

Note technique : validation des outils de monitoring du débit cardiaque

Technical note: validation of cardiac output monitoring tools

P. Squara · J.-D. Chiche

Reçu le 28 novembre 2010 ; accepté le 15 janvier 2011
© SRLF et Springer-Verlag France 2011

Résumé Jusqu'à récemment, les systèmes de monitoring continu du débit cardiaque (DC) étaient validés comme l'avaient été par le passé les outils de mesure discontinue, la plupart des auteurs se limitant à comparer deux technologies par la méthode de Bland & Altman. Par cette méthode, l'information se résume à une estimation minorée du biais moyen et à sa variabilité interpatients. Mais l'importance du temps et de la répétition des mesures n'est pas mise en lumière. Cette note technique propose un cadre de validation des outils de monitoring continu. Sept critères primaires de qualité sont retenus et étudiés : 1) l'exactitude (faible biais) ; 2) la linéarité (biais constant) ; 3) la précision (faible variabilité due aux erreurs de mesure liées au hasard) ; 4) la résolution (valeur de la plus petite modification significative) ; 5) le temps et l'amplitude de réponse ; 6) la stabilité (les critères précédents restent invariants dans le temps) ; 7) la plage de mesure (les critères précédents restent acceptables dans une plage de DC donnée). Une tolérance est obligatoirement définie pour chacun de ces critères. Cela rend nécessaire l'introduction d'un huitième critère : la capacité à détecter des changements significatifs de DC. D'autres conditions importantes à la validation des outils de monitoring continu du DC sont étudiées ici : choix de la population de patients, choix de la méthode de référence, acquisition, vraisemblance et segmentation des données. L'application de ce cadre de référence à la validation des outils de monitoring continu du DC conduit à souligner l'importance de la précision et du temps de réponse dans l'acceptabilité clinique de ces outils. *Pour citer cette revue : Réanimation 20 (2011).*

Mots clés Débit cardiaque · Monitoring · Biais · Précision · Variabilité · Répétabilité · Résolution

Abstract To date, tools for continuous cardiac output (CO) monitoring have been validated as if they were tools for snapshot measurements. Most authors compared variations in cardiac output between two time-points and used Bland & Altman representations to describe the agreement between these variations. The impact of repetitive measurements over time is not taken into consideration. This article proposes a conceptual framework for the validation of CO monitoring devices. The proposal consists of eight quality criteria: 1) accuracy (small bias); 2) linearity by comparing the regression line of the study and reference technology with the identity line; 3) precision (small random error of measurements); 4) least significant change; 5) stability (absence of time drift for the first four criteria); 6) measurement range where the first five criteria are fulfilled; 7) short response time; and 8) accurate amplitude response. Because tolerance is obviously admitted for each of these eight criteria, we propose to add as final criterion, the ability to detect significant CO directional changes. Other important issues in designing studies to validate CO monitoring tools are reviewed: choice of patient population to be studied, choice of the reference method, data acquisition method, data acceptability checking, data segmentation and final evaluation of reliability. Application of this framework will underline the importance of precision and time response for clinical acceptability of monitoring tools. *To cite this journal: Réanimation 20 (2011).*

P. Squara (✉)
Service de réanimation CERIC,
clinique Ambroise-Paré, 27, boulevard Victor-Hugo,
F-92200 Neuilly, France
e-mail : pierre.squara@wanadoo.fr

J.-D. Chiche
Service de réanimation médicale,
centre hospitalo-universitaire Cochin-Port-Royal,
23, rue du Faubourg-Saint-Jacques,
F-75013 Paris, France

Keywords Cardiac output · Monitoring · Bias · Precision · Variability · Repeatability · Least significant change

Introduction

L'analyse du débit cardiaque (DC) est un des paradigmes de la réanimation depuis ses débuts dans les années 1960, que

ce soit pour le diagnostic, la surveillance ou le traitement des défaillances circulatoires. Les premiers moyens de mesure disponibles en pratique clinique furent la méthode de Fick sur les gaz expirés, la thermodilution par cathétérisme du ventricule droit et la vélocimétrie Doppler. Afin de permettre une utilisation plus large et plus sûre du DC, des outils à la fois moins invasifs et continus ont été développés au cours des 20 dernières années. Cette avancée technologique a fait naître un besoin de validation clinique pour une gamme nouvelle d'appareils.

Bien que les informations issues d'un appareil de monitoring et d'un appareil de mesure soient clairement différentes, la mise au point de critères de validation spécifiques aux outils de monitoring est récente [1,2]. Il est notamment nécessaire de prendre en compte les effets du temps et la répétition des mesures. Cette note a pour but de faire le point sur les exigences de la validation d'un appareil de monitoring du DC et de proposer des solutions. Cette méthode pourra être souvent transposable au monitoring d'autres grandeurs.

Critères de qualité d'un système de monitoring du DC

Un transducteur est un instrument qui transforme une grandeur physique en une autre. La valeur affichée par un appareil de mesure du DC est toujours issue d'un signal électrique obtenu par transduction d'une composante directe ou indirecte de l'énergie hydraulique du cœur : température (thermodilution), pression artérielle (pulse contour), vitesse (Doppler), variation du volume de l'aorte (impédance, réactance), etc. La qualité de mesure du DC s'analyse donc comme celle d'une transduction. Il importe d'en définir les principaux termes, car leur acception peut varier d'un auteur à l'autre (le terme anglais est indiqué entre parenthèses).

