



HAL
open science

Méthodologies d'évaluation de l'optimalité des soins : exemples des délais diagnostiques et des infections bactériennes sévères de l'enfant

Elise Launay

► **To cite this version:**

Elise Launay. Méthodologies d'évaluation de l'optimalité des soins : exemples des délais diagnostiques et des infections bactériennes sévères de l'enfant. Santé publique et épidémiologie. Université Sorbonne Paris Cité, 2015. Français. NNT : 2015USPCB134 . tel-01578101

HAL Id: tel-01578101

<https://theses.hal.science/tel-01578101>

Submitted on 28 Aug 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



UNIVERSITE PARIS DESCARTES

ECOLE DOCTORALE DE SANTE PUBLIQUE (EDSP 570)

Thèse de doctorat en santé publique – Spécialité épidémiologie

Présentée par

Elise LAUNAY

Méthodologies d'évaluation de l'optimalité des soins :
exemples des délais diagnostiques
et des infections bactériennes sévères de l'enfant

Soutenue publiquement le 27 novembre 2015

Pour obtenir le grade de docteur de l'Université Paris Descartes

Devant le jury composé de :

Rapporteurs : Monsieur Alain Gervais

 Madame Leïla Moret

Examineurs : Monsieur Yves Gillet

 Monsieur Olivier Lortholary

 Monsieur Jean-Christophe Thalabard (Président)

Directeur de Thèse : Monsieur Martin Chalumeau

REMERCIEMENTS

Je remercie Madame **Leïla MORET** d'avoir accepté d'être rapporteur de cette thèse. Votre implication dans l'évaluation de la qualité des soins vous permet de porter sur mon travail un regard qui m'est précieux. Veuillez recevoir toute ma gratitude et mon plus profond respect.

Je remercie Monsieur **Alain GERVAIX** d'avoir accepté d'être rapporteur de cette thèse. Votre excellence dans le domaine de l'infectiologie et des urgences pédiatriques est pour moi un modèle. Veuillez recevoir toute ma gratitude et mon plus profond respect.

Je remercie Monsieur **Jean-Christophe THALABARD** d'avoir accepté de présider le jury de cette thèse. Votre expertise en biostatistiques et votre expérience de clinicien vous permettent de porter sur ce travail un regard critique qui m'est cher. Veuillez recevoir toute ma gratitude et mon plus profond respect.

Je remercie Monsieur **Olivier LORTHOLARY** d'avoir accepté d'être assesseur de cette thèse. Vos compétences en infectiologie constituent une référence internationale et votre engagement pour l'infectiologie pédiatrique est un soutien précieux. Veuillez recevoir toute ma gratitude et mon plus profond respect.

Je remercie Monsieur **Yves GILLET** d'avoir accepté d'être assesseur de cette thèse. Je suis admirative de tes connaissances dans le domaine de l'infectiologie et des urgences pédiatriques, de ta disponibilité et de ton humilité. Accepte de recevoir ici le témoignage de toute ma gratitude et de mon plus profond respect.

Je remercie Monsieur **Martin CHALUMEAU** de m'avoir proposé puis d'avoir dirigé ce travail doctoral. Je suis admirative de tes compétences cliniques et épidémiologiques et de ta volonté de faire de la pédiatrie générale une discipline d'excellence. Je te remercie pour ta rigueur, ton dynamisme et ton honnêteté dont j'ai beaucoup appris au cours de ce travail doctoral. Accepte de recevoir ici le témoignage de ma sincère reconnaissance et de mon plus profond respect.

Je tiens à remercier également également toutes les personnes qui m'ont permis de mener à bien ce travail de thèse :

Madame **Christèle GRAS- LE GUEN**. Merci infiniment pour ton soutien et tes conseils tout au long de ce travail (et au-delà). Ton excellence en infectiologie pédiatrique, ton dynamisme et ton inventivité sont pour moi des moteurs, ta confiance et ton amitié me sont précieuses.

Messieurs **Rémy ASSATHIANY et Alain MARTINOT**. Cette thèse n'aurait jamais vu le jour sans vos expertises, votre patience et vos précieux conseils qui nous ont permis d'optimiser les études DIABACT. Acceptez de recevoir ici le témoignage de ma sincère reconnaissance et de mon plus profond respect.

Monsieur **Jérémy COHEN**, Madame **Michèle MORFOUACE** et Madame **Claire SAULOU** pour leurs aides précieuses et amicales dans la réalisation de la première partie de cette thèse. Merci infiniment Jérémy de t'être rendu disponible alors que tes travaux personnels avaient largement de quoi remplir tes journées.

Madame **Catherine DENEUX-THARAUX** et Madame **Marie-Hélène BOUVIER-COLLE**. Vos travaux ont inspiré ce travail doctoral et vos conseils méthodologiques nous ont permis de le faire progresser.

Madame **Mariska LEEFLANG** et Messieurs **Patrick BOSSUYT, Pierre BUEKENS, Jon DEEKS, Timothy DYE, Richard FELTBOWER, Andrea FERRARI, Mikael KRAMER, David MOHER, Karel MOONS** et **Erik von ELM** pour m'avoir fait l'honneur d'accepter de constituer le comité scientifique d'une partie de ce travail doctoral.

Messieurs **Pierre-Yves ANCEL et François GOFFINET** pour m'avoir donné la chance de réaliser cette thèse au sein de l'équipe EPOPé.

Monsieur **Philippe RAVAUD** pour nous avoir fait partager son expérience et ses précieux conseils dans le domaine de la méthodologie de la recherche.

Monsieur **Jean-Christophe ROZE**. Merci infiniment pour votre soutien tout au long de ce travail. Votre engagement pour l'amélioration des soins, de l'enseignement jusqu'au « lit du malade », est pour moi un exemple de dynamisme et de courage. Veuillez recevoir ici le témoignage de mon plus profond respect.

Monsieur **Dominique GENDREL**. Merci pour votre bienveillance tout au long de mon parcours depuis mon stage d'interne à St Vincent de Paul jusqu'à ce travail doctoral. Je suis admirative de votre engagement pour l'infectiologie pédiatrique et la pédiatrie générale. Veuillez recevoir ici le témoignage de mon plus profond respect.

Madame **Pascale JOLLIET** et Monsieur **Jean-Michel ROGEZ**. Merci de la confiance que vous m'avez accordée en me nommant maître de conférence des universités. Ce temps universitaire m'a été précieux pour mener à bien ce travail de thèse.

Madame **Fanny HENAFF**, Madame **Elise MARTIN**, Monsieur **Pierre BOURGOIN**, Monsieur **Thomas BLANCHAIS**, Madame **Pauline POSTEC** et Monsieur **Xavier Mickaelli** et l'équipe du CIC Mère-Enfant de Nantes et notamment, Mesdames **Céline ROBIN**, **Evelyne JORAM** et **Chloé ADJAOU** et Monsieur **Arnaud LEGRAND** pour leur contribution aux études DIABACT.

Monsieur **Jean BOUYER**, Madame **Béatrice BLONDEL** et Messieurs **Babak KHOSHNOOD**, et **Antoine CHAMBAZ** pour m'avoir permis de bénéficier de leurs enseignements.

Merci à l'ensemble de l'équipe EPOPé et particulièrement à Mesdames **Edith CHOPLIN** et **Sophie GOUVAERT** pour leur disponibilité et leur bienveillance.

Merci à mes collègues et amis étudiants en thèse ou récemment thésés au sein de l'équipe et particulièrement à Madame **Nathalie BERTILLE**, Monsieur **Jean-François BRASME** et Mesdames **Pauline SCHREDEL** et **Anne-Sylvia SACRI**.

Enfin à mes collègues et amies du service de pédiatrie générale du CHU de Nantes pour leur soutien, leur bienveillance et leur bonne humeur.

Je dédie cette thèse

A mes parents, Jacqueline et Pierre, pour m'avoir appris à être têtue sans trop en avoir l'air...

A mes sœurs, Murielle, Fabienne et Caroline et à mes beaux-frères Etienne, Frédéric et Ludovic, pour me rappeler qu'il ne faut pas trop se prendre au sérieux... (et pour corriger mes fautes d'orthographe...)

A mes neveu et nièces Elliott, Ninon et Constance pour suivre l'exemple de leurs parents

A mes filleules Louise et Thaïs pour leur joie de vivre

A ma famille pour leur soutien à toute épreuve (et particulièrement à Solange et à ses soeurs pour m'avoir accueillie à Paris)

A mes amis de part et d'autre de l'Atlantique pour leur complicité

RESUME (1690 caractères espaces compris/1700)

Les objectifs de cette thèse étaient de produire des connaissances nouvelles sur les méthodologies d'évaluation de l'optimalité des soins avec l'exemple des délais diagnostiques et des infections bactériennes sévères de l'enfant (IBS).

Nous avons mis en évidence, dans deux revues systématiques de la littérature, que les points méthodologiques potentiellement associés à des risques de biais et d'obstacles à la transportabilité des résultats étaient rarement rapportés dans les études primaires sur les délais diagnostiques ou rarement évalués par les auteurs de méta-analyses. Nous avons donc construit et validé internationalement une *reporting guideline* pour aider les scientifiques à prendre en compte ces points méthodologiques critiques.

Nous avons montré par une enquête confidentielle avec comité d'experts en population que : (i) les prises en charge étaient suboptimales pour 76% des enfants décédés d'IBS, (ii) un retard au recours médical, une sous-évaluation de la gravité ou un retard à l'antibiothérapie étaient retrouvés dans la prise en charge de respectivement 20%, 20% et 24% des enfants atteints d'IBS, (iii) les soins suboptimaux étaient indépendamment et fortement associés au risque de décès et (iv) les soins suboptimaux étaient plus fréquents chez les enfants de moins d'un an ou lorsque qu'ils n'étaient pas administrés par un médecin spécifiquement formé.

La minimisation des biais dans la sélection des participants et la mesure de l'optimalité et la prise en compte de facteurs de confusion comme la sévérité intrinsèque de la maladie sont des éléments clés de l'évaluation de l'optimalité des soins afin de produire des messages cliniques correctifs valides.

MOTS-CLEFS

Soins suboptimaux, retard au diagnostic, délai diagnostique, infection bactérienne sévère, sepsis, enfant, *reporting guideline*, risque de biais, applicabilité, qualité méthodologique.

ABSTRACT (1568 characters with spaces)

Methods to assess the optimality of care: examples of time to diagnosis and serious bacterial infection in children

The aim of this thesis was to product new knowledge about the methodology on how to assess the optimality of care with the examples of time to diagnosis and serious bacterial infection (SBI).

In two systematic reviews, we found that the key methodological points potentially related to risks of bias or threats to transportability were rarely reported in the primary studies and rarely evaluated by authors of systematic reviews. Then, we developed and internationally validated a reporting guideline to help scientists to better take into consideration these critical methodological points.

In a population-based confidential inquiry, we found that: (i) care was suboptimal in 76% of the initial management of children who died from SBI, (ii) delayed first medical contact, undervaluation of severity or delayed antibiotic administration were detected in the management of 20%, 20% and 24% of children admitted to intensive care for a SBI, respectively, (iii) the total number of suboptimal cares delivered during the management was independently associated with death, and (iv) suboptimal cares were more frequent in children younger than one year old and if the care was delivered by a non specialist physician.

Minimizing the risks of bias both in the selection process of the study population and in the assessment of the optimality of care, and taking into account confounding factors such as the intrinsic severity of the disease are keys elements to ensure a reliable evaluation of optimality of care in order to produce effective corrective clinical messages.

KEY WORDS

Suboptimal care, delayed diagnosis, time to diagnosis, serious bacterial infection, sepsis, children, reporting guideline, risk of bias, threat to transportability, methodological quality.

UNITE DE RECHERCHE

Cette thèse a été préparée au sein de l'unité Inserm UMR S953
« Recherche épidémiologique en santé périnatale et en santé des femmes
et des enfants »

Directeur : Professeur François GOFFINET

Puis au sein de l'unité Inserm UMR 1153
«Centre de Recherche Épidémiologie et Statistiques Sorbonne Paris Cité
(CRESS)»

Directeur : Professeur Philippe RAVAUD

Equipe EPOPé - «Epidémiologie Périnatale, Obstétricale et Pédiatrique»

Responsable : Professeur Pierre-Yves ANCEL

Maternité de Port Royal

53, avenue de l'Observatoire 75014 PARIS

Téléphone : 01 42 34 55 70

Fax : 01 43 26 89 79

ABREVIATIONS

CépiDC :	Centre d'épidémiologie sur les causes médicales de Décès
CONSORT :	CONsolidated Standards Of Reporting randomized Trial
DAG :	Directed Acyclic Graph
IBS :	Infection Bactérienne Sévère
IC 95% :	Intervalle de Confiance à 95%
EIQ :	Espace InterQuartile
EQUATOR :	Enhancing the QUALity and Transparency Of Research
MOOSE :	Meta-analysis Of Observational Studies in Epidemiology
OR :	Odds Ratio
ORa :	Odds Ratio ajusté
PCR :	Polymerase Chain Reaction
PRISM :	Pediatric Risk of Mortality
PRISMA	Preferred Reporting Items for Systematic reviews and Meta-Analyses
PROSPERO :	international prospective register of systematic reviews
QUOROM :	QUality Of Reporting Of Meta-analyses
RMM :	Revue de Mortalité Morbidité
SAMU :	Service d'Aide Médicale d'Urgence
SIRS :	Systemic Inflammatory Response Syndrome
STARD :	STAndards for Reporting of Diagnostic accuracy
STROBE :	STrengthening the Reporting of OBServational studies in Epidemiology

TABLE DES MATIERES

PUBLICATIONS et COMMUNICATIONS	19
GLOSSAIRES	23
1. <i>Glossaire d'épidémiologie</i>	23
2. <i>Glossaire d'infectiologie</i>	31
INTRODUCTION GENERALE	33
1. <i>Etat des connaissances et justification du travail doctoral</i>	35
2. <i>Hypothèses</i>	41
3. <i>Objectifs et plan du travail doctoral</i>	42
PARTIE I - Méthodologie d'évaluation des délais diagnostiques	45
Introduction	47
CHAPITRE 1 Qualité du <i>reporting</i> des études sur les délais diagnostiques : revue systématique en pédiatrie	61
1. <i>Introduction</i>	63
2. <i>Matériel et méthodes</i>	64
3. <i>Résultats</i>	68
4. <i>Discussion</i>	69
5. <i>Conclusion : une reporting guideline spécifique est nécessaire</i>	72
6. <i>Figures et Tableaux</i>	75
CHAPITRE 2 Les outils d'évaluation de la qualité des études primaires dans les revues systématiques portant sur l'évaluation d'un délai diagnostique : revue systématique méthodologique	79
1. <i>Introduction</i>	81
2. <i>Méthodes</i>	82
3. <i>Résultats</i>	86
4. <i>Discussion</i>	89
5. <i>Conclusions</i>	93
6. <i>Figures et Tableaux</i>	94
CHAPITRE 3 Affinement et validation d'une <i>checklist</i> pour le <i>reporting</i> des études évaluant un délai diagnostique	101
1. <i>Introduction</i>	102
2. <i>Méthodes</i>	102
3. <i>Résultats</i>	105

4. <i>Discussion</i>	107
5. <i>Conclusions et perspectives</i>	110
6. <i>Figures et tableaux</i>	111
PARTIE II - Fréquence, déterminants et conséquences des soins suboptimaux dans la prise en charge initiale des enfants atteints d'infections bactériennes sévères	119
Introduction	121
Sources de données des études DIABACT	129
1. <i>Méthodologie générale</i>	131
2. <i>Population</i>	131
3. <i>Identification des cas</i>	133
4. <i>Collecte des données</i>	134
5. <i>Evaluation de l'optimalité des soins</i>	135
6. <i>Ethique et réglementations</i>	136
CHAPITRE 1 Soins suboptimaux dans la prise en charge initiale des enfants décédés d'infection bactérienne sévère	139
1. <i>Introduction</i>	141
2. <i>Méthodologie</i>	141
3. <i>Résultats</i>	142
4. <i>Discussion</i>	145
5. <i>Figures et Tableaux</i>	150
CHAPITRE 2 Fréquences, déterminants et conséquences des soins suboptimaux dans la prise en charge initiale des enfants atteints d'infection bactérienne sévère	153
1. <i>Introduction</i>	155
2. <i>Méthodes</i>	157
3. <i>Résultats</i>	159
4. <i>Discussion</i>	163
5. <i>Figures et tableaux</i>	171
SYNTHESE, CONCLUSIONS ET PERSPECTIVES	177
1. <i>Synthèse des résultats</i>	179
2. <i>Implications de nos résultats pour la méthodologie d'évaluation des soins</i>	181
3. <i>Perspectives de recherche</i>	185
REFERENCES BIBLIOGRAPHIQUES	193

ANNEXES	211
<i>Annexe 1 : Article issu de la Partie I - Chapitre 1</i>	213
<i>Annexe 2 : Manuscrit issu de la Partie I - Chapitre 2</i>	215
<i>Annexe 3 : Article issu de la Partie II - Chapitre 1</i>	217
<i>Annexe 4 : Article issu de la partie II – Chapitre 2</i>	219
<i>Annexe 5 : Articles didactiques en français</i>	221
<i>Annexe 6 : Résumé du protocole de l'étude DIABACT III</i>	223
<i>Annexe 7 : Protocole de l'étude DIABACT IV</i>	225

PUBLICATIONS et COMMUNICATIONS

PUBLICATIONS ORIGINALES EN LANGUE ANGLAISE

Article préparatoire (voir avertissements page 43)

Elise Launay, Christèle Gras-Le Guen, Alain Martinot, Rémy Assathiany, Thomas Blanchais, Nadjette Mourdi, Albertine Aouba, Marie-Hélène Bouvier-Colle, Jean-Christophe Rozé, Martin Chalumeau. Suboptimal care in the initial management of children who died from severe bacterial infection: a population-based confidential inquiry. *Pediatric Critical Care Medicine* 2010; 11:469-74 (Facteur d'impact = 2,3)

Articles publiés

Elise Launay, Michele Morfouace, Catherine Deneux-Tharoux, Christèle Gras Le-Guen, Philippe Ravaud et Martin Chalumeau. Quality of reporting of studies evaluating time to diagnosis: a systematic review in paediatrics. *Archives of Disease in Childhood*. 2014; 99:244-50 (Facteur d'impact = 2,9)

Elise Launay, Christèle Gras-Le Guen, Alain Martinot, Rémy Assathiany, Elise Martin, Thomas Blanchais, Catherine Deneux-Tharoux, Jean-Christophe Rozé, Martin Chalumeau. Why Children with Severe Bacterial Infection Die: A Population-Based Study of Determinants and Consequences of Suboptimal Care with a Special Emphasis on Methodological Issues. *PLoS One*. 2014; 9:e107286 (Facteur d'impact = 3,5)

Manuscrit soumis

Elise Launay, Jérémie F. Cohen, Michèle Morfouace, Christèle Gras-Le Guen, Philippe Ravaud, Martin Chalumeau. “Quality assessment” of primary studies in systematic reviews of time to diagnosis: a methodological systematic review and call for clarification.

Manuscrit en cours de rédaction

Elise Launay, Jérémie F. Cohen, Patrick Bossuyt, Pierre Buekens, Jon Deeks, Timothy Dye, Richard Feltbower, Andrea Ferrari, Mikael Kramer, Mariska Leeftang, David Moher, Karel Moons, Erik von Elm, Philippe Ravaud, Martin Chalumeau. Reporting studies on time to diagnosis and diagnosis delay: refinement and international validation of a guideline. *En cours de rédaction*.

COMMUNICATIONS (* si orales)

Elise Launay, Christèle Gras-Le Guen, Alain Martinot, Rémy Assathiany, Elise Martin, Thomas Blanchais, Catherine Deneux-Tharaux, Jean-Christophe Rozé, Martin Chalumeau. Why children with severe bacterial infection die? Consequences and determinants of suboptimal cares in a population based study. *Annual meeting of the European Society of Pediatric Infectious Disease*, Dublin, mai 2014.

Elise Launay, Michele Morfouace, Catherine Deneux-Tharaux, Christèle Gras-Le Guen, Philippe Ravaud et Martin Chalumeau. Quality of reporting of studies evaluating time to diagnosis: a systematic review in pediatrics. *International Forum on Quality and Safety in Healthcare*, Paris, avril 2014.

*Elise Launay, Elise Martin, Rémy Assathiany, Alain Martinot, Thomas Blanchais, Jean-Christophe Rozé, Martin Chalumeau, Christèle Gras-Le Guen. Evaluation de la prise en charge initiale des infections bactériennes sévères de l'enfant. *Congrès de la Société Française de pédiatrie*, Clermont-Ferrand, mai 2013.

*Elise Launay, Michele Morfouace, Catherine Deneux-Tharaux, Christèle Gras-Le Guen, Philippe Ravaud et Martin Chalumeau. Qualité du reporting des études évaluant les délais diagnostiques en pédiatrie : revue systématique. *Congrès de la Société Française de Pédiatrie*, Clermont-Ferrand, mai 2013.

CONFERENCES SUR INVITATION (si articles didactiques associés)**

Elise Launay, Christèle Gras-Le Guen, Martin Chalumeau. Comment analyser les soins suboptimaux : l'exemple de la prise en charge initiale du sepsis sévère. *Colloque du Réseau Mère-Enfant de la Francophonie*, Paris, juin 2014.

Elise Launay, Jean-François Brasme, Catherine Deneux-Tharaux, Dominique Gendrel, Jérémie F Cohen, Martin Chalumeau. Comment analyser les délais diagnostiques en pédiatrie. *Journées Parisiennes de Pédiatrie*, Paris, octobre 2013.

Elise Launay, Elise Martin, Thomas Blanchais, Rémy Assathiany, Alain Martinot, Martin Chalumeau, Christèle Gras-Le Guen. Evitabilité des infections bactériennes sévères de l'enfant. *Journées Parisiennes de Pédiatrie*, Paris, octobre 2011.

**Elise Launay, Michele Morfouace, Jean-François Brasme, Christèle Gras-Le Guen, Martin Chalumeau. Analyse des délais diagnostiques en pédiatrie : pièges méthodologiques. *Congrès de la Société Française de Pédiatrie*, Marseille, mai 2011. *Archives de Pédiatrie* 2011; 18:75-6

**Elise Launay, Christèle Gras-le Guen, Alain Martinot, Marie-Hélène Bouvier-Colle, Martin Chalumeau, Méthodologie de l'analyse des soins suboptimaux. *Congrès de la Société Française de Pédiatrie*, Toulouse, mai 2009. *Archives de Pédiatrie*. 2009; 16:646-8

**Elise Launay, Alain Martinot, Rémi Assathiany, Gérard Bréart, Martin Chalumeau, Christèle Gras-le Guen. Soins sub-optimaux dans la prise en charge initiale des infections bactériennes sévères. *Congrès de la Société Française de Pédiatrie*, Toulouse, mai 2009. *Archives de Pédiatrie*. 2009; 16:649-51

GLOSSAIRES

Note : la lecture de ces glossaires n'est pas indispensable à la compréhension de la thèse mais peut permettre au lecteur de retrouver rapidement la définition d'un terme, soit dans le glossaire d'épidémiologie, soit dans celui d'infectiologie. Les termes sont présentés par ordre alphabétique et identifiés dans la thèse par un astérisque lors de leur première apparition.

1. Glossaire d'épidémiologie

(Référence principale : J. Bouyer. Epidémiologie : Principes et méthodes quantitatives. Mars 2009)

Ajustement

Procédures statistiques - par exemple, analyses multivariées - permettant de prendre en compte l'effet de co-variables lors de l'étude de la relation entre un événement et une exposition d'intérêt.

Applicabilité

Cf. validité externe

Biais

Un biais est une erreur systématique responsable d'une déviation entre la valeur estimée d'un paramètre et sa valeur réelle, par exemple l'incidence d'une maladie ou la mesure d'une association entre deux variables. Cette déviation peut se faire dans les 2 sens : sous-évaluation ou sur-évaluation de la valeur réelle. Les biais peuvent également varier dans leur ampleur : certains biais ont peu d'impact, d'autres peuvent totalement fausser les résultats et expliquer entièrement le phénomène observé. Les biais affectent la validité interne* de l'étude. L'évaluation et la prise en compte des biais, que ce soit au moment de la planification de l'étude, de l'analyse ou de l'interprétation des résultats, ont donc une importance majeure. Il existe plusieurs types de biais.

Biais d'attrition

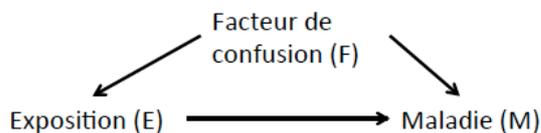
Le biais d'attrition est le biais induit par l'exclusion de patients au cours de l'étude ou du fait de patients perdus de vue. Cette exclusion est susceptible d'introduire un biais lorsqu'elle ne se fait pas au hasard mais de façon systématique selon certaines caractéristiques des sujets pouvant être associées au paramètre estimé.

Biais de classement

Ce biais est également appelé biais d'information ou biais de mesure. Un biais de classement est une erreur systématique dans la définition ou la mesure d'une exposition ou d'un état de santé. Il peut amener à classer à tort un sujet sain malade ou un sujet non-exposé exposé et inversement. Il peut résulter de la méthode de mesure notamment lorsque celle-ci varie de façon systématique selon les sujets par exemple lorsque la mesure d'une exposition (de l'optimalité d'un soin par exemple) varie selon l'issue pour le patient (survie ou décès par exemple).

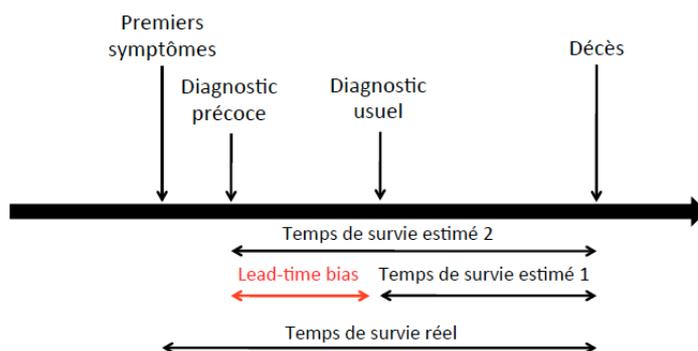
Biais lié à des facteurs de confusion

Il s'agit d'un biais provoqué par un tiers facteur modifiant la relation entre le facteur de risque étudié et la maladie. Un facteur F joue le rôle de facteur de confusion dans la relation entre E et M si dans la population étudiée : F est lié à E et à M ; F n'est pas une conséquence de E ; et la force de l'association entre E et M est différente selon que l'analyse est conduite de façon brute ou ajustée sur le facteur F. Ces relations peuvent être schématisées de la façon suivante dans le contexte de la recherche étiologique :



Biais de diagnostic précoce (Lead time bias)

Ce biais affecte les études évaluant le délai de survenue d'un événement de santé (durée de survie le plus souvent) et amène à conclure à tort que la durée de survie est plus longue chez les sujets ayant été diagnostiqués plus tôt (temps de survie 2 artificiellement supérieur au temps de survie 1). Le biais de diagnostic précoce peut être schématisé de la façon suivante :



Biais de sélection

Il existe un biais de sélection lorsque le processus de participation des sujets à l'étude entraîne une erreur systématique dans l'estimation des paramètres étudiés. Ce biais peut survenir lors du recrutement des sujets (recrutement dans un centre de référence d'une maladie rare par exemple), on parle alors de biais de recrutement, la population source* n'est alors pas représentative de la population cible*.

Chi-deux (χ^2) de tendance (test de)

Le test du χ^2 permet de tester l'indépendance entre deux variables qualitatives X et Y à n_x et n_y modalités respectivement. Le test de χ^2 de tendance permet de tester également la linéarité de l'association entre 2 variables quantitatives X et Y à 2 et n_y modalités respectivement. L'hypothèse nulle est comme pour le χ^2 que les valeurs de A et de B sont indépendantes alors que l'hypothèse H1 est que les valeurs de X augmentent ou diminuent de façon linéaire selon les modalités de Y.

Coefficient kappa de Cohen

C'est un test statistique non paramétrique utilisé pour évaluer la fiabilité de l'évaluation d'une variable catégorielle. Il s'agit d'une mesure de l'intensité de l'accord entre des jugements qualitatifs appariés, corrigée de l'effet du hasard. Pour ce faire, la statistique du Kappa permet de calculer une différence relative entre la proportion d'accord observée et la proportion d'accord attendue sous l'hypothèse aléatoire. Une échelle d'interprétation de la valeur obtenue de la statistique du test a été proposée par Landis et Koch.

Valeurs du KAPPA	Degré de concordance (ou d'agrément)
0	Dû au seul hasard
entre 0,01 et 0,19	Médiocre
entre 0,20 et 0,40	Faible
entre 0,41 et 0,60	Modéré
entre 0,61 et 0,80	Bon
entre 0,81 et 0,99	Très bon
1,0	Parfait

CONSORT

Cf. reporting guidelines

Delphi

La méthode Delphi est une méthode d'obtention de consensus par un groupe d'experts. Elle consiste à solliciter un groupe d'experts à qui l'on pose individuellement et de façon anonyme une ou plusieurs questions au cours de plusieurs « tours » successifs jusqu'à obtention éventuelle d'un consensus. A chaque tour (à partir du 2^{ème}), la ou les questions (éventuellement modifiées au vue des réponses au tour précédent) sont à nouveau posées avec la synthèse des réponses du tour précédent.

Délai diagnostique (délai patient, délai médical)

C'est l'intervalle de temps qui sépare la date d'apparition de l'un ou plusieurs des symptômes pouvant théoriquement conduire au diagnostic de la maladie de celle du diagnostic effectif. Cet intervalle peut se découper en plusieurs « sous-délais » notamment le délai patient correspondant à l'intervalle de temps entre la date d'apparition des premiers symptômes de la maladie et la date de la première consultation et le délai médical séparant la date de la première consultation et celle du diagnostic. Le délai médical peut également se découper en plusieurs intervalles notamment le délai pré-hospitalier et le délai hospitalier.

Etude primaire

Une étude primaire est une étude originale générant des résultats inédits par opposition aux revues de la littérature, et notamment les revues systématiques, qui visent à synthétiser les résultats d'études primaires.

EQUATOR

Cf. reporting guideline

Facteur intermédiaire

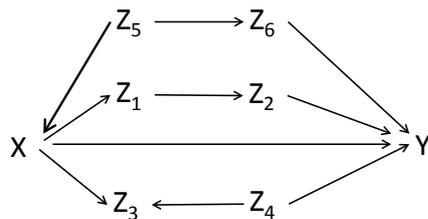
Il s'agit d'un facteur agissant sur la chaîne causale (par exemple, entre une exposition d'intérêt et la maladie étudiée). Par exemple, si les soins reçus à la phase initiale du sepsis (soins initiaux) conditionnent l'admission en

réanimation et que les soins reçus en réanimation conditionnent (en partie) le décès, alors la variable « soins reçus en réanimation » est un facteur intermédiaire sur la chaîne causale entre les soins initiaux et le décès.

Grphe acyclique orienté (Directed Acyclic Graph-DAG-)

C'est un mode de représentation des liens de causalité supposés entre des variables notamment entre une exposition d'intérêt (X) et l'événement d'intérêt (Y). Il est utilisé pour comprendre les phénomènes de biais de confusion et de sélection, sélectionner les bonnes variables d'ajustement ou encore comprendre les analyses d'effet direct et la prise en compte de variables non mesurées. Ce graphe obéit aux règles de construction suivantes : les nœuds du diagramme représentent des variables ou des événements, les flèches orientées représentent un lien causal entre 2 variables, le graphe est acyclique (aucun chemin orienté ne forme de boucle) avec des flèches ordonnées chronologiquement et pointant dans le sens « cause » (ancêtre) vers « conséquence » (descendant). La construction d'un tel modèle suppose d'avoir des hypothèses *a priori* sur les liens causaux entre les variables.

Exemple de graphe acyclique orienté : les variables Z_1 et Z_2 sont des variables intermédiaires (se situant sur le chemin entre X et Y), la variable Z_3 est un *collider* (variable affectée par X et par Z), la variable Z_5 est un facteur de confusion. Pour évaluer l'effet direct de X sur Y, il ne faut pas ajuster sur Z_3 (car c'est un *collider*) ni sur les variables intermédiaires Z_1 et Z_2 , il faut conditionner (ajuster) sur Z_5 car c'est une cause commune de X et de Y.



Imprécision

C'est une erreur non systématique liée à des variations aléatoires d'échantillonnage. L'imprécision traduit le fait que si l'on répétait une même mesure ou estimation dans des échantillons différents, la valeur estimée varierait autour de la valeur réelle. L'imprécision peut être évaluée par l'intervalle de confiance d'une estimation.

Modèle multiniveau

Un modèle multiniveau (auss appelé modèle hiérarchique, ou modèle mixte) a pour but de prendre en compte le fait que les données ont une structure hiérarchique. C'est par exemple le cas lorsqu'on étudie la qualité des soins reçus par des individus, on peut considérer plusieurs niveaux : un premier niveau correspondant à chaque soin et un deuxième correspondant à chaque individu recevant les soins. On peut en effet s'attendre à ce qu'il existe plusieurs niveaux de variabilité dans les mesures observées et dans les associations estimées : variabilité entre les soins pour un individu donné car il peut exister un certain degré de corrélation entre les soins chez un même patient et variabilité entre les individus.

Ce modèle s'écrit de la façon suivante dans l'exemple logistique et pour un modèle à intercept aléatoire

Niveau 1 (soins i)

$$\log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \beta_{0j} + \sum_{q=1}^Q \beta_{qj} X_{qij}$$

Niveau 2 (individus j)

$$\beta_{qj} = \gamma_{q0} + \sum_{s=1}^{S_q} \gamma_{qs} W_{sj} + u_{qj}$$

Où :

β_{0j} est l'intercept pouvant se décomposer en une partie fixe γ_{00} et une partie dépendante du niveau 2 u_{0j}

β_{qj} est le coefficient q de niveau 1 spécifique de la j^{ème} unité

X_{qij} est le prédicteur q de niveau 1 pour le cas i au sein de l'unité j

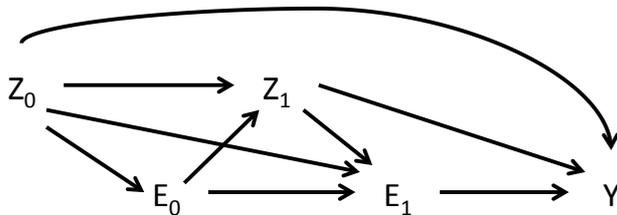
γ_{qs} est le coefficient de niveau 2

W_{sj} est le prédicteur de niveau 2

u_{qj} est l'effet aléatoire de niveau 2

Modèle structural marginal

Ces modèles sont utilisés dans les études longitudinales lorsque les facteurs de confusion impliqués dans l'association entre une exposition et un événement de santé sont dépendants du temps notamment lorsqu'une covariable (Z_1) prédit à la fois l'événement d'intérêt (Y) et l'exposition ultérieure (E_1) et l'exposition passée (E_0) prédit cette covariable (L_1) (cf. schéma ci-dessous). Un simple ajustement sur Z peut alors biaiser la mesure d'association. Le modèle structural marginal consiste à étudier la relation entre E et Y , en affectant à chaque individu i , et pour chaque temps t , un poids W_{it} égal à l'inverse de la probabilité de recevoir son propre niveau d'exposition conditionnellement aux valeurs observées pour cet individu pour les covariables L (jusqu'au temps t).



Polynôme fractionnaire

C'est une méthode mathématique permettant de mieux modéliser l'association entre deux variables continues X et Y dont au moins une est continue et lorsque cette association s'écarte de la linéarité (par exemple, lorsqu'elle suit une courbe en U ou en S). En ajoutant des puissances successives de X , où les exposants peuvent être négatifs et/ou non entiers, on obtient une fonction permettant de représenter la plupart des courbes observées. Par exemple, on étudie la relation entre âge au diagnostic (variable X , en jours) et délai diagnostique (variable Y , en jours). Si cette relation a une forme en S, elle sera probablement mieux décrite par un polynôme fractionnaire de type $Y = \alpha + \beta_1 X^n + \beta_2 X^p$ que par une relation linéaire de type $Y = \alpha + \beta X$.

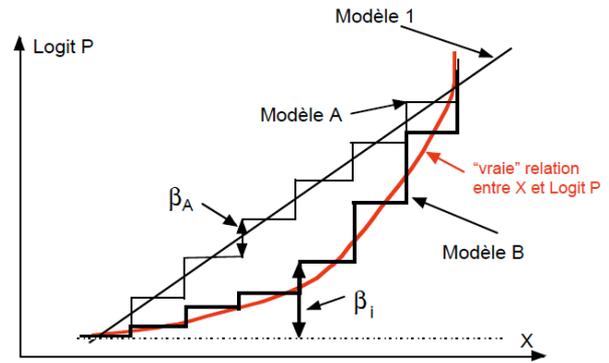
Pour tester si une association s'écarte de la linéarité dans un modèle de régression logistique*, il faut comparer par rapport des vraisemblances :

- un modèle A où X (par exemple l'âge) est transformé en une variable discrète X' à k classes codées de 0 à k-1 (par exemple, une classe par année) et qui s'écrit :

$$\text{Logit } P = \alpha_A + \beta_A X'$$

- et un modèle avec (k-1) variables X'_i indicatrices des classes de X' (X'_i = 1 si X = i et X'_i = 0 sinon) qui s'écrit :

$$\text{Logit } P = \alpha'_A + \sum_{i=1}^{k-1} \beta_i X'_i$$



Les modèles A et B sont identiques si $\beta_i = i \times \beta_A$

Population d'étude

C'est l'ensemble des sujets participant à l'étude (sujets inclus).

Population cible

La population cible d'une étude est la population sur laquelle on souhaiterait généraliser les résultats de l'étude. Elle est le plus souvent (très) théorique.

Population source

La population source est la population parmi laquelle ont été sélectionnés les sujets étudiés.

PRISMA

Cf. reporting guidelines

Régression logistique

La régression logistique est un modèle mathématique qui permet d'exprimer la relation entre plusieurs variables et une variable binaire, par exemple la présence ou l'absence de maladie. Les associations sont estimées sous forme d'odds ratio ajustés (ORA), ce qui permet d'identifier les prédicteurs associés à la variable prédite « indépendamment » des autres prédicteurs (c'est-à-dire, tout autre paramètre ayant été fixé). Sur le plan mathématique, on peut l'écrire ainsi :

$$P(M^+ | X_1, \dots, X_k) = \frac{1}{1 + \exp\left(-\left(\alpha + \sum_{i=1}^k \beta_i X_i\right)\right)} \quad \text{ou} \quad \text{Logit } P = \alpha + \sum_{i=1}^k \beta_i X_i$$

Reporting/Reporting guidelines

Le *reporting* est un terme anglais dont il n'existe pas d'équivalent synthétique en français qui désigne « la façon dont une étude est rapportée », c'est-à-dire comment les informations concernant la méthodologie, les résultats et l'interprétation d'une étude sont transcrites dans un article scientifique.

Les *reporting guidelines* sont des recommandations concernant la façon dont les études doivent être rapportées et les informations essentielles que le rapport de l'étude doit contenir. Le but d'une *reporting guideline* est que les articles rapportent de façon appropriée les informations essentielles à l'évaluation de la qualité de l'étude et donc de la pertinence des résultats. Il existe différentes *reporting guidelines* selon le type d'étude. Ces *reporting guidelines* ont été listées par le réseau EQUATOR (*Enhancing the QUAlity and Transparency Of Research*). Ce réseau édite des recommandations pour la construction de *reporting guidelines*, enregistre les *reporting guidelines* existantes et les diffuse. La première *reporting guideline* était CONSORT (*CONsolidated Standards Of reporting Randomized Trial*) pour les essais randomisés contrôlés. Il existe notamment des *reporting guidelines* pour les études observationnelles (*STrengthening the Reporting of OBservational studies in Epidemiology*, STROBE), pour les études évaluant les performances diagnostiques d'un test (*STAndards for Reporting of Diagnostic accuracy*, STARD) ou encore pour les revues systématiques (*Preferred Reporting Items for Systematic reviews and Meta-Analyses*, PRISMA).

Revue systématique

Une revue systématique est une démarche de revue exhaustive et de synthèse critique de la littérature. L'auteur tente de regrouper toute l'information scientifique disponible sur un sujet, afin d'éviter les biais de sélection inhérents à une revue non systématique ou « narrative ». La revue systématique doit suivre une méthodologie rigoureuse afin d'assurer son exhaustivité, son impartialité et sa reproductibilité. Dans une revue systématique, les points clés sont la méthode de recherche bibliographique et l'appréciation de la qualité des études notamment l'évaluation des risques de biais (lecture critique standardisée). La revue systématique peut être suivie d'une méta-analyse, dont elle est distincte.

Revue systématique méthodologique

Une revue systématique méthodologique est une catégorie particulière de revue systématique dont le but n'est pas de faire la synthèse des résultats mais de certaines composantes de la méthodologie comme la technique de randomisation pour les essais randomisés par exemple.

STARD

Cf. *reporting guideline*

STROBE

Cf. *reporting guideline*

Spin

C'est un concept émergent dans le domaine de la méthodologie de la recherche qui traduit le fait de rapporter ou de discuter les résultats d'une étude de façon inappropriée pour mettre l'emphase sur certains résultats, par exemple sur les effets bénéfiques du traitement ou de l'intervention étudiée.

Transportabilité

Cf. validité externe

Validité externe

La validité externe est l'ensemble des éléments d'une étude qui permettent de s'assurer que les résultats sont généralisables à l'ensemble de la population cible. On parle parfois de « généralisabilité » de l'étude, ce qui est un néologisme peu heureux phonétiquement mais pratique. La notion de validité externe implique les notions de transportabilité et d'applicabilité mais cette dernière est souvent utilisée pour résumer à elle seule la notion de validité externe. Certains auteurs utilisent également le terme de « facteurs de variations » pour évoquer les éléments nécessaires à l'évaluation de la validité externe.

La transportabilité d'une étude est le fait de pouvoir transposer les résultats à l'ensemble de la population cible. Pour évaluer la transportabilité des résultats d'une étude, il faut s'assurer que la prévalence de la maladie étudiée, les caractéristiques de la population de l'étude et du système de soins dans lequel s'est déroulée l'étude soient comparables à celles de la population cible.

L'applicabilité d'une étude est le fait de pouvoir appliquer les résultats d'une étude à un contexte particulier. Pour une étude évaluant une intervention, l'applicabilité consiste à s'assurer par exemple que l'intervention est facilement acceptable, que les ressources sont disponibles pour mettre en place l'intervention ou encore que l'intervention est suffisamment spécifique pour ne pas l'imposer à des sujets qui n'en tireraient aucun bénéfice.

Validité interne

La validité interne est l'ensemble des éléments d'une étude qui permettent de s'assurer que les résultats sont fiables et non biaisés. Évaluer la validité interne d'une étude, c'est évaluer les risques que des biais ou des facteurs de confusion aient faussés ses résultats.

2. Glossaire d'infectiologie

Choc septique

Selon la définition communément admise de la *Surviving sepsis Campaign*, un choc septique est un sepsis* associé à une défaillance hémodynamique persistante après le remplissage vasculaire initial (40 ml/kg en moins de 60 minutes). La défaillance hémodynamique étant définie par une pression artérielle systolique inférieure au 5^{ème} percentile pour l'âge, la nécessité d'administration de médicaments vasoactifs pour maintenir une pression artérielle systolique normale pour l'âge ou l'administration de dopamine à plus de 5 µg/kg/min ou de dobutamine, d'adrénaline ou de noradrénaline quelles que soient les doses ou 2 des critères suivants :

- acidose métabolique non expliquée (excès de bases de plus de 5 mEq/L),
- taux de lactate artériel supérieur à 2 fois la valeur normale,
- débit urinaire < 0,5 ml/kg/h,
- temps de recoloration capillaire supérieur à 5 secondes,
- différence entre températures centrale et périphérique de plus de 3°C.

Infection bactérienne sévère

Le terme d'infection bactérienne sévère (traduction de *serious bacterial infection*) désigne habituellement les infections bactériennes suivantes susceptibles d'engager le pronostic vital à des degrés variables : bactériémie, méningite et autres infections neuro-méningées, pneumonie (et pleuropneumonie), pyélonéphrite, ostéomyélite, arthrite, dermo-hypodermite, fasciite et péritonite. Dans ce travail doctoral nous restreindrons la définition des infections bactériennes sévères à celle très pragmatique d'infection bactérienne nécessitant une admission en réanimation. Le terme sévère est un anglicisme couramment employé qui peut être remplacé par grave.

Infection communautaire

Le terme d'infection communautaire s'oppose à celui d'infection liée aux soins (ou nosocomiale). Il désigne une infection qui a été acquise dans la communauté, c'est-à-dire en dehors d'une structure de soins.

Purpura fulminans

C'est une forme de choc septique associé à un purpura extensif et à une coagulation intravasculaire disséminée, il est en très grande majorité lié au méningocoque.

Sepsis

Le sepsis est défini par la présence d'une infection probable ou documentée et d'un syndrome de réponse inflammatoire systémique (*systemic inflammatory response syndrome –SIRS-*)

Sepsis sévère

Le sepsis sévère est défini comme un sepsis associé à une défaillance hémodynamique, un syndrome de détresse respiratoire aiguë ou à deux autres défaillances d'organe.

Systemic inflammatory response syndrome (SIRS)

Le SIRS est défini par la présence d'au moins 2 des 4 critères suivants, dont une anomalie de la température ou de la numération globulaire :

- température $\geq 38,5^{\circ}\text{C}$ ou $< 36^{\circ}\text{C}$,
- tachycardie ou bradycardie,
- fréquence respiratoire supérieure à + 2 écart-types pour l'âge ou nécessité de ventilation mécanique en dehors d'un contexte de maladie neuromusculaire connue ou d'anesthésie générale,
- numération leucocytaire anormale pour l'âge ou présence de plus de 10% de neutrophiles immatures (formes jeunes).

INTRODUCTION GENERALE

1. Etat des connaissances et justification du travail doctoral

1.1 Terminologie des soins suboptimaux

Les industries « à risque » comme l'aéronautique ou le nucléaire civil ont développé depuis des décennies des méthodes pour évaluer la sécurité de leurs activités et réduire les accidents dont les conséquences sont ou seraient souvent catastrophiques en termes de vies humaines. La médecine est une discipline où les défaillances peuvent aussi aboutir à des catastrophes (en termes de décès ou de morbidité) comme dans les industries « à risque ». La terminologie utilisée en médecine pour évoquer les défaillances dans la prise en charge des patients est variable selon les études : erreurs,(1,2) effets ou évènements indésirables,(1,2) soins suboptimaux (3) ou encore dommages évitables.(2) En 1999, l'*Institute of Medicine* américain publiait un rapport sur les erreurs médicales intitulé « *To Err is Human* ».(1) Dans ce rapport, une erreur était définie par le fait qu'une action planifiée n'ait pu être exécutée comme prévu (erreur d'exécution) ou comme le fait qu'une action ait été planifiée de façon erronée (erreur de planification). Nous pourrions également ajouter l'absence de planification d'une action comme source d'erreur. Le terme d'erreur peut cependant être culpabilisant et le terme d'erreur médicale réduit le système de soin au seul « médecin » occultant ainsi d'autres défaillances comme celles liées à la mauvaise interprétation des symptômes par le patient ou son entourage ou encore à l'accessibilité aux soins. Le terme de soin suboptimal, utilisé dans les enquêtes confidentielles avec comité d'experts sur la mortalité maternelle par exemple,(3) nous a semblé plus adapté à l'étude des défaillances d'un système de soin car il sous-entend également la notion de soin standard (soin optimal) sans se limiter au seul aspect médical du système de soin et sans connotation inutilement culpabilisante.

1.2 Soins suboptimaux et mortalité évitable

La plupart des travaux sur les soins suboptimaux sont sous-tendus par l'hypothèse que ces soins suboptimaux ont des conséquences néfastes et qu'y remédier pourrait alors contribuer à l'amélioration de la survie. Ainsi, le rapport de l'*Institute of Medicine* américain faisait état de 44 à 98 000 décès par an imputables aux erreurs médicales aux USA, plaçant les erreurs médicales comme la 8^{ème} cause de décès.(1) Dans une des études princeps de ce rapport publiée en 1991 par Leape et coll., il apparaissait que 70% des soins suboptimaux étaient jugés a posteriori évitables.(4) Une estimation plus récente publiée en 2013 par James et coll. faisait état de 400 000 décès prématurés associés aux soins suboptimaux chez des patients hospitalisés aux USA.(2) Une étude de synthèse sur 10 ans d'enquêtes confidentielles (1998 à 2007) rapportait que la moitié des décès maternels étaient jugés comme évitables.(5) Si des facteurs d'évitabilité peuvent être repérés dans la prise en charge, établir le lien causal entre ces facteurs et le décès est cependant difficile à démontrer car les études sont toujours observationnelles pour des raisons éthiques évidentes et donc soumises à un fort risque de biais*. Il a d'ailleurs été montré dans une étude publiée en 2001 que si les décès étaient souvent jugés évitables par des experts du fait de soins suboptimaux (pour un quart des 111 dossiers de décès intrahospitaliers d'adultes toutes causes confondues et étudiés rétrospectivement), la survie estimée par ces mêmes experts en cas de soins optimaux ne dépassait 3 mois que pour 0,5% des cas.(6) Si les soins suboptimaux peuvent donc être évitables, l'évitabilité du décès est quant à elle moins évidente à démontrer.

1.3 Les soins suboptimaux et les infections bactériennes sévères de l'enfant

La thématique des infections bactériennes sévères* (IBS) de l'enfant est un bon candidat à l'étude des soins suboptimaux tout d'abord parce que les IBS ont une morbi-mortalité non

* Les mots suivis d'un astérisque lors de leur première apparition sont explicités dans le glossaire

négligeable dans les pays occidentaux. En effet, selon les données du Centre d'Epidémiologie sur les Causes médicales de Décès (CépiDc) collectées d'après les certificats de décès, les infections bactériennes figurent parmi les 5 premières causes de mortalité chez les enfants de moins d'un an notamment dans la période périnatale.(7) Dans une étude américaine analysant les conclusions d'autopsies pédiatriques réalisées sur une période de 24 ans, un sepsis* bactérien était considéré comme responsable du décès à l'issue de l'autopsie dans 26% des cas.(8) Si la mortalité liée au sepsis sévère* a diminué notamment grâce à l'impact de certaines vaccinations comme la vaccination anti-*Haemophilus b*, l'incidence du sepsis sévère augmente.(9) Ainsi, Hartman et coll. ont mis en évidence, dans une population constituée de 7 Etats américains, une augmentation de l'incidence du sepsis sévère de 0,56 à 0,89 pour 1,000 enfants entre 1995 et 2005 avec un taux de mortalité de passant de 10,3 à 8,9%.(10) La prise en charge du sepsis reste donc fréquente en réanimation pédiatrique même si elle peut paraître rare dans l'exercice d'un médecin de soins primaires.(11) En effet les IBS sont des événements peu fréquents en comparaison aux infections virales bénignes de l'enfance dont elles partagent les symptômes initiaux et notamment la fièvre.(12) La rareté des IBS à l'échelle d'un médecin et le manque de spécificité des symptômes initiaux d'IBS en pédiatrie,(13) constituent un risque d'errance diagnostique et thérapeutique. Nous pouvons alors nous demander s'il existe un lien entre d'une part une incidence et une mortalité du sepsis en réanimation encore élevées malgré l'existence de thérapeutiques spécifiques (notamment l'antibiothérapie) et d'autre part les soins prodigués lors de la prise en charge initiale qui sont, comme nous venons de le voir, à risque de suboptimalité. L'étude d'un tel lien demande au préalable de pouvoir détecter les soins suboptimaux de façon adéquate.

1.4 Méthodologie d'évaluation des soins suboptimaux : de la clinique à la recherche

Prenons l'exemple de la prise en charge du patient décrite dans la **Figure 1** (page 44). Plusieurs éléments de sa prise en charge comme le délai d'arrivée du médecin, l'absence d'administration immédiate d'antibiotique ou le remplissage vasculaire insuffisant peuvent paraître suboptimaux. Cette prise en charge pourrait faire l'objet d'une revue de mortalité morbidité (RMM). La RMM a pour objectif l'analyse, le plus souvent au sein d'un service, de tous les décès et de certains accidents morbides prédéterminés ou non. L'analyse se fait avec l'ensemble des médecins du service et éventuellement ceux d'autres équipes extra-hospitalières et des membres du personnel non médical. Le groupe de professionnel doit rechercher des actions d'amélioration de la qualité des soins afin d'éviter la récurrence de l'évènement indésirable. La RMM est cependant difficilement utilisable en recherche clinique car ses résultats ne sont théoriquement pas généralisables à un autre service ou à un autre contexte de soin. (14)

Certains éléments de la prise en charge présentée en **Figure 1**, comme l'absence d'administration immédiate d'antibiotique devant un purpura ou le remplissage vasculaire insuffisant, pourrait faire l'objet d'une évaluation par audit car il existe des référentiels opposables contemporains : un avis du Conseil Supérieur d'Hygiène Publique et les recommandations de la *Surviving Sepsis Campaign* respectivement.(15,16) L'audit clinique est en effet une méthode d'évaluation des pratiques utilisant la comparaison à des références admises. Sa principale caractéristique est de mesurer les écarts entre la pratique observée et la pratique attendue.(17) L'audit clinique permet d'évaluer la prise en charge globale d'un patient, un thème transversal ou encore le fonctionnement d'un service. Lorsque l'audit prend un caractère national, il se rapproche d'une démarche scientifique, sa principale limite étant alors la nécessité de références admises pour les soins qui peuvent faire défaut pour des situations complexes.

D'autres éléments de la prise en charge présentée en **Figure 1** peuvent également paraître suboptimaux comme le délai trop long entre les premiers symptômes et la première consultation, cependant, il n'existe pas de référentiel concernant le délai de consultation devant une fièvre de l'enfant. L'expérience du clinicien peut alors être sollicitée pour pallier le manque de référence consensuelle lors de l'évaluation de l'optimalité des soins. L'enquête confidentielle avec comité d'experts est basée sur ce principe d'évaluation. Elle consiste en l'analyse de l'optimalité des soins par un groupe de spécialistes (comité d'experts) extérieurs à l'évènement mais concernés par la question (revue par les pairs). Cette méthode a été développée historiquement dans le domaine de l'obstétrique et notamment des décès maternels.(3) Utilisée à l'échelle d'une population, cette méthode peut permettre la détection de dysfonctionnements systématisés qui pourront faire l'objet d'actions correctrices structurées. Dans l'exemple des IBS, la méthodologie de l'enquête confidentielle pourrait permettre de répondre aux questions suivantes : les soins suboptimaux détectés dans la prise en charge de ce nourrisson sont-ils fréquents à l'échelle d'une population ? Quelles sont les variables associées à ces soins suboptimaux (déterminants) ? Et quelles sont les conséquences de ces soins suboptimaux en terme de survie ? Les réponses à ces questions pourraient justifier et guider la mise en place d'actions correctrices ciblées sur la typologie et les déterminants des soins suboptimaux.

1.5 Méthodologie d'évaluation de l'optimalité des soins : exemple du délai diagnostique

Plusieurs difficultés méthodologiques peuvent s'opposer à la fiabilité et à la transportabilité* des résultats d'une étude visant à évaluer les soins suboptimaux non plus à l'échelle individuelle ou d'un service mais dans un but de recherche. Nous avons cherché dans la littérature des éléments de réponse à ces problèmes méthodologiques en analysant quelques études primaires* évaluant les soins suboptimaux dans la prise en charge d'IBS (cf. partie II –

Introduction, page 123). Nous avons alors constaté des limites dans ces études notamment pour la sélection de la population, la définition et la mesure de l'optimalité des soins, ou encore l'étude de la relation entre les soins suboptimaux et la survenue du décès. Il nous a alors semblé pertinent de consacrer une partie de ce travail doctoral à la construction d'une *guideline* méthodologique pour évaluer les soins suboptimaux.

Il existe plusieurs types de soins suboptimaux (cf. classification proposée par l'*Institute of Medicine* américain **Figure 2**, page 44) pour lesquels le mode d'évaluation peut être différent. En effet, l'évaluation d'un défaut de prévention par absence de vaccination est différente de l'évaluation d'une non reconnaissance des signes précoces de sepsis par exemple. Nous avons donc décidé de focaliser notre réflexion méthodologique sur le retard diagnostique car il concerne différents éléments du système de soins à différents moments du parcours de soins traduisant bien la complexité de ce dernier. Un retard diagnostique peut ainsi être lié à une consultation trop tardive par le patient ou son entourage, à la désorganisation du système de soins et notamment à une mauvaise accessibilité aux soins, ou encore à une non reconnaissance ou une mauvaise interprétation des symptômes ou à un mésusage des tests diagnostiques par les médecins. Il est parfois difficile de dissocier le retard diagnostique du retard au traitement car si une maladie n'est pas diagnostiquée le traitement adapté ne peut être administré. Par exemple, si le diagnostic de sepsis sévère n'est pas fait, l'administration d'antibiotique et d'un support hémodynamique sera retardée. La notion de retard diagnostique est parfois très subjective car elle est toujours analysée *a posteriori* et elle représente en fait une interprétation d'un élément plus facilement objectivable : le délai diagnostique*.

Le délai diagnostique définit l'intervalle de temps séparant les premiers symptômes du diagnostic. Les auteurs anglo-saxons font souvent la différence entre le terme de « diagnosis delay » qui porte une connotation péjorative de retard et les termes plus objectifs de « *time to diagnosis* », « *lag time* », « *duration of symptoms* » ou encore « *pre-diagnosis symptomatic*

interval ». Dans la partie I de ce travail doctoral, nous nous sommes focalisés sur le délai diagnostique alors que nous avons plutôt évalué le retard diagnostique (ou thérapeutique) dans la partie II de la thèse. Nous verrons que ces deux évaluations se complètent. Une meilleure connaissance des particularités méthodologiques inhérentes à l'évaluation du délai diagnostique, et de façon plus générale à l'évaluation des soins, pourrait permettre d'éviter certains pièges méthodologiques sources de résultats biaisés et/ou non généralisables qui pourraient être eux-mêmes à l'origine de mises en place d'actions correctrices inadaptées.

2. Hypothèses

Notre hypothèse initiale était que les soins suboptimaux au cours de la prise en charge initiale des IBS de l'enfant étaient fréquents, potentiellement évitables et associés à une surmortalité et une surmorbidity et que l'identification de leur typologie standard et de leurs déterminants pourrait permettre la mise en place d'actions correctrices afin d'éviter certains décès ou des séquelles.

Dès les premières étapes de notre travail doctoral, nous avons rapidement fait le constat que l'évaluation des soins suboptimaux nécessitait une méthodologie complexe notamment pour l'analyse des délais diagnostiques. Ceci nous a amenés à notre seconde hypothèse qui était que la méthodologie d'évaluation des délais diagnostiques était souvent inadéquate, ce qui pouvait conduire à des résultats biaisés et donc des actions correctrices inutiles, voire contre-productives et qu'il était important de proposer des approches méthodologiques moins exposées aux biais.

Le travail doctoral repose donc sur deux hypothèses interconnectées. Nous avons choisi de présenter en premier les travaux essayant de répondre à notre deuxième hypothèse afin de progresser du cas général à l'exemple.

3. Objectifs et plan du travail doctoral

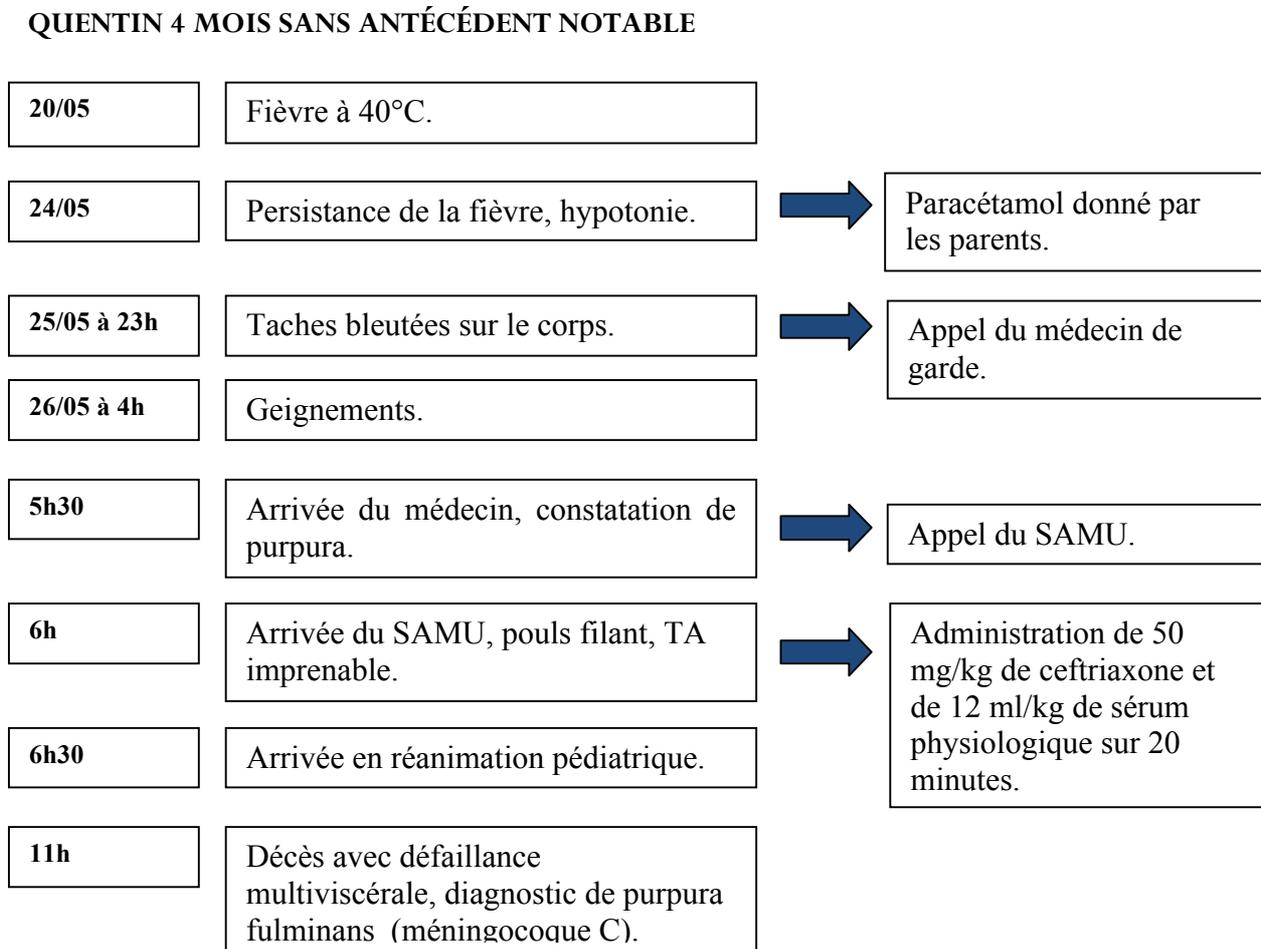
Dans la **Partie I** de cette thèse, pour tester l'hypothèse des difficultés méthodologiques rencontrées lors de l'évaluation de la fréquence, des déterminants et des conséquences des délais diagnostiques, nous avons eu pour objectif, dans le **chapitre 1**, d'évaluer systématiquement la qualité du *reporting** des études portant sur ces délais diagnostiques. Cette évaluation a nécessité la production d'une première *checklist* des éléments considérés comme critiques (car potentiellement liés à des risques de biais ou d'obstacles à la transportabilité) dans le *reporting* des études sur les délais diagnostiques. Le constat d'une qualité souvent insuffisante du *reporting* de ces points critiques nous a incités à transformer cette simple *checklist* en une véritable *reporting guideline** spécifique aux études sur les délais diagnostiques. L'objectif du **chapitre 2** a été de suivre les recommandations du réseau EQUATOR* (*Enhancing the Quality And Transparency Of Research*) pour la construction et la validation d'une *reporting guideline* et de confronter nos hypothèses concernant les risques de biais et d'obstacles à la transportabilité aux données de la littérature. Nous avons ainsi conduit une revue systématique méthodologique* des revues systématiques* sur les délais diagnostiques. En effet, une des étapes indispensables à la conduite d'une revue systématique est l'évaluation de la qualité et notamment des risques de biais et d'obstacles à la transportabilité des études primaires incluses. En analysant les outils utilisés par les auteurs de revue systématique pour évaluer ces risques nous voulions identifier les risques de biais et d'obstacles à la transportabilité reconnus par d'autres auteurs et les confronter à nos hypothèses afin de compléter notre *checklist* initiale. L'objectif du **chapitre 3** a été de finaliser le processus décrit par le réseau EQUATOR d'affinement et de validation de notre *reporting guideline* en la soumettant à deux comités d'experts successifs. Cette étape a permis aussi de commencer la diffusion de cette *reporting guideline* et nous a conduits à produire un document explicatif pour en améliorer l'impact sur les chercheurs.

Dans la **partie II** de cette thèse, pour tester notre hypothèse sur la fréquence, les effets délétères et l'évitabilité des soins suboptimaux au cours de la prise en charge initiale des IBS de l'enfant, nous avons eu comme objectif dans le **chapitre 1** d'évaluer la fréquence de ces soins suboptimaux en essayant d'éviter certains des écueils méthodologiques identifiés dans la partie I. Une enquête confidentielle en population avec comité d'experts a ainsi retrouvé une fréquence très élevée de soins suboptimaux dans la prise en charge initiale d'enfants décédés d'IBS sur deux départements français. Ce chapitre nous a aussi permis d'identifier une typologie des soins suboptimaux les plus fréquents. L'objectif du **chapitre 2** a été d'étudier les conséquences et les déterminants des soins suboptimaux en comparant un groupe d'enfants survivant à l'IBS au groupe des enfants décédés décrits dans le chapitre 1. Cette comparaison a nécessité d'analyser les soins suboptimaux non plus à l'échelle de la prise en charge globale d'un enfant mais à l'échelle de chaque soin. Cela nous a permis de démontrer une association entre la fréquence des soins suboptimaux et d'une part la survie des enfants et d'autre part certaines caractéristiques des patients ou du système de soins.

Avertissements

- 1. L'ordre chronologique des travaux de la thèse ne correspond pas à l'ordre dans lequel ils sont présentés.*
- 2. L'article publié correspondant au chapitre 1 de la partie II et portant sur la fréquence des soins suboptimaux dans la prise en charge initiale des enfants finalement décédés d'IBS s'inscrit dans la démarche scientifique de la thèse mais ne doit pas être comptabilisé dans la production scientifique de la thèse du fait de son antériorité à l'inscription en thèse. Cette première inscription était en effet initialement prévue plus tôt mais a été retardée par l'opportunité d'un fellowship de 2 ans en infectiologie pédiatrique au CHU Ste Justine de Montréal.*
- 3. Nous utiliserons malheureusement les anglicismes reporting, reporting guideline et checklist tout au long de ce travail doctoral car nous n'avons pas trouvé d'équivalents synthétiques en français.*

Figure 1. Cas clinique illustrant la problématique des soins suboptimaux et du délai diagnostique dans la prise en charge initiale des infections bactériennes sévères de l'enfant.



Abréviations : SAMU, service d'aide médicale urgente ; TA, tension artérielle

Figure 2 Classification des soins suboptimaux selon Leape et coll.(18)

<p>DIAGNOSTIQUE</p> <ul style="list-style-type: none"> Erreur ou retard diagnostique Défaut d'exécution d'un test diagnostique indiqué Utilisation de tests obsolètes Défaut d'interprétation d'un test ou de paramètres de surveillance <p>THERAPEUTIQUE</p> <ul style="list-style-type: none"> Erreur dans la performance d'une opération ; d'une procédure ou d'un test Erreur dans l'administration d'un traitement Erreur de posologie ou d'utilisation d'un médicament Retard évitable dans le traitement ou dans la réponse à un test anormal Administration de soins inappropriés (non indiqués) <p>PREVENTION</p> <ul style="list-style-type: none"> Absence de prescription d'un traitement prophylactique Surveillance ou suivi inadaptés d'un traitement <p>AUTRES</p> <ul style="list-style-type: none"> Défaut de communication Défaut d'équipement Autre défaillance du système de soins

**PARTIE I - Méthodologie d'évaluation des délais
diagnostiques**

Introduction

Caractéristiques des études sur les délais diagnostiques

Les délais diagnostiques sont définis comme nous venons de le voir dans l'introduction générale par l'intervalle de temps séparant la date d'apparition des premiers symptômes à celle du diagnostic. Le délai diagnostique implique donc à la fois les patients, les médecins, l'organisation du système de soins et les performances des procédures diagnostiques. La problématique des délais diagnostiques s'inscrit dans trois dimensions principales : médicale, psychologique et médico-légale. Un délai diagnostique long peut avoir des conséquences médicales sur la gravité initiale de la maladie, la lourdeur du traitement nécessaire et le pronostic pour le patient en termes de morbidité et de mortalité. Les délais diagnostiques peuvent également entraîner des répercussions psychologiques pour les patients et leur entourage mais aussi pour les médecins car la période de latence entre l'apparition des symptômes et le diagnostic peut être anxiogène et culpabilisante.(19) Les délais diagnostiques ont également des conséquences médico-légales puisque le retard diagnostique est une cause fréquente de poursuite judiciaire en médecine en particulier en pédiatrie où c'est la plus fréquente.(20,21) Ces différentes dimensions sont très probablement à l'origine de l'intérêt croissant retrouvé dans la littérature pour les études sur les délais diagnostiques objectivé par une augmentation du nombre absolu de publications référencées dans PubMed au cours des 15 dernières années et une augmentation relative de 19% (**Figure 3**, page 59).

Les études sur les délais diagnostiques peuvent avoir trois types d'objectifs non exclusifs : la mesure d'un délai diagnostique et de son évolution dans le temps, l'identification de déterminants de délais diagnostiques longs (retard diagnostique) et/ou la mesure de la relation entre les délais diagnostiques et le devenir des patients en terme de morbi-mortalité.(22,23) Les études sur les délais diagnostiques sont souvent menées par des cliniciens dans un but d'évaluation des pratiques pour conforter une impression de délais perçus comme trop longs. Ces études sont quasiment toujours observationnelles et basées sur une série de cas diagnostiqués. En

effet, si la randomisation peut s'envisager dans un but d'évaluation de stratégies de diagnostic précoce, elle est très rarement mise en œuvre en pratique en raison d'une faisabilité très faible, compte-tenu de la lourdeur de telles études. Un *design* observationnel prospectif de type cohorte peut être envisagé en utilisant des bases de données administratives dans lesquelles seraient renseignés tous les symptômes présentés à chaque recours médical pour une population donnée, ou pour un échantillon représentatif. Ces bases ont été utilisées pour étudier la nature des premiers symptômes des cancers de l'enfant.(24,25) Cependant le niveau de détail de ces bases ne permet pas d'obtenir l'heure de l'apparition des symptômes pour tous les recours ce qui limite leur usage pour les affections aiguës et la précision des symptômes codés peut être très aléatoire.

Points méthodologiques potentiellement critiques

A la lecture de certains articles rapportant des études observationnelles sur les délais diagnostiques, nous avons été interpellés par certains choix méthodologiques pouvant biaiser les résultats et menacer ainsi la validité interne* ou rendre difficile la transportabilité des résultats et compromettre alors la validité externe*. Nous avons également noté pour certains articles l'absence de discussion de ces limites méthodologiques et des interprétations erronées voire abusives de certains résultats.(26–35) En effet, l'expérience de notre équipe de recherche dans le domaine des délais diagnostiques et des soins suboptimaux nous a permis d'identifier des points méthodologiques nous semblant critiques en lien avec des risques de biais ou d'obstacles à la transportabilité.(5,19,22,23,36–38) Ces points critiques concernent les différentes étapes du parcours de soins entre les premiers symptômes et le diagnostic : la sélection de la population, la définition et la mesure du délai, l'étude éventuelle des relations entre le délai diagnostique et les caractéristiques des participants et l'interprétation des résultats dans l'objectif de la mise en place de plan d'action (**Figure 4**, page 59).

Nous allons donc détailler ces points critiques en nous appuyant sur des exemples issus d'une même étude publiée dans la plus grande revue pédiatrique (*Pediatrics*) en 2009 et dont l'objectif était de mesurer le délai diagnostique des accidents artériels ischémiques du nouveau-né et de l'enfant et d'identifier des caractéristiques des patients pouvant influencer ce délai.(33)

Sélection de la population : identification de la population éligible et mode de recrutement

Exemple (extrait de la section Méthodes):

« We identified retrospectively all children 0 to 18 years of age who were diagnosed as having a first episode of arterial ischemic stroke (AIS) during a 12.5-year period, from June 1, 1993, to January 30, 2006, at the Royal Children's Hospital (Melbourne, Australia). The Royal Children's Hospital is a tertiary pediatric referral center for the states of Victoria and Tasmania that treats 280 000 children each year. Cases were ascertained by using International Classification of Diseases (ICD) revisions 9 and 10 codes applied to discharge diagnoses (Table 1), on the basis of standardized methods recommended by the International Pediatric Stroke Study. These cases were cross-referenced with the institution's hematology database and stroke registry (which commenced in 2002) to ensure that no patients were missed. »

L'identification des cas à partir du diagnostic de sortie de patients hospitalisés peut générer un biais de sélection*. Premièrement, ces patients codés avec un diagnostic d'accident ischémique sont, par définition, ceux qui ont été diagnostiqués. Ils ont probablement des caractéristiques particulières par rapport à l'ensemble des patients avec un accident ischémique. En effet, ils sont potentiellement plus sévèrement atteints avec des symptômes plus bruyants et plus facilement identifiables et ayant des formes non fulminantes car non décédés avant l'admission à l'hôpital. De plus, l'organisation particulière des soins (avec la présence de spécialistes de la maladie concernée dans un centre de référence qui évoqueront plus rapidement le diagnostic) peut amener à raccourcir le délai diagnostique et à limiter la généralisabilité des résultats.

Un recrutement strictement hospitalier dans un centre de référence pourrait se justifier si l'organisation des soins dans la région et pour la période de l'étude était telle que tous les enfants avec le diagnostic d'accident vasculaire cérébral étaient systématiquement admis dans ce centre. Nous verrons dans la partie II que c'est la configuration qui prévalait pour les IBS dans la zone géographique étudiée. Dans le cas contraire, un recrutement multicentrique ou mieux en population permettrait de minimiser ce biais de recrutement. Une inclusion non systématique de

tous les patients diagnostiqués pourrait également entraîner un biais de sélection par inclusion sélective des patients pour lesquels le délai était jugé trop long. Il est donc important de préciser, comme dans l'exemple des accidents vasculaires que l'exhaustivité du recrutement a été vérifiée avec une triangulation des données. Pour sélectionner la population d'étude*, il est également important d'énoncer des critères diagnostiques précis permettant de bien définir la maladie d'intérêt, ce qui est parfois difficile.

Sélection de la population : critères diagnostiques et d'inclusion et d'exclusion

Exemple (extrait de la section Méthodes) :

« AIS was defined as (1) an acute neurologic deficit lasting \geq 24 hours and (2) parenchymal infarction on neuroimaging scans. Inclusion criteria included a first presentation of AIS in patients from the newborn period (defined as $<$ 28 days of age) up to 18 years of age, with radiologic confirmation of AIS. Patients with the following conditions were excluded: sinovenous thrombosis and hemorrhagic stroke subtypes, systemic vasculitides, birth before 36 weeks of gestation, other non-AIS disorders associated with focal deficits, and recurrent strokes. Subjects who did not have documentation of the times of clinical onset and neuroimaging were excluded. »

Si les auteurs énoncent ici les critères diagnostiques utilisés, aucune information n'est donnée concernant le caractère systématique ou non de la réalisation de la procédure diagnostique (imagerie cérébrale) devant le symptôme d'alerte (un déficit neurologique de plus de 24 heures) pendant la période d'étude. Il se peut donc que certains enfants avec un déficit neurologique n'aient pas été diagnostiqués du fait de l'absence de réalisation d'imagerie. De même, le manque de performance de la procédure diagnostique peut être responsable de cas non-diagnostiqués, il est donc important de justifier le choix de la procédure diagnostique. Enfin, il se peut que certaines présentations cliniques paucisymptomatiques n'alertent pas le clinicien et ne motivent donc pas la réalisation de la procédure diagnostique ou, au contraire et à nouveau, que des formes très sévères d'emblée conduisent à un décès avant qu'un diagnostic ne soit posé. L'impact de ces cas non-diagnostiqués sur la mesure du délai est difficilement anticipable dans les résultats, cependant il paraît essentiel de les prendre en compte pour leur interprétation prudente. Enfin, il peut être important d'identifier d'éventuels sous-groupes de patients avec un

risque inhérent de délai extrême. Dans cet exemple, les enfants avec des récives d'accidents artériels ischémiques, chez lesquels le délai diagnostique serait très probablement plus court du fait des antécédents, ont été exclus.

Définition et mesure du délai diagnostique bornes et mode d'évaluation du délai

Exemple (extrait de la section Méthodes) :

« The time of clinical onset was determined in a number of ways. The preferred method was to use the exact time of clinical onset (such as 9:00 AM), as documented in the medical chart by physicians or nurses. In the absence of such information, the clinical onset was extrapolated from other documented information. For example, if the physicians documented that “hemiparesis began 2 hours ago” at 11:00 AM, then stroke onset was estimated at 9:00 AM. If the symptoms were present at awakening, then the last time the patient was seen in normal condition was used. For the sedated population and for certain ICU patients, the time of clinical onset was determined from the detailed regular neurologic observations performed by ICU nursing staff members, which yielded a time of clinical onset within 15 minutes. In cases with discrepancies between the physicians and nurses, the order of preference was as follows: earliest precise time documented, neurologic observations charted by nurses, and information extrapolated (...) Focal neurologic deficits included focal neurologic signs and focal seizures. A focal sign was defined as a motor (hemiplegia, monoplegia, or ataxia), sensory (paresthesia or dysesthesia), visual (visual field deficit or eye deviation), or speech (dysarthria or dysphasia) deficit. (...) The time of neuroimaging confirming AIS was obtained from a time stamp recorded at the time of the procedure and might not have been equivalent to the time of initial neuroimaging, if a diagnosis of AIS was not made immediately after the first imaging study (for example, if the initial study was a computed tomographic [CT] study that did not reveal the AIS).»

Une définition et une mesure imprécise et/ou biaisée du délai diagnostique peuvent aboutir à des résultats approximatifs, peu reproductibles et difficilement transportables à une autre population.

