Received: 6 October 2021 | Accepted: 21 March 2022
DOI: 10.1111/iej.13738

REVIEW ARTICLE

INTERNATIONAL ENDODONTIC JOURNAL | WILEY

A critical analysis of research methods and experimental models to study working length determination and the performance of apex locators – A narrative review with recommendations for the future

Ashraf ElAyouti[1] | **Thomas Connert**[2] | **Paul Dummer**[3] | **Claus Löst**[4]

[1]Division of Endodontology, Department of Conservative Dentistry and Periodontology, University of Tübingen, Tübingen, Germany
[2]Department of Periodontology, Endodontology and Cariology, University Center for Dental Medicine Basel UZB, University of Basel, Basel, Switzerland
[3]Adult Dental Health, Cardiff University, Cardiff, UK
[4]Conservative Dentistry and Endodontology, University of Tübingen, Tübingen, Germany

Correspondence
Ashraf ElAyouti, Division of Endodontology, Department of Conservative Dentistry and Periodontology, University of Tübingen, Osianderstraße 2-8, 72076 Tübingen, Germany.
Email: Ashraf.ElAyouti@med.uni-tuebingen.de

Abstract
Outcome studies have repeatedly shown that the apical endpoint of root canal preparation and filling is a determinate factor for the outcome of root canal treatment. Accurate determination of root canal length enhances the efficacy of chemomechanical disinfection and prevents over-/under-instrumentation and over-/under-filling in relation to the canal terminus. Long and short root canal fillings are consistently reported to be associated with higher rates of post-treatment endodontic disease. Although standards for undertaking and reporting diagnostic accuracy studies are available, publications dealing with the determination of root canal length are highly heterogeneous and describe procedures inconsistently. The aim of this review is to critically assess the methodology of publications in the past three decades. The process of planning, performing and analysing working length studies are presented stepwise with suggestions to optimize research methods.

KEYWORDS
accuracy tolerance, endodontic working length, experimental design, review, root canal length determination, Study error

INTRODUCTION

The apical extent of root canal treatment remains an important prognostic factor for the success of root canal treatment (Kojima et al., 2004; Ng et al., 2008; Sjögren et al., 1990). In outcome studies, clinical success rates are correlated to the length of the root canal filling and not to the working length (WL). Therefore, the relevance of WL in terms of treatment success has not yet been established. Determining a specific endpoint for WL may not be suitable for all treatment protocols such as for teeth with vital and nonvital pulps or for primary and secondary root canal treatments. However, an accurate determination of an apical landmark provides clinicians with a reference from which the WL could be adjusted to adapt different root canal treatment procedures as well as to specific pulpal and periapical status of the teeth.

Unfortunately, studies investigating WL determination methods deviate from recommendations for diagnostic accuracy studies and have inconsistent experimental designs. The aim of this paper is not to summarize the literature of WL determination, but to provide a critical overview of the methodology used in WL studies. In reviewing the literature, criticism of specific publications and their authors has been avoided. Instead, a summary of research parameters is provided along with suggestions for designing and performing WL studies. In addition to the requirements described in the PRIDE guidelines for reporting and the STARD guidelines for manuscripts on diagnostic accuracy studies (Bossuyt et al., 2015; Nagendrababu et al., 2021), in the present review efforts were made to aid researchers in validating methodology protocols and statistical analysis.

Unlike electronic WL methods, radiographic methods are not covered extensively in this review as they are not the focus of the contemporary literature.

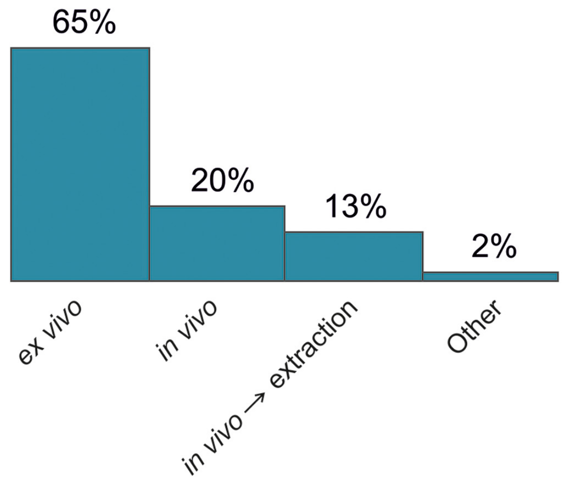

FIGURE 1 Types of studies on working length determination

REVIEW

Following a manual search of 621 publications in the PubMed and Web of Science databases, a total of 209 publications from 1990 to March 2021 were reviewed (File S1). Search terms were "root canal treatment", "working length" and "apex locators". Included were English-language publications that addressed WL determination and provided sufficient details of methodology to be summarized in this narrative review. The descriptive data of the present review are presented mainly in percentages without a margin of error. The frequencies presented should not be considered absolute values, as they may not include all publications, but are intended to represent the *status quo* of the literature on endodontic WL determination. The unit of analysis was not consistently the publication, but the dependent variable of interest reported.

TYPES OF WL STUDIES

Fortunately, *in vivo* studies constitute a considerable proportion of WL publications. In many *in vivo* studies, after determining the WL clinically, the teeth investigated were extracted with the measuring file fixed in the root canal and the distances to specific apical landmarks were then measured *ex vivo* (Figure 1). Although cone beam computed tomography (CBCT) is still not indicated for WL determination (Patel et al., 2019), when available, some studies used CBCT images to evaluate the *in vivo* determined WL (Ustun et al., 2016). Other *in vivo* studies compared WL measurements without an apical reference point. Although these studies may provide information about the agreement of different WL determination methods, they cannot provide information on the accuracy of tested methods.

The bulk of studies were performed *ex vivo*, and mostly confirmed the accuracy, precision and repeatability of electronic apex locators (EAL). It is rather peculiar that in some publications the less precise WL radiograph (Vandevoorde & Bjorndahl, 1969) was used to judge the accuracy of EALs. Some *in situ* studies were performed on cadavers, which offers better simulation of clinical circumstances and avoids the exposure of patients to ionizing radiation when radiographic techniques are used.

Recommendations:

- Investigating the agreement between EALs or CBCT measurements and the imprecise WL radiograph should be avoided.
- In vivo testing should be predominately used to assess WL methods that have been proven ex vivo to be accurate and precise.
- Evaluating WL measurements without the use of a suitable anatomical apical reference or controls should be avoided.

METHODS OF WL DETERMINATION

Clinically, endodontic WL is mainly determined radiographically or/and electronically (Zaugg et al., 2020). Over the past three decades, studies on WL accuracy focused on EALs due to their high accuracy and precision (Figure 2). However, only a small number of studies tested endodontic motors with integrated EALs.

Periapical radiographs, despite their inaccuracy, are still widely used to confirm electronic WL measurements as they may offer additional diagnostic information. The number of publications evaluating radiographic WL determination has dropped to 4% since the introduction of EALs.