- La justesse ou l'exactitude (*accuracy*) est la capacité à mesurer la grandeur sans erreur ou biais (*bias*) ;
- la linéarité (*linearity*) est la capacité à maintenir constante la sensibilité (rapport entre le signal physiologique d'entrée et le signal électrique de sortie). Le biais est donc constant ;
- la précision (*precision*) est la capacité à indiquer une même valeur lorsque la grandeur mesurée ne varie pas ;
- la résolution (*resolution*) est la plus petite modification de la grandeur mesurée détectable par l'appareil de mesure ;
- la stabilité (*stability*) est la capacité à conserver les propriétés précédentes sans dérive (drift) dans le temps ;
- la plage de mesure (*measurement range*) est l'enveloppe de valeurs extrêmes à l'intérieur de laquelle les critères précédents sont admis comme acceptables.

La validation d'un appareil de monitoring reprend les critères de qualité d'une mesure auxquels s'ajoute nécessairement une analyse des variations en fonction du temps, donc la capacité à indiquer des changements rapides. On définit ainsi un critère supplémentaire :

- la réactivité (*responsiveness*) est le délai (ou temps) de réponse entre la variation de la grandeur mesurée et la variation de la valeur affichée par l'appareil de monitoring. Couplé à l'exactitude, le temps de réponse détermine l'amplitude de réponse.

Lors d'une étude de validation, les critères ci-dessus sont analysés et des imperfections plus ou moins grandes sont observées. En fonction de l'utilisation clinique attendue, il faut déterminer si ces imperfections sont acceptables ou non. Une tolérance doit donc être définie a priori pour chaque critère. Même si, individuellement, chaque critère de qualité respecte la tolérance qui lui est assignée, on peut faire l'hypothèse que le cumul de ces tolérances peut aboutir à des erreurs inacceptables en pratique. Il est donc nécessaire d'imaginer un dernier critère, global, d'acceptabilité. Nous suggérons d'identifier ce critère comme :

- la capacité à identifier les changements significatifs de la grandeur mesurée. Cette dernière qualité s'évalue statistiquement en termes de valeur prédictive par rapport à une référence.

L'exactitude, la plage de mesure, la résolution, l'amplitude de réponse s'expriment en litre par minute. La précision est d'autant plus grande que la variabilité due aux erreurs de mesure est faible (Fig. 1). Elle s'évalue par l'écart-type (ET) de plusieurs mesures d'une même valeur et s'exprime en litre par minute ou souvent en pourcentage par le coefficient de variation en faisant le rapport entre l'ET et la valeur moyenne ($CV = ET/moy$). Lorsqu'on souhaite que la précision exprime un intervalle de confiance à 95 %, on utilise $2ET$ et $2ET/moyenne$ ($2ET/moy$).

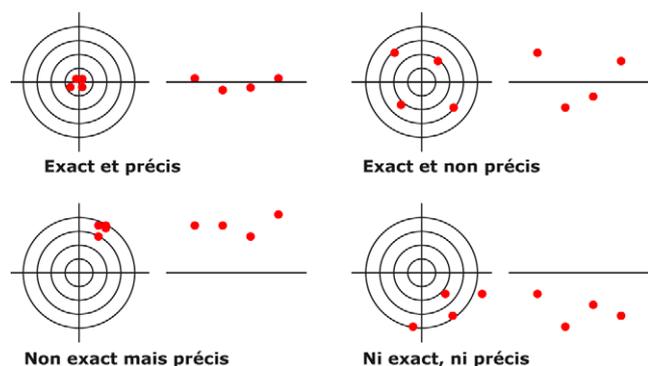


Fig. 1 La technologie étudiée est exacte lorsque la moyenne des mesures est proche du cœur de cible. La technologie est précise quand toutes les mesures sont proches.

Principes de validation d'un système de monitoring du DC

Les étapes suivantes doivent être envisagées pour évaluer un outil de monitoring du DC :

- choix des patients à étudier ;
- choix de la méthode de référence ;
- méthode d'acquisition des données ;
- vraisemblance des données ;
- segmentation des données ;
- analyse des critères de qualité.

Choix de la population de patients à étudier

Une étude de validation doit pouvoir permettre la généralisation raisonnable des données observées. Un système de monitoring du DC doit donc faire ses preuves dans des conditions représentatives des désordres circulatoires habituellement rencontrés, c'est-à-dire au minimum chez des patients septiques et non septiques, en ventilation mécanique ou spontanée, traités par inotropes, vasodilatateurs et vasopresseurs. De plus, les populations étudiées doivent avoir une forte probabilité de variation du DC pendant l'étude. Cela peut être assuré en prévoyant, dans le protocole d'étude, des interventions faisant varier le DC de manière très significative. De telles variations du DC peuvent être obtenues par des tests de remplissage, des variations de pression expiratoire positive (PEP) ou l'utilisation de substances.

Choix de la méthode de référence

Idéalement, la validation d'une technologie nécessiterait une validation indépendante de chaque critère de qualité listé auparavant avec une méthode de référence spécifique. Toutefois, en dehors d'études animales, les conditions éthiques et pratiques des études cliniques nous limitent souvent à tester une technologie nouvelle (TN) par rapport à une seule technologie de référence (TR) ; celle qui présente le meilleur compromis pour l'ensemble des critères de qualité.

Une valeur étalon (gold standard) du DC est nécessaire pour valider six des huit critères de qualité : l'exactitude, la réactivité, la linéarité, la stabilité, la plage de mesure, et la capacité d'identification des changements significatifs du DC. La précision et la résolution n'en ont pas besoin. Le principe de Fick énoncé dès 1870 [3], est toujours considéré comme la meilleure mesure du DC au laboratoire de physiologie [4]. En pratique clinique, en dehors de situations très particulières de chirurgie cardiaque permettant l'utilisation de la circulation extracorporelle ou de bague Doppler interne [5,6], un étalon du DC manque le plus souvent. On se limite

alors à l'utilisation d'une méthode largement utilisée dont l'exactitude, bien qu'imparfaite, est considérée comme acceptable. En pratique, la dilution d'un indicateur coloré ou thermique injecté par bolus, a été largement utilisée comme TR [7–11]. Cependant, cette technique requiert des interventions manuelles aboutissant à une variabilité intra- et interobservateurs [9].