Lorsque la définition du délai diagnostique ne peut se faire de façon totalement objective, comme dans l'exemple ci-dessus, par extraction de la date du début des symptômes et de celle du diagnostic renseignées dans une base de données ou d'un cahier d'observation *ad hoc*, ce délai peut être évalué par une ou plusieurs personnes grâce à une revue de dossiers. Plusieurs biais peuvent alors entraver la fiabilité des résultats notamment celui lié à l'absence d'aveugle vis-à-vis du devenir du patient. En effet, il existe un dogme fort dans la communauté médicale sur l'association entre délai diagnostique long et issue défavorable. Ainsi dans la définition de « *early diagnosis* », un des termes du Medical Subject Heading (MeSH) de la National Library of Medicine associé à la notion de délai diagnostique, il est mentionné que : « en général, un

diagnostic précoce améliore le pronostic »¹. Par ailleurs, il a été montré que la connaissance de l'issue d'une situation influençait l'évaluation de cette dernière.(39) Il y a donc un risque de surestimation du délai diagnostique si l'évaluation n'est pas à l'aveugle de l'issue. De la même façon certains *a priori* sur les caractéristiques des sujets de l'étude (par exemple l'âge, le genre, le niveau socio-économique) peuvent amener à un biais dans l'évaluation du délai, il est alors important que cette évaluation se fasse aussi à l'aveugle des caractéristiques du patient. Lorsque plusieurs personnes sont sollicitées pour évaluer le délai diagnostique, la non indépendance des experts pourrait avoir une influence sur leurs conclusions notamment s'il existe un lien hiérarchique entre eux. La spécialité des évaluateurs peut également avoir une influence sur l'estimation du délai. Ainsi des médecins spécialistes pourraient surestimer le délai diagnostique en considérant des symptômes peu spécifiques qu'ils associent à la maladie étudiée dont ils sont spécialistes comme des premiers signes de la maladie alors qu'un médecin n'ayant pas ce niveau d'expertise n'aurait pas retenu ces symptômes. Ce dernier point renforce également la nécessité de lister les premiers symptômes compatibles avec le diagnostic *a priori*.

Mesure du délai diagnostique : mode de traitement d'une variable quantitative

Exemple (extrait de la section Résultats) :

« *A median time of 24.8 hours (IQR: 10,2–67.0 hours) from clinical onset to radiologic confirmation of AIS was noted for the population of children. (...) Six (6.8%) of 88 patients were diagnosed as having AIS within 3 hours.* »

Dans cet exemple, les auteurs ont traité la variable « délai diagnostique » en variable continue.

C'est un choix important et pas uniquement pour des raisons statistiques. En effet, la dichotomisation d'emblée d'une variable par définition continue, comme le délai diagnostique, fait non seulement perdre de la puissance mais entraîne vers un glissement sémantique, source potentielle de *spin**. Ainsi, le groupe de patients avec des délais « plus longs » que le seuil choisi va devenir le groupe avec des délais « trop longs » ou des « retards diagnostiques » alors que les

¹ “Generally, early diagnosis improves prognosis and treatment outcome”

délais longs peuvent être le reflet de soins suboptimaux (retards diagnostiques) mais aussi de l'évolution naturelle plus lente et donc moins bruyante de la maladie.

Etude de l'association entre le délai diagnostique et les caractéristiques des patients

Exemple (extrait de la section Méthodes) :

« The associations between variables were analyzed by using Stata 9.2 (Stata, College Station, TX), with the 2-sample rank-sum (Mann-Whitney) test and Kruskal-Wallis test where appropriate. P values of <0.05 were considered statistically significant »

Dans cet exemple, les auteurs annoncent en introduction un objectif d'étude des facteurs influençant la date du diagnostic. Ils expliquent les tests statistiques utilisés pour évaluer ces associations mais ne donnent pas leur hypothèse sur la nature de ces facteurs, ni sur l'existence éventuelle de facteurs de confusion*. L'étude de l'association entre le délai diagnostique et certaines caractéristiques des malades intervenant en amont (« déterminants ») ou en aval (« conséquences ») de ce délai n'est pas obligatoire mais peut permettre de justifier la pertinence des actions correctrices et de les préparer. Des facteurs de confusion peuvent alors compromettre la fiabilité de la mesure des associations observées.

Préparation du plan d'action

Exemple (extrait de la section Discussion) :

« Not surprisingly, diagnosis took longer for neonates with AIS than for the childhood population. This is because localizing signs often are absent, which makes stroke onset difficult to determine accurately for neonates. Focal seizures can predominate over motor deficits. Stroke should be suspected for any neonate presenting with seizures, lethargy, or apnea, the latter 2 features being seen in neonatal stroke in the setting of encephalopathy. »

Ce dernier point concerne la mise en place du plan d'action qui n'appartient pas en soi à la méthodologie d'études des délais diagnostiques mais découle plutôt de l'interprétation des résultats. Cependant un plan d'action est très souvent évoqué dans les articles c'est pourquoi il nous semblait important de souligner quelques pièges dans l'interprétation des résultats. Dans cet exemple, les auteurs recommandent de suspecter un accident vasculaire cérébral néonatal devant une apnée alors qu'ils n'avaient pas identifié ce symptôme comme un symptôme d'alerte devant faire réaliser une imagerie cérébrale. Les auteurs n'ont d'ailleurs pas indiqué si le protocole local

en vigueur durant la période d'étude prévoyait que tous les enfants ayant une apnée aient une imagerie. Les auteurs ne discutent pas non plus la non spécificité de ce symptôme notamment chez le nouveau-né prématuré. La construction d'un diagramme de flux en partant de tous les sujets avec les symptômes d'alerte permettrait aux auteurs de réaliser que seuls les sujets ayant eu recours aux soins et ayant bénéficié de la procédure diagnostique ont été étudiés et que les conclusions faites sur la population d'étude ne sont donc pas forcément applicables à l'ensemble de la population avec les symptômes d'alerte.

Une *reporting guideline* pour les études sur les délais diagnostiques

Nous pensons que les éléments méthodologiques que nous venons d'illustrer doivent être clairement rapportés dans un article afin que le lecteur puisse évaluer le risque de biais ou d'obstacles à la transportabilité potentiellement associés à ces éléments. Il existe des recommandations, appelées *reporting guidelines*, pour guider les chercheurs dans la rédaction de leurs rapports d'étude (articles scientifiques) afin de mettre en lumière les points critiques concernant la méthodologie, les résultats et l'interprétation des résultats. Il existe plus de 280 *reporting guidelines* correspondant à des types d'études différents et listées par le réseau EQUATOR. Il a été montré que l'implémentation de *reporting guidelines* contribuait à l'amélioration de la qualité des articles publiés.(40–44)

Le réseau EQUATOR a émis des recommandations pour la construction de *reporting guideline*.(45) Selon ces recommandations, la première étape de construction consiste en une analyse de la littérature pour identifier des *reporting guidelines* déjà existantes. Il n'existe aucune *reporting guideline* enregistrée dans un EQUATOR spécifiquement dédiée aux études sur les délais diagnostiques. Nous pourrions alors nous demander si certaines *reporting guidelines* pourraient être adaptées. Les études sur les délais diagnostiques ont des points communs avec les études évaluant les performances des tests diagnostiques comme par exemple

les risques de biais liés à la sélection de la population ou l'importance de l'évaluation à l'aveugle de l'issue. Cependant, les études évaluant le délai diagnostique ont une étendue d'intérêt beaucoup plus large que la simple procédure diagnostique. Ainsi la *reporting guideline* STARD* (*STAndards for Reporting of Diagnostic accuracy*) ne mentionne pas de façon explicite (parce que cela ne concerne pas les études diagnostiques) la nécessité de lister les symptômes devant amener le clinicien à enclencher la procédure diagnostique.(46) Elle ne mentionne pas non plus la nécessité de rapporter si, et comment, les auteurs se sont assurés que tous les sujets avec les symptômes d'alerte avaient bénéficié de cette procédure. Les études sur les délais diagnostiques étant le plus souvent observationnelles, la *reporting guideline* STROBE* (*Strengthening The Reporting of OBservational studies in Epidemiology*) pourrait être utilisée. Cependant cette *reporting guideline* est généraliste et n'aborde pas l'ensemble des points critiques que nous venons de détailler plus haut.(47) Weller et coll. ont développé en 2012 des recommandations (*Aarhus statement*), obtenues par une méthode de consensus entre experts, pour améliorer la conduite et le *reporting* des études sur le diagnostic précoce en cancérologie. Une revue systématique des études évaluant un délai diagnostique avait été conduite préalablement pour dégager les définitions concernant les différents constituants du délai diagnostique et les différentes méthodes de recueil des données. Ces recommandations étaient donc centrées sur la définition et la mesure du délai diagnostique faisant abstraction de la sélection de la population, de l'analyse de ces délais et notamment de l'analyse de leur association avec les caractéristiques des patients.(48) De plus, ces recommandations étaient centrées sur la problématique du cancer et mêlaient *reporting* et *conducting* (c'est-à-dire conduite de l'étude).

Devant le constat qu'il n'existait pas de *reporting guideline* spécifique des délais diagnostiques comprenant des items sur toutes les dimensions des études sur les délais diagnostiques, nous avons entrepris d'en construire une. L'étape suivante était donc de confirmer

la nécessité d'une telle *reporting guideline* en évaluant la qualité du *reporting* de la littérature existante, ce qui a fait l'objet du chapitre 1 de cette partie I.

Figure 3. Evolution du nombre de publications sur les délais diagnostiques (identifiées avec les termes de recherche utilisés dans la partie I - chapitre 2) et du nombre total d'articles publiés dans PubMed entre 2000 et 2014.

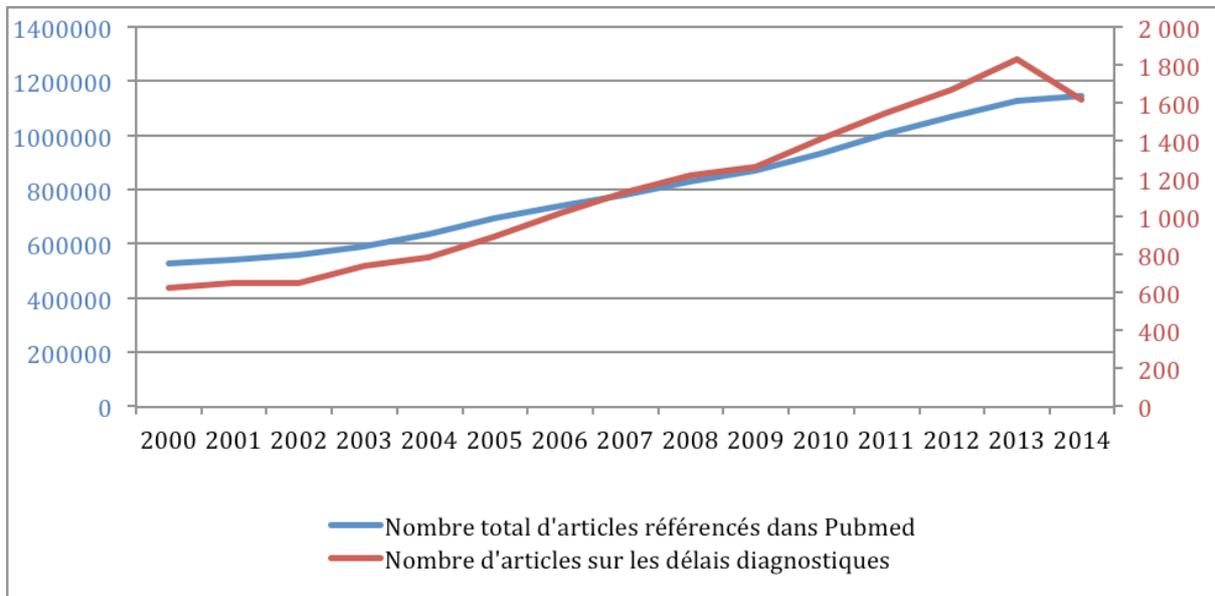
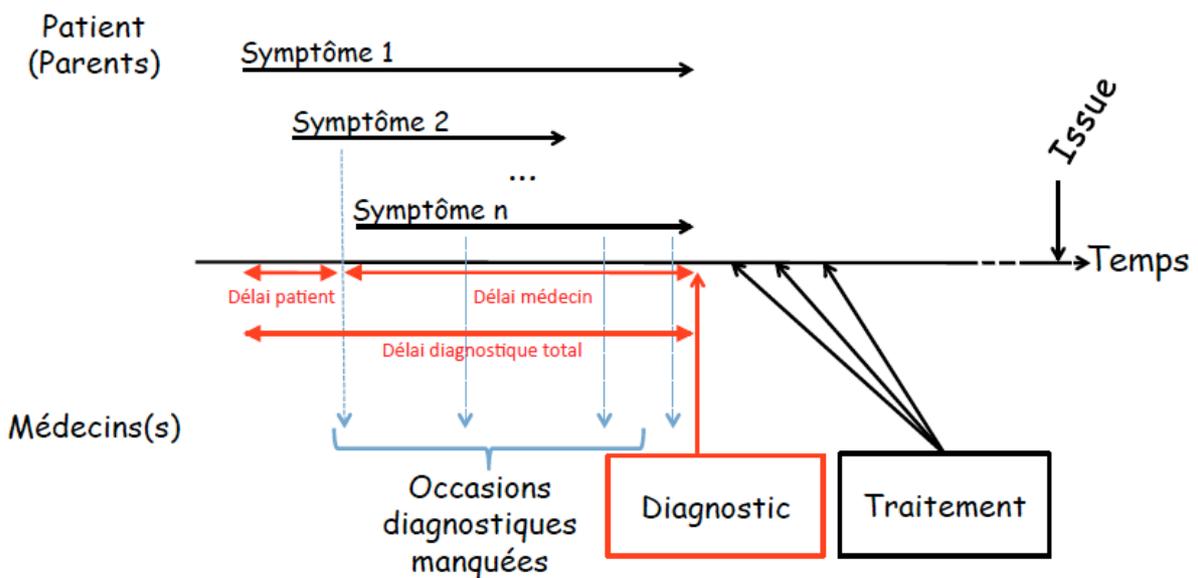


Figure 4. Schéma d'un parcours de soin type, des premiers symptômes à l'issue.



CHAPITRE 1

Qualité du *reporting* des études sur les délais diagnostiques : revue systématique en pédiatrie

Ce chapitre a fait l'objet de l'article suivant :

Quality of reporting of studies evaluating time to diagnosis: a systematic review in
paediatrics.

Elise Launay, Michele Morfouace, Catherine Deneux-Tharaux, Christèle Gras Le-Guen,

Philippe Ravaud, Martin Chalumeau

Archives of Disease in Childhood. 2014 ;99:244-50 (Facteur d'impact = 2,9)

(Annexe 1, page 213)

1. Introduction

Comme nous venons de le voir, l'évaluation de la qualité de la littérature existante est une des premières étapes de la construction d'une *reporting guideline*. Nous avons choisi la littérature pédiatrique car plusieurs arguments nous amènent à penser que cette population est particulièrement à risque de soins suboptimaux notamment dans le domaine diagnostique. En effet, les particularités notamment physiologiques de l'enfant, liées à l'immaturation initiale puis à la maturation progressive de ses grandes fonctions vitales, en font un groupe ayant des caractéristiques très spécifiques.(49) Ces enfants sont à risque de soins suboptimaux dans le domaine thérapeutique comme cela a été largement démontré (50–52) et ce qui a conduit à l'élaboration de recommandations pour la pratique clinique et la mise en place de programmes de prévention efficaces.(52,53) Les enfants sont probablement aussi exposés à plus de soins suboptimaux dans le domaine du diagnostic par comparaison aux adultes. Le sous-groupe des enfants d'âge pré-verbal ne pouvant détecter, rapporter ou préciser lui-même ses symptômes, il est dépendant des connaissances et des pratiques d'une tierce personne (le plus souvent les parents) puis du médecin.(54,55) Pour des raisons différentes, le groupe des adolescents a tendance à ne pas mentionner ses symptômes.(56) Par ailleurs, les enfants d'âge pré-scolaire sont souvent atteints de symptomatologie chronique (comme le reflux gastro-oesophagien) ou d'épisodes aigus récurrents pléomorphes (infections virales respiratoires et digestives) ou présentent un retentissement général (anorexie, douleurs abdominales, vomissement) pour de nombreuses affections bénignes. Ceci rend les symptômes d'appel des affections graves (comme la fièvre, la douleur, les vomissements, l'anorexie, la dyspnée) peu spécifiques, majorant le risque d'errance diagnostique.(57) La rareté des issues défavorables dans cette classe d'âge dont le taux de mortalité est le plus bas en comparaison aux nouveau-nés et aux adultes peut contribuer aussi à diminuer la vigilance des cliniciens. Enfin, les procédures diagnostiques ont souvent un niveau de validation inférieur à celui obtenu chez

l'adulte en raison de la difficulté à poursuivre des études cliniques en pédiatrie (réticence des parents et des médecins à l'inclusion de patients dans des études prospectives, difficultés de réaliser les tests de référence parfois invasifs, limitation du marché pédiatrique entraînant des difficultés de financement industriel des études).

L'objectif principal de ce premier chapitre était donc d'évaluer la qualité du *reporting* des études primaires investiguant un délai diagnostique par le moyen d'une revue systématique dans le champ particulier de la pédiatrie.

2. Matériel et méthodes

Nous avons évalué par une revue systématique la qualité du *reporting* de 50 articles rapportant une étude primaire investiguant un délai diagnostique en pédiatrie. Nous avons utilisé la méthodologie décrite par le « *Centre for Review and Dissemination* »(58) et nous avons rapporté cette revue systématique selon les recommandations préconisées par PRISMA* (*Preferred Reporting Items for Systematic reviews and Meta-Analyses*).(59) Le nombre d'articles inclus a été choisi de façon arbitraire par analogie au nombre médian d'articles (33, avec des extrêmes de 29 à 86) inclus dans des revues systématiques évaluant la qualité du *reporting* pour d'autres types d'études (notamment essais randomisés contrôlés et règles de décision clinique).(60–65)

2.1 Identification des études

Les études ont été systématiquement identifiées via une recherche électronique sur PubMed. La combinaison des termes utilisée pour la recherche était la suivante : (((pitfalls [Title] OR missed [Title]) OR delay* [Title]) OR interval* [Title]) AND diagnos* [Title/Abstract] NOT ((((((gene [Title/Abstract] OR chromosome [Title/Abstract]) OR antenatal [Title/Abstract]) OR review [Publication Type]) OR Case Reports [Publication Type]) OR editorial

[Publication type]) OR comment [Publication type]) OR letter [Publication type]). Nous avons également utilisé des limites d'âge (filtre « child » de PubMed) et de langue (filtre « English » de PubMed). Nous avons lancé la recherche à partir d'octobre 2011 et sommes remontés dans le temps jusqu'à atteindre 50 articles remplissant les critères d'inclusion. Nous avons vérifié la sensibilité de notre recherche électronique en comparant pour l'année 2008 les articles identifiés par cette recherche avec ceux identifiés par une recherche manuelle dans les tables des matières des 10 journaux généralistes et 20 journaux pédiatriques avec les plus forts facteurs d'impact pour leur discipline respective. Tous les articles identifiés par la recherche manuelle avaient également été retrouvés par la recherche électronique.

2.2 Critères d'inclusion et d'exclusion

Les études ont été sélectionnées par la lecture de leur titre puis de leur résumé et éventuellement du corps de l'article. Nous avons inclus toutes les études originales (primaires) dont l'objectif était de mesurer le délai diagnostique d'une maladie, d'évaluer la fréquence des retards diagnostiques ou d'étudier les déterminants ou conséquences de ceux-ci. Dans tous les articles, le délai diagnostique était défini par les auteurs comme l'intervalle de temps séparant la date du début de la maladie (apparition des premiers symptômes) et la date du diagnostic. Le retard diagnostique était défini comme un délai diagnostique jugé trop long par les auteurs des études incluses. Nous avons défini la pédiatrie comme l'ensemble des procédures médicales et chirurgicales concernant les patients âgés de 0 à 18 ans. Les articles rapportant des cas cliniques ont été exclus lorsqu'ils concernaient moins de 20 cas (seuil arbitraire). Les « *brief cases* » ou « *short reports* » ont été exclus car la quantité d'information disponible dans ce type d'article peut être limitée par les éditeurs et obliger les auteurs à restreindre la description de certains points méthodologiques, comme notre équipe a pu en faire l'expérience.⁽³⁶⁾ Les études concernant le diagnostic anténatal ont également été

exclues car elles s'apparentent plus à la problématique du dépistage que du diagnostic précoce. Les articles concernant à la fois une population pédiatrique et adulte n'ont été inclus que si au moins la moitié (seuil arbitraire) des sujets inclus était d'âge pédiatrique ou si les données pédiatriques étaient rapportées séparément.

2.3 Evaluation de la qualité du *reporting* : choix des items

Afin d'évaluer la qualité du *reporting*, nous avons mis au point une *checklist* dont les items ont été choisis en s'appuyant sur :

- (i) les critères de Wilson et Junger publiés par l'Organisation Mondiale de la Santé pour évaluer la pertinence d'actions de dépistage ;(66)
- (ii) les items utilisés dans d'autres *reporting guidelines* pour les études observationnelles (STROBE),(47) pour les études sur les tests diagnostiques (STARD)(46) ou pour les études sur les marqueurs pronostiques (*Reporting recommendations for tumour MARKer prognostic studies-REMARK*),(67) ou utilisés dans l'outil d'évaluation des risques de biais et de facteurs de variations pour les études sur les performances des tests diagnostiques (QUADAS 1 et 2) ;(68,69)
- (iii) une analyse préliminaire de 10 articles rapportant une étude investiguant un délai diagnostique en pédiatrie ;(26–35)
- (iv) l'expérience de notre équipe de recherche dans le domaine des délais diagnostiques.(19,22,23,36,37)

Une *checklist* de 35 items a ainsi été établie (cf. **Tableau 1**, pages 75) après discussion avec 3 épidémiologistes de l'équipe de recherche (MC, CDT et PR). Parmi ces 35 items, nous avons identifié empiriquement 21 items spécifiques de l'étude des délais diagnostiques potentiellement associés à des risques de biais ou d'obstacles à la transportabilité des résultats ou importants pour l'interprétation des résultats et leurs conséquences cliniques (mise en

place d'un plan d'action) comme expliqué en introduction de cette partie. Parmi ces 21 items nous avons mis en avant 11 items considérés comme spécifiques des études sur les délais diagnostiques et paraissant critiques pour une bonne qualité du *reporting* (cf. items en gras dans le **Tableau 1**, pages 75).

2.4 Extraction des données

Deux des co-auteurs (EL et MM) ont indépendamment identifié les articles potentiellement éligibles et extrait les informations correspondant aux 35 items de la *checklist*. Chaque item était présenté sous forme de question (« est-ce que les auteurs ont rapporté... ? ») et pouvait être codé selon 3 modalités : « oui », « non » ou « partiellement ». En cas de désaccord sur l'éligibilité d'un article ou sur le codage d'un item d'évaluation du *reporting*, une discussion était menée avec un 3^{ème} auteur (MC) afin d'atteindre un consensus.

2.5 Analyses

Nous avons décrit les caractéristiques générales des revues dans lesquelles les articles ont été publiés : spécialité de la revue (pédiatrique versus non pédiatrique) et facteur d'impact calculé selon la méthode du *Science Index Citation* (codé d'une part comme une variable continue et d'autre part dichotomisée autour du 90^{ème} percentile).(70) Nous avons également décrit les caractéristiques générales des articles : année de publication, maladie étudiée, type de recrutement (en population ou selon le niveau de soins) et mention de l'utilisation d'une *reporting guideline*. Nous avons décrit pour chaque item la fréquence avec laquelle il était rapporté dans les articles inclus et le nombre d'articles où l'ensemble des 11 items considérés comme critiques était rapporté. Pour éviter la multiplication des tests, nous avons sélectionné *a priori* 5 items parmi ces 11 items et analysé l'association entre la fréquence à laquelle ces items étaient rapportés et des déterminants potentiels de qualité du *reporting* : année de

publication et spécialité et facteur d'impact du journal. Les 5 items choisis étaient les suivants : « identification des participants » (item 7, **Tableau 1**, page 75), « définition des symptômes d'alerte » (c'est-à-dire la définition des symptômes qui, lorsqu'ils sont identifiés par un clinicien, devraient faire enclencher la procédure diagnostique) (item 8, **Tableau 1**), « définition de la date d'apparition des premiers symptômes détectables de la maladie (T0) » (item 13a, **Tableau 1**), « définition de la date du diagnostic (Tx) » (item 13b, **Tableau 1**) et « délai diagnostique moyen ou médian » (item 22a, **Tableau 1**). Pour ces analyses, nous avons regroupé les catégories « oui » et « partiellement » en opposition à la catégorie « non ». Nous avons utilisé selon la situation les tests suivants : χ^2 , χ^2 de tendance* ou test de Mann-Whitney via le logiciel SPSS v19 (SPSS, Chicago, Illinois, USA).

3. Résultats

3.1 Articles

Nous avons identifié 568 articles en remontant d'octobre 2011 jusqu'à janvier 2005. Nous avons retenu 50 articles remplissant les critères d'inclusion (cf. diagramme de flux, **Figure 5**, page 77). (26,28–31,33–35,37,71–111) Les affections étudiées étaient très variables mais 4 étaient récurrentes : cancer (n=10), accident vasculaire cérébral (n=5), lésions traumatiques (n=4) et maladie de Kawasaki (n=4). Les caractéristiques générales des revues et des articles sont détaillées dans le **Tableau 2** (page 77). Au total, 37 (74%) articles rapportaient dans leur conclusion qu'une action corrective était nécessaire pour réduire le délai diagnostique de l'affection étudiée. Dans aucun article, les auteurs ne mentionnaient l'utilisation d'une *reporting guideline* pour rapporter leur étude.

3.2 Qualité du *reporting*

Les résultats de l'évaluation de la qualité du *reporting* des 50 études incluses sont présentés dans le **Tableau 1** (page 75). Les items génériques étaient très souvent rapportés alors que les items plus spécifiques des études sur les délais diagnostiques étaient rarement rapportés. Moins de 50% des articles ont rapporté de façon complète la définition des dates T0 et Tx entourant le délai diagnostique. Quatorze articles (28%) ont listé les symptômes d'alerte pouvant conduire au diagnostic et 2 (4%) ont précisé que tous les sujets présentant le symptôme d'alerte avaient systématiquement bénéficié de la procédure diagnostique lors de la période de l'étude. Deux articles ont rapporté l'ensemble des 11 items considérés comme critiques.(37,101)

3.3 Déterminants potentiels de la qualité du *reporting*

Les items 22 (délai moyen ou médian) et 11 b (définition du moment précis du diagnostic Tx) étaient significativement plus renseignés au fil des années (de 0 et 1 article sur 3 en 2005 à 8 et 6 sur 9 articles en 2011 respectivement, $p < 0,02$). Nous n'avons pas mis en évidence de tendance significative pour les 3 autres items ($p > 0,2$). Nous n'avons mis en évidence aucune association significative ($p > 0,1$) entre les 5 items sélectionnés et le facteur d'impact ou la spécialité du journal (pédiatrique versus non pédiatrique).

4. Discussion

4.1 Principaux résultats

Cette revue systématique nous a permis d'identifier plusieurs faiblesses dans le *reporting* des études sur les délais diagnostiques. La qualité du *reporting* était bonne pour les items génériques mais bien moindre pour les items plus spécifiques aux études sur les délais diagnostiques. Seulement 14 articles (28%) ont listé les symptômes d'alerte et seulement 2

(4%) ont précisé que tous les sujets présentant le symptôme d'alerte avaient systématiquement bénéficié de la procédure diagnostique durant la période d'étude. Par ailleurs, la précision et la transportabilité de la mesure du délai diagnostique étaient limitées par l'absence de définition explicite des dates de début de la maladie et du diagnostic dans la moitié des articles. Ces défaillances dans le *reporting* limitent la possibilité d'évaluer les risques de biais et d'obstacles à la transportabilité des études au risque d'entraîner une interprétation erronée des résultats.

L'amélioration avec le temps de l'information disponible pour certains items de notre *checklist* (la date du diagnostic et le délai diagnostique moyen ou médian) indique qu'une optimisation du *reporting* est possible. De plus, le fait que 2 des articles rapportaient correctement l'ensemble des 11 items de notre *checklist* considérés comme critiques montre également qu'un *reporting* de bonne qualité (complet) est envisageable.

Nous n'avons pas mis en évidence d'association entre le facteur d'impact des journaux publiant les études incluses et la fréquence de *reporting* d'items clefs. Ce résultat pouvait suggérer l'existence d'un manque général de connaissance ou d'une sous-estimation des risques de biais et d'obstacles à la transportabilité liés à ce type d'étude mais un manque de puissance était aussi possible compte-tenu du petit effectif de notre étude.

Enfin, près de deux tiers des articles (37/50) concluaient qu'une action corrective était nécessaire pour réduire les délais mais très peu rapportaient les arguments pour justifier un tel plan d'action. Les arguments attendus étaient le fait que le délai diagnostique soit « long » ou qu'un retard diagnostique soit « trop fréquent » ou évitable notamment par l'identification de déterminants. Une action correctrice devrait également être justifiée par la mise en évidence de conséquences néfastes en terme de morbi-mortalité du délai ou retard diagnostique. Enfin pour garantir une action correctrice pertinente il faudrait que les auteurs rapportent une discussion concernant la faisabilité, la sensibilité et la spécificité de cette action. Un tel

glissement dans les articles relève du phénomène de *spin*, c'est-à-dire de surinterprétation des résultats, récemment décrit.(112)

4.2 Justification des items

La *checklist* que nous proposons est constituée d'items nouveaux et originaux pour évaluer la qualité du *reporting* d'aspects méthodologiques particuliers des études sur les délais diagnostiques. Parmi les 35 items que nous avons utilisés, certains sont communs à tous les types d'études observationnelles et tous ne sont pas reliés à des informations nécessaires pour évaluer les risques de biais ou d'obstacles à la transportabilité. Certains items génériques permettent de vérifier la pertinence de l'étude (items 2-4), de vérifier que les auteurs ont bien répondu à la question de recherche en évitant le *spin* (items 1, 26, 23 et 32) ou de vérifier la précision des résultats (item 17) ou la puissance de l'étude (item 18). D'autres items sont plus spécifiques de l'étude des délais diagnostiques. Nous avons ainsi spécifié 21 items parfois optionnels car dépendants des objectifs de l'étude mais permettant de souligner des points importants pour l'interprétation et l'application clinique des résultats (item 19, 24-25, 31 et 33). Enfin parmi ces 21 items, 11 nous ont semblé indispensables car reliés à des points méthodologiques critiques (items 7-9, 11, 13-15, 22, 28-30). Ces items avaient été décrits et illustrés dans l'introduction de cette partie I (page 47).

4.3 Limites de l'étude

Moher et coll. ont mis en évidence, en étudiant les *checklists* et échelles d'évaluation de la qualité du *reporting* des essais randomisés contrôlés, que les items qui les composaient étaient très peu souvent justifiés par leurs auteurs. Nous n'avons pas pu faire la démonstration de l'existence des biais dans les études sur les délais diagnostiques du fait du nombre limité d'études évaluant le délai diagnostique de la même maladie dans un contexte clinique

comparable et du fait que moins de deux tiers des études rapportaient de façon explicite un délai de façon quantitative (moyen ou médian) ou une fréquence de retard diagnostique. Nous avons donc justifié nos items en nous appuyant d'une part sur des risques de biais et d'obstacles à la transportabilité déduits d'autres types d'études partageant potentiellement ces éléments (113,114) et d'autre part sur l'expérience de notre équipe de recherche dans le domaine des délais diagnostiques (cf. introduction de cette partie I, page 47).(19,22,23,36–38,115) Certains items de notre *checklist* dépassaient le cadre du *reporting* des éléments méthodologiques pouvant influencer la qualité des études. En effet, les items concernant le plan d'action (items 33, 34, 35) se rapportent plus aux perspectives des études sur les délais diagnostiques et permettent plutôt d'évaluer le risque de sur-interprétation des résultats. Nous verrons dans le chapitre 3 de cette partie I que ces items seront critiqués par le panel international d'experts mis en place pour les valider.

Notre étude était limitée aux études publiées en langue anglaise ce qui a pu amener un biais de sélection dont le sens et la force ne peuvent être anticipés.(116,117) Par ailleurs, la mise en évidence des déterminants de la qualité a pu être limitée par un manque de puissance statistique liée au faible effectif. Enfin, notre étude était limitée à la littérature pédiatrique qui n'est probablement pas le reflet de l'ensemble de la littérature médicale. Il est cependant difficile de prévoir dans quel sens la restriction de l'étude à la population pédiatrique, particulièrement à risque de délais diagnostiques longs, aurait pu influencer sur l'estimation de la qualité du *reporting*.

5. Conclusion : une *reporting guideline* spécifique est nécessaire

Les faiblesses dans le *reporting* des 50 études analysées dans cette revue systématique soutiennent notre hypothèse initiale qu'une *reporting guideline* spécifique pour les études sur les délais diagnostiques doit être développée et implémentée. Ces résultats soulignent

également que les *reporting guidelines* existantes à l'époque de la publication des articles inclus (STARD,(46) et STROBE (47)) ne remplissent pas complètement ce besoin ou ne sont pas appliquées. En effet, l'utilisation d'une *reporting guideline* n'était mentionnée dans aucun des 50 articles. L'*Aarhus Statement* a été publié en 2012, donc après les articles inclus dans notre revue systématique et également après la mise au point de la première version de notre *checklist*. Comme nous l'avons dit en introduction de cette partie I, l'*Aarhus Statement* est focalisé sur la définition et la mesure du délai. Deux items concernant notamment le mode de collecte des données nous ont semblé pouvoir être complémentaires des items constituant notre *checklist*. Nous avons donc enrichi notre checklist des 2 items suivants avant de la soumettre aux experts pour l'étape de validation rapportée dans le chapitre 3 :

- « Décrire les méthodes utilisées pour évaluer T0 et Tx (par exemple, entretien, revue de dossiers, analyse de bases de données...) et

- « Donner les détails concernant la méthode d'estimation de T0 et de Tx :

a. pour les entretiens : existence d'une validation de l'instrument utilisé, délai entre l'entretien et le diagnostic, triangulation avec d'autres sources de données et

b. pour les revues de dossiers et les analyses de bases de données : décrire comment les informations ont été extraites, décrire la base de données utilisée. »

D'après les recommandations internationales du réseau EQUATOR,(45) l'étape de validation devrait théoriquement s'appuyer sur la démonstration empirique de biais et d'obstacles à la transportabilité liés à chaque item de la *checklist*. Cependant, comme nous venons de le voir, une telle démonstration est ici difficile. Nous avons donc souhaité explorer les méthodes utilisées actuellement par d'autres équipes pour évaluer les risques de biais et d'obstacles à la transportabilité dans les études sur les délais diagnostiques. Une telle évaluation est fortement recommandée dans les revues systématiques. Le chapitre 2 de cette

première partie concerne donc l'analyse des outils d'évaluation du risque de biais et d'obstacles à la transportabilité dans les revues systématiques sur les délais diagnostiques.

6. Figures et Tableaux

Tableau 1 Items utilisés pour l'analyse de la qualité du *reporting* des articles et fréquence à laquelle ils ont été renseignés dans les articles inclus.

Sections et sous sections	Est ce que les auteurs... ?	Nombre d'articles n (%)# où l'item était		
		rapporté	partiellement rapporté	non rapporté applicable
Introduction		50 (100)	0	0
1.	Ont rapporté l'objectif de l'étude	33 (66)	7 (14)	10 (20)
2.	Ont rapporté la fréquence et la morbidité de l'affection étudiée	28 (56)	2 (4)	20 (40)
3.	Ont rapporté l'existence d'un traitement	36 (72)	5 (10)	9 (18)
4.	Ont rapporté si un diagnostic ou un traitement précoce avait une influence sur l'issue de la maladie.	48 (96)	0	2 (4)
5.	Ont décrit la population d'étude	26 (52)	0	24 (48)
a.	Les critères d'inclusion	8(16)	0	42(84)
b.	Les critères d'exclusion	28 (56)	1 (2)	21 (42)
6.	Ont rapporté l'existence éventuelle de sous groupes particulièrement à risque de délai extrême et s'ils ont été exclus ou analysés séparément###	14 (28)	2 (4)	34 (68)
7.	Ont décrit comment les cas ont été identifiés	2 (4)	0	48 (96)
8.	Ont listé les symptômes déclenchant habituellement la procédure diagnostique par le clinicien (symptômes d'alerte) au cours de la période de l'étude			
9.	Ont rapporté si tous les patients avec les symptômes d'alerte ont bénéficié de la procédure diagnostique			
10.	Ont décrit le recrutement	47 (94)	0	3 (6)
a.	Type (mono ou multi centrique, en population)	44 (88)	0	6 (12)
b.	Lieux	17 (44)	0	22 (56)
c.	Niveau de soin (primaire, secondaire ou tertiaire)	45 (90)	1 (2)	4 (8)
d.	Période de l'étude			
Définition et mesure du délai				
11.	Ont rapporté la définition utilisée pour T0 de la maladie (symptômes, signes)	22 (44)	2 (4)	26 (52)
b.	Tx du diagnostic	24 (48)	3 (6)	23 (46)
12.	Ont rapporté la définition des différents intervalles composant le délai diagnostique (délai patient, délai médical)##	15 (40)	0	23 (60)
13.	Ont décrit			
a.	Combien d'experts ont évalués le délai diagnostique	10 (20)	0	40 (80)
b.	Si l'évaluation du délai diagnostique par les experts était indépendante	2 (20)	3 (30)	5 (50)
c.	Quel était le degré d'accord entre ces différents experts	1 (20)	0	4 (80)
d.	Comment les désaccords ont été résolus	1 (20)	0	4 (80)
14.	Ont décrit la qualification des experts qui ont évalué le délai diagnostique	9 (18)	0	41 (82)
15.	Ont précisé si les experts qui ont évalué le délai diagnostique étaient à l'aveugle de l'issue	1 (2)	0	49 (98)
16.	Ont donné la définition du retard diagnostique (ou dans résultats)##	20 (87)	0	3 (13)
a.	D'après la littérature	6		
b.	De façon arbitraire	12		
c.	Comme une variation de la distribution du délai diagnostique observé dans la population de l'étude	2		

Sections et sous sections	Est ce que les auteurs ... ?	Nombre d'articles n (%) * où l'item était			
		rapporté	partiellement rapporté	non rapporté	non applicable
<i>Analyses statistiques</i>	17. Ont décrit les méthodes statistiques utilisées pour analyser les déterminants et les conséquences du délai diagnostique	40 (87)	0	6 (3)	4
	18. Ont donné le rationnel pour justifier la taille de l'échantillon	1 (2)	0	49(98)	0
	19. Ont décrit comment les facteurs de confusion avaient été pris en compte pour analyser les déterminants et les conséquences du délai diagnostique ^{##}	21 (47)	0	24 (53)	5 ^e
Résultats	20. Ont rapporté le nombre de				
	a. Patients éligibles	33 (66)	1 (2)	16 (32)	0
	b. Patients inclus	50 (100)	0	0	0
	21. Ont rapporté les caractéristiques de la population d'étude	43 (86)	1 (2)	6 (12)	0
	22. Ont rapporté pour l'ensemble de la population étudiée				
	a. Le délai diagnostique moyen ou médian	30 (60)	0	20 (40)	0
	b. La distribution du délai diagnostique	31 (62)	0	19 (38)	0
	23. Ont rapporté la fréquence du retard diagnostique dans la population de l'étude	18 (78)	0	5 (22)	27 ^b
	24. Ont rapporté les déterminants du délai diagnostique ^{##}	37 (95)	0	2 (5)	11 ^b
	25. Ont rapporté les conséquences du délai diagnostique ^{##}	21 (95)	0	1 (5)	28 ^e
Discussion	26. Ont discuté les principaux résultats et donné leur interprétation	50 (100)	0	0	0
<i>Résultats principaux</i>	27. Ont comparé les résultats avec ceux précédemment publiés dans la littérature	43 (86)	0	7 (14)	0
<i>Limites</i>	28. Ont discuté les éventuels biais de recrutement	12 (30)	0	28 (70)	10 ^a
	29. Ont discuté l'existence éventuelle de cas non diagnostiqué	6 (12)	1 (2)	43 (86)	0
	30. Ont discuté l'existence de biais et d'imprécision de mesure	19 (38)	2 (4)	29 (58)	0
	31. Ont discuté un éventuel biais de « diagnostic précoce* » ^{##}	0	0	7 (100)	43
	32. Ont rapporté si le retard diagnostique était trop fréquent ou le délai diagnostique trop long ^{##}	37 (78)	0	10 (22)	3 ^g
<i>Application clinique</i>	33. Ont discuté l'évitabilité du délai ^{##}	19 (40)	6 (13)	22 (47)	3 ^h
	34. Ont proposé un plan d'action pour réduire le délai diagnostique étudié	17 (36)	21 (45)	9 (19)	3 ^h
	35. Ont discuté la sensibilité, la spécificité et la faisabilité de leur plan d'action proposé le cas échéant ^{##}	3 (8)	0	35 (92)	12

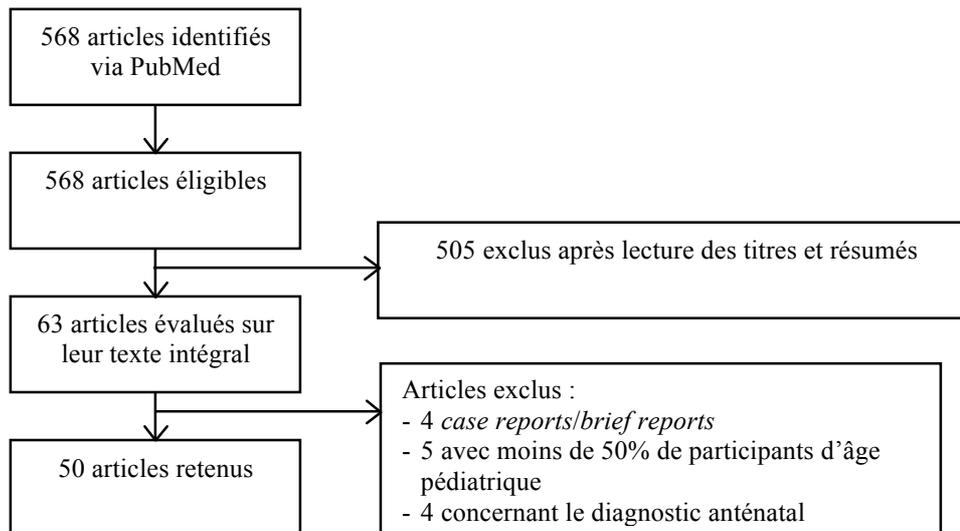
#. Le dénominateur des pourcentage est le nombre total d'articles où l'item est applicable (nombre total d'articles moins le nombre de « non applicable »),

items particulièrement important pour les articles mentionnant une action correctrice, ###items ne pouvant être évalués sans connaissance de la maladie étudiée.

Les items en gras sont les items identifiés comme critiques.

a, en population ; b, articles ne mentionnant pas le concept de retard diagnostique ; c, articles n'évaluant pas les déterminants ni les conséquences ; d, articles n'évaluant pas les déterminants ; e, articles n'évaluant pas les conséquences ; f, articles ne mentionnant aucune association entre le délai diagnostique et la survie ;

g, articles ne mentionnant pas que le délai diagnostique était trop long ; h, articles ne mentionnant aucune action correctrice

Figure 5. Diagramme de flux des études incluses dans la revue de la qualité du *reporting*.**Tableau 2.** Caractéristiques générales des journaux et des articles inclus.

Caractéristiques des journaux	
Facteur d'impact, médiane [extrêmes] ; % dans le top 10 de la spécialité	2.8 [0,6-8,2] ; 28%
Discipline, n (%)	
<i>Pédiatrie</i>	23 (46)
<i>Revue Généraliste</i>	5 (10)
<i>Cancer</i>	4 (8)
<i>Autre spécialité médicale</i>	12 (24)
<i>Autre spécialité chirurgicale</i>	5 (10)
<i>Epidémiologie</i>	1 (2)
Caractéristiques des articles	
Année de publication, n (%)	
2005	3 (6)
2006	4 (8)
2007	7 (14)
2008	9 (18)
2009	8 (16)
2010	10 (20)
2011	9 (18)
Type de recrutement, n (%)	
<i>En population</i>	10 (20)
<i>Dans un (des) centre(s) de référence</i>	14 (28)

CHAPITRE 2

Les outils d'évaluation de la qualité des études primaires dans les revues systématiques portant sur l'évaluation d'un délai diagnostique : revue systématique méthodologique

Ce chapitre a fait l'objet de l'article suivant :

“Quality assessment” of primary studies in systematic reviews of time to diagnosis: a
methodological systematic review and call for clarification

Elise Launay, Jérémie F. Cohen, Michèle Morfouace, Christèle Gras-Le Guen,

Philippe Ravaud, Martin Chalumeau

(Soumis)

(Annexe 2, page 215)

1. Introduction

Nous venons de voir dans le chapitre 1 de cette partie I que les points méthodologiques potentiellement critiques, comme les processus de sélection de la population ou de mesure du délai, n'étaient que peu rapportés dans les études sur les délais diagnostiques. Ce constat nous a conduits à proposer une première version d'une *reporting guideline* qui aiderait les auteurs à rapporter les points méthodologiques critiques potentiellement soumis à des risques de biais et d'obstacles à la transportabilité. La preuve de l'existence de risques de biais et d'obstacles à la transportabilité a pu être faite pour les essais randomisés contrôlés et pour les études sur les performances diagnostiques en comparant les résultats d'études ne différant que sur un point méthodologique précis (randomisation, aveugle, mode de recrutement ...).(113,118) Nous n'avons pu faire la démonstration de l'existence des biais dans les études sur les délais diagnostiques du fait du nombre limité d'études évaluant le délai diagnostique de la même maladie dans un contexte de soin comparable. Etudier la variabilité des délais calculés dans plusieurs études en fonction de variations méthodologiques ponctuelles est donc probablement impossible.

Nous avons souhaité savoir comment les autres chercheurs travaillant dans le domaine des délais diagnostiques avaient contourné cet obstacle. L'analyse des outils utilisés par les auteurs de revues systématiques pour évaluer les risques de biais dans les études primaires incluses constituait la manière la plus simple. En effet, l'évaluation des risques de biais est une étape indispensable à l'analyse critique de la littérature et notamment à la conduite de revues systématiques ou d'une revue systématique au singulier. Les *reporting guidelines* de revues systématiques PRISMA et QUOROM (*Quality of Reporting of Meta-analyses*, précurseur de PRISMA) et MOOSE (*Meta-analysis Of Observational Studies in Epidemiology*) recommandent ainsi de procéder à une analyse des risques de biais des études primaires incluses dans la revue systématique.(59,62,119) De telles méta-revues des revues

systematiques ont déjà été utilisées dans le domaine des études interventionnelles non randomisées par Deeks et coll. et celui des études observationnelles par Viswanathan. Elles avaient alors mis en évidence une grande variabilité des outils utilisés. (120,121)

Les objectifs spécifiques de ce chapitre étaient donc :

- de caractériser les outils utilisés pour évaluer la qualité méthodologique des études primaires incluses dans les revues systématiques sur les délais diagnostiques ;
- d'identifier les risques de biais ou d'obstacles à la transportabilité associés aux études sur les délais diagnostiques rapportés par les auteurs de revue systématique et de les confronter à ceux que nous avons discutés précédemment (cf. introduction de cette partie I) dans l'objectif de compléter notre *reporting guideline* proposée au Chapitre 1.

2. Méthodes

Pour cette revue systématique, nous nous sommes focalisés sur deux dimensions de l'évaluation de la qualité : l'évaluation des risques de biais (validité interne) et l'évaluation de la transportabilité (validité externe).(69,122,123) Nous avons conduit cette revue systématique en suivant les recommandations du *Centre for Review and Dissemination* (58) et nous avons rapporté cette revue systématique selon les recommandations préconisées par PRISMA.(59) Nous n'avons pas enregistré cette revue systématique dans PROSPERO (*international prospective register of systematic reviews*) car ce registre ne concerne pas les revues systématiques méthodologiques.(124)

2.1 Stratégie de recherche et critères d'inclusion

Nous avons conduit une recherche systématique dans MEDLINE via PubMed et dans le *Web of science* des articles, anglais ou français, publié jusque fin février 2015 inclus et rapportant une revue systématique, avec ou sans méta-analyse, ou son protocole. Nous avons fait évoluer

la stratégie de recherche par rapport à la revue systématique du chapitre 1 de cette partie I en s'inspirant des termes utilisés dans une revue systématique sur les délais diagnostiques des cancers de l'enfant conduite dans notre équipe.(19) La combinaison de recherche retenue contenait des termes associés à la notion de temps (*delay, interval, time, early, earlier, timely, timeliness, late*), d'accessibilité au soin (*seeking, barrier*), de diagnostic (*diagnosis, diagnoses, diagnostic*), de revue systématique (*systematic review, meta-analysis*) et de langues (*English, French*). La stratégie de recherche a été mise au point sur PubMed, elle a ensuite été adaptée pour *Web of Science* (cf. **Figure 6**, page 94). Nous avons complété ces recherches électroniques par une recherche manuelle des références des revues systématiques incluses ainsi que par une recherche sur *Google Scholar* pour repérer des articles citant les revues systématiques incluses. Les revues systématiques évaluant l'efficacité d'une intervention pour réduire les délais n'ont pas été incluses car des outils spécifiques existent pour évaluer les risques de biais dans les études interventionnelles.(122) Les revues systématiques n'incluant que des études primaires qualitatives ont également été exclues de notre recherche.

2.2 Sélection des études

Les titres et résumés des articles identifiés par la stratégie de recherche ont été analysés indépendamment par 2 des auteurs (MM et EL). En cas de doute sur la pertinence de l'inclusion nous avons lu le texte entier de l'article. En cas de désaccord sur l'inclusion, un troisième auteur (MC) a été sollicité.

2.3 Extraction des données

Pour chaque revue systématique incluse, 2 auteurs ont extrait les données suivantes de façon indépendante : titre de l'article, noms des auteurs, année de publication, facteur d'impact de la

revue,(70) type de revue systématique (avec ou sans méta-analyse), maladie étudiée, nombre d'études primaires incluses, qualification des auteurs basée sur l'unité de recherche ou le service d'affiliation (clinicien ou méthodologiste -biostatisticien, épidémiologiste ou spécialiste de santé publique-), objectifs de la revue systématique (mesure du délai, évaluation des déterminants ou des conséquences du délai), référence à une *reporting guideline* (PRISMA, MOOSE ou QUOROM),(59,62,119) mention d'une « évaluation de la qualité méthodologique » des études primaires dans le chapitre « Méthodes », utilisation des résultats de cette évaluation dans l'analyse ou l'interprétation des résultats et enfin méthode utilisée pour cette évaluation. En cas de désaccord, un consensus a été trouvé après discussion avec un 3^{ème} auteur (MC).

2.4 Analyses

Nous avons dans un premier temps décrit les caractéristiques générales des revues systématiques puis nous avons analysé « l'évaluation de la qualité » rapportée dans les articles. Pour ce faire nous avons calculé la fréquence de cette évaluation telle que mentionnée par les auteurs sans préjuger de ce que cette évaluation aurait dû être en théorie (c'est-à-dire une analyse des risques de biais et d'obstacles à la transportabilité) et nous avons également décrit la façon dont les auteurs avaient défini et mené cette évaluation. Nous avons caractérisé les outils utilisés pour l'évaluation de la qualité selon leur originalité à savoir : outils élaborés spécifiquement pour l'étude (« faits maison ») ou outils déjà publiés (échelle ou score d'évaluation de la qualité ou outil utilisé dans un autre revue systématique) utilisés tels quels ou adaptés pour l'étude. Nous avons également regardé si l'outil abordait des points spécifiques à l'étude des délais diagnostiques et, en cas d'outil « fait maison » ou adapté d'un outil existant, si le choix des items composant l'outil était justifié par les auteurs. Nous avons également décrit si les résultats de l'évaluation de la qualité étaient rapportés et comment ils

étaient utilisés dans l'analyse ou l'interprétation des résultats : exclusion des études de faible qualité, analyse en sous-groupe ou analyse de sensibilité.

Pour les revues systématiques qui rapportaient une revue de la qualité des études primaires ainsi que le ou les outil(s) utilisé(s), nous avons listé les items composant ces outils (un item étant une phrase interrogative ou non permettant d'évaluer un des aspects de la méthodologie de l'étude). Nous avons classé ces items en 5 catégories non exclusives en fonction de ce qu'ils évaluaient : risques de biais (de sélection, de classement*, de confusion ou d'attrition*), obstacles à la transportabilité (lié à la définition ou au spectre de la maladie ou aux caractéristiques de la population), analyses statistiques (puissance, données manquantes), qualité du *reporting* ou item ambigu. Pour mettre au point cette classification, nous nous sommes basés sur une analyse préliminaire de 20 items en utilisant la banque de biais pour les études observationnelles du *Research Triangle Institute* et les travaux de Deeks et coll. sur l'évaluation des risques de biais dans les études non-randomisées.(120,125) Nous avons également repéré les items plus spécifiques des délais diagnostiques en nous basant sur les points critiques identifiés au chapitre 1 de cette partie I, c'est-à-dire la définition et la mesure du délai, la problématique des cas non diagnostiqués et l'étude de la relation entre délai diagnostique et les caractéristiques des patients ou leur état en terme de morbi-mortalité.

Nous avons enfin étudié les déterminants d'une évaluation adéquate de la qualité que nous avons définie comme l'évaluation d'au moins un risque de biais ou obstacle à la transportabilité. Nous avons ensuite analysé l'association entre l'évaluation adéquate de la qualité dans une revue systématique et les déterminants potentiels suivants : année de publication de la revue systématique (en prenant comme césure arbitraire l'année 2010, soit un an après la publication de PRISMA), la présence d'un méthodologiste dans les co-auteurs et le facteur d'impact de la revue.

2.5 Analyses statistiques

Nous avons testé les associations en utilisant le test du χ^2 ou le test exact de Fisher pour les variables qualitatives et le test de Mann-Whitney pour les variables quantitatives. Nous avons utilisé pour ces analyses le logiciel STATA v11 (Statacorp, College Station, TX, USA).

3. Résultats

3.1 Description des revues systématiques

Nous avons identifié 1935 articles via la recherche électronique, après avoir éliminé les doublons. Un article supplémentaire a été identifié avec *Google Scholar*. Nous avons inclus au total 45 articles dont le processus de sélection est représenté sur le diagramme de flux (cf. **Figure 7**, page 95) ;(19,126–169) 8 de ces articles rapportaient une revue systématique avec méta-analyse et 1 rapportait un protocole de revue systématique. Le facteur d'impact médian des revues dans lesquelles étaient publiées les revues systématiques était de 2,5, un méthodologiste faisait partie des co-auteurs dans 33 articles (77%) et le cancer et la tuberculose étaient les maladies les plus souvent étudiées, respectivement dans 27 et 9 revues systématiques (cf. **Tableau 3**, page 96). Les objectifs non exclusifs des revues systématiques étaient de faire une synthèse des connaissances concernant la mesure ($n = 16$, 36%), les déterminants ($n = 31$, 69%) et/ou les conséquences ($n = 15$, 33%) d'un délai diagnostique. Dans 9 revues systématiques (20%), les auteurs ont mentionné qu'ils avaient suivi une *reporting guideline* pour les revues systématiques ou méta-analyses pour rapporter leurs résultats.

3.2 Processus d'évaluation de la qualité des études primaires

Vingt-trois articles (51%) rapportaient une évaluation de la qualité des études primaires. Dans 4 articles (9%), un processus d'évaluation de la qualité était mentionné dans le chapitre

Méthodes mais non détaillé dans l'article,(127,161) ni dans les références citées.(145,146) L'évaluation de la qualité était définie par les auteurs dans le chapitre « Méthodes » de 11 revues systématiques (48%) et l'objectif d'évaluation du risque de biais était mentionné dans 4 d'entre elles (36%) (cf. Appendix C, dans le manuscrit en annexe 2, page 215). Les auteurs de 2 revues systématiques ont mentionné le manque d'outil spécifique pour évaluer la qualité des études sur les délais diagnostiques.(19,160)

Parmi les 21 outils différents identifiés dans les revues systématiques, 6 (23%) étaient des *checklists*, des scores de qualité ou des *guidelines* pour le *reporting* déjà publiés et utilisés sans modification : STROBE, CONSORT* (*CONsolidated Standards Of reporting Randomized Trial*), *Aarhus statement*, *Guideline for critical appraisal* du *Critical appraisal skill program* pour les études transversales et les échelles de *Newcastle Ottawa* pour les études cas témoins et les études de cohortes.(47,48,118,170,171) Douze outils (46%) étaient « faits maison » et 8 (31%) autres étaient adaptés d'un outil utilisé dans une revue systématique précédemment publiée (n = 1), d'une *checklist* ou d'un score de qualité publiés (n = 5) ou des 2 (n = 2) (cf. **Tableau 3**, pages 96). Dans 4 revues systématiques, les outils utilisés étaient différents selon le *design* de l'étude primaire (études de cohorte, cas-témoins ou transversales).(147–149,151) Dans 6 revues systématiques, soit 26% des revues avec une évaluation de la qualité, les auteurs ont agrégé les résultats de leur évaluation de la qualité en un score global sans mentionner les propriétés psychométriques de ce score.(133,147–149,151,159) Pour les outils adaptés d'un outil existant ou élaborés spécifiquement pour les besoins de la revue systématique, la justification du choix des items composant l'outil n'était jamais rapportée. Le nombre de personnes qui avaient évalué la qualité et la stratégie pour obtenir un consensus en cas de désaccord était rapporté dans 7 revues systématiques.(133,140,144,150,153,154,159)

Les résultats de l'évaluation de la qualité étaient clairement rapportés dans 16 articles (72% de ceux rapportant une évaluation de la qualité en excluant le protocole), partiellement dans 2 (9%) et non rapportés dans 5 (23%) (cf. **Tableau 3**, pages 96). L'utilisation des résultats de l'évaluation était explicitée pour 17 revues systématiques (77% des revues systématiques avec évaluation de la qualité, à l'exclusion du protocole) : exclusion des études primaires de faible qualité (n = 6, 26%), modération de l'interprétation des résultats et/ou production de niveaux de preuve (n = 9, 39%) ou analyse de sensibilité (n = 3%).

3.3 Items utilisés pour évaluer la qualité des études primaires

Deux cent trente-deux items différents ont pu être identifiés dans les 21 outils utilisés pour évaluer la qualité des études primaires dans les revues systématiques. La classification de ces items est décrite dans la **Figure 7** (page 95). L'appendice D dans le manuscrit en Annexe 2 (page 215) donne également la répartition des items au sein des outils. Cent quarante-neuf items (64%) évaluaient la qualité du *reporting*, 85 (37%) le risque de biais, 14 (6%) les obstacles à la transportabilité et 13 (5%) les questions statistiques ; 1 item a été jugé ambigu (0,5%). Quarante-trois items (19%) ont été classés dans deux catégories, dont 21 items (9%) dans les catégories risques de biais et qualité du *reporting* et 6 items (3%) dans les catégories risques de biais et obstacle à la transportabilité. Parmi les 85 items évaluant un risque de biais, 25 évaluaient un risque de biais de sélection, 33 un risque de biais de classement* (dont 2 items issus du même outil -le *Newcastle Ottawa scale*- faisaient référence à la notion d'aveugle), (171) 25 un risque de biais lié à la présence de facteurs de confusion et 4 un risque de biais d'attrition*. Quatre items évaluaient à la fois un risque de biais de sélection et de classification. Dans 16 revues systématiques (36% de l'ensemble des revues systématiques incluses), au moins un des items utilisés évaluait un risque de biais ou d'obstacle à la transportabilité (cf. appendice D dans le manuscrit en annexe 2, page 215).

Vingt-six items (11%) ont été identifiés comme spécifiques des études sur les délais diagnostiques. Parmi eux, 6 évaluaient un risque de biais de classement lié à la définition ou au processus de mesure du délai, 24 évaluaient la qualité du *reporting* et 4 les 2 (cf. **Tableau 4**, pages 98).

3.4 Déterminants de l'évaluation adéquate de la qualité des études primaires

L'utilisation d'au moins un item concernant le risque de biais et/ou d'obstacle à la transportabilité était significativement associée avec la présence d'un méthodologiste parmi les co-auteurs ($p=0,003$) et avec un facteur d'impact élevé ($p=0,003$) mais nous n'avons pas mis en évidence d'association avec la période de publication (cf. **Tableau 5**, page 100).

4. Discussion

4.1 Résultats principaux

Les auteurs d'un faible nombre de revues systématiques confrontaient les études primaires à une grille d'évaluation des risques de biais et d'obstacles à la transportabilité. En effet, nous avons mis en évidence que 51% des revues systématiques étudiées rapportaient une évaluation de la qualité des études primaires qu'elles incluaient mais que seulement 36% évaluaient réellement le risque de biais ou d'obstacles à la transportabilité. Les auteurs des 45 revues systématiques étudiées ont utilisé 21 outils différents, le plus souvent élaborés pour les besoins de l'étude (46%) et sans justification du choix des items composant l'outil d'évaluation. Les items composant ces outils étaient d'ailleurs plus souvent dédiés à évaluer la qualité du *reporting* que le risque de biais ou d'obstacles à la transportabilité (64% vs 43%). De plus, parmi les revues systématiques évaluant la qualité seulement 23% rapportaient l'utilisation des résultats de cette évaluation. Enfin, un score de qualité était utilisé dans 23% des revues systématiques évaluant la qualité mais les propriétés psychométriques de ces

scores n'étaient jamais rapportées. Ces résultats contrastent avec les recommandations de PRISMA de spécifier si une évaluation des risques de biais a été menée et de la rapporter. L'utilisation d'un score de qualité est également controversée car il a été montré pour les essais randomisés contrôlés que de tels scores étaient peu reproductibles et n'évaluaient pas bien les risques de biais.(172)

4.2 Forces et limites de l'étude

Nous avons conduit cette revue systématique en suivant les recommandations sur la conduite de revue systématique du *Centre for Reviews and Dissemination* bien que notre revue systématique n'évaluait aucune intervention.(58) Nous avons construit notre stratégie de recherche sur celle utilisée dans le chapitre 1 de la partie I de ce travail doctoral ainsi que sur une publication antérieure qui avait fait preuve de sa sensibilité dans le domaine des délais diagnostiques.(19) Nous avons pu introduire un biais en limitant notre recherche aux articles en langue anglaise ou française mais la direction de ce biais est difficile à anticiper.(173) Les items utilisés pour évaluer la qualité ont été classés par 2 des auteurs en reprenant les listes de biais proposées pour les études observationnelles.(120,174) Cette classification était parfois difficile car, comme l'ont également constaté les auteurs de QUADAS-2, la frontière entre risque de biais et obstacle à la transportabilité est parfois étroite.(69) Par exemple, le processus de sélection de la population peut être source de biais de sélection (si les sujets sont inclus de façon systématique sur certaines caractéristiques associées au délai diagnostique) mais également d'obstacle à la transportabilité si la population incluse n'est pas comparable à la population cible*. Cependant, nous avons défini l'évaluation adéquate de la qualité méthodologique comme une évaluation du risque de biais et/ou d'obstacles à la transportabilité, les difficultés de classification n'influencent donc pas notre résultat principal de 36% de revues systématiques avec une évaluation adéquate. Nous avons fait le choix de

classer à la fois dans la catégorie biais et dans la catégorie *reporting* les items où le fait de ne pas rapporter un élément permettant d'évaluer un risque de biais était sanctionné par les auteurs (par exemple : « *Is the case definition adequate? a. yes, with independent validation ; b. yes, e.g., record linkage or based on self reports ;c. no description* »). Nous avons ainsi pu surévaluer le nombre d'items évaluant le *reporting* mais sans pour autant sous-évaluer ceux associés aux risques de biais. La classification dans la catégorie *reporting* était souvent plus facile du fait d'une terminologie propre avec par exemple l'utilisation par les auteurs des mots « *stated* », « *reported* » ou « *described* ».

4.3 Nécessité d'une clarification et d'outils spécifiques

Nos résultats sont comparables à ceux d'une revue systématique méthodologique menée par Deeks et coll. qui s'était intéressée à l'évaluation de la qualité des études primaires dans 511 revues systématiques d'études non randomisées. Deeks et coll. avaient montré que seulement 33% des auteurs des revues systématiques procédaient à une telle évaluation et seulement 20% évaluaient au moins un domaine se référant à la validité interne de l'étude.(120) Il est pourtant bien établi que l'absence d'évaluation ou une évaluation inadéquate des risques de biais des études primaires peut engendrer des conclusions erronées dans les revues systématiques.(122) Une des raisons pouvant expliquer ce manque d'évaluation adéquate des études primaires pourrait être le manque de prise de conscience de la part des auteurs et des éditeurs de l'importance de cette évaluation. En effet, nous avons ainsi montré que l'évaluation des risques de biais et d'obstacles à la transportabilité était plus fréquente lorsqu'un des co-auteurs était expert en méthodologie et lorsque le facteur d'impact du journal était élevé, c'est-à-dire lorsque les auteurs et les éditeurs étaient potentiellement plus avertis de l'importance de cette évaluation. Une autre raison pour expliquer le manque d'évaluation adéquate des études primaires pourrait être le manque d'outil spécifique. Cette

seconde hypothèse est confortée par le fait que plusieurs auteurs ont utilisé des *reporting guidelines* ou des outils d'évaluation des biais mis au point pour d'autres types d'études et que certains ont justifié cette attitude par l'absence d'outil spécifique disponible.(19,160)

Devant l'augmentation du nombre d'études publiées sur les délais diagnostiques et les conséquences potentielles des conclusions de ces études sur la pratique clinique et les politiques de santé publique, il est donc important qu'un outil spécifique d'évaluation de leur qualité méthodologique soit mis au point. Les *reporting guidelines* comme l'*Aarhus statement* et celle que nous avons développée dans le présent travail doctoral ne peuvent pas se substituer aux outils d'évaluation des risques de biais et d'obstacles à la transportabilité même si un *reporting* complet est un préalable indispensable à une telle évaluation.(48). Les outils existants pour l'évaluation des risques de biais et d'obstacles à la transportabilité des études observationnelles sont quant à eux trop génériques (174) et parfois non encore validés.(171,175) La construction d'un outil spécifique d'évaluation des risques de biais et d'obstacles à la transportabilité pourrait permettre d'éviter le « gaspillage » lié à la publication d'études biaisées et de moduler les résultats de façon appropriée afin de ne pas baser des actions correctrices sur des résultats d'études biaisées. (176)

4.4 Confrontation avec les risques de biais identifiés dans la première partie

Le faible nombre d'auteurs ayant correctement évalué les risques de biais et d'obstacles à la transportabilité dans les revues systématiques incluses a limité la réalisation de notre autre objectif qui était d'identifier ces risques et de les confronter à ceux que nous avons associés aux items dits critiques de notre *checklist* présentée dans le chapitre 1. Ainsi, nous avons identifié seulement 6 items permettant d'évaluer les risques de biais spécifiques aux études sur les délais diagnostiques. Ces items concernaient tous la définition du délai diagnostique ou son processus de mesure. Cependant, aucun item ne permettait d'évaluer les risques de

biais liés aux cas non diagnostiqués. Inversement, 5 items évaluaient le risque de biais d'attrition alors qu'aucun item de notre *checklist* ne permettait d'obtenir des informations précises sur les patients perdus de vue ou non analysés. Cependant le design des études sur les délais diagnostiques rend l'existence de perdus de vue peu probable car l'analyse se fait à partir des cas déjà diagnostiqués. Le manque de données nécessaires à l'évaluation du délai diagnostique pourrait entraîner une imprécision* de mesure si ces données étaient aléatoirement manquantes ou un biais d'attrition si ces données étaient manquantes en raison de certaines caractéristiques des participants. Ces patients non analysés pourraient être identifiés grâce à un diagramme de flux détaillé.

5. Conclusions

Dans ce chapitre, nous avons mis en évidence le manque d'évaluation adéquate de la qualité des études primaires dans les revues systématiques sur les délais diagnostiques ainsi que l'existence d'une confusion dans l'esprit des auteurs entre qualité du *reporting* et qualité méthodologique. Nous avons également mis en évidence une grande variabilité des outils utilisés et leur inadaptation fréquente ce qui souligne le besoin d'un outil spécifique.

Cette revue nous a permis d'identifier un risque de biais d'attrition que nous n'avions pas individualisé dans le chapitre 1 de cette partie I. Ce risque de biais sera pris en compte par une meilleure description des patients inclus non analysés par un diagramme de flux comme suggéré par certains experts dans le chapitre suivant. En effet, la prochaine étape recommandée dans la mise au point d'une *reporting guideline* spécifique aux délais diagnostiques est de faire critiquer, modifier et valider la *checklist* par des experts. Cette étape de validation a donc fait l'objet du chapitre 3 de cette partie I.

6. Figures et Tableaux

Figure 6. Stratégies de recherche utilisées pour la revue systématique des revues systématiques sur les délais diagnostiques.

PUBMED

Barrier [title] OR delay* [Title] OR interval* [Title] OR "time to diagnosis" [title] OR seeking [title] OR "early diagnosis" [title] OR "earlier diagnosis" [title] OR "timely diagnosis" [title] OR "timeliness" [title] OR "late diagnosis" [title]) AND diagnos* AND ((Meta-Analysis OR review [Publication Type]) OR meta-analysis as topic [MeSH] OR "systematic review" [title/abstract] AND (English [Language] OR French [Language])).

WEB OF SCIENCE

TI=(barrier OR delay* OR interval* OR "time to diagnosis" OR seeking OR early OR earlier OR timely) AND TS=diagnos* AND TI= ("systematic review" OR meta-analysis OR review)

Figure 7. Diagramme de flux des revues systématiques sur les délais diagnostiques incluses dans notre revue systématique méthodologique (au-dessus de la ligne en pointillés) et détail de l'évaluation de la qualité avec la classification des items (au-dessous de la ligne en pointillés).

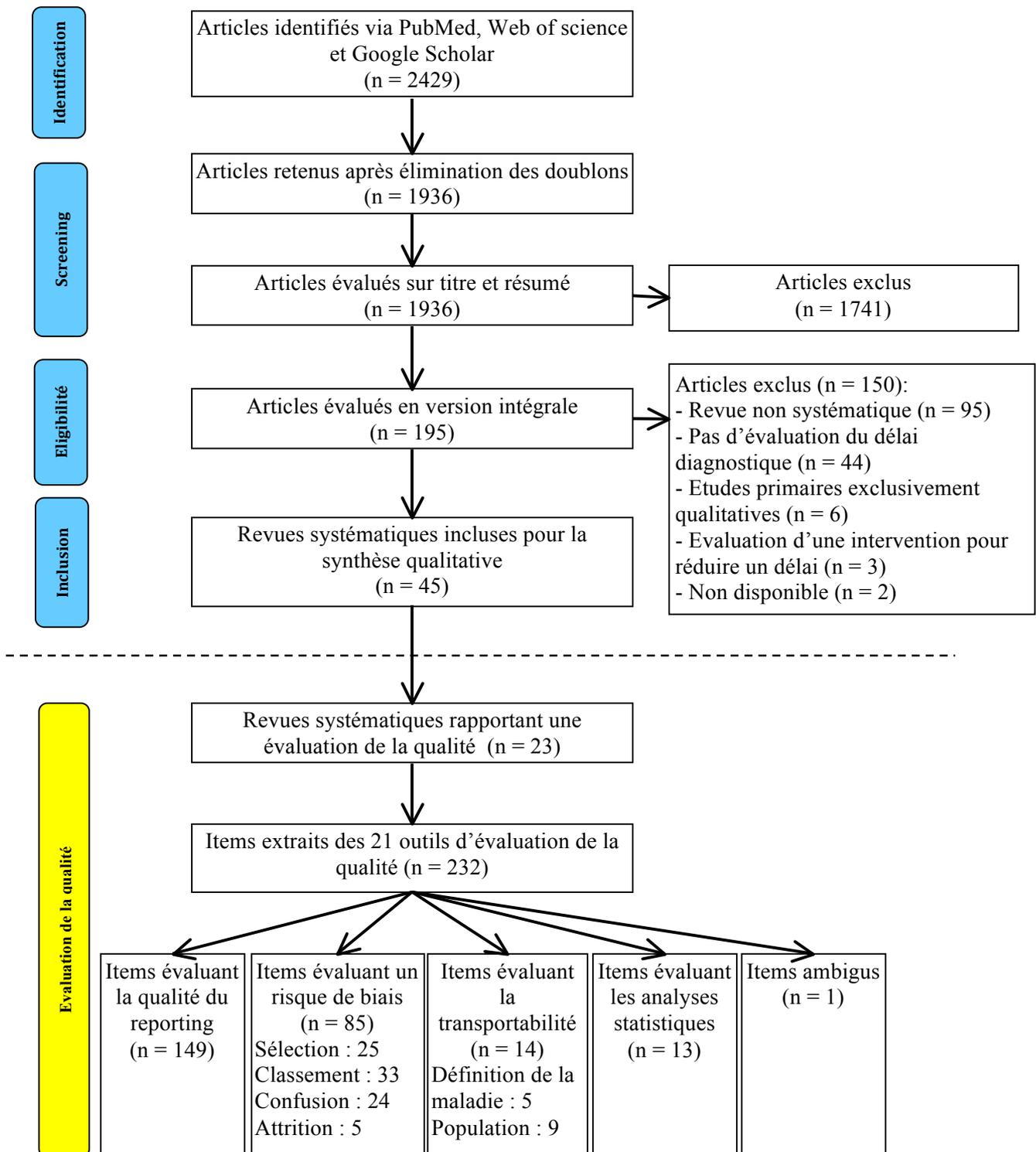


Tableau 3. Description des 45 revues systématiques incluses et des outils utilisés pour l'évaluation de la qualité des études primaires.

Auteur	Date	Maladie	Objectifs	Nb d'études incluses	Reporting guideline cité	Evaluation de la qualité des études primaires					
						Réalisé	Description	Risque de biais ou d'obstacle à la transportabilité	Au moins un item spécifique du délai	Résultats rapportés	Utilisation
Total	-	-	-	-	20% (9/45)	51% (23/45)	-	36% (16/45)	42% (9/19)	78% (17/22)	82% (18/22)
Abel §	2008	Hépatopathies chroniques malignes	D, C	56	Non	Non	-	-	-	-	-
Alhurishi*** §	2011	Cancer du sein	D	6	Non	Oui	Référence à une étude déjà publiée et à STROBE	Non	NA	Non	Exclusion des études de mauvaise qualité
Bradford	2009	Démence	M, D	40	Non	Non	-	-	-	-	-
Brasme §	2012	Cancer de l'enfant	M, D, C	98	PRISMA	Oui	Item isolé: ajustement sur des facteurs de confusion	Oui	Non	Non	Interprétation
Chatwin	2013	Cancer du poumon	D	58	Non	Non	-	-	-	-	-
Chen	2011	Tuberculose	M, D	14	PRISMA	Non	-	-	-	-	-
Chen*	2012	VIH	C	25	PRISMA	Non	-	-	-	-	-
Dang Tan §	2007	Cancer de l'enfant	M, D, C	23	Non	Non	-	-	-	-	-
Dubayova § #	2010	Cancer et infarctus du myocarde	D	15	Non	Oui	Liste d'items adaptés d'une checklist utilisée dans une précédente étude	Oui	Oui	Oui	Interprétation
Evenson §	2009	AVC	M	123	Non	Non	-	-	-	-	-
Evenson §	2001	AVC	M	48	Non	Non	-	-	-	-	-
Facione*	1993	Cancer du sein	M, D	12	Non	Non	-	-	-	-	-
Finnie* §	2011	Tuberculose	D	20	Non	Oui	Items isolés: type de recrutement et ajustement sur des facteurs confondants	Oui	Non	Non	Interprétation
Forrest*** §	2014	Cancer du poumon	D	-	Non	Oui	Liste d'items adaptés d'une étude précédente et du <i>Aar-hus Statement</i>	Oui	Oui	-	-
Fowler	2014	Cancer du sein	D	17	PRISMA	Non	-	-	-	-	-
Gomez*	2009	Cancer de la cavité buccale	C	9	MOOSE	Oui	Liste d'items adaptés de MOOSE	Oui	Non	Oui	Analyse de sensibilité
Goy §	2009	Cancer de la tête et du cou	C	27	Non	Non	-	-	-	-	-
Jayde	2009	Cancer de l'ovaire	D	56	Non	Non	-	-	-	-	-
Jensen §	2002	Cancer du poumon	C	16	Non	Non	-	-	-	-	-
Jones §	2014	Cancer du sein	D	18	PRISMA	Oui	Liste d'items adaptés d'une checklist publiée	Oui	Oui	Oui	Interprétation
Lefler***	2004	Infarctus du myocarde	D	48	Non	Oui	Référence aux critères de Cooper*	-	-	Partiellement	Interprétation
Lefler***	2002	Infarctus du myocarde	D	6	Non	Oui	Référence à une publication antérieure	-	-	Oui	Interprétation
Lethaby §#	2013	Cancer de l'enfant et du jeune adulte	M	32	Non	Oui	Liste d'items utilisés dans une précédente étude (cf. Macdonald) et <i>Aar-hus Statement</i>	Oui	Oui	Partiellement	Interprétation

Évaluation de la qualité des études primaires											
Auteur	Date	Maladie	Objectifs d'études incluses	Nb Reporting guideline cité	Réalisé	Description	Risque de biais ou d'obstacle à la transportabilité	Au moins un item spécifique du délai	Résultats rapportés	Utilisation	
Li*§,#	2013	Tuberculose	D	29	PRISMA	Oui	<i>Newcastle-Ottawa scale</i> pour les études cas/témoins et les cohortes et <i>Guideline for critical appraisal</i> pour les études transversales	Oui	Non	Oui	Non rapporté
Macdonald § #	2006	Cancer du tractus digestif haut	D	25	Non	Oui	Liste d'items élaborés spécifiquement selon le design de l'étude primaire	Non	Non	Oui	Exclusion des études de mauvaise qualité
Martins §	2013	Cancer	D	7	Non	Oui	Liste d'items adaptés de <i>Critical appraisal skills programme checklist</i> pour les études de cohortes	Oui	Oui	Oui	Interprétation
Mitchell § #	2008	Cancer colorectal	D	54	Non	Oui	Liste d'items utilisée dans une précédente étude (cf. Macdonald)	Non	Non	Oui	Exclusion des études de mauvaise qualité
Ngangro	2012	Tuberculose	D	73	Non	Non	-	-	-	-	-
Olsson §	2009	Cancer du poumon	M, C	49	Non	Oui	Liste d'items élaborés spécifiquement	Oui	Non	Oui	Exclusion des études de mauvaise qualité
Ramirez §	1999	Cancer du sein	D	23	Non	Oui	Liste d'items élaborés spécifiquement	Oui	Oui	Non	Exclusion des études de mauvaise qualité
Ramos* §	2008	Cancer colorectal	C	37	Non	Oui	Liste d'items adaptés d'une checklist publiée	Oui	Oui	Oui	Interprétation
Ramos**§	2007	Cancer colorectal	C	26	MOOSE	Oui	Liste d'items adaptés d'une checklist publiée	Oui	Oui	Oui	Exclusion des études de mauvaise qualité
Richards §	1999	Cancer du sein	C	87	Non	Oui	Liste d'items élaborés spécifiquement	Oui	Non	Non	Analyse de sensibilité
Scott	2006	Cancer de la cavité buccale	D	8	Non	Non	-	-	-	-	-
Seoane* § #	2012	Cancer de la tête et du cou	C	10	MOOSE	Oui	Liste d'items adaptés de <i>Newcastle-Ottawa scale</i>	Oui	Oui	Oui	Exclusion des études de mauvaise qualité
Sreeramareddy §	2014	Tuberculose	M, D	23	Non	Oui	Liste d'items élaborés spécifiquement	Oui	No	Oui	Non rapporté
Sreeramareddy	2009	Tuberculose	M, D	52	Non	Oui	Non rapporté	-	-	Non	Non rapporté
Storla §	2008	Tuberculose	M, D	58	Non	Non	-	-	-	-	-
Stoykoya §	1999	Oral cancer	D	22	Non	Non	-	-	-	-	-
Teuschl	2010	AVC	D	182	Non	Non	-	-	-	-	-
Thompson §	2009	Cancer de l'intestin	M, D, C	80	Non	Non	-	-	-	-	-
Unger §	2009	Cancer du sein	M, D, C	96	Non	Non	-	-	-	-	-
Villeneuve	2013	Arthrite inflammatoire	M	47	Non	Non	-	-	-	-	-
Vinas	2014	Cancer du poumon	M, C	13	Non	Non	-	-	-	-	-
Yang §	2014	Tuberculose	D	137	Non	Oui	STROBE et CONSORT	-	Non	Oui	Non rapporté

M: mesure, D: déterminants, C: conséquences, AVC: accident vasculaire cérébral, VIH: virus de l'immunodéficience humaine

§: revue systématique avec un méthodologiste dans les co-auteurs

* revue systématique avec méta-analyse, ** protocole, *** revue systématique où l'évaluation de la qualité est mentionnée mais non détaillée

#: revue systématique agrégeant les résultats de l'évaluation de la qualité en un score de qualité

Dans les revues systématiques de Lethaby, Mitchell et Macdonald les mêmes outils étaient utilisés (6 outils différents selon le *design* de l'étude primaire)

Tableau 4. Détails des 26 items spécifiques aux études sur les délais diagnostiques utilisés lors de l'évaluation de la qualité des études primaires dans les 23 revues systématiques.