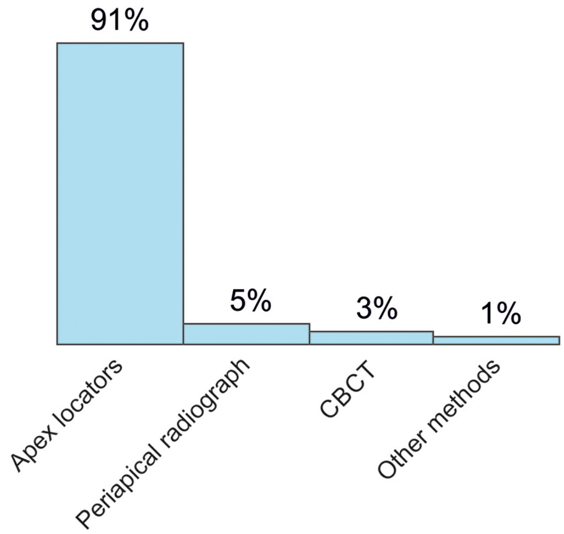

FIGURE 2 Methods of working length determination in reviewed publications

CBCT has been reported to provide accurate WL measurements (Connert et al., 2014; Janner et al., 2011; Jeger et al., 2012; Liang et al., 2013; Metska et al., 2014) but is still not indicated when the ALARA principles are followed. The hope is that in the future the radiation dose of CBCT will be reduced to allow the routine use of 3D diagnostics in endodontics.

Other WL assessment techniques, such as the paper points and tactile methods, have limited use either because of poor precision or the need for specific clinical circumstances (Jordan et al., 2014; Marcos-Arenal et al., 2009; Stabholz et al., 1995).

Recommendation:

- Further research should be performed on CBCT and EALs integrated within endodontic motors as these two WL methods are rarely reported.

ACCURACY AND PRECISION

Although the two terms are used interchangeably, they are two separate entities that differ from each other. Accuracy describes how far measurements are away from a specific target, such as the apical foramen. Precision describes how far the measurements are away from each other regardless of the target, that is, the scatter of the measurements. Indeed, a precise device can also be inaccurate. This occurs when its measurements are close to each other, but distant from the target. Accurate and precise methods provide measurements close to the target and close to each other.

Recommendations:

- Separate measures for accuracy and precision should be provided.
- Collecting WL data in categories should be avoided, as this obscures the precision.

ACCURACY TOLERANCE

To determine the "accuracy", a distance interval is specified within which all measurements are considered accurate. However, some studies did not report the accuracy tolerance. Figure 3 shows that the tolerance of ±0.5 mm was commonly used, probably because this is the tolerance accepted by clinicians, especially endodontists. The use of ±1 mm provides less differentiation between WL methods; however, this may be suitable for the less accurate radiographic measurements or as a complementary result. Some studies used a tolerance of 0 mm (Figure 3). The determined accuracy with a zero tolerance is not valid without a proof that the sum of all procedure errors, including examiner error, is equal to zero. Since this is unlikely to happen, any procedural error will inevitably result in measurements outside the accuracy tolerance range and will account for the inaccuracy of methods tested, which is incorrect.

The use of accuracy tolerance around the major foramen must be carefully interpreted. It is critical to consider measurements beyond the foramen as accurate, since it is well established that poorer outcomes are associated with over-filling of the root canal (Kojima et al., 2004; Ng et al., 2008; Sjögren et al., 1990). Therefore, it is recommended to differentiate between WLs beyond the foramen and those short of it. The question is, what is the cut-off point at which measurements beyond the foramen can be considered accurate? The answer to this question is related to the

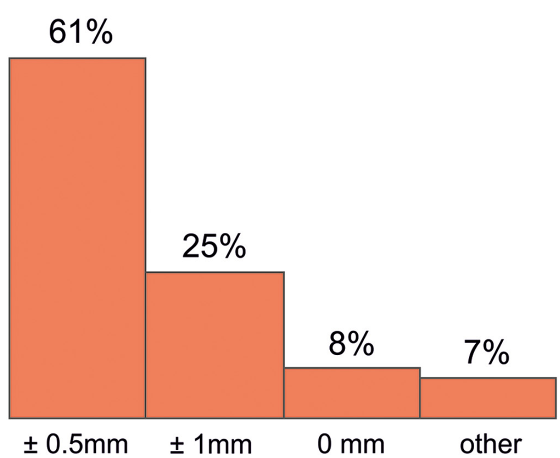

FIGURE 3 Distance intervals used for the determination of the accuracy of endodontic working lengths

quantity of procedural error. For instance, if it is around 0.25 mm, it is reasonable to consider measurements beyond the foramen but within this tolerance distance as accurate.

Recommendations:

- Using zero accuracy tolerance without summing up all procedural errors should be avoided.
- A ±0.5 mm tolerance should be used for better differentiation of WL methods and clinical acceptance.
- Wider accuracy intervals could be used for less accurate methods or as complementary results.

SAMPLE SIZE AND POWER CALCULATIONS

Only 3% of the publications reviewed reported sample size calculations. Setting the accuracy tolerance ahead of the study may facilitate sample size calculation. The accuracy tolerance can be regarded as the least significant difference to be detected, and together with the standard deviation (SD) of similar published studies, the power and sample size can be calculated. Also, the effect size, which is the difference between group means in terms of SD, is commonly used to calculate sample size. With a pooled SD of 1, the difference between the means of two groups is equal to the effect size.

In practical terms, statistical software is used to define these parameters and calculate the appropriate sample size for a specific power of the study. Figure 4a,b show the relation between sample size and power. There are far fewer samples required when the SD is reduced from 1 to 0.7. This underlines the benefits of minimizing procedural errors and comparing highly accurate and precise methods, as both will result in a narrow SD.

In the EAL publications reviewed, the distances between the tip of the measuring file and the target had a mean SD of 0.55 (99% CI: 0.4 – 0.7 mm). Therefore, it is reasonable to use Figure 4d to calculate the sample size for most EAL studies when the SD is unknown.

In some studies, groups with differences less than those used for sample size calculations were incorrectly considered statistically significantly different.

Recommendations:

- Sample size could be calculated using the SD of similar published studies along with the difference to detect or the effect size.
- For new methods with unknown SD, Figure 4d may be used for EAL accuracy studies; in this case, adjustments of sample size may be needed.

STUDY ERROR

All procedural errors that can be calculated are known as "determinate" errors. Figure 5 reveals that most publications did not report error parameters. In those publications, the study error is unknown and is inevitably added to that of the method tested, as in this publication (ElAyouti et al., 2002). The effect of summing up procedural errors to the determined accuracy is unpredictable and does not necessarily worsen the results, but may also lead to an improved accuracy depending on whether a negative or positive error value is added. This is another reason why the results of WL studies are poorly reproducible, even when performed with the same methodology.

Repeatability and reproducibility represent the agreement between replicated measurements. Repeatability requires that measurements be performed within a short period of time by the same operator, using the same tools and under the same conditions. Similar standards are applied for reproducibility, except that the time between replicated measurements is longer and the conditions may change. Repeatability may be more important in WL studies, as the clinically determined WL is generally used to shape or fill the root canal in the same visit. Usually, the WL is recapitulated and confirmed in subsequent visits.

In Figure 5, it is perplexing to see the cluster of different error parameters presented in WL publications. Fortunately, it is not necessary to calculate every single parameter, since one specific parameter, namely "experimental repeatability", can be calculated to include most determinate errors. The repeatability of the experimental setup would be calculated with a device of known high accuracy and precision. Pairs of measurements are taken on each subject going through all steps of the procedure, and two standard deviations are calculated below and above the mean difference of repeated measurements (Bland & Altman, 1986). This would include 95% of the determinable error of the experiment and must be well below the defined accuracy tolerance. If this value is above an acceptable range, for example 30–40% of the accuracy tolerance, the experimental setup is not precise enough and the main source of error must be identified by assessing the error parameters (Figure 5).