Quelle que soit la TR choisie, il est important de minimiser l'incertitude sur la valeur affichée. L'erreur standard (SE) dépend de la précision de la technologie choisie et du nombre de mesures élémentaires intégrées (n) dans la moyenne retenue pour effectuer la comparaison ($SE = ET/\sqrt{n}$). Pour un intervalle de confiance à 95 %, $SE = 2ET/\sqrt{n}$ [12]. Ainsi, si l'on admet, par exemple, que la précision de la thermodilution par bolus ($2ET/moy$) est de l'ordre de 10 %, il faudra faire la moyenne de cinq thermodilutions si on veut obtenir un intervalle de confiance inférieur à ± 5 % ($10\%/\sqrt{5} = 4,5\%$).

En conséquence du nombre de mesures moyennées pour obtenir la valeur de référence, il faudra comparer la valeur de la TN recueillie pendant la même période. Dans l'exemple précédent, la moyenne de cinq thermodilutions va prendre environ cinq minutes de manipulation. La valeur observée avec la TN devra également être moyennée sur la même période de cinq minutes.

Puisqu'aucun système de monitoring du DC n'a satisfait idéalement à tous les critères de qualité requis, il n'y a pas de TR consensuelle. S'agissant de monitoring, il est logique de tester la TN avec un système comparable pour juger d'un progrès éventuel, c'est-à-dire un outil automatique, donnant des valeurs en temps réel, dont les critères de qualité bien imparfaits sont connus. Dans cette optique, les versions du cathéter de Swan Ganz (PAC-CCO) permettent la mesure intermittente du DC par thermodilution bolus couplée à un monitoring continu avec deux périodes d'échantillonnages possibles : la première rafraîchit toutes les minutes (STAT) avec un temps de réponse assez bref (quelques minutes) mais une précision médiocre, l'autre avec un moyennage sur plus de cinq minutes résultant en un temps de réponse lent (10–15 min) et une précision plus grande [13–15]. Cette technologie est loin d'être parfaite mais reste à ce jour la plus utilisée [1,16–18] dans la mesure où le biais moyen a été trouvé assez faible pour être acceptable dans une grande variété de situations cliniques et sur une large plage de mesure [19–25]. De plus en plus d'études laissent présager que les solutions moins invasives qui font toutes appel à des technologies intrinsèquement plus précises et plus réactives atteindront bientôt un niveau d'exactitude suffisant pour satisfaire à tous les critères de qualité recommandés ici [12,26–30]. Toutefois, aucune d'entre elles n'a aujourd'hui fait l'objet d'une validation complète permettant une délimitation claire de son espace de fiabilité et de ses éventuelles restrictions. Il est en effet de toute première importance de

connaître les limites de chaque système que ce soit pour un usage clinique ou pour une utilisation éventuelle comme TR.

Acquisition des données

L'idéal est d'obtenir un enregistrement continu parallèle des données venant de la TR et de la TN étudiée pour éviter les erreurs inévitables lors du recueil d'une grande quantité de données. L'acquisition automatique permet aussi d'améliorer la répétabilité (faible variabilité intraobservateur) et la reproductibilité (faible variabilité interobservateurs) et de mieux cerner la précision inhérente à chaque technologie puisqu'il s'agit de la variabilité due seulement aux erreurs de mesure liées au hasard. L'échantillonnage peut poser problème, car les deux technologies peuvent afficher les données avec des cadences différentes. Chaque fabricant applique un traitement plus ou moins complexe au signal original pour aboutir à la valeur finale affichée. Ce traitement du signal détermine le compromis entre les deux exigences opposées que sont le temps de réponse et la précision. Plus l'échantillonnage est rapide, plus le temps de réponse est bref et la variabilité due aux erreurs de mesure grande, car la moyenne porte sur une durée moindre de signal, et vice versa [31]. Le compromis final dépend des propriétés intrinsèques de chaque technologie et du choix du fabricant. Idéalement, lorsqu'on compare deux technologies, le plus petit temps d'échantillonnage commun doit être appliqué pour comparer les compromis réponse/précision dans les mêmes conditions et pour une utilisation clinique la plus large possible. En effet, si la surveillance globale d'un patient stabilisé peut se satisfaire d'une valeur de DC toute les cinq minutes, le résultat d'un test de lever de jambe requiert un échantillonnage minute par minute et l'étude de la variation respiratoire du volume systolique nécessite des données battement par battement.

Vraisemblance des données

Avant d'analyser la base de données provenant des deux technologies, il est important de procéder à une étape de validation afin d'éviter que la comparaison ne soit biaisée par des artefacts, des aberrations dues à des problèmes techniques (agitation, déconnexion) ou à des situations où l'une des deux technologies est hors limites pour des raisons clairement connues, comme par exemple une régurgitation tricuspide pour la thermodilution. À défaut, les résultats de la validation ne dépendent plus des caractéristiques des deux technologies mais de la prévalence de cette situation particulière dans la période de l'étude.

Cette étape est particulièrement critique, car elle détermine la crédibilité des résultats. Il est nécessaire qu'elle soit assurée par un intermédiaire indépendant, si possible à l'aveugle, afin de garantir qu'une situation défavorable à l'une ou l'autre des

technologies ne sera pas soustraite indûment. Les périodes soustraites doivent être listées et justifiées. Elles peuvent faire l'objet d'une analyse a posteriori.

