Methods to Quantify Soft Tissue–Based Cranial Growth and Treatment Outcomes in Children: A Systematic Review

Sander Brons¹*, Machteld E. van Beusichem¹, Ewald M. Bronkhorst², Jos M. Draaisma³, Stefaan J. Bergé⁴, Jan G. Schols¹, Anne Marie Kuijpers-Jagtman¹

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

Abstract

**Context:** Longitudinal assessment of cranial dimensions of growing children provides healthcare professionals with information about normal and deviating growth as well as treatment outcome.

**Objective:** To give an overview of soft tissue–based methods for quantitative longitudinal assessment of cranial dimensions in children until age 6 years 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 manual search was performed to check for additional relevant studies.

**Study selection:** Primary publications on facial growth and treatment outcomes in children younger than age 6 years were included.

**Data extraction:** Independent data extraction was performed by two observers. A quality assessment instrument was used to determine 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, 165 studies were included, forming three groups of methods: head circumference anthropometry, direct anthropometry, and 2D photography and 3D imaging techniques (surface laser scanning and stereophotogrammetry). In general, the measurement error was below 2 mm, and correlation coefficients were very good.

**Conclusion:** Various methods for measuring cranial dimensions have shown to be reliable. Stereophotogrammetry is the most versatile method for quantitative longitudinal assessment of cranial dimensions and shapes in children. However, direct anthropometry continues to be the best method for routine clinical assessments of linear cranial dimensions in growing children until age 6 years.


Editor: Samuel J Lin, Harvard Medical School, United States of America

Received June 4, 2013; Accepted January 23, 2014; Published February 27, 2014

Copyright: © 2014 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 number 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 assessment of cranial dimensions of growing children provides healthcare professionals with information about normal and deviating growth as well as treatment outcome, for example in cases of deformational plagiocephaly and craniosynostosis [1,2]. Accurate quantitative evaluation of cranial 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, as in cases of autism and hydrocephalus [3,4].

Various methods for quantitative evaluation of craniofacial dimensions have been described for a variety of purposes. The standard technique is direct anthropometry, which has been extensively used for the study of craniofacial dimensions in the past century [3]. Two-dimensional (2D) x-ray cephalometry [6–8] and photography [9,10] also have been applied for decades and even today are the most commonly used records for dento-skeletal and facial diagnosis. Recent technological advancements have led craniofacial researchers and clinicians into the era of three dimensional (3D) digital imaging. Techniques like cone beam computed tomography [11,12], surface laser scanning [13,14], and
stereophotogrammetry [15–17] have become available for describing and comparing craniofacial dimensions and shapes, making a diagnosis or planning treatment, and evaluating growth and treatment outcomes.

In an earlier systematic review, we described various methods for quantitative evaluation of facial dimensions in children up to age 6 years for a variety of purposes [18]. This study describes the methods for quantitative evaluation of cranial dimensions. Its aims are to 1) give an overview of soft tissue–based methods for quantitative longitudinal assessment of cranial dimensions in children up to age 6 years; 2) assess the methodological quality of the studies using such approaches; and 3) assess the reliability of these methods applied 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. A registration number is not available for this review since PROSPERO [19] was still in development when it was performed.

**Eligibility criteria**

Eligible for inclusion were primary publications reporting on 1) soft tissue–based evaluation of the head and face; 2) children under age 6 years at the start of the study; 3) quantitative changes; and 4) longitudinal studies.

Excluded were publications describing 1) skeletal changes, 2) fetal growth, 3) animal studies, or 4) cross-sectional studies or featuring 5) 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, 2012. In addition, we manually 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 in which the full-text publication could not be retrieved, authors were requested by email to provide the article. The gray literature was not searched.

**Search strategy**

The search strategy was developed and databases 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 were directly derived from the former. The last search was performed on October 1, 2012.

The search strategy focused on four categories of terms, as follows: (1) terms to search for the population of interest (i.e., babies, infants, and pre-school children), for which 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 [20]; (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); and (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 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 the same two reviewers in a standardized manner. Again, in an additional step, disagreements were resolved by discussion and consensus. Results of both first and second independently performed assessments of eligibility were analyzed to assess inter-rater reliability. Third, the first author performed a manual search of the reference lists of the included studies. Finally, all included studies were categorized as describing facial or cranial evaluation of growth and treatment outcome. The plane connecting the glabella with the left and right euryon arbitrarily separates the cranium from the face. Measurements on or above this plane were considered to be cranial and those below this plane to be facial. Studies describing cranial evaluation of growth and treatment are included in this review.

**Quality assessment**

Study quality was assessed by the quality assessment instrument (QAI) for clinical trials used by Gordon et al. (Table 2) [21]. This instrument includes an assessment of study bias. A checklist was made 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, fewer 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 methods for measurement of cranial growth or treatment outcome. A mean quality score for each group of methods was calculated. Arbitrarily, a cut-off of 60% or higher was graded as good quality and below 60% as poor quality. To assess the inter-rater 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 inter-rater agreement for the process of study selection and for each criterion of the QAI. According to Landis and Koch, the level of inter-rater 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 [22].