Unlike determinate errors, random errors have an unknown source and are indeterminate, often due to unexpected changes in study conditions. Random errors may be reduced by taking the mean of multiple measurements on the same subject, as is often performed in WL studies (Cianconi et al., 2010; Duran-Sindreu et al., 2013; Saatchi et al., 2014; Williams et al., 2006).

Other statistically determinable errors are those that affect the results of an experiment, namely false positives

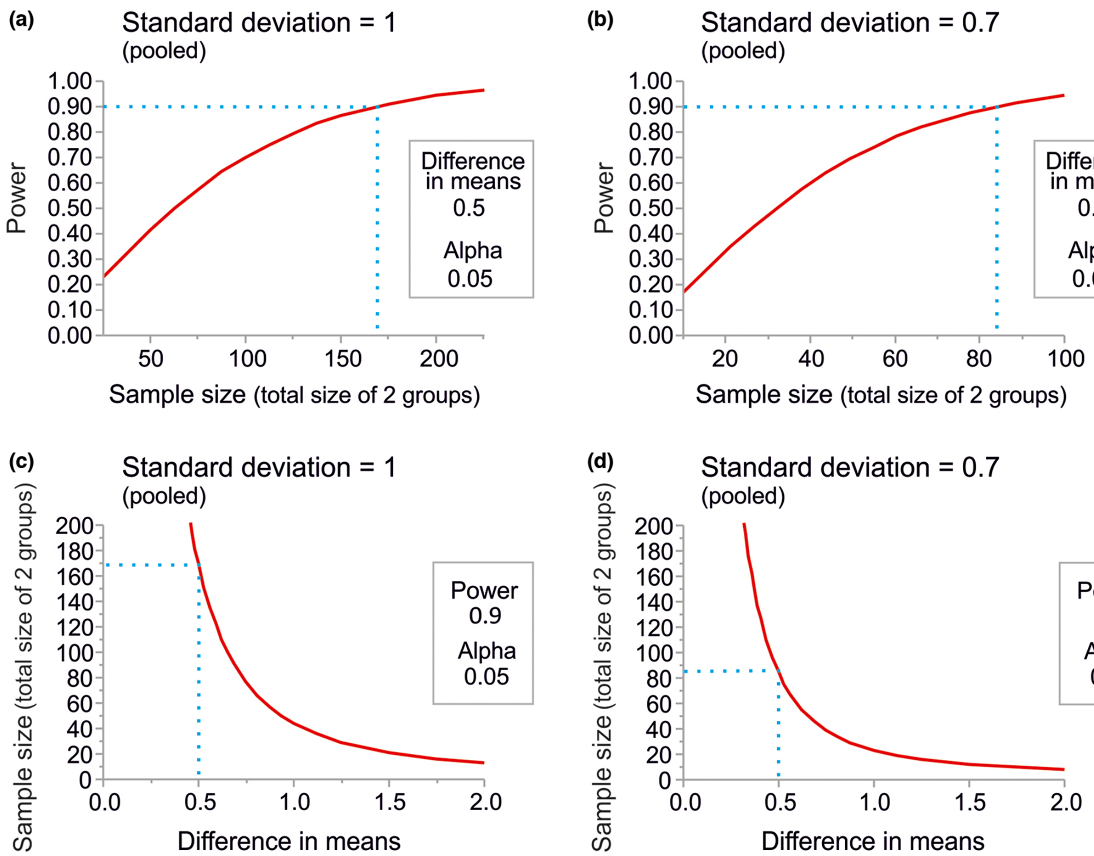

FIGURE 4 Relation between sample size and power calculations with standard deviation of 1 (a) and 0.7. (b), dotted lines show sample size using a power of 0.9 (a and b). Relation between sample size and difference to detect with standard deviation of 1 (c) and 0.7 (d), dotted lines show sample size using difference in means of 0.5 (c and b). Calculations for 2 groups, Alpha (Type I error) =0.05, statistical program JMP 15.2.0

and false negatives, also known as Type I (α) and Type II (β) errors.

Recommendations:

- The repeatability of the experimental setup should be determined to quantify systematic errors.
- If necessary, the magnitude of error parameters presented in Figure 5 could be evaluated to locate the main source of error.

ETHICAL APPROVAL

According to the PRILE 2021 reporting guidelines (Nagendrababu et al., 2021), the use of human materials must have ethical approval from review boards as well as informed consent from patients. In some publications, with these requirements fulfilled, patients including children were exposed to unnecessary ionizing radiation by obtaining WL radiographs prior to tooth extraction. Certainly, this is inappropriate, but it is also a reminder to critically judge the ethical aspects of publications aside from whether review board approval has been obtained. Recommendations: Unnecessary treatment steps, especially exposure to ionizing radiation, must be eliminated from study designs, even when informed consent was obtained. The validity of ethical approval should be checked by the journals on submission.

SAMPLE SELECTION

Generally, the sample should be representative of the population under study. In WL studies, the sample would preferably include all types of teeth requiring root canal treatment. Figure 6 shows that the majority of studies tested single-rooted teeth and straight root canals. Teeth indicated for root canal treatment were rarely included, but teeth extracted for periodontal or orthodontic reasons were frequently used. The evaluation of new WL methods may necessitate the use of easily accessible root canals to minimize measurement variability due to difficult anatomy. However, at some point, more information is required on the performance of these methods not only

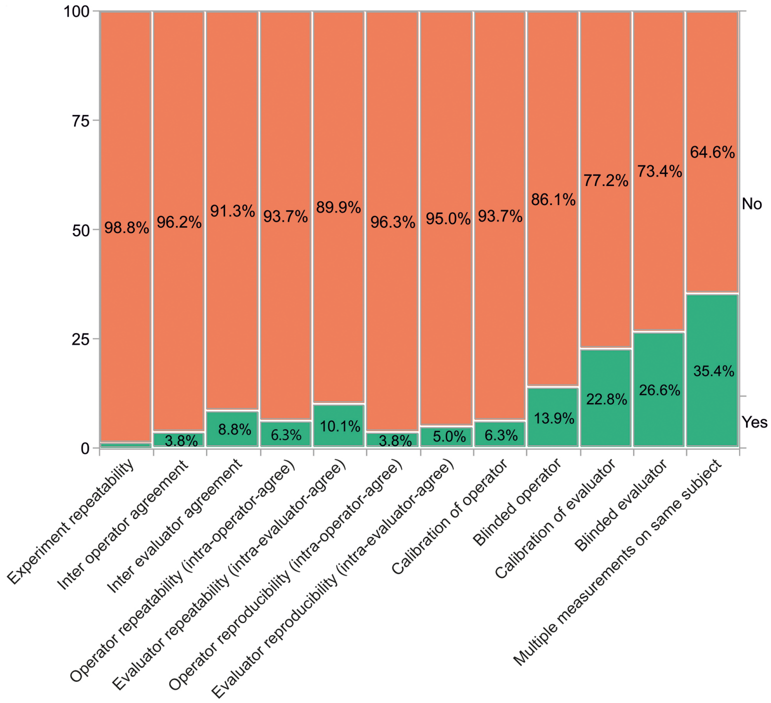

FIGURE 5 Frequency of different measures for error parameters presented in reviewed publications

