Adherence to guidelines strongly improves reproducibility of brachial artery flow-mediated dilation

Arno Greyling a, b, Anke C.C.M. van Mil a, c, d, Peter L. Zock b, c, Daniel J. Green e, Lorenzo Ghiadoni f, Dick H. Thijssen a, g, *, on behalf of the TIFN International Working Group on Flow Mediated Dilation

a Department of Physiology, Radboud University Medical Centre, Nijmegen, The Netherlands
b Unilever R&D Vlaardingen, Vlaardingen, The Netherlands
c TI Food and Nutrition, Wageningen, The Netherlands
d Maastricht University Medical Centre, Maastricht, The Netherlands
e School of Sports Science, Exercise and Health, The University of Western Australia, Crawley, Australia
f University of Pisa, Pisa, Italy
g Research Institute for Sport and Exercise Sciences, Liverpool John Moores University, Liverpool, United Kingdom

1. Introduction

The endothelium is a key regulator of vascular homeostasis and endothelial dysfunction is an early manifestation of atherosclerosis [1]. Currently, the most widely used technique to study endothelial function in vivo is the flow-mediated dilation (FMD) of the brachial artery. This is a non-invasive, ultrasound-based method which correlates with endothelial function of the coronary arteries [2, 3] and independently predicts cardiovascular disease (CVD) [4, 5]. The technique is attractive as a surrogate end-point, especially since changes in FMD can be detected across a relatively short timeframe [6]. Despite its popularity, minor changes in the methodological approach may critically impact variability and decrease reproducibility of the FMD response [7–9].

Previous expert consensus guidelines have made important contributions to standardize the technical approach and to set minimum standard requirements for FMD measurements [10, 11]. However, not all studies on FMD apply these recommendations, or...
only in part. The impact of adherence to these guidelines on the reproducibility of FMD measurements is currently unclear, but may importantly contribute to the measurement error of the FMD technique. Furthermore, little is known about the relative importance of the individual aspects of the expert-consensus guidelines to contribute to the reproducibility of the FMD. Better quantitative data on this matter can help reduce variation within and between studies, which will increase the statistical power of studies on FMD to detect changes and, subsequently, decrease chances for type II errors.

In light of these considerations, we hypothesized that adherence to expert consensus guidelines is related to better reproducibility of FMD measurements [10,11]. Therefore, we performed a systematic search for published studies that reported data on reproducibility of FMD measurements, and investigated the relation between (full or partial) adherence to current expert consensus guidelines and reproducibility of the FMD. Secondly, we explored which subject- and methodology-related factors were related to FMD reproducibility.

2. Methods

2.1. Search strategy

The MEDLINE bibliographic database was searched (January 2000 through July 2015) for studies that assessed the reproducibility of the FMD using the following search terms: “flow mediated dilation”, “flow mediated dilatation”, “flow mediated vasodilation”, “flow mediated vasodilatation”, “endothelial function”, “endothelial dysfunction”, “FMD”, “FMV”, “brachial artery”, “reproducibility”, “reliability”, “repeatability”, “coefficient of variation”, “CV”, and “variance”. The search was limited to studies in human adults published in the English language. Additionally, we supplemented the search by hand-searching references of included studies and relevant reviews and meta-analyses on this topic.

2.2. Selection of studies

Included studies were identified by means of a two-step selection process. During the first step, two reviewers (ACCMvM, AG) independently screened titles, abstracts and keywords of publications to identify potentially eligible studies. Studies were included if the mean difference and variance of repeated FMD measurements of the brachial artery were reported. During step 2 of the selection, both reviewers examined the full text of these publications to gauge eligibility based on two additional inclusion criteria: FMD was determined through noninvasive ultrasound imaging, and a reactive hyperaemia protocol (with an ischemia duration of 4–5 min) was used to elicit the shear stress stimulus required for FMD. Thus, studies that adopted (ischemic) hand-grip exercise, passive movement and/or skin warming protocols to elicit (brachial) artery dilation were not included in our analysis. In cases of discrepancy between the reviewers, eligibility was discussed along with a third reviewer (DHJT) until consensus was reached.

2.3. Data extraction

2.3.1. Study and subject characteristics

A standardized data collection sheet was used to extract general publication details (author, year of publication, country) and specific study- and subject characteristics: number of subjects, mean age (in years); CVD risk status of the study population (defined as presence of diagnosed CVD, hypertension or diabetes); baseline brachial artery diameter (in mm); % brachial artery FMD and its associated variance for each repeated measurement; and the mean absolute difference between repeated FMD measurements and its associated variance.