Analysis of variance and non-parametric Kruskal-Wallis tests were performed to evaluate differences in mean scores between groups of methods. SPSS version 21.0 was used for analysis.

**Results**

**Study selection**

The inter-rater kappa for screening on title and abstract was 0.76. For full-text assessment of eligibility, the kappa was 0.69. The reliability of both steps in the process of study selection is qualified as good [22].
The search of PubMed, EMBASE, Cochrane Library, Web of Science, Scopus, and CINAHL provided a total of 7027 citations, and the manual search provided 198 citations. After adjusting for duplicates, 5599 citations remained for screening of title and abstract. Of these, 4490 studies were discarded because they did not meet the eligibility criteria so that a total of 1109 studies remained for full-text assessment. Of these, 897 studies were excluded for different reasons, and 195 were discarded because the full-text publication could not be retrieved. We ultimately identified 212 studies that met the inclusion criteria. The last step in the inclusion process divided the studies into those on facial evaluation (n = 47) and studies on cranial evaluation (n = 165). A total of 188 studies originated from the electronic databases; the remaining 24 studies originated from the additional manual search of the references of the included studies. Figure 1 shows the PRISMA flow diagram, and Checklist S1 shows the PRISMA checklist [23]. The current systematic review is restricted to studies on cranial evaluation of growth and treatment outcomes in

Table 1. Search strategy PubMed.

<table>
<thead>
<tr>
<th>Search strategy PubMed</th>
</tr>
</thead>
</table>

Table 2. Quality assessment instrument [21].

<table>
<thead>
<tr>
<th>I. Study design (7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Objective—objective clearly formulated (☑)</td>
</tr>
<tr>
<td>B. Sample size—considered adequate (☑)</td>
</tr>
<tr>
<td>C. Sample size—estimated before collection of data (☑)</td>
</tr>
<tr>
<td>D. Selection criteria—clearly described (☑)</td>
</tr>
<tr>
<td>E. Baseline characteristics—similar baseline characteristics (☑)</td>
</tr>
<tr>
<td>F. Timing—prospective (☑)</td>
</tr>
<tr>
<td>G. Randomization—stated (☑)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>II. Study measurements (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>H. Measurement method—appropriate to the objective (☑)</td>
</tr>
<tr>
<td>I. Blind measurement—blinding (☑)</td>
</tr>
<tr>
<td>J. Reliability—adequate level of agreement (☑)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>III. Statistical analysis (5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>K. Dropouts—dropouts included in data analysis (☑)</td>
</tr>
<tr>
<td>L. Statistical analysis—appropriate for data (☑)</td>
</tr>
<tr>
<td>M. Confounders—confounders included in analysis (☑)</td>
</tr>
<tr>
<td>N. Statistical significance level—p value stated (☑)</td>
</tr>
<tr>
<td>O. Confidence intervals provided (☑)</td>
</tr>
</tbody>
</table>

Maximum number of ☑ = 15

doi:10.1371/journal.pone.0089602.t001
doi:10.1371/journal.pone.0089602.t002
children; a systematic review of studies on facial evaluation of growth and treatment outcomes in children is described in a separate publication [17].

Of the 165 included studies, 136 studies used direct anthropometry for head circumference, 19 studies used other direct cranial anthropometry approaches, 3 studies used 2D photography, and 7 studies used 3D imaging techniques (5 stereophotogrammetry and 2 surface laser scanning).

Study quality assessment

Inter-rater reliability values for all 15 criteria of the QAI were between kappa 0.19 and 1.00; 11 out of 15 criteria had a kappa of 0.50 or higher. Inter-rater agreement on criteria E (similar baseline characteristics), I (blind measurement), and K (dropouts included in data analysis) was below 0.20.

