	o	.	Y	.	o	.	o	.	o	o	30%
Daskalogiannakis	2006	Y	Y	o	Y	Y	o	.	Y	o	o	.	Y	Y	Y	o	62%
Graber	1977	Y	Y	o	Y	.	o	.	Y	o	Y	.	Y	.	Y	o	64%
Hanada	1975	Y	Y	.	Y	Y	o	.	Y	o	o	.	Y	.	o	o	55%
Hermann	2004	Y	Y	o	Y	Y	o	.	Y	o	Y	.	Y	.	Y	o	75%
Hermann	2003	Y	Y	o	Y	o	o	.	Y	o	Y	.	Y	Y	Y	o	62%
Hermann	2003	Y	Y	o	Y	o	o	.	Y	o	Y	.	Y	.	Y	o	58%
Hermann	2002	Y	Y	o	Y	o	o	.	Y	o	Y	.	Y	.	Y	o	58%
Hermann	1999	Y	Y	o	Y	Y	o	.	Y	o	o	.	Y	.	Y	o	58%
Padwa	1999	Y	Y	o	o	o	o	.	Y	o	o	.	Y	o	Y	o	38%
Posen	1967	Y	Y	o	Y	.	o	.	Y	.	o	.	Y	.	Y	o	60%
Sadowsky	1973	Y	Y	o	Y	.	o	.	Y	.	Y	.	Y	.	Y	Y	80%
Semb	1991	Y	Y	.	Y	.	o	.	Y	o	Y	.	Y	.	o	o	60%
Semb	1991	Y	Y	.	Y	.	o	.	Y	o	o	.	Y	.	o	o	50%
Smahel	1995	Y	Y	o	Y	.	o	.	Y	.	Y	.	Y	.	Y	o	70%
Subtelny	1959	Y	Y	.	Y	.	o	.	Y	.	o	.	Y	.	o	o	56%
Wen-Ching Ko	2000	Y	o	o	o	.	o	.	Y	.	Y	.	Y	.	Y	o	50%
Zettergren-Wijk	2006	Y	Y	o	Y	Y	o	.	Y	o	Y	.	Y	.	Y	o	75%
<i>2D photography</i>																	
Altug-Atac	2008	Y	o	.	o	.	o	.	Y	.	o	.	Y	.	o	o	33%
Castelo	2010	Y	Y	Y	Y	.	Y	.	Y	.	Y	o	Y	.	Y	o	82%
Cruz	2008	Y	o	o	Y	.	o	.	Y	o	Y	.	Y	.	Y	o	55%
Ko	2004	Y	o	.	o	.	Y	.	Y	o	Y	o	Y	.	o	o	45%
Kohout	1998	Y	o	o	Y	Y	Y	o	Y	o	o	o	Y	o	Y	o	47%
Liou	2007	Y	Y	o	Y	.	Y	.	Y	o	Y	.	Y	.	Y	Y	82%
Pai	2005	Y	Y	o	o	.	Y	.	Y	.	Y	o	Y	.	Y	o	64%
Schüler	2007	Y	Y	.	o	.	Y	.	Y	.	o	o	Y	.	o	o	50%
Sultana	2000	Y	Y	o	Y	.	o	.	Y	o	o	.	Y	.	Y	o	55%
<i>Direct anthropometry</i>																	
Bennun	1999	Y	Y	o	Y	o	Y	.	Y	o	o	o	Y	Y	Y	o	57%
Hansen	1997	Y	Y	o	Y	o	o	o	Y	Y	o	.	Y	o	Y	o	50%
Heimer	2008	Y	Y	Y	Y	.	Y	.	Y	Y	Y	Y	Y	.	Y	Y	100%
Ridgway	2011	Y	Y	o	Y	.	o	.	Y	.	o	.	Y	.	Y	o	60%
Vander Woude	1997	Y	Y	o	Y	.	Y	.	Y	.	o	o	Y	.	Y	o	64%
Yang	2009	Y	Y	o	Y	.	o	.	Y	.	o	.	Y	.	Y	o	60%
<i>Indirect anthropometry</i>																	
Ezzat	2007	Y	o	o	Y	.	Y	.	Y	Y	o	.	Y	.	Y	o	64%
<i>3D surface laser scanning</i>																	
Primozic	2009	Y	Y	o	Y	o	Y	.	Y	o	o	o	Y	.	Y	o	54%
Schwenzer	2008	Y	o	.	o	o	Y	.	Y	.	Y	.	Y	.	o	o	50%