Revue systématique Premier auteur (reference)	Item	Catégorie
Dubayova (133)	<i>1. Definition and operationalization of patient's delay. Patient's delay was defined exactly using a disease specific cut off point which divided the sample into delayers and non-delayers (+), or patient's delay was defined as a continuous variable (-)</i>	<i>Reporting</i>
Forrest (138)	<i>Are the beginning and end points of this interval clearly defined?</i>	<i>Reporting</i>
Lethaby (147) (Aarhus statement -48-)	<i>Are there precise, transparent and repeatable definitions, and is the complexity of time points such as the date of first symptom and date of first presentation addressed?</i>	<i>Biais de classement et reporting</i>
	<i>Do the researchers refer to a theoretical framework underpinning definition of this time point?</i>	<i>Reporting</i>
	<i>Is there a discussion of the different biases influencing measurement of this time point?</i>	<i>Reporting</i>
	<i>Do the researchers discuss the complexity of the date of first presentation?</i>	<i>Reporting</i>
	<i>Do the researchers discuss the nature of the referral and provide adequate detail – for example, whether it was for investigation or consultation by a colleague in secondary care?</i>	<i>Reporting</i>
	<i>Do the researchers use an existing hierarchical rationale for the date of diagnosis measurement?</i>	<i>Reporting</i>
	<i>Do the questions on time points and/or intervals clearly derive from stated definitions?</i>	<i>Biais de classement</i>
	<i>Do researchers acknowledge the need for theoretical validation and make reference to the theoretical framework(s) underpinning measurement and analysis of the time points?</i>	<i>Reporting</i>
	<i>Is there discussion of the different biases influencing measurement of the time points, such as how and when the question is asked and who is being asked?</i>	<i>Reporting</i>
	<i>Is the timing of the interview in relation to the date of diagnosis provided?</i>	<i>Biais de classement</i>
	<i>Is there a clear and precise description of how case-note data were used to ascertain time points with an acknowledgment of limitations of such data?</i>	<i>Reporting</i>
Martins (150)	<i>Was diagnostic intervals measured to Olsen and colleagues framework?</i>	<i>Biais de classement et reporting</i>

Revue systématique Premier auteur (reference)	Item	Catégorie
Ramirez (154)	<i>Studies were eligible for inclusion if they were based on a clear description of a discrete interval of delay— Reporting by patients or providers. We defined delay by patients as the time between the onset of initial symptoms and the first medical consultation. Delay by providers was any discrete interval between first medical consultation and treatment. We excluded studies if they were based on unclear definitions or looked at periods of delay by patients and providers that were indivisibly combined</i>	Reporting
Ramos 2007 (156)	<i>Interval of time definition 1. From beginning of symptoms to diagnosis; 2. From beginning of symptoms to treatment; 9. Not mentioned</i> <i>Manner of measuring the interval of time 1. Global; 2. Only a fragment (patient's delay or health system's delay)</i> <i>Measurements of the effects of the interval of time 1. Average; 2. Median; 3. <1 months; 4. <2 months; 5. <3 months; 6. <4 months; 7. <5 months; 8. <6 months; 9. Quartiles</i>	Reporting Reporting Reporting
Ramos 2008 (155)	<i>Definition of delay 1. From the onset of symptoms to diagnosis; 2. From the onset of symptoms to treatment</i> Reporting <i>Measures of the time interval 1. Overall; 2. Only a portion (delay due to patient, delay due to health system) Reporting</i> <i>Measures of the delay 1. Means; 2. Medians; 3. <1 month; 4. <2 months; 5. <3 months; 6. <4 months; 7. <5 months; 8. <6 months; 9. Quartiles</i> Reporting <i>Measurement of the delay by disease stage 1. Comparison of means or medians; 2. Frequencies; 3. Absolute Reporting figures; 4. Correlations; 5. Graphics; 9. Not given</i>	Reporting Reporting Reporting Reporting
Seoane (159)	<i>The first item assessed the source of the outcome date. One point was given if the date of death was ascertained through clinical history or death certificate and zero point if otherwise... (when the information reporting on a specific item was not provided by the authors, we graded this item as 'no')</i> <i>The fourth item assessed the definition of delay. One point was given if inception was clearly defined and zero if not... (when the information on a specific item was not provided by the authors, we graded this item as 'no')</i>	Biais de classement et reporting Biais de classement et reporting
Streeramreddy (160)	<i>Diagnostic delays separately reported as patient and health systems (y/n)</i> <i>Risk factors for diagnostic delays were reported (y/n)</i>	Reporting Reporting

Tableau 5. Déterminants de l'évaluation adéquate de la qualité (au moins un item évaluant un risque de biais ou d'obstacle à la transportabilité) des études primaires dans les revues systématiques sur les délais diagnostiques (n=45).

	Présence d'au moins un item évaluant un risque de biais ou d'obstacle à la transportabilité (n=16)	Pas d'évaluation du risque de biais ou d'obstacle à la transportabilité (n=29)	p
Période de publication, n (%)			
Avant 2010*	6 (24)	19 (76)	0,07
Après 2010	10 (50)	10 (50)	
Au moins un méthodologiste** parmi les co-auteurs, n (%)			
Non	1 (9)	10 (91)	0,03
Oui	15 (45)	18 (55)	
Facteur d'impact du journal, médiane [EIQ]	3,9 [2,3-8,6]	2,3 [1,8-3,7]	0,03

*un an après la publication de PRISMA (59)

**biostatisticien, épidémiologiste ou spécialiste de santé publique (la présence d'un méthodologiste n'était pas évaluable pour 1 article)

CHAPITRE 3

Affinement et validation d'une *checklist* pour le *reporting* des études évaluant un délai diagnostique

Ce chapitre fait l'objet de la préparation du manuscrit suivant:

Reporting studies on time to diagnosis and diagnosis delay: refinement and international validation of a guideline.

Elise Launay, Jérémie F. Cohen, Patrick Bossuyt, Pierre Buekens, Jon Deeks, Timothy Dye,

Richard Feltbower, Andrea Ferrari, Mikael Kramer, Mariska Leeftang, David Moher,

Karel Moons, Erik von Elm, Philippe Ravaud, Martin Chalumeau.

1. Introduction

Nous avons vu dans le chapitre 1 de cette partie I que le *reporting* des études sur le délai diagnostique souffrait de faiblesses notamment sur les points que nous avons considérés comme critiques (cf. introduction de la partie I). La revue systématique du chapitre 2 nous a permis de conforter ce constat de méconnaissance des risques de biais et d'obstacles à la transportabilité justifiant ainsi la construction d'une *reporting guideline* spécifique. Devant la difficulté de faire une démonstration des biais modifiant la mesure du délai diagnostique d'une part et la faible contribution de la revue systématique menée au chapitre 2 pour légitimer le choix des items de la *checklist* précédemment proposée d'autre part (cf. **Tableau 1**, pages 75), les items devaient être discutés avec des experts. Une telle discussion est recommandée par le réseau EQUATOR à l'étape de validation d'une *reporting guideline*.⁽⁴⁵⁾ Il est suggéré que cette validation repose sur la soumission de la *checklist* à un comité d'experts (méthodologistes, épidémiologistes, statisticien, éditeur de revues ou cliniciens-chercheurs ayant déjà mené et rapporté le type d'étude concernée) qui statuent sur la pertinence des items et leur intelligibilité.^(45–47,118)

L'objectif de ce chapitre était donc d'affiner et de valider la *checklist* proposée dans le chapitre 1.

2. Méthodes

2.1 Méthodologie générale

La validation de la *checklist* a été faite en 2 étapes : une première validation par un comité restreint d'experts (comité scientifique) et une deuxième étape de validation par un panel plus grand (panel de validation) constitué d'auteurs d'articles ayant rapporté une revue systématique ou une étude originale sur un délai diagnostique (cf. **Figure 8**, page 111). Nous avons déclaré le développement de cette *checklist* auprès du réseau EQUATOR où un résumé

du protocole a été publié.(177)

2.2 Checklist préliminaire

Les items génériques (objectif de l'étude, dates, lieux...) et communs à d'autres *reporting guidelines* comme STROBE, STARD ou CONSORT n'ont pas été soumis au comité scientifique lors de l'étape de première validation. Une *checklist* préliminaire de 25 items en langue anglaise a été soumise au comité scientifique. Cette liste a été constituée en reprenant les 21 items identifiés comme spécifiques des études sur les délais diagnostiques et considérés comme potentiellement associés à des risques de biais ou d'obstacles à la transportabilité au cours du chapitre 1 de cette partie. Nous y avons adjoint 2 autres items, également issus de la *checklist* établie dans le chapitre 1, qu'il nous semblait important d'inclure ainsi que 2 items issus de l'*Aarhus statement* qui détaillent la façon de rapporter les méthodes utilisées pour mesurer le délai diagnostique.(48) Nous n'avons pas ajouté d'item directement associé au *reporting* des perdus de vue ou des patients non analysés (risque de biais d'attrition) car il nous a semblé que ce biais pouvait être évalué grâce à un *reporting* précis du diagramme de flux de la population d'étude.

2.3 Première étape de validation

Nous avons identifié 16 experts dans le domaine de la méthodologie de la recherche (ayant déjà participé à la construction d'une *reporting guideline*), de l'épidémiologie (professeurs dans une unité d'épidémiologie) ou des délais diagnostiques (cliniciens chercheurs ayant déjà publié sur les délais diagnostiques) ou de l'édition scientifique et médicale (éditeur en chef). Nous les avons contactés individuellement par courriels. Une fois leur accord de participation confirmé, nous leur avons demandé de répondre par « oui » ou « non » à 2 questions pour chacun des 25 items de la *checklist* préliminaire : « est-ce que l'item est pertinent ? » et « est-

ce que l’item est intelligible ? », un espace de commentaires libres était également proposé avec chacune de ces 2 questions. Après chaque sous-section de la *checklist* (« *identification of study population* », « *definition and measurement of time to diagnosis* », « *determinants and consequences analyses* », « *results* », « *study limitations* », « *clinical application* »), il était également demandé aux experts s’ils souhaitaient ajouter un item manquant à la section. A l’issue de ce premier tour d’expertise, nous avons regroupé de façon anonyme dans un document les réponses avec les effectifs de « oui » et de « non » et les commentaires des experts et nous leur avons proposé une version révisée de la *checklist* prenant en compte leurs commentaires. Il leur a alors été demandé de donner leur appréciation globale sur cette version révisée de la *checklist* et éventuellement de faire des propositions de modifications via un forum de discussion. De nouvelles modifications ont alors été faites jusqu’à obtenir l’accord de tous les experts. La version révisée de la *checklist* a alors été soumise à une seconde étape de validation.

2.4 Seconde étape de validation

Pour la seconde étape de validation, le panel a été constitué de premiers et derniers auteurs des revues systématiques incluses dans chapitre 2 et d’auteurs des études primaires incluses dans le chapitre 1 de la partie I de la thèse. Les auteurs ont été invités successivement (d’abord les auteurs de revues systématiques puis les auteurs d’études primaires) à participer à une enquête en ligne via le site SurveyMonkey (<http://www.surveymonkey.com>). Comme pour la première étape de validation, il a été demandé à chaque membre de ce panel de validation de répondre aux questions « est-ce que l’item est pertinent ? » et « est-ce que l’item est intelligible ? » par « oui » ou par « non » ou par « je ne sais pas » (cette modalité étant possible uniquement pour la question concernant la pertinence) et de donner leurs commentaires éventuels. Afin de faciliter la compréhension des items spécifiques de cette

checklist, nous les avons illustrés d'exemples et/ou d'explications inspirés des commentaires et questions des experts du comité scientifique lors de la première étape de validation.

2.5 Version finale de la *checklist*

A l'issue de la deuxième étape de validation, nous avons procédé à une synthèse des réponses sous la forme de fréquence de réponse ainsi que des commentaires. Il était alors prévu de modifier les items de la *checklist* en fonction des commentaires sur l'intelligibilité et de soumettre au comité scientifique les items ayant recueilli moins de 50% (seuil arbitraire) de « oui » pour décider de leur exclusion ou non de la *checklist* finale. Nous en sommes actuellement à ce stade de validation. Une version finale de la *checklist* sera ensuite proposée à l'ensemble des experts du comité scientifique pour une validation sous forme d'un manuscrit original.

3. Résultats

3.1 Première étape de validation

Parmi les 16 experts sollicités, 1 n'a jamais répondu, 2 ont décliné l'invitation et 2 n'ont pas retourné leurs réponses malgré de multiples relances. Parmi les 11 experts ayant accepté de participer, un a renvoyé des commentaires généraux sur l'ensemble de la *checklist* et 10 ont répondu items par items. Ces 10 experts venaient de 6 pays différents (Canada, Etats-Unis, Italie, Pays-Bas, Royaume-Uni et Suisse) Au total, 20 des 25 items ont été évalués comme pertinents pour le *reporting* des études sur les délais diagnostiques par 7 experts ou plus dont 13 par 9 ou 10 experts. Le consensus sur l'intelligibilité était également bon (19 items jugés intelligibles par 7 experts ou plus) avec des commentaires libres des experts proposant une reformulation.

Trois items ont été supprimés car ils avaient obtenu moins de 5 « oui » à la question de la pertinence (« *Report if delayed diagnosis was too frequent or time to diagnosis too long** », « *Propose plan of action to avoid delayed diagnosis delay* » et « *Discuss the sensitivity, specificity and feasibility of this plan* »). Tous les items ont été reformulés. Dix items ont été fusionnés 2 par 2 (items 5 et 9, 13 et 14, 8 et 11, 22 et 23, 24 et 25, **Tableau 1**, page 75) en 5 items. Devant les remarques de plusieurs experts à ce sujet, les items génériques ont été réintégrés et un diagramme de flux (cf. **Figure 9**, page 113) a été proposé. Ce diagramme de flux nous permettait également de mieux faire apparaître le risque de biais d'attrition. Une première révision de la *checklist* contenant 26 items très développés avec des exemples illustratifs pour certains a alors été soumise aux 11 experts via un forum de discussion. Tous les experts ont participé à ce forum de discussion et se sont accordés sur le fait qu'il fallait plutôt privilégier une version courte de la *checklist* accompagnée d'un document explicatif. A l'issue de ces discussions, les items ont encore été remaniés. Une version révisée courte de la *checklist* comprenant ces modifications a lors été validée par l'ensemble des 11 experts. C'est cette version accompagnée d'un document explicatif, s'attachant à illustrer ou expliciter les items spécifiques de la *checklist* (cf. **Figure 10**, page 114), qui a été soumise au panel plus large pour la deuxième étape de validation.

3.2 Deuxième étape de validation

Avertissement : nous présentons les résultats préliminaires après 2 relances des auteurs de revues systématiques et une première invitation des auteurs d'études primaires (au 3 octobre 2015).

Quatre-vingt dix-huit auteurs de revues systématiques et quatre-vingt sept auteurs d'études primaires ont été sollicités. Vingt-neuf auteurs de revues systématiques (30%) et sept auteurs d'études primaires (8%) ont répondu à au moins une des questions. Parmi les 36 auteurs répondants, vingt-huit ont répondu à l'ensemble des questions concernant la pertinence des

items. Tous les items ont été jugés pertinents par 24 experts ou plus (86% des répondants) et 18 des 21 items ont été jugés pertinents par plus de 90% des répondants. Tous les items ont été jugés intelligibles par 23 experts ou plus (85% des répondants) et 15 items ont été jugés lisibles par plus de 90% des répondants. Les commentaires des experts visaient dans l'ensemble à demander des explications plus détaillées dans le document joint notamment pour les items 1, 9 et 19B (**Tableau 6**, page 112). Vingt-trois des vingt-six auteurs ayant répondu à cette question ont mentionné qu'ils utiliseraient potentiellement la *checklist* pour la rédaction d'un futur article. Trois auteurs ont commenté que l'*Aarhus statement* leur semblait plus adapté à leur champ d'exercice.

3.3 Version finale de la *checklist*

Nous n'avons pas modifié la *checklist* à ce stade par rapport à la version soumise aux experts de la deuxième étape de validation (**Tableau 6**, page 112) car tous les items ont été jugés pertinents et intelligibles par une large majorité des experts. Nous allons ajouter des précisions dans le document explicatif puis soumettre ce document et la *checklist* pour une ultime validation auprès du comité scientifique.

4. Discussion

Nous proposons une *reporting guideline* de 21 items pour guider les auteurs dans la rédaction des études primaires investiguant un délai diagnostique. Ces items ont été validés par des experts internationaux dans le domaine de la méthodologie de la recherche, de l'épidémiologie ou des délais diagnostiques et des auteurs de revue systématique sur les délais diagnostiques. Cette *reporting guideline* insiste sur les points considérés comme clés avec 6 items dédiés au *reporting* du processus de sélection de la population avec une proposition de diagramme de flux (items 3-6, 14-15 du **Tableau 6**, page 112), 6 items dédiés au *reporting* de

la définition et de la mesure du délai diagnostique et éventuellement du retard diagnostique (items 7-10, 16-17).

La méthodologie utilisée pour mettre au point cette *checklist* a suivi les recommandations du réseau EQUATOR avec un travail préalable d'analyse de la littérature et de réflexion sur les risques de biais et d'obstacles à la transportabilité de ce type d'étude. Cependant, nous n'avons pas pris l'option pour la validation de la *checklist* de la discussion lors d'une réunion présentielle, comme cela a été fait pour STARD, STROBE ou CONSORT, pour des raisons de faisabilité.(47,118,178) La proposition d'un premier tour de sondage anonyme auprès du comité scientifique a pu permettre une expression plus indépendante des différents experts sans risque d'influence ou d'argument d'autorité. Nous avons choisi de procéder à une validation en 2 étapes afin de renforcer dans un premier temps la légitimité des items sélectionnés en les soumettant à des experts dans le domaine de la méthodologie. L'utilisation d'une méthode de validation par un comité restreint d'experts avait également été choisie pour la mise au point de la *Research Triangle Institute bank of bias*. En effet, nous n'avons pu faire la démonstration de l'existence de biais liés à certains items. Comme nous l'avons évoqué précédemment, une telle démonstration nécessiterait de pouvoir comparer les résultats d'un grand nombre d'études ne variant que sur le point méthodologique associé au risque de biais à démontrer (caractère systématique ou non de la procédure diagnostique, ou évaluation du délai diagnostique par des experts à l'aveugle ou non de l'issue par exemple) et dans un contexte de soin similaire. Etudier la variabilité des délais calculés dans plusieurs études en fonction de variations méthodologiques ponctuelles est donc probablement impossible. La deuxième étape de validation nous a permis de confronter les items révisés à un public moins averti et peut-être plus proche des connaissances habituelles en méthodologie d'un auteur susceptible d'utiliser notre *guideline*. Cette étape de validation nous a également permis de vérifier que notre document facilitait l'utilisation de la *reporting*

guideline puisque sur les 28 répondants à ce jour, tous les items étaient jugés intelligibles par au moins 85% des répondants.

Le réseau EQUATOR propose que les items soient validés par une méthode de consensus type Delphi*. La première étape de validation est inspirée de cette méthode avec un premier tour anonyme, à l'issue du premier tour une large majorité des items (20/25) était jugée pertinente par 7 experts ou plus et seulement 3 étaient jugés non pertinents par au moins la moitié des experts. Les principales suggestions de modifications concernaient l'intelligibilité. Plutôt que de resoumettre un à un les items selon la méthode Delphi, nous avons alors choisi une validation globale à l'issue d'une discussion ouverte et non anonyme au fur et à mesure des modifications des items. Nous avons pu ainsi obtenir un accord de tous les experts ce qui n'aurait pas forcément été le cas avec une méthode Delphi. En effet l'expérience de Philipps et coll qui avait soumis à une validation par méthode Delphi les items composant une *reporting guideline* pour les études évaluant les interventions d'éducation à la pratique fondée sur les preuves (*evidence based*) (GREET) avait échoué à obtenir le taux de consensus fixé *a priori* de 80%. Une des limites de notre méthode par rapport à la méthode Delphi est cependant que certains experts étaient plus actifs que d'autres dans la discussion.

Lors des discussions avec le comité scientifique plusieurs auteurs nous ont suggéré de proposer une *conducting guideline* pour aider les investigateurs à construire une étude sur les délais diagnostiques plutôt qu'une *reporting guideline*. En effet, le concept des études sur les délais diagnostiques peut paraître difficile à appréhender pour les spécialistes des essais randomisés ou des études observationnelles au *design* plus conventionnel de type étude de cohorte ou cas/témoins. Le document explicatif que nous avons joint à la *reporting guideline* (cf. **Figure 10**, page 114) permet de préciser certaines notions propres aux études sur les délais diagnostiques avec des exemples comme la définition du délai et des sous-délais ou

l'importance du processus de sélection avec l'exemple de diagramme de flux (cf. **Figure 9**, page 113). Ce document ne se substitue pas à une *conducting guideline* mais peut permettre de sensibiliser les investigateurs d'études sur les points méthodologiques importants à prendre en considération lors de construction d'une étude sur les délais diagnostiques.

5. Conclusions et perspectives

Avec cette *reporting guideline*, nous espérons pouvoir sensibiliser les auteurs d'étude investiguant un délai diagnostique aux points importants à rapporter dans les articles et ainsi en améliorer la qualité. Cette *reporting guideline* devra être optimisée après une phase d'utilisation « pilote » comme cela a été fait pour d'autres *reporting guidelines* comme CONSORT.(118) Le processus de construction de cette *guideline* nous a permis de mettre en exergue des points méthodologiques potentiellement à risque de biais et d'obstacles à la transportabilité pouvant compromettre la fiabilité et l'applicabilité* des résultats des études sur les délais diagnostiques. Nous avons donc appliqué cette réflexion méthodologique à l'étude des soins suboptimaux dans la prise en charge initiale des IBS de l'enfant qui a fait l'objet de la partie II de cette thèse.

6. Figures et tableaux

Figure 8. Déroulement de la validation de la *checklist* pour le *reporting* des études primaires sur les délais diagnostiques.

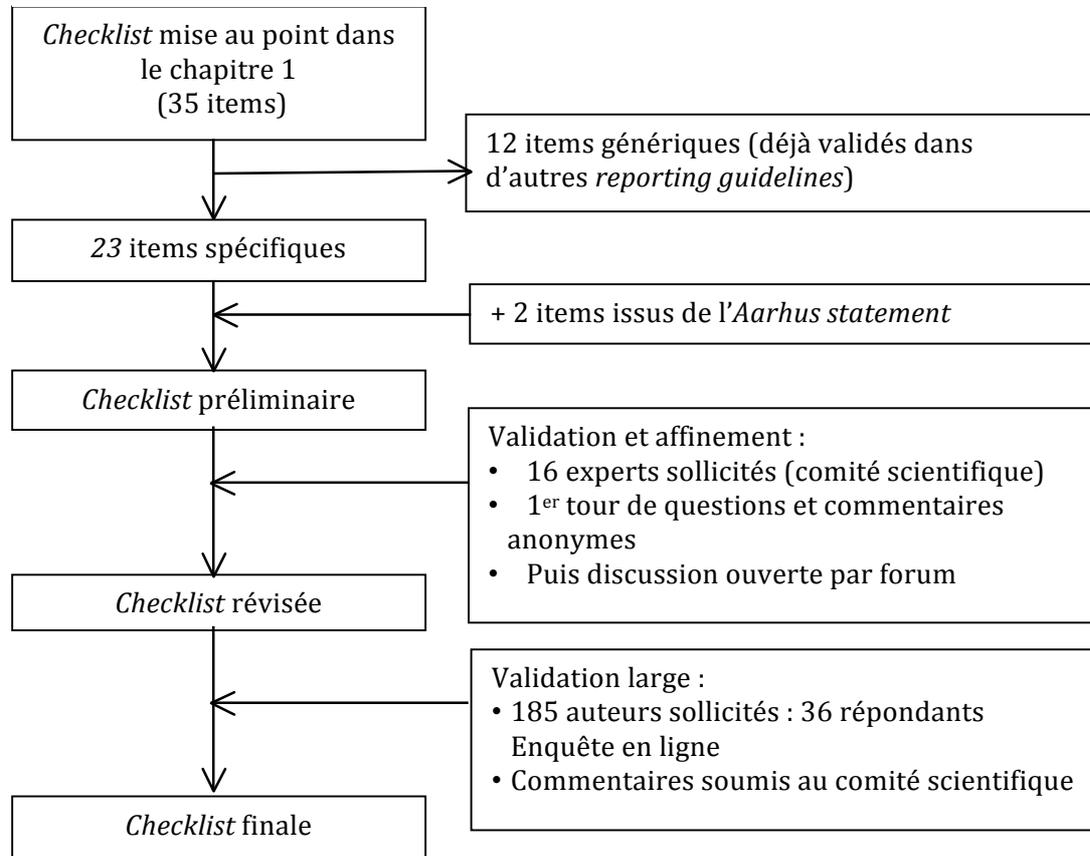


Tableau 6. Version finale de la *checklist*.

Items
Introduction
1. Identify the article as a study on time to diagnosis
2a. †Explain the scientific background and rationale for the study
2b. †State specific objective(s)
Methods
3. †Describe the setting, location(s), and relevant dates, including periods of recruitment
4. State eligibility criteria of participants (i.e., inclusion and exclusion criteria, especially diagnostic criteria)
5. Describe the source population [i.e., the population with signs and symptoms that usually trigger healthcare professionals to initiate the diagnostic procedure(s)] and how the participants were identified within it
6. *State how known sub-groups of participants with an inherent individual risk of short or long time to diagnosis were handled (e.g., by sub-group analysis, exclusion)
7. Define time points (e.g., time of first signs and symptoms, time of diagnosis) and *time intervals (e.g., patient or physician intervals)
8. State the methods used to collect study data
9. Describe how time points were assessed (e.g., number of assessors, their qualifications)
10a. *If the study aims to evaluate associations between participant characteristics and time to diagnosis, state whether assessors of time to diagnosis were blinded to these characteristics
10b. *If the study aims to evaluate associations between time to diagnosis and participant health outcomes (e.g., survival), state whether assessors of time to diagnosis were blinded to these outcomes
11. Describe the statistical methods used, including whether time to diagnosis was analysed as a continuous or categorized variable (e.g., delayed versus not delayed)
12. *If the study aims to evaluate associations between time to diagnosis and other factors (e.g., participant characteristics or health outcomes), describe which confounders were considered and how they were chosen, measured and analysed
13. †Give a rationale for the sample size
Results
14. Report the number of individuals at each step of the selection process between the source population and participants and provide a flowchart (see example). Give reasons for non-participation at each stage
15. †Report demographic and clinical characteristics of participants
16. Report the distribution of time to diagnosis
17. *If associations between time to diagnosis and other factors (e.g., participant characteristics or health outcomes) are described, report measures of association and their precision (e.g., confidence intervals)
Discussion
18. †Summarize key results with reference to study objectives and discuss their potential clinical implications
19a. Discuss sources of potential bias, including bias due to the selection of participants from the source population (e.g., undiagnosed cases) and to the assessment of time points
19b. *If association between time to diagnosis and survival was studied, discuss possible lead-time bias

Les items marqués par une croix « † » sont communs à d'autres *reporting guidelines* (CONSORT, STARD, STROBE), les items marqués par un astérisque « * » sont optionnels selon les objectifs de l'étude

Figure 9. Diagramme de flux proposé pour détailler le processus de sélection des participants dans une étude sur un délai diagnostique (cf. item 14 du Tableau 6).

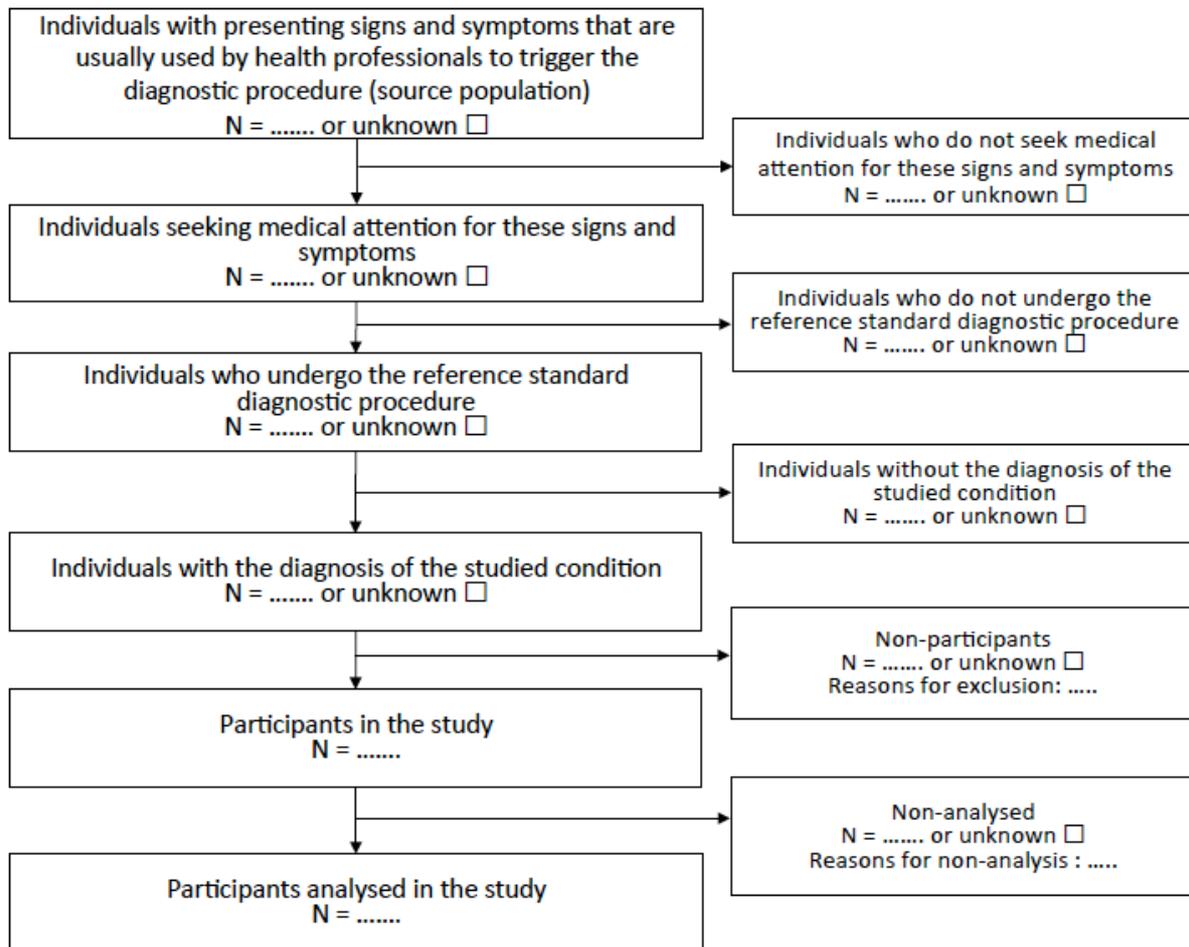


Figure 10. Explication des items spécifiques fournis avec la *checklist* pour guider les auteurs dans le *reporting* des études sur les délais diagnostiques.

ITEM 4 (Methods section). State eligibility criteria of participants (i.e., inclusion and exclusion criteria, especially diagnostic criteria).

Example: “All patients under the age of 18 years who presented with imported malaria to the emergency departments of the seven participating hospitals were included. Imported malaria was defined as the presence of *Plasmodium trophozoites* in blood after a stay in a malaria-endemic area, regardless of the clinical symptoms present. (...) If so, the search for *Plasmodium trophozoites* in thick and/or thin blood smears was ordered.” (Chalumeau, Eur J Clin Microbiol Infect Dis 2006)

Explanation: Studies of time to diagnosis by definition focus on patients with a positive diagnosis. Authors must detail the diagnosis criteria used in order to help readers to evaluate whether they are accurate.

ITEM 5 (Methods section). Describe the source population [i.e., the population with signs and symptoms that usually trigger healthcare professionals to initiate the diagnostic procedure(s)] and how the participants were identified within it.

Example: “In the participating emergency departments, the routine protocol for children with fever was to ask if they had recently visited a malaria-endemic area. If so, the search for *Plasmodium trophozoites* in thick and/or thin blood smears was ordered.” (Chalumeau, Eur J Clin Microbiol Infect Dis 2006)

Explanation: In this item, authors should describe the population presenting signs and symptoms that are usually used by healthcare professionals to trigger the diagnostic procedure (in the present example: children with fever and having recently visited a malaria-endemic area). They should also state whether the surveillance of these signs and symptoms was performed routinely or *ad hoc* (in this example, authors stated that all the children with fever and having visited an endemic area were tested for malaria during the study period). They should also state the mechanism for identifying participants among the source population and how participants were sampled (consecutive or random sampling, other).

ITEM 6 (Methods section). *State how known sub-groups of participants with an inherently short or long time to diagnosis were handled (e.g., by sub-group analysis, exclusion).

Example 1: “Clinical, diagnostic, and family history information is then reviewed by a clinical review committee to classify the patients into one of the following case definitions: definite, probable, possible, asymptomatic, female, or not Duchenne or Becker Muscular Dystrophy. For this study we included definite or probable cases of Duchenne Muscular Dystrophy without known family history of Duchenne Muscular Dystrophy prior to birth.” (Ciafaloni, J Pediatr 2009)

Example 2: The clinical charts of 64 consecutive patients who presented to the Memorial Sloan-Kettering Cancer Center with newly diagnosed retinoblastoma between November 11, 1993, and January 14, 1998, were reviewed. Seven patients with a family history of retinoblastoma were excluded from the results. (Butros, Pediatrics 2002)

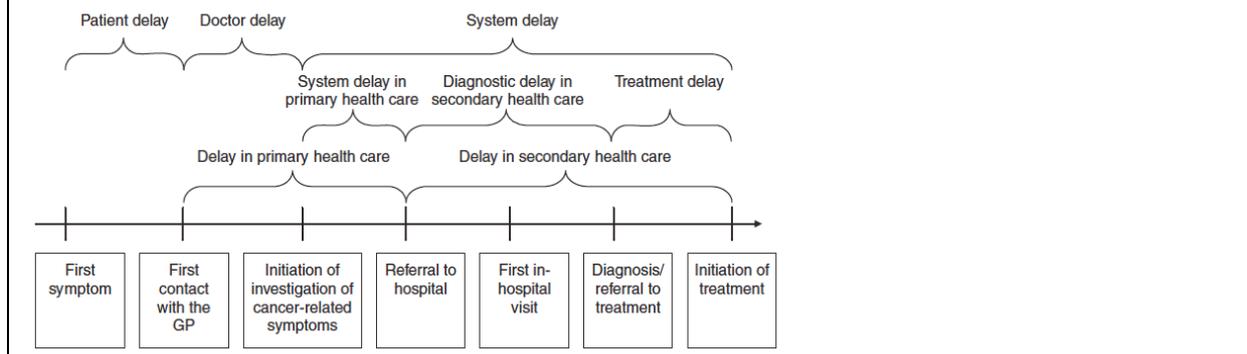
Explanation: In the first example, the studied condition is an X-linked disease; females (at risk of longer time to diagnosis) were identified as a separate group. In the first and the second example, people with known family history (at risk of shorter time to diagnosis) were excluded.

ITEM 7 (Methods section). Define time points (e.g., time of first signs and symptoms, time of diagnosis) and *time intervals (e.g., patient or physician intervals).

Example: “*Patient delay* was defined as the *time between fever onset and the first medical consultation*. *Doctor delay* was the *time between the first consultation and the parasitological diagnosis*. The *total delay* was the sum of the two. We defined a *late medical diagnosis* as a doctor delay of *longer than 12 h*.” (Chalumeau, Eur J Clin Microbiol Infect Dis 2006)

Explanation: Authors should detail how they defined time points (in this example, using a single symptom – fever) in order to help readers evaluate whether they are accurate. With several possible signs and symptoms, authors should list all of them. Sub-divisions composing the delay can also be defined, depending on the objective of the authors. Subdividing the diagnostic delay may help target corrective actions (health education and/or medical education) by measuring the role of each participant in the delay (patients/parents, physicians/healthcare system). The use of a framework may be useful to report these subdivisions.

Example of a framework (Olesen, Br J Cancer 2009)



ITEM 8 (Methods section). State the methods used to collect study data.

Explanation: Authors should report how data were collected (i.e., extraction from medical records, participant interview or questionnaires, analysis of an already existing database, other) Authors should also state whether the data were collected prospectively or retrospectively. Finally, authors should report whether they performed a triangulation with other types of data to enhance the validity of collected data.

ITEM 9 (Methods section). Describe how time points were assessed (e.g., number of assessors, their qualifications).

Example: “*Medical case notes, referral letters, neurosurgical records, histopathology reports, follow-up and survival outcomes were reviewed by two independent investigators (K.K. and M. D.).(...) Discrepancies in the PSI were noted for 32 patients. These cases were reexamined together with the last author, and a consensus was reached.*” (Kukal, Eur J Pediatr 2009)

Explanation: Authors should report whether time points could be easily retrieved (e.g., whether the precise dates of first symptoms and diagnosis are available in a database) or whether they were potentially ambiguous. If the assessment of time points is potentially ambiguous (e.g., retrospective assessment from medical charts), authors should report how many people assessed the time points with their qualifications, whether they were independent, and how any disagreements were resolved.

ITEM 10A (Methods section). *If the study aims to evaluate associations between participant characteristics and time to diagnosis, state whether assessors of time to diagnosis were blinded to these characteristics.

Explanation: Participant characteristics is a broad term that can include sociodemographic characteristics, disease characteristics (e.g., tumor stage, severity) or even healthcare system characteristics such as level of care, specialty of the physician in charge. Authors should state whether assessors of time to diagnosis were blinded to these characteristics in order to help readers evaluate the risk of measurement bias inherent in the lack of blinding (TTD may be systematically over- or underestimated when some characteristics are present).

ITEM 10B (Methods section). *If the study aims to evaluate associations between time to diagnosis and participant health outcomes (e.g., survival), state whether assessors of time to diagnosis were blinded to these outcomes.

Explanation: Participant health outcomes can be survival status, the occurrence of sequel or quality of life. Authors should state whether assessors of time to diagnosis were blinded to health outcomes because of the risk of measurement bias inherent to the lack of blinding (TTD may be systematically over- or underestimated with a given health outcome).

ITEM 11 (Methods section). Describe the statistical methods used, including whether time to diagnosis was analysed as a continuous or categorized variable (e.g., delayed versus not delayed)

Example: “*The primary outcome was a delay in AIS (arterial ischemic stroke) diagnosis defined as 6 or more hours elapsed between time of onset of AIS symptoms (based on the oral history taken from the parent/caregiver or child) and time of diagnosis (defined as time of CT scan or MRI confirming AIS diagnosis). Due to the imprecision in reporting the onset of symptoms, the time intervals were grouped as the following: <6 hours, 6-12 hours, 12-24 hours, 24-48 hours and >48 hours.(...) We used descriptive statistics to characterise the study population as appropriate. Exploratory analyses using Pearson’s χ^2 test and univariate logistic regression models were conducted to identify factors that may potentially explain inadequate management.*” (Martin, Swiss Med Wkly 2011)

Explanation: Time to diagnosis is a quantitative variable; it can be analyzed as a linear variable or transformed (e.g., in polynoma, logarithm) or categorized (as in the example).

ITEM 12 (Methods section). *If the study aims to evaluate associations between time to diagnosis and other factors (e.g., participant characteristics or health outcomes), describe which confounders were considered and how they were chosen, measured and analysed.

Examples: “*To describe disease outcome, the endpoints were first event-free survival and sarcoma-specific survival. Survival data were summarized using Kaplan-Meier plots and analyzed with multivariate Cox models, in which symptom interval was modeled as both a categorical and a continuous covariate; the patient’s age, the tumor’s site and size, histology, nodal status, and Intergroup Rhabdomyosarcoma Study stage were entered as adjustment factors and modeled as categorical covariates (Table 1). The results from the Cox model are reported for each covariate in terms of hazard ratios (HRs) and corresponding 95% confidence intervals (CIs), together with 2-sided Wald test P values. Additional exploratory analysis was performed to investigate whether the role of symptom interval varied in relation to the different histological subtypes, by including in the Cox model, together with all the Table 1 covariates, the symptom interval-histology interaction*” (Ferrari, Cancer 2010)

Explanation: Authors should describe the candidate variables for confounders, justify their choice of variables and state whether this choice was made *a priori*. Authors should define how they measured and analyzed confounders (as continuous linear, transformed or categorical variables).

ITEM 14 (Results section). Report the number of individuals at each step of the selection process between the source population and participants and provide a flowchart (see example). Give reasons for non-participation at each stage.

Suggestion of flowchart (cf. Figure 9, page 113)

Explanation: Authors are not expected to provide all numbers (N) because some may be unknown, but authors should identify each step of the selection process in order to help readers interpret the results. This flowchart follows the steps of population selection as defined in items 4 and 5.

ITEM 16 (Results section). Report the distribution of time to diagnosis.

Example: “*The mean patient delay was 3.1 days (median, 3 days; range, 0–11 days). The mean doctor delay was 1.5 days (median, 0 days; range, 0–19 days). The mean total delay was 4.7 days (median, 4 days; range, 0–19 days). (...) The medical diagnosis was made late (after 12 h) in 11 cases.*” (Chalumeau, Eur J Clin Microbiol Infect Dis 2006)

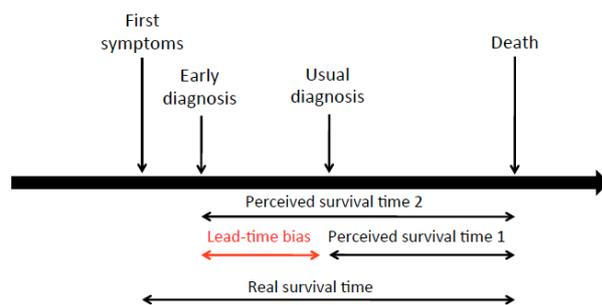
ITEM 19A (Discussion section). Discuss sources of potential bias, including bias due to the selection of participants from the source population (e.g., undiagnosed cases) and to the assessment of time points.

Example 1: “This exhaustiveness was verified for a 5-year period by the French National Paediatric Cancer Registry. As in any disease, the possibility of death before diagnosis cannot be ruled out.” (Brasme, Eur J Pediatr 2012)

Example 2: “Our study has some limitations: due to the retrospective design, we were unable to determine the exact time points when the first stroke symptoms occurred. Consequently, delay to AIS (arterial ischemic stroke) diagnosis could not be used a continuous variable, and instead we used a categorical variable with five groups. In addition, the registry data and information derived from medical records may have led to an underestimation of delays. It was also not possible for us to extract reliable data on the delays occurring in the 17 children already admitted to hospital at the time of AIS.” (Martin, Swiss Med Wkly 2009)

ITEM 19B (Discussion section). *If association between time to diagnosis and survival was studied, discuss possible lead-time bias.

Explanation:



In case of lead-time bias, authors would falsely conclude to longer survival time 1 than survival time 2

PARTIE II -

**Fréquence, déterminants et conséquences des soins
suboptimaux dans la prise en charge initiale des
enfants atteints d'infections bactériennes sévères**

Introduction

Comme nous l'avons vu en introduction générale, les IBS ont une morbi-mortalité non négligeable dans les pays occidentaux. Nous avons également vu en introduction de la partie I chapitre 1 que la population pédiatrique est particulièrement vulnérable et à risque de soins suboptimaux. Les IBS de l'enfant sont donc un bon candidat à l'étude des soins suboptimaux et leurs conséquences.

Les facteurs impliqués dans la physiopathologie des IBS sont multiples et sont liés à la fois à la bactérie en cause (inoculum, facteurs de pathogénicité, résistance aux antibiotiques) et à l'hôte infecté (adéquation de la réponse inflammatoire et immunitaire spécifique).(179) La mortalité dépend donc de ces facteurs mais également de l'administration d'un traitement spécifique.(179) Malgré les progrès des techniques de réanimation et l'existence d'antibiotiques efficaces sur les bactéries le plus souvent en cause dans les IBS communautaires* de l'enfant,(9,179) la mortalité reste élevée entre 3 à 10% au stade de sepsis et allant jusqu'à 30% en cas de choc septique*. Plusieurs pistes de recherche existent pour réduire la mortalité par IBS comme par exemple l'amélioration de la protection vaccinale pour éviter la survenue de l'IBS ou l'identification de nouvelles cibles de traitements modulateurs de la réponse inflammatoire de l'hôte.(179) Mais nous pouvons également nous demander si les traitements existants sont correctement et suffisamment précocement administrés et dans le cas contraire si une optimisation de l'administration de ces traitements pourrait permettre de réduire la mortalité et/ou les séquelles. Pour qu'un traitement soit administré de façon précoce et optimale, il faut que le diagnostic de la maladie ait été fait rapidement. Nous pouvons alors élargir la problématique du traitement à celle de l'ensemble de la prise en charge et nous demander si la qualité des soins reçus peut influencer la survie.

Pour répondre à une telle question, la première étape nécessaire nous a semblé être d'étudier les prises en charge des enfants décédés et de repérer et caractériser les soins suboptimaux. Deux études sont disponibles dans la littérature sur ce sujet. Des auteurs anglais

ont ainsi analysé les prises en charge en structure de soins primaires de 90 enfants décédés au Royaume-Uni en 2006 au moyen d'une enquête confidentielle. Ils ont montré que pour 18 enfants, les prises en charge n'avaient pas été optimales et que la non reconnaissance et le défaut de prise en charge de l'IBS étaient retrouvés dans 11 de ces prises en charge suboptimales (soit 12% des enfants décédés ayant reçu des soins en structure primaire et 44% des enfants décédés d'infections).(180) Cependant les décès survenus avant tout recours médical ou après prise en charge initiale en milieu hospitalier n'étaient pas étudiés au risque de sous-estimer l'étendue du problème. De plus, l'organisation du système de soins primaires extra-hospitaliers en Grande-Bretagne ne donne pas accès aux pédiatres contrairement au système de soins français ce qui peut rendre les résultats difficilement transposables à notre pays. Dans une revue rétrospective de 1779 autopsies d'enfants de 7 jours à 14 ans faites dans un centre régional d'anatomopathologie au Royaume-Uni (concernant un bassin de population de 18 millions d'habitants), Bamber et coll. ont mis en évidence que 134 enfants sans comorbidité identifiée avant le décès étaient décédés de cause potentiellement évitables dont 46% de sepsis.(181) Cependant, l'évitabilité du décès était définie de façon discutable comme un décès survenant chez un enfant sans comorbidité sous-jacente ni antécédent d'infection sévère. Il semble donc que les soins suboptimaux ne soient pas rares dans la prise en charge des IBS de l'enfant. Mais comme nous l'avons vu dans l'introduction générale, la détection *a posteriori* de soins suboptimaux dans la prise en charge d'enfants décédés ne préjuge pas du lien causal direct entre ces soins suboptimaux et le décès. Il faudrait pour cela comparer ces prises en charge d'enfants décédés à celles d'enfants survivants avec la même pathologie.

Cinq études se sont intéressées aux conséquences de soins suboptimaux sur la survie des enfants atteints d'IBS dans les pays développés. Nous avons résumé dans le **Tableau 7** (page 126) les résultats de ces études et leurs caractéristiques méthodologiques, en reprenant pour ces dernières les points critiques identifiés dans la partie I de ce travail doctoral.(29,182–

185) Ces études ont toutes montré une association entre les soins suboptimaux et le décès ou la survenue de séquelles. Cependant la fiabilité et la transportabilité des résultats de ces études ont pu être limitées par un recrutement monocentrique, une restriction à l'étude d'un seul pathogène, une définition arbitraire des soins suboptimaux, une évaluation de l'optimalité des soins non à l'aveugle de l'issue, une limitation de l'évaluation des soins aux soins hospitaliers ou au support hémodynamique, une non prise en compte ou une absence de justification du choix des facteurs de confusion ou encore par une absence de justification du choix de traitement des variables quantitatives. Enfin, aucune de ces études n'étudiait les déterminants des soins suboptimaux.

Les données fiables sur l'évaluation de la fréquence, des caractéristiques, des déterminants et des conséquences des soins suboptimaux dans la prise en charge initiale des enfants atteints d'IBS sont donc rares voire absentes pour la population française. Ces données sont cependant indispensables pour évaluer la pertinence de la mise en place d'actions correctrices. Au cours de cette partie II, nous avons donc dans un premier temps (chapitre 1) évalué l'étendue des soins suboptimaux chez les enfants décédés en mesurant la fréquence des soins suboptimaux dans la prise en charge initiale des enfants décédés d'IBS. Au vu de la fréquence élevée des prises en charge suboptimales, nous avons cherché à mieux comprendre leurs mécanismes de survenue (déterminants) et à mesurer leur association avec une issue défavorable (décès) (Chapitre 2), en prenant en compte les conclusions de la partie I pour essayer de minimiser les risques de biais et d'obstacles à la transportabilité.

Tableau 7. Résumé des éléments méthodologiques et des principaux résultats des études évaluant la relation entre soins suboptimaux et décès dans la prise en charge des IBS communautaires de l'enfant.

1 ^{er} auteur	Nadel (182)	Ninis (183)	Han (184)	McIntyre (29)	Inwald (185)
Pays	Royaume-Uni	Royaume-Uni	USA	Australie	Royaume-Uni
Période d'étude	1992-1994	1997-1999	1993-2001	1994-1999	2006-2007
Maladie étudiée	Méningococcie	Méningococcie	Choc septique*	Méningite à pneumocoque	Sepsis sévère
Type de recrutement	Monocentrique (centre de soins tertiaires)	En population	Monocentrique	En population	Multicentrique nationale
Processus d'identification de la population étudiée	Non précisé	Réseau national de santé publique, listings des laboratoires de bactériologie et des consultants en infectiologie	Listing de l'équipe mobile de soins d'urgence	Listings des laboratoires de la région et registres de morbidité établis à partir des codes de diagnostic d'hospitalisation	Listings de 17 unités de réanimations pédiatriques et de 2 unités de transport associés à la réanimation pédiatrique
Processus de sélection	Enfants admis successivement en réanimation pédiatrique pour méningococcémie	Tous les enfants avec un diagnostic de méningococcie dans la population source*	Tous les enfants transportés à l'exclusion des nouveau-nés prématurés	Tous les enfants avec un diagnostic de méningite à pneumocoque dans la population source	Tous les enfants admis dans une des 17 réanimations avec un diagnostic de sepsis dans les 12h après leur admission à l'hôpital
Critères diagnostiques	Culture positive (sang ou LCR) ou tableau clinique typique	Culture positive (sang ou LCR) ou tableau clinique typique	Critères de choc septique en référence à la littérature	Culture LCR positive ou méningite et hémoculture positive	Sepsis sévère défini en référence à la littérature
Recueil des données	Prospectif	Rétrospectif	Rétrospectif	Rétrospectif	Prospectif
Soins évalués					
Parentaux	Oui	Non	Non	Non	Non
Soins primaires	Oui	Non	Oui	Oui	Oui
Hospitaliers	Oui	Non	Oui	Non	Non
Catégories de soins					
Diagnostic	Oui	Oui	Non	Oui	Non
Antibiothérapie	Oui	Oui	Non	Non	Non
Support hémodynamique	Oui	Oui	Oui	Non	Oui
Définition des soins suboptimaux	Définis en référence aux recommandations nationales	Définis par un panel d'experts s'appuyant sur des recommandations nationales	Définis en référence aux recommandations du <i>Pediatric Advanced Life Support</i>	Absence d'hospitalisation après la première consultation (retard diagnostique)	Définis en référence aux recommandations du <i>Pediatric Advanced Life Support</i>
Modalité d'évaluation des soins suboptimaux					
Globale	Non	Non	Oui	Non	Oui
Par soins	Oui	Oui	Non	Oui (un seul soin)	Non
Nombre d'évaluateurs	Non précisé	4	Non précisé	Non précisé	Non précisé
Indépendance des évaluateurs	-	Non	-	-	-
Evaluation à l'aveugle de l'issue	Non précisé	Oui	Non précisé	Non précisé	Non précisé
Soin(s) dont l'association avec le décès a été étudiée	Délai pré-thérapeutique chez les patients avec septicémie (en heures)	Soins individuels d'une part et nombre de soins suboptimaux d'autre part	Délai de réversion du choc (en heures)	Retard diagnostique avec décès, séquelles ou décès et séquelles	Non réversion du choc à l'admission en réanimation

(Suite page suivante)

(.../...)

1 ^{er} auteur	Nadel (182)	Ninis (183)	Han (184)	McIntyre (29)	Inwald (185)
Test utilisé pour la mesure de l'association	Mann-Whitney	Régression logistique*	Régression logistique	Régression logistique	Régression logistique
Études des déterminants des soins suboptimaux	Non	Non	Non	Non	Oui
Prise en compte de facteurs de confusion	Non	Oui	Oui	Oui	Non
Justification des choix des facteurs de confusion	-	Oui	Non	Non	-
Traitement des variables quantitatives	Continu linéaire	Catégorisation	Continu linéaire	Non applicable	Continu linéaire
Nombre d'enfants analysés (dont décédés)	54 (13)	498 (143)	91 (26)	120 (15)	200 (34)
Résultats de la mesure de l'association entre soins suboptimaux et décès	Délai préthérapeutique médian de 18,3 heures chez les enfants décédés versus 12 heures chez les survivants (p < 0,01)	OR de décès en fonction du nombre de soins suboptimaux/décès (en référence à 0) 1 : 8,74 (IC 95%, 2,3-33) 2 : 34 (IC 95%, 5-198) >2 : 113 (IC 95%, 8,4-1510)	OR 2,29 (IC 95% 1,2-4,4) pour chaque heure de non réversion du choc	Association avec le décès OR 0,6 (IC 95%, 0,09-4.0) avec le décès ou les séquelles : OR 0,42 (IC 95%, 0,12-1.5) avec les séquelles OR 3,4 (IC 95%, 1,03-11,4)	OR 3,8 (IC 95%, 1,4-10,2)

Abréviations : IBS, infection bactérienne sévère ; IC 95%, intervalle de confiance à 95% ; LCR, liquide céphalorachidien

Sources de données des études DIABACT

Nous présentons ici les éléments méthodologiques communs aux études DIABACT I et II. Les éléments spécifiques notamment concernant les analyses statistiques seront présentés séparément dans chaque chapitre.

1. Méthodologie générale

La méthodologie des études DIABACT s'inspire de celle des enquêtes confidentielles avec comité d'experts. Comme nous l'avons vu dans l'introduction générale, l'enquête confidentielle avec comité d'experts est l'étude systématique d'un évènement grave précisément défini, réalisée à distance de l'évènement, pour en faire l'analyse médicale objective et notamment pour étudier le parcours de soins d'un patient.(3) Le caractère confidentiel est assuré durant l'analyse de l'optimalité des soins vis-à-vis des personnes (malades et soignants) et des structures de soins impliquées. L'optimalité est définie par comparaison à des soins dits "de référence".(3) Cette analyse de l'optimalité des soins est menée par un groupe de spécialistes (comité d'experts) extérieurs à l'évènement mais concernés par la question (revue par les pairs). L'objectif est d'établir un bilan, tirer des leçons et émettre un avis ou des recommandations destinés au public médical et aux décideurs de santé publique pour à l'avenir améliorer les soins.(3) Les points forts de cette méthodologie reposent sur un support totalement anonymisé, l'évaluation par plusieurs experts des soins (de façon collégiale ou indépendante) et également la possibilité de maintenir les experts à l'aveugle de l'évolution et de l'issue. Ces particularités en font une méthodologie intéressante pour l'étude des soins suboptimaux.

2. Population

2.1 Description des caractéristiques de la population source

La population source* était constituée de l'ensemble des enfants atteints d'une IBS communautaire sur une région géographique constituée de deux départements contigus, la Loire-Atlantique et la Vendée. L'IBS était définie comme une infection bactérienne

nécessitant une admission en réanimation ou responsable d'un décès avant l'admission en réanimation. Les deux départements de Loire-Atlantique et Vendée ont une superficie de 13 500 km². En 2005, ces 2 départements comprenaient 1,6 millions d'habitants dont 346 307 enfants de moins de 15 ans, un revenu moyen par ménage de 14 500 euros par an (moyenne nationale : 15 510 euros) et un taux de mortalité infantile de 3,3‰ (valeur nationale à 3,8‰). Lors de la période d'étude, les soins aux enfants y étaient prodigués par un centre hospitalier universitaire, 4 centres hospitaliers généraux, des pédiatres libéraux (29 pour 100 000 enfants), des médecins généralistes (145 pour 100 000 habitants) et 2 centres de régulations téléphoniques des urgences médicales pouvant envoyer des équipes médicales mobiles (dirigées par des médecins urgentistes) au domicile des patients. L'organisation des soins dans ces zones géographiques pour les 2 études prévoyait que tous les enfants présentant une IBS au-delà de la période néonatale soient transférés dans un des centres hospitaliers universitaires (Nantes).

2.2 Période d'études

Les études DIABACT I et II concernaient les enfants admis en réanimation ou décédés à domicile entre janvier 2000 et mars 2006.

2.3 Critères d'inclusion et d'exclusion

Nous avons inclus tous les enfants de 3 mois à 16 ans atteints d'infection bactérienne communautaire nécessitant une admission en réanimation ou à l'origine d'un décès avant l'admission en réanimation. L'infection bactérienne était définie comme un syndrome clinique très fortement suspect d'infection bactérienne (le purpura fulminans* par exemple) ou une infection prouvée par culture ou *polymerase chain reaction* (PCR) conformément à la conférence de consensus internationale sur le sepsis de l'enfant.(9,186) Nous avons choisi

comme critère arbitraire de sévérité l'admission en réanimation. Les enfants atteints d'infection virale ou fongique, de coqueluche ou d'infection nosocomiale n'étaient pas inclus dans l'étude. Nous avons défini l'infection nosocomiale comme une infection survenant plus de 48 heures après l'admission à l'hôpital. Les enfants dont les dossiers n'étaient pas suffisamment renseignés étaient secondairement exclus. Pour l'étude DIABACT I, seuls les enfants décédés dans un délai compatible avec un rapport de causalité avec l'IBS étaient étudiés.

3. Identification des cas

Nous avons identifié et évalué pour une éventuelle inclusion tous les enfants décédés ou admis en réanimation et dont le résumé d'unité médicale (diagnostic de sortie) comprenait un des mots-clés suivants : méningite, purpura fulminans, septicémie ou choc septique. Nous avons identifié dans les listings du service d'aide médicale urgente (SAMU) de Loire-Atlantique et de Vendée tous les enfants décédés à domicile dans un contexte de fièvre.

Au cours de la période d'étude, conformément aux protocoles du service, tous les enfants admis en réanimation avec un tableau de sepsis étaient suspects d'infection bactérienne et ont eu des investigations microbiologiques (hémocultures, ponction lombaire –en l'absence de contre-indication –, prélèvement d'urine pour culture bactériologique, ponction pleurale en cas d'épanchement, culture de liquide péritonéal en cas de péritonite). Le protocole du SAMU pour les enfants décédés à domicile prévoyait une anamnèse détaillée des parents recherchant notamment la présence de fièvre dans les jours précédant le décès. En cas de décès à domicile avant l'âge de 2 ans, le protocole du SAMU 44 et du SAMU 85 prévoyait un transfert vers le centre de référence de mort subite (service des urgences du CHU de Nantes) où des prélèvements bactériologiques étaient systématiquement réalisés (hémoculture, ponction lombaire, culture d'urine).

Pour vérifier l'exhaustivité des décès identifiés sur la période de l'étude, nous avons confronté nos effectifs à ceux de la base de données du CépiDc d'où nous avons extrait les cas de décès d'enfants âgés de 0 à 14 ans potentiellement reliés à une IBS de 2000 à 2005 (l'année 2006 n'était pas complète à l'époque) pour les 2 départements concernés. Les codes suivants de *l'International Statistical Classification of Diseases and Related Health Problems* dans sa version 9 et associés à la cause principale de décès ont été utilisés : A00–09, A30–49, B95, B96, B99, G00, G01, G03, G06, J15, J86.9 et R57.8. En cas de doute sur la cause du décès (infection nosocomiale ou non bactérienne), nous avons interrogé les données locales de décès (résumés d'unité médicale des services de pédiatrie –réanimation, néonatalogie, pédiatrie générale, onco-hématologie pédiatrique-, et données anonymisées des services de protection maternelle et infantile pour les moins de 6 ans).

4. Collecte des données

Nous avons extrait des dossiers les données concernant les caractéristiques des enfants (âge, sexe et antécédents), la présentation clinique initiale de leur IBS (premiers symptômes décrits par les parents, motif de la première consultation médicale, présences de signes de gravité à la première consultation médicale) et les IBS elles-mêmes (diagnostic final, bactérie en cause). Les signes suivants ont été considérés comme des signes de gravité initiale du sepsis : tachycardie, bradycardie et/ une polypnée, hypotonie, hypovigilance, détresse respiratoire, geignement, syndrome méningé ou purpura extensif. Nous avons également retracé le parcours de soins de chaque patient (nombre de consultations médicales, nature de la première consultation – cabinet libéral, urgences adultes ou pédiatriques, intervention du SAMU, spécialité du médecin -pédiatre, urgentiste, médecin généraliste-). Nous avons également relevé l'issue de santé des enfants selon 2 modalités : survie ou décès.

Pour la rédaction des fiches de synthèse soumise aux experts pour l'évaluation des soins (cf.

chapitre suivant), nous avons également extrait les dates et les heures des consultations ainsi que la nature et les dates et heures des actions parentales, des décisions d'examens paracliniques, d'orientation et de traitements prises par les médecins et des conclusions médicales.

5. Evaluation de l'optimalité des soins

Les dossiers des enfants étaient synthétisés par un des auteurs sous la forme de fiches horodatées sans le diagnostic final ni l'issue pour l'enfant (cf. **Figure 11**, page 137). Ces fiches étaient alors soumises à deux experts indépendants et à l'aveugle du diagnostic final et de l'issue : un pédiatre réanimateur universitaire et un pédiatre libéral, tous deux expérimentés. Les fiches étaient soumises de façon progressive par groupe de 20 à 40 fiches en mélangeant les fiches des enfants décédés avec ceux des enfants survivants allant être inclus dans l'étude DIABACT II. Les experts devaient évaluer l'optimalité des soins en référence aux recommandations françaises et internationales en vigueur durant la période de l'étude concernant notamment le délai d'administration de l'antibiothérapie en cas de suspicion de méningococcémie (théoriquement immédiate en cas de purpura extensif) et les modalités de remplissage vasculaire initiaux en cas de sepsis sévère avec défaillance hémodynamique (jusqu'à 40 mL/kg dans la première heure).(15,187,188) En référence à l'article de Nadel et coll., l'absence de consultation immédiate était considérée comme un retard au recours médical devant une fièvre accompagnée soit d'un rash purpurique soit d'un autre signe de gravité suivant : cyanose, geignements, convulsions, confusion, céphalées intenses, troubles de la vigilance, myalgies ou arthralgies intenses, asthénie marquée, vomissements incoercibles et/ou froideur des extrémités.(182) De façon arbitraire une absence de consultation après 48 heures de fièvre était considérée comme un retard au recours médical.

Il était demandé aux experts une évaluation globale avec 3 conclusions possibles concernant la prise en charge : suboptimale certaine (situation 1), suboptimale possible (situation 2) ou optimale (situation 3). Il était également demandé aux experts de justifier leurs conclusions en relevant chaque soin suboptimal selon la classification proposée par *l'Institute of Medicine* américain (cf. **Figure 2**, page 44). Nous avons ajouté une catégorie « retard au recours médical » à cette classification car ce type de soin avait été souligné dans les études sur les méningococcémies.(182,183) Le degré d'accord entre les experts était calculé par le coefficient kappa de Cohen*. L'interprétation du kappa était basée sur l'échelle proposée par Landis et Koch.(189) En cas de désaccord sur l'évaluation de la prise en charge globale ou d'un des soins de la prise en charge, un troisième expert d'un service de pédiatrie générale était sollicité.

6. Ethique et règlementations

L'étude DIABACT I a obtenu l'approbation du comité de protection des personnes d'Ile-de-France III et l'étude DIABACT II a obtenu l'approbation du groupe nantais d'éthique dans le domaine de la santé. Une dérogation à l'information des familles pour l'usage des données de routine a été accordée compte-tenu du sujet traité.

Figure 11. Exemple de fiche soumise aux experts pour les études DIABACT

N° du patient : 1 Age : 8 ans Sexe : F ATCD : aucun Vaccinations : à jour		
Date et heure	Evénements	Prise en charge
J0 en hiver (heure non précisée)	Fièvre, odynophagie et céphalées.	
J1	Persistance de la fièvre altération de l'état général (asthénie, anorexie) et vomissements.	Consultation chez le médecin traitant dans la soirée. Diagnostic d'angine et prescription de dompéridone, d'amoxicilline 500 mg x2 /J et de paracétamol.
Nuit de J1 à J2	Fièvre à 38,9°C avec céphalées, vomissements, geignements, troubles de la conscience (sommolence), hypotonie et modification du comportement (confusion).	Consultation aux urgences d'un premier hôpital.
J2 à 2h05	Agitation avec des phases de somnolence et de révolusion oculaire, Glasgow à 9, gémissements. Nuque raide, pas de déficit moteur. TA à 138/102mmHg. Tachycardie. Fièvre à 38,6°C, pas de marbrures, pas de purpura.	Hospitalisation aux lits portes. NFS CRP hémocultures.
J2 à 3h		1 ampoule de 10 mg de valium en IR et 1 g de ceftriaxone soit 33 mg/kg et perfusion par du G5% avec électrolytes 1 L sur 24h (33 ml/kg/24h).
J2 à 5h		Résultats NFS : leucocytes 26000/mm ³ , PNN : 23250/mm ³ , Plaquettes = 168 000/mm ³ ; CRP = 207 mg/L.
J2 à 9h		TDM cérébral normal
J2 à 10h30		Transfert en réanimation pédiatrique. Intubation et ventilation avant transport.
J2 à 11h		Arrivée en réanimation pédiatrique.

Abréviations : CRP, C reactive protein ; NFS, numération formule sanguine ; PNN, polynucléaires neutrophiles, TA, tension artérielle ; TDM, tomodensitométrie.

CHAPITRE 1

Soins suboptimaux dans la prise en charge initiale des enfants décédés d'infection bactérienne sévère : enquête confidentielle en population entre 2000 et 2006

Ce chapitre a donné lieu à la publication suivante (travail préparatoire à la thèse) :

Suboptimal care in the initial management of children who died from severe bacterial
infection: a population-based confidential inquiry

Elise Launay, Christèle Gras-Le Guen, Alain Martinot, Rémy Assathiany, Thomas Blanchais,

Nadjette Mourdi, Albertine Aouba, Marie-Hélène Bouvier-Colle, Jean-Christophe Rozé,

Martin Chalumeau

Pediatric Critical Care Medicine 2010; 11: 469-474 (Facteur d'impact = 2,3)

(Annexe 3, page 217)

1. Introduction

Nous venons de voir en introduction de cette partie II que deux études menées au Royaume-Uni avaient analysé la fréquence des prises en charge suboptimale chez des enfants décédés d'IBS mais que la portée de leurs résultats était limitée par une organisation différente du système de soins ou une définition arbitraire de l'optimalité.

Il a été montré, dans une étude récente menée dans un service de réanimation pédiatrique de référence en Grande-Bretagne, que près de 70% des 84 décès liés à un sepsis et survenus chez des enfants précédemment sains se produisaient dans les 24 premières heures d'admission. Cela peut laisser supposer que les actions d'amélioration doivent concerner la prise en charge précoce de l'IBS en espérant éviter ainsi l'évolution vers le choc septique. Nous avons décidé de focaliser l'analyse des soins suboptimaux sur la prise en charge initiale de l'IBS, c'est-à-dire avant l'admission en réanimation.(190)

L'objectif de cette première étude était donc de mesurer la fréquence des soins suboptimaux au cours de la prise en charge initiale des enfants d'IBS à l'échelle de la population de 2 départements français et de caractériser ces soins suboptimaux.

2. Méthodologie

La méthodologie utilisée concernant l'identification des cas, la vérification de l'exhaustivité des décès, la collecte des données et l'évaluation de l'optimalité a été détaillée précédemment (cf. PARTIE II- sources de données, page 129). Les analyses statistiques étaient purement descriptives pour cette étude. Le coefficient kappa de Cohen a été calculé sans et avec pondération.(191,192) Nous avons utilisé une méthode de pondération quadratique permettant de rendre compte d'un poids plus important de la discordance entre les conclusions « suboptimal certain » et « optimal » par rapport aux discordances entre les conclusions « suboptimal possible » et « optimal » ou « suboptimal certain » et « suboptimal possible ».

Le coefficient a été interprété selon l'échelle proposée par Landis et Koch.(189) Le logiciel Stata v11 (Satacorp, College Station, TX, USA) a été utilisé pour calculer le coefficient kappa de Cohen.

3. Résultats

3.1 Population

Durant la période de l'étude, 23 enfants avaient les critères d'inclusion sans critères d'exclusion. Deux dossiers (10%) n'ont pu être analysés du fait d'informations cliniques et bactériologiques insuffisantes : celui d'un garçon de 11 mois décédé d'un purpura fébrile et celui d'un garçon de 27 mois décédé d'une méningite à pneumocoque.

L'âge médian des 21 enfants décédés était de 2 ans (EIQ : 0,8–2,8), avec un sexe ratio (garçon/fille) à 1/0,8. Sept enfants avaient des antécédents notables à type de prématurité extrême, d'asthme, de purpura rhumatoïde récent, de cytopathie mitochondriale, d'anomalie chromosomique ou de méningite à salmonelle. Les diagnostics finaux étaient les suivants : purpura fulminans (n = 11), méningites (n = 7), pleuro-pneumonies (n = 2) et septicémie secondaire à une pyélonéphrite (n = 1). Les bactéries identifiées étaient : *Neisseria meningitidis* (n = 12 patients, soit 57%) dont sérotype B (n = 5), sérotype C (n = 3) et 4 pour lesquels le sérotype n'a pu être identifié ; *Streptococcus pneumoniae* (n = 3) ; et *Streptococcus pyogenes*, *Mycoplasma pneumoniae*, *Salmonella enterica* serovar Bredeney et *Escherichia coli* (n = 1 patient pour chaque) (cf. **Tableau 8**, page 150). Pour 2 enfants, aucune documentation bactériologique n'avait été retrouvée malgré des prélèvements et cultures adéquates : un avait un tableau typique de purpura fulminans et un autre une présentation clinique typique de méningite avec une analyse de liquide céphalorachidien très évocatrice d'une infection bactérienne. Treize enfants sont décédés le jour même de leur

admission en réanimation (jour 0), trois au jour 1 et les 5 autres sont décédés avant le jour 13. Tous les décès étaient reliés à l'IBS initiale et non à une complication iatrogénique.

3.2 Symptômes et parcours de soins

La fièvre était un des premiers symptômes pour tous les enfants. Les autres symptômes n'étaient pas spécifiques d'IBS à l'exception du purpura. Le motif de consultation le plus fréquent était l'apparition d'un rash purpurique pour 7 enfants (33% ; IC 95%, 12–54), et des geignements pour 2 enfants ; les autres motifs de consultation n'étaient pas spécifiques d'IBS. Pour un enfant, le motif de recours aux soins était l'arrêt cardio-respiratoire. Seize enfants (76% ; IC 95% 57–95) présentaient au moins un signe de gravité lors de la première consultation : un purpura extensif pour 9 enfants (43% ; 95% CI, 21–65), une défaillance hémodynamique pour 8 (38% ; 95% CI, 16–60), une altération de vigilance pour 5 (24% ; 95% CI, 5–43), une détresse respiratoire pour 4 (19% ; 95% CI, 1–37), un syndrome méningé pour 3 (14% ; 95% CI, 0–30) et une hypotonie pour 3 également (14% ; 95% CI, 0–30). Neuf enfants présentaient plusieurs signes de gravité. Les parcours de soins étaient très différents pour les 21 enfants. La première consultation était avec un médecin généraliste pour 12 enfants (57% ; IC 95%, 35–79), l'équipe du SAMU pour 6 et 3 enfants ont été amenés directement aux urgences. Quinze enfants (71% ; 95% CI, 50–92) ont été hospitalisés ou adressés en milieu hospitalier à l'issue de la première consultation. Six enfants (29% ; IC 95%, 8–50) ont été renvoyés à domicile après la première consultation.

3.3 Soins suboptimaux

Un des experts n'a pu rendre de conclusion sur l'optimalité d'une des prises en charge. Le degré d'accord (coefficient kappa de Cohen) entre les experts a donc été évalué pour les 20 dossiers restants. Les deux experts ont rendu des conclusions différentes pour 6 des 20 dossiers. La valeur du coefficient kappa « brut » était de 0,44 ($p = 0,003$) et le kappa pondéré

selon une méthode quadratique était de 0,73 ($p = 0,0005$). Il n'y avait aucun cas de désaccord de type « prise en charge suboptimale certaine » pour un expert et « prise en charge optimale » pour l'autre. Le troisième expert a donc été sollicité pour évaluer la prise en charge pour laquelle un des experts n'avait pu rendre de conclusion et pour les 6 prises en charge pour lesquelles les 2 experts n'avaient pas rendu les mêmes conclusions. Pour 16 des 21 enfants (76% ; IC 95%, 57–95) la prise en charge a été jugée suboptimale avec 13 prises en charge suboptimales certaines (62% ; IC 95%, 40–84) et 3 (14% ; IC 95%, 0–30) suboptimales possibles. Un retard au recours médical était présent pour les experts pour 7 enfants (33% ; IC 95%, 12–54) ; les justifications de ce retard étaient l'absence de consultation rapide devant un purpura pour 2 enfants, devant un changement de comportement pour 3 enfants, et devant une fièvre évoluant depuis plus de 48 heures pour 2 autres enfants. Le retard au recours médical était le seul soin suboptimal détecté depuis l'apparition des premiers symptômes jusqu'à l'admission en réanimation pour une des prises en charge jugée suboptimale.

La nature suboptimale de la prise en charge globale était reliée à la présence de soins suboptimaux médicaux pour les 15 autres prises en charge suboptimales. Parmi ces 15 prises en charge, 35 soins suboptimaux ont été détectés (cf. **Tableau 9**, page 151). Ces soins ont été répartis, conformément à la classification proposée par l'*Institute of Medicine* américain, en soins suboptimaux diagnostiques (31% ; IC 95%, 15–47), thérapeutiques (57% ; IC 95%, 40–74), et « autres » (12% ; IC 95%, 1–23). Parmi les soins suboptimaux de nature diagnostique, la sous-évaluation de la gravité est la plus fréquemment retrouvée avec notamment une absence de médicalisation du transport devant un purpura ou une non reconnaissance des signes de défaillance hémodynamique (tachycardie, insuffisance circulatoire périphérique) ($n = 8$; 38% des enfants ; IC 95%, 16–60). Parmi les soins suboptimaux thérapeutiques, un retard à l'administration du traitement antibiotique ou du support hémodynamique une fois le

diagnostic posé était le plus fréquent, concernant 8 enfants (38% des enfants; IC 95%, 16–60). L'administration immédiate d'une antibiothérapie conformément aux recommandations en vigueur n'était faite que pour un des 5 cas de purpura concernés par les recommandations. Les erreurs de posologie et notamment le remplissage vasculaire insuffisant étaient fréquentes (24% des patients ; IC 95%, 9–35). Le support hémodynamique était inadéquat (absence de ou retard au remplissage ou absence d'administration d'inotrope ou de vasoconstricteur) pour 9 enfants (43% ; IC 95%, 21–65). Tous les intervenants médicaux (médecins généralistes, urgentistes pédiatres et adultes, médecins du SAMU) étaient impliqués dans au moins un soin suboptimal : 8 soins suboptimaux étaient attribués à des médecins généralistes libéraux, 9 aux médecins du SAMU et 11 aux urgentistes (5 dans un service d'urgences polyvalentes et 6 dans un service d'urgences pédiatriques).

3.4 Vérification de l'exhaustivité des décès dans la population

La consultation des données du CépiDC concernant les départements de Loire-Atlantique et de Vendée entre 2000 et 2005 a permis d'identifier 6 autres décès non détectés par notre stratégie de recherche. Les recoupements avec les données locales de décès (concordance du sexe, du département et de la tranche d'âge) nous ont permis de retrouver ces 6 décès potentiellement éligibles. Tous ces décès avaient au moins un critère d'exclusion : âge inférieur à 3 mois (n = 1), aspergillose invasive (n = 1) et infection nosocomiale (n = 4).

4. Discussion

4.1 Principaux résultats

Dans cette première étude en population sur la fréquence des soins suboptimaux dans la prise en charge initiale des enfants atteints d'IBS tous germes confondus, 76% (IC 95%, 57–95) des prises en charge étaient jugées suboptimales par 2 experts indépendants. Les principaux

soins suboptimaux identifiés au sein de ces prises en charge étaient le retard à l'administration du traitement adéquat (notamment à l'antibiothérapie en cas de purpura extensif), des posologies inadaptées notamment concernant le remplissage vasculaire et une sous-estimation globale de la sévérité. Dans un tiers des cas, un retard au recours médical par les parents était également souligné par les experts.

4.2 Validité interne

Evaluation de l'optimalité

Comme nous l'avons vu lors de la première partie de cette thèse dans le domaine du diagnostic, la caractérisation de l'optimalité des soins est sujette à des risques de biais spécifiques, elle est donc très controversée.(193) Dans ce chapitre, les soins suboptimaux étaient définis indépendamment par 2 experts avec un bon degré d'accord comme le montre le coefficient de kappa pondéré (0,73). Pour renforcer l'objectivité des experts et s'approcher au plus près de la réalité clinique, ceux-ci évaluaient les prises en charge à l'aveugle du diagnostic final et de l'issue (décès ou survie). Cependant les experts avaient d'emblée accès à tout le déroulé de la prise en charge des premiers symptômes à l'admission en réanimation, la connaissance de l'évolution a donc pu influencer l'évaluation des premiers soins reçus. De la même façon, les experts, s'ils ne savaient pas quels étaient les enfants décédés des survivants, avaient connaissance des objectifs de l'étude et notamment du fait que tous les enfants avaient été au final admis en réanimation. Une évaluation au fur et à mesure des soins, comme cela avait été fait dans l'étude de Ninis et coll.,(183) aurait pu encore améliorer l'objectivité de l'évaluation de l'optimalité des soins. Cependant le risque de biais de classement lié à notre méthode d'évaluation de la prise en charge globale semble faible pour les 3 principaux soins suboptimaux mis en évidence au vu du vaste décalage existant entre ces soins et les recommandations.