Assessment of methodological quality of all reviewed studies resulted in scores ranging from 20% to 100%. A total of 118 studies qualified as good based on a methodological quality score equal to or above 60%. Score summaries of studies with good methodological quality using direct anthropometry for head
<table>
<thead>
<tr>
<th>First author</th>
<th>Year</th>
<th>Design</th>
<th>Measure</th>
<th>Statistics</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Belfort</td>
<td>2012</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td>64%</td>
</tr>
<tr>
<td>Bendersky</td>
<td>1998</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td>75%</td>
</tr>
<tr>
<td>Berry</td>
<td>1997</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td>64%</td>
</tr>
<tr>
<td>Bhalla</td>
<td>1993</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td>60%</td>
</tr>
<tr>
<td>Binns</td>
<td>1996</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td>70%</td>
</tr>
<tr>
<td>Bouthoom</td>
<td>2012</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td>75%</td>
</tr>
<tr>
<td>Bracewell</td>
<td>2007</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td>67%</td>
</tr>
<tr>
<td>Butz</td>
<td>2005</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td>69%</td>
</tr>
<tr>
<td>Cardoso</td>
<td>2007</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td>67%</td>
</tr>
<tr>
<td>Carlson</td>
<td>1996</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td>80%</td>
</tr>
<tr>
<td>Chaudhari</td>
<td>2012</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td>69%</td>
</tr>
<tr>
<td>De Bruin</td>
<td>1998</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td>64%</td>
</tr>
<tr>
<td>DeFeignier</td>
<td>1996</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td>69%</td>
</tr>
<tr>
<td>Desmytteere</td>
<td>2009</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td>79%</td>
</tr>
<tr>
<td>Desmytteere</td>
<td>2009</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td>71%</td>
</tr>
<tr>
<td>Donna</td>
<td>1997</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td>64%</td>
</tr>
<tr>
<td>D’Souza</td>
<td>1986</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td>62%</td>
</tr>
<tr>
<td>Durmus</td>
<td>2011</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td>85%</td>
</tr>
<tr>
<td>Ekblad</td>
<td>2010</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td>77%</td>
</tr>
<tr>
<td>Elwood</td>
<td>1987</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td>64%</td>
</tr>
<tr>
<td>Erasmus</td>
<td>2002</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td>80%</td>
</tr>
<tr>
<td>Ergiee</td>
<td>2001</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td>73%</td>
</tr>
<tr>
<td>Ernst</td>
<td>1990</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td>64%</td>
</tr>
<tr>
<td>Faroqqi</td>
<td>2006</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td>77%</td>
</tr>
<tr>
<td>Ford</td>
<td>2000</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td>85%</td>
</tr>
<tr>
<td>Ford</td>
<td>1986</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td>64%</td>
</tr>
<tr>
<td>Friel</td>
<td>1993</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td>73%</td>
</tr>
<tr>
<td>Fukumoto</td>
<td>2008</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td>60%</td>
</tr>
<tr>
<td>Gale</td>
<td>2006</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td>77%</td>
</tr>
<tr>
<td>Gale</td>
<td>2004</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td>70%</td>
</tr>
<tr>
<td>Georgieff</td>
<td>1995</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td>62%</td>
</tr>
<tr>
<td>Georgieff</td>
<td>1989</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td>60%</td>
</tr>
<tr>
<td>Gross</td>
<td>1983</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td>62%</td>
</tr>
<tr>
<td>Gross</td>
<td>1983</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td>75%</td>
</tr>
<tr>
<td>Guo</td>
<td>1988</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td>60%</td>
</tr>
<tr>
<td>Hagelberg</td>
<td>1990</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td>67%</td>
</tr>
<tr>
<td>Hansen-pupp</td>
<td>2011</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td>67%</td>
</tr>
<tr>
<td>Herrmann</td>
<td>2010</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td>60%</td>
</tr>
<tr>
<td>Ishikawa</td>
<td>1987</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td>69%</td>
</tr>
<tr>
<td>Jaffe</td>
<td>1992</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td>60%</td>
</tr>
<tr>
<td>Jaldin</td>
<td>2011</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td>73%</td>
</tr>
<tr>
<td>Jaruratanasirikul</td>
<td>1999</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td>62%</td>
</tr>
<tr>
<td>Kan</td>
<td>2007</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td>77%</td>
</tr>
<tr>
<td>Karatza</td>
<td>2003</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td>77%</td>
</tr>
<tr>
<td>Kiran</td>
<td>2007</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td>79%</td>
</tr>
<tr>
<td>First author</td>