2.3.2. Adherence to guidelines

We extracted information from the methods sections of the individual papers to assess the adherence to current expert-consensus guidelines. Based on recent guidelines [11], we scored each individual study on the reporting of 19 different factors which were divided over 4 categories. The categories were related to: 1. Subject preparation (10 items), 2. Image acquisition (4 items), 3. Data analysis (3 items), and 4. Laboratory (2 items). Before performing the systematic literature search, values were assigned to each factor proportional to its perceived importance for valid assessment of the FMD. This was done through expert consensus discussion within the Working Group (AG, LG and DHJT) [12]. The “Adherence Score” that could be assigned to a study ranged from 0 to 10 points, depending on how many of the 19 different factors that were reported. In addition, we counted the number of previous studies on FMD published by the principal author of each study included in the systematic review. This number served as a measure of the perceived experience in FMD measurements for each centre at the time of publication of the reproducibility data included here.

2.4. Statistical analysis

Reported measures of FMD reproducibility varied between studies. Many studies presented the coefficient of variation (CV) of repeated measurements, although this measure was calculated in a number of different ways, precluding direct comparisons. Measures of reproducibility included the technical error of the measurement (TEM), Pearson- and intraclass correlation coefficients (ICC), and limits of agreement. In order to make valid comparisons between studies, we defined as primary outcome measure the typical error of estimate (TEE) of FMD, which is calculated as standard deviation of the paired differences/√2 [13].

Data are presented as mean ± SD or median (range) as appropriate for continuous variables and as frequencies for categorical variables. FMD TEE data were highly skewed (Shapiro-Wilk test, P < 0.0001) and were log transformed prior to the analyses. Relations between log-FMD TEE and continuous variables were determined by Pearson correlation coefficients analysis. For categorical variables, the statistical significance of differences in FMD TEE between different levels were assessed by the Mann-Whitney U test. Significant correlates were entered in a multivariate linear regression analyses with backward elimination to identify independent predictors of FMD TEE. All analyses were conducted using JMP version 11.0 (SAS Institute Inc., Cary, NC, USA).

3. Results

Our systematic search identified 446 potentially relevant publications and an additional 12 were obtained through review of references of included studies, relevant reviews and meta-analyses. Twenty-seven studies [14–40] with 48 relevant study groups met our inclusion criteria and were included in our analysis (Fig. 1). Characteristics of the included study groups are presented in Table 1. The 48 study groups comprised a total of 1537 subjects (mean sample size 32; range, 8–135) with a mean age of 41.5 years (range, 22–79 years). Eleven study groups included subjects with increased CVD risk, i.e. presence of diagnosed CVD, hypertension or diabetes. The other remaining 37 study groups consisted of healthy subjects. The time between repeated FMD measurements ranged from 25 min to 9 months. Mean baseline brachial diameter was 3.9 mm (range, 3.5–4.7 mm) and mean baseline FMD (i.e. this first
of the two repeated measurements) was 7.1% (range, 1.8–19.9%). The FMD TEE ranged from 0.33 to 4.83% across study groups, with a mean value of 1.4%. The level of experience for each centre at the time of publication of the reproducibility study in question varied widely (number of previously published studies on FMD ranging from 0 to 71, median of 3).

3.1. Methodology-related factors versus variation in FMD

There was considerable variation in the methodological factors between studies. Adherence scores ranged from 2.4 to 9.2, with a mean of 5.3 (out of a maximum of 10). The adherence score was inversely correlated with log FMD TEE (adjusted $R^2 = 0.36$, $P < 0.01$, Fig. 2).

To explore the impact of the different aspects of the adherence score on the FMD TEE, we compared the FMD TEE between adherence (Yes vs No) to various methodological variables. Statistically significant differences in FMD TEE were found for use of the true peak diameter to calculate %FMD, continuous brachial artery diameter measurement over the cardiac cycle, use of automated edge detection software and smoking cessation prior to measurements (Table 2).

3.2. Subject-related factors versus variation in FMD

For the remaining methodology related factors and subject characteristics, there were weak, but statistically significant correlations of log FMD TEE with age (adjusted $R^2 = -0.18$, $P < 0.01$) and with baseline FMD (adjusted $R^2 = 0.11$, $P = 0.013$). In addition, FMD TEE was significantly smaller in the subgroup of studies that applied a stereostatic probe holder, and in studies performed by groups with more experience according to number of earlier publications on FMD. The %FMD TEE of studies above and below the median duration between repeated measurements (7 days) was not significantly different (Table 3), and there was no correlation between %FMD TEE and the time between repeated measurements (adjusted $R^2 = -0.02$, $P < 0.75$).

