



BU bibliothèque Lyon 1

<http://portaildoc.univ-lyon1.fr>

Creative commons : Paternité - Pas d'Utilisation Commerciale -
Pas de Modification 2.0 France (CC BY-NC-ND 2.0)



<http://creativecommons.org/licenses/by-nc-nd/2.0/fr>

Université Claude Bernard  Lyon 1

UFR de MEDECINE LYON-EST

Année 2017 N° 190

**Qualité du *reporting* dans les études diagnostiques de
réponse au remplissage vasculaire : revue systématique sur
dix ans.**

Thèse d'exercice en médecine

Présentée à l'Université Claude Bernard Lyon 1

Et soutenue publiquement **le 6 septembre 2017**

En vue d'obtenir le titre de Docteur en Médecine

Par

Guillaume Izaute

Né le 18 avril 1989 au Mans

Sous la direction du **Docteur Matthias Jacquet-Lagrèze**

UNIVERSITE CLAUDE BERNARD – LYON 1

Président

Frédéric FLEURY

Président du Comité de

Pierre COCHAT

Coordination des Etudes Médicales

Directrice Générale des Services

Dominique MARCHAND

Secteur Santé

UFR de Médecine Lyon Est

Doyen : Gilles RODE

UFR de Médecine Lyon Sud-

Doyen : Carole BURILLON

Charles Mérieux

Institut des Sciences Pharmaceutiques

Directrice : Christine VINCIGUERRA

Et Biologiques (ISPB)

UFR d'Odontologie

Directeur : Denis BOURGEOIS

Institut des Sciences et Techniques

Directeur : Xavier PERROT

De Réadaptation (ISTR)

Département de Biologie Humaine

Directrice : Anne-Marie SCHOTT

Secteur Sciences et Technologie

UFR de Sciences et Technologies

Directeur : Fabien de MARCHI

UFR de Sciences et Techniques des

Directeur : Yannick VANPOULLE

Activités Physiques et Sportives (STAPS)

Polytech Lyon

Directeur : Emmanuel PERRIN

I.U.T.

Directeur : Christophe VITON

Institut des Sciences Financières

Directeur : Nicolas LEBOISNE

Et Assurances (ISFA)

Observatoire de Lyon

Directrice : Isabelle DANIEL

Ecole Supérieure du Professorat

Directeur : Alain MOUGNIOTTE

Et de l'Education (ESPE)

Faculté de Médecine Lyon Est Liste des enseignants 2016/2017

Professeurs des Universités – Praticiens Hospitaliers Classe exceptionnelle Echelon 2

Blay	Jean-Yves	Cancérologie ; radiothérapie
Cochat	Pierre	Pédiatrie
Cordier	Jean-François	Pneumologie ; addictologie
Etienne	Jérôme	Bactériologie-virologie ; hygiène hospitalière
Gouillat	Christian	Chirurgie digestive
Guérin	Jean-François	Biologie et médecine du développement et de la reproduction ; gynécologie médicale
Mornex	Jean-François	Pneumologie ; addictologie
Ninet	Jacques	Médecine interne ; gériatrie et biologie du vieillessement ; médecine générale ; addictologie
Philip	Thierry	Cancérologie ; radiothérapie
Ponchon	Thierry	Gastroentérologie ; hépatologie ; addictologie
Revel	Didier	Radiologie et imagerie médicale
Rivoire	Michel	Cancérologie ; radiothérapie
Rudigoz	René-Charles	Gynécologie-obstétrique ; gynécologie médicale
Thivolet-Bejui	Françoise	Anatomie et cytologie pathologiques
Vandenesch	François	Bactériologie-virologie ; hygiène hospitalière

Professeurs des Universités – Praticiens Hospitaliers Classe exceptionnelle Echelon 1

Borson-Chazot	Françoise	Endocrinologie, diabète et maladies métaboliques ; gynécologie médicale
Chassard	Dominique	Anesthésiologie-réanimation ; médecine d'urgence
Claris	Olivier	Pédiatrie
D'Amato	Thierry	Psychiatrie d'adultes ; addictologie
Delahaye	François	Cardiologie
Denis	Philippe	Ophtalmologie
Disant	François	Oto-rhino-laryngologie
Douek	Philippe	Radiologie et imagerie médicale
Ducerf	Christian	Chirurgie digestive
Finet	Gérard	Cardiologie
Gaucherand	Pascal	Gynécologie-obstétrique ; gynécologie médicale
Guérin	Claude	Réanimation ; médecine d'urgence
Herzberg	Guillaume	Chirurgie orthopédique et traumatologique
Honorat	Jérôme	Neurologie
Lachaux	Alain	Pédiatrie
Lehot	Jean-Jacques	Anesthésiologie-réanimation ; médecine d'urgence
Lermusiaux	Patrick	Chirurgie thoracique et cardiovasculaire
Lina	Bruno	Bactériologie-virologie ; hygiène hospitalière
Martin	Xavier	Urologie
Mellier	Georges	Gynécologie-obstétrique ; gynécologie médicale
Mertens	Patrick	Anatomie
Michallet	Mauricette	Hématologie ; transfusion
Miossec	Pierre	Immunologie
Morel	Yves	Biochimie et biologie moléculaire

Moulin	Philippe	Nutrition
Négrier	Sylvie	Cancérologie ; radiothérapie
Neyret	Philippe	Chirurgie orthopédique et traumatologique
Nighoghossian	Norbert	Neurologie
Ninet	Jean	Chirurgie thoracique et cardiovasculaire
Obadia	Jean-François	Chirurgie thoracique et cardiovasculaire
Ovize	Michel	Physiologie
Rode	Gilles	Médecine physique et de réadaptation
Terra	Jean-Louis	Psychiatrie d'adultes ; addictologie
Zoulim	Fabien	Gastroentérologie ; hépatologie ; addictologie

Professeurs des Universités – Praticiens Hospitaliers Première classe

André-Fouet	Xavier	Cardiologie
Argaud	Laurent	Réanimation ; médecine d'urgence
Badet	Lionel	Urologie
Barth	Xavier	Chirurgie générale
Bessereau	Jean-Louis	Biologie cellulaire
Berthezene	Yves	Radiologie et imagerie médicale
Bertrand	Yves	Pédiatrie
Boillot	Olivier	Chirurgie digestive
Braye	Fabienne	Chirurgie plastique, reconstructrice et esthétique ; brûlologie
Breton	Pierre	Chirurgie maxillo-faciale et stomatologie
Chevalier	Philippe	Cardiologie
Colin	Cyrille	Epidémiologie, économie de la santé et prévention
Colombel	Marc	Urologie
Cottin	Vincent	Pneumologie ; addictologie
Devouassoux	Mojgan	Anatomie et cytologie pathologiques
Di Filippo	Sylvie	Cardiologie
Dumontet	Charles	Hématologie ; transfusion
Durieu	Isabelle	Médecine interne ; gériatrie et biologie du vieillissement ; médecine générale ; addictologie
Edery	Charles Patrick	Génétique
Fauvel	Jean-Pierre	Thérapeutique ; médecine d'urgence ; addictologie
Guenot	Marc	Neurochirurgie
Gueyffier	François	Pharmacologie fondamentale ; pharmacologie clinique ; addictologie
Guibaud	Laurent	Radiologie et imagerie médicale
Javouhey	Etienne	Pédiatrie
Juillard	Laurent	Néphrologie
Jullien	Denis	Dermato-vénéréologie
Kodjikian	Laurent	Ophtalmologie
Krolak Salmon	Pierre	Médecine interne ; gériatrie et biologie du vieillissement ; médecine générale ; addictologie
Lejeune	Hervé	Biologie et médecine du développement et de la reproduction ; gynécologie médicale
Mabrut	Jean-Yves	Chirurgie générale
Merle	Philippe	Gastroentérologie ; hépatologie ; addictologie
Mion	François	Physiologie
Morelon	Emmanuel	Néphrologie
Mure	Pierre-Yves	Chirurgie infantile
Négrier	Claude	Hématologie ; transfusion
Nicolino	Marc	Pédiatrie
Picot	Stéphane	Parasitologie et mycologie

Rouvière Roy	Olivier Pascal	Radiologie et imagerie médicale Biostatistiques, informatique médicale et technologies de communication
Rymlin Saoud Schaeffer Scheiber Schott-Pethelaz Tilikete Truy Turjman Vallée Vanhems Vukusic	Philippe Mohamed Laurent Christian Anne-Marie Caroline Eric Francis Bernard Philippe Sandra	Neurologie Psychiatrie d'adultes Biologie cellulaire Biophysique et médecine nucléaire Epidémiologie, économie de la santé et prévention Physiologie Oto-rhino-laryngologie Radiologie et imagerie médicale Anatomie Epidémiologie, économie de la santé et prévention Neurologie

Professeurs des Universités – Praticiens Hospitaliers Seconde Classe

Ader Aubrun Boussel Calender Chapurlat Charbotel Chêne Cotton Crouzet Dargaud David Di Rocco Dubernard Ducray Dumortier Fanton Fellahi Ferry Fourneret Gillet Girard Gleizal Henaine Hot Huissoud Jacquin-Courtois Janier Lesurtel Michel Million Monneuse Nataf Peretti Pignat Poncet Raverot Ray-Coquard	Florence Frédéric Loïc Alain Roland Barbara Gautier François Sébastien Yesim Jean-Stéphane Federico Gil François Jérôme Laurent Jean-Luc Tristan Pierre Yves Nicolas Arnaud Roland Arnaud Cyril Sophie Marc Mickaël Philippe Antoine Olivier Serge Noël Jean-Christian Gilles Gérald Isabelle	Maladies infectieuses ; maladies tropicales Anesthésiologie-réanimation ; médecine d'urgence Radiologie et imagerie médicale Génétique Rhumatologie Médecine et santé au travail Gynécologie-obstétrique ; gynécologie médicale Radiologie et imagerie médicale Urologie Hématologie ; transfusion Anesthésiologie-réanimation ; médecine d'urgence Neurochirurgie Gynécologie-obstétrique ; gynécologie médicale Neurologie Gastroentérologie ; hépatologie ; addictologie Médecine légale Anesthésiologie-réanimation ; médecine d'urgence Maladie infectieuses ; maladies tropicales Pédopsychiatrie ; addictologie Pédiatrie Pneumologie Chirurgie maxillo-faciale et stomatologie Chirurgie thoracique et cardiovasculaire Médecine interne Gynécologie-obstétrique ; gynécologie médicale Médecine physique et de réadaptation Biophysique et médecine nucléaire Chirurgie générale Epidémiologie, économie de la santé et prévention Chirurgie vasculaire ; médecine vasculaire Chirurgie générale Cytologie et histologie Nutrition Oto-rhino-laryngologie Chirurgie générale Endocrinologie, diabète et maladies métaboliques ; gynécologie médicale Cancérologie ; radiothérapie
--	---	--

Rheims	Sylvain	Neurologie
Richard	Jean-Christophe	Réanimation ; médecine d'urgence
Robert	Maud	Chirurgie digestive
Rossetti	Yves	Physiologie
Souquet	Jean-Christophe	Gastroentérologie ; hépatologie ; addictologie
Thaumat	Olivier	Néphrologie
Thibault	Hélène	Physiologie
Wattel	Eric	Hématologie ; transfusion

Professeur des Universités - Médecine Générale

Flori	Marie
Letrilliart	Laurent
Moreau	Alain
Zerbib	Yves

Professeurs associés de Médecine Générale

Lainé	Xavier
-------	--------

Professeurs émérites

Baulieux	Jacques	Cardiologie
Beziat	Jean-Luc	Chirurgie maxillo-faciale et stomatologie
Chayvialle	Jean-Alain	Gastroentérologie ; hépatologie ; addictologie
Daligand	Liliane	Médecine légale et droit de la santé
Droz	Jean-Pierre	Cancérologie ; radiothérapie
Floret	Daniel	Pédiatrie
Gharib	Claude	Physiologie
Mauguière	François	Neurologie
Neidhardt	Jean-Pierre	Anatomie
Petit	Paul	Anesthésiologie-réanimation ; médecine d'urgence
Sindou	Marc	Neurochirurgie
Touraine	Jean-Louis	Néphrologie
Trepo	Christian	Gastroentérologie ; hépatologie ; addictologie
Trouillas	Jacqueline	Cytologie et histologie
Viale	Jean-Paul	Réanimation ; médecine d'urgence

Maîtres de Conférence – Praticiens Hospitaliers

Hors classe

Benchaib	Mehdi	Biologie et médecine du développement et de la reproduction ; gynécologie médicale
Bringuier	Pierre-Paul	Cytologie et histologie
Dubourg	Laurence	Physiologie
Germain	Michèle	Physiologie
Jarraud	Sophie	Bactériologie-virologie ; hygiène hospitalière
Le Bars	Didier	Biophysique et médecine nucléaire
Normand	Jean-Claude	Médecine et santé au travail
Persat	Florence	Parasitologie et mycologie
Piaton	Eric	Cytologie et histologie

Sappey-Marinier	Dominique	Biophysique et médecine nucléaire
Streichenberger	Nathalie	Anatomie et cytologie pathologiques
Timour-Chah	Quadiri	Pharmacologie fondamentale ; pharmacologie clinique ; addictologie
Voiglio	Eric	Anatomie

Maîtres de Conférence – Praticiens Hospitaliers Première classe

Barnoud	Raphaëlle	Anatomie et cytologie pathologiques
Bontemps	Laurence	Biophysique et médecine nucléaire
Chalabreysse	Lara	Anatomie et cytologie pathologiques
Charrière	Sybil	Nutrition
Collardeau Frachon	Sophie	Anatomie et cytologie pathologiques
Confavreux	Cyrille	Rhumatologie
Cozon	Grégoire	Immunologie
Escuret	Vanessa	Bactériologie-virologie ; hygiène hospitalière
Hervieu	Valérie	Anatomie et cytologie pathologiques
Kolopp-Sarda	Marie Nathalie	Immunologie
Lesca	Gaëtan	Génétique
Lukaszewicz	Anne-Claire	Anesthésiologie-réanimation ; médecine d'urgence
Maucort Boulch	Delphine	Biostatistiques, informatique médicale et technologies de communication
Meyronet	David	Anatomie et cytologie pathologiques
Pina-Jomir	Géraldine	Biophysique et médecine nucléaire
Plotton	Ingrid	Biochimie et biologie moléculaire
Rabilloud	Muriel	Biostatistiques, informatique médicale et technologies de communication
Rimmele	Thomas	Anesthésiologie-réanimation ; médecine d'urgence
Ritter	Jacques	Epidémiologie, économie de la santé et prévention
Roman	Sabine	Physiologie
Tardy Guidollet	Véronique	Biochimie et biologie moléculaire
Tristan	Anne	Bactériologie-virologie ; hygiène hospitalière
Venet	Fabienne	Immunologie
Vlaeminck-Guillem	Virginie	Biochimie et biologie moléculaire

Maîtres de Conférences – Praticiens Hospitaliers Seconde classe

Casalegno	Jean-Sébastien	Bactériologie-virologie ; hygiène hospitalière
Curie	Aurore	Pédiatrie
Duclos	Antoine	Epidémiologie, économie de la santé et prévention
Lemoine	Sandrine	Physiologie
Marignier	Romain	Neurologie
Phan	Alice	Dermato-vénéréologie
Schluth-Bolard	Caroline	Génétique
Simonet	Thomas	Biologie cellulaire
Vasiljevic	Alexandre	Anatomie et cytologie pathologiques

Maîtres de Conférences associés de Médecine Générale

Farge	Thierry
Pigache	Christophe

Serment d'Hippocrate

Je promets et je jure d'être fidèle aux lois de l'honneur et de la probité dans l'exercice de la Médecine. Je respecterai toutes les personnes, leur autonomie et leur volonté, sans discrimination. J'interviendrai pour les protéger si elles sont vulnérables ou menacées dans leur intégrité ou leur dignité. Même sous la contrainte, je ne ferai pas usage de mes connaissances contre les lois de l'humanité.

J'informerai les patients des décisions envisagées, de leurs raisons et de leurs conséquences. Je ne tromperai jamais leur confiance. Je donnerai mes soins à l'indigent et je n'exigerai pas un salaire au dessus de mon travail. Admis dans l'intimité des personnes, je tairai les secrets qui me seront confiés et ma conduite ne servira pas à corrompre les mœurs.

Je ferai tout pour soulager les souffrances. Je ne prolongerai pas abusivement la vie ni ne provoquerai délibérément la mort. Je préserverai l'indépendance nécessaire et je n'entreprendrai rien qui dépasse mes compétences. Je perfectionnerai mes connaissances pour assurer au mieux ma mission.

Que les hommes m'accordent leur estime si je suis fidèle à mes promesses. Que je sois couvert d'opprobre et méprisé si j'y manque.

Remerciements

Tous mes remerciements et ma gratitude,

Au Professeur Jean-Luc Fellahi, pour nous faire l'honneur de présider ce jury de thèse, pour la confiance que vous avez eu dans notre travail, pour vos conseils et pour vos encouragements. Merci également pour le soin que vous apportez à la formation des internes, j'ai beaucoup appris au cours de mon passage dans votre service.

Au Professeur Vincent Piriou, pour avoir accepté de participer à mon jury de thèse, pour m'avoir accueilli déjà par deux fois dans votre service au sein duquel je me suis épanoui. Merci de m'avoir invité à poursuivre après mon internat au Centre hospitalier de Lyon Sud.

Au Professeur Jean-Christophe Richard, pour avoir accepté de participer à mon jury de thèse, pour votre expertise sur le sujet et pour les remarques que vous apporterez à notre travail.

Au Docteur Matthias Jacquet-Lagrèze, pour avoir encadré cette thèse. Malgré les difficultés rencontrées, travailler avec toi a été un réel plaisir, tu m'as beaucoup appris, sur la précharge dépendance, sur la méthodologie et sur la recherche en général. Merci pour ta disponibilité, ton aide, tes encouragements, ton humour. J'espère qu'on aura encore des occasions pour échanger. A bientôt, profite bien de l'année qui vient.

A Lolo, pour m'avoir soutenu tout au long du travail, pour ta patience, pour avoir accepté la réduction de notre temps-libre commun. Merci d'être là.

A mes parents, pour votre soutien et votre confiance. Merci pour vos relectures et vos conseils.

A mes amis, de Sarthe, d'Angers et de Rhône-Alpes. Merci à Alex et Nico pour votre soutien. Merci à Coco pour ton aide dans la préparation de l'ESA.

A mes co-internes, merci pour cette ambiance au sein de notre spécialité.

A mes frères Mathieu et Benoît.

Sommaire

SERMENT D'HIPPOCRATE	8
REMERCIEMENTS	9
SOMMAIRE	10
1. INTRODUCTION	13
1.1. LA REPONSE AU REMPLISSAGE VASCULAIRE	13
1.1.1. DE QUOI PARLONS-NOUS ?	13
1.1.1.1. La volémie	13
1.1.1.2. La précharge	15
1.1.1.3. Précharge dépendance ou réserve de précharge	17
1.1.1.4. Le remplissage vasculaire	19
1.1.2. LES INDICES STATIQUES DE PRECHARGE	21
1.1.3. LES INDICES DYNAMIQUES BASES SUR LES VARIATIONS DE LA PRESSION INTRATHORACIQUE	22
1.1.3.1. Variations de la pression intrathoracique pendant un cycle respiratoire.....	22
1.1.3.2. Variations de la pression intrathoracique lors d'une pause respiratoire.....	25
1.1.3.3. Variations de la pression intrathoracique lors d'une modification de la PEEP	26
1.1.3.4. Variations de la pression intrathoracique lors d'une modification du volume courant	26
1.1.3.5. Variations de la pression intrathoracique lors d'une manœuvre de recrutement	27
1.1.4. LES INDICES DYNAMIQUES NON BASES SUR LES INTERACTIONS CŒUR-POUMON	27
1.1.4.1. Le lever de jambes passif	27
1.1.4.2. La compression abdominale	28
1.1.4.3. Le mini-fluid challenge	29
1.1.4.4. La variation du temps de diastole post ESA.....	29
1.1.5. LIMITES ET UTILISATION PRATIQUE DE CES TESTS.....	29
1.2. LES ETUDES DIAGNOSTIQUES	32
1.2.1. DE QUOI PARLONS-NOUS ?	32
1.2.2. EVALUATION DES TESTS DIAGNOSTIQUES : LES GRANDS PRINCIPES	33
1.2.2.1. Méthodologie des études diagnostiques.....	34
1.2.2.2. Analyse des résultats des études diagnostiques.....	36
1.2.2.3. Intérêt et applicabilité du test étudié	40
1.3. QUALITE DU REPORTING DANS LES ETUDES DIAGNOSTIQUES	41

1.3.1.	L'EXISTENCE DE RISQUES POTENTIELS CONCERNANT LA VALIDITE INTERNE ET/OU EXTERNE DES ETUDES DE PRECISION DIAGNOSTIQUE.	41
1.3.2.	LA NAISSANCE D'UN GROUPE, LE PROJET STARD	42
1.3.2.1.	Leur méthodologie.....	42
1.3.2.2.	Leurs résultats.....	43
1.3.3.	LES RECOMMANDATIONS STARD.....	43
1.3.4.	DES RECOMMANDATIONS MISES A JOUR EN 2015.....	47
1.4.	HYPOTHESES ET OBJECTIFS	51
2.	<u>MATERIELS ET METHODES.....</u>	52
2.1.	CADRE METHODOLOGIQUE ET TERMINOLOGIE	52
2.1.1.	UNE REVUE SYSTEMATIQUE	52
2.1.2.	<i>THE PRISMA STATEMENT</i>	52
2.1.3.	<i>THE PICO STATEMENT</i>	53
2.1.4.	LES INVESTIGATEURS	54
2.1.5.	SOURCE DE FINANCEMENT	54
2.2.	SELECTION DES ETUDES	54
2.2.1.	SOURCE D'INFORMATION.....	54
2.2.2.	RECHERCHE DANS LA LITTERATURE.....	54
2.2.3.	SELECTION DES ETUDES.....	56
2.3.	COLLECTE DES DONNEES	56
2.3.1.	CARACTERISTIQUES DES ETUDES	56
2.3.2.	SCORE STARD 2003	57
2.3.3.	SCORE STARD 2015	59
2.4.	ANALYSE STATISTIQUE	59
3.	<u>RESULTATS</u>	61
3.1.	CARACTERISTIQUES DES ETUDES	61
3.2.	SCORE STARD 2003.....	64
3.2.1.	QUALITE GLOBALE DES ETUDES	64
3.2.2.	ANALYSE DU <i>REPORTING</i> EN FONCTION DES ITEMS.....	64
3.3.	SCORE STARD 2015.....	68
3.3.1.	QUALITE GLOBALE DES ETUDES	68
3.3.2.	ANALYSE DU <i>REPORTING</i> EN FONCTION DES ITEMS.....	68
3.4.	FACTEURS ASSOCIES A LA QUALITE DU <i>REPORTING</i>	71
4.	<u>DISCUSSION</u>	73

4.1. RESULTAT PRINCIPAL ET SON IMPLICATION.....	73
4.2. HYPOTHESES EXPLICATIVES DE NOS RESULTATS	78
4.3. VALIDITE EXTERNE.....	81
4.4. FORCES ET FAIBLESSES DE NOTRE ETUDE	84
4.4.1. LES ASPECTS POSITIFS	84
4.4.1.1. Le contexte	84
4.4.1.2. La méthodologie	85
4.4.1.3. Les résultats	86
4.4.2. LES LIMITES	86
4.5. NOUVELLES PERSPECTIVES	88
<u>5. CONCLUSION.....</u>	<u>90</u>
<u>6. ANNEXES</u>	<u>92</u>
6.1. ADHESION A PRISMA	93
6.2. RESUME PRESENTE A LA SRLF 2017	94
6.3. RESUME PRESENTE A L'ESA 2017	95
6.4. LETTRE ADRESSEE A ANESTHESIOLOGY	96
6.5. LISTE DES ARTICLES INCLUS DANS NOTRE ETUDE	99
<u>7. BIBLIOGRAPHIE</u>	<u>106</u>

1. Introduction

Le remplissage vasculaire est une thérapeutique quotidienne au bloc opératoire et en réanimation. C'est une pierre angulaire de la prise en charge hémodynamique des patients présentant une insuffisance circulatoire. Cependant, le choix de « remplir » ou non un patient peut être difficile : un excès ou un défaut de remplissage vasculaire ont des conséquences défavorables. Il est donc fondamental d'identifier les patients susceptibles de bénéficier d'un remplissage vasculaire. Cette nécessité a été abondamment pourvue ces dix dernières années par la publication de nombreuses études s'attachant à prédire la réponse au remplissage vasculaire. Les études diagnostiques nécessitent une méthodologie rigoureuse, sans cela elles s'exposent à un risque de biais. Nous proposons ici d'évaluer la qualité du *reporting* des études diagnostiques de réponse au remplissage vasculaire.

1.1. La réponse au remplissage vasculaire

Déterminer l'indication de remplissage vasculaire est un sujet de discussion (1). En augmentant le débit cardiaque, le remplissage vasculaire peut permettre une amélioration de la perfusion tissulaire lorsqu'il est administré dans les bonnes conditions. Non administré alors qu'il est indiqué, ou bien réalisé en excès, il peut avoir un effet défavorable. Il est important de détecter les patients pouvant en bénéficier. Afin de mieux comprendre et de cerner les enjeux de ce travail, il apparaît nécessaire de rappeler quelques fondamentaux. Dans un premier temps nous allons exposer les points essentiels de la précharge dépendance puis, nous chercherons à savoir comment elle peut être évaluée.

1.1.1. De quoi parlons-nous ?

1.1.1.1. La volémie

La volémie correspond à la masse sanguine totale de l'organisme. Elle comporte le plasma et les éléments figurés du sang. Sa valeur normale est d'environ 72,6ml/kg chez l'homme et

66,3 ml/kg chez la femme (2), ou bien 2566 ml/m² chez l'homme et 2245 ml/m² chez la femme (3). Ces valeurs, qu'elles soient indexées sur le poids ou bien sur la surface corporelle, sont cependant mises en défaut pour les poids et tailles extrêmes. Le volume sanguin est mieux corrélé au poids idéal (estimé à partir de la taille et du sexe), qu'au poids ou à la surface corporelle. La figure 1 montre la relation entre le volume sanguin en fonction du poids (ratio volume sanguin / poids) et l'écart par rapport au poids idéal (2) :

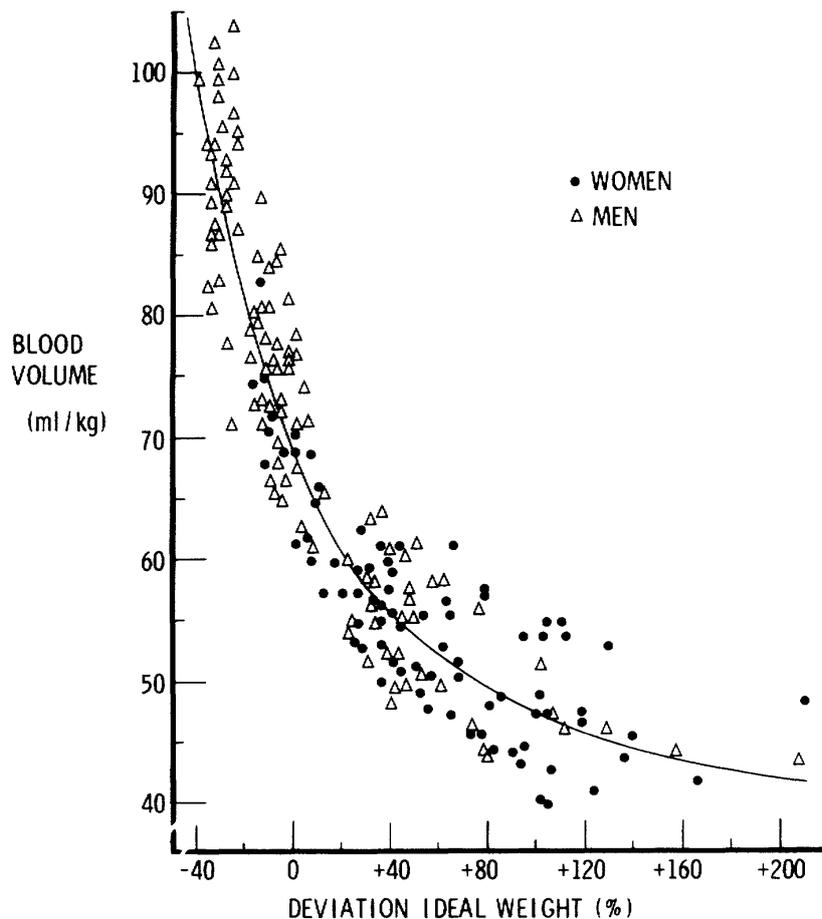


Figure 1 Relation entre le volume sanguin en fonction du poids (ratio volume sanguin / poids) et l'écart par rapport au poids idéal. (80 femmes : cercles ; 80 hommes : triangles)(2)

En situation clinique, les valeurs indexées sur le poids fournissent un haut niveau de discrimination prédictive pour classer l'état du volume sanguin idéal, avec un risque d'erreur de 8 % (4). Cependant, le rapport du volume sur le poids reste largement utilisé (4).

La volémie se répartie grossièrement entre la circulation veineuse systémique (65 %), la circulation artérielle systémique (15%), la circulation pulmonaire (10%) et le cœur en diastole (10%). Elle conditionne à la fois le retour veineux vers le cœur, le débit cardiaque, et in fine l'apport d'oxygène aux tissus.

1.1.1.2. La précharge

La précharge correspond à la tension appliquée à une fibre musculaire juste avant sa contraction. L'élongation du muscle strié avant sa contraction résulte de cette précharge. Par ailleurs, il existe une forte relation entre la force développée par la fibre musculaire et sa longueur. Celle-ci augmente avec l'élongation, jusqu'à un certain plateau, puis diminue au-delà. Ce plateau, ou bien la « taille idéale » du sarcomère, se situe entre 2,05 et 2,2 micromètres (5) (figures 2 et 3).

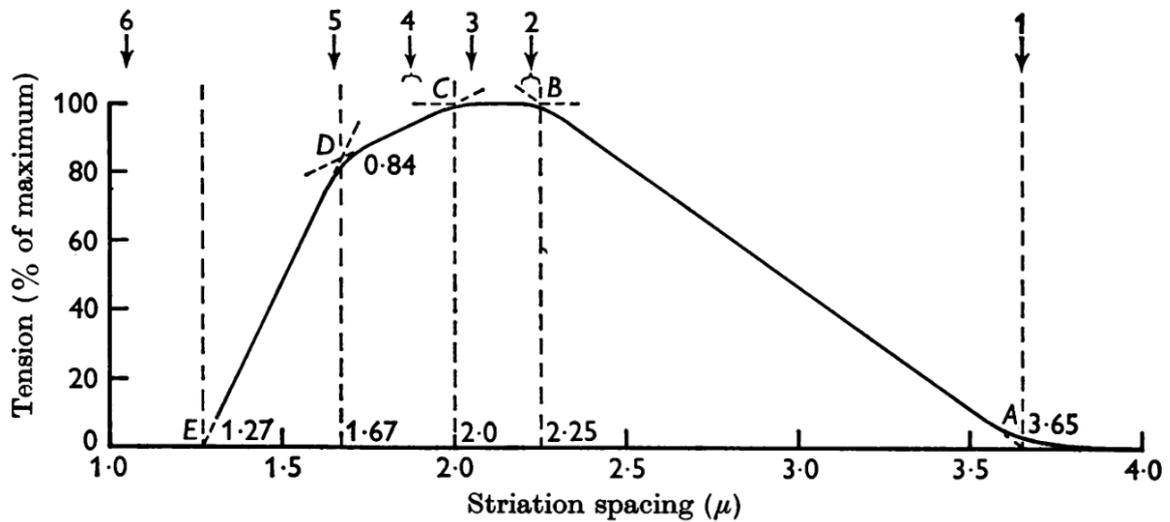


Figure 2 Relation entre la longueur du sarcomère et la force développée par la fibre musculaire. Les flèches numérotées correspondent à différents stades de chevauchement des fibres striées, cf figure 3 (5).

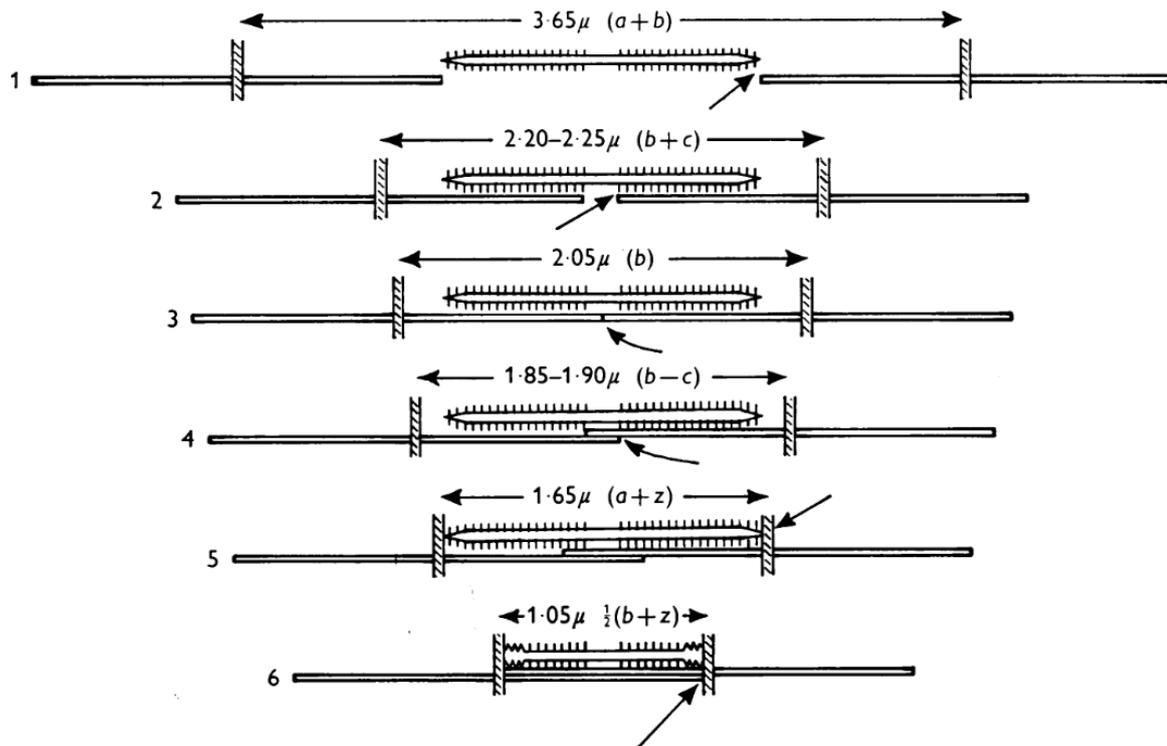


Figure 3 Différents états critiques de chevauchement au sein du sarcomère. On remarque que le recouvrement des fibres d'actine et de myosine est optimal pour une certaine élongation(5).

La transposition de ce concept au muscle strié cardiaque est rendue difficile par la conformation du cœur. Les fibres sont multiples, non linéaires, plus ou moins circulaires. La précharge ventriculaire correspond à la contrainte (ou stress) à laquelle les fibres sont soumises avant la contraction. Elle est estimée avec la formule suivante, en admettant que le cœur est une sphère (6) :

$$\sigma = P \times r / 2 h$$

avec :

σ : contrainte (stress) qui s'exerce par unité de surface en dynes/cm²

P : pression à l'intérieur de la sphère en mmHg

r : rayon de la sphère en cm

h : épaisseur de la sphère en cm

Cependant, le VG étant un ellipsoïde, une formule plus appropriée s'écrit(7) :

$$\sigma = (P \cdot D \cdot 0.33) / 4 h (1 + [h / D])$$

avec :

D : diamètre de l'ellipsoïde, en cm

On remarque que la précharge ventriculaire est proportionnelle à la pression et au rayon (et donc au volume) myocardique. Elle est inversement proportionnelle à l'épaisseur myocardique.

Schématiquement, les deux principaux déterminants de la précharge ventriculaire sont le retour veineux et la compliance cardiaque (7). Le retour veineux dépend du volume sanguin circulant, des pressions endothoracique et abdominale, du tonus veineux et de la résistance veineuse. La compliance cardiaque dépend de la pression intrapéricardique et de la fonction diastolique cardiaque (7).