<td>Year</td>
<td>Design</td>
<td>Measure</td>
<td>Statistics</td>
<td>Score</td>
</tr>
<tr>
<td>---------------</td>
<td>------</td>
<td>--------</td>
<td>---------</td>
<td>------------</td>
<td>-------</td>
</tr>
<tr>
<td>Kitchen</td>
<td>1992</td>
<td>Y Y o</td>
<td>Y o</td>
<td>Y o Y Y Y</td>
<td>67%</td>
</tr>
<tr>
<td>Koo</td>
<td>2006</td>
<td>Y Y Y Y Y Y Y Y Y</td>
<td>Y Y Y Y Y</td>
<td>87%</td>
<td></td>
</tr>
<tr>
<td>Lainhart</td>
<td>1997</td>
<td>Y Y o Y Y Y Y Y</td>
<td>Y Y Y</td>
<td>75%</td>
<td></td>
</tr>
<tr>
<td>Lasekan</td>
<td>2011</td>
<td>Y Y o Y Y Y Y Y</td>
<td>Y Y Y</td>
<td>73%</td>
<td></td>
</tr>
<tr>
<td>Lasekan</td>
<td>2006</td>
<td>Y Y Y Y Y Y Y Y</td>
<td>Y Y Y</td>
<td>87%</td>
<td></td>
</tr>
<tr>
<td>Lira</td>
<td>2009</td>
<td>Y Y o Y Y Y Y</td>
<td>Y Y Y</td>
<td>75%</td>
<td></td>
</tr>
<tr>
<td>Lucas</td>
<td>2001</td>
<td>Y Y Y Y Y Y Y Y</td>
<td>Y Y Y</td>
<td>87%</td>
<td></td>
</tr>
<tr>
<td>Maguire</td>
<td>2008</td>
<td>Y Y Y Y Y Y Y Y</td>
<td>Y Y Y</td>
<td>87%</td>
<td></td>
</tr>
<tr>
<td>Makrides</td>
<td>2000</td>
<td>Y Y Y Y Y Y Y Y</td>
<td>Y Y Y</td>
<td>93%</td>
<td></td>
</tr>
<tr>
<td>Makrides</td>
<td>1999</td>
<td>Y Y Y Y Y Y Y Y</td>
<td>Y Y Y</td>
<td>60%</td>
<td></td>
</tr>
<tr>
<td>Mambabolo</td>
<td>2004</td>
<td>Y Y o Y Y o</td>
<td>Y Y Y</td>
<td>60%</td>
<td></td>
</tr>
<tr>
<td>Marks</td>
<td>1979</td>
<td>Y Y o Y Y o</td>
<td>Y Y Y</td>
<td>60%</td>
<td></td>
</tr>
<tr>
<td>Maserei</td>
<td>2007</td>
<td>Y o Y Y Y Y Y Y</td>
<td>Y Y o</td>
<td>87%</td>
<td></td>
</tr>
<tr>
<td>Mathur</td>
<td>2009</td>
<td>Y Y o Y Y Y Y</td>
<td>Y Y Y</td>
<td>62%</td>
<td></td>
</tr>
<tr>
<td>McCowan</td>
<td>1999</td>
<td>Y Y o Y Y Y Y</td>
<td>Y Y o</td>
<td>70%</td>
<td></td>
</tr>
<tr>
<td>McLeod</td>
<td>2011</td>
<td>Y Y o Y Y o</td>
<td>Y Y o</td>
<td>64%</td>
<td></td>
</tr>
<tr>
<td>Mercuri</td>
<td>2000</td>
<td>Y Y o Y Y o</td>
<td>Y Y Y</td>
<td>64%</td>
<td></td>
</tr>
<tr>
<td>Meyer-Marcotty</td>
<td>2012</td>
<td>Y Y o Y Y Y</td>
<td>Y Y Y</td>
<td>62%</td>
<td></td>
</tr>
<tr>
<td>Moore</td>
<td>1995</td>
<td>Y Y o Y Y o</td>
<td>Y Y o</td>
<td>62%</td>
<td></td>
</tr>
<tr>
<td>Moje</td>
<td>1993</td>
<td>Y Y o Y Y o</td>
<td>Y Y o</td>
<td>69%</td>
<td></td>
</tr>
<tr>
<td>Nelson</td>
<td>1997</td>
<td>Y Y o Y Y o</td>
<td>Y Y o</td>
<td>62%</td>
<td></td>
</tr>
<tr>
<td>Ochiai</td>
<td>2008</td>
<td>Y Y o Y Y o</td>
<td>Y Y o</td>
<td>73%</td>
<td></td>
</tr>
<tr>
<td>Olivan</td>
<td>2003</td>
<td>Y Y o Y Y o</td>
<td>Y Y o</td>
<td>64%</td>
<td></td>
</tr>
<tr>
<td>Oliveira</td>
<td>2007</td>
<td>Y Y o Y Y o</td>
<td>Y Y o</td>
<td>67%</td>
<td></td>
</tr>
<tr>
<td>Padilla</td>
<td>2010</td>
<td>Y Y Y Y Y Y Y Y</td>
<td>Y Y Y</td>
<td>85%</td>
<td></td>
</tr>
<tr>
<td>Paul</td>
<td>2008</td>
<td>Y Y Y Y Y Y Y Y</td>
<td>Y Y Y</td>
<td>72%</td>
<td></td>
</tr>
<tr>
<td>Peng</td>
<td>2005</td>
<td>Y Y o Y Y o</td>
<td>Y Y o</td>
<td>71%</td>
<td></td>
</tr>
<tr>
<td>Piemontese</td>
<td>2011</td>
<td>Y Y Y Y Y Y Y Y</td>
<td>Y Y o</td>
<td>87%</td>
<td></td>
</tr>
<tr>
<td>Polberger</td>
<td>1999</td>
<td>Y Y Y Y Y Y Y Y</td>
<td>Y o o</td>
<td>73%</td>
<td></td>
</tr>
<tr>
<td>Rijken</td>
<td>2007</td>
<td>Y Y Y Y Y Y Y Y</td>
<td>Y Y o</td>
<td>77%</td>
<td></td>
</tr>
<tr>
<td>Roberfroid</td>
<td>2012</td>
<td>Y o Y Y Y Y Y Y</td>
<td>Y Y Y</td>
<td>87%</td>
<td></td>
</tr>
<tr>
<td>Roche</td>
<td>1987</td>
<td>Y Y Y Y Y Y Y Y</td>
<td>Y Y Y</td>
<td>60%</td>
<td></td>
</tr>
<tr>
<td>Rodriguez Garcia</td>
<td>2003</td>
<td>Y Y Y Y Y Y Y Y</td>
<td>Y Y Y</td>
<td>60%</td>
<td></td>
</tr>
<tr>
<td>Ross</td>
<td>2012</td>
<td>Y Y o Y Y Y Y Y Y</td>
<td>Y o</td>
<td>67%</td>
<td></td>
</tr>
<tr>
<td>Rothenberg</td>
<td>1999</td>
<td>Y Y o Y Y o Y Y</td>
<td>Y Y Y</td>
<td>82%</td>
<td></td>
</tr>
<tr>
<td>Saliba</td>
<td>1990</td>
<td>Y Y o Y Y o Y Y</td>
<td>Y Y o</td>
<td>69%</td>
<td></td>
</tr>
<tr>
<td>Sawada</td>
<td>2010</td>
<td>Y Y o Y Y Y Y</td>
<td>Y Y Y</td>
<td>64%</td>
<td></td>
</tr>
<tr>
<td>Schaefer</td>
<td>1994</td>
<td>Y Y o Y Y o Y Y</td>
<td>Y Y o</td>
<td>62%</td>
<td></td>
</tr>
<tr>
<td>Sharma</td>
<td>2011</td>
<td>Y o o Y Y Y Y</td>
<td>Y Y Y</td>
<td>64%</td>
<td></td>
</tr>
<tr>
<td>Shaw</td>
<td>1999</td>
<td>Y Y Y Y Y Y Y Y</td>
<td>Y Y Y</td>
<td>76%</td>
<td></td>
</tr>
<tr>
<td>Sheth</td>
<td>1995</td>
<td>Y Y o Y Y o Y Y</td>
<td>Y Y o</td>
<td>60%</td>
<td></td>
</tr>
<tr>
<td>Shortland</td>
<td>1998</td>
<td>Y Y o Y Y Y Y</td>
<td>Y Y Y</td>
<td>76%</td>
<td></td>
</tr>
<tr>
<td>Tan</td>
<td>2008</td>
<td>Y Y Y Y Y Y Y Y</td>
<td>Y Y Y</td>
<td>73%</td>
<td></td>
</tr>
<tr>
<td>Tinoco</td>
<td>2009</td>
<td>Y Y Y Y Y Y Y Y</td>
<td>Y Y Y</td>
<td>70%</td>
<td></td>
</tr>
<tr>
<td>Touwslager</td>
<td>2008</td>
<td>Y Y Y Y Y Y Y Y</td>
<td>Y Y Y</td>
<td>60%</td>
<td></td>
</tr>
<tr>
<td>Vaidya</td>
<td>2008</td>
<td>Y Y Y Y Y Y Y Y</td>
<td>Y Y Y</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>Van Daalen</td>
<td>2007</td>
<td>Y Y o Y Y Y Y</td>
<td>Y Y Y</td>
<td>70%</td>
<td></td>
</tr>
</tbody>
</table>