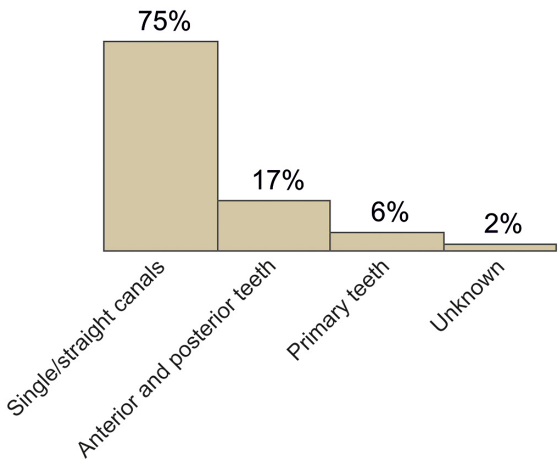

FIGURE 6 Type of teeth and root canals examined in studies dealing with working length determination

in anterior teeth but also in posterior teeth with more difficult root canal anatomy. Likewise, the routine exclusion of teeth with restorations or defects has led to a scarcity of data on the performance of WL methods in teeth with this regular clinical finding. In primary teeth, EALs have been shown to be accurate even in the presence of apical resorption (Beltrame et al., 2011; Nelson et al., 2010; Odabas et al., 2011); in permanent teeth, such information is rarely available as teeth with apical resorption were excluded routinely from WL studies. In addition, many studies lack detailed selection criteria such as tooth type.

The selection of comparison groups should be based on knowledge of the factors that affect the determination of the WL. For example, if it is assumed that root canal curvature in posterior teeth would affect WL results compared to straight root canals of anterior teeth, comparing posterior teeth to anterior teeth without describing differences in the angle and radius of root canal curvature is not sufficient.

Recommendations:

- Posterior teeth should be included in samples, as current data are based mainly on single-rooted teeth and straight root canals.
- Teeth indicated for root canal treatment and teeth with restorations and apical resorption should be included.

ALLOCATION OF SAMPLE

Whilst randomization and allocation requirements are illustrated in the PRILE 2021 guidelines, in EAL studies multiple methods/EALs may be examined in the same root canal, and thus, allocating the sample into groups may not be required. However, in this design, the sequence of using different WL methods in the same root canal must be randomized. Moreover, the different sequences should be equally distributed in the sample. In this powerful design, the use of pairwise comparisons eliminates the differences between subjects and thus offers a better differentiation between the methods tested. In clinical studies, it is sometimes difficult to obtain balance between test groups using randomization, especially when the criteria of patients are unknown prior to recruitment. In these instances, minimization (Scott et al., 2002) is more convenient than stratified randomization.

Recommendations:

- The sequence of WL measuring methods should be randomized when used in the same root canal.
- Randomization or minimization could be used for sample allocation.

ALTERATION OF TEETH FOR CONVENIENCE

Decoronation or sectioning of the crowns of teeth to be tested are usually justified by providing direct access to the root canal and attaining flat, horizontal surfaces for a reproducible coronal reference point. Some studies ground study teeth to a standardized length that is clinically uncommon Figure 7. Although these tooth alterations simplify the methodology and are convenient, they deviate from the clinical situation, and therefore, more careful interpretation of the results may be needed.

Recommendation:
- Sample and study conditions should be kept as close as possible to the clinical scenario, and major alterations to tooth anatomy should be avoided.

APICAL REFERENCE FOR WL

The majority of studies, including *in vivo* ones, verified the apical landmarks *ex vivo* on extracted teeth (Jung et al., 2011; Lucena et al., 2014; Vasconcelos et al., 2015; Vieyra & Acosta, 2011). Some studies used EALs to determine fixed distances from the foramen (Figure 8). This would

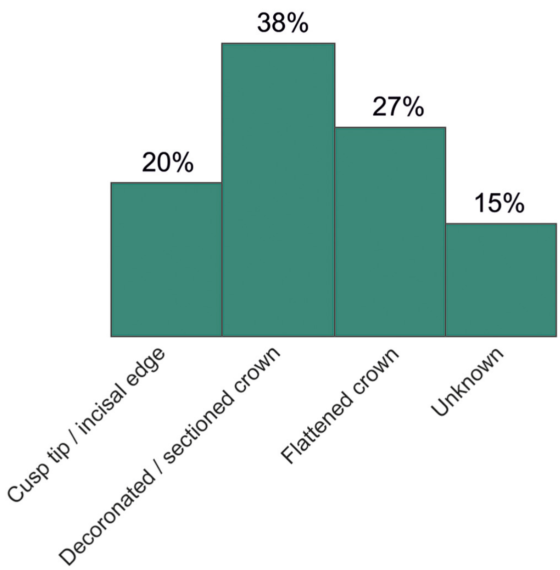

FIGURE 7 Distribution of coronal landmarks used as a reference for working length determination

be inappropriate because EALs measure relative impedances, not distances. A similar mistake is to expect apex locators to locate distances in mm according to the numerical display of the devices as "0.5" and "1". In fact, some companies replaced the numerical display with coloured LEDs to avoid this misunderstanding. According to the present principal function of EALs, two points can be located. The first point is the contact point between the periodontium and the root canal. It is not only represented by foramina, but also by resorptions, perforations, root fractures and all other contact points with the periodontium. EALs locate this point by utilizing the constant electrical resistance of soft tissues. Close to it is the second point of the narrowest root canal area, represented by the so-called apical constriction. At this point, EALs detect a remarkable change in the relative impedance measurements. In the case of tapered or parallel root canal walls, these two landmarks are very close to each other or nearly at the same point.

Apical foramen (major foramen)

Due to its location on the outer surface of the root tip, the foramen is the most consistent apical reference. It may also represent the junction between pulp and periodontal tissues. The WL to the apical foramen is the most frequently determined (Figure 8). It is commonly performed by introducing a file into the root canal until the tip is visualized flush with the major foramen. Whilst this procedure seems to be accurate, a measure of precision is often missing. And although many studies recorded the average of multiple measurements, the agreements between replicate measurements were rarely presented. A simple presentation of the limits of agreement of replicate measurements would be very informative (Bland & Altman, 1986).

Apical constriction

The constriction is the reference point of choice to preserve the apical anatomy. Retaining the WL to the constriction will minimize irritation to the periapical tissues and prevent over-instrumentation and over-filling (Ricucci & Langeland, 1998).