**Table 3. Cont.**

First author	Year	Design							Measure			Statistics					Score
		A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	
<i>3D stereophotogrammetry</i>																	
Hoefert	2010	Y	Y	o	o	o	Y	.	Y	.	o	o	Y	.	Y	Y	58%
Hood	2003	Y	Y	o	o	Y	o	.	Y	.	Y	.	Y	.	Y	o	64%
Ras	1995	Y	Y	o	o	o	Y	.	Y	.	o	o	Y	.	Y	o	50%
Singh	2005	Y	o	o	Y	.	Y	.	Y	o	Y	o	Y	.	Y	o	58%
<i>Conebeam computed tomography</i>																	
Seidenstricker	2008	Y	Y	.	o	.	o	.	Y	.	Y	.	Y	.	o	o	56%

Y = Fulfilled satisfactorily the methodological criteria;  
 o = Did not fulfil the methodological criteria;  
 . = Not applicable.  
 doi:10.1371/journal.pone.0041898.t003

**Table 4. Reliability of methods for soft tissue-based quantitative longitudinal assessment of facial dimensions in children until six years of age in studies with good methodological quality.**

First author	Year	Measurement error	Correlation coefficient
<i>2D x-ray cephalometry</i>			
Bishara	1998	0.5 mm/0.5°	.
Bishara	1997	0.5 mm/0.5°	.
Bishara	1985	0.5 mm/0.5°	.
Bishara	1985	0.2 mm/0.5°	.
Bishara	1984	0.2 mm/0.5°	.
Bongaarts	2009	.	0.655–0.989
Daskalogiannakis	2006	.	.
Graber	1977	ns	.
Hermann	2004	0.27–1.94 mm/0.36–2.97°	.
Hermann	2003	0.27–1.94 mm/0.36–2.97°	.
Posen	1967	.	.
Sadowsky	1973	.	.
Semb	1991	ns	.
Smahel	1995	.	0.95–0.97
Zettergren-Wijk	2006	0.86 mm/0.80°	.
<i>2D photography</i>			
Castelo	2010	0.01 (ratio)	.
Liou	2007	.	0.9956
Pai	2005	ns	.
<i>Direct anthropometry</i>			
Heimer	2008	.	0.96–1.0
Ridgway	2011	.	.
Vander Woude	1997	.	.
Yang	2009	.	.
<i>Indirect anthropometry</i>			
Ezzat	2007	.	.
<i>3D stereophotogrammetry</i>			
Hood	2003	0.5 mm	.

. = not reported.  
 ns = not significant.  
 doi:10.1371/journal.pone.0041898.t004

## Limitations

Failure to identify all relevant reports for a systematic review could result in bias [80]. For this reason highly sensitive search strategies were developed with the help of a senior librarian specialized in health sciences for a combination of both narrow and broad health science databases.

The process of study selection was performed in an independent blinded standardized manner by two reviewers to prevent unjustified exclusion of eligible studies. The hand search of the reference lists of the included studies was performed by only one reviewer. Possibly eligible studies could have been missed in this stage of the selection process. However, since only approximately one out of ten studies was retrieved by the hand search this might be negligible. Furthermore, failure to retrieve full text publications of possibly eligible studies ( $n=192$ ) was inevitable even though every effort was made to contact the authors by email in cases where online access was not permitted or the journal was not available in the library. It is estimated that approximately 8 additional studies would have been eligible for inclusion in this review.