**Table 3. Cont.**

**Head circumference anthropometry**

Methods for Cranial Growth in Children: A Review
Reliability

Scores for reliability of methods for soft tissue–based quantitative longitudinal assessment are shown in Table 6. Only 12 of the 118 studies with good methodological quality report data for the reliability of the method to quantify cranial dimensions; 8 studies used direct anthropometry for head circumference [25,27,35,58,63,71,100,103], two used other kinds of direct anthropometry [126,127], one study used 3D laser scanning [136], and one used 3D stereophotogrammetry [139].

Regarding direct anthropometry for head circumference, 5 studies with good methodological quality reported a measurement error equal to or below 1.7 mm. Two studies with good methodological quality using direct anthropometry reported correlation coefficients between repeated measurements of 0.87 and 0.99 and were both qualified as very good. Regarding other kinds of direct anthropometry, two studies with good methodological quality reported a measurement error of 1 mm.

Studies of good methodological quality using 2D photography and reporting the measurement error or correlation coefficients were not identified among the included studies. One study with good methodological quality using 3D laser scanning reported a measurement error of 0.5 mm, and another using 3D stereophotogrammetry reported a measurement error of 0.02–4.3 mm.

Discussion

Summary of evidence

The objectives of this systematic review were to 1) give an overview of soft tissue–based methods for quantitative longitudinal assessment of cranial dimensions in children until age 6 years; 2) assess the methodological quality of the studies using such approaches; and 3) assess reliability of these quantitative measurement methods used in studies with good methodological quality.