Utilisation de fiches de synthèse et exhaustivité

L'utilisation de fiches de synthèse peut également être moins objective que l'utilisation de l'ensemble du dossier comme dans les enquêtes confidentielles.(3) Les rédacteurs des fiches n'étaient pas à l'aveugle du diagnostic ni de l'issue, cependant les termes utilisés étaient le plus neutres possibles, et les soins qui auraient du être réalisés d'après les recommandations et qui ne l'ont pas été n'étaient pas notifiés (par exemple, il n'a pas été marqué « pas de bilan biologique », « pas de transfert médicalisé »). Tous les détails présents dans le dossier et concernant les soins avant l'admission en réanimation étaient retranscrits dans les fiches afin de ne pas perdre d'information. Enfin, la vérification de l'exhaustivité des inclusions était très satisfaisante puisque tous les décès remplissant les critères d'inclusion sans exclusion ont été identifiés. Cependant 2 dossiers (10%) n'ont pu être analysés du fait de données manquantes trop importantes.

4.3 Validité externe

L'organisation des soins primaires en France est particulière avec l'existence du SAMU sur tout le territoire français. Cependant la fréquence des soins suboptimaux est comparable à celle rapportée dans d'autres pays avec des économies avancées. Ainsi Nadel et coll. ont retrouvé 71% de prises en charge hospitalières suboptimales pour des enfants atteints de méningococcémie.(182) Ninis et coll. ont montré que 75% des enfants décédés de méningococcémies avaient été confrontés à au moins un soin suboptimal au cours de leur prise en charge.(183) La suboptimalité des soins semble donc due principalement à des facteurs extrinsèques à l'organisation des soins d'urgences. Cette hypothèse est renforcée par la typologie des soins suboptimaux avec 30% de retard à la première consultation pour l'étude de Nadel et coll. *versus* 33% dans notre étude. Les soins suboptimaux liés à la prise en charge médicale étaient également identiques à ceux retrouvés dans les études sur les enfants atteints

de méningococcémie. Ainsi, Ninis et coll. ont mis en évidence un défaut de reconnaissance des signes de gravité dans 38% des cas de décès et l'équipe de Riordan avait mis en évidence que 25% des patients atteints de méningococcémie n'avaient pas eu d'antibiothérapie avant l'admission à l'hôpital alors que le rash purpurique avait été constaté par un médecin.(183) L'étude de McIntyre et coll. retrouvait un délai entre l'admission hospitalière et l'antibiothérapie de plus de 6 heures pour 41 des 120 patients étudiés.(29) Dans notre étude, le traitement hémodynamique inadéquat était également fréquent (24% des prises en charge avec un remplissage insuffisant, 14% où un deuxième remplissage n'a pas été fait alors qu'il était indiqué et 5% de remplissage trop tardif) mais moindre que les 62% de traitement hémodynamique inadéquat retrouvé par l'équipe d'Inwald et coll.(185) Cela peut peut-être s'expliquer par des critères d'évaluation moins stricts par nos experts que dans l'audit anglais qui se référait aux recommandations très rigoureuses du *Pediatric advanced life support*. La majorité des enfants (76%; IC 95%, 57–95) présentait des signes de gravité lors de la première consultation, cela pourrait traduire soit des maladies très sévères d'emblée soit des consultations tardives. Il a en effet été montré dans une étude française qu'une minorité des parents (7% des 123 interrogés) connaissait les signes de gravité comme le purpura.(194) Cette information concernant le purpura est en effet absente des carnets de santé français alors qu'elle figure sur les carnets de santé des enfants britanniques où le test à la vitropression est décrit (*tumbler test*).

4.4 Conclusions et perspectives

Les soins suboptimaux concernaient un nombre important de prise en charge initiale avec 76% des 21 dossiers analysés. Nous avons pu détecter 35 soins suboptimaux et les caractériser pour dégager des pistes d'amélioration, notamment une sensibilisation des parents et des soignants aux signes de gravité devant une fièvre. Cependant, seuls les soins

suboptimaux étaient relevés par les experts, nous n'avons donc pas pu évaluer les variables associées aux soins suboptimaux en comparaison aux soins optimaux. La présente étude ne nous permettait pas non plus d'évaluer l'association entre la présence de soins suboptimaux et le décès puisque tous les enfants étaient décédés. Le chapitre suivant rapporte donc, au sein de la même population d'enfants élargie aux enfants survivants, l'analyse de la fréquence des soins suboptimaux, de leurs conséquences et de leurs déterminants potentiels.

5. Figures et Tableaux

Tableau 8. Description de l'âge, des diagnostics, de la documentation bactériologique et des causes de décès des 21 enfants inclus.

Patient	Age (ans)	Diagnostic final	Bactérie identifiée	Cause du décès
1	4,8	Purpura fulminans	<i>Neisseria meningitis</i> (sérotipe B)	Choc septique
2	0,3	Méningite	<i>Streptococcus pneumoniae</i>	Choc septique
3	0,4	Méningite	<i>Salmonella enterica</i>	Choc septique
4	2	Méningite	<i>Streptococcus pneumoniae</i>	Hypertension intracrânienne
5	1,6	Purpura fulminans	Aucune documentation bactériologique	Choc septique
6	2,4	Méningite	<i>Streptococcus pneumoniae</i>	Hypertension intracrânienne
7	0,6	Purpura fulminans	<i>Neisseria meningitis</i> (sérotipe B)	Choc septique
8	0,4	Purpura fulminans	<i>Neisseria meningitis</i>	Choc septique
9	10,3	Méningite	<i>Neisseria meningitis</i> (sérotipe inconnu)	Choc septique
10	3,9	Pleuro-pneumonie	<i>Mycoplasma pneumoniae</i>	Hypoxémie réfractaire et défaillance multiviscérale
11	2,8	Pleuro-pneumonie	<i>Streptococcus pyogenes</i>	Choc septique
12	2,8	Purpura fulminans	<i>Neisseria meningitis</i> (sérotipe C)	Choc septique
13	0,8	Purpura fulminans	<i>Neisseria meningitis</i> (sérotipe inconnu)	Choc septique
14	2	Méningite	<i>Neisseria meningitis</i> (sérotipe inconnu)	Choc septique
15	4,9	Purpura fulminans	<i>Neisseria meningitis</i> (sérotipe C)	Choc septique
16	0,9	Purpura fulminans	<i>Neisseria meningitis</i> (sérotipe C)	Choc septique
17	2,4	Purpura fulminans	<i>Neisseria meningitis</i> (sérotipe B)	Choc septique
18	1,2	Purpura fulminans	<i>Neisseria meningitis</i> (sérotipe B)	Choc septique
19	4,7	Méningite	Aucune documentation bactériologique	Choc septique et hypertension intracrânienne
20	0,3	Purpura fulminans	<i>Neisseria meningitis</i> (sérotipe B)	Choc septique
21	1,1	Choc septique sur pyélonéphrite	<i>Escherichia coli</i>	Choc septique

Tableau 9. Classification des soins suboptimaux détectés au cours de la prise en charge des 21 enfants décédés d'IBS.

Classification des soins suboptimaux selon l' <i>Institute of Medicine</i> américain	Soins suboptimaux identifiés par les experts dans les prises en charge	Nombre de patients (% des soins suboptimaux, n=35) [% des prises en charge, n=21]
Diagnostic	<p>Signes de pré-choc non pris en compte</p> <p>Absence de médicalisation du transport devant un purpura extensif</p> <p>Retard à la décision au transfert en réanimation</p> <p>Interprétation erronée de la CRP</p>	<p>5 (14) [24]</p> <p>4 (11) [19]</p> <p>1 (3) [5]</p> <p>1 (3) [5]</p>
Thérapeutique	<p>Remplissage vasculaire insuffisant</p> <p>Erreur de posologie de l'antibiothérapie</p> <p>Absence d'administration immédiate d'antibiotique face à un purpura extensif</p> <p>Retard à l'administration de l'antibiothérapie</p> <p>Retard à l'administration du remplissage</p> <p>Absence de deuxième remplissage</p> <p>Absence d'administration d'inotrope</p> <p>Absence d'intubation et de ventilation mécanique</p> <p>Echec à la mise en place d'une voie d'abord</p>	<p>5 (14) [24]</p> <p>1 (3) [5]</p> <p>4 (11) [19]</p> <p>3 (9) [14]</p> <p>1 (3) [5]</p> <p>3 (9) [14]</p> <p>2 (5) [10]</p> <p>1 (3) [5]</p> <p>1 (3) [5]</p>
Autres (autre défaillance du système)	Retard à l'arrivée de l'équipe du SAMU ou du médecin de garde	2 (5) [10]
Autre (défaut de communication)	Défaut de coordination des équipes médicales	1 (3) [5]

Abréviations : CRP, C reactive protein ; SAMU, Service d'aide médicale d'urgence

CHAPITRE 2

Fréquences, déterminants et conséquences des soins suboptimaux dans la prise en charge initiale des enfants atteints d'infection bactérienne sévère : enquête confidentielle rétrospective en population entre 2000 et 2006

Ce chapitre a donné lieu à la publication suivante :

Why children with severe bacterial infection die: a population-based study of determinants and consequences of suboptimal care with a special emphasis on methodological issues.

Elise Launay, Christèle Gras-Le Guen, Alain Martinot, Rémy Assathiany, Elise Martin, Thomas Blanchais, Catherine Deneux-Tharaux, Jean-Christophe Rozé, Martin Chalumeau

PLoS One. 2014 ; 9 : e107286 (Facteur d'impact = 3,5)

(Annexe 4, page 219)

1. Introduction

Nous avons vu dans le chapitre précédent que les prises en charge initiales d'enfants décédés d'IBS étaient fréquemment jugées suboptimales et que les soins suboptimaux les plus souvent détectés étaient le retard à la consultation, la sous-évaluation de la gravité par les médecins, le retard à l'antibiothérapie ainsi qu'un traitement hémodynamique inadapté (remplissage vasculaire trop tardif ou insuffisant). Cette fréquence élevée de soins suboptimaux justifie donc une meilleure compréhension de leurs mécanismes de survenue (déterminants) ainsi que de leurs conséquences potentielles. Cependant, le langage causal (« déterminants » et « conséquences ») doit être utilisé avec beaucoup de précautions dans une étude observationnelle. En effet, les études épidémiologiques ne permettent pas d'affirmer de façon certaine l'existence d'un lien causal entre une exposition et un événement de santé. Il est donc impossible d'affirmer que si un soin suboptimal n'avait pas eu lieu, le décès aurait été évité, sauf dans de très rares cas manifestes (étouffement a vacuo, administration erronée de médicaments...). Cependant, il existe des critères (critères de Hill) aidant à conforter la plausibilité du lien causal en épidémiologie, parmi eux : l'existence d'une association statistique, la force de cette association (valeur du risque relatif ou de l'OR), la relation dose-effet, le lien temporel (la cause précède l'effet), la reproductibilité des résultats ou encore la plausibilité biologique (ou physiopathologique).(195) Pour qu'un lien causal entre 2 variables statistiquement associées soit plausible, il faut également que cette association ne soit pas le fait d'un facteur de confusion. Dans les approches statistiques traditionnelles, les facteurs de confusion sont définis comme des variables associées à la fois à l'exposition et à l'événement de santé étudié et modifiant l'estimation de l'effet de l'exposition sur l'événement. Cependant cette méthode d'identification et d'ajustement* sur les facteurs de confusion peut introduire une association conditionnelle et amener à des conclusions biaisées (*conditioning* ou *collider bias**).(196) Des auteurs ont alors proposé une méthode graphique (graphe acyclique orienté*

-directed acyclic graph- DAG) consistant à représenter les relations causales supposées entre variables d'intérêt et permettant d'identifier de façon appropriée les variables sur lesquelles ajuster les analyses sans risquer de biaiser les résultats.(197–199) Aucune des études citées en introduction de la partie II et dont l'objectif était d'analyser l'association entre soins suboptimaux au cours d'une IBS et décès ne rapportait la construction de DAG pour justifier le choix des variables d'ajustement.

Une des difficultés pour évaluer la relation entre les soins suboptimaux dans la prise en charge initiale des IBS et le décès réside dans le caractère variable du nombre d'occasions de soins pour chaque enfant. En effet, dans le chapitre 1, les patients avec une évolution fulminante avaient très peu d'occasions de soin, donc un risque absolu de recevoir des soins suboptimaux moindre en comparaison à des enfants ayant des infections de progression plus lente et donc plus d'occasions de soins. Il y avait alors un risque de confusion sur la sévérité intrinsèque de la maladie. Pour se défaire de ce risque de confusion, il est possible d'évaluer des soins « obligatoires » c'est-à-dire des soins qui devraient être administrés à tous les patients atteints d'IBS comme cela a été fait pour l'évaluation des soins dans les hémorragies du post-partum.(200)

Enfin aucune des études présentées en introduction de cette partie II ne s'intéressait aux déterminants des soins suboptimaux. Cependant, mieux comprendre les mécanismes de survenue des soins suboptimaux permettrait de mieux cibler les actions correctrices. Une des difficultés dans l'analyse des déterminants des soins suboptimaux réside dans le fait que pour un même patient, les soins ne sont pas indépendants entre eux. L'étude des déterminants des soins suboptimaux à l'échelle individuelle demande donc de tenir compte de cette non-indépendance en utilisant un modèle multiniveau*.(201)

Les objectifs de ce chapitre étaient donc : (i) de préciser la fréquence de la suboptimalité des soins non pas à l'échelle de la prise en charge globale comme dans le chapitre 1 mais à l'échelle des soins individuels, (ii) d'évaluer la plausibilité du lien causal entre soins

suboptimaux et survenue du décès (conséquences), et (iii) d'étudier les caractéristiques des patients et/ou du système de soins associés aux soins suboptimaux (déterminants).

2. Méthodes

La méthodologie d'identification des patients, de collecte de données et d'évaluation de l'optimalité a été décrite précédemment (cf. PARTIE II- sources de données, page 129). Nous ne décrirons ici que les méthodes d'analyses des soins individualisés à l'intérieur des prises en charge ainsi que les analyses statistiques.

2.1 Optimalité des soins et structure des données

Nous avons retenu 6 soins dits obligatoires pour tous les enfants, c'est-à-dire six soins indiqués pour tout enfant avec une IBS. Ces soins étaient le délai de recours aux soins par les parents (« délai patient ») et 5 soins médicaux : l'évaluation initiale de la gravité (lors du premier contact médical), le délai et la posologie du remplissage vasculaire et le délai et la posologie de l'antibiothérapie. Etant donné qu'il avait été demandé aux experts d'identifier les soins suboptimaux, nous avons considéré que lorsqu'aucune remarque n'était faite à propos d'un de ces soins, il avait été optimal. Nous avons ainsi obtenu une structure de données à 2 niveaux avec un niveau enfant et un niveau soins, chaque enfant étant associé à 6 soins (cf. **Figure 12**, page 171).

2.2 Analyses

Comme pour l'étude précédente, nous avons décrit les caractéristiques des enfants, leurs symptômes, leurs diagnostics finaux et leurs parcours de soins. Nous avons mesuré le degré d'accord entre les experts pour chacun des 6 types de soins étudiés avec le coefficient kappa de Cohen avec cette fois seulement deux conclusions « soin optimal » ou « soin

suboptimal ».(189) Comme pour l'étude précédente, un troisième expert était sollicité en cas de désaccord. Nous avons étudié l'optimalité des 6 soins selon le devenir des enfants et la qualification des médecins ayant administré les soins. Nous avons également évalué la relation entre le nombre de soins suboptimaux par prise en charge et le décès et les déterminants individuels des soins suboptimaux médicaux (le délai patient n'était pas inclus dans cette analyse). Afin de n'avoir que des variables du même niveau (niveau enfant) pour expliquer la variable « décès », nous avons considéré comme variable « soins suboptimaux » la somme des soins suboptimaux pour chaque enfant, cette variable pouvait donc prendre toutes les valeurs entières entre 0 et 6. Nous avons analysé l'association entre les soins suboptimaux et la survenue du décès sans puis avec ajustement (régression logistique*) sur les facteurs de confusion potentiels. Pour identifier les variables confondantes potentielles sur lesquelles ajuster notre analyse, nous avons construit un DAG basé sur les concepts physiopathologiques publiés sur le sepsis sévère (cf. **Figure 13**, page 171).(179) Le DAG a été construit à l'aide du logiciel DAGitty qui indique également le « set minimal » de variables sur lesquelles l'analyse doit être ajustée (ici l'âge des enfants, le phénotype clinique et les antécédents).(202) Nous avons défini le phénotype clinique par plusieurs variables : le diagnostic final (méningite *versus* autre diagnostic) ainsi que deux autres variables reflétant la gravité intrinsèque de l'infection, la présence de signe de gravité à la première consultation et la prise en charge par une équipe du SAMU dès la première consultation.

Pour évaluer les déterminants de la qualité des soins, nous n'avons considéré que les 5 soins médicaux. Nous avons alors utilisé une régression logistique multiniveaux car ce modèle permettait de prendre en compte la structure hiérarchique des données (c'est-à-dire la non indépendance des soins entre eux pour un même enfant). Ce modèle nous a également permis d'inclure des variables reflétant les caractéristiques individuelles des soins, c'est-à-dire les variables de niveau 1 (comme la qualité du soin -optimal/suboptimal- et la qualification du médecin qui a administré le soin), ainsi que des variables reflétant les

caractéristiques des enfants, c'est-à-dire des variables de niveau 2 (comme l'âge des enfants, le diagnostic final et la présence de signes de gravité à la première consultation) (cf. **Figure 12**, page 171). Ces variables ont été identifiées comme celles théoriquement associées aux soins suboptimaux (cf. **Figure 13**, page 171). Nous avons d'abord testé un modèle à intercept aléatoire sans aucune variable (modèle vide) pour estimer la variance de base au niveau enfant. Nous avons ensuite inclus dans ce modèle les variables de niveau 1 puis de niveau 2 pour estimer l'association entre la qualité du soin et les caractéristiques des soins et des enfants. Nous avons évalué le pourcentage de variance expliquée par les variables de niveau 1 et par les variables de niveau 2 définies par les formules suivantes : (variance du modèle avec les variables de niveau 1 – variance du modèle vide)/variance du modèle vide et (variance du modèle avec les variables de niveau 1 et de niveau 2 – variance du modèle vide)/variance du modèle vide, respectivement.

L'hypothèse de gradient de risque linéaire a été testée pour les associations comprenant des variables quantitatives et ces variables ont été transformées en polynôme fractionnaire* du plus petit degré lorsqu'une déviation à la linéarité était observée.

Les analyses statistiques ont été faites avec le logiciel Stata 11 (StataCorp, College Station, TX, USA).

3. Résultats

3.1 Caractéristiques des patients et parcours de soins

Cent dix-neuf enfants étaient éligibles, 5 (4%) ont été exclus du fait d'un dossier incomplet (**Figure 14**, page 172). Vingt-et-un enfants étaient décédés (18% ; IC 95%, 11-25) avant l'admission (n = 1) ou lors du séjour en réanimation (n=20) et 93 étaient vivants à la sortie de réanimation. Les caractéristiques des enfants décédés ont été décrites au chapitre 1 de cette partie II. L'âge médian des 114 enfants inclus était de 2,4 ans (EIQ, 0,7-6,7), le sex-ratio était

de 1,3 (M/F) et la moitié des enfants avait des antécédents notables au moment du diagnostic d'IBS (cf. **Tableau 10**, page 173). Plus de la moitié des enfants (63%) avait des signes de gravité lors de la première consultation médicale. La méningite était le diagnostic le plus fréquent (57%), suivie par le purpura fulminans (35%). *Neisseria meningitidis* était retrouvé dans 47% des cas. La première consultation médicale était assurée par un médecin généraliste dans 66% des cas, un urgentiste dans 25%, un médecin du SAMU dans 9% ; 60% des enfants avaient été hospitalisés à l'issue de la première consultation. Les enfants pour lesquels la première consultation médicale était assurée par le SAMU étaient plus susceptibles de présenter des signes de gravité à la première consultation (100% *versus* 60%, $p = 0,01$).

3.2 Optimalité des soins

Le degré d'accord entre experts était « modéré » pour l'évaluation du délai du recours médical et du délai du remplissage, avec des coefficients kappa de Cohen respectifs à $0,40 \pm 0,06$ et $0,46 \pm 0,09$ ($p < 0,001$). Le degré d'accord était « bon » ou « très bon » pour les 4 autres types de soins (évaluation de la gravité, délai et posologie de l'antibiothérapie, et délai du remplissage vasculaire) avec des coefficients respectifs à $0,78 \pm 0,09$; $0,78 \pm 0,09$; $0,67 \pm 0,09$ et $0,88 \pm 0,09$ ($p < 0,001$). Cinquante-deux enfants ont reçu au moins un soin suboptimal au cours de leur prise en charge précédant l'admission en réanimation et 25% en ont reçu 2 ou plus (cf. **Tableau 10**, page 173). Parmi les 684 soins délivrés, 104 (15% ; IC 95%, 12-18%) étaient suboptimaux. Le retard au recours médical, la sous estimation de la gravité par les médecins et le retard à l'administration de l'antibiothérapie comptaient pour 70% de ces soins suboptimaux (respectivement 22%, 22% et 26%) (**Tableau 11**, page 174). La fréquence des soins suboptimaux dans leur ensemble n'a pas diminué de façon significative avec les années (de 19 à 15% entre 2000 et 2005 ; $p > 0,8$ - Chi-deux de tendance) pas plus que la fréquence de chaque type de soin ($p > 0,2$).

3.3 Variables associées au décès

En comparaison aux enfants décédés, les survivants étaient plus souvent âgés de plus de 5 ans ($p < 0,05$; **Tableau 10**, page 173) et étaient plus souvent atteints de méningite (62% vs 33%, $p = 0,02$). Les deux groupes n'étaient pas statistiquement différents pour les autres caractéristiques démographiques, cliniques ou bactériologiques ni pour le nombre de consultations avant l'admission en réanimation ($p > 0,1$, **Tableau 10**, page 173). La première consultation était plus souvent assurée par un médecin du SAMU pour les enfants décédés (vs un médecin généraliste ou un service d'urgences) : 29% vs 4% ($p < 0,001$). La proportion des soins suboptimaux parmi l'ensemble des soins était plus grande chez les enfants décédés que chez les survivants : 24 versus 13% (IC 95% de la différence de risque : 9-13%). En analyse bivariée, le remplissage vasculaire insuffisant était significativement associé au décès (OR = 8,8 ; IC 95% : 2,23-34,7). Il y avait une tendance statistique à l'association entre la survenue d'un décès et la sous-évaluation de la gravité d'une part et le retard au remplissage d'autre part (respectivement OR = 2,74 ; IC 95%, 0,82-7,03 et 2,92 ; IC 95%, 0,84-10,1) (cf. **Tableau 11**, page 174). Après ajustement sur les facteurs de confusion, chaque soin suboptimal (variable continue, sans déviance à la linéarité) augmentait l'OR de décès de 65% (ORa = 1,65 ; IC 95%, 1,07-2,54, $p = 0,02$) (cf. **Tableau 11**, page 174). Concernant l'âge des enfants, chaque année (variable continue, sans déviance à la linéarité) diminuait le rapport de cote de décès de 18% (ORa 0,82 ; IC 95%, 0,68-0,99, $p = 0,04$). Le diagnostic de méningite était également associé à un moindre risque de décès en comparaison aux autres diagnostics (ORa 0,31 ; IC 95%, 0,10-0,98, $p = 0,047$). A l'inverse, une première consultation assurée par le SAMU était associée à une issue défavorable (ORa 8,72 ; IC 95%, 1,76-43,28, $p = 0,008$).

3.4 Déterminants de la qualité des soins

La répartition des 570 soins médicaux reçus par les enfants avant l'admission en réanimation était différente selon la qualification des médecins (**Tableau 12**, page 175). La proportion de

soins suboptimaux était de 33% pour ceux administrés par les médecins généralistes, de 30% pour ceux administrés par des médecins urgentistes adultes, de 24% pour ceux administrés par des médecins au sein d'un service de pédiatrie, de 10% pour ceux administrés par des médecins d'urgences pédiatriques et de 7% pour ceux administrés par un médecin du SAMU ($p < 0,001$). Le pourcentage de sous-évaluation de la gravité était de 30% (17/56) pour les médecins généralistes, de 9% (4/45) pour les médecins des urgences pédiatriques et de 0% (0/11) pour les médecins du SAMU ($p = 0,001$). La proportion de retard à l'antibiothérapie était de 50% pour les médecins généralistes (10/20), de 20% (14/70) pour les médecins des urgences pédiatriques et de 0% (0/17) pour les médecins du SAMU ($p = 0,02$). Les proportions de soins suboptimaux ne différaient pas significativement selon les médecins impliqués pour les autres types de soins (posologie de l'antibiothérapie et posologie et délai du remplissage).

En analyse bivariée, les enfants de moins de 1 an avaient un risque accru de recevoir des soins suboptimaux en comparaison aux enfants de plus de 1 an (**Tableau 12**, page 175) et le fait d'avoir reçu un soin par un médecin des urgences pédiatriques ou un médecin du SAMU était associé à un risque moindre de soin suboptimal en comparaison au médecin généraliste (respectivement, OR = 0,23 ; IC 95%, 0,13-0,42 ; et OR = 0,15, IC 95%, 0,05-0,40) (**Tableau 12**, page 175). Nous n'avons pas mis en évidence d'association entre le risque de soins suboptimal et le diagnostic final (méningite *versus* autres), ni la présence de signes de gravité à la première consultation médicale. L'optimalité des soins variait de façon significative selon les enfants (nous avons donc observé un « effet enfant » avec le modèle vide, $p < 0,001$). Après ajustement, l'association persistait entre l'optimalité des soins et l'âge de l'enfant (dichotomisé en 4 classes ou transformé en polynôme) d'une part et la qualification des médecins d'autre part (**Tableau 12**, page 175). La variance du modèle vide était de 1,56, celle du modèle avec la variable de niveau 1 (qualification du médecin) était de 1,54 et celle du modèle complet (avec les variables de niveau 1 et de niveau 2) était de 1,21. La variable de

niveau 1 expliquait donc 2% de la variance et celles de niveau 2 expliquaient 21% de la variance.

4. Discussion

4.1 Principaux résultats

Cette étude nous a permis de montrer que les soins suboptimaux étaient associés de façon indépendante et avec un gradient de risque linéaire à un risque accru de décès avec un ORa de 1,65 (IC 95%, 1,07-2,54). Certains types de soins suboptimaux étaient associés au décès de façon significative comme le remplissage vasculaire insuffisant. Une tendance statistique (p compris entre 0,08 et 0,1) apparaissait entre un retard au remplissage vasculaire ou une sous-évaluation de la gravité par le médecin et la survenue d'un décès. Certaines caractéristiques des patients étaient également indépendamment associées à un risque accru de décès comme le jeune âge ou le fait d'avoir un diagnostic autre qu'une méningite (purpura fulminans notamment). Enfin, les soins suboptimaux étaient plus fréquents chez le jeune enfant de moins d'un an et lorsque que les soins étaient effectués par un autre médecin qu'un médecin exerçant au sein d'un service d'urgences pédiatriques ou du SAMU.

4.2 Validité interne

Sélection de la population

Avec un recrutement en population, nous avons minimisé le risque de biais de sélection et nous avons pu vérifier l'exhaustivité des décès grâce aux données du CépiDC (voir chapitre sources de données) et recouper 2 sources de données pour identifier les enfants non décédés atteints d'IBS et admis en réanimation. Cependant, il se peut que, de façon exceptionnelle et non systématique, certains enfants domiciliés dans les 2 départements de l'étude aient été admis dans une autre réanimation pédiatrique limitrophe (CHU d'Angers par exemple). Nous

n'avons pas inclus les enfants atteints d'IBS (au sens de la définition internationale) qui avaient survécu mais qui n'avaient pas été admis en réanimation. Nous pourrions émettre l'hypothèse que les enfants admis en réanimation étaient plus susceptibles d'avoir reçu des soins suboptimaux potentiellement responsables de leur dégradation clinique et que nous ayons ainsi surévalué la fréquence des soins suboptimaux. Un tel biais de sélection aurait comme conséquence une sous-évaluation de l'association retrouvée entre soins optimaux et décès.

Evaluation de l'optimalité

La mesure de l'optimalité était assurée par 2 experts indépendants et à l'aveugle du diagnostic et de l'issue pour renforcer leur objectivité. La robustesse de cette mesure est confirmée par un bon coefficient d'accord entre les experts. Cependant, tous les dossiers expertisés étaient ceux d'enfants admis en réanimation donc avec une infection particulièrement sévère. Les experts n'étaient pas à l'aveugle de l'admission en réanimation, ce qui a pu rendre leur jugement plus « sévère », c'est-à-dire qu'ils ont pu conclure plus souvent à des soins suboptimaux.

Evaluation des déterminants et des conséquences

Nous avons pris en compte la non-indépendance des soins pour un même enfant en utilisant un modèle multiniveau pour étudier les déterminants des soins suboptimaux. Cependant, nous n'avons pu utiliser de modèle multiniveau pour évaluer la relation entre les soins suboptimaux et le décès car ce modèle permet d'expliquer des variables de niveau 1 (soins) avec des variables de niveau 2 (enfants) mais pas l'inverse.(201) Nous avons vérifié l'hypothèse de gradient linéaire de risque pour les variables quantitatives et utilisé des polynômes fractionnaires en cas de déviance à la linéarité.(203)

La force de l'association entre soins suboptimaux et décès est restée stable après

ajustement sur les facteurs confondants potentiels identifiés par le DAG. Ces facteurs confondants étaient l'âge des enfants, les antécédents et le phénotype clinique approximé par le diagnostic final et la gravité initiale de la maladie (elle-même approximée par une première consultation assurée par le SAMU). Cependant, la mesure de certaines variables a pu être limitée du fait du caractère rétrospectif de notre étude et certaines variables pouvant être impliquées dans la relation entre soins suboptimaux et décès n'ont pas été mesurées. L'utilisation d'un parcours de soin comportant une première consultation par le SAMU comme indicateur de gravité est probablement une évaluation objective et fiable étant donné que la fonction de ces unités médicales mobiles en France est de prendre en charge les patients en détresse vitale. Cependant, la présence de signes de gravité lors de la première consultation médicale était une mesure moins robuste de la sévérité intrinsèque du fait du caractère rétrospectif de notre étude. Par exemple, les données concernant les signes vitaux, qui sont des éléments clés de l'évaluation d'un enfant fébrile,(204) étaient parfois manquantes, pouvant ainsi amener à une classification erronée de la gravité initiale utilisée pour les analyses statistiques. Ce manque de données concernant les signes vitaux a également pu rendre l'évaluation de la gravité difficile pour les experts. En effet, ceux-ci ne pouvaient évaluer la mauvaise interprétation de ces symptômes seulement lorsqu'ils étaient présents car il était impossible de savoir *a posteriori* si l'absence de renseignement des signes vitaux était le fait d'une absence de mesure de ces signes par les médecins ou d'un manque de transcription de l'information dans le dossier. Certaines variables connues pour leur rôle dans la physiopathologie et la gravité des IBS n'ont pas été mesurées comme l'existence de certains facteurs de virulence ou de certains polymorphismes génétiques de susceptibilité aux infections.(205,206) Cependant, ces variables interviennent en amont du phénotype clinique et sont reliées aux soins suboptimaux et au décès par l'intermédiaire de celui-ci et de la réponse au traitement. Avec une évaluation précise du phénotype clinique, il ne serait donc théoriquement pas justifié d'ajuster sur ces variables pour étudier le lien causal entre soins

suboptimaux et décès. Le graphe que nous avons construit ne tient pas compte de « l'effet temps ». En effet, un soin suboptimal donné à un temps t peut avoir une influence sur la gravité à un temps $t+1$ qui elle-même peut avoir une influence sur la qualité des soins reçus par la suite (cf. **Figure 15**, page 176). De plus, le nombre total de soins peut avoir une influence sur le risque de suboptimalité des soins et reflète également l'évolution naturelle de la maladie (plus l'évolution est lente et plus les occasions de soins sont nombreuses). Un modèle structural marginal* aurait permis de prendre en compte ces variations d'exposition et les facteurs de confusion potentiels (notamment de la sévérité). Le modèle structural marginal aurait alors consisté à étudier la relation entre les soins suboptimaux et le décès en affectant à chaque enfant (i) à un temps donné (t) un poids égal à l'inverse de la probabilité de recevoir son propre niveau d'exposition (aux soins suboptimaux) conditionnellement aux valeurs observées pour cet enfant pour les variables confondantes (dans le cas présent pour la variable S de la **Figure 15**). (207,208) Cependant, notre niveau de précision n'était pas suffisant pour définir la sévérité, paramètre qui ne peut bénéficier d'une évaluation objective qu'au moment d'une consultation médicale. De plus, l'échelle de temps n'est pas homogène pour tous les enfants ce qui rend ce modèle difficilement utilisable.

Enfin, nous avons défini l'issue des enfants de façon binaire (survie *versus* décès), alors qu'il est probable que la population des enfants survivants soit hétérogène avec des enfants ayant développé des séquelles dans les suites de leur IBS (amputation après un purpura fulminans, déficit neurologique ou surdité après une méningite). Il se peut donc qu'une analyse polytomique avec 3 issues (ou binaire composite) aurait permis une analyse plus fine des associations entre soins suboptimaux et issue défavorable. (209)

4.3 Validité externe

Modification de l'épidémiologie

L'évolution de l'épidémiologie bactérienne depuis la période de l'étude peut rendre les

résultats de notre étude difficilement généralisable. En France, les vaccinations conjuguées anti *Haemophilus influenzae*, *Neisseria meningitidis C* et *Streptococcus pneumoniae* à 7 puis 13 valences ont été recommandées pour tous les enfants à partir de 1992, 2009, 2002 et 2009, respectivement. Les infections invasives dues à *H. influenzae* avaient déjà presque totalement disparu à l'époque de l'étude. Les cas rapportés d'infections invasives (méningite et bactériémie) à pneumocoque chez les enfants de moins de 2 ans ont diminué depuis la généralisation de la vaccination avec des incidences à 29 pour 100 000 en 2001, 25 pour 100 000 en 2004 et 18 pour 100 000 en 2012. Cependant aucun changement significatif n'a été observé pour les plus de 2 ans.(210) La couverture vaccinale (c'est-à-dire le pourcentage d'enfants vaccinés parmi ceux concernés par l'indication vaccinale) contre le méningocoque C est trop faible en France pour permettre une diminution des infections invasives à méningocoque C depuis la période de l'étude.(211)

Des nouvelles recommandations

De nouvelles recommandations ont été publiées concernant la prise en charge des IBS et ont peut-être contribué à une diminution de la fréquence des soins suboptimaux.(16,186) Une étude observationnelle multicentrique internationale évaluant les prises en charge aux urgences et/ou en soins intensifs de 1794 adultes en sepsis sévère ou en choc septique a montré que le respect de ces recommandations était associé à un moindre risque de décès avec une réduction de l'odds ratio ajusté sur les facteurs confondants de 36% lorsque toutes les recommandations pour les 3 premières heures de prise en charge étaient respectées.(212) Cependant l'ensemble de ces recommandations n'était respecté que dans 19% des cas 10 ans après la première campagne « *surviving sepsis* ». Nous n'avons pas non plus observé de diminution de la fréquence des soins suboptimaux au cours des 5 années complètes de l'étude, cela indique que la simple diffusion de recommandations n'est pas suffisante pour changer les pratiques comme cela a déjà été démontré dans d'autres situations cliniques.(213)

4.4 Plausibilité du lien causal

Le *gold standard* pour démontrer la causalité en recherche médicale est l'essai randomisé contrôlé en double aveugle. Cependant, comme nous l'avons évoqué en introduction générale, une telle approche ne serait pas éthique dans le cas des IBS. Nous avons donc étudié les conséquences et les déterminants des soins suboptimaux dans la prise en charge initiale des IBS de l'enfant avec une étude observationnelle. En reprenant les critères de Hill pour évaluer la plausibilité du lien causal, nous pouvons conclure que : (i) il existe une association statistiquement significative entre chaque soin suboptimal et le décès avec un OR à 1,65, (ii) cette association persiste après ajustement sur les facteurs de confusion, (iii) il existe une relation dose-effet avec un gradient de risque linéaire, (iv) le lien temporel est probablement respecté même si la prise en compte de la gravité initiale de la maladie dans nos modèles n'était pas parfaite, (v) les résultats sont cohérents avec ceux de la littérature et (vi) il existe un support physiopathologique pour penser qu'une antibiothérapie précoce va limiter l'extension de l'infection et qu'un remplissage vasculaire adapté va permettre d'augmenter le débit cardiaque et le transport en oxygène. Nous n'avons pas évalué les soins administrés en réanimation qui agissent comme facteurs intermédiaires* entre les soins initiaux et l'issue. Nous ne pouvons donc pas évaluer l'effet direct des soins suboptimaux administrés avant l'admission en réanimation sur le décès mais simplement l'effet total. Le lien causal entre soins suboptimaux et décès est plausible, il semble donc justifier de construire des plans d'actions pour les éviter.

4.5 Les pistes pour optimiser les soins

L'analyse des déterminants des soins suboptimaux nous a permis d'identifier des cibles potentielles pour mettre en place des actions d'amélioration de la prise en charge initiale des enfants atteints d'IBS. Le jeune âge (moins de 1 an) était indépendamment associé à un risque accru de soins suboptimaux alors que les soins administrés par les médecins des urgences

pédiatriques ou du SAMU étaient moins à risque de suboptimalité. Dhamar et coll. avait fait les mêmes conclusions après avoir mené une étude rétrospective évaluant la qualité des soins reçus par 304 enfants présentant des signes de gravité à l'admission (détecté grâce à un score de triage à l'admission aux urgences) et pris en charge dans 5 services d'urgence en Californie entre 2000 et 2003. Après ajustement sur les facteurs de confusion et en utilisant un modèle multiniveau (un niveau soins et un niveau médecin), il avait en effet mis en évidence un risque accru de soins suboptimaux pour les jeunes enfants et les soins dispensés des médecins généralistes (*versus* par des pédiatres urgentistes).(214) Selon une étude qualitative de Gilleland et coll., le jeune âge apparaissait également comme un facteur de risque de soins suboptimaux dans la prise en charge des enfants sévèrement malades en milieu hospitalier non spécialisé en pédiatrie.(215) Les actions correctrices devraient donc cibler la formation des médecins généralistes à la reconnaissance des signes de gravité chez le nourrisson fébrile et à la prise en charge précoce des enfants suspects d'IBS (avis spécialisé urgent, médicalisation du transport, administration immédiate d'antibiotique en cas de purpura extensif). Les situations d'IBS étant relativement rares en comparaison aux épisodes fébriles liés à des infections virales bénignes, la mise en situation sous forme d'atelier de simulation peut-être un moyen de maintenir la vigilance chez les médecins en exercice.(216) L'*e-learning* est également un moyen de développer les compétences des professionnels de santé pour une meilleure reconnaissance de l'enfant sévèrement malade.(217)

Le délai entre l'apparition des premiers symptômes et le premier recours médical était considéré comme trop long dans 20% des cas et contribuait à 22% des soins suboptimaux. Nous n'avons pas pu étudier les déterminants des soins parentaux par manque d'information concernant les caractéristiques des parents. Il a été montré dans une étude française que les parents de jeunes enfants ne connaissaient pas la gravité du rash purpurique en cas de fièvre par exemple.(194) Cependant le purpura n'est pas le seul signe d'IBS et peut être tardif.(13,218) La difficulté dans la diffusion de messages concernant les IBS de l'enfant est

de ne pas alerter à tort sur des symptômes qui sont parfois peu spécifiques chez l'enfant, comme nous l'avons vu dans l'introduction du chapitre 1 de la partie I (page 63) et ainsi provoquer une angoisse et des consultations non justifiées. Ainsi la fièvre, les vomissements ou les céphalées sont des symptômes le plus souvent bénins mais qui peuvent également être les premiers signes de méningite. Des campagnes d'information autour de l'association de ces symptômes ont été développées au Royaume-Uni.(219) Une meilleure compréhension de ce qui inquiète ou non les parents face à la santé de leur enfant et notamment face à la fièvre pourrait permettre d'optimiser la détection précoce du sepsis.(220)

5. Figures et tableaux

Figure 12. Structure des données de l'étude DIABACT II.

a = co-variables utilisées pour l'étude des conséquences des soins suboptimaux, b = co-variables utilisées pour l'étude des déterminants des soins suboptimaux.

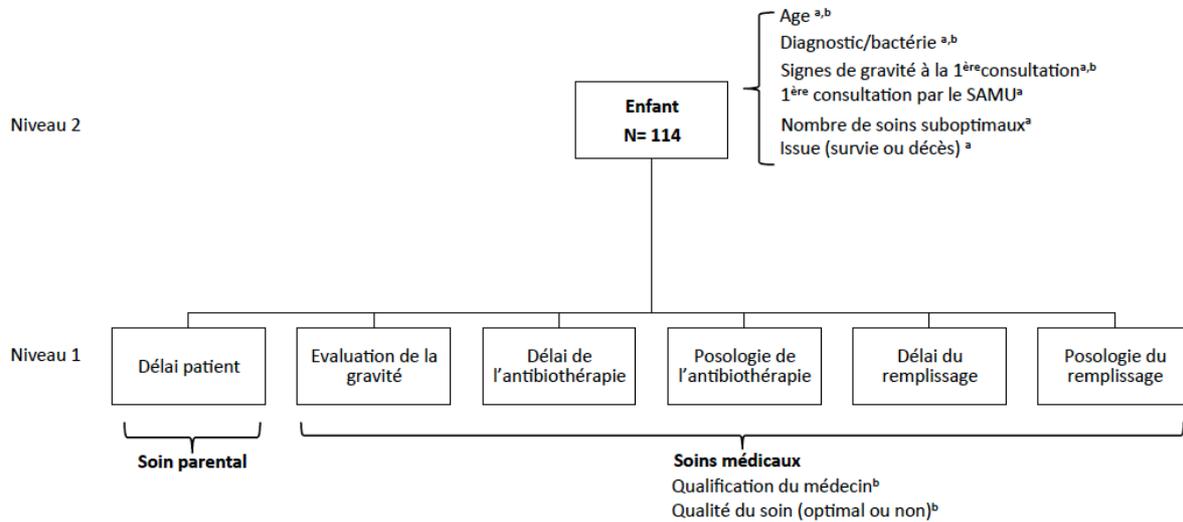


Figure 13. Graphe acyclique orienté (DAG) représentant les liens causaux possibles entre les variables étudiées et pouvant influencer sur la relation entre les soins suboptimaux (exposition) et le décès (*outcome*).

Les cercles roses représentent les variables associées à l'exposition et à l'outcome, les cercles verts représentent les variables associées l'exposition. Les cercles gris sont les variables non mesurées. Les variables entourées sont celles sur lesquelles la mesure de l'association entre soins suboptimaux et décès doit être ajustée.

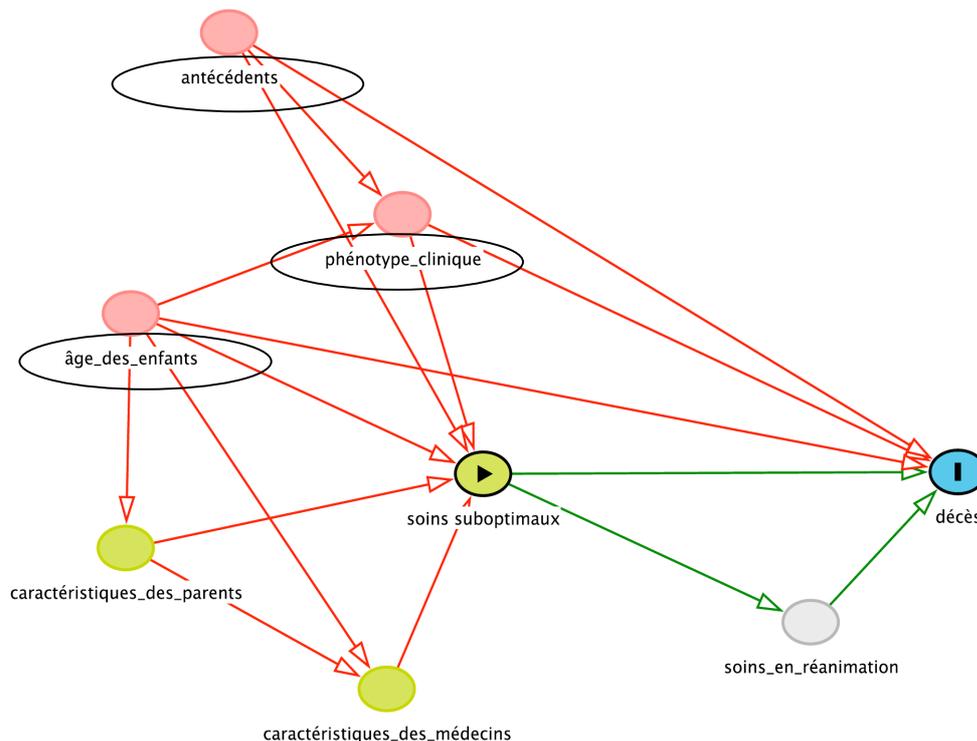


Figure 14. Diagramme de flux des patients inclus dans l'étude DIABACT II.

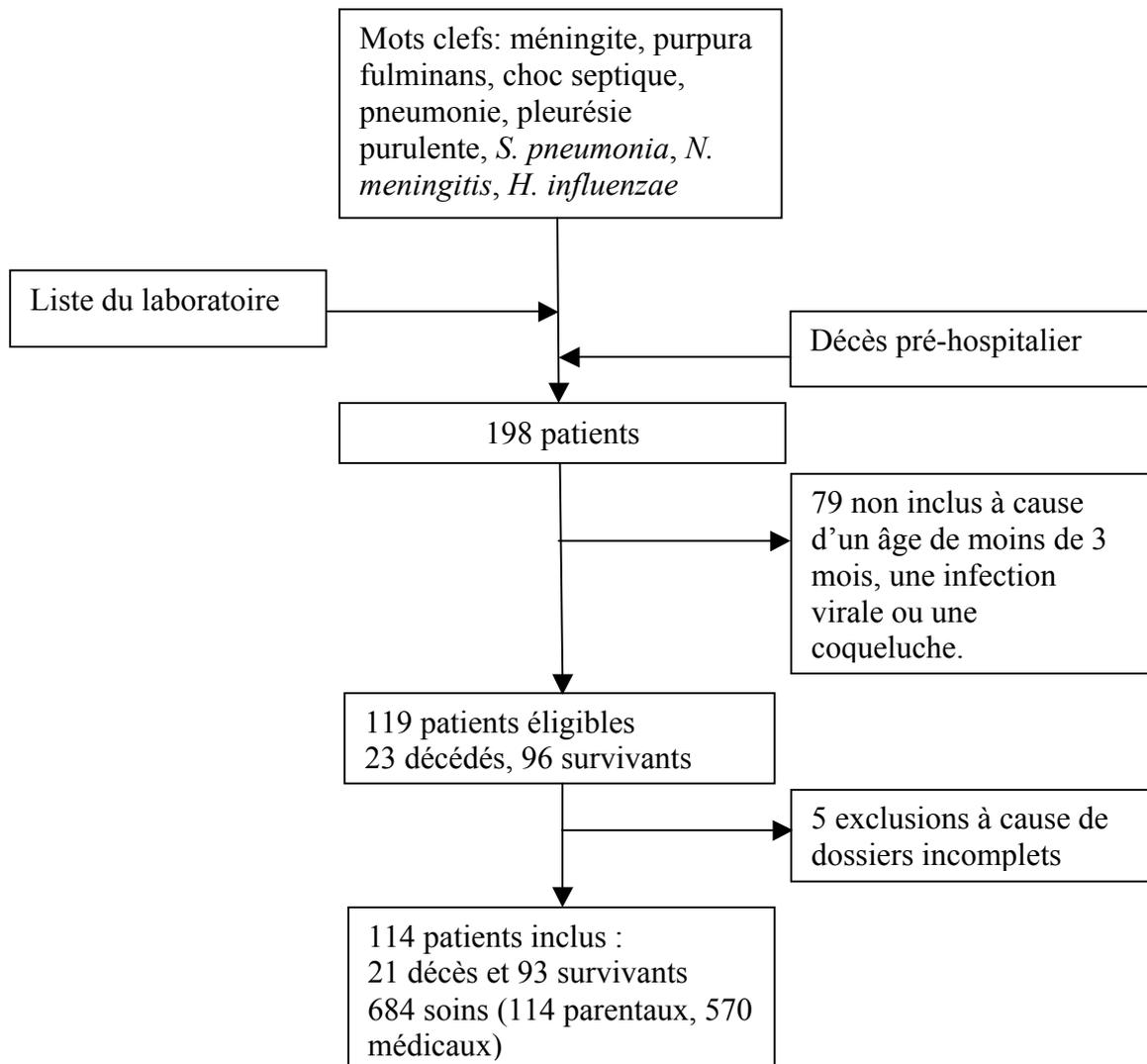


Tableau 10. Caractéristiques des patients, des parcours de soins et de l'optimalité des soins avec l'analyse bi-et multivariée de leur association avec l'issue (décès ou survie).

	Total n=114 (%)	Enfants décédés n=21 (%)	Enfants vivants n=93 (%)	Analyse bivariée			Analyse multivariée [§]		
				OR	IC 95%	p	ORa	IC 95%	p
PATIENTS									
Age, ans									
Médiane ^a [IQR]	2,4 [0,7-6,7]	2,0 [0,9-2,8]	2,9 [0,7-7,1]	0,85	0,73-1,0	0,06	0,82	0,68-0,99	0,04
< 1 an	32 (28)	7 (33)	25 (27)						
1 to 2 ans	21 (18)	4 (19)	17 (18)			0,01			
2 to 5 ans	27 (24)	9 (43)	18 (19)						
≥ 5 ans	34 (30)	1 (5)	33 (36)						
Sex-ratio M/F	1,28	1,33	1,27			0,92			
Antécédents notables, n (%)	39 (52)	7 (33)	32 (34)			0,93	1,16	0,34-3,96	0,81
Signes de gravité présents à la première consultation^b, n (%)	72 (63)	16 (76)	56 (60)			0,17			
Diagnostic final, n (%)									
Purpura fulminans et autres ^c	49 (43)	14 (67)	35 (38)	1	-	-	1	-	-
Méningite	65 (57)	7 (33)	58 (62)	0,30	0,11-0,85	0,02	0,31	0,10-0,98	0,047
Bactérie en cause, n (%)									
<i>Streptococcus pneumoniae</i>	31 (27)	3 (14)	28 (30)						
<i>Neisseria meningitidis</i>	54 (47)	12 (57)	42 (45)						
Autre ^d	11 (10)	4 (19)	7 (8)			0,17			
Sans documentation	18 (16)	2 (10)	16 (17)						
dont purpura fulminans	10 (55)	1 (50)	9 (56)						
PARCOURS DE SOINS									
Première consultation									
Médecin généraliste ou urgentiste, n (%)	104 (91)	15 (71)	89 (96)	1	-	-	1	-	-
SAMU	10 (9)	6 (29)	4 (4)	8,9	2,05-38,6	<0,001	8,72	1,76-43,28	0,008
Nb de consultations, n (%)									
1	68 (60)	15 (71)	53 (57)						
2	37 (32)	6 (29)	31 (33)			0,43			
> 2	9 (8)	0	9 (10)						
QUALITE DES SOINS									
Nb de soins suboptimaux par enfants									
Médiane [EIQ] ^a	1 [0-1]	1 [0-2]	0 [0-1]	1,55	1,06-2,26	0,03	1,65	1,07-2,54	0,02
0	55 (48)	7 (33)	48 (52)						
1	31 (27)	5 (24)	26 (28)						
2	17 (15)	5 (24)	12 (13)						
3	5 (4)	1 (5)	4 (4)						
4	6 (5)	3 (14)	3 (3)						
Nb soins suboptimaux/Nb soins total, % [IC9 5%]	15 [12-18]	24 [16-32]	13 [10-16]			0,003			

Nb, nombre.

[§]Modèle de régression logistique.^aAge et nombre de soins suboptimaux ont été traités comme des variables linéaires continues (pas de déviance à la linéarité).^bLes signes de gravité comprenaient : défaillance hémodynamique, purpura, altération de la conscience, détresse respiratoire syndrome méningé, changement du comportement ou hypotonie.^cLes autres diagnostics étaient : 2 pleuropneumonies et un choc septique sur pyélonéphrite chez un enfant avec uropathie dans le groupe des enfants décédés et 2 chocs septiques sur dermo-hypodermite bactérienne et une trachéite bactérienne dans le groupe des enfants survivants.^dLes bactéries autres étaient, pour les survivants, *Haemophilus influenzae* (n=3), *Streptococcus agalactiae* (n=1), *Staphylococcus aureus* (n=1), et pour les décédés, *Escherichia coli* (n=1), *Streptococcus pyogenes* (n=1), *Salmonella spp* (n=1) et *Mycoplasma pneumoniae* (n=1).

Tableau 11. Description de l'optimalité des 6 types de soins selon l'issue des enfants (décès ou survie) avec l'analyse bivariée de l'association avec l'issue.

Type de soins	Total n = 684 (%)	Enfants décédés n = 126 (%)	Enfants vivants n = 558 (%)	OR	IC 95%	p
Soin parental						
<i>Délai au recours médical (délai patient)</i>						
Suboptimal	23 (20)	6 (29)	17 (18)	1,79	0,60-5,34	0,29
Optimal	91 (80)	15 (71)	76 (82)	1	-	-
Soins médicaux						
<i>Evaluation de la gravité</i>						
Suboptimale	23 (20)	7 (33)	16 (17)	2,40	0,82-7,03	0,10
Optimale	91 (80)	14 (67)	77 (83)	1	-	-
<i>Délai de l'antibiothérapie</i>						
Suboptimal	27 (24)	5 (24)	22 (24)	1	0,33-3,08	0,98
Optimal	87 (76)	16 (76)	71 (76)	1	-	-
<i>Délai au remplissage</i>						
Suboptimal	14 (12)	5 (24)	9 (10)	2,92	0,84-10,1	0,08
Optimal	100 (88)	16 (76)	84 (90)	1	-	-
<i>Posologie du remplissage</i>						
Suboptimal	12 (11)	7 (33)	5 (5)	8,80	2,23-34,7	0,002
Optimal	102 (89)	14 (67)	88 (95)	1	-	-
<i>Posologie de l'antibiothérapie</i>						
Suboptimal	5 (4)	0 (0)	5 (5)	0,67	0,01-6,03	0,29
Optimal	109 (96)	21 (100)	88 (95)	1	-	-

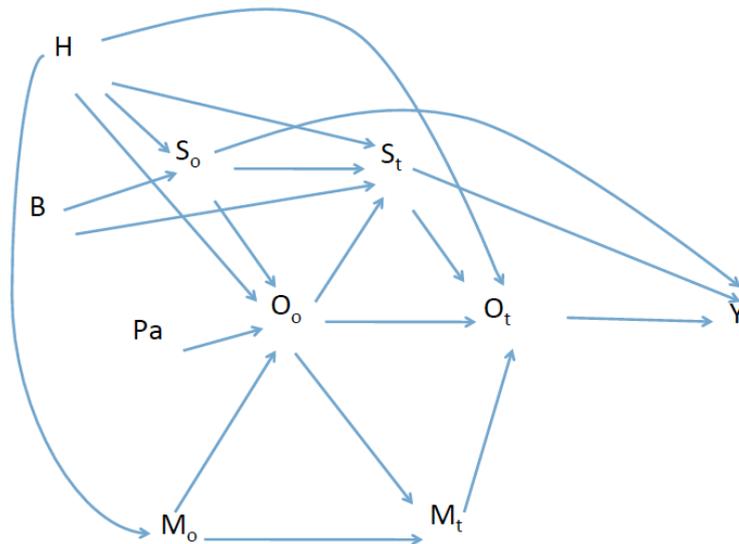
Tableau 12. Déterminants des soins suboptimaux médicaux

	Optimal n=489 (%)	Suboptimal n=81 (%)	Analyse bivariée			Analyse multivariée ***		
			OR	IC 95%	p	ORa	IC 95%	p
Age								
< 1 an	125 (26)	35 (43)	1			1		
1–2 ans	95 (19)	10 (12)	0,38	0,18-0,81	0,009	0,32	0,11-0,98	0,046
2–5 ans	119 (24)	16 (20)	0,48	0,25-0,92	0,02	0,37	0,14-0,98	0,045
≥5 ans	150 (31)	20 (25)	0,48	0,26-0,87	0,01	0,24	0,09-0,64	0,004
Qualification du médecin, n (%)								
Médecin généraliste	55 (11)	27 (33)	1			1		
Urgentiste adulte	16 (3)	7 (9)	0,90	0,33-2,44	0,82	0,63	0,15-2,62	0,53
Urgentiste pédiatrique	322 (66)	37 (46)	0,23	0,13-0,42	<0,001	0,16	0,08-0,35	<0,001
SAMU	83 (17)	6 (7)	0,15	0,05-0,40	<0,001	0,09	0,03-0,31	<0,001
Pédiatre	13 (3)	4 (5)	0,63	0,18-2,13	0,45	0,65	0,11-3,67	0,63
Signes de gravité à la 1^{ère} consultation, n (%)								
Non	182 (87)	28 (13)	1			1		
Oui	307 (85)	53(15)	1,12	0,68-1,84	0,6	1,3	0,59-2,90	0,51
Diagnostic final, n (%)								
Autres	210 (86)	35 (14)	1			1		
Méningite	279 (86)	46 (14)	0,99	0,62-1,59	0,9	0,73	0,34-1,59	0,43

*L'analyse multivariée a été faite avec une régression logistique multiniveaux à intercept aléatoire. **Les associations restaient significatives après transformation de l'âge en polynôme de type ($X = 10/[\text{âge} - 2,5]$), ORa pour l'âge à 1,04, 95% CI 1,01-1,07, $p=0,003$.

Figure 15 Représentation des relations entre les variables potentiellement impliquées dans la relation entre les soins suboptimaux (O) et le décès (Y) en prenant en compte le temps (indice 0 à t).

H, facteurs liés à hôte (âge, antécédents, prise de médicaments, susceptibilité génétique) ; B, facteurs liés à la bactérie ; S, gravité clinique ; M, caractéristiques des médecins ; Pa, caractéristiques des parents.



SYNTHESE, CONCLUSIONS ET PERSPECTIVES

1. Synthèse des résultats

Dans la **partie I** de la thèse portant sur la méthodologie d'évaluation de la suboptimalité d'un type de soin particulier, le diagnostic, nous avons émis l'hypothèse de difficultés méthodologiques rencontrées par les chercheurs lors de l'évaluation de la fréquence, des déterminants et des conséquences des délais diagnostiques. Nous avons alors proposé dans le **chapitre 1**, une liste d'items comprenant des éléments considérés comme critiques (car potentiellement liés à des risques de biais ou d'obstacles à la transportabilité) permettant d'évaluer systématiquement la qualité du *reporting* des études portant sur ces délais diagnostiques. Devant le constat que les points critiques n'étaient rapportés, pour la majorité d'entre eux, que dans moins de la moitié des articles, nous avons jugé nécessaire de poursuivre une démarche de sensibilisation sur ces points critiques en transformant cette simple *checklist* en une véritable *reporting guideline* spécifique aux études sur les délais diagnostiques. Afin de vérifier nos hypothèses concernant ces points critiques, nous avons mené **dans le chapitre 2** une revue systématique méthodologique des outils utilisés par les auteurs de revue systématique pour évaluer les risques de biais et d'obstacles à la transportabilité des études primaires qu'ils avaient incluses. Nous avons constaté que seulement la moitié (51%) des 45 revues systématiques étudiées rapportait une évaluation de la qualité et que celle-ci était le plus souvent inadéquate, c'est-à-dire qu'elle s'attachait plus à la qualité du *reporting* (169 des 232 items) qu'à l'évaluation des risques de biais (85 items) ou d'obstacles à la transportabilité (14 items). Parmi les risques de biais et d'obstacles à la transportabilité soulevés par les auteurs des revues systématiques pour évaluer la qualité des études primaires, seul le risque de biais d'attrition était absent de notre *checklist*. Nous avons pris en compte ce manque en proposant l'utilisation d'un diagramme de flux depuis la population source jusqu'à la population analysée. Ce diagramme de flux a été ajouté à notre *checklist* **dans le chapitre 3** qui était une étape d'affinement et de validation internationale.

Au cours de cette étape, nous avons validé les points qui nous semblaient critiques avec des experts internationaux des domaines concernés. Les discussions avec les experts nous ont également permis de modifier la formulation des items et de construire un document explicatif.

Ainsi, la *reporting guideline* constituée (*checklist* + document explicatif) pourra permettre de guider les auteurs d'étude sur les délais diagnostiques dans la rédaction des articles. Ce travail de construction d'une *reporting guideline* spécifique est la première étape d'un travail plus large de réflexion autour de la méthodologie des études sur les délais diagnostiques qui comportera notamment la construction d'un outil d'évaluation des risques de biais et d'obstacles à la transportabilité et une étude sur le risque de *spin* dans les études sur les délais diagnostiques (cf. perspectives de recherche, pages 185 et 186).

Dans la **partie II**, nous avons évalué l'optimalité des soins au cours de la prise en charge initiale des enfants atteints d'IBS. Nous avons montré **dans le chapitre 1**, par le moyen d'une enquête confidentielle en population, que 76% des prises en charge des enfants décédés d'IBS étaient suboptimales et que les soins suboptimaux les plus fréquents étaient le retard au recours médical, la sous-évaluation de la gravité, le retard à l'antibiothérapie et l'insuffisance du remplissage vasculaire. Devant la fréquence de la suboptimalité des prises en charge, nous avons alors cherché à mieux caractériser chaque soin considéré individuellement et à mesurer l'association entre suboptimalité et décès et étudier l'existence d'un lien causal. **Dans le chapitre 2**, nous avons montré que les soins suboptimaux étaient associés à un risque accru de mortalité avec un ORa à 1,65 et que selon les critères de Hill, un lien causal était plausible entre les soins suboptimaux et la survenue du décès. La mise en place d'actions correctrices semble donc justifiée. L'étude des déterminants des soins suboptimaux nous a permis d'identifier une cible pour ces actions correctrices : la formation

des médecins de soins primaires à la meilleure reconnaissance et prise en charge de l'IBS notamment chez les enfants de moins d'un an.

En confrontant notre expérience de construction d'une *reporting guideline* pour les études sur les délais diagnostiques et notre expérience d'évaluation des soins au travers des études DIABACT nous avons alors pu dégager des éléments de synthèse mais également de nouveaux questionnements et des perspectives de recherche pour l'évaluation des soins suboptimaux.

2. Implications pour la méthodologie d'évaluation des soins

2.1 Sélection de la population

En reprenant les items de la version finale de la *checklist* (**Tableau 6**, page 112) concernant la sélection de la population nous pourrions discuter les critères d'inclusion (item 4, **Tableau 6**) des études DIABACT I et II qui comprenaient l'admission en réanimation ou le décès avant admission en réanimation. Comme nous l'avons évoqué dans la partie II, ces critères d'inclusion ont pu être à l'origine d'une surestimation de la fréquence des soins suboptimaux en ne sélectionnant que les cas les plus graves et donc potentiellement plus à risque de soins suboptimaux et également plus à risque de décès. Cependant le choix de ces critères d'inclusion nous a permis d'assurer un recrutement en population et de n'inclure que les infections les plus graves. En effet, les IBS telles que définies habituellement comprennent par exemple les infections ostéo-articulaires à *Kingella kingae* dont nous savons que certaines formes peuvent guérir spontanément, rendant ainsi l'évaluation de l'optimalité des soins moins pertinente.(221)

Si nous appliquons maintenant le diagramme de flux proposé en **Figure 9** (page 113) aux études DIABACT, nous sommes confrontés à la difficulté de définir la population source. Nous avons défini cette population source comme l'ensemble des enfants présentant les

symptômes qui habituellement enclenchent la procédure diagnostique. Dans le cas des IBS, ces symptômes, comme la fièvre par exemple, sont parfois non spécifiques surtout lorsqu'ils sont isolés. Définir la population source comme l'ensemble des enfants fébriles sur la période donnée reviendrait à considérer que tous les enfants fébriles sont à risque d'IBS et devraient bénéficier d'explorations microbiologiques ce qui n'est pas réalisable, ni souhaitable en pratique. Nous pourrions alors définir la population source comme l'ensemble des enfants présentant des signes de sepsis, cependant certains de ces signes sont biologiques et ne peuvent donc pas être définis par le seul examen clinique. La population source pourrait alors être définie comme l'ensemble des enfants fébriles présentant les signes de gravité clinique proposés pour les études DIABACT (cf. sources de données, section population, page 131). Cependant il faudrait alors s'assurer que ces symptômes soient systématiquement recherchés et notifiés lors de chaque recours médical d'un enfant fébrile. La définition de la population source est donc parfois très difficile notamment dans le cas des IBS de l'enfant où la définition de la population source est elle-même le sujet de nombreuses recherches.^(57,222,223) L'utilisation du diagramme de flux ne peut alors pas permettre de quantifier la population à chaque étape de sélection mais elle peut servir de support de réflexion pour l'interprétation des résultats et éviter le *spin*. Par exemple, en ce qui concerne notre étude sur les IBS une interprétation abusive serait de conclure que : « pour réduire la fréquence des soins suboptimaux, il faut que tous les enfants fébriles soient vus en consultation aux urgences pédiatriques ». La construction d'un diagramme de flux mettant en évidence chaque étape de sélection permet cependant de réaliser que notre population d'étude n'était pas l'ensemble des enfants fébriles mais celle des enfants admis en réanimation pour une IBS.

2.2 Définition et mesure de l'optimalité

Dans la partie I, nous avons étudié la méthodologie d'évaluation du délai diagnostique principalement en tant que variable quantitative. Dans les études DIABACT, nous avons étudié le délai diagnostique de façon dichotomisée en utilisant plutôt le terme de retard diagnostique. En l'absence de critères objectifs et de niveaux de preuve suffisants dans la littérature pour définir le retard diagnostique, ce dernier (et plus généralement un soin suboptimal) a été défini par le jugement de deux cliniciens comme un délai diagnostique (ou un soin) qui, dans leur expérience, a été associé avec des issues défavorables. L'association entre les soins suboptimaux et l'issue défavorable retrouvée dans les études DIABACT pourrait donc traduire la non-indépendance entre ces deux variables mais également légitimer *a posteriori* le caractère non optimal du soin (puisque'il est effectivement associé à une issue défavorable). Pour définir de façon plus objective le retard diagnostique dans les IBS de l'enfant, nous pourrions alors étudier le délai diagnostique de façon quantitative en définissant un T0 et un Tx comme proposé dans notre *checklist* (item 7, **Tableau 6**). La définition du T0 peut cependant poser problème. En effet, si date de début du délai (T0) est définie comme celle de l'apparition chronologique des premiers symptômes alors les enfants avec une IBS d'évolution naturelle lente, du fait d'une bactérie moins virulente ou d'une réponse inflammatoire moins « explosive » par exemple, auront un délai diagnostique potentiellement plus long. Il serait donc préférable de définir le T0 du délai diagnostique comme la date d'apparition des symptômes d'alerte mais nous avons vu dans le chapitre précédent que ces symptômes pouvaient être difficiles à définir dans le cas des IBS. Le jugement de cliniciens expérimentés, que nous avons utilisé dans les études DIABACT pour définir l'optimalité des soins, pourrait alors permettre de définir le T0.

2.3 Etude de la relation entre optimalité du soin et l'issue

L'objectif de l'étude de la relation entre un soin donné et l'issue peut être, comme nous venons de le voir, de définir l'optimalité. Nous pouvons alors étudier la relation entre un soin « quantitatif » (comme le délai diagnostique ou thérapeutique) ou un soin « qualitatif » (comme le fait d'avoir reçu ou non un traitement) et l'issue de santé. L'objectif peut également être de vérifier qu'un soin supposé suboptimal est en effet suboptimal car associé à une issue défavorable. On étudiera alors si un soin supposé suboptimal (absence de consultation médicale rapide devant un enfant fébrile et hypotonique par exemple) est associé à une issue défavorable. Lorsque la « mesure » du soin n'est pas strictement objective (comme un délai diagnostique qui serait défini par l'intervalle séparant une date objective d'apparition d'un symptôme très spécifique de la maladie et une date d'examen complémentaire par exemple) ou lorsque la définition de la suboptimalité ne se fait pas en référence à un consensus basé sur un haut niveau de preuve, le recours à des experts pour l'étude de ces soins peut amener un biais de classement si les experts ne sont pas à l'aveugle de l'issue. Il est donc très important de maintenir les experts à l'aveugle des caractéristiques des patients pouvant influencer sur le jugement de la prise en charge (item 10 A et B, **Tableau 6**). Dans le cas de l'étude des soins dans la prise en charge initiale des IBS de l'enfant, nous pourrions améliorer l'objectivité du jugement en faisant évaluer les soins au fur et à mesure de la prise en charge.

L'identification et la prise en compte des facteurs de confusion sont également des éléments très importants dans l'évaluation de l'association entre un soin et l'issue de santé (item 12, **Tableau 6**). Dans le cas des IBS nous avons été confrontés à la difficulté d'évaluer et de mesurer de façon précise la sévérité intrinsèque de l'IBS. Un recueil prospectif des données, le calcul de scores de gravité utilisés en réanimation pédiatrique comme *Pediatric Risk of Mortality* (PRISM), le *Pediatric Index of Mortality* ou le *Pediatric Logistic Organ*

Dysfunction et l'étude des facteurs de virulence bactériens et de la réponse inflammatoire de l'hôte pourraient nous permettre de mieux prendre en compte ce facteur de confusion.

3. Perspectives de recherche

3.1 Outil spécifique d'évaluation des risques de biais

Dans le chapitre 2 de la partie I, nous avons montré qu'un faible nombre de revues systématiques évaluait la qualité des études primaires de façon adéquate. Nous avons alors émis l'hypothèse qu'un outil spécifique pour évaluer les risques de biais et d'obstacles à la transportabilité des études primaires sur les délais diagnostiques était nécessaire pour la conduite des revues systématiques ou plus généralement pour guider la lecture critique des études primaires. Nous avons aussi vu que la construction d'un tel outil se heurtait à des difficultés spécifiques aux études observationnelles. En effet, les outils comme QUADAS-II ont été bâtis sur la démonstration de l'effet de la variation de certains paramètres (comme la prévalence de la maladie par exemple) sur la performance des tests diagnostiques.(113) Une telle démonstration paraît plus difficile dans le domaine des délais diagnostiques car elle nécessiterait de pouvoir comparer les résultats d'un grand nombre d'études ne variant que sur le point méthodologique associé au risque de biais à démontrer (caractère systématique ou non de la procédure diagnostique, ou évaluation du délai diagnostique par des experts à l'aveugle ou non de l'issue par exemple) et dans un contexte de soin similaire. Cependant nous pensons pouvoir tenter une telle démonstration de l'impact de la variabilité de certains points méthodologiques sur les résultats. En effet, nous disposons de bases de données d'études primaires ($n > 200$) issues de ce travail doctoral et d'un autre travail doctoral.(19) Nous pourrions étudier l'impact de certains choix méthodologiques (aveugle, ajustement...) ou stratégiques (présence d'un méthodologiste ...) sur la fréquence d'une conclusion sur l'existence d'une relation entre un délai diagnostique long et une issue défavorable.

Une fois les risques de biais démontrés, il faudra alors suivre la même procédure de construction et de validation que celle décrite dans la partie I pour la *reporting guideline* mais en formulant les items de façon évaluative (par exemple : « la méthode employée pour mesurer le délai diagnostique entraîne-t-elle un risque de biais de classement ?). Une étape de test de la variabilité inter- et intra-évaluateur sera nécessaire avant toute diffusion et implémentation de l’outil.

3.2 *Spin* dans les études sur les délais diagnostiques

Nous avons été frappés lors de notre revue systématique du chapitre 1 par la discordance entre les conclusions et les résultats de certains articles. En effet, deux tiers des articles concluaient qu’une action correctrice de réduction des délais était nécessaire mais très peu rapportaient les éléments justifiant cette action comme le caractère évitable du délai ou l’impact du délai sur l’état de santé. Cette discordance est comparable à celle qui existe entre les conclusions de certains essais cliniques sur les effets bénéfiques d’un traitement et les résultats de l’essai. Boutron et coll. ont ainsi identifié 72 essais randomisés contrôlés dont les résultats concernant l’objectif principal ne montraient pas d’effet significatif ($p > 0,05$) de l’intervention étudiée sur le critère d’évaluation primaire. Ils ont montré que dans 40% des articles les auteurs utilisaient le *spin* dans au moins deux sections de leur article. Le *spin* était défini comme le fait de focaliser le titre, l’abstract, les résultats ou la discussion sur les seuls résultats significatifs comme ceux d’analyses en sous groupe par exemple, d’interpréter de façon erronée un résultat statistique non significatif concernant le critère d’évaluation principal (conclusion à une équivalence alors que l’objectif est de montrer la non infériorité par exemple) ou de conclure à un effet bénéfique d’une intervention alors que les résultats ne sont pas significatifs.(112) Il serait intéressant de définir les différents éléments de *spin* pour les études sur les délais diagnostiques et d’en évaluer l’importance dans la littérature sur les

délais diagnostiques en élargissant notre revue systématique faite au chapitre 1 à la littérature adulte.

3.3 Evaluation prospective de l'optimalité de la prise en charge des IBS de l'enfant (DIABACT III)

Les études DIABACT I et II étaient rétrospectives et basées sur 2 départements ce qui limitait la précision des données recueillies, la puissance statistique et la transportabilité des résultats à d'autres contextes de soin. Une partie du travail doctoral a consisté à rédiger, faire financer (PHRC inter-régional BRD 08-06-T, montant = 98 786 euros) et conduire une étude intitulée DIABACT III qui pourrait prendre en compte ces limites (protocole en annexe 6, page 223). Cette étude avait le même *design* d'enquête confidentielle avec comité d'experts et a inclus prospectivement tous les enfants de 1 mois à 16 ans avec une IBS telle que définie dans DIABACT I et II dans 10 départements de la région Grand-Ouest (Loire Atlantique, Vendée, Ile-et-Vilaine, Finistère, Côtes d'Armor, Morbihan, Indre, Indre et Loire, Vienne et Deux-Sèvres). Une collecte détaillée d'informations a été possible notamment concernant les caractéristiques des enfants, de leur parents, de leurs parcours de soin, des éléments cliniques permettant une meilleure mesure de la sévérité (éléments plus précis des examens cliniques successifs, mesure du score de sévérité PRISM-II à l'arrivée en réanimation) et également d'une évaluation plus précise de l'issue en prenant en compte les séquelles. Deux-cent soixante dix enfants ont été inclus entre août 2009 et janvier 2014 et les expertises des 270 dossiers n'ont été finalisées qu'en août 2015 ce qui n'a pas permis d'inclure les résultats comme un chapitre de thèse. Les premières analyses concernant la description des caractéristiques de la population d'étude ont cependant été faites. L'âge médian des enfants inclus était de 26,6 mois (EIQ : 7,1-68,8) et 49% des patients avaient moins de 2 ans ; 18% avaient des antécédents notables. Le diagnostic final était une méningite pour 31% des

enfants, un purpura fulminans pour 22% ; un méningocoque était isolé chez 39% et un pneumocoque chez 25%. A l'issue de l'hospitalisation en réanimation, 76% des enfants étaient vivants sans séquelles, 13% vivants avec séquelles immédiatement diagnostiquables et 11% étaient décédés. Des signes de gravité étaient présents lors de la première consultation pour 51% d'entre eux. La première consultation était réalisée par un médecin libéral dans 55% des cas, par un service d'urgences dans 36% des cas et par le SAMU dans 6% des cas. Le nombre de consultations avant hospitalisation en réanimation était supérieur à 2 dans 56% des cas. Les délais totaux médians de prise en charge (entre les premiers symptômes et l'admission en réanimation) pour les enfants survivants sans séquelle, survivants avec séquelles et décédés étaient respectivement de 36 (EIQ : 19,3-92), 75 (EIQ : 21-101) et 22,5 heures (15,5-65). Les premiers résultats concernant l'évaluation de l'optimalité de la prise en charge globale par nos experts montrent un coefficient kappa de Cohen avec pondération quadratique faible à 0,34 avec un taux d'accord de 80%. Vingt-et-une prises en charge ont été évaluées avec une discordance « prise en charge optimale » versus « prise en charge suboptimale ». L'analyse de la concordance soin par soin sera la prochaine étape de l'analyse.

3.4 Outil de repérage des parcours de soins à risque de suboptimalité (DIABACT IV)

La procédure d'évaluation des soins dans les prises en charge d'enfants atteints d'IBS utilisée dans les études DIABACT I, II et III demande un investissement très important et chronophage notamment pour les experts. Si cette démarche est tout à fait adaptée à un objectif de recherche, elle peut s'avérer irréalisable pour l'évaluation des soins en « routine ». Il serait pourtant important d'avoir un outil de mesure en routine de la qualité de la prise en charge des enfants atteints d'IBS notamment pour mesurer l'impact des campagnes de formation médicale continue et d'éducation à la santé. Nous proposons donc d'identifier des caractéristiques (indicateurs) extractibles des données du Programme Médicalisé des

Systèmes d'Informations (PMSI) et du Système National d'Information Inter-régimes de l'Assurance Maladie (SNIIRAM) et associées à un risque plus élevé de suboptimalité dans la prise en charge précoce (avant l'arrivée en réanimation) des enfants de 0 à 18 ans atteints d'IBS communautaires sévères en France. L'objectif n'est pas de se substituer à l'analyse détaillée des prises en charge mais de proposer des indicateurs d'alerte permettant de repérer *a posteriori* des prises en charge « à risque » d'être suboptimales et de déclencher un processus d'analyse de l'optimalité sur ces dossiers. Nous pourrions ainsi estimer en routine l'optimalité des prises en charge dans un but d'évaluation des pratiques et également de santé publique pour mesurer l'impact de certaines campagnes de formation et d'éducation à la santé. Une partie du travail doctoral a consisté à rédiger, faire financer (PREPS 2013 RC14_0119, montant = 247 835 euros) cette étude dont les inclusions ont débuté en mars 2015 (protocole en annexe 7, page 225). Un total de 512 enfants doit être inclus dans 13 centres de réanimations pédiatriques répartis sur tout le territoire français (Angers, Besançon, Bordeaux, Brest, Grenoble, Lille, Lyon, Montpellier, Nantes, St Etienne et 3 centres parisiens : Armand-Trousseau, Robert-Debré et Necker-Enfants malades). Nous avons pris en compte les limites des études DIABACT I et II mais également optimisé le recueil de données et le processus d'évaluation des soins qui se fera par les experts directement sur le cahier électronique de recueil de données au fur et à mesure de la prise en charge. Cela nous dispensera donc de faire des fiches de synthèse pour chaque enfant. La constitution d'une biocollection a également été prévue afin de pouvoir étudier certains facteurs génétiques de susceptibilité aux infections et mieux évaluer la virulence des souches. Nous pourrions ainsi compléter notamment les variables d'ajustement dans l'étude des conséquences des soins suboptimaux.