The instrument to assess methodological quality is adapted from Gordon et al. [28] and Lagravère et al. [81]. The majority of interrater disagreements arose in the assessment of applicability of criteria E, I and K to certain studies (similar baseline characteristics, blind measurement and dropouts included in data analysis respectively). This can be explained by the absence of adequate instructions of this QAI together with the presence of a wide variety of study designs. Therefore raters should test this QAI thoroughly and obtain consensus before scoring. In literature, there is not one single tool that is an obvious candidate for assessment of methodological quality of non-randomized studies [82]. Attempts to validate QAI's like the Newcastle-Ottawa [83] scale or the Jadad scale [84] are found to produce highly arbitrary results and are unable to demonstrate significant effects on quality

scores [85,86]. There is a need for a validated quality assessment instrument preferably applicable to a wide range of study designs. Furthermore, published studies are very often incomplete, cryptic, or written in a form unsuitable for quality assessment [87]. In order to overcome this drawbacks in future review studies, it should be recommended to publish only complete, unambiguous reports.

## Conclusions

Current 3D imaging techniques have not yet demonstrated their full potential to be used for quantitative longitudinal assessment of facial dimensions in children until six years of age. So far, stereophotogrammetry has been validated and has shown to be reliable and accurate. Its fast image capture, archival capabilities for subsequent morphometric studies, good-resolution color representation and no exposure to ionizing radiation make stereophotogrammetry at present the best 3D method for quantitative longitudinal assessment of facial dimensions in children until six years of age.

## Supporting Information

**Figure S1** PRISMA checklist. (DOC)

## Acknowledgments

We would like to thank senior librarian health sciences Elmie Peters for supporting the development of search strategies.

## Author Contributions

Conceived and designed the experiments: SB AK. Performed the experiments: SB MvB. Analyzed the data: SB EB. Wrote the paper: SB. Intellectual evaluation: JD TM SJB.

## References

- Ferring V, Pancherz H (2008) Divine proportions in the growing face. *Am J Orthod Dentofacial Orthop* 134(4): 472–479.
- van Loon B, Maal TJ, Plooi JM, Ingels KJ, Borstlap WA, et al. (2010) 3D Stereophotogrammetric assessment of pre- and postoperative volumetric changes in the cleft lip and palate nose. *Int J Oral Maxillofac Surg* 39(6): 534–540.
- Cole TJ, Williams AF, Wright CM (2011) Revised birth centiles for weight, length and head circumference in the UK-WHO growth charts. *Ann Hum Biol* 38(1): 7–11.
- Holdaway RA (1983) A soft-tissue cephalometric analysis and its use in orthodontic treatment planning. Part I. *Am J Orthod* 84(1): 1–28.
- Farkas LG (1994) Anthropometry of the head and face. second ed. New York: Raven Press.
- Aldridge K, Boyadjiev SA, Capone ST, DeLeon VB, Richtsmeier JT (2005) Precision and error of three-dimensional phenotypic measures acquired from 3dMD photogrammetric images. *Am J Med Genet* 138A(3): 247–253.
- Bartzela TN, Katsaros C, Bronkhorst EM, Rizell S, Halazonetis D, et al. (2011) A two-centre study on facial morphology in patients with complete bilateral cleft lip and palate at nine years of age. *Int J Oral Maxillofac Surg* 40(8): 782–789.