1.1.1.3. Précharge dépendance ou réserve de précharge

Approchons-nous de notre sujet avec un historique très bref. A la fin du 19ème siècle, Otto Frank observa une contraction ventriculaire accrue lorsque le ventricule était étiré avant sa contraction. Plus tard, Ernest Starling ajouta que l'augmentation des pressions de remplissage ventriculaire, par augmentation du retour veineux, conduisait à l'augmentation du volume d'éjection systolique. L'augmentation de la pression auriculaire droite (reflet de l'augmentation de la précharge ventriculaire) entraînait l'étirement des myocytes et l'élongation du sarcomère. Il nota que le phénomène s'accompagnait d'une augmentation de la force générée et par conséquent du volume d'éjection systolique. La précharge dépendance est donc une propriété contractile du cœur lui permettant d'augmenter son débit cardiaque lorsqu'il est soumis à une augmentation de précharge (8). La relation entre le volume d'éjection systolique et la précharge est décrite par la courbe de Franck-Starling, en figure 4.

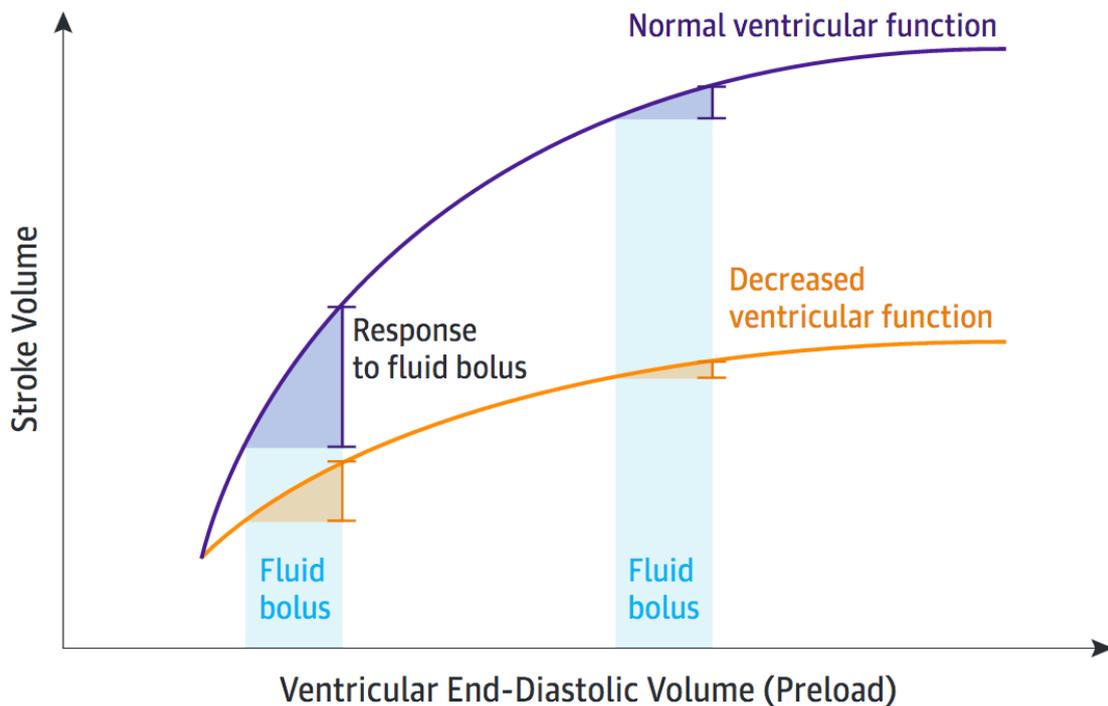


Figure 4 Effet de l'augmentation de la précharge sur le volume d'éjection systolique sur des ventricules avec une contractilité normale (courbe du haut, violette) ou diminuée (courbe du bas, orange). Les courbes de Frank-Starling illustrent le fait que l'augmentation du volume d'éjection systolique dépend à la fois de la contractilité ventriculaire et de la précharge basale(9).

Lorsque le cœur fonctionne sur la partie ascendante de la courbe, le volume d'éjection systolique augmente sensiblement lorsque l'on augmente la précharge (avec un remplissage vasculaire). Il y a une amélioration du débit cardiaque et du transport en oxygène. En revanche, lorsque le cœur fonctionne sur la partie plate de la courbe, l'augmentation de précharge par l'injection intraveineuse de fluides ne fera pas augmenter sensiblement le volume d'éjection systolique. La forme de la courbe Frank-Starling dépend de la contractilité cardiaque et de la postcharge du patient. Schématiquement, la pente de la courbe augmente avec la diminution de postcharge et avec l'augmentation de la contractilité. A l'inverse cette pente diminue avec l'augmentation de la postcharge (soit pour une augmentation importante soit en cas de fonction cardiaque altérée) et avec la diminution de la contractilité (10).

Les différentes formes que peut prendre la courbe de Frank-Starling varient donc d'un patient à l'autre et chez un même patient au cours du temps. Par exemple, lors d'un sevrage de CEC en chirurgie cardiaque ou lors de l'instauration d'un traitement par dobutamine, l'inotropisme cardiaque augmente et un patient qui n'était initialement non précharge dépendant peut le devenir. A la présentation initiale du choc septique, les patients hémodynamiquement instables sont généralement hypovolémiques et répondent souvent au remplissage vasculaire

avec une augmentation du débit cardiaque (11). Par la suite, le sepsis peut induire une dysfonction cardiaque, la courbe a alors tendance à s'aplatir, le patient n'est alors plus précharge dépendant (12).

Le défi clinique consiste à déterminer à un instant donné si le cœur du patient fonctionne sur la partie ascendante ou plate de la courbe Frank-Starling.

1.1.1.4. Le remplissage vasculaire

La première utilisation de solution salée intraveineuse pour le traitement des maladies est consignée dans une lettre de Thomas Latta soumise au Conseil central de la santé à Londres et publiée dans *The Lancet*, en 1832. L'histoire de cette première est détaillée dans l'*International Journal of Epidemiology* (13), la voici résumée.

Le choléra est apparu en Grande-Bretagne 1831, en provenance du Bengale et après avoir balayé l'Europe durant cinq ans. A son arrivée en Grande-Bretagne, les symptômes et la progression de la maladie étaient bien documentés, mais son traitement restait inconnu.

William O'Shaughnessy, un jeune diplômé en médecine, a étudié le sang des patients atteints de choléra et a constaté qu'il avait perdu de l'eau et de la solution saline neutre (13). Il a suggéré que « restaurer le sang à son état naturel » pourrait atténuer les symptômes de la maladie.

Il a alors mené une série d'expériences chez des chiens avec des sels dissous dans de l'eau tiède qu'il a rapporté au *Lancet* en 1831 (14). Il laisse le médecin Thomas Latta appliquer ses méthodes aux humains. Celui-ci a signalé ses conclusions dans une lettre adressée au Conseil central de la santé de Londres qui l'a transmise au *Lancet* (15).

Cette lettre de Latta est la première utilisation consignée de solution salée intraveineuse pour le traitement d'une maladie chez l'Homme. Cependant, les solutions intraveineuses de Latta, ont été largement ignorées jusqu'à ce que Ringer ait décrit une solution similaire 50 ans plus tard (16). Aujourd'hui la solution salée isotonique est un traitement omniprésent.

Le remplissage vasculaire vise à restaurer et à maintenir le volume intravasculaire et ainsi une perfusion tissulaire adéquate chez les patients présentant une insuffisance circulatoire aiguë. Comme vu précédemment, la perfusion tissulaire est alors maintenue grâce à l'augmentation du volume d'éjection systolique et donc du débit cardiaque.

Récemment, P Marik a constaté qu'au début du 21ème siècle, les patients atteints de sepsis sévère et de choc septique recevaient des quantités massives de cristaalloïdes pendant les 72

premières heures d'hospitalisation (17)(18). Cette approche agressive, appelée « thérapie précoce ciblée sur objectifs » pour *Early Goal-Directed Therapy* (EGDT) a été considérée comme la norme et approuvée par les recommandations internationales (19–21). Ces recommandations s'appuyaient pour partie sur l'étude de Rivers et ses collègues. Ils ont montré que l'administration massive de fluides pendant les six premières heures de réanimation de patients atteints de sepsis sévère et de choc septique était associée à une meilleure issue (11).

Cependant, trois essais cliniques multicentriques récents, ProCESS, ARISE et ProMiSe, ainsi qu'une méta-analyse sur l'EGDT, ont montré que cette approche ne parvient pas à améliorer le devenir des patients atteints de sepsis sévère et de choc septique (22–25). En développant « une approche rationnelle de la thérapie liquidienne dans le sepsis » (17), P Marik et R Bellomo s'appuient sur les résultats de ces études et sur la physiologie du sepsis pour remettre en cause l'EGDT.

La physiopathologie du sepsis est caractérisée par une vasoplégie avec perte du tonus artériel, une vasodilatation veineuse, une séquestration sanguine dans le compartiment non contraint et des altérations de la fonction ventriculaire cardiaque avec une réduction de la compliance et une précharge dépendance réduite (26–29). Ces données suggèrent que la composante principale du sepsis n'est pas une volémie abaissée. Une prise en charge comportant un apport de fluides important ne serait donc pas adaptée chez la plupart des patients.

Par ailleurs, il existe des preuves sur les effets néfastes de la réanimation liquidienne agressive. Une revue récente (30) met en perspective les données décrivant les complications associées à la surcharge de fluides chez les patients critiques. Elle propose des stratégies de gestion des fluides visant à prévenir la surcharge hydrique chez les patients en état de choc hémodynamique. La surcharge hydrique est une complication courante de la réanimation. Elle est associée à l'augmentation des coûts hospitaliers (31), de la morbidité au cours du syndrome de détresse respiratoire aiguë (32,33) et de la mortalité au cours du choc septique (34,35). La fuite liquidienne dans l'espace interstitiel entraîne un œdème au sein d'organes vitaux et augmente ainsi le risque de dysfonction d'organes. Les signes de surcharge hydrique ont été observés dans les systèmes pulmonaire, cardiaque, rénal et gastro-intestinal. Une augmentation de l'eau pulmonaire extravasculaire nuit à l'oxygénation : elle est associée à la mortalité (36). L'augmentation du liquide interstitiel rénal peut réduire le flux sanguin capillaire et entraîner une ischémie rénale, ce qui peut provoquer ou aggraver une insuffisance rénale aiguë (IRA) (37). La surcharge liquidienne pourrait également aggraver la fonction

myocardique et hépatique, altérer la coagulation, retarder la cicatrisation et favoriser l'hyperpression intra-abdominale (30).

Ces données suggèrent une approche conservatrice plus physiologique, incluant un monitoring hémodynamique, pour guider les thérapeutiques liquidiennes chez les patients de réanimation (30) en particulier en cas de sepsis ou choc septique (17). Le bénéfice potentiel de l'expansion volémique doit être mis en balance avec le risque d'aggraver l'œdème pulmonaire et tissulaire. Déterminer la position du patient sur sa propre courbe de Franck-Starling et donc diagnostiquer sa réponse au remplissage vasculaire devient alors un enjeu majeur. Si aucun test n'est réalisé, l'augmentation du débit cardiaque suite à un remplissage vasculaire ne se produit que chez la moitié des patients (38).

Pour prédire la réponse au remplissage vasculaire, la pression veineuse centrale a été utilisée pendant des décennies mais s'est révélée peu fiable. Des méthodes dynamiques, plus efficaces, ont été développées pour tester la précharge dépendance (9,39). Nous allons décrire ces principales méthodes dans la partie qui va suivre.

1.1.2. Les indices statiques de précharge

Les indices statiques sont nombreux. Ils reflètent la valeur de la précharge en condition statique (généralement lors de la période télé-expiratoire). On peut citer :

- La pression veineuse centrale (PVC), mesurée grâce à une voie veineuse centrale.
- La pression artérielle pulmonaire d'occlusion (PAPO), le volume télédiastolique du ventriculaire droit indexé (VTDVDI), tous deux mesurés grâce à un cathéter artériel pulmonaire.
- Le volume télédiastolique global indexé (VTDGI), mesuré par thermodilution transpulmonaire.
- Les indices doppler évaluant les pressions de remplissage, la surface télédiastolique du ventricule gauche, mesurés par échographie cardiaque.
- Le temps d'éjection ventriculaire gauche corrigé, mesuré par doppler œsophagien.

Ces différents marqueurs sont de bons indicateurs de la précharge cardiaque et des déterminants de la fonction cardiaque. Cependant, la PVC ne prédit pas la réponse au remplissage vasculaire (40)(41), au mieux elle en est un marqueur très médiocre (9). Ce constat est également valable pour les autres indices statiques de précharge (42).

La suite de l'exposé présente les indices dynamiques. Nous séparerons ceux qui sont basés sur les variations de la pression intrathoracique des autres.

1.1.3. Les indices dynamiques basés sur les variations de la pression intrathoracique

1.1.3.1. Variations de la pression intrathoracique pendant un cycle respiratoire

Le concept de ces indices repose sur l'étude de la variabilité respiratoire de certains paramètres hémodynamiques. Cette interaction cœur-poumons décrite sous ventilation mécanique est synthétisée sur la figure 5.

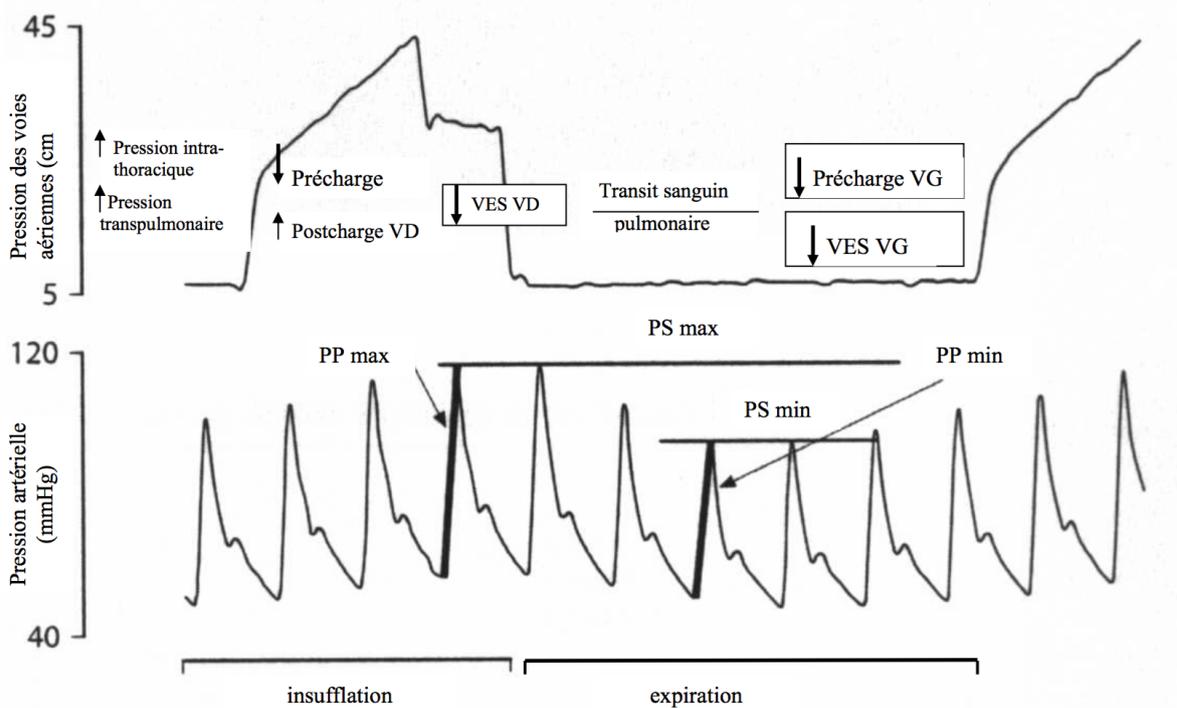


Figure 5 Effets des variations cycliques de la pression intrathoracique induite par la ventilation mécanique sur les conditions de charge ventriculaire et la courbe de pression artérielle. VD : ventricule droit ; VG : ventricule gauche ; VES : volume d'éjection systolique ; PP : pression artérielle pulsée ; PS : pression artérielle systolique ; max : valeur maximale sur un cycle respiratoire ; min : valeur minimale sur un cycle respiratoire (F. Saulnier, Réanimation, 2004)(43)

Lors de l'insufflation, l'augmentation de la pression intrathoracique provoque une baisse de la précharge ventriculaire droite, tandis que l'augmentation de la pression transpulmonaire entraîne une augmentation de la postcharge ventriculaire droite. Ces deux mécanismes

entraînent une diminution du volume d'éjection ventriculaire droit, maximale en fin d'insufflation. Ce phénomène est d'autant plus marqué que le ventricule droit est en situation de précharge-dépendance. Cette diminution du volume d'éjection ventriculaire droit a un effet sur le ventricule gauche deux à trois battements plus tard qui correspondent au temps de transit sanguin intrapulmonaire. On observe une diminution de la précharge et du volume d'éjection ventriculaire gauche au moment de l'expiration. Pareillement, ce phénomène est d'autant plus marqué que le ventricule gauche est en situation de précharge-dépendance. Différents outils permettent de mettre en évidence cette interaction cœur-poumons, on peut citer :

- La variation respiratoire de la pression dans l'oreillette droite (Δ POD) mesurée par une VVC. Sa diminution d'au moins 1 mmHg pourrait prédire la réponse à l'expansion volémique (44). Elle est également utilisable en ventilation spontanée.
- La variation de la pression pulsée (PPV), comme le Δ Down que nous ne développerons pas, s'appuie sur la variabilité respiratoire de la pression artérielle systolique, mesurée via un cathéter artériel. En calculant la différence entre les pressions artérielles systoliques maximale et minimale au cours d'un cycle respiratoire, on obtient une valeur de PPV absolue ou relative (%).

Variabilité de la pression artérielle pulsée (Δ PP)

$$\Delta PP (\%) = \frac{PP \text{ max} - PP \text{ min}}{(PP \text{ max} + PP \text{ min})/2} \times 100$$

La PPV prédit une augmentation du débit cardiaque au remplissage vasculaire (45). Ces résultats sont confirmés par plusieurs méta-analyses (9,46,47) qui retiennent un seuil de 12% (intervalle interquartile 10-13%) (47). La « zone grise » se situe entre 9 et 13% de variation. Dans cette zone, la sensibilité ou la spécificité sont inférieures à 90% (48) ; cela pourrait concerner 24% des valeurs habituellement rencontrées.

- La variation du volume d'éjection systolique (VES), ou *stroke volume variation* (SVV), s'appuie sur le même principe et la même formule. En ventilation positive, une diminution de la précharge ventriculaire droite, transmise au ventricule gauche, entraîne une baisse du VES. La mesure de la SVV nécessite une mesure continue du VES, principalement grâce à la thermodilution transpulmonaire couplée à l'analyse de l'onde de pouls. Comme pour le PPV, le seuil de positivité est à 13% de variation (9).

Variabilité du volume d'éjection systolique (Δ VES)

$$\Delta\text{VES} (\%) = \frac{\text{VES max}^* - \text{VES min}^*}{\text{VES moyen}^*} \times 100$$

- La variabilité respiratoire du pic de vélocité aortique (Δ Vpeak), mesurée par échographie transthoracique ou œsophagienne. Sur le même principe hémodynamique que la PPV et la SVV, les valeurs maximales de vélocité aortique (Vpeakmax) et les valeurs minimales de vélocité aortique (Vpeakmin) sont déterminées sur un cycle respiratoire. La valeur seuil de 12% est retenue (49).
- L'échographie permet aussi d'explorer la veine cave et de mesurer les variations respiratoires de ses dimensions. Les variations respiratoires des diamètres de la veine cave supérieur et inférieur (respectivement Δ SVC et Δ IVC) sont deux indices également basés sur les interactions cœur-poumon. Leur fonctionnement est différent des trois indices précédents (PPV, SVV et Δ Vpeak). Les variations de la pression intrathoracique induites par la ventilation mécanique influent sur le diamètre de la veine cave, sur sa collapsibilité, et ce d'autant plus que le volume sanguin contraint est faible. Les seuils retenus varient selon les études, nous en proposons deux pour chaque indice. Concernant l'indice de collapsibilité de la VCS, mesuré en ETO, on peut retenir les seuils de 21% (50) et 36% (51). Concernant la variation respiratoire du diamètre de la VCI, on peut retenir les seuils de 8% (50) et 12% (52). Ces deux indices ont été moins étudiés que la PPV et la SVV et semblent moins performants à prédire la précharge dépendance (53). De la même manière l'étude des variations de la VCS semble plus fiable que celle de la VCI (53).
- La variation respiratoire du flux sanguin aortique (Δ ABF pour *aortic blood flow*) peut également être mesurée grâce au doppler œsophagien et exprimée en pourcentage, selon la formule :

$$\Delta\text{ABF}(\%) = \frac{(\text{ABF}_{\text{max}} - \text{ABF}_{\text{min}})}{[(\text{ABF}_{\text{max}} + \text{ABF}_{\text{min}})/2]} \times 100$$

Le seuil retenu est de 18% (54).

1.1.3.2. Variations de la pression intrathoracique lors d'une pause respiratoire

Le test d'occlusion de fin d'expiration (EEO test pour *end-expiratory occlusion test*). La ventilation en pression positive gêne le retour veineux de manière cyclique. En empêchant ce phénomène pendant quelques secondes, en interrompant la ventilation mécanique, on augmente la précharge cardiaque de manière transitoire. On considère une durée de 15 secondes minimum pour que le test soit valide. La précharge dépendance est mise en évidence à partir d'une augmentation du débit cardiaque de 5% mesuré par analyse de l'onde de pouls (55). A noter que lors d'une étude sur les patients atteints de SDRA, le test est resté valide à un niveau PEEP de 5 cmH₂O ainsi que de 15 cmH₂O (56). Le reflet de la manœuvre sur la courbe de pression artérielle est montré sur la figure 6.

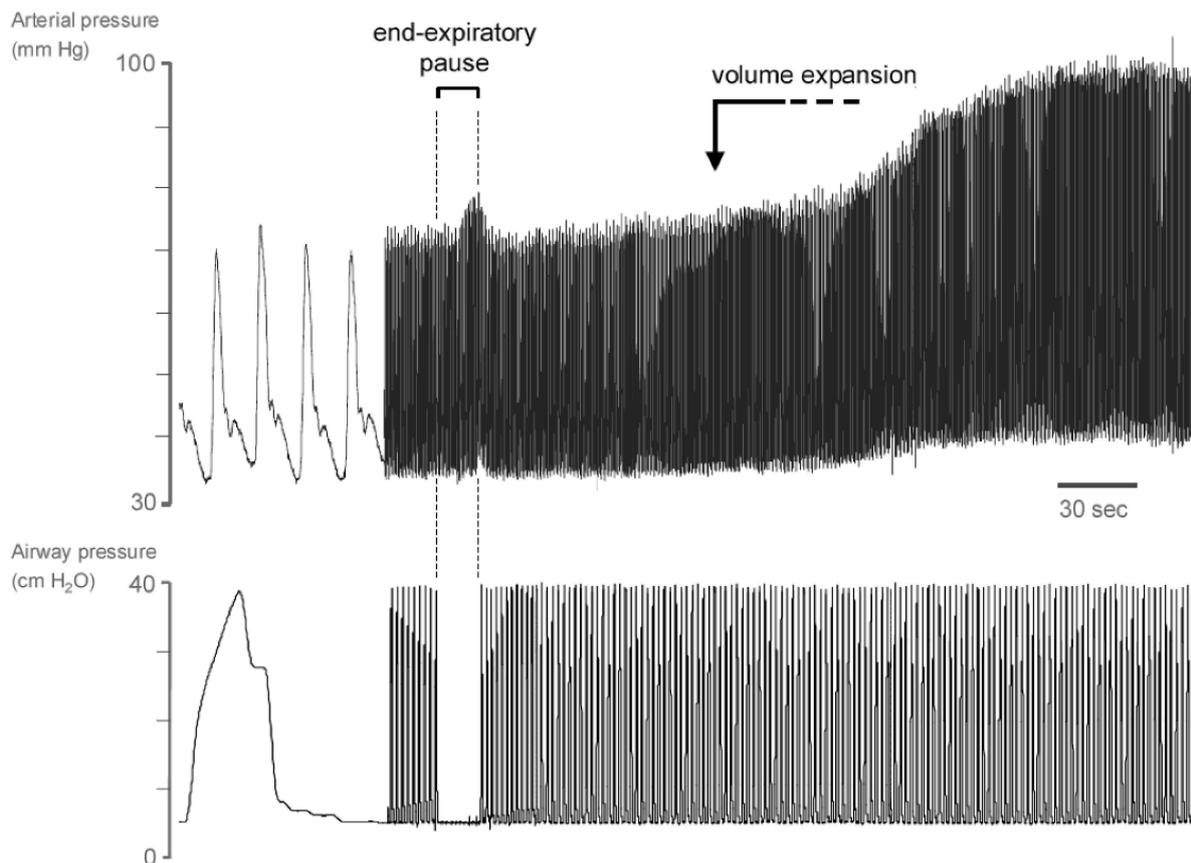


Figure 6 Enregistrement typique de la courbe de pression artérielle avant et pendant le test d'occlusion de fin d'expiration, puis avant et pendant l'expansion du fluide. L'augmentation de la pression pulsée induite par l'expansion volémique a été précédée d'une augmentation de 11% de la pression artérielle lors de l'occlusion de fin d'expiration (55).

1.1.3.3. Variations de la pression intrathoracique lors d'une modification de la PEEP

Dans une étude récente menée chez des patients de chirurgie cardiaque, la réponse au remplissage vasculaire a été prédite par les effets hémodynamiques d'une augmentation soudaine de la pression positive de fin d'expiration (PEEP) de 5 à 10 cmH₂O. Les effets de variation de PEEP ont été quantifiés en mesurant l'élimination du CO₂ (VCO₂) qui a été utilisée comme substitut du débit cardiaque. Le seuil retenu est une diminution du VCO₂ de 11% (57).

Une méthode similaire a été étudiée chez les patients en choc septique, en faisant passer la PEEP de 10 à 20 cm H₂O. La variation de pression artérielle moyenne était observée, avec pour valeur seuil, une baisse de 8 à 10 % (58).

1.1.3.4. Variations de la pression intrathoracique lors d'une modification du volume courant

Le test de variation systolique respiratoire (RSVT) quantifie la diminution de la pression systolique en réponse à une manœuvre standardisée consistant en trois respirations mécaniques consécutives avec une augmentation de la pression des voies aériennes. Les valeurs minimales de la pression artérielle systolique (PAS) en réponse à chaque inspiration sont enregistrées. Puis on calcule la pente de la diminution de PAS en fonction de la pression inspiratoire (figure 7). Le seuil retenu est une pente d'au moins 0.51 mmHg/cm H₂O (59).

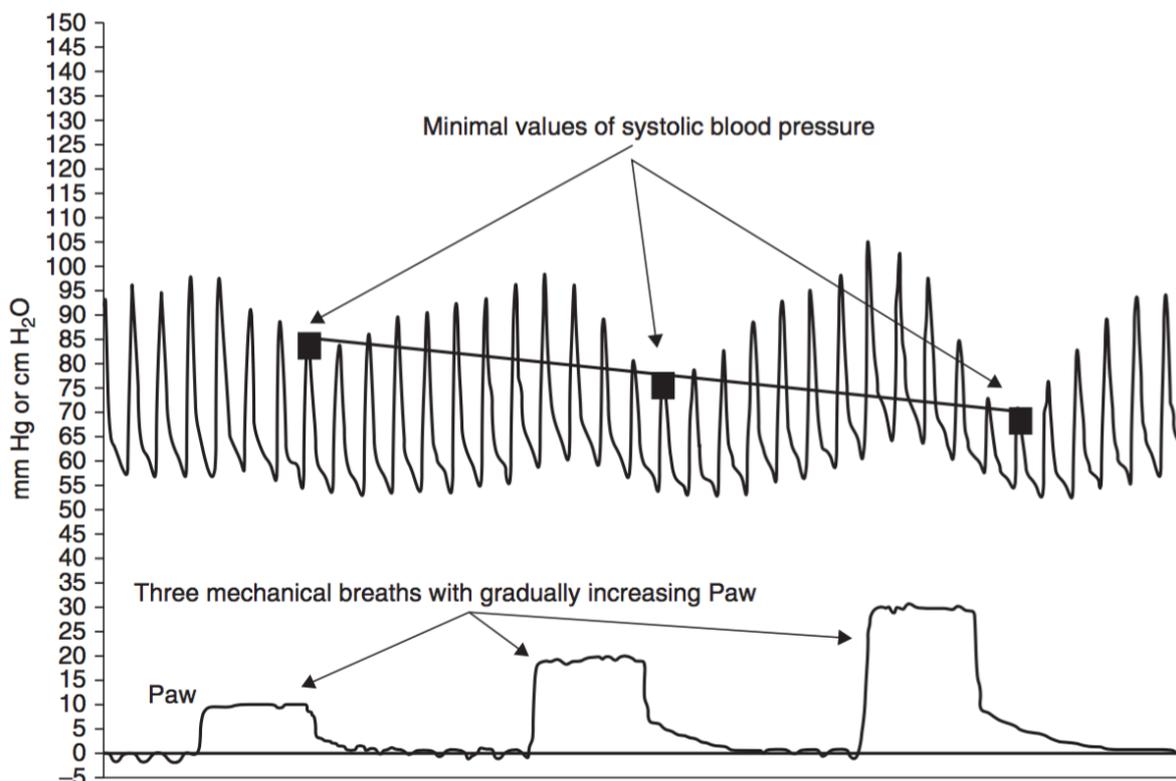


Figure 7 Réponse de la pression artérielle à la RSVT. Trois inspirations consécutives en pression contrôlée sont réalisées avec des pressions inspiratoires de 10, 20 et 30 cm H₂O(59).

1.1.3.5. Variations de la pression intrathoracique lors d'une manœuvre de recrutement

Les changements de volume d'éjection systolique induits par une manœuvre de recrutement alvéolaire pourraient prédire la réponse au remplissage vasculaire chez les patients intubés, au bloc opératoire (60). La valeur seuil retenue est une baisse de 30% du VES, mesurée par analyse de l'onde de pouls.

1.1.4. Les indices dynamiques non basés sur les interactions cœur-poumon

1.1.4.1. Le lever de jambes passif

Certains indices dynamiques de précharge dépendance ne s'appuient pas sur la physiologie des interactions cœur-poumon. Nous en présentons plusieurs. Le lever de jambes passif (PLR pour *passive leg rising*) est un test qui prédit si un remplissage vasculaire sera ou non suivi

d'une augmentation du débit cardiaque (61). Il mime l'effet d'un remplissage en transférant environ 300 ml de sang veineux du bas du corps vers le cœur droit. La méthode pour réaliser ce test varie selon les études. X Monnet en propose une qui suit cinq règles (62) (figure 8) :

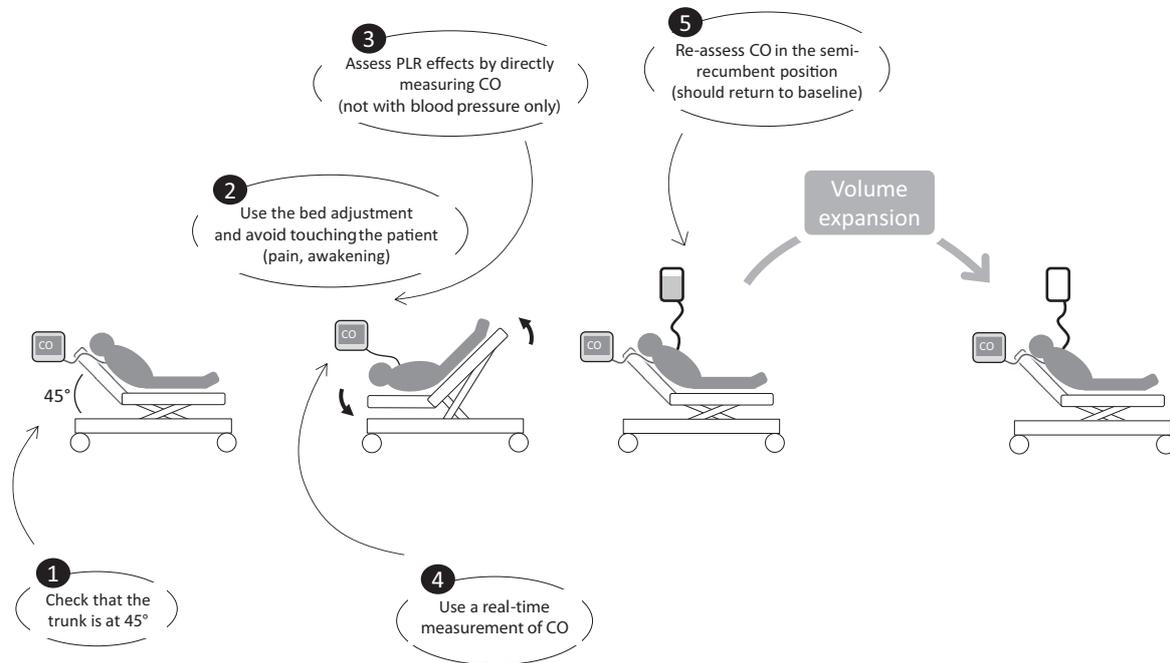


Figure 8 La méthode optimale pour le lever de jambes passif (62). CO, débit cardiaque ; PLR, lever de jambes passif. 1) Vérifier que le tronc est incliné à 45°, 2) Utiliser le réglage du lit et évitez de toucher le patient (douleur, réveil, 3) Évaluer les effets du PLR en mesurant directement le débit cardiaque (Pas seulement avec la tension artérielle) 4) Utiliser une mesure en temps réel du débit cardiaque 5) Réévaluer le débit cardiaque en position semi-assise (devrait revenir à la ligne de base).

On peut séparer deux sortes de PLR, l'une au cours de laquelle on observe les variations de débit cardiaque ou ses substituts, l'autre où l'on observe des modifications du PPV ou de ses dérivés. Il existe deux principales techniques permettant de monitorer le débit cardiaque battement par battement : la thermodilution transpulmonaire couplée à l'analyse de l'onde de pouls et l'échographie cardiaque transthoracique (9). Le seuil de variation permettant de prédire la réponse au remplissage vasculaire est de 10% (63). Les changements induits par le lever de jambes passif sur la variation de pression pulsée (PPV) ont également été étudiés. La précision du PLR pour prédire la réponse au remplissage vasculaire est alors moins bonne, avec un seuil retenu autour de 12% (63).

1.1.4.2. La compression abdominale

La compression abdominale pourrait prédire la précharge dépendance chez l'enfant en insuffisance circulatoire aiguë. On observe la variation du volume d'éjection systolique durant

une compression abdominale (Δ VES-CA) standardisée. Le seuil le plus discriminant est une Δ VES-CA de 9% (64) (65).

1.1.4.3. Le mini-fluid challenge

Dans une étude où un « *mini-fluid challenge* » (mini test de remplissage) a été réalisé avec 100 mL de colloïde, une variation de l'intégrale temps-vitesse à la sortie du ventricule gauche (mesurée par échographie) a prédit une réponse au remplissage vasculaire. Le seuil statistique était une augmentation de 6% de l'intégrale temps-vitesse. Cependant, celui-ci étant inférieur à la précision de l'échocardiographie, les auteurs ont suggéré un seuil de 10% (66).

1.1.4.4. La variation du temps de diastole post ESA.

L'utilisation des extrasystoles pour prédire la réponse au remplissage vasculaire a été proposée. Le battement postectopique a une précharge accrue en raison de la pause compensatoire, mais aussi de la reprise d'un battement cardiaque sinusal. Cela provoque un déplacement vers la droite sur la courbe Frank-Starling. Ce battement postectopique peut donc être considéré comme un *fluid challenge* sur un battement cardiaque du ventricule gauche en le comparant à une moyenne des battements précédents (67). En post-opératoire de chirurgie cardiaque, le seuil optimal retrouvé était une augmentation de la pression artérielle systolique de 5%(68). En étudiant la période de pré-éjection, soit l'intervalle de temps entre le début de la dépolarisation ventriculaire (signal ECG) et le début de l'éjection ventriculaire gauche (cathéter radial), la même équipe montre qu'une diminution de la période de pré-éjection de 19 ms lors du battement post-ectopique prédisait une réponse au remplissage vasculaire (68).

