



**HAL**  
open science

# Stratégies diagnostiques des pharyngites de l'enfant : du test de diagnostic rapide aux règles de décision clinique

Jérémie Cohen

## ► To cite this version:

Jérémie Cohen. Stratégies diagnostiques des pharyngites de l'enfant : du test de diagnostic rapide aux règles de décision clinique. Médecine humaine et pathologie. Université René Descartes - Paris V, 2014. Français. NNT : 2014PA05S011 . tel-01127375

**HAL Id: tel-01127375**

**<https://theses.hal.science/tel-01127375>**

Submitted on 7 Mar 2015

**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.

ECOLE DOCTORALE DE SANTE PUBLIQUE (ED 420)

---

Doctorat en Santé Publique

Spécialité Epidémiologie

---

Thèse présentée par  
Jérémy COHEN

# Stratégies diagnostiques des pharyngites de l'enfant

## Du test de diagnostic rapide aux règles de décision clinique

Soutenue publiquement le 2 octobre 2014

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

Devant le jury composé de :

Rapporteurs :	Madame Dominique COSTAGLIOLA Monsieur Patrick BOSSUYT
Examineurs :	Madame Claire POYART Monsieur Olivier LORTHOLARY Monsieur Didier GUILLEMOT (président)
Directeur de Thèse :	Monsieur Martin CHALUMEAU



## REMERCIEMENTS

Je remercie Madame **Dominique COSTAGLIOLA** d'avoir accepté d'être rapporteur de cette thèse de doctorat. Vos travaux en épidémiologie clinique et en biostatistique m'ont beaucoup appris. Veuillez recevoir toute ma gratitude et mon plus profond respect.

Je remercie Monsieur **Patrick BOSSUYT** d'avoir accepté d'être rapporteur de cette thèse de doctorat. Votre excellence dans le domaine de l'épidémiologie clinique et de la méthodologie des études diagnostiques est une source d'enseignement pour moi. Veuillez recevoir toute ma gratitude et mon plus profond respect.

Je remercie Madame **Claire POYART** d'avoir accepté de faire partie du jury de cette thèse de doctorat. J'ai eu l'occasion de lire certaines de vos nombreuses publications dans le domaine de l'infection par le streptocoque du groupe A et j'espère que mon intérêt pour ce pathogène m'amènera un jour à travailler avec vous. Veuillez recevoir toute ma gratitude et mon plus profond respect.

Je remercie Monsieur **Olivier LORTHOLARY** d'avoir accepté de faire partie du jury de cette thèse de doctorat. Votre expérience de clinicien et de chercheur dans le domaine des maladies infectieuses est un modèle pour mon exercice futur. Veuillez recevoir toute ma gratitude et mon plus profond respect.

Je remercie Monsieur **Didier GUILLEMOT** d'avoir accepté de présider le jury de cette thèse de doctorat. Vos compétences en épidémiologie des maladies infectieuses,

notamment dans le domaine de la consommation d'antibiotiques et des résistances bactériennes aux antibiotiques, vous permettent de porter sur cette thèse un regard critique qui m'est précieux. Veuillez recevoir toute ma gratitude et mon plus profond respect.

Je remercie Monsieur **Martin CHALUMEAU** de m'avoir proposé puis d'avoir dirigé ce travail doctoral. J'ai eu la chance de bénéficier de ton enseignement, de ton soutien et de ta disponibilité tout au long de ce travail et je t'en remercie sincèrement. J'espère avoir été à la hauteur de la qualité de ton encadrement et de ta grande générosité. Accepte de trouver ici le témoignage de ma sincère reconnaissance et de mon plus profond respect.

Je remercie tous ceux qui ont permis de mener à bien ce travail doctoral :

- Monsieur **Philippe RAVAUD**, directeur de l'unité Inserm U1153, notamment pour son soutien à la poursuite fructueuse de ces travaux ;
- Monsieur **Pierre-Yves ANCEL**, responsable de l'équipe EPOPé, unité Inserm U1153 ;
- Monsieur **François GOFFINET**, directeur de l'unité Inserm U953 ;
- Madame **Corinne LEVY** et Monsieur **Philippe BIDEZ**, qui ont eu un rôle primordial dans l'étude ACTIV TDR SGA et qui ont nourri le travail doctoral de leur expertise ;
- Les pédiatres investigateurs de l'étude ACTIV TDR SGA : les Dr. **Mohamed BENANI**, **François CORRARD**, **Patrice DEBERDT**, **Annie ELBEZ**, **Marie GOLDREY**, **Juliette GOSSELIN**, **Marc KOSKAS**, **Patrick MARTIN**, **Anne-Sylvestre MICHOT**, **Nadia PANIS**, **Douah QUTOB**, **Catherine ROMAIN**, **Olivier ROMAIN**, **Catherine SCHLEMMER**, **Franck THOLLOT** et **Alain WOLLNER** ;

- Les microbiologistes et techniciens de microbiologie de l'étude ACTIV TDR SGA :  
Madame **Patricia MARIANI-KURKDJIAN**, Mademoiselle **Camille D'HUMIÈRES**  
et Madame **Sandrine LIGUORI** ;
- Les membres de l'unité Inserm U953 et de l'équipe EPOPé, et en particulier : Madame  
**Jeanne-Sophie BERLIN**, Madame **Béatrice BLONDEL**, Madame **Marie-Hélène**  
**BOUVIER-COLLE**, Monsieur **Ricardo CARBAJAL**, Madame **Catherine**  
**DENEUX-THARAUX**, Madame **Nine GLANGEAUD**, Madame **Sophie**  
**GOUVAERT**, Madame **Françoise MAILLARD** et Madame **Jennifer ZEITLIN** ;
- Mes collègues et amis étudiants en thèse : **Nathalie BERTILLE**, **Marie-Pierre**  
**BONNET**, **Jean-François BRASME**, **Anne CHANTRY**, **Bénédicte COULM**,  
**Xavier DURRMEYER**, **Romain GUEDJ**, **Cyrille HUCHON**, **Elise LAUNAY** et  
**Karim TARARBIT**.

Je tiens à remercier particulièrement Monsieur **Robert COHEN**. Vous m'avez fait l'honneur de me proposer de travailler avec vous sur un sujet qui vous tient à cœur. J'admire votre énergie et votre enthousiasme et j'espère avoir la chance de pouvoir continuer à collaborer avec vous et à bénéficier de vos enseignements.

Je souhaite remercier chaleureusement Madame **Josette RAYMOND** et Monsieur **Dominique GENDREL**. Votre enseignement clinique et théorique en infectiologie pédiatrique, vos encouragements et votre bienveillance tout au long de mon internat et de ce travail doctoral m'ont aidé à surmonter les obstacles que j'ai pu rencontrer.

Mes remerciements vont à Monsieur **Babak KHOSHNOOD**. Les discussions que nous avons eues autour d'aspects méthodologiques et statistiques de ce travail doctoral me

sont chères. Cela m'a permis d'approfondir mes connaissances et d'éveiller ma curiosité.

Je remercie enfin Madame **Véronique ABADIE**. Vous avez accepté de me prendre comme interne dans le service de pédiatrie générale de l'Hôpital Necker-Enfants Malades au titre de l'Année Médaille de l'Internat et de m'accueillir prochainement comme Chef de Clinique Assistant. Je suis enthousiasmé par cette perspective clinique et espère qu'il s'agit du début d'une longue et fructueuse collaboration.

Ce travail est également dédié à la mémoire du Professeur **Edouard BINGEN**, chef du service de microbiologie de l'Hôpital Robert Debré à Paris, décédé soudainement en décembre 2012. Il restera une source d'enseignements inépuisables et travailler à ses côtés sur ce travail de recherche a constitué un grand honneur pour moi.

*À mes parents, Brigitte et Olivier,*

*À ma sœur, Léa et mon frère, Benjamin,*

*À toute ma famille, pour leur soutien indéfectible,*

*À ma compagne, Claire, et à notre fille, Clémence, pour leur patience et leur amour,*

*A Chantal et Yves pour leur grande générosité,*

*À mes collègues internes et chefs de clinique, pour tous nos bons moments passés ensemble, à l'hôpital et en dehors,*

*À mes amis de longue date,*

*Camille, Clara et Daniel, Faustine et Landry, François, Julien, Loriane et Mathieu, Maud et Antoine, Nomonde et Baptiste, Octavie et Mathias, Sarah et Thibault, Sophie et Sami, pour leur soutien, leur amitié et leur joie de vivre.*

**Je dédie cette thèse.**



## **Résumé (1594 caractères, espaces compris)**

**Introduction** – La place des tests de diagnostic rapide (TDR) et des règles de décision cliniques (RDC) pour le diagnostic des pharyngites à streptocoque du groupe A (SGA) chez l'enfant varie selon les recommandations internationales en raison de doutes sur la stabilité des performances diagnostiques du TDR et d'une validation insuffisante des RDC.

**Méthodes** – Dans une étude prospective multicentrique (n=17) ambulatoire réalisée au sein du réseau clinique pédiatrique ACTIV de 2009 à 2011, 1776 enfants avec pharyngite ou sains ont été soumis à des prélèvements de gorge pour réaliser un TDR et une mise en culture (test de référence). Nous avons étudié l'effet indépendant de variables liées aux patients et aux médecins sur les performances diagnostiques du TDR, exploré systématiquement les faux-positifs (FP) du TDR et réalisé une validation externe et une comparaison des RDC existantes.

**Résultats** – La sensibilité du TDR (en moyenne 87%) variait selon la présentation clinique (âge, signes cliniques), l'inoculum bactérien et le phénomène de portage (paramètres aussi liés entre eux), et selon des variables liées aux médecins (dont le type d'activité clinique). La valeur prédictive négative du TDR était élevée ( $\approx 90\%$ ) et stable. Les FP du TDR étaient positifs pour le SGA en PCR. Aucune RDC n'était satisfaisante en termes de calibration et de discrimination.

**Conclusion** – Le TDR est suffisant pour le diagnostic de pharyngite à SGA si les cliniciens évaluent leurs propres performances et les améliorent si besoin. Aucune RDC ne peut être recommandée en pratique clinique en pédiatrie.

**Mots-clés** – pharyngite, angine, streptocoque du groupe A, *Streptococcus pyogenes*, techniques et procédures diagnostiques, sensibilité et spécificité, valeur prédictive des tests, décision clinique, test rapide, règles de décision clinique, revue systématique, validation externe, enfant, adolescent



**Abstract (1495 characters, with spaces)**

**Background** – The roles of rapid antigen detection tests (RADT) and clinical prediction rules (CPR) for the diagnosis of group A streptococcus (GAS) in children with pharyngitis vary across international clinical guidelines. This might be related to unstable diagnostic accuracy of RADTs and insufficient validation of CPRs.

**Methods** – In a prospective multicenter (n=17) office-based study that took place in France within the ACTIV network between 2009 and 2011, 1776 children with pharyngitis or healthy controls underwent throat swabs to perform a RADT and a throat culture (reference standard). We assessed the independent effect of patient- and physician-level characteristics on the accuracy of a RADT, systematically re-analyzed RADT false-positive results, and externally validated and compared existing CPRs.

**Results** – RADT sensitivity (overall 87%) varied according to clinical signs and symptoms, bacterial inoculum size and GAS throat carriage (factors also related to each other), and according to physician-level characteristics (including type of clinical practice). RADT negative predictive value was high (about 90%) and stable. RADT false-positives were positive for GAS when using a new PCR technique. No CPR had sufficient performances regarding calibration and discrimination.

**Conclusions** – RADTs are sufficient for diagnosing GAS pharyngitis if clinicians accept diagnostic accuracy monitoring and adequate training when needed. No CPR can be recommended for use in pediatrics.

**Keywords** – pharyngitis, group A streptococcus, *Streptococcus pyogenes*, diagnostic techniques and procedures, sensitivity and specificity, predictive value of tests, decision making, rapid test, clinical decision rules, systematic review, external validation, child, adolescent

*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 : Docteur 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*

*Cette thèse a été préparée en collaboration étroite avec le groupe ACTIV (Association Clinique et Thérapeutique Infantile du Val-de-Marne) et avec l'équipe du laboratoire de microbiologie de l'hôpital Robert Debré, Paris.*

*Cette thèse est dédiée à la mémoire du Pr Edouard BINGEN (1946-2012), chef du service de bactériologie de l'hôpital Robert Debré à Paris, décédé soudainement en décembre 2012. Il restera une source d'enseignements inépuisables et travailler à ses côtés sur ce travail de recherche a constitué un grand honneur pour moi.*

## **ABREVIATIONS**

ACTIV – Association Clinique et Thérapeutique Infantile du Val-de-Marne

FP – faux-positif

OR – *odds ratio*

ORa – *odds ratio* ajusté

PCR – *polymerase chain reaction*

RDC – règle de décision clinique

ROC – *receiver operating characteristics*

SGA – streptocoque du groupe A

TDR – test de diagnostic rapide

# TABLE DES MATIERES

Publications et communications issues du travail doctoral	6
Glossaire	11
Introduction générale	25
Sources de données	35
Chapitre 1 Effet de l'inoculum bactérien et du spectre clinique sur les performances diagnostiques du TDR <i>PLoS One 2012; 7(6): e39085</i>	45
Chapitre 2 Effet de l'inoculum bactérien, du spectre clinique et des caractéristiques des médecins sur la sensibilité du TDR <i>Eur J Clin Microbiol Infect Dis. 2013;32(6):787-793</i>	63
Chapitre 3 Analyse des faux-positifs du TDR : une approche cas-témoins <i>J Pediatr. 2013;162(6):1282-1284</i>	81
Chapitre 4 Revue systématique et étude de validation externe des règles de décision clinique pour le diagnostic des pharyngites chez l'enfant <i>Canadian Medical Association Journal (en révision)</i>	95
Conclusion et perspectives de recherche	125
Bibliographie	135
Annexes	145

# PUBLICATIONS ET COMMUNICATIONS

## ISSUES DU TRAVAIL DOCTORAL

### **PUBLICATIONS ORIGINALES**

Cohen JF, Cohen R, Levy C, Thollot F, Benani M, Bidet P, Chalumeau M. Clinical prediction rules for group A streptococcus in children with pharyngitis: systematic review and external validation study. (*Canadian Medical Association Journal*, en révision)

Cohen JF, Cohen R, Chalumeau M. Rapid antigen detection test for group A streptococcus in children with pharyngitis (Protocol). *Cochrane Database of Systematic Reviews* 2013, Issue 4. Art. No.: CD010502. DOI:10.1002/14651858.CD010502.

Cohen JF, Cohen R, Bidet P, Levy C, Deberdt P, d'Humières C, Liguori S, Corrad F, Thollot F, Mariani-Kurkdjian P, Chalumeau M, Bingen E. Rapid-antigen detection tests for group A streptococcal pharyngitis: revisiting false-positive results using PCR testing. *The Journal of Pediatrics*. 2013;162(6):1282-1284.

Cohen JF, Chalumeau M, Levy C, Bidet P, Benani M, Koskas M, Bingen E, Cohen R. Effect of clinical spectrum, inoculum size and physician characteristics on sensitivity of a rapid antigen detection test for group A streptococcal pharyngitis. *European Journal of Clinical Microbiology and Infectious Diseases*. 2013;32(6):787-793.

Cohen JF, Chalumeau M, Levy C, Bidet P, Thollot F, Wollner A, Bingen E, Cohen R. Spectrum and inoculum size effect of a rapid antigen detection test for group A streptococcus in children with pharyngitis. *PLoS One* 2012; 7(6): e39085.

## COMMUNICATIONS

*\*Les communications orales sont signalées par un astérisque.*

### Communications dans des congrès internationaux à comité de sélection

Cohen JF, Cohen R, Levy C, Thollot F, Benani M, Bidet P, Chalumeau M. Selective testing strategies for group A streptococcus in children with pharyngitis: systematic review and prospective multicenter external validation study. *Annual Meeting of the European Society for Paediatrics Infectious Diseases*, Dublin, Irlande (mai 2014)

Cohen JF, Chalumeau M, Levy C, Bidet P, Benani M, Koskas M, Bingen E, Cohen R. Effect of clinical spectrum, inoculum size and physician characteristics on sensitivity of a rapid detection test for group A streptococcus pharyngitis. *Current Diagnostic and Therapeutic Dilemmas in the Clinical Management of Group A Streptococcal Infections*, Rome, Italie (mars 2013)

Cohen JF, Cohen R, Bidet P, Levy C, Deberdt P, d'Humières C, Liguori S, Corrad F, Thollot F, Mariani-Kurkdjian P, Chalumeau M, Bingen E. Rapid antigen detection tests for group A streptococcal pharyngitis: revisiting false-positive results with PCR. *Current Diagnostic and Therapeutic Dilemmas in the Clinical Management of Group A Streptococcal Infections*, Rome, Italie (mars 2013)

*\*Cohen JF*, Chalumeau M, Levy C, Bidet P, Thollot F, Wollner A, Bingen E, Cohen R. Spectrum and inoculum size effect of a rapid antigen detection test for streptococcal pharyngitis in children. *Annual Meeting of the European Society for Paediatrics Infectious Diseases*, Thessalonique, Grèce (juin 2012)

Cohen JF, Levy C, Bidet P, Benani M, Thollot F, Koskas M, Chalumeau M, Bingen E, Cohen R. Sensitivity of a rapid test for group A streptococcus in children with pharyngitis and healthy controls: a multi-level approach. *Annual Meeting of the European Society for Paediatrics Infectious Diseases*, Thessalonique, Grèce (juin 2012)

*\*Cohen JF*, Cohen R, Bidet P, Levy C, Deberdt P, d'Humières C, Liguori S, Corrad F, Mariani P, Chalumeau M, Bingen E. False-positive results of rapid antigen detection tests (RADT) for group A streptococcus (GAS) pharyngitis: a case-control study.

*Annual Meeting of the European Society for Paediatrics Infectious Diseases, Thessalonique, Grèce (juin 2012)*

*Communications dans des congrès nationaux sur invitation*

\*Cohen JF. Test de diagnostic rapide dans les angines de l'enfant. *Journée de Pathologie Infectieuse Pédiatrique Ambulatoire*, Paris, France (octobre 2014)

\*Cohen JF. Test de diagnostic rapide dans les angines de l'enfant. *Journée Annuelle du Groupe de Pathologie Infectieuse Pédiatrique*, Paris, France (juin 2014)

\*Cohen JF, Cohen R, Thollot F, Benani M, Bidet P, Levy C, Chalumeau M. Règles de décision clinique pour les angines de l'enfant. *Congrès National de la Société Française de Pédiatrie*, Lyon, France (mai 2014)

\*Cohen JF, Levy C, Bidet P, Thollot F, Wollner A, Mariani-Kurkdjian P, Chalumeau M, Bingen E, Cohen R. Test de diagnostic rapide du streptocoque du groupe A dans les angines de l'enfant : effet et biais de spectre. *Congrès National de la Société Française de Pédiatrie*, Bordeaux, France (juin 2012)

\*Cohen JF, Levy C, Bidet P, Benani M, Thollot F, Koskas M, Chalumeau M, Bingen E, Cohen R. Sensibilité du test de diagnostic rapide du streptocoque du groupe A chez les porteurs sains et les enfants ayant une angine. *Congrès National de la Société Française de Pédiatrie*, Bordeaux, France (juin 2012)

\*Cohen JF, Cohen R, Bidet P, Levy C, Deberdt P, d'Humières C, Liguori S, Corrad F, Thollot F, Mariani-Kurkdjian P, Chalumeau M, Bingen E. Faux-positifs du test de diagnostic rapide du streptocoque du groupe A dans les angines de l'enfant. *Congrès National de la Société Française de Pédiatrie*, Bordeaux, France (juin 2012)

\*Cohen JF, Levy C, Mariani-Kurkdjian P, Chalumeau M, Bingen E, Cohen R. Test de diagnostic rapide du streptocoque du groupe A dans les angines de l'enfant. *Journées Parisiennes de Pédiatrie*, Paris, France (octobre 2012)

# SOURCES DE FINANCEMENT

## Bourses :

- Prix Guigoz – Groupe de Recherches Epidémiologiques en Pédiatrie – Société Française de Pédiatrie, 2010 ;
- Année Recherche de l'inter-région Ile-de-France, 2011.

## Prix et concours :

Année Médaille de l'Internat de l'Assistance Publique - Hôpitaux de Paris, 2012.

---

# GLOSSAIRE

*Note : La lecture de ce glossaire n'est pas nécessaire à la compréhension de la thèse. Il permettra toutefois au lecteur de retrouver rapidement la définition d'un terme d'épidémiologie ou d'infectiologie utilisé au cours du travail doctoral. Il est séparé en deux parties : Epidémiologie et Infectiologie. Les termes présents dans le glossaire sont présentés par ordre alphabétique dans chacune des deux parties et sont identifiés par un astérisque dans la thèse.*

## **1. Epidémiologie**

### **Analyse stratifiée**

Une analyse stratifiée est une analyse répétée dans différents groupes de l'échantillon de population étudié. Ainsi, réaliser une analyse de l'effet de la variable A sur la variable d'intérêt Y en stratifiant sur la variable B, revient à répéter l'analyse de l'effet de A sur Y de manière séparée en fonction du niveau de la variable B. Par exemple, on a montré que l'âge (A) avait un effet sur Y et on se pose la question de savoir si l'effet de l'âge sur Y est différent chez les hommes et les femmes (B). On va alors étudier l'effet de A sur Y, au sein de la strate hommes de la variable B puis de la strate femmes.

### **Biais de mesure**

Il existe un biais de mesure lorsqu'il existe une erreur de mesure des variables étudiées et que cela entraîne une erreur systématique dans le paramètre estimé par l'étude. Il peut par exemple exister un biais de mesure dans une étude diagnostique portant sur un examen d'imagerie, si le test étudié (test index) n'est pas réalisé de la même manière chez les sujets malades que chez les sujets non-malades car le résultat du test index n'est pas mesuré à l'aveugle du résultat du test de référence. Par exemple, si la personne lisant l'examen d'imagerie passe plus de temps à lire les images des sujets malades, on s'attend à ce que cela entraîne une surestimation biaisée de la sensibilité du test index.

### **Biais de sélection**

Il existe un biais de sélection lorsque la sélection des sujets étudiés entraîne une erreur systématique dans le paramètre estimé par l'étude. Dans une étude diagnostique, il peut exister un biais de sélection si les sujets n'ont pas été échantillonnés au hasard ou s'il s'agit d'une étude cas-témoins dans laquelle la sélection des témoins est inappropriée, et que cela entraîne une distorsion dans la mesure des performances diagnostiques.

**Biais de vérification**

Le biais de vérification est un type de biais de mesure rencontré dans les études diagnostiques. On parle de biais de vérification partielle lorsque tous les sujets ne sont pas soumis au test étudié (test index) et au test de référence. Par exemple, si seuls les résultats négatifs du test étudié sont vérifiés par le test de référence, en considérant que le test index n'a pas de résultat faux-positif, la spécificité sera considérée de manière biaisée comme étant de 100%. On parle de biais de vérification différentielle lorsque les sujets ne reçoivent pas le même test de référence en fonction du résultat du test index. Par exemple, si les tests négatifs sont vérifiés par un test de référence A mais que les test positifs sont vérifiés par un test de référence B, plus invasif et plus performant que le test A, cela conduira à une surestimation biaisée de sensibilité et une sous-estimation biaisée de la spécificité du test index.

**Calibration et discrimination**

La calibration et la discrimination sont des mesures de performance d'un modèle de prédiction clinique. La calibration (ou adéquation) reflète dans quelle mesure les probabilités d'évènement observées dans la population concordent avec les probabilités d'évènement prédites par le modèle. Le test d'Hosmer-Lemeshow est fréquemment utilisé pour évaluer la calibration d'un modèle logistique. L'évaluation de la discrimination estime dans quelle mesure le modèle est capable de classer correctement les patients (habituellement en malades ou non malades). L'aire sous courbe ROC est une mesure classique de discrimination d'un modèle logistique.

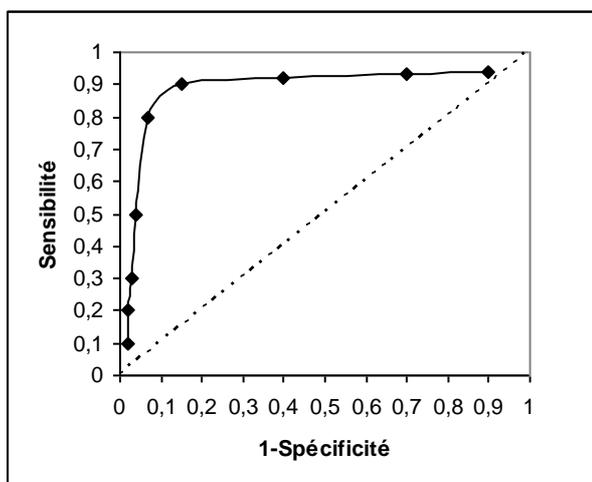
**Convergence**

Une procédure fréquemment utilisée en épidémiologie pour estimer les paramètres d'un modèle est la méthode du maximum de vraisemblance. L'identification des valeurs des paramètres maximisant la vraisemblance d'un modèle nécessite de dériver certaines fonctions et de trouver des valeurs de paramètres pour lesquelles la dérivée de ces fonctions est nulle. Pour la plupart des modèles, ces équations n'ont pas de solution analytique explicite et doivent alors être résolues grâce à des méthodes numériques. Ces méthodes numériques sont habituellement des procédures itératives, telles que l'algorithme de Newton-Raphson. Le processus est répété de manière itérative jusqu'à ce que la différence entre les estimations des paramètres d'une itération à l'autre passe sous un certain seuil, on parle alors de convergence. Dans certains cas, l'algorithme ne parvient pas à maximiser la fonction de vraisemblance et le modèle ne peut converger, ce qui empêche d'obtenir une estimation des paramètres du modèle. On peut parfois contourner ce problème en modifiant le nombre d'itérations ou le seuil de convergence.

**Courbe ROC**

Lorsque le test diagnostique étudié n'est pas dichotomique (positif/négatif) mais que le résultat est une variable continue, par exemple dans le cas de dosages biologiques, il existe une sensibilité et une spécificité pour chaque seuil de positivité potentiel du test. Lorsque l'on cherche un seuil pour utiliser le test en pratique, on peut s'aider d'une courbe (figure infra) représentant le taux de vrais-positifs

(Sensibilité) en fonction du taux de faux-positifs ( $1$ -Spécificité). Cette courbe est la courbe ROC (Receiver Operating Characteristics).



**Exemple de courbe ROC.** La ligne pointillée représente la courbe ROC d'un test non informatif. Chaque losange correspond à la sensibilité et à la spécificité du test pour un seuil donné et la ligne pleine est la courbe ROC du test.

### Effet médecin

Dans cette thèse, l'effet médecin a été défini par l'observation de variations de performances diagnostiques d'un test entre différents médecins réalisant le test. Ce concept est à rapprocher « l'effet centre » souvent observé dans les études interventionnelles multicentriques lorsque l'effet d'une intervention dépend du centre dans lequel elle a été expérimentée.

### Effet de spectre

Les performances d'un test diagnostique peuvent varier en fonction du spectre de la maladie, avec habituellement une sensibilité plus élevée chez les patients les plus symptomatiques ou ayant une forme de maladie plus avancée. Ainsi, les caractéristiques cliniques des patients dans un échantillon d'étude peuvent influencer les performances diagnostiques du test étudié.

### Facteurs contextuels et organisationnels

La pratique médicale et la prise de décision diagnostique et thérapeutique sont des phénomènes complexes influencés par de nombreux facteurs. Parmi ces facteurs, on distingue habituellement les facteurs liés directement au médecin (tels que ses caractéristiques démographiques, sa formation professionnelle et son expérience clinique), des facteurs contextuels et organisationnels. Les facteurs contextuels correspondent aux caractéristiques de la population que le médecin est amené à soigner, tant sur le plan socio-économique que sur le plan épidémiologique. Les facteurs organisationnels correspondent principalement au type de pratique clinique exercée par le médecin, par exemple cabinet individuel ou cabinet de groupe, accès à des ressources numériques de formation ou d'aide à la prescription, ou encore participation à un réseau ville-hôpital. Les facteurs contextuels et organisationnels caractérisent des groupes d'individus (par exemple un groupe de patients vus par un même médecin) et doivent être mesurés comme tels et pris en compte comme tels dans l'analyse.

**Interaction**

Il existe une interaction entre la variable A et la variable B lorsque l'effet de la variable A sur la variable d'intérêt Y dépend du niveau de la variable B. Par exemple, si l'effet de l'âge (A) sur Y est significativement plus important chez les femmes que chez les hommes (B), il existe une interaction entre l'âge et le genre. L'effet de l'âge sur Y ne peut alors être résumé par une seule mesure d'association.

**Loi binomiale**

La loi binomiale est une loi de probabilité de paramètres  $p$  et  $n$  décrivant le comportement d'une variable  $X$  de comptage du nombre de succès d'expériences binaires de Bernoulli répétées de manière indépendante, où  $p$  est la proportion de succès et  $n$  le nombre d'expériences indépendantes. Cette loi est fréquemment utilisée en épidémiologie lorsqu'on s'intéresse à des événements binaires et à des proportions, comme c'est le cas dans les études diagnostiques où la sensibilité et la spécificité peuvent être vues comme des proportions de succès du test index, respectivement chez les malades et chez les non-malades. On peut choisir de construire des modèles de sensibilité et de spécificité d'un test diagnostique en étudiant l'association de variables à ces quantités à une transformation près : transformation vers le logarithme de la cote (fonction de lien *logit*), transformation logarithmique (fonction de lien *log*), ou sans aucune transformation (fonction de lien *identité*).

**Méta-analyse**

Une méta-analyse est une technique statistique permettant de réaliser une synthèse quantitative des données issues d'une revue systématique. La méta-analyse diagnostique consiste à combiner les données issues d'études existantes (données issues des tableaux de contingence de chaque étude) puis à les réanalyser afin d'en tirer des valeurs résumées de performances diagnostiques (habituellement sensibilité et spécificité). Les premières techniques de méta-analyse impliquaient de simples calculs de moyennes pondérées tandis que les techniques les plus récentes impliquent des modèles hiérarchiques.

**Méthode binomiale exacte**

La loi de probabilité d'une variable aléatoire  $X$  égale au nombre de succès d'une épreuve de Bernoulli répétée de manière indépendante  $n$  fois est une loi binomiale. Il existe plusieurs méthodes pouvant être utilisées pour calculer l'intervalle de confiance d'une proportion de succès ( $p$ ). La méthode la plus simple et probablement la plus employée est l'approximation par la loi normale. Avec cette méthode, on estime la loi binomiale par une loi normale, bien connue et conduisant à des calculs simples. Parfois cette approximation n'est pas acceptable, notamment sur de petits échantillons ou lorsque la proportion estimée est proche de 0 ou de 1. Dans ces cas, il est plus adapté d'utiliser la méthode binomiale exacte, qui au lieu de passer par une approximation par une loi normale se base directement sur une véritable loi binomiale.

**Modèle hiérarchique**

Un modèle hiérarchique, ou modèle multiniveaux, a pour but de prendre en compte dans un modèle le fait que les données ont une structure hiérarchique. C'est par exemple le cas dans les études multicentriques dans lesquelles on peut considérer plusieurs niveaux : un premier niveau correspondant aux individus, et

un deuxième niveau correspondant aux centres dans lesquels les individus ont été recrutés. 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 individus et variabilité entre centres, car il existe un certain degré de corrélation entre individus au sein d'un centre. S'il existe une variabilité significative entre centres on parle habituellement d'hétérogénéité. Ce concept est également utilisé en méta-analyse lorsque l'on veut prendre en compte la variabilité intra-étude et la variabilité inter-études.

### **Performances diagnostiques d'un test**

Les performances diagnostiques (*diagnostic accuracy*) désignent la capacité d'un test à identifier correctement les sujets malades et les sujets non-malades, ce statut étant lui-même défini par le résultat d'un test de référence. Les performances d'un test diagnostiques sont habituellement mesurées par les indices de sensibilité et spécificité\*, valeurs prédictives\* positives et négatives, rapports de vraisemblance\* positifs et négatifs ou aire sous la courbe ROC\*.

### **Rapport de cote (*odds ratio*)**

La cote est une manière courante d'exprimer une probabilité, notamment dans le domaine des paris hippiques. La cote est le rapport entre la probabilité d'observer un événement (« cheval gagnant ») et la probabilité d'observer l'évènement complémentaire (« cheval perdant »). Par exemple, si la probabilité d'évènement est de 0,2, la cote sera de 0,2 : 0,8 soit 1:4. Le rapport de cote (*odds ratio*) est le rapport entre deux cotes, typiquement entre 2 groupes se distinguant par une caractéristique dichotomique, par exemple un groupe exposé à un certain facteur de risque et un groupe non-exposé au facteur de risque. Le rapport de cotes est une mesure classique d'association en épidémiologie.

### **Rapports de vraisemblance d'un test diagnostique**

Le principe des rapports de vraisemblance diagnostique est de relier la probabilité pré-test, c'est-à-dire la probabilité de la maladie avant d'avoir réalisé le test diagnostique, et la probabilité post-test, c'est-à-dire la probabilité de la maladie une fois qu'on dispose du résultat du test. Le théorème de Bayes permet d'estimer la probabilité de la maladie chez un patient en fonction de la prévalence de la maladie dans l'échantillon (probabilité pré-test) conditionnellement au résultat du test. En réalité, les rapports de vraisemblance n'expriment pas un rapport de probabilités mais un rapport de cotes.

Pour le rapport de vraisemblance en cas de test positif (RVP) :

$$RVP = \text{Cote post-test en cas de test positif} / \text{cote pré-test}$$

Et pour le rapport de vraisemblance en cas de test négatif (RVN) :

$$RVN = \text{Cote post-test en cas de test négatif} / \text{cote pré-test}$$

Les rapports de vraisemblance peuvent également être exprimés en tant que fonctions de la sensibilité et de la spécificité du test étudié. On enseigne donc classiquement que les rapports de vraisemblance ne varient pas avec la prévalence de la maladie dans l'échantillon. Nous verrons au cours de la thèse que cette affirmation est fautive et que les rapports de vraisemblance ne sont pas des caractéristiques fixes d'un test.

### Règles de décision clinique

Une règle de décision clinique (RDC) est un outil d'aide à la décision médicale basé sur la combinaison d'éléments de l'interrogatoire du patient, de l'examen clinique ou d'examen paracliniques de routine. Une RDC fournit au clinicien une estimation de risque et/ou une conduite à tenir à l'échelle du patient, par exemple la réalisation d'un examen complémentaire ou la mise en route d'un traitement.

### Régression logistique

La régression logistique est un modèle mathématique qui permet d'étudier l'association de variables à une variable d'intérêt binaire, par exemple la présence ou l'absence de maladie. Dans la régression logistique multivariée, les associations sont estimées sous forme d'*odds ratio* ajustés, ce qui permet d'identifier les prédicteurs associés à la variable prédite « indépendamment » des autres prédicteurs (c'est-à-dire tout autre paramètre fixé). La régression logistique est un outil classique pour la construction de modèles de prédiction clinique.

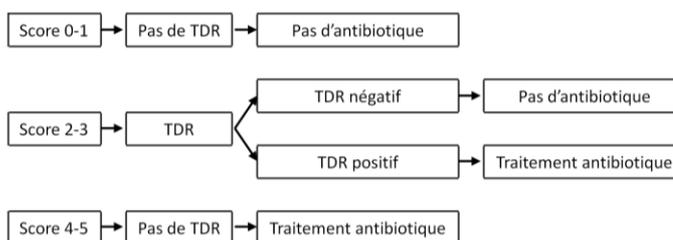
### Revue systématique

Une revue systématique est une démarche protocolisée de revue critique de la littérature. L'auteur tente de regrouper toute l'information scientifique disponible sur un sujet afin de minimiser les biais liés aux éléments arbitraires et subjectifs existants dans une revue non systématique (revue « narrative »). La revue systématique doit suivre une méthodologie rigoureuse afin d'assurer 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 (lecture critique standardisée). La revue systématique peut être suivie d'une méta-analyse dont elle est distincte (voir méta-analyse).

### Score de McIsaac

Le score de McIsaac est un score clinique prédictif de la présence du SGA en culture de prélèvement de gorge chez les patients ayant une pharyngite. Ce score est issu d'équations de régression logistique, et va de 0 à 5. Il a été montré qu'il existait une assez bonne corrélation entre le niveau de ce score et la probabilité d'isoler un SGA en culture de prélèvement de gorge. Il est basé sur les mêmes éléments cliniques que le score de Centor (destiné aux adultes) avec en plus un critère d'âge, et s'applique donc aux enfants et aux adultes. Une règle de décision clinique basée sur le score de McIsaac a également été développée :

Signe clinique	Point
Température >38°C	1
Absence de toux	1
Adénopathies cervicales sensibles	1
Gonflement / exsudat amygdalien	1
Age	
3-14 ans	1
15-44 ans	0
≥ 45 ans	-1



**Score de McIsaac et règle de décision clinique qui en est issue.** TDR, test de diagnostic rapide.

### Sensibilité et spécificité d'un test diagnostique

La sensibilité est la proportion de sujets ayant un test index positif parmi les sujets malades, c'est-à-dire la proportion de malades correctement détectés par le test étudié. La spécificité est la proportion de sujets ayant un test index négatif parmi les sujets non malades, c'est-à-dire la proportion de non-malades correctement détectés par le test étudié.

D'après le tableau de contingence (voir ci-dessous) on obtient :

$$\text{Sensibilité} = a / (a+c)$$

$$\text{Spécificité} = d / (b+d)$$

Il est classiquement enseigné que la sensibilité et la spécificité sont des caractéristiques fixes des tests diagnostiques. Nous verrons dans cette thèse qu'il existe des preuves mathématiques et empiriques que cette simplification est fautive.

### Tableau de contingence

Le tableau de contingence est une façon de représenter simultanément deux caractères observés sur une population. Dans une étude diagnostique, ces deux caractères sont le fait d'être malade (M+ ou M-), défini par la positivité du test de référence, et le résultat du test étudié (test index ; T+ ou T-).

	M+	M-
T+	<i>a</i>	<i>b</i>
T-	<i>c</i>	<i>d</i>

*a* : nombre de sujets malades correctement classés par le test index (vrais-positifs)

*b* : nombre de sujets non malades mal classés par le test index (faux-positifs)

*c* : nombre de sujets malades mal classés par le test index (faux-négatifs)

*d* : nombre de sujets non malades correctement classés par le test index (vrais-négatifs)

### Test index et test de référence

Pour évaluer les performances diagnostiques d'un test index, c'est-à-dire sa capacité à classer correctement les patients en malades ou non-malades, il faut que ce statut ait lui-même été défini par le résultat du test de référence. Les patients ayant un résultat positif pour le test de référence sont considérés comme malades et les patients ayant un résultat négatif pour le test de référence sont considérés comme non-malades.

### Tests du khi-2

Les tests du khi-2 sont des tests statistiques utilisés en épidémiologie pour comparer des proportions entre différents groupes d'observations indépendantes. Le test du khi-2 de Pearson est le test du khi-2 classique, le test du khi-2 de tendance est un test qui cherche à mettre en évidence une tendance dans des proportions entre différents groupes ordonnés (par exemple par âge). Le test du khi-2 de Cochran (test *Q*) n'est pas directement un test sur les proportions mais plutôt un test d'hétérogénéité utilisé en méta-analyse dont la formulation est proche de celle d'un test d'interaction sur des rapports de cote.

**Valeurs prédictives d'un test diagnostique**

La valeur prédictive positive (*VPP*) est la proportion de sujets malades parmi les sujets ayant test index positif :  $VPP = a / (a+b)$ .

La valeur prédictive négative (*VPN*) est la proportion de sujets non malades parmi les sujets ayant test index négatif :  $VPN = d / (c+d)$ .

La *VPP* et la *VPN* sont directement dépendantes de la prévalence de la maladie dans l'échantillon. La *VPP* augmente et la *VPN* diminue quand la prévalence de la maladie augmente (toutes choses égales par ailleurs).

**Validation interne et externe**

L'étape de validation interne consiste à appliquer un modèle de prédiction clinique à la population ayant servi à construire le modèle afin d'avoir une première estimation des performances du modèle. Il existe plusieurs méthodes de validation interne, principalement la partition de l'échantillon et les méthodes de ré-échantillonnage. L'étape de validation externe consiste à appliquer le modèle à une population différente et dans un contexte différent de ceux qui ont servi à la construction du modèle. Cela permet d'évaluer la "transportabilité" du modèle.

## 2. *Infectiologie*

### **Anticorps anti-streptolysine O (ASLO) et anti-streptodornase (ASD)**

Ces anticorps sont produits par le système immunitaire en réaction à une infection par le streptocoque du groupe A (SGA). Les ASLO ciblent la streptolysine O, un facteur de virulence du SGA qui confère à la bactérie une capacité d'hémolyse. Les ASD ciblent la streptodornase (ou désoxyribonucléase B, ou DNase B), enzyme du SGA qui permet de lyser l'ADN. Une élévation du titre des ASD est considérée comme étant spécifique d'une infection à SGA alors que les ASLO peuvent aussi être élevés lors d'infection dues à des streptocoques des groupes C et G.

### **Beta-hémolyse**

Le streptocoque du groupe A (SGA) est une bactérie capable de  $\beta$ -hémolyse. Cette caractéristique biologique est un facteur de pathogénicité du SGA et est utilisée au laboratoire de bactériologie pour identifier le SGA sur les milieux de culture contenant des hématies (« gélose au sang »). La  $\beta$ -hémolyse est définie par la destruction complète des globules rouges présents dans le milieu de culture autour de la colonie. L' $\alpha$ -hémolyse, définie par une destruction incomplète des globules rouges, donne une coloration verdâtre aux colonies car l'hémoglobine est oxydée en méthémoglobine. Des cas exceptionnels de souches pathogènes de SGA non- $\beta$ -hémolytiques ont été rapportés. Les streptocoques peuvent également être classifiés sur la base de leur antigène de Lancefield (polyoside C, un antigène de paroi), en streptocoques des groupes A à V. Le SGA est la bactérie la plus responsable de pharyngite, mais les streptocoques  $\beta$ -hémolytiques des groupes C et G sont également susceptibles d'entraîner des pharyngites.

### **Biofilm**

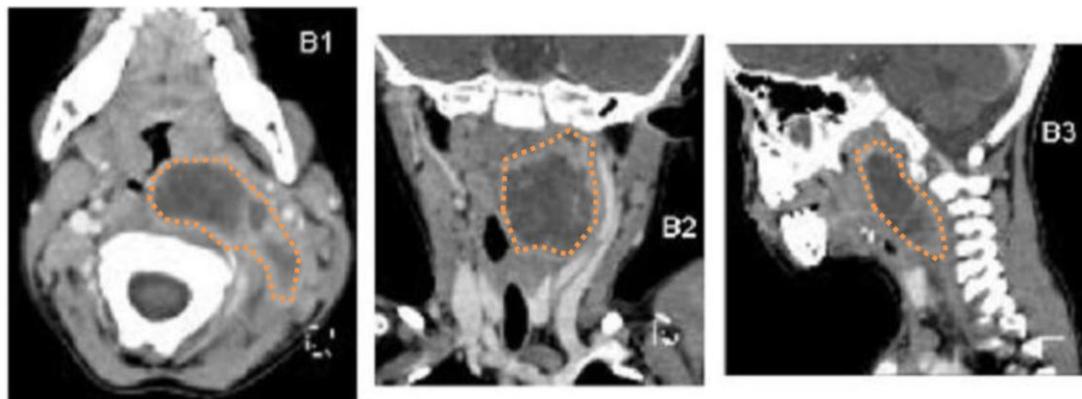
Un biofilm est une communauté de bactéries reliées entre elles et attachées à un substrat (vivant ou inerte) et incluse dans la matrice extracellulaire qu'elle a elle-même produite. Au sein d'un biofilm, les bactéries présentent un phénotype particulier, en particulier en ce qui concerne leur vitesse de croissance et l'expression de certains gènes et antigènes.

### **Complications locorégionales des pharyngites à SGA**

Ces complications suppuratives sont liées à la diffusion locale de l'infection et de l'inflammation. On distingue les complications suppuratives suivantes :

- adénite aiguë bactérienne ;
- phlegmon périamygdalien (*quinsy*) ;
- infections rétropharyngées et péripharyngées, collectées ou non ;
- cellulite cervicale profonde.

Il y a très peu de données récentes concernant l'incidence des complications locorégionales des pharyngites à SGA. Certaines études rapportent un taux d'incidence de phlegmon périamygdalien dans les suites d'une pharyngite de l'ordre de 0.1%.



*Tomodensitométrie avec injection de produit de contraste montrant un abcès rétropharyngé chez une fille de 3 ans (entouré en orange). B1 : coupe axiale ; B2 : coupe coronale ; B3 : coupe sagittale. D'après « Société Française d'Otorhinolaryngologie et de Chirurgie de la face et du cou. Complications locorégionales des pharyngites (recommandation pour la pratique clinique), 2008. »*

### **Complications non-suppuratives des pharyngites à SGA**

Ces complications sont dues à un mécanisme dysimmunitaire :

- rhumatisme articulaire aigu (RAA) : atteinte systémique avec fièvre, atteinte cardiaque (dont valvulopathie mitrale), polyarthrite, atteinte cutanée (nodules sous-cutanés de Meynet, érythème marginé) et atteinte neurologique (chorée de Sydenham) ;
- érythème noueux ;
- glomérulonéphrite aiguë post-streptococcique (GNA) : atteinte rénale avec insuffisance rénale aiguë, hématurie, protéinurie et hypertension artérielle.

L'incidence du RAA est actuellement inférieure à 5 cas /100000 enfants par an dans les pays industrialisés mais peut dépasser les 100 cas /100000 enfants par an dans certains pays en développement. En France, une étude menée entre 1995 et 1997 a rapporté un taux d'incidence nationale annuelle de l'ordre de 0,15/100000 enfants (29 cas en 3 ans). L'incidence des GNA est de l'ordre de 5 cas /100000 enfants par an dans les pays développés et 25 cas /100000 enfants par an dans les pays en développement.

### **Culture de prélèvement de gorge au laboratoire de bactériologie**

La culture est le test de référence pour mettre en évidence le SGA à partir d'un prélèvement de gorge. L'écouvillon, prélevé au niveau des amygdales, est mis en culture sur gélose d'agar additionnée de 5% de sang de cheval ou de mouton et incubé en atmosphère enrichie en CO<sub>2</sub> pendant 48 heures. L'identification des souches β-hémolytiques est complétée par un test au latex et une identification au spectromètre de masse (MALDI-TOF®). Le test au latex repose sur une réaction d'agglutination entre des particules de latex recouvertes d'anticorps spécifiques et l'antigène de Lancefield du groupe A. La spectrométrie de masse MALDI-TOF repose sur l'examen du profil d'extraits protéiques bactériens. Aux Etats-Unis, certains médecins de ville sont habilités à réaliser eux-mêmes des cultures de prélèvement de gorge au cabinet (*office culture*) mais cet examen est de moins en moins pratiqué en raison de l'arrivée des tests de diagnostic rapide (TDR) qui sont plus simples à réaliser et plus sensibles et du renforcement du cadre législatif encadrant la réalisation des examens de biologie.



*Culture de prélèvement de gorge sur gélose au sang montrant des colonies  $\beta$ -hémolytiques typiques du SGA. Ici les zones punctiformes blanchâtres représentent des colonies de SGA (Laboratoire de microbiologie de l'hôpital Robert-Debré, Paris).*

### **Génotype *emm***

La protéine M, facteur majeur de virulence, favorise la colonisation et l'invasion des tissus de l'hôte par le streptocoque du groupe A (SGA) tout en empêchant l'opsonisation et la phagocytose. Il existe environ 100 types différents de protéine M, chacun à l'origine d'une réponse immunitaire spécifique chez l'hôte. La caractérisation des différentes souches de SGA se fait par génotypage de la protéine M définissant des «*emm type*».

### **Inoculum bactérien**

Un inoculum bactérien désigne une quantité définie de bactéries. La taille d'un inoculum est exprimée de manière quantitative en unités formant colonies par ml (UFC/ml). La taille d'un inoculum peut également être mesurée de manière semi-quantitative par compte des colonies sur la boîte de culture. Le seuil de sensibilité des tests de diagnostic rapide est de l'ordre de  $10^5$  à  $10^6$  UFC/ml selon le test utilisé. A partir d'un prélèvement de gorge mis en culture sur une boîte de Pétri, on considère que l'inoculum mesuré sur le milieu de culture reflète la quantité de bactéries présente dans la gorge du patient lorsque le prélèvement de gorge a été correctement réalisé au niveau des amygdales.

### **Interférence bactérienne**

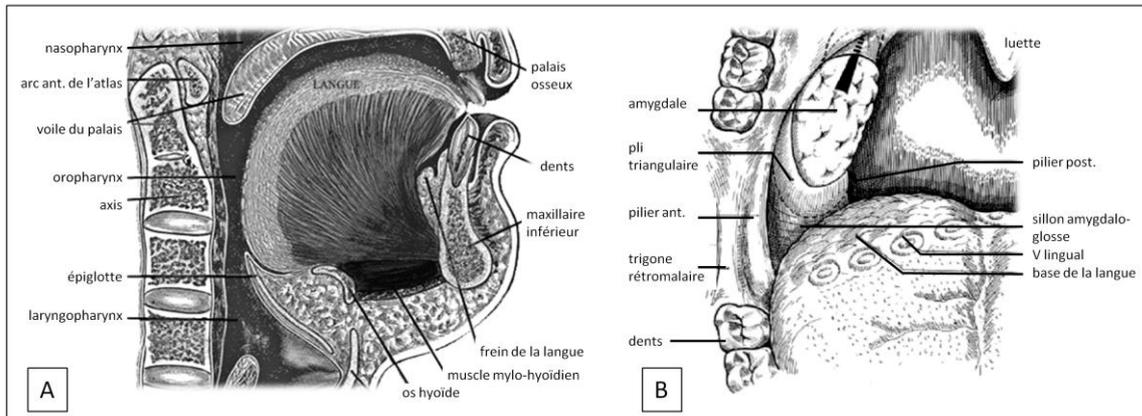
L'interférence bactérienne recouvre différents processus de compétition utilisés par les bactéries pour interagir entre elles afin de se maintenir dans leur environnement. L'interférence bactérienne inclut la production de substances toxiques pour les autres bactéries, la modification de paramètres physico-chimiques environnementaux ou des mécanismes de compétition nutritionnelle tels l'appauvrissement du milieu en certains nutriments indispensables à la survie des autres espèces.