Longitudinal sections of the root canals are commonly used to expose the constriction with the measuring file fixed in place. Longitudinal sections have been reported to be associated with substantial variability in determining the position of the apical constriction (Schell et al., 2017), but this still needs to be ascertained by other similar studies. Determination of the apical constriction is

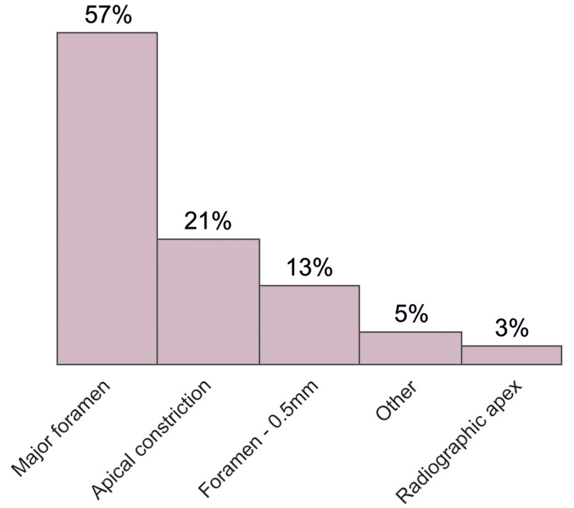

FIGURE 8 Apical reference landmarks used for working length determination in reviewed publications

extremely critical and needs a meticulous experimental setup. Comparing the areas of root canal cross sections, though laborious, may be an optimal method to locate the narrowest part of the root canal (Connert et al., 2018).

Minor foramen

When the major foramen has a funnel shape, the minor foramen is its smallest diameter more coronally, and is supposed to coincide with the apical extent of the constriction. Studies utilizing the minor foramen as an apical reference photographed the major foramen and used image software to measure the distance to the narrowest part of the root canal. In parallel and tapered root canal walls, the minor foramen is not identifiable. Also, the accuracy and precision of this procedure is unclear.

Radiographic apex

The radiographic apex is the only choice when evaluating periapical radiographs. It is well established that the radiographic apex can be misleading when used as a reference landmark (Liang et al., 2011). Nevertheless, a WL radiograph can provide valuable diagnostic information. However, due to the known imprecision of radiographic WL, it is not appropriate to use it to assess the more accurate EAL or CBCT measurements.

Recommendations:

- Apical reference landmarks should be defined and matched to the method of WL determination or the scale of the EAL.
- Using EALs to determine fixed distances from anatomical landmarks should be avoided.
- Investigating the agreement between EALs or CBCT measurements and the imprecise WL radiograph should be avoided.

CALIBRATION OF EXAMINERS

WL studies regularly involve operators performing WL measurements and evaluators rating them. Calibration allows examiners to become familiar with study procedures and improve their accuracy and precision. The level of attained examiner reliability should not be assumed, but must be presented by calculating the agreement of operators and evaluators. Figure 5 reveals that in the majority of studies, the examiners were not calibrated. Moreover, a measure of agreement is frequently missing.

Recommendations:

- Both operators and evaluators should be calibrated.
- A measure of agreement should be provided to check that calibration has been achieved.

BLINDING OF EXAMINERS

Blinding, for its part, reduces bias by keeping operators and evaluators unaware of the measuring parameters and measuring devices. Preferably, two examiners will be involved in the blinding process. One operator determines the WL and the second reads and records the measurements. To minimize transcription errors, another examiner monitors the entire process and checks the recorded data. These arrangements are rarely performed in WL studies. When blinding is reported, either operators are unaware of the measuring parameters or evaluators being blinded to the WL methods or devices; the latter scenario is more likely and was more frequently conducted in WL publications (Figure 5).

Recommendations:

- All examiners should be included in the process of blinding, not only evaluators.
- At least two examiners should conduct WL measurements: a blinded operator to perform the measurements and another operator to read and record them.

STOPPERS AND SPACERS

Stoppers are routinely used in WL determination to adjust and read the length of measuring files. For convenience, silicon or rubber stoppers are used that move easily along the shaft of endodontic instruments. To prevent over-instrumentation during shaping of the root canal, clinicians are constantly watching the stoppers to monitor the achieved preparation length and to instantly prevent the displacement of instruments when the stoppers touch a coronal reference point. The mean error of just adjusting and reading the length of hand files has been shown *ex vivo* to amount to 0.5 mm (ElAyouti & Löst, 2006); however, in practice, this error is likely to be much greater. It is much more dramatic when stoppers unmarketably become displaced upon hitting the coronal reference point, especially when rotating files are used. When considering the efforts needed to restrict study errors in relation to the defined accuracy tolerance, the error of adjusting and reading stoppers is large and would account for inconsistent WL measurements even when using highly accurate EALs and meticulous care during length determination. For that reason, many studies used resin to fix stoppers to the measuring file or to the tooth to reduce this error. In *ex vivo* studies, adopting automated tools (D'Assuncao et al., 2010; Hör et al., 2005) and utilizing micrometre gauges connected to a computer can reduce procedural errors arising from the movement of stoppers and reading the file length as well as eliminating operator and transcription errors. Yet, these simple mounting devices are rarely used in published studies. Clinically, instead of stoppers, the use of spacers resembling implant key heights (El Kholy et al., 2019; Sittikornpaiboon et al., 2021) eliminate errors of adjustment, reading and movement of stoppers, and spare clinicians the continuous monitoring of the position of stoppers.

Recommendations:

- In *ex vivo* studies, mounting and automated devices should be used to eliminate errors from adjusting stoppers and reading of measurements.
- *In vivo*, spacers should be used instead of stoppers to eliminate errors arising from the movement of stoppers.

THE USE OF EAL IN DRY ROOT CANALS

Dry root canals were a requirement for the first generations of EAL that measured simple resistance or impedance at the foramen (Nekoofar et al., 2006). Since the introduction of intracanal impedance gradient measurements, dry root canals are no longer a requirement; on the contrary, they cause unstable measurements. Modern EALs utilize relative impedance measurements, and therefore, the root canals should not be dried and must be kept moistened for EAL measurements. However, some studies tested EALs in dry root canals and, as expected, reported inconsistent results.

Recommendation:

- The root canal should be kept moistened when using EALs to ensure consistent measurements.

DESIGN OF STUDY

The methodology of WL studies is frequently designed to explore either the accuracy of WL methods or the effect of clinical parameters on the determined accuracy. For accuracy determination, commonly percentages of measurements within a tolerance interval are given as well as the mean and SD of the differences. The margin of error, for example confidence intervals, and the repeatability of tested methods are frequently missing. The repeatability is clinically relevant because electronic WL measurements can be repeated in the same root canal, particularly when EAL are integrated in endodontics motors. Repeatability can be calculated easily by pairwise measurements in the same root canal; multiple measurements in the same root canal can also be analysed (Bland & Altman, 2007).

To explore the factors influencing WL measurements, the simplest and most intuitively understandable design is to maintain all WL parameters constant, except for the specific parameter of interest. The effect of combined parameters should be explored later when the effect of each influencing factor is known. Then, a meaningful combination of those parameters having the greatest impact could be evaluated. However, combination of many parameters may be difficult to report. For example, evaluating three EALs using multiple irrigants in many types of root canals often results in more than 20 groups of combinations. Statistically, it is possible to analyse the separate and combined parameters. However, the results are often difficult to comprehend, and the sample size must be enormous to reach a satisfactory power.

Recommendations:

- Repeatability of the WL methods tested should be provided.
- Accuracy should be presented with the margin of error to estimate the precision, for example CI.
- Combining multiple influencing factors should be avoided; it is far better to present one to two parameters in the same study for optimal power and clarity of the results.