In the literature, various terms to describe 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 [142]. More often in the literature, reliability is used to describe landmark identification error of a method. Reliability can be expressed by the measurement error or a correlation coefficient between repeated measurements [9,143,144] and represents the ability of observers to make a consistent analysis. In this systematic review, reliability in studies with good methodological quality was assessed and expressed by duplicate measurement errors and correlation coefficients between repeated measurements. Direct measurement of head circumference is the most often used method for soft tissue–based evaluation of cranial growth and treatment outcomes. The use of growth charts in the clinical assessment of growing infants and children in pediatric nutritional screening and epidemiologic assessments has already been recommended for decades [27]. For this purpose, length and weight also are recorded in many countries from birth onwards. Direct anthropometry for head circumference seems to be a generally accepted method for most researchers because only 8 out of 95 studies with good methodological quality reported on its reliability. Measurement errors varied from 0.2 to 1.7 mm, and correlation coefficients were very good. Other kinds of direct anthropometry yielded a measurement error of 1 mm.

Normal growth of head circumference shows an increase for mean head circumference from 34–36 cm at birth to 51–52 by age 6 years [145]. The measurement errors for head circumference anthropometry presented in this review are within 1% of the values of normal growth, which seems to be negligible. Direct anthropometry is a reliable and cheap method to study linear dimensions and has been regarded as the gold standard for many decades, but it requires patient cooperation and precludes archiving [146].

Photographic techniques, on the other hand, can capture images for data storage. The most common imaging technique is 2D photography, which has the advantages of being safe, relatively cheap, and user friendly. However, because none of the included studies reported a measurement error or correlation coefficient, its reliability for evaluation of cranial growth and treatment outcome is uncertain.

Various 3D imaging techniques have been recently introduced and were applied in 6 out of 165 studies included in this systematic review. Only two studies with good methodological quality reported the measurement error (Table 6). The measurement error in one study using 3D laser scanning was 0.5 mm; in one study using 3D stereophotogrammetry, it was 0.02 mm for head width, 0.04 mm for head circumference, and 4.3 mm for vertex height. Therefore, 3D imaging might be a reliable method to quantify cranial dimensions. Reliability of measurements from 3D imaging seems to be more related to the exact anatomical region of interest than to the method itself.

When reviewing the literature for this study, we found only six included studies with good methodological quality using 3D imaging to quantify soft tissue–based cranial growth and treatment

Table 3. Cont.

<table>
<thead>
<tr>
<th>First author</th>
<th>Year</th>
<th>Design</th>
<th>Measure</th>
<th>Statistics</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whitehouse</td>
<td>2010</td>
<td></td>
<td>A B C D</td>
<td></td>
<td>67%</td>
</tr>
<tr>
<td>Wood</td>
<td>2003</td>
<td></td>
<td>A B C D</td>
<td></td>
<td>75%</td>
</tr>
<tr>
<td>Zabaneh</td>
<td>2011</td>
<td></td>
<td>A B C D</td>
<td></td>
<td>67%</td>
</tr>
</tbody>
</table>

A = fulfilled the methodological criteria satisfactorily.  
B = did not fulfill the methodological criteria.  
C = not applicable.  
D = not applicable.

doi:10.1371/journal.pone.0089602.t003

circumference [24–118] are shown in Table 3 (n = 95); those for other direct anthropometry [119–132] are shown in Table 4 (n = 14); and those for indirect 2D and 3D imaging techniques [133–141] are shown in Table 5 (n = 9).

Analysis of variance (p = 0.14) and Kruskal-Wallis test (p = 0.16) revealed no statistically significant difference for methodological quality, expressed as a percentage between groups of methods.
Methods for Cranial Growth in Children: A Review

Table 4. Methodological quality scores of studies using direct anthropometry reporting on soft tissue–based quantitative longitudinal assessment of cranial dimensions in children until age 6 years with a score equal to or above 60% (n = 14) [119–132].