### **Internalisation cellulaire**

Une des premières étapes de la colonisation ou de l'infection par des bactéries ayant un tropisme cutanéomuqueux, tel le streptocoque du groupe A (SGA), est la phase d'adhésion durant laquelle la bactérie peut se fixer à l'épithélium concerné. Après cette phase d'adhésion, le SGA peut être internalisé, c'est-à-dire qu'il peut pénétrer dans les cellules épithéliales de l'hôte et y demeurer dans une phase de quiescence prolongée sans causer d'infection d'aiguë.

### Pharyngite : rappels anatomiques

La pharyngite est l'inflammation aiguë du pharynx et/ou des amygdales. La terminologie française distingue classiquement les angines (avec atteinte des amygdales) et les pharyngites (sans atteinte des amygdales). Cependant, les anglo-saxons ne font pas cette distinction et utilisent le terme générique de *pharyngitis*. Par souci de clarté, nous n'emploierons que le terme « pharyngite » au cours du travail doctoral.



**Anatomie du pharynx.** A. Coupe sagittale (d'après Gray's anatomy of the Human body). B. Vue frontale de la loge amygdalienne. D'après Legent et Pelemuter, dans : « Société Française d'Otorhinolaryngologie et de Chirurgie de la face et du cou. Complications locorégionales des pharyngites. Recommandation pour la pratique clinique, 2008. »)

### Portage

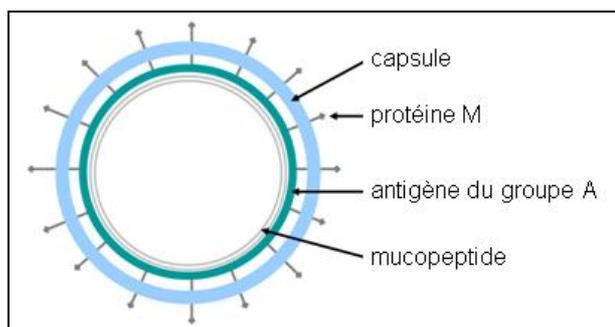
Environ 10% des enfants d'âge scolaire sont des porteurs asymptomatiques de SGA au niveau pharyngé. Le portage de SGA est défini par la mise en évidence de SGA à partir d'une culture de prélèvement de gorge alors que le patient ne présente aucun signe d'inflammation pharyngée, ni de réponse inflammatoire biologique spécifique du SGA. Ce portage peut être chronique et ne nécessite pas de traitement antibiotique. En cas de pharyngite, la culture de prélèvement de gorge ne permet pas de faire la distinction entre une pharyngite due au SGA et une pharyngite virale chez un patient ayant un portage de SGA. Le seul moyen de faire cette distinction est de réaliser une sérologie SGA (anticorps anti-streptodornase et anticorps anti-streptolysine) sur deux prélèvements sanguins réalisés à deux semaines d'intervalle. Cependant, l'interprétation des sérologies SGA est complexe et le rôle des sérologies SGA comme test de référence pour faire la distinction entre portage et infection à SGA est actuellement remis en question.

**SpeB**

*SpeB*, pour *Streptococcus pyrogenic exotoxin B*, est un gène spécifique du streptocoque du groupe A (présent uniquement chez le SGA et non chez d'autres espèces bactériennes) et ubiquitaire (présent chez toutes les souches de SGA). Ce gène chromosomique code pour un facteur de virulence appartenant à la famille des exotoxines streptococciques appelé cystéine protéase. Cette protéine facilite l'invasion des tissus de l'hôte en clivant des composants de la matrice extracellulaire.

**Streptocoque du groupe A (SGA)**

Le streptocoque du groupe A (SGA), ou *Streptococcus pyogenes*, est un cocci à Gram positif se regroupant habituellement en diplocoque ou en chaîne. Le SGA peut être responsable chez l'homme d'infections variées. Il est principalement responsable d'infections cutanéomuqueuses bénignes telles que la pharyngite (environ 80% des infections à SGA) mais le SGA peut également être responsable d'infections sévères (par exemple dermo-hypodermite bactériennes nécrosantes, pleuropneumonies, méningites, endocardites, chocs toxiques streptococciques, arthrites) et de complications suppuratives ou non-suppuratives.



**Structure schématique du SGA.** La protéine M et la capsule favorisent la colonisation et l'invasion des tissus de l'hôte tout en empêchant l'opsonisation et la phagocytose. L'antigène du groupe A est habituellement considéré comme étant spécifique du SGA.

**Streptocoques du groupe milleri**

Le groupe *milleri* est un groupe de streptocoques comprenant trois espèces distinctes : *Streptococcus constellatus*, *Streptococcus intermedius* et *Streptococcus anginosus*. Les streptocoques du groupe *milleri* partagent des caractéristiques communes :

- ils font partie de la flore commensale des voies aériennes supérieures, du tractus digestif, des voies urinaires et génitales ;
- cliniquement, ils sont impliqués dans des infections cutanées, des bactériémies et ont une tendance à former des abcès (notamment abcès dentaires, hépatiques, cérébraux) ;
- en culture, ils forment habituellement des colonies dites « minutes » (diamètre <0,5 mm) et produisent une odeur caractéristique dite « de caramel » en raison de la production de diacetyl ;
- ils partagent un profil biochimique commun sur les galeries d'identification de type API.

Ils sont habituellement  $\alpha$ -hémolytique ou non-hémolytique mais jusqu'à 25% des souches peuvent être  $\beta$ -hémolytiques. Certaines souches peuvent présenter des antigènes de Lancefield du groupe A, C, F ou G.

### **Variants nutritionnels**

La plupart des souches de streptocoque du groupe A (SGA) croissent sans difficulté sur une gélose d'agar additionnée de 5% de sang de mouton ou de cheval. Toutefois, certaines souches de SGA appelés variants nutritionnels poussent très lentement sur ce milieu et ne peuvent donc pas être identifiés par les techniques de culture classiques. Ces variants nutritionnels nécessitent des milieux enrichis pour leur croissance tels que le milieu TGY ou le milieu de Todd-Hewitt.

# INTRODUCTION GENERALE

*Note : les termes accompagnés d'un astérisque sont définis dans le glossaire de la thèse.*

## **1. Pharyngites de l'enfant**

Les pharyngites de l'enfant constituent un problème majeur de santé publique à l'échelle internationale. Plus d'un milliard de cas de pharyngites surviennent chaque année chez l'enfant dans le monde, entraînant une importante consommation de soins et d'antibiotiques [1]. Ces pharyngites surviennent à 95% dans des pays en voie de développement. Aux Etats-Unis, environ 10 millions d'enfants consultent chaque année un pédiatre ou un médecin généraliste pour une pharyngite et 60% à 70% de ces consultations se concluent par une prescription d'antibiotiques [2-4]. Des taux similaires de prescription d'antibiotiques sont observés en France [5].

La surconsommation d'antibiotiques est le principal facteur de l'émergence de bactéries résistantes aux antibiotiques et la majorité de la consommation humaine d'antibiotiques a lieu en dehors de l'hôpital (consommation communautaire). En Europe, les taux de consommation d'antibiotiques varient entre 10 doses définies journalières pour 1000 habitants par jour (Pays-Bas) à environ 30 doses définies journalières pour 1000 habitants par jour (France) [6]. Parallèlement, environ 15% des souches de *Staphylococcus aureus* isolées lors d'infections invasives sont devenues résistantes à la méthicilline et environ 10% des souches d'*Escherichia coli* isolées lors d'infections invasives sont devenues résistantes aux céphalosporines de 3<sup>ème</sup> génération [7].

Le streptocoque du groupe A (SGA)\* est la bactérie la plus fréquemment retrouvée dans les pharyngites de l'enfant, dans 30 à 40% des cas, et le reste des cas est considéré comme étant d'origine virale [8]. Les estimations les plus récentes montrent

qu'environ 15% des enfants d'âge scolaire sont sujets à un épisode de pharyngite à SGA par an dans les pays industrialisés (soit une incidence de 15 cas pour 100 personnes-années). D'autres bactéries peuvent être responsables de pharyngite chez l'enfant mais elles sont beaucoup plus rares et sont à l'origine de tableaux cliniques spécifiques (par exemple l'angine ulcéro-nécrotique de Vincent liée à l'association de *Fusobacterium necrophorum* et *Borrelia vincentii*).

Les pharyngites virales sont des infections bénignes qui évoluent spontanément vers la guérison en quelques jours. A l'inverse, les pharyngites à SGA peuvent entraîner des complications suppuratives\* et non suppuratives\* parfois sévères, telles que le rhumatisme articulaire aigu [9, 10]. Il est largement admis que le traitement antibiotique est donc uniquement justifié en cas de pharyngite à SGA. Les objectifs du traitement antibiotique sont de réduire la contagiosité du cas index, la durée des symptômes et la fréquence des complications suppuratives et non suppuratives [11]. Une revue systématique\* avec méta-analyse\* publiée par la collaboration Cochrane a démontré l'effet protecteur important des antibiotiques sur la survenue de phlegmon périamygdalien avec un risque relatif de 0,15 (IC95% 0,05–0,47), soit un nombre de sujets à traiter par antibiotique pour éviter un cas de phlegmon de l'ordre de 50. Un des présupposés du travail doctoral est donc qu'il y a un bénéfice individuel et de santé publique à correctement identifier et traiter les patients ayant une pharyngite à SGA, mais cela sera remis en question dans la discussion de la thèse.

## ***2. Problématique d'ordre diagnostique***

La problématique diagnostique dans les pharyngites de l'enfant est habituellement considérée de manière binaire, l'objectif étant de distinguer les cas liés au SGA des cas d'origine virale [10, 12]. La stratégie diagnostique choisie doit être suffisamment

sensible car les patients ayant un résultat faux-négatif risquent de développer des complications, et elle doit être suffisamment spécifique car les résultats faux-positifs risquent d'entraîner une consommation d'antibiotiques inutiles, participant à l'émergence des résistances bactériennes et générant des coûts de santé inutiles. Les différents moyens de faire le diagnostic de pharyngite à SGA sont la mise en culture de prélèvement de gorge\*, l'examen clinique et les tests de diagnostic rapide (TDR)\*.

## **2.1 Culture de prélèvement de gorge**

La mise en culture est le test de référence mais son défaut majeur est que son résultat n'est obtenu qu'au bout de 48 heures. Cela risque d'entraîner des prescriptions inutiles d'antibiotiques avant l'obtention des résultats. Dans une étude de pratique portant sur 948 médecins nord-américains, 42% des médecins déclaraient prescrire un antibiotique en attendant le résultat de la culture mais ne pas rappeler les familles pour arrêter le traitement antibiotique en cas de résultat négatif [13]. Nous considérons pour cette thèse la culture de prélèvement de gorge comme test de référence mais il s'agit d'un test de référence imparfait. La mise en culture sur une unique boîte de gélose au sang standard a elle-même une sensibilité de l'ordre de 90-95% en comparaison à des tests plus sensibles tels que la mise en culture combinée sur plusieurs boîtes de Pétri (notamment en utilisant des techniques d'enrichissement ou des milieux sélectifs du SGA) [14].

## **2.2 Examen clinique**

Il a été démontré que les éléments individuels de l'examen clinique tels que la présence de fièvre, d'un exsudat des amygdales (« angine blanche ») ou d'adénopathies cervicales douloureuses ne permettent pas de distinguer de manière suffisamment

sensible et spécifique les pharyngites virales des pharyngites à SGA [15]. Pour cette raison, plusieurs auteurs ont tenté de combiner les éléments de l'examen clinique sous la forme de règles de décision clinique (RDC)\* telles que la règle de Centor ou de McIsaac. Le score de McIsaac\* est un score clinique prédictif de la présence du SGA qui peut varier de 0 à 5, chacun des items suivants étant côté 1 point : âge entre 3 et 15 ans, absence de toux, fièvre, adénopathies cervicales antérieures sensibles, gonflement ou exsudat des amygdales [16]. Le frein majeur à l'utilisation de ce type de RDC chez l'enfant est qu'elles ont été insuffisamment validées dans des populations différentes de leur population de construction (validation externe\*) et que leurs performances n'ont jamais été directement comparées, ce qui empêche la possibilité d'évaluer leurs performances diagnostiques relatives [16-20].

### **2.3 Test de diagnostic rapide**

Les tests de diagnostic rapide (TDR) permettent d'identifier le SGA en moins de 10 minutes à partir d'un prélèvement de gorge. Ces TDR sont basés sur la détection de l'antigène de Lancefield du groupe A (polyoside C spécifique du groupe A) grâce à différentes techniques immunologiques. Ces tests sont simples d'utilisation et peuvent être réalisés dans le cadre d'une consultation de médecine ambulatoire.

Les principaux avantages du TDR sont de fournir un résultat immédiat et d'être hautement spécifiques (spécificité\* autour de 95%) mais leur principal défaut est leur sensibilité\* qui est jugée trop basse par certains experts (autour de 85%) et instable (variant entre 52,6% et 98,9%) [21]. En effet, il a été montré qu'au-delà des variations de performances diagnostiques pouvant exister entre différents kits commerciaux de TDR, il pouvait exister des variations de performances du TDR en fonction de nombreux paramètres. Tout d'abord, la sensibilité du TDR est susceptible de varier en

fonction du type de signes cliniques (ou présentation clinique) du patient, un phénomène habituellement appelé effet de spectre\*. Ainsi, la sensibilité du TDR est habituellement plus élevée chez les patients les plus symptomatiques, par exemple ceux ayant un score de McIsaac élevé [22-24]. Ensuite, certaines études ont mis en évidence des variations de sensibilité du TDR en fonction d'éléments bactériologiques tels que l'inoculum bactérien\*, la sensibilité du TDR augmentant généralement lorsque la quantité de bactéries présentes dans la gorge du patient augmente [25-27]. Enfin, la sensibilité du TDR peut varier en fonction du médecin réalisant le test, les médecins les plus expérimentés obtenant habituellement de meilleures performances, phénomène appelé effet médecin\* [28, 29].

#### **2.4 Portage pharyngé de SGA**

Par ailleurs, bien que la plupart des sociétés savantes considèrent le diagnostic des pharyngites de manière binaire (présence ou absence de SGA), un autre élément à prendre en compte dans les stratégies pour le diagnostic des pharyngites chez l'enfant est le phénomène de portage\*. En effet, environ 10% des enfants d'âge scolaire sont des porteurs asymptomatiques du SGA au niveau pharyngé [8]. Le portage est défini par un résultat positif de la culture de prélèvement de gorge pour le SGA en l'absence de toute réaction immunitaire spécifique du SGA (anticorps anti-streptolysine\* [ASLO] et anti-streptodornase\* [ASD]) [30]. Ni la culture de prélèvement de gorge, ni les TDR ne permettent de faire la distinction entre les enfants ayant une véritable pharyngite à SGA et les enfants ayant un portage de SGA et une pharyngite virale concomitante. Ainsi, réaliser une culture de prélèvement de gorge ou un TDR chez tous les enfants ayant une pharyngite risquerait de conduire à des prescriptions inutiles d'antibiotiques chez les enfants ayant un portage chronique de SGA et une pharyngite virale concomitante. Le

fait que la culture de prélèvement de gorge soit un test de référence imparfait en raison de sa sensibilité imparfaite et du phénomène de portage est souvent négligé dans les études diagnostiques sur les pharyngites à SGA. Nous tenterons de discuter les implications du portage au cours du travail doctoral.

### ***3. Des recommandations hétérogènes à l'échelle internationale***

Il existe actuellement une controverse internationale quant à la stratégie optimale pour le diagnostic des pharyngites chez l'enfant. Il semble que cette hétérogénéité puisse être expliquée par les différentes limites de la culture, de la clinique et du TDR, ainsi que par le poids des habitudes cliniques et des traditions académiques [31]. Ainsi, il semble que les mêmes processus de revue de la littérature aboutissent à des recommandations très différentes. Ainsi, il est recommandé dans certains pays comme le Royaume-Uni de se baser sur une RDC pour sélectionner les patients devant bénéficier d'un traitement antibiotique [32] alors qu'il est recommandé en France et en Europe de réaliser un TDR chez tous les enfants ayant une pharyngite indépendamment de leur présentation clinique, sans confirmation des TDR négatifs par une mise en culture [33]. Aux Etats-Unis, les recommandations sont hétérogènes en ce qui concerne l'utilisation des RDC chez l'enfant mais les recommandations s'accordent sur la nécessité de confirmer les résultats négatifs du TDR par une mise en culture [34-36].

	<b>France</b> SFP-GPIP-SPILF 2011 [33]	<b>Europe</b> ESCMID 2012 [37]	<b>USA</b> AAP 2012 [34]	<b>USA</b> AAP 2013 [38]	<b>Royaume-Uni</b> NICE 2008 [32]
Score clinique	non	non	non	McIsaac	Centor
TDR	systématique	systématique	selon clinique	si score $\geq 2$	non
Culture	non	non	si TDR -	si TDR -	non
Antibiotique	si TDR +	si TDR +	si TDR ou culture +	si TDR ou culture +	si score $\geq 3$

*Synthèse et comparaison des stratégies diagnostiques dans différentes recommandations de pratique clinique. SFP, Société Française de Pédiatrie ; GPIP, Groupe de Pathologie Infectieuse Pédiatrique; SPILF, Société de Pathologie Infectieuse de Langue Française ; ESCMID, European Society of Clinical Microbiology and Infectious Diseases ; AAP, American Academy of Pediatrics ; NICE, National Institute for health and Care Excellence (adapté de Matthys et al.,2007 [31])*

Au total, la problématique diagnostique et thérapeutique des pharyngites à SGA semble habituellement abordée de manière sur-simplificatrice (problématique diagnostique binaire, test de référence fiable, test de diagnostic rapide stable, règles de décision clinique établies et fiables, bénéfice attendu du traitement antibiotique), alors qu'il s'agit en réalité d'une situation clinique bien plus complexe. Au vu des données actuelles, il est difficile de définir la place possible du TDR et des RDC dans la stratégie diagnostique chez l'enfant. En ce qui concerne les TDR, aucune étude n'a exploré l'effet indépendant des facteurs de variabilité de ses performances diagnostiques, liés aux patients (effet de spectre et effet de l'inoculum bactérien) ou liés aux médecins réalisant le test (effet médecin). De plus, les facteurs susceptibles d'expliquer les résultats faux-positifs du TDR (taux de faux-positifs de l'ordre de 5%) ont été très peu explorés dans la littérature. En ce qui concerne les RDC, leur place dans les recommandations est variable parce qu'elles ont été insuffisamment validées chez

l'enfant et qu'aucune évaluation systématique ni comparaison directe prospective de leurs performances diagnostiques n'a été réalisée.

#### **4. *Étapes du travail doctoral***

L'objectif de ce travail doctoral était de produire des connaissances nouvelles permettant d'éclairer la controverse internationale actuelle quant à la place du TDR et des RDC dans le diagnostic des pharyngites chez l'enfant. Nous avons conduit ce travail en quatre étapes.

Dans le **Chapitre 1**, nous avons évalué l'effet de différents facteurs liés aux patients sur la sensibilité du TDR et discuté leur impact en pratique clinique.

Dans le **Chapitre 2**, nous avons évalué l'effet de différents facteurs liés aux patients et aux médecins sur la sensibilité du TDR et discuté leur impact en pratique clinique.

Dans le **Chapitre 3**, nous avons exploré de manière systématique les causes susceptibles d'expliquer les résultats faux-positifs du TDR et avons discuté leur impact en pratique clinique.

Dans le **Chapitre 4**, nous avons tenté d'identifier l'ensemble des RDC publiées à ce jour pour le diagnostic des pharyngites à SGA, nous avons évalué leur qualité méthodologique puis nous avons évalué leur validité externe et comparé leurs performances diagnostiques.

Dans la **conclusion** du travail doctoral nous avons essayé de mettre en perspective les connaissances nouvelles produites. Enfin, nous avons présenté les **perspectives de recherche** potentielles au-delà du travail doctoral.



## SOURCES DE DONNEES

## **1. Méthodologie générale de l'étude ACTIV TDR SGA**

Le travail doctoral est en majeure partie basé sur l'analyse des données de l'étude ACTIV TDR SGA 2009-2011. Il s'agit d'une étude observationnelle transversale multicentrique prospective réalisée en contexte de soins primaires dont les objectifs étaient d'étudier les performances diagnostiques d'un TDR (StreptAtest) dans différentes situations cliniques (pharyngites, portage pharyngé asymptomatique, infections cutanées, infections périnéales). Seules les données concernant les pharyngites et le portage pharyngé asymptomatique ont été utilisées dans la thèse.

## **2. Médecins investigateurs**

Les patients ont été inclus par 17 pédiatres de ville de 2 régions de France faisant partie du réseau ACTIV (Association Clinique et Thérapeutique Infantile du Val-de-Marne - <http://www.activ-france.com/>) [20, 39]. Il s'agit d'un réseau professionnel associatif créé en 1988 dédié à la recherche et à l'enseignement dans le domaine de la pédiatrie de ville et en particulier dans le domaine de l'épidémiologie des maladies infectieuses rencontrées en pédiatrie ambulatoire (pharyngite, otite, méningite, pneumonie aiguë communautaire). L'association collabore activement avec la Société Française de Pédiatrie (SFP) et l'Association Française de Pédiatrie Ambulatoire (AFPA).

A l'issue de l'étude ACTIV TDR SGA 2009-2011, les 17 pédiatres investigateurs ayant participé à l'étude ont répondu à un questionnaire en ligne afin de recueillir des informations concernant leurs caractéristiques sociodémographiques, leur type d'exercice médical et leurs connaissances et croyances concernant la pathologie à SGA (voir détails au Chapitre 2 page 66 et annexe 8 page 225). Le questionnaire n'était pas anonyme mais les réponses ont été anonymisées après chaînage avec la base contenant les données individuelles des patients.

### **3. Patients**

#### **3.1 Groupe pharyngite**

Les critères d'inclusion étaient :

- enfants âgés de 3 à 15 ans ;
- vus par leur pédiatre de ville pour pharyngite clinique, définie par la présence de signes d'inflammation du pharynx et/ou des amygdales lors de l'examen clinique ;
- n'ayant pas reçu d'antibiotiques depuis 1 semaine.

#### **3.2 Groupe d'enfants asymptomatiques**

Les critères d'inclusion étaient :

- enfants âgés de 3 à 15 ans ;
- bien portants le jour de la consultation (visite de contrôle systématique ou vaccination) ;
- n'ayant pas eu de pharyngite dans le mois précédent ;
- n'ayant pas reçu d'antibiotiques dans la semaine précédente ;
- dont la fratrie ne présente pas d'infection à streptocoque du groupe A identifiée dans la semaine précédente.

### **4. Données cliniques**

Pour chaque enfant inclus dans le groupe pharyngite, le médecin investigateur devait remplir une fiche comportant des données démographiques de base et des données cliniques décrivant la pharyngite (voir les fiches de recueil de données cliniques en

annexe 8 page 225). Les données cliniques recueillies étaient celles habituellement décrites comme associées à la présence du SGA d'après plusieurs revues de la littérature sur le sujet [12, 40] :

- variables démographiques : âge et sexe de l'enfant ;
- signes fonctionnels : contagion familiale de pathologie à SGA, début brutal, odynophagie, toux, rhinorrhée, enrouement, conjonctivite, céphalées, douleurs abdominales, nausées, vomissements, diarrhée ;
- signes physiques : température maximale observée depuis le début de l'épisode, érythème pharyngé, gonflement des amygdales, pharynx érythémato-pultacé, purpura du voile du palais, présence d'adénopathies sous-angulo-maxillaires, taille des adénopathies sous-angulo-maxillaires si présentes, adénopathies sous-angulo-maxillaires sensibles si présentes, éruption cutanée, éruption scarlatiniforme si éruption présente.

Les variables « âge » et « température maximale rapportée au cours de l'épisode infectieux » étaient des variables continues, la variable « taille des adénopathies sous-angulo-maxillaires » était une variable discrète en 3 classes et les autres variables cliniques étaient dichotomiques. Pour les analyses portant sur l'effet de l'âge, cette variable a été catégorisée en 2 classes de même étendue (3–8 ans ; 9–14 ans).

Pour chaque enfant asymptomatique inclus, une fiche clinique simplifiée était également remplie. Dans tous les cas, les données cliniques étaient recueillies à l'aveugle des résultats du TDR et de la culture de prélèvement de gorge.

## 5. Données bactériologiques

Tous les patients inclus ont été soumis à un TDR et à une mise en culture, excluant un éventuel biais de sélection\* ou de vérification partielle\* présent dans d'autres études [22, 23].

Chaque enfant a été soumis à un prélèvement de gorge à visée bactériologique. Les prélèvements de gorge ont été réalisés avec un écouvillon double (Venturi Transystem Amies agar gel, COPAN, Corona, CA, USA). Il était recommandé aux investigateurs de réaliser les prélèvements de gorge en frottant l'écouvillon sur les deux amygdales, la paroi postérieure du pharynx et toute zone inflammatoire observée à l'examen du pharynx, tout en évitant de toucher la langue et les parois de la cavité buccale.



*Double écouvillon utilisé dans l'étude ACTIV TDR SGA (photo adaptée du catalogue du fabricant, accessible sur : <http://www.copanusa.com/products/aerobic-anaerobic/>).*

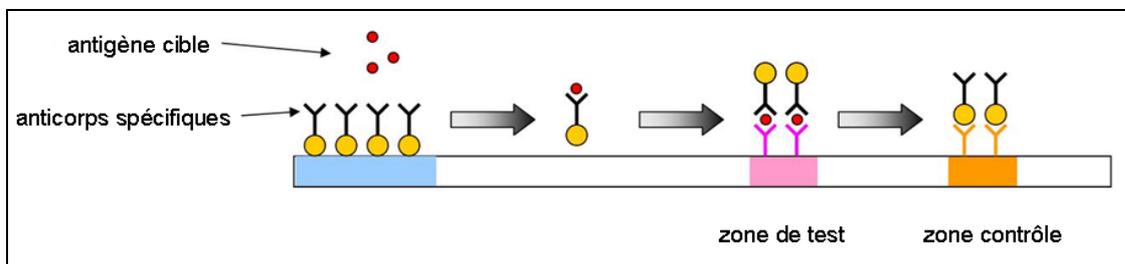
### 5.1 Test de diagnostic rapide

Le premier écouvillon servait au pédiatre pour réaliser le TDR durant la consultation (StreptAtest, Dectrapharm, France). Le résultat du TDR était lu comme positif, positif faible ou négatif. Dans le cadre de cette étude, les résultats positifs et faiblement positifs du TDR ont été regroupés afin d'obtenir une variable dichotomique (TDR positif ou négatif), comme cela est souvent fait dans ce type d'études [27]. Le StreptAtest est une bandelette immunochromatographique. Il s'agit du TDR recommandé en France en raison de son couple performance *in vitro* - praticabilité élevée [41]. *In vitro*, ce TDR est capable de détecter le SGA à partir d'un inoculum  $>10^5$  UFC/ml. Depuis 2002, la Caisse Nationale d'Assurance Maladie le met gratuitement à disposition des médecins

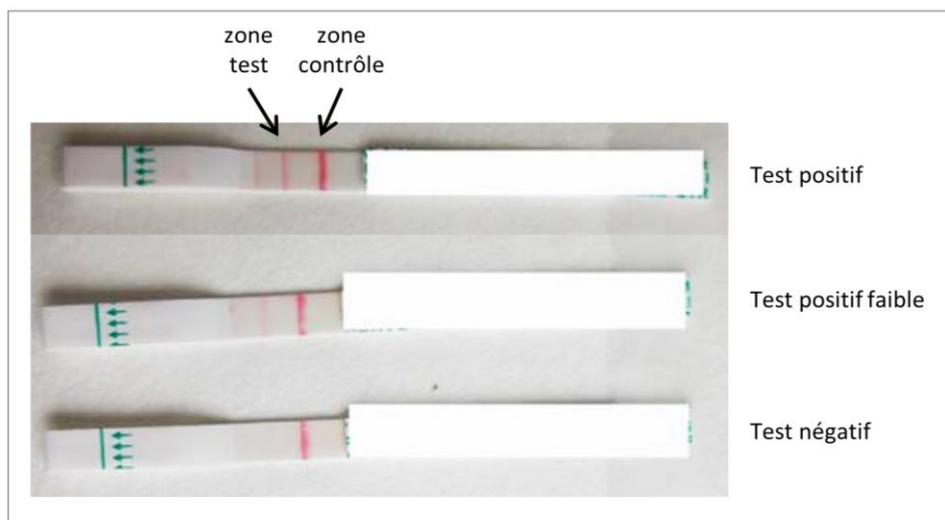
libéraux et des services hospitaliers en faisant la demande. Le coût facturé par le fabricant à l'Assurance Maladie est inférieur à 1 €/test. Les étapes que doit suivre le clinicien pour réaliser un StreptAtest sont : 1/ réaliser un prélèvement de gorge en écouvillonnant les deux amygdales et les zones inflammatoires du pharynx, 2/ réaliser l'extraction antigénique en mettant en présence pendant une minute l'écouvillon avec 4 gouttes de chacun des 2 réactifs fournis, 3/ immerger la bandelette réactive dans le mélange et 4/ lire la bandelette au bout de 5 minutes.



*Eléments contenus dans une boîte de StreptAtest, TDR utilisé dans l'étude ACTIV TDR et SGA. 1/ réactifs chimiques, 2/ abaisse langue, 3/ écouvillon servant à réaliser le prélèvement de gorge, 4/ bandelette réactive (d'après la documentation fournie par le fabricant du StreptAtest, accessible sur : <http://www.testangine.com/>)*



**Principe de fonctionnement du TDR immunochromatographique.** Le complexe antigène-anticorps migre le long d'une bandelette de cellulose et est révélé par une réaction colorée grâce à un second anticorps dit « de marquage ». La zone de contrôle positif permet de vérifier que la migration par capillarité le long de la bandelette a bien eu lieu.



*Bandelettes de StreptAtest montrant différents types de résultat. Le TDR implique une étape de lecture de la bandelette comprenant une part de subjectivité (adapté de Charlier-Bret et al., 2004 [41]).*

## 5.2 Mise en culture

Le deuxième écouvillon était acheminé par coursier dans un délai théorique de 72 heures au laboratoire de bactériologie de l'hôpital Robert-Debré à Paris (laboratoire associé au centre national de référence du SGA) pour mise en culture selon une technique de référence [42]. Durant l'année 1 (2009-2010), le prélèvement de gorge était ensemencé directement sur une gélose de type trypticase-caséine-soja additionnée de 5% de sang de mouton [14]. Pour les besoins de l'étude correspondant au **Chapitre 3** du travail doctoral et en raison de la régulation des marchés de l'Assistance Publique, le protocole de mise en culture de prélèvement de gorge a été modifié pour la deuxième année d'étude. Durant l'année 2 (2010-2011), l'écouvillon était dilué dans 1 ml d'eau stérile puis une goutte du liquide extrait de l'écouvillon était ensemencée sur une gélose sélective contenant de l'acide nalidixique et de la colistine (Columbia ANC+5% sheep blood, Biomérieux, Lyon, France). Les milieux sélectifs tels que cette gélose Columbia ANC ont pour objectif de sélectionner le SGA en inhibant la croissance des espèces de la flore commensale de la sphère oro-pharyngée [14], ce qui permet d'augmenter la

sensibilité de la technique de culture en maximisant la détection du SGA. Dans les deux cas la boîte de Pétri était mise en incubation à 37°C en anaérobiose pendant 48 heures et le groupe de Lancefield des colonies  $\beta$ -hémolytiques était déterminé par agglutination au latex (Prolex, Pro-Lab Diagnostics, Canada). Le résultat de la culture était binaire (positif/négatif). En cas de culture positive pour le SGA, l'inoculum bactérien était estimé de manière semi-quantitative en fonction du nombre de colonies  $\beta$ -hémolytiques présentes dans la boîte de Pétri après 48 heures de culture (1+, <10 colonies ; 2+, 11–50 colonies ; 3+, >50 colonies par boîte). Pour l'analyse, les deux premières classes d'inoculum (1+ et 2+) ont été regroupées afin d'obtenir une variable dichotomique (inoculum faible ou fort) [43]. Les analyses réalisées au laboratoire de bactériologie étaient faites à l'aveugle des données cliniques individuelles et du résultat du TDR.

### 5.3 PCR

Durant l'année 2 (2010-2011) et au vu des objectifs du **Chapitre 3**, certains prélèvements de gorge ont également été testés pour la présence du SGA par une technique de biologie moléculaire de type réaction de polymérisation en chaîne (*polymerase chain reaction* - PCR). Cette technique est basée sur l'identification non pas d'un antigène bactérien mais d'un fragment de génome bactérien. Dans cette étude, le fragment de gène cible était une portion du gène *speB\** codant pour l'exotoxine B, spécifique du SGA et considéré ubiquitaire pour cette espèce (voir détails au Chapitre 3, page 84).

## 6. *Nombre de sujets nécessaires*

Dans le groupe d'enfants avec pharyngite, le nombre de sujets nécessaires a été calculé pour que l'intervalle de confiance à 95% de la sensibilité du TDR soit une estimation

ayant une précision de +/-5%. En se basant sur une prévalence attendue de SGA de 35% et une sensibilité attendue minimale de 75%, 714 enfants avec pharyngite devaient être inclus dans l'étude.

Chez les enfants asymptomatiques, le nombre de sujets nécessaires a été estimé pour que l'intervalle de confiance à 95% de la fréquence du portage asymptomatique de SGA soit une estimation ayant une précision de +/-4%. En se basant sur une fréquence de portage attendue minimale de 10%, 250 enfants asymptomatiques étaient nécessaires à l'étude.

## **7. Période d'étude**

Le recrutement était initialement prévu entre le 1<sup>er</sup> octobre 2009 et le 31 mai 2010 mais l'étude a été reconduite l'année suivante du 1<sup>er</sup> octobre 2010 au 31 mai 2011 dans les deux groupes en raison des difficultés à inclure des enfants asymptomatiques. Au vu de la nécessité de prolonger l'étude pour une deuxième année, le comité scientifique de l'étude a décidé de rajouter des objectifs secondaires à l'étude, correspondant aux Chapitres 3 et 4 de la thèse.

## **8. Ethique**

Les parents ont été informés par écrit et par oral des objectifs, méthodes, bénéfices escomptés ainsi que des risques potentiels de l'étude et des désagréments qui pourraient en résulter pour leur enfant. Une affiche était également mise dans la salle d'attente des médecins investigateurs afin d'informer les parents de cette étude. Pour chaque individu du groupe d'enfants asymptomatiques, les investigateurs ont obtenu le consentement libre et éclairé d'au moins un parent par écrit avant l'inclusion de l'enfant dans l'étude. Pour les enfants atteints de pharyngite, une simple non-opposition à l'utilisation des

données de routine a été obtenue car il s'agissait d'une étude non-interventionnelle dans laquelle l'obtention d'un prélèvement de gorge correspondait à la prise en charge standard des patients au moment de l'étude. Cette étude a reçu un avis favorable du Comité de Protection des Personnes Ile-de-France XI (n°09016), de la CNIL (n°1354254) et de l'Afssaps (n°2009-A00086-51).

## ***9. Monitoring***

Une base de données spécifique à l'étude a été créée grâce au logiciel 4D version 6.4. Seul un nombre restreint de personnes habilitées avaient un droit d'accès sécurisé à la base de données. Cette base était située sur un poste informatique au siège d'ACTIV, sous la responsabilité de deux personnes chargées de la conception et de la mise en place de cette base et de la réception des formulaires de recueils de données. Toutes les données ont été entrées dans la base de données 4D par double saisie, permettant des contrôles de saisie par détection automatique d'incohérences. La récupération des données manquantes et/ou erronées était effectuée en temps réel au fur et à mesure de l'arrivée des formulaires de recueils de données et de leur saisie puis deux vérifications générales du contenu de la base ont été réalisées, au terme de la première année d'étude puis au terme de la deuxième année d'étude.

## ***10. Analyses statistiques***

Toutes les analyses statistiques ont été réalisées sur le logiciel Stata/SE 12 (StataCorp, College Station, TX, USA).

# CHAPITRE 1

## Effet de l'inoculum bactérien et du spectre clinique sur les performances diagnostiques du TDR

Jérémie Cohen, Martin Chalumeau, Corinne Levy, Philippe Bidet, Franck Thollot,

Alain Wollner, Edouard Bingen, Robert Cohen

*PLoS One* 2012; 7(6): e39085

Impact factor 2012 : 3,730 ; rang 1/8

(voir publication en annexe 1, page 147)

## 1. Introduction

Nous avons vu dans l'introduction que les TDR ont été proposés comme une alternative à la mise en culture de prélèvement de gorge depuis les années 1980. En comparaison à la culture, les TDR ont habituellement une spécificité élevée, de l'ordre de 95% et ont l'avantage de fournir un résultat immédiat. Leurs principales limites sont une sensibilité jugée trop basse par certains auteurs, de l'ordre de 85% (de 52,6% à 98,9%) [21, 44] et des variations de sensibilité en fonction du spectre clinique appelées effet de spectre\* [45-48].

La mise en évidence d'un effet de spectre pose la question de la généralisabilité des performances diagnostiques observées dans une étude. Tout d'abord, les performances observées au sein d'une population dans son ensemble ne sont pas forcément applicables à des sous-groupes avec un spectre particulier de la maladie cette même population. Deuxièmement, les performances observées dans une population ne sont pas forcément applicables à une autre population avec un de spectre là aussi particulier de la maladie.

Trois études pédiatriques ont mis en évidence un effet de spectre significatif sur la sensibilité du TDR pour le diagnostic des pharyngites à SGA [22-24]. Dans ces trois études, il a été montré que la sensibilité du TDR augmentait avec le score de McIsaac\*. Par exemple, dans l'étude d'Edmonson *et al.* portant sur 1184 patients nord-américains âgés de moins de 24 ans, la probabilité pré-test de SGA augmentait de 9% à 50% et la sensibilité du TDR de 67% à 88% chez les enfants ayant un score de McIsaac à 1 et 4, respectivement [2]. Toutefois, la portée de ces trois études est limitée car elles présentaient toutes des défauts méthodologiques. Par exemple, dans l'étude d'Edmonson *et al.*, les TDR n'étaient pas réalisés de manière systématique en cas de pharyngite et cela a probablement conduit à une sélection des patients les plus

symptomatiques (biais de sélection\*) [2]. Dans l'étude de Hall *et al.*, seuls les TDR négatifs étaient contrôlés par culture, ce qui introduit un biais de vérification partielle\* ayant tendance à surévaluer la prévalence du SGA et la sensibilité du TDR [3]. Dans l'étude de Tanz *et al.*, les auteurs concluaient à un effet de spectre sur la sensibilité du TDR sur la base d'une sensibilité plus élevée chez les patients ayant un score de McIsaac > 2 mais cela n'était soutenu par aucun test statistique [4]. De plus, aucune étude n'a évalué l'effet du spectre clinique sur les performances du TDR en s'intéressant à la fois à des variations de sensibilité, de spécificité, de valeurs prédictives et de rapports de vraisemblance [22-24].

Au-delà de l'effet de spectre, la sensibilité du TDR est également affectée par l'importance de l'inoculum\* bactérien [25-27]. Dans une étude américaine chez 746 patients atteints de pharyngite, Dale *et al.* ont démontré l'existence d'une relation dose-effet entre inoculum et sensibilité du TDR, avec une sensibilité augmentant de 17% à 96% en fonction de l'inoculum bactérien estimé de manière semi-quantitative (de 1+ à 4+) [49].

A ce jour, plusieurs études ont exploré les effets du spectre clinique et de l'inoculum bactérien sur les performances du TDR mais aucune étude n'a évalué l'effet indépendant du spectre clinique et de l'inoculum, ni les relations potentielles entre ces deux facteurs de variations. Notre objectif était d'étudier l'effet indépendant du spectre clinique évalué par le score de McIsaac et de l'inoculum bactérien sur les performances diagnostiques du TDR chez les enfants atteints de pharyngite.

## **2. Méthodes**

### **2.1 Patients et prélèvements de gorge**

Cette étude est basée sur l'analyse intermédiaire des données de l'étude ACTIV TDR

---

SGA de la première année d'inclusion (2009-2010) du groupe pharyngite (voir Sources de données, page 37).

## **2.2 Données cliniques**

Nous avons utilisé les données cliniques permettant de calculer le score de McIsaac pour chaque patient. Les items du score de McIsaac incluent l'âge, la fièvre ( $T^{\circ} > 38^{\circ}\text{C}$ ), la présence d'une inflammation des amygdales (gonflement des amygdales ou exsudat des amygdales), la présence d'adénopathies cervicales antérieures sensibles et l'absence de toux. Comme cela a été fait dans d'autres études [23, 50], le critère de McIsaac concernant la présence d'adénopathies sensibles a été modifié et un point a été attribué en cas d'adénopathie cervicale antérieure  $\geq 1$  cm et/ou sensible car la présence d'adénopathies cervicales antérieures sensibles peut être difficile à préciser chez l'enfant. Comme dans d'autres études [22-24], les scores extrêmes ont été regroupés (scores 1 et 2 d'une part, et 4 et 5 d'autre part).

## **2.3 Nombre de sujets nécessaires**

Aucun calcul supplémentaire de nombre de sujets nécessaires n'a été réalisé pour cette étude par rapport au calcul fait pour l'étude princeps.

## **2.4 Analyses statistiques**

Le résultat de la culture au laboratoire de bactériologie a été considéré comme le test de référence de l'étude.

Premièrement, nous avons décrit les caractéristiques générales de la population d'étude et la fréquence des différents streptocoques  $\beta$ -hémolytiques\* retrouvés en culture (streptocoques du groupe A, C ou G). Dans la suite des analyses, le résultat de la

culture a été considéré de manière binaire et les patients ayant un streptocoque du groupe C\* ou G identifié en culture ont été considérés comme négatifs pour le SGA. Ensuite, nous avons décrit la distribution des scores de McIsaac dans la population d'étude ainsi que la prévalence du SGA en fonction du score de McIsaac et de chaque item du score de McIsaac.

Deuxièmement, nous avons utilisé des tests du  $\chi^2$ \* ( $\chi^2$  de Pearson\*, de Cochran\* ou de tendance\*) pour étudier l'effet du spectre clinique sur les performances du TDR en termes de sensibilité, spécificité, valeurs prédictives et rapports de vraisemblance, en fonction du score de McIsaac et de chaque item du score.

Troisièmement, nous avons utilisé des tests du  $\chi^2$  pour étudier la distribution des inocula bactériens en fonction du délai entre prélèvement et mise en culture ( $\leq 48$  contre  $> 48$  heures) et pour étudier l'effet de l'inoculum\* sur la sensibilité du TDR.

Quatrièmement, la distribution des inocula en fonction du score de McIsaac a été étudiée par des tests du  $\chi^2$  de tendance. L'association entre signes cliniques et inoculum a été étudiée en comparant la fréquence des inocula forts en fonction de chaque item du score de McIsaac et de l'âge.

Cinquièmement, nous avons utilisé un modèle binomial\* avec fonction de lien identité pour évaluer l'effet indépendant du spectre clinique et de l'inoculum bactérien sur la sensibilité du TDR [51]. Le modèle permet d'estimer les mesures d'association sous formes de différences de sensibilité en se restreignant aux patients ayant une culture positive et en utilisant le résultat du TDR comme variable prédite par le modèle :

$$\text{Sensibilité} = P(\text{TDR}+) = \beta_0 + \sum_{i=1}^k \beta_i X_i$$

Où  $\beta_0$  est l'intercept du modèle et représente la sensibilité du TDR de base, toutes les variables du modèle étant égales à 0, et les  $\beta_i$  sont les paramètres du modèle et

représentent les variations de sensibilité du TDR associées à l'augmentation d'une unité des variable  $X_i$ .

La sensibilité du TDR a été étudiée en analyse univariée en fonction de chaque critère clinique puis en analyse multivariée en ajustant sur chaque item clinique et sur l'inoculum. Les interactions potentielles ont été testées en introduisant les termes d'interaction\* séparément dans un modèle binomial vide. Pour l'analyse multivariée, les variables ont été ajoutées dans le modèle en suivant l'ordre ascendant de degré de significativité de chaque variable en analyse univariée jusqu'à non-convergence\* du modèle. Les interactions ont été gérées en introduisant les termes d'interaction dans le modèle, ou en cas de non-convergence, en stratifiant\* sur l'inoculum.

### **3. Résultats**

#### **3.1 Caractéristiques de la population, prévalence du SGA**

Au total, 807 enfants ont été inclus sur la première période d'étude. Un patient ayant un TDR ininterprétable, 2 patients chez qui le prélèvement de gorge avait été perdu et 19 patients ayant des données cliniques manquantes empêchant le calcul du score de McIsaac ont été exclus secondairement (Figure 1, page 61). Ainsi, l'analyse a porté sur 785 enfants (dont 351 filles [44,7%]), ayant un âge moyen (écart-type) de 6,1 (2,5) ans. La prévalence globale du SGA était de 36,3% (IC95, 32,9%–39,8%). Au total, 4 (< 1%) streptocoques du groupe C et 7 (< 1%) streptocoques du groupe G ont été identifiés en culture. L'âge moyen des patients secondairement exclus était de 5,9 ans (3,0) et le SGA a été retrouvé en culture dans 4/20 cas (20,0% [7,9%–39,2%]).

La prévalence du SGA augmentait avec le score de McIsaac ( $p < 0,001$ ) (Tableau 1, page 57 et Tableau 2, page 58), de 24,3% chez les enfants ayant un score  $\leq 2$  à 41,2% chez les enfants ayant un score  $\geq 4$ . Des éléments cliniques individuels étaient

également significativement associés à la présence du SGA : l'âge < 9 ans, l'absence de toux, et la présence de fièvre (Tableau 2, page 58).

### 3.2 Effet de spectre

La sensibilité globale du TDR était de 86,7% (82,2%–90,4%), la spécificité de 94,6% (92,2%–96,4%), le rapport de vraisemblance positif de 16,1 (11,1–23,2), le rapport de vraisemblance négatif de 0,14 (0,10–0,19), la valeur prédictive positive de 90,1% (86,0%–93,4%) et la valeur prédictive négative de 92,6% (89,9%–94,7%).

Nous avons mis en évidence des variations significatives de sensibilité et de spécificité du TDR en fonction du score de McIsaac en analyse univariée (Tableau 2, page 58). Ainsi, la sensibilité augmentait de 75,0% (55,1%–89,3%) à 84,3% (75,0%–91,1%) puis 89,9% (84,3%–94,0% ;  $p = 0,02$ ), et la spécificité diminuait de 97,7% (91,9%–99,7%) à 96,0% (91,8%–98,4%) puis 92,5% (88,4%–95,5% ;  $p = 0,04$ ) en cas de score de McIsaac  $\leq 2$ , 3 et  $\geq 4$  (Tableau 2, page 58).

Il existait également des variations de rapports de vraisemblance et de valeurs prédictives du TDR en fonction du score du McIsaac, mais celles-ci n'étaient pas statistiquement significatives. Les performances du TDR variaient de manière importante en fonction des éléments du score de McIsaac pris séparément ainsi qu'en fonction de l'âge. Par exemple, la sensibilité du TDR était plus basse chez les enfants de plus de 9 ans que chez les enfants de moins de 9 ans (66,7% [47,2%–82,7%] vs. 89,0% [84,5%–92,6%],  $p < 0,01$ ). La valeur prédictive négative du TDR variait de manière non significative en fonction de score de McIsaac ou des différents éléments cliniques considérés de manière séparée, allant de 88,8% (80,8%–94,3%) chez les enfants sans atteinte des amygdales à 94,6% (90,0%–97,5%) chez les enfants sans fièvre.

### 3.3 Effet de l'inoculum sur la sensibilité du TDR

Parmi les 285 cultures de prélèvement de gorge positives pour le SGA, 243 (85% [80,6%–89,2%]) présentaient un inoculum fort. Il n'y avait pas de différence de distribution des inocula en fonction du délai de mise en culture ( $\leq 48$  vs.  $> 48$  heures,  $p = 0,97$ ). La sensibilité du TDR variait de manière importante en fonction de l'inoculum, avec une sensibilité 40,5% [25,6%–56,7%] en cas d'inoculum faible contre une sensibilité de 94,7% [91,0%–97,1%] en cas d'inoculum fort ( $p < 0,001$ ). Par ailleurs, un inoculum faible était retrouvé chez 65,8% (50,0%–81,6%) des faux-négatifs du TDR contre 6,9% (3,7%–10,1%) des vrais positifs du TDR ( $p < 0,001$ ). La fréquence des inocula forts augmentait avec le score de McIsaac, mais de manière non significative ( $p = 0,09$ , tableau 1, page 57). Les inocula forts étaient également plus fréquents chez les enfants de moins de 9 ans que chez les enfants plus âgés (87,1% [82,9%–91,2%] contre 70,0% [52,6%–87,4%],  $p < 0,05$ ).

### 3.4 Modélisation de l'effet du spectre clinique et de l'inoculum sur la sensibilité du TDR

Les résultats de la modélisation univariée étaient très proches de ceux de l'analyse stratifiée (Tableaux 2, page 58 et Tableau 3, page 59) : l'âge  $\geq 9$  ans et la présence de toux étaient associés de manière significative à une diminution de sensibilité ( $p < 0,05$ ). Après stratification sur l'inoculum, nous avons mis en évidence des variations importantes et significatives de sensibilité du TDR chez les patients ayant un inoculum faible (par exemple différence de 37,4% de sensibilité en fonction de l'âge,  $p < 0,01$ ), tandis que ces variations étaient de faible amplitude ( $< 6\%$ ) et non significatives chez les patients ayant un inoculum fort. Il existait des interactions entre inoculum et âge ( $p = 0,03$ ), inoculum et adénopathies cervicales antérieures ( $p = 0,01$ ), et une tendance entre

inoculum et signes d'inflammation des amygdales ( $p = 0,07$ ).