- Broadbent BH (1931) A new X-ray technique and its application to orthodontia. *Angle Orthod* 1(2): 45–66.
- Nollet PJ, Katsaros C, Huyskens RW, Borstlap WA, Bronkhorst EM, et al. (2008) Cephalometric evaluation of long-term craniofacial development in unilateral cleft lip and palate patients treated with delayed hard palate closure. *Int J Oral Maxillofac Surg* 37(2): 123–130.
- Davis JP, Valentine T, Davis RE (2010) Computer assisted photo-anthropometric analyses of full-face and profile facial images. *Forensic Sci Int* 200(1–3): 165–176.
- Farkas LG, Bryson W, Klotz J (1980) Is photogrammetry of the face reliable? *Plast Reconstr Surg* 66(3): 346–355.
- Cevidane LH, Alhadidi A, Paniagua B, Styner M, Ludlow J, et al. (2009) Three-dimensional quantification of mandibular asymmetry through cone-beam computerized tomography. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod.* 2011;111(6): 757–770.
- Nada RM, Maal TJ, Breuning KH, Berge SJ, Mostafa YA (2011) Accuracy and reproducibility of voxel based superimposition of cone beam computed tomography models on the anterior cranial base and the zygomatic arches. *PLoS One*; 6(2): e16520.
- Djordjevic J, Toma AM, Zhurov AI, Richmond S (2011) Three-dimensional quantification of facial symmetry in adolescents using laser surface scanning. In: *Eur J Orthod.*; Epub ahead of print.
- Toma AM, Zhurov A, Playle R, Richmond S (2008) A three-dimensional look for facial differences between males and females in a British-Caucasian sample aged 15/2 years old. *Orthod Craniofac Res.* 11(3): 180–185.
- Kau CH, Kamel SG, Wilson J, Wong ME (2011) New method for analysis of facial growth in a pediatric reconstructed mandible. *Am J Orthod Dentofacial Orthop* 139(4): e285–290.
- Maal TJ, van Loon B, Plooi JM, Rangel F, Ettema AM, et al. (2010) Registration of 3-dimensional facial photographs for clinical use. *J Oral Maxillofac Surg* 68(10): 2391–2401.
- van Loon B, van Heerbeck N, Maal TJ, Borstlap WA, Ingels KJ, et al. (2011) Postoperative volume increase of facial soft tissue after percutaneous versus endonasal osteotomy technique in rhinoplasty using 3D stereophotogrammetry. *Rhinology* 49(1): 121–126.
- Littlefield TR, Kelly KM, Cherney JC, Beals SP, Pomatto JK (2004) Technical strategies: development of a new three-dimensional cranial imaging system. *J Craniofac Surg* 15(1): 175–181.
- Hajeer MY, Millett DT, Ayoub AF, Siebert JP (2004) Applications of 3D imaging in orthodontics: part I. *J Orthod* 31(1): 62–70.
- Albarakati SF, Kula KS, Ghoneima AA. (2012) The reliability and reproducibility of cephalometric measurements: a comparison of conventional and digital methods. *Dentomaxillofac Radiol* 41(1): 11–17.
- Aynechi N, Larson BE, Leon-Salazar V, Beiraghi S (2011) Accuracy and precision of a 3D anthropometric facial analysis with and without landmark labeling before image acquisition. *Angle Orthod* 81(2): 245–252.
- El-Beialy AR, Fayed MS, El-Bialy AM, Mostafa YA (2011) Accuracy and reliability of cone-beam computed tomography measurements: Influence of head orientation. *Am J Orthod Dentofacial Orthop* 140(2): 157–165.
- De Menezes M, Rosati R, Ferrario VF, Sforza C (2010) Accuracy and reproducibility of a 3-dimensional stereophotogrammetric imaging system. *J Oral Maxillofac Surg* 68(9): 2129–2135.