Segmentation des données

Lorsqu'on dispose de courbes de tendances comportant un nombre suffisant de valeurs simultanées provenant de la TR et de la TN étudiée, une corrélation croisée permet de voir si la forme des deux tendances est comparable ou non. De plus, il est important d'étudier les deux technologies lorsque la tendance est soit inchangée, soit ascendante, ou descendante. En effet, un désaccord unidirectionnel peut advenir (défaut de linéarité).

L'étude de la précision ne peut se faire que pendant des périodes de stabilité du DC. À défaut, on augmente artificiellement la variabilité due aux erreurs de mesure en ajoutant la variabilité naturelle du DC.

Par ailleurs, une différence de temps de réponse entre la TN étudiée et la TR entraîne des différences de valeur instantanées qui seront interprétées à tort comme un biais de la TN étudiée. Ainsi, il est donc nécessaire de segmenter la base de données. On peut utiliser facilement la pente de la tendance de la TR à cet effet. L'identification d'un point d'inflexion entre deux pentes peut être obtenue automatiquement en déterminant la valeur minimale de la somme des

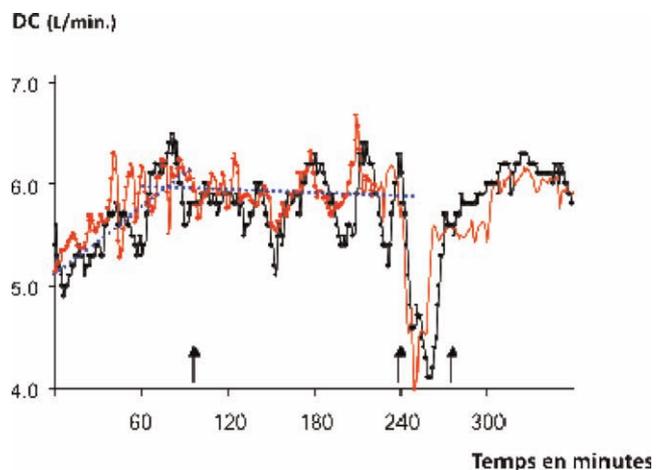


Fig. 2 Segmentation de la courbe de tendance du débit cardiaque (DC) en différentes phases. En rouge une technologie nouvelle (TN), en noir la technologie de référence (TR). De la minute 1 à 90 (flèche), la pente de TR (pointillé noir) est ascendante (+25 %). De la minute 90 à 235, la pente est plate (-2 %). Entre la minute 235 et 250, un test de recrutement alvéolaire par une PEP est réalisé, aboutissant à une chute brutale du DC, puis remontée rapide à l'arrêt de la PEP. De la minute 275 à 360, la pente est ascendante (+20 %). Après la minute 235, les pentes ne sont plus représentées pour plus de visibilité. Dans l'étude d'où cette figure est tirée, chaque segment était analysé séparément [1].

résiduelles proposée par John-Alder et Bennet [32]. La Figure 2 montre un exemple d'étude de la tendance du DC d'une TN simultanément comparée à la thermodilution continue prise comme TR. Dans cet exemple, la tendance et les ruptures de pentes étaient assez bien superposables, bien que les variations de la TN apparaissent plus précocement qu'avec la thermodilution.

Analyse des critères de qualité

Chaque critère vu auparavant est détaillé ici avec sa tolérance habituelle qui pourra varier en fonction de l'usage attendu.

Exactitude

Le biais entre la TN et la TR est la différence moyenne observée pour un patient sur une période de temps donnée. Si la TR ne peut pas être pas considérée comme exacte, le biais ne correspond pas à une erreur de la TN mais simplement à une différence entre les deux méthodes. Dans ce cas particulier, et dans ce cas seulement, Bland et al. ont proposé une représentation schématique qui permet de visualiser rapidement le biais moyen et sa variabilité par l'intervalle de confiance encadrant le biais moyen. Quand le biais est exprimé en litre par minute, la variabilité correspond aux limites d'agrément ($\pm 2\text{ET}$) [33]. Quand le biais est exprimé en pourcentage en faisant le rapport entre le biais et la moyenne des deux technologies, la variabilité correspond à un coefficient de variation ($\pm 2\text{ET}/\text{moy}$) [34].

En revanche, si la TR est considérée comme exacte ou acceptablement exacte, le biais observé entre TR et TN correspond à une erreur de la TN. Il est donc préférable de calculer l'erreur absolue (différence entre les deux valeurs) ou l'erreur relative (valeur absolue de la différence entre les deux technologies/valeur de la TR) pour chaque patient puis de décrire la fréquence de distribution de ces erreurs sur la population de patient étudiée [1]. Cette solution est donc plus exigeante pour la TN étudiée, car elle prend comme référence la valeur de la TR et non la moyenne des deux technologies.

Dans la mesure ou une différence de temps de réponse ou d'amplitude de réponse peut affecter le biais instantané, le biais moyen ne sera le reflet de l'exactitude que si l'impact du temps et de l'amplitude de réponse peut être minimisé. C'est pourquoi il est nécessaire de restreindre l'analyse du biais à des périodes de stabilité maximale du DC. Par exemple, dans l'étude originale de validation de la bioréactance, l'exactitude était analysée sur des périodes de stabilité du DC définies par une pente de la TR (PAC-CCO) inférieure à 10 % avec une variabilité du DC ($2\text{ET}/\text{moy}$) inférieure à 20 %. En procédant ainsi, l'SE de la TR peut être rendue inférieure à 2 % avec 25 mesures moyennées ($\text{SE} = 10 \%/ \sqrt{25}$).

Malgré les limites de la TR utilisée, une telle période de stabilité peut donc servir de référence stable, juste, appropriée pour l'évaluation de l'exactitude de la TN. Le biais peut être calculé d'abord pour l'échantillonnage minimum (par exemple, une minute) puis pour des échantillonnages plus long permettant d'estimer le biais pour différents usages cliniques.