En raison de ces interactions multiples entre inoculum et éléments cliniques et en raison de la non convergence du modèle après avoir introduit les termes d'interaction correspondant, nous avons construit deux modèles multivariés en stratifiant sur l'inoculum (Tableau 3, page 59). Les résultats de ces modèles multivariés étaient stables par rapport à la modélisation univariée : il existait des différences ajustées significatives de sensibilité du TDR de grande amplitude en cas d'inoculum faible (de 23,6% à 40,9%), mais des différences ajustées de sensibilité de faible amplitude (<6%) et non significatives en cas d'inoculum fort.

#### **4. Discussion**

Nous avons confirmé que le spectre clinique a un impact important sur la sensibilité du TDR et avons mis en évidence pour la première fois le fait que le spectre clinique affecte également la spécificité, les rapports de vraisemblance et la valeur prédictive positive du TDR. Nous avons également montré que la valeur prédictive négative du TDR ne variait pas en fonction du spectre clinique. Dans cette étude, la valeur prédictive négative du TDR restait stable, autour de 92%, indépendamment du score de McIsaac et de ses composants. Les cliniciens ne souhaitent pas exposer leurs patients aux complications des pharyngites à SGA et recherchent pour cela un test ayant une valeur prédictive négative élevée et stable. Dans cet échantillon ayant une prévalence de SGA de 36% et en utilisant le TDR recommandé et fourni par l'Assurance Maladie, la probabilité de SGA en cas de TDR négatif était stable, de l'ordre de 8%, indépendamment de la présentation clinique des patients. La fréquence du taux de portage asymptomatique de SGA étant d'environ 10% [8], la valeur prédictive négative que nous rapportons semble satisfaisante en pratique clinique courante.

Cette étude confirme également l'effet majeur de l'inoculum bactérien sur la sensibilité du TDR, comme cela a déjà été rapporté dans d'autres études, et elle décrit pour la première fois la relation existant entre effet de spectre et effet de l'inoculum. Nos résultats suggèrent qu'un inoculum bactérien plus important est associé à une symptomatologie plus bruyante (score de McIsaac plus élevé), même si nous avons probablement manqué de puissance pour confirmer cette association ( $p = 0,09$ ). Une hypothèse permettant potentiellement d'expliquer cette relation entre effet de l'inoculum et effet de spectre, est celle d'une distribution bimodale des inocula, avec d'une part les inocula faibles et d'autre part les inocula forts. Selon cette hypothèse, si un signe clinique est associé à une baisse d'inoculum, l'effet sur la sensibilité du TDR sera plus important en cas d'inoculum faible qu'en cas d'inoculum fort (Figure 2, page 62).

Cette étude ne nous a pas permis de tester l'hypothèse selon laquelle les patients ayant un score de McIsaac bas et un inoculum faible seraient en majorité des porteurs ayant une pharyngite virale concomitante car nous n'avons pas été en mesure de différencier les patients réellement infectés par les SGA des patients porteurs du SGA. Cette différenciation nécessite classiquement la mesure de la réponse immunitaire au SGA (anticorps anti-streptolysine O [ASLO]\* et anti-streptodornase [ASD]\*) par des prélèvements sanguins répétés qui n'étaient pas justifiés au regard des objectifs de l'étude [25]. Cependant, l'interprétation des sérologies SGA est complexe et le rôle des sérologies SGA comme test de référence pour faire la distinction entre portage et infection à SGA est actuellement remis en question. En effet, les résultats d'une étude ayant suivi 160 enfants nord-américains pendant 2 ans, avec prélèvements de gorge et prélèvements sanguins répétés dans le temps a montré que :

- suite à l'acquisition asymptomatique (portage) d'un nouveau sérotype de SGA,

une augmentation des ASLO et des ASD était observée dans 35% des cas ;

- certains patients ayant un portage chronique asymptomatique de SGA gardaient un titre élevé d'ASLO et d'ASD pendant plusieurs mois ;
- l'acquisition d'un streptocoque du groupe C ou G était accompagnée d'une augmentation des ASLO dans 37% des cas [52].

De plus, l'interprétation de ces sérologies est complexe car les normes varient en fonction de l'âge et de la zone géographique [52]. L'utilisation des sérologies SGA en clinique courante est de toute manière difficile à proposer en raison de la nécessité de prélèvements sanguins répétés.

Les résultats de cette étude sont proches de ceux rapportés précédemment dans la littérature en ce qui concerne la prévalence du SGA (36,3% contre 37% dans une méta-analyse récente) [8] et la sensibilité du TDR (86,7% contre 85% dans une récente méta-analyse) [44]. Une des limites de cette étude est que nous avons dû modifier l'item « adénopathie » de McIsaac en raison de données manquantes concernant le caractère sensible de l'adénopathie, probablement car cela est plus difficile à étudier chez l'enfant que chez l'adulte. Toutefois, nous avons pu vérifier la stabilité des résultats de l'analyse univariée et de la modélisation de sensibilité du TDR par rapport à l'analyse utilisant le critère modifié dans une analyse de sensibilité (Tableau 4, page 60).

L'Académie américaine de pédiatrie recommande que “les enfants présentant des signes suggérant une infection virale [...] ne nécessitent pas, en général, la réalisation d'un test complémentaire cherchant à identifier le SGA”[34]. Ces recommandations ne peuvent être formellement validées puisque les critères proposés pour identifier les patients suspects d'infection virale ne devant pas recevoir de TDR ou de mise en culture ne sont pas explicités sous la forme d'une RDC claire (l'Académie américaine de pédiatrie propose une simple liste de critères faisant suspecter une

virose : rhinorrhée, conjonctivite, voix enrouée, toux, stomatite, ulcérations mineures de la muqueuse bucco-pharyngée ou diarrhée). En raison du phénomène de portage qui touche environ 10% des enfants et parce que les complications non suppuratives des pharyngites à SGA sont devenues rares en Amérique du Nord et dans les pays d'Europe de l'Ouest [1], il n'est sûrement pas nécessaire de chercher à mettre en place une stratégie ayant 100% de sensibilité pour le diagnostic des pharyngites à SGA chez l'enfant. Il est également important d'insister sur le fait que les discussions ne devraient pas se focaliser uniquement sur des problématiques de sensibilité mais également prendre en compte les autres indices de performances diagnostiques tels que rapports de vraisemblance et valeurs prédictives. Nous n'avons pas mis en évidence de variations significatives de valeur prédictive négative du TDR au long du spectre clinique. Les variations de prévalence du SGA semblaient compensées de manière complexe par des variations concomitantes de sensibilité et de spécificité.

En conclusion, en considérant qu'environ 10% des enfants d'âge scolaire sont des porteurs pharyngés de SGA et que ces patients ont fréquemment des inocula faibles, les résultats de cette étude suggèrent que les cliniciens peuvent se fier au résultat du TDR pour exclure le diagnostic de pharyngite à SGA.

## 5. Tableaux et figures

Tableau 1. Distribution des scores de McIsaac et fréquences des TDR positifs, résultats de culture positifs pour le SGA et inocula forts correspondants (N = 785).

Score de McIsaac <sup>a</sup>	n (%)			
	Total	TDR positif	Culture SGA positive	Inoculum fort <sup>b</sup>
1	12 (1)	2/12 (17)	3/12 (25)	2/3 (67)
2	103 (13)	21/103 (20)	25/103 (24)	20/25 (80)
3	262 (33)	82/262 (31)	89/262 (34)	73/89 (82)
4	286 (36)	115/286 (40)	114/286 (40)	100/114 (88)
5	122 (16)	54/122 (44)	54/122 (44)	48/54 (89)

<sup>a</sup> Tous les patients avaient un score  $\geq 1$  puisque tous les patients étaient âgés de 3 à 15 ans.

<sup>b</sup>  $p = 0.09$ ,  $\chi^2$  de tendance pour la fréquence des inocula forts en fonction du score de McIsaac.

Tableau 2. Etude d'effet de spectre : performances du TDR en fonction du spectre clinique.

	n (%)	SGA (%)	$p^a$	Sensibilité (IC95)	$p^a$	Spécificité (IC95)	$p^a$	RVP (IC95)	$p^b$	RVN (IC95)	$p^b$	VPP (IC95)	$p^a$	VPN (IC95)	$p^a$
Score de McIsaac															
≤2	115 (14)	24,3	<0,001 <sup>c</sup>	75,0 (55,1–89,3)	0,02 <sup>c</sup>	97,7 (91,9–99,7)	0,04 <sup>c</sup>	32,6 (8,2–130,5)	0,23	0,26 (0,13–0,49)	0,09	91,3 (72,0–98,9)	0,61 <sup>c</sup>	92,4 (84,9–96,9)	0,83 <sup>c</sup>
3	262 (33)	34,0		84,3 (75,0–91,1)		96,0 (91,8–98,4)		20,8 (10,0–43,3)		0,16 (0,10–0,27)		91,5 (83,2–96,5)		92,2 (87,3–95,7)	
≥4	408 (51)	41,2		89,9 (84,3–94,0)		92,5 (88,4–95,5)		12,0 (7,7–18,7)		0,11 (0,07–0,17)		89,3 (83,7–93,6)		92,9 (88,9–95,8)	
Age (années)															
3–8	663 (84)	38,5	<0,01	89,0 (84,5–92,6)	<0,001	94,1 (91,4–96,2)	0,45	15,1 (10,2–22,4)	0,62	0,12 (0,08–0,17)	<0,001	90,4 (86,1–93,8)	0,48	93,2 (90,3–95,4)	0,26
9–14	122 (16)	24,6		66,7 (47,2–82,7)		96,7 (90,8–99,3)		20,4 (6,5–64,0)		0,34 (0,21–0,57)		87,0 (66,4–97,2)		89,9 (82,2–95,0)	
Absence de toux															
Oui	428 (55)	44,4	<0,001	90,0 (84,8–93,9)	0,02	94,1 (90,3–96,7)	0,65	15,3 (9,2–25,5)	0,89	0,11 (0,07–0,16)	0,02	92,4 (87,6–95,8)	0,07	92,2 (88,1–95,2)	0,76
Non	357 (45)	26,6		80,0 (70,5–87,5)		95,0 (91,7–97,3)		16,1 (9,4–27,7)		0,21 (0,14–0,31)		85,4 (76,3–92,0)		92,9 (89,2–95,7)	
Adénopathie cervicale antérieure															
Non	407 (52)	34,6	0,32	83,0 (75,7–88,8)	0,07	97,7 (95,2–99,2)	0,001	36,8 (16,6–81,4)	0,003	0,17 (0,12–0,25)	0,12	95,1 (89,7–98,2)	0,01	91,5 (87,7–94,5)	0,33
Oui	378 (48)	38,1		90,3 (84,2–94,6)		91,0 (86,6–94,4)		10,1 (6,7–15,2)		0,11 (0,06–0,18)		86,1 (79,5–91,2)		93,8 (89,9–96,6)	
Gonflement ou exsudat des amygdales															
Non	157 (20)	42,0	0,10	83,3 (72,1–91,4)	0,36	95,6 (89,1–98,8)	0,80	19,0 (7,2–49,7)	0,71	0,17 (0,10–0,30)	0,38	93,2 (83,5–98,1)	0,47	88,8 (80,8–94,3)	0,11
Oui	628 (80)	34,9		87,7 (82,6–91,7)		94,4 (91,7–96,4)		15,6 (10,5–23,3)		0,13 (0,09–0,19)		89,3 (84,4–93,1)		93,5 (90,6–95,6)	
Fièvre															
Non	246 (31)	31,3	0,049	88,3 (79,0–94,5)	0,62	93,5 (88,7–96,7)	0,43	13,6 (7,6–24,2)	0,47	0,13 (0,07–0,23)	0,66	86,1 (76,5–92,8)	0,15	94,6 (90,0–97,5)	0,22
Oui	539 (69)	38,6		86,1 (80,6–90,5)		95,2 (92,3–97,2)		17,8 (11,0–28,8)		0,15 (0,10–0,21)		91,8 (87,0–95,2)		91,6 (88,1–94,3)	
Total	785	36,3		86,7 (82,2–90,4)		94,6 (92,2–96,4)		16,1 (11,1–23,2)		0,14 (0,10–0,19)		90,1 (86,0–93,4)		92,6 (89,9–94,7)	

Abréviations: SGA, streptocoque du groupe A ; RVP, rapport de vraisemblance positif ; RVN, rapport de vraisemblance négatif ; VPP, valeur prédictive positive ; VPN, valeur prédictive négative.

<sup>a</sup> test du  $\chi^2$  ; <sup>b</sup>  $\chi^2$  de Cochran ; <sup>c</sup>  $\chi^2$  de tendance.

Tableau 3. Analyse de l'effet de l'inoculum et du spectre clinique sur la sensibilité du TDR : modélisation

Variable	Analyse univariée <sup>a</sup>				Analyse multivariée <sup>a</sup>						
	Tout inoculum (n = 285)	<i>p</i>	Inoculum faible (n = 42)	<i>p</i>	Inoculum fort (n = 243)	<i>P</i>	<i>p</i> <sup>b</sup>	Inoculum faible <sup>c</sup> (n = 42)	<i>p</i>	Inoculum fort <sup>d</sup> (n = 243)	<i>p</i>
Age (années)											
3–8	<i>ref</i>	0,01	<i>ref</i>	<0,01	<i>ref</i>	0,49	0,03	<i>ref</i>	<0,001	<i>ref</i>	0,36
9–14	-22,4 (-39,7 ; -5,1)		-37,4 (-64,1 ; -10,7)		-4,6 (-17,4 ; 8,3)			-40,9 (-57,9 ; -24,0)		-5,7 (-17,9 ; 6,4)	
Absence de toux											
Oui	<i>ref</i>	0,03	<i>ref</i>	0,09	<i>ref</i>	0,14	0,19	<i>ref</i>	<0,001	<i>ref</i>	0,08
Non	-10,0 (-19,1 ; -9,0)		-25,0 (-53,6 ; 3,6)		-5,2 (-12,1 ; 1,7)			-23,6 (-23,7 ; -23,5)		-6,0 (-12,5 ; 0,6)	
Adénopathie cervicale antérieure											
Non	<i>ref</i>	0,07	<i>ref</i>	0,01	<i>ref</i>	0,88	0,01	<i>ref</i>	0,01	<i>ref</i>	0,37
Oui	7,3 (-0,6 ; 15,2)		36,1 (7,7 ; 64,5)		-0,4 (-6,1 ; 5,2)			40,5 (16,7 ; 64,3)		2,0 (-2,3 ; 6,2)	
Gonflement ou exsudat des amygdales											
Non	<i>ref</i>	0,40	<i>ref</i>	0,12	<i>ref</i>	0,21	0,07	n/a <sup>e</sup>		<i>ref</i>	0,29
Oui	4,3 (-5,6 ; 14,3)		-25,0 (-56,2 ; 6,2)		5,4 (-3,1 ; 13,9)					4,0 (-3,4 ; 11,4)	
Fièvre											
Non	<i>ref</i>	0,61	<i>ref</i>	0,27	<i>ref</i>	0,77	0,26	n/a <sup>e</sup>		<i>ref</i>	0,41
Oui	-2,3 (-10,8 ; 6,3)		-19,1 (-53,0 ; 14,8)		1,0 (-5,6 ; 7,6)					2,3 (-3,3 ; 8,0)	

Les résultats sont exprimés sous forme de différence de sensibilité brute ou ajustée (et intervalle de confiance à 95%), sauf mention contraire.

Abréviations: *ref*, classe de référence; n/a, non applicable.

<sup>a</sup> Modèle binomial avec fonction de lien identité.

<sup>b</sup> Test d'interaction entre inoculum et variable clinique (test de Wald).

<sup>c</sup> Modèle ajusté sur l'âge, l'absence de toux et la présence d'adénopathie cervicale antérieure.

<sup>d</sup> Modèle ajusté sur l'âge, l'absence de toux, la présence d'adénopathie cervicale antérieure, la présence d'un gonflement ou exsudat des amygdales et la présence de fièvre.

<sup>e</sup> Variable non incluse dans le modèle car le modèle ne pouvait converger.

Tableau 4. Analyse de sensibilité en prenant le critère adénopathie tel que décrit par McIsaac (« adénopathie sensible ») plutôt que le critère modifié utilisé dans l'étude (« d'adénopathie cervicale antérieure  $\geq 1$  cm et/ou sensible »)

Variable	Analyse univariée <sup>a</sup>				Analyse multivariée <sup>a</sup>					
	Tout inoculum (n = 254)	<i>p</i>	Inoculum faible (n = 37)	<i>p</i>	Inoculum fort (n = 217)	<i>P</i>	Inoculum faible <sup>c</sup> (n = 37)	<i>p</i>	Inoculum fort <sup>d</sup> (n = 243)	<i>p</i>
Age (années)										
3–8							<i>ref</i>	<0,001	<i>ref</i>	
9–14							-43,0 (-62,7 ; -23,3)		-4,1 (-15,3 ; 7,1)	0,474
Absence de toux										
Oui							<i>ref</i>	0,007	<i>ref</i>	0,101
Non							-18,8 (-32,3 ; -5,2)		-5,4 (-11,8 ; 1,0)	
Adénopathie cervicale antérieure										
Non	<i>ref</i>	0,263	<i>ref</i>	0,036	<i>ref</i>	0,720	<i>ref</i>		<i>ref</i>	0,283
Oui	4,6 (-3,4 ; 12,6)		34,7 (2,3 ; 67,0)		-1,1 (-7,2 ; 5,0)		38,9 (10,8 ; 67,0)	0,007	1,7 (-1,4 ; 4,9)	
Gonflement ou exsudat des amygdales										
Non							n/a <sup>e</sup>		<i>ref</i>	0,298
Oui									4,2 (-3,7 ; 12,2)	
Fièvre										
Non							n/a <sup>e</sup>		<i>Ref</i>	0,485
Oui									1,8 (-3,2 ; 6,8)	

Les résultats sont exprimés sous forme de différence de sensibilité brute ou ajustée (et intervalle de confiance à 95%), sauf mention contraire.

Abréviations: Ref, classe de référence; n/a, non applicable.

<sup>a</sup> Modèle binomial avec fonction de lien identité.

<sup>b</sup> Test d'interaction entre inoculum et variable clinique (test de Wald).

<sup>c</sup> Modèle ajusté sur l'âge, l'absence de toux et la présence d'adénopathie cervicale antérieure.

<sup>d</sup> Modèle ajusté sur l'âge, l'absence de toux, la présence d'adénopathie cervicale antérieure, la présence d'un gonflement ou exsudat des amygdales et la présence de fièvre.

<sup>e</sup> Variable non incluse dans le modèle car le modèle ne pouvait converger.

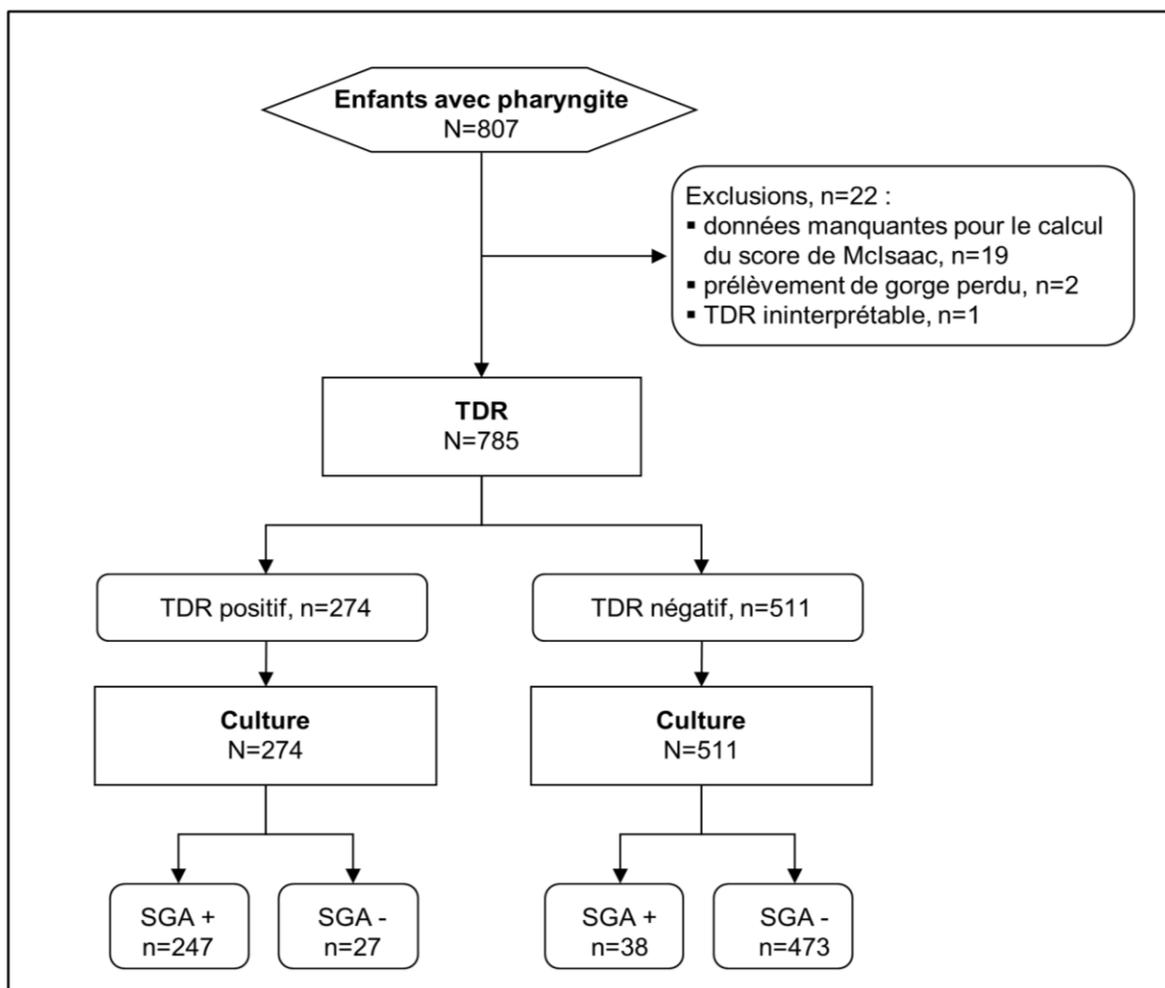


Figure 1. Flux des patients dans l'étude correspondant au Chapitre 1.

Abréviations : TDR : test de diagnostic rapide ; SGA : streptocoque du groupe A.

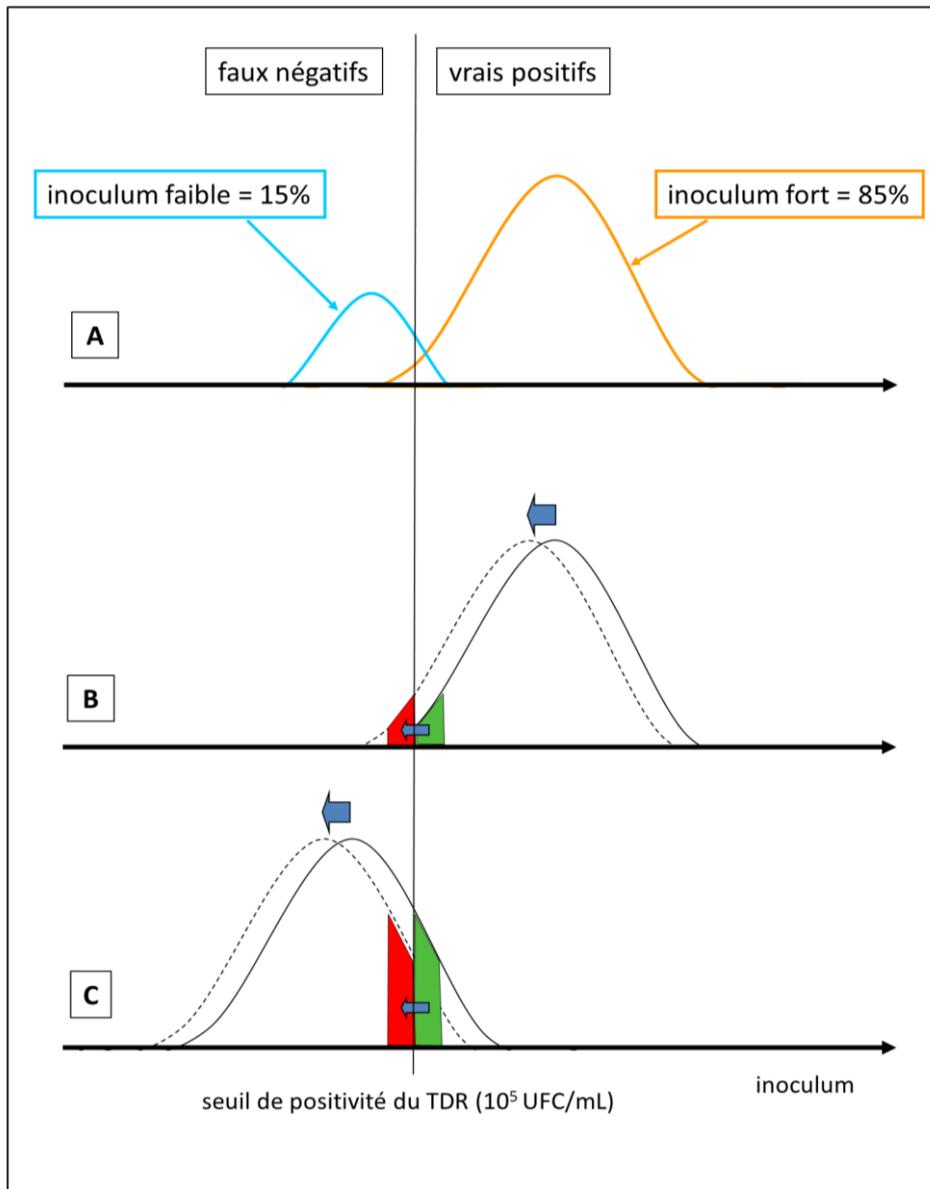


Figure 2. Hypothèse de distribution bimodale des inocula et effet sur la sensibilité du TDR

- A. Distribution bimodale des inocula. En cas d'inoculum faible, la majeure partie des patients se situent sous le seuil de positivité du TDR (sensibilité de l'ordre de 40%) ; en cas d'inoculum fort, la majorité des patients se situent au-dessus du seuil de positivité du TDR (sensibilité de l'ordre de 95%).
- B. En cas d'inoculum fort et en présence d'un signe clinique associé à une baisse d'inoculum, l'augmentation du nombre de faux-négatifs du TDR (en rouge) est modérée.
- C. En cas d'inoculum faible et en présence du même signe clinique associé à une même baisse d'inoculum, l'augmentation du nombre de faux-négatifs (en rouge) est plus importante qu'en cas d'inoculum fort.

## CHAPITRE 2

### Effet de l'inoculum bactérien, du spectre clinique et des caractéristiques des médecins sur la sensibilité du TDR

Jérémie Cohen, Martin Chalumeau, Corinne Levy, Philippe Bidet, Mohamed Benani,

Marc Koskas, Edouard Bingen, Robert Cohen

*European Journal of Clinical Microbiology and Infectious Diseases.* 2013;32(6):787-793

Impact factor 2012 : 3,024 ; rang 1/8

(voir publication en Annexe 2, page 155)

## ***1. Introduction***

Dans le **Chapitre 1** du travail doctoral, nous avons montré dans quelle mesure la présentation clinique et l'inoculum bactérien des patients ayant une pharyngite pouvaient faire varier les performances diagnostiques du TDR en nous basant sur le score de McIsaac puis nous en avons discuté la portée clinique. Nous allons, dans ce **Chapitre 2**, étendre l'étude des facteurs de variations de performances du TDR en nous intéressant d'une part à l'exploration de l'effet du portage sain de SGA au niveau pharyngé et d'autre part à l'effet éventuel du médecin réalisant le test.

En effet, environ 10% des enfants d'âge scolaire sont porteurs asymptomatiques du SGA au niveau pharyngé [8]. Un traitement antibiotique n'est pas recommandé chez ces enfants sauf circonstances exceptionnelles (antécédent personnel ou contact direct avec un cas de rhumatisme articulaire aigu, de glomérulonéphrite aiguë post-streptococcique ou d'infection invasive à SGA, patients ou personnels de communautés fermées tels qu'hôpitaux, centres de soins de suite et de réadaptation ou maisons de retraite, familles dont les membres se réinfectent régulièrement entre eux – infections « ping-pong ») [30]. Le phénomène de portage est important à prendre en compte car on peut raisonnablement estimer qu'en cas de pharyngite, environ 10% des enfants ayant une culture positive pour le SGA sont en fait des porteurs ayant une pharyngite virale concomitante chez qui un traitement antibiotique n'aurait aucun bénéfice. Ainsi, une stratégie diagnostique dans les pharyngites ayant 100% de sensibilité conduirait à de nombreuses antibiothérapies inutiles. Il est classiquement proposé pour distinguer entre pharyngite à SGA "vraie" et pharyngite virale chez un porteur de réaliser des tests sérologiques spécifiques du SGA (anticorps anti-streptolysine O [ASLO] et anti-streptodornase [ASD]) mais cela n'est pas réalisable en routine car cela nécessiterait deux prélèvements sanguins à 15 jours d'intervalle [25, 52]. Par ailleurs, comme nous

---

l'avons vu précédemment (Chapitre 1, page 54), les sérologies SGA (ASLO et ASD) posent des problèmes d'interprétation qui mettent en doute leur capacité à différencier entre infection aiguë et portage de manière fiable [53].

Etudier les performances du TDR chez des porteurs asymptomatiques pourrait être intéressant car cela permettrait d'approcher les performances du TDR chez les porteurs ayant une pharyngite virale. Les porteurs sains ont habituellement de faibles inocula bactériens et plusieurs auteurs ont émis l'hypothèse que les faux-négatifs du TDR en cas de pharyngite, qui ont habituellement des inocula faibles, seraient des porteurs ayant une pharyngite virale, minimisant ainsi le poids clinique de ces faux-négatifs. Plusieurs études ont évalué les performances du TDR chez des porteurs sains [54-57] mais ces études ont rapporté des sensibilités hétérogènes (de 20% à 100%) et la portée de leurs résultats étaient limitée par l'utilisation de TDR maintenant obsolètes (TDR par agglutination au latex) [54], un test de référence non consensuel (mise en culture avec lecture des boîtes de Pétri au bout de 24 heures au lieu des 48 heures recommandées) [55, 56], et une faible fréquence de portage asymptomatique (1,6%), suggérant un important biais de sélection ou des problèmes techniques lors de l'écouvillonnage, du transport des prélèvements au laboratoire de microbiologie ou de la mise culture [57].

Comme cela été montré dans le **Chapitre 1**, les performances du TDR sont susceptibles de varier en fonction de la présentation clinique des patients et de l'inoculum bactérien. De même, plusieurs auteurs ont montré que la sensibilité du TDR variait en fonction de la personne réalisant le test (effet médecin\*) [28, 29]. A ce jour, aucune étude n'a été étudié l'effet indépendant de ces facteurs de variations sur la sensibilité du TDR, ni étudié les caractéristiques des médecins susceptibles d'expliquer l'effet médecin. Les objectifs de ce chapitre étaient donc d'étudier les effets

indépendants du spectre clinique pris de manière binaire et extrême (pharyngite contre portage asymptomatique), de l'inoculum bactérien et du médecin réalisant le TDR et de tenter d'identifier des caractéristiques des médecins pouvant expliquer l'effet médecin.

## **2. Méthodes**

### **2.1 Patients et prélèvements de gorge**

Cette analyse est basée sur l'ensemble des données de l'étude ACTIV TDR SGA pour les deux années d'inclusion, 2009-2010 et 2010-2011, et pour les deux groupes de patients éligibles, enfants avec pharyngite et enfants asymptomatiques (voir Sources de données, page 37).

### **2.2 Questionnaire destiné aux investigateurs**

Les 17 pédiatres ayant participé à cette étude font partie du réseau ACTIV qui a été décrit précédemment (Voir Sources de données, page 36). Un an après la fin de l'étude clinique, tous les pédiatres ayant participé à l'étude ont été invités à répondre à un questionnaire en ligne. Le questionnaire a été développé sur la base d'une revue de la littérature concernant l'influence des facteurs contextuels et organisationnels\* sur les pratiques des médecins [13, 58, 59]. Le questionnaire a initialement été construit selon quatre domaines concernant 1/ les caractéristiques générales des médecins (âge [variable continue], genre, année d'obtention de la thèse de médecine [variable continue]) ; 2/ leur type d'exercice médical et les caractéristiques du cabinet (nombre de médecins dans le cabinet [ $\leq 2$  ou  $> 2$ ], nombre d'heures d'activité clinique au cabinet par semaine [variable continue], estimation personnelle du pourcentage de patients couverts par la CMU [variable continue], existence d'une activité clinique hospitalière en plus de leur activité clinique au cabinet de ville) ; 3/ leurs habitudes concernant la réalisation du

TDR (personne réalisant le prélèvement de gorge [médecin ou autre, par exemple assistante], site préférentiel de prélèvement de gorge [une amygdale ou les 2], le nombre de gouttes de chaque réactif nécessaire à la bonne réalisation du TDR [le fabricant recommande 4 gouttes], l'utilisation d'un chronomètre pour mesurer les temps de réaction chimique nécessaires à la bonne réalisation du TDR ; 4/ leurs connaissances et croyances concernant la pathologie à SGA (estimation de la prévalence du SGA dans les pharyngites de l'enfant [ $<20\%$ ,  $20-40\%$  ou  $>40\%$ ] et leur estimation de l'incidence du rhumatisme articulaire aigu en France chez l'enfant de 3 à 15 ans [variable continue]).

Dans la version finale du questionnaire (voir annexe 8 page 225), les questions ont été réorganisées de manière aléatoire afin d'éviter que les participants ne découvrent pas les 4 domaines du questionnaire et que cela n'influence leurs réponses, puis l'ordre des questions a été figé afin que tous les médecins soient soumis au même questionnaire. Les données concernant les investigateurs ont ensuite été anonymisées et chaînées avec les données individuelles des patients.

### **2.3 Analyses statistiques**

La culture au laboratoire de bactériologie a été considérée comme le test de référence. Premièrement, nous avons décrit les caractéristiques générales de la population d'étude, la fréquence du SGA chez les patients avec pharyngite et chez les enfants asymptomatiques (taux de portage asymptomatique) ainsi que les performances diagnostiques globales du TDR (sensibilité et spécificité). Deuxièmement, les caractéristiques des médecins ayant participé à l'étude ont été décrites. Troisièmement, nous avons réalisé des analyses univariées utilisant des tests du  $\chi^2$  afin d'identifier les facteurs liés aux patients et liés aux médecins susceptibles d'affecter la sensibilité du

TDR. Quatrièmement, nous avons utilisé un modèle logistique multiniveaux à intercept aléatoire pour prendre en compte la structure hiérarchique\* des données, avec les patients au niveau 1 et les médecins au niveau 2 [60-62]. En limitant l'analyse aux patients ayant une culture positive à SGA et en prenant le résultat du TDR comme variable prédite par le modèle, cela nous a permis d'étudier l'effet indépendant des facteurs liés aux patients et liés aux médecins sur la sensibilité du TDR.

Le modèle estime les associations sous forme d'odds ratio d'avoir un TDR positif chez les patients ayant une culture positive pour le SGA [61]. Les facteurs liés aux patients étaient le statut clinique (pharyngite ou enfant asymptomatique), l'âge dichotomisé en deux catégories de même étendue (<9 ans contre  $\geq 9$  ans) et l'inoculum bactérien (inoculum faible ou fort). Les facteurs liés aux médecins étaient ceux recueillis dans le questionnaire décrit précédemment, en dichotomisant les variables continues autour de leur médiane. Tous les modèles ont également été ajustés sur l'année d'inclusion (2009-2010 vs 2010-2011).

La stratégie de modélisation était basée sur la construction de modèles de complexité croissante. Tout d'abord, un modèle sans variable explicative (modèle vide) a été ajusté afin d'estimer la variance de base au niveau médecins ( $\tau_{00}$ ), c'est-à-dire l'effet médecin\*. Ensuite, les variables individuelles liées aux patients (niveau 1) ont été entrées en bloc dans le modèle (Modèle 1). Enfin, les variables liées aux médecins (niveau 2) associées à la sensibilité du TDR en analyse univariée avec un  $p \leq 0,2$  ont été entrées une par une dans le modèle (Modèles 2-5). Pour chaque modèle nous avons pu calculer la variance résiduelle de niveau 2 ( $\tau_{00}^*$ ) et la part de variance expliquée ( $PVE = [\tau_{00} - \tau_{00}^*] / \tau_{00}$ ), ce qui nous a permis d'étudier dans quelle mesure l'effet médecin pouvait être expliqué par des différences de distribution des variables liées aux

patients ou aux médecins. La probabilité  $p_{ij}$  que le résultat du TDR soit positif chez le patient  $i$  vu par le médecin  $j$  a donc été modélisée ainsi :

	Modèle vide	Modèle 1	Modèles 2–5
Niveau 1	$Logit(p_{ij}) = \beta_{0j}$	$Logit(p_{ij}) = \beta_{0j} + \sum_{q=1}^Q \beta_{qj} X_{qij}$	$Logit(p_{ij}) = \beta_{0j} + \sum_{q=1}^Q \beta_{qj} X_{qij}$
Niveau 2	$\beta_{0j} = \gamma_{00} + u_{0j}$	$\beta_{0j} = \gamma_{00} + u_{0j}$	$\beta_{0j} = \gamma_{00} + \gamma_{01} W_{1j} + u_{0j}$

Où  $\beta_{0j}$  est l'intercept aléatoire variable d'un médecin  $j$  à l'autre,  $\beta_{qj}$  est le coefficient de la  $q^{\text{ème}}$  variable de niveau 1 spécifique au médecin  $j$ ,  $X_{qij}$  est la  $q^{\text{ème}}$  variable de niveau 1 du patient  $i$  chez le médecin  $j$ ,  $\gamma_{00}$  est l'intercept moyen de sensibilité du TDR,  $\gamma_{01}$  est le coefficient de la variable de niveau 2 spécifique du médecin  $j$ ,  $W_{1j}$  est la variable de niveau 2 pour le médecin  $j$ , et  $u_{0j}$  est l'effet aléatoire de niveau 2 avec  $u_{0j} \sim N(0, \tau_{00})$  [63].

### 3. Résultats

#### 3.1 Caractéristiques des patients, prévalence du SGA et performances globales du TDR

Nous avons inclus un total de 1781 enfants. Deux patients au TDR ininterprétable et 3 patients pour qui le prélèvement de gorge a été perdu ont été secondairement exclus. L'analyse a donc porté sur 1776 patients (790 filles, 44%) dont 1482 enfants avec pharyngite et 294 enfants asymptomatiques. L'âge moyen (DS) des enfants inclus était de 6,3 (2,6) ans. La prévalence du SGA dans le groupe pharyngite était de 38% (IC95 36%–41%) et le taux de portage asymptomatique de SGA dans le groupe d'enfants sains était de 11% (7%–14%). La sensibilité globale du TDR dans la population étudiée était de 87% (84%–90%) et la spécificité globale était de 93% (91%–94%).

### 3.2 Caractéristiques des médecins investigateurs

L'âge moyen (écart-type) des 17 médecins ayant participé à l'étude était de 53,4 (7,1) ans et 8 (47%) étaient des femmes (Tableau 1). Sur les 17 médecins ayant participé à l'étude, 10 (59%) ont déclaré avoir une activité clinique hospitalière en plus de leur activité clinique au cabinet. Tous les médecins sauf un ont déclaré réaliser eux-mêmes les prélèvements de gorge et tous ont déclaré lire le résultat du TDR eux-mêmes.

### 3.3 Effet du spectre clinique et de l'inoculum bactérien

En analyse univariée, la sensibilité du TDR était plus élevée chez les enfants ayant une pharyngite que chez les porteurs asymptomatiques (89% [87%–92%] contre 41% [24%–59%],  $p < 0,001$  ; Tableau 2, page 78). Il y avait également une tendance à une sensibilité plus élevée chez les enfants ayant un âge inférieur à 9 ans que chez les enfants plus âgés (88% [85%–91%] contre 79% [70%–89%],  $p=0,053$ ).

Parmi les 602 cultures positives pour le SGA, 501 (83% [80%–86%]) étaient des inocula forts et les inocula forts étaient moins fréquents chez les porteurs asymptomatiques que chez les enfants ayant une pharyngite (34% [17%–52%] contre 86% [83%–89%],  $p < 0,001$ ). En analyse univariée, la sensibilité du TDR augmentait avec l'inoculum bactérien (53% [44%–63%] en cas d'inoculum faible contre 94% [91%–96%] en cas d'inoculum fort,  $p < 0,001$  ; Tableau 2, page 78). La sensibilité du TDR était plus élevée durant la deuxième année d'étude que pendant la première (92% [89%–95%] contre 82% [78%–86%],  $p < 0,001$ ).

En analyse multivariée, la sensibilité du TDR était plus élevée en cas de pharyngite que chez les porteurs asymptomatiques (odds ratio ajusté [ORa] 5,0 [1,9–13,0],  $p=0,001$ ), chez les enfants de moins de 9 ans par rapport aux enfants de plus de 9 ans (ORa 2,3 [1,1–5,1],  $p=0,04$ ) et chez les enfants ayant un inoculum fort retrouvé en

culture (ORa 10,2 [5,7–18,5],  $p < 0,001$  ; Modèle 1, Tableau 3, page 79). Ces associations et leur force étaient stables après ajustement sur les caractéristiques des médecins (Tableau 3, page 79).

### 3.4 Effet médecin

Nous avons mis en évidence des variations significatives de sensibilité du TDR en fonction du médecin ayant réalisé le test, ce qui correspond à un effet médecin significatif (sensibilité variant de 56% à 96%,  $p=0,01$  ; Figure 1). En analyse univariée sur les caractéristiques des médecins, la sensibilité du TDR était plus élevée chez les médecins ayant  $\leq 5\%$  de leurs patients couverts par la CMU (89% [86%–92%] contre 83% [78%–88%],  $p=0,03$  ; Tableau 2) et ceux ayant une activité clinique hospitalière en plus de leur activité clinique au cabinet (91% [88%–94%] contre 78% [72%–84%],  $p < 0,001$ ). La sensibilité du TDR était légèrement plus élevée mais de manière non statistiquement significative chez les médecins ayant plus d'années d'exercice clinique (89% [85%–92%] vs. 84% [79%–89%],  $p=0,10$ ) et ceux ayant déclaré utiliser le bon nombre de gouttes de chaque réactif pour la réalisation du TDR 88% [85%–90%] contre 81% [70%–91%],  $p=0,15$ ).

En analyse multiniveau, le modèle vide a confirmé l'existence d'un effet médecin significatif ( $\tau_{00}=0,19$ ,  $p=0,03$ ) et cet effet médecin ne semblait pas expliqué par des différences de distribution des variables liées aux patients d'un médecin à l'autre (Modèle 1,  $\tau_{00}^*=0,21$  ; Tableau 3, page 79). En analyse multivariée multiniveaux prenant en compte les variables liées aux patients et aux médecins, le fait d'avoir une activité clinique hospitalière en plus de l'activité clinique au cabinet était la seule caractéristique des médecins associée de manière significative à des variations de

sensibilité (ORa 3,4 [1,9–6,3],  $p < 0,001$ ), et cette variable expliquait la totalité de l'effet médecin (Modèle 3,  $\tau_{00}^* < 0,001$ , PVE >99%).

#### **4. Discussion**

A notre connaissance, il s'agit de la première étude cherchant à explorer l'effet indépendant de ces facteurs de variations sur la sensibilité du TDR. Nos résultats suggèrent que certaines caractéristiques des médecins qui réalisent le test sont susceptibles d'affecter la sensibilité du TDR indépendamment des caractéristiques des patients et de l'inoculum bactérien.

Les résultats de cette étude confirment l'effet du spectre clinique sur les performances du TDR, en montrant une sensibilité plus élevée chez les enfants avec pharyngite que les porteurs asymptomatiques. Dans le **Chapitre 1**, nous avons montré l'effet du spectre clinique mesuré par le score de McIsaac et ses composantes (adénopathies cervicales antérieures sensibles, fièvre, inflammation des amygdales et toux) [19] sur la sensibilité du TDR. Dans la présente étude, le spectre clinique a été considéré de manière extrême et binaire par le statut clinique (pharyngite ou asymptomatique). En effet, le score de McIsaac n'a pas de sens chez les enfants asymptomatiques puisque le score a été construit comme un score prédictif de la présence du SGA en cas de pharyngite. En ce qui concerne les hypothèses physiopathologiques, l'effet indépendant du statut clinique sur la sensibilité du TDR pourrait être expliquée par des variations locorégionales de la muqueuse pharyngée dues à l'infection et à l'inflammation en cas de pharyngite telles la formation d'un biofilm\*, des modifications d'exposition ou de structure antigénique, des phénomènes d'internalisation cellulaire\* en cas de portage, ou encore des différences de distribution des génotypes *emm*\* entre enfants atteints de pharyngite et porteurs asymptomatiques

[64].

Nos résultats confirment également l'importance de l'inoculum bactérien et de l'effet médecin sur la sensibilité du TDR. D'habitude, l'effet de l'inoculum et l'effet médecin sont conçus comme étant liés, les médecins les plus expérimentés obtenant plus souvent un fort inoculum bactérien d'une part et les inocula forts étant associés à une sensibilité plus élevée d'autre part. Cependant, nos résultats montrent que l'effet médecin sur la sensibilité du TDR persiste de manière indépendante de l'inoculum. Nous avons mis en évidence de fortes variations de sensibilité en fonction du médecin réalisant le TDR (sensibilité allant de 56% à 96%), et cette variabilité ne diminuait pas de manière significative en analyse multivariée multiniveaux prenant en compte l'inoculum bactérien (Modèle 1, Tableau 3).

Parmi les caractéristiques des médecins associées à des variations de sensibilité du TDR en analyse univariée, le fait d'avoir une activité clinique hospitalière en plus de l'activité clinique au cabinet était la seule variable restant associée à la sensibilité du TDR en analyse multivariée multiniveaux et cette variable expliquait la totalité de l'effet médecin (part de variance expliquée >99%). Cette variable ne semble pas pouvoir expliquer en soit les variations de sensibilité observées mais pourrait capturer de manière indirecte l'effet d'autres variables qui n'ont pas été mesurées dans cette étude, faisant de cette variable un probable facteur intermédiaire. Par exemple, il serait plausible que l'effet de cette variable passe par le fait que les médecins hospitaliers bénéficient de plus de formation médicale et d'interactions avec des pairs que les médecins ayant une activité strictement libérale. Ces résultats sont cohérents avec ceux d'une précédente étude qui avait montré que l'adéquation aux recommandations dans la prise en charge des enfants avec pharyngite dépendait de caractéristiques du médecin telles que la spécialité (généraliste ou pédiatre) ou le type d'exercice (cabinet

comportant un ou 2 médecins ou cabinet de groupe ou encore cabinet situé en zone rurale ou urbaine) [13].

En termes de validité externe, les résultats de notre étude concernant certains indicateurs de base sont cohérents avec les résultats de la littérature. Ainsi, la prévalence du SGA en cas de pharyngite était de 38% (IC95 36%–41%) et la fréquence du portage sain de SGA était de 11% (7%–14%), ce qui est très proche des résultats d'une récente méta-analyse (37% [32%–43%] et 12% [9%–14%], en cas de pharyngite ou de portage respectivement) [8].

A notre connaissance, il s'agit de la première étude ayant utilisé une modélisation multiniveaux pour étudier les variations de performances d'un test diagnostique. Cette extension multiniveaux du modèle logistique proposé par Coughlin *et al.* [61] permet de prendre en compte la structure hiérarchique des données issues d'études diagnostiques multicentriques. Dans le **Chapitre 1**, nous avons utilisé un modèle binomial avec lien identité plutôt que logistique afin d'étudier les facteurs associés à des variations de sensibilité du TDR. Ce modèle avait l'avantage de mesurer les associations sous la forme de différence de sensibilité, ce qui est très explicite pour le lecteur, en particulier les cliniciens. Cependant, nous avons fait face à des problèmes de convergence du modèle nous empêchant d'ajuster certains modèles contenant plus de 3 variables et nous obligeant à stratifier nos analyses sur l'inoculum pour prendre en compte l'interaction existant entre certains signes cliniques et l'inoculum bactérien [65]. Le modèle logistique utilisé ici est particulièrement flexible en termes de covariables pouvant être intégrées dans le modèle [51]. A l'inverse du modèle binomial avec fonction de lien identité, le modèle logistique estime les associations sous la forme d'*odds ratio* qui sont peu intuitifs à interpréter dans le cas de la modélisation des performances d'un test diagnostique. Les *odds ratios* risquent d'exagérer l'effet des

---

covariables s'ils sont mal interprétés, c'est-à-dire interprétés comme des risques relatifs. Il est en effet classique de considérer que plus l'évènement d'intérêt est fréquent (ici la positivité du TDR), plus l'*odds ratio* sera éloigné du risque relatif.

Cette étude présente plusieurs limites. Premièrement elle est basée sur un petit échantillon de pédiatres français faisant partie d'un réseau dédié à la recherche et à l'enseignement et pratiquant tous dans des zones urbaines (réseau ACTIV). Ces pédiatres participent régulièrement à des programmes de formation médicale continue et à des études épidémiologiques observationnelles, interventionnelles ou diagnostiques [20]. Ainsi, même si notre population d'enfants semble représentative en termes de prévalence du SGA, il est possible que notre population de pédiatres investigateurs ne soit pas représentative des médecins de ville exerçant en France. Nos résultats doivent donc être confirmés sur de plus larges échantillons de médecins, généralistes et pédiatres. Cependant, même dans ce groupe de pédiatres semblant homogène, nous avons mis en évidence un effet médecin cliniquement important et statistiquement significatif sur la sensibilité du TDR, ainsi que des caractéristiques des médecins susceptibles de l'expliquer.