DATA ANALYSIS

Analysis of WL studies is simple and straightforward because continuous data arise that frequently have a normal distribution. Another important point is that the reference values or controls are known, as the majority of studies validate the length determined *ex vivo* directly on extracted teeth, and the same applies for most *in vivo* WL measurements. Although the length of the root canal is determined, the main interest is in the distances to an apical reference point. Some studies did not present these distances, instead the mean of root canal lengths was used to compare study groups. Whilst this is statistically correct, it does not provide the more clinically relevant information about how close the WL measurements were to the apical reference. However, the determined root canal lengths may be further analysed to see if there is a relation between the magnitude of length and the distance to the apical reference.

Mean of positive and negative values

In WL studies, positive and negative values arise, as for example measurements beyond and short of the foramen. The mean of such measurements is often around zero, as WL data typically show a normal distribution. This can be misleading when the measurements are far away from the target landmark. Whilst the SD will show the dispersion of the data, it is helpful to calculate the mean of absolute measurements or to separately present data beyond or short of the reference. Here, graphical presentation of the raw data will solve this problem.

Recommendations:

- WL measurements beyond and short of the reference landmark should be analysed and presented, for example, by plotting the raw data.
- The mean of absolute distances to the apical reference should be provided.

CONFIDENCE INTERVAL (CI) AND SD

Figure 9 reveals that the minority of WL studies provided CIs. In WL studies, the SDs of the means are frequently presented, but in some studies, as in this publication (ElAyouti et al., 2005), the SD is missing. SD is an important parameter that describes the dispersion of the data around the mean. Further, pooled SD can be used to calculate the effect size of the study, which is a useful parameter in meta-analysis. However, CIs of the mean are strongly recommended (Altman, 2005) as they provide an estimate of the precision when similar studies are performed and therefore provide an interval where the true mean of the population would lie. Because CIs are dependent on sample size, they are wide with a small sample size and narrow with a large one. CIs have a link to p-values and can also be used to compare test groups.

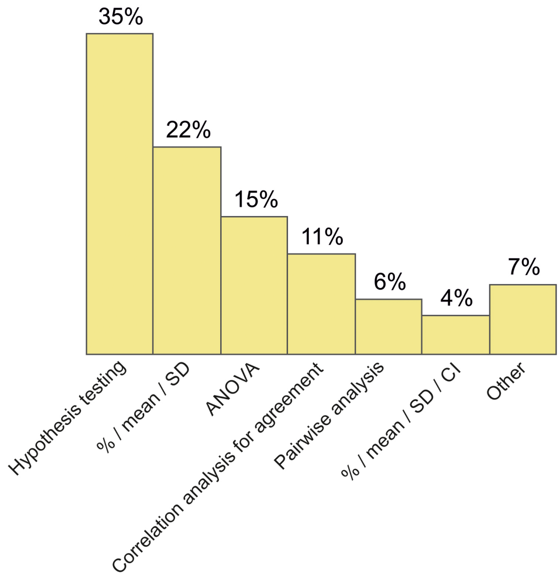

FIGURE 9 Frequency of different statistical methods implemented in reviewed working length accuracy studies

When CIs of the means of tested groups do not overlap, it can be inferred that they are statistically significantly different. Commonly, a 95% CI or 99% CI are calculated. Another merit of CIs is that they have the same unit of the mean (mm in WL studies); therefore, they are easier to comprehend than p-values (Gardner & Altman, 1986).

Recommendation:

- SDs and CIs should be presented as they provide valuable information about the dispersion and precision of the results. Both have the same unit of data and are easy to comprehend.

Hypothesis testing

Hypothesis testing is unnecessarily and excessively used (Gigerenzer, 2004). Whilst both CI and hypothesis testing utilize an approximation of the sampling distribution, CI is clearly better than testing a hypothesis. In WL studies, the differences are clearly measured, and these differences need to be quantified, not just to detect their presence. If only p-values are presented, the magnitude of the determined factors is lost. Similarly, when using paired t-tests, it is not appropriate to conclude that WL methods agree simply because they are not significantly different. In fact, a dispersion of the data will commonly lead to a nonsignificant difference between the means. When performed, hypothesis testing has to be sufficiently presented and carefully interpreted (Souza, 2014).

Recommendation:

- Hypothesis testing of WL data should be avoided as it obscures the magnitude of the differences and is not appropriate for assessing the agreement of WL methods.

Correlation analysis

In WL studies, correlation is often erroneously used to determine the agreement of measurements from different methods or different operators. Correlation is a measure of association and cannot be used to determine the agreement. In some situations, methods of poor agreement may have a high correlation coefficient. The continuous character of the data in WL measurements and the possibility for performing repeated measurements in the same root canal is optimal for eliminating inter-subject variation by performing pairwise analysis and determination of the limits of agreement (Bland & Altman, 1999).

Recommendation:

- The use of correlation analysis to determine the agreement between different WL methods should be avoided.

Raw data

Raw data are rarely presented in WL studies (Figure 10). A simple graphical presentation of the raw data can be much more powerful than any statistical analysis and is the best proof of the conclusions. WL studies rarely include huge volumes of data, and the results are mainly continuous numbers of distances to a target landmark. The raw data can be simply presented by plotting data points in Box Plots of different groups, or when pairwise analysis is performed, differences can be presented in relation to the equality line (Mancini et al., 2014; Miletic et al., 2011; Stoll et al., 2010; Uzun et al., 2008).

Recommendations for reporting:

- The raw data should be presented in the form of graphs as they are more comprehensible than tables.
- Accuracy can be presented by giving the percentage of measurements within the accuracy tolerance. Provide mean distances to the apical references with SD and CI as margin of error. CI can be used for comparisons.
- The precision of experimental setup and examiners should be presented by providing the repeatability of study procedures and agreement of examiners.
- The repeatability of investigated WL devices/methods should be provided.
- Depending on sampling distribution, and when needed, hypotheses testing or nonparametric comparisons can be presented.

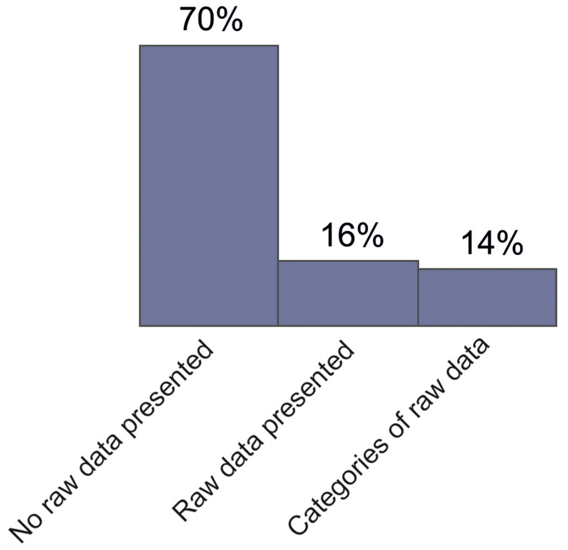

FIGURE 10 Frequency of raw data presentation in publications dealing with the accuracy of working length determination

JUST BECAUSE IT IS PUBLISHED DOES NOT MEAN IT MUST BE CORRECT

Mistakes do not really matter as long as they do not influence the methodology and results, for example typographical errors or incorrect wording. However, inappropriate procedures or analysis have a major impact on the results and need to be addressed, especially if they are repeatedly published over many years. The best way to avoid such mistakes is to question published methodology and revisit procedures and analysis before adopting them. An example of a repeated simple mistake that keeps occurring in publications over the past years can be seen in studies testing the Root ZX device (J. Morita MFG Corp, Tokyo, Japan), which is the most tested EAL. The use of the display scale "Flashing bar between 0.0 and 1" is ambiguous because the flashing bar is an adjustable display that can take any value between "0.0" and "1". A small mistake but one which has a major impact because in all these studies, the reference scale used was not clearly defined.