Tolérance : l'erreur observée avec la thermodilution continue varie entre 5 et 15 % suivant les études [7–9]. Lorsqu'on considère la valeur de la TR comme inexacte, un coefficient de variation interpatients ($2\text{ET}/\text{moy}$) inférieur à 30 % a pu être jugé acceptable [34]. Lorsqu'on considère la TR comme exacte, des tolérances plus restrictives ont été proposées (erreur relative < 20%) [1]. Tout dépend de l'utilisation clinique qu'on veut faire. De telles tolérances sont souvent considérées comme acceptables, car si l'on utilise la valeur du DC comme partie d'un équilibre multifactoriel propre à chaque patient, une erreur sur un des facteurs a des conséquences souvent limitées [35,36]. En revanche, si le clinicien vise une valeur prédéfinie du DC plutôt qu'un équilibre multifactoriel, le défaut d'exactitude du DC pourra avoir plus de conséquences.

Linéarité

En pratique clinique, la linéarité de la TN est étudiée en comparant la droite de régression de TR et TN avec la droite d'identité. Elle permet de vérifier notamment l'absence de surestimation des bas DC et de sous-estimation des hauts DC [1].

Tolérance : la linéarité est acceptable lorsque la différence statistique entre la régression TN-TR est non significativement différente de la droite d'identité.

Précision

La précision est grande quand la variabilité due aux erreurs de mesure est faible. On a vu qu'il s'agit là d'une qualité intrinsèque du système, qui peut coexister avec une exactitude bonne ou faible. Elle ne nécessite pas de TR pour être étudiée. La représentation de Bland-Altman ne permet pas d'estimer la précision telle que nous l'avons définie, que ce soit par les limites d'agrément du biais (2ET interpatients), ou son coefficient de répétabilité du biais (2ET intrapartients). Ce sont là des moyens d'analyser la variabilité du biais qui ne représente pas la précision (Fig. 3). La précision de chaque TN peut être estimée, pour chaque patient, par la variabilité du DC autour de sa propre ligne de tendance. C'est donc l'équivalent d'un bruit électronique autour d'une valeur centrale. On peut ensuite faire une moyenne interpatients, une étude de sa distribution ou toute autre statistique descriptive. Comme pour le biais, la précision est évaluée au mieux lorsque les autres formes de variabilité

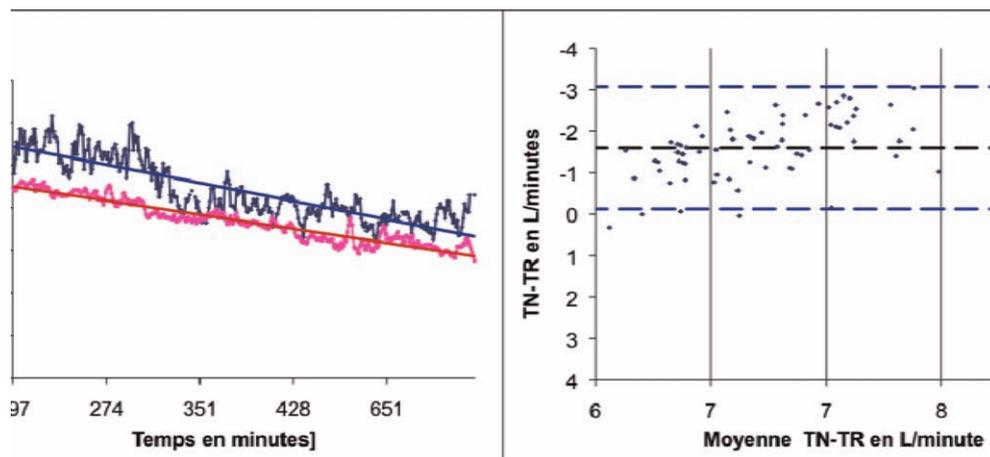


Fig. 3 À gauche, les tendances d'un même débit cardiaque (DC) enregistrées avec deux technologies différentes en rouge la technologie nouvelle (TN), en bleu la technologie de référence (TR). À droite, la représentation de Bland-Altman correspondante. Puisque les pentes à gauche sont presque parallèles (TN = -9 %, TR = -11 %) le biais peut être vu comme la distance entre les deux lignes de pentes sur l'axe Y ou comme la différence moyenne observée sur la représentation de Bland-Altman à droite (-1,50 l/min). On voit à gauche que la précision est plus grande (variabilité autour de la pente plus faible) pour la TN (7 %) que pour la TR (27 %). Ces pentes et la large variabilité de la TR induisent une très large variabilité du biais instantané et donc du pourcentage d'erreur donné par la représentation de Bland-Altman à droite, $\pm 2,50$ l ou ± 24 %. Cela illustre que la variabilité du biais telle que le montre le Bland-Altman d'une part, et variabilité due aux erreurs de mesure liées au hasard (précision) d'autre part, sont deux dimensions différentes de l'évaluation d'un système de monitoring du DC.

sont minimisées et donc pendant les phases de stabilité du DC. Lorsque la pente de la tendance est plate, la variabilité autour de la ligne de tendance est alors équivalente au coefficient de variation ($2ET/moy$) puisque la pente nulle représente la valeur moyenne.

Tolérance : la précision détermine la capacité d'un appareil à identifier les changements de DC. La précision doit donc être a priori déterminée en fonction de l'usage clinique attendu et du compromis entre précision, résolution et capacité de réponse comme on le verra aussi plus loin. Pour information, la précision in vitro de la thermodilution continue par PAC-CCO ($2ET/moy$) est estimée entre 18,4 et 23,2 % avec différents appareils [9]. En pratique clinique, la précision habituellement observée est faible avec une variabilité allant jusqu'à 40 % [13,36]. D'autres technologies de monitoring du DC ont montré des précisions entre 5 et 10 % [37].