Deuxièmement, bien que nos résultats confirment que les inocula forts soient moins fréquents en cas de portage que de pharyngite à SGA (34% vs. 86%) [66], nos résultats ne nous permettent pas d'appuyer l'hypothèse selon laquelle les résultats faux-négatifs du TDR en cas de pharyngite seraient en majorité des porteurs ayant une pharyngite virale concomitante [25]. Nous avons inclus des enfants asymptomatiques dans cette étude dans l'idée que la sensibilité du TDR observée chez ces patients approcherait la sensibilité du TDR chez les porteurs avec pharyngite virale concomitante. Cependant, puisque la sensibilité du TDR est affectée par le statut clinique (pharyngite contre portage), cette approximation n'est pas valable.

En conclusion, les résultats de cette étude élargissent la vision que nous pouvons avoir des facteurs susceptibles de faire varier les performances du TDR. Le diagnostic médical, comme tout processus de décision clinique, est soumis à l'influence de nombreux facteurs [58, 59]. Cette étude confirme que la sensibilité du TDR ne doit pas être appréhendée comme un paramètre fixe intrinsèque au test, mais plutôt un paramètre affecté de manière indépendante par la présentation clinique des patients ainsi que par le médecin réalisant le test. Le fait que la sensibilité du TDR varie en fonction du médecin réalisant le test indépendamment du spectre clinique et de l'inoculum bactérien devrait être pris en compte dans les programmes de formation initiale ou continue des médecins dans le but de tirer les performances des médecins vers le haut et de réduire les variations de sensibilité pouvant exister entre eux (*benchmarking*).

En conclusion, nos résultats vont dans le sens d'une recommandation récente de l'Académie Américaine de Pédiatrie selon laquelle les médecins qui se basent uniquement sur le résultat du TDR afin de décider à quels patients ils prescriront des antibiotiques, comme cela est habituellement fait en France, devraient évaluer les performances qu'ils obtiennent avec le TDR en comparaison à la culture [67]. Les médecins obtenant une sensibilité insuffisante pourraient bénéficier de formations ciblées sur la réalisation du TDR.

## 5. Tableaux et figures

Tableau 1. Caractéristiques des pédiatres ayant participé à l'étude (n=17)

Caractéristiques	
Age, en années, moyenne (DS)	53,4 (7,1)
Femmes	8 (47)
Année d'obtention de la thèse d'exercice, moyenne (range)	1987 (1979–2002)
Type de pratique clinique	
Nombre de médecins dans le cabinet	
1 ou 2	6 (35)
Cabinet de groupe (>2 cliniciens)	11 (65)
Volume d'activité clinique au cabinet, en heures (médiane, range)	37 (20–55)
Proportion estimée de patients (%) couverts par la CMU, médiane (range)	5 (1–15)
Activité clinique hospitalière	10 (59)
Habitudes du médecin dans la réalisation du TDR	
Site de prélèvement de gorge	
1 amygdale	5 (29)
les 2 amygdales	12 (71)
Utilisation du nombre correct de gouttes pour chaque réactif servant à réaliser le TDR <sup>a</sup>	13 (76)
Utilisation d'un chronomètre pour contrôler les temps de réaction nécessaires au TDR <sup>a</sup>	6 (35)
Connaissances et croyances concernant la pathologie à SGA	
Prévalence estimée du SGA dans les pharyngitis de l'enfant	
<20%	2 (12)
20–40%	12 (71)
>40%	3 (18)
Incidence estimée du RAA chez les enfants de 3 à 15 ans en France, en nombre de cas/an, moyenne (DS)	2,6 (3,7)

Les résultats sont des nombres (%), sauf mention contraire.

Abréviations: TDR, test de diagnostic rapide ; SGA, streptocoque du groupe A ; RAA, rhumatisme articulaire aigu.

<sup>a</sup>Les instructions du fabricant sont les suivantes: 1) placer 4 gouttes du réactif A et du réactif B dans le tube ; 2) plonger l'écouvillon dans le tube et le tourner 10 fois ; 3) laisser tremper l'écouvillon dans le tube pendant 1 minute ; 4) presser l'écouvillon contre les parois du tube afin d'en exprimer le contenu ; 5) retirer l'écouvillon ; 6) placer la bandelette réactive dans le tube ; 7) lire le résultat de la bandelette au bout de 5 minutes.

Tableau 2. Sensibilité du TDR en fonction des variables de niveau patient et de niveau médecin : analyse univariée

	Nombre <sup>a</sup> (%)	Sensibilité, % (IC95)	p
<b>VARIABLES DE NIVEAU PATIENT</b>			
Statut clinique			
Asymptomatique	32 (5)	41 (24–59)	<0,001
Pharyngite	570 (95)	89 (87–92)	
Age, années			
3–8	534 (89)	88 (85–91)	0,053
9–14	68 (11)	79 (70–89)	
Inoculum <sup>b</sup>			
Faible	101 (17)	53 (44–63)	<0,001
Fort	501 (83)	94 (91–96)	
<b>VARIABLES DE NIVEAU MEDECIN</b>			
Proportion estimée de patients (%) couverts par la CMU			
≤5	10 (59)	89 (86–92)	0,03
>5	7 (41)	83 (78–88)	
Activité clinique hospitalière			
Non	7 (41)	78 (72–84)	<0,001
Oui	10 (59)	91 (88–94)	
Année d'obtention de la thèse d'exercice			
≤1987	10 (59)	89 (85–92)	0,10
>1987	7 (41)	84 (79–89)	
Utilisation du nombre correct de gouttes pour chaque réactif servant à réaliser le TDR <sup>c</sup>			
Non	4 (24)	81 (70–91)	0,15
Oui	13 (76)	88 (85–90)	

IC95, intervalle de confiance à 95%.

<sup>a</sup> Nombre de patients (n=602) ou de médecins (n=17) ; <sup>b</sup> faible, ≤50 colonies de SGA par boîte de Pétri ; fort, >50 colonies de SGA par boîte de Pétri ; <sup>c</sup> Les instructions du fabricant sont d'utiliser 4 gouttes du réactif A et 4 gouttes du réactif B.

Tableau 3. Sensibilité du TDR en fonction des variables de niveau patient et de niveau médecin : analyse multivariée multiniveaux

	Nombre <sup>a</sup> (%)	Modèle 1		Modèle 2		Modèle 3		Modèle 4		Modèle 5	
		ORa (IC95)	p	ORa (IC95)	p	ORa (IC95)	p	ORa (IC95)	p	ORa (IC95)	p
<b>VARIABLES DE NIVEAU PATIENT</b>											
Statut clinique											
Asymptomatique	32 (5)	1	0,001	1	0,001	1	0,001	1	0,001	1	0,001
Pharyngite	570 (95)	5,0 (1,9–13,0)		4,8 (1,8–12,4)		4,5 (1,8–11,4)		5,0 (1,9–12,9)		5,1 (2,0–13,1)	
Age, années											
3–8	534 (89)	2,3 (1,1–5,1)	0,04	2,4 (1,1–5,4)	0,03	2,5 (1,1–5,5)	0,02	2,3 (1,1–5,2)	0,04	2,3 (1,1–5,1)	0,04
9–14	68 (11)	1		1		1		1		1	
Inoculum <sup>b</sup>											
Faible	101 (17)	1	<0,001	1	<0,001	1	<0,001	1	<0,001	1	<0,001
Fort	501 (83)	10,2 (5,7–18,5)		10,3 (5,7–18,6)		11,0 (6,1–20,0)		10,1 (5,6–18,3)		10,1 (5,6–18,3)	
<b>VARIABLES DE NIVEAU MEDECIN</b>											
Proportion estimée de patients (%) couverts par la CMU											
≤5	10 (59)	-	-	1,5 (0,7–3,3)	0,35	-	-	-	-	-	-
>5	7 (41)	-	-	1		-	-	-	-	-	-
Activité clinique hospitalière											
Non	7 (41)	-	-	-	-	1	<0,001	-	-	-	-
Oui	10 (59)	-	-	-	-	3,4 (1,9–6,3)		-	-	-	-
Année d'obtention de la thèse d'exercice											
≤1987	10 (59)	-	-	-	-	-	-	1,2 (0,5–2,6)	0,67	-	-
>1987	7 (41)	-	-	-	-	-	-	1		-	-
Utilisation du nombre correct de gouttes pour chaque réactif servant à réaliser le TDR <sup>c</sup>											
Non	4 (24)	-	-	-	-	-	-	-	-	1	0,75
Oui	13 (76)	-	-	-	-	-	-	-	-	1,2 (0,4–3,2)	
Variance résiduelle de niveau 2		0,21		0,21		<0,001		0,18		0,21	
PVE (%) <sup>*</sup>		n/a <sup>**</sup>		n/a <sup>**</sup>		>99%		14%		n/a <sup>**</sup>	

Abréviations: ORa, odds ratio ajusté ; IC95, intervalle de confiance à 95% ; PVE, part de variance expliquée ; n/a, non applicable.

Tous les modèles sont ajustés sur l'année d'inclusion (année 1 contre année 2).

<sup>a</sup> Nombre de patients (n=602) ou de médecins (n=17) ; <sup>b</sup> faible, ≤50 colonies de SGA par boîte de Pétri ; fort, >50 colonies de SGA par boîte de Pétri ; <sup>c</sup> Les instructions du fabricant sont d'utiliser 4 gouttes du réactif A et 4 gouttes du réactif B ; \* la PVE est calculée sur la base de la variance de niveau 2 du modèle vide (0,19 ; p=0,03) pour le modèle 1 et sur la variance de niveau 2 du modèle 1 pour les modèles 2–5 ; \*\*non applicable car la variance de niveau 2 augmentait entre les 2 modèles comparés.

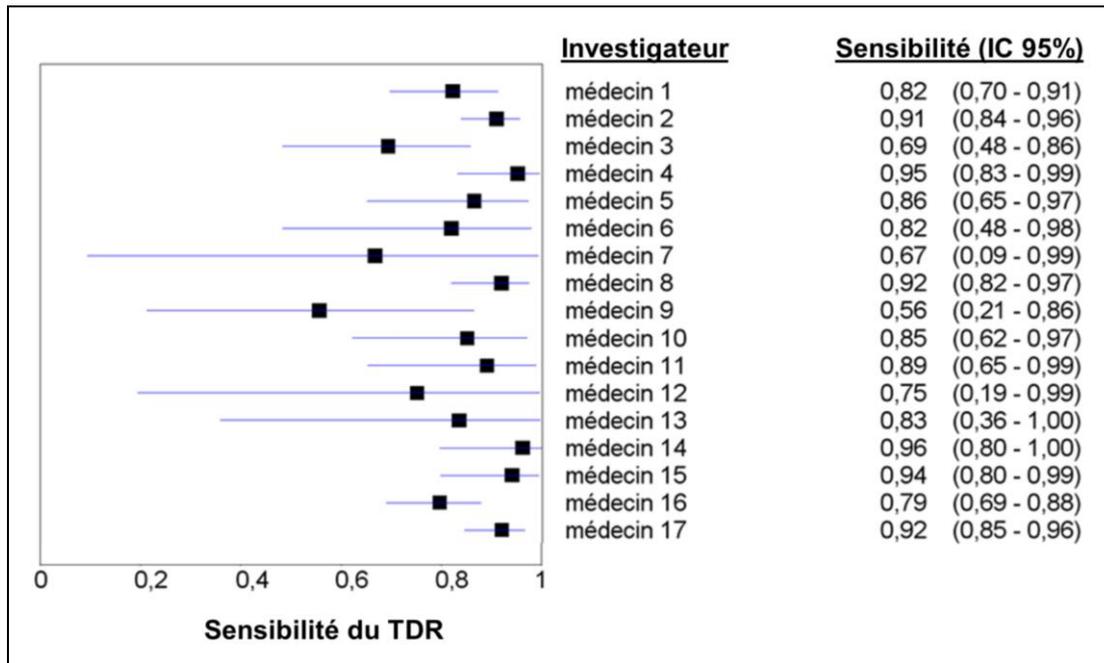


Figure 1. Effet médecin : variations de sensibilité du TDR en fonction du médecin ayant réalisé le test ( $p=0,01$ ).

## CHAPITRE 3

### **Analyse des faux-positifs du TDR : une approche cas-témoins**

Jérémie Cohen, Robert Cohen, Philippe Bidet, Corinne Levy, Patrice Deberdt,

Camille d'Humières, Sandrine Liguori, Francois Corrard, Franck Thollot,

Patricia Mariani-Kurkdjian, Martin Chalumeau, Edouard Bingen

*The Journal of Pediatrics*. 2013 Jun;162(6):1282-1284

Impact factor 2012 : 4,035 ; rang 1/12

(voir publication en Annexe 3, page 163)

## 1. Introduction

Au cours des deux premiers chapitres du travail doctoral, nous avons étudié différents facteurs liés aux patients ou aux médecins susceptibles de faire varier les performances du TDR en nous focalisant principalement sur les facteurs de variation de sensibilité du TDR. Dans le **Chapitre 3** nous tenterons d'explorer les facteurs de variation de la spécificité du TDR en évaluant les différentes raisons pouvant potentiellement expliquer les résultats faux-positifs (FP) du TDR, c'est-à-dire les cas de prélèvements de gorge avec TDR positif et culture négative pour le SGA.

En pratique clinique, les TDR ont habituellement une spécificité de l'ordre de 95% [21, 37]. Plusieurs raisons susceptibles d'expliquer les résultats FP du TDR ont été rapportées dans la littérature. Les FP peuvent être dus à une réaction croisée avec des streptocoques du groupe *milleri*\* [53, 68] ou à des variants nutritionnels\* ou non- $\beta$ -hémolytiques du SGA qui ne sont alors pas identifiés par les techniques habituelles de culture sur gélose au sang [53, 69, 70]. Une autre hypothèse serait la présence de SGA non viables lors de la mise en culture, en raison d'un délai prolongé entre le prélèvement de gorge et la mise en culture ou la prise récente d'antibiotiques [71]. Enfin, les FP du TDR pourraient être dus à la présence de bactéries pouvant empêcher la pousse du SGA en culture, par exemple *Staphylococcus aureus*, via des mécanismes d'interférence bactérienne\* [72].

Une précédente étude a cherché à expliquer les résultats FP du TDR par la présence de SGA non viables en utilisant une technique de réaction par polymérisation en chaîne (PCR) chez 200 enfants atteints de pharyngite. La PCR SGA était positive chez 8/18 (44%) des FP du TDR [73]. Dans cette étude, les auteurs avaient utilisé une amorce de PCR détectant le gène *MF* (également appelé *speF* par certains auteurs) qui n'est pas présent chez toutes les souches de SGA, contrairement au gène *speB*\*, connu

pour être ubiquitaire chez toutes les souches de SGA [74]. De plus, cette étude n'avait pas exploré les différents mécanismes susceptibles d'expliquer la présence de SGA non viables.

Les objectifs de cette étude ancillaire étaient donc d'explorer les différentes hypothèses pouvant expliquer les résultats FP du TDR, tout en utilisant une technique de PCR SGA basée sur une amorce originale du gène *speB*.

## **2. Méthodes**

### **2.1 Patients**

Cette étude ancillaire était basée sur les données du groupe pharyngite de la deuxième année d'inclusion de l'étude ACTIV TDR SGA (2010-2011).

### **2.2 TDR et mise en culture**

Comme cela a déjà été décrit précédemment, un double prélèvement de gorge a été obtenu chez chaque patient (Venturi Transystem Amies agar gel, COPAN, Corona, CA, USA). Le premier écouvillon servait à réaliser le TDR directement au cabinet du pédiatre (StreptAtest, Dectrapharm, Strasbourg, France) et le deuxième écouvillon était envoyé au laboratoire de bactériologie de l'Hôpital Robert Debré pour une mise en culture dans les 72 heures. Dès réception, l'écouvillon était dilué dans 1 ml d'eau stérile puis une goutte du liquide extrait de l'écouvillon étaitensemencée sur une gélose sélective contenant de l'acide nalidixique et de la colistine (Columbia ANC+5% sheep blood, Biomérieux, Lyon, France). La boîte de Pétri était mise en incubation à 37°C en anaérobiose pendant 48 heures et le groupage des colonies  $\beta$ -hémolytiques était effectué par agglutination sur latex (Prolex, Pro-Lab Diagnostics, Canada). Le résultat de la culture était binaire (positif ou négatif pour le SGA). Ensuite, l'écouvillon était placé

dans un bouillon d'enrichissement de type TGY d'un volume de 9 ml pour incubation à 37°C et, si la boîte de Pétri ne mettait pas en évidence de colonie  $\beta$ -hémolytique, une goutte de ce bouillon d'enrichissement TGY était mis en culture sur une seconde boîte de Pétri selon la technique décrite précédemment (Figure 1, page 94).

### 2.3 PCR SGA

Si les deux boîtes de cultures (culture standard et culture après enrichissement en bouillon TGY) étaient négatives pour le SGA, les 9 ml de bouillon TGY restants étaient centrifugés pendant 5 minutes à 5000g puis le centrifugat était resuspendu dans 180  $\mu$ L de tampon G2 et 20  $\mu$ L de lysozyme (50 mg/mL) en utilisant l'EZ1 DNA Tissue Kit (Qiagen, Courtaboeuf, France). Après incubation à 37°C pendant 30 minutes, l'extraction d'ADN était réalisée sur l'EZ1 BioRobot (Qiagen, Courtaboeuf, France). L'extrait d'ADN était ensuite utilisé pour réaliser une PCR spécifique du SGA en utilisant une paire d'amorces servant à amplifier un fragment de 240 paires de bases du gène *speB* : SGA-*speB*-F (5'-ATA CCA ACA TCA GCC ATC AAT TC-3') et SGA-*speB*-R (5'-TGC TAC TGC AAC TGC TCA AAT TA-3'). Pour cela, 5  $\mu$ l d'extrait d'ADN étaient mélangés à 45  $\mu$ l de 2 $\times$ AmpliTaq Gold PCR Master Mix (Applied Biosystems, Foster City, CA, USA) et 1  $\mu$ M de chaque amorce et l'amplification du gène *speB* était réalisée sur un iCycler Thermal Cycler (Bio-Rad, Marnes-la-Coquette, France) selon les étapes suivantes : dénaturation de l'ADN et activation de la polymérisation pendant 15 minutes à 95°C ; 35 cycles de 15 secondes à 95°C, 30 secondes à 56°C, 30 secondes à 72°C puis une phase finale d'extension de 10 minutes à 72°C. Toutes les PCR incluaient un témoin positif (ADN extrait d'une souche de SGA ATCC 12344) et un témoin négatif (eau distillée stérile). L'ADN était ensuite révélé

après migration sur gel d'agarose 2% et la PCR était considérée comme positive si l'on mettait en évidence le fragment de 240 paires de base correspondant au gène *speB*.

Dans des études préliminaires réalisées au laboratoire de bactériologie, il a été montré que cette technique de PCR SGA était capable de mettre en évidence l'ensemble des différents géotypes *emm* de SGA à partir d'une collection microbiologique contenant 250 souches différentes de SGA [75]. Il a également été montré que cette technique ne conduisait à aucune réaction croisée avec différentes espèces de la flore oro-pharyngée (streptocoques  $\beta$ -hémolytiques des groupes B, C et G, *Streptococcus pneumoniae*, streptocoques  $\alpha$ -hémolytiques, *Enterococcus faecalis*, *Staphylococcus aureus*, staphylocoques coagulase-négative, *Micrococcus* sp., *Neisseria* sp., *Moraxella catarrhalis* et *Haemophilus influenzae*).

#### **2.4 Identification de *S. aureus* et *S. milleri***

Tous les prélèvements de gorge ont également été testés afin de rechercher la présence de *S. aureus* en ensemencant une boîte d'agar hypertonique au mannitol (Becton Dickinson, Le-Pont-de-Claix, France) par une goutte de liquide extrait de l'écouvillon humidifié (cf supra). La présence de *S. aureus* était ensuite confirmée par des tests enzymatiques spécifiques (production de catalase et de coagulase). En cas de colonie d'un diamètre <0,5 mm (colonie dite « minute ») donnant des réactions positives au test au latex spécifique du groupe A [76], *S. milleri* était recherché par tests biochimiques sur galerie (API Rapid ID32 Strep, Biomérieux, Lyon, France). En cas de culture positive à *S. aureus* ou à *S. milleri*, l'inoculum bactérien était quantifié comme faible (<50 colonies par boîte de Pétri) ou fort ( $\geq$ 50 colonies par boîte de Pétri). Toutes les analyses faites au laboratoire de microbiologie étaient réalisées à l'aveugle des données cliniques et du résultat du TDR.

## 2.5 Etude cas-témoins nichée dans l'étude ACTIV TDR SGA

Nous avons étudié les différents facteurs susceptibles d'expliquer les résultats FP du TDR en utilisant un schéma d'étude cas-témoins. Les cas et les témoins ont été recrutés dans cette étude ancillaire de manière consécutive parmi les patients de l'étude princeps jusqu'à obtention du nombre de sujets nécessaires. Les FP étaient les patients ayant un TDR positif et une culture (standard ou enrichie) négative pour le SGA. Les témoins vrais-positifs étaient les patients ayant un TDR positif et une culture positive pour le SGA et les témoins vrais-négatifs étaient les patients ayant un TDR négatif et une culture négative pour le SGA. En se basant sur un taux attendu de positivité de PCR SGA chez les FP de 60% et de 20% chez les témoins vrais-négatifs et une fréquence attendue de *S. aureus* de 60% chez les FP et de 30% chez les témoins vrais-positifs [77], un échantillon de 47 FP, 52 témoins vrais-négatifs et 94 témoins vrais-positifs était nécessaire (en prenant  $\alpha=0,05$ ,  $\beta=0,10$ ).

## 2.6 Analyses statistiques

Premièrement, nous avons décrit les patients de l'étude princeps ainsi que la prévalence du SGA et la sensibilité et spécificité du TDR. Deuxièmement, nous avons vérifié la comparabilité des cas et des témoins en ce qui concerne les caractéristiques démographiques de base en comparant l'âge par un test  $H$  de Kruskal-Wallis et le pourcentage de filles par un test du  $\chi^2$ . Troisièmement, nous avons comparé le taux de PCR SGA positives, de cultures positives à *S. aureus* et l'importance de l'inoculum de *S. aureus* entre les cas et les témoins par des tests du  $\chi^2$ . Quatrièmement, nous avons comparé le délai moyen entre prélèvement de gorge et mise en culture au laboratoire

entre cas et témoins par un test *t* de Student. Les intervalles de confiance pour les proportions ont été construits en utilisant la méthode binomiale exacte\*.

### 3. Résultats

Entre le 1er octobre 2010 et le 31 mai 2011, 678 enfants avec pharyngite ont été inclus dans l'étude princeps. Un enfant ayant un TDR ininterprétable et un patient chez qui le prélèvement de gorge a été perdu ont été exclus secondairement et l'analyse a donc porté sur 676 enfants dont 313 filles (46%), ayant une moyenne d'âge (écart-type) de 6,1 (2,5) ans. La prévalence du SGA dans l'étude princeps était de 41,4% (IC95 37,7–45,2%), la sensibilité du TDR était de 92,5% (88,8–95,3%) et la spécificité du TDR était de 88,4% (84,8–91,4%), en prenant le résultat de la culture standard comme test de référence\*.

Au total, nous avons échantillonné 41 FP, 58 témoins vrais-négatifs et 116 témoins vrais-positifs pour cette étude ancillaire. Nous n'avons pas mis en évidence de différence significative d'âge entre les cas et les témoins (médiane de 5,8 ans contre 5,5 ans contre 5,7 ans ;  $p=0,70$ ), ni de sex-ratio (59% de filles contre 45% contre 48% ;  $p=0,37$ ).

Toutes les cultures après enrichissement sur bouillon TGY sont restées négatives pour le SGA. La PCR SGA était positive chez 76% des FP (31/41 [60–88%]) contre 2% chez les témoins vrais-négatifs (1/58 [0–9%]) ;  $p<0,001$  ; Tableau 1).

*S. aureus* a été retrouvé chez 73% des FP (30/41 [57–86%]) contre 14% chez les témoins vrais-positifs (16/116 [8–21%]) ;  $p<0,001$ ). Au total, 61% des FP (25/41 [45–76%]) étaient positifs en PCR SGA et positifs pour la culture de *S. aureus*. Chez les patients ayant une culture positive pour *S. aureus*, les inocula forts de *S. aureus* étaient plus fréquents chez les FP que chez les témoins vrais positifs (63% [44–80%] contre

12,5% [2–38%],  $p=0,002$ ). Aucune colonie de *S. milleri* n'a été identifiée. Nous n'avons pas mis en évidence de différence significative dans le délai moyen entre prélèvement et mise en culture entre les FP et les contrôles vrais positifs (2,3 contre 2,6 jours,  $p=0,22$ ).

#### **4. Discussion**

Les TDR sont utilisés largement en pratique clinique afin d'identifier les pharyngites à SGA et tenter de réduire les prescriptions inutiles d'antibiotiques. Cette étude est la 1<sup>ère</sup> à explorer de manière systématique les différentes hypothèses retrouvées dans la littérature pour tenter d'expliquer les résultats FP du TDR. Au total, 76% des FP du TDR avaient une PCR SGA positive et ces enfants étaient donc réellement infectés ou colonisés par le SGA.

Nous avons montré que 61% des FP avaient à la fois une PCR SGA positive et une culture *S. aureus* positive, la plupart du temps avec un inoculum fort. Cela suggère que les FP pourraient être dus en grande partie à des enfants infectés ou colonisés à SGA chez qui le SGA n'a pu être retrouvé en culture en raison d'une co-infection à *S. aureus*. Nos résultats suggèrent également un effet dose entre l'inoculum de *S. aureus* et l'inhibition de la pousse du SGA en culture. Le SGA et *S. aureus* sont des germes faisant partie du microbiome nasopharyngé et ils interagissent de manière complexe entre eux ainsi qu'avec de nombreuses autres espèces bactériennes commensales [78]. D'un point de vue physiopathologique, l'effet inhibiteur de *S. aureus* sur la croissance du SGA en culture pourrait être expliqué par une interférence bactérienne\* impliquant différents mécanismes tels que la production de bactériocines ou d'autres substances bactériolytiques ou encore des mécanismes de compétition vis-à-vis des nutriments présents dans le milieu de culture [72].

Au total, 6 FP avaient une PCR SGA positive, montrant la présence de SGA non viables mais une culture négative pour *S. aureus*. Ces 6 FP pourraient tout simplement être des sujets infectés ou colonisés à SGA mais chez qui un très faible nombre de cellules de SGA viables étaient présentes sur le prélèvement de gorge, en quantité insuffisante pour pouvoir pousser sur les milieux de culture habituels. Cependant, cette hypothèse nous paraît peu probable car il serait attendu que de tels prélèvements de gorge donnent des résultats positifs en culture après passage en milieu d'enrichissement TGY, or tous les prélèvements qui ont étéensemencés après enrichissement sont restés négatifs pour le SGA. De plus, les faibles inocula de SGA semblent plus propices à donner des résultats faux-négatifs du TDR que des résultats FP car le seuil de détection du TDR est assez élevé, de l'ordre de  $10^5$  à  $10^7$  unités formant colonies in vitro [41, 65].

Parmi les 41 FP que nous avons pu analyser, 10 avaient une PCR SGA négative. Ces résultats pourraient toujours être expliqués par l'hypothèse de SGA non viables en raison de la présence d'autres germes ayant une action inhibitrice sur le SGA. En effet, d'autres protocoles de PCR pourraient nous avoir permis d'avoir une technique de PCR encore plus sensible. Par exemple, nous aurions pu utiliser de la mutanolysine au lieu d'utiliser du lysozyme dans l'étape de dégradation de la paroi cellulaire précédant l'extraction d'ADN ou encore utiliser une technique de PCR en temps réel plutôt qu'une technique de PCR conventionnelle. Autrement, ces 10 cas de FP avec PCR négative pourraient être réellement des FP chez qui le SGA n'était pas présent. Une autre hypothèse pourrait être alors celle de réactions croisées du TDR avec d'autres espèces bactériennes, comme cela a déjà été montré par d'autres auteurs [53, 68].

Contrairement à ce que nous attendions, nous n'avons pas mis en évidence de souche de *S. milleri* dans cette étude. Toutefois, cette absence de *S. milleri* reste cohérente avec les résultats d'une étude du même type dans laquelle une seule souche

de *S. milleri* portant un antigène du groupe A avait été identifiée sur 77 patients adultes ayant une pharyngite avec un résultat FP du TDR [53]. Une troisième hypothèse pouvant expliquer ces résultats serait une simple discordance entre les deux écouvillons constituant le double prélèvement de gorge. En effet, il est possible que le médecin ait prélevé la plus grande partie de l'inoculum bactérien avec le premier écouvillon, permettant ainsi de positiver le TDR, mais que l'inoculum restant sur le deuxième écouvillon soit insuffisant pour une détection en culture ou en PCR.

Dans cette étude, la sensibilité du TDR était plus élevée que celle habituellement observée (92,5% vs. 85%) et la spécificité du TDR était plus basse que celle habituellement observée (88,4% vs. 95%) [21, 44]. Une hypothèse pouvant expliquer cette différence est celle d'un biais lié aux attentes du lecteur du TDR (*expectation bias*, *clinical review bias*, *context bias*) [79, 80]. En effet, les résultats de la première année d'étude (2009-2010) avaient été transmis aux médecins investigateurs en ce qui concerne la prévalence du SGA, de l'ordre de 36%. Devant cette prévalence située dans la fourchette haute de ce qui est habituellement rencontré [8], on peut supposer que les investigateurs aient tenté d'optimiser la mise en évidence du SGA et la sensibilité du TDR au cours de la deuxième année d'étude (2010-2011) et qu'ils aient pour cela interprété les TDR comme étant positifs de manière plus fréquente et systématique. Cela a pu entraîner une surestimation de la sensibilité et une sous-estimation de la spécificité du TDR.

Nous avons utilisé une nouvelle technique de PCR spécifique du SGA afin de chercher à expliquer les FP du TDR. Dans notre série de patients, la PCR SGA était positive chez 32% des patients ayant une culture négative pour le SGA (32/99), ce qui pourrait mettre en question le statut de test de référence de la culture pour le diagnostic de pharyngite à SGA. La PCR semble avoir plusieurs avantages par rapport à la culture :

1/ une sensibilité plus élevée, 2/ l'identification des souches de SGA non  $\beta$ -hémolytiques, 3/ l'absence de réaction croisée avec d'autres espèces bactériennes [81]. Toutefois, l'utilisation de la PCR est pour l'instant plutôt destinée à des fins de recherche en épidémiologie ou en infectiologie fondamentale plutôt qu'en pratique clinique courante car sa réalisation nécessite un personnel et un équipement hautement spécialisés et parce que son coût est plus élevé que celui de la culture [82, 83].

Notre étude a plusieurs limites. Tout d'abord, notre approche basée sur la PCR ne nous a pas permis de faire la distinction entre des SGA non viables et des variants de SGA non  $\beta$ -hémolytiques. Le SGA produit 2 types d'hémolysines : la streptolysine S, qui est à l'origine de la majeure partie de l'hémolyse observée sur les géloses au sang [84], et la streptolysine O. Il y a eu à ce jour très peu de cas rapportés dans la littérature de souches de SGA non  $\beta$ -hémolytiques pathogènes chez l'homme [85]. Récemment, 3 cas de FP du TDR associés à la présence de SGA non hémolytiques ou  $\alpha$ -hémolytiques ont été rapportés [69, 70]. Dans notre étude, nous n'avons pas exploré de manière systématique toutes les colonies non  $\beta$ -hémolytiques présentes sur les boîtes de Pétri car cela aurait nécessité un travail supplémentaire pour les microbiologistes qui ne nous paraissait pas justifié au regard de la faible fréquence attendue de ce phénomène et de leur signification clinique incertaine [53]. Toutefois, afin de maximiser l'identification du SGA, les cultures de prélèvement de gorge ont été réalisées en milieu strictement anaérobie et dépourvu de CO<sub>2</sub>, soit des conditions de culture permettant d'augmenter l'expression de la streptolysine O et de majorer ainsi la réaction d'hémolyse chez les souches de SGA déficientes en streptolysine S [14]. Ainsi, nous avons considéré que les résultats de PCR SGA positifs étaient dus à des SGA non viables plutôt qu'à des variants non  $\beta$ -hémolytiques du SGA.

Deuxièmement, nous ne disposons pas d'argument biologique direct pour soutenir notre hypothèse d'interférence bactérienne entre *S. aureus* et le SGA, tels que la mise en évidence de bactériocines ou de substances bactériolytiques dans le milieu de culture [86]. D'autres études complémentaires au laboratoire de bactériologie seraient nécessaires pour confirmer cette hypothèse. Cependant, notre approche basée sur un schéma cas-témoins nichée dans une cohorte et la mise en évidence d'un effet-dose en ce qui concerne l'inoculum de *S. aureus* nous paraissent être suffisantes pour explorer cette hypothèse.

En conclusion, cette étude nous a permis de montrer que la majorité des résultats FP du TDR pourrait être associée à de réelles infections ou colonisations à SGA qui ne peuvent être mises en évidence grâce aux techniques classiques de culture en raison d'une co-infection avec d'autres germes tels que *S. aureus*. Ces résultats minimisent le poids clinique des FP dans les stratégies basées sur l'utilisation du TDR.

## 5. Tableaux et figures

Tableau 1. Etude bactériologique des faux-positifs du TDR

	Cas Faux-positifs (n=41)	Témoins		<i>p</i>
		Vrais-négatifs (n=58)	Vrais-positifs (n=116)	
Résultats des tests				
TDR	+	-	+	
Culture standard pour le SGA	-	-	+	
PCR SGA positive	31/41 (76)	1/58 (2)	n/a	<0.001
<i>Staphylococcus aureus</i>	30/41 (73)	n/a	16/116 (14)	<0.001
avec inoculum faible	11/30 (37)	n/a	14/16 (87,5)	0.002
avec inoculum fort	19/30 (63)	n/a	2/16 (12,5)	

Les données présentées sont des nombre/total (%), sauf mention contraire.

Abréviations: SGA, streptocoque du groupe A ; +, positif ; -, négatif ; n/a: non applicable.

Le test exact de Fisher a été utilisé à la place du test du  $\chi^2$  en cas d'effectif faible ( $n < 5$ ).

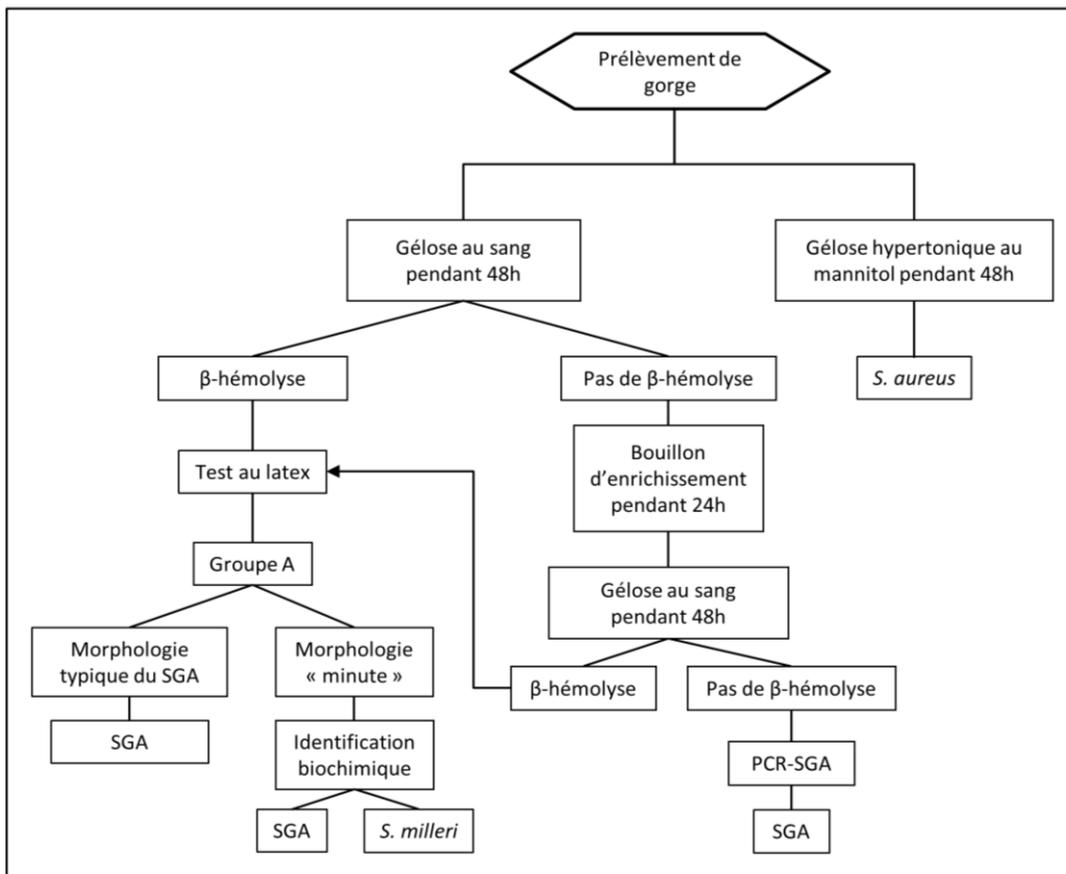


Figure 1. Analyses complémentaires de bactériologie pour cette étude ancillaire

# CHAPITRE 4

## Revue systématique et étude de validation externe des règles de décision clinique pour le diagnostic des pharyngites chez l'enfant

Jérémie Cohen, Robert Cohen, Corinne Levy, Franck Thollot, Mohamed Benani,

Philippe Bidet, Martin Chalumeau

*Canadian Medical Association Journal (en révision )*

## ***1. Introduction***

Le test de référence pour le diagnostic des pharyngites chez l'enfant est la mise en culture d'un prélèvement de gorge sur une gélose au sang au laboratoire de microbiologie [87] mais son résultat est obtenu au bout de 48 heures alors que le clinicien a besoin de prendre une décision quant à l'indication d'un traitement antibiotique à la fin de la consultation. Les alternatives sont les règles de décision clinique (RDC) et les tests de diagnostic rapide (TDR). Nous avons étudié dans les **Chapitres 1 à 3** le TDR et notamment les facteurs susceptibles de faire varier ses performances diagnostiques. Nous avons vu dans l'introduction qu'une RDC est un outil d'aide à la décision combinant des éléments de l'anamnèse de l'examen clinique fournissant au clinicien une conduite à tenir à l'échelle du patient.

La situation du diagnostic étiologique des pharyngites à SGA est une situation propice au développement d'une RDC puisqu'aucun élément de l'anamnèse ou de l'examen clinique ne permet de faire à lui seul la distinction entre pharyngite virale ou à SGA [12, 88, 89]. Des RDC combinant des éléments de l'anamnèse et de l'examen clinique ont donc été développées afin d'aider les cliniciens à décider quels patients devraient être soumis à un test diagnostique (TDR ou mise en culture) et/ou traités par antibiotiques [16, 90, 91]. Les RDC définissent habituellement 3 groupes de risque de SGA : les patients à bas risque reçoivent uniquement un traitement symptomatique, les patients à risque intermédiaire sont soumis à un test diagnostique complémentaire (culture ou TDR) et reçoivent un traitement antibiotique en cas de test positif, les patients à haut risque reçoivent un traitement antibiotique empirique.

Il existe actuellement une controverse internationale quant à la place des RDC [31]. Ces variations entre recommandations peuvent être en partie expliquées par le fait que la plupart des RDC n'ont pas été suffisamment validées ni directement comparées

---

entre elles. La validation externe permet d'évaluer dans quelle mesure un outil de prédiction clinique est utilisable dans un contexte et dans une population différents de ceux de sa construction (transportabilité) [92-97]. La comparaison directe des performances diagnostiques dans une même population est le moyen le plus fiable pour comparer différents modèles de prédiction et identifier celui ayant les meilleures performances diagnostiques [97, 98].

Quatre revues de la littérature ont tenté d'évaluer l'intérêt de l'examen clinique et/ou des RDC pour le diagnostic des pharyngites chez l'enfant [12, 40, 88, 89]. Toutes ont conclu qu'aucun élément de l'anamnèse ou de l'examen clinique n'était suffisamment discriminant pour distinguer à lui seul entre pharyngites virales ou à SGA et trois d'entre elles ont conclu que certaines RDC étaient suffisamment performantes pour pouvoir être utilisées. Ces revues de la littérature avaient des faiblesses méthodologiques limitant la portée de leurs résultats : 1/ revues non systématiques ou stratégie de recherche bibliographique insuffisante [12, 40, 89], 2/ absence d'analyse séparée entre adultes et enfants [12], 3/ absence d'évaluation des RDC combinant des éléments cliniques avec un TDR [40, 88, 89], 4/ comparaison indirecte de performances diagnostiques [12, 88, 89] (Tableau 1). Le Marechal *et al.* ont conclu leur revue systématique en soulignant qu'une validation externe avec comparaison directe dans une même population pédiatrique était nécessaire avant de pouvoir conclure quant à la place souhaitable pour les RDC en pratique clinique [89].

Les objectifs de l'étude étaient donc de réaliser une revue systématique de la littérature concernant les RDC pour le diagnostic des pharyngites à SGA chez l'enfant puis de réaliser une validation externe avec comparaison directe des différentes RDC identifiées.

## 2. Méthodes

### 2.1 Revue systématique de la littérature

Toutes les études ayant rapporté de construction ou de mise à jour de RDC ayant pour affection cible les pharyngites à SGA chez l'enfant étaient éligibles. Une RDC a été définie comme un outil d'aide à la décision destiné aux cliniciens et incluant au moins deux éléments de l'anamnèse de l'examen physique ou d'examens complémentaires de routine et fournissant au clinicien soit une probabilité d'évènement soit une conduite à tenir à l'échelle du patient [92-95]. Toutes les études basées sur plus de 50 patients (seuil choisi arbitrairement a priori) étaient éligibles s'il était possible d'extraire des données pédiatriques. Pour les RDC ayant été mises à jour par les auteurs ayant eux-mêmes construit la RDC, les deux études étaient éligibles pour la revue systématique mais seules les mises à jour des RDC ont été validées.

Nous avons réalisé une recherche bibliographique sur MEDLINE de janvier 1975 à août 2012 en utilisant la stratégie décrite en Annexe 1 (page 122). Nous avons utilisé un filtre dédié aux RDC [99] combiné à un filtre destiné à identifier les études pédiatriques [100], tout en nous limitant à la langue anglaise. La stratégie de recherche a ensuite été adaptée à la base de données EMBASE. Nous avons complété cette recherche en utilisant Science Citation Index et Google Scholar, en utilisant la fonction "related articles" de PubMed (20 premières références) et en examinant les références des études et des revues pertinentes sur le sujet.

Deux auteurs (JC, MC) ont évalué de manière indépendante les études pour inclusion. Nous nous sommes d'abord basés sur un examen du titre et du résumé. Ensuite, les études potentiellement éligibles ont été récupérées en texte complet puis évaluées. Les différences de jugement ont été résolues par discussion jusqu'au consensus.

Pour chaque étude, nous avons extrait un ensemble de variables spécifié a priori en utilisant une fiche de recueil standardisé. Les données extraites concernaient le type d'étude et son design (site d'étude, pays d'étude, objectifs de l'étude, étude prospective ou rétrospective, étude monocentrique ou multicentrique), les critères d'inclusion et d'exclusion (critères cliniques de sélection, limites d'âge), la taille de l'échantillon de la population d'étude, la prévalence du SGA dans l'ensemble de la population et dans chaque niveau de risque défini par la RDC, le type de RDC (score, arbre décisionnel, liste d'items, équation de régression), les différents prédicteurs cliniques présents dans la RDC et leur nombre et enfin la conduite à tenir en suivant la RDC en fonction du niveau de risque du patient.

Il n'existe pas d'outil validé pour l'évaluation de la qualité méthodologique d'études de construction de RDC. Nous avons combiné différents éléments liés aux RDC [92-95], dérivés de QUADAS-2 (*Quality Assessment of Diagnostic Accuracy Studies*) [101, 102] et de recommandations pour la construction et la validation de modèles de prédiction clinique [98, 103-109] (Annexe 2, page 123). Nous avons coté chaque item comme présent, absent, pas clair ou non applicable. Pour les études renvoyant le lecteur à de précédents articles, nous avons recherché les articles correspondant et nous nous sommes basés sur leur contenu pour évaluer la qualité méthodologique des études. Nous avons également classé les RDC selon leur niveau de preuve tel que défini par McGinn *et al.* [94], en nous basant sur les résultats de précédentes revues de la littérature [12, 40, 88, 89]. L'extraction des données, l'évaluation de la qualité méthodologique et le classement selon le niveau de preuve ont été faits par un premier auteur (JC) en utilisant une fiche de recueil de données standardisé puis vérifiées par un deuxième auteur (MC). Les différences ont été réglées par discussion jusqu'à obtention d'un consensus.

## **2.2 Etude de validation externe**

### *2.2.1 Cohorte de validation*

Cette étude de validation est basée sur les données de l'année 2010-2011 du groupe d'enfants avec pharyngite (voir Sources de données, page 37). Cette étude avait été prévue a priori et les données cliniques qui paraissaient susceptibles d'être nécessaires à la validation de RDC avaient été identifiées sur la base de revues de la littérature sur le sujet [12, 40].

### *2.2.2 Stratégie d'analyse*

La validité externe des RDC a été évaluée en termes de calibration et de discrimination. La calibration reflète dans quelle mesure les probabilités d'évènement observées dans la cohorte de validation sont en accord avec les probabilités prédites par la RDC. Pour chaque niveau de risque défini par la RDC, nous avons comparé les probabilités de SGA observées dans la cohorte aux probabilités prédites par la RDC. Pour tenir compte de la différence de risque de base de SGA entre la population de construction et notre population de validation, nous avons calculé des probabilités de SGA prédites après ajustement sur la prévalence. Les données des études princeps ont été réanalysées et les rapports de vraisemblance ont été calculés pour chaque niveau de risque de la RDC, puis ces rapports de vraisemblance ont été appliqués à la prévalence de SGA dans la cohorte de validation pour calculer les probabilités prédites ajustées [95, 110]. La calibration a ensuite été évaluée à l'aide de graphes de calibration [98].

La discrimination reflète dans quelle mesure une RDC est capable de classer correctement les sujets comme malades ou non malades. Pour cette étude de validation, la RDC a été considérée comme "positive" si la conclusion de la RDC pour le patient

était que le SGA était hautement probable et qu'un traitement antibiotique était indiqué, et comme "négative" si la conclusion de la RDC était que le SGA était peu probable et qu'un traitement antibiotique n'était pas nécessaire. Dans le cas où les auteurs avaient construit un score sans décrire de seuil de décision, nous avons validé deux RDC à partir du score :

- règle A : RDC basée uniquement sur la clinique avec un seul seuil de décision (bas risque, pas d'antibiotique ; haut risque, antibiotique d'emblée) ;
- règle B : RDC basée sur deux seuils définissant trois groupes de risque avec utilisation d'un TDR dans le groupe à risque intermédiaire (bas risque, pas d'antibiotique ; risque intermédiaire, traitement antibiotique selon le résultat du TDR ; haut risque, antibiotique d'emblée).

Les seuils décisionnels ont été choisis en accord avec la littérature sur la base de précédentes études de validation. Chaque RDC a été appliquée à chaque patient et son résultat a été comparé au résultat de la culture de prélèvement de gorge (test de référence). Cela nous a permis de calculer la sensibilité, la spécificité, l'aire sous courbe ROC et le nombre de TDR utilisés pour chaque RDC. Nous avons représenté graphiquement les performances des RDC dans un espace ROC.

### **3. Résultats**

#### **3.1 Revue systématique**

##### *3.1.1 Sélection et caractéristiques des études*

A partir de 1106 articles identifiés, 30 ont été retenus pour examen du texte complet et 13 ont finalement été inclus pour analyse (Tableau 2 page 111, Figure 1 page 118). Ces études ont été publiées entre 1975 [111] et 2010 [112] et provenaient de pays et de contextes d'études variés (Tableau 2, page 111). Deux études mélangeaient des enfants

et des adultes mais permettaient d'extraire les données pédiatriques [16, 111] et les autres études étaient strictement pédiatriques. L'ensemble de ces 13 études comprenait au total 7893 patients, avec une médiane de 271 patients par étude (range 94–3877) et la prévalence médiane du SGA était de 29% (range 13%–48%).

### 3.1.2 Description des règles

Nous avons identifié 16 RDC différentes, car trois études rapportaient la construction de deux RDC à partir du même jeu de données [112-114] (Tableau 3, page 112). Six RDC proposaient clairement de réaliser un test diagnostique complémentaire (culture ou TDR) chez certains patients sélectionnés en fonction du niveau du score [16, 111, 112, 114-116].

Lorsque nous avons évalué le nombre de prédicteurs de chaque RDC, nous avons dû subdiviser certains prédicteurs. Par exemple, la variable « absence de signes d'atteinte des voies aériennes supérieures » (*no upper respiratory symptoms*) du score de Wald nécessitait l'évaluation des 3 variables cliniques « rhinorrhée, toux ou conjonctivite » (*rhinorrhea, cough or conjunctivitis*) [91]. Au total, nous avons identifié 26 prédicteurs différents. Le nombre médian de prédicteurs par RDC était de neuf (range 3–14). Les 5 signes les plus fréquemment présents dans les différentes RDC étaient la présence d'adénopathies cervicales sensibles (12/16 RDC), la fièvre (9/16), l'âge (9/16), une rhinite (9/16) et la présence d'un érythème du pharynx (7/16). Seul le score de Breese contenait une variable biologique (compte des globules blancs) mais les auteurs avaient anticipé qu'une prise de sang ne soit pas faite en routine et avaient attribué trois points pour cette variable en cas de donnée manquante [90].

