ORIGINAL ARTICLE

Are "virtual" paediatric weight estimation studies valid?

Mike Wells*, Lara Goldstein

Division of Emergency Medicine, Faculty of Health Sciences, University of the Witwatersrand, South Africa

ARTICLE INFO

ABSTRACT

Introduction: "Virtual" studies account for nearly one-third of all published weight estimation articles, but the validity of these virtual studies has never been evaluated. It is important to establish this validity in order to decide whether the results of these studies can be applied to real-world usage. The objectives of this study were to evaluate the accuracy of virtual weight estimates using the Broselow and PAWPER tapes and compare these to actual real-life estimates from the tapes.

Methods: Virtual weights were generated for the Broselow and PAWPER tapes using anthropometric data from a sample of 1385 children for whom actual Broselow and PAWPER tape weights were available. The accuracy of the virtual and real-life estimates was compared against each child's actual weight. The agreement of the virtual and real estimates was also evaluated.

Results: The percentage of weight estimates within 10% of actual weight were 57.9% and 59.3% for the real and virtual Broselow tapes respectively and 76.6% and 78.4% for the real and virtual PAWPER tapes respectively. The Cohen's kappa for the real and virtual Broselow and PAWPER tapes was 0.65 and 0.64 respectively, which indicated substantial agreement.

Conclusions: The virtual and real weight estimates had very similar accuracy outcomes for both tapes in this study. However, if virtual studies are used, they should be followed by real-life studies in order to assess the impact of human and patient factor errors on the accuracy of the weight estimation systems.

African relevance

- Virtual weight estimation studies are commonly-used in studies from LMICs, without any evidence of their validity.
- Virtual studies are useful as initial studies to evaluate the potential accuracy of weight estimation systems in LMICs.
- Accurate weight estimation is important in resource-poor settings where accurate scales might not be available.

Introduction

Many paediatric weight estimation studies are "virtual" studies, but we do not know if there is scientific validity in the assumption that the findings from virtual studies can be applied in the real world. A virtual weight estimation study is generally a retrospective study in which demographic or anthropometric data is used to calculate an estimated weight, but the method is not used in real children (e.g. measured length is used to determine Broselow tape weight instead of using the Broselow tape itself). In a systematic review published in 2016, 30/46 studies on the Broselow tape were virtual studies, and in a meta-analysis published in 2017, 28/98 of all types of studies with usable data were virtual studies [1,2]. These studies frequently appear disproportionately impressive because of the large numbers of children that can be included for evaluation in the study.

While it has been assumed that virtual studies are equivalent to prospective, real-life studies, there is no evidence to support this assumption. In fact, there is good evidence that human factor errors and patient factor errors may contribute significantly to differences in the performance of weight estimation systems under virtual, real and emergency circumstances [3–5]. It is also quite possible that technical considerations may contribute substantially to differences in outcomes between real and virtual studies (e.g. the difference between stadiometer-measured height and Broselow tape-measured length).

Since these studies are widely used to develop and validate certain weight estimation systems (often in new populations), it would be of value for clinicians, policy- and decision-makers to know whether these studies are scientifically valid.

The primary objective of this study was to compare weight estimates acquired using the Broselow tape and PAWPER tape in real life with virtual Broselow tape and PAWPER tape estimates generated from

anthropometric measurements in the same children. The secondary objective was to review the literature to identify other factors that might impact on the validity and generalizability of virtual weight estimation studies.

Methods

This was a secondary analysis of pooled data from four previous prospective paediatric weight estimation studies (three with published data and one with unpublished data) [6–8]. The original studies were approved by the Human Research Ethics Committee of the University of the Witwatersrand, and a new approval was obtained for the secondary analysis (M1511107).

Each included study was a prospective, observational, cross-sectional study in Johannesburg. Two of the studies were conducted in low-income communities and two in middle-income communities. The first study enrolled 453 children from September 2008 to October 2008, the second study enrolled 332 children from July 2014 to December 2014, the third study enrolled 300 children from August 2014 to January 2015 and the final study enrolled 300 children from June 2017 to January 2018.

These studies made use of convenience samples of children who did not require emergency medical care. Children aged 1 month to 12 years (Studies 1 and 3) or 1 month to 16 years (Studies 2 and 4) were included. Children with congenitally abnormal stature and children whose length could not be measured were excluded. Parents signed informed consent for all participants and children over the age of seven years signed assent.