25. Wong JY, Oh AK, Ohta E, Hunt AT, Rogers GF, et al. (2008) Validity and reliability of craniofacial anthropometric measurement of 3D digital photogrammetric images. *Cleft Palate Craniofac J* 45(3): 232–239.
26. PROSPERO website. Available: <http://www.crd.york.ac.uk/prosperto/>. Accessed 2012 Jul 16.
27. Boluyt N, Tjosvold L, Lefeuvre C, Klassen TP, Offringa M (2008) Usefulness of systematic review search strategies in finding child health systematic reviews in MEDLINE. *Arch Pediatr Adolesc Med* 162(2): 111–116.
28. Gordon JM, Rosenblatt M, Witmans M, Carey JP, Heo G, et al. (2009) Rapid palatal expansion effects on nasal airway dimensions as measured by acoustic rhinometry. A systematic review. *Angle Orthod* 79(5): 1000–1007.
29. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. *Biometrics* 33: 159–174.
30. The PRISMA Statement website. Available from: <http://www.prisma-statement.org>. Accessed 2012 Jul 16.
31. Bishara SE, Jakobsen JR, Hession TJ, Treder JE (1998) Soft tissue profile changes from 5 to 45 years of age. *Am J Orthod Dentofacial Orthop* 114(6): 698–706.
32. Bishara SE, Jakobsen JR, Vorhies B, Bayati P (1997) Changes in dentofacial structures in untreated Class II division 1 and normal subjects: a longitudinal study. *Angle Orthod* 67(1): 55–66.
33. Bishara SE, Hession TJ, Peterson LC (1985) Longitudinal soft-tissue profile changes: a study of three analyses. *Am J Orthod* 88(3): 209–223.
34. Bishara SE, Jakobsen JR (1985) Longitudinal changes in three normal facial types. *Am J Orthod* 88(6): 466–502.
35. Bishara SE, Peterson LC, Bishara EC (1984) Changes in facial dimensions and relationships between the ages of 5 and 25 years. *Am J Orthod* 85(3): 233–252.
36. Bongaarts CA, Prah-Andersen B, Bronkhorst EM, Prah C, Ongkosuwito EM, et al. (2009) Infant orthopedics and facial growth in complete unilateral cleft lip and palate until six years of age (Dutchcleft). *Cleft Palate Craniofac J* 46(6): 654–663.
37. Coccato PJ, Pruzansky S (1965) Longitudinal study of skeletal and soft tissue profile in children with unilateral cleft lip and palate. *Cleft Palate Craniofac J* 45: 1–12.
38. Daskalogiannakis J, Dijkman G, Kuijpers-Jagtman AM, Ross RB (2006) Comparison of facial morphology in two populations with complete unilateral cleft lip and palate from two different centers. *Cleft Palate-Craniofac J* 43(4): 471–476.
39. Graber LW (1977) Chin cup therapy for mandibular prognathism. *Am J Orthod* 72(1): 23–41.
40. Hanada K, Krogman WM (1975) A longitudinal study of postoperative changes in the soft-tissue profile in bilateral cleft lip and palate from birth to 6 years. *Am J Orthod* 67(4): 363–376.
41. Hermann NV, Darvann TA, Jensen BL, Dahl E, Bolund S, et al. (2004) Early craniofacial morphology and growth in children with bilateral complete cleft lip and palate. *Cleft Palate Craniofac J* 41(4): 424–438.
42. Hermann NV, Kreiborg S, Darvann TA, Jensen BL, Dahl E, et al. (2003) Craniofacial morphology and growth comparisons in children with Robin Sequence, isolated cleft palate, and unilateral complete cleft lip and palate. *Cleft Palate Craniofac J* 40(4): 373–396.
43. Hermann NV, Kreiborg S, Darvann TA, Jensen BL, Dahl E, et al. (2003) Early craniofacial morphology and growth in children with nonsyndromic Robin Sequence. *Cleft Palate Craniofac J* 40(2): 131–143.
44. Hermann NV, Kreiborg S, Darvann TA, Jensen BL, Dahl E, et al. (2002) Early craniofacial morphology and growth in children with unoperated isolated cleft palate. *Cleft Palate Craniofac J* 39(6): 604–622.
45. Hermann NV, Jensen BL, Dahl E, Bolund S, Darvann TA, et al. (1999) Craniofacial growth in subjects with unilateral complete cleft lip and palate, and unilateral incomplete cleft lip, from 2 to 22 months of age. *J Craniofac Genet Dev Biol* 19(3): 135–147.
46. Padwa BL, Sonis A, Bagheri S, Mulliken JB (1999) Children with repaired bilateral cleft lip/palate: effect of age at premaxillary osteotomy on facial growth. *Plast Reconstr Surg* 104(5): 1261–1269.