### 3.1.3 Qualité méthodologique et niveau de preuve

La qualité méthodologique des études retenues est résumée dans le tableau 4 (page 113). D'une manière générale, la qualité méthodologique des études retenue était plutôt bonne. Toutes les études étaient prospectives et la majorité (8/13 soit 62%) décrivaient de manière précise le test de référence [16, 91, 112, 114-118]. Les critères d'inclusion étaient décrits dans 10/13 études (77%) [16, 91, 111-115, 117-119]. Une seule des 13 études retenues a mentionné de manière explicite que la culture avait été interprétée à l'aveugle des prédicteurs [91]. Parmi les 8/13 (62%) études ayant rapporté des valeurs de sensibilité et spécificité [16, 90, 112-114, 117-119], seules 3/8 (38%) présentaient les intervalles de confiance correspondant [112, 114, 118].

Seul McIsaac a présenté les résultats d'une réelle étape de validation interne (partage de l'échantillon en 2 parties de taille égale, la moitié des données servant à la construction de la règle et l'autre à la validation) [16]. Forsyth *et al.* ont présenté une étape qu'ils ont appelé « validation interne » dans laquelle les performances obtenues par un groupe de médecin étaient comparées aux performances d'un autre groupe de médecins, mais cela ne correspond à aucune des méthodes de validation interne recommandées [111]. Plusieurs études de validation externe ont été publiées pour le score de Breese [40, 90, 120, 121], de Wald [20, 40] et le score de McIsaac [18-20, 22, 24, 40, 65, 122] (niveau de preuve 2) ; les RDC d'Attia et Smeesters ont été mises à jour et soumises à validation externe une fois [112, 118] (niveau de preuve 3), et les autres RDC n'ont jamais été validées (niveau de preuve 4). Aucune des RDC identifiées n'a été soumise à étude d'impact (niveau 1) [123].

Les méthodes statistiques employées pour construire les modèles de prédiction étaient insuffisantes. Cinq des 13 (38%) études ont décrit le processus de sélection des variables finalement retenues [16, 115, 116, 118, 119] et seulement 6/13 (46%) études

ont rapporté avoir utilisé des analyses multivariées [16, 114-116, 118, 119]. Les variables continues telles que l'âge ou la température de l'enfant étaient toujours traitées comme des variables discrètes dans les analyses. Seule une étude a précisé avoir exploré l'existence d'éventuelles interactions\* entre les prédicteurs [116]. Aucune étude n'a présenté de résultat concernant la discrimination (aire sous courbe ROC) ou la calibration du modèle (test d'adéquation d'Hosmer-Lemeshow).

## 3.2 Validation externe

### 3.2.1 Cohorte de validation

Comme cela a déjà été mentionné au **Chapitre 3**, un total de 678 enfants a été inclus sur la période d'étude et deux patients ont été secondairement exclus en raison d'un TDR ininterprétable et d'un prélèvement de gorge perdu (voir Chapitre 3, page 87). De plus, 145 (21%) patients ayant des données cliniques manquantes ont été exclus de cette étude et l'analyse a donc finalement porté sur 531 enfants (243 filles [46%]), d'un âge moyen (DS) de 6,2 (2,5) ans. La prévalence du SGA dans la population retenue pour analyse était de 40,3% (intervalle de confiance à 95% [IC95%] 36,1%–44,6%), la sensibilité du TDR était de 93,9% (89,8%–96,7%) et la spécificité de 88,0% (83,9%–91,4%). L'âge moyen des patients exclus pour données manquantes était de 5,7 ans ( $p=0,02$  en comparaison aux patients retenus pour analyse, test  $t$  de Student) et la prévalence du SGA chez ces enfants était de 45,5% [37,3%–53,7%], ( $p=0,26$  en comparaison aux patients retenus, test du  $\chi^2$ ). Les caractéristiques cliniques des patients retenus pour analyse n'étaient pas différentes de celles des patients exclus pour données manquantes sauf en ce qui concerne la présence d'une inflammation des amygdales (gonflement ou exsudat) (75% contre 84%,  $p=0,02$ ), la présence d'une éruption

scarlatiniforme (9% contre 15%,  $p=0,03$ ) et la présence de diarrhée (4% contre 10%,  $p=0,004$  ; Tableau 5, page 114).

### 3.2.2 Performances des RDC

Certaines RDC ont dû être modifiées afin de pouvoir être validées car les auteurs n'avaient pas défini les prédicteurs cliniques de manière précise ou parce que nos variables cliniques ne correspondaient pas exactement à celles décrites par les auteurs (Tableau 6, page 115).

Pour toutes les RDC, la fréquence du SGA augmentait avec le score ou le niveau de risque défini par la règle. La calibration de trois RDC n'a pu être évaluée car les auteurs ne précisaient pas la fréquence du SGA en fonction des différents niveaux de risque [113, 116]. Pour les RDC évaluables, la calibration était globalement faible (Tableau 7 page 116, Figure 2 page 119). Aucune RDC n'était en mesure d'identifier des patients pouvant être considérés comme réellement à bas risque de SGA (risque de SGA  $\leq 12\%$ , fréquence moyenne du portage pharyngé asymptomatique de SGA chez l'enfant) [88] et seule la règle Attia II était capable d'identifier des patients à haut risque (risque de SGA  $\geq 85\%$ , valeur prédictive positive minimale du TDR) [88], mais ce groupe de patients ne comptait que 33/531 (6%) de l'ensemble de la population de validation. Pour les autres RDC, la plus haute probabilité de SGA observée était de 53,1% (chez les patients ayant un score de Wald égal à 6).

La sensibilité des RDC allait de 56,5% (IC95 49,6%–63,3% pour la règle Breese A) à 96,3% (92,8%–98,4% pour la règle de McIsaac et 92,8%–98,4% pour la règle Sahin II) ; la spécificité allait de 5,0% (2,9%–8,1% pour la règle Sahin II) à 87,4% (83,2%–90,8% pour la règle Attia II) ; et l'aire sous courbe ROC allait de 0,50 (0,47–0,53 pour Sahin I) à 0,89 (0,86–0,91 pour Attia II) (Tableau 8 page 117, Figure 3 page

120). Les RDC basées uniquement sur la clinique, avec un seul seuil et suggérant une action binaire, avaient toutes un faible pouvoir discriminant (aire sous courbe ROC de 0,50 à 0,56). A l'inverse, les RDC basées sur deux seuils de décision et suggérant de réaliser un TDR dans le groupe à risque intermédiaire, avaient globalement un bon pouvoir discriminant (aire sous courbe ROC systématiquement  $\geq 0,70$ , sauf la règle Smeesters IIB). Par ailleurs, l'aire sous courbe ROC était parfaitement corrélée avec le nombre de TDR réalisés en suivant la règle (test de corrélation sur les rangs de Spearman,  $\rho=1$ ,  $p<0,0001$  ; Figure 4 page 121, Tableau 8 page 117).

#### **4. Discussion**

De nombreuses RDC ont été construites pour le diagnostic des pharyngites chez l'enfant en comparaison aux autres pathologies de l'âge pédiatrique [124]. La plupart des RDC que nous avons identifiées sont des scores basés sur la combinaison des mêmes signes cliniques classiquement associés à la présence du SGA [12]. La qualité méthodologique globale des études de construction de RDC était bonne, mais la méthodologie statistique sur laquelle reposait la construction des RDC était d'une manière générale plutôt faible. Dans notre étude prospective de validation externe, la plupart des RDC étaient incapables d'identifier des patients à bas risque ou à haut risque de SGA. Les RDC basées uniquement sur des critères cliniques n'étaient clairement pas en mesure de classer correctement les patients tandis que les RDC intégrant la réalisation d'un TDR chez les patients à risque intermédiaire avaient des performances diagnostiques directement proportionnelles au nombre de TDR réalisés en suivant la RDC.

Notre cohorte de validation est constituée d'enfants vus en pédiatrie ambulatoire de ville et est donc représentative de la population cible des RDC que nous avons validées. Cependant, ce recrutement pourrait expliquer en partie les faibles

performances des RDC dans notre population car ces RDC ont été principalement construites sur des patients vus aux urgences pédiatriques (Tableau 2, page 111). Il est en effet possible que les patients vus aux urgences aient une présentation clinique plus bruyante que les patients vus en ville, et que cela ait pu conduire à surestimer l'amplitude des associations entre prédicteurs cliniques et présence du SGA en culture (biais de sélection\*). En terme de validité externe, la prévalence du SGA dans notre population de validation était de 40,3%, ce qui est proche des résultats d'une récente méta-analyse (37% [IC95% 32%–43%]) [8].

Une des limites de cette étude de validation est le nombre important de données manquantes [145/678 (21%)], ce qui nous a conduits à exclure secondairement une part importante de notre population d'étude. Nous n'avons pas utilisé de technique d'imputation multiple [98, 108] car nous n'avons pas mis en évidence de différence significative dans la plupart des caractéristiques cliniques ni dans la prévalence du SGA entre les patients secondairement exclus et les patients retenus pour analyse (Tableau 5, page 114).

Nous avons montré que la calibration des RDC était faible pour la plupart des RDC. Ce manque de transportabilité pourrait être expliqué par des différences de population d'étude entre l'étude de construction et cette étude de validation [98] ainsi que par le manque de reproductibilité des signes et symptômes entrant dans les RDC (faible accord inter-juges) [125-127]. Nos résultats sont cohérents avec ceux d'une étude rétrospective de validation à large échelle du score de McIsaac réalisée chez 65000 enfants Nord-Américains âgés de 3 à 14 ans qui a montré que, pour une prévalence de base de SGA de 27%, la probabilité de SGA était de 17% en cas de score de McIsaac égal à 1 et de 23% en cas de score égal à 2 [122], ce qui semble supérieur à ce que les cliniciens attendent d'un RDC afin d'exclure la présence du SGA.

Nous avons montré que les RDC basées uniquement sur des combinaisons de signes cliniques avaient de faibles performances alors que les RDC combinant des signes cliniques avec un TDR chez les patients à risque intermédiaire avaient des performances plus acceptables. En explorant les raisons susceptibles d'expliquer les variations de performances au sein de ce groupe de règles combinant clinique et TDR, nous avons mis en évidence une corrélation parfaite entre aire sous courbe ROC et nombre de TDR ( $p < 0,0001$ ). Ainsi, plus la RDC suggère de réaliser de TDR, plus la RDC est discriminante. Cela peut être expliqué par le fait que les TDR ont des rapports de vraisemblance positifs et négatifs très élevés en comparaison aux signes cliniques (de l'ordre de 20 contre 1 à 4 pour les rapports de vraisemblance positifs et de 0,15 contre 0,40 à 1 pour les rapports de vraisemblance négatifs) [44, 88, 89].

Les prochaines études de construction de RDC devraient se baser sur les recommandations méthodologiques les plus récentes afin de tenter d'optimiser à la fois la calibration et la discrimination des règles proposées. Par ailleurs, il persiste actuellement une incertitude quant aux attentes des cliniciens face aux RDC en ce qui concerne leurs performances (sensibilité et spécificité) et le nombre acceptable de TDR réalisés.

L'examen clinique seul, même rationalisé sous forme de RDC a des performances insuffisantes pour le diagnostic des pharyngites chez l'enfant [88]. La seule RDC combinant clinique et TDR qui semblait intéressante en termes de performances diagnostiques nécessitait l'utilisation d'un TDR chez quasiment tous les patients (88%, règle Attia II). Les TDR étant peu invasifs et peu coûteux, il semble peu intéressant pour les cliniciens d'utiliser une RDC impliquant le calcul d'un score pour réduire le nombre de TDR utilisés de seulement 12%.

En conclusion, les résultats de cette étude renforcent la pertinence clinique de la stratégie non-sélective actuellement recommandée en France qui est de réaliser un TDR chez tous les enfants avec pharyngite.

## 5. Tableaux et figures

Tableau 1. Résumé des revues de la littérature précédentes sur les performances des RDC pour le diagnostic des pharyngites à SGA chez l'enfant

Auteur (année de publication)	Type de revue	Revue		Nb de RDC	Conclusions	Limites de la revue
		Signes et symptômes	RDC			
Ebell (2000)	RN	oui	oui	9	Aucun élément clinique ne permet à lui seul de confirmer ou d'exclure le diagnostic de pharyngite à SGA. Certaines RDC sont suffisamment performantes pour être utilisées en pratique clinique.	Pas de revue systématique de la littérature. Pas d'évaluation de la qualité des études. Pas d'analyse séparée entre adultes et enfants. Comparaison indirecte de performances.
Fischer Walker (2006)	RN, VE	non	oui	7	Certaines RDC sont suffisamment performantes pour être utilisées en pratique clinique.	Pas de revue systématique de la littérature. Recherche bibliographique limitée à une seule base de données (MEDLINE). La validation externe était limitée aux RDC binaires, sans utilisation de test complémentaire (culture ou TDR).
Shaikh (2012)	RS-MA	oui	oui	5	Les éléments de l'anamnèse et de l'examen clinique, pris seuls ou combinés sous la forme de RDC ne peuvent être utilisés pour confirmer ou exclure le diagnostic de pharyngite à SGA.	Etude limitée aux RDC ayant passé une validation externe. Méta-analyse limitée aux RDC binaires, sans utilisation de test complémentaire (culture ou TDR). Comparaison indirecte de performances.
Le Marechal (2013)	RS-MA	oui	oui	7	Aucun élément clinique ne permet à lui seul de confirmer ou d'exclure le diagnostic de pharyngite à SGA. Certaines RDC sont suffisamment performantes pour être utilisées en pratique clinique.	Stratégie de recherche bibliographique probablement insuffisante (combinaison de 3 termes MeSH). Méta-analyse limitée aux RDC binaires, sans utilisation de test complémentaire (culture ou TDR). Comparaison indirecte de performances.

Abréviations: RDC, règle de décision clinique ; TDR, test de diagnostic rapide ; RN, revue narrative ; VE, validation externe ; RS-MA, revue systématique avec méta-analyse

Tableau 2. Description des études incluses dans la revue systématique

Auteur [Reference]	Période	Pays	Site d'étude	Objectifs	Type d'étude	Test de référence	Critères d'inclusion			
							Signes cliniques et symptômes	Age	n	SGA+ (%)
Forsyth [111]	non rapportée	EUA	non précisé	CONST, VI	PRO, MULTI	culture	Mal de gorge évoluant depuis moins d'une semaine	≤14 ans	234	30,3
Breese [90]	1973–1975	EUA	cabinets de ville	MAJ, VE	PRO, MULTI	culture	Infection respiratoire haute aiguë	pas clair	3877	pas clair
Fujikawa [128]	1982–1983	JPN	Service de pédiatrie (hôpital)	CONST	PRO, MONO	culture	Enfants ayant des signes et symptômes d'infection respiratoire haute et suspects d'être infectés par une bactérie	pas clair	271	39,9
Edmond [116]	1994–1995	AUS	Service d'urgences (hôpital)	CONST	PRO, MONO	culture	Enfants ayant des signes de pharyngite lors de l'examen clinique	pas clair	271	21
McIsaac [16]	1995–1997	CAN	Centre de médecine de famille (hôpital)	CONST, VI	PRO, MONO	culture	Nouvelle infection respiratoire haute avec un diagnostic enregistré de pharyngite ou angine	≥3 ans	94*	36,2*
Wald [91]	1990–1992	EUA	Service d'urgences (hôpital) et clinique de consultation sans rendez-vous	CONST	PRO, MONO	culture	Episode aigu de mal de gorge (moins de 48 heures d'évolution) et notion de fièvre (tactile) ou température documentée > 38,3°C en prise buccale lors de la consultation ou dans les 24 heures précédentes	2–16 ans	365	48
Attia I [115]	1996	EUA	Service d'urgences (hôpital)	CONST	PRO, MONO	culture	Enfants ayant une pharyngite et ayant eu un prélèvement de gorge pour mise en culture en raison d'une suspicion clinique de SGA	6 mois–18 ans	297	29
Attia II [118]	1999–2000	EUA	Service d'urgences (hôpital) et deux cliniques de pédiatrie ambulatoire	MAJ, VE	PRO, MULTI	culture	Patients ayant des signes et symptômes de pharyngite aiguë	1–18 ans	587	37
Nandi [117]	non rapportée	IND	visites à domicile	CONST	PRO, n/a	culture	Enfants se plaignant de mal de gorge	5–15 ans	536	13,4
Sahin [113]	2001–2002	TUR	service de pédiatrie ambulatoire (hôpital)	CONST	PRO, MULTI	culture	Enfants ayant des signes et symptômes d'infection respiratoire haute	0–17 ans	245	26,5
Steinhoff [119]	1992–1993	EGY	service de pédiatrie ambulatoire (hôpital)	CONST	PRO, MONO	culture	Enfants se présentant avec une notion de mal de gorge et avec un érythème pharyngé	2–13 ans	410	24,6
Smeesters I [114]	2004	BRE	3 services d'urgences (hôpital)	CONST	PRO, MULTI	culture	Enfants ayant des signes et symptômes de pharyngite aiguë	0–15 ans	220	26
Smeesters II <sup>§</sup> [112]	2007–2008	BRE	3 services d'urgences (hôpital) et une "unité médicale"	MAJ	PRO, MULTI	culture ou TDR	Enfants ayant des symptômes de pharyngite aiguë	0–15 ans	356	33

Abréviations: MAJ, mise à jour ; CONST, construction ; VE, validation externe ; VI, validation interne ; PRO, prospective ; MONO, étude monocentrique ; MULTI, étude multicentrique ; TDR, test de diagnostic rapide ; SGA, streptocoque du groupe A ; Pays: EUA, Etats-Unis d'Amérique ; JPN, Japon ; AUS, Australie ; CAN, Canada ; IND, Inde ; TUR, Turquie, EGY, Egypte ; BRE, Brésil.

\*521 patients au total (3–14 ans, n=94, SGA+=36,2% ; 15–44 ans, n=346, SGA+=10,7% ; ≥45 ans, n=79, SGA+=1,3% ; donnée manquante pour l'âge, n=2).

<sup>§</sup> Le premier auteur est Joachim mais il s'agit explicitement d'une mise à jour de la règle de Smeesters *et al.*, réalisée par l'équipe de Smeesters.

Tableau 3. Description des RDC

Auteur	Type de RDC	Prédicteurs cliniques	Nb de prédicteurs	Nb de groupes de risque	Conduite à tenir suggérée
Forsyth	liste	température, adénopathies cervicales sensibles (1 or $\geq 2$ ), exsudat (1 point, $\geq 2$ points ou points confluents, exsudat purulent), syndrome mononucléosique (exsudat muqueux et adénopathies multiples de petite taille), syndrome grippal (fièvre et myalgies mais sans exsudat ni adénite)	5	3	cliniquement “non-SGA” : culture ; traitement symptomatique cliniquement “douteux” : culture ; traitement symptomatique cliniquement “SGA” : pas de culture ; pénicilline orale
Breese	score	Mois durant lequel le patient est vu, âge, compte des globules blancs, fièvre $\geq 38^\circ\text{C}$ , mal de gorge, toux, céphalées, anomalies du pharynx (rougeur, gonflement, exsudat, pétéchies, “lésions en anneau”, ulcérations), adénopathies cervicales (très augmentées de volume non sensibles ou palpables et sensibles)	14	n/a	pas de conduite à tenir claire
Fujikawa	score	fièvre $>37,5^\circ\text{C}$ , mal de gorge, nausées ou vomissements, anorexie, absence de toux et/ou rhinorrhée, pharynx (rougeur intense ou pétéchies), gonflement de ganglions cervicaux, exsudat amygdalien (points blancs), rash cutané (scarlatiniforme, érythème ou urticaire), langue framboisée ou papilles marquées	13	n/a	pas de conduite à tenir claire
Edmond	arbre décisionnel	âge, rash scarlatiniforme, pharyngo-tonsillite (pharynx érythémateux et gonflé, œdème des amygdales), adénopathies cervicales sensibles	5	3	risque $<20\%$ : traitement symptomatique risque $20\%–60\%$ : culture et/ou traitement par pénicilline risque $>60\%$ : traitement par pénicilline
McIsaac	score	température $>38^\circ\text{C}$ , pas de toux, adénopathie cervicale antérieure sensible, gonflement ou exsudat des amygdales, âge	6	3	0-1 : pas de culture ni d’antibiotique 2-3 : culture ; antibiotique en cas de culture positive 4 : culture ou antibiotique
Wald	score	âge, saison, fièvre $\geq 38,3^\circ\text{C}$ , adénopathie (ganglion lymphatique cervical $\geq 1$ cm ou sensible à la palpation), pharyngite (érythème, gonflement ou exsudat du pharynx ou des amygdales), absence de signe d’atteinte des voies aériennes supérieures (rhinorrhée, toux ou conjonctivite)	11	n/a	pas de conduite à tenir claire
Attia I	diagramme de décision	rash scarlatiniforme, gonflement des amygdales (modéré à sévère), adénopathies cervicales sensibles (modéré à sévère), absence de coryza (modéré à sévère)	5	3	risque $<15\%$ : pas d’antibiotique, pas de test risque $15\%–65\%$ : test (TDR et/ou culture) risque $65\%–95\%$ : antibiotique, pas de test
Attia II	score	Idem	5	n/a	pas de conduit à tenir claire
Nandi	score	âge, saisons, fièvre, érythème du pharynx (modéré en cas de rougeur ; sévère en cas de pétéchies ou de saignement après écouvillonnage), taille des amygdales, exsudat pharyngé, adénopathie +/- sensible, douleur de gorge	11	2	pas de conduite à tenir claire
Sahin I*	liste	mal à la gorge, érythème pharyngé, exsudat pharyngé	3	2	prescrire de la pénicilline si $\geq 2$ signes
Sahin II*	liste	mal à la gorge, érythème pharyngé, fièvre	3	2	prescrire de la pénicilline si $\geq 2$ signes
Steinhoff	score	adénopathies cervicales ( $\geq 1,5$ cm), pas de rhinite, pas de rash cutané	3	2	prescrire de la pénicilline si $\geq 2$ signes
Smeesters IA**	score	âge, signes viraux (conjonctivite, coryza, toux, diarrhée, exanthème d’allure virale), signes bactériens (adénopathie cervicale sensible, céphalées, pétéchies sur le palais, fièvre $>38,5$ , douleurs abdominales, début brutal [ $<12$ h])	12	2	$\geq 8$ : traitement symptomatique $<8$ : antibiotique
Smeesters IB**	score	Idem	12	3	$\geq 8$ : traitement symptomatique 5–7 : antibiotique en cas de culture positive $<5$ : antibiotique
Smeesters IIA**	score	âge, signes bactériens (adénopathie cervicale sensible, céphalées, pétéchies sur le palais, douleurs abdominales, début brutal [ $<12$ h]), signes viraux (conjonctivite, coryza, diarrhée)	9	2	score $\leq 2$ : traitement symptomatique score $\geq 3$ : traitement antibiotique
Smeesters IIB**	score	Idem	9	3	score $\leq 2$ : pas de TDR, traitement symptomatique score 3 : antibiotique en cas de TDR positif score $\geq 4$ : pas de TDR, traitement antibiotique

n/a, non applicable par ce que les auteurs n’ont pas défini clairement de seuil décisionnel et de conduit à tenir. \*Les auteurs ont proposé 2 RDC (désignées Sahin I et Sahin II par la suite). \*\*Les auteurs ont proposé 2 RDC : une avec un seuil décisionnel, 2 groupes de risque et pas de test (règle A) et une avec 2 seuils décisionnels, 3 groupes de risque et un test (culture ou TDR) dans le groupe à risque intermédiaire (règle B).

Tableau 4. Qualité méthodologique et niveau de preuve des RDC

	Forsyth	Breese	Fujikawa	Edmond	McIsaac	Wald	Attia I	Attia II	Nandi	Sahin	Steinhoff	Smeesters I	Smeesters II	no. (%)
Variable prédite														
Définition	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	13/13 (100)
Description du test de référence*	non	non	non	oui	oui	oui	oui	oui	oui	non	non	oui	oui	8/13 (62)
Evaluation à l'aveugle	?	?	?	?	?	oui	?	?	?	?	?	?	?	1/13 (8)
Prédicteurs														
Définition	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	13/13 (100)
Catégorisation des prédicteurs continus	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	13/13 (100)
Mesure de reproductibilité	non	non	non	non	non	non	non	oui	non	non	non	non	non	1/13 (8)
Evaluation à l'aveugle	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	?	11/13 (85)
Sélection des patients														
Description des critères de sélection**	oui	non	non	non	oui	oui	oui	oui	oui	oui	oui	oui	oui	10/13 (77)
Description des caractéristiques des patients (âge et sexe)	non	non	non	oui	oui	non	oui	oui	non	oui	oui	oui	oui	8/13 (62)
Site d'étude														
Pays d'étude	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	13/13 (100)
Dates d'inclusion	non	non	oui	oui	oui	oui	oui	oui	non	oui	oui [129]	oui	oui	10/13 (77)
Description du site d'étude	non	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	oui	12/13 (92)
Résultats de la règle														
Sensibilité et spécificité	non	oui	non	non	oui	non	non	oui	oui	oui	oui	oui	oui	8/13 (62)
Intervalles de confiance	n/a	non	n/a	n/a	non	n/a	n/a	oui	non	non	non	oui	oui	3/8 (38)
Pertinence clinique														
Facilité d'utilisation***	oui	non	non	oui	oui	non	oui	oui	non	oui	oui	non	oui	8/13 (62)
Probabilité de SGA	oui	oui	oui	non <sup>§</sup>	oui	oui	non <sup>§§</sup>	oui	oui	non	oui	oui	oui	10/13 (77)
Conduite à tenir	oui	non	non	oui	oui	non	oui	non	non	oui	oui	oui	oui	8/13 (62)
Validation interne	oui	non	non	non	oui	non	non	non	non	non	non	non	non	2/13 (15)
Validation externe	non	oui [40, 90, 120, 121]	non	non	oui [18-20, 22, 24, 40, 65]	oui [20, 40]	non	oui [118]	non	non	non	oui [112]	non	5/13 (38)
Etude prospective d'impact	non	non	non	non	non	non	non	non	non	non	non	non	non	0/13 (0)
Niveau de preuve	3	2	4	4	2	2	4	3	4	4	4	3	4	
Analyse statistique														
Description de la méthode de sélection des prédicteurs	non	non	non	oui	oui	non	oui	oui	non	non	oui	non	non	5/13 (38)
Analyse multivariée	non	non	non	oui	oui	non	oui	oui	non	non	oui	oui	?	6/13 (46)
Régression logistique	n/a	n/a	n/a	oui	oui	n/a	oui	oui	n/a	n/a	oui	?	n/a	5/6 (83)
Puissance suffisante <sup>§</sup>	n/a	n/a	n/a	oui	oui	n/a	oui	oui	n/a	n/a	oui	non	n/a	5/6 (83)
Tests d'interaction	n/a	n/a	n/a	oui	non	n/a	non	non	n/a	n/a	non	non	n/a	1/6 (17)
Discrimination du modèle <sup>†</sup>	n/a	n/a	n/a	non	non	n/a	non	non	n/a	n/a	non	non	n/a	0/6 (0)
Calibration du modèle <sup>††</sup>	n/a	n/a	n/a	non	non	n/a	non	non	n/a	n/a	non	non	n/a	0/6 (0)

\* ≥2 critères parmi milieu, durée, atmosphère, température de mise en culture et technique de confirmation du SGA ; \*\* au moins les signes et symptômes et les limites d'âge pour inclusion ; \*\*\* ≤9 prédicteurs cliniques (médiane du nombre de prédicteurs) et/ou existence d'un calculateur automatisé de risque ; § ≥10 évènements/prédicteur dans la règle.

† Aire sous la courbe ROC du modèle ; †† Graphe de calibration ou test d'Hosmer-Lemeshow par exemple.

§ Les auteurs ont défini 3 niveaux de risque (<20%, 20%–60%, >60%) mais n'ont pas présenté la fréquence précise du SGA dans ces groupes. §§ Les auteurs ont défini 3 niveaux de risque (<15%, 15%–65%, 65%–95%) mais n'ont pas présenté la fréquence précise du SGA dans ces groupes.

Tableau 5. Caractéristiques cliniques des patients de la population de validation (N=678)

Signes cliniques et symptômes	Patients retenus pour analyse (n=531)	Patients exclus (données manquantes) (n=145)	<i>p</i> *
Erythème du pharynx	472/531 (89)	127/141 (90)	0,688
Mal à la gorge	451/531 (85)	91/113 (81)	0,244
Fièvre >38°C	399/531 (75)	120/145 (83)	0,054
Gonflement ou exsudat des amygdales	397/531 (75)	108/128 (84)	<b>0,021</b>
Adénopathies cervicales ≥1 cm et/ou sensible	218/531 (41)	63/141 (45)	0,438
Toux	205/531 (39)	57/133 (43)	0,370
Rhinorrhée	201/531 (38)	51/142 (36)	0,672
Douleurs abdominales	170/531 (32)	54/136 (40)	0,090
Céphalées	157/531 (30)	29/131 (22)	0,090
Pétéchies du palais	97/531 (18)	19/103 (18)	0,966
Rash scarlatiniforme	47/531 (9)	22/145 (15)	<b>0,026</b>
Diarrhée	20/531 (4)	13/132 (10)	<b>0,004</b>
Exanthème viral	14/531 (3)	3/136 (2)	1,000
Conjonctivite	13/531 (2)	6/140 (4)	0,254

\*Test du  $\chi^2$  ou test exact de Fisher.

Tableau 6. Validation externe des RDC : adaptations par rapport aux études originales

Règle	Adaptations pour cette étude de validation
Forsyth	Le degré de positivité de critères cliniques tels que l'adénite (nombre d'adénopathies) ou l'exsudat (exsudat punctiforme ou exsudat confluent) et la présence de myalgies n'ont pas été mesurés dans la cohorte de validation et cette RDC n'a donc pas pu être validée.
Breese	<i>Abnormal cervical glands (very enlarged without tenderness or if they are palpable and tender)</i> : ce critère a été considéré comme positif en cas d'adénopathie $\geq 2$ cm et/ou sensible. L'étude originale ne précisait pas de conduit à tenir claire en fonction du score. Les règles suivantes ont été adaptées de l'article original : Breese A [40, 120, 130] : <ul style="list-style-type: none"> <li>▪ score <math>&lt; 30</math> : pas de test, traitement symptomatique</li> <li>▪ score <math>\geq 30</math> : pas de test, traitement antibiotique</li> </ul> Breese B : <ul style="list-style-type: none"> <li>▪ score <math>\leq 25</math> : pas de test, traitement symptomatique</li> <li>▪ score 26–31 : antibiotique en cas de TDR positif</li> <li>▪ score <math>\geq 32</math> : pas de test, traitement antibiotique</li> </ul>
Fujikawa	<i>Strawberry tongue or marked papillae</i> : cette variable n'a pas été mesurée dans la cohorte de validation et cette RDC n'a donc pas pu être validée.
Edmond	<i>Tender cervical nodes</i> : ce critère a été considéré comme positif en cas d'adénopathie $\geq 1$ cm et/ou sensible, comme dans le score de Wald [91].
McIsaac	<i>Tender anterior cervical adenopathy</i> : ce critère a été considéré comme positif en cas d'adénopathie $\geq 1$ cm et/ou sensible, comme dans le score de Wald [91]. La règle suivante a été adaptée de l'article original [18, 19] : <ul style="list-style-type: none"> <li>▪ score 0–1 : pas de test, traitement symptomatique</li> <li>▪ score 2–3 : antibiotique en cas de TDR positif</li> <li>▪ score <math>\geq 4</math> : pas de test, traitement antibiotique</li> </ul>
Wald	L'étude originale ne précisait pas de conduit à tenir claire en fonction du score. Les règles suivantes ont été adaptées de l'article original : Wald A [40] : <ul style="list-style-type: none"> <li>▪ score 0–2 : pas de test, traitement symptomatique</li> <li>▪ score <math>\geq 3</math> : pas de test, traitement antibiotique</li> </ul> Wald B : <ul style="list-style-type: none"> <li>▪ score 0–1 : pas de test, traitement symptomatique</li> <li>▪ score 2–4 : antibiotique en cas de TDR positif</li> <li>▪ score <math>\geq 5</math> : pas de test, traitement antibiotique</li> </ul>
Attia II	<i>Moderate to severe cervical lymphadenopathy</i> : ce critère a été considéré comme positif en cas d'adénopathie $\geq 1$ cm et/ou sensible, comme dans le score de Wald [91]. Cette RDC a été initialement publiée sous la forme d'un diagramme de décision [115], puis la règle a été ensuite mise à jour et publiée sous la forme d'un score. Cependant, dans cette mise à jour, les auteurs ne précisait pas de conduit à tenir claire en fonction du score. Les règles suivantes ont été adaptées de l'article de mise à jour de la règle [118] : <ul style="list-style-type: none"> <li>▪ score 0 : pas de test, traitement symptomatique</li> <li>▪ score 1–3 : antibiotique en cas de TDR positif</li> <li>▪ score <math>\geq 4</math> : pas de test, traitement antibiotique</li> </ul>
Nandi	<i>Size of tonsils</i> : le score prévoyait l'attribution de 3 points en cas d'amygdales modérément augmentées de volume et 5 points en cas d'amygdales franchement augmentées de taille. Afin de simplifier cet item, 4 points ont été attribués en cas d'augmentation de taille des amygdales. <i>Lymphadenopathy</i> : le score original prévoyait l'attribution de 4 points en cas d'adénopathies non sensible et de 5 points en cas d'adénopathies sensibles. Afin de simplifier cet item, 4 points ont été attribués en cas d'adénopathie $\geq 1$ cm et/ou sensible, comme dans le score de Wald [91].
Sahin I/II	Aucune adaptation n'a été nécessaire
Steinhoff	<i>Enlarged cervical nodes (<math>&gt;1.5</math> cm)</i> : ce critère a été considéré comme positif en cas d'adénopathie 1 cm. <i>No rash</i> : ce critère a été interprété comme "pas de rash d'allure virale" car dans l'article original de Steinhoff <i>et al.</i> cette variable était associée à la présence du SGA, alors que le rash scarlatiniforme est habituellement considéré comme le signe le plus spécifique de pharyngite à SGA [88].
Smeesters II	<i>Tender cervical node</i> : ce critère a été considéré comme positif en cas d'adénopathie $\geq 1$ cm et/ou sensible, comme dans le score de Wald [91].

Tableau 7. Validation externe des RDC : calibration\* (N=531)

RDC	Nombre de patients (%)	Probabilité de SGA		
		Prédite par la règle (étude originale)	Prédite par la règle (après ajustement sur la prévalence)	Observée dans la population de validation
<b>Breese*</b>				
18–25	26 (5)	6,1	3,6	26,9
26–29	242 (46)	36,0	24,3	35,5
30–31	159 (30)	56,4	42,5	44,0
32–38	104 (20)	84,4	75,5	49,0
<b>Edmond</b>				
Risque <20%	168 (32)	n/a	n/a	28,6
Risque 20%–60%	249 (47)	n/a	n/a	42,2
Risque >60%	114 (21)	n/a	n/a	53,5
<b>McIsaac</b>				
-1 ou 0	0 (0)	-	-	-
1	10 (2)	12,5	14,5	30,0
2	77 (15)	21,0	24,0	32,5
3	162 (31)	29,2	32,9	39,5
4-5	282 (53)	51,3	55,6	43,3
<b>Wald</b>				
1	7 (1)	0	0,0	28,6
2	29 (5)	24	18,8	34,5
3	136 (26)	22	17,1	30,9
4	177 (33)	41	33,7	41,8
5	133 (25)	59	51,3	45,1
6	49 (9)	75	68,7	53,1
<b>Attia II</b>				
0	30 (6)	12	13,5	36,7
1–3	468 (88)	36	39,9	37,4
≥4	33 (6)	79	81,2	84,8
<b>Nandi</b>				
<15	82 (15)	1,6	6,6	35,4
≥15	449 (85)	48,5	80,4	41,2
<b>Sahin A</b>				
<2	79 (15)	n/a	n/a	39,2
≥2	452 (85)	n/a	n/a	40,5
<b>Sahin B</b>				
<2	24 (5)	n/a	n/a	33,3
≥2	507 (95)	n/a	n/a	40,6
<b>Steinhoff</b>				
<2	138 (26)	7	13,5	36,2
≥2	393 (74)	32	49,3	41,7
<b>Smeesters II</b>				
≤2	53 (10)	13	18,6	24,5
3	123 (23)	26	35,0	38,2
≥4	355 (67)	41	51,5	43,4

n/a, non applicable parce que les auteurs n'ont pas précisé la probabilité de SGA en fonction des niveaux de risque de la règle.

\*Données de l'étude de validation externe de Breese (publiée avec l'étude de construction du score) [90].

Tableau 8. Validation externe des RDC : discrimination\* (N=531)

Auteur	Sensibilité, % (IC95)	Spécificité, % (IC95)	ASC ROC (IC95)	Nb de TDR, n/N (%)
Règles avec un seul seuil et sans utilisation de TDR				
Breese A	56,5 (49,6–63,3)	55,2 (49,5–60,8)	0,56 (0,52–0,60)	n/a
Wald A	94,4 (90,4–97,1)	7,6 (4,9–11,1)	0,51 (0,49–0,53)	n/a
Nandi	86,4 (81,1–90,7)	16,7 (12,8–21,3)	0,52 (0,48–0,55)	n/a
Sahin I	85,5 (80,1–89,9)	15,1 (11,4–19,6)	0,50 (0,47–0,53)	n/a
Sahin II	96,3 (92,8–98,4)	5,0 (2,9–8,1)	0,51 (0,49–0,52)	n/a
Steinhoff	76,6 (70,4–82,1)	27,8 (22,9–33,0)	0,52 (0,48–0,56)	n/a
Smeesters IIA	93,9 (89,8–96,7)	12,6 (9,2–16,8)	0,53 (0,51–0,56)	n/a
Règles avec deux seuils et utilisation du TDR dans le groupe à risque intermédiaire				
Breese B	92,1 (87,6–95,3)	75,1 (69,9–79,7)	0,84 (0,81–0,87)	401/531 (76)
Wald B	94,9 (91,0–97,4)	61,8 (56,2–67,2)	0,78 (0,75–0,81)	342/531 (64)
Attia II	89,7 (84,8–93,4)	87,4 (83,2–90,8)	0,89 (0,86–0,91)	468/531 (88)
Edmond	75,2 (68,9–80,9)	76,7 (71,6–81,2)	0,76 (0,72–0,80)	249/531 (47)
McIsaac	96,3 (92,8–98,4)	44,2 (38,6–49,8)	0,70 (0,67–0,73)	239/531 (45)
Smeesters IIB	92,5 (88,1–95,7)	35,3 (30,1–40,9)	0,64 (0,61–0,67)	123/531 (23)

IC95, intervalle de confiance à 95% ; ASC ROC, aire sous la courbe ROC.

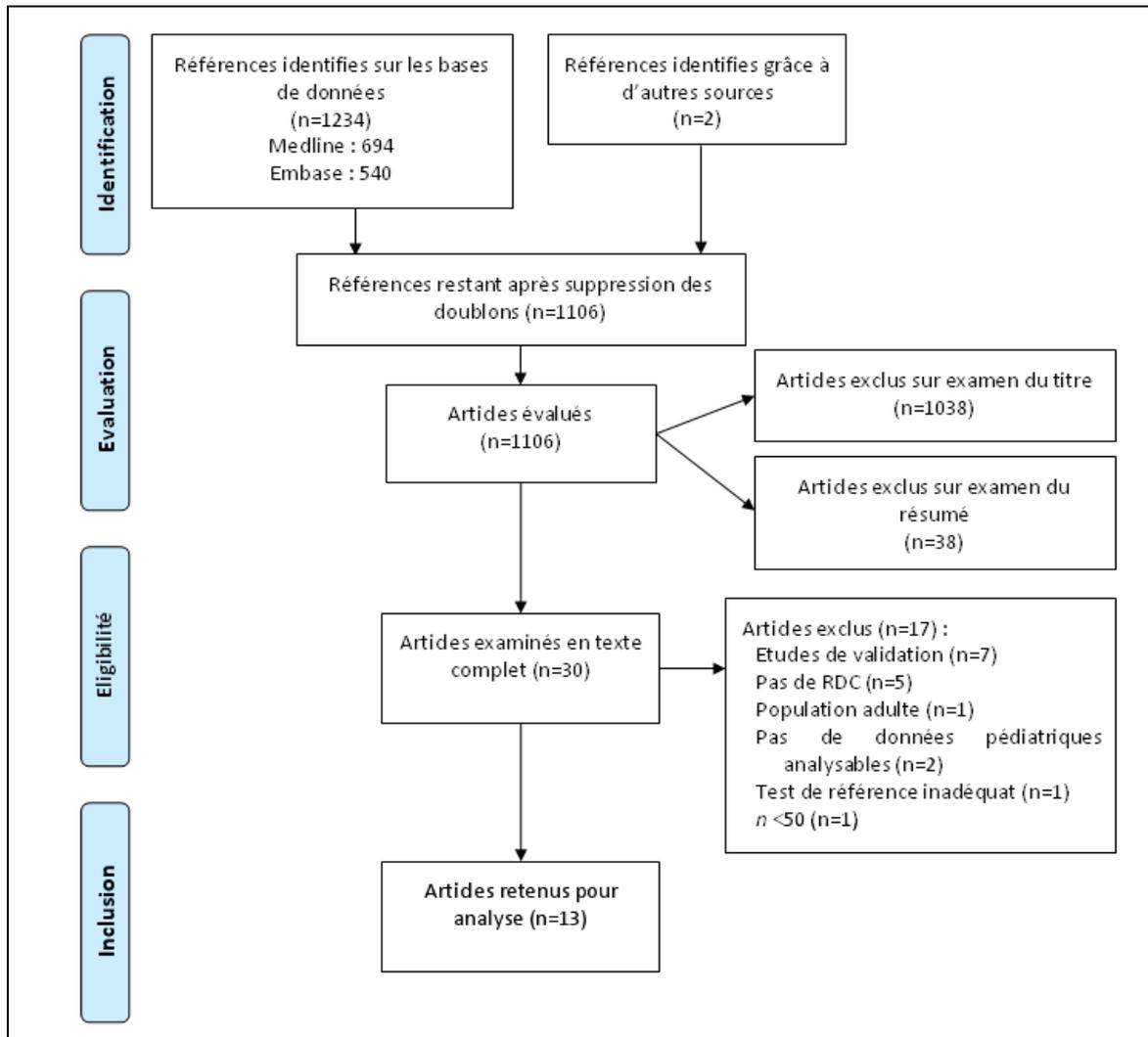


Figure 1. Flux des études dans la revue systématique\*

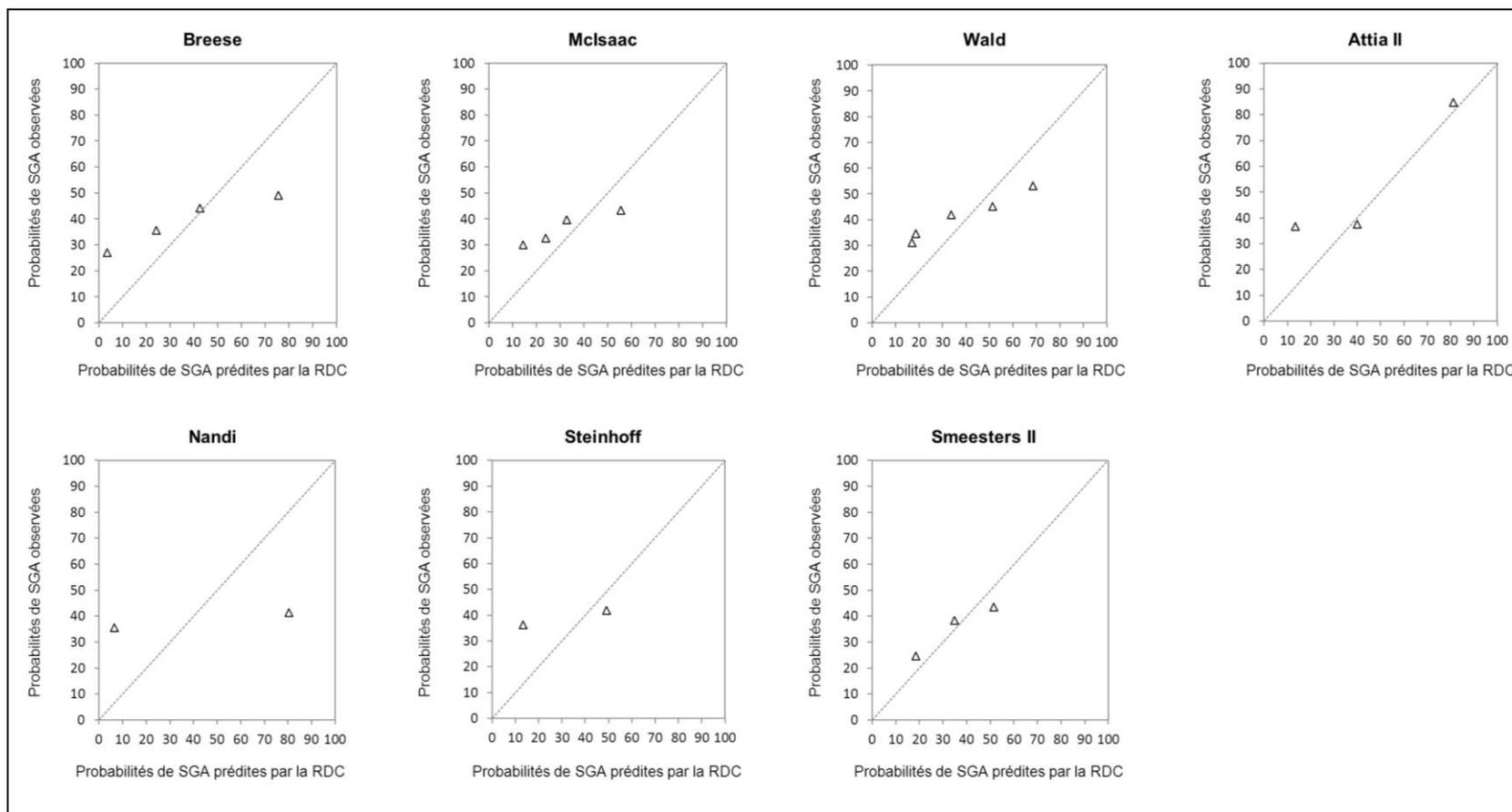


Figure 2. Validation externe des RDC : graphes de calibration. Confrontation des probabilités de SGA observées contre les probabilités de SGA prédites par la règle (ajustées pour la prévalence). La ligne diagonale correspond à une calibration parfaite

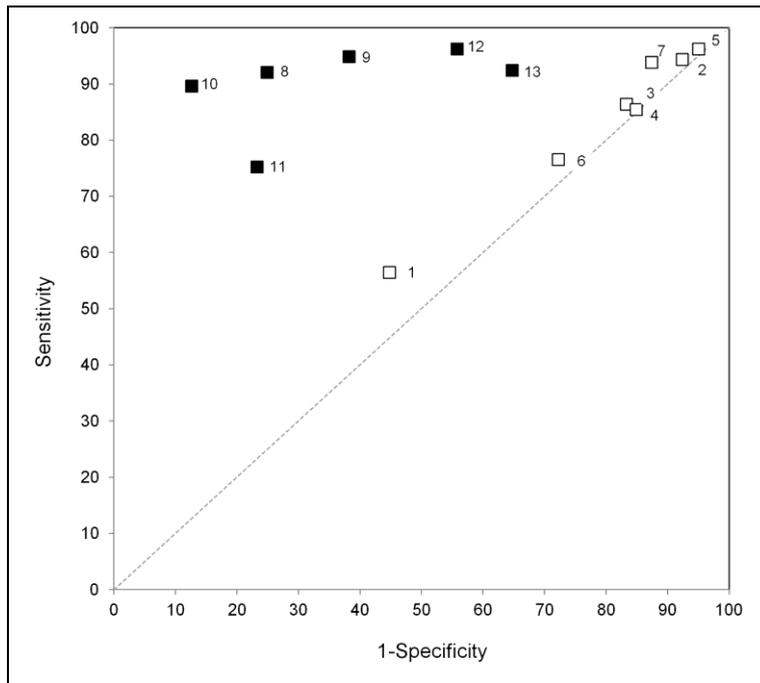


Figure 3. Validation externe des RDC : sensibilité et spécificité des RDC représentées dans l'espace ROC. Les carrés blancs représentent les RDC sans utilisation de TDR (1, Breese A ; 2, Wald A ; 3, Nandi ; 4, Sahin I ; 5, Sahin II ; 6, Steinhoff ; 7, Smeesters IIA). Les carrés noirs représentent les RDC avec utilisation de TDR dans le groupe à risque intermédiaire (8, Breese B ; 9, Wald B ; 10, Attia II ; 11, Edmond ; 12, McIsaac ; 13, Smeesters IIB). La ligne diagonale correspond à l'absence de pouvoir discriminant (aire sous courbe ROC=0,50).

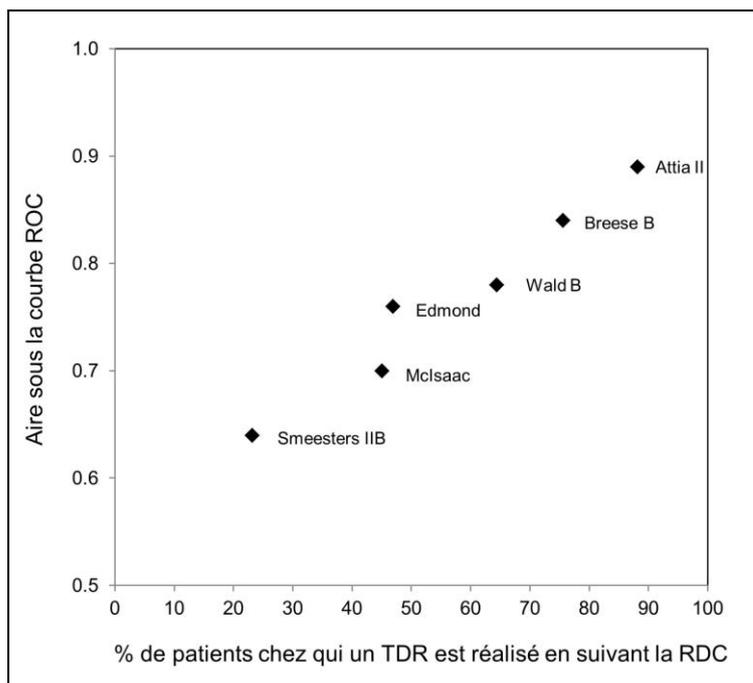


Figure 4. Aire sous courbe ROC de chaque RDC en fonction du nombre de TDR utilisés en suivant la règle (test de Spearman,  $\rho=1$ ,  $p<0,0001$ ).