1.1.5. Limites et utilisation pratique de ces tests

Comme nous l'avons vu, les tests statiques ne doivent plus être utilisés pour évaluer la précharge dépendance. En revanche, les tests dynamiques permettent de prédire la réponse au remplissage vasculaire, dans certaines conditions, avec les limites suivantes :

- Concernant la PPV, la SVV et probablement leurs dérivées, on retiendra : les mouvements respiratoires spontanés, les arythmies cardiaques, un volume courant bas ou une compliance pulmonaire altérée, le thorax ouvert, une pression intra-abdominale

élevée, une fréquence respiratoire très élevée ($FC/FR < 3,6$), une insuffisance cardiaque droite (42).

- Les variations du diamètre de la veine cave (supérieure et inférieure) ne peuvent pas être utilisées en cas de ventilation spontanée, à faible volume courant, ou lors d'une faible compliance pulmonaire.
- Le lever de jambes passif a peu de limitations, on peut retenir les fractures du rachis et du bassin ainsi que l'hyperpression intra-abdominale.
- Le test d'occlusion de fin d'expiration est dépendant de la tolérance du patient à une pause respiratoire d'au moins 15 secondes.

En 2016, X Monnet résume certaines méthodes pour prédire la précharge dépendance avec leurs valeurs seuils et leurs limitations (tableau 1). Il propose également un arbre décisionnel pour faire le choix entre les techniques d'évaluation de la précharge dépendance en fonction de l'état du patient et des techniques de surveillance disponibles (figure 9).

Tableau 1 Résumé des méthodes pour prédire la précharge dépendance, avec leurs valeurs seuils et leurs limitations (42)

Method	Threshold	Main limitations
Pulse pressure/stroke volume variations [22]	12%	Cannot be used in case of spontaneous breathing, cardiac arrhythmias, low tidal volume/ lung compliance
Inferior vena cava diameter variations [44]	12%	Cannot be used in case of spontaneous breathing, low tidal volume/lung compliance
Superior vena caval diameter variations [44]	36%*	Requires performing transesophageal Doppler Cannot be used in case of spontaneous breathing, low tidal volume/lung compliance
Passive leg raising [55]	10%	Requires a direct measurement of cardiac output
End-expiratory occlusion test [75]	5%	Cannot be used in non-intubated patients Cannot be used in patients who interrupt a 15-s respiratory hold
"Mini"-fluid challenge (100 mL) [84]	6%**	Requires a precise technique for measuring cardiac output
"Conventional" fluid challenge (500 mL) [81]	15%	Requires a direct measurement of cardiac output Induces fluid overload if repeated

* Thresholds from 12 to 40% have been reported

** 10% is more compatible with echography precision. Citations indicate the most important reference regarding the test

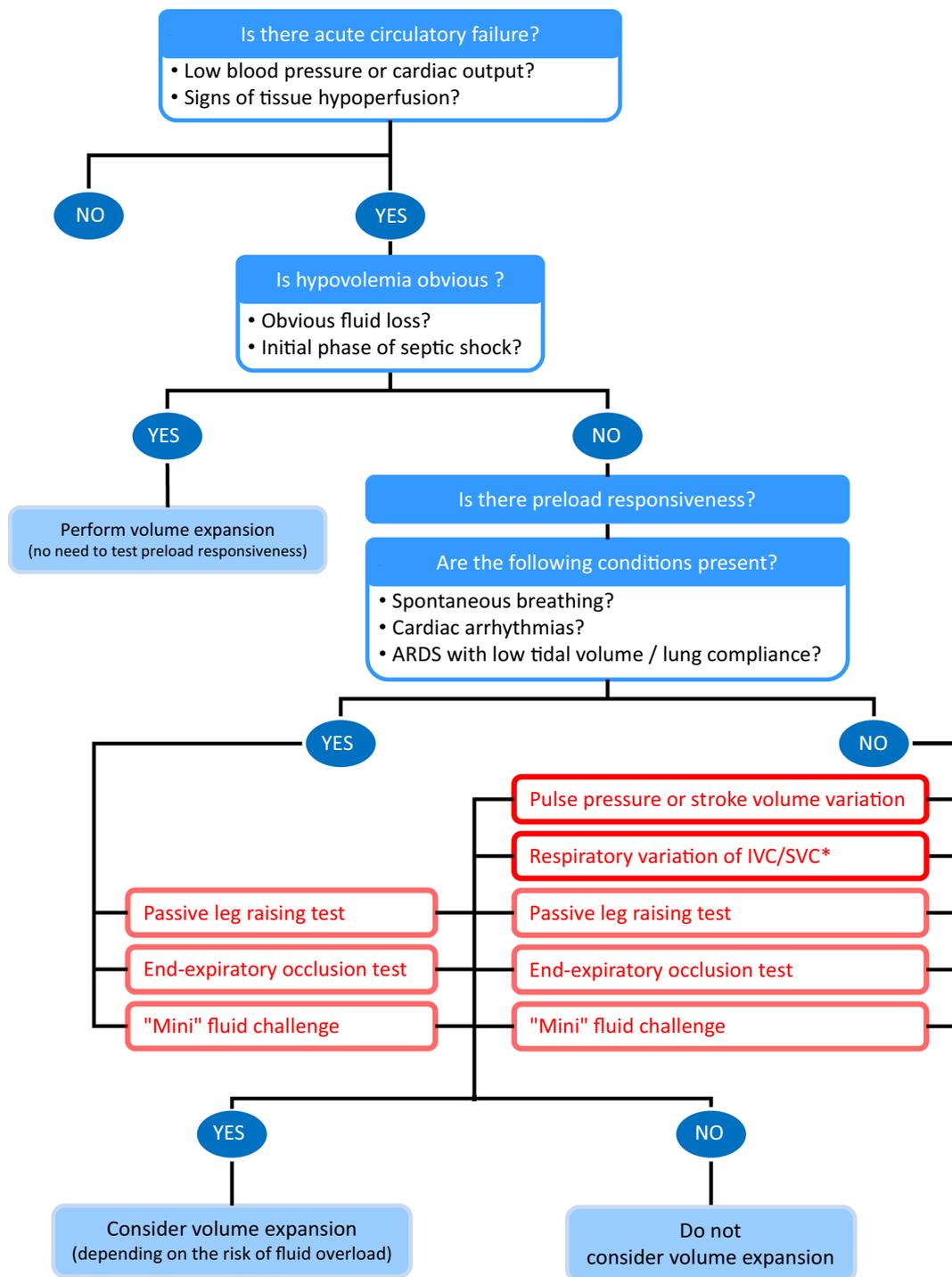


Figure 9 Stratégie d'utilisation du remplissage vasculaire. * La variation des diamètres inférieurs / supérieurs de la veine cave peut être utilisée en cas d'arythmie cardiaque. ARDS : Syndrome de détresse respiratoire aiguë, IVC : Veine cave inférieure, SVC : veine cave supérieure.

Nous venons de présenter un certain nombre d'indices pour diagnostiquer la précharge dépendance. Il en existe d'autres, le but n'est pas d'être exhaustif mais d'avoir une vision

globale du sujet que l'on traite. La précharge dépendance est une problématique quotidienne de l'anesthésiste-réanimateur. C'est un sujet qui est donc très sensible aux différentes publications qui le traitent. On remarque que la littérature sur le sujet est foisonnante. A titre d'exemple, la recherche « *fluid responsiveness* » dans PubMeb retrouve 3333 entrées (recherche réalisée le 27/04/2017). En effet, si nous utilisons des indices de précharge dépendance au quotidien, il nous paraît intéressant de savoir comment ils ont été évalués, quelle est la méthodologie des études qui les ont validés et surtout quelle est la qualité de ces études. D'où la question majeure pour nous qui est celle de **la qualité de la littérature relative aux études diagnostiques de précharge dépendance**.

1.2. Les études diagnostiques

1.2.1. De quoi parlons-nous ?

Avant toute chose, précisons que par la suite nous utiliserons le terme « d'étude diagnostique » pour faire référence à « une étude sur la précision diagnostique ».

Dans les études diagnostiques, les résultats d'un test évalué (parfois plusieurs) sont comparés aux résultats d'un test de référence. Les deux tests sont réalisés chez des patients soupçonnés d'avoir la condition d'intérêt (souvent une maladie). Après avoir donné quelques définitions des termes (69)(70), nous expliquerons ensuite l'évaluation des tests diagnostiques.

- Le terme test est défini comme toute méthode permettant d'obtenir des informations supplémentaires sur l'état de santé d'un patient. Il s'agit principalement d'informations données par l'anamnèse, l'examen physique, des prélèvements biologiques ou histologiques, de l'imagerie, des tests fonctionnels.
- La condition d'intérêt ou la condition cible peut se référer à une maladie ou à toute autre condition pouvant entraîner une action clinique (nouveaux tests, traitements...)
- Le test de référence est considéré comme la meilleure méthode disponible pour établir la présence ou l'absence de la condition d'intérêt. Il peut être unique ou une combinaison de plusieurs tests. Il peut s'agir d'un suivi clinique, de prélèvements biologiques ou histologiques, d'imagerie ou de tests fonctionnels.
- L'exactitude (*accuracy* en anglais) est un concept qui recouvre à la fois la justesse et la fidélité. Un test est d'autant plus exact que les résultats qu'il indique coïncident avec la « valeur vraie » (par définition théorique) que l'on cherche à mesurer. L'exactitude

ne s'exprime pas par une valeur chiffrée, il s'agit d'une appréciation qualitative des résultats, elle est liée à la justesse et à la fidélité. Elle est souvent définie par l'erreur de mesure qui s'exprime en unité de grandeur (erreur absolue) ou en pourcentage (erreur relative).

- La justesse (*trueness*) est définie par l'erreur de justesse, c'est à dire l'erreur globale résultant de toutes les causes, pour chacun des résultats de mesure, prises isolément. C'est donc l'aptitude à donner des résultats qui ne contiennent pas d'erreurs.
- La fidélité (*precision*) est l'aptitude d'un test à donner des mesures exemptes d'erreurs accidentelles. Elle définit la dispersion des résultats. Elle représente la probabilité qu'une seule mesure soit représentative du résultat moyen (obtenu par une infinité de mesures).

Ces trois notions sont représentées dans la figure 10

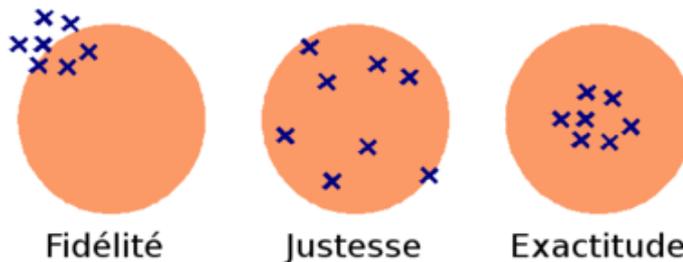


Figure 10 Représentation graphique de la fidélité, de la justesse et de l'exactitude (71).

- Ce que l'on va appeler par la suite la précision diagnostique correspond en fait à l'exactitude, traduite en anglais par *accuracy*. Elle peut être exprimée de plusieurs façons : *sensibilité, spécificité, rapport de vraisemblance, odds ratios, aires sous la courbes ROC* (de l'anglais *receiver operating characteristic* également appelée courbe sensibilité/spécificité).

1.2.2. Evaluation des tests diagnostiques : les grands principes

Le nombre de tests diagnostiques disponibles dans la pratique médicale a augmenté rapidement au cours des dernières décennies et devrait continuer d'augmenter (72). Par conséquent, il existe une crainte quant à la surutilisation de ces tests, avec des conséquences potentiellement négatives pour la santé et une hausse des coûts (73). Pour les éviter, nous avons besoin d'une évaluation minutieuse des nouveaux tests (74).

Cette évaluation de la précision diagnostique de chaque test ou procédure diagnostique doit répondre à un certain nombre de questions : Le test sera-t-il utilisé dans un cadre clinique ou de dépistage ? Dans quelle partie du « parcours clinique » ce test sera-t-il placé ? Quelle sera la capacité du test à discriminer la maladie de l'absence de maladie ? Quelle est la variabilité des résultats liés à l'événement étudié ? De quelle capacité discriminative avons-nous besoin dans notre pratique clinique ? Comment réaliser le test dans de bonnes conditions pour qu'il soit efficace ? L'utilisateur doit connaître ces informations pour que le test soit utilisé en pratique courante. Cette évaluation nécessite des règles que nous allons développer.

1.2.2.1. Méthodologie des études diagnostiques

- Dans un premier temps, l'objectif de l'étude doit être clair. Il s'agit de : « Déterminer la valeur informationnelle apportée par un test dans une démarche diagnostique. »(75)
- Le nouveau test est comparé à un test de référence. Il cherche à identifier les malades (test positif) et les non-malades (test négatif). Ces deux groupes de patients sont identifiés par un test de référence appelé *gold standard*. Le test de référence doit pouvoir affirmer le diagnostic et cela de manière reconnue par tous. Il doit idéalement associer fiabilité, validité et détermination *a priori* (75).
- La méthode de sélection des patients doit être adaptée à la question posée. Les résultats de l'étude s'appliquent à un échantillon de patients. Pour pouvoir extrapoler les résultats, l'échantillon doit être pertinent. Cette sélection des patients, définie par les critères d'inclusion et d'exclusion, influence la validité externe de l'étude. Il en va de même pour le lieu de l'étude qui déteint sur la prévalence de la maladie, elle-même modifiant la performance des tests diagnostiques. Bien sûr la sélection *a posteriori*, ou l'analyse de sous-groupes de patients, influe sur la validité du test.
- La fréquence de la maladie dans l'échantillon étudié doit être connue et en accord avec les données épidémiologiques connues. Si ce n'est pas le cas, on assiste à une modification de la validité externe.
- La normalité ou la non-maladie doit être définie. En général, elle est « définie par l'identification d'une valeur seuil au-delà de laquelle la maladie est présente avec une probabilité connue » (75), figure 11.

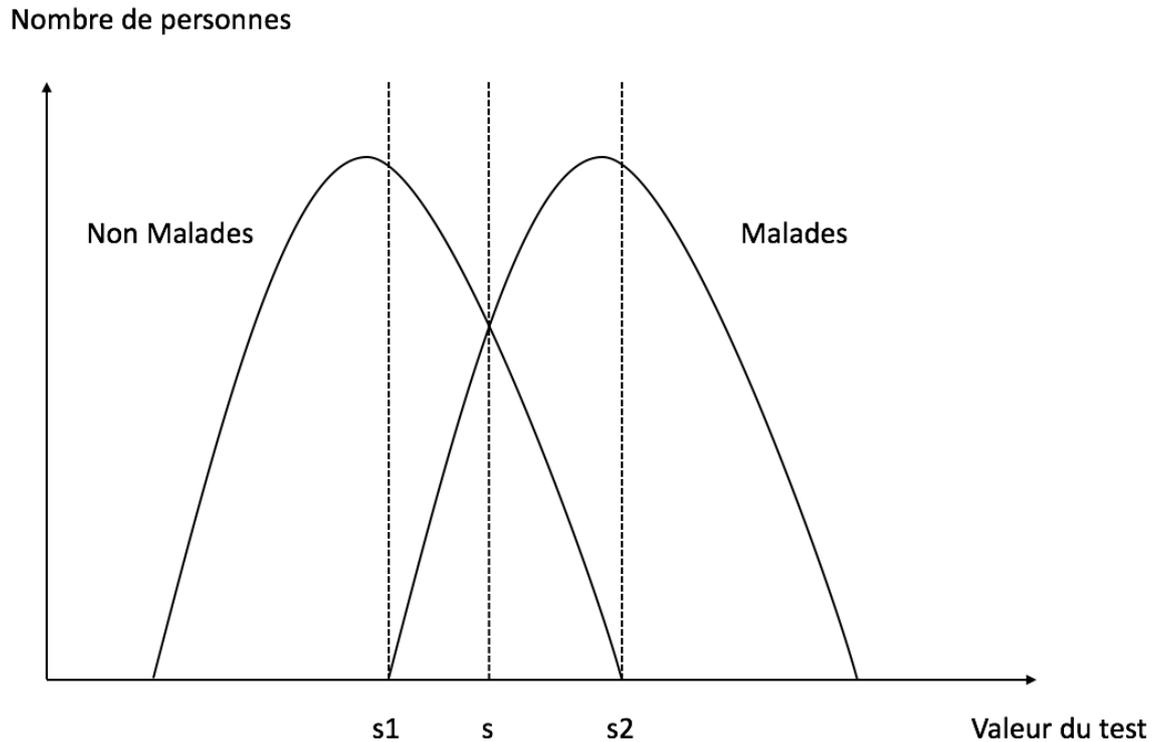


Figure 11 Répartition des malades et les non-malades en fonction d'un seuil. On y remarque une zone de recouvrement à l'intérieur de laquelle, pour une même valeur du test, le patient peut être malade ou non.

Plusieurs valeurs de seuil peuvent être choisies :

s : identifie un maximum de résultats corrects

$s1$: élimine la présence de la maladie

$s2$: certifie la présence de la maladie

- L'aveugle est une composante essentielle à la limitation des biais. Le résultat du nouveau test doit être inconnu aux personnes réalisant le test de référence, et vice-versa.
- Le nombre d'échantillons de patients (en général un ou deux) est principalement conditionné par la fréquence de la maladie. En cas de maladie fréquente, un seul échantillon de patients représentatif de la population étudiée est idéal. Le test de référence identifie alors les malades des non-malades. En cas de maladie rare, les chercheurs ont généralement recours à deux échantillons indépendants, le premier comportant les malades (test de référence positif), l'autre les non-malades. Dans ce deuxième cas, la prévalence ne pouvant être calculée, les indicateurs qui en dépendent ne pourront pas être utilisés.

- Quel que soit le résultat du nouveau test, le test de référence doit être réalisé.

1.2.2.2. Analyse des résultats des études diagnostiques

Il existe quatre indicateurs de base qui permettent d'en construire d'autres (tableau 2).

Tableau 2 Indicateurs de base des études diagnostiques, VP : vrai positif, FP : faux positif, FN : faux négatif, VN : vrai négatif.

	Malades	Non malades
Test positif	VP	FP
Test négatif	FN	VN

Les vrais positifs (VP) représentent le nombre d'individus malades avec un test positif,

Les faux positifs (FP) représentent le nombre d'individus non malades avec un test positif,

Les faux négatifs (FN) représentent le nombre d'individus malades avec un test négatif,

Les vrais négatifs (VN) représentent le nombre d'individus non malades avec un test négatif.

Ces quatre paramètres permettent de mesurer la validité intrinsèque d'un test par sa sensibilité et sa spécificité. Ces deux valeurs sont indépendantes de la prévalence de la maladie dans la population évaluée.

La sensibilité correspond à la probabilité que le test soit positif si la maladie est présente,

$$\text{soit : } \frac{VP}{VP+FN}$$

La spécificité correspond à la probabilité d'obtenir un test négatif chez les non-malades, soit :

$$\frac{VN}{VN+FP}$$

La probabilité de faire une erreur de classement correspond au nombre de tests erronés par

$$\text{rapport au nombre total de test, soit : } \frac{FP+FN}{VP+FP+VN+FN}$$

La précision diagnostique globale (souvent appelée « précision diagnostique », *diagnostic accuracy*)

correspond à la proportion de sujets correctement classés : $\frac{VP+VN}{VP+FP+VN+FN}$. Pour la même sensibilité et la même spécificité, la précision diagnostique d'un test augmente à mesure que la prévalence de la maladie diminue (72).

Une sensibilité élevée signifie que la proportion de sujets malades identifiée est élevée, permettant d'écarter le diagnostic si le test est négatif. A contrario, une spécificité élevée

signifie que la proportion de sujets non malades identifiée est élevée, permettant d'affirmer le diagnostic si le test est positif. Ces deux valeurs permettent de comparer deux tests. Si un test possède à la fois une sensibilité et une spécificité plus élevées, il sera meilleur que l'autre. Cependant, si l'un des tests est plus sensible mais moins spécifique que l'autre, il faut prendre en compte l'erreur de classement de chacun des tests. Celle-ci est dépendante de la prévalence de la maladie. La figure 12 rend compte de cette situation, le test A est plus sensible mais moins spécifique que le test B.

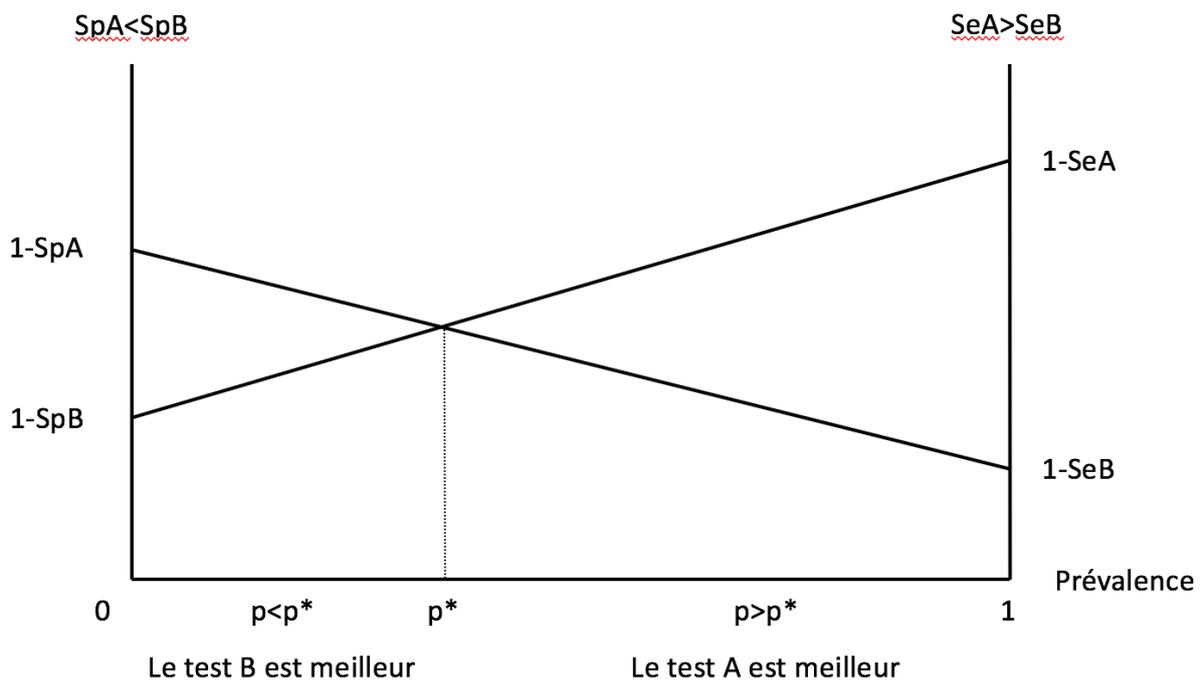


Figure 12 Evaluation de deux tests diagnostiques A et B lorsque $SeA > SeB$ et $SpA < SpB$ en fonction de la prévalence de la maladie. Se : sensibilité, Sp : spécificité, p : prévalence(75).

Pour utiliser un test, nous devons donc connaître la probabilité que le test donne le diagnostic correct. La sensibilité et la spécificité ne nous donnent pas cette information. Nous devons pour cela utiliser les valeurs prédictives d'un test, elles indiquent la probabilité de la présence ou de l'absence de la condition étudiée une fois le résultat du test connu (76).

La valeur prédictive positive est la probabilité que la condition soit présente lorsque le test est positif, soit : $\frac{VP}{VP+FP}$

La valeur prédictive négative est la probabilité que la condition ne soit pas présente lorsque le test est négatif, soit : $\frac{VN}{VN+FN}$

La valeur prédictive est fonction de la sensibilité et de la spécificité du test ainsi que de la prévalence de la condition étudiée.

L'interprétation des valeurs prédictives doit être précautionneuse. En effet, alors que la sensibilité et la spécificité sont des valeurs intrinsèques d'un test, les valeurs prédictives dépendent de la prévalence de la maladie dans la population. Ainsi, pour des valeurs données de sensibilité et spécificité, la valeur prédictive positive du test va s'améliorer d'autant que la maladie est fréquente alors que la valeur prédictive négative va quant à elle s'améliorer d'autant que la maladie est rare.

Le rapport de vraisemblance (RV) est un indicateur qui permet de modifier la probabilité d'un diagnostic en fonction du résultat du test, distinguant une probabilité pré-test (correspondant souvent à la prévalence de la maladie) à la probabilité post-test (probabilité une fois le test réalisé). Il s'agit du rapport entre la probabilité du résultat d'un test spécifique chez les personnes malades sur la probabilité du résultat de ce test chez les personnes non malades (77).

Le RV positif est le rapport entre la probabilité d'obtenir un test positif chez un sujet malade et la probabilité d'obtenir un test positif chez un sujet non malade, soit : $\frac{Se}{(1-Sp)}$. Il mesure l'augmentation de la probabilité d'être malade quand le test est positif.

Le RV négatif est le rapport entre la probabilité d'obtenir un test négatif chez un sujet malade et la probabilité d'obtenir un test négatif chez un sujet non-malade, soit : $\frac{1-Se}{Sp}$. Il mesure la diminution de la probabilité d'être malade quand le test est négatif.

Ainsi, pour affirmer un diagnostic, on choisira un test avec un RV positif élevé. Inversement, on choisira un test avec un RV négatif le plus faible pour éliminer un diagnostic.

On note que les rapports de vraisemblance sont indépendants de la prévalence de la maladie. Le nomogramme de Fagan (figure 13) est un outil permettant d'obtenir la probabilité post-test d'être malade lorsque l'on connaît la probabilité pré-test ainsi que le RV du test. On remarque que pour un test positif, un RV supérieur à 1 renforce la probabilité diagnostique. A l'inverse, pour un test négatif, un RV inférieur à 1 rend le diagnostic moins probable qu'initialement.

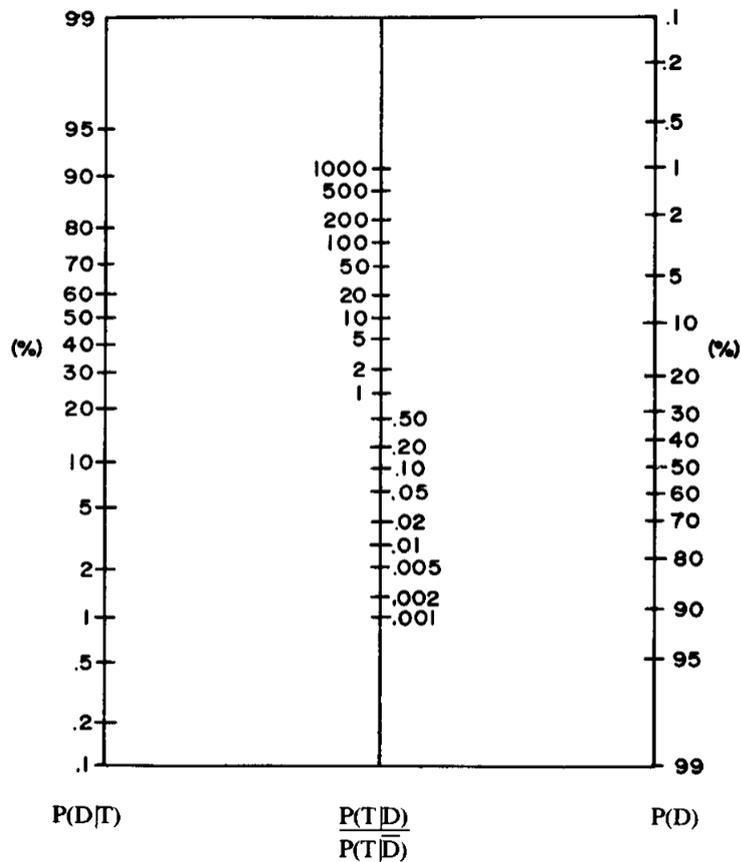


Figure 13 Nomogramme de Fagan (78)

Lorsque le résultat du test est une variable continue, il est essentiel de déterminer le seuil de cette valeur de manière à ce que la sensibilité du test ainsi que sa spécificité soient les meilleurs. A partir de tableaux de contingence réalisés pour chaque valeur du test, on détermine la sensibilité et la spécificité du test pour chacune de ces valeurs. On peut alors tracer une courbe correspondant au couple $Se ; 1 - Sp$ (ou Se : sensibilité et Sp : spécificité) pour chaque valeur du test. Cette courbe ROC (pour *Receiver-Operating Characteristic*) résume l'ensemble de ces valeurs. Elle permet de mesurer la performance du test grâce à l'aire sous la courbe. Plus l'aire sous la courbe est proche de 1 (cadran supérieur gauche), plus la sensibilité et la spécificité sont élevées, plus le test est performant.

La courbe ROC permet de comparer la performance diagnostique de plusieurs tests, toujours au travers de la valeur de l'air sous la courbe de chaque test (figure 14).

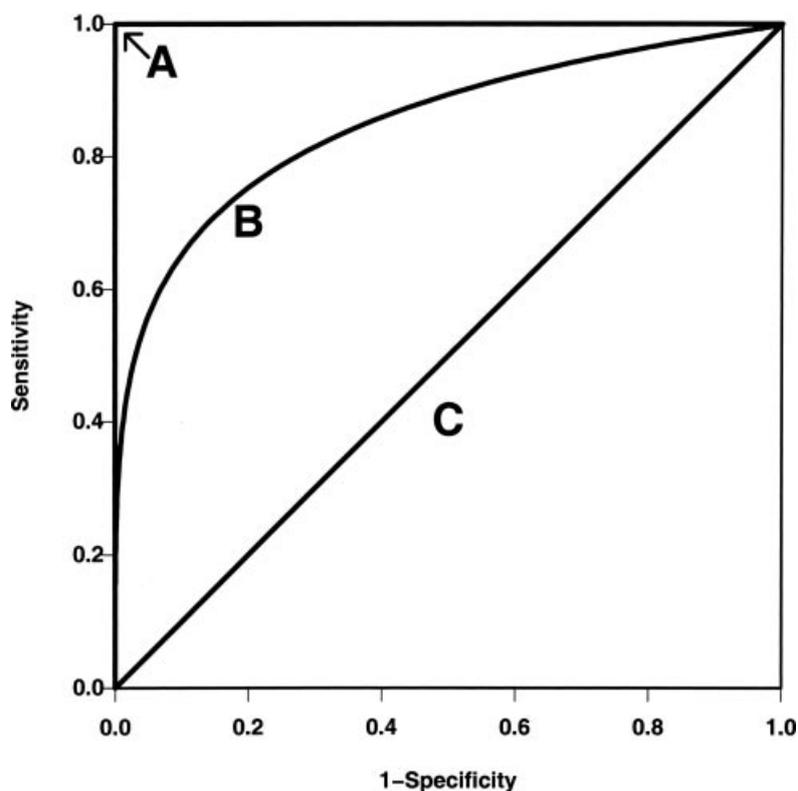


Figure 14 Trois courbes ROC hypothétiques représentant la précision diagnostique du gold standard (lignes A ; AUC = 1), une courbe ROC typique (courbe B, AUC = 0.85) et une ligne diagonale correspondant au hasard (ligne C, AUC 0,5). Au fur et à mesure que la précision du test diagnostique s'améliore, la courbe ROC se déplace vers A et l'AUC s'approche de 1(79).

Enfin, il est primordial de déclarer des mesures de variabilité ou d'incertitude pour les résultats de précision diagnostique avec l'intervalle à 95% (IC 95%).

1.2.2.3. Intérêt et applicabilité du test étudié

Le nouveau test doit apporter un bénéfice au patient par rapport au test de référence. Il est souhaitable de connaître le devenir des patients mal classés (faux positifs et faux négatifs), quelles sont les conséquences d'un mauvais classement : mauvais traitement, complications, coût inutile, etc. Il faut connaître les conditions de réalisation du test afin qu'il soit reproduit dans ces mêmes conditions en pratique courante.

1.3. Qualité du *reporting* dans les études diagnostiques

1.3.1. L'existence de risques potentiels concernant la validité interne et/ou externe des études de précision diagnostique.

Le monde des tests diagnostiques est dynamique. De nouveaux tests sont développés à un rythme rapide et la technologie des tests existants est régulièrement améliorée. Des résultats exagérés ou biaisés tirés d'études diagnostiques mal conçues et mal rapportées peuvent déclencher la dissémination prématurée de ces tests et conduire les médecins à prendre des décisions de traitement incorrectes. Une évaluation rigoureuse des tests diagnostiques avant leur introduction dans la pratique clinique pourrait réduire le nombre de conséquences cliniques indésirables ainsi que limiter les coûts de soins de santé en diminuant les tests inutiles (69).

Dès 1986, G H Guyatt préconise « d'évaluer systématiquement et avant sa mise en œuvre clinique : la validité technologique, les applications possibles, la précision diagnostique, l'incidence sur les personnes et les organismes chargés des soins sanitaires, sur les traitement et sur le devenir de malades » (80) Les tests diagnostiques ne devraient être rendus disponibles que s'ils sont « moins chers, au moins aussi précis que les tests existants, diminuent le recours à d'autres explorations et permettent la mise en œuvre d'un traitement » (80)

Dans une étude publiée en 1995 dans le JAMA, MC Reid et ses collègues évaluent la qualité méthodologique de 112 études diagnostiques publiées dans quatre revues médicales majeures (*Lancet*, *British Medical Journal*, *New England Journal of Medicine*, and *Journal of the American Medical Association (JAMA)*) entre 1978 et 1993. En s'intéressant à la présence de sept critères méthodologiques ainsi qu'à l'évolution de leur *reporting* au fil des ans, ils montrent que la qualité méthodologique des articles était médiocre au mieux et que les tests diagnostiques sont évalués de manière inadéquate (81).

Nous introduisons ici la notion de *reporting* (en anglais) qui correspond à la communication de données. Il s'agit de l'opération consistant, pour une équipe ou une entreprise, à faire le

rapport de son activité, dans notre cas à rapporter les différents éléments du travail de recherche. Par la suite, pour plus de lisibilité, nous utiliserons les termes « *reporting* » ou « rapporter » et parfois « déclaration ».

Comme dans tout autre type de recherche, les défauts dans la conception de l'étude peuvent conduire à des résultats biaisés. Une étude publiée par JG Lijmer et ses collègues dans le *JAMA* en 1999 tenta de déterminer l'effet d'un défaut de conception sur la précision diagnostique estimée. Pour 184 études publiées entre 1996 et 1997, ils évaluèrent l'association entre les caractéristiques des études et la précision diagnostique. Leur résultats montrèrent de manière empirique que les études diagnostiques comportant des défauts méthodologiques ou des défauts de conception peuvent être associées à des estimations biaisées, optimistes, de la précision diagnostique par rapport aux études sans ces carences (82).

1.3.2. La naissance d'un groupe, le projet STARD

A partir de 1999, et suite à l'initiative CONSORT (pour *CONsolidated Standards of Reporting Trials*) qui était considérée comme réussie (83–85), le Groupe de travail sur les méthodes de test de diagnostic et de dépistage Cochrane (*Cochrane Diagnostic and Screening Test Methods Working Group*) voulu améliorer la qualité du *reporting* des études diagnostiques. Ce groupe de travail développa une liste d'items qui devraient être inclus dans le *reporting* des « études diagnostiques ».

Ce travail se concrétise en 2003 au travers de l'initiative STARD, pour *Standards for Reporting of Precision Diagramme* (STARD), dont l'objectif est d'améliorer la qualité du *reporting* des études diagnostiques.

1.3.2.1. Leur méthodologie

Le comité directeur de STARD a identifié les publications sur la conduite et le *reporting* des études diagnostiques. Il en a extrait une liste étendue d'éléments pouvant potentiellement être inclus sur une liste ou grille de contrôle. En juillet 2000, s'est tenue une réunion de consensus de deux jours entre les experts de groupes d'intérêts suivants : chercheurs, éditeurs, méthodologues et organisations professionnelles. Une première ébauche de la liste STARD a été faite, puis elle a été discutée et modifiée par les experts. Elle a été testée sur place par des

utilisateurs potentiels. A la suite du congrès, elle a été publiée en ligne sur le site CONSORT, des commentaires pouvant être faits. Après avoir discuté les commentaires, le comité directeur a rassemblé la liste finale (69).