Un schéma d'étude articulé sur 2 phases d'investigation est prévu. Dans un premier temps, nous testerons les associations éventuelles entre des candidats indicateurs

SNIIRAM/PMSI d'intérêt (variables prédictives) et le caractère suboptimal de la prise en charge (variable prédite), via la cohorte constituée par l'étude DIABACT III. Les variables prédictives candidates et disponibles dans les bases de données du SNIIRAM sont les suivantes : le statut de survie à l'issue du séjour en réanimation –survie ou décès, l'âge de l'enfant, la présence d'une affection de longue durée, le nombre de consultations médicales précédant immédiatement l'hospitalisation, la présence et le nombre de transferts inter-hospitaliers au cours de la prise en charge de l'infection, le nombre de consultations médicales par an, le nombre, la nature et la date des vaccinations prescrites ainsi que le nombre et la nature des prescriptions d'antibiotiques antérieures ou encore la date de la première consultation pour l'épisode en cours. Les prédicteurs identifiés seront éventuellement combinés dans un outil prédictif composite. Puis nous validerons ces prédicteurs et l'éventuel outil prédictif composite à partir des données d'une étude prospective multicentrique nationale incluant des enfants admis en réanimation pour infection sévère.

3.5 Etude expérimentale des facteurs de virulence

La prise en compte des facteurs de virulence des bactéries et de la réponse inflammatoire est également un moyen de comprendre la variabilité dans la sévérité des IBS. Nous avons donc prévu de caractériser, sur un modèle murin de pneumopathie, les facteurs de virulence de souches de pneumocoque isolées dans un contexte de pneumonie ou de pleuro-pneumonie chez des enfants de la cohorte DIABACT III. Une fois l'infection pulmonaire installée, nous caractériserons la virulence de ces souches en évaluant la dissémination des souches bactériennes et l'ampleur de la réponse inflammatoire qu'elles induisent. Nous évaluerons la réponse inflammatoire locale, par l'étude anatomo-pathologique de coupes de poumons infectés, par détermination de l'activité myéloperoxydase des polynucléaires neutrophiles, et dosage des cytokines pro-inflammatoires : Tumor Necrosis Factor α (TNF- α), interleukine 1 β

(IL-1 β) et la chimiokine Macrophage Inflammatory Protein 2-alpha (MIP 2- α). Nous pourrions alors étudier les associations entre les degrés de virulence des souches (définis par la capacité de dissémination et les différents niveaux de réponses cytokiniques) et l'optimalité des soins d'une part et l'issue des enfants infectés par ces souches d'autre part. Cette étude sera réalisée par l'équipe d'accueil UPRES EA 3826 « Thérapeutiques cliniques et expérimentales des infections » à l'université de Nantes.

REFERENCES BIBLIOGRAPHIQUES

1. Kohn LT, Corrigan J, Donaldson MS. To err is human: building a safer health system. Washington, D.C.: National Academy Press; 2000.
2. James JT. A new, evidence-based estimate of patient harms associated with hospital care. *J Patient Saf.* 2013;9:122-8.
3. Bouvier-Colle MH. Enquêtes confidentielles avec comités d'experts : une méthode d'évaluation des soins. L'exemple de l'Obstétrique. *Rev Epidemiol Sante Publique.* 2002;50:203-17.
4. Leape LL, Brennan TA, Laird N, Lawthers AG, Localio AR, Barnes BA, Hebert L, Newhouse JP, Weiler PC, Hiatt H. The nature of adverse events in hospitalized patients. Results of the Harvard Medical Practice Study II. *N Engl J Med.* 1991;324:377-84.
5. Saucedo M, Deneux-Tharoux C, Bouvier-Colle M-H, French National Experts Committee on Maternal Mortality. Ten years of confidential inquiries into maternal deaths in France, 1998-2007. *Obstet Gynecol.* 2013;122:752-60.
6. Hayward RA, Hofer TP. Estimating hospital deaths due to medical errors: preventability is in the eye of the reviewer. *JAMA.* 2001;286:415-20.
7. Centre d'Epidémiologie sur les Causes médicinales de Décès (CépiDc) [Internet]. <http://www.cephdc.inserm.fr>
8. Craver R, Springer J, Begue R. Infections in a children's hospital autopsy population. *Fetal Pediatr Pathol.* 2014;33:135-44.
9. Plunkett A, Tong J. Sepsis in children. *BMJ.* 2015;350:h3017-h3017.
10. Hartman ME, Linde-Zwirble WT, Angus DC, Watson RS. Trends in the epidemiology of pediatric severe sepsis. *Pediatr Crit Care M.* 2013;14:686-93.
11. Franc C, Le Vaillant M, Rosman S, Pelletier-Fleury N. La prise en charge des enfants en médecine générale : une typologie des consultations et visites [Internet]. 2007. <http://www.drees.sante.gouv.fr/IMG/pdf/er588.pdf>
12. Rudinsky SL, Carstairs KL, Reardon JM, Simon LV, Riffenburgh RH, Tanen DA. Serious bacterial infections in febrile infants in the post-pneumococcal conjugate vaccine era. *Acad Emerg Med.* 2009;16:585-90.
13. Thompson MJ, Ninis N, Perera R, Mayon-White R, Phillips C, Bailey L, Harnden A, Mant D, Levin M. Clinical recognition of meningococcal disease in children and adolescents. *Lancet.* 2006;367:397-403.
14. Wilson A, Grimshaw G, Baker R, Thompson J. Differentiating between audit and research: postal survey of health authorities' views. *BMJ.* 1999;319:1235.

15. Conseil supérieur d'hygiène publique. Avis et Recommandations. Avis du Conseil Supérieur d'Hygiène Publique de France du 10 mars 2000 [Internet]. http://www.sante.gouv.fr/dossiers/cshpf/a_mt_100300_meningite_01.htm
16. Dellinger RP, Levy MM, Rhodes A, Annane D, Gerlach H, Opal SM, Sevransky JE, Sprung CL, Douglas IS, Jaeschke R, Osborn TM, Nunnally ME, Townsend SR, Reinhart K, Kleinpell RM, Angus DC, Deutschman CS, Machado FR, Rubenfeld GD, Webb SA, Beale RJ, Vincent J-L, Moreno R, Surviving Sepsis Campaign Guidelines Committee including the Pediatric Subgroup. Surviving sepsis campaign: international guidelines for management of severe sepsis and septic shock: 2012. *Crit Care Med.* 2013;41:580-637.
17. Agence Nationale d'Accréditation et d'Evaluation en Santé. Méthodes et Outils des démarches qualité pour les établissements de santé. [Internet]. 2000. <http://www.has-sante.fr/portail/upload/docs/application/pdf/methodes.pdf>
18. Leape LL, Lawthers AG, Brennan TA, Johnson WG. Preventing medical injury. *Qual Rev Bull.* 1993;19:144-9.
19. Brasme J-F, Morfouace M, Grill J, Martinot A, Amalberti R, Bons-Letouzey C, Chalumeau M. Delays in diagnosis of paediatric cancers: a systematic review and comparison with expert testimony in lawsuits. *Lancet Oncol.* 2012;13:e445-59.
20. Wallace E, Lowry J, Smith SM, Fahey T. The epidemiology of malpractice claims in primary care: a systematic review. *BMJ Open.* 2013;3.
21. Najaf-Zadeh A, Dubos F, Pruvost I, Bons-Letouzey C, Amalberti R, Martinot A. Epidemiology and aetiology of paediatric malpractice claims in France. *Arch Dis Child.* 2011;96:127-30.
22. Brasme J-F, Chalumeau M, Doz F, Lacour B, Valteau-Couanet D, Gaillard S, Delalande O, Aghakhani N, Sainte-Rose C, Puget S, Grill J. Interval between onset of symptoms and diagnosis of medulloblastoma in children: distribution and determinants in a population-based study. *Eur J Pediatr.* 2012;171:25-32.
23. Brasme J-F, Grill J, Doz F, Lacour B, Valteau-Couanet D, Gaillard S, Delalande O, Aghakhani N, Puget S, Chalumeau M. Long time to diagnosis of medulloblastoma in children is not associated with decreased survival or with worse neurological outcome. *PLoS ONE.* 2012;7:e33415.
24. Ahrensberg JM, Fenger-Grøn M, Vedsted P. Use of primary care during the year before childhood cancer diagnosis: a nationwide population-based matched comparative study. *PLoS ONE.* 2013;8:e59098.
25. Dommett RM, Redaniel MT, Stevens MCG, Hamilton W, Martin RM. Features of childhood cancer in primary care: a population-based nested case-control study. *Br J Cancer.* 2012;106:982-7.
26. Ciafaloni E, Fox DJ, Pandya S, Westfield CP, Puzhankara S, Romitti PA, Mathews KD, Miller TM, Matthews DJ, Miller LA, Cunniff C, Druschel CM, Moxley RT. Delayed diagnosis in duchenne muscular dystrophy: data from the Muscular Dystrophy

- Surveillance, Tracking, and Research Network (MD STARnet). *J Pediatr.* 2009;155:380-5.
27. Cappendijk VC, Hazebroek FW. The impact of diagnostic delay on the course of acute appendicitis. *Arch Dis Child.* 2000;83:64-6.
 28. Hartman AL, Lunney KM, Serena JE. Pediatric stroke: do clinical factors predict delays in presentation? *J Pediatr.* 2009;154:727-32.
 29. McIntyre PB, Macintyre CR, Gilmour R, Wang H. A population based study of the impact of corticosteroid therapy and delayed diagnosis on the outcome of childhood pneumococcal meningitis. *Arch Dis Child.* 2005;90:391-6.
 30. Pais PJ, Kump T, Greenbaum LA. Delay in diagnosis in poststreptococcal glomerulonephritis. *J Pediatr.* 2008;153:560-4.
 31. Rafay MF, Pontigon A-M, Chiang J, Adams M, Jarvis DA, Silver F, Macgregor D, Deveber GA. Delay to diagnosis in acute pediatric arterial ischemic stroke. *Stroke.* 2009;40:58-64.
 32. Säwendahl L, Davenport ML. Delayed diagnoses of Turner's syndrome: proposed guidelines for change. *J Pediatr.* 2000;137:455-9.
 33. Srinivasan J, Miller SP, Phan TG, Mackay MT. Delayed recognition of initial stroke in children: need for increased awareness. *Pediatrics.* 2009;124:e227-34.
 34. Urschel S, Kayikci L, Wintergerst U, Notheis G, Jansson A, Belohradsky BH. Common variable immunodeficiency disorders in children: delayed diagnosis despite typical clinical presentation. *J Pediatr.* 2009;154:888-94.
 35. Wilder MS, Palinkas LA, Kao AS, Bastian JF, Turner CL, Burns JC. Delayed diagnosis by physicians contributes to the development of coronary artery aneurysms in children with Kawasaki syndrome. *Pediatr Infect Dis J.* 2007;26:256-60.
 36. Chalumeau M, Holvoet L, Chéron G, Minodier P, Foix-L'Hélias L, Ovetckine P, Moulin F, Nouyrigat V, Bréart G, Gendrel D. Delay in diagnosis of imported Plasmodium falciparum malaria in children. *Eur J Clin Microbiol Infect Dis.* 2006;25:186-9.
 37. Gascoin-Lachambre G, Brauner R, Duche L, Chalumeau M. Pituitary stalk interruption syndrome: diagnostic delay and sensitivity of the auxological criteria of the growth hormone research society. *PLoS ONE.* 2011;6:e16367.
 38. Philibert M, Deneux-Tharoux C, Bouvier-Colle M-H. Can excess maternal mortality among women of foreign nationality be explained by suboptimal obstetric care? *Br J Obstet Gynaecol.* 2008;115:1411-8.
 39. Bradfield A, Wells GL. Not the same old hindsight bias: outcome information distorts a broad range of retrospective judgments. *Mem Cognit.* 2005;33:120-30.
 40. Kane RL, Wang J, Garrard J. Reporting in randomized clinical trials improved after adoption of the CONSORT statement. *J Clin Epidemiol.* 2007;60:241-9.

41. 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*. 2001;285:1992-5.
42. Plint AC, Moher D, Morrison A, Schulz K, Altman DG, Hill C, Gaboury I. Does the CONSORT checklist improve the quality of reports of randomised controlled trials? A systematic review. *Med J Aust*. 2006;185:263-7.
43. Smidt N, Rutjes AWS, van der Windt D a. WM, Ostelo RWJG, Bossuyt PM, Reitsma JB, Bouter LM, de Vet HCW. The quality of diagnostic accuracy studies since the STARD statement: has it improved? *Neurology*. 2006;67:792-7.
44. Panic N, Leoncini E, de Belvis G, Ricciardi W, Boccia S. Evaluation of the endorsement of the preferred reporting items for systematic reviews and meta-analysis (PRISMA) statement on the quality of published systematic review and meta-analyses. *PLoS ONE*. 2013;8:e83138.
45. Moher D, Schulz KF, Simera I, Altman DG. Guidance for developers of health research reporting guidelines. *PLoS Med*. 2010;7:e1000217.
46. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, Lijmer JG, Moher D, Rennie D, de Vet HCW, Standards for Reporting of Diagnostic Accuracy. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD Initiative. *Ann Intern Med*. 2003;138:40-4.
47. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *J Clin Epidemiol*. 2008;61:344-9.
48. Weller D, Vedsted P, Rubin G, Walter FM, Emery J, Scott S, Campbell C, Andersen RS, Hamilton W, Olesen F, Rose P, Nafees S, van Rijswijk E, Hiom S, Muth C, Beyer M, Neal RD. The Aarhus statement: improving design and reporting of studies on early cancer diagnosis. *Br J Cancer*. 2012;106:1262-7.
49. European Medicines Agency. Guideline on the role of pharmacokinetics in the development of medicinal products in the paediatric population [Internet]. http://cerped.fr/IMG/pdf/PK_14701304en.pdf
50. Kozer E. Medication errors in children. *Paediatr Drugs*. 2009;11:52-4.
51. McD Taylor D, Robinson J, MacLeod D, MacBean CE, Braitberg G. Therapeutic errors among children in the community setting: nature, causes and outcomes. *J Paediatr Child Health*. 2009;45:304-9.
52. Otero P, Leyton A, Mariani G, Ceriani Cernadas JM, Patient Safety Committee. Medication errors in pediatric inpatients: prevalence and results of a prevention program. *Pediatrics*. 2008;122:e737-43.
53. Fortescue EB, Kaushal R, Landrigan CP, McKenna KJ, Clapp MD, Federico F, Goldmann DA, Bates DW. Prioritizing strategies for preventing medication errors and adverse drug events in pediatric inpatients. *Pediatrics*. 2003;111:722-9.

54. Kai J. What worries parents when their preschool children are acutely ill, and why: a qualitative study. *BMJ*. 1996;313:983-6.
55. Lioffi C, Noble G, Franck LS. How parents make sense of their young children's expressions of everyday pain: A qualitative analysis. *Eur J Pain*. 2012; 16:1166-75
56. Klein-Geltink J, Shaw AK, Morrison HI, Barr RD, Greenberg ML. Use of paediatric versus adult oncology treatment centres by adolescents 15-19 years old: the Canadian Childhood Cancer Surveillance and Control Program. *Eur J Cancer*. 2005;41:404-10.
57. Claudius I, Baraff LJ. Pediatric emergencies associated with fever. *Emerg Med Clin North Am*. 2010;28:67-84, vii - viii.
58. Centre for Reviews and Dissemination. Systematic Reviews: CRD's guidance for undertaking systematic reviews in health care [Internet]. http://www.york.ac.uk/inst/crd/pdf/Systematic_Reviews.pdf
59. Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Med*. 2009;6:e1000097.
60. Sacks HS, Berrier J, Reitman D, Ancona-Berk VA, Chalmers TC. Meta-analyses of randomized controlled trials. *N Engl J Med*. 1987;316:450-5.
61. Laupacis A, Sekar N, Stiell IG. Clinical prediction rules. A review and suggested modifications of methodological standards. *JAMA*. 1997;277:488-94.
62. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, Moher D, Becker BJ, Sipe TA, Thacker SB. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *JAMA*. 2000;283:2008-12.
63. Wasson JH, Sox HC, Neff RK, Goldman L. Clinical prediction rules. Applications and methodological standards. *N Engl J Med*. 1985;313:793-9.
64. Mulrow CD. The medical review article: state of the science. *Ann Intern Med*. 1987;106:485-8.
65. Bouwmeester W, Zuithoff NPA, Mallett S, Geerlings MI, Vergouwe Y, Steyerberg EW, Altman DG, Moons KGM. Reporting and methods in clinical prediction research: a systematic review. *PLoS Med*. 2012;9:1-12.
66. Wilson J, Jungner G. Principles and practice of screening for disease. WHO. Public health papers n 34. [Internet]. 1968. www.who.int/php/WHO_PHP_34.pdf
67. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM, Statistics Subcommittee of the NCI-EORTC Working Group on Cancer Diagnostics. REporting recommendations for tumour MARKer prognostic studies (REMARK). *Br J Cancer*. 2005;93:387-91.

68. Whiting P, Rutjes AWS, Reitsma JB, Bossuyt PMM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol*. 2003;3:25.
69. Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, Leeflang MMG, Sterne JAC, Bossuyt PMM, QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155:529-36.
70. Journal Citation Report. [Internet]. <http://about.jcr.incites.thomsonreuters.com>
71. Aghamohammadi A, Bahrami A, Mamishi S, Mohammadi B, Abolhassani H, Parvaneh N, Rezaei N. Impact of delayed diagnosis in children with primary antibody deficiencies. *J Microbiol Immunol Infect*. 2011;44:229-34.
72. Anderson MS, Todd JK, Glodé MP. Delayed diagnosis of Kawasaki syndrome: an analysis of the problem. *Pediatrics*. 2005;115:e428-33.
73. Azarsiz E, Gulez N, Edeer Karaca N, Aksu G, Kutukculer N. Consanguinity rate and delay in diagnosis in Turkish patients with combined immunodeficiencies: a single-center study. *J Clin Immunol*. 2011;31:106-11.
74. Bai S, Ren R, Li B, Xu X, Zhao B, Gao F, Li L, Jonas JB. Delay in the diagnosis of retinoblastoma in China. *Acta Ophthalmol*. 2011;89:e72-4.
75. Bayne AP, Madden-Fuentes RJ, Jones EA, Cisek LJ, Gonzales ET, Reavis KM, Roth DR, Hsieh MH. Factors associated with delayed treatment of acute testicular torsion-do demographics or interhospital transfer matter? *J Urol*. 2010;184:1743-7.
76. Braun KPJ, Kappelle LJ, Kirkham FJ, Deveber G. Diagnostic pitfalls in paediatric ischaemic stroke. *Dev Med Child Neurol*. 2006;48:985-90.
77. Crawford JR, Santi MR, Vezina G, Myseros JS, Keating RF, LaFond DA, Rood BR, MacDonald TJ, Packer RJ. CNS germ cell tumor (CNSGCT) of childhood: presentation and delayed diagnosis. *Neurology*. 2007;68:1668-73.
78. Dang-Tan T, Trottier H, Mery LS, Morrison HI, Barr RD, Greenberg ML, Franco EL. Delays in diagnosis and treatment among children and adolescents with cancer in Canada. *Pediatr Blood Cancer*. 2008;51:468-74.
79. Dang-Tan T, Trottier H, Mery LS, Morrison HI, Barr RD, Greenberg ML, Franco EL. Determinants of delays in treatment initiation in children and adolescents diagnosed with leukemia or lymphoma in Canada. *Int J Cancer*. 2010;126:1936-43.
80. England RJ, Crabbe DCG. Delayed diagnosis of appendicitis in children treated with antibiotics. *Pediatr Surg Int*. 2006;22:541-5.
81. Fehr S, Downs J, Bebbington A, Leonard H. Atypical presentations and specific genotypes are associated with a delay in diagnosis in females with Rett syndrome. *Am J Med Genet A*. 2010;152A:2535-42.

82. Ferrari A, Miceli R, Casanova M, Meazza C, Favini F, Luksch R, Catania S, Fiore M, Morosi C, Mariani L. The symptom interval in children and adolescents with soft tissue sarcomas. *Cancer*. 2010;116:177-83.
83. Foster HE, Eltringham MS, Kay LJ, Friswell M, Abinun M, Myers A. Delay in access to appropriate care for children presenting with musculoskeletal symptoms and ultimately diagnosed with juvenile idiopathic arthritis. *Arthritis Rheum*. 2007;57:921-7.
84. Haimi M, Perez-Nahum M, Stein N, Ben Arush MW. The role of the doctor and the medical system in the diagnostic delay in pediatric malignancies. *Cancer Epidemiol*. 2011;35:83-9.
85. Jackson A, Foley L, Daly L, Fitzpatrick P, Harrington M, Zhou S, Kelleher C. Delayed cystic fibrosis presentation in children in the absence of newborn screening. *Ir Med J*. 2010;103:113-6.
86. Juan C-C, Hwang B, Lee P-C, Lin Y-J, Chien J-C, Lee H-Y, Meng CCL. The clinical manifestations and risk factors of a delayed diagnosis of Kawasaki disease. *J Chin Med Assoc*. 2007;70:374-9.
87. Kim MS, Lee S-Y, Cho WH, Song WS, Koh J-S, Lee JA, Yoo JY, Shin DS, Jeon D-G. Prognostic effects of doctor-associated diagnostic delays in osteosarcoma. *Arch Orthop Trauma Surg*. 2009;129:1421-5.
88. Kukal K, Dobrovoljac M, Boltshauser E, Ammann RA, Grotzer MA. Does diagnostic delay result in decreased survival in paediatric brain tumours? *Eur J Pediatr*. 2009;168:303-10.
89. Letton RW, Worrell V, Tuggle DW, American Pediatric Surgical Association Committee on Trauma Blunt Intestinal Injury Study Group. Delay in diagnosis and treatment of blunt intestinal perforation does not adversely affect prognosis in the pediatric trauma patient. *J Trauma*. 2010;68:790-5.
90. Majak P, Bąk-Walczak E, Stelmach I, Jerzyńska J, Krakowiak J, Stelmach W. An increasing trend of the delay in asthma diagnosis after the discontinuation of a population-based intervention. *J Asthma*. 2011;48:414-8.
91. Mann JR, Crawford S, Wilson L, McDermott S. Does race influence age of diagnosis for children with developmental delay? *Disabil Health J*. 2008;1:157-62.
92. McCormick J, Sims EJ, Mehta A. Delayed diagnosis of females with respiratory presentation of cystic fibrosis did not segregate with poorer clinical outcome. *J Clin Epidemiol*. 2006;59:315-22.
93. McGlennan C, Ganesan V. Delays in investigation and management of acute arterial ischaemic stroke in children. *Dev Med Child Neurol*. 2008;50:537-40.
94. Minich LL, Sleeper LA, Atz AM, McCrindle BW, Lu M, Colan SD, Printz BF, Klein GL, Sundel RP, Takahashi M, Li JS, Vetter VL, Newburger JW, Pediatric Heart Network Investigators. Delayed diagnosis of Kawasaki disease: what are the risk factors? *Pediatrics*. 2007;120:e1434-40.

95. Mistry PK, Sadan S, Yang R, Yee J, Yang M. Consequences of diagnostic delays in type 1 Gaucher disease: the need for greater awareness among hematologists-oncologists and an opportunity for early diagnosis and intervention. *Am J Hematol*. 2007;82:697-701.
96. Pan KL, Chan WH, Chia YY. Initial symptoms and delayed diagnosis of osteosarcoma around the knee joint. *J Orthop Surg (Hong Kong)*. 2010;18:55-7.
97. Pawłowicz M, Birkholz D, Niedźwiecki M, Balcerska A. Difficulties or mistakes in diagnosing type 1 diabetes in children?--demographic factors influencing delayed diagnosis. *Pediatr Diabetes*. 2009;10:542-9.
98. Perno JF, Schunk JE, Hansen KW, Furnival RA. Significant reduction in delayed diagnosis of injury with implementation of a pediatric trauma service. *Pediatr Emerg Care*. 2005;21:367-71.
99. Purper-Ouakil D, Cortese S, Wohl M, Asch M, Acquaviva E, Falissard B, Michel G, Gorwood P, Mouren MC. Predictors of diagnostic delay in a clinical sample of French children with attention-deficit/hyperactivity disorder. *Eur Child Adolesc Psychiatry*. 2007;16:505-9.
100. Rahme D, Comley A, Foster B, Cundy P. Consequences of diagnostic delays in slipped capital femoral epiphysis. *J Pediatr Orthop B*. 2006;15:93-7.
101. Ravichandiran N, Schuh S, Bejuk M, Al-Harthy N, Shouldice M, Au H, Boutis K. Delayed identification of pediatric abuse-related fractures. *Pediatrics*. 2010;125:60-6.
102. Shilon Y, Pollak Y, Benarroch F, Gross-Tsur V. Factors influencing diagnosis delay in children with Tourette syndrome. *Eur J Paediatr Neurol*. 2008;12:398-400.
103. Shlizerman L, Mazzawi S, Rakover Y, Ashkenazi D. Foreign body aspiration in children: the effects of delayed diagnosis. *Am J Otolaryngol*. 2010;31:320-4.
104. Stefan DC, Siemonsma F. Delay and causes of delay in the diagnosis of childhood cancer in Africa. *Pediatr Blood Cancer*. 2011;56:80-5.
105. Steinraths M, Vallance HD, Davidson AGF. Delays in diagnosing cystic fibrosis: can we find ways to diagnose it earlier? *Can Fam Physician*. 2008;54:877-83.
106. Timmer A, Behrens R, Buderus S, Findeisen A, Hauer A, Keller K-M, Kliemann G, Lang T, Lohr W, Rzehak P, Koletzko S, CEDATA-GPGE Study Group. Childhood onset inflammatory bowel disease: predictors of delayed diagnosis from the CEDATA German-language pediatric inflammatory bowel disease registry. *J Pediatr*. 2011;158:467-73.e2.
107. Viani RM, Lopez G, Chacón-Cruz E, Hubbard P, Spector SA. Poor outcome is associated with delayed tuberculosis diagnosis in HIV-infected children in Baja California, Mexico. *Int J Tuberc Lung Dis*. 2008;12:411-6.
108. Weibel L, Laguda B, Atherton D, Harper JJ. Misdiagnosis and delay in referral of children with localized scleroderma. *Br J Dermatol*. 2011;165:1308-13.

109. Weigall P, Vladusic S, Torode I. Slipped upper femoral epiphysis in children--delays to diagnosis. *Aust Fam Physician*. 2010;39:151-3.
110. Weiss SL, Fitzgerald JC, Pappachan J, Wheeler D, Jaramillo-Bustamante JC, Salloo A, Singhi SC, Erickson S, Roy JA, Bush JL, Nadkarni VM, Thomas NJ. Global epidemiology of pediatric severe sepsis: the sepsis prevalence, outcomes, and therapies study. *Am J Respir Crit Care Med*. 2015;191:1147-57.
111. Williams BG, Hlaing T, Aaland MO. Ten-year retrospective study of delayed diagnosis of injury in pediatric trauma patients at a level II trauma center. *Pediatr Emerg Care*. 2009;25:489-93.
112. Boutron I, Dutton S, Ravaud P, Altman DG. Reporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes. *JAMA*. 2010;303:2058-64.
113. 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*. 2013;66:1093-104.
114. 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.
115. Bouvier-Colle MH, Ould El Joud D, Varnoux N, Goffinet F, Alexander S, Bayoumeu F, Beaumont E, Fernandez H, Lansac J, Lévy G, Palot M. Evaluation of the quality of care for severe obstetrical haemorrhage in three French regions. *Br J Obstet Gynaecol*. 2001;108:898-903.
116. Egger M, Zellweger-Zähner T, Schneider M, Junker C, Lengeler C, Antes G. Language bias in randomised controlled trials published in English and German. *Lancet*. 1997;350:326-9.
117. Moher D, Pham B, Klassen TP, Schulz KF, Berlin JA, Jadad AR, Liberati A. What contributions do languages other than English make on the results of meta-analyses? *J Clin Epidemiol*. 2000;53:964-72.
118. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. *Ann Intern Med*. 2010;152:726-32.
119. 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. *Lancet*. 1999;354:1896-900.
120. Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovich C, Song F, Petticrew M, Altman DG, International Stroke Trial Collaborative Group, European Carotid Surgery Trial Collaborative Group. Evaluating non-randomised intervention studies. *Health Technol Assess*. 2003;7:iii - x, 1-173.

121. Viswanathan M, Berkman ND. Development of the RTI item bank on risk of bias and precision of observational studies. *J Clin Epidemiol*. 2012;65:163-78.
122. The Cochrane Collaboration. Assessing risk of bias in included studies. In: Julian PT Higgins, Douglas G Altman, Sterne JA, editors. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0*. [Internet]. <http://www.cochrane-handbook.org/>
123. The Cochrane Collaboration. Applicability and transferability. In: Julian PT Higgins, Douglas G Altman, Sterne JA, editors. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0*. [Internet]. <http://www.cochrane-handbook.org/>
124. University of York. Centre for Reviews and Dissemination. PROSPERO. Inclusion criteria [Internet]. <http://www.crd.york.ac.uk/PROSPERO>
125. Viswanathan M, Berkman ND, Dryden DM, Hartling L. Assessing Risk of Bias and Confounding in Observational Studies of Interventions or Exposures: Further Development of the RTI Item Bank [Internet]. Rockville (MD): Agency for Healthcare Research and Quality (US); 2013. <http://www.ncbi.nlm.nih.gov/books/NBK154461/>
126. Abel GA, Friese CR, Magazu LS, Richardson LC, Fernandez ME, De Zengotita JJ, Earle CC. Delays in referral and diagnosis for chronic hematologic malignancies: A literature review. *Leuk Lymphoma*. 2008;49:1352-9.
127. Alhurishi S, Lim JNW, Potrata B, West R. Factors influencing late presentation for breast cancer in the middle East: a systematic review. *Asian Pac J Cancer Prev*. 2011;12:1597-600.
128. Bradford A, Kunik ME, Schulz P, Williams SP, Singh H. Missed and Delayed Diagnosis of Dementia in Primary Care: Prevalence and Contributing Factors. *Alzheimer Dis Assoc Disord*. 2009;23:306-14.
129. Chatwin J, Sanders C. The influence of social factors on help-seeking for people with lung cancer: Help-seeking in lung cancer. *Eur J Cancer Care*. 2013;22:709-13.
130. Chen T-C, Lu P-L, Lin C-Y, Lin W-R, Chen Y-H. Fluoroquinolones are associated with delayed treatment and resistance in tuberculosis: a systematic review and meta-analysis. *Int J Infect Dis*. 2011;15:e211-6.
131. Chen NE, Gallant JE, Page KR. A Systematic Review of HIV/AIDS Survival and Delayed Diagnosis Among Hispanics in the United States. *J Immigr Minor Health*. 2012;14:65-81.
132. Dang-Tan T, Franco EL. Diagnosis delays in childhood cancer: A review. *Cancer*. 2007;110:703-13.
133. Dubayova T, van Dijk JP, Nagyova I, Rosenberger J, Havlikova E, Gdovinova Z, Middel B, Groothoff JW. The impact of the intensity of fear on patient's delay regarding health care seeking behavior: a systematic review. *Int J Public Health*. 2010;55:459-68.

134. Evenson KR, Foraker RE, Morris DL, Rosamond WD. A comprehensive review of prehospital and in-hospital delay times in acute stroke care. *Int J Stroke*. 2009;4:187-99.
135. Evenson KR, Rosamond WD, Morris DL. Prehospital and in-hospital delays in acute stroke care. *Neuroepidemiology*. 2001;20:65-76.
136. Facione NC. Delay versus help seeking for breast cancer symptoms: a critical review of the literature on patient and provider delay. *Soc Sci Med*. 1993;36:1521-34.
137. Finnie RKC, Khoza LB, van den Borne B, Mabunda T, Abotchie P, Mullen PD. Factors associated with patient and health care system delay in diagnosis and treatment for TB in sub-Saharan African countries with high burdens of TB and HIV: Patient and system delay in diagnosis and treatment for TB in sub-Saharan Africa. *Trop Med Int Health*. 2011;16:394-411.
138. Forrest LF, Sowden S, Rubin G, White M, Adams J. Socio-economic inequalities in patient, primary care, referral, diagnostic, and treatment intervals on the lung cancer care pathway: protocol for a systematic review and meta-analysis. *Syst Rev*. 2014;3:30.
139. A Fowler B. Neighborhood-level Influences on Delays in Diagnostic Follow-up from Mammography Screening in African-American Women: A Systematic Review. *Journal of Women's Health Care [Internet]*. 2014;03. <http://www.omicsgroup.org/journals/neighborhoodlevel-influences-on-delays-in-diagnostic-followup-from-mammography-screening-in-africanamerican-women-a-systematic-review-2167-0420.1000151.php?aid=25078>
140. Gómez I, Seoane J, Varela-Centelles P, Diz P, Takkouche B. Is diagnostic delay related to advanced-stage oral cancer? A meta-analysis. *Eur J Oral Sci*. 2009;117:541-6.
141. Goy J, Hall SF, Feldman-Stewart D, Groome PA. Diagnostic delay and disease stage in head and neck cancer: A systematic review. *Laryngoscope*. 2009;119:889-98.
142. Jayde V, White K, Blomfield P. Symptoms and diagnostic delay in ovarian cancer: a summary of the literature. *Contemp Nurse*. 2009;34:55-65.
143. Jensen AR, Mainz J, Overgaard J. Impact of delay on diagnosis and treatment of primary lung cancer. *Acta Oncol*. 2002;41:147-52.
144. Jones CE, Maben J, Jack RH, Davies EA, Forbes LJ, Lucas G, Ream E. A systematic review of barriers to early presentation and diagnosis with breast cancer among black women. *BMJ Open*. 2014;4:e004076-e004076.
145. Lefler LL, Bondy KN. Women's delay in seeking treatment with myocardial infarction: a meta-synthesis. *J Cardiovasc Nurs*. 2004;19:251-68.
146. Lefler L. The advanced practice nurse's role regarding women's delay in seeking treatment with myocardial infarction. *J Am Acad Nurse Pract*. 2002;14:449-56.
147. Lethaby CD, Picton S, Kinsey SE, Phillips R, van Laar M, Feltbower RG. A systematic review of time to diagnosis in children and young adults with cancer. *Arch Dis Child*. 2013;98:349-55.

148. Li Y, Ehiri J, Tang S, Li D, Bian Y, Lin H, Marshall C, Cao J. Factors associated with patient, and diagnostic delays in Chinese TB patients: a systematic review and meta-analysis. *BMC Medicine*. 2013;11:156.
149. Macdonald S, Macleod U, Campbell NC, Weller D, Mitchell E. Systematic review of factors influencing patient and practitioner delay in diagnosis of upper gastrointestinal cancer. *Br J Cancer*. 2006;94:1272-80.
150. Martins T, Hamilton W, Ukoumunne OC. Ethnic inequalities in time to diagnosis of cancer: a systematic review. *BMC Fam Pract*. 2013;14:197.
151. Mitchell E, Macdonald S, Campbell NC, Weller D, Macleod U. Influences on pre-hospital delay in the diagnosis of colorectal cancer: a systematic review. *Br J Cancer* 2008;98:60-70.
152. Ndeikoundam Ngangro N, Chauvin P, Halley des Fontaines V. Les déterminants du délai de diagnostic de la tuberculose dans les pays aux ressources limitées. *Revue Epidemiol Sante Publique*. 2012;60:47-57.
153. Olsson JK, Schultz EM, Gould MK. Timeliness of care in patients with lung cancer: a systematic review. *Thorax*. 2009;64:749-56.
154. Ramirez AJ, Westcombe AM, Burgess CC, Sutton S, Littlejohns P, Richards MA. Factors predicting delayed presentation of symptomatic breast cancer: a systematic review. *Lancet*. 1999;353:1127-31.
155. Ramos M, Esteva M, Cabeza E, Llobera J, Ruiz A. Lack of association between diagnostic and therapeutic delay and stage of colorectal cancer. *Eur J Cancer*. 2008;44:510-21.
156. Ramos M, Esteva M, Cabeza E, Campillo C, Llobera J, Aguiló A. Relationship of diagnostic and therapeutic delay with survival in colorectal cancer: A review. *Eur J Cancer*. 2007;43:2467-78.
157. Richards MA, Westcombe AM, Love SB, Littlejohns P, Ramirez AJ. Influence of delay on survival in patients with breast cancer: a systematic review. *Lancet*. 1999;353:1119-26.
158. Scott SE, Grunfeld EA, McGurk M. Patient's delay in oral cancer: A systematic review. *Community Dent Oral Epidemiol*. 2006;34:337-43.
159. Seoane J, Takkouche B, Varela-Centelles P, Tomás I, Seoane-Romero JM. Impact of delay in diagnosis on survival to head and neck carcinomas: a systematic review with meta-analysis: Diagnostic delay and head and neck cancer. *Clinical Otolaryngology*. 2012;37:99-106.
160. Sreeramareddy CT, Qin ZZ, Satyanarayana S, Subbaraman R, Pai M. Delays in diagnosis and treatment of pulmonary tuberculosis in India: a systematic review. *Int J Tuberc Lung Dis*. 2014;18:255-66.

161. Sreeramareddy CT, Kishore PV, Menten J, Van den Ende J. Time delays in diagnosis of pulmonary tuberculosis: a systematic review of literature. *BMC Infect Dis.* 2009;9:91.
162. Storla DG, Yimer S, Bjune GA. A systematic review of delay in the diagnosis and treatment of tuberculosis. *BMC Public Health.* 2008;8:15.
163. Stoykova M. Delayed diagnosis of cancer with emphasis on oral cavity cancers. *Folia Med (Plovdiv).* 1999;41:132-5.
164. Teuschl Y, Brainin M. Stroke education: discrepancies among factors influencing prehospital delay and stroke knowledge: Reviews. *Int J Stroke.* 2010;5:187-208.
165. Thompson MR, Heath I, Swarbrick ET, Faulds Wood L, Ellis BG. Earlier diagnosis and treatment of symptomatic bowel cancer: can it be achieved and how much will it improve survival?: Earlier diagnosis of symptomatic bowel cancer. *Colorectal Dis.* 2011;13:6-16.
166. Unger-Saldaña K, Infante-Castañeda C. Delay of medical care for symptomatic breast cancer: a literature review. *Salud Publica Mex.* 2009;51 Suppl 2:s270-85.
167. Villeneuve E, Nam JL, Bell MJ, Deighton CM, Felson DT, Hazes JM, McInnes IB, Silman AJ, Solomon DH, Thompson AE, White PHP, Bykerk VP, Emery P. A systematic literature review of strategies promoting early referral and reducing delays in the diagnosis and management of inflammatory arthritis. *Ann Rheum Dis.* 2013;72:13-22.
168. Vinas F, Ben Hassen I, Jabot L, Monnet I, Chouaid C. Delays for diagnosis and treatment of lung cancers: a systematic review. *Clin Respir J.* 2014;
169. Yang W-T, Gounder CR, Akande T, De Neve J-W, McIntire KN, Chandrasekhar A, de Lima Pereira A, Gummadi N, Samanta S, Gupta A. Barriers and delays in tuberculosis diagnosis and treatment services: does gender matter? *Tuberc Res Treat.* 2014;2014:461935.
170. Critical Appraisal Skill Program. Critical appraisal Skill Program Checklists. [Internet]. www.casp-uk.net/#!/checklists/cb36
171. Wells G, Shea B, O'Connell D, Peterson J, Welch V, Losos M, Tugwell P. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. [Internet]. http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp
172. Jüni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA.* 1999;282:1054-60.
173. Juni P. Direction and impact of language bias in meta-analyses of controlled trials: empirical study. *Int J Epidemiol.* 2002;31:115-23.
174. Viswanathan M, Berkman ND, Dryden DM, Hartling L. Assessing Risk of Bias and Confounding in Observational Studies of Interventions or Exposures: Further

- Development of the RTI Item Bank [Internet]. Rockville (MD): Agency for Healthcare Research and Quality (US); 2013. <http://www.ncbi.nlm.nih.gov/books/NBK154461/>
175. Critical Appraisal Skill Program. CASP case control checklist. [Internet]. http://media.wix.com/ugd/dded87_63fb65dd4e0548e2bfd0a982295f839e.pdf
176. Glasziou P, Altman DG, Bossuyt P, Boutron I, Clarke M, Julious S, Michie S, Moher D, Wager E. Reducing waste from incomplete or unusable reports of biomedical research. *Lancet*. 2014;383:267-76.
177. Development of a reporting guideline for reporting studies on time to diagnosis [Internet]. <http://www.equator-network.org/wp-content/uploads/2009/02/Reporting-studies-on-time-to-diagnosis-summary.pdf>.
178. Bossuyt PM, Reitsma JB, Standards for Reporting of Diagnostic Accuracy. The STARD initiative. *Lancet*. 2003;361:71.
179. Angus DC, van der Poll T. Severe sepsis and septic shock. *N Engl J Med*. 2013;369:840-51.
180. Harnden A, Mayon-White R, Mant D, Kelly D, Pearson G. Child deaths: confidential enquiry into the role and quality of UK primary care. *Br J Gen Pract*. 2009;59:819-24.
181. Bamber AR, Mifsud W, Wolfe I, Cass H, Pryce J, Malone M, Sebire NJ. Potentially preventable infant and child deaths identified at autopsy; findings and implications. *Forensic Science, Medicine, and Pathology [Internet]*. 2015; <http://link.springer.com/10.1007/s12024-015-9681-9>
182. Nadel S, Britto J, Booy R, Maconochie I, Habibi P, Levin M. Avoidable deficiencies in the delivery of health care to children with meningococcal disease. *J Accid Emerg Med*. 1998;15:298-303.
183. Ninis N, Phillips C, Bailey L, Pollock JI, Nadel S, Britto J, Maconochie I, Winrow A, Coen PG, Booy R, Levin M. The role of healthcare delivery in the outcome of meningococcal disease in children: case-control study of fatal and non-fatal cases. *BMJ*. 2005;330:1475.
184. Han YY, Carcillo JA, Dragotta MA, Bills DM, Watson RS, Westerman ME, Orr RA. Early reversal of pediatric-neonatal septic shock by community physicians is associated with improved outcome. *Pediatrics*. 2003;112:793-9.
185. Inwald DP, Tasker RC, Peters MJ, Nadel S, Paediatric Intensive Care Society Study Group (PICS-SG). Emergency management of children with severe sepsis in the United Kingdom: the results of the Paediatric Intensive Care Society sepsis audit. *Arch Dis Child*. 2009;94:348-53.
186. Goldstein B, Giroir B, Randolph A, International Consensus Conference on Pediatric Sepsis. International pediatric sepsis consensus conference: definitions for sepsis and organ dysfunction in pediatrics. *Pediatr Crit Care Med*. 2005;6:2-8.
187. Control of meningococcal disease: Guidance for consultants in communicable disease control. *Commun Dis Rep CDR Rev*. 1995;5:R189-94.

188. Société de réanimation de langue française. Utilisation des catécholamines au cours du choc septique (adultes - enfants). XVème conférence de consensus de la Société de Réanimation de Langue Française. 1996 [Internet]. <http://www.sfar.org/article/33/utilisation-des-catecholamines-au-cours-du-choc-septique-adultes-enfants-cc-1996>
189. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159-74.
190. Cvetkovic M, Lutman D, Ramnarayan P, Pathan N, Inwald DP, Peters MJ. Timing of death in children referred for intensive care with severe sepsis: implications for interventional studies. *Pediatr Crit Care Med*. 2015;16:410-7.
191. Cohen J. A Coefficient of agreement for nominal Scales. *Educ Psychol Meas*. 1960;20:37-46.
192. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bull*. 1968;70:213-20.
193. McDonald CJ, Weiner M, Hui SL. Deaths due to medical errors are exaggerated in Institute of Medicine report. *JAMA*. 2000;284:93-5.
194. Aurel M, Dubos F, Motte B, Pruvost I, Leclerc F, Martinot A. Recognising haemorrhagic rash in children with fever: a survey of parents' knowledge. *Arch Dis Child*. 2011;96:697-8.
195. Hill AB. The environment and disease: association or causation? *Proc R Soc Med*. 1965;58:295-300.
196. Hernandez-Diaz S, Schisterman EF, Hernan MA. The birth weight « paradox » uncovered? *Am J Epidemiol*. 2006;164:1115-20.
197. Shrier I, Platt RW. Reducing bias through directed acyclic graphs. *BMC Med Res Methodol*. 2008;8:70.
198. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology*. 1999;10:37-48.
199. Pearl J. Causality: models, reasoning, and inference. 9. print. Cambridge: Univ. Press; 2009. 384 p.
200. Driessen M, Bouvier-Colle M-H, Dupont C, Khoshnood B, Rudigoz R-C, Deneux-Tharoux C. Postpartum hemorrhage resulting from uterine atony after vaginal Delivery: factors associated with severity. *Obstet Gynecol*. 2011;117:21-31.
201. Snijders TAB, Bosker RJ. Multilevel analysis: an introduction to basic and advanced multilevel modeling. 2. ed. Los Angeles, Calif.: Sage; 2012. 354 p.
202. Textor J, Hardt J, Knüppel S. DAGitty: a graphical tool for analyzing causal diagrams. *Epidemiology*. 2011;22:745.

203. Altman DG, Royston P. The cost of dichotomising continuous variables. *BMJ*. 2006;332:1080.
204. Thompson M, Mayon-White R, Harnden A, Perera R, McLeod D, Mant D. Using vital signs to assess children with acute infections: a survey of current practice. *Br J Gen Pract*. 2008;58:236-41.
205. Casanova J-L, Abel L. Human genetics of infectious diseases: a unified theory. *EMBO J*. 2007;26:915-22.
206. Peterson JW. Bacterial Pathogenesis. In: Baron S, éditeur. *Medical Microbiology [Internet]*. 4th éd. Galveston (TX): University of Texas Medical Branch at Galveston; 1996. <http://www.ncbi.nlm.nih.gov/books/NBK8526/>
207. Dumas O, Siroux V, Le Moual N, Varraso R. Approches d'analyse causale en épidémiologie. *Rev Epidemiol Sante Publique*. 2014;62:53-63.
208. Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology*. 2000;11:550-60.
209. Prieto-Merino D, Smeeth L, Staa TP v., Roberts I. Dangers of non-specific composite outcome measures in clinical trials. *BMJ*. 2013;347:f6782-f6782.
210. Varon E, Janoir C, Gutmann L. Centre national de référence du pneumocoque. Rapport d'activités 2013. [Internet]. 2012. <http://www.cnr-pneumo.fr/docs/rapports/CNRP2013.pdf>
211. Stahl J-P, Cohen R, Denis F, Gaudelus J, Lery T, Lepetit H, Martinot A. Vaccination against meningococcus C. vaccinal coverage in the French target population. *Med Mal Infect*. 2013;43:75-80.
212. Rhodes A, Phillips G, Beale R, Cecconi M, Chiche JD, De Backer D, Divatia J, Du B, Evans L, Ferrer R, Girardis M, Koulenti D, Machado F, Simpson SQ, Tan CC, Wittebole X, Levy M. The Surviving Sepsis Campaign bundles and outcome: results from the International Multicentre Prevalence Study on Sepsis (the IMPReSS study). *Intensive Care Med*. 2015;41:1620-8.
213. Giguère A, Légaré F, Grimshaw J, Turcotte S, Fiander M, Grudniewicz A, Makosso-Kallyth S, Wolf FM, Farmer AP, Gagnon M-P. Printed educational materials: effects on professional practice and healthcare outcomes. *Cochrane Database Syst Rev*. 2012;10:CD004398.
214. Dharmar M, Marcin JP, Romano PS, Andrada ER, Overly F, Valente JH, Harvey DJ, Cole SL, Kuppermann N. Quality of care of children in the emergency department: association with hospital setting and physician training. *J Pediatr*. 2008;153:783-9.
215. Gilleland J, McGugan J, Brooks S, Dobbins M, Ploeg J. Caring for critically ill children in the community: a needs assessment. *BMJ Qual Saf*. 2014;23:490-8.
216. Cheng A, Goldman RD, Aish MA, Kisson N. A simulation-based acute care curriculum for pediatric emergency medicine fellowship training programs. *Pediatr Emerg Care*. 2010;26:475-80.

-
217. Spotting the sick child [Internet]. www.spottingthesickchild.com
 218. Mant D, Van den Bruel A. Should we promote the tumbler test? *Arch Dis Child*. 2011;96:613-4.
 219. Meningitis Research Foundation. Signs of septicaemia and meningitis. [Internet]. <http://www.meningitis.org/symptoms>
 220. Bertille N, Fournier-Charrière E, Pons G, Chalumeau M. Managing fever in children: a national survey of parents' knowledge and practices in France. *PLoS ONE*. 2013;8:e83469.
 221. Lebel E, Rudensky B, Karasik M, Itzhaki M, Schlesinger Y. *Kingella kingae* infections in children. *J Pediatr Orthop B*. 2006;15:289-92.
 222. Baraff LJ. Management of infants and young children with fever without source. *Pediatr Ann*. 2008;37:673-9.
 223. Thompson M, Van den Bruel A, Verbakel J, Lakhanpaul M, Haj-Hassan T, Stevens R, Moll H, Buntinx F, Berger M, Aertgeerts B, Oostenbrink R, Mant D. Systematic review and validation of prediction rules for identifying children with serious infections in emergency departments and urgent-access primary care. *Health Technol Assess*. 2012;16:1-100.

ANNEXES

Annexe 1 : Article issu de la Partie I - Chapitre 1

Quality of reporting of studies evaluating time to diagnosis: a systematic review in paediatrics

Elise Launay,^{1,2} Michele Morfouace,^{1,3} Catherine Deneux-Tharaux,¹
 Christèle Gras le-Guen,^{3,4} Philippe Ravaud,⁵ Martin Chalumeau^{1,3}

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/archdischild-2013-304778>).

¹Unité Inserm 953, Maternité Port-Royal, Paris, France

²Clinique Médicale Pédiatrique, Hôpital Mère-Enfant, CHU NANTES, Nantes, France

³Service de pédiatrie générale, hôpital Necker Enfants Malades; AP-HP; Université Paris Descartes, Paris, France

⁴Urgences pédiatriques, Hôpital Mère-Enfant, CHU NANTES, Nantes, France

⁵Unité Inserm 738, Université Paris Descartes, Paris, France

Correspondence to

Dr Elise Launay, Unité Inserm 953, Maternité Port-Royal, 53 avenue de l'Observatoire, Paris 75014, France; launay.elise@gmail.com

Received 3 July 2013

Revised 31 October 2013

Accepted 1 November 2013

ABSTRACT

Objective An ever-increasing number of studies analyses the distribution, determinants and consequences of time to diagnosis and delays. Weaknesses in their reporting can impede the assessment of the risks of bias and variation and thus create a risk of invalid conclusions and counterproductive clinical and public health efforts. This study sought to assess systematically the quality of reporting of articles about time to diagnosis in paediatrics.

Design Two authors identified and analysed the quality of reporting of 50 consecutive articles assessing these intervals published from 2005 through October 2011, according to a checklist we developed of 35 items potentially associated with risks of bias and variation.

Main outcome measure Frequency of articles reporting each item.

Results Symptoms that should trigger a diagnostic procedure were reported in 28% of the articles; only two articles reported whether all patients with these symptoms underwent that procedure. Only 44% of the articles defined the beginning of the illness, 46% the date of diagnosis and 60% the distribution of time to diagnosis. Two studies met the criteria for all 11 items considered essential for assessing the risks of bias and variation in this type of study.

Interpretation This study identified many weaknesses in the quality of reporting of studies of time to diagnosis in paediatrics, especially for items potentially related to risks of bias and variation. This finding underlines the need for the development of new (or the refinement of existing) guidelines for reporting this type of study.

INTRODUCTION

According to a study published in 2013, more than 400 000 deaths per year in the USA might be attributable to suboptimal care.¹ The physiological immaturity and the continuing maturation of the paediatric population give them very specific characteristics² that have already been shown to put them at special risk for suboptimal drug treatment.^{3 4} These characteristics may also expose children to a higher risk of suboptimal diagnostic care, compared with adult populations. Preverbal children cannot identify, report or describe their symptoms themselves and are thus dependent on the knowledge and practices of third persons (most often, their parents and physicians).⁵ Teenagers, on the contrary, tend not to report their symptoms spontaneously.⁶ Moreover, preschool children can often appear very ill with benign conditions and frequently have chronic symptoms (eg, gastro-oesophageal reflux) or recurrent polymorphic acute episodes (viral respiratory or intestinal infections).

What is already known on this topic

- The number of published studies on time to diagnosis in paediatrics continues to grow, and most of them conclude that long intervals are frequent.
- Assessing the quality of reporting of studies is an indirect way of evaluating their risk of bias and variation.
- Existing guidelines (STARD, STROBE, Aarhus statement) do not provide a complete framework for reporting studies of time to diagnosis.

What this study adds

- The quality of reporting in publications studying time to diagnosis was poor, notably for specific items potentially associated with a high risk of bias.
- There is a need for specific reporting guidelines for studies of time to diagnosis.

These characteristics mean that warning symptoms (such as fever, pain, vomiting, anorexia or dyspnoea) of severe diseases are less specific and the risk of diagnostic error and delay accordingly higher.^{7 8} The rarity of unfavourable outcomes in this age group with the lowest mortality rate (compared with neonates and adults) might also contribute to reducing clinicians' awareness. Finally, the level of evidence for diagnostic procedures is often lower in paediatrics, because of the difficulties inherent in clinical research in children (including parents' and physicians' reluctance to include them in prospective studies, difficulties in performing reference tests requiring invasive diagnostic investigations,⁹ and the limited size of the paediatric market, which discourages industrial funding of studies).

Diagnostic delays and diagnostic errors are thus presumably frequent in paediatric practice, and the literature about them continues to grow (see online supplementary appendix 1). Many of these studies conclude that adverse outcomes are associated with diagnostic delay and that it is thus essential to take corrective clinical or public health measures (eg, parental education). However, there are numerous ways in which a study of the distribution,

To cite: Launay E, Morfouace M, Deneux-Tharaux C, et al. *Arch Dis Child* Published Online First: [please include Day Month Year] doi:10.1136/archdischild-2013-304778

determinants and consequences of diagnostic delay might be biased or subject to variation and could therefore lead to unnecessary or counterproductive public health efforts. As the community of biomedical researchers became aware of the need to improve the quality of reporting for other specific types of studies, methodological guidelines were developed for their reporting,¹⁰ and these guidelines have helped to improve their methodological quality.^{11–13} According to Moher *et al*,¹⁴ one initial step in developing reporting guidelines is to seek evidence on the quality of reporting of published studies and to identify potential sources of bias.^{15–17}

Our aim therefore was to systematically evaluate the quality of reporting of publications about time to diagnosis in the paediatric population. To review reporting quality, we developed criteria specific to this kind of study based on their potential association with the risks of bias and variation.

MATERIAL AND METHODS

By analysis of the quality of reporting, we mean specifically analysis of the reporting of factors potentially related to the risk of bias and variation (eg, variation of the population, setting, protocol or definition of the target disorder) and of their consequences on future corrective policies. We evaluated the quality of the reporting in 50 studies of time to diagnosis in children. This arbitrarily chosen number was fairly close to the median number (33, range 29–86) of studies used to assess reporting quality in other fields of health research.^{16 18–22}

Identification of studies

Studies were systematically identified by an electronic search on the PubMed website. The search formula was: (((pitfalls[Title] OR missed[Title]) OR delay*[Title]) OR interval*[Title]) AND diagnos* [Title/Abstract] NOT (((((gene[Title/Abstract] OR chromosome[Title/Abstract]) OR antenatal[Title/Abstract]) OR review[Publication Type]) OR Case Reports[Publication Type]) OR editorial[publication type]) OR comment[publication type]) OR letter[publication type]). We used the following limiting criteria: child (age) and English (language). Starting in October 2011 and working backwards, we included papers meeting our criteria until we found 50. To verify the sensitivity of our electronic search criteria, we conducted a manual search limited to 1 year (2008) in the tables of contents of the 20 paediatric journals and 10 general medical journals with the highest impact factors.²³ Our electronic search criteria found every 2008 article identified manually.

Inclusion and exclusion criteria

Studies were chosen for inclusion based on the article titles, and then on a review of the abstract, and if necessary the full article. We included all original studies that evaluated at least one time to diagnosis. The time to diagnosis was arbitrarily defined as the time from the onset of a disease (ie, reporting of the first symptoms) to its diagnosis. Diagnostic delay was defined as a time to diagnosis that was judged as too long by the authors of the published study. We defined paediatrics as medical and surgical procedures involving patients from birth to 18 years. Case reports with fewer than 20 patients (arbitrary) were excluded. Brief or short papers were also excluded because we have observed that the format of these articles forces authors to delete important methodological information.²⁴ Studies of prenatal diagnosis were excluded because their topic is closer to screening rather than diagnosis. Articles reporting studies in which fewer than half the patients were children were excluded (arbitrary) unless the paediatric data could be extracted.

Evaluation of reporting quality: specific checklist

We established a specific checklist based on (1) the criteria used for disease screening by Wilson and Jugner and published by WHO,²⁵ (2) the list of items used in STROBE (STrengthening the Reporting of OBServational studies in Epidemiology),²⁶ STARD (STAndards for the Reporting of Diagnostic accuracy studies),²⁷ REMARK (REporting recommendations for tumour MARKer prognostic studies),²⁸ and QUADAS (QUALity Assessment of studies of Diagnostic Accuracy included in Systematic reviews) I and II,^{29 30} (3) a preliminary analysis of 10 articles about time to diagnosis in paediatrics^{31–40} and (4) our research team's previous experience in the field of time to diagnosis.^{8 24 41–44} A first checklist was reviewed and discussed by three experienced epidemiologists (MC, CD-T, PR) and we finally obtained a 35-item checklist (table 1). Table 2 presents more thorough descriptions of the 21 checklist items that appear to us to be most specific and most critical for assessing the quality of reporting, specifically, the elements potentially associated with risks of bias and variation. Among these 21 items, 11 were considered very important (see bold item in table 1 or web table 3 for justifications).

Data extraction

Two of the authors (EL and MM) independently conducted the search, identified the potentially eligible articles and extracted the items necessary for evaluating reporting quality (see below). Each item could be scored in three categories: 'yes', 'no' or 'partially'. A third author (MC) participated in the discussion to help reach a consensus in disagreements about eligibility or data extraction.

Analyses

We described the general characteristics of journals from which articles were extracted (specialty as paediatric or non-paediatric and impact factor as a continuous variable or dichotomised around the 90th centile) and general characteristics of the articles (publication year, main clinical topics and recruitment type, ie, population-based or ordinary or referral level of care). We described the frequency of reporting of the 35 checklist items and the number of studies that reported all 11 items considered very important. To avoid multiple comparisons, we selected 5 of the 11 items in advance and analysed the relations between the frequency of their reporting and the potential determinants (publication year, journal specialty and impact factor), using χ^2 , χ^2 for trend and Mann-Whitney tests as appropriate (SPSS V.19.0, SPSS, Chicago, Illinois, USA). These five items were: (1) patient identification (item 7, table 1), (2) definition of warning symptoms (ie, symptoms that should trigger procedures leading to diagnosis) (item 8, table 1), (3) definition of first detectable symptoms of the condition (T0, item 13a, table 1), (4) definition of the precise time of diagnosis (Tx, item 11b, table 1) and (5) mean or median interval (item 22a, table 1). For these analyses, we considered two rather than three possible categories: yes or no, by grouping 'yes' and 'partially'.

RESULTS

Articles

Our search strategy identified 568 articles; working chronologically backwards from October 2011 to January 2005, we selected 50 articles (see appendix for the article references and flow chart). Specific topics varied widely, but four were recurrent: cancer (n=10), stroke (n=5), trauma consequences (n=4) and Kawasaki disease (n=4). The general characteristics of journals and articles are detailed in table 2. In all, 37 (74%) articles

Table 1 Items (by section) used for the analysis of the quality of the reporting and their frequency of completion

Section and topic	Do the authors...	Number of publications n (%)* in which this item was				
		Reported	Partially reported	Not reported	Not applicable	
Introduction	1. State the aim of the study	50 (100)	0	0	0	
	2. State frequency, morbidity and mortality of the studied condition	33 (66)	7 (14)	10 (20)	0	
	3. State if there exists a treatment	28 (56)	2 (4)	20 (40)	0	
	4. State if early diagnosis/treatment potentially changes the outcome	36 (72)	5 (10)	9 (18)	0	
Methods <i>Identification of study population</i>	5. Describe study population					
	a. Inclusion criteria	48 (96)	0	2 (4)	0	
	b. Exclusion criteria	26 (52)	0	24 (48)	0	
	6. State if subgroups at risk of extreme time to diagnosis were excluded or analysed separately†	8 (16)	0	42 (84)	0	
	7. Describe how cases were identified	28 (56)	1 (2)	21 (42)	0	
	8. List the symptoms allowing the physician to trigger diagnostic procedure (alert symptoms) during the study period	14 (28)	2 (4)	34 (68)	0	
	9. State if all patients with alert symptoms underwent the reference diagnostic procedure	2 (4)	0	48 (96)	0	
	10. Describe recruitment					
	a. Type (single centre, multicentre, population-based)	47 (94)	0	3 (6)	0	
	b. Locations	44 (88)	0	6 (12)	0	
	c. Setting (primary, secondary or tertiary centre)	17 (44)	0	22 (56)	11 ^a	
	d. Period	45 (90)	1 (2)	4 (8)	0	
	<i>Definition and measurement of time to diagnosis</i>	11. State the definition used for				
		a. T0 of illness (symptoms, signs)	22 (44)	2 (4)	26 (52)	0
		b. Tx of diagnosis used in the study	24 (48)	3 (6)	23 (46)	0
		12. State the definition of the different time intervals within the time to diagnosis studied (patient interval, doctor interval) ‡	15 (40)	0	23 (60)	12
13. Describe						
a. how many assessors evaluated time to diagnosis		10 (20)	0	40 (80)	0	
b. if evaluations were independent		2 (20)	3 (30)	5 (50)	40	
c. the degree of agreement about time to diagnosis between the different assessors		1 (20)	0	4 (80)	45	
d. how potential disagreements were resolved		1 (20)	0	4 (80)	45	
14. Describe the qualification of the assessors who evaluated time to diagnosis		9 (18)	0	41 (82)	0	
15. State whether the persons who determined time to diagnosis were blinded to the outcome	1 (2)	0	49 (98)	0		
<i>Statistical analyses</i>	16. Definition of delayed diagnosis (or in results)‡	20 (87)	0	3 (13)	27 ^b	
	a. Defined with reference to literature	6				
	b. Arbitrarily defined	12				
	c. Defined as variation in the distribution of diagnostic delay observed in the study population	2				
	17. Describe the statistical methods used to analyse the determinants and/or consequences of diagnostic delay	40 (87)	0	6 (3)	4	
	18. Give a rationale for the sample size	1 (2)	0	49 (98)	0	
Results	19. Describe how confounders were taken into account in the analysis of the determinants or consequences of diagnostic delay‡	21 (47)	0	24 (53)	5 ^c	
	20. Report number of					
	a. Eligible patients	33 (66)	1 (2)	16 (32)	0	
	b. Included patients	50 (100)	0	0	0	
	21. Report the characteristics of the study population	43 (86)	1 (2)	6 (12)	0	
	22. Report for the entire group					
	a. Mean/median diagnosis interval	30 (60)	0	20 (40)	0	
	b. Distribution of time to diagnosis	31 (62)	0	19 (38)	0	
23. Report the frequency of delayed diagnosis in the study population	18 (78)	0	5 (22)	27 ^d		
24. Report the determinants of diagnostic delay‡	37 (95)	0	2 (5)	11 ^b		
25. Report the consequences of diagnosis delay‡	21 (95)	0	1 (5)	28 ^e		
Discussion <i>Main results</i>	26. Discuss key results and interpret them	50 (100)	0	0	0	
	27. Compare results with those of previously published studies	43 (86)	0	7 (14)	0	

Continued

Table 1 Continued

Section and topic	Number of publications n (%)* in which this item was			
	Reported	Partially reported	Not reported	Not applicable
<i>Study limitation</i>				
28. Discuss possible recruitment bias	12 (30)	0	28 (70)	10 ^a
29. Discuss possible undiagnosed cases	6 (12)	1 (2)	43 (86)	0 ^f
30. Discuss the bias in and the precision of the measurement of interval	19 (38)	2 (4)	29 (58)	0
31. Discuss possible lead time bias†	0	0	7 (100)	43
<i>Clinical applicability</i>				
32. Report if delayed diagnosis was too frequent or time to diagnosis too long‡	37 (78)	0	10 (22)	3 ^g
33. Discuss avoidability of the delay‡	19 (40)	6 (13)	22 (47)	3 ^h
34. Propose plan of action to reduce diagnostic delay	17 (36)	21 (45)	9 (19)	3 ^h
35. Discuss the sensitivity, specificity and feasibility of this plan‡	3 (8)	0	35 (92)	12

*Denominator of percentage is the total number of concerned articles (total articles minus those 'not applicable').

†Items particularly important for articles mentioning a corrective action.

‡Items for which the number of reported or inapplicable items cannot be evaluated without medical knowledge of the relevant condition.

Items in bold are the critical items specific to time to diagnosis and potentially linked to risk of bias and variation.

a=in population; b=articles that do not mention the concept of delayed diagnosis.

c=articles that do not evaluate determinants and consequences.

d=articles that do not evaluate determinants.

e=articles that do not evaluate consequences.

f=articles that do not mention the relation between delay and survival.

g=articles that do not report that the diagnostic delay is too long.

h=articles that do not mention any corrective action.

concluded that corrective action was needed to reduce time to diagnosis.

Quality of reporting

Table 1 lists the 35 checklist items and the frequency of their reporting in the 50 articles studied. Overall, the frequency of reporting was high for generic items but low for items more specific to studies of time to diagnosis. Two articles reported all 11 of the items considered very important.^{44 45}

Potential determinants of the quality of reporting

Completion of items 22 (mean or median time to diagnosis) and 11b (definition of the precise time of diagnosis, Tx) rose significantly over the study period (from 0 and 1/3 in 2005 to 8/9 and 6/9 in 2011, $p < 0.02$). No significant trend was found for the other three items selected ($p > 0.2$). We found no significant relation ($p > 0.1$) between the completion of the five selected items and either the journal impact factor or type (paediatric vs non-paediatric) ($p > 0.1$).

DISCUSSION

Main results

In this systematic review, we have identified several weaknesses in the quality of reporting of studies of time to diagnosis; these shortcomings prevent the evaluation of risks of bias and variation that might directly affect potential corrective actions. The quality of reporting was good for generic items but poorer for items more specific to studies of time to diagnosis. Only 14 (28%) articles specifically listed warning symptoms, and only two (4%) stated whether the alert symptoms systematically triggered performance of the diagnostic procedure. We note that we were not assessing the quality of the studies directly, but rather the quality of their reports. We may assume that some items were studied but not explicitly stated in the study report. However, it has been shown in other areas that the quality of the reporting of the methods more frequently reflects flaws in the actual study methodology than in the reporting.⁴⁶

The accuracy and applicability of the measurements of time to diagnosis were limited by the lack of an explicit definition of this interval in more than half the articles. On the other hand, the trend over the study period towards a rise in the frequency of reporting of some checklist items (definition of time of diagnosis and mean or median time to diagnosis) indicates that improvement of reporting quality in this area is possible. Our finding that two articles even reported all 11 critical items further demonstrates that a high quality of reporting is feasible. We found no association between journal impact factors and the frequency of reporting of key items. This lack of relation suggests that the reporting weaknesses could be linked more to lack of awareness of these issues or to ubiquitous underestimation of the risk of bias and variation in these studies than to the quality of the editorial process in the relevant journals. Finally, nearly three quarters of the articles (37/50) concluded that an action plan to reduce delays was required. Unfortunately, however, the relevance of such action plans was limited by failure to discuss the avoidability of excessive time to diagnosis or to assess its determinants or consequences, or to specify the sensitivity, specificity and feasibility of the proposed corrective action.

Need for specific reporting guidelines

The poor quality of reporting of the 50 studies we reviewed supports our initial hypothesis—that specific reporting guidelines must be developed and implemented (or existing guidelines extended). Studies of time to diagnosis share some particularities

Table 2 General characteristics of journals and articles

<i>Journals' general characteristics</i>	
Median impact factor (range)	2.8 (0.6–8.2)
Number of journals published in the top 10 of the discipline (%)	14 (28)
<i>Discipline (%)</i>	
Paediatrics	23 (46)
Generalist	5 (10)
Cancer	4 (8)
Other medical specialty	12 (24)
Other surgical specialty	5 (10)
Epidemiology	1 (2)
<i>Articles' general characteristics</i>	
<i>Year of publication (%)</i>	
2005	3 (6)
2006	4 (8)
2007	7 (14)
2008	9 (18)
2009	8 (16)
2010	10 (20)
2011	9 (18)
<i>Recruitment (%)</i>	
Population-based	10 (20)
Referral centre	14 (28)

with studies evaluating diagnostic test accuracy, such as the risk of bias related to population selection or the need to assess time to diagnosis blinded to the outcome. On the other hand, studies assessing diagnostic tests differ from studies of time to diagnosis, for the latter concern the entire process from the first signs and symptoms to diagnosis, and not simply a single test. Thus, for example, STARD and QUADAS do not explicitly mention (nor do they need to) a requirement for reporting the list of symptoms that call for the physician to order diagnostic procedures (warning symptoms) or the procedure used to assess whether all patients with one or more warning symptoms underwent the standard diagnostic procedure.

Similarly studies of time to diagnosis might be considered a particular type of observational study that should be reported according to the STROBE statement. Application of those guidelines would not, however, improve the reporting of critical items concerning the definition and measurements of time to diagnosis. The STROBE guideline states: “give diagnostic criteria, if available”. For studies of time to diagnosis, however, listing the diagnostic criteria is not enough; the time of diagnosis and of the first symptoms (item 11) must be reported precisely, for they are the bounds of the measured interval. These items were clearly reported in less than half of our studies. A more explicit item formulation than that of STROBE might result in more frequent reporting.

In 2012, Weller *et al* published the Aarhus statement, intended to improve the design and reporting of studies of early cancer diagnosis. This statement provides very precise indications on how to report the definition and measurement of time to diagnosis, but it does not provide a reporting checklist for critical points such as the population selection process, the methods used to study determinants and consequences of diagnosis intervals, and the design of the corrective action planned.⁴⁷

Justification of potentially critical items

Our checklist uses new and original items for assessing the quality of the reporting of methodological aspects of studies of

time to diagnosis. Moher *et al*⁴⁸ studied scales and checklists for assessing the quality of randomised controlled trials and showed that most authors do not justify the development of their tools. Of the 35 items we used, some were common to all observational studies, 21 were more specific to the report of studies evaluating time to diagnosis (web table 3) and 11 were identified as critical items (7–9, 11, 13–15, 22, 28–30) that might affect the internal and external quality of the results. These 11 items are discussed below and in more detail in web table 3. Other items are optional but address the quality of interpretation and potential clinical applications (19, 24–25, 31, 33).

Population selection (items 7–9, 28–29)

Studies of time to diagnosis are very sensitive to selection bias. For example, recruitment from a hospital (or still worse, a referral centre) creates a risk of bias related to the patients' phenotypical characteristics and the organisation of care pathways. It is then important to report how patients were identified and to discuss recruitment bias for non-population based studies (items 7 and 28).

Studies of time to diagnosis analyse only diagnosed cases. Depending on the natural history of the disease, authors must consider the possibility of (1) death before diagnosis, (2) minimally symptomatic undiagnosed cases and (3) undiagnosed cases due to potential problems with the routine diagnostic process during the study period (item 29). Because these procedures apply to patients with at least one alert symptom (item 9), unlike diagnostic screening tests, the alert symptoms should be reported (item 8). For example, authors reporting the time to diagnosis of neonatal strokes have not discussed the possibility of either minor forms or rapidly lethal massive forms that remained undiagnosed (item 29).³⁸ The authors did not indicate, among infants experiencing at least one of the symptoms that can lead to the diagnosis of stroke (item 8), the proportion of those who had the diagnostic reference procedure (brain imaging) (item 9).

Definition and measurement of time to diagnosis (items 11,13–15, 22, 30)

Imprecise definition and measurement of time to diagnosis also leads to inaccuracy. As emphasised in the Aarhus statement, the various methods that can be used to determine date of first symptoms and date of diagnosis should be clearly reported to allow assessment of the risk of bias and variation (item 11).⁴⁷ However, the Aarhus statement does not recommend the reporting of other potential sources of bias, such as the lack of independence of the experts (item 13), the procedure used to manage discordance (item 13 a and b), the expert's qualifications and type of practice (item 14). Without a blinding procedure that must be reported, analysis of the association between diagnosis intervals and outcome or determinants may be subjected to measurement bias (item 15).

Study weaknesses

Moher recommended that one of the initial steps in developing health research reporting guidelines be to identify key information related to the potential sources of bias in studies. For example, Schulz *et al*⁴⁹ tested associations between estimates of treatment effects and inadequate allocation concealment, exclusions after randomisation and lack of double-blinding to prove the influence of bias linked to these three methodological points. This was not possible in our study because of the limited number of studies evaluating the interval for the same disease and because less than two-thirds of the studies explicitly

reported a quantitative interval or a frequency of delayed diagnosis that could have been the object of a comparison. The items selected were therefore deduced, in part from other types of studies that reviewed these risks of bias,^{50 51} and in part from the field experience of the research team and validated by three experienced epidemiologists.^{24 41–44 52–56}

Our study was limited to articles published in English, which might well have led to selection bias, the direction and strength of which cannot be anticipated.^{57 58} The statistical power of our study might not be sufficient to highlight the determinants of reporting quality. Finally, it is likely that the quality of articles about time to diagnosis in paediatrics does not reflect the quality in other areas of medicine. It is nonetheless difficult to predict the magnitude and direction of bias related to the restriction to this age group, particularly at risk of long time to diagnosis.

Conclusion and perspectives

The quality of reporting of studies about time to diagnosis in paediatric practice is globally poor. Only two articles among the 50 reported all 11 critical items. Less than 30% reported the list of alert symptoms and less than 10% mentioned whether all patients with alert symptoms underwent the standard diagnostic procedure. Less than a half reported the definition of the bounds of the diagnosis interval. This highlights the need for reporting guidelines. Existing guidelines (STARD,²⁷ STROBE²⁶ and the Aarhus statement⁴⁷) do not completely meet this need, and we have completed them with more specific items, intended to focus on the risk of bias, factors of variation and their impact on the corrective actions that might be proposed. These proposed items should now be reinforced by a review for evidence of their association with the particular risks of bias and variation. This review could compare articles reporting different methodologies and studying intervals for the same diagnosis.⁵¹ This would require enlarging our review beyond the paediatric population. Further steps in the development of specific reporting guidelines would include the submission of these items to a larger panel of experts by a Delphi procedure, face-to-face meetings and pilot testing.¹⁴

Acknowledgements Inserm Unit 953 has received a grant from the Bettencourt Foundation (Coups d'élan pour la recherche française) in support of its research activities.

Contributors Study concept and design: EL, MM, CD-T, CGLeG, MC, PR. Analysis and interpretation of data: EL, MM, MC. Drafting of the manuscript: EL, MC. Critical revision of the manuscript for important intellectual content: MM, CD-T, CGLeG, PR. Statistical analysis: EL, MC. Study supervision: MC, CD-T, PR.

Competing interests None.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

- James JT. A new, evidence-based estimate of patient harms associated with hospital care. *J Patient Saf* 2013;9:122–8.
- E.M.A. Guideline on the role of pharmacokinetics in the development of medicinal products in the paediatric population. http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003066.pdf (accessed 28 Aug 2012).
- McD Taylor D, Robinson J, MacLeod D, et al. Therapeutic errors among children in the community setting: nature, causes and outcomes. *J Paediatr Child Health* 2009;45:304–9.
- Otero P, Leyton A, Mariani G, et al. Medication errors in pediatric inpatients: prevalence and results of a prevention program. *Pediatrics* 2008;122:e737–43.
- Liozzi C, Noble G, Franck LS. How parents make sense of their young children's expressions of everyday pain: A qualitative analysis. *Eur J Pain* 2012;16:1166–75.
- Klein-Geltink J, Shaw AK, Morrison HI, et al. Use of paediatric versus adult oncology treatment centres by adolescents 15–19 years old: the Canadian Childhood Cancer Surveillance and Control Program. *Eur J Cancer* 2005;41:404–10.
- Claudius I, Baraff LJ. Pediatric emergencies associated with fever. *Emerg Med Clin North Am* 2010;28:67–84, vii–viii.
- Launay E, Gras-Le Guen C, Martinot A, et al. Suboptimal care in the initial management of children who died from severe bacterial infection: a population-based confidential inquiry. *Pediatr Crit Care Med* 2010;11:469–74.
- Rose MV, Kimaro G, Nissen TN, et al. QuantiFERON(R)-TB gold in-tube performance for diagnosing active tuberculosis in children and adults in a high burden setting. *PLoS ONE* 2012;7:e37851.
- The EQUATOR Network website—the resource centre for good reporting of health research studies. <http://www.equator-network.org/> (accessed 6 Sep 2011).
- Kane RL, Wang J, Garrard J. Reporting in randomized clinical trials improved after adoption of the CONSORT statement. *J Clin Epidemiol* 2007;60:241–9.
- Moher D, Jones A, Lepage L. Use of the CONSORT statement and quality of reports of randomized trials: a comparative before-and-after evaluation. *JAMA* 2001;285:1992–5.
- Plint AC, Moher D, Morrison A, et al. Does the CONSORT checklist improve the quality of reports of randomised controlled trials? A systematic review. *Med J Aust* 2006;185:263–7.
- Moher D, Schulz KF, Simera I, et al. Guidance for developers of health research reporting guidelines. *PLoS Med* 2010;7:e1000217.
- Berwanger O, Ribeiro RA, Finkelsztein A, et al. The quality of reporting of trial abstracts is suboptimal: survey of major general medical journals. *J Clin Epidemiol* 2009;62:387–92.
- Sacks HS, Berrier J, Reitman D, et al. Meta-analyses of randomized controlled trials. *N Engl J Med* 1987;316:450–5.
- Altman DG, Dore CJ. Randomisation and baseline comparisons in clinical trials. *Lancet* 1990;335:149–53.
- Laupacis A, Sekar N, Stiell IG. Clinical prediction rules. A review and suggested modifications of methodological standards. *JAMA* 1997;277:488–94.
- Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *JAMA* 2000;283:2008–12.
- Wasson JH, Sox HC, Neff RK, et al. Clinical prediction rules. Applications and methodological standards. *N Engl J Med* 1985;313:793–9.
- Mulrow CD. The medical review article: state of the science. *Ann Intern Med* 1987;106:485–8.
- Bouwmeester W, Zuihoff NP, Mallett S, et al. Reporting and methods in clinical prediction research: a systematic review. *PLoS Med* 2012;9:1–12.
- Journal citation reports. Web of knowledge. <http://admin-apps.webofknowledge.com/JCR/JCR?PointOfEntry=Home&SID=U2GfKkciCP@BFFmPJJ> (accessed 11 Apr 2012).
- Chalumeau M, Holvoet L, Cheron G, et al. Delay in diagnosis of imported Plasmodium falciparum malaria in children. *Eur J Clin Microbiol Infect Dis* 2006;25:186–9.
- Wilson JMG, Jungner G. *Principles and practice of screening for disease*. WHO. Public health papers n 34. 1968. http://whqlibdoc.who.int/php/WHO_PHP_34.pdf (accessed 19 Jun 2013).
- von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *J Clin Epidemiol* 2008;61:344–9.
- Bossuyt PM, Reitsma JB. The STARD initiative. *Lancet* 2003;361:71.
- McShane LM, Altman DG, Sauerbrei W, et al. Reporting recommendations for tumor MARKER prognostic studies (REMARK). *Nat Clin Pract Urol* 2005;2:416–22.
- Whiting P, Rutjes AW, Reitsma JB, et al. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol* 2003;3:25.
- Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;155:529–36.
- Ciafaloni E, Fox DJ, Pandya S, et al. Delayed diagnosis in duchenne muscular dystrophy: data from the Muscular Dystrophy Surveillance, Tracking, and Research Network (MD STARnet). *J Pediatr* 2009;155:380–5.
- Cappendijk VC, Hazebroek FW. The impact of diagnostic delay on the course of acute appendicitis. *Arch Dis Child* 2000;83:64–6.
- Hartman AL, Lunney KM, Serena JE. Pediatric stroke: do clinical factors predict delays in presentation? *J Pediatr* 2009;154:727–32.
- McIntyre PB, Macintyre CR, Gilmour R, et al. A population based study of the impact of corticosteroid therapy and delayed diagnosis on the outcome of childhood pneumococcal meningitis. *Arch Dis Child* 2005;90:391–6.
- Pais PJ, Kump T, Greenbaum LA. Delay in diagnosis in poststreptococcal glomerulonephritis. *J Pediatr* 2008;153:560–4.
- Rafay MF, Pontigon AM, Chiang J, et al. Delay to diagnosis in acute pediatric arterial ischemic stroke. *Stroke* 2009;40:58–64.
- Savendahl L, Davenport ML. Delayed diagnoses of Turner's syndrome: proposed guidelines for change. *J Pediatr* 2000;137:455–9.
- Srinivasan J, Miller SP, Phan TG, et al. Delayed recognition of initial stroke in children: need for increased awareness. *Pediatrics* 2009;124:e227–34.
- Urschel S, Kayikci L, Wintergerst U, et al. Common variable immunodeficiency disorders in children: delayed diagnosis despite typical clinical presentation. *J Pediatr* 2009;154:888–94.