## 6. Annexes

### Annexe 1. Stratégie de recherche bibliographique dans MEDLINE (PubMed)

((clinical[tiab] AND predict\*[tiab]) OR (clinical[tiab] AND model\*[tiab]) OR (clinical[tiab] AND score\* [tiab]) OR (decision [tiab] AND rule\*[tiab]) OR (derive\*[tiab] AND validat\*[tiab]) OR (diagnos\*[tiab] AND accura\*[tiab]) OR (diagnos\*[tiab] AND rule\*[tiab]) OR (diagnos\*[tiab] AND score\*[tiab]) OR (diagnos\*[tiab] AND value[tiab]) OR (predict\*[tiab] AND outcome\*[tiab]) OR (predict\*[tiab] AND rule\*[tiab]) OR(predict\*[tiab] AND score\*[tiab])OR (predict\*[tiab]AND validat\*[tiab])OR (predict\* [tiab] AND value\*[tiab]) OR (risk\*[tiab] AND assessment\*[tiab]) OR (risk[tiab] AND score\*[tiab]) OR (sensitivity [tiab] AND specificity [tiab]) OR (symptoms [tiab] AND signs [tiab]) OR (validat\*[tiab] AND decision\*[tiab]) OR (validat\*[tiab] AND rule\*[tiab]) OR (validat\*[tiab] AND score\*[tiab]) OR (predict\*[tiab] AND risk\*[tiab]))

**AND** (“pharyngitis/diagnosis”[Mesh] OR “pharyngitis/microbiology”[Mesh] OR “pharyngitis”[tiab] OR “tonsillitis”[Mesh] OR “tonsillopharyngitis”[tw] OR “pharyngotonsillitis”[tw] OR sore throat\*[tw] OR "streptococcal infections/diagnosis"[Mesh] OR "streptococcal infections/microbiology"[Mesh] OR group a beta hemolytic streptococc\*[tw] OR group a hemolytic streptococc\*[tw] OR group a beta haemolytic streptococc\*[tw] OR group a haemolytic streptococc\*[tw] OR “gabhs”[tw] OR "streptococcus pyogenes/isolation and purification"[MeSH] OR “streptococcus pyogenes”[tw] OR “s. pyogenes”[tw])

**AND** (“infant”[Mesh] OR infant\* OR infancy OR newborn\* OR baby\* OR babies OR neonat\* OR preterm\* OR prematur\* OR postmatur\* OR child[Mesh] OR child\* OR schoolchild\* OR school age\* OR preschool\* OR kid or kids OR toddler\* OR adolescent[Mesh] OR adoles\* OR teen\* OR boy\* OR girl\* OR minors[Mesh] OR minors\* OR puberty[Mesh] OR pubert\* OR pubescen\* OR prepubescen\* OR pediatrics[Mesh] OR paediatric\* OR paediatric\* OR peadiatric\* OR schools[Mesh] OR nursery school\* OR kindergar\* OR primary school\* OR secondary school\* OR elementary school\* OR high school\* OR highschool\*)

**AND** ("humans"[Mesh] AND english[Language])

**NOT** (pneumonia\*[tiab] OR pneumococc\*[tiab] OR influenza\*[ti] OR group b streptococc\*[tiab] OR “tonsillectomy”[ti] OR review[pt] OR case reports[pt] OR editorial[pt] OR comment[pt] OR letter[pt])

## Annexe 2. Critères utilisés pour l'évaluation de la qualité des études et des RDC

<b>Variable prédite</b>	
Définition	La variable prédite devait être la présence du SGA dans le pharynx.
Description du test de référence	Les auteurs devaient préciser au moins 2 des items suivant : milieu de culture, atmosphère d'incubation, durée d'incubation, température d'incubation et technique de confirmation du SGA.
Evaluation à l'aveugle	Le test de référence devait être interprété à l'aveugle des prédicteurs potentiels.
<b>Prédicteurs</b>	
Définition	Tous les prédicteurs prédictifs potentiels devaient avoir été décrits de manière suffisamment précise pour pouvoir être utilisé en pratique clinique de la même manière qu'elles ont été utilisées dans la construction de la règle.
Catégorisation des prédicteurs continus	La transformation des variables continues en variables catégorielles conduit à une perte d'information et devait être évitée.
Mesure de reproductibilité	L'évaluation des prédicteurs devrait donner le même résultat lors de différentes mesure par le même observateur ou lors d'une même mesure par différents observateurs. L'accord inter-juges est habituellement évalué par le coefficient kappa.
Evaluation à l'aveugle	Les prédicteurs devaient être mesurés à l'aveugle des résultats du test de référence pour éviter un biais de mesure.
<b>Sélection des patients</b>	
Description des critères de sélection	Des variations dans le spectre clinique de la maladie peuvent entraîner des variations de performances de la RDC, avec en général de meilleures performances diagnostiques chez les patients ayant une présentation clinique plus grave (effet de spectre). Les auteurs devaient avoir au moins détaillé les signes cliniques et les limites d'âge pour l'inclusion.
Description des caractéristiques des patients (âge et sexe)	Une même RDC peut avoir des performances différentes dans différents échantillons de patients en raison de variations des caractéristiques démographiques des patients. Cela est particulièrement important dans le cas des pharyngites puisque la prévalence du SGA varie avec l'âge et que les performances diagnostiques peuvent varier avec la prévalence de la maladie dans l'échantillon.
<b>Site d'étude</b>	
Pays d'étude	Le contexte spatio-temporel de l'étude devait avoir été rapporté car il est possible d'observer des variations de spectre de patients, de gravité de la maladie et de techniques de références en fonction de l'endroit et du moment où a lieu l'étude.
Dates d'inclusion	Idem
Description du site d'étude	Le site d'étude (cabinets de pédiatrie ambulatoire, services d'Urgences, consultation hospitalière) devait avoir été décrit car une même RDC peut avoir des performances diagnostiques différentes dans différents sites d'études en raison d'un recrutement différent et de pratiques cliniques différentes.
<b>Résultats de la règle</b>	
Sensibilité et spécificité	La sensibilité et la spécificité sont des mesures universellement connues de performances diagnostiques. Elles devaient avoir été présentées pour que le lecteur puisse estimer la fréquence des faux-négatifs (nombre de cas de SGA manqués) et de faux-positifs (antibiothérapies inutiles) auxquelles il peut s'attendre s'il utilise la RDC.
Intervalles de confiance	La sensibilité et la spécificité devaient avoir été présentées avec leurs intervalles de confiance respectifs (habituellement intervalle de confiance à 95%) afin que le lecteur puisse évaluer l'incertitude entourant les estimations ponctuelles de performances diagnostiques.
<b>Pertinence clinique</b>	
Facilité d'utilisation	La RDC devait être suffisamment facile à utiliser pour pouvoir être utilisée en pratique clinique de routine. Cela implique un nombre limité de variables et/ou la disponibilité d'un calculateur automatique de risque sur support informatique en cas de calcul complexe (par exemple équation de régression logistique). Ici nous avons arbitrairement considéré qu'une RDC était facile d'utilisation si elle se basait sur un nombre de prédicteurs inférieur ou égal au nombre médian de prédicteurs.

Probabilité de SGA	Le résultat d'une RDC à l'échelle d'un patient devait être soit une probabilité de risque de SGA individuelle soit le positionnement du patient dans un groupe à risque de SGA (faible, intermédiaire ou élevé par exemple).
Conduite à tenir	Après que la RDC ait fourni une estimation du risque, la règle suggère habituellement une conduite à tenir diagnostique et/ou thérapeutique. Dans le cas des pharyngites, la RDC peut suggérer la réalisation d'un test diagnostique complémentaire (mise en culture ou TDR) et/ou la mise en route d'une antibiothérapie.
<b>Validation</b>	
Validation interne	Une étape de validation interne est nécessaire pour évaluer si une règle est reproductible dans une population proche de celle ayant servi à la construction de la RDC.
Validation externe	La validation externe permet d'évaluer dans quelle mesure la RDC peut être appliquée à des patients et dans des contextes différents de ceux de l'étude de construction.
Etude prospective d'impact	L'impact de l'utilisation de la RDC en pratique clinique devait avoir été évalué prospectivement afin d'évaluer dans quelle mesure la RDC est susceptible de modifier les pratiques cliniques (nombre de tests prescrits, consommation d'antibiotiques) et d'améliorer la santé des patients.
Niveau de preuve	Les RDC peuvent être classées selon leur niveau de preuve en utilisant l'échelle développée par l'Evidence-Based Medicine Working Group.
<b>Analyse statistique</b>	
Description de la méthode de sélection des prédicteurs	La construction d'une RDC nécessite de sélectionner des variables prédictives parmi un ensemble plus large de prédicteurs potentiels. Un processus inapproprié de sélection des prédicteurs peut compromettre la transportabilité de la RDC en raison de du risque d' <i>overfitting</i> , qui correspond au fait que le modèle dépendrait plus de la structure des données que de véritables associations entre prédicteurs et SGA.
Analyse multivariée	Les auteurs devaient avoir décrit s'ils ont utilisé des analyses multivariées afin d'estimer l'effet indépendant des différents prédicteurs.
Régression logistique	La plupart des RDC ont pour objectif de prédire le résultat d'une variable binaire (présence ou absence du SGA dans notre cas). Dans de telles circonstances, le modèle logistique est particulièrement utilisé car il estime les mesures d'associations sous la forme d'odds ratios qui sont facilement compréhensibles par la plupart des lecteurs. D'autres approches peuvent être utilisées, tels que les arbres de régression (CART), l'analyse discriminante ou les réseaux neuronaux.
Puissance suffisante <sup>s</sup>	En régression logistique multivariée, le nombre de prédicteurs dans le modèle doit être adapté au nombre d'événements observés. Une règle classique consiste à retenir une variable dans le modèle pour 10 participants dans le groupe de plus petit effectif.
Tests d'interaction	La formulation du modèle logistique implique l'additivité des effets des prédicteurs. Si cette hypothèse n'est pas vérifiée, les résultats du modèle peuvent être faux. Une manière de vérifier que l'hypothèse d'additivité des effets est respectée est de tester les éventuelles interactions qui peuvent exister entre prédicteurs. Il existe une interaction entre la variable A et la variable B si l'effet de la variable A sur la variable prédite dépend du niveau de la variable B.
Discrimination du modèle	La discrimination est une mesure de performance du modèle logistique. La discrimination permet d'estimer dans quelle mesure le modèle est capable de classer correctement les patients en malades ou non malades. L'aire sous courbe ROC est une mesure classique de discrimination d'un modèle logistique.
Calibration du modèle	La calibration est une autre mesure de performance d'un modèle logistique. La calibration reflète dans quelle mesure les probabilités d'événement observées dans la population concordent avec les probabilités d'événement prédites par le modèle. Le test d'Hosmer-Lemeshow est le test de calibration le plus répandu pour évaluer les performances d'un modèle logistique.

**CONCLUSION ET  
PERSPECTIVES DE RECHERCHE**

## ***1. Synthèse du travail doctoral***

Depuis une vingtaine d'années, les principes de l'évaluation des tests diagnostiques ont beaucoup évolué afin de s'intégrer à la médecine basée sur les preuves. A l'instar de ce qui a été fait dans le domaine des études interventionnelles, la manière d'envisager la place des tests diagnostiques et la méthodologie des études diagnostiques ont été repensées et formalisées. Ces nouveaux concepts et apports méthodologiques sont appelés "diagnostic basé sur les preuves". Au cours du travail doctoral, nous avons appliqué les principes les plus récents du "diagnostic basé sur les preuves" à la problématique du diagnostic des pharyngites chez l'enfant. Nous avons ainsi été amenés à utiliser des méthodes avancées de recherche bibliographique, d'évaluation du niveau de preuve et de la qualité, de modélisation d'indices de performances diagnostiques et de validation de règles de décision clinique. Dans les perspectives du travail doctoral détaillées ci-après, nous avons également développé un protocole de construction d'un nouvel outil de prédiction clinique et celui d'une méta-analyse diagnostique.

Dans le **Chapitre 1**, nous avons montré que l'effet de spectre touche la sensibilité et la spécificité du TDR mais que la valeur prédictive négative du TDR est stable. Nous avons également montré que l'effet de spectre concernant la sensibilité n'a lieu que chez les patients ayant un faible inoculum bactérien. Ainsi nos résultats suggèrent que les cliniciens peuvent se fier aux résultats négatifs du TDR indépendamment du profil clinique du patient et de son score de McIsaac. Un résultat négatif du TDR rend très improbable le diagnostic de pharyngite à SGA de manière fiable et stable. En considérant que 10% des enfants sont des porteurs de SGA et que ces patients ont le plus souvent un faible inoculum bactérien, la confirmation systématique des TDR négatifs par une mise en culture de prélèvement de gorge ne nous semble pas nécessaire.

Dans le **Chapitre 2**, nous avons montré qu'il existait d'importantes variations de la sensibilité du TDR en fonction du médecin ayant réalisé le test. Ainsi, les pédiatres qui décident de baser leur décision thérapeutique uniquement sur le résultat d'un TDR devraient d'abord s'assurer de la qualité de leur technique de réalisation du test. Ceux qui obtiennent les moins bons résultats devraient avoir accès à des formations ciblées leur permettant d'améliorer leur technique et leurs performances diagnostiques. En pratique, cela impliquerait que tous les médecins généralistes et pédiatres qui utilisent le TDR réalisent sur une période donnée un double prélèvement de gorge et qu'ils envoient un écouvillon dans un laboratoire de microbiologie pour évaluer la sensibilité et la spécificité qu'ils obtiennent avec le TDR. Cela peut sembler intéressant d'un point de vue théorique mais difficile à mettre en place. En effet, plusieurs études ont montré qu'un des principaux obstacles à la réalisation du TDR était le manque de temps de consultation et il est peu probable que des médecins qui estiment déjà manquer de temps soient prêts à se soumettre à ce type d'évaluation [131, 132].

Dans le **Chapitre 3**, nous avons montré que les résultats faux-positifs du TDR sont rares en pratique courante et que la plupart des résultats faux-positifs du TDR semblent expliqués par un phénomène d'interférence bactérienne entre *S. aureus* et le SGA qui empêche la croissance du SGA en culture. Ainsi, les faux-positifs du TDR surviennent chez des enfants réellement infectés ou colonisés par le SGA et cela confirme que la spécificité du TDR peut être considérée comme étant proche de 100% en pratique clinique. Cela confirme également que le poids des faux-positifs du TDR dans le nombre des antibiothérapies inutiles prescrites dans le cadre des pharyngites de l'enfant est minime.

Dans le **Chapitre 4**, nous avons montré que les RDC actuelles pour le diagnostic des pharyngites à SGA ne semblent pas pouvoir être recommandées en pratique clinique

chez l'enfant. Malgré les efforts des cliniciens et des épidémiologistes depuis plus de 40 ans, aucune combinaison de signes cliniques ne semble suffisamment performante pour faire le diagnostic de pharyngite à SGA de manière fiable. En ce qui concerne la combinaison de signes cliniques avec un TDR, nous avons montré que les performances de ces RDC étaient directement liées au nombre de TDR utilisés.

En conclusion, les résultats des différentes étapes du travail doctoral amènent des éléments de réponse à certains aspects de la controverse internationale quant à la place du TDR et des RDC dans le diagnostic des pharyngites de l'enfant. En outre, nos résultats apportent des arguments permettant d'appuyer les recommandations françaises publiées en 2011 mais qui étaient jusqu'à présent basées sur un faible niveau de preuve [33] :

- le clinicien peut se fier aux résultats négatifs et positifs du TDR sans avoir à réaliser de mise en culture de prélèvements de gorge (« les cultures [...] ne sont pas recommandées en dehors de rares indications ») ;
- il existe des variations importantes de sensibilité du TDR d'un médecin à l'autre et les médecins qui utilisent le TDR devraient y être correctement formés (« L'entraînement du préleveur est déterminant et conditionne la performance de ces techniques de diagnostic ») ;
- la présentation clinique ne permet pas de faire la distinction entre pharyngites liées au SGA et pharyngites virales de manière suffisamment discriminante (« Aucun signe ni symptôme n'est spécifique des angines à SGA ») ;
- la stratégie la plus sensible et la plus spécifique consiste à utiliser un TDR chez tous les enfants ayant une pharyngite afin de limiter les prescriptions d'antibiotiques au cas confirmés de pharyngites à SGA (« Sur un plan pratique, le TDR doit être réalisé de façon systématique chez l'enfant de plus de 3 ans »).

## ***2. Perspectives de recherche***

Au cours du travail doctoral, nous avons pu produire des connaissances nouvelles permettant de préciser la place du TDR et des RDC dans le diagnostic des pharyngites chez l'enfant. Toutefois, plusieurs questions auxquelles nous n'avons pu répondre ont retenu notre attention et sont l'objet de recherches en cours ou à venir.

### **2.1 Revue systématique et méta-analyse des performances du TDR**

La question des performances diagnostiques du TDR et de ses facteurs de variations n'a été que partiellement réglée. Au cours des **Chapitres 1 à 3** du travail doctoral, nous avons montré que plusieurs facteurs sont susceptibles de faire varier les performances du TDR au sein d'une étude. Cependant, nous avons constaté qu'au-delà des variations de performances intra-étude, il semblait exister des facteurs de variation de performances d'une étude à l'autre. En effet, la sensibilité du TDR rapportée dans la littérature varie entre 52,6% et 98,9% [21], et il est donc difficile de tirer des conclusions en termes de santé publique face à un test dont les performances semblent si fluctuantes. Nous sommes en train de réaliser une revue systématique avec méta-analyse diagnostique des performances du TDR dans le cadre de la collaboration Cochrane. Les objectifs de ce travail de recherche sont d'estimer les performances diagnostiques moyennes du TDR chez l'enfant et d'explorer les sources de l'hétérogénéité observée entre études y compris celles liées à la qualité méthodologie des études. Le protocole pour cette revue systématique a été publié dans la revue de la collaboration Cochrane et il est présenté en annexe 5, page 185.

## **2.2 Construction et validation d'un nouvel outil de prédiction clinique pour le diagnostic des pharyngites chez l'enfant**

Nous avons montré dans le **Chapitre 4** de la thèse que les RDC actuellement disponibles pour le diagnostic des pharyngites de l'enfant avaient un pouvoir discriminant (performances diagnostiques) et une calibration (transportabilité) faibles. Notre hypothèse est que cela pourrait être dû à des faiblesses méthodologiques concernant la construction et la validation des modèles de prédiction clinique sous-jacents. Aussi, nous avons rédigé un protocole de construction et de validation d'un nouvel outil de prédiction clinique pour le diagnostic des pharyngites chez l'enfant en nous basant sur les données issues de l'étude ACTIV TDR SGA. Ce projet d'étude a été retenu au PHRC régional 2012 (AOR 12089) et le protocole de recherche est présenté en annexe 6, page 203.

## **2.3 Comparaison du génotype *emm* entre enfants avec pharyngite et porteurs asymptomatiques**

Nous avons montré au **Chapitre 2** qu'il existait un effet indépendant du portage sain sur la sensibilité du TDR, et en particulier indépendant de l'inoculum bactérien. Cela nous a amené à émettre l'hypothèse que cette sensibilité moindre du TDR en cas de portage pourrait être liée à des différences de distribution des génotypes *emm* entre porteurs et enfants atteints de pharyngite. En effet, il est possible que certains types *emm* de SGA soient plus fréquents en cas de portage mais que leur affinité pour le TDR soit moindre que celle des types *emm* que l'on retrouve en cas de pharyngite à SGA. Une première étape serait de génotyper l'ensemble des souches de SGA identifiées dans l'étude ACTIV TDR SGA et de comparer la distribution des génotypes *emm* entre le groupe des porteurs asymptomatiques et le groupe des enfants avec pharyngite. Si cette

distribution est différente, cela rendrait pertinente l'exploration biologique de la relation entre génotype *emm* et affinité pour le TDR.

#### **2.4 Validation du test $Q$ de Cochran pour explorer l'hétérogénéité de rapports de vraisemblance entre sous-groupes de patients d'une même étude**

D'un point de vue statistique, nous avons utilisé au **Chapitre 1** une approche originale afin de rechercher à mettre en évidence des variations significatives de rapports de vraisemblance entre différents sous-groupes de patients en nous basant sur le test  $Q$  de Cochran. Il existe à ce jour peu d'approches pour explorer ce type d'effet de spectre et nous avons prolongé le travail dans cette direction en validant cette approche sur les données d'un échantillon d'études diagnostiques publiées explorant l'effet de spectre d'un test et à travers une étude de simulation (article accepté pour publication dans le *Journal of Clinical Epidemiology*, voir annexe 7 page 223).

#### **2.5 Revue systématique et méta-analyse de l'effet de l'implémentation du TDR et des RDC pour réduire les prescriptions inutiles d'antibiotiques**

Le travail doctoral a porté sur la question des performances diagnostiques de différentes stratégies pour la prise en charge des pharyngites de l'enfant. Cependant, au-delà des performances diagnostiques, l'objectif à l'échelle individuelle et en termes de santé publique est la diminution du nombre d'antibiothérapies inutiles. Il serait notamment intéressant de savoir si l'utilisation du TDR ou des RDC a un impact sur les pratiques médicales et permet de diminuer les prescriptions d'antibiotiques inutiles. Plusieurs publications rapportent les résultats de telles études d'impact mais leurs résultats sont

variables [133-137]. Une revue systématique avec méta-analyse aurait donc sa place. S'il était démontré que l'implémentation du TDR et/ou des RDC permet une réduction cliniquement significative de la consommation d'antibiotiques inutiles, cela apporterait un argument en faveur de leur utilisation. Cela permettrait également d'apporter de nouvelles estimations de l'impact attendu des TDR et des RDC qui pourraient être utilisées dans des études coût-efficacité.

## **2.6 Effet des antibiotiques sur la survenue des complications des pharyngites chez l'enfant**

L'ensemble du travail doctoral était basé sur le présupposé qu'il existe un intérêt à diagnostiquer et à traiter les pharyngites a SGA par antibiotiques. Ce présupposé est notamment appuyé par une revue systématique avec méta-analyse publiée par la collaboration Cochrane [11], ce qui est habituellement considéré comme un haut niveau de preuve. Toutefois, les résultats de cette méta-analyse méritent d'être discutés.

En ce qui concerne les complications non-suppuratives, le risque relatif de rhumatisme articulaire aigu est de 0,27 (IC95 0,12–0,60) chez les patients ayant reçu un traitement antibiotique en comparaison au groupe ayant reçu un placebo. Toutefois, ces résultats sont totalement attribuables à 10 essais cliniques réalisés dans les années 1950, pour la plupart dans des bases de l'armée américaine. Dans les 6 essais cliniques publiés après 1975, il est impossible de conclure quant à l'effet protecteur des antibiotiques car il n'y avait aucun cas de rhumatisme articulaire aigu dans les deux bras de ces essais [11]. De même, les résultats des cinq études d'incidence du rhumatisme articulaire aigu dans des pays européens les plus récentes sont hétérogènes et difficiles à interpréter [138-142]. Dans l'étude grecque rétrospective hospitalière monocentrique de Kafetzis *et al.* (1992–2002), 6 des 7 (86%) patients atteints de rhumatisme articulaire aigu avaient

reçu des antibiotiques lors de l'épisode infectieux initial [142]. Dans l'enquête hospitalière nationale française d'Olivier *et al.* (1995-1997), 11 des 32 (34%) patients atteints de rhumatisme articulaire aigu avaient reçu des antibiotiques lors de l'épisode infectieux initial [141]. La prise d'antibiotique n'avait pas été documentée dans les 3 autres études [138-140].

En ce qui concerne le phlegmon périamygdalien (*quinsy*), la méta-analyse Cochrane a montré un risque relatif de 0,15 (IC95% 0,05–0,47) en faveur des antibiotiques, ce qui correspond à un nombre de sujets à traiter de l'ordre de 50 [11]. Cependant, les résultats de cette méta-analyse d'essais randomisés sont contradictoires avec ceux d'une grande étude rétrospective analysant les données de la base d'information de soins primaires au Royaume-Uni (*General Practice Research Database-GPRD*) [143]. Dans cette étude, l'incidence du phlegmon dans les suites d'une pharyngite était de l'ordre de 0,1%, le risque relatif de développer un phlegmon en prenant des antibiotiques était de 0,84 (IC95% 0,73–0,97), soit un nombre de sujets à traiter par antibiotiques pour éviter un cas de phlegmon de l'ordre de 4000. Les résultats dans le groupe des enfants âgés de 5 à 15 ans montraient un nombre de sujets à traiter de l'ordre de 15000. C'est en raison de cette faible incidence et de ce doute concernant l'effet protecteur des antibiotiques que certains pays tels la Belgique et les Pays-Bas recommandent de ne pas traiter les pharyngites par antibiotiques [144, 145].

En termes de recherche clinique, dans l'hypothèse d'une incidence de phlegmon de 0,15% et d'un risque relatif de 0,5 en faveur des antibiotiques (nombre de sujets à traiter de 1000), un nouvel essai randomisé antibiotiques contre placebo avec une allocation 1 :1 nécessiterait d'inclure environ 20000 patients dans chaque bras (avec  $\alpha=0,05$  et  $\beta=0,8$ ), ce qui semble difficilement faisable. L'alternative est l'approche cas-témoins, si possible nichée dans une cohorte pour éviter les biais de mémorisation. Il

serait envisageable de mettre en place une étude prospective de pharmaco-épidémiologie de type cas-témoins *ad hoc* afin d'évaluer l'effet protecteur des antibiotiques sur la survenue du phlegmon périamygdalien. S'il est confirmé que les antibiotiques ne permettent pas de diminuer la survenue des complications suppuratives des pharyngites, et notamment la survenue de phlegmons périamygdaliens, il s'agirait d'un argument solide contre l'utilisation des antibiotiques et l'ensemble de la stratégie de prise en charge des enfants avec pharyngite serait remise en cause.

Les autres bénéfices attendus du traitement antibiotique sont de diminuer la durée des symptômes liés à la pharyngite et de diminuer la contagiosité en permettant une éradication rapide du SGA au niveau du pharynx. L'efficacité des antibiotiques pour réduire la durée des symptômes douloureux est cliniquement et statistiquement significative (RR 0,68 [ 0,59–0,79 ] pour le mal de gorge au troisième jour d'évolution et RR 0,49 [ 0,32–0,76 ] à une semaine d'évolution ; réduction moyenne de la durée du mal de gorge de 16 heures) et semble stable dans le temps (1951–2003) d'après la revue systématique Cochrane [11]. En ce qui concerne la contagiosité, l'effet protecteur des antibiotiques a été démontré sur la base du taux d'éradication du SGA au niveau pharyngé. Ainsi, alors qu'en l'absence de traitement antibiotique le SGA peut être retrouvé au niveau pharyngé deux semaines après le début de l'épisode infectieux dans environ 70% des cas [146], 80% des enfants traités par antibiotique ont une culture de prélèvement de gorge négative 24 heures après le début de l'antibiothérapie [147].

---

## BIBLIOGRAPHIE

1. Carapetis JR, Steer AC, Mulholland EK, Weber M. The global burden of group A streptococcal diseases. *Lancet Infect Dis* 2005;5:685-94.
2. Grijalva CG, Nuorti JP, Griffin MR. Antibiotic prescription rates for acute respiratory tract infections in US ambulatory settings. *JAMA* 2009;302:758-66.
3. McCaig LF, Besser RE, Hughes JM. Trends in antimicrobial prescribing rates for children and adolescents. *JAMA* 2002;287:3096-102.
4. Linder JA, Bates DW, Lee GM, Finkelstein JA. Antibiotic treatment of children with sore throat. *JAMA* 2005;294:2315-22.
5. Allemand H, editor Etat des lieux de la consommation d'antibiotiques en France. Résistances aux antibiotiques : Quels enjeux aujourd'hui pour la médecine de ville?; Table ronde de la CNAM, 19 mai 2010.
6. Goossens H, Ferech M, Vander Stichele R, Elseviers M. Outpatient antibiotic use in Europe and association with resistance: a cross-national database study. *Lancet* 2005;365:579-87.
7. European Centre for Disease Prevention and Control. Antimicrobial resistance surveillance in Europe 2012. Annual Report of the European Antimicrobial Resistance Surveillance Network (EARS-Net). Stockholm: ECDC, 2013.
8. Shaikh N, Leonard E, Martin JM. Prevalence of streptococcal pharyngitis and streptococcal carriage in children: a meta-analysis. *Pediatrics* 2010;126:e557-64.
9. Gerber MA. Diagnosis and treatment of pharyngitis in children. *Pediatr Clin North Am* 2005;52:729-47, vi.
10. Wessels MR. Clinical practice. Streptococcal pharyngitis. *N Engl J Med* 2011;364:648-55.
11. Spinks A, Glasziou PP, Del Mar CB. Antibiotics for sore throat. *Cochrane Database Syst Rev* 2013;11:CD000023.
12. Ebell MH, Smith MA, Barry HC, Ives K, Carey M. The rational clinical examination. Does this patient have strep throat? *JAMA* 2000;284:2912-8.
13. Park SY, Gerber MA, Tanz RR, Hickner JM, Galliher JM, Chuang I, et al. Clinicians' management of children and adolescents with acute pharyngitis. *Pediatrics* 2006;117:1871-8.
14. Kellogg JA. Suitability of throat culture procedures for detection of group A streptococci and as reference standards for evaluation of streptococcal antigen detection kits. *J Clin Microbiol* 1990;28:165-9.
15. Shaikh N, Swaminathan N, Hooper EG. Accuracy and Precision of the Signs and Symptoms of Streptococcal Pharyngitis in Children: A Systematic Review. *J Pediatr* 2011.
16. McIsaac WJ, White D, Tannenbaum D, Low DE. A clinical score to reduce unnecessary antibiotic use in patients with sore throat. *CMAJ* 1998;158:75-83.
17. Centor RM, Witherspoon JM, Dalton HP, Brody CE, Link K. The diagnosis of strep throat in adults in the emergency room. *Med Decis Making* 1981;1:239-46.
18. McIsaac WJ, Goel V, To T, Low DE. The validity of a sore throat score in family practice. *CMAJ* 2000;163:811-5.

19. McIsaac WJ, Kellner JD, Aufricht P, Vanjaka A, Low DE. Empirical validation of guidelines for the management of pharyngitis in children and adults. *JAMA* 2004;291:1587-95.
20. Cohen R, Levy C, Ovetchkine P, Boucherat M, Weil-Olivier C, Gaudelus J, et al. Evaluation of streptococcal clinical scores, rapid antigen detection tests and cultures for childhood pharyngitis. *Eur J Pediatr* 2004;163:281-2.
21. Gerber MA, Shulman ST. Rapid diagnosis of pharyngitis caused by group A streptococci. *Clin Microbiol Rev* 2004;17:571-80.
22. Edmonson MB, Farwell KR. Relationship between the clinical likelihood of group A streptococcal pharyngitis and the sensitivity of a rapid antigen-detection test in a pediatric practice. *Pediatrics* 2005;115:280-5.
23. Hall MC, Kieke B, Gonzales R, Belongia EA. Spectrum bias of a rapid antigen detection test for group A beta-hemolytic streptococcal pharyngitis in a pediatric population. *Pediatrics* 2004;114:182-6.
24. Tanz RR, Gerber MA, Kabat W, Rippe J, Seshadri R, Shulman ST. Performance of a rapid antigen-detection test and throat culture in community pediatric offices: implications for management of pharyngitis. *Pediatrics* 2009;123:437-44.
25. Gerber MA, Randolph MF, Chanatry J, Wright LL, DeMeo KK, Anderson LR. Antigen detection test for streptococcal pharyngitis: evaluation of sensitivity with respect to true infections. *J Pediatr* 1986;108:654-8.
26. Kuhn S, Davies HD, Katzko G, Jadavji T, Church DL. Evaluation of the Strep A OIA assay versus culture methods: ability to detect different quantities of group A *Streptococcus*. *Diagn Microbiol Infect Dis* 1999;34:275-80.
27. Kurtz B, Kurtz M, Roe M, Todd J. Importance of inoculum size and sampling effect in rapid antigen detection for diagnosis of *Streptococcus pyogenes* pharyngitis. *J Clin Microbiol* 2000;38:279-81.
28. Gerber MA, Tanz RR, Kabat W, Dennis E, Bell GL, Kaplan EL, et al. Optical immunoassay test for group A beta-hemolytic streptococcal pharyngitis. An office-based, multicenter investigation. *JAMA* 1997;277:899-903.
29. Fox JW, Cohen DM, Marcon MJ, Cotton WH, Bonsu BK. Performance of rapid streptococcal antigen testing varies by personnel. *J Clin Microbiol* 2006;44:3918-22.
30. Tanz RR, Shulman ST. Chronic pharyngeal carriage of group A streptococci. *Pediatr Infect Dis J* 2007;26:175-6.
31. Matthys J, De Meyere M, van Driel ML, De Sutter A. Differences among international pharyngitis guidelines: not just academic. *Ann Fam Med* 2007;5:436-43.
32. NICE. Respiratory tract infections – antibiotic prescribing 2008 [updated July 2008/January 2011]. Available from: <http://www.nice.org.uk/CG69>.
33. SFP-SPILF-GPIP. Antibiothérapie par voie générale en pratique courante dans les infections respiratoires hautes de l'adulte et de l'enfant 2011 [November 15, 2013]. Available from: <http://www.infectiologie.com>.
34. American Academy of Pediatrics. Group A Streptococcal Infections. In: Pickering L, Baker C, Long S, McMillan J, editors. *Red Book: 2012 Report of the Committee on Infectious Disease*. 29th ed. Elk Grove Village, IL: American Academy of Pediatrics; 2012.
35. Shulman ST, Bisno AL, Clegg HW, Gerber MA, Kaplan EL, Lee G, et al. Clinical practice guideline for the diagnosis and management of group A streptococcal pharyngitis: 2012 update by the Infectious Diseases Society of America. *Clin Infect Dis* 2012;55:e86-e102.

36. Choby BA. Diagnosis and treatment of streptococcal pharyngitis. *Am Fam Physician* 2009;79:383-90.
37. Pelucchi C, Grigoryan L, Galeone C, Esposito S, Huovinen P, Little P, et al. Guideline for the management of acute sore throat. *Clin Microbiol Infect* 2012;18 Suppl 1:1-28.
38. Hersh AL, Jackson MA, Hicks LA. Principles of judicious antibiotic prescribing for upper respiratory tract infections in pediatrics. *Pediatrics* 2013;132:1146-54.
39. Cohen R, Levy C, Corrad F, Boucherat M, de La Rocque F. Les réseaux spécialisés: pédiatrie de ville. *Arch Pediatr* 2006;13:634-5.
40. Fischer Walker CL, Rimoin AW, Hamza HS, Steinhoff MC. Comparison of clinical prediction rules for management of pharyngitis in settings with limited resources. *J Pediatr* 2006;149:64-71.
41. Charlier-Bret N, Boucher B, Poyart C, Quesne G, Bingen E, Doit C, et al. Tests de diagnostic rapide des angines a streptocoque du groupe A : étude comparative de la sensibilité et de la praticabilité de 16 dispositifs médicaux de diagnostic in vitro dans le cadre d'un contrôle du marché effectuée par l' Afssaps en juillet 2002. *Pathol Biol (Paris)* 2004;52:438-43.
42. Bidet P, Plainvert C, Doit C, Mariani-Kurkdjian P, Bonacorsi S, Lepoutre A, et al. Infections a *Streptococcus pyogenes* ou streptocoque du groupe A chez l'enfant : données du Centre national de référence (CNR). *Arch Pediatr* 2010;17:201-8.
43. Gunnarsson RK, Holm SE, Soderstrom M. The prevalence of beta-haemolytic streptococci in throat specimens from healthy children and adults. Implications for the clinical value of throat cultures. *Scand J Prim Health Care* 1997;15:149-55.
44. Ruiz-Aragon J, Rodriguez Lopez R, Molina Linde JM. Evaluation of rapid methods for detecting *Streptococcus pyogenes*. systematic review and meta-analysis. *An Pediatr* 2010;72:391-402.
45. Mulherin SA, Miller WC. Spectrum bias or spectrum effect? Subgroup variation in diagnostic test evaluation. *Ann Intern Med* 2002;137:598-602.
46. Goehring C, Perrier A, Morabia A. Spectrum bias: a quantitative and graphical analysis of the variability of medical diagnostic test performance. *Stat Med* 2004;23:125-35.
47. Elie C, Coste J. A methodological framework to distinguish spectrum effects from spectrum biases and to assess diagnostic and screening test accuracy for patient populations: application to the Papanicolaou cervical cancer smear test. *BMC Med Res Methodol* 2008;8:7.
48. Ransohoff DF, Feinstein AR. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. *N Engl J Med* 1978;299:926-30.
49. Dale JC, Vetter EA, Contezac JM, Iverson LK, Wollan PC, Cockerill FR, 3rd. Evaluation of two rapid antigen assays, BioStar Strep A OIA and Pacific Biotech CARDS O.S., and culture for detection of group A streptococci in throat swabs. *J Clin Microbiol* 1994;32:2698-701.
50. Dimatteo LA, Lowenstein SR, Brimhall B, Reiquam W, Gonzales R. The relationship between the clinical features of pharyngitis and the sensitivity of a rapid antigen test: evidence of spectrum bias. *Ann Emerg Med* 2001;38:648-52.
51. Wacholder S. Binomial regression in GLIM: estimating risk ratios and risk differences. *Am J Epidemiol* 1986;123:174-84.
52. Johnson DR, Kurlan R, Leckman J, Kaplan EL. The human immune response to streptococcal extracellular antigens: clinical, diagnostic, and potential pathogenetic implications. *Clin Infect Dis* 2010;50:481-90.

53. Johnson DR, Kaplan EL. False-positive rapid antigen detection test results: reduced specificity in the absence of group A streptococci in the upper respiratory tract. *J Infect Dis* 2001;183:1135-7.
54. Lieu TA, Fleisher GR, Schwartz JS. Clinical evaluation of a latex agglutination test for streptococcal pharyngitis: performance and impact on treatment rates. *Pediatr Infect Dis J* 1988;7:847-54.
55. Nsanze H, Dawson KP, Ameen AS, Mustafa N. Group A streptococcal antigen detection in schoolchildren. *East Mediterr Health J* 1998;4:260-4.
56. Dawson KP, Ameen AS, Nsanze H, Bin-Othman S, Mustafa N. The prevalence of group A streptococcal throat carriage in Al Ain, United Arab Emirates. *Ann Trop Paediatr* 1996;16:123-7.
57. Gupta R, Kalia A, Rattan A, Kumar R, Gupta SK. Comparative evaluation of two indigenously developed tests for rapid detection of group-A streptococci directly from throat swabs. *Indian J Med Res* 1997;105:200-5.
58. Landon BE, Reschovsky J, Reed M, Blumenthal D. Personal, organizational, and market level influences on physicians' practice patterns: results of a national survey of primary care physicians. *Med Care* 2001;39:889-905.
59. Tracy CS, Dantas GC, Moineddin R, Upshur RE. Contextual factors in clinical decision making: national survey of Canadian family physicians. *Can Fam Physician* 2005;51:1106-7.
60. Diez-Roux AV. Multilevel analysis in public health research. *Annu Rev Public Health* 2000;21:171-92.
61. Coughlin SS, Trock B, Criqui MH, Pickle LW, Browner D, Tefft MC. The logistic modeling of sensitivity, specificity, and predictive value of a diagnostic test. *J Clin Epidemiol* 1992;45:1-7.
62. Grammatico-Guillon L, Thiolet JM, Bernillon P, Coignard B, Khoshnood B, Desenclos JC. Relationship between the prevalence of methicillin-resistant *Staphylococcus aureus* infection and indicators of nosocomial infection control measures: a population-based study in French hospitals. *Infect Control Hosp Epidemiol* 2009;30:861-9.
63. Snijders TAB, Bosker RJ. *Multilevel Analysis: An Introduction to Basic and Applied Multilevel Analysis*. 2nd ed. London, UK: Sage; 2012.
64. Blandino G, Puglisi S, Speciale A, Musumeci R. *Streptococcus pyogenes* emm types and subtypes of isolates from paediatric asymptomatic carriers and children with pharyngitis. *New Microbiol* 2011;34:101-4.
65. Cohen JF, Chalumeau M, Levy C, Bidet P, Thollot F, Wollner A, et al. Spectrum and inoculum size effect of a rapid antigen detection test for group A streptococcus in children with pharyngitis. *PLoS One* 2012;7:e39085.
66. Bell SM, Smith DD. Quantitative throat-swab culture in the diagnosis of streptococcal pharyngitis in children. *Lancet* 1976;2:62-3.
67. American Academy of Pediatrics. Group A Streptococcal Infections. In: Pickering L, Baker C, Long S, McMillan J, editors. *Red Book: 2006 Report of the Committee on Infectious Disease*. 27th ed. Elk Grove Village, IL: American Academy of Pediatrics; 2006. p. 616-28.
68. Rubin LG, Kahn RA, Vellozzi EM, Isenberg HD. False positive detection of group A *Streptococcus* antigen resulting from cross-reacting *Streptococcus intermedius* (*Streptococcus milleri* group). *Pediatr Infect Dis J* 1996;15:715-7.
69. Rubin LG, Mirkin GS. Apparent false positive detection of group A *Streptococcus* antigen resulting from pharyngeal infection with a nonhemolytic *Streptococcus pyogenes*. *Pediatr Infect Dis J* 2000;19:672-4.

70. Schroeder S, Procop GW. False positive strep A antigen test. *Pediatr Infect Dis J* 2000;19:1114-5.
71. Gerber MA, Randolph MF, DeMeo KK. Streptococcal antigen in the pharynx after initiation of antibiotic therapy. *Pediatr Infect Dis J* 1987;6:489-91.
72. Brook I. Bacterial interference. *Crit Rev Microbiol* 1999;25:155-72.
73. Kaltwasser G, Diego J, Welby-Sellenriek PL, Ferrett R, Caparon M, Storch GA. Polymerase chain reaction for *Streptococcus pyogenes* used to evaluate an optical immunoassay for the detection of group A streptococci in children with pharyngitis. *Pediatr Infect Dis J* 1997;16:748-53.
74. Hsueh PR, Wu JJ, Tsai PJ, Liu JW, Chuang YC, Luh KT. Invasive group A streptococcal disease in Taiwan is not associated with the presence of streptococcal pyrogenic exotoxin genes. *Clin Infect Dis* 1998;26:584-9.
75. Bidet P, Liguori S, Plainvert C, Bonacorsi S, Courroux C, d'Humieres C, et al. Identification of group A streptococcal emm types commonly associated with invasive infections and antimicrobial resistance by the use of multiplex PCR and high-resolution melting analysis. *Eur J Clin Microbiol Infect Dis* 2012;31:2817-26.
76. Hamrick HJ, Mangum ME. Beta-hemolytic *Streptococcus milleri* group misidentified as *Streptococcus pyogenes* on throat culture. *Pediatr Infect Dis J* 1999;18:75-6.
77. Bogaert D, van Belkum A, Sluijter M, Luijendijk A, de Groot R, Rumke HC, et al. Colonisation by *Streptococcus pneumoniae* and *Staphylococcus aureus* in healthy children. *Lancet* 2004;363:1871-2.
78. Bogaert D, Keijser B, Huse S, Rossen J, Veenhoven R, van Gils E, et al. Variability and diversity of nasopharyngeal microbiota in children: a metagenomic analysis. *PLoS One* 2011;6:e17035.
79. Willis BH. Empirical evidence that disease prevalence may affect the performance of diagnostic tests with an implicit threshold: a cross-sectional study. *BMJ Open* 2012;2:e000746.
80. Eggin TK, Feinstein AR. Context bias. A problem in diagnostic radiology. *JAMA* 1996;276:1752-5.
81. Tyler SD, Johnson WM, Huang JC, Ashton FE, Wang G, Low DE, et al. Streptococcal erythrogenic toxin genes: detection by polymerase chain reaction and association with disease in strains isolated in Canada from 1940 to 1991. *J Clin Microbiol* 1992;30:3127-31.
82. Bidet P, Lesteven E, Doit C, Liguori S, Mariani-Kurkdjian P, Bonacorsi S, et al. Subtyping of emm1 group A Streptococci causing invasive infections in France. *J Clin Microbiol* 2009;47:4146-9.
83. Bidet P, Courroux C, Salgueiro C, Carol A, Mariani-Kurkdjian P, Bonacorsi S, et al. Molecular epidemiology of the sil streptococcal invasive locus in group A streptococci causing invasive infections in French children. *J Clin Microbiol* 2007;45:2002-4.
84. Wannamaker LW. Streptococcal toxins. *Rev Infect Dis* 1983;5 Suppl 4:S723-32.
85. James L, McFarland RB. An epidemic of pharyngitis due to a nonhemolytic group A streptococcus at Lowry Air Force Base. *N Engl J Med* 1971;284:750-2.
86. Clawson CC, Dajani AS. Effect of Bactericidal Substance from *Staphylococcus aureus* on Group A Streptococci II. Structural Alterations. *Infect Immun* 1970;1:491-8.
87. American Academy of Pediatrics. Group A Streptococcal Infections. In: Pickering L, Baker C, Long S, McMillan J, editors. *Red Book: 2009 Report of the Committee*

- on Infectious Disease. 28th ed. Elk Grove Village, IL: American Academy of Pediatrics; 2009. p. 616-28.
88. Shaikh N, Swaminathan N, Hooper EG. Accuracy and precision of the signs and symptoms of streptococcal pharyngitis in children: a systematic review. *J Pediatr* 2012;160:487-93 e3.
  89. Le Marechal F, Martinot A, Duhamel A, Pruvost I, Dubos F. Streptococcal pharyngitis in children: a meta-analysis of clinical decision rules and their clinical variables. *BMJ Open* 2013;3.
  90. Breese BB. A simple scorecard for the tentative diagnosis of streptococcal pharyngitis. *Am J Dis Child* 1977;131:514-7.
  91. Wald ER, Green MD, Schwartz B, Barbadora K. A streptococcal score card revisited. *Pediatr Emerg Care* 1998;14:109-11.
  92. Wasson JH, Sox HC, Neff RK, Goldman L. Clinical prediction rules. Applications and methodological standards. *N Engl J Med* 1985;313:793-9.
  93. Laupacis A, Sekar N, Stiell IG. Clinical prediction rules. A review and suggested modifications of methodological standards. *JAMA* 1997;277:488-94.
  94. McGinn TG, Guyatt GH, Wyer PC, Naylor CD, Stiell IG, Richardson WS. How to use articles about clinical decision rules. Evidence-Based Medicine Working Group. *JAMA* 2000;284:79-84.
  95. Stiell IG, Wells GA. Methodologic standards for the development of clinical decision rules in emergency medicine. *Ann Emerg Med* 1999;33:437-47.
  96. Steyerberg EW, Moons KG, van der Windt DA, Hayden JA, Perel P, Schroter S, et al. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. *PLoS Med* 2013;10:e1001381.
  97. Bouwmeester W, Zuithoff NP, Mallett S, Geerlings MI, Vergouwe Y, Steyerberg EW, et al. Reporting and methods in clinical prediction research: a systematic review. *PLoS Med* 2012;9:1-12.
  98. Steyerberg EW. *Clinical Prediction Models: a practical approach to development, validation, and updating*. 1st ed. New York, NY: Springer; 2009. 500 p.
  99. Keogh C, Wallace E, O'Brien KK, Murphy PJ, Teljeur C, McGrath B, et al. Optimized retrieval of primary care clinical prediction rules from MEDLINE to establish a Web-based register. *J Clin Epidemiol* 2011;64:848-60.
  100. Boluyt N, Tjosvold L, Lefebvre C, Klassen TP, Offringa M. Usefulness of systematic review search strategies in finding child health systematic reviews in MEDLINE. *Arch Pediatr Adolesc Med* 2008;162:111-6.
  101. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, 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.
  102. Whiting PF, Rutjes AW, 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;155:529-36.
  103. Hosmer D, Lemeshow S. *Applied Logistic Regression*. 2nd ed. New York, NY: John Wiley & Sons; 2000.
  104. Harrell FE, Jr., Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996;15:361-87.
  105. Concato J, Feinstein AR, Holford TR. The risk of determining risk with multivariable models. *Ann Intern Med* 1993;118:201-10.