The details of the study procedures can be found in the individual publications [6–8]. Briefly, each child had a weight estimation using the Broselow tape (edition 2007B for the first study and 2011A for the others) and the PAWPER tape (or PAWPER XL tape for the latter two studies) according to the manufacturer's instructions. Each child also had their recumbent length measured using a standard tape measure. After the anthropometric measurements, each child's actual weight was measured on a digital scale and recorded to the nearest 0.1 kg.

The records from each of the contributing studies were pooled for the secondary analysis. The data that were included from each of the children in this analysis were age, recumbent length, Broselow tape weight, Broselow tape colour zone, PAWPER tape weight, PAWPER tape habitus score, BMI-for-age Z-scores and actual measured weight.

The virtual Broselow tape weights were generated in a Microsoft Excel spreadsheet using the measured recumbent length of each child. The formulas used in the spreadsheet were custom-developed based on precise measurements of the appropriate Broselow tape edition which were obtained from a previous study [9]. The corresponding Broselow tape colour zones were determined from the weights.

The virtual PAWPER tape weights were generated in a similar fashion, using the measured recumbent length and the original recorded habitus score (the habitus score for the PAWPER tape cannot be generated from anthropometric data but is based on a visual assessment of habitus by the user). The spreadsheet with the data and formulas is included in the Supplementary material.

In the data analysis, the performance of the real and virtual methods was compared against the children's actual measured weight. The principal indicators of this performance were:

- mean percentage error (MPE) which is a measure of bias;
- the Bland & Altman 95% limits of agreement of the MPE;
- the root mean squared percentage error (RMSPE) which is a measure of precision;
- the percentage of weight estimations falling within 10% and 20% of actual weight (p10 and p20 respectively) which is a measure of overall accuracy.

The percentages of weight estimations falling within 10% and 20%

is a standard descriptor used in weight estimation studies. It is clinically relevant as a substantial proportion of children receiving a weight estimation with > 10% error are at risk of harm from a medication error [10].

Differences in continuous data were analyzed using the paired t-test and between proportions using the Fisher exact test. Secondly, the actual differences in weight estimations between the real and the virtual methods, ignoring actual measured weight, were categorized in percentage categories. A similar analysis was used to determine differences in prediction of colour zones for the Broselow tape. Finally, the agreement between real and virtual weight estimations were analyzed in terms of inter-rater assessment statistics using Cohen's kappa. A percentage difference of > 10% between the real and the virtual systems was considered to be a meaningful difference. This cutoff was chosen because of evidence that a weight estimate error > 10% is associated with patient harm from medication error [11].

The 95% level was used to determine significance for all statistical analyses (p < 0.05). The differences between the real and virtual weight estimations were also evaluated separately in each dataset, to identify any potential changes over time and discrepancies between populations.

Microsoft excel (Microsoft Excel for Mac version 16.14.1) and GraphPad Prism (GraphPad Prism version 8.00 for Mac, GraphPad Software, La Jolla California USA, www.graphpad.com) were used for all data management and statistical analysis.

Results

There were 1385 children in the pooled data sample. All 1385 were able to receive a weight estimation using the PAWPER tape system, but only 1279 could receive a weight estimation using the Broselow tape, as 106 (7.7%) were too tall for the Broselow tape.

The demographic information for the study sample is shown in Table 1. There was a large number of underweight and very underweight children in the study sample because two of the studies from which data were obtained were conducted in populations of low socioeconomic status with a very high prevalence of underweight children.

The outcomes of the comparisons between the real and virtual weight estimation methods and actual measured weight are shown in Table 2.

The comparisons between the real and virtual tape outcomes i.e. the actual differences in weight estimations between the real and the virtual methods are shown in Fig. 1.

With respect to the Broselow colour zones, 1116/1279 (87.3%)

Table 1

Demographic information for the study sample. IQR = interquartile range; BMI-for-age Z-scores used. The cutoffs for the weight categories were based on the World Health Organization standards and definitions [12]. The category of "slightly underweight" children was added as an intermediate category to include children between 1 and 2 standard deviations below the mean BMI-for-age.

Variable	Value
n	1385
Age (months) [median (IQR)]	62 (26, 122)
Sex = male [n (%)]	718 (51.8)
Length (cm) [median (IQR)]	109 (87, 128)
Weight (kg) [median (IQR)]	17.6 (12.5, 25.8)
BMI [median (IQR)]	15.8 (14.2, 17.5)
Z-score [median (IQR)]	−0.1 (−1.2, 0.8)
Slightly underweight (−2.0 < Z-score ≤ −1.3) [n (%)]	117 (8.4)
Very underweight (Z-score ≤ −2.0) [n (%)]	112 (8.1)
Normal weight (−1.4 < Z-score < 2.0)	999 (72.1)
Obese (2.5 > Z-score ≥ 2.0) [n (%)]	47 (3.4)
Severely obese (Z-score ≥ 2.5) [n (%)]	39 (2.8)

IQR, interquartile range.