1.3.2.2. Leurs résultats

Après avoir récolté 33 listes de recommandations, le comité directeur a rassemblé une liste de 75 items. Lors d'une seconde réunion en septembre 2000, certains items ont été améliorés et d'autres éliminés pour aboutir à une liste de 25 éléments. Les différents commentaires faits, au cours de l'élaboration et après, ont été pris en compte pour aboutir à la version finale de la checklist. Un diagramme des flux a été élaboré en parallèle de la checklist. Il fournit des informations sur la méthode de recrutement des patients, l'ordre de réalisation des tests et le nombre de patients soumis au test évalué ainsi qu'au test de référence.

Ces recommandations furent publiées simultanément en janvier 2003 dans 8 journaux médicaux (*Radiology*, *American Journal of Clinical Pathology*, *Annals of Internal Medicine*, *British Medical Journal*, *Clinical Biochemistry*, *Clinical Chemistry*, *Clinical Chemistry of Laboratory Medicine*, et *Lancet*) et mises en ligne sur le site internet des critères CONSORT.

1.3.3. Les recommandations STARD

Nous présentons dans le tableau 3 les recommandations STARD publiées en janvier 2003 (69,86–88). Elles sont disponibles à cette adresse : <http://www.stard-statement.org/>.

Tableau 3 Recommandation STARD publiées en 2003 (69)

Section and topic	Item	Description
Title, abstract, and keywords	1	Identify the article as a study of diagnostic accuracy (recommend MeSH heading "sensitivity and specificity")
Introduction	2	State the research questions or aims, such as estimating diagnostic accuracy or comparing accuracy between tests or across participant groups
Methods		
Participants	3	Describe the study population: the inclusion and exclusion criteria and the settings and locations where the data were collected
	4	Describe participant recruitment: was this based on presenting symptoms, results from previous tests, or the fact that the participants had received the index tests or the reference standard?
	5	Describe participant sampling: was this a consecutive series of

		participants defined by selection criteria in items 3 and 4? If not, specify how participants were further selected
	6	Describe data collection: was data collection planned before the index tests and reference standard were performed (prospective study) or after (retrospective study)?
Test methods	7	Describe the reference standard and its rationale
	8	Describe technical specifications of material and methods involved, including how and when measurements were taken, or cite references for index tests or reference standard, or both
	9	Describe definition of and rationale for the units, cut-off points, or categories of the results of the index tests and the reference standard
	10	Describe the number, training, and expertise of the persons executing and reading the index tests and the reference standard
	11	Were the readers of the index tests and the reference standard blind (masked) to the results of the other test? Describe any other clinical information available to the readers.
Statistical methods	12	Describe methods for calculating or comparing measures of diagnostic accuracy and the statistical methods used to quantify uncertainty (eg 95% confidence intervals)
	13	Describe methods for calculating test reproducibility, if done
Results		
Participants	14	Report when study was done, including beginning and ending dates of recruitment
	15	Report clinical and demographic characteristics (eg age, sex, spectrum of presenting symptoms, comorbidity, current treatments, and recruitment centre)
	16	Report how many participants satisfying the criteria for inclusion did or did not undergo the index tests or the reference standard, or both; describe why participants failed to receive either test (a flow diagram is strongly recommended)
Test results	17	Report time interval from index tests to reference standard, and any treatment administered between
	18	Report distribution of severity of disease (define criteria) in those with the target condition and other diagnoses in participants without the target condition
	19	Report a cross tabulation of the results of the index tests (including indeterminate and missing results) by the results of the reference standard; for continuous results, report the distribution of the test results by the results of the reference standard
	20	Report any adverse events from performing the index test or the reference standard

Estimates	21	Report estimates of diagnostic accuracy and measures of statistical uncertainty (e.g. 95% confidence intervals)
	22	Report how indeterminate results, missing responses, and outliers of index tests were handled
	23	Report estimates of variability of diagnostic accuracy between readers, centres, or subgroups of participants, if done
	24	Report estimates of test reproducibility, if done
Discussion	25	Discuss the clinical applicability of the study findings

Un traduction française a été validée en 2015 (89) (tableau 4) :

Tableau 4 STARD 2003 traduction française (89)

Section et sujet	Item	Description
Titre, résumé, mots clefs	1	Identifier l'article comme une étude sur la précision d'un test diagnostique (se référer à la rubrique MeSH « sensibilité et spécificité »).
Introduction	2	Formuler les questions de recherche ou les objectifs de l'étude, comme évaluer la précision diagnostique ou comparer la précision entre plusieurs tests ou sur divers groupes de participants.
Méthodes		
Population	3	Population étudiée : critères d'inclusion et d'exclusion, contexte et lieux où les données ont été collectées.
	4	Recrutement des participants : le recrutement était-il basé sur la présentation de symptômes, sur des résultats d'exams antérieurs, ou sur le fait que les participants avaient déjà subi les tests étudiés ou le test de référence ?
	5	Échantillonnage des participants : la population de l'étude était-elle une série consécutive de participants définis selon les critères de sélection précisés dans les items 3 et 4 ? Si non, spécifier la manière dont les participants ont alors été sélectionnés.
	6	Collecte des données : le recueil des données était-il planifié avant que le test étudié et le test de référence aient été effectués (étude prospective) ou après (étude rétrospective) ?
Méthodes des tests	7	Description et justification du test de référence.
	8	Spécifications techniques du matériel et des méthodes utilisés, y compris comment et quand les mesures ont été prises, et/ou citer les références relatives aux tests étudiés et au test de référence.
	9	Définition et justification des unités, des seuils et/ou des catégories de résultats du test étudié et du test de référence.
	10	Nombre, formation et niveau d'expertise des personnes exécutant et interprétant les tests étudiés et le test de référence.

	11	Préciser si les lectures des tests étudiés et du test de référence ont été faites en aveugle (insu) des résultats de l'autre test, et décrire toute autre information clinique à la disposition des personnes interprétant le test.
Analyses statistiques	12	Méthodes de calcul ou de comparaison des mesures de précision diagnostique et analyses statistiques utilisées pour quantifier l'incertitude (par exemple, intervalles de confiance de 95 %)
	13	Méthodes de calcul de la reproductibilité du test, si réalisé.
Résultats		
Population	14	Quand l'étude a été réalisée, y compris dates de début et de fin de recrutement
	15	Caractéristiques cliniques et démographiques de la population de l'étude (au minimum : âge, sexe, type de symptômes présents)
	16	Nombre de participants répondant aux critères d'inclusion qui ont subi ou non les tests étudiés et/ ou le test de référence ; expliquer pourquoi les participants n'ont pas subi l'un ou l'autre test (un diagramme de flux est fortement recommandé).
Résultats pour les test	17	17 Délai entre les tests étudiés et le test de référence, et de tout traitement administré entre deux.
	18	Répartition des différents degrés de gravité de la maladie (définir les critères) des sujets remplissant les conditions ciblées ; autres diagnostics chez les participants ne remplissant pas les conditions ciblées.
	19	Tableau croisé comparant les résultats du test étudié (y compris les résultats indéterminés et manquants) aux résultats du test de référence ; pour les données continues, distribution des résultats du test étudié par rapport aux résultats du test de référence.
	20	Événements indésirables liés à la réalisation des tests étudiés ou du test de référence.
Estimations	21	Estimations de la précision diagnostique et des mesures de l'incertitude statistique (par exemple, intervalles de confiance de 95 %).
	22	Gestion des résultats indéterminés, des valeurs manquantes ou aberrantes des tests étudiés.
	23	Estimations de la variabilité de la précision diagnostique entre les sous-groupes de participants, entre les personnes interprétant le test ou entre les centres d'étude, si réalisées.
	24	Estimations de la reproductibilité du test, si réalisées.
Discussion	25	Discuter l'applicabilité clinique des résultats de l'étude.

Les recommandations publiées préconisent également l'utilisation d'un diagramme de flux (figure 15). Il doit fournir des informations sur la méthode de recrutement des patients, l'ordre de l'exécution des tests et le nombre de patients soumis au test évalué et au test de référence.



Figure 15 Modèle de diagramme des flux pour la réalisation d'une étude diagnostique (69).

1.3.4. Des recommandations mises à jour en 2015

Après la publication des critères STARD en 2003, plusieurs évaluations ont mis en évidence des améliorations, mineures, mais statistiquement significatives dans la qualité du *reporting* des études diagnostiques, avec un gain moyen de 1,4 items (IC 95%: 0,7 à 2,2) (90,91). En 2013, le comité directeur de STARD a décidé de remettre à jour la liste, avec pour

objectifs d'incorporer des données récentes sur les sources de biais, d'améliorer l'applicabilité, de faciliter l'analyse de l'interprétabilité des tests, rendre la liste plus facile à utiliser et l'harmoniser avec d'autres recommandations, en particulier CONSORT 2010 (92). Les membres du groupe STARD 2003 ainsi que de nouveaux participants ont participé au processus de mise à jour. La structure générale de STARD 2015 est similaire à celle de STARD 2003. Cette fois-ci, 30 éléments sont regroupés sous des sections qui suivent la structure IMRAD d'un article scientifique (tableau 5). Certains items sont identiques à ceux de la version 2003, d'autres ont été reformulés, combinés ou divisés, et enfin neuf items ont été ajoutés. Le diagramme des flux recommandé est légèrement modifié (figure 16). Les recommandations sont publiées conjointement dans plusieurs journaux (93–95). Elles sont disponibles, ainsi qu'une aide à leur utilisation, sur le site : <http://www.equator-network.org/reporting-guidelines/stard/>.

*Tableau 5 Mise à jour des recommandation STARD publiées en 2015, * : nouveaux items.*

Section & Topic	No	Item	Reported on page #
TITLE OR ABSTRACT			
	1	Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values, or AUC)	
ABSTRACT			
	2*	Structured summary of study design, methods, results, and conclusions (for specific guidance, see STARD for Abstracts)	
INTRODUCTION			
	3*	Scientific and clinical background, including the intended use and clinical role of the index test	
	4*	Study objectives and hypotheses	
METHODS			
Study design	5	Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study)	
Participants	6	Eligibility criteria	
	7	On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry)	
	8	Where and when potentially eligible participants were identified (setting, location and dates)	
	9	Whether participants formed a consecutive, random or	

		convenience series	
Test methods	10a	Index test, in sufficient detail to allow replication	
	10b	Reference standard, in sufficient detail to allow replication	
	11	Rationale for choosing the reference standard (if alternatives exist)	
	12a	Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory	
	12b	Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory	
	13a	Whether clinical information and reference standard results were available to the performers/readers of the index test	
	13b	Whether clinical information and index test results were available to the assessors of the reference standard	
Analysis	14	Methods for estimating or comparing measures of diagnostic accuracy	
	15	How indeterminate index test or reference standard results were handled	
	16	How missing data on the index test and reference standard were handled	
	17	Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory	
	18*	Intended sample size and how it was determined	
RESULTS			
Participants	19	Flow of participants, using a diagram	
	20	Baseline demographic and clinical characteristics of participants	
	21a	Distribution of severity of disease in those with the target condition	
	21b	Distribution of alternative diagnoses in those without the target condition	
	22	Time interval and any clinical interventions between index test and reference standard	
Test results	23	Cross tabulation of the index test results (or their distribution) by the results of the reference standard	
	24	Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)	
	25	Any adverse events from performing the index test or the reference standard	
DISCUSSION			
	26*	Study limitations, including sources of potential bias, statistical uncertainty, and generalisability	

	27*	Implications for practice, including the intended use and clinical role of the index test	
OTHER INFORMATION			
	28*	Registration number and name of registry	
	29*	Where the full study protocol can be accessed	
	30*	Sources of funding and other support; role of funders	

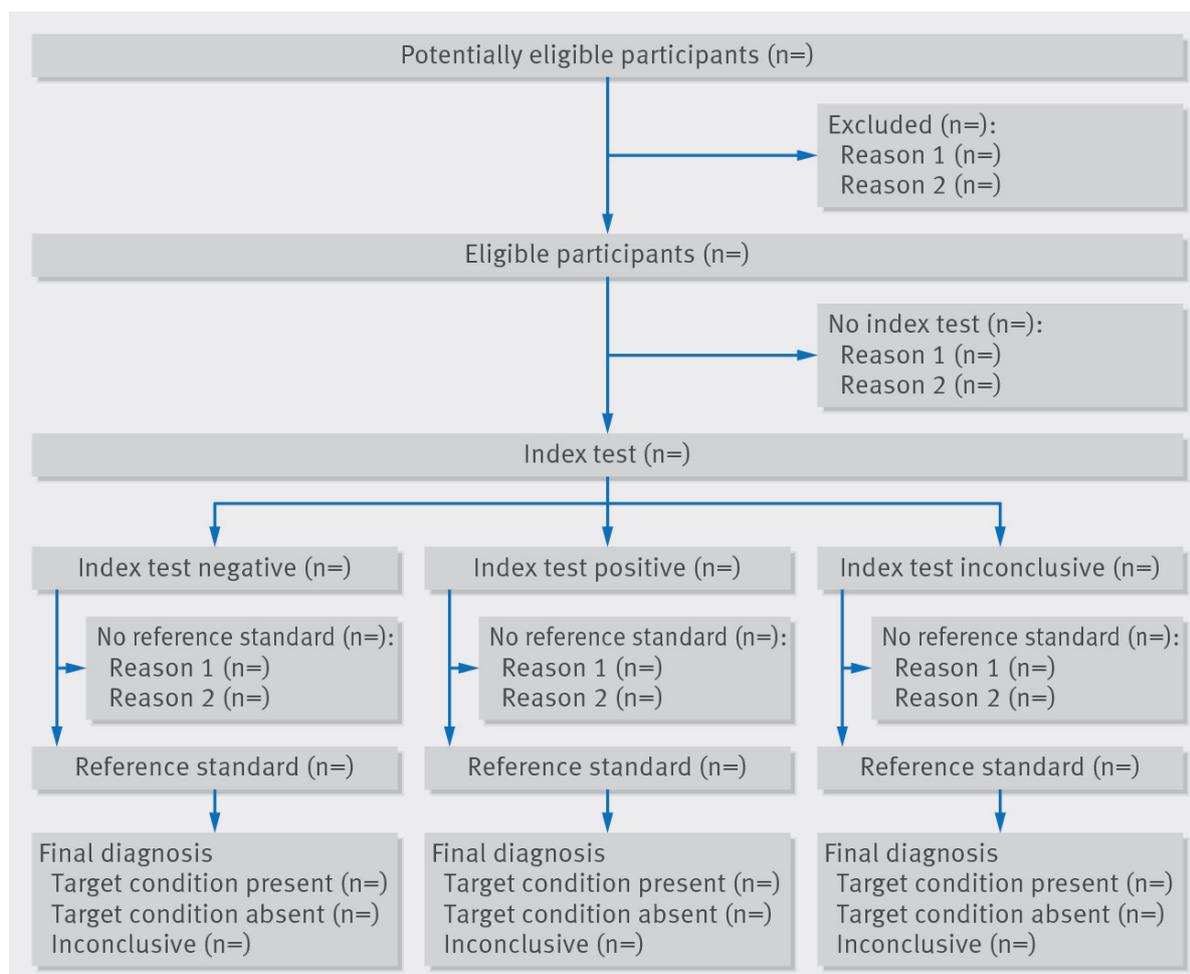


Figure 16 Modèle de diagramme des flux pour la réalisation d'une étude diagnostique, version de 2015.

Les éléments STARD ne sont pas spécifiques d'une discipline, ils s'appliquent à tous les types de recherche sur la précision diagnostique, l'imagerie, la microbiologie, les tests de laboratoire, les examens physiques et les questionnaires de dépistage.

1.4. Hypothèses et objectifs

La qualité du *reporting* et en particulier l'adhésion aux critères STARD des études diagnostiques de précharge dépendance ne sont pas connues. L'évaluation de l'adhésion de base à STARD nous paraît essentielle : elle permettra des interventions ciblées face aux déficiences qui pourraient exister. Par ailleurs, établir un point de comparaison est indispensable pour mesurer l'amélioration au fil du temps.

Notre hypothèse est que la qualité de la méthodologie et du *reporting* des études diagnostiques de précharge dépendance est moyenne ou faible, mais qu'elle s'améliore avec le temps.

Notre objectif principal est d'évaluer l'adhésion des études diagnostiques de précharge dépendance avec les critères STARD 2003.

Nos objectifs secondaires sont la recherche des caractéristiques des études associées à un meilleur *reporting*, l'observation de l'évolution du *reporting* au cours du temps ainsi que l'évaluation avec la mise à jour STARD 2015.

Pour y répondre, nous avons entrepris une revue systématique de la littérature sur dix ans.

2. Matériels et méthodes

2.1. Cadre méthodologique et terminologie

2.1.1. Une revue systématique

Nous avons adopté la définition présentée dans le *Cochrane handbook for systematic reviews of interventions* (96). « Une revue systématique tente de rassembler toutes les preuves empiriques qui correspondent à des critères d'éligibilité pré-spécifiés afin de répondre à une question de recherche spécifique. Elle utilise des méthodes explicites et systématiques qui sont choisies pour minimiser les biais, fournissant ainsi des résultats plus fiables à partir desquels des conclusions peuvent être tirées et des décisions prises. »

Voici les principales caractéristiques d'une revue systématique :

- Un ensemble clairement défini d'objectifs avec des critères d'éligibilité prédéfinis pour les études incluses ;
- Une méthodologie explicite et reproductible ;
- Une recherche systématique qui tente d'identifier toutes les études qui répondraient aux critères d'éligibilité ;
- Une évaluation de la validité des résultats des études incluses, par exemple en évaluant le risque de biais ;
- Une présentation systématique et une synthèse des caractéristiques et des résultats des études incluses.

2.1.2. The PRISMA Statement

Nous avons développé précédemment l'intérêt de l'utilisation d'une grille de *reporting* des caractéristiques lors de la réalisation d'études diagnostiques. Il en va de même pour les revues de la littérature ou les méta-analyses, dont les caractéristiques sont peu rapportées (97) malgré une amélioration (98). Dès 1996, pour améliorer la qualité du *reporting* des méta-analyses des

essais randomisés contrôlés, un groupe international a élaboré des instructions appelées *QUOROM Statement (Quality Of Reporting Of Meta-analysis)* (99). Cependant, ces recommandations ne concernaient pas les revues de la littérature bien que leur *reporting* soit également insuffisant (100). Elles ont été mises à jour et améliorées en 2009 (101), englobant cette fois-ci à la fois les revues systématiques et les méta-analyses, renommées alors PRISMA (*Preferred Reporting Items for Systematic reviews and Meta-Analyses*). Pour ce travail nous avons donc suivi ces recommandations (101). Elles comportent une liste de 27 items ainsi qu'un diagramme des flux (102). La liste complète des items ainsi que le *reporting* concernant notre étude sont présentés en Annexe 1.

2.1.3. The PICO statement

Définir la question clinique que l'on se pose en tant que problème spécifique pour le patient est le point de départ à la recherche de preuves cliniquement pertinentes dans la littérature. L'utilisation du formulaire PICO est une aide recommandée pour déclarer explicitement la question de recherche traitée. Le formulaire PICO se réfère aux participants, aux interventions, aux comparaisons et aux résultats (*PICO : Patients/Problem, Intervention, Comparison, Outcome*). Nous l'avons donc utilisé pour notre revue, le voici détaillé :

- Patients : Patient en insuffisance circulatoire, **suspecté de répondre à un remplissage vasculaire** par une augmentation de son débit cardiaque, le contexte peut être varié (réanimation, bloc opératoire, service d'urgence ou autre).
- Intervention : Nouveau test de **précharge dépendance**.
- Comparaison : **Diagnostic de précharge dépendance validé** par la mesure du débit cardiaque ou un substitut après un remplissage vasculaire ou une manœuvre modifiant la précharge cardiaque. L'utilisation d'un *gold standard* n'est pas une obligation.
- Résultats : **Précision diagnostique**, avec des valeurs de sensibilité, de spécificité et l'aire sous la courbe ROC du nouveau test.

La question ainsi dégagée et à laquelle les études incluses doivent répondre est la suivante :

Quelle est la précision diagnostique d'un test (le test évalué) pour prédire la réponse au remplissage vasculaire dans une population de patients suspectés de précharge dépendance ?

2.1.4. Les investigateurs

Notre travail réunit trois investigateurs. Nous les présentons ainsi que leur rôle dans le processus de recherche :

- Pr Jean-Luc Fellahi (JLF) est professeur au CHU de Lyon, chef du service d'anesthésie-réanimation de l'hôpital cardiologique Louis Pradel aux Hospices Civils de Lyon. Il coordonne le travail de recherche, apporte son expertise sur le sujet de recherche et la méthodologie, participe aux discussions.
- Dr Matthias Jacquet-Lagrèze (MJL) est chef de clinique assistant dans le service d'anesthésie-réanimation de l'hôpital cardiologique Louis Pradel. Il est l'initiateur du projet de recherche, participe à la mise en place de l'étude ainsi qu'à sa méthodologie, participe à la sélection des études, participe au recueil de données pour le calcul de la variabilité inter opérateur, conduit l'analyse statistique, rédige l'article.
- Guillaume Izaute (GI) est interne en anesthésie-réanimation au CHU de Lyon. Il participe à la mise en place de l'étude ainsi qu'à sa méthodologie, participe à la sélection des études, effectue le recueil de données, rédige cette thèse.

2.1.5. Source de financement

Cette étude n'est supportée par aucun financement extérieur.

2.2. Sélection des études

2.2.1. Sourced'information

Nous avons effectué notre recherche sur la base de données MEDLINE uniquement, via le moteur de recherche PubMed (103).

2.2.2. Recherche dans la littérature

Deux des investigateurs (GI et MJL) ont effectué une revue systématique sur la base de données MEDLINE *via* PubMed, pour rechercher toute publication dont la question principale était la suivante : « Quelle est la précision diagnostique d'un test (le test évalué)

pour prédire la réponse au remplissage vasculaire dans une population de patients suspectés de précharge dépendance ? »

Les études ont été incluses si elles répondaient aux critères d'inclusion suivants :

- 1) Etude évaluant la précision diagnostique d'un ou plusieurs tests ou indice permettant de prédire la réponse au remplissage vasculaire.
- 2) Le nouveau test est comparé à un index de précharge dépendance comportant une mesure du débit cardiaque ou un substitut après un remplissage vasculaire ou une manœuvre modifiant la précharge cardiaque.
- 3) L'étude rapporte au moins une estimation de précision diagnostique (exemple : sensibilité, spécificité, valeur prédictive positive, valeur prédictive négative, aire sous la courbe ROC).

Nous avons exclu les études qui comportaient un des critères suivants : les études non publiées en langue anglaise, les études pédiatriques, les études animales, les études ne présentant pas de résumé.

Le contexte n'était pas un facteur limitant, il pouvait être la période périopératoire ou bien une défaillance circulatoire. Les études pouvaient donc avoir lieu au bloc opératoire ainsi que les services de réanimation ou d'urgences.

Nous avons mis en place un algorithme de recherche puis nous l'avons mis à l'épreuve et amélioré en le confrontant à trois revues de la littérature et/ou méta-analyse récentes sur le sujet (46,47,104). Les articles analysés dans ces revues de la littérature devaient apparaître dans notre recherche sur PubMed.

Finalement, les mots clés les plus pertinents que nous avons conservés et utilisés pour la sélection des études étaient :

(« fluid responsiveness » OR « volume responsiveness » OR « preload OR hypovolemia ») AND (« critically ill » OR « shock » OR « intensive care » or « anesthesia » OR « critical care » OR « perioperative » OR « resuscitation »).

Nous avons choisi de manière arbitraire d'effectuer nos recherches sur 10 années consécutives, de 2005 à 2014, soit peu de temps après l'apparition des critères STARD 2003.

2.2.3. Sélection des études

Dans un second temps, deux des auteurs, GI et MJL ont examiné de manière indépendante les titres et les résumés des articles afin d'identifier les études potentiellement éligibles. En cas de concordance entre les deux auteurs, l'article était soit sélectionné pour l'étape suivante, soit éliminé. En cas de discordance (article uniquement sélectionné par un seul auteur), le texte intégral de l'article était évalué par les deux auteurs. Les revues de la littérature et les méta-analyses étaient exclues à ce moment-là.

Ensuite, furent exclues les études publiées dans un journal dont *l'impact factor* était inférieur à 2. Nous avons décidé de ce seuil pour plusieurs raisons :

- Plusieurs études publiées dans des revues à faible *impact factor* avait une conception ou méthodologie atypique et un *reporting* très mauvais. La recherche de la présence ou de l'absence des critères était rendue difficile.
- Les études concernant les critères diagnostiques de précharge dépendance sont très nombreuses. Nous avons préféré privilégier un meilleur étalement dans le temps de notre analyse plutôt que de proposer une analyse complète de la littérature mais sur une période plus courte.
- La plupart des études incluses dans les méta-analyses sur le sujet sont publiées dans des revues à *impact factor* supérieur à 2.

Dans un dernier temps, c'est à dire au moment de la lecture complète de l'article pour en extraire les données, certaines études furent exclues si elles ne satisfaisaient pas aux critères d'inclusion, après consensus entre GI et MJL.

2.3. Collecte des données

2.3.1. Caractéristiques des études

Les caractéristiques de base suivantes ont été extraites par GI : année de publication, équipe du premier auteur (afin de mesurer le nombre de publications par équipe), région où se déroule l'étude, nom du journal, impact factor, adoption de STARD (les critères STARD doivent être mentionnés dans les recommandations faites aux auteurs), source de financement (présente, absente ou non mentionnée), technique utilisée pour définir les répondants, aire

sous la courbe ROC du nouveau test, taille de l'échantillon, milieu de l'étude (réanimation, anesthésie, autre).

Le statut d'adoption de STARD a été évalué en examinant les instructions aux auteurs de chaque journal. La simple mention des critères STARD suffisait, le journal n'avait pas besoin de recommander explicitement l'adhésion à STARD et/ou STARD 2015.

2.3.2. Score STARD 2003

La grille STARD 2003 (69) présentée en table 3 (chapitre 1.3.3.) a été utilisée pour évaluer la qualité du *reporting*. Un investigateur (GI) a lu l'ensemble des études. Pour chacune d'elle, il a appliqué la grille STARD 2003 en relevant la présence de chacun des items. Un second investigateur (MJL) a lu les études de l'année 2005 afin de calculer l'agrément inter-investigateur. Lorsque l'item était bien rapporté, il était noté 1 point, s'il était non rapporté ou si sa présence n'était pas claire il était noté 0 point. Il est important de remarquer que ce n'est pas la qualité de l'étude ou le risque de biais qui est évalué mais uniquement la qualité du *reporting*. Par exemple il ne s'agit pas de s'avoir si l'étude est réalisée en aveugle ou pas, mais si le fait qu'elle soit en l'aveugle ou non est mentionné dans le manuscrit.

Avant le recueil des scores STARD 2003, deux auteurs (GI et MJL) ont effectué un score pilote sur cinq études afin d'affiner la définition de chaque item.

La notation a été modifiée pour certains items, par consensus entre GI, MJL et JLF :

- Les items intéressant à la fois le nouveau test et le test de référence étaient divisés en deux, 0,5 point était attribué par test. Par exemple voici comment était évalué l'item 8 : « Spécifications techniques... relatives aux tests étudiés (0,5 points) et au test de référence (0,5 points) ». Les items 8, 9, 10, 11, 13 et 24 sont concernés par cette situation et ont été notés de la même manière.
- Nous avons divisé par 3 l'item 3 : « Population étudiée : critères d'inclusion (0,33 point) et d'exclusion (0,33 point), contexte et lieux (0,33 point) où les données ont été collectées », les points étaient attribués s'il était rapporté au moins 3 critères d'inclusion et 3 critères d'exclusion.
- Nous avons divisé par 4 l'item 4 : « Recrutement des participants : le recrutement était-il basé sur la présentation de symptômes (0,25 point), sur des résultats d'examen antérieurs (0,25 point), ou sur le fait que les participants avaient déjà subi les tests étudiés (0,25 point) ou le test de référence (0,25 point) ? ».

La définition a été précisée pour certains items :

- Nous avons défini l’item 7 « Description et justification du test de référence » comme suit : soit nous considérons le test de référence comme un *gold standard*, une simple référence suffisait à ce nous le jugions « justifié » (1 point), soit ce n’était pas un *gold standard* et nous devions retrouver une justification argumentée du test de référence au sein de l’article. Le test de référence était considéré comme le *gold standard* si deux évaluations du débit cardiaque étaient effectuées, avant et après expansion volémique, la variation de débit cardiaque qui en découlait devait être mesurée par thermodilution (intra ou transpulmonaire) ou par échocardiographie. Dans tous les autres cas, le test de référence n’était pas considéré comme le gold standard.
- L’item 9 « Définition et justification des unités, des seuils et/ou des catégories de résultats du test étudié et du test de référence » était divisé en deux. Concernant le nouveau test, son seuil optimal de sensibilité et de spécificité était déterminé *après* que le test soit effectué, en général à partir d’une courbe ROC. La méthode pour choisir le seuil devait être citée (0,5 point). Concernant le test de référence, le choix de son seuil devait être justifié par la valeur du *least significant change* (LSC ou variation minimale significative) soit la plus petite différence entre 2 valeurs du test qui peuvent être considérées comme un changement réel et non attribuable au hasard. 0,5 point était attribué si le *least significant change* était cité, ou une de ces deux études citées : *Precision of the transpulmonary thermodilution measurements* de X Monnet ou *Reliability of the Thermodilution Method in the Determination of Cardiac Output in Clinical Practice* de C W Stetz (105,106).
- Pour l’item 15 « Caractéristiques cliniques et démographiques de la population de l’étude », le point était attribué si au moins trois caractéristiques étaient présentes.
- Pour l’item 17 « Délai entre les tests étudiés et le test de référence, et de tout traitement administré entre les deux », il fallait que la non modification de traitement soit spécifiée par au moins une phrase (par exemple : « pendant l’étude, les traitements sédatifs, inotrope, vasopresseur et les réglages du respirateur sont restés inchangés »)
- Pour l’item 19 « Tableau croisé comparant les résultats du test étudié aux résultats du test de référence ; pour les données continues, distribution des résultats du test étudié par rapport aux résultats du test de référence », nous attendions une figure présentant les résultats croisés des deux tests.

- Pour l’item 21 « Estimations de la précision diagnostique et des mesures de l’incertitude statistique (par exemple, intervalles de confiance de 95 %) », le point n’était attribué qu’en présence de l’IC à 95%.
- Enfin, les item 13, 23 et 24 ne sont à rapporter que « si réalisés » ; nous avons fait le choix de les inclure systématiquement. S’ils n’étaient pas réalisés, ils ne rapportaient pas de point. Par exemple pour l’item 13 « Méthodes de calcul de la reproductibilité du test, si réalisé », si la reproductibilité est calculée et la méthode décrite, le point est attribué, si la reproductibilité n’est pas évaluée, le point n’est pas attribué.

Si le premier investigateur (GI) avait des doutes dans le *reporting* de certains articles, un deuxième investigateur (MJL) lisait l’étude et les points difficiles étaient résolus par consensus.

2.3.3. Score STARD 2015

Les items de la grille STARD 2015 étaient recueillis de manière identique et concomitante à la grille STARD 2003. GI a recueilli les items de la grille pour l’ensemble des études, MJL a recueilli l’année 2005 pour le calcul de l’agrément inter-investigateur. La notation était la même (bien rapporté : 1 point, mal ou non rapporté : 0 point).

La grille 2015 est une amélioration de celle de 2003, elle est en pratique plus claire et plus facile d’utilisation. Les items intéressant à la fois le nouveau test et le test de référence étaient d’emblée divisés en deux, il nous a suffi d’attribuer 0,5 point par demi item (item 10a et 10b par exemple). Pour les items que nous avons dû détailler ou préciser lors du recueil STARD 2003, nous avons attribué les points de la même manière.

Un des nouveaux items a été précisé : l’item 26 « *Study limitations, including sources of potential bias, statistical uncertainty, and generalisability* », il fallait qu’à la fois 1) une source de biais potentiel, 2) une incertitude statistique et 3) et une évocation de la généralisabilité pour que le point soit attribué.

2.4. Analyse statistique

L’adhésion globale à STARD 2003 a été mesurée en utilisant le nombre total d’items STARD 2003 présents pour chaque étude, ce qui peut aller de 0 (aucun des items n’est signalé) à 25 (tous les items correctement signalés). L’adhésion globale à STARD 2015 a été mesurée en

utilisant le nombre total d'items STARD 2015 présents pour chaque étude, ce qui peut aller de 0 à 30. Les items avec plusieurs sous-éléments ont été notés avec un total maximum de 1 point avec une fraction de point pour chaque sous-item. Les données recueillies ont été résumées à l'aide de statistiques descriptives : moyenne, écart-type et l'intervalle de confiance. L'analyse du *reporting* spécifique de chaque item a été réalisée sur la cohorte entière. L'agrément inter-évaluateur a été calculé.

Un test de Cochran-Armitage a été réalisé pour identifier une variation significative au cours du temps du score STARD 2003 et de chacun des items de la grille STARD 2003.

Une analyse univariée des caractéristiques des études sélectionnées qui influencent la valeur du score STARD 2003 a été réalisée sur un modèle de régression linéaire. Ensuite, une analyse multivariée a été réalisée pour identifier des caractéristiques indépendantes associées à un meilleur score STARD 2003. Les caractéristiques utilisées dans l'analyse sont les suivantes : année de publication, technique utilisée pour définir les répondants, milieu de l'étude, source de financement, journal, nombre de publication par équipe, *impact factor*, région où se déroule l'étude, taille de l'échantillon, aire sous la courbe ROC du nouveau test, STARD cité dans instructions aux auteurs.

Tous les tests étaient bilatéraux (*two sided*), $P < 0,05$ était considéré comme statistiquement significatif. L'analyse statistique a été effectuée avec le logiciel R. Nous avons utilisé plusieurs paquets du projet CRAN R.

3. Résultats

3.1. Caractéristiques des études

Sur les 725 études initialement passées en revue, 104 répondaient aux critères d'inclusion et d'exclusion, le diagramme des flux est présenté en figure 17. Elles ont été incluses dans notre analyse, leurs caractéristiques sont présentées sur le tableau 6. La liste complète des études analysées est présentée en annexe 5.