106. Moons KG, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. *BMJ* 2009;338:b606.
107. Altman DG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research: validating a prognostic model. *BMJ* 2009;338:b605.
108. Royston P, Moons KG, Altman DG, Vergouwe Y. Prognosis and prognostic research: Developing a prognostic model. *BMJ* 2009;338:b604.
109. Moons KG, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? *BMJ* 2009;338:b375.
110. Poses RM, Cebul RD, Collins M, Fager SS. The importance of disease prevalence in transporting clinical prediction rules. The case of streptococcal pharyngitis. *Ann Intern Med* 1986;105:586-91.
111. Forsyth RA. Selective utilization of clinical diagnosis in treatment of pharyngitis. *J Fam Pract* 1975;2:173-7.
112. Joachim L, Campos D, Jr., Smeesters PR. Pragmatic scoring system for pharyngitis in low-resource settings. *Pediatrics* 2010;126:e608-14.
113. Sahin F, Ulukol B, Aysev D, Suskan E. The validity of diagnostic criteria for streptococcal pharyngitis in Integrated Management of Childhood Illness (IMCI) guidelines. *J Trop Pediatr* 2003;49:377-9.
114. Smeesters PR, Campos D, Jr., Van Melder L, de Aguiar E, Vanderpas J, Vergison A. Pharyngitis in low-resources settings: a pragmatic clinical approach to reduce unnecessary antibiotic use. *Pediatrics* 2006;118:e1607-11.
115. Attia M, Zaoutis T, Eppes S, Klein J, Meier F. Multivariate predictive models for group A beta-hemolytic streptococcal pharyngitis in children. *Acad Emerg Med* 1999;6:8-13.
116. Edmond KM, Grimwood K, Carlin JB, Chondros P, Hogg GG, Barnett PL. Streptococcal pharyngitis in a paediatric emergency department. *Med J Aust* 1996;165:420-3.
117. Nandi S, Kumar R, Ray P, Vohra H, Ganguly NK. Clinical score card for diagnosis of group A streptococcal sore throat. *Indian J Pediatr* 2002;69:471-5.
118. Attia MW, Zaoutis T, Klein JD, Meier FA. Performance of a predictive model for streptococcal pharyngitis in children. *Arch Pediatr Adolesc Med* 2001;155:687-91.
119. Steinhoff MC, Walker CF, Rimo AW, Hamza HS. A clinical decision rule for management of streptococcal pharyngitis in low-resource settings. *Acta Paediatr* 2005;94:1038-42.
120. Ulukol B, Gunlemez A, Aysev D, Cin S. Alternative diagnostic method for streptococcal pharyngitis: Breese scoring system. *Turk J Pediatr* 2000;42:96-100.
121. Karacan M, Karakelleoglu C, Orbak Z. Diagnosis of group A beta-hemolytic Streptococcus using the Breese clinical scoring system. *South Med J* 2007;100:1192-7.
122. Fine AM, Nizet V, Mandl KD. Large-scale validation of the Centor and McIsaac scores to predict group A streptococcal pharyngitis. *Arch Intern Med* 2012;172:847-52.
123. McGinn TG, Guyatt GH, Wyer PC, Naylor CD, Stiell IG, Richardson WS. Users' guides to the medical literature: XXII: how to use articles about clinical decision rules. Evidence-Based Medicine Working Group. *JAMA* 2000;284:79-84.
124. Maguire JL, Kulik DM, Laupacis A, Kuppermann N, Uleryk EM, Parkin PC. Clinical prediction rules for children: a systematic review. *Pediatrics* 2011;128:e666-77.

125. Jensen DM, Brousseau DC, Tumpach EA, Gorelick MH. Intervention to improve interobserver agreement in the assessment of children with pharyngitis. *Pediatr Emerg Care* 2005;21:238-41.
126. Schwartz K, Monsur J, Northrup J, West P, Neale AV. Pharyngitis clinical prediction rules: effect of interobserver agreement: a MetroNet study. *J Clin Epidemiol* 2004;57:142-6.
127. Donner-Banzhoff N, Beck C, Meyer F, Werner JA, Baum E. Clinical findings in patients presenting with sore throat. A study on inter-observer reliability. *Fam Pract* 2002;19:466-8.
128. Fujikawa S, Ito Y, Ohkuni M. A new scoring system for diagnosis of streptopharyngitis. *Jpn Circ J* 1985;49:1258-61.
129. Steinhoff MC, Abd el Khalek MK, Khallaf N, Hamza HS, el Ayadi A, Orabi A, et al. Effectiveness of clinical guidelines for the presumptive treatment of streptococcal pharyngitis in Egyptian children. *Lancet* 1997;350:918-21.
130. Reed BD, Huck W, French T. Diagnosis of group A beta-hemolytic *Streptococcus* using clinical scoring criteria, Directigen 1-2-3 group A streptococcal test, and culture. *Arch Intern Med* 1990;150:1727-32.
131. Société Française d'Oto-Rhino-Laryngologie et de Chirurgie de la Face et du Cou. Recommendation pour la pratique clinique : Complications locorégionales des pharyngites Octobre 2008 [cited 2013 Novembre]. Available from: <http://www.orlfrance.org>.
132. Dunn N, Lane D, Everitt H, Little P. Use of antibiotics for sore throat and incidence of quinsy. *Br J Gen Pract* 2007;57:45-9.
133. Llor C, Madurell J, Balague-Corbella M, Gomez M, Cots JM. Impact on antibiotic prescription of rapid antigen detection testing in acute pharyngitis in adults: a randomised clinical trial. *Br J Gen Pract* 2011;61:e244-51.
134. Ayanruoh S, Waseem M, Quee F, Humphrey A, Reynolds T. Impact of rapid streptococcal test on antibiotic use in a pediatric emergency department. *Pediatr Emerg Care* 2009;25:748-50.
135. Maltezou HC, Tsagris V, Antoniadou A, Galani L, Douros C, Katsarolis I, et al. Evaluation of a rapid antigen detection test in the diagnosis of streptococcal pharyngitis in children and its impact on antibiotic prescription. *J Antimicrob Chemother* 2008;62:1407-12.
136. Worrall G, Hutchinson J, Sherman G, Griffiths J. Diagnosing streptococcal sore throat in adults: randomized controlled trial of in-office aids. *Can Fam Physician* 2007;53:666-71.
137. McIsaac WJ, Goel V, To T, Permaul JA, Low DE. Effect on antibiotic prescribing of repeated clinical prompts to use a sore throat score: lessons from a failed community intervention study. *J Fam Pract* 2002;51:339-44.
138. Bolz D, Tyndall A. Rhumatisme articulaire aigu – encore actuel en Suisse ? *Forum Médical Suisse* 2006;6:642-6.
139. Breda L, Marzetti V, Gaspari S, Del Torto M, Chiarelli F, Altobelli E. Population-based study of incidence and clinical characteristics of rheumatic fever in Abruzzo, central Italy, 2000-2009. *J Pediatr* 2012;160:832-6 e1.
140. Giannoulia-Karantana A, Anagnostopoulos G, Kostaridou S, Georgakopoulou T, Papadopoulou A, Papadopoulos G. Childhood acute rheumatic fever in Greece: experience of the past 18 years. *Acta Paediatr* 2001;90:809-12.
141. Olivier C, Portier H, Cohen R, Schlemmer B, Boucot I, Peyramond D. Rhumatisme articulaire aigu : résultats d'une enquête nationale (1995-1997). *Bulletin Epidémiologique Hebdomadaire* 1999;12:45-7.

142. Kafetzis DA, Chantzi FM, Grigoriadou G, Vougiouka O, Liapi G. Incidence and clinical profile of acute rheumatic fever in Greece. *Eur J Clin Microbiol Infect Dis* 2005;24:68-70.
143. Petersen I, Johnson AM, Islam A, Duckworth G, Livermore DM, Hayward AC. Protective effect of antibiotics against serious complications of common respiratory tract infections: retrospective cohort study with the UK General Practice Research Database. *BMJ* 2007;335:982.
144. De Meyere M, Matthys J. Mal de gorge aigu: Recommandations de Bonne Pratique. 2001. Available from: <http://www.inami.fgov.be/care/fr/doctors/promotion-quality/guidelines-antibiotics/gorge-keelpijn/pdf/rbpcomplet.pdf>.
145. Dagnelie CF, Zwart S, Balder FA, Romeijnders ACM, Geije RMM. The Dutch College of General Practitioners (NHG) Practice Guideline: Acute sore throat 1999. Available from: [iskra.bfm.hr/upload/AcuteSoreThroat.doc](http://iskra.bfm.hr/upload/AcuteSoreThroat.doc).
146. Zwart S, Rovers MM, de Melker RA, Hoes AW. Penicillin for acute sore throat in children: randomised, double blind trial. *BMJ* 2003;327:1324.
147. Snellman LW, Stang HJ, Stang JM, Johnson DR, Kaplan EL. Duration of positive throat cultures for group A streptococci after initiation of antibiotic therapy. *Pediatrics* 1993;91:1166-70.



# ANNEXES

Annexe 1. Article correspondant au Chapitre 1

Annexe 2. Article correspondant au Chapitre 2

Annexe 3. Article correspondant au Chapitre 3

Annexe 4. Articles de synthèse en français

Annexe 5. Protocole de revue systématique Cochrane

Annexe 6. Protocole pour la construction d'un nouvel outil prédictif

Annexe 7. Validation du test  $Q$  de Cochran

Annexe 8. Fiches de recueil de données



---

ANNEXE 1. ARTICLE CORRESPONDANT AU CHAPITRE 1

# Spectrum and Inoculum Size Effect of a Rapid Antigen Detection Test for Group A Streptococcus in Children with Pharyngitis

J r mie F. Cohen<sup>1,2\*</sup>, Martin Chalumeau<sup>1,2</sup>, Corinne Levy<sup>3</sup>, Philippe Bidet<sup>4</sup>, Franck Thollot<sup>5</sup>, Alain Wollner<sup>3</sup>, Edouard Bingen<sup>4</sup>, Robert Cohen<sup>3,6</sup>

**1** Unit  de Recherche Epid miologique en Sant  P rinatale et Sant  des Femmes et des Enfants (U953), Inserm, Paris, France, **2** Service de P diatrie G n rale, H pital Necker-Enfants-Malades, AP-HP et Universit  Paris-Descartes, Paris, France, **3** Association Clinique et Th rapeutique du Val-de-Marne (ACTIV), Saint-Maur-des-Foss s, France, **4** Service de Microbiologie, H pital Robert Debr , AP-HP et Universit  Paris-Diderot, Paris, France, **5** Association Fran aise de P diatrie Ambulatoire (AFPA), Essey-l s-Nancy, France, **6** Service de Microbiologie, Centre Hospitalier Intercommunal de Cr teil, Cr teil, France

## Abstract

**Background:** The stability of the accuracy of a diagnostic test is critical to whether clinicians can rely on its result. We aimed to assess whether the performance of a rapid antigen detection test (RADT) for group A streptococcus (GAS) is affected by the clinical spectrum and/or bacterial inoculum size.

**Methods:** Throat swabs were collected from 785 children with pharyngitis in an office-based, prospective, multicenter study (2009–2010). We analysed the effect of clinical spectrum (i.e., the Mclsaac score and its components) and inoculum size (light or heavy GAS growth) on the accuracy (sensitivity, specificity, likelihood ratios and predictive values) of a RADT, with laboratory throat culture as the reference test. We also evaluated the accuracy of a Mclsaac-score-based decision rule.

**Results:** GAS prevalence was 36% (95CI: 33%–40%). The inoculum was heavy for 85% of cases (81%–89%). We found a significant spectrum effect on sensitivity, specificity, likelihood ratios and positive predictive value ( $p < 0.05$ ) but not negative predictive value, which was stable at about 92%. RADT sensitivity was greater for children with heavy than light inoculum (95% vs. 40%,  $p < 0.001$ ). After stratification by inoculum size, the spectrum effect on RADT sensitivity was significant only in patients with light inoculum, on univariate and multivariate analysis. The Mclsaac-score-based decision rule had 99% (97%–100%) sensitivity and 52% (48%–57%) specificity.

**Conclusions:** Variations in RADT sensitivity only occur in patients with light inocula. Because the spectrum effect does not affect the negative predictive value of the test, clinicians who want to rule out GAS can rely on negative RADT results regardless of clinical features if they accept that about 10% of children with negative RADT results will have a positive throat culture. However, such a policy is more acceptable in populations with very low incidence of complications of GAS infection.

**Citation:** Cohen JF, Chalumeau M, Levy C, Bidet P, Thollot F, et al. (2012) Spectrum and Inoculum Size Effect of a Rapid Antigen Detection Test for Group A Streptococcus in Children with Pharyngitis. PLoS ONE 7(6): e39085. doi:10.1371/journal.pone.0039085

**Editor:** Bernard Beall, Centers for Disease Control & Prevention, United States of America

**Received:** March 7, 2012; **Accepted:** May 18, 2012; **Published:** June 29, 2012

**Copyright:**   2012 Cohen 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.

**Funding:** This study was funded by Dectrapharm, manufacturer of the RADT (Rapid Antigen Detection Test). JFC is a masters student supported by educational grants from Agence R gionale de Sant  d' le-de-France and Laboratoires Guigoz - Soci t  Fran aise de P diatrie - Groupe de P diatrie G n rale - Groupe de Recherches Epid miologiques en P diatrie. This work was supported by a research grant from the Assistance Publique-H pitaux de Paris (PHRC R gional 2012, AOR 12089). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** This study was funded by Dectrapharm, manufacturer of the RADT (Rapid Antigen Detection Test). Dectrapharm had no role in the design and conduct of the study, the collection, analysis, and interpretation of the data, or the review or approval of the manuscript. There are no patents, products in development or marketed products to declare. This does not alter the authors' adherence to all the PLoS ONE policies on sharing data and materials.

\* E-mail: jeremie.cohen@inserm.fr

## Introduction

Group A streptococcus (GAS) is found in 20% to 40% of cases of childhood pharyngitis; the remaining cases are considered viral [1]. Clinical examination cannot distinguish accurately between viral and GAS pharyngitis [2], and a diagnostic test based on a throat swab is recommended in most countries [3]. Throat culture on a blood agar plate in a microbiology laboratory for 48 hours is the reference test for diagnosis of GAS pharyngitis [4–7]. Rapid antigen detection tests (RADTs) have been proposed as an alternative to throat culture. Compared to laboratory culture, RADTs have high specificity ( $\approx 95\%$ ) and results are immediate.

Their main drawbacks are low sensitivity ( $\approx 85\%$ , range 65.6% to 98.9%) [8,9] and variations in sensitivity by clinical spectrum of the disease (spectrum effect, or spectrum bias) [10–12]. The major issue for the spectrum effect is the generalisability of test performance. First, the overall population estimate might not be generalisable to patient subpopulations; second, the diagnostic accuracy that was observed in one study might not be applicable to other patients. Three studies have shown a significant spectrum effect on RADT sensitivity, but these studies had methodological limitations suggesting selection, indication, partial verification, and measurement biases [13–15]. No studies investigated the effect of

clinical spectrum on sensitivity, specificity, likelihood ratios and predictive values of RADTs at the same time [13–15]. As well, RADT sensitivity is affected by inoculum size [16–18] – the amount of GAS colonies identified on throat culture and considered a proxy of the bacterial load on the swab. In vitro, the threshold of positivity of RADTs is between  $10^5$  and  $10^7$  colony-forming units per ml [19]. To date, several studies evaluated the effects of clinical spectrum and inoculum size on RADT sensitivity, but these effects were studied only separately, and no study has analysed the potential relation of clinical spectrum and inoculum size.

Because of the need to reduce antibiotics consumption, clinical decision rules were developed to help clinicians determine which patients should undergo testing and/or treatment with antibiotics. A decision rule based on the McIsaac score was recently proposed for adults and children [20–22], but this McIsaac-score-strategy was insufficiently validated in children [21–23]. In one validation study by McIsaac *et al.*, the reader cannot evaluate the risk of selection bias because the distribution of scores in children was not reported [21]. Another validation study did not include children with low-risk scores (score  $\leq 1$ ) [22]. A third study aimed to validate the McIsaac score in children, but the rule was modified for application in low-resource settings and no longer integrated the use of a throat culture or RADT [23].

We aimed to determine whether the diagnostic accuracy of a RADT for GAS pharyngitis is affected by the clinical spectrum effect and/or bacterial inoculum size and to validate a McIsaac-score-based decision rule.

## Methods

### Study Design

This is a secondary analysis of data from an office-based, multicenter, prospective study that took place in France between October 2009 and June 2011 (unpublished data). The aims of the princeps study were to evaluate the frequency of GAS carriage in healthy children and to compare the performance of a RADT between children with pharyngitis and healthy children, with throat culture in a microbiology laboratory as the reference test (intermediate results presented at the 28<sup>th</sup> Annual Meeting of The European Society for Paediatric Infectious Diseases, Nice, France, 2010; abstract A-229-0021-00920). This ancillary analysis was restricted to patients with pharyngitis from the first year of inclusion. The STARD statement was followed for reporting the results of the study [24].

### Patients

Eligible patients were 3 to 15 years old who were evaluated by their paediatrician between October 2009 and May 2010, who had a diagnosis of pharyngitis and did not receive antibiotics for 7 days before inclusion. Pharyngitis was defined by inflammation of the pharynx and/or tonsils. In total, 17 French paediatricians who are part of a research and teaching network (ACTIV) participated in the study [25].

### Throat Swabs

Throat samples were obtained by use of a double-swab collection-transportation system (Venturi Transystem Amies agar, COPAN Diagnostics, Corona, CA, USA). The RADT (StreptAtest, Dectrapharm, France), a GAS-specific immuno-chromatographic strip assay, was performed immediately with swab #1 collected in the paediatrician's office. Swab #2 was held at ambient temperature and sent within 72 hr to the Robert Debré Hospital laboratory by an express messenger service. On receipt,

the swab was rolled over one-quarter of a trypticase soy agar plate with 5% sheep blood, and the inoculum was further distributed on the plate by streaking with a sterile wire loop. The plates were incubated anaerobically at 37°C and read after 18 to 24 hr. Negative cultures were reexamined after an additional 24 hr of incubation (48 hr total).  $\beta$ -haemolytic colonies were further investigated by latex agglutination (Prolex, Pro-Lab Diagnostics, Richmond Hill, ON, Canada). With GAS positivity, inoculum size was estimated as follows: 1+, <10 colonies; 2+, 11–50 colonies; and 3+, >50 colonies per plate. Data for patients with 1+ and 2+ cultures were combined to produce a dichotomized variable (light or heavy GAS growth). Microbiologists were blinded to individual clinical data and RADT results.

### Clinical Data

We collected clinical data needed to calculate the McIsaac score for each patient. The McIsaac score predicts GAS pharyngitis and ranges from 0 to 5 [20–22]. McIsaac criteria include age 3 to 15 years, history of fever, tonsillar swelling or exudate, tender anterior cervical adenopathy, and absence of cough. As for other studies [13,26], the adenopathy criterion was modified to include any anterior cervical adenopathy  $\geq 1$  cm and/or presence of tenderness because tender lymph nodes can be difficult to assess in children. As for other authors [13–15], extreme McIsaac scores were combined (scores 1 and 2, and 4 and 5), and for analysis of clinical components of the score, age was re-coded into 2 categories of equal range (3–8 years; 9–14 years).

### Sample Size

In the original study, sample size was estimated so that the 95% confidence interval (95CI) for RADT sensitivity would be a +/–5% estimation. Assuming sensitivity to vary between 75% and 95% and a GAS prevalence of 35%, a sample of 714 children with pharyngitis was needed. No other sample size calculation was performed for this secondary analysis.

### Ethics

The study protocol was approved by the Institutional Review Board Comité de Protection des Personnes Ile-de-France XI (n°09016) and the French administrative authorities (CNIL, n°1354254; Afssaps, n°2009-A00086-51). Parent and patient approval for participation was obtained before inclusion. Data were double entered into 4D software version 6.4 (4D) and the database was fully anonymized.

### Statistical Analysis

Hospital laboratory throat culture was considered the reference test. First, we described patient demographic features and overall prevalence of groups A, C and G  $\beta$ -haemolytic streptococci. Patients with group C or G  $\beta$ -haemolytic streptococci were considered GAS-negative in the analysis. Then, GAS prevalence by McIsaac score and each McIsaac criterion were described. Second, we used chi-square tests to investigate the clinical spectrum effect on RADT diagnostic accuracy (sensitivity, specificity, likelihood ratios and predictive values) by McIsaac score and each McIsaac criterion. Third, chi-square tests were used to compare the distribution of GAS inoculum size by office-to-laboratory delay ( $\leq 48$  vs.  $>48$  hr) and to study the effect of inoculum size (light or heavy GAS growth) on RADT sensitivity. Fourth, the distribution of heavy inocula by McIsaac score was explored by a chi-square test for trend, and the association of clinical signs and inoculum size was evaluated by comparing the frequency of heavy inoculum according to each clinical criterion

by chi-square tests. Fifth, a binomial model with an identity link was used to evaluate the combination of clinical spectrum effect and inoculum size on RADT sensitivity [27]. Differences in RADT sensitivity were estimated by restricting the analysis to patients with a positive throat culture and using the result of the RADT as the outcome in the model. RADT sensitivity was investigated by univariate modeling according to each McIsaac criterion and then, multivariate modeling by all McIsaac criteria and inoculum size. Potential interaction between inoculum size and each McIsaac criterion was assessed by use of a Wald test. For multivariate modeling, variables were added one by one following the crescent univariate statistical significance order until non-convergence of the model. Significant interactions were managed by introducing corresponding interaction terms in the multivariate model and, in case of non-convergence, by stratification on inoculum size. Sixth, the accuracy of the following McIsaac-score-based decision rule was studied in terms of sensitivity and specificity: (a) scores <2, no further testing or antibiotic; (b) scores 2–3, culture all, antibiotics only with positive culture; (c) scores ≥4, treat empirically with antibiotics [22]. Statistical analysis involved use of Stata/SE 12 (StataCorp, College Station, TX, USA).

## Results

### Population Characteristics, GAS Prevalence

In total, 807 children met the inclusion criteria; 1 patient with unreadable RADT results, 2 with lost throat swabs and 19 with missing data for McIsaac score calculation were secondarily excluded. Therefore, analysis involved 785 children (351 girls [44.7%]), with mean (SD) age 6.1 (2.5) years. Overall, GAS prevalence was 36.3% (95CI, 32.9%–39.8%). In all, 4 (<1%) Group C and 7 (<1%) Group G β-haemolytic streptococci were identified on BAP. The mean age of secondarily excluded children was 5.9 (3.0) years and GAS was found in throat cultures for 4/20 (20.0% [7.9%–39.2%]).

GAS prevalence increased with increasing McIsaac score ( $p < 0.001$ ) (Tables 1 and 2), from 24.3% to 41.2% for children with ≤2 and ≥4 McIsaac score, respectively. Some clinical criteria were also associated with GAS: age <9 years, absence of cough, and presence of fever (Table 2).

**Table 1.** Distribution of the McIsaac scores and corresponding frequencies of positive results of rapid antigen detection test (RADT), group A streptococcus (GAS) and heavy GAS inocula (N = 785).

McIsaac score <sup>a</sup>	n (%)			
	Total	RADT -positive	GAS -positive <sup>b</sup>	Heavy inoculum <sup>c</sup>
1	12 (1)	2/12 (17)	3/12 (25)	2/3 (67)
2	103 (13)	21/103 (20)	25/103 (24)	20/25 (80)
3	262 (33)	82/262 (31)	89/262 (34)	73/89 (82)
4	286 (36)	115/286 (40)	114/286 (40)	100/114 (88)
5	122 (16)	54/122 (44)	54/122 (44)	48/54 (89)

<sup>a</sup>All scores were ≥1 because all included patients were 3 to 15 years old.

<sup>b</sup>Laboratory throat culture positive for GAS.

<sup>c</sup> $p = 0.09$ ,  $\chi^2$  test for trend for the frequency of heavy inoculum by McIsaac score.

doi:10.1371/journal.pone.0039085.t001

### Spectrum Effect

Overall, the sensitivity and specificity of the RADT was 86.7% (82.2%–90.4%) and 94.6% (92.2%–96.4%), respectively; positive and negative likelihood ratios 16.1 (11.1–23.2) and 0.14 (0.10–0.19), respectively; and positive and negative predictive values 90.1% (86.0%–93.4%) and 92.6% (89.9%–94.7%), respectively.

We found significant variations in RADT sensitivity and specificity by McIsaac score (Table 2). Sensitivity increased from 75% (55.1%–89.3%) to 84.3% (75.0%–91.1%) and to 89.9% (84.3%–94.0%;  $p = 0.02$ ), and specificity decreased from 97.7% (91.9%–99.7%) to 96.0% (91.8%–98.4%) and to 92.5% (88.4%–95.5%;  $p = 0.04$ ) with McIsaac scores ≤2, 3 and ≥4, respectively (Table 2). RADT likelihood ratios and predictive values varied, but not significantly, by McIsaac score.

The RADT performance varied significantly by McIsaac criteria. For example, sensitivity was significantly lower in children ≥9 years old than in younger children (66.7% [47.2%–82.7%] vs. 89.0% [84.5%–92.6%],  $p < 0.01$ ). The negative predictive value varied, but not significantly, by McIsaac score or clinical criteria, ranging from 88.8% (80.8%–94.3%) for children without tonsillar swelling or exudate to 94.6% (90.0%–97.5%) for children without fever.

### Effect of Inoculum Size on RADT Sensitivity

Among 285 GAS-positive throat cultures, 243 (85% [80.6%–89.2%]) showed heavy inoculum. The distribution of inoculum size did not differ by office-to-laboratory delay (≤48 vs. >48 hr,  $p = 0.97$ ). RADT sensitivity varied widely and significantly between patients with light and heavy inoculum (40.5% [25.6%–56.7%] vs. 94.7% [91.0%–97.1%],  $p < 0.001$ ). Light inoculum was found in 65.8% (50.0%–81.6%) of false-negative RADT cases versus 6.9% (3.7%–10.1%) of true-positive cases ( $p < 0.001$ ).

Frequency of heavy inocula increased with increasing McIsaac score, although not significantly ( $p = 0.09$ , table 1). Heavy inoculum was significantly more frequent in children <9 years old than in older children (87.1% [82.9%–91.2%] vs. 70.0% [52.6%–87.4%],  $p < 0.05$ ).

### Combined Effect of Clinical Spectrum and Inoculum Size

The results of univariate binomial modeling for RADT sensitivity were almost identical to those of stratified analysis (Tables 2 and 3): age ≥9 years and presence of cough were significantly associated with decreased RADT sensitivity ( $p < 0.05$ ). After stratification on inoculum size, wide and statistically significant variations in RADT sensitivity were found in patients with light inoculum (e.g., 37.4% variation in sensitivity by age,  $p < 0.01$ ), but not widely (<6%) nor significantly in patients with heavy inoculum. Interactions were found between inoculum size and age ( $p = 0.03$ ), anterior cervical adenopathy ( $p = 0.01$ ), and tonsillar swelling or exudate ( $p = 0.07$ ).

Because of these multiple interactions and because of no convergence in the multivariate model after introducing interaction terms, 2 multivariate models of RADT sensitivity were fit by stratifying on inoculum size (Table 3). Results of multivariate modeling were close to those of univariate modeling. Light inoculum was associated with wide and statistically significant adjusted variation in RADT sensitivity (e.g., from 23.6% to 40.9%), but heavy inoculum was not widely (≤6%) or significantly associated.

**Table 2.** Spectrum effect analysis: performance of rapid antigen detection test by clinical spectrum.

	n (%)	GAS (%)	Sensitivity (95CI)	Specificity (95CI)	PLR (95CI)	NLR (95CI)	PPV (95CI)	NPV (95CI)
Mdisaac score								
≤2	115 (14)	24.3	75.0 (55.1–89.3)	97.7 (91.9–99.7)	32.6 (8.2–130.5)	0.23	91.3 (72.0–98.9)	92.4 (84.9–96.9)
3	262 (33)	34.0	84.3 (75.0–91.1)	96.0 (91.8–98.4)	20.8 (10.0–43.3)	0.23	91.5 (83.2–96.5)	92.2 (87.3–95.7)
≥4	408 (51)	41.2	89.9 (84.3–94.0)	92.5 (88.4–95.5)	12.0 (7.7–18.7)	0.11	89.3 (83.7–93.6)	92.9 (88.9–95.8)
Age (years)								
3–8	663 (84)	38.5	89.0 (84.5–92.6)	94.1 (91.4–96.2)	15.1 (10.2–22.4)	0.62	90.4 (86.1–93.8)	93.2 (90.3–95.4)
9–14	122 (16)	24.6	66.7 (47.2–82.7)	96.7 (90.8–99.3)	20.4 (6.5–64.0)	0.34	87.0 (66.4–97.2)	89.9 (82.2–95.0)
Lack of cough								
Yes	428 (55)	44.4	90.0 (84.8–93.9)	94.1 (90.3–96.7)	15.3 (9.2–25.5)	0.89	92.4 (87.6–95.8)	92.2 (88.1–95.2)
No	357 (45)	26.6	80.0 (70.5–87.5)	95.0 (91.7–97.3)	16.1 (9.4–27.7)	0.21	85.4 (76.3–92.0)	92.9 (89.2–95.7)
Anterior cervical adenopathy								
No	407 (52)	34.6	83.0 (75.7–88.8)	97.7 (95.2–99.2)	36.8 (16.6–81.4)	0.003	95.1 (89.7–98.2)	91.5 (87.7–94.5)
Yes	378 (48)	38.1	90.3 (84.2–94.6)	91.0 (86.6–94.4)	10.1 (6.7–15.2)	0.11	86.1 (79.5–91.2)	93.8 (89.9–96.6)
Tonsillar swelling or exudates								
No	157 (20)	42.0	83.3 (72.1–91.4)	95.6 (89.1–98.8)	19.0 (7.2–49.7)	0.71	93.2 (83.5–98.1)	88.8 (80.8–94.3)
Yes	628 (80)	34.9	87.7 (82.6–91.7)	94.4 (91.7–96.4)	15.6 (10.5–23.3)	0.13	89.3 (84.4–93.1)	93.5 (90.6–95.6)
Fever								
No	246 (31)	31.3	88.3 (79.0–94.5)	93.5 (88.7–96.7)	13.6 (7.6–24.2)	0.47	86.1 (76.5–92.8)	94.6 (90.0–97.5)
Yes	539 (69)	38.6	86.1 (80.6–90.5)	95.2 (92.3–97.2)	17.8 (11.0–28.8)	0.15	91.8 (87.0–95.2)	91.6 (88.1–94.3)
Total	785	36.3	86.7 (82.2–90.4)	94.6 (92.2–96.4)	16.1 (11.1–23.2)	0.14	90.1 (86.0–93.4)	92.6 (89.9–94.7)

Abbreviations: GAS, group A streptococcus; PLR, positive likelihood ratio; NLR, negative likelihood ratio; PPV, positive predictive value; NPV, negative predictive value.  
<sup>a</sup>χ<sup>2</sup> test; <sup>b</sup> Mantel-Haenszel's χ<sup>2</sup> test; <sup>c</sup> χ<sup>2</sup> test for trend.  
 doi:10.1371/journal.pone.0039085.t002

**Table 3.** Analysis of combined spectrum and inoculum size effect on sensitivity of rapid antigen detection test (RADT).

Variable	Univariate analysis <sup>a</sup>						Multivariate analysis <sup>a</sup>					
	All inocula		Light inoculum		Heavy inoculum		Light inoculum <sup>c</sup>		Heavy inoculum <sup>d</sup>		$p^b$	
	(n = 285)	$p$	(n = 42)	$p$	(n = 243)	$p$	(n = 42)	$p$	(n = 243)	$p$		
Age (years)												
3–8	<i>ref</i>	0.01	<i>ref</i>	<0.01	<i>ref</i>	0.49	0.03	<i>ref</i>	<0.001	<i>ref</i>	0.36	
9–14	–22.4 (–39.7; –5.1)		–37.4 (–64.1; –10.7)		–4.6 (–17.4; 8.3)			–40.9 (–57.9; –24.0)		–5.7 (–17.9; 6.4)		
Lack of cough												
Yes	<i>ref</i>	0.03	<i>ref</i>	0.09	<i>ref</i>	0.14	0.19	<i>ref</i>	<0.001	<i>ref</i>	0.08	
No	–10.0 (–19.1; –9.0)		–25.0 (–53.6; 3.6)		–5.2 (–12.1; 1.7)			–23.6 (–23.7; –23.5)		–6.0 (–12.5; 0.6)		
Anterior cervical adenopathy												
No	<i>ref</i>	0.07	<i>ref</i>	0.01	<i>ref</i>	0.88	0.01	<i>ref</i>	0.01	<i>ref</i>	0.37	
Yes	7.3 (–0.6; 15.2)		36.1 (7.7; 64.5)		–0.4 (–6.1; 5.2)			40.5 (16.7; 64.3)		2.0 (–2.3; 6.2)		
Tonsillar swelling or exudate												
No	<i>ref</i>	0.40	<i>ref</i>	0.12	<i>ref</i>	0.21	0.07	n/a <sup>e</sup>		<i>ref</i>	0.29	
Yes	4.3 (–5.6; 14.3)		–25.0 (–56.2; 6.2)		5.4 (–3.1; 13.9)					4.0 (–3.4; 11.4)		
Fever												
No	<i>ref</i>	0.61	<i>ref</i>	0.27	<i>ref</i>	0.77	0.26	n/a <sup>e</sup>		<i>ref</i>	0.41	
Yes	–2.3 (–10.8; 6.3)		–19.1 (–53.0; 14.8)		1.0 (–5.6; 7.6)					2.3 (–3.3; 8.0)		

Data RADT sensitivity difference (95% confidence interval) unless indicated.

Abbreviations: *ref*, reference; n/a, not applicable.

<sup>a</sup>Binomial modelling.

<sup>b</sup>Interaction test between clinical variables and inoculum size (Wald test).

<sup>c</sup>Model adjusted on age, lack of cough, anterior cervical adenopathy.

<sup>d</sup>Model adjusted on age, lack of cough, anterior cervical adenopathy, tonsillar swelling or exudate and fever.

<sup>e</sup>Not included in the model because of non-convergence.

doi:10.1371/journal.pone.0039085.t003

### Performance of the McIsaac-score-based Decision Rule

Among 785 children, 522 (66%) would have received antibiotics according to the decision rule. The rule had 98.9% (97.0%–99.8%) sensitivity and 52.0% (47.5%–56.5%) specificity.

### Discussion

Accuracy of a diagnostic test can vary across patient subgroups within a population, a phenomenon referred to as spectrum effect (or spectrum bias) [10–12]. In assessing the relation of clinical spectrum and inoculum size with RADT performance for GAS pharyngitis, we confirm that the spectrum effect has an impact on RADT sensitivity and report for the first time that it also affects the specificity, likelihood ratios and positive predictive value of the test. However, the spectrum effect had no impact on the negative predictive value of the RADT, which remained stable at about 92%, regardless of McIsaac score or its components. Clinicians do not want to expose their patients to GAS suppurative and nonsuppurative complications by withholding antibiotic treatment and therefore need a diagnostic test with high, stable negative predictive value. In this population with a GAS prevalence of 36%, a patient with a negative RADT result had a probability of GAS-positive throat culture close to 8% regardless of clinical features.

This study confirms the important effect of inoculum size on RADT sensitivity, already described in other studies, and reports for the first time an association of clinical spectrum and inoculum size. Our data suggest that having more streptococci in the throat

(greater inoculum size) might be associated with more intense symptoms (higher McIsaac score) ( $p = 0.09$ ), although we might not have had enough statistical power to validate this hypothesis. However, because we were not able to differentiate GAS carriers from truly infected patients (i.e., by assessment of streptococcal antibody response) [16,28], we cannot conclude on whether these patients with low McIsaac scores and light inocula were more likely to be GAS carriers rather than truly infected patients.

We chose to investigate the McIsaac score in 3 categories rather than compare the accuracy of the RADT above and below a defined breakpoint [15] because this corresponds to the original aims of the score – to stratify patients into 3 levels of risk and to suggest a subsequent course of action (low-risk: no test, no antibiotics; intermediate risk: culture or RADT, antibiotics only for positive results; high-risk: empirical antibiotic treatment without testing) [20–22]. According to our results, the McIsaac score alone does not allow for ruling out or affirming the diagnosis of GAS pharyngitis because the reported GAS prevalence ranged from 25% (too-high level to rule out) to 44% (too-low level to affirm) in children with a score of 1 and 5, respectively. Moreover, the McIsaac-score-based decision rule had only 52.0% (47.5%–56.5%) specificity, which does not seem consistent with the current need to reduce antibiotic consumption to contain antimicrobial resistance.

One of the strengths of this study is its prospective, multicenter design, which limits the selection bias potentially present in other studies [13,14]. All included children underwent RADT and laboratory culture, which excluded the possible indication and

verification biases of other studies [13,14]. The results of our study are close to those reported in the literature. GAS prevalence was 36.3% (32.9%–39.8%), which is close to that from a recent meta-analysis (37%, 32%–43%) [1]. The overall RADT sensitivity was 86.7% (82.2%–90.4%), which is close to that from another recent meta-analysis (85%, 84%–87%) [9]. A limitation is that we used a modified adenopathy criterion for the analyses. However, univariate and multivariate modeling results and the performance of the McIsaac-score-based decision rule were stable in a sensitivity analysis involving the original McIsaac adenopathy criterion instead of our modified criterion.

Another limitation to our study is that some throat swabs were plated >48 hr after collection. This delay had no significant effect on the distribution of GAS inocula, but prolonged or inadequate shipping conditions could have resulted in the loss of viability of GAS if the original swab had only small numbers of bacteria. Therefore, some swabs with light inocula could have been falsely read as negative on throat culture. Because RADTs are more likely to give negative results with light inocula, these swabs might have led to a systematic decrease in number of RADT false-negative results and systematic increase in number of RADT true-negative results, with over-estimation of the negative predictive value of the RADT as a result. We can assume that this bias occurred at random because shipping and all bacteriologic investigations were blinded to clinical data and RADT results – non-differential measurement bias. Such a bias usually leads to loss of power, and our results regarding the stability of the negative predictive value of the RADT should be considered with caution until they are confirmed with a larger sample of patients.

The American Academy of Pediatrics advises that “*Children with manifestations highly suggestive of viral infection [...] generally should not be tested for GAS infection*” [7]. Such guidelines cannot be formally validated because the criteria proposed to determine patients who should be tested are not integrated in an explicit decision rule and remain subject to personal interpretation. Because of GAS carriage and because GAS complications have become uncommon in North America and Western Europe [29], we advocate that the target sensitivity of diagnostic strategies for GAS

pharyngitis should not be 100%. Moreover, discussions should not focus on sensitivity alone. We showed that the negative predictive value of a test can be stable across patient subgroups because variations in prevalence can be balanced by concomitant variations in sensitivity and specificity. Therefore, clinicians who want to rule out GAS can rely on negative RADT results regardless of clinical features if they accept that about 10% of children with negative RADT results will have a positive throat culture. However, these results should be considered with caution because they depend highly on the study population, the training of clinical and laboratory personnel and the microbiological devices and protocols used. Moreover, available data suggest that false-negative RADT results are as likely to occur in truly infected patients as in GAS carriers [16,28]. Therefore, such a policy would be more acceptable in populations with very low incidence of suppurative and non-suppurative complications of GAS infection [30]. The McIsaac-score-based decision rule is insufficiently accurate for children, and efforts are needed to develop more specific selective testing strategies for GAS pharyngitis in children.

## Acknowledgments

We thank A. Liboz (Department of Microbiology, Robert Debré Hospital), M. Boucherat, MD (database design), F. de La Rocque, MD (study conception), I. Ramay, D. Menguy, S. Tortorelli and M. de Pereira (ACTIV), Pr A. Morabia, MD, PhD (spectrum effect analysis methodological support) and all the physician investigators who participated in the RADT GAS study: M. Benani, MD, F. Corrad, MD, P. Deberdt, MD, A. Elbez, MD, M. Goldrey, MD, J. Gosselin, MD, M. Koskas, MD, P. Martin, MD, A.S. Michot, MD, N. Panis, MD, D. Qutob, MD, C. Romain, MD, O. Romain, MD, C. Schlemmer, MD.

## Author Contributions

Conceived and designed the experiments: RC EB AW FT MC. Performed the experiments: RC AW FT. Analyzed the data: JFC CL RC MC. Contributed reagents/materials/analysis tools: PB EB. Wrote the paper: JFC MC CL EB RC. Approved the final version of the manuscript: JFC MC CL PB FT AW EB RC.

## References

1. Shaikh N, Leonard E, Martin JM (2010) Prevalence of streptococcal pharyngitis and streptococcal carriage in children: a meta-analysis. *Pediatrics* 126: e557–564.
2. Shaikh N, Swaminathan N, Hooper EG (2011) Accuracy and Precision of the Signs and Symptoms of Streptococcal Pharyngitis in Children: A Systematic Review. *J Pediatr*.
3. Matthys J, De Meyere M, van Driel ML, De Sutter A (2007) Differences among international pharyngitis guidelines: not just academic. *Ann Fam Med* 5: 436–443.
4. Bisno AL, Gerber MA, Gwaltney JM, Jr., Kaplan EL, Schwartz RH (2002) Practice guidelines for the diagnosis and management of group A streptococcal pharyngitis. *Infectious Diseases Society of America. Clin Infect Dis* 35: 113–125.
5. Gerber MA (2005) Diagnosis and treatment of pharyngitis in children. *Pediatr Clin North Am* 52: 729–747, vi.
6. Choby BA (2009) Diagnosis and treatment of streptococcal pharyngitis. *Am Fam Physician* 79: 383–390.
7. American Academy of Pediatrics (2009) Group A Streptococcal Infections. In: Pickering L, Baker C, Long S, McMillan J, editors. *Red Book: 2009 Report of the Committee on Infectious Disease*. 28th ed. Elk Grove Village, IL: American Academy of Pediatrics. 616–628.
8. Gerber MA, Shulman ST (2004) Rapid diagnosis of pharyngitis caused by group A streptococci. *Clin Microbiol Rev* 17: 571–580.
9. Ruiz-Aragon J, Rodriguez Lopez R, Molina Linde JM (2010) Evaluation of rapid methods for detecting *Streptococcus pyogenes*. systematic review and meta-analysis. *An Pediatr* 72: 391–402.
10. Mulherin SA, Miller WC (2002) Spectrum bias or spectrum effect? Subgroup variation in diagnostic test evaluation. *Ann Intern Med* 137: 598–602.
11. Goehring C, Perrier A, Morabia A (2004) Spectrum bias: a quantitative and graphical analysis of the variability of medical diagnostic test performance. *Stat Med* 23: 125–135.
12. Elie C, Coste J (2008) A methodological framework to distinguish spectrum effects from spectrum biases and to assess diagnostic and screening test accuracy for patient populations: application to the Papanicolaou cervical cancer smear test. *BMC Med Res Methodol* 8: 7.
13. Hall MC, Kieke B, Gonzales R, Belongia EA (2004) Spectrum bias of a rapid antigen detection test for group A beta-hemolytic streptococcal pharyngitis in a pediatric population. *Pediatrics* 114: 182–186.
14. Edmonson MB, Farwell KR (2005) Relationship between the clinical likelihood of group A streptococcal pharyngitis and the sensitivity of a rapid antigen-detection test in a pediatric practice. *Pediatrics* 115: 280–285.
15. Tanz RR, Gerber MA, Kabat W, Rippe J, Seshadri R, et al. (2009) Performance of a rapid antigen-detection test and throat culture in community pediatric offices: implications for management of pharyngitis. *Pediatrics* 123: 437–444.
16. Gerber MA, Randolph MF, Chanatry J, Wright LL, DeMeo KK, et al. (1986) Antigen detection test for streptococcal pharyngitis: evaluation of sensitivity with respect to true infections. *J Pediatr* 108: 654–658.
17. Kuhn S, Davies HD, Katzko G, Jadavji T, Church DL (1999) Evaluation of the Strep A OIA assay versus culture methods: ability to detect different quantities of group A *Streptococcus*. *Diagn Microbiol Infect Dis* 34: 275–280.
18. Kurtz B, Kurtz M, Roe M, Todd J (2000) Importance of inoculum size and sampling effect in rapid antigen detection for diagnosis of *Streptococcus pyogenes* pharyngitis. *J Clin Microbiol* 38: 279–281.
19. Charlier-Bret N, Boucher B, Poyart C, Quesne G, Bingen E, et al. (2004) Rapid antigen detection tests for diagnosis of group A streptococcal pharyngitis: comparative evaluation of sensitivity and practicability of 16 in vitro diagnostics medical devices performed in July 2002 by the French health products safety agency (Afsaps) as part of its market control mission. *Pathol Biol (Paris)* 52: 438–443.
20. McIsaac WJ, White D, Tannenbaum D, Low DE (1998) A clinical score to reduce unnecessary antibiotic use in patients with sore throat. *CMAJ* 158: 75–83.

21. McIsaac WJ, Goel V, To T, Low DE (2000) The validity of a sore throat score in family practice. *CMAJ* 163: 811–815.
22. McIsaac WJ, Kellner JD, Aufricht P, Vanjaka A, Low DE (2004) Empirical validation of guidelines for the management of pharyngitis in children and adults. *JAMA* 291: 1587–1595.
23. Fischer Walker CL, Rimoin AW, Hamza HS, Steinhoff MC (2006) Comparison of clinical prediction rules for management of pharyngitis in settings with limited resources. *J Pediatr* 149: 64–71.
24. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, et al. (2003) The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Ann Intern Med* 138: W1–12.
25. Cohen R, Levy C, Ovetchkine P, Boucherat M, Weil-Olivier C, et al. (2004) Evaluation of streptococcal clinical scores, rapid antigen detection tests and cultures for childhood pharyngitis. *Eur J Pediatr* 163: 281–282.
26. Dimatteo LA, Lowenstein SR, Brimhall B, Reiquam W, Gonzales R (2001) The relationship between the clinical features of pharyngitis and the sensitivity of a rapid antigen test: evidence of spectrum bias. *Ann Emerg Med* 38: 648–652.
27. Wacholder S (1986) Binomial regression in GLIM: estimating risk ratios and risk differences. *Am J Epidemiol* 123: 174–184.
28. Johnson DR, Kurlan R, Leckman J, Kaplan EL (2010) The human immune response to streptococcal extracellular antigens: clinical, diagnostic, and potential pathogenetic implications. *Clin Infect Dis* 50: 481–490.
29. Carapetis JR, Steer AC, Mulholland EK, Weber M (2005) The global burden of group A streptococcal diseases. *Lancet Infect Dis* 5: 685–694.
30. Pelucchi C, Grigoryan L, Galeone C, Esposito S, Huovinen P, et al. (2012) Guideline for the management of acute sore throat. *Clin Microbiol Infect* 18 Suppl 1: 1–28.

---

ANNEXE 2. ARTICLE CORRESPONDANT AU CHAPITRE 2

# Effect of clinical spectrum, inoculum size and physician characteristics on sensitivity of a rapid antigen detection test for group A streptococcal pharyngitis

J. F. Cohen · M. Chalumeau · C. Levy · P. Bidet ·  
M. Benani · M. Koskas · E. Bingen · R. Cohen

Received: 23 October 2012 / Accepted: 18 December 2012  
© Springer-Verlag Berlin Heidelberg 2013

**Abstract** We aimed to assess the independent effect of clinical spectrum, bacterial inoculum size and physician characteristics on the sensitivity of a rapid antigen detection test (RADT) for group A streptococcus (GAS) in children. Double throat swabs were collected from 1,482 children with pharyngitis and 294 asymptomatic children in a French prospective, office-based, multicenter ( $n=17$ ) study, from October 2009 to May 2011. Patient- and physician-level factors potentially affecting RADT sensitivity were studied by univariate and multivariate multilevel analysis, with laboratory throat culture as the reference test. In children with

pharyngitis and asymptomatic children, the prevalence of GAS was 38 % (95 % confidence interval 36–41 %) and 11 % (7–14 %), respectively. Overall, RADT sensitivity was 87 % (84–90 %). On stratified and multivariate multilevel analysis, RADT sensitivity was higher for children with pharyngitis than asymptomatic children (89 % vs. 41 %), children <9 than  $\geq 9$  years old (88 % vs. 79 %) and those with heavy than light inoculum (94 % vs. 53 %). RADT sensitivity was influenced by the physician performing the test (range 56–96 %,  $p=0.01$ ) and was higher for physicians with hospital-based clinical activity in addition to office-based practice (adjusted odds ratio 3.4 [95 % confidence interval 1.9–6.3],  $p<0.001$ ); inter-physician variations in RADT sensitivity were largely explained by this variable (proportional change in variance >99 %). The sensitivity of the RADT is independently affected by patient- and physician-level factors. Physicians who base their diagnosis of GAS pharyngitis on the results of a RADT alone should consider diagnostic accuracy monitoring and adequate training when needed.

Prof. Edouard Bingen, \**In memoriam*

J. F. Cohen (✉) · M. Chalumeau  
INSERM, UMR S953, Epidemiological Research Unit on  
Perinatal Health and Women's and Children's Health, Maternité  
de Port Royal, 6e étage, 53 avenue de l'Observatoire,  
75014 Paris, France  
e-mail: jeremie.cohen@inserm.fr

J. F. Cohen · M. Chalumeau  
Department of Pediatrics, Necker-Enfants-Malades Hospital,  
Assistance Publique-Hôpitaux de Paris, Univ Paris Descartes,  
Sorbonne Paris Cité, Paris, France

C. Levy · M. Benani · M. Koskas · R. Cohen  
Association Clinique et Thérapeutique  
du Val-de-Marne (ACTIV),  
Saint-Maur-des-Fossés, France

C. Levy  
Centre de Recherche Clinique (CRC), Centre Hospitalier  
Intercommunal de Créteil, Créteil, France

P. Bidet · E. Bingen  
Department of Microbiology, Robert Debré Hospital,  
Assistance Publique-Hôpitaux de Paris, Univ Paris Diderot,  
Sorbonne Paris Cité, Paris, France

R. Cohen  
Department of Microbiology, Centre Hospitalier  
Intercommunal de Créteil, Créteil, France

## Introduction

Group A streptococcus (GAS) is found in 20–40 % of childhood pharyngitis cases; the remaining cases are considered mainly of viral origin [1]. The reference test for the diagnosis of GAS pharyngitis is throat culture on a blood agar plate (BAP) in a microbiology laboratory for 48 h. Rapid antigen detection tests (RADTs) can provide immediate results and are generally considered highly specific as compared to laboratory culture (about 95 %) [2]. However, the American Academy of Pediatrics and the Infectious Diseases Society of America recommend the confirmation of negative RADT results by a throat culture because the sensitivity of RADTs is estimated to be insufficient, at about 85 % [3, 4]. Furthermore, RADT sensitivity is subject to variations by clinical spectrum of the disease (spectrum

effect) [5–10], bacterial inoculum size (inoculum size effect) [10–13] and the physician who performed the test (physician effect) [14, 15].

About 10 % of school-age children are asymptomatic pharyngeal GAS carriers [1], for which antibiotic treatment is usually not required [16]. Studying the sensitivity of RADTs in asymptomatic carriers might be important because the sensitivity might approach that in children with GAS carriage and intercurrent viral pharyngitis. Previous studies that evaluated a RADT in asymptomatic children [17–20] reported heterogeneous sensitivities (from 20 % to 100 %) and were limited by their use of an obsolete RADT technique (latex agglutination) [17], a non-consensual reference test (reading of cultures after 24 h) [18, 19], and a low GAS carriage rate (1.6 %), which suggested selection bias [20].