47. Posen JM (1967) A longitudinal study of the growth of the nose. *Am J Orthod* 58(10): 746–756.
48. Sadowsky C, Aduss H, Pruzansky S (1973) The soft tissue profile in unilateral clefts. *Angle Orthod* 43(3): 233–246.
49. Semb G (1991) A study of facial growth in patients with bilateral cleft lip and palate treated by the Oslo CLP Team. *Cleft Palate Craniofac J* 28(1): 22–39; discussion 46–28.
50. Semb G (1991) A study of facial growth in patients with unilateral cleft lip and palate treated by the Oslo CLP Team. *Cleft Palate Craniofac J* 28(1): 1–21; discussion 46–28.
51. Smahel Z, Mullerova Z (1995) Facial growth and development in unilateral cleft lip and palate from the time of palatoplasty to the onset of puberty: a longitudinal study. *J Craniofac Genet Dev Biol* 15(2): 72–80.
52. Subtelny JD (1959) A longitudinal study of soft tissue facial structures and their profile characteristics, defined in relation to underlying skeletal structures. *Am J Orthod* 45(7): 481–507.
53. Wen-Ching Ko E, Figueroa AA, Polley JW (2000) Soft tissue profile changes after maxillary advancement with distraction osteogenesis by use of a rigid external distraction device: a 1-year follow-up. *J Oral Maxillofac Surg* 58(9): 959–969; discussion 969–970.
54. Zettergren-Wijk L, Forsberg CM, Linder-Aronson S (2006) Changes in dentofacial morphology after adeno-/tonsillectomy in young children with obstructive sleep apnoea—a 5-year follow-up study. *Eur J Orthod* 28(4): 319–326.
55. Altug-Atac AT, Grayson BH, McCarthy JG (2008) Comparison of skeletal and soft-tissue changes following unilateral mandibular distraction osteogenesis. *Plast Reconstr Surg* 121(5): 1751–1759.
56. Castelo PM, Pereira LJ, Bon Jardim LR, Gavião MB (2010) Changes in bite force, masticatory muscle thickness, and facial morphology between primary and mixed dentition in preschool children with normal occlusion. *Ann Anat* 192(1): 23–26.
57. Cruz AAV, Akaiishi PMS, Arnaud E, Marchac D, Renier D (2008) Palpebral fissure changes after monoblock frontofacial advancement in faciocraniosynostosis. *Journal of Craniofacial Surgery* 19(1): 106–109.
58. Ko EW, Hung KF, Huang CS, Chen PK (2004) Correction of facial asymmetry with multiplanar mandible distraction: a one-year follow-up study. *Cleft Palate Craniofac J* 41(1): 5–12.
59. Kohout MP, Aljaro LM, Farkas LG, Mulliken JB (1998) Photogrammetric comparison of two methods for synchronous repair of bilateral cleft lip and nasal deformity. *Plast Reconstr Surg* 102(5): 1339–1349.
60. Liou EJ, Subramanian M, Chen PK (2007) Progressive changes of columella length and nasal growth after nasoalveolar molding in bilateral cleft patients: a 3-year follow-up study. *Plast Reconstr Surg* 119(2): 642–648.
61. Pai BC, Ko EW, Huang CS, Liou EJ (2005) Symmetry of the nose after presurgical nasoalveolar molding in infants with unilateral cleft lip and palate: a preliminary study. *Cleft Palate Craniofac J* 42(6): 658–663.
62. Schuler G (2007) Head proportion and shape of the head of children between 2 and 7 years—results of a longitudinal study. *Anthropol Anz* 65(2): 203–212.
63. Sultana R, Matsuo K, Yuzuriha S, Kushima H (2000) Disinsertion of the levator aponeurosis from the tarsus in growing children. *Plast Reconstr Surg* 106(3): 563–570.
64. Bennun RD, Perandones C, Sepiarsky VA, Chantiri SN, Ines Ulfe Aguirre M, et al. (1999) Nonsurgical correction of nasal deformity in unilateral complete cleft lip: A 6-year follow-up. *Plastic and Reconstructive Surgery* 104(3): 616–630.
65. Hansen M, Padwa BL, Scott RM, Stieg PE, Mulliken JB (1997) Synostotic frontal plagiocephaly: anthropometric comparison of three techniques for surgical correction. *Plast Reconstr Surg* 100(6): 1387–1395.
66. Heimer MV, Tornisiello Katz CR, Rosenblatt A (2008) Non-nutritive sucking habits, dental malocclusions, and facial morphology in Brazilian children: a longitudinal study. *Eur J Orthod* 30(6): 580–585.