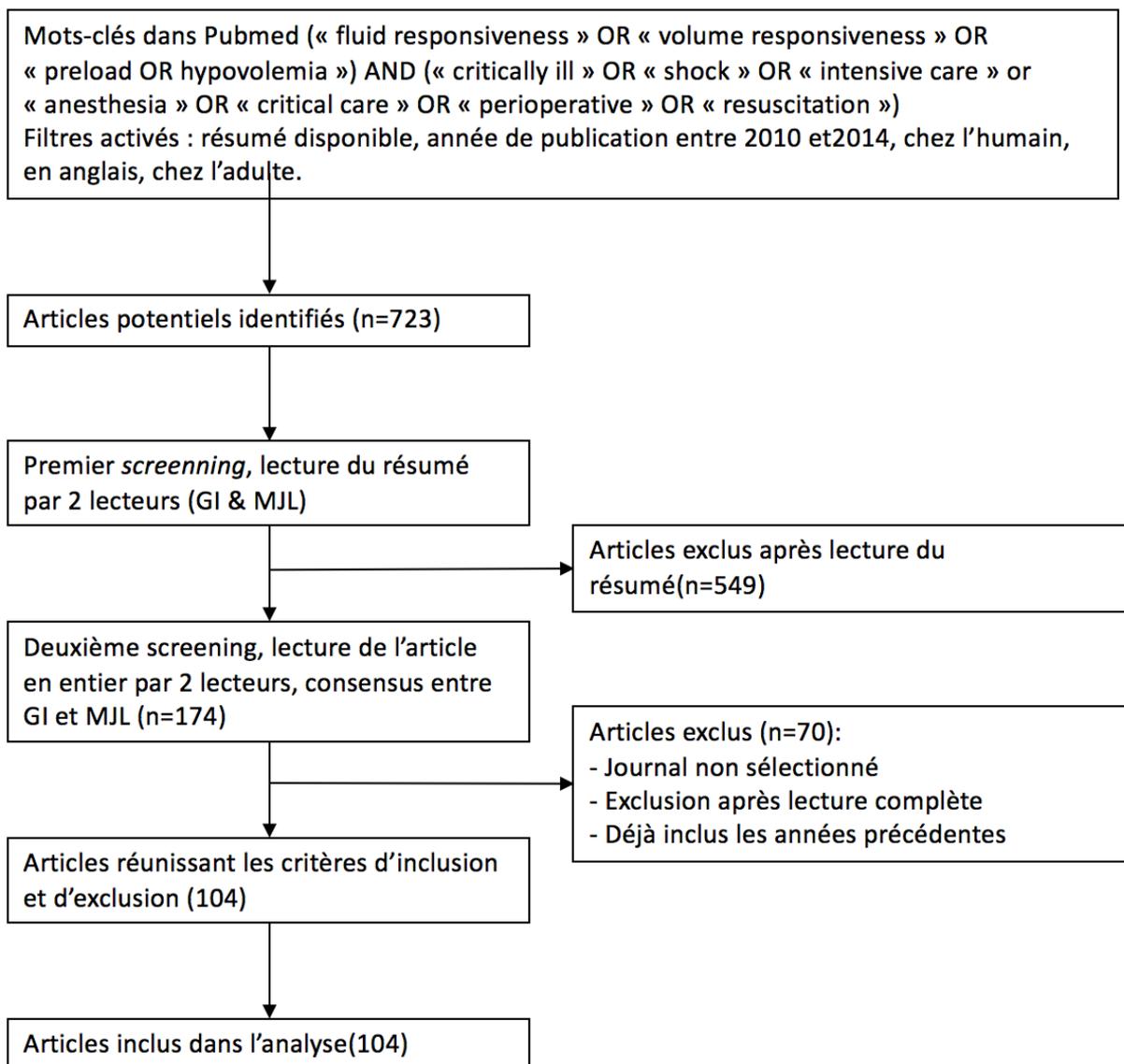


Figure 17 Diagramme des flux. Sélection des articles inclus dans l'analyse, réalisée par deux lecteurs : GI = G Izaute ; MJL = M Jacquet-Lagrèze.

La majorité des études ont été publiées en Europe (87, soit 83,7%), en France principalement (47, soit 45,2%). 40 (38,5%) études avaient un financement uniquement institutionnel, 14 (13,5%) sont au moins partiellement subventionnées par l'industrie, mais cette information n'est mentionnée que dans 54 (51,9%) des articles. L'*impact factor* moyen des journaux est de 4,8 (IC95% : 4,4-5). Les trois principaux journaux ayant publié des études diagnostiques sur la réponse au remplissage vasculaire sont : *British Journal of Anesthesia* (24 études soit 23,1%), *Intensive Care Medicine* (15 études soit 14,4%) et *Critical Care* (13 études soit 12,5%). Les principaux tests évalués sont basés sur l'interaction cœur-poumon (70 études soit 67,3%) avec principalement la variation de pression pulsée (15 études) et la variation du VES (10 études). Viennent ensuite les tests basés sur des indices statiques de précharge dépendance (12 études soit 11,5%) et les tests incluant un lever de jambes passif (11 études soit 10,6%). Les principaux tests de référence utilisés mesurent la variation de débit cardiaque, ou bien un indice dérivé, après un remplissage vasculaire. Les principales techniques de mesures sont : la thermodilution transpulmonaire (26 études soit 25%), le cathétérisme artériel pulmonaire (19 études soit 18,3%), le doppler œsophagien (17 études soit 16,3%), l'analyse de l'onde de pouls (17 études soit 16,3%) et l'échographie cardiaque (16 études soit 15,4%).

Trois journaux citent les recommandations STARD dans leurs instructions pour les auteurs : *Anesthesia Analgesia*, *Critical Care* et *European Journal of Anaesthesiology*. Cela représente 29,8% des études incluses.

La taille médiane des échantillons est de 33,0 [23,7-47,5] patients, pour une médiane de 35,0 [25,0-51,0] remplissages vasculaires étudiés. Il y avait une tendance à une augmentation de la taille de l'échantillon dans le temps (de 21,0 [20,0-40,0] en 2005 à 50,0 [42,0-62,0] en 2014), cette augmentation est significative en analyse univariée, $p=0,014$.

La précision des tests évalués, exprimée par l'aire sous la courbe ROC, est en moyenne de 0,84, avec un intervalle de confiance à 95% moyen entre 0,70 et 0,93).

Tableau 6 Caractéristiques des études. N=104

Caractéristiques	No	%
Année de publication		
2005	9	8,7
2006	11	10,6
2007	10	9,6
2008	11	10,6
2009	13	12,5
2010	10	9,6
2011	13	12,5
2012	7	6,7
2013	11	10,6
2014	9	8,7
Technique utilisée pour définir les répondeurs		
Echographie	34	32,7
Thermodilution	45	43,3
Pression artérielle	4	3,8
Onde pouls	17	16,3
Bioréactance	2	1,9
Multiple	2	1,9
Milieu		
Anesthésie	34	32,7
Soins intensifs	66	63,5
Mixte	3	2,9
Autre	1	1,0
Source de financement		
Financement institutionnel	40	38,5
Financement par l'industrie	14	13,5
Inconnu	50	48,1
Journal		
<i>Anesthesia-Analgesia</i>	10	9,6
<i>British Journal of Anaesthesia</i>	24	23,1
<i>Critical Care Medicine</i>	12	11,5
<i>Critical Care</i>	13	12,5
<i>Intensive Care Medicine</i>	15	14,4
Autre	30	28,8
Nombre de publication par équipe		
Median [IQR]	3 [1-4]	
Impact factor		
<3	19	18,3
>3 And< 5	53	51,0
>5	32	30,8
Lieu de l'étude		
America	5	4,8
Europe	88	84,6
Asia	11	10,6
Taille de l'échantillon		
Median [IQR]	33 [24-48]	
Résultat avec l'AUC		
Median [IQR]	0.89 [0.77-0.94]	
STARD cité dans instructions aux auteurs		
Oui	31	29,8
Non	71	68,2

IQR : interquartile range

3.2. Score STARD 2003

Les données de neuf premières études (année 2005), soit 324 items, ont été évaluées indépendamment et en aveugle par deux investigateurs (MJL et GI). Parmi ces 324 items, 30 divergences ont été identifiées, elles ont toutes été résolues avec succès par consensus [16 Items en faveur de l'extracteur principal (GI) et 14 contre]. Le taux d'erreur du principal extracteur de données était donc de 4,3% (14/324).

3.2.1. Qualité globale des études

Le score STARD 2003 moyen pour tous les items était de 12,0 sur une échelle de 0 à 25 [intervalle de confiance à 95% (IC95%) : 11,4-12,5]. En moyenne 47,9% des items des recommandations STARD 2003 étaient correctement rapportés.

La moyenne des scores STARD 2003 est restée stable pendant les 10 années étudiées, de 2005 à 2014, $p=0.546$ en analyse univariée.

Aucune étude n'a eu un score maximal, c'est à dire de 25. Les scores minimum et maximum étaient respectivement 4,8 et 19,5 sur 25. 60,1% des études avait un score STARD 2003 inférieur à la moyenne, c'est à dire inférieur à 12,5 points.

3.2.2. Analyse du *reporting* en fonction des items

Si l'on s'intéresse de manière individuelle à la fréquence du *reporting* des critères STARD 2003, on remarque d'importantes disparités entre les items. Les résultats sont présentés dans le tableau 7.

Les items les moins bien rapportés étaient les suivants : le recrutement des patients (conformément rapporté dans 26,4 % des études), l'échantillonnage des patients (32,7%), le test de référence et son rationnel (44,2%), la manière dont sont choisies les valeurs seuil des test (41,4%), les personnes effectuant les tests (9,1%), si les tests sont réalisés en aveugle (18,3%), la période pendant laquelle est réalisée l'étude (30,8%), la présence de résultats intermédiaires ou manquant (25,0%). Par ailleurs, le diagramme des flux n'est présent que dans 4,8% des études.

Les items les mieux rapportés étaient les suivants : la population étudiée (conformément rapporté dans 72,3% des études), les spécificités techniques des tests étudiés (74,5%), la méthode utilisée pour calculer ou comparer les mesures des performances des tests (91,3%), les caractéristiques cliniques et démographiques de la population étudiée (80,8%), la distribution de la sévérité de la maladie chez les malades et les autres diagnostics chez les non-malades (70,2%), les performances diagnostics et les mesures d'incertitudes (69,2%), l'application clinique des résultats (87,5%).

Un test de tendance Cochran-Armitage a identifié deux critères STARD 2003 ayant évolué au cours du temps de manière significative (Figure 18). La présence d'une analyse croisée entre les résultats des tests étudiés et les tests de référence était moins rapportée, passant de 66,7% de *reporting* en 2005 à 44,4% en 2014 ($p=0.030$). A l'inverse, les performances diagnostics et les mesures d'incertitudes des tests étaient mieux rapportées, passant de 55,5% de *reporting* en 2005 à 100% en 2014 ($p<0.001$).

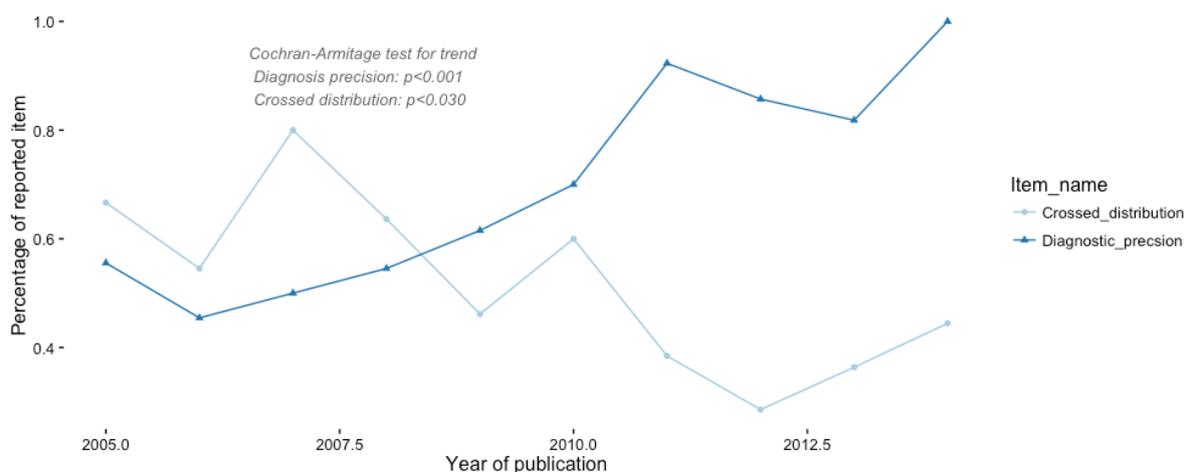


Figure 18 Evolution de l'analyse croisée (item 19, bleu clair) et de la précision diagnostique (item 21, bleu foncé) sur dix ans. Pourcentage de reporting par année de publication.

Tableau 7 Reporting en fonction des items STARD 2003

Section et sujet	Item	Critère	Description	Items correctement rapportés (No & %)	
Titre, résumé, mots clefs	1	Identification	Identifier l'article comme une étude sur la précision d'un test diagnostique (se référer à la rubrique MeSH « sensibilité et spécificité »).	64	61,5
Introduction	2	Objectifs	Formuler les questions de recherche ou les objectifs de l'étude, comme évaluer la précision diagnostique ou comparer la précision entre plusieurs tests ou sur divers groupes de participants.	95	91,3
Méthodes					
Population	3	Population étudiée	Critères d'inclusion et d'exclusion, contexte et lieux où les données ont été collectées.	75,24	72,3
	4	Recrutement des participants	Le recrutement était-il basé sur la présentation de symptômes, sur des résultats d'examens antérieurs, ou sur le fait que les participants avaient déjà subi les tests étudiés ou le test de référence ?	27,5	26,4
	5	Échantillonnage des participants	La population de l'étude était-elle une série consécutive de participants définis selon les critères de sélection précisés dans les items 3 et 4 ? Si non, spécifier la manière dont les participants ont alors été sélectionnés.	34	32,7
Méthodes des tests	6	Collecte des données	Le recueil des données était-il planifié avant que le test étudié et le test de référence aient été effectués (étude prospective) ou après (étude rétrospective) ?	100	96,2
	7	Test de référence	Description et justification du test de référence.	46	44,2
	8	Spécifications techniques	Spécifications techniques du matériel et des méthodes utilisés, y compris comment et quand les mesures ont été prises, et/ou citer les références relatives aux tests étudiés et au test de référence.	77,5	74,5
	9	Unités	Définition et justification des unités, des seuils et/ou des catégories de résultats du test étudié et du test de référence.	44	42,3
	10	Personnes réalisant les tests	Nombre, formation et niveau d'expertise des personnes exécutant et interprétant les tests étudiés et le test de référence.	9,5	9,1
	11	Aveugle	Préciser si les lectures des tests étudiés et du test de référence ont été faites en aveugle (insu) des résultats de l'autre test, et décrire toute autre information clinique à la disposition des personnes interprétant le test.	19	18,3

Analyses statistiques	12	Méthode statistique	Méthodes de calcul ou de comparaison des mesures de précision diagnostique et analyses statistiques utilisées pour quantifier l'incertitude (par exemple, intervalles de confiance de 95 %)	95	91,3
	13	Reproductibilité	Méthodes de calcul de la reproductibilité du test, si réalisé.	14,5	13,9
Resultats					
Population	14	Quand	Quand l'étude a été réalisée, y compris dates de début et de fin de recrutement	32	30,8
	15	Population étudiée	Caractéristiques cliniques et démographiques de la population de l'étude (au minimum : âge, sexe, type de symptômes présents)	84	80,8
	16	Diagramme des flux	Nombre de participants répondant aux critères d'inclusion qui ont subi ou non les tests étudiés et/ ou le test de référence ; expliquer pourquoi les participants n'ont pas subi l'un ou l'autre test (un diagramme de flux est fortement recommandé).	5	4,8
Résultats pour les test	17	Délai et traitement	Délai entre les tests étudiés et le test de référence, et de tout traitement administré entre deux.	50	48,1
	18	Distribution de la maladie ; autres diagnostics	Répartition des différents degrés de gravité de la maladie (définir les critères) des sujets remplissant les conditions ciblées ; autres diagnostics chez les participants ne remplissant pas les conditions ciblées.	73	70,2
	19	Analyse croisée	Tableau croisé comparant les résultats du test étudié (y compris les résultats indéterminés et manquants) aux résultats du test de référence ; pour les données continues, distribution des résultats du test étudié par rapport aux résultats du test de référence.	54	51,9
	20	Évènements indésirables	Évènements indésirables liés à la réalisation des tests étudiés ou du test de référence.	9	8,7
Estimations	21	Précision diagnostique	Estimations de la précision diagnostique et des mesures de l'incertitude statistique (par exemple, intervalles de confiance à 95 %).	72	69,2
	22	Résultats indéterminés	Gestion des résultats indéterminés, des valeurs manquantes ou aberrantes des tests étudiés.	26	25,0
	23	Analyse en sous-groupes	Estimations de la variabilité de la précision diagnostique entre les sous-groupes de participants, entre les personnes interprétant le test ou entre les centres d'étude, si réalisées.	28	26,9
	24	Reproductibilité	Estimations de la reproductibilité du test, si réalisées.	21	20,2
Discussion	25	Applicabilité	Discuter l'applicabilité clinique des résultats de l'étude.	91	87,5

3.3. Score STARD 2015

3.3.1. Qualité globale des études

Le score STARD 2015 moyen pour tous les items était de 14,94 sur une échelle de 0 à 30 (IC95% = 14,35 à 15,54). En moyenne 49,8% des items des STARD 2015 étaient correctement rapportés.

Les scores maximum et minimum étaient respectivement 6,5 et 21,5 sur 30.

Le calcul de la corrélation entre les score STARD 2003 et 2015 montrait un coefficient de corrélation linéaire $r=0,859$.

3.3.2. Analyse du *reporting* en fonction des items

Parmi les nouveaux items de la grille STARD 2015, le *reporting* est hétérogène. Les résultats sont présentés dans le tableau 8. Le taux de *reporting* de ces neuf nouveaux items est de 56,7%.

Un résumé structuré est rapporté dans 93,3% des études, la taille de l'échantillon prévue (le nombre de sujets nécessaires) dans 27,9% des études, les implications du nouveau test dans la pratique clinique dans 84,6% des études, les sources de financement et leur rôle dans 50,0% des études.

Tableau 8 Reporting en fonction des items STARD 2015. * : nouveaux items

Section & Topic	No	Item	Reporting of the item (No & %)	
TITLE OR ABSTRACT				
	1	Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values, or AUC)	93	89,4
ABSTRACT				
	2*	Structured summary of study design, methods, results, and conclusions (for specific guidance, see STARD for Abstracts)	97	93,3
INTRODUCTION				
	3*	Scientific and clinical background, including the intended use and clinical role of the index test	101	97,1
	4*	Study objectives and hypotheses	96	92,3
METHODS				
Study design	5	Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study)	98	94,2
Participants	6	Eligibility criteria	52	50,0
	7	On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry)	37	35,6
	8	Where and when potentially eligible participants were identified (setting, location and dates)	28	26,9
	9	Whether participants formed a consecutive, random or convenience series	33	31,7
Test methods	10a	Index test, in sufficient detail to allow replication	89	85,6
	10b	Reference standard, in sufficient detail to allow replication	64	61,5
	11	Rationale for choosing the reference standard (if alternatives exist)	45	43,3
	12a	Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory	69	66,4
	12b	Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory	25	24,0
	13a	Whether clinical information and reference standard results were available to the performers/readers of the index test	19	18,3
	13b	Whether clinical information and index test results were available to the assessors of the reference standard	16	15,4
Analysis	14	Methods for estimating or comparing measures of diagnostic accuracy	94	90,4
	15	How indeterminate index test or reference standard results	7	6,7

		were handled		
	16	How missing data on the index test and reference standard were handled	25	24,0
	17	Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory	33	31,7
	18*	Intended sample size and how it was determined	29	27,9
RESULTS				
Participants	19	Flow of participants, using a diagram	7	6,7
	20	Baseline demographic and clinical characteristics of participants	83	79,8
	21a	Distribution of severity of disease in those with the target condition	77	74,0
	21b	Distribution of alternative diagnoses in those without the target condition	74	71,2
	22	Time interval and any clinical interventions between index test and reference standard	44	42,3
Test results	23	Cross tabulation of the index test results (or their distribution) by the results of the reference standard	50	48,1
	24	Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)	71	68,3
	25	Any adverse events from performing the index test or the reference standard	7	6,7
DISCUSSION				
	26*	Study limitations, including sources of potential bias, statistical uncertainty, and generalisability	46	44,2
	27*	Implications for practice, including the intended use and clinical role of the index test	88	84,6
OTHER INFORMATION				
	28*	Registration number and name of registry	18	17,3
	29*	Where the full study protocol can be accessed	4	3,9
	30*	Sources of funding and other support; role of funders	52	50,0

3.4. Facteurs associés à la qualité du *reporting*

En analyse univariée, les caractéristiques des études associées à un meilleur score STARD 2003 étaient les études réalisées en soins intensifs, les études publiées dans les revues *Critical Care Medicine* et *Intensive Care Medicine*, les études publiées dans des revues d'*impact factor* supérieur à 5 et les études publiées dans des revues citant STARD dans leurs instructions aux auteurs (tableau 9).

En analyse multivariée seul le lieu où se déroule l'étude était associé au score STARD 2003, avec un score plus élevé si l'étude se déroule en soins intensifs. De manière inattendue, l'*impact factor* n'était plus associé à un meilleur *reporting* en analyse multivariée (tableau 9).

Tableau 9 Analyses univariée et multivariée des caractéristiques des études diagnostique de précharge dépendance pour prédire le score STARD 2003

Caractéristiques	STARD 2003 (écart-type)	Analyse univariée			Analyse multivariée		
		Estimate	SE	P-value	Estimate	SE	P-value
Année de publication	-	0.06	0.10	0.546	0.06	0.12	0.776
Technique utilisée pour définir les réponders							
Pression artérielle	10,35 (2,37)	Referent		0.432	Referent		0.151
Echographie	12,81 (3,31)	2.46	1.55		1.95	1.93	
Thermodilution	11,74 (2,51)	1.40	1.53		0.64	1.95	
Onde de pouls	11,44 (2,66)	1.10	1.63		0.44	2.10	
Bioréactance	11,87 (7,83)	1.52	2.53		4.30	3.52	
Mixte	11,32 (3,06)	0.98	2.53		-0.43	2.83	
Milieu							
Anesthésie	10,70 (2,99)	Referent		0,002*	Referent		0,003*
Soins intensifs	12,70 (2,67)	2.00	0.58		1.48	0.76	
Mixte	12,74 (0,92)	2.04	1.66		1.31	1.95	
Autre	6,33 (NA)	-3.45	2.80		-4.66	2.90	
Source de financement							
Industrielle	11,06 (2,64)	Referent		0.13	Referent		0.560
Institutionnelle	12,10 (3,13)	-0.13	0.90		-0.44	1.03	
Inconnue	12,42 (2,82)	1.06	0.88		0.19	1.03	
Journal							
<i>Anesthesia-Analgesia</i>	11,28 (2,51)	Referent		0,007*	Referent		0.299
<i>British Journal of Anaesthesia</i>	11,41 (2,74)	0.13	1.04		0.65	1.81	
<i>Critical Care Medicine</i>	13,69 (3,12)	2.40	1.19		-0.53	2.92	
<i>Critical Care</i>	10,88 (2,55)	-0.40	1.17		0.08	1.25	
<i>Intensive Care Medicine</i>	13,96 (2,54)	2.67	1.13		-0.33	2.83	
Autre	11,48 (2,92)	0.19	1.01		0.71	2.00	
Nombre de publications par équipe	-	0.12	0.08	0.581	-0.03	0.10	0.836
Impact factor							
<3	11,51 (2,58)	Referent		0,001*	Referent		0.245
>3 et < 5	11,19 (2,62)	-0.32	0.74		0.18	1.68	
>5	13,57 (3,03)	2.06	0.80		2.56	1.84	
Lieu de l'étude							0.800
Amérique	11,54 (4,03)	Referent		0.935	Referent		
Asie	11,89 (1,77)	0.35	1.59		-0.73	1.51	
Europe	12,02 (3,00)	0.48	1.36		-0.39	1.77	
Taille de l'échantillon	-	0.01	0.00	0.217	0.00	0.00	0.777
AUC ROC	-	2.95	2.07	0.158	0.81	2.51	0.753
STARD cité dans instructions aux auteurs							
Non	12,39 (3,01)	Referent			Referent		
Oui	11,03 (2,51)	-1.36	0.62	0,029*	-0.13	1.43	0.928

* p<0,05

4. Discussion

4.1. Résultat principal et son implication

La littérature concernant l'évaluation de la précharge dépendance est abondante. Cependant, pour que son utilisation soit faite avec discernement, il est nécessaire d'en apprécier la qualité. Nous avons souhaité évaluer la qualité méthodologique des études diagnostiques de précharge dépendance. En utilisant les critères STARD, nous avons réalisé une revue de la littérature sur dix ans.

Nous avons étudié 104 études, de 2005 à 2014. Le score STARD 2003 moyen pour tous les items était de 12,0 sur une échelle de 0 à 25, avec des scores minimum et maximum respectivement de 4,8 et 19,5. En moyenne 47,9% des items des recommandations STARD 2003 étaient correctement rapportés.

Le score STARD 2015 est similaire, avec en moyenne 49,8% des items correctement rapportés. Il y a d'ailleurs une bonne corrélation entre les score STARD 2003 et 2015 puisque le coefficient de corrélation linéaire est égal à 0,859.

Ces scores, bien que proches de la moyenne sont assez bas et certains items sont très peu rapportés. Ces résultats étaient attendu puisque ce sont nos doutes concernant à la qualité globale du *reporting* qui nous ont poussés à réaliser ce travail. Cette faiblesse de *reporting* implique différentes choses :

1. Lorsqu'une information manque dans un article publié, se pose la question de savoir si cette information est connue des auteurs ou non. S'agit-il d'une information non disponible, dont les auteurs de l'article ne disposent pas ? Ou bien s'agit-il d'une information qu'ils possèdent ? Dans ce cas, la non-mention de cette information est-elle involontaire ou bien réalisée sciemment avec une volonté de cacher certaines informations ?
2. Nous avons constaté une mauvaise conformité avec STARD dans de nombreuses études de notre revue de la littérature. Souvent, il n'apparaissait pas clairement si l'étude était conforme ou non à l'élément censé être rapporté. Il est montré qu'une

qualité de *reporting* faible est associée à la présence de biais plus nombreux, aussi bien au sein des études diagnostiques (82) que dans les essais contrôlés (107).

3. Les informations réclamées par la grille STARD permettent d'identifier la présence de biais potentiels au sein d'une étude. Leur absence empêche cette analyse.

a. Pour que la qualité des tests (*Integrity test*) soit conservée, il faut que la connaissance du statut du sujet (précharge dépendant ou non) n'influence pas l'évaluation du nouveau test, et vice versa. Même si pour plusieurs tests que nous avons étudiés, l'insu du test de référence n'est a priori pas indispensable, on dit que : « L'intégrité de ce test est inhérente à son fonctionnement ». Il est cependant préférable dans ce cas de le mentionner et dans le cas contraire de veiller à ce que l'aveugle (insu des résultats de l'autre test) soit respecté. Or il n'est mentionné que dans 18% des études analysées. Il est également nécessaire que le délai entre le test étudié et le test de référence et de tout traitement administré entre deux soit précisé. C'est seulement le cas dans 48% des études analysées.

b. Les erreurs sur la référence apparaissent lorsqu'il est impossible de déterminer le statut du patient avec certitude. Pour cela, la justification du test de référence utilisé par les auteurs est importante. Dans notre étude, plusieurs tests différents ont servi de référence pour définir la précharge dépendance. Leur agrément n'est pas parfait et certains sont plus précis que d'autres. Il existe donc un intérêt à bien justifier son utilisation dans le manuscrit. Nous avons considéré comme le *gold standard* la variation de débit cardiaque mesurée par thermodilution (intra ou transpulmonaire) ou par échocardiographie après une expansion volémique. Ces techniques nous paraissaient plus robustes pour détecter des changements de débit cardiaque. La thermodilution mesure le débit cardiaque avec une précision acceptable lorsque trois mesures (par bolus froids) sont répétées (105,106), avec un *least significant change* de 12% pour la mesure de l'index cardiaque, c'est-à-dire inférieur à la limite de 15% habituellement considérée comme cliniquement pertinente. Concernant l'échographie, si elle n'est pas interchangeable avec la thermodilution pour mesurer débit cardiaque, elle l'est pour évaluer une variation (directionnelle) du débit cardiaque (108). Pour limiter les erreurs sur la référence, il est important que le choix du test de référence soit justifié, or il ne l'est que dans

44% des études. De plus nous voulions une justification de son seuil de positivité par la valeur de son *least significant change* (*LSC* ou variation minimale significative), soit en citant une étude l'ayant calculé, soit en l'estimant d'une quelconque manière. Or cette justification n'est retrouvée que dans 22% des études (données non présentées).

- c. Un biais de recrutement peut apparaître lorsque les sujets malades ou non-malades (précharge-dépendant ou non dans le cas présent) ne sont pas représentatifs de la population étudiée. Or, si la population étudiée est bien rapportée dans 72% des études analysées, le recrutement des patients et l'échantillonnage des participants ne le sont que dans respectivement 26% et 33% des études analysées. Il peut alors y avoir un risque de surestimation du paramètre étudié et donc de la précision diagnostique du nouveau test.
 - d. Les données ininterprétables chez certains sujets doivent être prises en compte, dans le cas contraire elles peuvent surestimer la validité d'un test. Si l'on peut supposer que dans les articles étudiés il y a peu de résultats indéterminés, leur présence ou absence n'est mentionnée que dans 25% des études analysées.
 - e. Enfin, certains facteurs influencent l'extrapolation des résultats (biais d'extrapolation). Les caractéristiques démographiques sont bien rapportées (80% des études analysées), mais les personnes réalisant les tests ne sont presque jamais décrites (seulement 9% des études analysées).
4. Le lecteur est impliqué par les manquements dans le *reporting*. En effet, outre le risque d'introduire certains biais, le manque de *reporting* rend l'utilisation du nouveau test pour le médecin plus difficile. Il a besoin des spécifications techniques des tests, qui sont ici plutôt bien rapportées, dans 75% des études analysées. Cette information est nécessaire pour l'utilisation quotidienne du test, mais aussi pour reproduire d'autres études comportant ce test. Dans ce sens, la reproductibilité des tests est une donnée importante, bien décrite uniquement dans 20% des études analysées.

Ces résultats moyens sont à mettre en balance avec des points méthodologiques bien rapportés.

1. Les objectifs des études sont formulés clairement dans 91% dans études. La spécification a priori de l'hypothèse de l'étude limite les chances de faire ressortir des résultats fallacieux à partir de données post hoc ou des conclusions prématurées sur la performance du test. Ce résultat peut cependant en partie être lié au fait qu'une étude

aux objectifs non clairement identifiés avait plus de probabilité d'être exclue de notre revue.

2. Nous l'avons vu, la population étudiée est bien décrite. Les critères d'inclusion, d'exclusion et les caractéristiques démographiques sont bien rapportés. L'identification des patients admissibles est indispensable puisque le test évalué n'a de sens que dans des conditions/circonstances particulières. Ces critères sont liés à l'utilisation future prévue du nouveau test. L'exclusion des patients pour une condition affectant négativement la manière dont le test fonctionne peut conduire à une surestimation de l'efficacité diagnostique (109).
3. De la même manière, les spécifications techniques des tests sont bien rapportées. Des différences dans l'exécution des tests (le nouveau et celui de référence) sont une source potentielle de variation de la précision diagnostique. Il a déjà été observé une variabilité entre des études qui évaluaient la précision d'un même test en raison de différences dans le protocole du test (109)(110).
4. La description de la collecte de données est bien reportée permettant de savoir si elle a bien été planifiée avant ou après que les tests ne soient effectués. En définissant la question d'étude avant l'évaluation des tests, les auteurs peuvent prendre les mesures appropriées pour optimiser le protocole d'étude et la collecte de données. Les études rétrospectives peuvent ne pas identifier tous les patients admissibles et entraîner des données recueillies de moindre qualité ou pour le moins avec des données manquantes.
5. Les caractéristiques démographiques et cliniques de la population sont bien rapportées ; elles sont indispensables car la précision diagnostique d'un test ne s'applique qu'à la population décrite (111). Ces informations permettent au lecteur de juger si l'étude peut traiter adéquatement la question posée et si les résultats s'appliquent à la question posée.
6. La méthode statistique pour mesurer la précision diagnostique et pour quantifier son incertitude est correctement rapporté dans 91% des études analysées. La précision diagnostique en elle-même est bien rapportée dans 69% des études. Puisqu'en définitive, les valeurs de sensibilité et de spécificité du test ne sont que des estimations ; plus le nombre de participants à l'étude est faible, moins les estimations seront précises. L'intervalle de confiance à 95% évalue cette imprécision, son absence doit rendre l'interprétation des résultats prudente.

7. La distribution de la maladie est bien rapportée à 70%. Connaître les différents degrés de gravité de la maladie est indispensable car plus la maladie est sévère plus la sensibilité sera importante. Dans notre cas, le fait que les patients soient « très » précharge dépendant augmentera la sensibilité du test.
8. Enfin, les auteurs rapportent souvent l'application pratique possible du nouveau test évalué puisque l'applicabilité clinique des résultats est discutée dans 88% des études analysées.

Certains items qui nous paraissent importants sont moyennement rapportés (autour de 50%) :

1. Le délai et les traitements entre la réalisation des deux tests est rapporté à 48%. Il existe toujours un délai (plus ou moins important) entre la réalisation du nouveau test et le test de référence. Dans notre cas, ils ne peuvent quasiment jamais être réalisés de manière concomitante. Durant ce laps de temps, la condition du patient peut changer, « passivement », ou bien via la modification des traitements qui lui sont appliqués. Il est important de mentionner l'absence de modification : dans les paramètres de ventilation, dans les posologies des médicaments vasoactifs ou inotropes (si présents), dans le débit des solutés de remplissage, dans la position du patient, etc. De tels changements, lorsqu'ils surviennent, peuvent modifier l'agrément entre les deux tests et induire une estimation biaisée de leur performance.
2. Un tableau croisé des résultats entre le nouveau test et le test de référence, rapporté à 52%, permet au lecteur de recalculer les mesures de précision du diagnostic. Sa présence est indispensable pour rendre les travaux de recherche reproductibles et vérifiables par d'autres. Ces données permettent aussi d'effectuer des analyses complémentaires ou des méta-analyses. Pour les données continues, un graphique représentant les valeurs du nouveau test en fonction des valeurs du test de référence permet de vérifier la distribution de ces valeurs. Il est possible d'observer une distribution bimodale, avec deux groupes prédominants, des patients « absolument pas » précharge dépendants et des patients « très » précharge dépendants (illustration en figure 19). Une distribution bimodale peut faire évoquer un biais de sélection : les patients de la zone grise sont exclus, volontairement ou non, ce qui améliore artificiellement la capacité discriminative du test et son aire sous la courbe ROC (Annexe 4 : Correspondance dans la revue *Anesthesiology*).

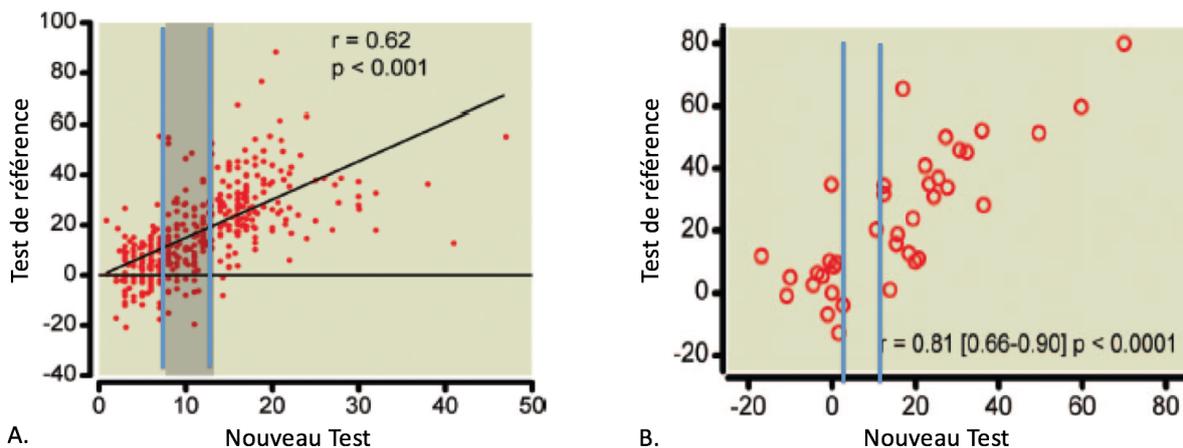


Figure 19 Exemple de distribution. A : la distribution des patients apparait homogène, B : la distribution des patients est bimodale, avec deux groupes prédominant, des patients « absolument pas » précharge dépendants et des patients « très » précharge dépendants selon le nouveau test.

Nous nous attendions à observer une amélioration du *reporting* au cours des dix ans, cela n'a pas été le cas. Bien que les moyennes des scores STARD soient restées stables, deux critères ont évolué au cours des dix ans : la présence d'une analyse croisée entre les résultats des tests étudiés et les tests de référence s'est réduite, tandis que les performances diagnostiques et les mesures d'incertitudes des tests étaient mieux rapportées.

STARD 2015 a introduit neuf nouveaux items (93). Il est intéressant de noter que le taux de *reporting* de ces neuf nouveaux items semble meilleur (56,7%) que les scores globaux.

4.2. Hypothèses explicatives de nos résultats

Même s'il est difficile d'établir la ou les causes de nos résultats, nous pouvons formuler plusieurs hypothèses ou questions.

Tout d'abord, nous pouvons nous demander si la méthodologie des études diagnostiques est parfaitement connue des auteurs ? S'il existe des lacunes méthodologiques, il est logique qu'il en découle des lacunes dans le *reporting* des points de méthodologie.