Résolution

C'est, comme on l'a vu, la plus petite variation de la grandeur mesurée détectable par l'appareil de mesure. Mais en ce qui concerne le monitoring du DC, cette plus petite variation, même si elle est enregistrée par l'appareil de mesure, est noyée dans le bruit de fond que constitue la variabilité due aux erreurs de mesure, la précision donc. Ce qui intéresse le clinicien est donc de savoir quelle est la plus petite variation de la mesure observée avec la TN qui corresponde à une variation réelle du DC. On peut montrer que le plus petit seuil de

significativité (LSC, pour *least significant change*) avec un intervalle de confiance à 95 % est donné par la formule $LSC = 2 \sqrt{2} SE$, où $SE = ET/\sqrt{n}$. On voit donc qu'un outil de monitoring peut compenser une précision faible par l'importance du nombre de mesure.

Tolérance : ce critère de qualité n'a pas fait jusqu'à présent l'objet d'une analyse de tolérance. Il a été jusqu'ici peu utilisé dans la littérature en réanimation [12,38]. Sur la base de la tolérance admise pour l'exactitude, on peut difficilement accepter d'utiliser un outil de monitoring du DC dont la LSC soit supérieure à 1 l/min ou à 20 % de sa valeur de base.

Stabilité

Les quatre critères listés ci-dessus, ainsi que la réactivité étudiée ci-après peuvent faire l'objet d'une dérive dans le temps. Une étude de validation devra donc être suffisamment prolongée et répétée pour mettre en évidence un tel phénomène. Habituellement, l'existence d'une dérive requiert le changement ou le repositionnement des dispositifs de recueil du signal ou une recalibration du système.

Tolérance : ici encore, tout dépend de l'usage clinique. Pour un usage en anesthésie, une stabilité de quelques heures peut suffire. En réanimation, dans un but de surveillance, une stabilité de 12-24 heures est souhaitable pour ne pas alourdir excessivement la charge de travail.

Plage de mesure

L'étude de la plage de mesure repose sur la population de patients étudiés. Il faut que le DC observé avec la TR couvre une plage la plus large possible. On pourra vérifier, par les moyens que nous venons d'exposer, pour quelle plage de mesure les critères étudiés sont à l'intérieur de la tolérance définie.

Tolérance : la plage de mesure habituelle du DC en réanimation est $1-12 \pm 1$ l/min. C'est à l'intérieur de cette plage que les tolérances définies pour les autres critères sont requises. Au-delà, les tolérances pourront être élargies sans grande conséquence clinique.

Réactivité

La réactivité de la TN peut être comparée à celle de la TR au moment où des challenges hémodynamiques sont réalisés. Ainsi, une baisse brutale du DC est attendue en cas de réalisation d'un recrutement alvéolaire par application d'un haut niveau de PEP. De même, le retour à une PEP nulle induit une remontée rapide du DC. Une augmentation du DC peut aussi être obtenue par un test de lever de jambe, un remplissage vasculaire ou l'utilisation de substances inotropes. Pour mesurer précisément le délai et l'amplitude de réponse, il faut déterminer les points d'inflexion de départ et de fin sur la courbe de tendance des deux technologies (TN et TR). Le temps de réponse est le délai entre les deux points d'inflexion et l'amplitude est la différence de DC entre ces deux mêmes points (Fig. 4). Les points d'inflexion peuvent

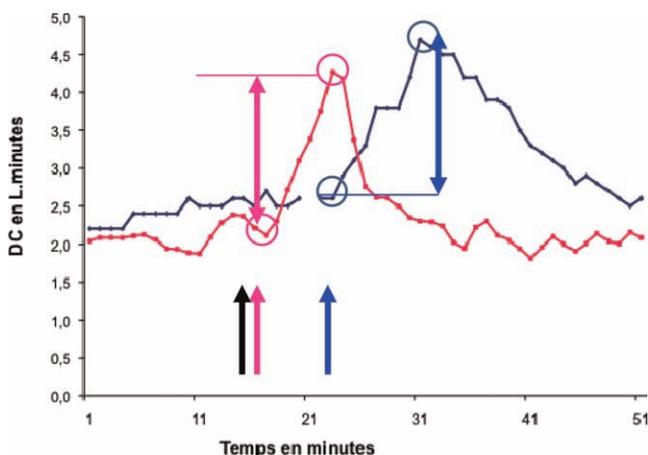


Fig. 4 Courbes de tendance du débit cardiaque (DC) avec deux technologies de son monitoring. La flèche noire indique de début d'un challenge hémodynamique. Les cercles de couleur indiquent les points d'inflexion des courbes correspondantes. Les flèches simples indiquent le délai de réponse par report sur l'axe des X et les flèches doubles l'amplitude de réponse des courbes correspondantes par report sur l'axe des Y.

être déterminés automatiquement par la méthode décrite par John-Alder [32]. Toutefois, l'amplitude de la réponse observée après un challenge hémodynamique dépend aussi de la précision de chaque technologie. Une estimation indépendante de l'amplitude de réponse peut être obtenue en comparant les deux pentes de variation du DC lorsque le DC change de manière régulière pendant une période de temps suffisante. Dans ce cas, comme pour l'étude de l'exactitude, un nombre approprié de valeurs permettent de considérer la pente de la TR comme la pente réelle. Une autre possibilité est de comparer la TN avec la TR après stabilisation du DC à deux moments différents et de comparer la différence de variation du DC moyen entre ces deux points dans le temps.

