identified a pooled AUROC of 0.91 for the MFC method.[10] Since the systematic review, more MFC studies have been published, all pointing to the same compelling conclusion: that the method is accurate in predicting fluid responsiveness.[11–15]

In 2018, we published a correspondence debating the way MFC studies were designed.[16] The correspondence raised clinical and statistical issues with the most adopted methodology. Yet, the notion that optimal MFC methodology may not be completely settled has hardly influenced methodology in subsequent publications. In this paper, we will:

- explain in simple terms the problems with the most frequently used MFC method
- demonstrate, by secondary analysis of an existing study and by simulations, the potential magnitude of the problem
- discuss strengths and limitations of less frequently used designs
- give recommendations on the way forward for researching this otherwise compelling method.

1.1 | A representative MFC study design

To simplify the key message, we will consider and discuss a representative MFC study design as depicted in Figure 1: 100 ml fluid is infused within 1 min (the MFC), the haemodynamic response (relative SV change) of that MFC is evaluated, and subsequently another 400 ml fluid (totalling 500 ml) is infused over 15 min. The final response (outcome) is evaluated as a relative SV change from baseline (i.e. before any fluid administration) to after the full amount of 500 ml. While we use SV in the examples, the arguments can be generalised to any method for estimating SV or cardiac output.

2 | METHOD

Figure 1 identifies that calculations of the haemodynamic response to the MFC (ΔSV_{100}) and the response to the *full* fluid challenge (ΔSV_{500}) both include the haemodynamic variable measured at *baseline*, that is before the MFC. Specifically, ΔSV_{100} and ΔSV_{500} are calculated as

$$\Delta SV_{100} = \frac{SV_{100} - SV_{baseline}}{SV_{baseline}}$$

and

$$\Delta SV_{500} = \frac{SV_{500} - SV_{baseline}}{SV_{baseline}}.$$

This shared baseline causes the problem.[1] It introduces two effects that, in addition to a true classification accuracy, can explain the high classification accuracy found in several MFC studies:

1. The predictor and the outcome share measurement error, creating a spurious correlation.

Editorial Comment

This review presents a detailed assessment of methodological aspects of studies assessing clinical effects of a form of intravascular fluid administration challenge. Findings are presented which demonstrate how many clinical reports in this area of inquiry can contain bias related to the choice of assessment variables, which must be considered when interpreting results. The authors suggest possible means to improve reliability for results related to methodological choices.

2. The predictor (ΔSV_{100}) is also a part of the outcome we try to predict (ΔSV_{500}).

2.1 | Shared error

Any measurement is associated with uncertainty (error). This can be subdivided into a systematic error (often referred to as bias) and a random error (often referred to as variance and defining *precision*).[17,18] It is useful to think of a 'true' SV and a random error around this value. The 'true' SV is what the clinician wants to measure, and what they hope to increase with a fluid infusion. The random error comprises both the imprecision of the monitoring equipment *and* minor temporal (minute-wise) physiologic changes in haemodynamics that are effectively noise in the context of evaluating a fluid response. It is the random error on the baseline measurement that causes the problem. In the following equations, each measured SV is divided into a 'true' SV and a random measurement error.

$$\Delta SV_{100} = \frac{(SV_{100} + \varepsilon_{SV,100}) - (SV_{baseline} + \varepsilon_{SV,\,baseline})}{SV_{baseline} + \varepsilon_{SV,\,baseline}} = \frac{SV_{100} + \varepsilon_{SV,100}}{SV_{baseline} + \varepsilon_{SV,\,baseline}} - 1,$$

and

$$\Delta SV_{500} = \frac{SV_{500} + \varepsilon_{SV,500}}{SV_{baseline} + \varepsilon_{SV,\,baseline}} - 1.$$

These two equations essentially depict the problem: the random error ($\varepsilon_{SV,\,baseline}$) is part of the denominator in the calculation of both the predictor (ΔSV_{100}) and the outcome (ΔSV_{500}), making them spuriously correlated, and therefore more likely to agree.[19]

2.2 | The predictor (ΔSV_{100}) is also a part of the outcome we try to predict (ΔSV_{500})

The MFC should be used as a *predictive* method, that is to decide whether to administer the remaining 400 ml fluid or not. Thus,