Il existe très probablement une méconnaissance des critères STARD pour une partie des auteurs d'études diagnostiques sur la précharge dépendance. Il semble en tout cas que les critères STARD aient moins de notoriété que les critères CONSORT pour les essais randomisés. C'est ce que l'on pourrait déduire du fait que les critères CONSORT sont mieux

rapportés, que ce soit dans la littérature oncologique (112) ou que dans celle oto-rhino-laryngologique (113).

Concernant les recommandations faites aux auteurs de la part des revues, se pose la question de la présence ou non des recommandations STARD. Sont-elles évoquées, recommandées, expliquées et si oui, comment ? Si des biais méthodologiques peuvent justifier la non-publication d'un article dans une revue, un manque d'informations sur la méthodologie pourrait justifier que des corrections soient apportées au manuscrit.

En 2007, N Smidt et ses collègues (114) ont déterminé dans quelle mesure les journaux ont adopté la déclaration STARD et incorporé les recommandations dans leurs « instructions pour les auteurs ». Ils se sont intéressés aux 50 principales revues qui publient fréquemment des études de précision diagnostique. La déclaration STARD a été mentionnée dans les instructions pour les auteurs dans 19 des 50 revues (38%). Dans les revues ayant adopté STARD, ils notaient une large variation dans les indications expliquant comment utiliser la déclaration STARD. Cinq revues ont clairement décrit leurs attentes concernant l'utilisation de la déclaration STARD, demandant aux auteurs de soumettre la liste de contrôle STARD et / ou d'inclure un diagramme de flux dans leur manuscrit. Mais dans plupart des revues l'ayant adopté, STARD est seulement mentionné, sans description des attentes concernant l'utilisation de la grille. Cette variabilité pourrait être une raison du *reporting* incomplet encore observé dans les revues majeures. Ils ont interrogé les rédacteurs de journaux « non adoptifs ». Les principaux arguments avancés justifiant la non-adoption des critères STARD sont : leurs examinateurs et leurs rédacteurs sont bien adaptés aux éléments énumérés dans la liste de vérification STARD et leurs revues publient relativement peu d'études diagnostiques. Par ailleurs, certains éditeurs jugeaient qu'ils en demandaient déjà trop aux auteurs, pour ne pas ajouter une checklist et un diagramme de flux STARD. En plus d'inclure la déclaration STARD dans les instructions pour les auteurs, N Smidt concluait que les revues pouvaient améliorer la qualité du *reporting* en recherche diagnostique en communiquant plus clairement leurs attentes concernant l'utilisation de la déclaration STARD et en fournissant aux auteurs une version imprimable des instructions. La liste de contrôle STARD devait être utilisée comme outil au long du processus d'évaluation du manuscrit, à la fois par les auteurs et par les examinateurs.

Plus récemment en 2016, M T Simset et son équipe (115) ont analysé les exigences en matière de *reporting* qui étaient contenues dans les instructions pour les auteurs des revues de médecine d'urgence. Il s'agissait de déterminer si chaque journal nécessite, recommande, ou

ne mentionne pas les consignes de *reporting* suivantes : EQUATOR Network, ICMJE, ARRIVE, CARE, CONSORT, STARD, TRÉPIED, CHEERS, MOOSE, STROBE, COREQ, SRQR, SQUIRE, PRISMA-P, SPIRIT, PRISMA et QUOROM. Sur les 27 revues, 40,7% n'ont pas mentionné une seule recommandation, les 59,3% restantes ont mentionné une ou plusieurs recommandations. Les directives ICMJE (*International Committee of Journal Medical Editors*) (66,7%) et la déclaration CONSORT (55,6%) ont été mentionnées le plus souvent. Concernant les critères STARD, ils étaient mentionnés dans 40,7% des journaux. Les auteurs estimaient que le fait qu'autant de revues ne mentionnent aucune recommandation nuisait aux autres revues pour améliorer l'intégralité et la transparence du *reporting* dans la recherche.

Parmi les journaux que nous avons inclus dans notre étude, seulement trois citaient les critères STARD dans leurs recommandations faites aux auteurs (cela représente 29,8% des études incluses). Pour *Anesthesia Analgesia* il s'agit d'une recommandation « *should be* » et non d'une obligation, de même pour *Critical Care* qui recommande fortement « *strongly recommend* » leur utilisation, enfin *European Journal of Anaesthesiology* est plus catégorique puisque l'utilisation des critères STARD est requise « *authors are requested* ». Les critères et leur explication ne sont pas détaillés, un lien vers le site du groupe STARD est proposé pour télécharger la grille et les instructions.

Lorsqu'on observe les études des auteurs qui publient beaucoup, le *reporting* n'était pas meilleur, or ce sont souvent ces auteurs qui participent au *reviewing* des études de précharge dépendance puisqu'ils sont experts sur le sujet. On peut supposer que peu de *reviewers* exigent la présence des critères STARD dans les manuscrits.

Il nous paraît difficile d'émettre des hypothèses concernant le taux de *reporting* de chacun des items de la grille STARD en particulier. Nous nous sommes intéressés à certains items apparaissant dans la nouvelle grille STARD 2015. Les items 2 (« un résumé structuré »), 3 (« le contexte scientifique et clinique »), 4 (« objectifs de l'étude et hypothèses ») et 27 (« implications pour la pratique ») sont particulièrement bien rapportés. Cela pourrait indiquer que, malgré leur absence des recommandations STARD 2003, les auteurs avaient déjà adopté certaines bonnes pratiques dans le *reporting* des études diagnostiques de précharge dépendance. Cependant, certains nouveaux items STARD ont été rapportés peu fréquemment et nécessitent une amélioration : l'item 18 (« taille prévue de l'échantillon et comment elle a été calculée »), l'item 28 (« numéro d'enregistrement et nom du registre ») et l'item 29 (« où le protocole d'étude complet peut être consulté »). Le calcul de la taille prévue de l'échantillon

est généralement rapporté dans les essais randomisés, une revue publiée en 2009 (116) montrait que seulement 5% des articles n'avaient rapporté aucun calcul de la taille de l'échantillon, alors que 53% avaient rapporté tous les paramètres requis pour le calcul. Pourtant, le calcul de la taille de l'échantillon s'applique également aux études diagnostiques. Concernant l'enregistrement de l'étude, il est obligatoire pour les essais randomisés, avec un taux d'enregistrement de 95.8% (117). Le faible taux d'enregistrement rapporté est peut-être dû au fait qu'il n'est pas requis par l'ICMJE. Il n'y a pas de registre spécifique aux études diagnostiques, elles sont enregistrées sur le site ClinicalTrials.gov (118). Ces résultats concernant les nouveaux items de la grille sont très similaires à ceux retrouvés par PJ Hong et son équipe (119)

Un autre facteur explicatif de score bas est qu'il existe parfois une confusion entre études physiologiques et diagnostiques. Des études forment leur problématique comme l'explication d'un mécanisme physiologique alors qu'il s'agit d'une étude diagnostique. Cela a une influence négative sur la méthodologie. Nous avons exclu les études physiologiques mais conservé les études diagnostiques non identifiées comme telles.

4.3. Validité externe

Nous avons comparé nos résultats à ceux d'études utilisant les critères STARD pour évaluer la qualité du *reporting* dans des articles d'autres disciplines.

Tout d'abord, des études ont observé s'il y avait une évolution du score entre avant et après la parution de la première version des recommandations STARD.

En médecine de la reproduction, SFPJ Coppus et ses collègues (120) ont analysé 27 études en 1999 et 24 en 2004, avec un nombre moyen d'items déclarés allant de 12,1 en 1999 à 12,4 en 2004. Moins de la moitié des études avaient rapporté au moins 50% des items, il y avait une grande variation dans le *reporting* des items, il n'était pas retrouvé d'évolution entre avant et après la publication des critères STARD. Ces résultats sont en accord avec les nôtres.

N Smidt (121) et ses collègues ont analysé 124 articles publiés en 2000 et 141 articles publiés en 2004 dans 12 revues médicales (six revues généralistes, six revues spécialisées). Le nombre moyen d'item STARD déclarés était de 11,9 en 2000 et de 13,6 (de 4,0 à 21,0) en 2004, soit une augmentation de 1,81 items (IC à 95% : 0,61 à 3,01). Les auteurs notent donc une légère amélioration dans la qualité du *reporting* avec encore une importante possibilité d'amélioration. Ils encouragent les éditeurs à mentionner l'utilisation de la déclaration STARD

comme une exigence dans leurs directives pour les auteurs et demandent aux *reviewers* de vérifier la présence des items STARD.

NL Wilczynski (122) a analysé 240 études publiées dans 12 revues (six soutenant les critères STARD, six ne les soutenant pas), pour les années 2001, 2002, 2004 et 2005. Elle ne retrouvait pas d'interaction significative entre le score total moyen et les journaux STARD vs non-STARD, de même pour l'année de publication. Le *reporting* des items est resté relativement constant avant et après la publication des recommandations STARD. A noter que contrairement à nous, N L Wilczynski n'a utilisé que 13 des 25 items.

D'autres auteurs ont simplement voulu évaluer la qualité du *reporting* les études diagnostiques de leur spécialité.

TJ Selman et ses collègues (123) ont analysé 300 études (195 en obstétrique, 105 en gynécologie). Les études obstétriques avaient un *reporting* supérieur à 50% pour 62,1% (18/29) des items, contre 51,7% (15/29) pour les études gynécologiques, avec un meilleur *reporting* en obstétrique. Ils notent une corrélation positive entre la taille de l'échantillon et la qualité du *reporting*, ainsi qu'une amélioration au fil du temps, mais cette donnée correspond à une évolution sur le long terme puisque l'analyse s'étend sur 40 ans (de 1977 à 2007) avec seulement 17% des études publiées après l'apparition des critères STARD.

En 2017, L Gallo et son équipe (124) ont déterminé dans quelle mesure les études diagnostiques publiées dans 8 revues spécialisées de médecine d'urgence respectaient les critères STARD 2003. Ils ont sélectionné au hasard et ont analysé 20% des 114 études incluses entre 2010 et 2014. Un peu plus de la moitié des critères STARD sont rapportés dans plus de 80% d'études.

En 2016, ATM Grob et son équipe (125) ont évalué les études diagnostiques d'échographie 3D du plancher pelvien, avec les directives STARD. Quarante études ont été analysées, le score moyen des articles inclus était de 16,0 sur un maximum de 22 points (3 items avaient été écartés), soit 73%. Ce score est meilleur que celui que nous retrouvons. La corrélation inter-observateur était de 0,77 (ICC), il n'y avait pas de différence de score entre les journaux ayant adopté ou non STARD. L'amélioration entre les périodes 2003-2009 et 2010-2015 n'est pas significative, ce résultat est similaire au notre, sur une période d'étude très similaire.

Au regard de ces différentes études, nos résultats concernant l'adhérence aux critères STARD 2003 apparaissent similaire ou très légèrement inférieurs à ceux d'autres auteurs, dans des disciplines différentes.

La mise à jour des critères STARD faite fin 2015 est récente et peu de revues les ont utilisés pour l'instant.

Très récemment en 2017, PJ Hong et son équipe (119) ont évalué l'adhésion à STARD 2015 des études diagnostiques en imagerie publiées en 2016. Ils ont également cherché à identifier les différences dans le *reporting* pour les études d'imagerie par résonance magnétique (IRM). L'adhésion (n=142 études) était de 55%, sans différence importante pour les études sur l'IRM. Ces résultats sont meilleurs que ceux que nous retrouvons. Concernant leurs résultats secondaires, de manière similaire à ce que nous retrouvons, la description du nouveau test (et son interprétation) ont été fréquemment rapportées (> 66% des études). Leurs résultats concernant les items rapportés peu fréquemment sont semblables aux nôtres, en particulier en ce qui concerne la généralisation des résultats (par exemple les lieu et date de l'étude) et l'évaluation des biais (par exemple l'aveugle des évaluateurs). Concernant les nouveaux articles STARD 2015 : le calcul de la taille de l'échantillon, le lieu où le protocole d'étude complet peut être consulté et l'enregistrement ont été signalés peu fréquemment, ce qui a été le cas pour nous aussi. Ils obtiennent un meilleur *reporting* parmi les revues issues d'un journal impact factor plus élevé. Enfin, les journaux ayant adopté STARD ont rapporté plus d'éléments que les autres. Notre analyse univariée fait également ressortir une association entre le score STARD et l'impact factor de la revue ou l'adoption des critères STARD par la revue. Cette association disparaît en analyse multivariée.

YJ Choi et son équipe (127) ont identifié une adhésion plus importante à STARD 2015 dans des études d'imagerie, avec un *reporting* moyen de 20/27 (74%). Il est important de noter que leur évaluation s'intéressait à des études plus récentes (2011 à 2015), dans un seul journal, et utilisait une liste de contrôle STARD 2015 tronquée.

En 2014, DA Korevaar et ses collègues ont publié ce qui nous semble être la première méta-analyse s'intéressant à la qualité du *reporting* des critères STARD dans les études diagnostiques (91). Ils ont inclus 16 études, pour analyser 1496 articles au total. Trois études ont étudié l'adhésion dans un échantillon général d'études diagnostiques les autres l'ont fait dans un domaine de recherche spécifique. Le nombre moyen total d'items déclarés variait de 9,1 à 14,3 parmi les 13 études qui ont évalué les 25 items STARD. Six études se sont intéressées à l'évolution qualitative des articles post-STARD et pré-STARD. La méta-analyse a révélé une augmentation modeste mais significative de l'adhérence après l'introduction de STARD, avec une différence moyenne de 1,41 item (IC 95% 0,65 à 2,18).

Cette méta-analyse comporte d'autres informations. La plupart des études incluses ont révélé un agrément inter-évaluateur "substantiel" concernant les items, mais avec des différences marquées entre les études dans les fréquences de *reporting* de certains items spécifiques. Ils retrouvent une hétérogénéité considérable dans l'adhérence pré-STARD et post-STARD, ils l'expliquent, au moins en partie, par des différences entre les études dans la notation de certains items. Seuls sept études ont spécifiquement rapporté à quelle fréquence un item a été jugé non applicable aux articles évalués, tandis que les autres ne l'ont pas fait. Pour eux il était difficile de dire si les différences entre les études dans les scores des items ont été causées par une grande diversité d'adhésion dans les domaines de recherche respectifs ou bien par hétérogénéité dans les méthodes de notation ou les deux.

4.4. Forces et faiblesses de notre étude

4.4.1. Les aspects positifs

4.4.1.1. Le contexte

Le contexte dans laquelle notre étude a été réalisée la rend d'autant plus pertinente, puisque le nombre d'études publiées dans le domaine des tests diagnostiques augmente. De 2700 études référencées sur PubMed en 1996, nous sommes passés à 6200 en 2006 et 12800 en 2016 pour la recherche « *diagnostic accuracy* ». Les tests diagnostiques sont indispensables dans la gestion des patients car de nombreuses décisions cliniques dépendent de leurs résultats. L'utilisation correcte d'un test dans un contexte clinique donné devrait s'appuyer sur une évaluation de son coût, de sa sécurité, de ses performances cliniques et de son utilité. Des études de précision diagnostique de haute qualité sont cruciales dans cette perspective. Il est probable que par rapport à d'autres formes de recherche, les études de précision diagnostique sont probablement plus sensibles aux biais (82,128).

En anesthésie-réanimation, l'évaluation de la précharge dépendance est une pierre angulaire de la prise en charge des patients. De nombreux outils ont été développés pour la diagnostiquer, la littérature sur le sujet s'enrichit très régulièrement. La qualité de ces études est donc primordiale, cependant elle n'a pas été évaluée jusqu'à présent. À notre connaissance, il s'agit de la première étude s'intéressant à la conformité des études de

précharge dépendance aux critères STARD. Ce travail est novateur en ce sens et trouve son utilité dans le premier aperçu qualitatif sur ce sujet.

4.4.1.2. La méthodologie

La méthodologie de notre étude s'appuie sur des recommandations internationales concernant les revues de la littérature : PRISMA. Nous avons pu rapporter 24 éléments de méthodologie sur les 27 demandés soit 88,9 % (annexe 1).

Afin de réduire le biais de sélection des études analysées, la sélection des études à partir des résumés a été réalisée de manière indépendante par deux auteurs. En cas de désaccord, une discussion avait lieu et un consensus était trouvé.

Afin de réduire le biais d'observation, la qualité du *reporting* a été évaluée indépendamment et en aveugle par deux investigateurs sur la première année d'étude (2005). Le taux de divergence était de 9,3%, et le taux d'erreur du principal extracteur de données de 4,3%, ce qui est satisfaisant. Ce résultat est conforme aux résultats de N Smidt qui retrouve un agrément inter-opérateur global de 85% (129). N Smidt justifie ces désaccords inter-opérateurs non pas par des différences d'interprétation des items par les examinateurs, mais par des difficultés à repérer la présence des items en raison d'un manque de clarté du manuscrit. Nous avons pu constater ce phénomène au cours de notre recueil.

Notre étude s'est appuyée sur les deux scores STARD publiés, la première version de 2003 et sa mise à jour et amélioration de 2015. Nous avons choisi comme critère de jugement principal le score STARD 2003 (69) car toutes les études analysées se sont déroulées après la première publication du score, mais avant sa mise à jour. Les auteurs ne pouvaient donc pas tenir compte du nouveau score dans leur méthodologie. Cependant nous avons aussi calculé les scores STARD 2015 (93), et ce de manière très précoce. En effet l'utilisation de cette mise à jour est maintenant recommandée par la plupart des revues ayant adopté STARD, en faisant d'elle la nouvelle référence pour l'élaboration et l'évaluation méthodologique des études diagnostiques. Si une nouvelle évaluation est faite dans plusieurs années, l'utilisation du score STARD 2015 permettra désormais de comparer les scores afin d'y déceler une évolution.

L'analyse a porté sur dix ans, ce qui nous paraît suffisant pour détecter une évolution du score si celle-ci existe.

4.4.1.3. Les résultats

Nous avons inclus 104 études, sur une période de 10 ans, ce qui est conforme à ce que nous avons prévu. Ce nombre nous paraît suffisant pour avoir un aperçu qualitatif du *reporting* dans les études de précharge dépendance. De plus, les études que nous avons analysées sont publiées dans un large éventail de revues pour plus d'hétérogénéité.

Les lacunes identifiées pourraient guider les auteurs, les éditeurs et les *reviewers* concernant la façon d'optimiser le *reporting* des études diagnostiques de précharge dépendance.

Les résultats intermédiaires de notre étude ont déjà subi une certaine validation scientifique puisqu'ils ont été acceptés puis présentés sous forme de résumé lors de deux congrès. Ce résumé (annexe 2) a été présenté lors du congrès de la Société de Réanimation de Langue Française (SRLF) à Paris en janvier 2017, celui-ci (annexe 3) lors du congrès européen d'anesthésie *Euroanaesthesia* à Genève en juin 2017. Ces deux présentations ont permis de commencer à interpeller d'autres cliniciens sur le sujet.

Nous espérons influencer positivement la littérature sur ce sujet par la publication de notre travail. Nous avons déjà adressé une lettre, en cours de publication, à la revue *Anesthesiology* concernant une étude de M Biais publiée en février 2017 (60). Les auteurs suggèrent que l'amplitude de la diminution du volume d'éjection systolique pendant une manœuvre de recrutement pulmonaire pourrait prédire la précharge dépendance chez les patients en ventilation mécanique au bloc opératoire. Nous tenions cependant à indiquer des manquements méthodologiques. Des items clés de la grille STARD (qui n'est pas mentionnée) ne sont pas rapportés : l'aveugle, le diagramme de flux, la reproductibilité, de rationnel dans le choix du test de référence et le choix de son seuil. L'étude se retrouve ainsi exposée à un risque de biais important. La lettre est présentée en annexe 4.

4.4.2. Les limites

Notre étude comporte plusieurs limitations.

La première concerne la méthode de notation. Nous avons décidé de diviser en deux les items concernant à la fois le nouveau test et le test de référence, en accordant 0,5 point si l'item était bien rapporté, pour chacun des deux tests, il s'agissait des items 8, 9, 10, 11, 13 et 24. De plus, litem 3 concernant la population étudiée a été divisée en 3 (0,33 points par tiers d'item), l'item 4 a quant à lui été divisé en 4 (0,25 point par quart d'item bien rapporté). Certains auteurs ont employé un système de notation permettant de noter les items partiellement

rapportés (130), un item totalement rapporté valant 2 points, un item partiellement rapporté valant 1 point. Nous avons choisi de seulement accorder 0 ou 1 point par item, en étant « sévère » puisqu'un item partiellement rapporté n'entraînait aucun point. Une évaluation à trois possibilités (0, 1 ou 2) aurait sans doute fourni plus de détails.

Une deuxième limitation de notre étude est liée la subjectivité de certains items qui sont ouverts à l'interprétation de l'examineur. Nous avons essayé de minimiser cette limitation en définissant chaque élément en détail, en faisant des essais sur quelques études, en ajustant à nouveau nos définitions pour qu'elles soient claires et explicites, cela au cours de plusieurs réunions de consensus.

Troisièmement, nous avons conservé les items 13, 23 et 24 portant la notions « si réalisé » (« *if done* ») malgré le fait qu'il ne soit pas toujours présents dans les études. Leur absence était systématiquement notée 0, ce qui a forcément abaissé les scores globaux. Des études similaires à la nôtre avaient choisi d'exclure ces items, puisque non applicables à toutes les études (125,126), d'autres avaient choisi d'attribuer systématiquement le point aux études pour lequel l'item n'était pas applicable (119). Cependant nous avons préféré garder l'ensemble des items pour être plus proche de la grille originale, quitte à pénaliser certaines études.

Quatrièmement, nous avons légèrement modifié la grille originale en précisant la définition de certains items et en exigeant la présence de certains points méthodologiques absents de la grille originale. Nous avons choisi d'adapter certains critères à notre champ de recherche afin de diminuer la subjectivité du recueil. Les auteurs des directives STARD suggèrent que l'on peut adapter la grille à son champ de recherche.

Cinquièmement, notre travail peut paraître tronqué. Il exclut de nombreuses études publiées dans des revue d'*impact factor* inférieur à deux, ainsi que toutes les études non référencées dans PubMed, puisqu'il s'agissait de notre seule base de données. Nous ne souhaitons pas réaliser une revue exhaustive, mais plutôt nous focaliser sur un échantillon représentatif d'études qui influencent réellement les cliniciens dans leur pratique.

Sixièmement, le choix de nos critères d'exclusion est critiquable. En se limitant aux études en anglais et chez l'adulte, nous avons réduit le nombre de manuscrits sélectionnés. Il s'agit d'un choix pratique, mais nous ne pensons pas qu'il modifie nos conclusions. Par contre, nos conclusions auraient sans doute été modifiées en intégrant tous les journaux d'anesthésie réanimation, sans distinction d'*impact factor*. Au début de la période de recueil, des études publiées dans des revues de faible impact factor avaient été étudiées. L'évaluation était

parfois rendue difficile par le manque de clarté du manuscrit, ne permettant pas de déterminer la présence des items. Nous pouvons penser a priori, que les journaux à impact factor élevé ont des exigences méthodologiques et de *reporting* plus strictes. Peut-être que le rejet des études diagnostiques mal rapportées par les journaux de haute qualité peut expliquer leur publication dans les revues d'*impact factor* inférieur. Les études ainsi exclues auraient sans doute eu des scores de *reporting* équivalents, ou bien plus faible.

Enfin, nous n'avons pas calculé *a priori* le nombre d'études à inclure. En fixant la limite d'*impact factor* à deux, nous espérons atteindre une centaine d'études analysées, ce qui fut le cas. Nous avons prévu d'analyser onze caractéristiques d'études. Pour que la taille de l'échantillon soit suffisante pour pouvoir procéder à une régression logistique, on suggère au minimum 10 observations par variable indépendante analysée (131).

4.5. Nouvelles perspectives

Nous pensons que l'utilisation des critères STARD est une nécessité pour les études diagnostiques, en particulier en anesthésie-réanimation. Il faut une transmission de l'information les concernant, à la fois sur leur existence et sur leur utilisation. Cette information doit être globale, d'abord auprès des éditeurs de journaux et des *reviewers*. L'ensemble des journaux devraient requérir le score STARD pour toutes les études diagnostiques qu'ils publient. Le score en lui-même ainsi que la façon de l'utiliser devraient ainsi être présents dans les recommandations aux auteurs. Enfin, lors des premiers échanges entre les auteurs et les *reviewers*, les critères STARD devraient être évoqués et si besoin expliqués.

Du côté des auteurs, qu'il s'agisse des professeurs qui encadrent les travaux de recherche ou des personnes qui effectuent le recueil de données, les médecins et les internes, les aspects méthodologiques des travaux devraient être au centre des discussions. Aujourd'hui les chercheurs disposent d'une plateforme sur internet qui propose les outils nécessaires à l'élaboration rigoureuse de leurs travaux. Le réseau EQUATOR, pour *Enhancing the Quality and Transparency Of health Research* (132), propose une bibliothèque contenant une base de données consultable complète sur les recommandations en matière de *reporting* et des liens vers d'autres ressources concernant la manière d'exposer ses recherches. Les outils (133) proposés s'adressent à la fois aux auteurs, aux éditeurs, ainsi qu'aux professeurs enseignant la recherche. Ils couvrent les principaux types d'étude, proposant ainsi les recommandations

CONSORT (92) pour les essais randomisés, ou les critère STARD (134) pour les études diagnostiques.

Enfin, tous les internes étant amenés à réaliser des travaux de recherche (mémoire, thèse ...) nous pensons que la formation concernant le *reporting* des différents types d'études pourrait se faire lors des cours aux internes, aussi bien via les cours de DES que lors de la formation hospitalière.

Concernant la suite de ce travail, il devra faire l'objet d'une publication scientifique. Les résultats préliminaires ont déjà été présentés lors du congrès de la SRLF en janvier 2017 et lors du congrès de l'ESA en juin 2017 (annexes 2 et 3).

Une autre suite logique à ce travail serait une nouvelle évaluation dans plusieurs années afin de rechercher une évolution, si possible une amélioration, dans la qualité du *reporting* des études de précharge dépendance. Le fait d'avoir dès maintenant évalué la mise à jour de 2015 des critères STARD permettra d'avoir un outil de comparaison récent avec une évaluation future.

Enfin, comme nous l'avons fait avec la lettre adressée à M. Biais dans *Anesthesiology* (annexe 4), nous garderons une vigilance particulière pour influencer positivement la littérature à venir sur ce sujet.

5. Conclusion

De nombreux outils ont été développés pour diagnostiquer la réponse au remplissage vasculaire et la littérature sur le sujet est abondante. Cependant, même si les études diagnostiques semblent plus faciles à réaliser que les études thérapeutiques, leur qualité méthodologique est plus faible. Nous rapportons dans ce travail de thèse une revue systématique de la littérature sur une période de dix ans afin d'étudier la qualité du *reporting* des études évaluant la réponse au remplissage vasculaire.

Nous avons utilisé les critères STARD, pour « *STAndards for Reporting Diagnostic accuracy* », une liste de 25 items rassemblant l'ensemble des éléments qui devraient être rapportés dans les études diagnostiques. Ces critères ont été publiés en 2003, puis mis à jour en 2015. Nous avons recherché dans PubMed les études évaluant la capacité de prédire la réponse au remplissage vasculaire ou à diagnostiquer la précharge dépendance. Après avoir élaboré une définition consensuelle pour chaque item de la liste STARD, nous avons apprécié la présence de chacun d'entre eux dans toutes les études sélectionnées. Pour chaque étude, nous avons ainsi obtenu un score de qualité qui s'étend de 0 à 25 pour les critères STARD 2003 et de 0 à 30 pour les critères STARD 2015. Les caractéristiques des études associées à un meilleur *reporting* ont été étudiées.

Sur les 725 études initialement sélectionnées, 104 répondaient aux critères d'inclusion et d'exclusion et ont été analysées. Il s'agissait majoritairement d'études européennes, publiées dans des revues dont l'impact factor moyen est de 4,8 (IC 95% : 4,4-5,0), 50% de ces études ayant été publiées dans le *British Journal of Anesthesia*, *Intensive Care Medicine* et *Critical Care*. La majorité (67,3%) des tests évalués étaient basés sur les interactions cœur-poumons.

Le score STARD 2003 moyen pour tous les items était de 12,0 sur une échelle de 0 à 25 (IC 95% : 11,4-12,5). Les scores minimum et maximum étaient respectivement 4,8 et 19,5 sur 25. En moyenne 47,9% des items étaient correctement rapportés. Le score STARD 2015 moyen pour tous les items était de 14,9 sur une échelle de 0 à 30 (IC 95% : 14,3 à 15,5). Les scores minimum et maximum étaient respectivement 6,5 et 21,5 sur 30. En moyenne 49,8% des items étaient correctement rapportés. Le calcul de la corrélation entre les scores STARD 2003 et STARD 2015 montrait un coefficient de corrélation linéaire ($r=0,859$).

Les moyennes des scores STARD 2003 et STARD 2015 sont restées stables sur les dix années étudiées. Seuls deux critères STARD 2003 ont évolué au cours du temps. La présence d'une analyse croisée entre les résultats des tests étudiés et les tests de référence était moins rapportée, passant de

66,7% en 2005 à 44,4% en 2014. A l'inverse, les performances diagnostiques et les mesures d'incertitude des tests étaient mieux rapportées, passant de 55,5% en 2005 à 100% en 2014.

En analyse univariée, les caractéristiques des études associées à un meilleur score STARD 2003 étaient : le milieu dans lequel se déroule l'étude, le journal de publication, l'*impact factor* du journal, et la présence de STARD dans les recommandations faites aux auteurs de la revue. En analyse multivariée, seul le fait que l'étude se déroule dans une unité de soins intensifs était associé à un meilleur score STARD 2003. De manière inattendue, l'*impact factor* des revues ou le nombre de publications de l'auteur sur le sujet n'étaient pas associés à un meilleur *reporting* en analyse multivariée.

En conclusion, les scores STARD 2003 et STARD 2015 sont bas. Plusieurs items sont insuffisamment reportés. La qualité du *reporting* dans les études évaluant la précharge dépendance et la réponse au remplissage vasculaire est globalement faible et ne semble pas progresser au fil du temps. L'amélioration de la qualité des études diagnostiques est un prérequis fondamental à la détection des états pathologiques et à l'administration des thérapeutiques adéquates en pratique clinique. Nous recommandons l'utilisation systématique des critères STARD dans l'élaboration et le *reporting* des futures études évaluant la précharge dépendance et la réponse au remplissage vasculaire.

Le Président de la thèse,

Nom et Prénom du Président

Signature

R. Feuthi Jean Luc

Vu :

**Pour Le Président de l'Université
Le Doyen de l'UFR de Médecine Lyon Est**



Gilles Rode

Professeur Gilles RODE



Vu et permis d'imprimer

Lyon, le **06 JUIL. 2017**

6. Annexes

6.1. Adhésion à PRISMA

Section/Topic	#	Checklist Item	Reported Yes / No
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	Yes
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	Yes
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	Yes
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	Yes
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	No
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	Yes
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	Yes
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	Yes
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	Yes
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	Yes
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	Yes
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	Yes
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	Yes
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	No
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	Yes
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	Yes
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	Yes
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	Yes
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome-level assessment (see Item 12).	Yes
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group and (b) effect estimates and confidence intervals, ideally with a forest plot.	No
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	Yes
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	Yes
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	Yes
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., health care providers, users, and policy makers).	No
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review level (e.g., incomplete retrieval of identified research, reporting bias).	Yes
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	Yes
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	Yes

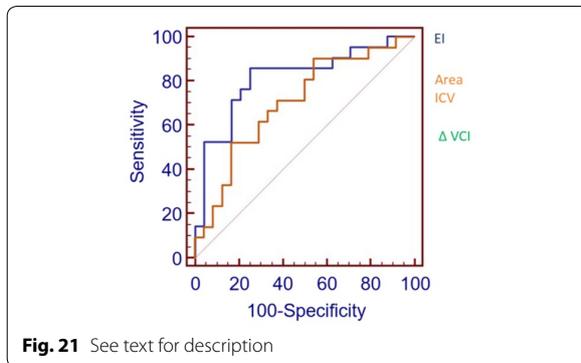
doi:10.1371/journal.pmed.1000097.t001

Total : 24 / 27 (88,9%)

6.2. Résumé présenté à la SRLF 2017

Ann. Intensive Care 2017, 7(Suppl 1):7

Page 188 of 207



the necessity of being very perpendicular at the risk of falsifying the EI. For now it is just a static index but an ovoid vena cava (EI: 0.85) may be predicative of fluid responsiveness. The strangely big size of the VCI can be attributed to a deformation of the in-depth image because of the probe but this should not modify in theory the report of EI. It is a question in the future of testing this parameter at the spontaneous breathing patient, study his variability with respiration and strengthening its external validity.

Competing interests

None.

S88

Quality of reporting of fluid responsiveness evaluation studies: a five year systematic review

Izaute Guillaume¹, Matthias Jacquet-Lagrèze¹, Jean-Luc Fellahi¹

¹Anesthésie réanimation, Hôpital Louis Pradel, Bron, France

Correspondence: Matthias Jacquet-Lagrèze - matthias.jl@gmail.com

Annals of Intensive Care 2017, **7**(Suppl 1):S88

Introduction The Standards for the Reporting of Diagnostic Accuracy Studies (STARD) statement was developed in 2003 to improve the quality of reporting of diagnosis studies (1). STARD is a list of 25 essential items, it was updated in 2015 (2). Fluid responsiveness evaluation is a cornerstone of hemodynamic management of patient with circulatory failure. Numerous studies have evaluated tools for assessing fluid responsiveness. The purpose of this work was to evaluate the quality of reporting of studies assessing fluid responsiveness, using the STARD criteria.

Materials and methods We used MEDLINE via PubMed to search all publication of studies assessing the ability to predict or diagnose fluid responsiveness in a perioperative or circulatory failure context between the 1st of January 2004 and the 1st of January 2014. We have presented herein the last 5 years (2010–2014). Paediatric studies, publication with no abstract available or in another language than English, review, meta-analysis, and journals with impact factor less than 2 were excluded. We have checked these studies for all the 25 STARD criteria. The rating methodology of each item was discussed and double checked (GI and MJL). The 2003 and 2015 STARD quality score (SQS) was the integer of the present item. Univariate and multivariate analysis were conducted to find out if some characteristic of the studies were linked with a better SQS. Following items were studied: year of publication, positive or negative study, location of study, expert team, funding source, conflict of interest, the use of a new device, instruction to the author including STARD statement, impact factor of the journal. Every covariate associated with a p -value lower than 0.20 in the univariate analysis were used in the multivariate analysis. Statistical analysis was performed with R software. All the test were two-sided, $p < 0.05$ was considered statistically significant.

Results After a double screening (GI, MJL) of 430 articles (2010–2014), 97 were selected, then 46 were excluded because of an IF < 2. Fifty-one

studies were selected and reviewed. The mean 2003 overall quality score was 11.85 on a scale of 0–25, whereas the mean 2015 overall quality score was 15.34 on a scale of 0–30. The 2003 SQS remained stable between 2010 and 2014 ($p = 0.173$). Correlation between 2003 and 2015 SQS was good ($r = 0.83$; $p < 0.001$). Some items are insufficiently reported, especially: participant sampling (adequately reported in 19.6% of studies), executing and reading the tests (11.7%), blindness (11.8%) and, the description of test reproducibility (19.1%). Some items are better reported: technical specifications and/or references to tests (65.1%), methods for calculating or comparing (90.2%), clinical and demographic characteristics (82.3%), diagnostic accuracy (90.2%). Only 16% of the journals studied required authors to use STARD. A high Impact factor and the year of the study were the items associated with a better SQS the presence of a conflict of interest was associated with a lower SQS in univariate analysis. A higher impact factor (> 5), was the only independent factors statistically significantly ($p = 0.037$) associated with higher SQS in a multivariate regression model.

Discussion Our study showed that the SQS were very low. Assessment of a study depends on quality of reporting. Blindness and participant sampling are the cornerstone to evaluate such bias as spectrum, verification, review and selection bias of a study, and were unfortunately scarcely reported compared to existing data in diagnosis accuracy reporting. One of the limitation is the 5 years sample of the study. We have planned to continue the analysis for a 10-year review starting just after the 2003 STARD publication.