Tolérance : la tolérance de l'amplitude de réponse doit logiquement être équivalente à celle de l'exactitude. Toutefois, cette amplitude doit être obtenue dans un délai acceptable qui détermine la tolérance du temps de réponse. Le meilleur compromis dépend des caractéristiques intrinsèques du signal, la plupart du temps inaccessibles au clinicien. La thermodilution continue par PAC-CCO à un temps de réponse proche de dix minutes [20,39,40]. C'est bien trop long pour la plupart des challenges hémodynamiques effectués couramment comme un test de lever de jambes [41]. Une tolérance proche de la minute apparaît souhaitable pour une utilisation clinique large.

Ainsi donc, tout dépend de l'usage clinique, du temps de réponse, de la précision et du nombre de mesures effectuées. Si on a par exemple un appareil qui détermine le DC toutes les minutes et qu'on considère qu'un changement brutal de DC de 20 % doit être obligatoirement dépisté en cinq minutes, alors l'équation $LSC = 2 \times \sqrt{2} \times ET/\sqrt{n}$, devient $20 \% = 2,8 \times ET/\sqrt{5}$ et donc $ET = 16 \%$. La précision ($2ET/\text{moy}$) du DC doit donc être inférieure à 32 %. On s'aperçoit que l'étude de la précision est indissociable de celle du temps et de l'amplitude de réponse et du nombre de mesure effectuées. En effet, un système qui moyennait les mesures sur une longue période pourrait minimiser la variabilité due aux erreurs instantanées de mesure et donc améliorer sa précision. Cela se ferait cependant au détriment de son temps et de son amplitude de réponse, puisque tout changement ne serait enregistré pleinement qu'au décours de la période de moyennage.

Capacité à identifier les changements significatifs

On doit donc commencer par identifier ce qu'est un changement significatif du DC. On peut définir un seuil de variation du DC observé par la TR à cet effet. Le seuil étant fonction de l'usage clinique. En tout état de cause ce seuil devra être supérieur alors au LSC de la TR [38]. On peut ensuite vérifier si la TN est capable d'identifier comme la TR les changements significatifs de DC, en utilisant le LSC de la TN.

Puisqu'on définit des seuils qualitatifs, l'accord entre TN et TR s'analyse donc en termes de sensibilité et de spécificité et peut faire l'objet d'une analyse dans l'espace ROC [1]. On prendra sensibilité = $VP/(VP + FN)$ et spécificité = $VN/(FP + VN)$, où VP = variation parallèles des deux systèmes, VN = aucun changement pour les deux systèmes, FP = variation pour VN seulement, et FN = variation pour TR seulement.

Tolérance : une technologie fiable de monitoring du DC doit détecter les changements significatifs avec une sensibilité et une spécificité proche de 1.

Conclusion

Comparer un outil de monitoring à un outil de mesure est un peu comme comparer une vidéo avec une photo. Dans un système de monitoring, l'augmentation considérable du nombre de mesure élémentaire et leur cinétique en temps réel ou quasi réel autorisent une plus grande tolérance dans l'exactitude. Cela donne une importance particulière à la précision et au temps de réponse ; deux caractéristiques opposées qui résultent toujours d'un compromis technologique. C'est pourquoi tout nouvel outil de monitoring du DC doit bénéficier d'une validation suivant la grille présentée ici.

Conflit d'intérêt : les auteurs déclarent ne pas avoir de conflit d'intérêt.

