Pre read 2. Validation of Cuffless Blood Pressure (BP) Devices

Background information and proposal for the BIHS/IQVIA BP Tech Summit

Background

Validation of cuff-based BP devices is well established through standardised protocols that compare device readings with reference values obtained by auscultation, using a pre-specified number of participants across a range of different BPs.

By contrast, validation of the rapidly emerging **cuffless technologies** presents greater challenges. Most wearable devices use **optical photoplethysmography (PPG)** to derive pulse waveforms based on light absorption by red blood cells passing through skin capillaries. **Machine learning algorithms** then estimate BP from waveform features.

However, because the PPG signal is only *indirectly* related to BP, results can be influenced by factors such as **skin colour** and **haemodynamic variations** that are independent of BP. These may arise from physiological or pathological states including pregnancy, ageing and arterial stiffening, circadian variations, ambient temperature, antihypertensive therapy, and cardiac conditions such as valve disease or heart failure.

The **European Society of Hypertension (ESH)** has proposed detailed recommendations for validation of medically approved cuffless BP monitors. These protocols are, however, complex and have not yet been fully implemented for any cuffless device. Many manufacturers consider the ESH requirements impractical and more stringent than those applied to cuff-based devices, calling instead for a **pragmatic and scalable approach**. Notably, current frameworks often omit key population characteristics - such as age, skin colour, and medication use - that must be represented to infer accuracy across the general population.

General Principles of Validation

Accuracy and Sample Size

The required level of accuracy for any BP monitor (cuff-based or cuffless) should be guided by a **risk:benefit analysis**. The most recent **ISO Universal Standard** provides a starting benchmark, with a validation sample size of $n = 85$, determined by the probability that devices of differing accuracy would pass or fail the protocol.

Cuffless devices introduce additional complexity: the influence of demographic and clinical or environmental factors (e.g., age, skin tone, temperature, antihypertensive drug use) on BP estimation is not yet well quantified. Until these effects are characterised, a **conservative approach** would require larger sample sizes and more diverse recruits and environmental conditions than those in existing standards.

Other sources of variation of cuffless devices may include:

- Variations in "fit" e.g. of tightness of band and ring fit and the interactin of the sensor with local skin and tissue characteristics.
- Ambient temperature variations altering blood flow- especially in fingers and in subjects with Raynaud's.
- Problems with cuff-related validation related to excess arm size where distal PPG may be much more acceptable

Intended Use and Validation Approach

Validation must reflect the **intended clinical role** of the device:

1. **Wellness Devices** – These may indicate the *risk* of hypertension rather than provide a definitive measurement. Validation typically relies on machine learning performance within manufacturer-supplied datasets, assessed via **sensitivity, specificity**, and associated confidence intervals for hypertension detection. Independent de novo validation against reference BP measurements in large cohorts may offer limited added value beyond critical appraisal of these datasets.
2. **Medically Approved Devices** – These may provide:
    - **Snapshot readings** akin to conventional cuff-based monitors;
    - **24-hour measurements** substituting for ambulatory BP monitoring (ABPM);
    - **Continuous tracking** over days or weeks to assess BP variability or treatment response.

Validation by Application

1. Snapshot Devices

In principle, existing validation protocols for cuff-based devices can be adapted, provided samples adequately represent:

- Both sexes
- A range of skin tones
- Diverse age groups
- Individuals on and off antihypertensive therapy

If intended for specific populations (e.g. pregnancy, valve disease), separate validation in those groups is required.

Furthermore, validation needs to be shown to be robust across the range of environmental conditions (ambient temperature, lighting etc) that will occur in practice.

The large sample size required to include meaningful strata with differing characteristics and with measurements under varying environmental conditions means that a pragmatic validation process might require 2 stages: first to demonstrate in a small sample size (approx. 40 people)

that agreement with a reference is sufficiently high to justify an "in service evaluation" where the test device is paired with standard care to obtain comparative data in a large sample size (approx. 400).

2. 24-hour Monitoring

This represents a key potential use case for wearable technology, with the promise of replacing conventional cuff-based ABPM - currently burdensome for patients, resource-intensive for health systems, and a barrier to timely diagnosis.

Wrist and ring-based devices are typically **calibrated** using oscillometric BP readings, after which they provide continuous BP estimates. Validation through direct comparison with manual auscultation is not feasible for such prolonged measurements. The **BIHS** is therefore developing a **comparative protocol** using standard 24-hour cuff-based monitors as references, ensuring demographic and clinical representativeness as above and with a two-stage approach.

3. Continuous Tracking and Treatment Response

Evaluating long-term tracking accuracy requires assessing both **trend fidelity** (accuracy in detecting BP change) and **calibration stability**. A practical approach may involve comparison of device-estimated BP before and after antihypertensive treatment.

Claims Ladder

For cuffless wearables, **validation must be proportionate to the claim**. We will use a simple **claims ladder** to align evidence with intended purpose:

- **Wellness (no BP claim):** general wellbeing; no mmHg values; evidence = UX/comprehension and no medical inference.
- **Risk alert (possible hypertension):** device may flag risk; requires 7-day cuff or ABPM **confirmation**; evidence = sensitivity/specificity/positive predictive value (PPV)/negative predictive value (NPV) for hypertension detection, alert burden, and completion of confirmation.
- **Measurement (BP, heart rate ± AF alert):** shows systolic/diastolic values for decisions; evidence = accuracy vs validated cuff (bias/SD, subgroup analysis (skin tone/age/BMI/therapy/AF), usability, and drift checks.
- **24-hour/continuous (ABPM-substitute):** day–night means, dipping, load; evidence = agreement vs 24-h cuff ABPM, missing-data/artifact handling, motion robustness, calibration stability.

(If a product automates treatment or triage, it sits above these tiers and needs prospective safety/clinical-effect evidence.) This ladder anchors **what can be said**, the **minimum evidence** to say it, and the plain-English limits that must accompany each tier.

This means "Say only what you can prove."

- ✓ **Wellness**: feel better, stay active.
- ✓ **Risk alert**: "You might have high BP — please check with a cuff for 7 days."
- ✓ **Measurement**: shows BP numbers (mmHg) and has passed accuracy tests.
- ✓ **24-hour**: tracks patterns like day vs night; tested against standard monitors

Evolution of Validation Protocols

As understanding of cuffless BP technology and its influencing factors advances, validation methods can be progressively refined. Initial studies will likely require larger and more diverse samples, but sample sizes may later be reduced if specific factors (e.g. age, temperature or skin tone) prove to have negligible effects on accuracy.

A flexible, iterative validation framework is therefore essential. Early protocols will require significant investment, and sustainable funding mechanisms for these validation studies must be explored.