- 40 Wilder MS, Palinkas LA, Kao AS, *et al.* Delayed diagnosis by physicians contributes to the development of coronary artery aneurysms in children with Kawasaki syndrome. *Pediatr Infect Dis J* 2007;26:256–60.
- 41 Brasme JF, Chalumeau M, Doz F, *et al.* Interval between onset of symptoms and diagnosis of medulloblastoma in children: distribution and determinants in a population-based study. *Eur J Pediatr* 2012;171:25–32.
- 42 Brasme JF, Grill J, Doz F, *et al.* Long time to diagnosis of medulloblastoma in children is not associated with decreased survival or with worse neurological outcome. *PLoS ONE* 2012;7:e33415.
- 43 Brasme JF, Morfouace M, Grill J, *et al.* Delays in diagnosis of paediatric cancers: a systematic review and comparison with expert testimony in lawsuits. *Lancet Oncol* 2012;13:e445–59.
- 44 Gascoin-Lachambre G, Brauner R, Duche L, *et al.* Pituitary stalk interruption syndrome: diagnostic delay and sensitivity of the auxological criteria of the growth hormone research society. *PLoS ONE* 2011;6:e16367.
- 45 Ravichandiran N, Schuh S, Bejuk M, *et al.* Delayed identification of pediatric abuse-related fractures. *Pediatrics* 2010;125:60–6.
- 46 Liberati A, Himel HN, Chalmers TC. A quality assessment of randomized control trials of primary treatment of breast cancer. *J Clin Oncol* 1986;4:942–51.
- 47 Weller D, Vedsted P, Rubin G, *et al.* The Aarhus statement: improving design and reporting of studies on early cancer diagnosis. *Br J Cancer* 2012;106:1262–7.
- 48 Moher D, Jadad AR, Nichol G, *et al.* Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. *Control Clin Trials* 1995;16:62–73.
- 49 Schulz KF, Chalmers I, Hayes RJ, *et al.* Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995;273:408–12.
- 50 Whiting P, Rutjes AW, Reitsma JB, *et al.* Sources of variation and bias in studies of diagnostic accuracy: a systematic review. *Ann Intern Med* 2004;140:189–202.
- 51 Lijmer JG, Mol BW, Heisterkamp S, *et al.* Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA* 1999;282:1061–6.
- 52 Bouvier-Colle MH, Breart G. Surveillance of maternal mortality in France: why and how? *Gynecol Obstet Fertil* 2004;32:925–6.
- 53 Deneux-Tharaux C, Berg C, Bouvier-Colle MH, *et al.* Underreporting of pregnancy-related mortality in the United States and Europe. *Obstet Gynecol* 2005;106:684–92.
- 54 Philibert M, Deneux-Tharaux C, Bouvier-Colle MH. Can excess maternal mortality among women of foreign nationality be explained by suboptimal obstetric care? *BJOG* 2008;115:1411–18.
- 55 Boutron I, Moher D, Tugwell P, *et al.* A checklist to evaluate a report of a nonpharmacological trial (CLEAR NPT) was developed using consensus. *J Clin Epidemiol* 2005;58:1233–40.
- 56 Boutron I, Ravaud P. Classification systems to improve assessment of risk of bias. *J Clin Epidemiol* 2012;65:236–8.
- 57 Egger M, Zellweger-Zahner T, Schneider M, *et al.* Language bias in randomised controlled trials published in English and German. *Lancet* 1997;350:326–9.
- 58 Moher D, Pham B, Klassen TP, *et al.* What contributions do languages other than English make on the results of meta-analyses? *J Clin Epidemiol* 2000;53:964–72.

Appendices for on-line publication.

Appendix 1. List of articles reviewed for quality evaluation of reporting.

1. Aghamohammadi A, Bahrami A, Mamishi S, et al. Impact of delayed diagnosis in children with primary antibody deficiencies. *J Microbiol Immunol Infec* 2011;44:229-34.
2. Anderson MS, Todd JK, Glode MP. Delayed diagnosis of Kawasaki syndrome: an analysis of the problem. *Pediatric* 2005;115:e428-33.
3. Azarsiz E, Gulez N, Edeer Karaca N, Aksu G, Kutukculer N. Consanguinity rate and delay in diagnosis in Turkish patients with combined immunodeficiencies: a single-center study. *J Clin Immuno* 2011;31:106-11.
4. Bai S, Ren R, Li B, et al. Delay in the diagnosis of retinoblastoma in China. *Acta Ophthalmol* 2011;89:e72-4.
5. Bayne AP, Madden-Fuentes RJ, Jones EA, et al. Factors associated with delayed treatment of acute testicular torsion-do demographics or interhospital transfer matter? *J Urol* 2010;184:1743-7.
6. Braun KP, Kappelle LJ, Kirkham FJ, Deveber G. Diagnostic pitfalls in paediatric ischaemic stroke. *Dev Med Child Neuro*. 2006;48:985-90.
7. Ciafaloni E, Fox DJ, Pandya S, et al. Delayed diagnosis in duchenne muscular dystrophy: data from the Muscular Dystrophy Surveillance, Tracking, and Research Network (MD STARnet). *J Pediat* 2009;155:380-5.
8. Crawford JR, Santi MR, Vezina G, et al. CNS germ cell tumor (CNSGCT) of childhood: presentation and delayed diagnosis. *Neurology*. 2007;68:1668-73.
9. Dang-Tan T, Trottier H, Mery LS, et al. Delays in diagnosis and treatment among children and adolescents with cancer in Canada. *Pediatr Blood Cancer* 2008;51:468-74.

10. Dang-Tan T, Trottier H, Mery LS, et al. Determinants of delays in treatment initiation in children and adolescents diagnosed with leukemia or lymphoma in Canada. *Int J Cancer* 2010;126:1936-43.
11. England RJ, Crabbe DC. Delayed diagnosis of appendicitis in children treated with antibiotics. *Pediatr Surg Int* 2006;22:541-5.
12. Fehr S, Downs J, Bebbington A, Leonard H. Atypical presentations and specific genotypes are associated with a delay in diagnosis in females with Rett syndrome. *Am J Med Genet A* 2010;152A:2535-42.
13. Ferrari A, Miceli R, Casanova M, et al. The symptom interval in children and adolescents with soft tissue sarcomas. *Cancer* 2010;116:177-83.
14. Foster HE, Eltringham MS, Kay LJ, Friswell M, Abinun M, Myers A. Delay in access to appropriate care for children presenting with musculoskeletal symptoms and ultimately diagnosed with juvenile idiopathic arthritis. *Arthritis Rheum.* 2007;57:921-7.
15. Gascoin-Lachambre G, Brauner R, Duche L, Chalumeau M. Pituitary stalk interruption syndrome: diagnostic delay and sensitivity of the auxological criteria of the growth hormone research society. *PLoS One* 2011;6:e16367.
16. Haimi M, Perez-Nahum M, Stein N, Ben Arush MW. The role of the doctor and the medical system in the diagnostic delay in pediatric malignancies. *Cancer Epidemiol* 2011;35:83-9.
17. Hartman AL, Lunney KM, Serena JE. Pediatric stroke: do clinical factors predict delays in presentation? *J Pediatr* 2009;154:727-32.
18. Jackson A, Foley L, Daly L, et al. Delayed cystic fibrosis presentation in children in the absence of newborn screening. *Ir Med J* 2010;103:113-6.
19. Juan CC, Hwang B, Lee PC, et al. The clinical manifestations and risk factors of a delayed diagnosis of Kawasaki disease. *J Chin Med Assoc* 2007;70:374-9.

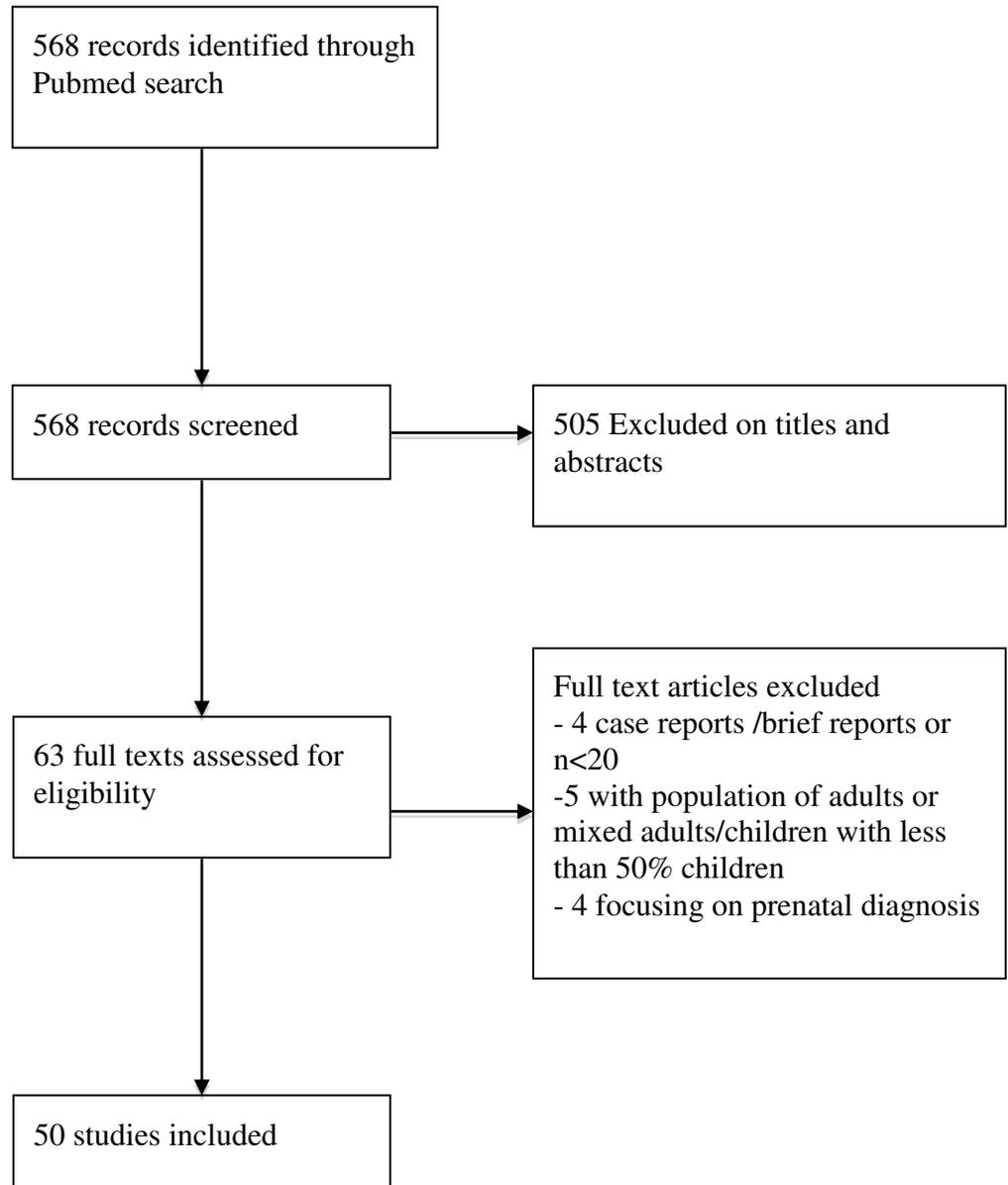
20. Kim MS, Lee SY, Cho WH, et al. Prognostic effects of doctor-associated diagnostic delays in osteosarcoma. *Arch Orthop Trauma Surg* 2009;129:1421-5.
21. Kukal K, Dobrovoljac M, Boltshauser E, Ammann RA, Grotzer MA. Does diagnostic delay result in decreased survival in paediatric brain tumours? *Eur J Pediatr* 2009;168:303-10.
22. Letton RW, Jr., Worrell V, Tuggle DW. Delay in diagnosis and treatment of blunt intestinal perforation does not adversely affect prognosis in the pediatric trauma patient. *J Trauma* 2010;68:790-5.
23. Majak P, Bak-Walczak E, Stelmach I, Jerzynska J, Krakowiak J, Stelmach W. An increasing trend of the delay in asthma diagnosis after the discontinuation of a population-based intervention. *J Asthma* 2011;48:414-8.
24. Mann JR, Crawford S, Wilson L, McDermott S. Does race influence age of diagnosis for children with developmental delay? *Disabil Health J* 2008;1:157-62.
25. McCormick J, Sims EJ, Mehta A. Delayed diagnosis of females with respiratory presentation of cystic fibrosis did not segregate with poorer clinical outcome. *J Clin Epidemiol* 2006;59:315-22.
26. McGlennan C, Ganesan V. Delays in investigation and management of acute arterial ischaemic stroke in children. *Dev Med Child Neurol* 2008;50:537-40.
27. McIntyre PB, Macintyre CR, Gilmour R, Wang H. A population based study of the impact of corticosteroid therapy and delayed diagnosis on the outcome of childhood pneumococcal meningitis. *Arch Dis Child* 2005;90:391-6.
28. Minich LL, Sleeper LA, Atz AM, et al. Delayed diagnosis of Kawasaki disease: what are the risk factors? *Pediatrics* 2007;120:e1434-40.
29. Mistry PK, Sadan S, Yang R, Yee J, Yang M. Consequences of diagnostic delays in type 1 Gaucher disease: the need for greater awareness among hematologists-oncologists and an opportunity for early diagnosis and intervention. *Am J Hematol* 2007;82:697-701.

30. Pais PJ, Kump T, Greenbaum LA. Delay in diagnosis in poststreptococcal glomerulonephritis. *J Pediatr* 2008;153:560-4.
31. Pan KL, Chan WH, Chia YY. Initial symptoms and delayed diagnosis of osteosarcoma around the knee joint. *J Orthop Surg (Hong Kong)* 2010;18:55-7.
32. Pawlowicz M, Birkholz D, Niedzwiecki M, Balcerska A. Difficulties or mistakes in diagnosing type 1 diabetes in children?--demographic factors influencing delayed diagnosis. *Pediatr Diabetes* 2009;10:542-9.
33. Perno JF, Schunk JE, Hansen KW, Furnival RA. Significant reduction in delayed diagnosis of injury with implementation of a pediatric trauma service. *Pediatr Emerg Care* 2005;21:367-71.
34. Purper-Ouakil D, Cortese S, Wohl M, et al. Predictors of diagnostic delay in a clinical sample of French children with attention-deficit/hyperactivity disorder. *Eur Child Adolesc Psychiatry* 2007;16:505-9.
35. Rafay MF, Pontigon AM, Chiang J, et al. Delay to diagnosis in acute pediatric arterial ischemic stroke. *Stroke* 2009;40:58-64.
36. Rahme D, Comley A, Foster B, Cundy P. Consequences of diagnostic delays in slipped capital femoral epiphysis. *J Pediatr Orthop B* 2006;15:93-7.
37. Ravichandiran N, Schuh S, Bejuk M, et al. Delayed identification of pediatric abuse-related fractures. *Pediatrics* 2010;125:60-6.
38. Shilon Y, Pollak Y, Benarroch F, Gross-Tsur V. Factors influencing diagnosis delay in children with Tourette syndrome. *Eur J Paediatr Neurol* 2008;12:398-400.
39. Shlizerman L, Mazzawi S, Rakover Y, Ashkenazi D. Foreign body aspiration in children: the effects of delayed diagnosis. *Am J Otolaryngol* 2010;31:320-4.
40. Srinivasan J, Miller SP, Phan TG, Mackay MT. Delayed recognition of initial stroke in children: need for increased awareness. *Pediatrics* 2009;124:e227-34.

41. Stefan DC, Siemonsma F. Delay and causes of delay in the diagnosis of childhood cancer in Africa. *Pediatr Blood Cancer* 2011;56:80-5.
42. Steinraths M, Vallance HD, Davidson AG. Delays in diagnosing cystic fibrosis: can we find ways to diagnose it earlier? *Can Fam Physician* 2008;54:877-83.
43. Timmer A, Behrens R, Buderus S, et al. Childhood onset inflammatory bowel disease: predictors of delayed diagnosis from the CEDATA German-language pediatric inflammatory bowel disease registry. *J Pediatr* 2011;158:467-73 e2.
44. Urschel S, Kayikci L, Wintergerst U, Notheis G, Jansson A, Belohradsky BH. Common variable immunodeficiency disorders in children: delayed diagnosis despite typical clinical presentation. *J Pediatr* 2009;154:888-94.
45. Viani RM, Lopez G, Chacon-Cruz E, Hubbard P, Spector SA. Poor outcome is associated with delayed tuberculosis diagnosis in HIV-infected children in Baja California, Mexico. *Int J Tuberc Lung Dis* 2008;12:411-6.
46. Weibel L, Laguda B, Atherton D, Harper JJ. Misdiagnosis and delay in referral of children with localized scleroderma. *Br J Dermatol* 2011;165:1308-13.
47. Weigall P, Vladusic S, Torode I. Slipped upper femoral epiphysis in children--delays to diagnosis. *Aust Fam Physician* 2010;39:151-3.
48. Weiss B, Skourikhin Y, Modan-Moses D, Broide E, Fradkin A, Bujanover Y. Is adult height of patients with celiac disease influenced by delayed diagnosis? *Am J Gastroenterol* 2008;103:1770-4.
49. Wilder MS, Palinkas LA, Kao AS, Bastian JF, Turner CL, Burns JC. Delayed diagnosis by physicians contributes to the development of coronary artery aneurysms in children with Kawasaki syndrome. *Pediatr Infect Dis J* 2007;26:256-60.

50. Williams BG, Hlaing T, Aaland MO. Ten-year retrospective study of delayed diagnosis of injury in pediatric trauma patients at a level II trauma center. *Pediatr Emerg Care* 2009;25:489-93.

Appendix 2. Flow chart from identification to inclusion of the articles studied.



Appendix 3. Table 3: Items most likely to be associated with risks of bias and variations, with illustrative examples.

Domain	Do the authors... (items from checklist -table 1-)	Illustrative explanations
Patient identification and selection	6. State if patient subgroups at risk of extreme diagnostic delay were excluded or analysed separately	In cases where the disease of interest is genetically transmitted or susceptibility is family-linked, failure to specify whether patients with a family history of the disease are excluded or analysed separately results in a risk of underestimating delay. ²⁹ In contrast, when a disease rarely affects only one particular group, failure to limit the study to this particular group may lead to overestimating delay (for example, including women in the study of a recessive genetic disease linked to the X chromosome). ⁷
	7. Describe how cases were identified* 28. Discuss possible recruitment bias*	Knowing the method of identification of the population enables the risk of selection bias and the exhaustiveness of the population to be assessed. The inclusion of patients identified by discharge codes ^{19 24} or by voluntary participation in a trial ^{28 34} can strongly bias the measurement of diagnostic delay. When a study is not population-based, the limitation of inclusion to patients on a register kept by specialists, by patient associations or referral services can lead to recruitment bias. Not discussing the effects of this bias can lead to misinterpretation of results by generalising results about a specific population to the general population. ^{29 35}
	8. List the symptoms allowing the physician to trigger diagnostic procedures (warning symptoms) during the study period* 9. State if all patients with warning symptoms underwent the reference diagnostic procedure* 29. Discuss potential undiagnosed cases*	There is a risk of bias if all symptoms that may alert the clinician to start the diagnostic procedure (warning symptoms) are not listed. Some more serious symptoms (e.g., motor deficit in stroke) could in fact lead to a more systematic diagnostic procedure and therefore to diagnosis while other less specific symptoms (such as apnea) may be associated with a longer diagnostic delay; failure to include those patients creates a risk of overstating the consequences of delay. ^{35 40} When patients die before any medical referral (cancer ⁴¹ or cystic fibrosis ⁴²), there is a risk of underestimating the consequences of diagnostic delay. On the contrary, there is a risk of overestimating the consequences of delay in diagnosis when patients remain undiagnosed for a long time without any major health problem or when spontaneous recovery is possible.

Table 3 continues.

Table 3 Continued.

Domain	Do the authors... (items from checklist -table 1-)	Illustrated explanation
Diagnostic delay measurement	<p>11. Describe the definition used in the study for T0 of illness (symptoms, signs) and Tx of diagnosis*</p> <p>30. Discuss the quality of the measurement of delay*</p>	<p>If the boundaries of the diagnostic delay, that is, the threshold beyond which delay is considered to be present, is not clearly defined, there is a risk of bias, of imprecision, and non-reproducibility, especially when:</p> <ul style="list-style-type: none"> - an investigator alone decides while reviewing records which symptoms were the first;^{5 7} - the patient is asked to remember later on when the disease or symptoms began (recall bias);⁴⁴ - Tx (time of diagnosis) may correspond either to the first mention by a clinician or to the confirmation by further investigations, which may be shifted in time.⁷
	<p>13. Describe*</p> <ul style="list-style-type: none"> a. how many assessors evaluated diagnostic delay b. if evaluations were independent c. the degree of agreement about diagnostic delay between the different assessors d. how potential disagreements were resolved <p>14. Describe the qualification of the assessors who evaluated diagnostic delay*</p> <p>15. State whether the persons who determined diagnostic delays were blinded to the outcome*</p>	<p>When diagnostic delay can result in variations in the measurement (e.g., when several symptoms can be considered as the first symptoms (T0) because for example they were not defined in advance), the evaluation of diagnostic delay by a single person can result in a more subjective measure than after discussion between several evaluators.^{5 7} This subjectivity is reduced if several people assess the delay independently.^{42 37} A low degree of agreement suggests that the reproducibility of the measurement of delay will be poor. The qualifications of the assessors may also influence the measurement. Specialists will choose the earliest symptoms of the disease, of which non-specialists may be unaware (when these symptoms are not defined in advance), or a subjective date of symptom onset (when these symptoms are nonspecific and fluctuate, such as headaches).^{14 30} Failure to blind evaluators to the outcome risks overestimating diagnostic delay and inducing differential bias.^{7 48}</p>
	<p>16. Definition of delayed diagnosis (<i>or in results section</i>)</p> <p>22. Report for the entire group*</p> <ul style="list-style-type: none"> a. Mean/median diagnostic delay b. Distribution of diagnostic delay 	<p>Diagnostic delay is a continuous time variable that can also be studied as dichotomised (delayed versus not delayed).^{19 28 49} Giving the mean or median diagnostic delay with standard deviation or range is essential for describing the major endpoint of study on diagnostic delay. Detecting possible extreme delay that might be linked to a particular phenotype of the disease is useful for interpreting the results appropriately.</p>

Table 3 continues.

Tableau 3 continued.

Domain	Do the authors... (items from checklist -table 1-)	Illustrated explanation
Preparation of corrective plan**	12. State the definition of the different time intervals within the diagnostic delay studied (patient delay, doctor delay)	Subdividing the diagnostic delay can help to target corrective actions (health education and/or medical education) by measuring the role of each participant in the delay (patients/parents, physicians/healthcare system). ^{14 45}
	19. Describe how confounders were taken into account to analyse the determinants or consequences of diagnostic delays 24. Report the determinants of diagnostic delay	The evaluation of the determinants of delay is not mandatory in studies evaluating diagnostic delays but it is a prerequisite for the preparation of corrective action. Proposing an action plan to reduce this delay without knowing the determinants creates the risk of developing an imprecise and ineffective plan. ^{7 42} Failure to take confounding factors into account in the analysis of the determinants can lead to an erroneous conclusion that an association exists between some characteristics of the patients or the health care system and the diagnostic delay. ^{17 35}
	19. Describe how confounders were taken into account to analyse the determinants or consequences of diagnostic delays 25. Report the consequences of diagnostic delay 31. Discuss possible lead time bias	The evaluation of the consequences of delays is not mandatory in studies evaluating diagnostic delays but it is a prerequisite for the preparation of corrective action. Proposing an action plan to reduce this delay without demonstrating its consequences first creates the risk of developing an ineffective plan. ^{7 45} Failure to take confounding factors into account in the analysis of the effects of delay can lead to an erroneous conclusion that an association exists between diagnostic delay and unfavourable outcome and to the initiation of corrective actions that will have no benefit for patients (or might even prove harmful). When a relation has been demonstrated between diagnostic delay and survival, failure to take the possibility of lead-time bias into account (in cases where the order of magnitude of diagnostic delay and survival were similar) produces a risk of overstating the consequences of diagnostic delay on survival. ^{14 35 45}
	33. Discuss the avoidability of the delay	Assessing the avoidability of delays is an essential prerequisite to corrective action. ^{15 45}
	35. Discuss the sensitivity, specificity and feasibility of this plan	If a plan provides the systematic implementation of a diagnostic procedure (e.g., radiological or laboratory) in response to one or more warning symptoms, it must be ensured that these symptoms are specific enough to the disease and especially prevalent in its early stages, and that the validity of the diagnostic procedure at this stage is known. ^{35 40 44}

*very important items for all articles

**very important items for articles mentioning a corrective plan

(see Appendix 1 for references)

Annexe 2 : Manuscrit issu de la Partie I - Chapitre 2

1 **“Quality assessment” of primary studies in systematic reviews of time to diagnosis: a**
2 **methodological systematic review and call for clarification**

3

4 Elise Launay^{1,2}, Jérémie F. Cohen^{1,3}, Michèle Morfouace^{1,3}, Christèle Gras-Le Guen²,
5 Philippe Ravaud⁴, Martin Chalumeau^{1,3}

6

7 1: Inserm UMR 1153, Obstetrical, Perinatal and Pediatric Epidemiology Research Team
8 (Epopé), Center for Epidemiology and Statistics Sorbonne Paris Cité, Paris Descartes
9 University, Paris, France

10 2: CHU de Nantes, Hôpital Mère-Enfant, services de pédiatrie générale et d’urgences
11 pédiatriques, Nantes, France

12 3: Service de pédiatrie générale, hôpital Necker-Enfants malades ; AP-HP ; Université Paris
13 Descartes, Paris, France

14 4: Inserm UMR 1153, METHODS Team, Center for Epidemiology and Statistics Sorbonne
15 Paris Cité. Paris Descartes University, Paris, France

16

17 Corresponding author: Elise Launay, INSERM U1153, Maternité de Port-Royal, 53 avenue de
18 l’observatoire, 75014 Paris, email: elise.launay@chu-nantes.fr, phone: +33171722991, fax:
19 +33143268979

20

21

1 **Abstract (201)**

2 **Introduction:** Studies of time to diagnosis (TTD) are observational by nature and subject to
3 risk of bias and threats to transportability. We aimed to analyse tools used in systematic
4 reviews (SRs) of studies investigating TTD to “assess the quality” of primary studies.

5 **Methods:** We systematically searched MEDLINE via PubMed and Web of Science for SRs
6 of TTD published up to the end of February 2015; we identified and characterised tools used
7 for “quality assessment” and classified their items.

8 **Results:** From 1,936 articles identified, we included 45 SRs that aimed to summarize the
9 available information on the length (n = 16), determinants (n = 31) and/or consequences (n =
10 15) of TTD. For the 23 SRs (51%) reporting a “quality assessment” process, 21 different
11 tools were used, with 232 items assessing quality of reporting (64%), risk of bias or threats to
12 transportability (43%), statistical issues (5%) and/or an unclear domain (0.5%); 11% were
13 specific to TTD issues.

14 **Conclusions:** Assessment of risk of bias and threats to transportability in primary studies
15 included in SRs of TTD was infrequent, non-standardised and rarely concerned TTD study
16 specificities. These findings highlight the need for guidance on assessment of the
17 methodological quality of studies of TTD.

18

19 **Key words:** time to diagnosis, early diagnosis, systematic review, risk of bias,
20 methodological quality

21

22

23

24

25

1 **What is new?**

2 *Key findings*

3 - In this methodological systematic review, we systematically evaluated the “assessment of
4 quality” in primary studies included in systematic reviews of TTD and found that only 51%
5 of the 45 included systematic reviews reported a “quality assessment” of the primary studies
6 and only 36% reported an assessment of risk of bias and/or threats to transportability.

7 - Among the 232 items composing the “quality assessment” tools, 64% assessed the quality
8 of reporting whereas only 43% assessed risk of bias and/or threats to transportability.

9

10 *What this adds to what was known*

11 - The necessity of assessing methodological quality of primary studies included in systematic
12 review is well accepted but the content of the quality assessment was frequently found
13 inadequate in the systematic reviews of time to diagnosis included in our study with important
14 confusion between quality of reporting and risk of bias and threats to transportability
15 assessment.

16

17 *What is the implication, what should change now*

18 - Our findings highlight the need for a specific tool to assess risk of bias and threat to
19 transportability for studies of time to diagnosis.

20

21

22

1 **1. Introduction**

2 Time to diagnosis (TTD) is defined as the interval from the first symptoms to the diagnosis of
3 a disease.[1,2] TTD may be related to patient behaviours, physician skills, diagnostic test
4 performance and/or healthcare system organisation, particularly health care accessibility.[2]
5 Delayed diagnosis may have direct medical and psychological consequences for patients and
6 their relatives and physicians and also medico-legal consequences: diagnosis delays are a
7 major cause of lawsuits in healthcare.[3,4] All these reasons probably explain why TTD is a
8 growing field of research with a marked increase of the number of related publications in the
9 last 15 years (Appendix A).

10 Studies of TTD have 3 main non-exclusive purposes: measuring length of TTD and its
11 evolution over time, identifying the determinants of long TTD, and/or evaluating the
12 relationship between short and long TTD and patient outcomes.[1,2] Long TTD is usually *a*
13 *priori* considered associated with worse outcome. For example, the definition for “early
14 diagnosis”, the Medical Subject Heading (MeSH) term for TTD studies in PubMed, is
15 “Generally, early diagnosis improves prognosis and treatment outcome”.[5] In most published
16 studies of TTD, authors have concluded a high prevalence of diagnosis delay associated with
17 worse outcome and the need for urgent corrective action.[1] Because most studies evaluating
18 TTD are observational and retrospective, they are at risk of bias and threats to transportability,
19 notably during the participant selection process, TTD measurement and study of the
20 association of TTD and participant characteristics or health outcomes.[1,2] It was shown that
21 authors of studies on TTD rarely address some critical issues related to the risk of
22 undiagnosed cases among the source population (12%), or to the lack of blinding to patient
23 outcomes while assessing time points defining TTD (2%).[1] These risks of bias and threats to
24 transportability need to be assessed to draw reliable conclusions from studies, notably during
25 the synthesis of available evidence in systematic reviews (SRs).

1 The internationally accepted reporting guideline for SRs, the Preferred Reporting
2 Items for Systematic Reviews and Meta-Analyses (PRISMA), its ancestor Quality of
3 Reporting of Meta-analyses (QUOROM) and the reporting guideline Meta-analysis Of
4 Observational Studies in Epidemiology (MOOSE) require that articles specify and present the
5 results of an assessment of risk of bias that may affect the cumulative evidence.[6–8] Risk of
6 bias tools were developed and validated to conduct SRs of randomised control trials and
7 diagnosis test accuracy studies;[9,10] however, those available for observational studies are
8 generic, mostly developed for studies of interventions,[11] or are not fully validated.[12,13]
9 As a first step to develop a specific risk of bias tool for studies of TTD, we aimed to
10 systematically identify and characterize tools used by authors of SRs to assess the
11 methodological quality of primary studies of TTD.

12

13 **2. Methods**

14 For the present SR, we defined methodological quality assessment as the assessment of the
15 risk of bias and threats to transportability according to the definition used by the authors of
16 the revised tool for the Quality Assessment of Diagnostic Accuracy Studies (QUADAS II) but
17 we acknowledge that the scope of methodological quality is larger than this definition
18 including ethical concerns, statistical issues and even quality of reporting as defined in the
19 Cochrane Handbook for Systematic Reviews of Interventions.[10,14] We conducted a
20 methodological SR following the methodology proposed by the Centre for Reviews and
21 Dissemination and reported it using PRISMA guidelines.[6,15] This SR was not registered in
22 PROSPERO (international prospective register of systematic reviews) because it does not
23 cover SRs of exclusively methodological issues.[16]

24

25

1 *2.1 Search strategy and selection criteria*

2 We systematically searched MEDLINE via PubMed and Web of Science for protocols for in-
3 progress SRs or articles for completed SRs, with or without meta-analyses, of studies of TTD
4 published in English or French up to the end of February 2015. The search strategy was
5 developed from that used in previously published SRs of studies of TTD,[1,2] including terms
6 referring to time (delay, interval, time, early, earlier, timely, timeliness, late), healthcare
7 accessibility (seeking, barrier), diagnosis (diagnosis, diagnoses, diagnostic), SR (systematic
8 review, meta-analysis) and language (English, French). The search strategies were adapted to
9 each database (Appendix B). We also hand-searched reference lists of included SRs and used
10 Google Scholar to search for reports that cited included SRs. We excluded SRs evaluating the
11 impact of an intervention to reduce TTD because a specific tool exists to evaluate risk of bias
12 for this type of study.[9] SRs including only qualitative primary studies were also excluded.

13

14 *2.2 Study selection*

15 Two authors (MM and EL) independently screened the titles and abstracts of retrieved reports
16 and the full text in case of uncertainty. In case of disagreement, a third author (MC) was
17 consulted.

18

19 *2.3 Data extraction*

20 For each included SR, two authors independently extracted: title of the article, authors, year of
21 publication and impact factor of the journal,[17] type of SR (with or without meta-analysis),
22 disease studied, number of primary studies included, qualification of the authors based on
23 affiliation (clinician, or “methodologist” [biostatistician, epidemiologist or public health
24 specialist]), objective of the SR (study length, determinants and/or consequences of TTD),
25 citation of a reporting guideline (PRISMA, MOOSE or QUOROM)[6–8] as well as reporting

1 of “quality assessment” in the Methods section, use of the results of the “quality
2 assessment” in the Results or Discussion sections, and the method used for this quality
3 assessment. In case of disagreement concerning data extraction, a third author (MC) was
4 consulted.

5

6 *2.4 Analyses*

7 We described the general characteristics of the SRs. We calculated the frequency of “quality
8 assessment” (i.e., any mention of a quality assessment process without predefining what this
9 quality assessment should be) and described its purpose and process as declared by the
10 authors. We characterized the “originality” of the method used for “quality assessment”:
11 published tool, homemade tool, or tool customised from that proposed in another SR or from
12 a published checklist or score. We analysed whether the tool was specifically customized for
13 TTD studies and whether the steps of elaboration of the homemade or customized tools were
14 explained and justified. We described whether the results of the “quality assessment” were
15 reported, if they were aggregated in a score, and how authors used them (e.g., to exclude poor
16 quality studies, perform subgroup or sensitivity analyses or balance the interpretation of
17 results).

18 For SR in which a “quality assessment” tool was used, we listed items reported by
19 authors to “assess the quality” of primary studies. These items were classified by 2 authors
20 (EL and MC) in 5 non-exclusive categories according to what they assessed: risk of bias
21 (selection, classification, confusion, and/or attrition), threats to transportability (i.e., transfer
22 to clinical practice, because of disease definition or spectrum or characteristics of the
23 population), statistical issues (power or missing data), quality of reporting, or unclear. This
24 classification was based on the preliminary analyses of 20 items by the risk of bias described
25 in the Research Triangle Institute (RTI) bank of items of bias for observational studies and in

1 an evaluation of non-randomized intervention studies published by Deeks et al.[11,18]. From
2 our experience,[1,2] we also distinguished items related to issues specific to TTD studies (i.e.,
3 items related to the issue of the undiagnosed cases, the definition or the assessment of the time
4 points for the TTD or the study of the association between the TTD and participant
5 characteristics or health outcomes).

6 We analysed the association between the use of at least one item evaluating risk of
7 bias or threats to transportability (versus no “quality assessment” or “quality assessment”
8 without any item evaluating risk of bias or threats to transportability) and the following
9 potential determinants: year of publication of the SR (before or after 2010: 1 year [arbitrary]
10 after the publication of PRISMA), the presence of a methodologist among the co-authors, and
11 impact factor of the journal.

12

13 *2.5 Statistical analysis*

14 We used chi-square or Fisher’s exact test to compare qualitative variables and Mann-Whitney
15 U-test for quantitative variables. Analyses involved use of Stata 11 (StataCorp, College
16 Station, TX, USA).

17

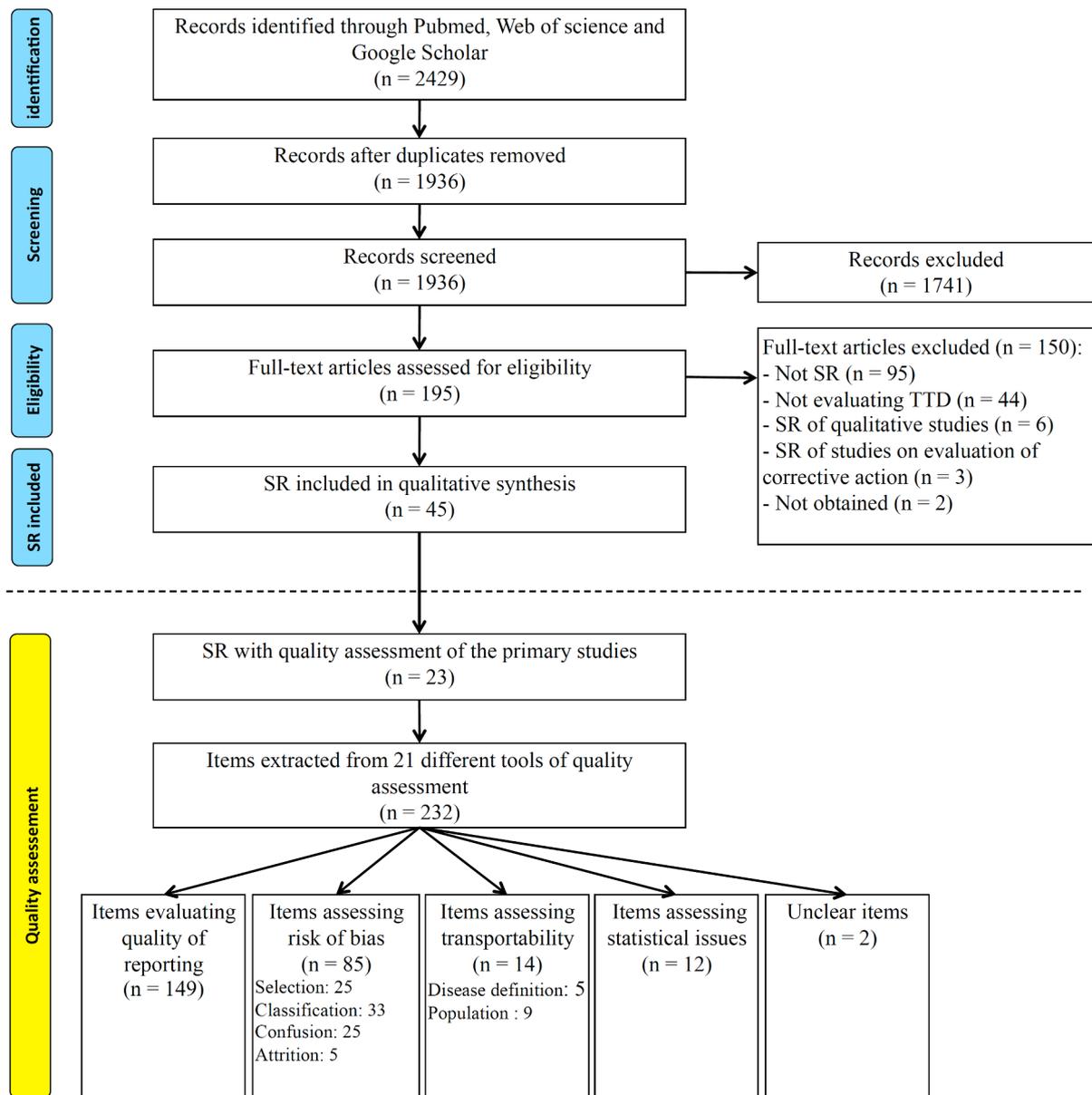
18 **3. Results**

19 *3.1 Systematic reviews*

20 We identified 1,935 articles (after removing duplicates) by electronic searches. One article
21 was identified by searching Google Scholar. We finally included 45 articles (Figure 1);[2,19–
22 62] 8 reported results of an SR with meta-analysis, and one a protocol for an SR. The median
23 impact factor for the publishing journals was 2.5 (range 0.65–39.21). A methodologist was a
24 co-author for 33 articles (73%). Cancer and tuberculosis were the most frequently studied
25 diseases, for 27 and 9 SRs, respectively (Table 1). The non-exclusive objectives of the SR

1 were to summarize the available information on length (n = 16, 36%), determinants (n = 31,
 2 69%), and/or consequences of TTD (n = 15, 33%) (Table 1). Nine SRs (20%) declared
 3 following a reporting guideline for SRs or meta-analyses (Table 1).

4 **Figure 1.** Flowchart of systematic reviews (SRs) of primary studies of time to diagnosis
 5 (TTD) in the study (above the dotted line) and ‘‘quality assessment’’ process with items
 6 classification (below the dotted line).



7
8

Table 1. Description of the 45 included systematic reviews (SRs) on time to diagnosis and the tools used for “quality assessment”

Authors	Year	Disease	Objective	No. of primary studies	Reporting guideline cited	“Quality assessment”					
						Performed	Description	Risk of bias and/or threats to transportability	At least one item specific to TTD studies	Results reported	Use
Total	-	-	-	-	20% (9/45)	51% (23/45)	-	36% (16/45)	42% (9/19)	78% (17/22)	82% (18/22)
Abel § [19]	2008	Chronic hematologic malignancies	D, C	56	No	No	-	-	-	-	-
Alhurishi*** § [20]	2011	Breast cancer	D	6	No	Yes	Reference to a previously published study and to STROBE	No	NA	No	Exclusion of quality studies
Bradford [21]	2009	Dementia	M, D	40	No	No	-	-	-	-	-
Brasme § [2]	2012	Childhood cancer	M, D, C	98	PRISMA	Yes	Isolated items: adjustment on confounding factors	Yes	No	No	Interpretat
Chatwin [22]	2013	Lung cancer	D	58	No	No	-	-	-	-	-
Chen [23]	2011	Tuberculosis	M, D	14	PRISMA	No	-	-	-	-	-
Chen*[24]	2012	HIV	C	25	PRISMA	No	-	-	-	-	-
Dang Tan §[25]	2007	Childhood cancer	M, D, C	23	No	No	-	-	-	-	-
Dubayova § # [26]	2010	Cancer and myocardial infarction	D	15	No	Yes	List of items adapted from checklists used in previously published studies	Yes	Yes	Yes	Interpretat
Evenson § [27]	2009	Stroke	M	123	No	No	-	-	-	-	-
Evenson §[28]	2001	Stroke	M	48	No	No	-	-	-	-	-
Facione* [29]	1993	Breast cancer	M, D	12	No	No	-	-	-	-	-
Finnie* § [30]	2011	Tuberculosis	D	20	No	Yes	Isolated items: type of recruitment and adjustment on confounding factors	Yes	No	No	Interpretat
Forrest***§ [31]	2014	Lung cancer	D	-	No	Yes	List of items adapted from a previously published study and from Aarhus statement	Yes	Yes	-	-
Fowler [32]	2014	Breast cancer	D	17	PRISMA	No	-	-	-	-	-
Gomez* [33]	2009	Oral cancer	C	9	MOOSE	Yes	List of items adapted from MOOSE	Yes	No	Yes	Sensitivity an
Goy § [34]	2009	Head and neck cancer	C	27	No	No	-	-	-	-	-
Jayde [35]	2009	Ovarian cancer	D	56	No	No	-	-	-	-	-
Jensen § [36]	2002	Lung cancer	C	16	No	No	-	-	-	-	-
Jones § [37]	2014	Breast cancer	D	18	PRISMA	Yes	Adapted from a previously published checklist	Yes	Yes	Yes	Interpretat
Lefler*** [38]	2004	Myocardial infarction	D	48	No	Yes	Reference to Cooper’s criteria	-	-	Partially	Interpretat
Lefler*** [39]	2002	Myocardial infarction	D	6	No	Yes	Reference to previous publication	-	-	Yes	Interpretat
Lethaby §# [40]	2013	Childhood and young adult cancer	M	32	-	Yes	Lists of items used in a previously published study (See MacDonald) and Aarhus Statement	Yes	Yes	Partially	Interpretat
Li*§,# [41]	2013	Tuberculosis	D	29	PRISMA	Yes	Newcastle-Ottawa scale for case-control and cohort studies and Guideline for critical appraisal of cross-sectional studies	Yes	No	Yes	Not report

Authors	Year	Disease	Objective	No. of primary studies	Reporting guideline cited	"Quality assessment"					
						Performed	Description	Risk of bias and/or threats to transportability	At least one item specific to TTD studies	Results reported	Use
Macdonald § # [42]	2006	Upper gastrointestinal cancer	D	25	No	Yes	List of items specifically elaborated according to the design of the primary study	No	No	Yes	Exclusion of quality stuc
Martins § [43]	2013	Cancer	D	7	No	Yes	List of items adapted from Critical appraisal skills programme checklist for cohort studies	Yes	Yes	Yes	Interpretat
Mitchell § # [44]	2008	Colorectal cancer	D	54	No	Yes	List of items from a previously published study (See MacDonald)	No	No	Yes	Exclusion of quality stuc
Ngangro [45]	2012	Tuberculosis	D	73	No	No	-	-	-	-	-
Olsson § [46]	2009	Lung cancer	M, C	49	No	Yes	List of items specifically elaborated	Yes	No	Yes	Exclusion of quality stuc
Ramirez § [47]	1999	Breast cancer	D	23	No	Yes	List of items specifically elaborated	Yes	Yes	No	Exclusion of quality stuc
Ramos* § [48]	2008	Colorectal cancer	C	37	No	Yes	List of items adapted from previously published checklist	Yes	Yes	Yes	Interpretat
Ramos*§ [49]	2007	Colorectal cancer	C	26	MOOSE	Yes	List of items adapted from previously published checklist	Yes	Yes	Yes	Exclusion of quality stuc
Richards § [50]	1999	Breast cancer	C	87	No	Yes	List of items specifically elaborated	Yes	No	No	Sensitivity an
Scott [51]	2006	Oral cancer	D	8	No	No	-	-	-	-	-
Seoane* § # [52]	2012	Head and neck cancer	C	10	MOOSE	Yes	List of items adapted from Newcastle-Ottawa scale	Yes	Yes	Yes	Exclusion of quality stuc
Sreeramareddy § [53]	2014	Tuberculosis	M, D	23	No	Yes	List of items specifically elaborated	Yes	No	Yes	Not report
Sreeramareddy [54]	2009	Tuberculosis	M, D	52	No	Yes	Not reported	-	-	No	Not report
Storla § [55]	2008	Tuberculosis	M, D	58	No	No	-	-	-	-	-
Stoykoya § [56]	1999	Oral cancer	D	22	No	No	-	-	-	-	-
Teuschl [57]	2010	Stroke	D	182	No	No	-	-	-	-	-
Thompson § [58]	2009	Bowel cancer	M, D, C	80	No	No	-	-	-	-	-
Unger §[59]	2009	Breast cancer	M, D, C	96	No	No	-	-	-	-	-
Villeneuve [60]	2013	Inflammatory arthritis	M	47	No	No	-	-	-	-	-
Vinas [61]	2014	Lung cancer	M, C	13	No	No	-	-	-	-	-
Yang § [62]	2014	Tuberculosis	D	137	No	Yes	STROBE and CONSORT	-	No	Yes	Not report

M: measurement, D: determinants, C: consequences

§: SR with a methodologist among the co-authors

* SR with meta-analysis, ** protocol, *** SR with a quality assessment reported in Methods but not detailed

#: SR aggregating results of quality assessment as a global score

In Lethaby, Mitchell and Mac Donald the same 6 tools were used according to the design of the primary study

1 3.2 “Quality assessment” purpose and process

2 In total, 23 SRs (51%) reported an “assessment of the quality” of primary studies. In 4,
3 “quality assessment” was mentioned in the Methods section but not described in the
4 article [20,54] or the cited reference.[38,39] “Quality assessment” was defined by
5 authors in the Methods section in 11 SRs (48%), and a purpose of “risk of bias
6 assessment” was explicit in 4 (see Appendix C). The authors of 2 SRs highlighted the
7 lack of specific guidelines to “assess quality” of studies of TTD.[2,53] Among the 21
8 described tools used to “assess quality”, 6 (23%) were published checklists or scores
9 (Strengthening the Reporting of Observational Studies in Epidemiology [STROBE],
10 CONSolidated Standards Of Reporting Trials [CONSORT], Aarhus Statement,
11 guideline for critical appraisal of cross-sectional studies, Newcastle-Ottawa scales for
12 case-control and cohort studies).[12,13,63–65] A total of 12 tools (46%) were
13 homemade, and 8 (31%) were customised from a tool proposed in a published SR (n =
14 1) or from a published score or checklist (n = 5) or both (n = 2) (Table 1). In 4 SRs,
15 different tools were used for “quality assessment” according to the design of the
16 primary study (cohort, case-control, cross sectional or case series).[40–42,44] In 6 SRs
17 (26% of those with “quality assessment”), the authors aimed to aggregate the results of
18 the “quality assessment” in a global score.[26,40–42,44,52] None of the authors who
19 described a homemade tool or modified an existing tool explained how they selected or
20 modified the items for their tool or provided a justification for the psychometric quality
21 the score used. The authors of 7 SRs reported the number of people who “assessed the
22 quality” of primary studies and how disagreement was resolved.[26,33,37,43,46,47,52]

23 The results of “quality assessment” were clearly reported in 16 SRs (72% of
24 those with “quality assessment”, protocol excluded), partially in 2 (9%) and not
25 reported in 5 (23%) (Table 1). “Quality assessment” was used to exclude poor-quality

1 studies in 6 SRs (26%), balance the interpretation of results and/or generate a level of
2 evidence in 9 (39%), or conduct a sensitivity analysis in 2 (9%); 5 SRs (22%) provided
3 no clear statement on how “quality assessment” was used.

4

5 *3.3 Items used for quality assessment*

6 In the 23 SRs reporting a “quality assessment” process, 232 items could be identified
7 from the 21 different tools used (Fig. 1; Appendix D); 149 (64%) assessed quality of
8 reporting, 85 (37%) risk of bias, 14 (6%) threats to transportability and 12 (5%)
9 statistical issues, and 1 (0.5%) was unclear. A total of 43 items were classified in 2 or
10 more categories (Appendix D). Among the 85 items assessing risk of bias, 25 were
11 related to selection, 33 classification (2 in the same SR related to blinding),[41] 25
12 confusion and 5 attrition; 4 items assessed both selection and classification biases.
13 Overall, 16 SRs (36% of all SRs) reported assessment of at least one item on risk of bias
14 and/or threats to transportability (Table 1). A total of 26 items (11%) were specifically
15 related to TTD: 6 (2%) assessed risk of classification bias during the process of
16 definition or assessment of time points and/or TTD, 24 (8%) quality of reporting and 4
17 both (Table 2).

Table 2. Items related to time to diagnosis (n=26) among the 232 used for “quality assessment”

First author name and SR [reference]	Item	Category
Dubayova [26]	<i>I. Definition and operationalization of patient’s delay. Patient’s delay was defined exactly using a disease specific cut off point which divided the sample into delayers and non-delayers (+), or patient’s delay was defined as a continuous variable (-)</i>	Reporting
Forrest [31] Lethaby [40] (Aarhus statement [65])	<i>Are the beginning and end points of this interval clearly defined?</i>	Reporting
	<i>Are there precise, transparent and repeatable definitions, and is the complexity of time points such as the date of first symptom and date of first presentation addressed?</i>	Classification bias and reporting
	<i>Do the researchers refer to a theoretical framework underpinning definition of this time point?</i>	Reporting
	<i>Is there a discussion of the different biases influencing measurement of this time point?</i>	Reporting
	<i>Do the researchers discuss the complexity of the date of first presentation?</i>	Reporting
	<i>Do the researchers discuss the nature of the referral and provide adequate detail – for example, whether it was for investigation or consultation by a colleague in secondary care?</i>	Reporting
	<i>Do the researchers use an existing hierarchical rationale for the date of diagnosis measurement?</i>	Reporting
	<i>Do the questions on time points and/or intervals clearly derive from stated definitions?</i>	Classification bias
	<i>Do researchers acknowledge the need for theoretical validation and make reference to the theoretical framework(s) underpinning measurement and analysis of the time points?</i>	Reporting
	<i>Is there discussion of the different biases influencing measurement of the time points, such as how and when the question is asked and who is being asked?</i>	Reporting
	<i>Is the timing of the interview in relation to the date of diagnosis provided?</i>	Classification bias
<i>Is there a clear and precise description of how case-note data were used to ascertain time points with an acknowledgment of limitations of such data?</i>	Reporting	
Martins [43]	<i>Was diagnostic intervals measured to Olsen and colleagues framework?</i>	Classification bias and reporting

First author name and SR [reference]	Item	Category
Ramirez [47]	<i>Studies were eligible for inclusion if they were based on a clear description of a discrete interval of delay—by patients or providers. We defined delay by patients as the time between the onset of initial symptoms and the first medical consultation. Delay by providers was any discrete interval between first medical consultation and treatment. We excluded studies if they were based on unclear definitions or looked at periods of delay by patients and providers that were indivisibly combined</i>	Reporting
Ramos 2007 [49]	<i>Interval of time definition 1. From beginning of symptoms to diagnosis; 2. From beginning of symptoms to treatment; 9. Not mentioned</i>	Reporting
	<i>Manner of measuring the interval of time 1. Global; 2. Only a fragment (patient's delay or health system's delay)</i>	Reporting
	<i>Measurements of the effects of the interval of time 1. Average; 2. Median; 3. <1 months; 4. <2 months; 5. <3 months; 6. <4 months; 7. <5 months; 8. <6 months; 9. Quartiles</i>	Reporting
Ramos 2008 [48]	<i>Definition of delay 1. From the onset of symptoms to diagnosis; 2. From the onset of symptoms to treatment</i>	Reporting
	<i>Measures of the time interval 1. Overall; 2. Only a portion (delay due to patient, delay due to health system)</i>	Reporting
	<i>Measures of the delay 1. Means; 2. Medians; 3. <1 month; 4. <2 months; 5. <3 months; 6. <4 months; 7. <5 months; 8. <6 months; 9. Quartiles</i>	Reporting
	<i>Measurement of the delay by disease stage 1. Comparison of means or medians; 2. Frequencies; 3. Absolute figures; 4. Correlations; 5. Graphics; 9. Not given</i>	Reporting
Seoane [52]	<i>The first item assessed the source of the outcome date. One point was given if the date of death was ascertained through clinical history or death certificate and zero point if otherwise... (when the information on a specific item was not provided by the authors, we graded this item as 'no').</i>	Classification bias and reporting
	<i>The fourth item assessed the definition of delay. One point was given if inception was clearly defined and zero if not... (when the information on a specific item was not provided by the authors, we graded this item as 'no').</i>	Classification bias and reporting
Sreeramareddy [53]	<i>Diagnostic delays separately reported as patient and health systems (y/n)</i>	Reporting
	<i>Risk factors for diagnostic delays were reported (y/n)</i>	Reporting

1 *3.4 Determinants of methodological quality assessment*

2 Assessment of at least one item on risk of bias and/or threats to transportability was
 3 significantly associated with the presence of a methodologist among co-authors ($p = 0.03$) and
 4 a high journal impact factor ($p = 0.03$) but not a more recent period of publication (Table 3).

5 **Table 3.** *Determinants of assessment of risk of bias or threat to transportability in systematic*
 6 *reviews of time to diagnosis (n=45)*
 7

	Presence of at least one item assessing risk of bias or threat to transportability (n=16)	No assessment of risk of bias or threat to transportability (n=29)	P value
Period of publication, n (%)			
Before 2010*	6 (24)	19 (76)	0.07
After 2010	10 (50)	10 (50)	
At least one methodologist** among the co-authors, n (%)			
No	1 (9)	10 (91)	0.03
Yes	15 (45)	18 (55)	
Journal impact factor, median [IQR]	3.9 [2.3-8.6]	2.3 [1.8-3.7]	0.03

8 *one year after the publication of PRISMA [6]

9 **biostatistician, epidemiologist or public health specialist (presence of a methodologist was not ascertainable
 10 for 1 systematic review)

11 IQR: interquartile range

12

13 **4. Discussion**

14 *4.1 Main results*

15 Our methodological SR revealed that more than a half (64%) of SRs of TTD did not report an
 16 assessment of risk of bias or threats to transportability of included primary studies. The
 17 authors of the included SRs used 21 different tools that were often homemade (46%) with no
 18 justification of the items used. Items used for “quality assessment” assessed quality of
 19 reporting more frequently than risk of bias or threats to transportability (64% vs 43%).
 20 Authors did not report how results of the “quality assessment” were used in 23% of the SRs
 21 with “quality assessment”. Scoring systems were used in 26% of SRs, but psychometrical
 22 qualities of scoring systems were never reported. These results strongly contrast with
 23 PRISMA international recommendations to specify and present the results of any assessment
 24 of risk of bias that may affect the cumulative evidence.[6] No item explored previously

1 identified key issues in the selection process between the source population and participants,
2 such as risk of undiagnosed cases,[1] and only one SR explored risk of bias due to the lack of
3 blinding during the assessment of time points.[41]

4

5 *4.2 Strengths and limitations*

6 We followed the methodology proposed by the Centre for Reviews and Dissemination for
7 SRs even if the present SR did not aim to evaluate health interventions.[15] The search
8 strategy was based on previously published strategies with proven sensitivity in the field of
9 TTD research.[1,2] A potential bias may be related to the restriction to SRs in English or
10 French.[66] Two authors classified items used to assess the “quality” of primary studies with
11 comparison to those already proposed for observational studies.[11,18] Classification of items
12 was sometimes difficult because the line between reporting risk of bias and threats to
13 transportability may be thin, as was mentioned in the development of the QUADAS-II
14 tool.[10] For example, assessing the selection process of participants in primary studies and
15 its consequences for target populations may be related to risk of selection bias and threats to
16 transportability. Classification of items in the “quality of reporting” category was less
17 subjective because authors of SRs commonly used dedicated terms such as “stated”,
18 “reported” or “described”.

19

20 *4.3 Need for clarification and tools*

21 Our results agree with those of a large meta-review of observational studies finding that
22 “quality assessment” of the primary study was reported in only 33% of SRs included.[18]
23 Lack or inadequate assessment of risk of bias and threats to transportability may lead to
24 inadequate conclusions in SRs. The first reason for this lack or inadequate assessment of risk
25 of bias and threats to transportability could be related to the lack of awareness of the

1 importance of such quality assessment among authors, reviewers and editors. Indeed, we
2 found assessment of risk of bias and threats to transportability more frequent when authors
3 of SRs included a methodologist or for SRs published in a journal with a high impact factor.
4 A second reason could be the lack of an adequate tool to assess risk of bias and threats to
5 transportability in studies of TTD. This second hypothesis is supported by our finding that
6 several authors of SRs used reporting checklists and risk of bias tools designed for other types
7 of research because they were unable to identify a tool for assessing methodological quality of
8 TTD studies. As well, only 2% of the items for the identified tools of quality assessment were
9 related to risk of bias specific to TTD and none of those items aimed to assess the previously
10 identified key issue in the selection process because of undiagnosed cases.

11 Because TTD studies are becoming popular and their results may have major clinical
12 and public health impacts, adequate tools are needed to evaluate their methodological quality.
13 Two reporting guidelines have been proposed for studies of TTD.[1,65] Complete reporting is
14 a prerequisite to the assessment of methodological quality but does not protect against bias.
15 Available tools to evaluate the risk of bias and threats to transportability in observational
16 studies are designed for diagnostic-test accuracy studies or are generic and evaluate studies of
17 interventions or exposure.[10–13] Among observational studies, those investigating TTD
18 have strong specificities, such as participant selection from the source population and TTD
19 measurement.[1,2] These specificities were detailed previously in an explanatory and
20 elaboration document for a reporting guideline [1] and are highlighted by the differences
21 between reporting guidelines developed specifically for TTD studies and STROBE.[1,63,65]
22 Therefore, a specific risk of bias tool for TTD studies may be needed. Building a specific tool
23 to assess risk of bias and threat to transportability for studies of TTD may help avoid the
24 waste of research and inappropriate clinical and public health decisions based on poor-quality
25 studies.[67]

5. References

1. Launay E, Morfouace M, Deneux-Tharoux C, Gras le-Guen C, Ravaud P, Chalumeau M. Quality of reporting of studies evaluating time to diagnosis: a systematic review in paediatrics. *Arch Dis Child*. 2014;99: 244–250. doi:10.1136/archdischild-2013-304778
2. Brasme J-F, Morfouace M, Grill J, Martinot A, Amalberti R, Bons-Letouzey C, et al. Delays in diagnosis of paediatric cancers: a systematic review and comparison with expert testimony in lawsuits. *Lancet Oncol*. 2012;13: e445–459. doi:10.1016/S1470-2045(12)70361-3
3. Wallace E, Lowry J, Smith SM, Fahey T. The epidemiology of malpractice claims in primary care: a systematic review. *BMJ Open*. 2013;3. doi:10.1136/bmjopen-2013-002929
4. Najaf-Zadeh A, Dubos F, Pruvost I, Bons-Letouzey C, Amalberti R, Martinot A. Epidemiology and aetiology of paediatric malpractice claims in France. *Arch Dis Child*. 2011;96: 127–130. doi:10.1136/adc.2010.189209
5. National Library of Medicine. Early diagnosis. Medical Subject Headings. National Library of Medicine, Bethesda, USA [Internet]. [cited 28 Jun 2015]. Available: <http://www.ncbi.nlm.nih.gov/mesh/?term=early+diagnosis>
6. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ*. 2009;339: b2535.
7. 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. *Lancet*. 1999;354: 1896–1900. doi:10.1016/S0140-6736(99)04149-5
8. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *JAMA*. 2000;283: 2008–2012.
9. The Cochrane Collaboration. Applicability and transferability. In: Julian PT Higgins, Douglas G Altman, Sterne JA, editors. *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0. [Internet]. [cited 8 Jul 2015]. Available: <http://www.cochrane-handbook.org/>

10. Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med.* 2011;155: 529–536. doi:10.7326/0003-4819-155-8-201110180-00009
11. Viswanathan M, Berkman ND, Dryden DM, Hartling L. Assessing Risk of Bias and Confounding in Observational Studies of Interventions or Exposures: Further Development of the RTI Item Bank [Internet]. Rockville (MD): Agency for Healthcare Research and Quality (US); 2013. Available: <http://www.ncbi.nlm.nih.gov/books/NBK154461/>
12. Wells G, Shea B, O’Connell D, Peterson J, Welch V, Losos M, et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. [Internet]. [cited 13 Jun 2015]. Available: http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp
13. Critical Appraisal Skill Program. Critical appraisal Skill Program Checklists. [Internet]. [cited 13 Jun 2015]. Available: www.casp-uk.net/#!/checklists/cb36
14. The Cochrane Collaboration. Assessing risk of bias in included studies. In: Julian PT Higgins, Douglas G Altman, Sterne JA, editors. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0.* [Internet]. [cited 8 Jul 2015]. Available: <http://www.cochrane-handbook.org/>
15. Centre for Reviews and Dissemination. *Systematic Reviews: CRD’s guidance for undertaking systematic reviews in health care* [Internet]. [cited 17 Dec 2014]. Available: http://www.york.ac.uk/inst/crd/pdf/Systematic_Reviews.pdf
16. University of York. Centre for Reviews and Dissemination. PROSPERO. Inclusion criteria [Internet]. [cited 4 Aug 2015]. Available: <http://www.crd.york.ac.uk/PROSPERO>
17. Journal Citation Report. [Internet]. [cited 4 Jun 2015]. Available: <http://about.jcr.incites.thomsonreuters.com>
18. Deeks JJ, Dinnes J, Amico R D’, Sowden AJ, Sakarovitch C, Song F, et al. Evaluating non-randomised intervention studies. *Health Technol Assess Winch Engl.* 2003;7: iii–x, 1–173.
19. Abel GA, Friese CR, Magazu LS, Richardson LC, Fernandez ME, De Zengotita JJ, et al. Delays in referral and diagnosis for chronic hematologic malignancies: A literature review. *Leuk Lymphoma.* 2008;49: 1352–1359. doi:10.1080/10428190802124281

20. Alhurishi S, Lim JNW, Potrata B, West R. Factors influencing late presentation for breast cancer in the middle East: a systematic review. *Asian Pac J Cancer Prev APJCP*. 2011;12: 1597–1600.
21. Bradford A, Kunik ME, Schulz P, Williams SP, Singh H. Missed and Delayed Diagnosis of Dementia in Primary Care: Prevalence and Contributing Factors. *Alzheimer Dis Assoc Disord*. 2009;23: 306–314. doi:10.1097/WAD.0b013e3181a6bebc
22. Chatwin J, Sanders C. The influence of social factors on help-seeking for people with lung cancer: Help-seeking in lung cancer. *Eur J Cancer Care (Engl)*. 2013;22: 709–713. doi:10.1111/ecc.12078
23. Chen T-C, Lu P-L, Lin C-Y, Lin W-R, Chen Y-H. Fluoroquinolones are associated with delayed treatment and resistance in tuberculosis: a systematic review and meta-analysis. *Int J Infect Dis*. 2011;15: e211–e216. doi:10.1016/j.ijid.2010.11.008
24. Chen NE, Gallant JE, Page KR. A Systematic Review of HIV/AIDS Survival and Delayed Diagnosis Among Hispanics in the United States. *J Immigr Minor Health*. 2012;14: 65–81. doi:10.1007/s10903-011-9497-y
25. Dang-Tan T, Franco EL. Diagnosis delays in childhood cancer: A review. *Cancer*. 2007;110: 703–713. doi:10.1002/cncr.22849
26. Dubayova T, van Dijk JP, Nagyova I, Rosenberger J, Havlikova E, Gdovinova Z, et al. The impact of the intensity of fear on patient's delay regarding health care seeking behavior: a systematic review. *Int J Public Health*. 2010;55: 459–468. doi:10.1007/s00038-010-0149-0
27. Evenson KR, Foraker RE, Morris DL, Rosamond WD. A comprehensive review of prehospital and in-hospital delay times in acute stroke care. *Int J Stroke*. 2009;4: 187–199. doi:10.1111/j.1747-4949.2009.00276.x
28. Evenson KR, Rosamond WD, Morris DL. Prehospital and in-hospital delays in acute stroke care. *Neuroepidemiology*. 2001;20: 65–76. doi:54763
29. Facione NC. Delay versus help seeking for breast cancer symptoms: a critical review of the literature on patient and provider delay. *Soc Sci Med* 1982. 1993;36: 1521–1534.
30. Finnie RKC, Khoza LB, van den Borne B, Mabunda T, Abotchie P, Mullen PD. Factors associated with patient and health care system delay in diagnosis and treatment for TB in sub-Saharan African countries with high burdens of TB and HIV: Patient and system delay in diagnosis and treatment for TB in sub-Saharan Africa. *Trop Med Int Health*. 2011;16: 394–411. doi:10.1111/j.1365-3156.2010.02718.x

31. Forrest LF, Sowden S, Rubin G, White M, Adams J. Socio-economic inequalities in patient, primary care, referral, diagnostic, and treatment intervals on the lung cancer care pathway: protocol for a systematic review and meta-analysis. *Syst Rev.* 2014;3: 30. doi:10.1186/2046-4053-3-30
32. A Fowler B. Neighborhood-level Influences on Delays in Diagnostic Follow-up from Mammography Screening in African-American Women: A Systematic Review. *J Womens Health Care.* 2014;03. doi:10.4172/2167-0420.1000151
33. Gómez I, Seoane J, Varela-Centelles P, Diz P, Takkouche B. Is diagnostic delay related to advanced-stage oral cancer? A meta-analysis. *Eur J Oral Sci.* 2009;117: 541–546. doi:10.1111/j.1600-0722.2009.00672.x
34. Goy J, Hall SF, Feldman-Stewart D, Groome PA. Diagnostic delay and disease stage in head and neck cancer: A systematic review. *Laryngoscope.* 2009;119: 889–898. doi:10.1002/lary.20185
35. Jayde V, White K, Blomfield P. Symptoms and diagnostic delay in ovarian cancer: a summary of the literature. *Contemp Nurse.* 2009;34: 55–65.
36. Jensen AR, Mainz J, Overgaard J. Impact of delay on diagnosis and treatment of primary lung cancer. *Acta Oncol Stockh Swed.* 2002;41: 147–152.
37. Jones CE, Maben J, Jack RH, Davies EA, Forbes LJ, Lucas G, et al. A systematic review of barriers to early presentation and diagnosis with breast cancer among black women. *BMJ Open.* 2014;4: e004076–e004076. doi:10.1136/bmjopen-2013-004076
38. Lefler LL, Bondy KN. Women's delay in seeking treatment with myocardial infarction: a meta-synthesis. *J Cardiovasc Nurs.* 2004;19: 251–268.
39. Lefler L. The advanced practice nurse's role regarding women's delay in seeking treatment with myocardial infarction. *J Am Acad Nurse Pract.* 2002;14: 449–456.
40. Lethaby CD, Picton S, Kinsey SE, Phillips R, van Laar M, Feltbower RG. A systematic review of time to diagnosis in children and young adults with cancer. *Arch Dis Child.* 2013;98: 349–355. doi:10.1136/archdischild-2012-303034
41. Li Y, Ehiri J, Tang S, Li D, Bian Y, Lin H, et al. Factors associated with patient, and diagnostic delays in Chinese TB patients: a systematic review and meta-analysis. *BMC Med.* 2013;11: 156. doi:10.1186/1741-7015-11-156
42. Macdonald S, Macleod U, Campbell NC, Weller D, Mitchell E. Systematic review of factors influencing patient and practitioner delay in diagnosis of upper gastrointestinal cancer. *Br J Cancer.* 2006;94: 1272–1280. doi:10.1038/sj.bjc.6603089

43. Martins T, Hamilton W, Ukoumunne OC. Ethnic inequalities in time to diagnosis of cancer: a systematic review. *BMC Fam Pract*. 2013;14: 197. doi:10.1186/1471-2296-14-197
44. Mitchell E, Macdonald S, Campbell NC, Weller D, Macleod U. Influences on pre-hospital delay in the diagnosis of colorectal cancer: a systematic review. *Br J Cancer*. 2008;98: 60–70. doi:10.1038/sj.bjc.6604096
45. Ndeikoundam Ngangro N, Chauvin P, Halley des Fontaines V. [Determinants of tuberculosis diagnosis delay in limited resources countries]. *Rev Epidemiol Sante Publique*. 2012;60: 47–57. doi:10.1016/j.respe.2011.08.064
46. Olsson JK, Schultz EM, Gould MK. Timeliness of care in patients with lung cancer: a systematic review. *Thorax*. 2009;64: 749–756. doi:10.1136/thx.2008.109330
47. Ramirez AJ, Westcombe AM, Burgess CC, Sutton S, Littlejohns P, Richards MA. Factors predicting delayed presentation of symptomatic breast cancer: a systematic review. *Lancet*. 1999;353: 1127–1131.
48. Ramos M, Esteva M, Cabeza E, Llobera J, Ruiz A. Lack of association between diagnostic and therapeutic delay and stage of colorectal cancer. *Eur J Cancer*. 2008;44: 510–521. doi:10.1016/j.ejca.2008.01.011
49. Ramos M, Esteva M, Cabeza E, Campillo C, Llobera J, Aguiló A. Relationship of diagnostic and therapeutic delay with survival in colorectal cancer: A review. *Eur J Cancer*. 2007;43: 2467–2478. doi:10.1016/j.ejca.2007.08.023
50. Richards MA, Westcombe AM, Love SB, Littlejohns P, Ramirez AJ. Influence of delay on survival in patients with breast cancer: a systematic review. *Lancet*. 1999;353: 1119–1126.
51. Scott SE, Grunfeld EA, McGurk M. Patient's delay in oral cancer: A systematic review. *Community Dent Oral Epidemiol*. 2006;34: 337–343. doi:10.1111/j.1600-0528.2006.00290.x
52. Seoane J, Takkouche B, Varela-Centelles P, Tomás I, Seoane-Romero JM. Impact of delay in diagnosis on survival to head and neck carcinomas: a systematic review with meta-analysis: Diagnostic delay and head and neck cancer. *Clin Otolaryngol*. 2012;37: 99–106. doi:10.1111/j.1749-4486.2012.02464.x
53. Sreeramareddy CT, Qin ZZ, Satyanarayana S, Subbaraman R, Pai M. Delays in diagnosis and treatment of pulmonary tuberculosis in India: a systematic review. *Int J Tuberc Lung Dis Off J Int Union Tuberc Lung Dis*. 2014;18: 255–266. doi:10.5588/ijtld.13.0585

54. Sreeramareddy CT, Kishore PV, Menten J, Van den Ende J. Time delays in diagnosis of pulmonary tuberculosis: a systematic review of literature. *BMC Infect Dis.* 2009;9: 91. doi:10.1186/1471-2334-9-91
55. Storla DG, Yimer S, Bjune GA. A systematic review of delay in the diagnosis and treatment of tuberculosis. *BMC Public Health.* 2008;8: 15. doi:10.1186/1471-2458-8-15
56. Stoykova M. Delayed diagnosis of cancer with emphasis on oral cavity cancers. *Folia Med (Plovdiv).* 1999;41: 132–135.
57. Teuschl Y, Brainin M. Stroke education: discrepancies among factors influencing prehospital delay and stroke knowledge: Reviews. *Int J Stroke.* 2010;5: 187–208. doi:10.1111/j.1747-4949.2010.00428.x
58. Thompson MR, Heath I, Swarbrick ET, Faulds Wood L, Ellis BG. Earlier diagnosis and treatment of symptomatic bowel cancer: can it be achieved and how much will it improve survival?: Earlier diagnosis of symptomatic bowel cancer. *Colorectal Dis.* 2011;13: 6–16. doi:10.1111/j.1463-1318.2009.01986.x
59. Unger-Saldaña K, Infante-Castañeda C. Delay of medical care for symptomatic breast cancer: a literature review. *Salud Pública México.* 2009;51 Suppl 2: s270–285.
60. Villeneuve E, Nam JL, Bell MJ, Deighton CM, Felson DT, Hazes JM, et al. A systematic literature review of strategies promoting early referral and reducing delays in the diagnosis and management of inflammatory arthritis. *Ann Rheum Dis.* 2013;72: 13–22. doi:10.1136/annrheumdis-2011-201063
61. Vinas F, Ben Hassen I, Jabot L, Monnet I, Chouaid C. Delays for diagnosis and treatment of lung cancers: a systematic review. *Clin Respir J.* 2014; doi:10.1111/crj.12217
62. Yang W-T, Gounder CR, Akande T, De Neve J-W, McIntire KN, Chandrasekhar A, et al. Barriers and delays in tuberculosis diagnosis and treatment services: does gender matter? *Tuberc Res Treat.* 2014;2014: 461935. doi:10.1155/2014/461935
63. Elm E von, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *J Clin Epidemiol.* 2008;61: 344–349. doi:10.1016/j.jclinepi.2007.11.008
64. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. *Ann Intern Med.* 2010;152: 726–732. doi:10.7326/0003-4819-152-11-201006010-00232

65. Weller D, Vedsted P, Rubin G, Walter FM, Emery J, Scott S, et al. The Aarhus statement: improving design and reporting of studies on early cancer diagnosis. *Br J Cancer*. 2012;106: 1262–1267. doi:10.1038/bjc.2012.68
66. Juni P. Direction and impact of language bias in meta-analyses of controlled trials: empirical study. *Int J Epidemiol*. 2002;31: 115–123. doi:10.1093/ije/31.1.115
67. Glasziou P, Altman DG, Bossuyt P, Boutron I, Clarke M, Julious S, et al. Reducing waste from incomplete or unusable reports of biomedical research. *Lancet*. 2014;383: 267–276. doi:10.1016/S0140-6736(13)62228-X

Appendix A. Number of publications on time to diagnosis (TTD) in PubMed over time.

Year	A. No. of articles identified using the filter tailored for studies on TTD*	B. Total no. of articles referenced	A/B ratio
2000	624	529859	0.00118
2001	650	544388	0.00119
2002	651	561993	0.00116
2003	737	591926	0.00125
2004	781	636714	0.00123
2005	896	697179	0.00129
2006	1 017	743541	0.00137
2007	1 124	781806	0.00144
2008	1 217	830799	0.00146
2009	1 259	871055	0.00145
2010	1 407	934419	0.00151
2011	1 545	1009410	0.00153
2012	1 665	1072394	0.00155
2013	1 831	1131349	0.00162
2014	1 617	1150252	0.00141
2014/2000 increase	+ 159%	+ 117 %	+ 19%

Represents absolute and relative number of articles referenced in PubMed between 2000 and 2014 identified by a tailored filter* for studies of TTD [1].

articles reporting TTD studies were identified with the following search terms: barrier[title] OR delay[Title] OR interval*[Title] OR "time to diagnosis"[title] OR seeking[title] OR "early diagnosis"[title] OR "earlier diagnosis"[title] OR "timely diagnosis"[title] OR "timeliness"[title] OR "late diagnosis"[title]) AND diagnos*

Appendix B. Electronic search strategies

PUBMED

barrier[title] OR delay*[Title] OR interval*[Title] OR "time to diagnosis"[title] OR seeking[title] OR "early diagnosis"[title] OR "earlier diagnosis"[title] OR "timely diagnosis"[title] OR "timeliness"[title] OR "late diagnosis"[title]) AND diagnos* AND ((Meta-Analysis OR review [Publication Type]) OR meta-analysis as topic[MeSH] OR "systematic review"[title/abstract] AND (English[Language] OR French[Language])).