Références

- Squara P, Denjean D, Estagnasie P, et al (2007) Non-invasive cardiac output monitoring (NICOM): a clinical validation. *Intensive Care Med* 33:1191-4
- Squara P, Cecconi C, Singer M, et al (2009) Tracking changes in cardiac output; Methodological considerations for the validation of monitoring devices. *Intensive Care Med* 35:1801-8
- Fick A (1870) Ueber die Messung des Blutquantums in den Herzventrikeln: Würzburg
- Stickland MK, Welsh RC, Haykowsky MJ, et al (2006) Effect of acute increases in pulmonary vascular pressures on exercise pulmonary gas exchange. *J Appl Physiol* 100:1910-7
- Botero M, Kirby D, Lobato EB, et al (2004) Measurement of cardiac output before and after cardiopulmonary bypass: comparison among aortic transit-time ultrasound, thermodilution, and noninvasive partial CO₂ rebreathing. *J Cardiothorac Vasc Anesth* 18:563-72
- Keren H, Burkhoff D, Squara P (2007) Evaluation of a non-invasive continuous cardiac output monitoring system based on thoracic bioreactance. *Am J Physiol Heart Circ Physiol* 293: H583-H9
- Stetz CW, Miller RG, Kelly GE, Raffin TA (1982) Reliability of the thermodilution method in the determination of cardiac output in clinical practice. *Am Rev Respir Dis* 126:1001-4
- Hillis LD, Firth BG, Winniford MD (1985) Analysis of factors affecting the variability of Fick versus indicator dilution measurements of cardiac output. *Am J Cardiol* 56:764-8
- Rubini A, Del Monte D, Catena V, et al (1995) Cardiac output measurement by the thermodilution method: an in vitro test of accuracy of three commercially available automatic cardiac output computers. *Intensive Care Med* 21:154-8
- De Backer D, Moraine JJ, Berre J, et al (1994) Effects of dobutamine on oxygen consumption in septic patients. Direct versus indirect determinations. *Am J Respir Crit Care Med* 150:95-100
- Lu Z, Mikkamala R (2006) Continuous cardiac output monitoring in humans by invasive and non-invasive peripheral blood pressure waveform analysis. *J Appl Physiol* 101:598-608
- Cecconi M, Dawson D, Grounds R, Rhodes A (2009) Lithium dilution cardiac output measurement in the critically ill patient: determination of precision of the technique. *Intensive Care Med* 35:498-504
- Le Tulzo Y, Belghith M, Seguin P, et al (1996) Reproducibility of thermodilution cardiac output determination in critically ill patients: comparison between bolus and continuous method. *J Clin Monit* 12:379-85
- Zollner C, Goetz AE, Weis M, et al (2001) Continuous cardiac output measurements do not agree with conventional bolus thermodilution cardiac output determination. *Can J Anaesth* 48:1143-7
- Bendjelid K, Schutz N, Suter PM, Romand JA (2006) Continuous cardiac output monitoring after cardiopulmonary bypass: a comparison with bolus thermodilution measurement. *Intensive Care Med* 32:919-22
- Maxwell RA, Gibson JB, Slade JB, et al (2001) Non-invasive cardiac output by partial CO₂ rebreathing after severe chest trauma. *J Trauma* 51:849-53
- Della Rocca G, Costa MG, Coccia C, et al (2003) Cardiac output monitoring: aortic transpulmonary thermodilution and pulse contour analysis agree with standard thermodilution methods in patients undergoing lung transplantation. *Can J Anaesth* 50:707-11
- Kotake Y, Moriyama K, Innami Y, et al (2003) Performance of non-invasive partial CO₂ rebreathing cardiac output and continuous thermodilution cardiac output in patients undergoing aortic reconstruction surgery. *Anesthesiology* 99:283-8
- Boldt J, Menges T, Wollbruck M, et al (1994) Is continuous cardiac output measurement using thermodilution reliable in the critically ill patient? *Crit Care Med* 22:1913-8
- Haller M, Zollner C, Briegel J, Forst H (1995) Evaluation of a new continuous thermodilution cardiac output monitor in critically ill patients: a prospective criterion standard study. *Crit Care Med* 23:860-6
- Burchell SA, Yu M, Takiguchi SA, et al (1997) Evaluation of a continuous cardiac output and mixed venous oxygen saturation catheter in critically ill surgical patients. *Crit Care Med* 25:388-91
- Luchette FA, Porembka D, Davis K Jr, et al (2000) Effects of body temperature on accuracy of continuous cardiac output measurements. *J Invest Surg* 13:147-52
- Mihm FG, Gettinger A, Hanson CW 3rd, et al (1998) A multi-center evaluation of a new continuous cardiac output pulmonary artery catheter system. *Crit Care Med* 26:1346-50
- Medin DL, Brown DT, Wesley R, et al (1998) Validation of continuous thermodilution cardiac output in critically ill patients with analysis of systematic errors. *J Crit Care* 13:184-9
- Sun Q, Rogiers P, Pauwels D, Vincent JL (2002) Comparison of continuous thermodilution and bolus cardiac output measurements in septic shock. *Intensive Care Med* 28:1276-80
- Su NY, Huang CJ, Tsai P, et al (2002) Cardiac output measurement during cardiac surgery: esophageal Doppler versus pulmonary artery catheter. *Acta Anaesthesiol Sin* 40:127-33
- Dark PM, Singer M (2004) The validity of transesophageal Doppler ultrasonography as a measure of cardiac output in critically ill adults. *Intensive Care Med* 30:2060-6

28. Chew M, Poelaert J (2003) Accuracy and repeatability of pediatric cardiac output measurement using Doppler: 20-year review of the literature. *Intensive Care Med* 29:1889–94
29. de Waal EE, Kalkman CJ, Rex S, Buhre WF (2007) Validation of a new arterial pulse contour-based cardiac output device. *Crit Care Med* 35:1904–9
30. Friesecke S, Heinrich A, Abel P, Felix S (2009) Comparison of pulmonary artery and aortic transpulmonary thermodilution for monitoring of cardiac output in patients with severe heart failure: validation of a novel method. *Crit Care Med* 37:119–23
31. Singh A, Juneja R, Mehta Y, Trehan N (2002) Comparison of continuous, stat, and intermittent cardiac output measurements in patients undergoing minimally invasive direct coronary artery bypass surgery. *J Cardiothorac Vasc Anesth* 16:186–90
32. John-Alder H, Bennet A (1981) Thermal dependence of endurance and locomotory energetics in a lizard. *Am J Physiol* 241:R342–9
33. Bland JM, Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1(8476):307–10
34. Critchley LA, Critchley JA (1999) A meta-analysis of studies using bias and precision statistics to compare cardiac output measurement techniques. *J Clin Monit Comput* 15:85–91
35. Squara P, Fourquet E, Jacquet L, et al (2003) A computer program for interpreting pulmonary artery catheterization data: results of the European HEMODYN Resident Study. *Intensive Care Med* 29:735–41
36. Squara P (2004) Matching total body oxygen consumption and delivery: a crucial objective? *Intensive Care Med* 30:2170–9
37. Marqué S, Cariou A, Chiche J, Squara P (2009) Non-invasive cardiac output monitoring (NICOM) Compared to minimally invasive monitoring (VIGILEO). *Crit Care* 13:R73
38. Benomar B, Ouattara A, Estagnasie P, et al (2010) Fluid responsiveness predicted by non-invasive bioimpedance-based passive leg raise test. *Intensive Care Med* 16:1875–81
39. Poli de Figueiredo LF, Malbouisson LM, Varicoda EY, et al (1999) Thermal filament continuous thermodilution cardiac output delayed response limits its value during acute hemodynamic instability. *J Trauma* 47:288–93
40. Siegel LC, Hennessy MM, Pearl RG (1996) Delayed time response of the continuous cardiac output pulmonary artery catheter. *Anesth Analg* 83:1173–7
41. Monnet X, Teboul J (2008) Passive leg raising. *Intensive Care Med* 34:659–63