67. Ridgway EB, Wu JK, Sullivan SR, Vasudavan S, Padwa BL, et al. (2011) Craniofacial growth in patients with FGFR3Pro250Arg mutation after fronto-orbital advancement in infancy. *J Craniofac Surg* 22(2): 455–461.
68. Vander Woude DL, Mulliken JB (1997) Effect of lip adhesion on labial height in two-stage repair of unilateral complete cleft lip. *Plast Reconstr Surg* 100(3): 567–572; discussion 573–564.
69. Yang X, Mu X, Gu Q, Yu Z, Ding M, et al. (2009) Oriental surgical protocol in orbital hypertelorism. *J Craniofac Surg* 20(3): 718–723.
70. Ezzat CF, Chavarria C, Teichgraber JF, Chen J, Stratmann RG, et al. (2007) Presurgical nasoalveolar molding therapy for the treatment of unilateral cleft lip and palate: a preliminary study. *Cleft Palate-Craniofac J* 44(1): 8–12.
71. Primozic J, Ovsenik M, Richmond S, Kau CH, Zhurov A (2009) Early crossbite correction: a three-dimensional evaluation. *Eur J Orthod* 31(4): 352–356.
72. Schwenzer-Zimmerer K, Chaitidis D, Berg-Boerner I, Krol Z, Kovacs L, et al. (2008) Quantitative 3D soft tissue analysis of symmetry prior to and after unilateral cleft lip repair compared with non-cleft persons (performed in Cambodia). *J Craniomaxillofac Surg* 36(8): 431–438.
73. Hoefert CS, Bacher M, Herberts T, Krimmel M, Reinert S, et al. (2010) 3D Soft Tissue Changes in Facial Morphology in Patients with Cleft Lip and Palate and Class III Malocclusion under Therapy with Rapid Maxillary Expansion and Delaire Facemask. *J Orofac Orthop* 71(2): 136–151.
74. Hood CA, Bock M, Hosey MT, Bowman A, Ayoub AF (2003) Facial asymmetry—3D assessment of infants with cleft lip & palate. *Int J Paediatr Dent* 13(6): 404–410.
75. Ras F, Habets LL, van Ginkel FC, Prah-Andersen B (1995) Longitudinal study on three-dimensional changes of facial asymmetry in children between 4 to 12 years of age with unilateral cleft lip and palate. *Cleft Palate Craniofac J* 32(6): 463–468.
76. Singh GD, Levy-Bercowski D, Santiago PE (2005) Three-dimensional nasal changes following nasoalveolar molding in patients with unilateral cleft lip and palate: Geometric morphometrics. *Cleft Palate-Craniofac J* 42(4): 403–409.
77. Seidenstricker-Kink LM, Becker DB, Govier DP, DeLeon VB, Lo LJ, et al. (2008) Comparative osseous and soft tissue morphology following cleft lip repair. *Cleft Palate Craniofac J* 45(5): 511–517.
78. Ayoub A, Garrahy A, Hood C, White J, Bock M, et al. (2003) Validation of a vision-based, three-dimensional facial imaging system. *Cleft Palate Craniofac J* 40(5): 523–529.
79. Swennen GR, Grimaldi H, Berten JL, Kramer FJ, Dempf R, et al. (2004) Reliability and validity of a modified lateral cephalometric analysis for evaluation of craniofacial morphology and growth in patients with clefts. *J Craniofac Surg* 15(3): 399–412.
80. Robinson KA, Dickersin K (2002) Development of a highly sensitive search strategy for the retrieval of reports of controlled trials using PubMed. *Int J Epidemiol* 31(1): 150–153.

81. Lagravère MO, Major PW, Flores-Mir C (2005) Long-term skeletal changes with rapid maxillary expansion: a systematic review. *Angle Orthod* 75(6): 1046–1052.
82. Sanderson S, Tatt ID, Higgins JP (2007) Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: a systematic review and annotated bibliography. *Int J Epidemiol* 36(3): 666–676.
83. The Newcastle-Ottawa Scale website. Available: [http://www.ohri.ca/programs/clinical\\_epidemiology/oxford.asp](http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp). Accessed 2012 Jul 16.
84. Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, et al. (1996) Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials* 17(1): 1–12.
85. Stang A (2010) Critical evaluation of the Newcastle-Ottawa scale for the assessment of the quality of nonrandomized studies in meta-analyses. *Eur J Epidemiol* 25(9): 603–605.
86. Clark HD, Wells GA, Huet C, McAlister FA, Salmi LR, et al. (1999) Assessing the quality of randomized trials: reliability of the Jadad scale. *Control Clin Trials* 20(5): 448–452.
87. Hodges JR (2004) Are quality assessment instruments any good? *J Evid Based Dent Pract* 4: 24–31.