We constructed a stepwise multivariate regression model with log FMD TEE as the dependent variable and all factors that significantly influenced FMD TEE based on the individual analyses (adherence score, age, baseline FMD, probe holder and previous experience). The stepwise multivariate regression model predicted 51% of the variability in log FMD TEE. Adherence score ($\beta = -0.16$), age ($\beta = -0.01$) and probe holder ($\beta = -0.19$) remained as statistically significant ($P < 0.05$) predictors in the model (Table 4).

4. Discussion

In the recent years the measurement of FMD in the brachial artery has been associated with predictive capacity for future CVD events. Despite this, and the relatively straightforward and non-invasive approach to its measurement, the clinical use of the FMD is hampered by its sensitivity to variations in methodology.

Our systematic analysis of previous studies that explored FMD reproducibility provides us with a number of novel observations. First, we found considerable variation in the methodology applied to measure FMD and consequently, differences between studies in the adherence to current expert consensus guidelines. Secondly, these data show a robust inverse association between adherence to the guidelines and FMD reproducibility, with higher adherence to guidelines being related to smaller variation in FMD. Thirdly, we identified methodological factors that were associated with smaller variation in FMD. Specifically, the use of automated edge detection software, continuous measurement of brachial artery diameter across the cardiac cycle, calculating %FMD by means of the true peak diameter and use of a stereostatic probe holder were related to a better reproducibility. Taken together, our study provides strong scientific data that highlight the importance of rigorous application of standardized contemporary methodology to reduce measurement error of FMD and, consequently, improve its use in (pre)clinical studies.

To our knowledge, no previous study has explored the (relative) importance of adherence to expert consensus guidelines for measures of vascular health, including frequently used techniques like intima-media thickness, pulse wave velocity, and finger photoplethysmography. Taking all studies on the reproducibility of the FMD together, involving 1537 subjects, we found a TEE of 1.4% based on an average FMD of 7.1%. This indicates an overall good-to-acceptable reproducibility of the FMD. However, significant variation was observed between studies, with adherence to the expert consensus guidelines representing an important determinant of this variation. Our data suggests that roughly 36% of the variation in
FMD reproducibility can be explained through adherence to the guidelines alone. The presence of a linear relation between adherence to the guidelines and variation of the FMD suggests that measurement error would be further reduced with stricter adherence to the guidelines. Our data also indicate that even with full adherence to current expert consensus guidelines, some level of measurement error remains present. Nonetheless, a significant amount of variation in the FMD can be prevented by strong adherence to guidelines.

Our analysis provides further insight into methodological factors that determine within-person error of the FMD measurement. For example, we found that taking the true peak artery diameter (rather than a fixed time point), continuous diameter measurement and automated edge-detection contribute to minimizing measurement error. The importance of these methodological factors have already been acknowledged in previous work. For example, Black et al. found that the peak diameter following cuff release differs between young and older subjects [7]. Consequently, calculating the FMD% at an arbitrary time point (e.g. 60 s) may lead to misleading conclusions compared to an approach in which diameter of the brachial artery is recorded continuously, allowing for the detection of the true peak dilation. Furthermore, previous work demonstrated that the adoption of edge-detection software to perform observer-independent analysis leads to smaller variation compared to the application of manual calipers, the latter being highly prone to measurement bias [36,41,42]. Whilst these studies highlight the importance of considering these factors for valid use of FMD, the present study also highlights the importance of considering these factors to lower variation. Therefore, our study provides an additional rationale to perform continuous assessment.
regardless of the use of a probe holder [19]. One may speculate that sonographers’ learning curves will likely differ depending on whether a probe holder is used or not and also depending on the design and construction of the probe holder itself. Therefore, despite the significant inverse association in our analysis, it remains difficult to ascertain whether use of a probe holder leads to a smaller variation in FMD per se. Further studies are needed to confirm the importance of using a probe holder to reduce variability of the FMD.

Of the subject-related factors (age, diameter and baseline FMD), only age contributed independently to the variation in FMD. Notably, higher age of subjects was associated with a smaller variation in FMD. Older age is typically associated with a lower FMD [43,44], which may contribute to a smaller (biological) variation and/or less ability to change in response to hemodynamic stimuli, consequently leading to a smaller measurement error. However, at least a previous work suggests the presence of larger variability for measurements of vascular health in clinical groups. For example, Craiem et al. found that subjects with CVD, despite comparable baseline FMD% values, demonstrate a larger coefficient of variation compared to healthy controls [22]. Our data suggest that the reproducibility of the FMD may differ between (clinical) groups.