CALL IT APEX LOCATOR, FORAMEN LOCATOR, CONSTRICTION LOCATOR, ROOT CANAL LENGTH MEASURING DEVICE, ...

Apex locators are devices that initially measured the electrical resistance and later the relative impedance through the root canal. Some authors name them "Foramen locator" or "root canal length measuring device" to describe what these devices determine. Actually, these devices can determine not only the apical foramen, but also every egress to the periodontium, so they can locate root fractures, resorptions, perforations and the level of root end resection. In addition, and because they measure intracanal relative impedance, they can locate the change of canal cross-sectional area next to the root canal egress, that is to say the apical constriction. So, if we want a name that best describes contemporary devices, this would be "root canal-relative-impedance measuring device". Actually, it does not matter which name is chosen, as long as everyone understands what is meant by it and more importantly, that one accepts the names others have chosen. It just happened that these devices are best known as apex locators.

CONCLUDING REMARKS

In vivo studies comprise around one third of the publications dealing with working length determination. The majority of studies evaluated electronic apex locators.

Most studies tested straight root canals and single-rooted teeth that did not have an indication for root canal treatment. Information on the accuracy of working length determination in more difficult root canal anatomies is scarce.

Procedural error is rarely calculated. In the context of the narrow accuracy tolerance of endodontic working length, the results of many studies may be questioned.

Hypothesis testing and correlation analysis were improperly used to determine the agreement between WL methods. Raw data presentation would have been more valuable than sophisticated statistical analysis.

ACKNOWLEDGEMENT
Open Access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST
The authors state that there are no conflicts of interest in connection with this article.

AUTHORS CONTRIBUTION
All authors contributed to the accomplishment of this review and approved the Manuscript.

ORCID
Ashraf ElAyouti https://orcid.org/0000-0003-3146-1806
Thomas Connert https://orcid.org/0000-0003-4461-5835

REFERENCES
Altman, D.G. (2005) Why we need confidence intervals. *World Journal of Surgery*, 29, 554–556.

Beltrame, A.P.C.A., Triches, T.C., Sartori, N. & Bolan, M. (2011) Electronic determination of root canal working length in primary molar teeth: an in vivo and ex vivo study. *International Endodontic Journal*, 44, 402–406.

Bland, J.M. & Altman, D.G. (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, 1, 307–310.

Bland, J.M. & Altman, D.G. (1999) Measuring agreement in method comparison studies. *Statistical Methods in Medical Research*, 8, 135–160.

Bland, J.M. & Altman, D.G. (2007) Agreement between methods of measurement with multiple observations per individual. *Journal of Biopharmaceutical Statistics*, 17, 571–582.

Bossuyt, P.M., Reitsma, J.B., Bruns, D.E., Gatsonis, C.A., Glasziou, P.P., Irwig, L. et al. (2015) STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *British Medical Journal 351*.

Cianconi, L., Angotti, V., Felici, R., Conte, G. & Mancini, M. (2010) Accuracy of three electronic apex locators compared with digital radiography: an *ex vivo* study. *Journal of Endodontics*, 36, 2003–2007.

Connert, T., Hulber, M., Godt, A., Löst, C. & ElAyouti, A. (2014) Accuracy of endodontic working length determination using cone beam computed tomography. *International Endodontic Journal*, 47, 698–703.

Connert, T., Judenhofer, M.S., Hulber-J, M., Schell, S., Mannheim, J.G., Pichler, B.J. et al. (2018) Evaluation of the accuracy of nine electronic apex locators by using Micro-CT. *International Endodontic Journal*, 51, 223–232.

D'Assuncao, F.L.C., Albuquerque, D.S., Salazar-Silva, J.R., Dos Santos, V.C. & Sousa, J.C.N. (2010) Ex vivo evaluation of the accuracy and coefficient of repeatability of three electronic apex locators using a simple mounting model: a preliminary report. *International Endodontic Journal*, 43, 269–274.

Duran-Sindreu, F., Gomes, S., Stober, E., Mercade, M., Jane, L. & Roig, M. (2013) In vivo evaluation of the iPex and Root ZX electronic apex locators using various irrigants. *International Endodontic Journal*, 46, 769–774.

El Kholy, K., Janner, S.F.M., Schimmel, M. & Buser, D. (2019) The influence of guided sleeve height, drilling distance, and drilling key length on the accuracy of static computer-assisted implant surgery. *Clinical Implant Dentistry and Related Research*, 21, 101–107.

ElAyouti, A., Kimionis, I., Chu, A.L. & Löst, C. (2005) Determining the apical terminus of root-end resected teeth using three modern apex locators: a comparative ex vivo study. *International Endodontic Journal*, 38, 827–833.

ElAyouti, A. & Löst, C. (2006) A simple mounting model for consistent determination of the accuracy and repeatability of apex locators. *International Endodontic Journal*, 39, 108–112.

ElAyouti, A., Weiger, R. & Löst, C. (2002) The ability of Root ZX apex locator to reduce the frequency of overestimated radiographic working length. *Journal of Endodontics*, 28, 116–119.

Gardner, M.J. & Altman, D.G. (1986) Confidence-intervals rather than *p*-values - estimation rather than hypothesis-testing. *British Medical Journal*, 292, 746–750.

Gigerenzer, G. (2004) Mindless statistics. *The Journal of Socio-Economics*, 33, 587–606.

Hör, D., Krusy, S. & Attin, T. (2005) Ex vivo comparison of two electronic apex locators with different scales and frequencies. *International Endodontic Journal*, 38, 855–859.

Janner, S.F.M., Jeger, F.B., Lussi, A. & Bornstein, M.M. (2011) Precision of endodontic working length measurements: a pilot investigation comparing cone-beam computed tomography scanning with standard measurement techniques. *Journal of Endodontics*, 37, 1046–1051.

Jeger, F.B., Janner, S.F.M., Bornstein, M.M. & Lussi, A. (2012) Endodontic working length measurement with preexisting cone-beam computed tomography scanning: a prospective, controlled clinical study. *Journal of Endodontics*, 38, 884–888.

Jordan, R.A., Holzner, A.L., Markovic, L., Brueckner, I. & Zimmer, S. (2014) Clinical effectiveness of basic root canal treatment after 24 months: a randomized controlled trial. *Journal of Endodontics*, 40, 465–470.

Jung, I.Y., Yoon, B.H., Lee, S.J. & Lee, S.J. (2011) Comparison of the reliability of "0.5" and "APEX" mark measurements in two frequency-based electronic apex locators. *Journal of Endodontics*, 37, 49–52.

Kojima, K., Inamoto, K., Nagamatsu, K., Hara, A., Nakata, K. et al. (2004) Success rate of endodontic treatment of teeth with vital and nonvital pulps. A meta-analysis. *Oral Surgery Oral Medicine Oral Pathology Oral Radiology and Endodontology*, 97, 95–99.

Liang, Y.H., Jiang, L., Chen, C., Gao, X.J., Wesselink, P.R., Wu, M.-K. et al. (2013) The validity of cone-beam computed tomography in measuring root canal length using a gold standard. *Journal of Endodontics*, 39, 1607–1610.