WEB OF SCIENCE

TI=(barrier OR delay* OR interval* OR "time to diagnosis" OR seeking OR early OR earlier OR timely) AND TS=diagnos* AND TI= ("systematic review" OR meta-analysis OR review)

The search was constructed by associating terms referring to time (delay, interval, time, early, earlier, timely, timeliness, late), healthcare accessibility (seeking, barrier), diagnosis (diagnosis, diagnoses, diagnostic), and SR (systematic review, meta-analysis).

GOOGLE SCHOLAR

The reference for each SR included through previous electronic searches was entered in the search box of Google Scholar to retrieve a citation that would not have been identified by electronic search.

Appendix C. Authors' definition of quality assessment according to the Methods section of the systematic review (SR).

SR First author name [reference]	Definition of quality assessment by authors of SR (extracts from Methods section)
Brasme [2]	<i>"In view of the absence of standards such as quality assessment of diagnostic accuracy studies (QUADAS) for assessment of the risk of bias in studies of diagnostic delay, we judged that the results of studies that included multivariate analyses to adjust for potential confounders in the relations between time to diagnosis and outcome were less exposed to bias than were those of crude relations."</i>
Jones[37]	<i>"Methodological quality was appraised using six attributes of research design that impact directly on the veracity of research findings irrespective of the particular research design"</i>
Lefler [38]	<i>"Cooper described evaluation of synthesis studies as judging the adequacy of each individual study by examining the validity of the study's methods.(...) Each of the retrieved studies was judged valid for inclusion in the review by Cooper's criteria."</i>
Lefler [39]	<i>"In the validation phase, Stetler (1994) states that an assessment of the studies is due in order to decide whether the findings of the studies are usable for utilization in practice."</i>
Lethaby [40]	<i>"The methodological quality of the papers was assessed using the scoring systems outlined in Macdonald et al and against the Aarhus checklist"</i>
Mac Donald [42]	<i>"Methodological quality of experimental studies (randomised controlled trial, clinical trials, controlled before and after studies) was assessed for sample formation, baseline differences between groups, unit of allocation to groups, measures of outcome and follow-up (Appendix A1). Owing to the lack of controlled trials in this field, a method of scoring descriptive studies (cross-sectional, cohort, case-control, before and after studies) was also used, allowing interpretation of useful papers that would be discarded by strict adherence to Cochrane standards. Each study was scored on six criteria relevant to its design (e.g. case-control was assessed for a clear research question, source of cases/controls, clear inclusion/exclusion criteria, sampling method and comparability of the groups) (Appendix B1). (...) Assessment of studies that had already been rated using the systems outlined above was based on those previous scores. Evidence was rated as 'strong' if a paper had an adequate sample size, used a rigorous methodology to ascertain data (i.e. not open to selection bias) and reported statistically significant differences in relation to the factors identified (or used appropriate analytic techniques if qualitative)"</i>
Martins [43]	<i>"Each retrieved article was independently appraised by two reviewers (TM and OU) and was classified as either "satisfactory", "medium" or "high" quality paper depending on the extent to which the checklist items were met, and also on the level of concerns raised about their methodology in general. For instance, a concern was raised where authors classified ethnic groups differently to the contemporary ethnic classifications in their population."</i>
Mitchell [44]	<i>"Previously developed scoring systems were extended and applied to assess the methodological adequacy of studies. Studies providing strong evidence were those with an adequate sample size, rigorous methods to ascertain data (i.e. not open to selection bias) and reporting statistically significant differences in relation to the delay-related factors identified (or using appropriate analytic techniques if qualitative)."</i>
Ramirez [47]	<i>"We developed criteria, including quality criteria, to assess whether data related to the different putative factors were adequate for inclusion in the review"</i>
Ramos [49]	<i>"For critical reading, we utilised criteria used to review non experimental studies and those used in other reviews of the same topic (...)"</i>
Sreeramareddy [54]	<i>"The selected studies were assessed for methodological quality by two of the authors (CTS and PVK)."</i>

Appendix D. Classification of the 232 items used in SRs to evaluate methodological quality.

Tools are identified by the first author of the SR when specifically elaborated or customised or by the name of the scale or checklist when published as such.

Tools	Total	Risk of bias				Transportability		Statistical issues	Reporting	Unclear	Specific to TTD
		Selection	Classification	Confusion	Attrition	Disease definition	Population				
Aarhus Statement [65] used by Lethaby [40] and Forrest [31]	19	0	5	0	1	0	0	0	14	0	12
Brasme [2]	1	0	0	1	0	0	0	0	0	0	0
CONSORT [64] used by Yang [62]	37	0	0	0	0	0	0	0	37	0	0
Dubayova [26]	5	0	1	0	0	0	0	3	2	0	1
Finnie [30]	2	1	0	1	0	0	0	0	0	0	0
Forrest [31]	21	1	2	5	0	0	2	2	12	0	0
Gomez [33]	3	0	0	2	0	0	0	1	0	0	0
Guideline for critical appraisal of cross-sectional studies [13]	6	2	2	1	0	0	0	1	0	1	0
Jones [37]	10	3	3	2	0	0	1	0	2	0	0
Martins [43]	6	2	2	1	1	0	1	0	1	0	1
Newcastle-Ottawa scale for case-control studies [12]	7	3	3	1	0	1	1	0	4	1	0
Newcastle-Ottawa scale for cohort studies [12]	8	4	3	0	0	0	0	1	5	0	0
Olsson* [46]	10	2	2	2	0	0	0	0	5	0	0
Ramirez [47]	5	2	1	1	0	0	0	1	2	0	1
Ramos 2007 [49]	21	1	1	5	2	1	2	2	9	0	3
Ramos 2008 [48]	14	1	1	1	0	2	2	1	5	0	4
Richards [50]	5	1	2	0	1	0	0	0	2	0	0
Scales for before–after time studies, case–control, case report, case series, cohort studies or cross sectional used by Lethaby, Mitchell and MacDonalds [40,42,44]	17	0	0	0	0	0	0	0	17	0	0
Seoane [52]	5	0	4	1	0	0	0	0	5	0	2
Sreeramareddy 2014* [53]	5	2	1	1	0	1	0	0	2	0	2
STROBE [63] used by Yang [62]	25	0	0	0	0	0	0	0	25	0	0
TOTAL	232	25	33	25	5	5	9	12	149	2	26

*tools for which one of the items assessed 2 types of risk of bias.

Annexe 3 : Article issu de la Partie II - Chapitre 1

Suboptimal care in the initial management of children who died from severe bacterial infection: A population-based confidential inquiry*

Elise Launay, MD; Christèle Gras-Le Guen, MD, PhD; Alain Martinot, MD, PhD; Rémy Assathiany, MD; Thomas Blanchais, MD; Nadjette Mourdi, MPH; Albertine Aouba, MD; Marie-Hélène Bouvier-Colle, PhD; Jean-Christophe Rozé, MD, PhD; Martin Chalumeau, MD, PhD

Objectives: To study the frequency and types of suboptimal care and medical errors in children who died of severe bacterial infection as the first-stage procedure intended to improve quality of care.

Design: Population-based confidential inquiry.

Setting: Two adjoining administrative districts in France.

Patients: Children older than 3 months dead from severe bacterial infection from 2000 through 2006.

Interventions: The medical files were summarized on standardized forms and then evaluated independently by two experts, who determined whether the initial management before the patients' arrival in intensive care was or was not optimal, in comparison with current guidelines.

Measurements and Main Results: Of 23 deaths from severe bacterial infection, 21 could be analyzed; management was considered suboptimal in 76%. The coefficient of agreement between

the experts was high, with a weighted κ of 0.73. The types of errors identified included parental delay in seeking medical care (33%; 95% confidence interval, [12–54]), physicians' delay in administering appropriate treatment (antibiotic therapy in the case of purpura; 38%; 95% confidence interval, 16–60), insufficient doses of or failure to repeat fluid resuscitation (24%; 95% confidence interval, [9–35]), and overall underestimation of disease severity (38%; 95% confidence interval, [16–60]).

Conclusion: This study found a high frequency of suboptimal care in the initial management of children who died of severe bacterial infection, with four separate types of errors. Other studies are needed to assess the potential avoidability of this type of death. (*Pediatr Crit Care Med* 2010; 11:469–474)

KEY WORDS: severe bacterial infection; suboptimal care; medical error; pediatric death

In 1999, the U.S. Institute of Medicine reported that medical errors and suboptimal care accounted for more deaths than the third (high estimate) or the eighth (low estimate) lead-

ing cause of death in the United States and suggested that these errors be analyzed to help improve the global health status of the population (1). Severe bacterial infections (SBI) are a good candidate for studying suboptimal care and medical errors in pediatrics. They are among the top three causes of death in Western countries (2, 3) and their early symptoms are unspecific, especially in the youngest patients. This can cause delay in diagnosis and treatment (4). Furthermore, standardized guidelines for the diagnosis and treatment of these SBI have been published and widely diffused for many years (5–7). Finally, it has been demonstrated that the early detection and rapid appropriate management of SBI (especially the early administration of appropriate antibiotic therapy and hemodynamic support) significantly reduce the morbidity and mortality in adults and children (8–11).

Only sparse data are available about suboptimal care of SBI in children (11–14). These studies have shown that suboptimal care is very frequent (71%–75%), is associated with increased mortality (odds ratio, 9; 95% confidence interval [CI], 2–33) (11, 12), and could be improved by campaigns to teach both professionals and the general public how to

respond (14). Nonetheless, several factors limit the usefulness of the results of these studies. First, three studies were restricted to SBI associated only with meningococci (11, 12, 14), although this microorganism accounts for only 34% of SBI deaths in France (15). One other study by McIntyre et al (13) considered only pneumococci, and adopted an arbitrary definition of suboptimal care (more than one consultation before diagnosis). Finally, the study by Nadel et al (12) used a hospital-based recruitment, which increased the likelihood of selection bias.

No population-based study of the frequency of suboptimal care in the management of SBI in children, regardless of the microorganism, has thus far been published. The aim of this preliminary work was to assess the frequency of suboptimal care in the initial management of children who died of SBI, on the scale of a geographically defined population and by means of a confidential inquiry, and to identify the types of mistakes made.

MATERIALS AND METHODS

General Methodology

We conducted a retrospective, population-based study of the initial care received by

*See also p. 525.

From the Clinique médicale pédiatrique (EL), Urgences pédiatriques (CG-LG), and Réanimation pédiatrique et néonatale (J-CR), Hôpital de la Mère et de l'Enfant, CHU Nantes, Nantes, France; Unité d'urgences pédiatriques et de maladies infectieuses (AM), Hôpital R. Salengro, CHU de Lille, Université de Lille-Nord de France, Lille, France; Association pour le Recherche et l'Enseignement en Pédiatrie Générale (RA), Cabinet de Pédiatrie, Issy-les-Moulineaux, France; Service de pédiatrie générale (TB), Centre Hospitalier Départemental de la Roche-sur-Yon, France; Unité INSERM U953 (NM, M-HB-C, MC), Recherche épidémiologique en santé périnatale et santé des femmes et des enfants, Paris, France; CépIDC-Inserm (AA), Centre d'épidémiologie sur les causes médicales de décès, Le Vésinet, France; and Service de pédiatrie générale (MC), Hôpital Saint-Vincent-de-Paul, Paris Descartes University, AP-HP, Paris, France.

Drs. Launay and Gras-Le Guen contributed equally to this work.

The authors have not disclosed any potential conflicts of interest.

For information regarding this article, E-mail: christele.grasleguen@chu-nantes.fr

Copyright © 2010 by the Society of Critical Care Medicine and the World Federation of Pediatric Intensive and Critical Care Societies

DOI: 10.1097/PCC.0b013e3181ce752e

children who died of SBI from January 2000 through March 2006. The geographic zone comprised two adjoining administrative districts, Loire-Atlantique and Vendée. This area in western France covers 13,500 km² and has a population of 1.6 million people, a mean household income per consumer unit of 14,500 € per year, and an infant mortality rate of 3.3‰ (3). Pediatric care in this area is provided by one university hospital center (in Nantes), four general hospitals, pediatricians in private practice (29 per 100,000 children), general practitioners (145 per 100,000 inhabitants), and two call centers for medical emergencies that can send emergency mobile medical teams (including physicians specialized in emergency medicine) to patients' homes. The organization of care in this geographic zone called for all children >3 months old and requiring hospitalization for SBI to be transferred to the Nantes University Hospital Center. The Institutional Review Committee (Comité de Protection des Personnes Ile de France III) stated that "this research was found to conform to generally accepted scientific principles and research ethical standards and to be in conformity with the laws and regulations of France, where the research experiment was performed." Written informed consent of the parents of the patients was not judged necessary for this kind of retrospective study.

Study Population

The study included all children aged 3 months to 14 yrs living in the geographic zone who died of bacterial infection during the study period. We defined infection according to the criteria of the international consensus conference on sepsis in pediatrics (16). SBI was then defined as bacterial infection leading to admission to the pediatric intensive care unit (ICU) in Nantes. Children who died of severe nonbacterial infections, whooping cough, or nosocomial infection were not included. Children whose files were too incomplete to trace the clinical history with sufficient precision were excluded.

Identification of Cases

We considered for inclusion all patients who died in the pediatric intensive care department of the Nantes University Hospital Center with at least one of the following discharge codes, or words in their electronic medical file: meningitis, purpura fulminans, septicemia, or septic shock. Also considered for possible inclusion were all children who died at home with fever, according to the records of the emergency medical call centers.

To assess the exhaustiveness of our study, our population was compared with the anonymized national cause of death database (CépiDc)

(17) from which we extracted all deaths related to SBI from 2000 through 2005 (data were not available for 2006 at the time of our study) of children aged 0 to 14 yrs who lived in the study area. The following International Statistical Classification of Diseases and Related Health Problems, Revision 9 codes listed as the principal cause of death were used: A00–09, A30–49, B95, B96, B99, G00, G01, G03, G06, J15, J86.9, and R57.8. We also questioned the local database manager directly in suspicious cases.

Data Collected

Data were collected from the complete patient medical file by three physicians from the pediatric intensive care department and one physician not working in the ICU. A preestablished template reconstructed the timed and dated medical observations, which were anonymized and expressed as neutrally as possible. It included neither the final diagnosis nor the outcome. This standardized report retraced the patient's history, the sequence of the initial management (arbitrarily defined as the period from the onset of the first symptom compatible with starting SBI until arrival in the pediatric intensive care department), the symptoms presented by the child, the clinical signs noted by parents or physicians, and the actions taken by parents and physicians. Treatment doses administered were reported as mg or mL per kg.

Analyses

We described the children studied, their demographic characteristics, and their final diagnoses, especially bacteriologic. We also analyzed the clinical presentation, focusing on early symptoms (first symptoms described by the parents), reasons for the first medical consultation (symptoms that were not necessarily present at the beginning but the onset of which led the parents to seek medical care), the signs of sepsis (tachycardia, bradycardia, and tachypnea) (16), and the presence of tonus disorders, impaired vigilance, respiratory distress, moaning, or other signs of potentially SBI, such as meningism or extensive purpura. We studied the sequence of care for each patient.

We calculated the frequency of suboptimal care or errors in the management of the children who died of SBI. The suboptimal character of the initial management was first determined independently by two experts, blinded to final diagnosis and outcome. One of these experts was an experienced pediatrician in private practice and the other was a pediatric intensive care specialist who supervises a pediatric emergency department. The experts were asked to assess whether the care was or was not optimal in comparison with specific criteria selected from French and interna-

tional clinical practice guidelines applicable during the study period (6, 18, 19), concerning the timing of administration of antibiotics in meningococemia (immediate in case of extensive purpura) and the modality of administration of hemodynamic support in septic shock (bolus up to 40 mL/kg in the first hour). Using the study by Nadel et al (12) that evaluated suboptimal care in meningococcal disease, we defined delay in seeking medical care by parents as the absence of immediate consultation in cases of fever with either a purpuric rash or accompanied by other signs of severity: cyanosis, moaning, convulsions, confusion, impairment of higher functions, intense headaches, intense muscle or articular pain, marked asthenia, persistent vomiting, or cold hands or feet. We also arbitrarily considered the failure to seek medical care when a high fever lasted >48 hrs as a delay in seeking medical care. The experts could reach one of three possible conclusions: certainly suboptimal management (situation 1); possibly suboptimal management (situation 2); and optimal management (situation 3). The degree of agreement between the experts was then assessed by calculating κ and weighted κ according to a quadratic method (Stata software, Stata Corp., College Station, TX). We chose a quadratic weighting system because the discordance between situations 1 and 3, or 2 and 3 seemed greater than between situations 1 and 2. Interpretation of κ was based on the Landis and Koch scale (20). In cases of disagreement, the optimal nature of the care was determined by a third expert, a hospital-staff pediatrician in a general pediatric department; the frequency of suboptimal care was then calculated. We described the distribution of the different categories of suboptimal care identified by the experts within the categories proposed by the Institute of Medicine (1), to which we added a category "parental delay in seeking medical care" because this type of suboptimal care has been underlined in several studies of the management of meningococcal disease in children (11, 12). Because some cases involved several categories of suboptimal care, the total number may exceed the number of children.

RESULTS

Patients

In total, 23 children who met all the inclusion criteria and none of the exclusion criteria died from SBI during the study period. Two medical files (10%) were finally excluded because the clinical and bacteriologic information was insufficient: an 11-month-old boy who died with febrile purpura and a 27-month-old boy who died of pneumococcal meningitis.

Table 1. Description of the age, the final diagnosis, bacterial documentation, and cause of death for each patient

Patient	Age (yrs)	Final Diagnosis	Bacteria Identified	Cause of Death
1	4.8	Purpura fulminans	<i>Neisseria meningitidis</i> (serotype B)	Septic shock
2	0.3	Meningitis	<i>Streptococcus pneumoniae</i>	Septic shock
3	0.4	Meningitis	<i>Salmonella enterica</i>	Septic shock
4	2	Meningitis	<i>Streptococcus pneumoniae</i>	Intracranial hypertension
5	1.6	Purpura fulminans	No bacteria documented	Septic shock
6	2.4	Meningitis	<i>Streptococcus pneumoniae</i>	Intracranial hypertension
7	0.6	Purpura fulminans	<i>Neisseria meningitidis</i> (serotype B)	Septic shock
8	0.4	Purpura fulminans	<i>Neisseria meningitidis</i>	Septic shock
9	10.3	Meningitis	<i>Neisseria meningitidis</i> (undefined serotype)	Septic shock
10	3.9	Pneumonia with empyema	<i>Mycoplasma pneumoniae</i>	Refractory hypoxemia and multiple organ failure
11	2.8	Pneumonia with empyema	<i>Streptococcus pyogenes</i>	Septic shock
12	2.8	Purpura fulminans	<i>Neisseria meningitidis</i> (serotype C)	Septic shock
13	0.8	Purpura fulminans	<i>Neisseria meningitidis</i> (undefined serotype)	Septic shock
14	2	Meningitis	<i>Neisseria meningitidis</i> (undefined serotype)	Septic shock
15	4.9	Purpura fulminans	<i>Neisseria meningitidis</i> (serotype C)	Septic shock
16	0.9	Purpura fulminans	<i>Neisseria meningitidis</i> (serotype C)	Septic shock
17	2.4	Purpura fulminans	<i>Neisseria meningitidis</i> (serotype B)	Septic shock
18	1.2	Purpura fulminans	<i>Neisseria meningitidis</i> (serotype B)	Septic shock
19	4.7	Meningitis	No bacteria documented	Septic shock and intracranial hypertension
20	0.3	Purpura fulminans	<i>Neisseria meningitidis</i> (serotype B)	Septic shock
21	1.1	Pyelonephritis and septic shock	<i>Escherichia coli</i>	Septic shock

The median age of the 21 children studied was 2 yrs (interquartile range, [0.8–2.8]), with a 1:0.8 ratio of boys to girls. Seven children had significant medical histories: extreme preterm birth, asthma, rheumatoid purpura, mitochondrial cytopathy, chromosomal abnormality, or salmonella meningitis. The final diagnoses were: purpura fulminans (n = 11), meningitis (n = 7), pleuropulmonary disease (n = 2), and septicemia from acute pyelonephritis (n = 1). The microorganisms identified were: *Neisseria meningitidis* (n = 12 patients; 57%); serotype B (n = 5); serotype C and 4 (n = 3) for which a serovar could not be identified; *Streptococcus pneumoniae* (n = 3); and *Streptococcus pyogenes*, *Mycoplasma pneumoniae*, *Salmonella enterica* serovar Bredeney and *Escherichia coli* (n = 1 patient for each microorganism) (Table 1). Extensive testing was unable to identify any bacteria in two children; one had a clinical picture typical of purpura fulminans and the other showed typical clinical and cerebrospinal fluid findings for bacterial meningitis. Thirteen children died the day they arrived in intensive care (day 0), three died on day 1, and the other five died before day 13. All the deaths were due to the initial SBI and not to iatrogenic complications (Table 1).

Symptoms and Stages of Care

All the children had fever as one of the first symptoms. The other initial symptoms were not specific for SBI except for one purpuric rash. The most frequent reason for the first consultation was the

appearance of a purpuric rash, for seven children (33%; 95% CI, 12–54), and moaning for two children; the other reasons for seeking medical care were non-specific SBI symptoms. Cardiopulmonary arrest was the cause of the first medical intervention for one child. Sixteen children (76%; 95% CI, 57–95) had at least one sign of severity at the first consultation: extensive purpura for nine children (43%; 95% CI, 21–65); hemodynamic disorders for eight; impaired consciousness for five; respiratory distress for four; and meningeal syndrome and hypotonia for three children each. Nine children presented several signs of severity.

The sequence of care for these 21 children varied substantially. Initial medical care was sought from a general practitioner for 12 (57%; 95% CI, 35–79), a mobile emergency medical service team was called for six and three went directly to a pediatric emergency department. Fifteen (71%; 95% CI, 50–92) children were hospitalized or referred to a hospital at the time of their first medical consultation. Six (29%; 95% CI, 8–50) children went home after their first consultation.

Suboptimal Care

One of the primary two experts could not reach a conclusion about the quality of care for one patient. The coefficient of agreement was thus assessed for the remaining 20 files. The two experts reached different conclusions for six of these 20 files. The κ value was 0.44 ($p = .003$) and the weighted κ , according to a quadratic

method, was 0.73 ($p = .0005$). In no case did one expert find suboptimal management, whereas the other found optimal management. The third expert assessed management for the case in which one of the first two had abstained and the six discordant cases. Sixteen (76%; 95% CI, 57–95) of 21 patients were judged to have been managed less than optimally: 13 (62%; 95% CI, 40–84) certainly suboptimally, and three (14%; 95% CI, 0–30) possibly suboptimally.

Parental delay in seeking medical care was noted in seven patients (33%; 95% CI, 12–54), because of the absence of a rapid consultation for two children with purpura, for three other patients with worrisome changes in behavior, and for two more patients with a fever developing for >48 hrs before the first consultation. One of these seven delays in seeking medical care was the only defect in care identified for the child from the onset of disease until admission to the pediatric intensive care department. The suboptimal nature of the management of 16 children was associated with medical errors in 15 and included 35 specific errors. These defects in management were distributed, according to the Institute of Medicine grid (Table 2), into diagnostic errors (31%; 95% CI, 15–47), treatment errors (57%; 95% CI, 40–74), and “other” errors (12%; 95% CI, 1–23). Among the diagnostic errors, underassessment of initial severity was the most frequent. These errors were expressed most often by the failure to call the mobile emergency medical service team for transportation to the hospital or failure to recognize signs of

Table 2. Classification of errors in management causing suboptimal care of 21 children who died of a severe bacterial infection

Classification According to the Institute Of Medicine		Errors in Management Identified by Experts	Number of Patients (% of Total Errors, n = 35) [% of Deaths, n = 21]
Diagnostic errors	Underestimation of severity	Signs of preshock not taken into account	5 (14) [24]
		Lack of use of mobile medical teams despite signs of sepsis or purpura	4 (11) [19]
		Delay in decision to transfer to intensive care	1 (3) [5]
Treatment errors	Failure to act on results of test	Erroneous interpretation of CRP	1 (3) [5]
		Error in the dose or method of using a drug or treatment	5 (14) [24]
	Avoidable delay in treatment after diagnosis	Error in dose of antibiotic treatment	1 (3) [5]
		Antibiotics not administered to patients with purpura	4 (11) [19]
		Delay in antibiotic administration	3 (9) [14]
		Delay in administration of fluid resuscitation	1 (3) [5]
		Absence of second fluid resuscitation	3 (9) [14]
Other (other system failure)	Error in the performance of a procedure	Failure to administer inotropic agent	2 (5) [10]
		Absence of intubation and mechanical ventilation	1 (3) [5]
		Failure in placement of a catheter	1 (3) [5]
Other (failure of communication)		Excessive delay in arrival of physician (private practice or mobile medical teams)	2 (5) [10]
		Error in coordination of medical teams	1 (3) [5]

CRP, C-reactive protein.

Some cases involved several different errors in any of the three categories (diagnostic, treatment, or other).

preshock (n = 8; 38% of patients; 95% CI, 16–60). Among the treatment errors, delay after diagnosis in administering appropriate treatment (antibiotic therapy or fluid resuscitation) was the most frequent error, concerning eight children (38% of patients; 95% CI, 16–60). Immediate antibiotic therapy was administered in only one of the five cases covered by the guidelines for purpura fulminans. Dose errors and especially the insufficiency of fluid resuscitation were frequent (24% of patients; 95% CI, 9–35). Hemodynamic support (absence, delay or insufficiency of fluid resuscitation, and failure to administer inotropes or vasoconstrictors) was inadequate in nine cases (43%; 95% CI, 21–65).

Every type of participant committed at least one error: eight errors were attributed to private practitioners, nine to the mobile emergency medical teams, and 11 to emergency department physicians (five in general and six in pediatric emergency departments).

Exhaustiveness of the Geographic Base

Six children in the study area died from 2000 through 2005 from a cause of death

code compatible with an SBI according to the CépiDc data but had not been identified, using our search strategy. After complete investigation, all of them met the exclusion criteria: age <3 months (n = 1); invasive aspergillosis (n = 1); and nosocomial infection (n = 4).

DISCUSSION

Principal Results

We report the first population-based study assessing the frequency of suboptimal care in the initial management of children who died from any SBI, regardless of the microorganism. The frequency of cases with at least one error or suboptimal decision before the study patients' arrival in the ICU was high (76%; 95% CI, 57–95). The types of errors identified included delay in administering appropriate treatment (emergency antibiotic therapy for patients with a purpuric rash), dose errors (especially insufficient or unrepeated fluid resuscitation), and overall underestimation of disease severity. In a third of the cases, the experts also considered that the parents had waited too long to seek medical care.

Internal Validity

Assessing the suboptimal quality of care is difficult and controversial (21). In our study, this determination was based on evaluation by two independent experts, and the level of agreement was high, as the weighted κ showed. To reinforce the objectivity of the experts' judgment and to work as closely as possible in the conditions of daily clinical practice, the experts were not informed of the patient's outcome or final diagnosis. Nonetheless, the experts had a complete clinical history through arrival in intensive care and, thus, could assess the entire course of treatment until then to assess its quality. This procedure might have led to bias, because knowledge of the subsequent course can modify an expert's perception (22). Alternatives are possible to avoid this bias. Ninis et al (11) studied 143 children who died from and 355 who survived meningococcal infection in Great Britain and compared their care during the first 24 hrs of hospitalization. The four experts in that study had access to the clinical course hour by hour and assessed the management on an hourly basis, in response to events of the previ-

ous hour. The suboptimal character of care was thus determined in an hourly comparison between the care actually provided and the optimal care proposed by the expert panel. Nonetheless, the potential bias associated with not using this method seemed slight to us for the three principal types of error, in view of the large gap between the practices observed and those expected according to the guidelines.

The use of synthetic medical observations is less objective than a complete copy of the medical file, as in confidential inquiries about maternal deaths, in which experts receive entire files for review (23). Nonetheless, in our study, which focused principally on care received before admission to the ICU, most of the patients did not have a complete medical file but only a summary of their visit to the emergency department. The impairment of the available information due to its summary in a form is therefore probably minimal.

Finally, the completeness of our study is excellent, because we identified all the files eligible according to C epiDc data, and only two (10%) files could not be analyzed because of insufficient data for tracing the clinical history.

External Validity

Despite the recruitment pattern of our study (a single geographic zone), the prevalence of suboptimal care we found was very close to that reported in other countries. A prospective study by Nadel et al (12) in a London tertiary pediatric center showed a frequency of 71% of suboptimal care and errors in the hospital management of 54 children with meningococcal disease. In the study by Ninis et al (11), one or more errors were identified in the files of 75% of the children who died and 45% of those who survived. The types of suboptimal care in our study were identical to those in studies of meningococcal disease. For example, Nadel et al (12) observed parental delay in seeking medical care in 30% of the patients studied; this finding is similar to the percentage (33%) in our study. These high rates of parental delays suggest that the strategy to improve outcomes should include public education.

Results for medical management were also consistent with previous reports. Ninis et al (11) found a failure to recognize signs of severity in 38% of the deaths, whereas Riordan noted a lack of prehospital antibiotic administration in patients

with purpuric rash in 25% of the patients with meningococcal infection (14). In our study, of the five children who had a purpuric rash and who consulted a private practice physician first, four did not receive antibiotic therapy, contrary to the guidelines sent to all French physicians by the Ministry of Health in 2001 (6). The study by McIntyre et al (13) in Australia, which assessed the impact of early management of pneumococcal meningitis on outcome in children, found that administration of antibiotic therapy was delayed >6 hours after admission in 41 of 120 patients. The inadequacy of hemodynamic support (fluid resuscitation or inotropic agents) during the initial management of sepsis was demonstrated by an audit by a London group: 62% of children had not received the care recommended in guidelines in terms of fluid resuscitation and inotropes vs. 43% in our study (24).

Most children (76%; 95% CI, 57–95) presented signs of severity at the time of their first consultation. This might suggest that their prognosis was already very poor by the time that they received medical care. This also reflects the importance of the information given to parents about signs for the early recognition of SBI, through brochures distributed to the population, as in Great Britain (25). This information is absent from the “health education” information included in the personal health record booklet that parents receive at each child’s birth in France.

Other Issues

The methods used in our study do not allow us to examine or suggest a possible causal relationship between quality of care and favorable outcome, in the absence of a comparison group of SBI survivors. The study by Ninis et al (11) showed that the presence of at least one error in management increased the odds of mortality in meningococcal infections (odds ratio, 9; 95% CI, 2–33). The study by McIntyre et al (13) showed an association between increased morbidity in survivors of pneumococcal meningitis and delayed hospitalization (odds ratio, 3.4; 95% CI, 1.0–11). In situations of septic shock, mortality is thought to increase by 40% for each hour of delay in fluid resuscitation (9). Since the period covered by our study, the publication of several sets of guidelines (16, 26, 27) has provided physicians with the opportunity to implement them and to improve care. An ob-

servational prospective study (28) in an adult ICU in Italy recently showed that good compliance with one of the Surviving Sepsis Campaign guidelines for early management of sepsis was associated with higher survival rate. A pediatric study (29) conducted in Brazil showed a reduction in deaths when fluid resuscitation was administered <30 mins after diagnosis. This study also identified some characteristics, such as lack of a local clinical protocol, associated with delayed fluid resuscitation. We can thus suppose that improving the quality of care of SBI in pediatric patients should reduce its related mortality, as suggested by the scientific committee of the Surviving Sepsis Campaign (26, 27). Ongoing large-scale prospective studies (30) will probably help to confirm this hypothesis.

REFERENCES

1. To Err Is Human: Building a Safer Health System. Kohn LT, Corrigan JM, Donaldson MS (Eds). Washington, DC, Committee on Quality of Health Care in America, Institute of Medicine, National Academy Press, 2000
2. Watson RS, Carcillo JA: Scope and epidemiology of pediatric sepsis. *Pediatr Crit Care Med* 2005; 6:S3–S5
3. INSEE: Causes of infant death by department and areas 2005. Leading of death for children and young people 2005. Incomes of Ligerian in 2004. Available at <http://www.insee.fr>. Accessed on December 24, 2007
4. Thompson MJ, Ninis N, Perera R, et al: Clinical recognition of meningococcal disease in children and adolescents. *Lancet* 2006; 367: 397–403
5. American College of Chest Physicians/Society of Critical Care Medicine Consensus Conference: Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. *Crit Care Med* 1992; 20:864–874
6. Conseil sup erieur d’hygi ne publique. Opinion from the French High Committee on Public Health, Mars 10, 2000. Available at http://www.sante.gouv.fr/html/dossiers/cshpf/a_mt_100300_meningite_01.htm. Accessed on December 22, 2007
7. Guidelines for the management of severe sepsis and septic shock. The International Sepsis Forum. *Intensive Care Med* 2001; 27: S1–S134
8. Garnacho-Montero J, Aldabo-Pallas T, Garnacho-Montero C, et al: Timing of adequate antibiotic therapy is a greater determinant of outcome than are TNF and IL-10 polymorphisms in patients with sepsis. *Crit Care Med* 2006; 10:2742–2751
9. Han YY, Carcillo JA, Dragotta MA, et al: Early reversal of pediatric-neonatal septic shock by community physicians is associated with im-

- proved outcome. *Pediatrics* 2003; 112: 793–799
10. Carcillo JA: Reducing the global burden of sepsis in infants and children: A clinical practice research agenda. *Pediatr Crit Care Med* 2005; 6:S157–S164
 11. Ninis N, Phillips C, Bailey L, et al: The role of healthcare delivery in the outcome of meningococcal disease in children: Case-control study of fatal and non-fatal cases. *BMJ* 2005; 330:1475–1480
 12. Nadel S, Britto J, Booy R, et al: Avoidable deficiencies in the delivery of health care to children with meningococcal disease. *J Accid Emerg Med* 1998; 15:298–303
 13. McIntyre PB, Macintyre CR, Gilmour R, et al: A population based study of the impact of corticosteroid therapy and delayed diagnosis on the outcome of childhood pneumococcal meningitis. *Arch Dis Child* 2005; 90:391–396
 14. Riordan FA: Improving promptness of antibiotic treatment in meningococcal disease. *Emerg Med J* 2001; 18:162–163
 15. Floret D: Pediatric deaths due to community-acquired bacterial infection. Survey of French pediatric intensive care units. *Arch Pediatr* 2001; 8:S705–S711
 16. Goldstein B, Giroir B, Randolph A: International pediatric sepsis consensus conference: Definitions for sepsis and organ dysfunction in pediatrics. *Pediatr Crit Care Med* 2005; 6:2–8
 17. Centre d'épidémiologie sur les causes médicales de décès, Institut National de la Santé et de la Recherche Médicale. Causes of death between 1979 and 2005. Available at <http://www.cepidc.vesinet.inserm.fr/inserm/html/index2.htm>. Accessed on December 22, 2007
 18. Société de Réanimation de Langue Française. Use of catecholamines during septic shock (adults and children). XV consensus conference of the French Society of Intensive Care 1996. Available at <http://www.sfar.org/srlfsfar/catecholcscons.html>. Accessed on February 15, 2007
 19. Control of meningococcal disease: Guidance for consultants in communicable disease control. PHLS Meningococcal Infections Working Group and Public Health Medicine Environmental Group. *Commun Dis Rep CDR Rev* 1995; 5:R189–R194
 20. Landis JR, Koch GG: The measurement of observer agreement for categorical data. *Biometrics* 1977; 33:159–174
 21. McDonald CJ, Weiner M, Hui SL: Deaths due to medical errors are exaggerated in Institute of Medicine report. *JAMA* 2000; 284:93–95
 22. Bradfield A, Wells GL: Not the same old hindsight bias: Outcome information distorts a broad range of retrospective judgments. *Mem Cognit* 2005; 33:120–130
 23. Bouvier-Colle MH, Varnoux N, Breart G: Maternal deaths and substandard care: The results of a confidential survey in France. *Medical Experts Committee. Eur J Obstet Gynecol Reprod Biol* 1995; 58:3–7
 24. Inwald DP, Tasker R, Peters MJ, et al: Emergency management of children with severe sepsis in the United Kingdom—The results of the Paediatric Intensive Care Society sepsis audit. *Arch Dis Child* 2009; 94:348–353
 25. Meningitis Research Foundation. Meningitis Research Foundation's Web site. Available at <http://www.meningitis.org>. Accessed on January 23, 2009
 26. Dellinger RP, Carlet JM, Masur H, et al: Surviving Sepsis Campaign guidelines for management of severe sepsis and septic shock. *Intensive Care Med* 2004; 30:536–555
 27. Martin C, Garnier F, Vallet B: Recommendations for management of severe sepsis and septic shock. Surviving sepsis campaign. *Ann Fr Anesth Reanim* 2005; 24:440–443
 28. Zambon M, Ceola M, Almeida-de-Castro R, et al: Implementation of the Surviving Sepsis Campaign guidelines for severe sepsis and septic shock: We could go faster. *J Crit Care* 2008; 23:455–460
 29. Oliveira CF, Nogueira de Sa FR, Oliveira DS, et al: Time- and fluid-sensitive resuscitation for hemodynamic support of children in septic shock: Barriers to the implementation of the American College of Critical Care Medicine/Pediatric Advanced Life Support Guidelines in a pediatric intensive care unit in a developing world. *Pediatr Emerg Care* 2008; 24:810–815
 30. Launay E, Martinot A, Assathiany R, et al: [Sub-optimal care during the initial treatment of severe bacterial infection in children]. [Article in French.] *Arch Pediatr* 2009; 16:649–651

Annexe 4 : Article issu de la partie II – Chapitre 2



Why Children with Severe Bacterial Infection Die: A Population-Based Study of Determinants and Consequences of Suboptimal Care with a Special Emphasis on Methodological Issues

Elise Launay^{1,2*}, Christèle Gras-Le Guen^{1,3}, Alain Martinot⁴, Rémy Assathiany⁵, Elise Martin¹, Thomas Blanchais¹, Catherine Deneux-Tharoux², Jean-Christophe Rozé⁶, Martin Chalumeau^{2,7}

1 CHU Nantes, Hôpital de la Mère et de l'Enfant, Clinique médicale pédiatrique, Faculté de médecine de Nantes, Nantes, France, **2** Inserm U1153, Obstetrical, Perinatal and Pediatric Epidemiology Research Team, Research Center for Epidemiology and Biostatistics Sorbonne Paris Cité (CRESS), Paris Descartes University, Paris, France, **3** CHU Nantes, Hôpital de la Mère et de l'Enfant, Urgences pédiatriques, Faculté de médecine de Nantes, Nantes, France, **4** CHU de Lille, Hôpital R. Salengro, Unité d'urgences pédiatriques et de maladies infectieuses, Université de Lille-Nord de France, Lille, France, **5** Association pour le Recherche et l'Enseignement en Pédiatrie Générale (AREPEGE); Association Française de Pédiatrie Ambulatoire (AFPA), Cabinet de Pédiatrie, Issy-les-Moulineaux, France, **6** CHU Nantes, Hôpital de la Mère et de l'Enfant, Réanimation pédiatrique et néonatale, Faculté de médecine de Nantes, Nantes, France, **7** Hôpital Necker Enfants Malades, AP-HP, Service de pédiatrie générale, Paris Descartes University, Paris, France

Abstract

Introduction: Suboptimal care is frequent in the management of severe bacterial infection. We aimed to evaluate the consequences of suboptimal care in the early management of severe bacterial infection in children and study the determinants.

Methods: A previously reported population-based confidential enquiry included all children (3 months- 16 years) who died of severe bacterial infection in a French area during a 7-year period. Here, we compared the optimality of the management of these cases to that of pediatric patients who survived a severe bacterial infection during the same period for 6 types of care: seeking medical care by parents, evaluation of sepsis signs and detection of severe disease by a physician, timing and dosage of antibiotic therapy, and timing and dosage of saline bolus. Two independent experts blinded to outcome and final diagnosis evaluated the optimality of these care types. The effect of suboptimal care on survival was analyzed by a logistic regression adjusted on confounding factors identified by a causal diagram. Determinants of suboptimal care were analyzed by multivariate multilevel logistic regression.

Results: Suboptimal care was significantly more frequent during early management of the 21 children who died as compared with the 93 survivors: 24% vs 13% ($p=0.003$). The most frequent suboptimal care types were delay to seek medical care (20%), under-evaluation of severity by the physician (20%) and delayed antibiotic therapy (24%). Young age (under 1 year) was independently associated with higher risk of suboptimal care, whereas being under the care of a paediatric emergency specialist or a mobile medical unit as compared with a general practitioner was associated with reduced risk.

Conclusions: Suboptimal care in the early management of severe bacterial infection had a global independent negative effect on survival. Suboptimal care may be avoided by better training of primary care physicians in the specifics of pediatric medicine.

Citation: Launay E, Gras-Le Guen C, Martinot A, Assathiany R, Martin E, et al. (2014) Why Children with Severe Bacterial Infection Die: A Population-Based Study of Determinants and Consequences of Suboptimal Care with a Special Emphasis on Methodological Issues. PLoS ONE 9(9): e107286. doi:10.1371/journal.pone.0107286

Editor: Susanna Esposito, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Università degli Studi di Milano, Italy

Received: June 10, 2014; **Accepted:** August 12, 2014; **Published:** September 23, 2014

Copyright: © 2014 Launay et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability: The authors confirm that all data underlying the findings are fully available without restriction. All relevant data are within the paper and its Supporting Information files.

Funding: The authors received no specific funding for this work. Inserm Unit 1153 received a grant from the Bettencourt Foundation (Coups d'élan pour la recherche française) in support of its research activities. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* Email: elise.launay@chu-nantes.fr

☞ These authors contributed equally to this work.

Introduction

Bacterial infection remains a major cause of childhood mortality in industrialised countries. [1] In 2009, Harndern et al. reviewed pediatric deaths in 5 regions of the United Kingdom and found that among the 15% of deaths related to infection, failure to recognise and manage severe bacterial infection (SBI) was the most common avoidable primary care factor. [2] In 2010, we published a population-based study evaluating optimality of care for 21 children who died due to SBI: the initial medical management was suboptimal in 76% of cases, with a delay in seeking medical care in 33%. [3] These alarming frequencies in suboptimal initial care in pediatric patients with SBI do not allow for drawing conclusions on the relationship between suboptimal care and outcomes because both studies focused on patients who died.

The consequences of suboptimal care in pediatric patients with SBI have been examined in 4 studies. [4,5,6,7] All found clinically meaningful and statistically significant associations between suboptimal care and morbidity and mortality. [4,5,6,7] However, the results were limited by methodological concerns such as selection bias related to hospital-based recruitment, [6] classification bias related to arbitrary definition of diagnosis delay as consultation more than once before hospitalisation, [5] non-independent evaluation of the optimality of care, [7] non-justified use of continuous variables in multivariable models, [4,5,7] and/or selection of non-appropriate variables for adjustment (without using a causal diagram that could help deal with co-variables that could be confounders or intermediate variables). [8,9] No study examined the determinants of this suboptimal care to inform corrective actions for parents and healthcare workers.

The aim of the present study was to evaluate the determinants and consequences of suboptimal care in the initial management of SBI in children, using appropriate methodological approaches, to evaluate the relevance of future targeted corrective actions for parents and healthcare workers.

Methods

General methodology

The present study is an extension of a previously published population-based confidential enquiry into the quality of initial care in children age from 3 months to 16 years who died of SBI from January 2000 to March 2006 in a geographic zone of France comprising two adjoining administrative districts. [3] The definition of SBI (bacterial infection leading to admission to the pediatric intensive care unit [PICU]), the strategy of identification of cases, and the assessment of exhaustiveness were described in detail in the previous publication. Pediatric care in this area was provided by one university hospital center (in Nantes), four general hospitals, pediatricians in private practice, general practitioners (GPs), and two call centers for medical emergencies that could send emergency mobile medical teams (including physicians specialized in emergency medicine) to the patient's home. For the present study, we defined a control group of pediatric patients who survived a SBI during the same period and in the same geographic region. The organization of care called for all children older than 3 months and requiring hospitalization for SBI to be transferred to the PICU of the Nantes university hospital. Thus, controls were all pediatric patients hospitalised for SBI in the PICU of the hospital during the study period. Controls were identified by discharge codes and the microbiology laboratory electronic files as described previously. [3] The initial research was approved by Institutional Review Committee (Comité de Protection des Personnes Ile de France III) and this extension was

approved by the ethics committee of the Nantes university hospital (Groupe Nantais d'Ethique dans le Domaine de la Santé), which approved a waiver of the need for consent. The results were reported according to the STROBE checklist for reporting observational studies. [10]

Data were collected as previously described from the complete patient medical file: a pre-established template reconstructed the timed and dated medical observations with blinding to final diagnosis or outcome. [3] Children whose files were too incomplete to trace the clinical history with sufficient precision were identified and excluded.

Optimality of care evaluation

Two experts (an experienced pediatrician in private practice and a pediatric intensive care specialist who supervises a pediatric emergency department), blinded to final diagnosis and outcome, independently determined the suboptimal character of the initial management as described previously. [3] These two experts were not involved in the management of any included children. Experts had to justify their final conclusion by giving details on the optimality (optimal or not optimal) of each care in terms of specific criteria selected from national and international clinical practice guidelines applicable during the study period: [11,12,13] the timing of administration of antibiotics for meningococemia (immediate in case of extensive purpura) and the modality of administration of hemodynamic support in septic shock (bolus up to 40 mL/kg in the first hour). As in the study by Nadel *et al.*, [6] which evaluated suboptimal care for meningococcal disease, we defined delay in seeking medical care by parents as the absence of immediate consultation in cases of fever with a purpuric rash or accompanied by other signs of severity: cyanosis, moaning, convulsions, confusion, impairment of higher functions, intense headaches, intense muscle or articular pain, marked asthenia, persistent vomiting, or cold hands or feet. We also arbitrarily considered the failure to seek medical care when a high fever lasted more than 48 hr as a delay in seeking medical care. For each child, we were then able to evaluate the optimality of 6 different key types of care: 1) seeking medical care by parents; 2) evaluation of sepsis signs and detection of severe disease by a physician, 3) timing of antibiotic therapy, 4) dosage of antibiotic therapy, 5) timing of saline bolus, and 6) dosage of saline bolus.

Analyses

We described the children studied, their demographic characteristics, and their final diagnoses, especially bacteriologic. We analyzed signs of severe disease: signs of sepsis (tachycardia, bradycardia, and tachypnea), [14] presence of tonus disorders, impaired vigilance, respiratory distress, moaning, or other signs of potential SBI, such as meningism or extensive purpura. We described the sequence of care of children (first medical contact and number of consultations before hospitalization).

We assessed the degree of agreement between the two experts for each type of care by calculating the kappa coefficient interpreted with the Landis and Koch scale. [15] In cases of disagreement, the optimal nature of the care was determined by a third expert. We analyzed the 6 categories of suboptimal care by outcome and physicians' qualification. We also analyzed risk factors for death (relation between suboptimal care and death) and determinants of suboptimal medical care (excluding seeking medical care).

We analyzed the crude and adjusted association between number of suboptimal care and death. The number of suboptimal care was the sum of the 6 above-mentioned types for each child and thus ranged from 0 to 6. To identify confounding variables,

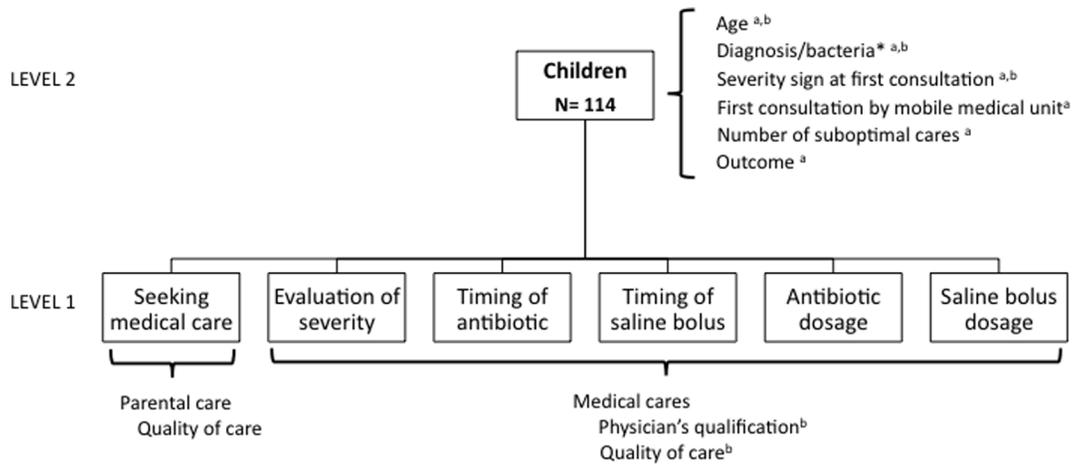


Figure 1. Structure of data. a = co-variables used in the study of the consequences of suboptimal care. b = co-variables used in the study of the determinants of medical sub-optimal care. doi:10.1371/journal.pone.0107286.g001

we built a theoretical causal diagram between optimality of care and outcome (dead/alive at discharge from hospital) based on the published pathophysiological concepts of severe sepsis (Figure S1) and adapted from this a “realistic” causal diagram between optimality of initial care (before admission to a PICU) and outcome considering the available data and using DAGitty software (Figure S2). [16,17] Clinical phenotype was defined by diagnosis (meningitis versus other diagnosis) and two other variables reflecting the measurable intrinsic severity of the disease: presence of severity sign at the first consultation and first consultation by a mobile medical unit (this unit is reserved for patients with the most severe condition in France). Covariables tested on univariate analysis were age of children, diagnosis, sign of severe disease at the first consultation, and first consultation by a mobile medical unit (Figure 1). Relevant variables according to the causal diagram were included in multivariate analyses.

To evaluate the determinants of the quality of initial medical care, we considered each of the 5 medical care types by children. We used a hierarchical regression model that took into account the hierarchical structure of the data (i.e., non-independence of the variables for the 5 care types), and allowed us to include characteristics of care at the care level (level 1; i.e., quality of each care [optimal/suboptimal] and qualification of physician giving the care) and characteristics of the children at the level of the child (level 2; i.e., age of children, diagnosis, and presence of signs of severe disease at first consultation [Figure 1]). We included variables considered associated with suboptimal care (Figure S2). First, we estimated a random intercept model without any variable (“empty” model) to obtain the baseline children-level variance and to test the effect of children. Then, we included care and children characteristics and estimated the association of these variables and quality of care. We calculated the proportion of the model’s variance explained by level 1 and level 2 variables defined as (variance of the model with level 1 variables – variance of the empty model)/variance of empty model and (variance of the model with level 1 and level 2 variables – variance of the empty model)/variance of empty model, respectively. Quantitative variables were tested for linearity and transformed into polynomials of the smallest degree when deviation was observed. Analyses involved use of Stata 11 (StataCorp, College Station, TX, USA).

Results

Patients and care pathway

In total, 119 patients were eligible; five (4%) were excluded because of incomplete charts, for 114 patients analysed (Figure S3). Overall, 21 children died (18%, 95% confidence interval [95% CI] 11–25) before PICU admission (n = 1) or PICU discharge (n = 20), and 93 survived. The clinical characteristics of the 21 children who died were described elsewhere. [3] The median age of the 114 included children was 2.4 years [interquartile range 0.7–6.7 years], the sex ratio was 1.3 (M/F) and one half had known serious medical conditions at the time of diagnosis (Table 1). More than a half of the children (63%) presented signs of severe disease at the first medical contact. Meningitis was the most frequent diagnosis (57%), followed by purpura fulminans (35%). Meningococcus was found in 47% of cases. The first medical contact was a GP in 66% of cases, an emergency physician in 25%, and a mobile medical unit in 9%; 60% of children were hospitalized after this first medical contact. Children whose first medical contact was the mobile medical unit were more likely to have severity signs at this first medical contact (100% versus 60%, p = 0.01).

Optimality of care

Agreement between experts was “moderate” for evaluation of the optimality of the delay to seek medical care and for saline bolus dosage, with a κ coefficient of 0.40 ± 0.06 and 0.46 ± 0.09 , respectively ($p < 0.001$). Agreement with the optimality of the 4 other medical care types (severity evaluation, antibiotic therapy timing and dosage, and saline bolus timing) was “substantial” or “almost perfect,” with κ 0.78 ± 0.09 ; 0.78 ± 0.09 ; 0.67 ± 0.09 and 0.88 ± 0.09 , respectively ($p < 0.001$). Overall, 52% of children received at least one care type evaluated as suboptimal, and 25% received two or more suboptimal care types (Table 1). Among the 684 individual care types delivered, 104 (15%, 95% CI 12–18%) were suboptimal. Parental delay in seeking medical care and physician underestimation of severity and delayed antibiotic administration accounted for 70% of this suboptimal care (22%, 22% and 26%, respectively). The frequency of suboptimal care in the initial management did not significantly decrease over the years (19% to 15% from 2000 to 2005; p for trend > 0.8) nor did the frequency of each type of care ($p > 0.2$).

Table 1. Patient characteristics and care pathways before admission to a pediatric intensive care unit, quality of care and their association with outcome by dead and alive children and univariate and multivariate analysis.

	Total n = 114 (%)		Dead n = 21 (%)		Alive n = 93 (%)		Univariate analysis		Multivariate analysis [§]		
							OR	95% CI	aOR	95% CI	P
PATIENTS											
Age, yr											
Median ^a [IQR]	2.4 [0.7–6.7]	2.0 [0.9–2.8]	2.9 [0.7–7.1]	0.85	0.73–1.0	0.057	0.82	0.68–0.99	0.04		
<1 yr	32 (28)	7 (33)	25 (27)								
1 to 2 yr	21 (18)	4 (19)	17 (18)			0.01					
2 to 5 yr	27 (24)	9 (43)	18 (19)								
≥5 yr	34 (30)	1 (5)	33 (36)								
Sex ratio M/F	1.28	1.33	1.27			0.92					
Underlying medical conditions, n (%)											
Severity signs at first medical contact^b, n (%)	39 (52)	7 (33)	32 (34)			0.93					
72 (63)	16 (76)	56 (60)			0.17						
Final diagnosis, n (%)											
Purpura fulminans and others ^c	49 (43)	14 (67)	35 (38)	1	-	-	1	-	-		
Meningitis	65 (57)	7 (33)	58 (62)	0.30	0.11–0.85	0.02	0.31	0.10–0.98	0.047		
Bacteria involved, n (%)											
<i>Streptococcus pneumoniae</i>	31 (27)	3 (14)	28 (30)								
<i>Neisseria meningitidis</i>	54 (47)	12 (57)	42 (45)								
Other ^d	11 (10)	4 (19)	7 (8)			0.17					
No documentation with purpura fulminans	18 (16)	2 (10)	16 (17)								
10 (55)	1 (50)	9 (56)									
CARE PATHWAYS											
First medical contact											
GP or emergency physician, n (%)	104 (91)	15 (71)	89 (96)	1	-	-	1	-	-		
Mobile medical unit	10 (9)	6 (29)	4 (4)	8.9	2.05–38.6	<0.001	8.72	1.76–43.28	0.008		
No of medical contacts, n (%)											
1	68 (60)	15 (71)	53 (57)								
2	37 (32)	6 (29)	31 (33)			0.43					
>2	9 (8)	0	9 (10)								
QUALITY OF CARES											
No. of suboptimal care, by children											
Median [IQR] ^e	1 [0–1]	1 [0–2]	0 [0–1]	1.55	1.06–2.26	0.025	1.65	1.07–2.54	0.022		
0	55 (48)	7 (33)	48 (52)								
1	31 (27)	5 (24)	26 (28)								
2	17 (15)	5 (24)	12 (13)								

Table 1. Cont.

	Total n = 114 (%)		Dead n = 21 (%)		Alive n = 93 (%)		Univariate analysis			Multivariate analysis [§]		
							OR	95% CI	P	aOR	95% CI	P
3	5 (4)	1 (5)	4 (4)									
4	6 (5)	3 (14)	3 (3)									
No suboptimal care/no care, % [95% CI]	15 [12–18]	24 [16–32]	13 [10–16]						0.003			
Care types												
Parental care, n (%)												
<i>Seeking medical care</i>												
Suboptimal	23 (20)	6 (29)	17 (18)				1.79	0.60–5.34	0.29			
Optimal	91 (80)	15 (71)	76 (82)				1	-	-			
Medical care, n (%)												
<i>Evaluation of severity</i>												
Suboptimal	23 (20)	7 (33)	16 (17)				2.40	0.82–7.03	0.10			
Optimal	91 (80)	14 (67)	77 (83)				1	-	-			
<i>Antibiotic therapy timing</i>												
Suboptimal	27 (24)	5 (24)	22 (24)				1	0.33–3.08	0.98			
Optimal	87 (76)	16 (76)	71 (76)				1	-	-			
<i>Saline bolus timing</i>												
Suboptimal	14 (12)	5 (24)	9 (10)				2.92	0.84–10.1	0.08			
Optimal	100 (88)	16 (76)	84 (90)				1	-	-			
<i>Saline bolus dosage</i>												
Suboptimal	12 (11)	7 (33)	5 (5)				8.80	2.23–34.7	0.002			
Optimal	102 (89)	14 (67)	88 (95)				1	-	-			
<i>Antibiotic therapy dosage</i>												
Suboptimal	5 (4)	0 (0)	5 (5)				0.67	0.01–6.03	0.29			
Optimal	109 (96)	21 (100)	88 (95)				1	-	-			

aOR, adjusted odds ratio; 95% CI, 95% confidence interval; IQR, interquartile range.

[§]Logistic regression model.

^aAge and no. of suboptimal care were treated as continuous variables (no deviation to linearity).

^bSeverity signs were hemodynamic failure, purpura, consciousness impairment, respiratory distress, meningism, behavioural changes or hypotonia.

^cOthers were 2 pneumoniae with pleural effusion and a septic shock following pyelonephritis in a child with malformative uropathy in the deceased group, and 2 septic shock on bacterial cellulitis and a bacterial tracheitis in the survivor group.

^dOthers were, for survivors, *Haemophilus influenzae* (n = 3), Group B *Streptococcus* (n = 1), *Staphylococcus aureus* (n = 1), and for deceased children, *E.coli* (n = 1), Group A *Streptococcus* (n = 1), *Salmonella spp* (n = 1) and *Mycoplasma pneumoniae* (n = 1).

doi:10.1371/journal.pone.0107286.t001

Table 2. Risk factors for medical suboptimal care.

	Optimal n = 489 (%)		Suboptimal n = 81 (%)		Univariate analysis			Multivariate analysis **, **		
					OR	95% CI	P	aOR	95% CI	P
Age										
<1 yr	125 (26)	35 (43)	1					1		
1–2	95 (19)	10 (12)	0.38	0.18–0.81	0.009		0.32	0.11–0.98	0.046	
2–5 yr	119 (24)	16 (20)	0.48	0.25–0.92	0.02		0.37	0.14–0.98	0.045	
≥5 yr	150 (31)	20 (25)	0.48	0.26–0.87	0.01		0.24	0.09–0.64	0.004	
Physician qualification, n (%)										
General practitioner	55 (11)	27 (33)	1					1		
Adult emergency	16 (3)	7 (9)	0.90	0.33–2.44	0.82		0.63	0.15–2.62	0.53	
Pediatric emergency	322 (66)	37 (46)	0.23	0.13–0.42	<0.001		0.16	0.08–0.35	<0.001	
Mobile medical unit	83 (17)	6 (7)	0.15	0.05–0.40	<0.001		0.09	0.03–0.31	<0.001	
Pediatric ward	13 (3)	4 (5)	0.63	0.18–2.13	0.45		0.65	0.11–3.67	0.63	
Severity signs at first consultation, n (%)										
No	182 (87)	28 (13)	1					1		
Yes	307 (85)	53 (15)	1.12	0.68–1.84	0.6		1.3	0.59–2.90	0.51	
Final diagnosis, n (%)										
Other	210 (86)	35 (14)	1					1		
Meningitis	279 (86)	46 (14)	0.99	0.62–1.59	0.9		0.73	0.34–1.59	0.43	

*Multivariate analysis involved a hierarchical logistic regression model with random intercept and effects.

**Significant associations remained when age was transformed into polynomials (X = 10/[age – 2.5]), aOR for age 1.04, 95% CI 1.01–1.07, p = 0.003. doi:10.1371/journal.pone.0107286.t002

Factors associated with outcome

As compared with children who died, survivors were more frequently older than 5 years ($p < 0.05$; Table 1) and diagnosed as having meningitis (62% vs 33%). The two groups did not differ in other demographic, clinical or bacteriologic characteristics or total number of medical contacts before admission to the ICU ($p > 0.1$, Table 1). For children who died, the first medical contact was frequently a mobile medical unit (vs GP office or hospital emergency department): 29% vs 4% ($p < 0.001$). The proportion of suboptimal care among all care types during the initial management was higher for children who died than survived: 24% vs 13% (95% CI of the risk difference: 9–13%). On univariate analysis, insufficient saline bolus was significantly associated with death (OR = 8.8; 95% CI: 2.23–34.7), under-evaluation of severity and delay to administer saline bolus was associated but not significantly with death (OR = 2.74; 95% CI: 0.82–7.03 and 2.92; 95% CI: 0.84–10.1 respectively) (Table 1).

After adjustment for confounders, each suboptimal care (continuous variable, no deviance to linearity) increased the odds of death by 65% (adjusted odds ratio [aOR] 1.65, 95% CI 1.08–2.54, $p = 0.02$) (Table 1). Each year of age (continuous variable, no deviation to linearity) decreased the odds of death (aOR 0.82, 95% CI 0.68–0.99, $p = 0.04$), as did having meningitis as compared with other diagnoses (aOR 0.31, 95% CI 0.10–0.98, $p = 0.047$). A first medical contact by the mobile medical unit was associated with an adverse outcome (aOR 8.72, 95% CI 1.76–43.28, $p = 0.008$) (Table 1).

Determinants of optimality of medical care

Among the 570 cares received by children during their initial management (Table 2), the repartition of suboptimal care differed by physician qualification. The proportion of suboptimal care was 33% for those provided by a GP, 30% for those in adult emergency settings, 24% for those in pediatric wards, 10% for those in pediatric emergency care and 7% for those in the mobile medical unit ($p < 0.001$). The proportion of under-evaluation of severity was 30% for a GP, 9% for pediatric emergency care and 0% for the mobile medical unit ($p = 0.001$). The proportion of delayed antibiotic therapy was 50% for a GP, 20% for pediatric emergency care and 0% for the mobile medical unit ($p = 0.02$). The other types of care (antibiotic therapy dosage and timing and dosage of saline bolus) did not differ by physician qualification.

On univariate analysis, younger children (<1 year) were at increased risk of suboptimal care (see Table 2), and odds of suboptimal care were lower with pediatric emergency or mobile medical unit care than GP care (OR 0.23, 95% CI 0.13–0.42; and OR 0.15, 95% CI 0.05–0.40, respectively) (Table 2). We found no association between final diagnosis or presence of severity sign at first medical contact and quality of care. The optimality of care varied significantly between children (i.e., children effect, empty model, $p < 0.001$). After adjustment in a multilevel multivariate model, the association between optimality of care and age of children (dichotomised in 4 classes or transformed in polynomials) and physician qualification remained stable (Table 2). The variance of the empty model was 1.56, that of the model with a level 1 variable (physician qualification) was 1.54 and that of the full model (level 1 and 2 variables) was 1.21. Level 1 variables explained 2% of the variance, whereas level 1 and 2 variables explained 21% of the variance.

Discussion

We found a strong association of suboptimal medical care and death for children with an SBI: each suboptimal care increased the

odds of death by 65%. Some types of care, particularly the dosage of saline bolus, were associated more with death than others. The gold standard to demonstrate causality in medical research is a controlled double blind randomised trial, but such studies are obviously not ethical in the case of SBI. Observational study analysis of causal association requires being aware of the risk of bias. [18] Here, we studied the determinants and consequences of suboptimal care in the early management of SBI in pediatric patients using an adequate approach to deal with the structure and type of data, including multilevel analysis and fractional polynomials [19,20,21] while minimizing the selection bias for children who died by using a population recruitment pattern with exhaustivity checking. As recommended by methodological standards, [22] suboptimality of care was evaluated by two independent experts who were blinded to the final diagnosis and outcome, with an overall high level of agreement between experts.

The strength of the significant association between suboptimal care and death remained nearly unchanged after adjustment for potential confounders: age of children, final diagnosis and initial severity of disease (represented as having a first medical contact by a mobile medical unit). However, some variables were inaccurately measured and/or some explanatory variables were lacking in the model. Indeed, we show a gap between the number of variables in the theoretical causal diagram and the one used (Figure S1 and Figure S2). For example, we could have explained more accurately the risk of death by considering genetic susceptibility to infection or bacterial virulence. [23,24] We were also limited in the evaluation of initial severity of the infection because of the retrospective design of the study. First medical contact by a mobile medical unit is an objective and reliable evaluation of clinical severity because in France, the mobile medical unit aims to care for children with the most severe disease who could not be transported to the hospital before receiving emergency care. Nevertheless, the presence of a severity sign at the first medical examination is a more arguable reflection of intrinsic severity because of the retrospective design of the study. For example, data on vital signs at the first medical contact, which are a key point to evaluate clinical severity in children in the context of SBI, [25] were sometimes missing, which led to inaccurate evaluation of severity of the disease for statistical analysis and also difficulty for the expert to accurately assess the optimality of the severity evaluation. Experts evaluated only misinterpretation of vital signs when they were mentioned. Here, we demonstrated the significant and independent global effect of suboptimal care on outcome, but we cannot affirm that suboptimal care in the early management was directly responsible for death because suboptimal care in the PICU was not assessed. Moreover, we analyzed only six types of care for each management and not all types. We could not examine the time effect and the total number of care. The total number of care could be the result of intrinsic severity (severely ill children requiring more care and sometimes showing a fulminant evolution) or the result of previous suboptimal care (inadequate care could lead to worsened disease, which then requires more care). The time effect could have been considered in a marginal structural model (Figure S1), but such a model requires timely detailed information that cannot be obtained with a retrospective study. [26]

We did not include children with SBI who survived but were not hospitalized in a PICU. It could be argued that we over-evaluated the frequency of suboptimal care because children with SBI admitted to a PICU may have received more suboptimal care, which caused clinical worsening and then admission to a PICU as compared with children who received adequate care and would not have required admission to a PICU. Thus, this selection bias

could have led to an under-evaluation of the association of suboptimal care and death because these children not hospitalized in a PICU and having received potentially more optimal care would most probably have survived.

The generalization of our results may be limited because the bacterial epidemiology may have changed since the study period. In France, conjugate vaccines against *Haemophilus influenzae*, *Neisseria meningitidis C* and *Streptococcus pneumoniae* with 7 and 13 valences were routinely recommended for all children by health authorities in 1992, 2009, 2002 and 2009, respectively. Invasive infection due to *H. influenzae* had almost disappeared during the study period. Reported cases of invasive pneumococcal infection decreased after vaccination introduction for only children younger than 2 years old. Incidences were 29 per 100 000 in 2001, 25 per 100 000 in 2004 and 18 per 100 000 in 2012 (meningitis and bacteremia). No significant changes were observed for older children. [27] Vaccine against meningococcus C had a too low coverage in the pediatric population to evidence a decrease in invasive infection due to *N. meningitidis C* since study period. [28] Thus, since the study period, the pattern of SBI may have changed for invasive pneumococcal infection in children less than 2 years old.

We did not observe a significant decrease in suboptimality of care across the years even though French recommendations concerning immediate administration of antibiotic therapy with purpura fulminans were largely diffused in 2000 and the Surviving Sepsis campaign began in 2003. [29,30] This finding highlights that simple diffusion of written recommendations are not enough to quickly modify practices of a large healthcare professional public. [31]

The analysis of suboptimal care determinants allowed us to identify potential targets for corrective actions. Young age (<1 year) was independently associated with increased risk of suboptimal care, whereas being under the care of a paediatric emergency specialist or a mobile medical unit physician was associated with reduced risk. Similar conclusions were reached by Dhamar *et al.* in a retrospective review of the quality of care received by 304 children with serious illnesses receiving treatment in 5 emergency departments in California between 2000 and 2003: after adjustment for confounding factors with a hierarchical model, younger children were at increased risk of receiving suboptimal care, and quality of care was better when provided by pediatric emergency physicians as compared with a GP. [32] Young age of children also appeared to be a barrier to optimal management of critical illness in community hospitals according to a qualitative study. [33] Corrective actions should then target the GP, in training and established, and focus on clinical evaluation of the youngest children.

Seeking medical care was considered delayed in 20% of our cases and accounted for 22% of the suboptimal care. We could not study the determinants of this delay, but French parents were previously found to poorly recognise purpuric rash. [34] Nevertheless, methods to recognize purpuric rash are warranted for not missing severe bacterial infections. [35] Parents worrying about their child's health has also been identified as a good marker of severe infection, although this sign is often missing, even with severe infection. [36] The better understanding of why parents are worried or not could be helpful to optimize early detection of sepsis.

Conclusions

Thanks to an adequate strategy for data analysis, we showed a significant association of suboptimal care for children with SBI and death. We identified determinants that could be acted on to optimize early management of SBI in children and then hopefully reduce the incidence of death. Physicians who are in charge of febrile children should pay particular attention to children younger than 1 year and systematically evaluate vital signs (pulse, respiratory rate, consciousness, capillary refill) that allow for early recognition of severe sepsis. Physicians and parents could be warned via widely distributed flyers or even television, as has been efficient in United Kingdom by the meningitis research foundation. [37] Physicians who rarely experience vital emergency situations could also benefit from a simulated training program. [38]

Supporting Information

Figure S1 Theoretical causal diagrams between optimality of care and death reflecting time-dependance of exposure and confounding factors. a: summarized diagram with C representing confounding factors; E, exposition (optimality of care); F, risk factors for exposition (determinants of optimality); and Y, outcome (survival status). b: more complete diagram with H representing host factors (age, genetic and non-genetic susceptibility to infection); B, bacterial factors (type of infection, bacterial specie/serotype, virulence, inoculum); O, optimality of care; S, clinical severity; P, physician characteristics (qualification, clinical experience etc.); Pa, parent characteristics (educational/socioeconomic status, facility of access to health care systems etc.). Indices represent different time points (from 0 to k) (Inspired by Robins *et al*, *Epidemiology*, September 2000, Vol. 11 No. 5). (TIF)

Figure S2 “Realistic” causal diagram between optimality of care before admission to a pediatric intensive care unit (PICU) and death. This diagram was established with DAGitty considering available variables. [16] The green circle with triangle inside represents exposure; blue circle with stick inside, outcome; green circles, exposure ancestors; pink circles, confounding factors; pink vectors, biasing pathway; green vectors, causal pathways; grey vectors, ancestor pathway. Clinical phenotype was represented by final diagnosis, severity signs at the first medical contact and first medical contact by a medical mobile unit. (TIF)

Figure S3 Study flowchart. (TIF)

Acknowledgments

This article is dedicated to the memory of Ms Albertine Aouba, MD, C epiDe-Inserm, Centre d' epid emiologie sur les causes m edicales de d ec es, Le V esinet, France.

Author Contributions

Conceived and designed the experiments: EL CGL CDT JCR MC. Performed the experiments: AM RA. Analyzed the data: EL CGL EM TB MC. Contributed reagents/materials/analysis tools: EL CGL MC. Contributed to the writing of the manuscript: EL MC. Critical revision of the manuscript for important intellectual concept: CGL AM RA CDT JCR.

References

1. Hartman ME, Linde-Zwirble WT, Angus DC, Watson RS (2013) Trends in the epidemiology of pediatric severe sepsis. *Pediatr Crit Care Med* 14: 686–693.
2. Harnden A, Mayon-White R, Mant D, Kelly D, Pearson G (2009) Child deaths: confidential enquiry into the role and quality of UK primary care. *Br J Gen Pract* 59: 819–824.
3. Launay E, Gras-Le Guen C, Martinot A, Assathiany R, Blanchais T, et al. (2010) Suboptimal care in the initial management of children who died from severe bacterial infection: a population-based confidential inquiry. *Pediatr Crit Care Med* 11: 469–474.
4. Han YY, Carcillo JA, Dragotta MA, Bills DM, Watson RS, et al. (2003) Early reversal of pediatric-neonatal septic shock by community physicians is associated with improved outcome. *Pediatrics* 112: 793–799.
5. McIntyre PB, Macintyre CR, Gilmour R, Wang H (2005) A population based study of the impact of corticosteroid therapy and delayed diagnosis on the outcome of childhood pneumococcal meningitis. *Arch Dis Child* 90: 391–396.
6. Nadel S, Britto J, Booy R, Maconochie I, Habibi P, et al. (1998) Avoidable deficiencies in the delivery of health care to children with meningococcal disease. *J Accid Emerg Med* 15: 298–303.
7. Ninis N, Phillips C, Bailey L, Pollock JI, Nadel S, et al. (2005) The role of healthcare delivery in the outcome of meningococcal disease in children: case-control study of fatal and non-fatal cases. *BMJ* 330: 1475.
8. Hernan MA, Cole SR (2009) Invited Commentary: Causal diagrams and measurement bias. *Am J Epidemiol* 170: 959–962; discussion 963–954.
9. Ahrens KA, Schisterman EF (2013) A time and place for causal inference methods in perinatal and paediatric epidemiology. *Paediatr Perinat Epidemiol* 27: 258–262.
10. von Elm E, Altman DG, Egger M, Pocock SJ, Gotsche PC, et al. (2008) The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *J Clin Epidemiol* 61: 344–349.
11. Conseil supérieur d’hygiène publique Opinion from the French High Committee on Public Health, Mars 10, 2000. Available: http://www.sante.gouv.fr/htm/dossiers/cshpf/a_mt_100300_meningite_01.htm. Accessed 2007 Dec 22.
12. Health Protection Agency (2006) Guidance for public health management of meningococcal disease in the UK. Health Protection Agency Meningococcus Forum.
13. Société de Réanimation de Langue Française Use of catecholamines during septic shock (adults and children). XV consensus conference of the French Society of Intensive cares 1996. Available: <http://www.sfar.org/article/33/utilisation-des-catecholamines-au-cours-du-cho-c-septique-adultes-enfants-cc-1996>. Accessed 2014 Jan 8.
14. Goldstein B, Giroir B, Randolph A (2005) International pediatric sepsis consensus conference: definitions for sepsis and organ dysfunction in pediatrics. *Pediatr Crit Care Med* 6: 2–8.
15. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. *Biometrics* 33: 159–174.
16. Textor J, Hardt J, Knuppel S (2011) DAGitty: a graphical tool for analyzing causal diagrams. *Epidemiology* 22: 745.
17. Angus DC, van der Poll T (2013) Severe sepsis and Septic Shock. *N Engl J Med* 369: 840–851.
18. Launay E, Morfouace M, Deneux-Tharaux C, Gras le-Guen C, Ravaud P, et al. (2013) Quality of reporting of studies evaluating time to diagnosis: a systematic review in paediatrics. *Arch Dis Child* doi:101136.
19. Altman DG, Royston P (2006) The cost of dichotomising continuous variable. *BMJ* 332: 1080.
20. Ambler G, Royston P (2001) Fractional polynomial model selection procedures: investigation of type I error rate. *J Statist Comput Simul* 69: 89–108.
21. Snijders T, Bosker R (1999) *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London: Sage Publications Ltd.
22. Bouvier-Colle MH (2002) Confidential enquiries and medical expert committees: a method for evaluating healthcare. The case of Obstetrics. *Rev Epidemiol Sante Publique* 50: 203–217.
23. Casanova JL, Abel L (2007) Human genetics of infectious diseases: a unified theory. *EMBO J* 26: 915–922.
24. Peterson JW (1996) Bacterial Pathogenesis. In: Baron S, editor. *Medical Microbiology* 4th edition. 2011/03/18 ed.
25. Thompson M, Mayon-White R, Harnden A, Perera R, McLeod D, et al. (2008) Using vital signs to assess children with acute infections: a survey of current practice. *Br J Gen Pract* 58: 236–241.
26. Robins JM, Hernan MA, Brumback B (2000) Marginal Structural Models and Causal Inference in Epidemiology. *Epidemiology* 11: 550–560.
27. Varon E, Janoir C, Gutmann L (2012) Centre national de référence du pneumocoque. Rapport d’activités 2013. Available: <http://www.cnr-pneumo.fr/docs/rapports/CNRP2013.pdf>. Accessed 2014 July 22.
28. Stahl J, Cohen R, Denis F, Gaudelus J, Lery T, et al. (2013) Vaccination against meningococcus C. Vaccinal coverage in the French target population. *Med Mal Infect* 43: 75–80.
29. Ministère des affaires sociales et de la santé Aide mémoire sur les infections invasives à méningocoque. Available: http://www.sante.gouv.fr/IMG/pdf/Annexe_-_Fiche_aide_memoire_sur_les_infections_invasives_a_meningocoques.pdf. Accessed 2014 Jan 8.
30. Slade E, Tamber PS, Vincent JL (2003) The Surviving Sepsis Campaign: raising awareness to reduce mortality. *Crit Care* 7: 1–2.
31. Giguere A, Legare F, Grimshaw J, Turcotte S, Fiander M, et al. (2012) Printed educational materials: effects on professional practice and healthcare outcomes. *Cochrane Database Syst Rev* 10: CD004398.
32. Dharmar M, Marcin JP, Romano PS, Andrada ER, Overly F, et al. (2008) Quality of care of children in the emergency department: association with hospital setting and physician training. *J Pediatr* 153: 783–789.
33. Gilleland J, McGugan J, Brooks S, Dobbins M, Ploeg J (2013) Caring for critically ill children in the community: a needs assessment. *BMJ Qual Saf*.
34. Aurel M, Dubos F, Motte B, Pruvost I, Leclerc F, et al. (2011) Recognising haemorrhagic rash in children with fever: a survey of parents’ knowledge. *Arch Dis Child* 96: 697–698.
35. Mant D, Van den Bruel A (2011) Should we promote the tumbler test? *Arch Dis Child* 96: 613–614.
36. Van den Bruel A, Haj-Hassan T, Thompson M, Buntinx F, Mant D (2010) Diagnostic value of clinical features at presentation to identify serious infection in children in developed countries: a systematic review. *Lancet* 375: 834–845.
37. Meningitis Research Fondation Meningitis research fondation’s website. Available: <http://www.meningitis.org>. Accessed 2014 July 22.
38. Katznelson J, Mills W, Forsythe C, Shaikh S, Tolleson-Rinehart S (2014) Project CAPE: a high-fidelity, in situ simulation program to increase critical access hospital emergency department provider comfort with seriously ill pediatric patients. *Pediatr Emerg Care* 30: 397–402.