Conclusion Our study showed that several items remain poorly reported. We recommend systematic use of STARD criteria in the elaboration and reporting of future studies that evaluates the preload dependence.

Competing interests

None.

References

1. Bossuyt PM. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ*. 2015;351:h5527.
2. Bossuyt PM. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD initiative. *Ann Intern Med*. 2003;138(1):40–4.

S89

Neurological impairment in cirrhotic patients admitted to ICU: hepatic versus drug-induced encephalopathy

Julie Assaraf¹, Simona Tripon², Maxime Mallet³, Marika Rudler⁴, Julien Mayaux⁵, Dominique Thabut⁶, Nicolas Weiss⁷

¹Unité de soins intensifs d'hépatogastroentérologie, Pitié-Salpêtrière Hospital, Paris, France;

²Unité de soins intensifs d'hépatogastroentérologie, Hôpital Pitié-Salpêtrière, Paris, France;

³Unité de Soins Intensifs. Service d'Hépatogastro-entérologie du Pr Poynard, CHU Pitié Salpêtrière, Paris, France;

⁴Hepatology and gastroenterology, Pitié-Salpêtrière Hospital, Paris, France;

⁵Réanimation médicale, Hôpital Pitié-Salpêtrière, Paris, France;

⁶Brain Liver Pitié-Salpêtrière Study Group (BLIPS), Hôpital Pitié-Salpêtrière, Paris, France;

⁷Unité de réanimation neurologique, Hôpital Pitié-Salpêtrière, Paris, France

Correspondence: Julie Assaraf - julie.assaraf@gmail.com

Annals of Intensive Care 2017, **7**(Suppl 1):S89

Introduction Neurological impairment, i.e. encephalopathy, is commonly observed in patients with decompensated cirrhosis and/or portosystemic shunts admitted in ICU. Often ascribed to high plasmatic levels of ammonia, encephalopathy could also be induced by drugs or infection, due to altered blood-brain barrier (BBB) permeability. This latter setting is often underdiagnosed and encephalopathy related to hyperammonemia (so called hepatic encephalopathy-HE) being pointed out as the culprit of all neurological symptoms in cirrhotic patients. Quinolones and betalactams were recently found in the cerebrospinal fluid of HE patients and it has been shown that the expression of efflux pumps, responsible for drugs passing through the BBB, was altered in animal models of HE. The purpose of this study was to assess the incidence of neurological impairment, i.e. encephalopathy, in cirrhotic patients hospitalized in

6.3. Résumé présenté à l'ESA 2017

Quality of Reporting of Fluid Responsiveness Evaluation Studies: A five years Systematic Review

Guillaume Izaute¹, Matthias Jacquet-Lagrèze^{1,2,*}, Jean-Luc Fellahi^{1,2}.

Background and goal of study:

The Standards for the Reporting of Diagnostic Accuracy Studies (STARD) statement was developed in 2003 to improve the quality of reporting of diagnosis studies [1]. It was updated in 2015 [2]. The purpose of this work was to evaluate the quality of reporting of studies assessing fluid responsiveness, using the STARD criteria.

Materials and methods:

We used MEDLINE via PubMed to search all publication of studies assessing the ability to predict or diagnose fluid responsiveness between 2004 and 2014. We have presented herein the last five years (2010-2014). We have checked these studies for all the 25 STARD criteria. The rating methodology of each item was discussed and double checked (GI and MJL). The 2003 and 2015 STARD quality score (SQS) was the sum of the items present. Univariate and multivariate analysis were conducted to find out if some characteristic of the studies were linked with a better SQS. Statistical analysis was performed with R software. $p < 0.05$ was considered statistically significant.

Results and Discussion:

After a double screening (GI, MJL) of 430 articles, 97 were selected, then 46 were excluded because an IF < 2 . Fifty-one studies were selected and reviewed. The mean 2003 overall quality score was 11.8 on a scale of 0 to 25, whereas the mean 2015 overall quality score was 15.3 on a scale of 0 to 30. The 2003 SQS remained stable between 2010 and 2014 ($p=0.173$). Correlation between 2003 and 2015 SQS was good ($r=0.83$; $p<0.001$). Some items are insufficiently reported: participant sampling (adequately reported in 19.6% of studies), executing and reading the tests (11.7%), blindness (11.8%). Some items are better reported: methods for calculating or comparing (90.2%), clinical and demographic characteristics (82.3%), diagnostic accuracy (90.2%). Only 16 % of the journals studied required authors to use STARD. An higher impact factor (>5), was the only independent factors statistically significantly ($p=0.037$) associated with higher SQS in a multivariate regression model. Our study show that the SQS are very low. Assessment of a study depends on quality of reporting. Blindness and participant sampling are the cornerstone to evaluate bias, and were unfortunately scarcely reported. One of the limitation is the five years sample of the study.

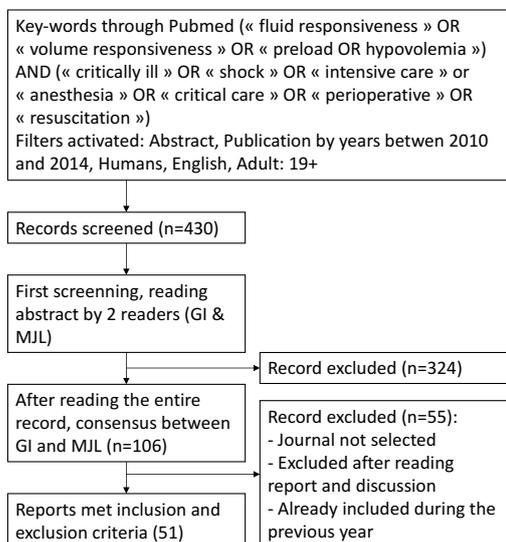


Fig. 1 Flowchart of study selection

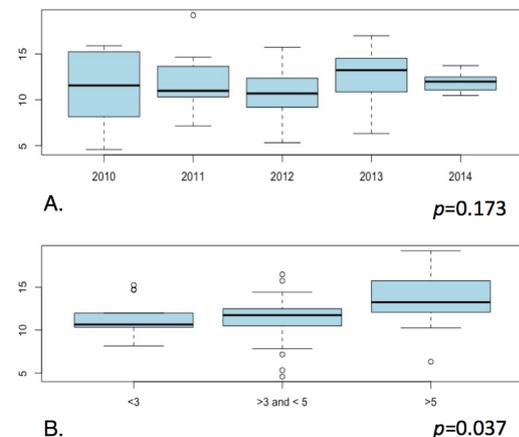


Fig. 2 A. Evolution of 2003 STARD quality score between 2010 and 2014. B. 2003 STARD quality score according to impact factor of the journals.

Conclusion:

Our study showed that several items remain poorly reported. We recommend systematic use of STARD criteria in the elaboration and reporting of future studies that evaluates the preload dependence.

References:

- Bossuyt PM. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD Initiative. *Ann Intern Med.* 2003 ; 7 : 138(1):40-4.
- Bossuyt PM. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ.* 2015 28;351:h5527.

1. Service d'Anesthésie Réanimation, Hôpital Louis Pradel, Hospices civils de Lyon, 59 Boulevard Pinel, 695394 Lyon Cedex, France.
2. Université Claude-Bernard Lyon 1, Campus Lyon Santé Est, 8 avenue Rockefeller, 69008 Lyon, France.
*Corresponding author



6.4. Lettre adressée à ***Anesthesiology***

Diagnostic Accuracy Studies: The Methodologic Approach Matters!

To the Editor:

With regard to the recent *ANESTHESIOLOGY* article by Biaï et al.,¹ we acknowledge the overall quality of the report and consider the relevancy of the topic of underlying research. However, we have found several methodologic concerns that we would like to address.

First, Standards for Reporting Diagnostic Accuracy Studies have been developed as a list of items^{2,3} that contribute to the completeness, transparency, and quality of reporting of diagnostic accuracy studies. We found that key items were lacking in the study from Biaï et al.¹ The study should have included a degree of blinding that describes whether clinical information and index test results were available to the assessors of the reference standard. A flow diagram is also required to evaluate the risk of selection bias. The reproducibility of the index test and the reference standard should also have been reported. Moreover, the studies cited in the article that support the rationale of the reference standard and its cutoff do not to support an increase of 10% of stroke volume after volume expansion measured by proAQT (Pulsion Medical Systems, Feldkirchen, Germany) that defined fluid responders.

Second, the threshold to differentiate between responders and nonresponders should be chosen above and close to the least significant change (LSC) of the stroke volume measurement by their considered device. The LSC is defined as the minimum change that can be recognized as a significant change, not a measurement of random variation. Although LSC has been reported previously with transpulmonary thermodilution,⁴ no data were reported using the proAQT system. Therefore, LSC for the proAQT system should have been calculated and reported by the authors. Because there was no threshold of stroke volume variation after a volume expansion to differentiate responders and nonresponders that can be supported by a solid clinical or physiologic background, another strategy would have been to provide data for several thresholds.⁵ To address this last point, we collected data of the scatterplot given in figure 2 of the article by Biaï et al.¹ using the software ImageJ. This allowed us to recover raw data of variations of stroke volume after lung recruitment maneuver and after volume expansion and enabled us to perform subsequent analysis. We explored 16 thresholds between 5% and 20% using the R software and pROC package (<https://www.r-project.org/>. Accessed XXX). We computed 95% CI using the bootstrap technique with thousand repetitions. In our point of view, the area under the

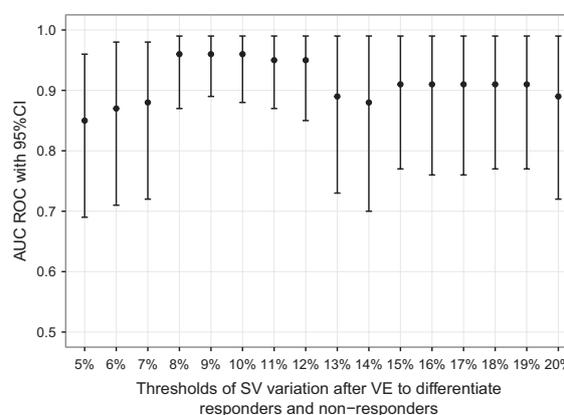


Fig. 1. Relationship between the areas under the receiver operating characteristic (ROC) curves (points) with 95% CI (error bars) of the lung recruitment maneuver and the thresholds of stroke volume (SV) variation after volume expansion (VE) between 5% and 20% to differentiate between responders and nonresponders.

receiver operating characteristic curve was overestimated with the chosen threshold of 10% (fig. 1). Because the threshold increases beyond the LSC of the measurement system, the area under the receiver operating characteristic curve should remain constant or increase with the threshold, which was not the case in the study by Biaï et al.¹ For all of these reasons, we strongly suspect that some recruitment biases could have occurred.

Finally, some studies have previously evaluated the diagnostic accuracy of a transient positive end-expiratory pressure elevation, identified as a recruitment maneuver, to diagnose preload responsiveness. Such diagnostic approaches were similar to those proposed by the authors and should have been discussed.⁶⁻⁸ Diagnostic studies are at a high risk for biases,⁹ and the methodologic considerations above highlight a risk of bias in the study by Biaï et al.¹

Competing Interests

The authors declare no competing interests.

Matthias Jacquet-Lagrèze, M.D., Guillaume Izaute, M.D., Jean-Luc Fellahi, M.D., Ph.D. Hôpital Louis Pradel, Hospices Civils de Lyon, Lyon, France (M.J.-L.). matthias.jl@gmail.com

References

- Biaï M, Lanchon R, Sesay M, Le Gall L, Pereira B, Futier E, Nouette-Gaulain K: Changes in stroke volume induced by lung recruitment maneuver predict fluid responsiveness in mechanically ventilated patients in the operating room. *ANESTHESIOLOGY* 2017; 126:260-7
- Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, Lijmer JG, Moher D, Rennie D, de Vet HC: Standards for Reporting of Diagnostic Accuracy: Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD initiative. *BMJ* 2003; 326:41-4

3. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, Lijmer JG, Moher D, Rennie D, Vet HCW de, Kressel HY, Rifai N, Golub RM, Altman DG, Hooft L, Korevaar DA, Cohen JF: STARD 2015: An updated list of essential items for reporting diagnostic accuracy studies. *BMJ* 2015; 351:h5527
4. Monnet X, Persichini R, Ktari M, Jozwiak M, Richard C, Teboul JL: Precision of the transpulmonary thermodilution measurements. *Crit Care* 2011; 15:R204
5. Cannesson M, Le Manach Y, Hofer CK, Goarin JP, Lehot JJ, Vallet B, Tavernier B: Assessing the diagnostic accuracy of pulse pressure variations for the prediction of fluid responsiveness: A “gray zone” approach. *ANESTHESIOLOGY* 2011; 115:231–41
6. Wilkman E, Kuitunen A, Pettilä V, Varpula M: Fluid responsiveness predicted by elevation of PEEP in patients with septic shock. *Acta Anaesthesiol Scand* 2014; 58:27–35
7. Tusman G, Groisman I, Maidana GA, Scandurra A, Arca JM, Bohm SH, Suarez-Sipmann F: The sensitivity and specificity of pulmonary carbon dioxide elimination for noninvasive assessment of fluid responsiveness. *Anesth Analg* 2016; 122:1404–11
8. Broca B De, Garnier J, Fischer M-O, Archange T, Marc J, Abou-Arab O, Dupont H, Lorne E, Guinot P: Stroke volume changes induced by a recruitment maneuver predict fluid responsiveness in patients with protective ventilation in the operating theater. *Medicine (Baltimore)* 2016; 95:e4259
9. Lijmer JG, Mol BW, Heisterkamp S, Bossel GJ, Prins MH, van der Meulen JH, Bossuyt PM: Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA* 1999; 282:1061–6

(Accepted for publication July 7, 2017.)

6.5. Liste des articles inclus dans notre étude

2005

1. Hofer CK, Furrer L, Matter-Ensner S, Maloigne M, Klaghofer R, Genoni M, et al. Volumetric preload measurement by thermodilution: a comparison with transoesophageal echocardiography. *Br J Anaesth.* 6 janv 2005;94(6):748-55.
2. Preisman S. Predicting fluid responsiveness in patients undergoing cardiac surgery: functional haemodynamic parameters including the Respiratory Systolic Variation Test and static preload indicators. *Br J Anaesth.* 14 oct 2005;95(6):746-55.
3. Hofer CK, Müller SM, Furrer L, Klaghofer R, Genoni M, Zollinger A. Stroke volume and pulse pressure variation for prediction of fluid responsiveness in patients undergoing off-pump coronary artery bypass grafting. *Chest.* août 2005;128(2):848-54.
4. Wiesenack C, Fiegl C, Keyser A, Laule S, Prasser C, Keyl C. Continuously assessed right ventricular end-diastolic volume as a marker of cardiac preload and fluid responsiveness in mechanically ventilated cardiac surgical patients. *Crit Care Lond Engl.* juin 2005;9(3):R226-233.
5. Feissel M, Badie J, Merlani PG, Faller J-P, Bendjelid K. Pre-ejection period variations predict the fluid responsiveness of septic ventilated patients. *Crit Care Med.* nov 2005;33(11):2534-9.
6. Wiesenack C, Fiegl C, Keyser A, Prasser C, Keyl C. Assessment of fluid responsiveness in mechanically ventilated cardiac surgical patients: *Eur J Anaesthesiol.* sept 2005;22(9):658-65.
7. Monnet X, Rienzo M, Osman D, Anguel N, Richard C, Pinsky MR, et al. Esophageal Doppler monitoring predicts fluid responsiveness in critically ill ventilated patients. *Intensive Care Med.* 1 sept 2005;31(9):1195-201.
8. Vallée F, Fourcade O, Soyres OD, Angles O, Sanchez-Verlaan P, Pillard F, et al. Stroke output variations calculated by esophageal Doppler is a reliable predictor of fluid response. *Intensive Care Med.* 1 oct 2005;31(10):1388-93.
9. Backer DD, Heenen S, Piagnerelli M, Koch M, Vincent J-L. Pulse pressure variations to predict fluid responsiveness: influence of tidal volume. *Intensive Care Med.* 1 avr 2005;31(4):517-23.

2006

10. Perner A, Faber T. Stroke volume variation does not predict fluid responsiveness in patients with septic shock on pressure support ventilation. *Acta Anaesthesiol Scand.* 1 oct 2006;50(9):1068-73.
11. Natalini G, Rosano A, Franceschetti ME, Facchetti P, Bernardini A. Variations in Arterial Blood Pressure and Photoplethysmography During Mechanical Ventilation: *Anesth Analg.* nov 2006;103(5):1182-8.
12. Charron C, Fessenmeyer C, Cosson C, Mazoit J-X, Hebert J-L, Benhamou D, et al. The influence of tidal volume on the dynamic variables of fluid responsiveness in critically ill patients. *Anesth Analg.* mai 2006;102(5):1511-7.
13. Natalini G, Rosano A, Taranto M, Faggian B, Vittorielli E, Bernardini A. Arterial versus plethysmographic dynamic indices to test responsiveness for testing fluid administration in hypotensive patients: a clinical trial. *Anesth Analg.* déc 2006;103(6):1478-84.
14. Mekontso-Dessap A, Tual L, Kirsch M, D'Honneur G, Loisanche D, Brochard L, et al. B-type natriuretic peptide to assess haemodynamic status after cardiac surgery. *Br J Anaesth.* 12 janv 2006;97(6):777-82.
15. Solus-Biguenet H, Fleyfel M, Tavernier B, Kipnis E, Onimus J, Robin E, et al. Non-invasive prediction of fluid responsiveness during major hepatic surgery. *Br J Anaesth.* 12 janv 2006;97(6):808-16.

16. Cannesson M, Sliker J, Desebbe O, Farhat F, Bastien O, Lehot J-J. Prediction of fluid responsiveness using respiratory variations in left ventricular stroke area by transoesophageal echocardiographic automated border detection in mechanically ventilated patients. *Crit Care*. 2006;10:R171.
17. Lafanechère A, Pène F, Goulenok C, Delahaye A, Mallet V, Choukroun G, et al. Changes in aortic blood flow induced by passive leg raising predict fluid responsiveness in critically ill patients. *Crit Care*. 2006;10:R132.
18. Heenen S, De Backer D, Vincent J-L. How can the response to volume expansion in patients with spontaneous respiratory movements be predicted? *Crit Care*. 2006;10:R102.
19. Monnet X, Rienzo M, Osman D, Anguel N, Richard C, Pinsky MR, et al. Passive leg raising predicts fluid responsiveness in the critically ill. *Crit Care Med*. mai 2006;34(5):1402-7.
20. Westphal GA, Silva E, Caldeira Filho M, Roman Gonçalves AR, Poli-de-Figueiredo LF. Variation in amplitude of central venous pressure curve induced by respiration is a useful tool to reveal fluid responsiveness in postcardiac surgery patients. *Shock Augusta Ga*. août 2006;26(2):140-5.

2007

21. Wyffels PAH, Durnez P-J, Helderweirt J, Stockman WMA, De Kegel D. Ventilation-induced plethysmographic variations predict fluid responsiveness in ventilated postoperative cardiac surgery patients. *Anesth Analg*. août 2007;105(2):448-52.
22. Cannesson M, Attof Y, Rosamel P, Desebbe O, Joseph P, Metton O, et al. Respiratory Variations in Pulse Oximetry Plethysmographic Waveform Amplitude to Predict Fluid Responsiveness in the Operating Room. *J Am Soc Anesthesiol*. 1 juin 2007;106(6):1105-11.
23. Lee J-H, Kim J-T, Yoon SZ, Lim Y-J, Jeon Y, Bahk J-H, et al. Evaluation of corrected flow time in oesophageal Doppler as a predictor of fluid responsiveness. *Br J Anaesth*. 9 janv 2007;99(3):343-8.
24. Osman D, Ridel C, Ray P, Monnet X, Anguel N, Richard C, et al. Cardiac filling pressures are not appropriate to predict hemodynamic response to volume challenge. *Crit Care Med*. janv 2007;35(1):64-8.
25. Monnet X, Chemla D, Osman D, Anguel N, Richard C, Pinsky MR, et al. Measuring aortic diameter improves accuracy of esophageal Doppler in assessing fluid responsiveness. *Crit Care Med*. févr 2007;35(2):477-82.
26. Cannesson M, Desebbe O, Hachemi M, Jacques D, Bastien O, Lehot J-J. Respiratory variations in pulse oximeter waveform amplitude are influenced by venous return in mechanically ventilated patients under general anaesthesia. *Eur J Anaesthesiol*. mars 2007;24(3):245-51.
27. Feissel M, Teboul J-L, Merlani P, Badie J, Faller J-P, Bendjelid K. Plethysmographic dynamic indices predict fluid responsiveness in septic ventilated patients. *Intensive Care Med*. 1 juin 2007;33(6):993-9.
28. Maizel J, Airapetian N, Lorne E, Tribouilloy C, Massy Z, Slama M. Diagnosis of central hypovolemia by using passive leg raising. *Intensive Care Med*. 1 juill 2007;33(7):1133-8.
29. Soubrier S, Saulnier F, Hubert H, Delour P, Lenci H, Onimus T, et al. Can dynamic indicators help the prediction of fluid responsiveness in spontaneously breathing critically ill patients? *Intensive Care Med*. 1 juill 2007;33(7):1117-24.
30. Lamia B, Ochagavia A, Monnet X, Chemla D, Richard C, Teboul J-L. Echocardiographic prediction of volume responsiveness in critically ill patients with spontaneously breathing activity. *Intensive Care Med*. 1 juill 2007;33(7):1125-32.

2008

31. Muller L, Louart G, Bengler C, Fabbro-Peray P, Carr J, Ripart J, et al. The Intrathoracic Blood Volume Index as an Indicator of Fluid Responsiveness in Critically Ill Patients with Acute Circulatory Failure: A Comparison with Central Venous Pressure. *Anesth Analg*. août 2008;107(2):607-13.
32. Auler JO, Galas F, Hajjar L, Santos L, Carvalho T, Michard F. Online Monitoring of Pulse Pressure Variation to Guide Fluid Therapy After Cardiac Surgery. *Anesth Analg*. avr 2008;106(4):1201-6.
33. Deflandre E, Bonhomme V, Hans P. Delta down compared with delta pulse pressure as an indicator of volaemia during intracranial surgery. *Br J Anaesth*. 2 janv 2008;100(2):245-50.

34. Biais M, Nouette-Gaulain K, Cottenceau V, Revel P, Sztark F. Uncalibrated pulse contour-derived stroke volume variation predicts fluid responsiveness in mechanically ventilated patients undergoing liver transplantation. *Br J Anaesth.* 12 janv 2008;101(6):761-8.
35. Cannesson M, Desebbe O, Rosamel P, Delannoy B, Robin J, Bastien O, et al. Pleth variability index to monitor the respiratory variations in the pulse oximeter plethysmographic waveform amplitude and predict fluid responsiveness in the operating theatre. *Br J Anaesth.* 8 janv 2008;101(2):200-6.
36. Hofer CK, Senn A, Weibel L, Zollinger A. Assessment of stroke volume variation for prediction of fluid responsiveness using the modified FloTrac™ and PiCCOplus™ system. *Crit Care.* 2008;12(3):R82.
37. Keller G, Cassar E, Desebbe O, Lehot J-J, Cannesson M. Ability of pleth variability index to detect hemodynamic changes induced by passive leg raising in spontaneously breathing volunteers. *Crit Care.* 2008;12(2):R37.
38. Huang C-C, Fu J-Y, Hu H-C, Kao K-C, Chen N-H, Hsieh M-J, et al. Prediction of fluid responsiveness in acute respiratory distress syndrome patients ventilated with low tidal volume and high positive end-expiratory pressure*: *Crit Care Med.* oct 2008;36(10):2810-6.
39. García MIM, Cano AG, Monrové JCD. Arterial pressure changes during the Valsalva maneuver to predict fluid responsiveness in spontaneously breathing patients. *Intensive Care Med.* 1 janv 2009;35(1):77.
40. Jabot J, Teboul J-L, Richard C, Monnet X. Passive leg raising for predicting fluid responsiveness: importance of the postural change. *Intensive Care Med.* 1 janv 2009;35(1):85.
41. Donati A, Nardella R, Gabbanelli V, Scarcella M, Romanelli M, Romagnoli L, et al. The ability of PiCCO versus LiDCO variables to detect changes in cardiac index: a prospective clinical study. *Minerva Anestesiol.* août 2008;74(7-8):367-74.

2009

42. Vistisen ST, Struijk JJ, Larsson A. Automated pre-ejection period variation indexed to tidal volume predicts fluid responsiveness after cardiac surgery. *Acta Anaesthesiol Scand.* avril 2009;53(4):534-42.
43. Cannesson M, Musard H, Desebbe O, Boucau C, Simon R, Hénaine R, et al. The Ability of Stroke Volume Variations Obtained with Vigileo/FloTrac System to Monitor Fluid Responsiveness in Mechanically Ventilated Patients: *Anesth Analg.* févr 2009;108(2):513-7.
44. Kobayashi M, Koh M, Irinoda T, Meguro E, Hayakawa Y, Takagane A. Stroke volume variation as a predictor of intravascular volume depression and possible hypotension during the early postoperative period after esophagectomy. *Ann Surg Oncol.* mai 2009;16(5):1371-7.
45. Lahner D, Kabon B, Marschalek C, Chiari A, Pestel G, Kaider A, et al. Evaluation of stroke volume variation obtained by arterial pulse contour analysis to predict fluid responsiveness intraoperatively. *Br J Anaesth.* 9 janv 2009;103(3):346-51.
46. Derichard A, Robin E, Tavernier B, Costecalde M, Fleyfel M, Onimus J, et al. Automated pulse pressure and stroke volume variations from radial artery: evaluation during major abdominal surgery. *Br J Anaesth.* 11 janv 2009;103(5):678-84.
47. Monge García MI, Gil Cano A, Díaz Monrové JC. Brachial artery peak velocity variation to predict fluid responsiveness in mechanically ventilated patients. *Crit Care.* 2009;13(5):R142.
48. Thiel SW, Kollf MH, Isakow W. Non-invasive stroke volume measurement and passive leg raising predict volume responsiveness in medical ICU patients: an observational cohort study. *Crit Care.* 2009;13:R111.
49. Mahjoub Y, Pila C, Friggeri A, Zogheib E, Lobjoie E, Tinturier F, et al. Assessing fluid responsiveness in critically ill patients: False-positive pulse pressure variation is detected by Doppler echocardiographic evaluation of the right ventricle. *Crit Care Med.* sept 2009;37(9):2570-5.
50. de Waal EEC, Rex S, Kruitwagen CLJJ, Kalkman CJ, Buhre WF. Dynamic preload indicators fail to predict fluid responsiveness in open-chest conditions. *Crit Care Med.* févr 2009;37(2):510-5.
51. Breukers R-MBGE, de Wilde RBP, van den Berg PCM, Jansen JRC, Faes TJC, Twisk JWR, et al. Assessing fluid responses after coronary surgery: role of mathematical coupling of global end-diastolic volume to cardiac output measured by transpulmonary thermodilution. *Eur J Anaesthesiol.* nov 2009;26(11):954-60.

52. Breukers R-MBGE, Trof RJ, Wilde RBP de, Berg PCM van den, Twisk JWR, Jansen JRC, et al. Relative value of pressures and volumes in assessing fluid responsiveness after valvular and coronary artery surgery. *Eur J Cardiothorac Surg*. 1 janv 2009;35(1):62-8.
53. Muller L, Louart G, Bousquet P-J, Candela D, Zoric L, Coussaye J-E de L, et al. The influence of the airway driving pressure on pulsed pressure variation as a predictor of fluid responsiveness. *Intensive Care Med*. 1 mars 2010;36(3):496-503.
54. Vallée F, Richard JCM, Mari A, Gallas T, Arsac E, Verlaan PS, et al. Pulse pressure variations adjusted by alveolar driving pressure to assess fluid responsiveness. *Intensive Care Med*. 1 juin 2009;35(6):1004-10.

2010

55. Wyffels P a. H, Sergeant P, Wouters PF. The value of pulse pressure and stroke volume variation as predictors of fluid responsiveness during open chest surgery. *Anaesthesia*. juill 2010;65(7):704-9.
56. Biais M, Bernard O, Ha JC, Degryse C, Sztark F. Abilities of pulse pressure variations and stroke volume variations to predict fluid responsiveness in prone position during scoliosis surgery. *Br J Anaesth*. 4 janv 2010;104(4):407-13.
57. Wyler von Ballmoos M, Takala J, Roeck M, Porta F, Tueller D, Ganter CC, et al. Pulse-pressure variation and hemodynamic response in patients with elevated pulmonary artery pressure: a clinical study. *Crit Care*. 2010;14(3):R111.
58. Mahjoub Y, Touzeau J, Airapetian N, Lorne E, Hijazi M, Zogheib E, et al. The passive leg-raising maneuver cannot accurately predict fluid responsiveness in patients with intra-abdominal hypertension. *Crit Care Med*. sept 2010;38(9):1824-9.
59. Préau S, Saulnier F, Dewavrin F, Durocher A, Chagnon J-L. Passive leg raising is predictive of fluid responsiveness in spontaneously breathing patients with severe sepsis or acute pancreatitis. *Crit Care Med*. mars 2010;38(3):819-25.
60. Zimmermann M, Feibicke T, Keyl C, Prasser C, Moritz S, Graf BM, et al. Accuracy of stroke volume variation compared with pleth variability index to predict fluid responsiveness in mechanically ventilated patients undergoing major surgery. *Eur J Anaesthesiol*. juin 2010;27(6):555-61.
61. Benomar B, Ouattara A, Estagnasie P, Brusset A, Squara P. Fluid responsiveness predicted by noninvasive Bioreactance-based passive leg raise test. *Intensive Care Med*. nov 2010;36(11):1875-81.
62. Qiao H, Zhang J, Liang W. Validity of pulse pressure and systolic blood pressure variation data obtained from a Datex Ohmeda S/5 monitor for predicting fluid responsiveness during surgery. *J Neurosurg Anesthesiol*. oct 2010;22(4):316-22.
63. Moretti R, Pizzi B. Inferior vena cava distensibility as a predictor of fluid responsiveness in patients with subarachnoid hemorrhage. *Neurocrit Care*. août 2010;13(1):3-9.
64. Lee C-H, Wang J-Y, Wu Y-K, Chiu H-W, Lan C-C, Chang H, et al. Pulse Pressure Power Spectrum Predicts Volume Responsiveness in Shock Patients without Sedation: *Shock*. nov 2009;1.

2011

65. Høiseth LØ, Hoff IE, Skare Ø, Kirkebøen KA, Landsverk SA. Photoplethysmographic and pulse pressure variations during abdominal surgery. *Acta Anaesthesiol Scand*. 1 nov 2011;55(10):1221-30.
66. Broch O, Bein B, Gruenewald M, Höcker J, Schöttler J, Meybohm P, et al. Accuracy of the pleth variability index to predict fluid responsiveness depends on the perfusion index. *Acta Anaesthesiol Scand*. 1 juill 2011;55(6):686-93.
67. Biais M, Stecken L, Ottolenghi L, Rouillet S, Quinart A, Masson F, et al. The ability of pulse pressure variations obtained with CNAPTM device to predict fluid responsiveness in the operating room. *Anesth Analg*. sept 2011;113(3):523-8.
68. Hood JA, Wilson RJT. Pleth variability index to predict fluid responsiveness in colorectal surgery. *Anesth Analg*. nov 2011;113(5):1058-63.

69. Muller L, Toumi M, Bousquet P-J, Riu-Poulenc B, Louart G, Candela D, et al. An Increase in Aortic Blood Flow after an Infusion of 100 ml Colloid over 1 Minute Can Predict Fluid Responsiveness: The Mini-fluid Challenge Study. *Anesthesiology*. sept 2011;115(3):541-7.
70. Cannesson M, Le Manach Y, Hofer CK, Goarin JP, Lehot J-J, Vallet B, et al. Assessing the Diagnostic Accuracy of Pulse Pressure Variations for the Prediction of Fluid Responsiveness: A “Gray Zone” Approach. *Anesthesiology*. août 2011;115(2):231-41.
71. Desgranges F-P, Desebbe O, Ghazouani A, Gilbert K, Keller G, Chiari P, et al. Influence of the site of measurement on the ability of plethysmographic variability index to predict fluid responsiveness. *Br J Anaesth*. 9 janv 2011;107(3):329-35.
72. Geerts BF, Aarts LPHJ, Groeneveld AB, Jansen JRC. Predicting cardiac output responses to passive leg raising by a PEEP-induced increase in central venous pressure, in cardiac surgery patients. *Br J Anaesth*. 8 janv 2011;107(2):150-6.
73. Velissaris D, Pierrakos C, Scolletta S, De Backer D, Vincent JL. High mixed venous oxygen saturation levels do not exclude fluid responsiveness in critically ill septic patients. *Crit Care*. 2011;15(4):R177.
74. Trof RJ, Danad I, Reilingh MW, Breukers R-MB, Groeneveld AJ. Cardiac filling volumes versus pressures for predicting fluid responsiveness after cardiovascular surgery: the role of systolic cardiac function. *Crit Care*. 2011;15(1):R73.
75. Loupec T, Nanadoumgar H, Frasca D, Petitpas F, Laksiri L, Baudouin D, et al. Pleth variability index predicts fluid responsiveness in critically ill patients. *Crit Care Med*. févr 2011;39(2):294-9.
76. Geerts BF, Maas J, de Wilde RBP, Aarts LPHJ, Jansen JRC. Arm occlusion pressure is a useful predictor of an increase in cardiac output after fluid loading following cardiac surgery. *Eur J Anaesthesiol*. nov 2011;28(11):802-6.
77. Lee J-H, Jeon Y, Bahk J-H, Gil N-S, Hong DM, Kim JH, et al. Pulse pressure variation as a predictor of fluid responsiveness during one-lung ventilation for lung surgery using thoracotomy: randomised controlled study. *Eur J Anaesthesiol*. janv 2011;28(1):39-44.

2012

78. Høiseith LØ, Hoff IE, Myre K, Landsverk SA, Kirkebøen KA. Dynamic variables of fluid responsiveness during pneumoperitoneum and laparoscopic surgery. *Acta Anaesthesiol Scand*. juill 2012;56(6):777-86.
79. Monnet X, Dres M, Ferré A, Teuff GL, Jozwiak M, Bleibtreu A, et al. Prediction of fluid responsiveness by a continuous non-invasive assessment of arterial pressure in critically ill patients: comparison with four other dynamic indices. *Br J Anaesth*. 9 janv 2012;109(3):330-8.
80. Lansdorp B, Lemson J, Putten MJAM van, Keijzer A de, Hoeven JG van der, Pickkers P. Dynamic indices do not predict volume responsiveness in routine clinical practice. *Br J Anaesth*. 3 janv 2012;108(3):395-401.
81. Muller L, Bobbia X, Toumi M, Louart G, Molinari N, Ragonnet B, et al. Respiratory variations of inferior vena cava diameter to predict fluid responsiveness in spontaneously breathing patients with acute circulatory failure: need for a cautious use. *Crit Care*. 2012;16(5):R188.
82. Monnet X, Bleibtreu A, Ferré A, Dres M, Gharbi R, Richard C, et al. Passive leg-raising and end-expiratory occlusion tests perform better than pulse pressure variation in patients with low respiratory system compliance. *Crit Care Med*. janv 2012;40(1):152-7.
83. Khwannimit B, Bhurayanontachai R. Prediction of fluid responsiveness in septic shock patients: comparing stroke volume variation by FloTrac/Vigileo and automated pulse pressure variation. *Eur J Anaesthesiol*. févr 2012;29(2):64-9.
84. Cecconi M, Monti G, Hamilton MA, Puntis M, Dawson D, Tuccillo ML, et al. Efficacy of functional hemodynamic parameters in predicting fluid responsiveness with pulse power analysis in surgical patients. *Minerva Anesthesiol*. mai 2012;78(5):527-33.