Liang, Y.H., Li, G., Wesselink, P.R. & Wu, M.K. (2011) Endodontic Outcome Predictors Identified with Periapical Radiographs and Cone-beam Computed Tomography Scans. *Journal of Endodontics*, 37, 326–331.

Lucena, C., Lopez, J.M., Martin, J.A., Robles, V. & Gonzalez-Rodriguez, M.P. (2014) Accuracy of working length measurement: electronic apex locator versus cone-beam computed tomography. *International Endodontic Journal*, 47, 246–256.

Mancini, M., Palopoli, P., Iorio, L., Conte, G. & Cianconi, L. (2014) Accuracy of an electronic apex locator in the retreatment of teeth obturated with plastic or cross-linked gutta-percha carrier-based materials: an ex vivo study. *Journal of Endodontics*, 40, 2061–2065.

Marcos-Arenal, J.L., Rivera, E.M., Caplan, D.J. & Trope, M. (2009) Evaluating the paper point technique for locating the apical foramen after canal preparation. *Oral Surgery Oral Medicine Oral Pathology Oral Radiology and Endodontology*, 108, E101–E105.

Metska, M.E., Liem, V.M.L., Parsa, A., Koolstra, J.H., Wesselink, P.R. & Ozok, A.R. (2014) Cone-beam computed tomographic scans in comparison with periapical radiographs for root canal length measurement: an in situ study. *Journal of Endodontics*, 40, 1206–1209.

Miletic, V., Beljic-Ivanovic, K. & Ivanovic, V. (2011) Clinical reproducibility of three electronic apex locators. *International Endodontic Journal*, 44, 769–776.

Nagendrababu, V., Murray, P.E., Ordinola-Zapata, R., Peters, O.A., Rocas, I.N. et al. (2021) PRILE 2021 guidelines for reporting laboratory studies in Endodontology: explanation and elaboration. *International Endodontic Journal*, 54, 1491–1515.

Nekoofar, M.H., Ghandi, M.M., Hayes, S.J. & Dummer, P.M. (2006) The fundamental operating principles of electronic root canal length measurement devices. *International Endodontic Journal*, 39, 595–609.

Nelson, P., Lucisano, M.P., Leonardo, M.R., da Silva, R.A.B. & da Silva, L.A.B. (2010) Electronic working length determination in primary teeth by Propex and digital signal processing. *Australian Endodontic Journal*, 36, 105–108.

Ng, Y.L., Mann, V., Rahbaran, S., Lewsey, J. & Gulabivala, K. (2008) Outcome of primary root canal treatment: systematic review of the literature - Part 2. Influence of clinical factors. *International Endodontic Journal*, 41, 6–31.

Odabas, M.E., Bodur, H., Tulunoglu, O. & Alacam, A. (2011) Accuracy of an electronic apex locator: a clinical evaluation in primary molars with and without resorption. *Journal of Clinical Pediatric Dentistry*, 35, 255–258.

Patel, S., Brown, J., Semper, M., Abella, F., Mannocci, F. et al. (2019) European society of endodontology position statement: use of cone beam computed tomography in endodontics European society of endodontology (ESE) developed by. *International Endodontic Journal*, 52, 1675–1678.

Ricucci, D. & Langeland, K. (1998) Apical limit of root canal instrumentation and obturation, part 2. A histological study. *International Endodontic Journal*, 31, 394–409.

Saatchi, M., Aminozarbian, M.G., Hasheminia, S.M. & Mortaheb, A. (2014) Influence of apical periodontitis on the accuracy of 3 electronic root canal length measurement devices: an in vivo study. *Journal of Endodontics*, 40, 355–359.

Schell, S., Judenhofer, M.S., Mannheim, J.G., Hulber-J, M., Lost, C., Pichler, B.J. et al. (2017) Validity of longitudinal sections for determining the apical constriction. *International Endodontic Journal*, 50, 706–712.

Scott, N.W., McPherson, G.C., Ramsay, C.R. & Campbell, M.K. (2002) The method of minimization for allocation to clinical trials: a review. *Controlled Clinical Trials*, 23, 662–674.

Sittikornpaiboon, P., Arunjaroensuk, S., Kaboosaya, B., Subbalekha, K., Mattheos, N. et al. (2021) Comparison of the accuracy of implant placement using different drilling systems for static computer-assisted implant surgery: A simulation-based experimental study. *Clinical Implant Dentistry and Related Research*, 23, 635–643.

Sjögren, U., Hagglund, B., Sundqvist, G. & Wing, K. (1990) Factors affecting the long-term results of endodontic treatment. *Journal of Endodontics*, 16, 498–504.

Souza, E. (2014) Research that matters: setting guidelines for the use and reporting of statistics. *International Endodontic Journal*, 47, 115–119.

Stabholz, A., Rotstein, I. & Torabinejad, M. (1995) Effect of preflaring on tactile detection of the apical constriction. *Journal of Endodontics*, 21, 92–94.

Stoll, R., Urban-Klein, B., Roggendorf, M.J., Jablonski-Momeni, A., Strauch, K. & Frankenberger, R. (2010) Effectiveness of four electronic apex locators to determine distance from the apical foramen. *International Endodontic Journal*, 43, 808–817.

Ustun, Y., Aslan, T., Sekerci, A.E. & Sagsen, B. (2016) Evaluation of the reliability of cone-beam computed tomography scanning and electronic apex locator measurements in working length determination of teeth with large periapical lesions. *Journal of Endodontics*, 42, 1334–1337.

Uzun, O., Topuz, O., Tinaz, C., Nekoofar, M.H. & Dummer, P.M.H. (2008) Accuracy of two root canal length measurement devices integrated into rotary endodontic motors when removing gutta-percha from root-filled teeth. *International Endodontic Journal*, 41, 725–732.

Vandevoorde, H.E. & Bjorndahl, A.M. (1969) Estimating endodontic working length with paralleling radiographs. *Oral Surgery Oral Medicine Oral Pathology Oral Radiology and Endodontology*, 27, 106–110.

Vasconcelos, B.C., Frota, L.M., Souza Tde, A., Bernardes, R.A. & Duarte, M.A. (2015) Evaluation of the maintenance of the apical limit during instrumentation with hybrid equipment in rotary and reciprocating modes. *Journal of Endodontics*, 41, 682–685.

Vieyra, J.P. & Acosta, J. (2011) Comparison of working length determination with radiographs and four electronic apex locators. *International Endodontic Journal*, 44, 510–518.

Williams, C.B., Joyce, A.P. & Roberts, S. (2006) A comparison between in vivo radiographic working length determination and measurement after extraction. *Journal of Endodontics*, 32, 624–627.

Zaugg, L.K., Savic, A., Amato, M., Amato, J., Weiger, R., Connert, T. et al. (2020) Endodontic treatment in Switzerland - a national survey. *Swiss Dental Journal*, 130, 18–29.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: ElAyouti, A., Connert, T., Dummer, P & Löst, C. (2022) A critical analysis of research methods and experimental models to study working length determination and the performance of apex locators – A narrative review with recommendations for the future. *International Endodontic Journal*, 55(Suppl. 2), 281–294. Available from: https://doi.org/10.1111/iej.13738