Annexe 5 : Articles didactiques en français

Analyse des délais diagnostiques en pédiatrie : pièges méthodologiques

E. Launay^a, M. Morfouace^b, J.-F. Brasme^b, C. Gras-le Guen^a, M. Chalumeau^{b,c,*}

^a Clinique médicale et urgences pédiatriques, Hôpital Mère-Enfant, 44000 Nantes

^b Inserm U953, hôpital Saint-Vincent-de-Paul, 75014 Paris

^c Service de pédiatrie générale, hôpital Necker, AP-HP ; Université Paris-Descartes, 75015 Paris

Les articles rapportant des études ayant analysé les délais diagnostiques d'une maladie doivent suivre les référentiels méthodologiques usuels des études épidémiologiques : STROBE [1] et REMARK [2]. Certains éléments sont cependant particulièrement importants ou spécifiques au problème des délais diagnostiques : biais de sélection, calcul des délais, étude de la relation délai-gravité et spécificité des conclusions pour la pratique.

1. Biais de sélection

Les études portant sur les délais diagnostiques sont très sensibles aux biais de sélection et notamment sur trois points.

Premièrement, un recrutement strictement hospitalier (ou pire dans un centre de référence) va biaiser l'évaluation des délais diagnostiques en raison de la gravité des patients et de la formation des professionnels organisés dans un éventuel réseau. Les résultats de telles études ne seront que difficilement extrapolables. Une alternative à ce type de biais est de se rapprocher le plus possible d'un recrutement « en population » c'est-à-dire de s'assurer de l'exhaustivité des cas durant une période et dans une zone géographique définie, ce qui peut être rendu possible par l'existence d'un registre [3] ou de données médico-administratives enregistrées en routine comme celles du CepiDC [4].

Deuxièmement, il est important de rester conscient que les études sur les délais diagnostiques ne peuvent analyser que les cas ... diagnostiqués. Il faut donc réfléchir à l'éventualité de formes létales ayant entraîné un décès avant diagnostic ou de formes mineures ayant pu ne pas être diagnostiquées. Ce risque dépend donc de l'histoire naturelle de la maladie, du caractère systématique et sensible des procédures diagnostiques appliquées aux patients potentiellement atteints de cette maladie durant la période d'étude. Par exemple, des auteurs analysant le délai diagnostique des accidents vasculaires de l'enfant ne discutent pas la possibilité de non diagnostic de formes mineures ou de formes massives très rapidement létales avant diagnostic [5]. La description de la méthode d'identification des patients potentiellement éligibles est donc très importante et, si les patients éligibles ne sont que les patients diagnostiqués, l'interprétation des résultats devra prendre en compte ce biais de sélection potentiel.

Troisièmement, il est important d'identifier et de prendre en compte (comme un critère d'exclusion ou d'analyse en sous-groupe) des populations à haut risque bénéficiant d'un dépistage systématique ou d'une surveillance spécifique qui vont modifier le délai diagnostique. Cela a été fait par exemple dans l'étude des délais diagnostiques des médulloblastomes de l'enfant, où les patients atteints d'une prédisposition génétique aux cancers ont été exclus [3].

2. Définition des délais

Le délai diagnostique est l'intervalle séparant les premiers symptômes qui auraient pu permettre de faire un diagnostic d'une maladie du moment où il est fait. Pour calculer un délai diagnostique il est donc nécessaire de définir un To correspondant aux premiers symptômes compatibles avec le début de la maladie. Cette définition est facile s'il existe un consensus de la communauté médicale. Par exemple, pour le calcul des délais diagnostiques du déficit en hormone de croissance avec interruption de tige, des auteurs ont utilisé comme To l'apparition des critères auxologiques de la *Growth Hormone Research Society* [6]. Si les symptômes d'alerte sont plus insidieux ou peuvent se mélanger avec ceux d'une affection intercurrente ou pré-existante, la détermination du To peut être délicate. Par exemple, pour la détermination du To des infections bactériennes mortelles, les symptômes d'une virose précédant l'infection bactérienne ont pu rendre la détermination du To difficile [4]. La définition du To du diagnostic des médulloblastomes a pu, elle, être compliquée par la préexistence d'un strabisme [3]. Une solution est alors de faire évaluer indépendamment ce délai par deux personnes, si possible de formation et de type d'exercice éloignés, et de trancher les cas discordants par un troisième expert ou par discussion. Le calcul d'un coefficient d'accord entre les juges permettra d'évaluer de la robustesse de l'évaluation des délais [4].

Une autre difficulté dans le calcul du délai diagnostique peut apparaître si celui-ci ne peut être évoqué que sur l'accumulation d'arguments cliniques et paracliniques. Le piège est alors de surévaluer le délai diagnostique en considérant le moment du diagnostic comme celui du premier élément apparu. Par exemple, des auteurs ont étudié le délai diagnostique du déficit immunitaire commun variable avec un To au premier épisode infectieux, le plus souvent une bronchite [7]. Si on peut légitimement attribuer ce

* Auteur correspondant.
e-mail : martin.chalumeau@nck.aphp.fr

premier épisode au déficit immunitaire, il n'est en revanche pas possible que ce premier épisode définisse le To qui doit être défini par la récurrence d'épisodes infectieux.

3. Relation délais / gravité

Les auteurs rapportant l'analyse des délais diagnostiques d'une affection sont souvent enclins, après avoir rapporté des délais longs, à essayer de démontrer que ces délais sont associés à une issue défavorable pour le patient en termes de gravité initiale de la maladie, de survie, ou de séquelles. La validité de cette démonstration dépend premièrement d'un calcul du délai diagnostique à l'aveugle de l'issue. Par exemple, dans une étude de la relation entre le délai diagnostique et la gravité initiale et la survie des médulloblastomes de l'enfant [3], le calcul des délais est décrit comme difficile pour certains patients présentant un strabisme ancien. Il est alors important que ce calcul soit fait à l'aveugle de l'issue afin d'éviter d'introduire un biais de classement différentiel qui induirait une fausse association significative. Deuxièmement, il est important que les facteurs de confusion soient pris en compte. Dans le même exemple des médulloblastomes, un délai diagnostique long était associé à une survie plus longue, mais après ajustement sur un âge élevé (variable associée à la fois à un délai plus long et une survie plus longue) cette relation protectrice paradoxale disparaissait [3].

4. Propositions pour réduire les délais

La spécificité et la faisabilité des recommandations pour la pratique clinique souvent proposées dans la discussion des articles analysant les délais diagnostiques doivent être évaluées de manière critique. Les auteurs des études ont souvent basé leur analyse sur une série de patients atteints d'une affection rare et grave et font souvent des recommandations pour la pratique clinique pour un ensemble de patients présentant un des symptômes qui peut aboutir au diagnostic de cette affection ou d'autres plus bénignes. Par exemple, des auteurs analysant le délai diagnostique des accidents vasculaires de l'enfant indiquent, dans un article publié en 2009 dans la revue *Pediatrics* : « *Stroke*

should be suspected for any neonate presenting with apnea » [5]. La spécificité et la faisabilité d'une telle proposition mériteraient probablement une évaluation critique.

5. Conclusion

L'analyse des délais diagnostiques nécessite de suivre une méthodologie rigoureuse si l'objectif recherché est l'amélioration de l'état de santé des enfants via la diminution de ces délais. La production d'un référentiel complétant STROBE et REMARK offrirait une ligne directrice aux chercheurs et aux cliniciens pour mener rigoureusement ce type d'analyse.

Références

- [1] Von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *J Clin Epidemiol* 2008;61:344-9.
- [2] McShane LM, Altman DG, Sauerbrei W, et al. REporting recommendations for tumor MARKer prognostic studies (REMARK). *Nat Clin Pract Urol* 2005;2:416-22.
- [3] Brasme J-F, Grill J, Doz F, et al. Diagnosis delay of medulloblastoma in children: distribution, determinants and consequences in a population-based study. 14th International Symposium on Pediatric Neuro-Oncology, Vienna. 2010.
- [4] Launay E, Gras-Le Guen C, Martinot A, et al. Suboptimal care in the initial management of children who died from severe bacterial infection: a population-based confidential inquiry. *Pediatr Crit Care Med* 2010;11:469-74.
- [5] Srinivasan J, Miller SP, Phan TG, et al. Delayed recognition of initial stroke in children: need for increased awareness. *Pediatrics*. 2009;124:e227-34.
- [6] Gascoin-Lachambre G, Brauner R, Duche L, et al. Pituitary stalk interruption syndrome: diagnostic delay and sensitivity of the auxological criteria of the Growth Hormone Research Society. *PLoS One* 2011;6:e16367.
- [7] Urschel S, Kayikci L, Wintergerst U, et al. Common variable immunodeficiency disorders in children: delayed diagnosis despite typical clinical presentation. *J Pediatr* 2009;154:888-94.

Méthodologie de l'analyse des soins sub-optimaux

Methodological approaches to study sub-optimal care in children

E. Launay¹, C. Gras-le Guen¹, A. Martinot², M.-H. Bouvier-Colle³, M. Chalumeau^{3,4*}

¹Clinique médicale et service d'urgences pédiatriques, Hôpital Mère-Enfant, 8 quai Moncoussu, 44000 Nantes, France

²Unité d'urgences pédiatriques et de maladies infectieuses, Hôpital R. Salengro, CHU de Lille, Université de Lille-2, 2, avenue Oscar-Lambret, 59037 Lille, France

³Inserm U149, hôpital Saint-Vincent-de-Paul, 74, avenue Denfert-Rochereau, 75014 Paris, France

⁴Service de pédiatrie générale, hôpital Saint-Vincent-de-Paul, AP-HP, Université Paris-Descartes, 74, avenue Denfert-Rochereau, 75014 Paris, France

Disponible en ligne sur
 ScienceDirect
www.sciencedirect.com

Mots clés : Soins sub-optimaux,
Mortalité-morbidité, Audit clinique, Enquête
confidentielle avec comité d'experts

Les industries « à risque » comme l'aéronautique ou le nucléaire civil ont développé depuis des décennies des méthodes pour réduire les accidents dont les conséquences sont souvent catastrophiques en termes de vies humaines. La principale méthode utilisée est appelée AMDE (analyse des modes de défaillance et de leurs effets) et se décompose en quatre grandes étapes : recensement des défaillances, recherche des causes, étude des effets et recensement des moyens de détection [1]. L'objectif de cette méthode est de définir un plan d'action pour éviter à l'avenir des accidents, tout en maintenant une analyse de l'impact des actions menées pour les ajuster au fur et à mesure. La médecine est une discipline où les défaillances peuvent aussi aboutir à des catastrophes (en termes de décès ou de morbidité), comme dans les industries « à risque ». Ainsi, un rapport de l'*Institute of Medicine* publié en 2000 indiquait que les soins sub-optimaux étaient entre la 3^e et la 8^e cause de mortalité aux USA et suggérait une analyse systémique de ces soins sub-optimaux afin d'améliorer l'état de santé global de la population [2]. Une telle entreprise nécessite, comme pour l'industrie, une méthodologie adaptée aux particularités de la discipline, ici médicale.

1. La nécessité de critères explicites

En 1996, Westerling et al. ont analysé 65 études des facteurs de mortalité évitable dans différents domaines de la médecine et ont souligné la disparité des méthodes utilisées pour évaluer l'évitabilité des décès avec pour conséquence des difficultés d'interprétation des résultats et/ou une impossibilité de comparaison [3]. Ces difficultés ont abouti à une controverse sur les conclusions du rapport de l'*Institute of Medicine* sur l'évitabilité des décès [4]. Ainsi la démarche d'évaluation des soins nécessite au préalable le choix d'un référentiel de soins dits « optimaux ». Ceux-ci doivent être fondés dans la mesure du possible sur le plus haut niveau de preuve reconnu par les professionnels de santé établi à partir de données récentes, validées, tenant compte de la réglementation

et selon des critères concrets et observables, ou à défaut sur un accord professionnel [1].

2. Trois principaux types d'approche

Les démarches d'amélioration de la qualité des soins existent dans tous les établissements de santé et souvent une approche empirique est suffisante. Cependant, certains problèmes, notamment liés à des exigences de qualité et de sécurité très élevées, requièrent le recours à des méthodes structurées. Ces différentes méthodes et outils d'amélioration de la qualité ont été décrits en détail par l'ANAES puis l'HAS [1]. Nous présenterons très synthétiquement deux des principales méthodes utilisables couramment dans le cadre de l'activité de soin (la revue de mortalité-morbidité et l'audit clinique) et nous détaillerons celle correspondant plutôt à une activité de recherche (l'enquête confidentielle).

2.1. Revue de mortalité-morbidité (RMM)

La RMM a pour objectif l'analyse, le plus souvent au sein d'un service, de tous les décès et de certains accidents morbides pré-déterminés ou non. L'analyse se fait avec l'ensemble des médecins du service et éventuellement ceux d'autres équipes et des membres du personnel non médical. L'objectif est de « porter un regard critique sur la façon dont le patient a été pris en charge, de s'interroger sur le caractère évitable de l'événement et rechercher collectivement les causes de la défaillance survenue lors de la prise en charge » [1]. En pratique, chaque dossier de patient est présenté à la RMM après qu'un membre de l'équipe en ait réalisé la synthèse. Le groupe doit rechercher des actions d'amélioration de la qualité des soins afin d'éviter que l'évènement indésirable ne se reproduise.

2.2. Audit clinique

L'audit clinique est une méthode d'évaluation des pratiques par comparaison à des références admises. Sa principale caractéristique est de mesurer les écarts entre la pratique observée et la pratique attendue [1]. Bien que l'audit puisse être global et alors souvent mené par des pairs extérieurs, il est recommandé de réaliser des audits sur

* Auteur correspondant.
e-mail : martin.chalumeau@svp.aphp.fr

des pathologies ou des actes homogènes avec des critères pré-définis [1]. L'audit clinique permet d'évaluer la prise en charge globale d'un patient, un thème transversal ou encore le fonctionnement d'un service. L'audit se déroule en 6 étapes : choix du thème, choix des critères, choix de la méthode de mesure, recueil des données, analyse des résultats et plan d'actions d'amélioration et réévaluation. Bien que de nombreuses publications médicales se réfèrent à la méthodologie de l'audit, celui-ci ne devrait pas avoir de visée de recherche notamment car ses résultats ne sont pas généralisables à d'autres services [5]. Cependant, lorsque l'audit prend un caractère national, il se rapproche d'une démarche scientifique. Par exemple, une équipe londonienne a audité prospectivement les pratiques de prise en charge du sepsis chez l'enfant en comparaison avec les recommandations du *Pediatric Advanced Life Support manual* sur l'ensemble du territoire anglais [6]. Parmi les 200 enfants étudiés, 62 % n'avaient pas reçu les soins recommandés dans les référentiels en termes de remplissages vasculaires et d'inotropes.

2.3. L'enquête confidentielle avec comité d'experts (ECCE)

L'ECCE est l'étude systématique d'un événement grave précisément défini, réalisée à distance de l'évènement, pour en faire l'analyse médicale objective et notamment pour étudier précisément le parcours de soins d'un patient [7]. Le caractère confidentiel est assuré vis-à-vis des personnes (malades et soignants) et des structures de soins impliquées, et durant l'analyse de l'optimalité des soins réalisée par comparaison à des soins de références [7]. Cette analyse de l'optimalité des soins est menée par un groupe de spécialistes (comité d'experts) extérieurs à l'évènement mais concernés par la question (pairs) [7]. Comme pour les méthodes précédentes, l'objectif est d'établir un bilan, tirer des leçons et émettre un avis ou des recommandations destinés à un large public médical et de décideurs de santé publique pour à l'avenir améliorer les soins [7]. L'un des apports spécifiques de l'ECCE est son caractère didactique car elle permet à des soignants, qui sont par définition peu ou pas confrontés à des événements rares, de se préparer à affronter ces situations. Cette méthode a été développée historiquement dans le domaine de l'obstétrique [7]. La lourdeur de la mise en place d'enquête confidentielle ne permet de les envisager que pour des événements rares et graves, même hétérogènes, et dans le cadre de programme de recherche et de politiques nationales comme cela est le cas pour les décès maternels.

3. Quelques aspects méthodologiques des enquêtes confidentielles

3.1. Le support

L'ECCE sur les décès maternels se base sur un questionnaire détaillé pré-défini visant à reconstituer le parcours d'une patiente, associé à une copie anonymisée des éléments clefs du dossier médical. Ce support est préparé par un binôme anesthésiste-obstétricien après entretien avec les acteurs sur le lieu du décès [7]. Ce système évite la perte et la déformation des informations pouvant se produire lors de la retranscription des dossiers médicaux en fiche de synthèse comme cela a été fait par exemple dans une étude anglaise sur l'évaluation de la prise en charge des méningococcies chez l'enfant [8].

3.2. Indépendance des experts

Les experts, d'origines professionnelles diverses, pourront être regroupés en panel [7, 8], les conclusions seront alors collégiales après discussion autour du dossier, comme dans l'ECCE sur les

décès maternels. Cela permet une bonne cohérence de l'expertise à condition que la discussion soit libre entre les experts sans « pression hiérarchique » entre eux. Les experts peuvent également être indépendants les uns des autres, pour éviter l'influence des experts sur leurs évaluations respectives. Cependant, pour être concluant, ce système demande une bonne cohérence entre les experts qui peut être améliorée a priori par un *training* sur quelques dossiers et par la formalisation des critères qui définiront les soins sub-optimaux, et être évaluée a posteriori par un coefficient d'accord.

3.3. Évolution et issue à l'aveugle

Il a été montré que la connaissance de l'évolution et de l'issue pouvait influencer sur le jugement de l'optimalité des soins [9]. Afin d'être le plus objectif possible, il convient donc de laisser si possible les experts à l'aveugle de l'évolution et de l'issue. En ce qui concerne l'issue, il faut réfléchir à cet aveugle très tôt, dès les premiers contacts avec les experts, qui ne doivent pas savoir s'ils analysent des issues favorables ou défavorables pour les patients. Quant à l'aveugle concernant l'évolution, les événements cliniques peuvent être dévoilés aux experts heure par heure afin qu'ils rendent leurs conclusions sur l'optimalité des soins au fur et à mesure. Cela a par exemple été fait dans l'étude de Ninis et al. sur la prise en charge des méningococcies de l'enfant en Grande-Bretagne [8].

4. Préparer l'action

L'objectif de la démarche d'analyse des soins sub-optimaux est l'amélioration de la santé des populations par l'amélioration des soins délivrés principalement par des actions de formation (des parents ou des soignants) toujours longues et difficiles. Il est donc nécessaire d'identifier les cibles pertinentes, c'est-à-dire des soins sub-optimaux « assez » fréquents et dont on peut espérer que la correction sera associée à une amélioration de la mortalité ou de la morbidité.

4.1. Identifier une typologie de soins sub-optimaux

Il est donc important dans un premier temps de classer les soins sub-optimaux pour en identifier une typologie. L'*Institute of Medicine* propose une classification permettant de distinguer les défaillances diagnostiques, thérapeutiques, préventives ou encore logistiques [2, 10]. Cependant la distinction entre ces différents types de soins sub-optimaux n'est pas toujours facile a posteriori et elle est en partie empirique. Par exemple, l'absence de remplissage vasculaire adéquat devant des signes de sepsis sévère doit-elle être considérée comme une sous-évaluation de la gravité de l'état clinique de l'enfant (sepsis non diagnostiqué) ou un soin sub-optimal de nature thérapeutique (sepsis diagnostiqué mais insuffisamment traité) ? Ces différentes possibilités ciblent des mécanismes de défaillance et donc des actions de corrections différents. La distinction entre ces mécanismes suppose que les histoires cliniques et les actions thérapeutiques soient le plus détaillées possibles dans les dossiers évalués et dans l'idéal que les soignants ayant exécuté ces actions soient re-interviewés a posteriori ce qui pose des problèmes éthiques.

4.2. Étudier la relation soins sub-optimaux/issue

Dans un deuxième temps, il faut essayer d'apporter des éléments de preuve dans la relation causale entre la présence de soins sub-optimaux et l'issue défavorable pour le patient. La prudence est ici nécessaire car seule une méthodologie expérimentale de type

essai randomisé permettrait d'établir une relation causale, ce qui est éthiquement impossible. Une approche observationnelle par comparaison des groupes à issue favorable ou défavorable peut être utilisée mais elle est sujette à des biais principalement liés à l'état de gravité initiale. Ces biais peuvent être pris en compte en partie soit a priori par des méthodes d'appariement, soit a posteriori par des méthodes d'ajustement si des critères de gravité initiaux sont connus, comme cela a été fait dans l'étude de Ninis et al. où les patients étaient appariés selon un score de gravité [8].

5. Conclusion

Comme pour toutes les études épidémiologiques observationnelles, la méthodologie de l'évaluation des soins sub-optimaux et notamment celle des études confidentielles avec comité d'experts est une recherche du moins mauvais compromis entre des contraintes opposées de rigueur scientifique et de faisabilité. Certains éléments sont cependant absolument nécessaires : respect total de la confidentialité, choix de référentiels indiscutables, adhésion des équipes soignantes et objectif d'amélioration des soins par des actions pratiques.

Références

1. Agence Nationale d'Accréditation et d'Évaluation en Santé. Méthodes et Outils des démarches qualité pour les établissements de santé. 2000. Disponible en ligne à l'adresse : <http://www.has-sante.fr/portail/upload/docs/application/pdf/methodes.pdf>. Consulté le 15/12/2008.
2. Committee on Quality of Health Care in America. Kohn LT, Corrigan J, Donaldson MS, Institute of Medicine. To err is human: building a safer health system. National Academy Press 2000.
3. Westerling R. Studies of avoidable factors influencing death: a call for explicit criteria. *Qual Health Care* 1996;5:159-65.
4. McDonald CJ, Weiner M, Hui SL. Deaths due to medical errors are exaggerated in Institute of Medicine report. *JAMA* 2000;284:93-5.
5. Wilson A, Grimshaw G, Baker R, et al. Differentiating between audit and research: postal survey of health authorities' views. *Br Med J* 1999;319:1235.
6. Inwald DP, Tasker R, Peters MJ, et al. Emergency management of children with severe sepsis in the United Kingdom – the results of the Paediatric Intensive Care Society sepsis audit. *Arch Dis Child* 2009 Published Online First: 8 January 2009. doi:10.1136/adc.2008.153064.
7. Bouvier-Colle MH. Enquêtes confidentielles avec comités d'experts : une méthode d'évaluation des soins. L'exemple de l'obstétrique. *Rev Epidemiol Santé Publique* 2002;50:203-17.
8. Ninis N, Phillips C, Bailey L, et al. The role of healthcare delivery in the outcome of meningococcal disease in children: case-control study of fatal and non-fatal cases. *Br Med J* 2005;330:1475-82.
9. Bradfield A, Wells GL. Not the same old hindsight bias: outcome information distorts a broad range of retrospective judgments. *Mem Cognit* 2005;33:120-30.
10. Rubin G, George A, Chinn DJ, et al. Errors in general practice: development of an error classification and pilot study of a method for detecting errors. *Qual Saf Health Care* 2003;12:443-7.

Disponible en ligne sur
 ScienceDirect
www.sciencedirect.com

Mots clés : Soins sub-optimaux,
Infections bactériennes

Soins sub-optimaux dans la prise en charge initiale des infections bactériennes sévères

Sub-optimal care during the initial treatment of severe
bacterial infection in children

E. Launay^{1,5}, A. Martinot², R. Assathiany³, G. Bréart⁴, M. Chalumeau^{4,5*},
C. Gras-le Guen¹

¹Clinique médicale et service d'urgences pédiatriques, Hôpital Mère-Enfant, 8, quai Moncoussu, 44000 Nantes, France

²Unité d'urgences pédiatriques et de maladies infectieuses, Hôpital R.-Salengro, CHU de Lille, Université de Lille-2, 2, avenue Oscar-Lambret, 59037 Lille, France

³AREPEGE, Association pour la Recherche et l'Enseignement en Pédiatrie Générale, Issy-les-Moulineaux, France

⁴Inserm U149, hôpital Saint-Vincent-de-Paul, 74, avenue Denfert-Rochereau, 75014 Paris, France

⁵Service de pédiatrie générale, hôpital Saint-Vincent-de-Paul, AP-HP, Université Paris-Descartes, 74, avenue Denfert-Rochereau, 75014 Paris, France

1. Pourquoi étudier les infections bactériennes sévères ?

En 1999, l'*Institute of Medicine* estimait que les soins sub-optimaux étaient entre la 3^e et la 8^e cause de mortalité aux USA et suggérait une analyse de ces soins sub-optimaux afin d'améliorer l'état de santé global de la population [1]. Malheureusement, ce rapport n'offrait pas une analyse des causes de décès liés à des soins sub-optimaux par groupe d'âge. Les infections bactériennes sévères (IBS) constituent un groupe d'affection qui est un bon candidat à l'étude des soins sub-optimaux en pédiatrie. En effet, les IBS sont l'une des premières causes de mortalité des patients de moins de 15 ans en France [2], et la deuxième cause de morbidité et de mortalité chez l'enfant aux USA [3]. Par ailleurs, les premiers symptômes des IBS sont peu spécifiques, se limitent souvent à une fièvre isolée notamment chez les plus jeunes patients, et pourraient donc être source de retards diagnostiques et thérapeutiques [4]. De plus, il existe des recommandations standardisées pour le diagnostic et le traitement de ces IBS qui ont été publiées et largement diffusées depuis de nombreuses années. Enfin, il a été démontré que la détection précoce et la rapidité de la prise en charge adaptée des IBS (notamment l'administration précoce d'une antibiothérapie adaptée et d'un soutien hémodynamique) réduisaient très significativement leur morbidité et leur mortalité chez l'adulte et l'enfant [5, 6].

2. Référentiels existants

L'évaluation de l'optimalité des soins nécessite au préalable la définition de soins de référence. Dans le cas des IBS, il existe des référentiels nationaux et internationaux concernant le purpura

fulminans [7] et le sepsis de l'enfant [8]. Le conseil supérieur d'hygiène publique a ainsi envoyé une circulaire à tous les médecins français recommandant l'administration immédiate, devant « tout malade présentant des signes infectieux et à l'examen clinique, lorsqu'il a été totalement dénudé, un purpura comportant au moins un élément nécrotique ou ecchymotique de diamètre supérieur ou égal à 3 millimètres », d'une première dose par voie parentérale d'un traitement antibiotique approprié aux infections à méningocoques (amoxicilline, céfotaxime ou ceftriaxone) [7]. Grâce à un comité international d'experts [8], le « sepsis » est maintenant clairement défini (présence d'au moins 2 des signes suivant dans un contexte de suspicion d'infection : tachycardie ou bradycardie, fièvre > 38,5° C ou hypothermie < 36° C, polypnée, hyperleucocytose ou leucopénie) ainsi que sa prise en charge (stabilisation hémodynamique par remplissage par des bolus de 20 ml/kg de cristalloïdes sur 5 à 10 minutes, vasopresseurs en cas d'échec et antibiothérapie adaptée dans l'heure suivant le diagnostic) [6]. Ces différents référentiels peuvent donc être une base à l'évaluation de l'optimalité des soins dans la prise en charge précoce des IBS de l'enfant.

3. Trois études anglo-saxonnes

Nadel et al. ont conduit une étude de cohorte prospective monocentrique dans un service pédiatrique londonien de 1992 à 1994 qui a évalué l'optimalité de la prise en charge initiale (avant l'arrivée en réanimation) des enfants atteints de méningococcie [9]. Les données étaient recueillies par l'interrogatoire direct des parents quelques heures après l'admission et par relecture des dossiers médicaux. Les auteurs ont préalablement défini, en utilisant les référentiels disponibles à l'époque, les symptômes et les signes cliniques de méningococcie qui devraient alerter les parents et les médecins, et les pratiques optimales une fois le diagnostic

* Auteur correspondant.
e-mail : martin.chalumeau@svp.aphp.fr

fait. Parmi les 54 enfants inclus, 30 % avaient un retard au premier recours médical en raison de la non-reconnaissance des signes d'alerte par les parents, 70 % de ceux qui avaient un purpura à la première consultation médicale n'avaient pas reçu d'antibiotique en pré-hospitalier et 54 % ont reçu un traitement hospitalier initial sub-optimal (remplissage inadapté...). Les auteurs concluaient à la fréquence élevée des soins sub-optimaux dans la prise en charge des méningococcies de l'enfant mais étaient très prudents sur une éventuelle relation causale avec l'issue pour ces patients.

La deuxième étude concernant les méningococcies a été conduite par la même équipe anglaise entre 1997 et 1999 [10] et visait plus particulièrement à évaluer les conséquences des soins sub-optimaux sur le devenir à court terme des enfants atteints de méningococcie. Cette étude nationale rétrospective a analysé les dossiers des enfants hospitalisés pour méningococcie et a comparé les enfants décédés à 3 témoins survivants à leur infection, appariés sur l'âge. Les dossiers médicaux anonymisés des enfants ont été soumis à 4 experts sous forme de retranscription heure par heure des symptômes et des constantes vitales pour les 24 premières heures après l'admission à l'hôpital, à l'aveugle de l'issue pour l'enfant. À chaque heure, en s'appuyant sur des référentiels de la littérature (par exemple le *Pediatric advance life support manual*), les experts devaient proposer leurs diagnostics, statuer sur les complications éventuellement présentes et suggérer *a posteriori* une prise en charge considérée comme optimale. Les experts devaient également définir le degré de gravité à l'admission en utilisant le *Glasgow meningococcal septicaemia prognostic score*, afin de permettre d'ajuster les résultats sur la gravité initiale des patients. L'évaluation des dossiers se faisait alors par comparaison entre la prise en charge recommandée par les experts et celle observée heure par heure. Par exemple, un décalage d'une heure entre le temps du diagnostic fait *a posteriori* par les experts et le temps du diagnostic fait par les médecins en charge de l'enfant était considéré comme un soin sub-optimal. La fréquence des soins sub-optimaux était beaucoup plus élevée parmi les 143 enfants décédés que les 355 enfants survivants, par exemple pour l'absence de diagnostic des complications de la méningococcie (40 % vs 22 %), le remplissage insuffisant (24 % vs 11 %), ou l'absence de recours aux inotropes (44 % vs 14 %). Ces deux premiers types de soins sub-optimaux restaient significativement associés à une issue défavorable après ajustement sur le sérotype du méningococque et la gravité initiale mesurée par le score pronostique. Au moins un soin sub-optimal était présent dans la prise en charge de 75 % des enfants décédés versus 45 % des enfants survivants. Les auteurs concluaient qu'une relation entre l'optimalité des soins et l'issue des patients était établie dans leur étude.

Des auteurs américains ont également cherché à évaluer l'impact de la qualité des soins sur le devenir des enfants atteints d'infections bactériennes sévères au stade de choc septique sans se limiter aux méningococcies [5]. Il s'agissait d'une étude rétrospective à partir des dossiers de transfert par une équipe de transports pédiatriques spécialisés d'hôpitaux périphériques vers un centre hospitalier pédiatrique en Pennsylvanie entre 1993 et 2001. Les soins prodigués étaient évalués en fonction de leur adéquation au *Pediatric advanced life support manual*. Parmi les 91 enfants en choc septique étudiés, 45 % avaient reçu un remplissage approprié au cours de la prise en charge avant leur arrivée dans le centre de référence et seulement 30 % avaient été réanimés de manière optimale. Les enfants dans les situations cliniques les plus graves (en choc persistant, c'est-à-dire en absence de normalisation

de la pression artérielle systolique et du temps de recoloration capillaire) avaient reçu moins souvent des remplissages optimaux (25 % vs 45 %). Après ajustement sur la gravité évaluée par le score PRISM, chaque heure de choc persistant était associée à une augmentation de la probabilité de décès avec un *odd-ratio* de 1,5 (intervalle de confiance à 95 % 1,0-2,2).

Ces trois études ont mis en évidence une fréquence élevée des soins sub-optimaux dans certaines situations d'IBS de l'enfant, ont permis de définir quelques-unes de leurs caractéristiques, et suggèrent une relation avec l'issue de ces IBS. Ces études ont cependant plusieurs limites : restriction à la méningococcie [9, 10], à la prise en charge hospitalière [5, 10], recrutement monocentrique [5, 9], absence d'experts indépendants pour évaluer l'optimalité des prises en charge [5, 9, 10], restriction à la prise en charge hémodynamique pour définir le caractère optimal des soins [5].

4. L'étude nantaise

Une enquête confidentielle rétrospective en population a évalué la fréquence des soins sub-optimaux dans la prise en charge d'enfants décédés d'IBS entre janvier 2000 et mars 2006 dans deux départements : Vendée et Loire Atlantique. L'optimalité des soins a été évaluée indépendamment par un pédiatre libéral et un pédiatre hospitalo-universitaire qui ont analysé des fiches anonymes de synthèse horodatées mentionnant les symptômes présentés par les enfants ainsi que les actions menées en réponse par les parents et les médecins. Ces fiches ne comportaient aucune information sur le diagnostic final, ni sur l'issue pour les patients et les critères d'inclusion des patients n'étaient pas connus des experts. Parmi les 21 enfants décédés étudiés, l'âge médian était de 2 ans (écart inter-quartile 0,8-2,8) et les germes en cause : *Neisseria meningitis* (57 %), *Streptococcus pneumoniae* (14 %), et *Streptococcus pyogenes*, *Mycoplasma pneumoniae*, *Salmonella enterica*, *Escherichia coli*. L'évaluation de l'optimalité des prises en charge a conclu à un retard des parents au recours médical dans 33 % des cas, principalement en raison de l'absence de consultation devant un purpura, une altération de la vigilance, une hypotonie ou une fièvre élevée persistante depuis plus de 48 heures. Les prises en charge médicales ont été considérées sub-optimales dans 71 % des cas, en raison de 35 défauts différents : 31 % de nature diagnostique, 57 % thérapeutique. Les typologies d'erreur les plus fréquentes étaient la sous-évaluation de la gravité initiale (29 %), le retard à l'administration du traitement adapté (38 %). L'antibiothérapie n'était administrée que dans 20 % des cas concernés par la circulaire sur le purpura fulminans et le remplissage vasculaire en cas de sepsis était absent (24 %), insuffisant (24 %), et/ou non renouvelé malgré son échec (14 %). Cette étude préliminaire se poursuit par la comparaison à un groupe d'enfants atteints d'IBS mais non décédés et par une étude prospective multicentrique dans la région Grand-Ouest.

5. Conclusion

La fréquence des soins sub-optimaux dans la prise en charge initiale des IBS est élevée dans les rares études disponibles, sans qu'il soit possible d'établir un lien de causalité entre la non optimalité initiale des soins et l'issue pour les patients. Si ce lien de causalité était démontré par les études en cours, la typologie assez standardisée de ces soins sub-optimaux devrait inciter à des campagnes de formation pour les parents (importance

d'une consultation rapide en cas de symptômes précoces de sepsis), et les médecins ambulatoires (administration précoce d'antibiotiques) et hospitaliers (identification précoce du sepsis et remplissage précoce).

Références

1. Committee on Quality of Health Care in America, Institute of Medicine. To err is human: building a safer health system. National Academy Press 2000.
2. INSEE. Principales causes de décès des jeunes et des enfants. Disponible à l'adresse : http://www.insee.fr/fr/themes/tableau.asp?reg_id=0&ref_id=natccjo62o6. Consulté le 15 février 2007.
3. Watson R, Carcillo J. Scope and epidemiology of pediatric sepsis. *Pediatr Crit Care Med* 2005;6:53-5.
4. Thompson M, Ninis N, Perera R, et al. Clinical recognition of meningococcal disease in children and adolescents. *Lancet* 2006;367:397-403.
5. Han Y, Carcillo J, Dragotta M, et al. Early reversal of pediatric-neonatal septic shock by community physicians is associated with improved outcome. *Pediatrics* 2003;112:793-9.
6. Dellinger R, Levy M, Carlet J, et al. Surviving Sepsis Campaign: international guidelines for management of severe sepsis and septic shock: 2008. *Crit Care Med* 2008;36:296-327.
7. Conseil supérieur d'hygiène publique. Avis du Conseil Supérieur d'Hygiène Publique de France du 10 mars 2000. Disponible en ligne : http://www.sante.gouv.fr/htm/dossiers/cshpf/a_mt_100300_meningite_01.htm. Consulté le 24 décembre 2007.
8. Goldstein B, Giroir B, Randolph A. International pediatric sepsis consensus conference: definitions for sepsis and organ dysfunction in pediatrics. *Pediatr Crit Care Med* 2005;6:2-8.
9. Nadel S, Britto J, Booy R, et al. Avoidable deficiencies in the delivery of health care to children with meningococcal disease. *J Accid Emerg Med* 1998;15:298-303.
10. Ninis N, Phillips C, Bailey L, et al. The role of healthcare delivery in the outcome of meningococcal disease in children: case-control study of fatal and non-fatal cases. *Br Med J* 2005;330:1475.

Annexe 6 : Résumé du protocole de l'étude DIABACT III

Résumé du protocole de l'étude DIABACT III (PHRC inter-régional 2008)

Titre de l'étude	Fréquence, conséquences et déterminants des soins sub-optimaux dans la prise en charge des infections bactériennes sévères de l'enfant : enquête confidentielle prospective en population dans la région Ouest (étude DIABACT III).
Mots clés	infections bactériennes communautaires sévères, pédiatrie, évaluations des pratiques, optimisation des soins.
Promoteur de l'étude	CHU de Nantes
Investigateur coordonnateur (si étude multicentrique)	<p><i>Dr Christèle GRAS-LE GUEN</i> Réanimation Pédiatrique et Néonatale Hôpital Mère-Enfant 38 Boulevard Jean Monnet 44093 Nantes Cedex</p> <p><i>Dr Elise LAUNAY</i> Clinique Médicale Pédiatrique Hôpital Mère Enfant CHU NANTES</p>
Méthodologiste	<p><i>Dr Martin CHALUMEAU</i> Laboratoire d'Epidémiologie Clinique, Service de Pédiatrie Générale, Unité INSERM U149, France Hôpital Saint Vincent de Paul, AP-HP, 82 avenue Denfert Rochereau 75014 Paris</p>
Nombre de centres prévus	5
Planning de l'étude	Durée totale : 2 ans Période de recrutement : 2ans Durée de traitement par patients :NA Durée de suivi par patient : NA

Design de l'étude	<ul style="list-style-type: none"> - Enquête confidentielle - Prospective - Multicentrique
Objectifs de l'étude	<p>Principal : Déterminer la fréquence des soins sub-optimaux dans la prise en charge initiale des infections bactériennes sévères de l'enfant à l'échelle d'une population (région ouest).</p> <p>Secondaires</p> <ul style="list-style-type: none"> ● Identifier les caractéristiques des soins sub-optimaux en établissant une classification des défauts de prises en charge à la recherche d'erreurs systémiques telles que définies par l'Institute of Medicine. (annexe) ● Identifier les déterminants des soins sub-optimaux (par exemple : lieu de prise en charge, qualification et expérience du médecin, moment de la prise en charge, délais de prise en charge) ● Identifier les conséquences des soins sub-optimaux initiaux sur le devenir des enfants atteints d'IBS.
Nombre de cas (effectif de l'échantillon)	N = 270
Critères principaux de sélection, d'inclusion, de non-inclusion et d'exclusion	<ul style="list-style-type: none"> ● Inclusion : <ul style="list-style-type: none"> ○ Enfant âgé de 1 mois à 16 ans ○ admis en réanimation pour infection bactérienne sévère ○ parents sont domiciliés dans la région de l'étude ● Non inclusion : <ul style="list-style-type: none"> ○ Enfant âgés de moins d'1 mois et de plus de 16 ans ○ enfants atteints d'une infection bactérienne nosocomiale ○ enfants admis en réanimation pour coqueluche ○ enfants non admis en réanimation, enfants atteints d'infection virale ou fongique. ● Critères d'exclusion des personnes qui se prêtent à la recherche : <ul style="list-style-type: none"> ○ dossiers incomplets ○ infection bactérienne non prouvée a posteriori.

Traitement, acte interventionnel à l'étude	Etude Soins Courants
Traitement de référence (si applicable)	NA
Critère de jugement principal	<p><u>Fréquence</u> du caractère optimal ou non de la prise en charge médicale initiale des infections bactérienne sévères de l'enfant.</p> <p>Le caractère optimal ou non de la prise en charge sera déterminé par 2 experts indépendants, à l'aveugle l'un de l'autre et de l'issue. En cas de désaccord initial, un accord a été trouvé par un troisième expert. <u>Le degré d'accord</u> entre les juges sera ensuite évalué par le calcul du kappa et du kappa pondéré selon une méthode quadratique qui permet de tenir compte du degré de désaccord. Trois possibilités de conclusions seront proposées aux juges : prise en charge sub-optimale certaine (1), prise en charge sub-optimale possible (2) et prise en charge optimale (3).</p>
Critères de jugement secondaires	<ul style="list-style-type: none"> ● Caractéristiques des soins sub-optimaux relevés par les experts: celles-ci seront analysées grâce à la <u>classification des erreurs médicales</u> établie par l'IOM permettant notamment de différencier les erreurs diagnostiques des erreurs thérapeutiques. ● Déterminants des soins sub-optimaux : ceux-ci seront analysés en dégagant plusieurs éléments à partir des observations médicales (CRF) notamment le lieu de prise en charge initiale, la date et l'heure de prise en charge initiale, la qualification des médecins impliqués dans la prise en charge initiale, la présentation clinique initiale ou encore l'âge de l'enfant. ● Analyse des délais : les différents délais (délai du patient et délais médicaux) seront calculés et analysés à la lumière du caractère optimal ou non de la prise en charge. ● Conséquences des soins sub-optimaux : celles-ci seront analysées par comparaison des enfants décédés versus les enfants non décédés afin de dégager un éventuel lien statistique entre prise en charge sub-optimale et décès, et délais de prise en

	charge optimale et décès (schéma récapitulatif de ces différents délais en annexe).
Autres évaluations	NA
Analyse statistiques	<ul style="list-style-type: none"> ● Description des caractéristiques générales de la population par le calcul de moyennes avec leurs intervalles de confiance à 95%, de médianes avec précision des extrêmes et des écarts interquartiles. ● Evaluation du coefficient d'accord entre les experts par le calcul du kappa et du kappa pondéré selon un système quadratique. ● Comparaison de la distribution des prises en charge non optimale entre les enfants décédés et les enfants non décédés grâce à des tests non paramétriques afin de dégager un lien statistiques entre soins sub-optimaux et mortalité. ● Analyse uni et multivariée de certains paramètres comme l'âge, le germe en cause , les délais de prise en charge, le caractère optimal ou non de la prise en charge initiale, la qualification des médecins à la prise en charge initiale, le lieu de prise en charge initiale, l'heure de prise en charge initiale, afin de dégager des facteurs de risque de mortalité ● Comparaison des conclusions des expertises d'après l'eCRF et d'après les copies des dossiers ● Comparaison des conclusions des expertises faites à des temps différents (1 an d'intervalle)

Annexe 7 : Protocole de l'étude DIABACT IV

Protocole DIABACT IV

PREPS 2013

Réf. : RC14_0119

« Construction et validation d'un outil de repérage automatique des parcours de soins à risque de sub-optimalité dans la prise en charge des infections sévères de l'enfant »

Investigateur Coordonnateur ou personne qui dirige et surveille la réalisation de la recherche :

Pr Christèle Gras Le Guen
Service de néonatalogie et réanimation pédiatrique
CHU de Nantes, site HME, 1 place Alexis Ricordeau – 44093 Nantes cedex 1
Tel : 02.40.08.76.71
Mail : christele.gras-leguen@chu-nantes.fr

Méthodologiste :

Pr Martin Chalumeau
INSERM U953, recherches épidémiologiques en santé périnatale, santé des femmes et des enfants
Maternité de Port Royal – 63, avenue de l'observatoire – 75014 Paris
Mail : martin.chalumeau@nck.aphp.com

Epidémiologiste :

Matthieu Hanf
EC-CIC – CHU de Nantes, 5, allée de l'île Gloriette – 44093 Nantes cedex 1
Tel : 02.53.48.28.77
Mail : matthieu.hanf@chu-nantes.fr

Etablissement responsable de la recherche :



CHU de Nantes
Contact : Anne Omnès
Département promotion, Direction de la Recherche
5, allée de l'île Gloriette
44 093 Nantes cedex 01 (FRANCE)
Contact :
Tel : 02 53 48 28 35
Fax : 02 53 48 28 36

RESUME

Titre de l'étude	Construction et validation d'un outil de repérage automatique des parcours de soins à risque de sub-optimalité dans la prise en charge des infections sévères de l'enfant
Mots clés	IBS, sub-optimalité, données SNIIR-AM, pédiatrie
Responsable de la recherche	CHU DE NANTES
Investigateur coordonnateur	Pr Christèle Gras Le Guen
Nombre de centres prévus	13 centres nationaux : Nantes, Hôpital Trousseau (AP-HP), Hôpital Robert Debré (AP-HP), Hôpital Necker-Enfants malades (AP-HP), Bordeaux, Besançon, Montpellier, Lyon, Grenoble, Lille, Saint-Etienne, Brest et Angers.
Type d'étude	Recherche non Interventionnelle
Planning de l'étude	<ul style="list-style-type: none"> ❖ Durée totale : 36 mois ❖ Période de recrutement : 24 mois ❖ Durée de suivi par patient : NA
Design de l'étude	<ul style="list-style-type: none"> ❖ Multicentrique nationale ❖ Prospective et rétrospective (données de DIABACT III)
Objectifs de l'étude	Nous proposons ici d'identifier des caractéristiques extractibles des données du PMSI et SNIIRAM et associées à un risque plus élevé de suboptimalité dans la prise en charge précoce (avant l'arrivée en réanimation) des enfants de 0 à 16 ans atteints d'infections communautaires sévères en France (défini comme une infection amenant à l'admission en réanimation) afin de proposer des indicateurs d'alerte permettant de repérer <i>a posteriori</i> des prises en charge « à risque » d'être suboptimale et de déclencher un processus d'audit sur ces dossiers
Nombre de cas prévisionnel	782 cas (512 prospectifs et 270 rétrospectifs — DIABACT III)

<p>Critères principaux de sélection, d'inclusion, de non-inclusion et d'exclusion</p>	<p>Critères d'inclusion :</p> <p><u>Pour la construction du score (données de l'étude DIABACT III):</u></p> <ul style="list-style-type: none"> - Enfants de 1 mois à 15 ans et 3 mois admis en réanimation pédiatrique ; - avec une infection bactérienne communautaire sévère <p><u>Pour la validation du score (inclusions prospectives)</u></p> <ul style="list-style-type: none"> - Les critères d'inclusion de DIABACT III seront élargis en incluant des enfants de 0 à 1 mois. - Parents et patient (le cas échéant) ayant accepté de participer à l'étude et ayant signé le consentement de biocollection pour la participation à l'étude ancillaire. <p>Critères de non inclusion :</p> <ul style="list-style-type: none"> - Infections nosocomiales - Bronchiolites - Infections virales
<p><input type="checkbox"/> Pharmacogénétique</p> <p><input type="checkbox"/> Pharmacocinétique</p> <p><input type="checkbox"/> Pharmacodynamique</p> <p><input type="checkbox"/> Pharmaco-économique</p> <p><input checked="" type="checkbox"/> Autres analyses : analyses génétiques sur prélèvements</p>	
<p>Analyse statistiques</p>	<p>L'analyse statistique des données sera réalisée avec les logiciels SAS ou R.</p> <p>Le seuil de signification des tests a été fixé à 5 %.</p>
<p>Soumission au Groupe Nantais d'Ethique dans le Domaine de la Santé (GNEDS)</p>	<p>Avis favorable obtenu le 10 Juin 2014</p>

1. INTRODUCTION

L'infection bactérienne communautaire demeure à ce jour une cause fréquente de morbi-mortalité chez l'enfant, dont le caractère évitable constitue un véritable défi pour les cliniciens. Lors d'un précédent travail, nous avons constaté que 76% des prises en charge des enfants admis en réanimation pour une infection bactérienne sévère étaient expertisées comme suboptimales, et significativement associées à un risque accru de décès. Dans ce contexte, nous cherchons à identifier des indicateurs extractibles des données du PMSI et SNIIR-AM associées à un risque plus élevé de prise en charge précoce suboptimale des enfants présentant une infection bactérienne sévère afin de les combiner et de les utiliser sous forme de score ou d'arbre décisionnel que nous validerons à partir des données d'une étude prospective multicentrique nationale incluant 512 enfants admis en réanimation pour infection sévère. Nous proposerons alors un score associé à un risque de suboptimalité des soins permettant d'évaluer la performance du système de soins.

2. JUSTIFICATION DE L'ETUDE

2.1. RATIONNEL : CONTEXTE, HYPOTHESES ET ENJEUX

Contexte : les infections bactériennes sont une cause de morbi-mortalité non négligeable en pédiatrie.

Selon les données du Centre d'épidémiologie sur les causes médicales de décès (CépiDC), 70 enfants sont décédés en 2010 en France à cause d'un processus infectieux (autre que VIH/SIDA, tuberculose ou hépatites virales). Une étude pilote menée en 2006 en Grande Bretagne sur les causes des décès des enfants montrait que les infections représentaient la 1^{ère} cause de décès dus à une maladie aigue (1). Dans une étude publiée en 2009, les infections étaient responsables de 15% des causes de décès (2). Dans un état des lieux récent des chocs septiques dans 16 départements d'urgences pédiatriques, les chocs septiques représentaient jusqu'à 27 % des hospitalisations.

Hypothèse (1) : la suboptimalité des soins a une part importante de responsabilité dans les décès par infections bactériennes sévères (IBS) en pédiatrie et l'amélioration de la qualité des soins pourrait donc améliorer le pronostic des IBS.

Dans une étude publiée en 2009 par Harndern et al, les dossiers de 168 enfants morts dans 5 régions du Royaume-Uni étaient analysés pour mettre en évidence d'éventuelles défaillances évitables dans leur prise en charge en soins primaires. Quinze pour cent des décès étaient liés

à une infection. Les défaillances de prise en charge de soins primaires les plus fréquentes étaient la non reconnaissance du sepsis et son retard de prise en charge et concernaient 11 décès (7%) (2). Des études plus anciennes sur les méningococcémies (3, 4), le choc septique (5) ou les méningites à pneumocoque (6) de l'enfant ont mis en évidence une fréquence élevée des soins suboptimaux : de 40 à 75% des prises en charge. Ces études ont également montré une association entre soins suboptimaux et morbi-mortalité (3-6).

Hypothèse (2) : l'analyse des bases de données nationales des systèmes de santé est un outil d'évaluation de la santé et des soins.

La recherche épidémiologique à partir des grandes bases de données nationales peut permettre en les croisant d'évaluer le parcours de soins des patients (7) ou d'évaluer la consommation de soins voire même l'observance au traitement d'une population définie (8). Ces études ont l'avantage de pouvoir assurer un recrutement en population évitant ainsi les biais de sélection. En pédiatrie, ces bases de données ont été utilisées pour définir des indicateurs de qualité d'hospitalisation à partir des données sur complications nosocomiales (9). Elles permettent également d'évaluer les coûts spécifiques de certaines prescriptions (10), de comparer les prescriptions de différents professionnels de santé (11), d'évaluer la prescription d'antibiotiques chez les enfants asthmatiques (12), évaluer l'impact de la vaccination anti-varicelleuse de masse sur les hospitalisations pour varicelle (13) ou encore définir les déterminants de récurrence d'invagination intestinale aiguë (14).

Enjeux : détecter et caractériser la suboptimalité des soins dans la prise en charge des infections bactériennes sévères pour cibler les actions correctrices et améliorer la qualité des soins.

Nous avons montré précédemment par une enquête confidentielle en population (Diabact 1) que 75% (IC de 57% à 95%) des prises en charge précoces (pré-admission en réanimation) des enfants décédés d'infection bactérienne sévère (IBS) dans la population étudiée étaient considérées comme sub-optimales par 2 experts indépendants (15). En incluant les enfants survivants admis en réanimation pédiatrique pour IBS au cours de la même période (étude Diabact 2) nous avons mis en évidence une relation entre le nombre de soins suboptimaux par prise en charge et le décès (OR ajusté à 1,5 ; p=0,038) et dégagé certains déterminants de ces soins suboptimaux (jeune âge, qualification du médecin notamment) [Launay, manuscrit en préparation]. Le caractère rétrospectif de ces 2 études limitait cependant la précision de la mesure des délais ainsi que de certaines caractéristiques potentiellement déterminantes de la qualité des soins et du parcours de soins. Nous avons donc mis en place une étude prospective dans un plus grand bassin de population afin de préciser la fréquence, les conséquences et les déterminants des soins sub-optimaux dans la prise en charge des IBS [DIABACT 3, PHRC

interrégional 2008]. Cette étude, dont les inclusions se se sont terminées en Janvier 2014 a permis d'inclure à ce jour 270 enfants atteints d'IBS dont 20 décédés dans 6 centres hospitaliers universitaires du grand ouest (Nantes, Rennes, Poitiers, Brest, Angers et Tours).

2.2. JUSTIFICATION SCIENTIFIQUE DU PROJET

Les 3 études précédemment citées (Diabact 1, 2 et 3) reprenaient la méthodologie des enquêtes confidentielles à comité d'expert, et nécessitaient un processus lourd avec analyse de tout le dossier médical. Parmi les éléments pris en compte pour évaluer l'optimalité de soins, certains sont disponibles dans les bases de données du PMSI et du SNIIRAM (nombre d'actes de soins avant hospitalisation, nombre de vaccins reçus notamment anti pneumococcique ou méningococcique grâce aux vignettes CIP des médicaments, lieux de prise en charge avant admission en réanimation, durée de séjours). D'autres éléments disponibles dans ces bases de données (prise en charge en affection de longue durée (ALD), prescriptions antérieures d'antibiotiques, prise d'anti-inflammatoire) pourraient également constituer des éléments de « repérage » des enfants « à risque » de soins suboptimaux. Un système de détection (par exemple autour d'un score établi à partir de données SNIIRAM) des enfants, atteints d'IBS , à risque de soins suboptimaux pourrait permettre de déclencher une analyse plus approfondie de ces dossiers et aboutir à la reconnaissance d'erreurs dans la prise en charge pouvant amener à des actions correctrices à l'échelle des établissements de santé (par exemple, problème d'orientation des enfants ou problème d'accessibilité aux soins...) ou à plus large échelle (problème de sous vaccination des enfants à risque (en ALD)...).

Ces 3 études étaient limitées à un secteur géographique restreint (2 départements pour Diabact 1 et 2, Grand-Ouest pour Diabact 3) et ne tenaient pas compte d'éventuelles variations nationales dans l'épidémiologie bactérienne, l'accès aux soins ou encore les politiques de santé régionale. De plus, si les infections bactériennes sévères sont une cause relative importante de mortalité, elles demeurent un phénomène rare dans l'absolu dans une population avec une faible mortalité globale. L'élargissement du recrutement au territoire national permettrait de vérifier les hypothèses émises dans Diabact 3 à l'échelon national.

Enfin nous avons exclu les infections virales et la coqueluche de nos précédentes études. Cependant certaines infections virales comme la grippe ou la gastro-entérite aigue sont encore responsables de décès en France malgré des mesures de prévention existantes (vaccination contre la grippe et le rotavirus, mesure précoce de réhydratation orale pour la gastro-entérite aigue). La problématique de prise en charge de ces infections peut différer sur certains points (pas d'enjeu de l'antibiothérapie précoce comme dans les infections bactériennes sévères) cependant certains facteurs d'évitabilité sont les mêmes comme la vaccination, l'importance de l'éducation parentale pour la reconnaissance des signes d'alerte devant amener à une

consultation rapide, l'importance du suivi et de l'accessibilité aux soins. Evaluer le parcours de soins des enfants atteints d'infection sévère communautaire (la problématique des infections nosocomiales est différente et peut déjà être identifiée par des indicateurs du PMSI (9)) peut permettre de dégager des pistes d'optimisation commune (vaccination, éducation parentale, Formation Médicale Continue/Développement Professionnel Continu...).

La prise en charge en réanimation ne fera pas ici l'objet de notre évaluation car il existe déjà des recommandations précises de prise en charge du sepsis en réanimation (surviving sepsis campaign) et des indicateurs de complications nosocomiales (infection sur cathéter central, pneumonie secondaire à la ventilation ...).

3. OBJECTIFS ET CRITERES DE JUGEMENT

3.1. OBJECTIF ET CRITERE D'EVALUATION PRINCIPAL

Nous proposons ici d'identifier des caractéristiques (indicateurs) extractibles des données du PMSI et SNIIRAM et associées à un risque plus élevé de suboptimalité dans la prise en charge précoce (avant l'arrivée en réanimation) des enfants de 0 à 16 ans atteints d'infections communautaires sévères en France (défini comme une infection amenant à l'admission en réanimation (15)) afin de proposer des indicateurs d'alerte permettant de repérer *a posteriori* des prises en charge « à risque » d'être suboptimale et de déclencher un processus d'audit sur ces dossiers. Par exemple, s'il apparaît qu'un grand nombre de consultations avant l'admission en réanimation (dans les 7 jours avant par exemple) est lié à une fréquence plus élevée de prise en charge suboptimale, l'analyse systématique des dossiers des enfants avec grand nombre de consultations avant l'admission en réanimation pourra permettre de mettre en évidence des erreurs systémiques évitables en soins primaires. Ce processus de repérage évite une analyse systématique de tous les dossiers des enfants admis pour infection sévère en réanimation.

Un schéma d'étude articulé sur 2 phases d'investigation est proposé. Dans un premier temps, nous testerons les associations éventuelles entre des candidats indicateurs SNIIRAM/PMSI d'intérêt (variables prédictives) et le caractère suboptimal de la prise en charge (variable prédite), via la cohorte DIABACT III (d'ores et déjà disponible avec 270 patients exploitables en Janvier 2014). Les prédicteurs identifiés seront éventuellement combinés dans un outil prédictif soit score composite, soit dans un arbre décisionnel. Puis nous validerons ces prédicteurs et l'éventuel outil prédictif combiné à partir des données d'une étude prospective multicentrique nationale incluant des enfants admis en réanimation pour infection sévère. Nous proposerons alors un score associé à un risque de suboptimalité des soins permettant d'évaluer la performance du système de soins.

3.2. OBJECTIFS ET CRITERES D'EVALUATION SECONDAIRES

L'objectif secondaire de cette étude est de décrire l'épidémiologie et les caractéristiques des infections communautaires sévères de l'enfant en France, les dernières études datant de plus de 10 ans (16).

3.3. OBJECTIF ET CRITERES D'EVALUATION DES ETUDES ANCILLAIRES

(17) Une étude ancillaire à partir d'une bio-collection est également prévue dans le but d'étudier plus précisément les facteurs de l'hôte (susceptibilité génétique aux infections) et de la bactérie (facteur de virulence) dans le pronostic des infections bactériennes sévères. La constitution de la bio-collection ne concernera que les enfants atteints d'infections bactériennes sévères, un échantillon sanguin (2 à 5 mL) sera prélevé au cours d'une prise de sang s'intégrant dans les soins courants sous réserve d'un consentement parental (et de l'enfant selon l'âge), et conservé à -80°C. Les souches bactériennes des enfants correspondants seront quant à elles envoyées aux centres nationaux de référence en fonction de la bactérie comme prévu dans les règles de bonnes pratiques devant une infection bactérienne sévère de l'enfant (18, 19). Cette biocollection est intitulée « Pédiatrie » et a été déclarée au Ministère de la recherche et autorisée par le CPP ainsi que le consentement de biocollection associé. Cette biocollection (ADN des patients et bactéries) sera analysée et valorisée (renforcement de notre modèle explicatif du risque de décès pour IBS) via un autre appel d'offre spécifique des biocollections.

Nous utiliserons des modèles prédictifs multivariés pour construire un score de risque de non optimalité sur les données de l'étude DIABACT III (270 enfants inclus en janvier 2014).

4. POPULATION ETUDIEE

4.1. DESCRIPTION DE LA POPULATION

Les patients inclus dans cette étude seront des enfants admis en réanimation pédiatrique pour une infection bactérienne communautaire sévère.

Les inclusions rétrospectives (données de l'étude DIABACT III) concernant des enfants de 1 mois à 15 ans et 3 mois et les inclusions prospectives concerneront des enfants de 0 mois à 15 ans et 3 mois.

4.2. CRITERES D'INCLUSION

Pour la construction du score (données de l'étude DIABACT III):

- tous les enfants de 1 mois à 15 ans et 3 mois admis en réanimation pédiatrique ;
- avec une infection bactérienne communautaire sévère définie par une infection bactérienne documentée par la présence d'une bactérie dans un site normalement stérile (liquide pleural, sang, liquide céphalorachidien, liquide articulaire, biopsie osseuse ...), une culture d'urine positive sur un prélèvement fait dans des bonnes conditions d'asepsie accompagnée d'un syndrome fébrile (ex : pyélonéphrite), ou d'un purpura fulminans même en l'absence de documentation.

Pour la validation du score (inclusions prospectives)

- Les critères d'inclusion de DIABACT III seront élargis en incluant des enfants de 0 à 1 mois.
- Parents et patient (le cas échéant) ayant accepté de participer à l'étude et ayant signé le consentement de biocollection pour la participation à l'étude ancillaire.

Remarque : ce qui importe pour l'analyse du critère de jugement principal (prise en charge optimale ou non), c'est de connaître le parcours de soin pré hospitalier et la façon dont des familles et les soignants ont réagi face aux symptômes présentés par l'enfant. Il est donc important de ne considérer que les enfants dont les premiers symptômes ont commencé à domicile, après la sortie de maternité mais non durant l'hospitalisation. Le parcours de soin des mères en anténatal par contre n'interviendra pas, et ne sera pas pris en compte dans l'analyse de la prise en charge, même en cas d'infection materno-fœtale.

4.3. CRITERES DE NON INCLUSION

Pour la construction et la validation du score :

- Infections nosocomiales définies selon la définition du Ministère de la Santé dans son dossier de janvier 2007 sur les infections nosocomiales à savoir : les infections nosocomiales sont les infections contractées dans un établissement de santé. Une infection est considérée nosocomiale si elle était absente au moment de l'admission du patient dans l'établissement de santé. Lorsque l'état infectieux du patient à l'admission est inconnu, l'infection est généralement considérée comme nosocomiale si elle apparaît après un délai d'au moins 48 heures d'hospitalisation ou un délai supérieur à la période d'incubation de l'infection. En cas d'infection du site opératoire, le délai communément admis est de 30 jours, ou, s'il y a mise en place d'une prothèse ou d'un implant, d'une année après l'intervention.
- Enfants atteints de coqueluche, de bronchiolites ou d'infections virales (16)

5. DEROULEMENT DE L'ETUDE

5.1. METHODOLOGIE GENERALE DE LA RECHERCHE

La recherche présente les caractéristiques suivantes :

- ❖ Etude épidémiologique
- ❖ Etude **multicentrique** nationale,
- ❖ Etude **prospective** et **rétrospective**.

5.2. TECHNIQUES D'ETUDES ET D'ANALYSES

5.2.1. Description détaillée des paramètres d'évaluation

Nous proposons ici d'identifier des caractéristiques (indicateurs) extractibles des données du PMSI et SNIIR-AM et associées à un risque plus élevé de suboptimalité dans la prise en charge précoce (avant l'arrivée en réanimation) des enfants de 0 à 16 ans atteints d'infections communautaires sévères en France (défini comme une infection amenant à l'admission en réanimation (16)) afin de proposer des indicateurs d'alerte permettant de repérer *a posteriori* des prises en charge « à risque » d'être suboptimale et de déclencher un processus d'audit sur ces dossiers. Par exemple, s'il apparaît qu'un grand nombre de consultations avant l'admission en réanimation (dans les 7 jours avant par exemple) est lié à une fréquence plus élevée de prise en charge suboptimale, l'analyse systématique des dossiers des enfants avec grand nombre de consultations avant l'admission en réanimation pourra permettre de mettre en évidence des erreurs systémiques évitables en soins primaires. Ce processus de repérage évite une analyse systématique de tous les dossiers des enfants admis pour infection sévère en réanimation.

Un schéma d'étude articulé sur 2 phases d'investigation est proposé. Dans un premier temps, nous testerons les associations éventuelles entre des candidats indicateurs SNIIR-AM d'intérêt (variables prédictives) et le caractère suboptimal de la prise en charge (variable prédite), via la cohorte DIABACT III (d'ores et déjà disponible avec 270 patients exploitables au 1er Janvier 2014). Les prédicteurs identifiés seront éventuellement combinés dans un outil prédictif soit score composite, soit dans un arbre décisionnel. Puis nous validerons ces prédicteurs et l'éventuel outil prédictif combiné à partir des données d'une étude prospective multicentrique nationale incluant des enfants admis en réanimation pour infection sévère. Nous proposerons alors un score associé à un risque de suboptimalité des soins permettant d'évaluer la performance du système de soins.

5.2.2. Description des techniques et analyses

La variable prédite

La variable prédite sera la non optimalité de la prise en charge précoce de l'infection sévère. La non optimalité sera définie selon la méthodologie des enquêtes confidentielles avec comité d'experts utilisée dans les études DIABACT I et II. Cette méthodologie s'appuie sur l'élaboration systématique par le réanimateur pédiatre en charge de l'enfant d'une fiche synthétique de l'histoire clinique horodatée exhaustive sur la séquence et les modalités de prise en charge (site, acteur, dates et heures des symptômes, diagnostic, traitement etc...), cette synthèse est soumise à 2 experts indépendants chargés de statuer sur le caractère optimal de la prise en charge, en aveugle de l'issue (survie ou décès). Les experts évaluent la prise en charge globale de façon argumentée avec 3 conclusions possibles : prise en charge suboptimale certaine, optimale ou ne peut conclure. La validité de la suboptimalité est testée par le calcul d'un coefficient d'accord inter-juge (coefficient kappa). En cas de discordance un 3^{ème} expert est sollicité. Pour les analyses uni et multivarié, la variable prédite sera dichotomisée en prise en charge suboptimale certaine versus optimale ou ne peut juger car le but est de proposer un outil prédictif associé à un risque élevé de suboptimalité certaine.

Les variables prédictives

Les variables prédites « candidates » seront les variables issues des bases de données SNIIR-AM (données de consommation inter-régime- DCIR-et du programme de médicalisation des systèmes d'information-PMSI-) ci-dessous :

- le décès : le statut décédé est disponible via le PMSI, la date précise du décès constitue une des données dites « sensibles » justifiant d'une autorisation spécifique de la CNIL.
- l'âge de l'enfant : qui est un facteur identifié de risque de décès et également de suboptimalité (DIABACT II, résultats non publiés) est également disponible via le PMSI et constitue une donnée sensible (mois et année de naissance)
- la présence d'une ALD avec le code CIM 10 correspondant qui reflète les caractéristiques du patient et d'éventuelles fragilités intrinsèques aux infections est disponible via le DCIR
- le nombre de consultations médicales précédant immédiatement l'hospitalisation (dans les 14 jours précédents) qui est un reflet du parcours de soins primaires est disponible via le DCIR
- la présence et le nombre de transferts inter-hospitaliers au cours de la prise en charge de l'infection qui reflète le parcours de soins secondaire est disponible via le PMSI
- le nombre de consultations médicales par an (qui sera rapporté au nombre de consultations attendues en fonction de l'âge de l'enfant et de la présence d'une ALD) qui donne un reflet de l'adhésion des parents face au système de soins est disponible via le DCIR.

- le nombre, la nature et la date (retard) des vaccinations prescrites ainsi que le nombre et la nature des prescriptions d'antibiotiques antérieures qui donnent un reflet de l'adhésion du médecin traitant aux recommandations de bonnes pratiques émises par les autorités de santé. Ces données sont disponibles via le DCIR pour le type et la date de l'acte et les vignettes CIP des médicaments remboursables (vaccin, antibiotiques). La date précise de l'acte constitue une donnée sensible.

- la date de la première consultation pour l'épisode en cours est disponible via le DCIR (données sensibles), la date et la durée d'hospitalisation et de séjour en réanimation pour l'infection en cours disponible via le PMSI et permettant le calcul des délais médicaux pré et intra hospitaliers et qui reflètent la cinétique du parcours de soins (dépendant à la fois du phénotype clinique- maladie fulminante par exemple- et de la réactivité du système de soins).

Parmi ces variables, certaines ne sont pas disponibles dans les données de DIABACT III (statut ALD, nombre de consultations par an, prescriptions antibiotiques) et nécessite donc une interrogation de la base SNIIR-AM (DCIR et PMSI).

5.2.3. Rattachement à une biocollection

L'étude ancillaire nécessite un recueil de sang qui sera réalisé lors du prélèvement sanguin effectué dans le cadre de la prise en charge habituelle de ces enfants (tube de sang supplémentaire).

Ce prélèvement supplémentaire pour la recherche sera associé à la biocollection intitulée « Pédiatrie » et déclarée au ministère de la recherche sous le n° DC-2011-1399, soumise au CPP avec avis favorable en date du 09/11/2011, de même que le formulaire d'information-consentement proposé aux parents/patients pour la conservation de leurs échantillons.

Chaque centre conservera les échantillons à -80°C dans son service. Critères d'arrêt prématuré de la participation d'une personne à la recherche

Pour les modalités et la durée du suivi des personnes ayant arrêté prématurément l'étude, se reporter à la section statistique.

5.3. STATISTIQUES

Les données seront nettoyées et consolidées par des data-managers du CHU de Nantes et transmises à l'unité INSERM U953 pour analyses.

L'analyse statistique des données sera réalisée avec les logiciels SAS ou R.

5.3.1. Description des méthodes statistiques prévues, y compris du calendrier des analyses intermédiaires prévues

Analyse univariée :

Ces indicateurs potentiels seront testés en analyse univariée par un test du χ^2 pour les variables en classes ou dichotomiques et par un test non-paramétrique de Mann-Whitney pour les variables continues afin de décrire les associations brutes.

Analyse multivariée :

Nous utiliserons un modèle de régression logistique multivarié afin d'identifier les variables associées de manière indépendante à la suboptimalité de la prise en charge (SSO) après ajustement sur les co-variables d'intérêt.

Les modèles finaux seront de la forme : score = logit de la probabilité de SSO = $\alpha + \sum \beta_i x_i$; où x_i représente les variables choisies dans le modèle après analyse univariée (facteurs de risque ou variable d'ajustement) et β_i représente le coefficient associé à une variable i donnée (c'est à dire l'accroissement du logit de la probabilité de SSO par unité de x_i).

La probabilité prédite par chaque modèle sera alors donnée par :

$$P(\text{SSO}) = 1/[1+e(-\text{score})].$$

En cas d'écart significatif à la linéarité, les variables continues seront traitées en continu dans le modèle multivarié en utilisant une transformation en polynômes fractionnaires d'ordre le plus bas (≤ 2).

Additivité, interactions :

Les interactions ayant un substratum physiopathologique (par exemple entre nombre de consultations par an et ALD ; entre le décès et l'ALD ou entre l'âge et le décès) seront testées directement dans un modèle logistique contenant les 2 variables d'intérêt et leur terme d'interaction.

En présence d'interactions multiples, nous choisirons une approche de partition récursive en remplacement de l'approche score.

Sélection de variables :

Nous utiliserons une technique de sélection de variables de type pas-à-pas afin d'éviter le surajustement et d'obtenir un modèle comportant un minimum de variable et donc facilement utilisable en routine. Certaines variables (comme l'âge ou le décès) pourront être forcées dans le modèle en raison de leur pertinence clinique ou bibliographique et/ou de leur notoriété.

Performances des modèles

Le pouvoir discriminant des modèles finaux sera apprécié par l'ASC ROC (Aire sous la courbe ROC). Les performances diagnostiques du modèle à différents seuils (terciles ou quartiles) seront décrites par la sensibilité (Se), spécificité (Sp), les rapports de vraisemblance positif ($RVP = \text{sensibilité}/(1-\text{spécificité})$) et négatif ($RVN=(1-\text{sensibilité})/\text{spécificité}$) et leurs intervalles de confiance. Nous n'avons pas fixé de seuil de sensibilité ou de spécificité *a priori* pour le choix du cut-off du modèle prédictif de la probabilité de SSO (score). En effet ces seuils pourront être personnalisés selon les moyens et les objectifs de chaque établissement. Par exemple si un établissement veut (et à les moyens de) détecter toutes les prises en charges suboptimales pour organiser des revues de morbi-mortalité (RMM), le cut-off du score devra être celui qui assure la meilleure sensibilité (au dépend d'une faible spécificité). A l'inverse, si l'objectif est de cibler les RMM sur les prises en charge très probablement suboptimales, il conviendra de choisir une forte spécificité.

La calibration des modèles sera évaluée par un test d'adéquation Hosmer-Lemeshow et représentée graphiquement par un graphe de calibration (calibration plot).

Validation interne des modèles prédictifs

La validation interne de chaque modèle sera réalisée par partition (split-sample) en réalisant une partition en un effectif de construction de 2/3 et de validation de 1/3 par exemple ou par validation croisée de type jackknife (10-fold cross-validation).

Comparaison des performances des nouveaux modèles prédictifs

La comparaison des modèles entre eux se fera par comparaison d'ASC ROC par un test de DeLong.

5.3.2. Méthode de validation du ou des indicateurs(s)

Validation externe multicentrique

La validation externe de l'outil prédictif précédemment construit sera faite sur la base d'un recrutement national (13 centres français de réanimation pédiatrique répartis sur le territoire national métropolitain). La calibration et le pouvoir discriminant du modèle seront étudiés sur cette nouvelle population en référence à la suboptimalité certaine définie par une expertise suivant la même méthodologie que celle citée ci-haut (variable prédite par l'expertise DIABACT). L'outil prédictif sera éventuellement affiné en cas de performances diagnostiques insuffisantes puis re-validé par bootstrap.

Vérification de l'exhaustivité du recrutement pour la construction et la validation du score

Le recrutement de chaque service sera recoupé aux données du CépiDC et du PMSI pour vérifier l'exhaustivité du recrutement.

5.3.3. Justification statistique du nombre d'inclusions

Pour la construction du score

« Les deux cent soixante-dix patients prévus pour l'étude DIABACT III ont été inclus en date de Janvier 2014. Le tableau ci-dessous représente les OR minimaux détectables en analyse univariée à un risque alpha de 10% et beta de 80% pour 270 enfants inclus dans l'hypothèse d'une suboptimalité à 60% (taux observé dans l'étude DIABACT II de 58%) en fonction de la fréquence des indicateurs candidats pour l'établissement du score. Les OR représentent la force de l'association entre chaque indicateur et la suboptimalité pour l'analyse univariée. »

Fréquence de l'indicateur (%)	10	20	30	40	50	60	70	80	90
OR minimal détectable	2,50	2,10	1,95	1,90	1,91	2	2,17	2,63	5,28

Dans l'étude DIABACT II, les variables associées à un risque de suboptimalité des soins avaient des OR de cet ordre de grandeur (3,6 à 9)

Pour la validation du score

L'enjeu de la recherche est d'élaborer un outil prédictif performant de jugement de la qualité d'un parcours de soins. Il devra permettre un repérage automatisé, rapide et discriminant de dossiers médicaux qui seront prioritaires pour subir ensuite un audit clinique et/ou revue de morbi-mortalité. Le seuil de sensibilité et de spécificité dépend des objectifs des établissements de santé comme expliqués plus haut. Pour un seuil de sensibilité arbitraire supérieure ou égale à 80% (seuil arbitraire correspond à un établissement de santé qui voudrait passer en revue la grande majorité des prises en charge suboptimales ou évaluer sa performance en terme de prise en charge d'infection sévère de l'enfant) de l'outil prédictif avec un intervalle de confiance à 95% cible de +/- 5%, il faut inclure 309 cas de prises en charge suboptimales certaines. Considérant une fréquence des prises en charge suboptimales d'environ 60% (76 % dans DIABACT I, avec une hypothèse d'amélioration des soins au cours du temps), la taille de la population cible visée est de 512 patients.

6. REFERENCES BIBLIOGRAPHIQUES

1. Pearson G, Ward-Platt M, Kelly D. How children die: classifying child death Archives of disease in childhood 2010;Epub ahead of print.
2. Harndern A, Mayon-White R, Mant D, Kelly D, Pearson G. Child deaths: confidential enquiry into the role and quality of UK primary care. British Journal of general practice 2009;59:819-24.
3. Nadel S, Britto J, Booy R, et al. Avoidable deficiencies in the delivery of health care to children with meningococcal disease. J Accid Emerg Med 1998;15:298-303.
4. Ninis N, Phillips C, Bailey L, et al. The role of healthcare delivery in the outcome of meningococcal disease in children: case-control study of fatal and non-fatal cases. BMJ 2005;330:1475.
5. Han YY, Carcillo JA, Dragotta MA, et al. Early reversal of pediatric-neonatal septic shock by community physicians is associated with improved outcome. Pediatrics 2003;112:793-9.
6. McIntyre PB, Macintyre CR, Gilmour R, Wang H. A population based study of the impact of corticosteroid therapy and delayed diagnosis on the outcome of childhood pneumococcal meningitis. Arch Dis Child 2005;90:391-6.
7. Chan CL, You HJ, Huang HT, Ting HW. Using an integrated COC index and multilevel measurements to verify the care outcome of patients with multiple chronic conditions. BMC Health Serv Res 2012;12:405.
8. Chen CC, Tseng CH, Cheng SH. Continuity of care, medication adherence, and health care outcomes among patients with newly diagnosed type 2 diabetes: a longitudinal analysis. Med Care 2013;51:231-7.
9. Scanlon MC, Harris JM, 2nd, Levy F, Sedman A. Evaluation of the agency for healthcare research and quality pediatric quality indicators. Pediatrics 2008;121:e1723-31.
10. Chalumeau M, Bouvier-Colle MH, Breart G. Can clinical risk factors for late stillbirth in West Africa be detected during antenatal care or only during labour? Int J Epidemiol 2002;31:661-8.
11. Bocquet A, Chalumeau M, Bollotte D, et al. [Comparison of prescriptions by pediatricians and general practitioners: a population-based study in Franche-Comte from the database of Regional Health Insurance Fund]. Arch Pediatr 2005;12:1688-96.
12. De Boeck K, Vermeulen F, Meyts I, et al. Coprescription of antibiotics and asthma drugs in children. Pediatrics 2011;127:1022-6.
13. Chang LY, Huang LM, Chang IS, Tsai FY. Epidemiological characteristics of varicella from 2000 to 2008 and the impact of nationwide immunization in Taiwan. BMC Infect Dis 2011;11:352.
14. Hanf M, Quantin C, Farrington P, et al. Validation of the French national health insurance information system as a tool in vaccine safety assessment: Application to febrile convulsions after pediatric measles/mumps/rubella immunization. Vaccine 2013;31:5856-62.
15. Chen SC, Wang JD, Hsu HY, et al. Epidemiology of childhood intussusception and determinants of recurrence and operation: analysis of national health insurance data between 1998 and 2007 in Taiwan. Pediatr Neonatol 2010;51:285-91.

16. Launay E, Gras-Le Guen C, Martinot A, et al. Suboptimal care in the initial management of children who died from severe bacterial infection: a population-based confidential inquiry. *Pediatr Crit Care Med* 2010;11:469-74.
17. Floret D. Pediatric deaths due to community-acquired bacterial infection. Survey of French pediatric intensive care units. *Arch Pediatr* 2001;8:S705-11.
18. Kempf M, Baraduc R, Bonnabau H, et al. Epidemiology and antimicrobial resistance of *Streptococcus pneumoniae* in France in 2007: data from the pneumococcus surveillance network. *Microb Drug Resist* 2011;17:31-6.
19. Ministère des affaires sociales et de la santé. Aide mémoire sur les infections invasives à méningocoque. http://www.sante.gouv.fr/IMG/pdf/Annexe__Fiche_aide_memoire_sur_les_infections_invasives_a_meningocoques.pdf.