<table>
<thead>
<tr>
<th>First author</th>
<th>Year</th>
<th>Design</th>
<th>Measure</th>
<th>Statistics</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agrawal</td>
<td>2006</td>
<td>Y Y o</td>
<td>Y o</td>
<td>Y Y Y Y Y Y</td>
<td>70%</td>
</tr>
<tr>
<td>Chatterjee</td>
<td>2009</td>
<td>Y Y o</td>
<td>Y o</td>
<td>Y Y Y Y Y Y</td>
<td>60%</td>
</tr>
<tr>
<td>Fearon</td>
<td>2009</td>
<td>Y Y o</td>
<td>Y o</td>
<td>Y Y Y Y Y Y</td>
<td>60%</td>
</tr>
<tr>
<td>Fearon</td>
<td>2006</td>
<td>Y Y o</td>
<td>Y o</td>
<td>Y Y Y Y Y Y</td>
<td>60%</td>
</tr>
<tr>
<td>Kelly</td>
<td>1999</td>
<td>Y Y o</td>
<td>Y o</td>
<td>Y Y Y Y Y Y</td>
<td>64%</td>
</tr>
<tr>
<td>Lee</td>
<td>2006</td>
<td>Y Y o</td>
<td>Y o</td>
<td>Y Y Y Y Y Y</td>
<td>70%</td>
</tr>
<tr>
<td>Lee</td>
<td>2008</td>
<td>Y Y Y Y Y Y Y Y o</td>
<td>o Y Y Y o</td>
<td>77%</td>
<td></td>
</tr>
<tr>
<td>Littlefield</td>
<td>1998</td>
<td>Y Y o</td>
<td>Y o</td>
<td>Y Y Y Y Y Y</td>
<td>64%</td>
</tr>
<tr>
<td>Mulliken</td>
<td>1999</td>
<td>Y Y o</td>
<td>Y o</td>
<td>Y Y Y Y Y Y</td>
<td>62%</td>
</tr>
<tr>
<td>Pedroso</td>
<td>2008</td>
<td>Y Y o</td>
<td>Y o</td>
<td>Y Y Y Y Y Y</td>
<td>64%</td>
</tr>
<tr>
<td>Teichgraeber</td>
<td>2004</td>
<td>Y Y o</td>
<td>Y o</td>
<td>Y Y Y Y Y Y</td>
<td>62%</td>
</tr>
<tr>
<td>Teichgraeber</td>
<td>2002</td>
<td>Y Y o</td>
<td>Y o</td>
<td>Y Y Y Y Y Y</td>
<td>64%</td>
</tr>
<tr>
<td>Van Vlimmeren</td>
<td>2008</td>
<td>Y Y Y Y Y Y Y Y Y Y o</td>
<td>o Y Y Y o</td>
<td>78%</td>
<td></td>
</tr>
<tr>
<td>Van Vlimmeren</td>
<td>2007</td>
<td>Y Y o</td>
<td>Y o</td>
<td>Y Y Y Y Y Y</td>
<td>69%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>= fulfilled the methodological criteria satisfactorily.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>o = did not fulfill the methodological criteria.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>= not applicable.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>doi:10.1371/journal.pone.0089602.t004</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The explanation is that only recently have techniques become available to capture the full 360° image needed to study cranial dimensions. Most studies using 3D imaging concern facial growth and treatment outcome because this technique has been available for two decades. Therefore, it is expected that within the next decade, more studies using 3D imaging of cranial dimensions will be published. Advantages of these 3D techniques are millisecond fast image capture, archival capabilities, a good-resolution color representation, and no exposure to ionizing radiation. Furthermore, assessment of linear dimensions and cranial size and shape can be made three-dimensionally. These are major advantages compared to more simplistic analyses performed

Table 5. Methodological quality scores of studies using indirect 2D and 3D imaging techniques reporting on soft tissue–based quantitative longitudinal assessment of cranial dimensions in children until age 6 years with a score equal to or above 60% (n = 9) [133–141].

<table>
<thead>
<tr>
<th>First author</th>
<th>Year</th>
<th>Design</th>
<th>Measure</th>
<th>Statistics</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>2D photography</td>
<td>2010</td>
<td>Y Y Y</td>
<td>Y Y Y Y Y Y Y Y</td>
<td>Y o o</td>
<td>73%</td>
</tr>
<tr>
<td>Hutchinson</td>
<td>2004</td>
<td>Y Y Y</td>
<td>Y Y Y Y Y Y Y</td>
<td>Y o o</td>
<td>75%</td>
</tr>
<tr>
<td>Hutchinson</td>
<td>2011</td>
<td>Y Y Y</td>
<td>Y Y Y Y Y Y Y</td>
<td>Y o o</td>
<td>67%</td>
</tr>
<tr>
<td>3D surface laser scanning</td>
<td>2006</td>
<td>Y Y Y</td>
<td>Y Y Y</td>
<td>Y Y Y Y Y Y Y Y</td>
<td>62%</td>
</tr>
<tr>
<td>3D stereophotogrammetry</td>
<td>2012</td>
<td>Y Y Y</td>
<td>Y Y Y</td>
<td>Y Y Y Y Y Y Y Y</td>
<td>92%</td>
</tr>
<tr>
<td>Collet</td>
<td>2010</td>
<td>Y Y Y</td>
<td>Y Y Y</td>
<td>Y Y Y Y Y Y Y</td>
<td>62%</td>
</tr>
<tr>
<td>Lipira</td>
<td>2012</td>
<td>Y Y Y</td>
<td>Y Y Y</td>
<td>Y Y Y Y Y Y Y</td>
<td>62%</td>
</tr>
<tr>
<td>Meyer-Marcotty</td>
<td>2010</td>
<td>Y Y Y</td>
<td>Y Y Y</td>
<td>Y Y Y Y Y Y Y</td>
<td>64%</td>
</tr>
<tr>
<td>Schaf</td>
<td>2010</td>
<td>Y Y Y</td>
<td>Y Y Y</td>
<td>Y Y Y Y Y Y Y</td>
<td>64%</td>
</tr>
<tr>
<td>Toma</td>
<td>2010</td>
<td>Y Y Y</td>
<td>Y Y Y</td>
<td>Y Y Y Y Y Y Y</td>
<td>64%</td>
</tr>
</tbody>
</table>

Y = fulfilled the methodological criteria satisfactorily. 
O = did not fulfill the methodological criteria. 
= not applicable. 
doi:10.1371/journal.pone.0089602.t005
with direct anthropometry. A shortcoming of 3D imaging is its inability to capture surface morphology in the presence of cranial hair. The reliability of 3D imaging techniques for soft tissue-based evaluation of cranial growth and treatment outcomes needs to be investigated further.