Table 2
Comparisons on the weight estimation accuracy of the real and virtual weight estimation systems.

	MPE (95% LOA) (%)	RMSPE (%) Mean (SD)	p10 (%)	p20 (%)
Broselow tape – real	1.3 (−26.7, 29.2)	10.9 (9.3)	57.9	85.8
Broselow tape – virtual	2.4 (−25.5, 30.3)	10.7 (9.7)	59.3	86.6
N = 1339	Paired t-test	Paired t-test	Fisher exact test	Fisher exact test
	p < 0.001	p = 0.48	p = 0.78	p = 0.91
PAWPER tape – real	1.1 (−18.9, 21.1)	7.4 (7.1)	76.6	95.0
PAWPER tape – virtual	1.6 (−17.5, 20.8)	7.0 (7.0)	78.4	95.3
N = 1339	Paired t-test	Paired t-test	Fisher exact test	Fisher exact test
	p < 0.001	p = 0.001	p = 0.19	p = 0.76

MPE = mean percentage error, LOA = limits of agreement, RMSPE = root mean square percentage error, SD = standard deviation, p10 = proportion of weight estimates within 10% of actual weight, p20 = proportion of weight estimates within 20% of actual weight. A negative MPE is indicative of a bias to underestimation of weight.

children were predicted in the same zone by the real and virtual tapes, 112/1279 (8.8%) children were predicted one zone too high by the virtual tape and 50/1279 (3.9%) children were predicted one zone too low by the virtual tape.

The Cohen's kappa for both the Broselow tape and PAWPER tape real and virtual weights showed substantial agreement (0.65 and 0.64 respectively). The agreement in terms of Broselow colour zones was also substantial, with a Cohen's kappa of 0.78. Fig. 1 shows the percentage difference categories. A negative value indicates a virtual tape weight that was greater than the real tape weight.

There were no significant differences in the real and virtual measurements between the individual datasets for either the Broselow tape or the PAWPER tape. There were also no differences found when comparing the virtual and real estimations for the Broselow tape 2007B with the 2011A.

Discussion

Virtual studies have been used to evaluate age-based weight estimation formulas, length-based weight estimation formulas, the Broselow tape, other length-based tapes, the Mercy method and the PAWPER XL-MAC method [13–17]. This has been done despite the absence of evidence supporting the validity of the underlying assumption: that the use of the method in real-life can be fully reproduced in a virtual study. There are two considerations in this regard – can the virtual methodology replicate the real methodology satisfactorily in terms of accuracy; and is the impact of human factor errors and patient factor errors sufficiently inconsequential to allow for definitive

conclusions to be drawn from the virtual study about the accuracy and usability of the weight estimation method during emergency care?

What degree of correspondence between virtual and real weight estimations would be sufficient to draw conclusions about the methodological accuracy of the system? The limited data available on the inter-observer reliability of the Broselow tape suggests that there may be a 10–20% relative difference in accuracy between experienced and novice users of the Broselow tape under non-emergency conditions [3]. A virtual weight estimation performance that was proven to be within 10% of the accuracy of the system used in real life would, therefore, have the potential to provide information that could be extrapolated to real-life accuracy in non-emergency settings.

In this study the accuracy of the virtual Broselow tape was indistinguishable from the accuracy of the real Broselow tape. Although the virtual estimates of weight were slightly higher than the real estimates, the difference was negligible. A similar pattern was seen with the PAWPER tape. The differences in accuracy were small enough to conclude that virtual studies using these two methods could provide valuable data on their maximal potential methodological accuracy.