2013

85. Fischer M-O, Coucoravas J, Truong J, Zhu L, Gérard J-L, Hanouz J-L, et al. Assessment of changes in cardiac index and fluid responsiveness: a comparison of Nexfin and transpulmonary thermodilution. *Acta Anaesthesiol Scand*. 1 juill 2013;57(6):704-12.
86. Bubenek-Turconi SI, Craciun M, Miclea I, Perel A. Noninvasive Continuous Cardiac Output by the Nexfin Before and After Preload-Modifying Maneuvers: A Comparison with Intermittent Thermodilution Cardiac Output. *Anesth Analg*. août 2013;117(2):366-72.
87. Trepte CJC, Eichhorn V, Haas SA, Stahl K, Schmid F, Nitzschke R, et al. Comparison of an automated respiratory systolic variation test with dynamic preload indicators to predict fluid responsiveness after major surgery. *Br J Anaesth*. 11 janv 2013;111(5):736-42.
88. Freitas FGR, Bafi AT, Nascente APM, Assunção M, Mazza B, Azevedo LCP, et al. Predictive value of pulse pressure variation for fluid responsiveness in septic patients using lung-protective ventilation strategies. *Br J Anaesth*. 3 janv 2013;110(3):402-8.
89. Guinot P-G, Broca B de, Arab OA, Diouf M, Badoux L, Bernard E, et al. Ability of stroke volume variation measured by oesophageal Doppler monitoring to predict fluid responsiveness during surgery. *Br J Anaesth*. 1 janv 2013;110(1):28-33.
90. Monnet X, Guérin L, Jozwiak M, Bataille A, Julien F, Richard C, et al. Pleth variability index is a weak predictor of fluid responsiveness in patients receiving norepinephrine. *Br J Anaesth*. 2 janv 2013;110(2):207-13.
91. Marik PE, Levitov A, Young A, Andrews L. The Use of Bioreactance and Carotid Doppler to Determine Volume Responsiveness and Blood Flow Redistribution Following Passive Leg Raising in Hemodynamically Unstable Patients. *Chest*. févr 2013;143(2):364-70.
92. Silva S, Jozwiak M, Teboul J-L, Persichini R, Richard C, Monnet X. End-expiratory occlusion test predicts preload responsiveness independently of positive end-expiratory pressure during acute respiratory distress syndrome. *Crit Care Med*. juill 2013;41(7):1692-701.
93. Lakhil K, Ehrmann S, Perrotin D, Wolff M, Boulain T. Fluid challenge: tracking changes in cardiac output with blood pressure monitoring (invasive or non-invasive). *Intensive Care Med*. nov 2013;39(11):1953-62.
94. Monnet X, Bataille A, Magalhaes E, Barrois J, Le Corre M, Gosset C, et al. End-tidal carbon dioxide is better than arterial pressure for predicting volume responsiveness by the passive leg raising test. *Intensive Care Med*. janv 2013;39(1):93-100.

2014

95. Brun C, Zieleskiewicz L, Textoris J, Muller L, Bellefleur J-P, Antonini F, et al. Prediction of fluid responsiveness in severe preeclamptic patients with oliguria. *Intensive Care Med*. avr 2013;39(4):593-600.
96. Biais M, Ehrmann S, Mari A, Conte B, Mahjoub Y, Desebbe O, et al. Clinical relevance of pulse pressure variations for predicting fluid responsiveness in mechanically ventilated intensive care unit patients: the grey zone approach. *Crit Care Lond Engl*. 2014;18(6):587.
97. MacDonald N, Ahmad T, Mohr O, Kirk-Bayley J, Moppett I, Hinds CJ, et al. Dynamic preload markers to predict fluid responsiveness during and after major gastrointestinal surgery: an observational substudy of the OPTIMISE trial. *Br J Anaesth*. avr 2015;114(4):598-604.
98. Guarracino F, Ferro B, Forfori F, Bertini P, Magliacano L, Pinsky MR. Jugular vein distensibility predicts fluid responsiveness in septic patients. *Crit Care Lond Engl*. 2014;18(6):647.
99. Høiseth LØ, Hoff IE, Hagen OA, Landsverk SA, Kirkebøen KA. Dynamic variables and fluid responsiveness in patients for aortic stenosis surgery. *Acta Anaesthesiol Scand*. août 2014;58(7):826-34.
100. Hong DM, Lee JM, Seo JH, Min JJ, Jeon Y, Bahk JH. Pulse pressure variation to predict fluid responsiveness in spontaneously breathing patients: tidal vs. forced inspiratory breathing. *Anaesthesia*. juill 2014;69(7):717-22.
101. Song Y, Kwak YL, Song JW, Kim YJ, Shim JK. Respiratory phase carotid artery peak velocity variation as a predictor of fluid responsiveness in mechanically ventilated patients with coronary artery disease. *Br J Anaesth*. 7 janv 2014;113(1):61-6.

102. Guinot P-G, Godart J, de Broca B, Bernard E, Lorne E, Dupont H. End-expiratory occlusion manoeuvre does not accurately predict fluid responsiveness in the operating theatre. *Br J Anaesth.* juin 2014;112(6):1050-4.
103. Guinot P-G, Broca B de, Bernard E, Arab OA, Lorne E, Dupont H. Respiratory stroke volume variation assessed by oesophageal Doppler monitoring predicts fluid responsiveness during laparoscopy. *Br J Anaesth.* 4 janv 2014;112(4):660-4.
104. Wilkman E, Kuitunen A, Pettilä V, Varpula M. Fluid responsiveness predicted by elevation of PEEP in patients with septic shock. *Acta Anaesthesiol Scand.* 1 janv 2014;58(1):27-35.
-

7. Bibliographie

1. De Backer D. Prédiction de la réponse au remplissage vasculaire: que faire de tous ces indices proposés ? Réanimation. mars 2012;21(2):123-7.
2. Feldschuh J, Enson Y. Prediction of the normal blood volume. Relation of blood volume to body habitus. Circulation. 1 oct 1977;56(4):605-12.
3. Samet P, Fritts HW, Fishman AP, Cournand A. The blood volume in heart disease. Medicine (Baltimore). 1 mai 1957;36(2):211-35.
4. Feldschuh J, Katz S. The Importance of Correct Norms in Blood Volume Measurement. Am J Med Sci. 1 juill 2007;334(1):41-6.
5. Gordon AM, Huxley AF, Julian FJ. The variation in isometric tension with sarcomere length in vertebrate muscle fibres. J Physiol. mai 1966;184(1):170-92.
6. Norton JM. Toward Consistent Definitions for Preload and Afterload. Adv Physiol Educ. 1 mars 2001;25(1):53-61.
7. Déterminants de la fonction systolique [Internet]. PAC - Précis d'anesthésie cardiaque. 2015 [cité 8 août 2017]. Disponible sur: <http://www.pac4.ch/Chapitre%205/determfonctysyst.html>
8. Katz AM. Ernest Henry Starling, His Predecessors, and the "Law of the Heart". Circulation. 3 déc 2002;106(23):2986-92.
9. Bentzer P, Griesdale DE, Boyd J, MacLean K, Sirounis D, Ayas NT. Will This Hemodynamically Unstable Patient Respond to a Bolus of Intravenous Fluids? JAMA. 27 sept 2016;316(12):1298-309.
10. Cherpanath TGV, Geerts BF, Lagrand WK, Schultz MJ, Groeneveld ABJ. Basic concepts of fluid responsiveness. Neth Heart J. déc 2013;21(12):530-6.
11. Rivers E, Nguyen B, Havstad S, Ressler J, Muzzin A, Knoblich B, et al. Early goal-directed therapy in the treatment of severe sepsis and septic shock. N Engl J Med. 2001;345(19):1368-1377.
12. Rudiger A, Singer M. Mechanisms of sepsis-induced cardiac dysfunction: Crit Care Med. juin 2007;35(6):1599-608.
13. The First Use of Intravenous Saline for the Treatment of Disease: Letter from Thomas Latta submitted to the Central Board of Health, London and published in The Lancet, 1832. Preface by Jane Ferrie. Int J Epidemiol. 1 avr 2013;42(2):387-90.
14. O'Shaughnessy WB. PROPOSAL OF A NEW METHOD OF TREATING THE BLUE EPIDEMIC CHOLERA BY THE INJECTION OF HIGHLY-OXYGENISED SALTS INTO THE VENOUS SYSTEM. The Lancet. 10 déc 1831;17(432):366-71.
15. MALIGNANT CHOLERA. The Lancet. 2 juin 1832;18(457):274-80.
16. Awad S, Allison SP, Lobo DN. The history of 0.9% saline. Clin Nutr. 1 avr 2008;27(2):179-88.
17. Marik P, Bellomo R. A rational approach to fluid therapy in sepsis. BJA Br J Anaesth. 1 mars

2016;116(3):339-49.

18. Jones DAE, Shapiro DNI, Trzeciak DS, Arnold DRC, Claremont MHA, Kline DJA. Lactate Clearance vs Central Venous Oxygen Saturation as Goals of Early Sepsis Therapy: A Randomized Clinical Trial. *JAMA J Am Med Assoc.* 24 févr 2010;303(8):739.
19. Dellinger RP, Carlet JM, Masur H, Gerlach H, Calandra T, Cohen J, et al. Surviving Sepsis Campaign guidelines for management of severe sepsis and septic shock: *Crit Care Med.* mars 2004;32(3):858-73.
20. Dellinger RP, Levy MM, Carlet JM, Bion J, Parker MM, Jaeschke R, et al. Surviving Sepsis Campaign: International guidelines for management of severe sepsis and septic shock: 2008: *Crit Care Med.* janv 2008;36(1):296-327.
21. Dellinger RP, Levy MM, Rhodes A, Annane D, Gerlach H, Opal SM, et al. Surviving Sepsis Campaign: International Guidelines for Management of Severe Sepsis and Septic Shock. *Crit Care Med.* févr 2013;41(2):580-637.
22. ProCESS Investigators, Yealy DM, Kellum JA, Huang DT, Barnato AE, Weissfeld LA, et al. A randomized trial of protocol-based care for early septic shock. *N Engl J Med.* 1 mai 2014;370(18):1683-93.
23. ARISE Investigators, ANZICS Clinical Trials Group, Peake SL, Delaney A, Bailey M, Bellomo R, et al. Goal-directed resuscitation for patients with early septic shock. *N Engl J Med.* 16 oct 2014;371(16):1496-506.
24. Mouncey PR, Osborn TM, Power GS, Harrison DA, Sadique MZ, Grieve RD, et al. Trial of early, goal-directed resuscitation for septic shock. *N Engl J Med.* 2 avr 2015;372(14):1301-11.
25. Angus DC, Barnato AE, Bell D, Bellomo R, Chong C-R, Coats TJ, et al. A systematic review and meta-analysis of early goal-directed therapy for septic shock: the ARISE, ProCESS and ProMISE Investigators. *Intensive Care Med.* sept 2015;41(9):1549-60.
26. Landry DW, Oliver JA. The pathogenesis of vasodilatory shock. *N Engl J Med.* 23 août 2001;345(8):588-95.
27. Peters J, Mack G, Lister G. The importance of the peripheral circulation in critical illnesses. *Intensive Care Med.* 1 sept 2001;27(9):1446-58.
28. Ognibene FP, Parker MM, Natanson C, Shelhamer JH, Parrillo JE. Depressed Left Ventricular Performance. *Chest.* 1 mai 1988;93(5):903-10.
29. Sanfilippo F, Corredor C, Fletcher N, Landesberg G, Benedetto U, Foex P, et al. Diastolic dysfunction and mortality in septic patients: a systematic review and meta-analysis. *Intensive Care Med.* 1 juin 2015;41(6):1004-13.
30. Ogbu OC, Murphy DJ, Martin GS. How to Avoid Fluid Overload. *Curr Opin Crit Care.* août 2015;21(4):315.
31. Child DL, Cao Z, Seiberlich LE, Brown H, Greenberg J, Swanson A, et al. The costs of fluid overload in the adult intensive care unit: is a small-volume infusion model a proactive solution? *Clin Outcomes Res CEOR.* 2015;7:1.
32. Sakr Y, Vincent J-L, Reinhart K, Groeneveld J, Michalopoulos A, Sprung CL, et al. High tidal volume and positive fluid balance are associated with worse outcome in acute lung injury. *Chest.* nov 2005;128(5):3098-108.

33. Murphy CV, Schramm GE, Doherty JA, Reichley RM, Gajic O, Afessa B, et al. The importance of fluid management in acute lung injury secondary to septic shock. *Chest*. juill 2009;136(1):102-9.
34. Boyd JH, Forbes J, Nakada T, Walley KR, Russell JA. Fluid resuscitation in septic shock: a positive fluid balance and elevated central venous pressure are associated with increased mortality. *Crit Care Med*. févr 2011;39(2):259-65.
35. Sirvent J-M, Ferri C, Baró A, Murcia C, Lorenzo C. Fluid balance in sepsis and septic shock as a determining factor of mortality. *Am J Emerg Med*. févr 2015;33(2):186-9.
36. Cordemans C, Laet ID, Regenmortel NV, Schoonheydt K, Dits H, Huber W, et al. Fluid management in critically ill patients: the role of extravascular lung water, abdominal hypertension, capillary leak, and fluid balance. *Ann Intensive Care*. 2012;2(Suppl 1):S1.
37. Prowle JR, Echeverri JE, Ligabo EV, Ronco C, Bellomo R. Fluid balance and acute kidney injury. *Nat Rev Nephrol*. févr 2010;6(2):107-15.
38. Michard F, Teboul J-L. Predicting fluid responsiveness in ICU patients: a critical analysis of the evidence. *Chest*. juin 2002;121(6):2000-8.
39. Marik PE, Monnet X, Teboul J-L. Hemodynamic parameters to guide fluid therapy. *Ann Intensive Care*. 21 mars 2011;1(1):1.
40. Eskesen TG, Wetterslev M, Perner A. Systematic review including re-analyses of 1148 individual data sets of central venous pressure as a predictor of fluid responsiveness. *Intensive Care Med*. 1 mars 2016;42(3):324-32.
41. Marik PE, Cavallazzi R. Does the central venous pressure predict fluid responsiveness? An updated meta-analysis and a plea for some common sense. *Crit Care Med*. juill 2013;41(7):1774-81.
42. Monnet X, Marik PE, Teboul J-L. Prediction of fluid responsiveness: an update. *Ann Intensive Care*. 17 nov 2016;6(1):111.
43. Indices dynamiques invasifs prédictifs de la réponse au remplissage vasculaire [Internet]. [cité 25 avr 2017]. Disponible sur: https://www.srlf.org/wp-content/uploads/2015/11/0406-Reanimation-Vol13-N4-p299_305.pdf
44. Magder S, Georgiadis G, Cheong T. Respiratory variations in right atrial pressure predict the response to fluid challenge. *J Crit Care*. juin 1992;7(2):76-85.
45. Michard F, Boussat S, Chemla D, Anguel N, Mercat A, Lecarpentier Y, et al. Relation between Respiratory Changes in Arterial Pulse Pressure and Fluid Responsiveness in Septic Patients with Acute Circulatory Failure. *Am J Respir Crit Care Med*. juillet 2000;162(1):134-8.
46. Marik PE, Cavallazzi R, Vasu T, Hirani A. Dynamic changes in arterial waveform derived variables and fluid responsiveness in mechanically ventilated patients: a systematic review of the literature. *Crit Care Med*. sept 2009;37(9):2642-7.
47. Yang X, Du B. Does pulse pressure variation predict fluid responsiveness in critically ill patients? A systematic review and meta-analysis. *Crit Care*. 2014;18:650.
48. Cannesson M, Le Manach Y, Hofer CK, Goarin JP, Lehot J-J, Vallet B, et al. Assessing the Diagnostic Accuracy of Pulse Pressure Variations for the Prediction of Fluid Responsiveness: A “Gray Zone” Approach. *Anesthesiology*. août 2011;115(2):231-41.
49. Feissel M, Michard F, Mangin I, Ruyer O, Faller JP, Teboul JL. Respiratory changes in aortic

- blood velocity as an indicator of fluid responsiveness in ventilated patients with septic shock. *Chest*. mars 2001;119(3):867-73.
50. Vignon P, Repessé X, Bégot E, Léger J, Jacob C, Bouferrache K, et al. Comparison of Echocardiographic Indices Used to Predict Fluid Responsiveness in Ventilated Patients. *Am J Respir Crit Care Med*. 21 sept 2016;195(8):1022-32.
51. Vieillard-Baron A, Chergui K, Rabiller A, Peyrouset O, Page B, Beauchet A, et al. Superior vena caval collapsibility as a gauge of volume status in ventilated septic patients. *Intensive Care Med*. sept 2004;30(9):1734-9.
52. Feissel M, Michard F, Faller J-P, Teboul J-L. The respiratory variation in inferior vena cava diameter as a guide to fluid therapy. *Intensive Care Med*. sept 2004;30(9):1834-7.
53. Charbonneau H, Riu B, Faron M, Mari A, Kurrek MM, Ruiz J, et al. Predicting preload responsiveness using simultaneous recordings of inferior and superior vena cavae diameters. *Crit Care [Internet]*. 2014 [cité 26 avr 2017];18(5). Disponible sur: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4175634/>
54. Monnet X, Rienzo M, Osman D, Anguel N, Richard C, Pinsky MR, et al. Esophageal Doppler monitoring predicts fluid responsiveness in critically ill ventilated patients. *Intensive Care Med*. 1 sept 2005;31(9):1195-201.
55. Monnet X, Osman D, Ridet C, Lamia B, Richard C, Teboul J-L. Predicting volume responsiveness by using the end-expiratory occlusion in mechanically ventilated intensive care unit patients. *Crit Care Med*. mars 2009;37(3):951-6.
56. Silva S, Jozwiak M, Teboul J-L, Persichini R, Richard C, Monnet X. End-expiratory occlusion test predicts preload responsiveness independently of positive end-expiratory pressure during acute respiratory distress syndrome. *Crit Care Med*. juill 2013;41(7):1692-701.
57. Tusman G, Groisman I, Maidana GA, Scandurra A, Arca JM, Bohm SH, et al. The Sensitivity and Specificity of Pulmonary Carbon Dioxide Elimination for Noninvasive Assessment of Fluid Responsiveness. *Anesth Analg*. mai 2016;122(5):1404-11.
58. Wilkman E, Kuitunen A, Pettilä V, Varpula M. Fluid responsiveness predicted by elevation of PEEP in patients with septic shock. *Acta Anaesthesiol Scand*. 1 janv 2014;58(1):27-35.
59. Preisman S. Predicting fluid responsiveness in patients undergoing cardiac surgery: functional haemodynamic parameters including the Respiratory Systolic Variation Test and static preload indicators. *Br J Anaesth*. 14 oct 2005;95(6):746-55.
60. Biais M, Lanchon R, Sesay M, Le Gall L, Pereira B, Futier E, et al. Changes in Stroke Volume Induced by Lung Recruitment Maneuver Predict Fluid Responsiveness in Mechanically Ventilated Patients in the Operating Room. *Anesthesiology*. févr 2017;126(2):260-7.
61. Monnet X, Rienzo M, Osman D, Anguel N, Richard C, Pinsky MR, et al. Passive leg raising predicts fluid responsiveness in the critically ill. *Crit Care Med*. mai 2006;34(5):1402-7.
62. Monnet X, Teboul J-L. Passive leg raising: five rules, not a drop of fluid! *Crit Care [Internet]*. 2015 [cité 27 avr 2017];19(1). Disponible sur: <https://www.ncbi.nlm.nih.gov/doi/10.1186/s13054-015-0982-2>
63. Monnet X, Marik P, Teboul J-L. Passive leg raising for predicting fluid responsiveness: a systematic review and meta-analysis. *Intensive Care Med*. déc 2016;42(12):1935-47.

64. Tiberghien N. Efficacité diagnostique de la compression abdominale pour la prédiction de la précharge dépendance chez l'enfant en insuffisance circulatoire aigüe [Thèse]. [Lyon]: Université Claude Bernard Lyon 1; 2017.
65. Jacquet-Lagrèze M. Diagnostic accuracy of abdominal compression for predicting fluid responsiveness in children. *Eur J Anaesthesiol.* juin 2017;34(e-Supplement 55):141.
66. Muller L, Toumi M, Bousquet P-J, Riu-Poulenc B, Louart G, Candela D, et al. An Increase in Aortic Blood Flow after an Infusion of 100 ml Colloid over 1 Minute Can Predict Fluid Responsiveness: The Mini-fluid Challenge Study. *Anesthesiology.* sept 2011;115(3):541-7.
67. Vistisen ST, Juhl-Olsen P. Where are we heading with fluid responsiveness research?: *Curr Opin Crit Care.* août 2017;23(4):318-25.
68. Vistisen ST, Andersen KK, Frederiksen CA, Kirkegaard H. Variations in the pre-ejection period induced by ventricular extra systoles may be feasible to predict fluid responsiveness. *J Clin Monit Comput.* août 2014;28(4):341-9.
69. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *BMJ.* 4 janv 2003;326(7379):41-4.
70. Griner PF, Mayewski RJ, Mushlin AI, Greenland P. Selection and interpretation of diagnostic tests and procedures. Principles and applications. *Ann Intern Med.* avr 1981;94(4 Pt 2):557-92.
71. Jacquet-Lagrèze M. La variation du dioxyde de carbone télé-expiratoire au cours d'une épreuve de remplissage permet d'évaluer la réponse au remplissage au bloc opératoire. [Université Claude Bernard Lyon 1]; 2015.
72. Eusebi P. Diagnostic Accuracy Measures. *Cerebrovasc Dis.* 2013;36(4):267-72.
73. Welch HG, Black WC. Overdiagnosis in cancer. *J Natl Cancer Inst.* 5 mai 2010;102(9):605-13.
74. Koffijberg H, van Zaane B, Moons KG. From accuracy to patient outcome and cost-effectiveness evaluations of diagnostic tests and biomarkers: an exemplary modelling study. *BMC Med Res Methodol.* 31 janv 2013;13:12.
75. Gilles Bouvenot, Pierre Ambrosi , Pascal Auquier , David Debensason , Collectif. S'entraîner à la lecture critique d'un article médical. Med-Line Editions. (Med-Line ECN).
76. Altman DG, Bland JM. Statistics Notes: Diagnostic tests 2: predictive values. *BMJ.* 9 juill 1994;309(6947):102.
77. Deeks JJ, Altman DG. Diagnostic tests 4: likelihood ratios. *BMJ.* 15 juill 2004;329(7458):168-9.
78. Nomogram for Bayes's Theorem. *N Engl J Med.* 31 juill 1975;293(5):257-257.
79. Zou KH, O'Malley AJ, Mauri L. Receiver-Operating Characteristic Analysis for Evaluating Diagnostic Tests and Predictive Models. *Circulation.* 6 févr 2007;115(5):654-7.
80. Guyatt GH, Tugwell PX, Feeny DH, Haynes RB, Drummond M. A framework for clinical evaluation of diagnostic technologies. *CMAJ Can Med Assoc J J Assoc Médicale Can.* 15 mars 1986;134(6):587-94.
81. Reid MC, Lachs MS, Feinstein AR. Use of Methodological Standards in Diagnostic Test

Research: Getting Better but Still Not Good. JAMA. 23 août 1995;274(8):645-51.

82. Lijmer JG, Mol BW, Heisterkamp S, Bossel GJ, Prins MH, Meulen JHP van der, et al. Empirical Evidence of Design-Related Bias in Studies of Diagnostic Tests. JAMA. 15 sept 1999;282(11):1061-6.

83. Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. JAMA. 28 août 1996;276(8):637-9.

84. Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. Lancet Lond Engl. 14 avr 2001;357(9263):1191-4.

85. Moher D, Jones A, Lepage L, CONSORT Group (Consolidated Standards for Reporting of Trials). Use of the CONSORT statement and quality of reports of randomized trials: a comparative before-and-after evaluation. JAMA. 18 avr 2001;285(15):1992-5.

86. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD Initiative. Ann Intern Med. 7 janv 2003;138(1):40-4.

87. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD Initiative. Radiology. janv 2003;226(1):24-8.

88. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. Standards for Reporting of Diagnostic Accuracy. Clin Chem. janv 2003;49(1):1-6.

89. Gedda M. Traduction française des lignes directrices STARD pour l'écriture et la lecture des études sur la précision des tests diagnostiques. Kinésithérapie Rev. janv 2015;15(157):45-9.

90. Korevaar DA, Wang J, van Enst WA, Leeftang MM, Hooft L, Smidt N, et al. Reporting Diagnostic Accuracy Studies: Some Improvements after 10 Years of STARD. Radiology. 27 oct 2014;274(3):781-9.

91. Korevaar DA, Enst WA van, Spijker R, Bossuyt PMM, Hooft L. Reporting quality of diagnostic accuracy studies: a systematic review and meta-analysis of investigations on adherence to STARD. Evid Based Med. 4 janv 2014;19(2):47-54.

92. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials | The EQUATOR Network [Internet]. [cité 1 août 2017]. Disponible sur: <http://www.equator-network.org/reporting-guidelines/consort/>

93. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. BMJ. 28 oct 2015;h5527.

94. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, et al. STARD 2015: An Updated List of Essential Items for Reporting Diagnostic Accuracy Studies. Radiology. déc 2015;277(3):826-32.

95. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, et al. STARD 2015: An Updated List of Essential Items for Reporting Diagnostic Accuracy Studies. Clin Chem. déc 2015;61(12):1446-52.

96. Cochrane Handbook for Systematic Reviews of Interventions | Cochrane Training [Internet]. [cité 12 août 2017]. Disponible sur: <http://training.cochrane.org/handbook>
97. Sacks HS, Berrier J, Reitman D, Ancona-Berk VA, Chalmers TC. Meta-Analyses of Randomized Controlled Trials. *N Engl J Med*. février 1987;316(8):450-5.
98. Sacks HS, Reitman D, Pagano D, Kupelnick B. Meta-analysis: an update. *Mt Sinai J Med N Y*. sept 1996;63(3-4):216-24.
99. Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. *The Lancet*. 27 nov 1999;354(9193):1896-900.
100. Moher D, Tetzlaff J, Tricco AC, Sampson M, Altman DG. Epidemiology and Reporting Characteristics of Systematic Reviews. *PLOS Med*. 27 mars 2007;4(3):e78.
101. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ*. 21 juill 2009;339:b2535.
102. Gedda M. Traduction française des lignes directrices PRISMA pour l'écriture et la lecture des revues systématiques et des méta-analyses. *Kinésithérapie Rev*. janv 2015;15(157):39-44.
103. pubmeddev. Home - PubMed - NCBI [Internet]. [cité 12 août 2017]. Disponible sur: <https://www.ncbi.nlm.nih.gov.docelec.univ-lyon1.fr/pubmed/>
104. Cavallaro F, Sandroni C, Marano C, La Torre G, Mannocci A, De Waure C, et al. Diagnostic accuracy of passive leg raising for prediction of fluid responsiveness in adults: systematic review and meta-analysis of clinical studies. *Intensive Care Med*. sept 2010;36(9):1475-83.
105. Monnet X, Persichini R, Ktari M, Jozwiak M, Richard C, Teboul J-L. Precision of the transpulmonary thermodilution measurements. *Crit Care Lond Engl*. 27 août 2011;15(4):R204.
106. Stetz CW, Miller RG, Kelly GE, Raffin TA. Reliability of the Thermodilution Method in the Determination of Cardiac Output in Clinical Practice. *Am Rev Respir Dis*. décembre 1982;126(6):1001-4.
107. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical Evidence of Bias: Dimensions of Methodological Quality Associated With Estimates of Treatment Effects in Controlled Trials. *JAMA*. 1 févr 1995;273(5):408-12.
108. Wetterslev M, Møller-Sørensen H, Johansen RR, Perner A. Systematic review of cardiac output measurements by echocardiography vs. thermodilution: the techniques are not interchangeable. *Intensive Care Med*. août 2016;42(8):1223-33.
109. Philbrick JT, Horwitz RI, Feinstein AR. Methodologic problems of exercise testing for coronary artery disease: groups, analysis and bias. *Am J Cardiol*. nov 1980;46(5):807-12.
110. Detrano R, Gianrossi R, Froelicher V. The diagnostic accuracy of the exercise electrocardiogram: a meta-analysis of 22 years of research. *Prog Cardiovasc Dis*. déc 1989;32(3):173-206.
111. Whiting PF, Rutjes AWS, Westwood ME, Mallett S. A systematic review classifies sources of bias and variation in diagnostic test accuracy studies. *J Clin Epidemiol*. 1 oct 2013;66(10):1093-104.
112. Péron J, Maillet D, Gan HK, Chen EX, You B. Adherence to CONSORT Adverse Event Reporting Guidelines in Randomized Clinical Trials Evaluating Systemic Cancer Therapy: A Systematic Review. *J Clin Oncol*. nov 2013;31(31):3957-63.

113. Assessment of the Quality of Reporting of Randomised Controlled Trials in Otorhinolaryngologic Literature – Adherence to the CONSORT Statement [Internet]. [cité 4 août 2017]. Disponible sur: <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0122328>
114. Smidt N, Overbeke J, de Vet H, Bossuyt P. Endorsement of the STARD Statement by biomedical journals: survey of instructions for authors. *Clin Chem*. 2007;53(11):1983–1985.
115. Sims MT, Henning NM, Wayant CC, Vassar M. Do emergency medicine journals promote trial registration and adherence to reporting guidelines? A survey of “Instructions for Authors”. *Scand J Trauma Resusc Emerg Med*. 24 nov 2016;24(1):137.
116. Charles P, Giraudeau B, Dechartres A, Baron G, Ravaud P. Reporting of sample size calculation in randomised controlled trials: review. *BMJ*. 12 mai 2009;338:b1732.
117. Huser V, Cimino JJ. Evaluating adherence to the International Committee of Medical Journal Editors’ policy of mandatory, timely clinical trial registration. *J Am Med Inform Assoc*. juin 2013;20(e1):e169-74.
118. Korevaar DA, Ochodo EA, Bossuyt PMM, Hooft L. Publication and Reporting of Test Accuracy Studies Registered in ClinicalTrials.gov. *Clin Chem*. 1 avr 2014;60(4):651–9.
119. Hong PJ, Korevaar DA, McGrath TA, Ziai H, Frank R, Alabousi M, et al. Reporting of imaging diagnostic accuracy studies with focus on MRI subgroup: Adherence to STARD 2015. *J Magn Reson Imaging*. :n/a-n/a.
120. Coppus SFPJ, van der Veen F, Bossuyt PMM, Mol BWJ. Quality of reporting of test accuracy studies in reproductive medicine: impact of the Standards for Reporting of Diagnostic Accuracy (STARD) initiative. *Fertil Steril*. nov 2006;86(5):1321–9.
121. Smidt N, Rutjes AWS, Van der Windt D, Ostelo R, Bossuyt PM, Reitsma JB, et al. The quality of diagnostic accuracy studies since the STARD statement Has it improved? *Neurology*. 2006;67(5):792–797.
122. Wilczynski NL. Quality of Reporting of Diagnostic Accuracy Studies: No Change Since STARD Statement Publication—Before-and-after Study. *Radiology*. 1 sept 2008;248(3):817–23.
123. Selman TJ, Morris RK, Zamora J, Khan KS. The quality of reporting of primary test accuracy studies in obstetrics and gynaecology: application of the STARD criteria. *BMC Womens Health*. 23 mars 2011;11:8.
124. Gallo L, Hua N, Mercuri M, Silveira A, Worster A. Adherence to Standards for Reporting Diagnostic Accuracy in Emergency Medicine Research. *Acad Emerg Med*. :n/a-n/a.
125. Grob A t. m., van der Vaart L r., Withagen M i. j., van der Vaart C h. The quality of reporting of diagnostic accuracy studies in pelvic floor transperineal three-dimensional ultrasound: a systematic review. *Ultrasound Obstet Gynecol*. 1 déc 2016;n/a-n/a.
126. Walther S, Schueler S, Tackmann R, Schuetz GM, Schlattmann P, Dewey M. Compliance with STARD Checklist among Studies of Coronary CT Angiography: Systematic Review. *Radiology*. 10 janv 2014;271(1):74–86.
127. Choi YJ, Chung MS, Koo HJ, Park JE, Yoon HM, Park SH. Does the Reporting Quality of Diagnostic Test Accuracy Studies, as Defined by STARD 2015, Affect Citation? *Korean J Radiol*. oct 2016;17(5):706.
128. Whiting P, Rutjes AW, Reitsma JB, Glas AS, Bossuyt PM, Kleijnen J. Sources of Variation and Bias in Studies of Diagnostic Accuracy A Systematic Review. *Ann Intern Med*. 2004;140(3):189–

202.

129. Smidt N, Rutjes AW, Windt DA van der, Ostelo RW, Bossuyt PM, Reitsma JB, et al. Reproducibility of the STARD checklist: an instrument to assess the quality of reporting of diagnostic accuracy studies. *BMC Med Res Methodol*. 15 mars 2006;6(1):12.

130. Zafar A, Khan GI, Siddiqui MR. The quality of reporting of diagnostic accuracy studies in diabetic retinopathy screening: a systematic review. *Clin Experiment Ophthalmol*. 1 août 2008;36(6):537-42.

131. Hosmer DW, Lemeshow S. *Applied Logistic Regression*. John Wiley & Sons; 2004. 397 p.

132. The EQUATOR Network | Enhancing the QUAlity and Transparency Of Health Research [Internet]. Disponible sur: <http://www.equator-network.org/>

133. Toolkits | The EQUATOR Network [Internet]. Disponible sur: <http://www.equator-network.org/toolkits/>

134. STARD 2015: An Updated List of Essential Items for Reporting Diagnostic Accuracy Studies | The EQUATOR Network [Internet]. Disponible sur: <http://www.equator-network.org/reporting-guidelines/stard/>

Guillaume Izaute

Qualité du *reporting* dans les études diagnostiques de réponse au remplissage vasculaire : revue systématique sur dix ans.

RESUME :

Introduction : Les recommandations STARD (*Standards for the Reporting of Diagnostic Accuracy Studies*) ont été élaborées en 2003 pour améliorer la qualité du *reporting* des études diagnostiques. Elles ont été mises à jour en 2015. Le but de ce travail était d'évaluer la qualité du *reporting* des études évaluant la réponse au remplissage vasculaire, en utilisant les critères STARD.

Matériels et méthodes : Nous avons recherché dans MEDLINE *via* PubMed les études diagnostiques de précharge dépendance publiées entre 2005 et 2014. Après avoir élaboré une définition consensuelle pour chaque item de la liste STARD, nous avons apprécié leur présence dans les études sélectionnées. Chaque étude obtenait un score de qualité qui s'étend de 0 à 25 pour les critères STARD 2003 et de 0 à 30 pour les critères STARD 2015. Les caractéristiques des études associées à un meilleur *reporting* étaient analysées.

Résultats et discussion : 104 études ont été analysées, majoritairement européennes et publiées dans des revues d'impact factor moyen de 4,8 (IC 95% : 4,4-5,0). Le score STARD 2003 moyen était de 12,0 sur une échelle de 0 à 25 (IC 95% : 11,4-12,5). Le score STARD 2015 moyen était de 14,9 sur une échelle de 0 à 30 (IC 95% : 14,3 à 15,5). Le coefficient de corrélation linéaire entre les score STARD 2003 et STARD 2015 était $r=0,859$. Les moyennes des scores sont restées stables sur les dix années étudiées. En analyse univariée, les caractéristiques des études associées à un meilleur score STARD 2003 étaient : le milieu dans lequel se déroule l'étude, le journal de publication, l'*impact factor* du journal, et la présence de STARD dans les recommandations faites aux auteurs de la revue. En analyse multivariée, seul le fait que l'étude se déroule dans une unité de soins intensifs était associé à un meilleur score STARD 2003.

Conclusion : Les scores STARD 2003 et STARD 2015 sont bas. Plusieurs items sont insuffisamment reportés. Nous recommandons l'utilisation systématique des critères STARD dans l'élaboration et le *reporting* des futures études évaluant la précharge dépendance et la réponse au remplissage vasculaire.

MOTS CLES : STARD, études diagnostiques, *reporting*, précharge dépendance.

JURY : Président : Monsieur le Professeur Jean-Luc Fellahi
Membres : Monsieur le Professeur Vincent Piriou
Monsieur le Professeur Jean-Christophe Richard
Monsieur le Docteur Matthias Jacquet-Lagrèze

DATE DE SOUTENANCE : 6 septembre 2017

ADRESSE DE L'AUTEUR : Guillaume Izaute
316 rue Duguesclin
69007 Lyon
gizaute@gmail.com