Another important observation in our study was that previous experience of a laboratory with the FMD resulted in a smaller variation. A potential explanation for this finding is that experienced laboratories are more likely to demonstrate better adherence to the expert consensus guidelines. Indeed, when all factors were included in the final regression analysis (including adherence to the guidelines), previous experience of a laboratory with FMD did not emerge as an independent predictor of FMD reproducibility. Another factor that contributed to a smaller variation of the FMD was the use of a probe holder. The use of such devices is largely dependent on the personal preference of the laboratory and the effect on measurement reproducibility is a complex topic, since highly skilled operators with years of experience are able to conduct FMD measurements with exceptional reproducibility.

Interestingly, the time between repeated measurements did not significantly affect FMD reproducibility in our analyses. This might seem counter-intuitive, as poorer reproducibility is expected as the time between repeated measurements increases. Indeed, a recent study specifically designed to determine FMD reproducibility over short (48 h), medium (3 months) and long (9 months) time frames did find poorer reproducibility at 9 months between repeated measurements [15]. Reproducibility was comparable for the shorter time periods however, which is in agreement with a recent Italian multicenter study which found no differences in FMD reproducibility up to 30 days between measurements [16]. It should be noted that there was a large heterogeneity in time between measurements in the included study groups, with the majority ranging between one and 15 days (n = 32) and some up to 30 (n = 11), 90 (n = 4) and 270 days (n = 1). Excluding these last 16 studies from the analyses did not appreciably change our findings our findings of the diameter and the adoption of edge-detection software when performing valid and reproducible assessments of the FMD.

Another important observation in our study was that previous experience of a laboratory with the FMD resulted in a smaller variation. A potential explanation for this finding is that experienced laboratories are more likely to demonstrate better adherence to the expert consensus guidelines. Indeed, when all factors were included in the final regression analysis (including adherence to the guidelines), previous experience of a laboratory with FMD did not emerge as an independent predictor of FMD reproducibility. Another factor that contributed to a smaller variation of the FMD was the use of a probe holder. The use of such devices is largely dependent on the personal preference of the laboratory and the effect on measurement reproducibility is a complex topic, since highly skilled operators with years of experience are able to conduct FMD measurements with exceptional reproducibility.
however (data not shown).

4.1. Limitations

An obvious limitation of our systematic review is that the degree of adherence to expert consensus guidelines was assessed from information as provided in the papers. If a methodological description omitted one or more of the 19 different scoring factors, no points were assigned for those factors. As a consequence some studies with sparse methodological descriptions received lower scores. Poor methodological reporting might therefore have confounded our outcomes. It should also be acknowledged that our estimation of the experience of a laboratory with FMD measurements does not necessarily reflect the experience of an individual sonographer. However, a laboratory more experienced in performing FMD measurements will generally require a level of skill and training for their sonographers that will meet at least the standard of their previous work. This highlights the importance of the level of experience in performing studies with FMD as an outcome variable. Another limitation is that our analysis on the relative importance of independent subject- and/or methodological-related factors could only be based on a between-study comparison of factors contributing to the reproducibility of the FMD. Various other factors may have influenced this analysis. Therefore, future studies are necessary to further explore the importance of (some of) the methodological-related factors, including the effects of factors which we could not examine with the current dataset such as the observer/analyst, the time of cuff occlusion and changes in baseline brachial artery diameter.

In conclusion, this systematic review shows that adherence to current expert consensus guidelines significantly reduces measurement error when assessing brachial artery FMD in humans. Moreover, when adopting the guidelines, we found that the use of contemporary techniques (i.e. continuous diameter recording, edge-detection and wall-tracking software and under some circumstances the use of a probe holder) is crucial to improve reproducibility of the FMD measurement. Considering these factors will importantly decrease measurement error of the FMD and, consequently, decrease chances for type II errors in studies that rely on FMD as their primary outcome parameter. In other words, ignoring current expert-consensus guidelines causes significant variability of the FMD and, consequently, may lead to spurious conclusions. This study delivers important insight that should be taken into account when developing future updates to expert-consensus guidelines.

Sources of funding

Ms Anke van Mil is financially supported by a Top Institute for Food and Nutrition-grant.

Dr. Dick Thijsen is financially supported by the Netherlands Heart Foundation (E Dekker-stipend, 2009T064). Professor Green receives Fellowship and grant funding from the National Heart Foundation of Australia (APP1045204).

Disclosures

AG and PLZ are employed by Unilever R&D Vlaardingen B.V. No conflicts of interest, financial or otherwise, are declared by the remaining authors.

References