A previous study evaluating the inter-observer error between prehospital and emergency department use of the Broselow tape showed a good agreement for colour zone selection (Cohen's kappa of 0.74) [18]. This was similar to the agreement between the virtual and real Broselow tapes found in this study (Cohen's kappa of 0.78). This adds weight to the validity of virtual tapes as predictors of the maximum potential accuracy of weight estimation systems. Real-life accuracy may not be optimum, however, because of real-world human factor and patient factor confounders.
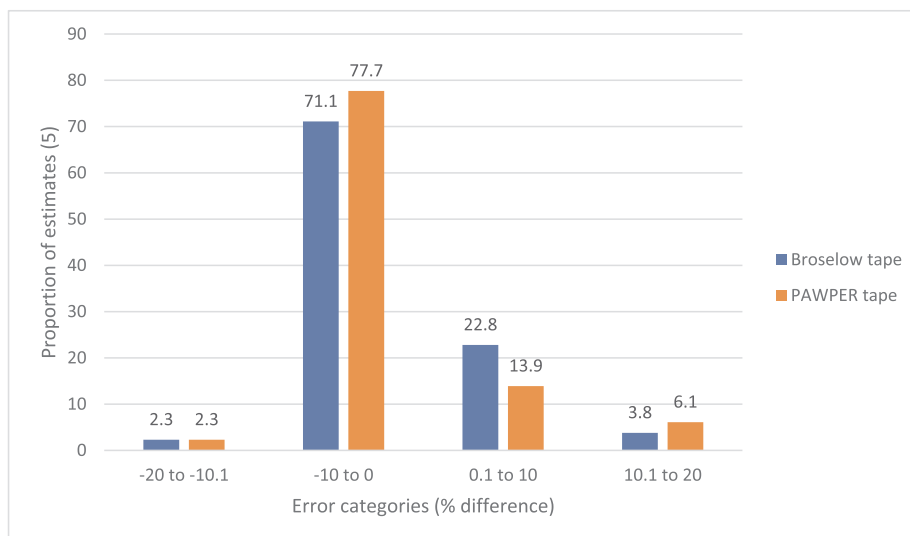

Fig. 1. Differences between real and virtual weight estimates for the Broselow and PAWPER tapes.

The assumption that other real-life considerations, such as human errors and patient-related factors (especially patient position and degree of cooperation) have a negligible impact on the performance of a weight estimation system is naïve and not supported by current evidence [4].

Virtual weight estimation studies and those conducted in non-clinical (or non-simulated clinical) conditions may not provide an accurate indication of the usability of a weight estimation system or its accuracy under adverse conditions. In a recent meta-analysis, only seven of the 150 weight estimation studies included were conducted in real or simulated paediatric emergencies with two further articles published subsequently [2,4,5]. This means that there is very little evidence on the actual functioning of weight estimation systems when exposed to human and patient factor errors. The fact that the accuracy of weight estimation systems evaluated during real or simulated emergencies was lower than the same systems tested in non-clinical settings is compelling evidence that further research is needed to identify and quantify the influence of method-factor, human-factor and patient-factor errors [2,9].

The limited evidence that is available for the Broselow tape, the Mercy method and the PAWPER XL tape suggests that undertrained users are substantially less accurate than experienced users during simulated emergencies [4]. There is also evidence that age-based formulas cannot be remembered during emergencies, the calculations cannot be performed accurately and might be unusable as a result [5]. Similarly, adverse patient position and cooperation – common during real emergencies – can have significant effects on the accuracy of certain weight estimation systems, particularly the Broselow tape and the Mercy method [4]. Even outside of emergency use, human factor errors have been shown to have a significant impact on the accuracy of weight estimation systems [3]. These findings suggest that both virtual studies as well as real-life studies in non-emergency settings (real or simulated) might not provide sufficient information about the real-world accuracy and usability of weight estimation systems. It is clear that one of the reasons that the virtual studies are able to emulate real-life studies is that most real-life studies are not performed in emergency situations. Both these types of studies are thus somewhat artificial and should be interpreted as being indicative of the maximum potential accuracy within any given population. The studies remain useful but must be considered as preliminary and should be followed by more realistic research.

Both the real and virtual Broselow and the PAWPER tapes had very similar accuracy outcomes in this study. This suggests that virtual weight estimation studies can indeed have a valuable role to play in the initial development or validation of length-based weight estimation methodologies. What is equally clear from the evidence available from other studies, is that virtual studies *must* be followed by real-life studies in which an evaluation should be conducted of the performance of the systems under the conditions similar to those in which they would be used (e.g. simulated paediatric emergencies). This would be essential to identify the impact and magnitude of human and patient factor errors on the accuracy and usability of weight estimation systems.

The virtual PAWPER tape was not completely virtual in this study as the original "real-life" habitus scores were used. The differences between the real and virtual results of the length component of the tape were nonetheless interesting enough to warrant inclusion in the study.

The children included in this study had their recumbent length measured using a measuring tape, so the results might not be generalizable for children who have their height measured using a stadiometer. This findings in this study can also not be generalized to age-based formula systems: although virtual studies may be able to evaluate the potential accuracy of these formulas, not enough work has been done to even speculate meaningfully on the impact of human and patient factor errors on the use of age-formulas in emergencies. Furthermore, current evidence does not support the use of age-formulas in terms of both accuracy and usability [5,19].

Dissemination of results

This paper has not been disseminated beyond this publication.