**Limitations**

Failure to identify all relevant reports for a systematic review is likely to result in bias [147]. 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 standardized manner by two reviewers to prevent unjustified exclusion of eligible studies. The manual 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, because only approximately one out of ten studies was retrieved by the manual search, this omission might be negligible. Furthermore, 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. Nevertheless, failure to retrieve full-text publications of possibly eligible studies (n = 195) was inevitable.

The instrument used to assess methodological quality was adapted from Lagravère et al. [148] who developed a methodological quality checklist to assess study design, study measurements, and statistical analysis in clinical trials. Since the introduction the checklist has been modified and used by Gordon et al. [21] and Van Vlijmen et al. [149]. Scientific use of the checklist has been accepted because the criteria check for generally accepted reasons for bias, despite a lack of validation of the QAI. The majority of inter-rater 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 greater frequency 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 the literature, no single tool is an obvious candidate for assessment of methodological quality of non-randomized studies [150]. Attempts to validate QAI’s like the Newcastle-Ottowa [151] scale or the Jadad scale [152] produce highly arbitrary results and cannot demonstrate significant effects on quality scores [153,154]. There is a need for a validated QAI that is preferably applicable to a wide range of study designs.

**Conclusions**

Direct anthropometrical measurement of head circumference in growing children below age 6 years is a reliable method for assessing cranial dimensions. The non-invasive 3D surface laser scanning and 3D-stereophogrammetry techniques can assess size and shape three-dimensionally. However, their reliability for assessing cranial dimensions needs to be investigated further.

**Supporting Information**

Checklist S1 PRISMA checklist.

(DOCX)

**Acknowledgments**

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

**Author Contributions**

Conceived and designed the experiments: S. Brons AK. Performed the experiments: S. Brons MvB AK. Analyzed the data: S. Brons MvB. Contributed reagents/materials/analysis tools: S. Brons EB. Wrote the paper: S. Brons. Intellectual contribution: JD JS S. Berge.

**References**


**Table 6. Reliability of methods for soft tissue–based quantitative longitudinal assessment of cranial dimensions in children until age 6 years in studies with good methodological quality (methodological quality score equal to or above 60%).**

<table>
<thead>
<tr>
<th>First author</th>
<th>Year</th>
<th>Measurement error</th>
<th>Correlation coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Head circumference, direct anthropometry</td>
<td>Bendersky 1998</td>
<td>0.99</td>
<td></td>
</tr>
<tr>
<td>Dhalla 1993</td>
<td>0.9 mm</td>
<td></td>
<td></td>
</tr>
<tr>
<td>De Bruin 1998</td>
<td>ns</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Guo 1988</td>
<td>1.6 mm</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jaffe 1992</td>
<td>1–2 mm</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Laingart 1997</td>
<td>0.87</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Roche 1987</td>
<td>0.9 mm</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rothenberg 1999</td>
<td>0.8–1.7 mm</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Direct anthropometry</td>
<td>Littlefield 1998</td>
<td>1 mm</td>
<td></td>
</tr>
<tr>
<td>Mullikin 1999</td>
<td>1 mm</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3D laser scanning</td>
<td>Plank 2006</td>
<td>0.5 mm</td>
<td></td>
</tr>
<tr>
<td>3D stereophogrammetry</td>
<td>Meyer-Marcotty 2012</td>
<td>0.02–4.3 mm</td>
<td></td>
</tr>
</tbody>
</table>

= not reported.
ns = not significant.
doi:10.1371/journal.pone.0089602.t006
22. Landis JR, Koch GG (1977) The measurement of observer agreement for
5. Bhalla AK, Malhotra BN (1990) Head circumference growth charts for preterm and
1. Bhalla AK, Malhotra BN (1979) Head circumference charts for preterm and premature

64. Jaldin MM, Pinheiro FS, Dos Santos AM, Muniz NC, Brito LMO (2011) Head
55. PROSPERO: International prospective register of systematic reviews.
46. Ernst JA, Bull MJ, Rickard KA, Brady MS, Lemons JA (1990) Growth outcome and feeding practices of the very low birth weight infant (less than 1500 grams) within the first year of life J Pediatr 117 815–816.
44. Ernst JA, Bull MJ, Rickard KA, Brady MS, Lemons JA (1990) Growth outcome and feeding practices of the very low birth weight infant (less than 1500 grams) within the first year of life J Pediatr 117 815–816.
80. Marks KH, Maisels MJ, Moore E, Friedman Z (1979) Head growth
77. Makrides M, Neumann MA, Jeffrey B, Lien EL, Gibson RA (2000) A