Author contribution

Authors contributed as follow to the conception or design of the work; or the acquisition, analysis, or interpretation of data for the work; drafting the work or revising it critically for important intellectual content; and final approval of the version to be published: MW 70% and LG 30%. Both authors agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Conflicts of interest

Prof Mike Wells is the developer of the PAWPER tape but derives no financial benefit from it. Prof Wells is an editor of the African Journal of Emergency Medicine. Prof Wells was not involved in the editorial workflow for this manuscript. The African Journal of Emergency Medicine applies a double blinded process for all manuscript peer reviews. The authors declared no further conflict of interest.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.afjem.2019.01.008.

References

[1] Young KD, Korotzer NC. Weight estimation methods in children: a systematic review. Ann Emerg Med 2016;68(4):441–51. e10.

[2] Wells M, Goldstein L, Bentley A. The accuracy of emergency weight estimation systems in children – a systematic review and meta-analysis. Int J Emerg Med 2017;10(29):1–43. https://doi.org/10.1186/s12245-017-0156-5.

[3] Abdel-Rahman SM, Jacobsen R, Watts JL, Doyle SL, O'Malley DM, Hefner TD, et al. Comparative performance of pediatric weight estimation techniques: a human factor errors analysis. Pediatr Emerg Care 2017;33(8):548–52.

[4] Wells M, Goldstein L, Bentley A. The accuracy of paediatric weight estimation during simulated emergencies: the effects of patient position, patient cooperation and human errors. Afr J Emerg Med 2018;8(2):43–50.

[5] Marlow RD, Wood DLB, Lyttle MD. Comparing the usability of paediatric weight estimation methods: a simulation study. Arch Dis Child 2018.

[6] Wells M, Coovadia A, Kramer E, Goldstein L. The PAWPER tape: a new concept tape-based device that increases the accuracy of weight estimation in children through the inclusion of a modifier based on body habitus. Resuscitation 2013;84(2):227–32.

[7] Georgoulas V, Wells M. The PAWPER tape and the Mercy Method outperform other methods of weight estimation in children in South Africa. S Afr Med J 2016;106(9):933–9.

[8] Wells M, Goldstein L, Bentley A. A validation study of the PAWPER XL tape: accurate estimation of both total and ideal body weight in children up to 16 years of age. Trauma Emerg Care 2017;2(4):1–8.

[9] Wells M, Goldstein L, Bentley A, Basnett S, Monteith I. The accuracy of the Broselow tape as a weight estimation tool and a drug-dosing guide – a systematic review and meta-analysis. Resuscitation 2017;121:9–17. https://doi.org/10.1016/j.resuscitation.2017.09.026.

[10] Hirata KM, Kang AH, Ramirez GV, Kimata C, Yamamoto LG. Pediatric weight errors and resultant medication dosing errors in the Emergency Department. Pediatr Emerg Care 2017.

[11] Murugan S, Parris P, Wells M. Drug preparation and administration errors during simulated paediatric resuscitations. Arch Dis Child 2018. (in press).

[12] Technical report on WHO child growth standards-length/height-for-age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age – methods and development 2006 [accessed 9.11.2018]. Available from: http://www.who.int/childgrowth/standards/technical_report/en/index.html.

[13] Marlow R, Lo D, Walton L. Accurate paediatric weight estimation by age: mission impossible? Arch Dis Child 2011;96(Suppl. 1):A1–2.

[14] Jang HY, Shin SD, Kwak YH. Can the Broselow tape be used to estimate weight and

endotracheal tube size in Korean children? Acad Emerg Med 2007;14(5):489–91.

[15] Both CP, Schmitz A, Buehler PK, Weiss M, Schmidt AR. How accurate are pediatric emergency tapes? A comparison of 4 emergency tapes with different length-based weight categorization. Pediatr Emerg Care 2017.

[16] Abdel-Rahman SM, Ridge AL. An improved pediatric weight estimation strategy. Open Med Dev J 2012;4:87–97.

[17] Wells M, Goldstein L, Bentley A. Development and validation of a method to estimate bodyweight in critically ill children using length and mid-arm circumference measurements – the PAWPER XL-MAC system. S Afr Med J 2017;107(11):1015–21.

[18] Heyming T, Bosson N, Kurobe A, Kaji AH, Gausche-Hill M. Accuracy of paramedic Broselow tape use in the prehospital setting. Prehospital Emerg Care 2012;16(3):374–80.

[19] Wells M, Goldstein L, Bentley A. It is time to abandon age-based emergency weight estimation in children! A failed validation of 20 different age-based formulas. Resuscitation 2017;116(7):73–83.