To date, several studies have evaluated the effects of clinical spectrum, inoculum size and person performing the test on the sensitivity of the RADT, but no study analyzed these factors concomitantly. Furthermore, no study included asymptomatic carriers, nor analyzed physician-level factors that might explain the physician effect. We assessed the independent effect of patient- and physician-level factors on the sensitivity of a RADT.

## Materials and methods

The study protocol was approved by the Ile-de-France XI institutional review board (n°09016). Parents and children gave their informed approval for participation before inclusion. The STARD statement was followed for reporting [21].

### Patients

Eligible patients were 3–14 years old who were evaluated by their pediatrician and who did not receive antibiotics for 7 days before inclusion. Recruitment of a first case-cohort of symptomatic children was based on a clinical diagnosis of acute pharyngitis at the time of consultation (i.e., clinical signs of inflammation of the pharynx and/or tonsils). We recruited a second group of asymptomatic children with no symptoms at the time of consultation (i.e., well-child visits for evaluation of physical and cognitive development and/or vaccination), nor pharyngitis in the previous month.

### Sample size

In children with pharyngitis, the sample size was estimated so that the 95 % confidence interval (95CI) for RADT sensitivity would be a  $\pm 5\%$  estimation. Assuming sensitivity to vary between 75 % and 95 % and a GAS prevalence of

35 %, a sample of 714 children with pharyngitis was needed. In asymptomatic children, sample size was estimated so that the 95% CI for the frequency of GAS carriage would be a  $\pm 4\%$  estimation. Assuming a GAS carriage frequency of 10 %, a sample of 250 asymptomatic children was needed. Enrollment was initially planned from October 1, 2009 to May 31, 2010 (year 1), but the study was extended for both groups from October 1, 2010 to May 31, 2011 (year 2) because of difficulties in sampling asymptomatic children.

### Throat swabs

Throat samples were obtained by use of a contiguous double-swab collection-transportation system (Venturi Transystem Amies agar, COPAN Diagnostics, Corona, CA, USA). Throat samples were taken as recommended by the manufacturer of the RADT, that is, by use of a tongue depressor to avoid contact with any area of the mouth and by vigorously rubbing the surface of the tonsils, the posterior pharynx and any area of inflammation of the pharynx. The RADT (StreptAtest, Dectrapharm, Strasbourg, France) was performed immediately with swab #1 collected in the pediatrician's office. Swab #2 was held at ambient temperature and sent to arrive within 72 h to the hospital laboratory by an express messenger service. On receipt, swab #2 was used to perform a throat culture. During year 1, the swab was directly inoculated onto a trypticase soy agar plate with 5 % sheep blood [10], and during year 2, the swab was moistened with 1 mL sterile distilled water, and one drop of the fluid expressed from the swab was streaked on a BAP containing colistin and nalidixic acid (Columbia CNA+5 % sheep blood, Biomérieux, Lyon, France). The plates were incubated anaerobically at 37 °C and read at 24 and 48 h.  $\beta$ -hemolytic colonies were investigated by latex agglutination (Prolex, Pro-Lab Diagnostics, Richmond Hill, ON, Canada). Inoculum size was estimated as follows: light,  $\leq 50$  GAS colonies per plate; heavy,  $> 50$  GAS colonies. Microbiologists were blinded to individual clinical data and RADT results.

### Office-based pediatricians and survey tool

The 17 participating pediatricians were part of a French research-and-teaching network (Association Clinique et Thérapeutique du Val-de-Marne [ACTIV]) [22]. After completion of the clinical study, all participating pediatricians were invited to answer an online survey. The questionnaire was developed after review of previous publications aiming to explore the influence of contextual and organizational factors on physician practices [23–25]. The questionnaire was designed as a four-domain survey tool including questions about physician general characteristics (age [continuous], gender, year of medical graduation [continuous]),

practice type and setting (number of physicians in the office [solo/2-person practice vs. >2 clinicians], hours of weekly clinical activity in the office [continuous], estimated proportion of patients covered by the French universal health insurance [continuous], whether the physician had a hospital-based clinical activity in addition to office-based practice [yes/no]), physician habits in performing the RADT (person collecting the throat sample and reading the result of the RADT [physician vs. office staff], site of throat sampling [1 tonsil vs. both tonsils], following manufacturer's instructions in terms of number of drops for each chemical reagent [4 drops vs. other] and use of a timer to control chemical reactions [yes/no]), and knowledge and beliefs of pharyngitis and GAS disease (estimated GAS prevalence in childhood pharyngitis [<20 % vs. 20–40 % vs. >40 %] and estimated incidence of acute rheumatic fever in France in children 3–14 years old [continuous]). The final version of the survey tool consisted of 14 questions organized in a random fashion. Physician data without identifier variables were entered into the database and linked to patient data.

#### Statistical analysis

Hospital laboratory throat culture was considered the reference test. First, the general characteristics of the study population, the frequency of GAS recovery in children with pharyngitis and asymptomatic children and the overall diagnostic accuracy of the RADT (sensitivity and specificity) were described. Second, the characteristics of the pediatricians who participated in the study were described. Third, we used univariate analysis with chi-square test to examine patient- and physician-level characteristics that may affect RADT sensitivity. Fourth, we used a multilevel logistic model to account for the hierarchical structure of data, with patients (level 1) nested within physician offices (level 2) [26, 27]. Restricting the analysis to patients with a positive throat culture for GAS and using the result of the RADT as the outcome in the model allowed us to study the independent effect of patient- and physician-level variables on the sensitivity of the RADT. The model estimated the odds ratio (OR) of having a positive RADT result with a positive throat culture result for GAS. Patient-level variables were clinical status (pharyngitis vs. asymptomatic), age dichotomized into 2 categories of equal range (<9 year vs. ≥9 year) and inoculum size (light or heavy GAS growth). Physician-level variables were those from the questionnaire described above, with continuous variables dichotomized at their median.

First, an empty model that included only the random effect was adjusted (model 0) to estimate the baseline physician-level variance ( $var^{(0)}$ ), i.e., the physician effect. Next, we adjusted a multivariate multilevel model that included patient-level covariates (model 1). It allowed us to assess the association between RADT sensitivity and patient-level

variables and to calculate the residual physician-level variance after adjustment for patient-level variables ( $var^{(1)}$ ). We used the proportional change in the variance (PCV =  $[var^{(0)} - var^{(1)}]/var^{(0)}$ ) to assess the extent to which the physician effect may be explained by differences in the distribution of patient-level characteristics. Then, we adjusted multivariate multilevel models that included physician-level covariates (models 2–5). Physician-level characteristics that were associated with RADT sensitivity with  $p \leq 0.2$  on univariate analysis were entered stepwise in multilevel models. We also calculated the PCV for each physician-level variable. Data were double entered into 4D software v6.4, and statistical analysis involved use of Stata/SE 12 (StataCorp, College Station, TX, USA).

## Results

### Patient characteristics, GAS prevalence and overall RADT accuracy

In total, 1,781 children met the inclusion criteria. Two patients with uninterpretable RADT results and three with a lost throat swab were secondarily excluded. Therefore, analysis involved 1,776 children (790 girls, 44 %), including 1,482 with pharyngitis and 294 asymptomatic children. The mean (SD) age was 6.1 (2.5) and 7.0 (2.7) years for children with pharyngitis and asymptomatic children, respectively ( $p < 0.001$ ). The prevalence of GAS in children with pharyngitis was 38 % (95 CI 36–41 %), and the frequency of GAS carriage was 11 % (7–14 %) in asymptomatic children ( $p < 0.001$ ). Overall, the sensitivity and specificity of the RADT was 87 % (84–90 %) and 93 % (91–94 %), respectively.

### Physician characteristics

The mean (SD) age of physicians was 53.4 (7.1) years and 8 (47 %) were women (Table 1). In all, ten physicians (59 %) had a hospital-based clinical activity in addition to their office-based practice. All but one physician declared collecting throat samples themselves and all physicians declared reading the result of the RADT themselves.

### Effect of clinical spectrum and inoculum size

On univariate analysis, the sensitivity of the RADT was higher in children with pharyngitis than asymptomatic children (89 % [95% CI 87–92 %] vs. 41 % [24–59 %],  $p < 0.001$ ; Table 2). RADT sensitivity was higher but not significantly in children <9 than ≥9 years (88 % [85–91 %] vs. 79 % [70–89 %],  $p = 0.053$ ). Among 602 GAS-positive throat cultures, 501 (83 % [80–86 %]) showed heavy inoculum. Heavy inoculum was less frequent in asymptomatic

**Table 1** Characteristics of physicians who participated in the study ( $n=17$ )

Characteristics	Value
Age, years, mean (SD)	53.4 (7.1)
Women	8 (47)
Year of medical graduation, median (range)	1987 (1979–2002)
Practice type and setting	
Physicians in the office	
Solo/2-person practice	6 (35)
Group practice (>2 clinicians)	11 (65)
Weekly clinical activity in the office, hours (median, range)	37 (20–55)
Estimated proportion (%) of patients covered by the French universal health insurance, median (range)	5 (1–15)
Hospital-based clinical activity	10 (59)
Physician's habits for performing the RADT	
Site of throat sampling	
One tonsil	5 (29)
Both tonsils	12 (71)
Use of the correct quantity of each reagent for the RADT <sup>a</sup>	13 (76)
Use of a timer to control time of reactions for the RADT <sup>a</sup>	6 (35)
Knowledge and beliefs of GAS disease	
Estimated GAS prevalence in childhood pharyngitis	
<20 %	2 (12)
20–40 %	12 (71)
>40 %	3 (18)
Estimated incidence of ARF in France in children aged 3–14 years, cases/year, mean (SD)	2.6 (3.7)

Data are no. (%) unless indicated

RADT rapid antigen detection test, GAS group A streptococcus, ARF acute rheumatic fever

<sup>a</sup> Manufacturer's instructions are as follows: (1) place four drops of reagent A and four drops of reagent B in the tube; (2) add the throat swab in the tube and turn the swab ten times; (3) leave the swab in the tube for 1 min; (4) press the swab against the side of the tube; (5) discard the swab; (6) place the test strip into the tube; (7) leave the strip in the tube and read the result at 5 min

carriers than in children with pharyngitis (34 % [17–52 %] vs. 86 % [83–89 %],  $p<0.001$ ). On univariate analysis, RADT sensitivity increased with inoculum size (53 % [44–63 %] vs. 94 % [91–96 %],  $p<0.001$ ; Table 2). Sensitivity was higher during year 2 than year 1 (92 % [89–95 %] vs. 82 % [78–86 %],  $p<0.001$ ). On multivariate analysis, RADT sensitivity was higher for children with pharyngitis than asymptomatic children (adjusted odds ratio [aOR] 5.0 [95% CI 1.9–13.0],  $p=0.001$ ), children <9 than  $\geq 9$  years (aOR 2.3 [1.1–5.1],  $p=0.04$ ) and children with heavy than light inocula (aOR 10.2 [5.7–18.5],  $p<0.001$ ; model 1; Table 3). These associations were stable after adjusting for physician-level variables (Table 3).

**Table 2** Sensitivity of the RADT by patient- and physician-level characteristics: univariate analysis

Characteristics	No. <sup>a</sup> (%)	Sensitivity, % (95CI)	$p$
Patient-level characteristics			
Clinical status			
Asymptomatic	32 (5)	41 (24–59)	<0.001
Pharyngitis	570 (95)	89 (87–92)	
Age, years			
3–8	534 (89)	88 (85–91)	0.053
9–14	68 (11)	79 (70–89)	
Inoculum <sup>b</sup>			
Light	101 (17)	53 (44–63)	<0.001
Heavy	501 (83)	94 (91–96)	
Physician-level characteristics			
Estimated proportion of patients covered by the French universal health insurance, %			
$\leq 5$	10 (59)	89 (86–92)	0.03
>5	7 (41)	83 (78–88)	
Hospital-based clinical activity			
No	7 (41)	78 (72–84)	<0.001
Yes	10 (59)	91 (88–94)	
Year of medical graduation			
$\leq 1987$	10 (59)	89 (85–92)	0.10
>1987	7 (41)	84 (79–89)	
Use of the correct quantity of each reagent for the RADT <sup>c</sup>			
No	4 (24)	81 (70–91)	0.15
Yes	13 (76)	88 (85–90)	

95CI 95 % confidence interval

<sup>a</sup> Number of patients ( $n=602$ ) or physicians ( $n=17$ )

<sup>b</sup> Light,  $\leq 50$  GAS colonies per plate; heavy,  $>50$  GAS colonies

<sup>c</sup> Manufacturer's instructions are to use four drops of reagent A and four drops of reagent B

### Physician effect

RADT sensitivity was significantly affected by the physician performing the test (range 56–96 %,  $p=0.01$ ). On univariate analysis, RADT sensitivity was higher for physicians with  $\leq 5$  % of their patients covered by the French universal health insurance system (89 % [86–92 %] vs. 83 % [78–88 %],  $p=0.03$ ; Table 2) and those with hospital-based clinical activity in addition to their office-based practice (91 % [88–94 %] vs. 78 % [72–84 %],  $p<0.001$ ). RADT sensitivity was higher but not significantly for physicians who had more clinical years in practice (89 % [85–92 %] vs. 84 % [79–89 %],  $p=0.10$ ) and those who declared they used the correct number of drops of each chemical reagent needed for the RADT (88 % [85–90 %] vs. 81 % [70–91 %],  $p=0.15$ ). On multilevel analysis, the empty model confirmed a significant physician effect (Model 0,  $var^{(0)}=0.19$ ,  $p=0.03$ ). This physician effect was not explained by differences in distribution of patient-level variables (Model 1,  $var^{(1)}=0.21$ ;

**Table 3** Sensitivity of the RADT by patient- and physician-level characteristics: multivariate multilevel analysis

Characteristics	No. <sup>a</sup> (%)	Model 1		Model 2		Model 3		Model 4		Model 5	
		aOR (95CI)	<i>p</i>								
Patient-level characteristics											
Clinical status											
Asymptomatic	32 (5)	1	0.001	1	0.001	1	0.001	1	0.001	1	0.001
Pharyngitis	570 (95)	5.0 (1.9–13.0)		4.8 (1.8–12.4)		4.5 (1.8–11.4)		5.0 (1.9–12.9)		5.1 (2.0–13.1)	
Age, years											
3–8	534 (89)	2.3 (1.1–5.1)	0.04	2.4 (1.1–5.4)	0.03	2.5 (1.1–5.5)	0.02	2.3 (1.1–5.2)	0.04	2.3 (1.1–5.1)	0.04
9–14	68 (11)	1		1		1		1		1	
Inoculum <sup>b</sup>											
Light	101 (17)	1	<0.001	1	<0.001	1	<0.001	1	<0.001	1	<0.001
Heavy	501 (83)	10.2 (5.7–18.5)		10.3 (5.7–18.6)		11.0 (6.1–20.0)		10.1 (5.6–18.3)		10.1 (5.6–18.3)	
Physician-level characteristics											
Estimated proportion of patients covered by the French universal health insurance, %											
≤5	10 (59)	–	–	1.5 (0.7–3.3)	0.35	–	–	–	–	–	–
>5	7 (41)	–	–	1	–	–	–	–	–	–	–
Hospital-based clinical activity											
No	7 (41)	–	–	–	–	1	<0.001	–	–	–	–
Yes	10 (59)	–	–	–	–	3.4 (1.9–6.3)	–	–	–	–	–
Year of medical graduation											
≤1987	10 (59)	–	–	–	–	–	–	1.2 (0.5–2.6)	0.67	–	–
>1987	7 (41)	–	–	–	–	–	–	1	–	–	–
Use of the correct quantity of each reagent for the RADT <sup>c</sup>											
No	4 (24)	–	–	–	–	–	–	–	–	1	0.75
Yes	13 (76)	–	–	–	–	–	–	–	–	1.2 (0.4–3.2)	–
Residual physician-level variance		0.21		0.21		<0.001		0.18		0.21	
PCV (%) <sup>*</sup>		n/a**		n/a**		>99 %		14 %		n/a**	

All models adjusted for year of inclusion (year 1 vs. year 2)

aOR adjusted odds ratio, 95CI 95 % confidence interval, PCV proportional change in variance, n/a not applicable

<sup>a</sup> Number of patients (*n*=602) or physicians (*n*=17)

<sup>b</sup> Light, ≤50 GAS colonies per plate; heavy, >50 GAS colonies

<sup>c</sup> Manufacturer's instructions are to use four drops of reagent A and four drops of reagent B

\* PCV is calculated on the basis of the physician-level variance of model 0 ( $var^{(0)}=0.19$ ,  $p=0.03$ ) for model 1 and physician-level variance of model 1 ( $var^{(1)}=0.21$ ) for models 2–5

\*\* Not applicable because of equal or increasing physician-level variance between the two models compared

Table 3). Hospital-based clinical activity was the only physician-level variable significantly associated with RADT sensitivity (aOR 3.4 [1.9–6.3],  $p<0.001$ ), and it accounted for 99 % of the physician effect (Model 3,  $var^{(3)}<0.001$ , PCV >99 %).

## Discussion

The sensitivity of a RADT is a key determinant defining its role in diagnostic strategies for pharyngitis. Several authors have suggested that the RADT sensitivity can vary by patient-level characteristics, as well as by the person

performing the test [7–15]. To our knowledge, this is the first study to examine the independent effect of patient- and physician-level factors on the sensitivity of a RADT. Our results suggest that physician-level factors may affect the sensitivity of the RADT independently of patient clinical spectrum and bacterial inoculum size.

The results from this study confirm the effect of patient clinical spectrum on the sensitivity of the RADT, with increased sensitivity for children with pharyngitis than asymptomatic carriers. The variation in sensitivity of the RADT by clinical status may be explained by locoregional changes due to infection and inflammation, such as biochemical changes, bacterial biofilm formation, increased

antigen exposure, internalization phenomenon, changes in antigenic structures and differences in the distribution of GAS genotype (*emm* gene) in children with pharyngitis and asymptomatic carriers [28].

Our results also confirm the importance of inoculum size and physician effect on the sensitivity of the RADT. In general, inoculum size effect and the physician effect are thought to be linked through sampling variation, with the most experienced operators more frequently obtaining heavy inocula, and heavy inocula being associated with higher RADT sensitivity. However, our results show that RADT sensitivity may vary significantly depending on the physician who performed the test, independent of inoculum size. We found wide variation in RADT sensitivity by physician (range 56–96 %), which did not decrease when taking inoculum size into account on multivariate multilevel analysis (model 1, Table 3).

Among several physician-level factors associated with variations in RADT sensitivity on univariate analysis, having a hospital-based clinical activity in addition to office-based practice was the only variable remaining significant on multivariate multilevel analysis and explained most of the physician effect (PCV >99 %). However, we do not believe that having a hospital-based clinical activity is likely to increase the sensitivity of the RADT *per se*, but rather, this variable might indirectly capture the effect of other physician characteristics that we did not assess. As well, previously, the appropriate management of childhood pharyngitis was found to depend on physician characteristics such as specialty (family physicians vs. pediatricians) and type of practice (solo/2-person vs. group practice and rural vs. urban location) [25].

Our study has several limitations. First, the study is based on a small sample of French pediatricians ( $n=17$ ), all of whom are involved in a research-and-teaching network (ACTIV) and practice in urban areas. Therefore, our data might not be representative of the practices of office-based physicians, and our results need to be confirmed in larger samples of physicians and in different practice types and settings.

Second, although our results confirm that heavy inocula are less frequent in asymptomatic carriers than in children with pharyngitis (34 % vs. 86 %) [29], we were not able to investigate whether RADT false-negative results would be associated with GAS carriage in children with pharyngitis [11]. We included asymptomatic children in an attempt to approach the sensitivity of the RADT in GAS carriers who have an intercurrent viral pharyngitis. However, because the sensitivity of the RADT is affected by clinical status (pharyngitis vs. asymptomatic children) independent of inoculum size, this approximation does seem reliable. Most experts consider that the only way to differentiate true GAS infection and GAS carriage is by use of GAS-specific antibodies

(i.e., anti-streptolysin O and anti-DNase B antibodies) [11]. However, the assessment of streptococcal antibody response would have required repeated blood samples that seemed untenable because of the cross-sectional design of the study. Moreover, the interpretation of streptococcal antibody titers is debated with regard to most recent available data [30].

Third, we cannot exclude that sample discordance between the two swabs might explain some discrepancies between results from the RADT and from throat culture, but this hypothesis could not be further investigated. Indeed, sample discordance was previously reported to be 5–10 % in studies involving duplicate throat cultures [2]. Although we used a contiguous double-swab collection system to minimize sample discordance, RADT sensitivity may have been low for some physicians because of asymmetrical throat sampling.

In conclusion, the results of this large prospective study expand our understanding of the factors affecting RADT sensitivity for detecting GAS. Diagnosis, like any decision-making process, is affected by various factors [23, 24]. Our results confirm that the sensitivity of a RADT should not be considered a fixed value, but rather, several patient- and physician-level factors may independently affect the sensitivity of the test. That physician characteristics affect the results of the RADT independent of patient clinical features and inoculum size should be emphasized in medical educational programs to improve physician practices and reduce inter-physician variations in RADT sensitivity. Our results support the American Academy of Pediatrics recommendation for physicians who have chosen to base their diagnosis of GAS pharyngitis solely on the result of a RADT to first evaluate the performance of this RADT in their office [3] and consider taking adequate educational and training programs when needed.

**Acknowledgments** The authors thank A. Liboz (Department of Microbiology, Robert Debré Hospital); M. Boucherat, MD (database design); F. de La Rocque, MD (study conception); I. Ramay, D. Menguy, S. Tortorelli and M. de Pereira (ACTIV); B. Khoshnood, MD, PhD (multilevel analysis methodological support); and all the physician investigators who participated in the RADT GAS study: F. Corrad, MD, P. Deberdt, MD, A. Elbez, MD, M. Goldrey, MD, J. Gosselin, MD, P. Martin, MD, A.S. Michot, MD, N. Panis, MD, D. Qutob, MD, C. Romain, MD, O. Romain, MD, C. Schlemmer, MD, F. Thollot, MD, A. Wollner, MD.

**Funding sources** This study was internally funded by ACTIV and externally by Dectrapharm, manufacturer of the RADT (Rapid Antigen Detection Test). JFC was supported by educational grants from Agence Régionale de Santé d'Ile-de-France, Laboratoires Guigoz - Société Française de Pédiatrie - Groupe de Pédiatrie Générale - Groupe de Recherches Epidémiologiques en Pédiatrie, and a research grant from the French Ministry of Health (PHRC Régional 2012, AOR 12089). The external funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Conflicts of interest** The authors declare that they have no conflict of interest.

**Previous presentations** This work was presented in part at the 30th Annual Meeting of the European Society for Pediatric Infectious Diseases, Thessaloniki, Greece, May 2012 (Abstract A-435-0007-01213).

## References

1. Shaikh N, Leonard E, Martin JM (2010) Prevalence of streptococcal pharyngitis and streptococcal carriage in children: a meta-analysis. *Pediatrics* 126:e557–e564
2. Gerber MA, Shulman ST (2004) Rapid diagnosis of pharyngitis caused by group A streptococci. *Clin Microbiol Rev* 17:571–580
3. American Academy of Pediatrics (2006) Group A streptococcal infections. In: Pickering L, Baker C, Long S, McMillan J (eds) Red book: 2006 report of the committee on infectious disease, 27th edn. American Academy of Pediatrics, Elk Grove Village, pp 616–628
4. Shulman ST, Bisno AL, Clegg HW, Gerber MA, Kaplan EL, Lee G, Martin JM, Van Beneden C (2012) Clinical practice guideline for the diagnosis and management of group A streptococcal pharyngitis: 2012 update by the Infectious Diseases Society of America. *Clin Infect Dis* 55:e86–e102
5. Ransohoff DF, Feinstein AR (1978) Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. *N Engl J Med* 299:926–930
6. Goehring C, Perrier A, Morabia A (2004) Spectrum bias: a quantitative and graphical analysis of the variability of medical diagnostic test performance. *Stat Med* 23:125–135
7. Hall MC, Kieke B, Gonzales R, Belongia EA (2004) Spectrum bias of a rapid antigen detection test for group A beta-hemolytic streptococcal pharyngitis in a pediatric population. *Pediatrics* 114:182–186
8. Edmonson MB, Farwell KR (2005) Relationship between the clinical likelihood of group A streptococcal pharyngitis and the sensitivity of a rapid antigen-detection test in a pediatric practice. *Pediatrics* 115:280–285
9. Tanz RR, Gerber MA, Kabat W, Rippe J, Seshadri R, Shulman ST (2009) Performance of a rapid antigen-detection test and throat culture in community pediatric offices: implications for management of pharyngitis. *Pediatrics* 123:437–444
10. Cohen JF, Chalumeau M, Levy C, Bidet P, Thollot F, Wollner A, Bingen E, Cohen R (2012) Spectrum and inoculum size effect of a rapid antigen detection test for group A streptococcus in children with pharyngitis. *PLoS One* 7:e39085
11. Gerber MA, Randolph MF, Chanatry J, Wright LL, DeMeo KK, Anderson LR (1986) Antigen detection test for streptococcal pharyngitis: evaluation of sensitivity with respect to true infections. *J Pediatr* 108:654–658
12. Kuhn S, Davies HD, Katzko G, Jadavji T, Church DL (1999) Evaluation of the strep A OIA assay versus culture methods: ability to detect different quantities of group A Streptococcus. *Diagn Microbiol Infect Dis* 34:275–280
13. Kurtz B, Kurtz M, Roe M, Todd J (2000) Importance of inoculum size and sampling effect in rapid antigen detection for diagnosis of Streptococcus pyogenes pharyngitis. *J Clin Microbiol* 38:279–281
14. Gerber MA, Tanz RR, Kabat W, Dennis E, Bell GL, Kaplan EL, Shulman ST (1997) Optical immunoassay test for group A beta-hemolytic streptococcal pharyngitis. An office-based, multicenter investigation. *JAMA* 277:899–903
15. Fox JW, Cohen DM, Marcon MJ, Cotton WH, Bonsu BK (2006) Performance of rapid streptococcal antigen testing varies by personnel. *J Clin Microbiol* 44:3918–3922
16. Tanz RR, Shulman ST (2007) Chronic pharyngeal carriage of group A streptococci. *Pediatr Infect Dis J* 26:175–176
17. Lieu TA, Fleisher GR, Schwartz JS (1988) Clinical evaluation of a latex agglutination test for streptococcal pharyngitis: performance and impact on treatment rates. *Pediatr Infect Dis J* 7:847–854
18. Nsanze H, Dawson KP, Ameen AS, Mustafa N (1998) Group A streptococcal antigen detection in schoolchildren. *East Mediterr Health J* 4:260–264
19. Dawson KP, Ameen AS, Nsanze H, Bin-Othman S, Mustafa N (1996) The prevalence of group A streptococcal throat carriage in Al Ain, United Arab Emirates. *Ann Trop Paediatr* 16:123–127
20. Gupta R, Kalia A, Rattan A, Kumar R, Gupta SK (1997) Comparative evaluation of two indigenously developed tests for rapid detection of group-A streptococci directly from throat swabs. *Indian J Med Res* 105:200–205
21. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, Moher D, Rennie D, de Vet HC, Lijmer JG (2003) The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Ann Intern Med* 138:W1–W12
22. Cohen R, Levy C, Ovetchkine P, Boucherat M, Weil-Olivier C, Gaudelus J, de la Rocque F, Bingen E (2004) Evaluation of streptococcal clinical scores, rapid antigen detection tests and cultures for childhood pharyngitis. *Eur J Pediatr* 163:281–282
23. Landon BE, Reschovsky J, Reed M, Blumenthal D (2001) Personal, organizational, and market level influences on physicians' practice patterns: results of a national survey of primary care physicians. *Med Care* 39:889–905
24. Tracy CS, Dantas GC, Moineddin R, Upshur RE (2005) Contextual factors in clinical decision making: national survey of Canadian family physicians. *Can Fam Physician* 51:1106–1107
25. Park SY, Gerber MA, Tanz RR, Hickner JM, Galliher JM, Chuang I, Besser RE (2006) Clinicians' management of children and adolescents with acute pharyngitis. *Pediatrics* 117:1871–1878
26. Diez-Roux AV (2000) Multilevel analysis in public health research. *Annu Rev Public Health* 21:171–192
27. Coughlin SS, Trock B, Criqui MH, Pickle LW, Browner D, Tefft MC (1992) The logistic modeling of sensitivity, specificity, and predictive value of a diagnostic test. *J Clin Epidemiol* 45:1–7
28. Blandino G, Puglisi S, Speciale A, Musumeci R (2011) Streptococcus pyogenes emm types and subtypes of isolates from paediatric asymptomatic carriers and children with pharyngitis. *New Microbiol* 34:101–104
29. Bell SM, Smith DD (1976) Quantitative throat-swab culture in the diagnosis of streptococcal pharyngitis in children. *Lancet* 2:62–63
30. Johnson DR, Kurlan R, Leckman J, Kaplan EL (2010) The human immune response to streptococcal extracellular antigens: clinical, diagnostic, and potential pathogenetic implications. *Clin Infect Dis* 50:481–490

---

ANNEXE 3. ARTICLE CORRESPONDANT AU CHAPITRE 3

## Rapid-Antigen Detection Tests for Group A *Streptococcal* Pharyngitis: Revisiting False-Positive Results Using Polymerase Chain Reaction Testing

Jérémie F. Cohen, MD, MSc<sup>1,2</sup>, Robert Cohen, MD<sup>3,4</sup>, Philippe Bidet, MD, PhD<sup>5</sup>, Corinne Levy, MD<sup>3,6</sup>, Patrice Deberdt, MD<sup>3</sup>, Camille d'Humières, MD<sup>5</sup>, Sandrine Liguori, MD<sup>5</sup>, François Corrad, MD<sup>3</sup>, Franck Thollot, MD<sup>7</sup>, Patricia Mariani-Kurkdjian, MD, PhD<sup>5</sup>, Martin Chalumeau, MD, PhD<sup>1,2,\*</sup>, and Edouard Bingen, PharmD, PhD<sup>5,\*,\*\*</sup>

We investigated mechanisms of the false-positive test results on rapid-antigen detection test (RADT) for group A *Streptococcal* (GAS) pharyngitis. Most RADT false-positives (76%) were associated with polymerase chain reaction-positive GAS results, suggesting that RADT specificity could be considered close to 100%. Finding that 61% of GAS culture-negative but RADT-positive cases were positive on both GAS polymerase chain reaction and *Staphylococcus aureus* testing, we posit bacterial inhibition as causative. (*J Pediatr* 2013;162:1282-4).

False-positive (FP) test results of rapid-antigen detection tests (RADTs) (ie, positive RADT results with negative throat culture results for group A *Streptococcus* [GAS]) may be due to cross-reactivity with bacteria from the *Streptococcus milleri* (*S milleri*) group<sup>1,2</sup> or to GAS nutritional variants or non- $\beta$ -hemolytic variants<sup>2-4</sup> or non-viable GAS. Non-viable GAS can be explained by suboptimal or prolonged transport conditions of the swab specimen or recent antibiotic therapy.<sup>5</sup> RADT FPs also could be attributed to the growth inhibiting presence of *Staphylococcus aureus* (*S aureus*).<sup>6</sup> A previous study used a polymerase chain reaction (PCR) assay in 200 children with pharyngitis to resolve discrepant results between RADT and throat culture.<sup>7</sup> However, the authors detected a segment of the *mitogenic factor* gene, which is not present in all GAS strains, contrary to encoding for *Streptococcal* pyrogenic exotoxin B (*speB*).<sup>8</sup>

The present case-control study was an ancillary study of data from a French prospective office-based study assessing the performance of a RADT (Table; available at www.jpeds.com). Here, we aimed to explain RADT FP results and included systematic PCR assay for *speB*.

### Methods

Eligible patients were children aged 3-14 years who had a diagnosis of pharyngitis and did not receive antibiotics for 7 days before inclusion (October 2010-May 2011). The primary study protocol was approved by the Ile-de-France

XI Institutional Review Board. For the present case-control study, all FP cases, and consecutive true-negative controls (ie, negative RADT results with negative culture results for GAS) and true-positive controls (ie, positive RADT results with positive culture results for GAS) were recruited among patients included in the primary study.

Throat samples were obtained by use of a double-swab collection-transport system. Swab 1 was used for the RADT (StreptAtest; Dectrapharm, Strasbourg, France) performed in the pediatrician's office. Swab 2 was transferred within  $\leq 72$  hours to the microbiology laboratory. On receipt, swab 2 was streaked on a blood agar plate (BAP; Columbia ANC + 5% sheep blood; Biomérieux, Lyon, France) and culture was performed as described elsewhere.<sup>9</sup> The swab was then transferred to a 9-mL tryptone-glucose-yeast (TGY) enrichment broth. When no  $\beta$ -hemolytic colony was found on the BAP, 1 drop of the TGY broth previously incubated at 37°C for 24 hours was used for subsequent culture on a BAP.

If the TGY-enriched plate remained negative for GAS, the 9 mL of the TGY broth were used for DNA extraction on

BAP	Blood agar plate
FP	False-positive
GAS	Group A <i>Streptococcus</i> ( <i>Streptococcal</i> )
PCR	Polymerase chain reaction
RADT	Rapid-antigen detection test
<i>S aureus</i>	<i>Staphylococcus aureus</i>
<i>S milleri</i>	<i>Streptococcus milleri</i>
<i>speB</i>	<i>Streptococcal</i> pyrogenic exotoxin B
TGY	Tryptone-glucose-yeast

From the <sup>1</sup>INSERM, UMR S953, Epidemiological Research Unit on Perinatal Health and Women's and Children's Health, Paris, France; <sup>2</sup>Department of Pediatrics, Necker-Enfants-Malades Hospital, Assistance Publique-Hôpitaux de Paris, Univ Paris Descartes, Sorbonne Paris Cité, Paris, France; <sup>3</sup>Association Clinique et Thérapeutique du Val-de-Marne (ACTIV), Saint-Maur-des-Fossés, France; <sup>4</sup>Department of Microbiology, Centre Hospitalier Intercommunal de Créteil, Créteil, France; <sup>5</sup>Department of Microbiology, Robert Debré Hospital, Assistance Publique-Hôpitaux de Paris, Univ Paris Diderot, Sorbonne Paris Cité, Paris, France; <sup>6</sup>Centre de Recherche Clinique (CRC), Centre Hospitalier Intercommunal de Créteil, Créteil, France; and <sup>7</sup>Association Française de Pédiatrie Ambulatoire (AFPA), Essey-lès-Nancy, France

\*Contributed equally.

\*\*Deceased.

Internally funded by ACTIV and externally by Dectrapharm, manufacturer of the RADT. J.C. is supported by educational grants from Agence Régionale de Santé d'Ile-de-France, Laboratoires Guigoz, Société Française de Pédiatrie, Groupe de Pédiatrie Générale, Groupe de Recherches Epidémiologiques en Pédiatrie, and a research grant from the French Ministry of Health (PHRC Régional 2012, AOR 12089). The authors declare no conflicts of interest.

Portions of this study were presented as an abstract at the annual meeting of the European Society for Pediatric Infectious Diseases, Thessaloniki, Greece, May 8-12 2012.

0022-3476/\$ - see front matter. Copyright © 2013 Mosby Inc.  
All rights reserved. <http://dx.doi.org/10.1016/j.jpeds.2013.01.050>

the EZ1 BioRobot (Qiagen, Courtaboeuf, France). The DNA extract was then used for GAS-specific PCR assay (GAS-PCR) with a pair of primers to amplify a 240-bp fragment of the *speB* gene: SGA-*speB*-F (5'-ATA CCA ACA TCA GCC ATC AAT TC-3') and SGA-*speB*-R (5'-TGC TAC TGC AAC TGC TCA AAT TA-3'). All PCR runs included a positive control (DNA extracted from GAS strain ATCC 12344) and negative control (sterile distilled water). In preliminary experiments, this GAS-PCR method could detect all major *emm*-types from a collection of 250 GAS strains,<sup>10</sup> and results were uniformly negative in testing various other strains of bacteria belonging to the oropharyngeal flora.

Throat swab samples were screened for the presence of *S aureus* by inoculating mannitol salt agar with one drop of the original swab fluid. *S aureus* was identified by gram staining, catalase and coagulase production, and *S aureus* growth quantified by colony plate count as light (<50 colonies) or heavy ( $\geq$ 50 colonies). In case of minute colonies giving group A-positive reactions on latex agglutination, we searched for the presence of *S milleri* by using biochemical procedures (API Rapid ID32 Strep; Biomérieux, Lyon, France). Microbiologists were blinded to individual clinical data and RADT results.

## Results

Analysis involved 41 FP cases and 58 true-negative and 116 true-positive controls. Cases and controls did not differ in age or sex ( $P > .35$ ).

All TGY-enriched BAP cultures remained negative for GAS. GAS-PCR results were positive for 76% (95 CI 60%-88%) of FP cases versus 2% (0%-9%) of true-negative controls ( $P < .001$ ). *S aureus* was recovered in 73% (57%-86%) of FP cases versus 14% (8%-21%) of true-positive controls ( $P < .001$ ). In total, 61% (45%-76%) of FP cases were positive on both GAS-PCR and *S aureus* testing. Heavy inocula for *S aureus* were more frequent for FP cases than true-positive controls (63% [44%-80%] vs 12.5% [2%-38%],  $P = .002$ ). No *S milleri* colony was identified.

Mean office-to-laboratory delay did not differ between FP cases and true-positive controls (2.3 vs 2.6 days,  $P = .22$ ).

## Discussion

In all, 76% of cases of apparent RADT FP results were positive for GAS by PCR, indicating that these patients showed true colonization or infection with GAS. This suggests that RADT specificity may be considered close to 100%. Our hypothesis is that throat culture failure could possibly be due to GAS being non-viable, inhibited, or non-detectable because of the presence of other bacterial species.

We found 61% of FP cases had both GAS-PCR assay positive and growth of *S aureus*, usually with heavy density. These results suggest that RADT FP results are mostly found in patients for whom GAS was not recovered in culture because of *S aureus* co-colonization. The possible inhibitory

effect of *S aureus* might be explained by bacterial interference involving the production of bacteriocins and bacteriolytic substances and competition for nutrients.<sup>6</sup> However, conclusive biological evidence (eg, direct evidence of the presence of staphylococcal bacteriocins in the culture media or results from co-culture experiments) is needed to support this hypothesis.

That GAS-PCR results were positive for 32% of patients with GAS-negative BAP culture results (32/99) challenges the classic role of conventional throat culture as the reference test for diagnosis of GAS pharyngitis. We did not perform serologic testing to confirm that PCR-positive, culture-negative cases represented true infection versus colonization/carriage. The PCR target we used seems to be the most appropriate for the identification of GAS because it was previously shown that the *speB* gene is identifiable by PCR in all GAS clinical isolates, contrary to *mitogenic factor*.<sup>7,8</sup> However, our PCR technique might not be the most sensitive. For example, we could have used mutanolysin instead of lysozyme for cell wall degradation prior to DNA extraction<sup>7</sup> and real-time PCR instead of conventional PCR. ■

*The authors thank A. Liboz (Department of Microbiology, Robert Debré Hospital), M. Boucherat, MD (database design), F. de La Rocque, MD (study conception), I. Ramay, D. Menguy, S. Tortorelli and M. de Pereira (ACTIV), and all the physician investigators who participated in the RADT GAS study: M. Benani, MD, A. Elbez, MD, M. Goldrey, MD, J. Gosselin, MD, M. Koskas, MD, P. Martin, MD, A.S. Michot, MD, N. Panis, MD, D. Qutob, MD, C. Romain, MD, O. Romain, MD, C. Schlemmer, MD, A. Wollner, MD.*

Submitted for publication Aug 15, 2012; last revision received Dec 21, 2012; accepted Jan 24, 2013.

Reprint requests: Jérémie F. Cohen, MD, INSERM, UMR S953, Maternité de Port Royal 53, avenue de l'Observatoire, Paris 75014, France. E-mail: jeremie.cohen@inserm.fr

## References

1. Rubin LG, Kahn RA, Vellozzi EM, Isenberg HD. False positive detection of group A *Streptococcus* antigen resulting from cross-reacting *Streptococcus intermedius* (*Streptococcus milleri* group). *Pediatr Infect Dis J* 1996;15:715-7.
2. Johnson DR, Kaplan EL. False-positive rapid antigen detection test results: reduced specificity in the absence of group A *Streptococci* in the upper respiratory tract. *J Infect Dis* 2001;183:1135-7.
3. Rubin LG, Mirkin GS. Apparent false positive detection of group A *Streptococcus* antigen resulting from pharyngeal infection with a nonhemolytic *Streptococcus pyogenes*. *Pediatr Infect Dis J* 2000;19:672-4.
4. Schroeder S, Procop GW. False positive strep A antigen test. *Pediatr Infect Dis J* 2000;19:1114-5.
5. Gerber MA, Randolph MF, DeMeo KK. *Streptococcal* antigen in the pharynx after initiation of antibiotic therapy. *Pediatr Infect Dis J* 1987;6:489-91.
6. Brook I. Bacterial interference. *Crit Rev Microbiol* 1999;25:155-72.
7. Kaltwasser G, Diego J, Welby-Sellenriek PL, Ferrett R, Caparon M, Storch GA. Polymerase chain reaction for *Streptococcus pyogenes* used to evaluate an optical immunoassay for the detection of group A *Streptococci* in children with pharyngitis. *Pediatr Infect Dis J* 1997;16:748-53.
8. Hsueh PR, Wu JJ, Tsai PJ, Liu JW, Chuang YC, Luh KT. Invasive group A *Streptococcal* disease in Taiwan is not associated with the

- presence of *Streptococcal* pyrogenic exotoxin genes. Clin Infect Dis 1998;26:584-9.
9. Cohen JF, Chalumeau M, Levy C, Bidet P, Thollot F, Wollner A, et al. Spectrum and inoculum size effect of a rapid antigen detection test for group A *Streptococcus* in children with pharyngitis. PLoS One 2012;7:e39085.
  10. Bidet P, Liguori S, Plainvert C, Bonacorsi S, Courroux C, d'Humieres C, et al. Identification of group A *Streptococcal* emm types commonly associated with invasive infections and antimicrobial resistance by the use of multiplex PCR and high-resolution melting analysis. Eur J Clin Microbiol Infect Dis 2012;31:2817-26.

**Table.** Characteristics of included children, GAS prevalence, and diagnostic accuracy of RADT in the primary study (n = 676)

	% (95 CI)	Mean (SD)
Age, years		6.1 (2.5)
Girls	46.3 (42.5-50.0)	
GAS prevalence*	41.4 (37.7-45.2)	
RADT sensitivity*	92.5 (88.8-95.3)	
RADT specificity*	88.4 (84.8-91.4)	

\*With conventional throat culture as the reference test.



---

ANNEXE 4. ARTICLES DE SYNTHÈSE EN FRANÇAIS

## Sensibilité du test de diagnostic rapide du streptocoque du groupe A chez les porteurs sains et les enfants ayant une angine

J.-F. Cohen<sup>a,\*</sup>, C. Levy<sup>b</sup>, P. Bidet<sup>c</sup>, M. Benani<sup>b</sup>, F. Thollot<sup>d</sup>, M. Koskas<sup>b</sup>,  
M. Chalumeau<sup>a,e</sup>, E. Bingen<sup>c</sup>, R. Cohen<sup>b,f</sup>

<sup>a</sup>Inserm U953, Recherche épidémiologique en santé périnatale et santé des femmes  
et des enfants, Hôpital Saint-Vincent-de-Paul, Paris, France

<sup>b</sup>Association Clinique et Thérapeutique du Val-de-Marne (ACTIV), Saint-Maur-des-Fossés, France

<sup>c</sup>Laboratoire de Bactériologie, Hôpital Robert Debré, AP-HP, Faculté de médecine Paris Diderot,  
Sorbonne Paris Cité EA 3105, F-75505 Paris, France

<sup>d</sup>Association Française de Pédiatrie Ambulatoire (AFPA), Essey-lès-Nancy, France

<sup>e</sup>Service de Pédiatrie Générale, Hôpital Necker-Enfants Malades, AP-HP, Université Paris  
Descartes, Paris, France

<sup>f</sup>Laboratoire de Bactériologie, Centre Hospitalier Intercommunal, Créteil, France

Des tests de diagnostic rapide (TDR) sont disponibles depuis les années 1980 pour aider au diagnostic des angines à streptocoque du groupe A (SGA). Une des critiques faites aux TDR est leur sensibilité proche de 85 % en moyenne. Il a été montré qu'une part importante des faux négatifs du TDR avait un inoculum bactérien faible [1] et que les porteurs sains de SGA avaient plus fréquemment un inoculum faible que les enfants ayant une angine [2]. Ceci a conduit à émettre l'hypothèse que les faux négatifs du TDR surviendraient chez des enfants porteurs de SGA ayant une angine virale intercurrente, ce qui minimiserait leur poids clinique.

### 1. Portage sain de SGA

Environ 5 % à 20 % des enfants d'âge scolaire sont des porteurs sains de SGA au niveau pharyngé [3]. Le portage sain est défini par la présence de SGA en culture de prélèvement de gorge chez un sujet asymptomatique et en l'absence de réponse inflammatoire biologique. Ces enfants ne nécessitent pas de traitement antibiotique. Lorsqu'un enfant porteur de SGA a une angine, il peut s'agir d'une vraie angine à SGA pour laquelle une antibiothérapie est recommandée, ou d'une angine virale ne justifiant pas une antibiothérapie.

### 2. Sensibilité du TDR et portage

Pour explorer l'hypothèse selon laquelle les faux négatifs du TDR seraient des porteurs de SGA plutôt que des enfants ayant une infection active à SGA, 2 approches ont été utilisées : l'une basée sur les sérologies SGA et l'autre sur l'utilisation d'un groupe témoin d'enfants asymptomatiques.

#### 2.1. Étude basée sur les sérologies SGA

Les sérologies SGA (anticorps anti-streptolysine (ASLO) et anti-streptodornase (ASD)) ont longtemps été proposées pour faire la distinction entre infection vraie à SGA et portage de SGA. Il est en effet souvent avancé qu'une augmentation significative du titre de ces anticorps ne s'observait qu'en cas d'infection active à SGA. Une seule étude a évalué les performances du TDR en utilisant les sérologies SGA pour faire la distinction entre infections actives à SGA et porteurs [4]. Dans cette étude menée sur 313 enfants américains, une augmentation des ASLO ou des ASD était constatée dans une proportion comparable chez les faux négatifs et chez les vrais positifs du TDR (45 % contre 51 % respectivement). Les auteurs concluaient que la moitié des faux négatifs du TDR étaient des enfants ayant une infection active à SGA et recommandaient la réalisation systématique d'une culture en cas de TDR négatif. Cette étude est limitée car ce TDR au latex est aujourd'hui techniquement obsolète, la fréquence du SGA était particulièrement élevée (82 %) et le spectre clinique des patients particulièrement grave (100 % de fièvre), conduisant à un biais de sélection. De plus l'intérêt des sérologies SGA a été récemment remis en cause par les résultats d'une étude ayant suivi 160 enfants nord-américains pendant 2 ans, avec

\*Auteur correspondant.  
e-mail : jeremie.cohen@inserm.fr

prélèvements de gorge et prélèvements sanguins répétés [5], qui ont notamment montré que :

- suite à l'acquisition asymptomatique d'un nouveau sérotype de SGA, une augmentation des ASLO et des ASD était observée dans 35 % des cas ;
- certains patients ayant un portage chronique asymptomatique de SGA gardaient un titre élevé d'ASLO et d'ASD pendant plusieurs mois ;
- l'acquisition d'un streptocoque du groupe C ou G était accompagnée d'une augmentation des ASLO dans 37 % des cas.

Enfin, l'interprétation de ces sérologies est complexe, car les normes varient en fonction de l'âge et de la zone géographique [5]. L'utilisation des sérologies SGA en clinique courante est de toute manière difficile à proposer en raison de la nécessité de prélèvements sanguins répétés.

### 2.2. Étude basée sur un groupe témoin d'enfants asymptomatiques

En raison des limites des sérologies SGA, il est intéressant d'étudier la sensibilité du TDR chez des porteurs sains car elle peut probablement être considérée comme étant proche de la sensibilité du TDR chez les enfants ayant un portage de SGA et une angine virale. Si la sensibilité du TDR chez les porteurs sains était très basse, cela appuierait l'hypothèse selon laquelle les faux négatifs du TDR en cas d'angine sont en grande partie des porteurs ayant une angine virale.

Une seule étude a comparé la sensibilité d'un TDR proche de celui utilisé en France entre des enfants avec angine et des porteurs sains [6]. La sensibilité du TDR était plus élevée en cas d'angine que de portage sain (100 % contre 87 %), mais les résultats sont limités car la lecture des cultures n'était pas faite à l'aveugle du TDR et la fréquence du SGA chez les enfants atteints d'angine et chez les enfants asymptomatiques était particulièrement basse (12 % et 2 % respectivement), suggérant un test de référence insuffisamment sensible.

### 2.3. Étude ACTIV

Dans une étude chez 1482 enfants avec angine et 294 enfants sains recrutés par les pédiatres d'ACTIV, la fréquence du portage sain de SGA était de 11 %, proche de celle rapportée dans une méta-analyse récente [3]. En analyse univariée, la sensibilité du TDR était plus élevée en cas d'angine que de portage sain (89 % contre 41 %,  $p < 0,001$ ), et augmentait avec l'inoculum bactérien évalué selon la densité des colonies de SGA en culture (53 % contre 94 % en cas d'inoculum faible ou fort respectivement,  $p < 0,001$ ). Les enfants avec angine avaient plus fréquemment un inoculum fort que les porteurs sains mais à inoculum égal, la sensibilité du TDR était plus élevée en cas d'angine que de portage sain ( $p = 0,01$ ). Nous avons également mis en évidence des variations importantes de sensibilité du TDR en fonction du médecin ayant réalisé le test (de 56 % à 96 %,  $p = 0,01$ ). En analyse multivariée multi-niveau, la sensibilité du TDR était toujours plus élevée chez les enfants avec angine que chez les porteurs sains indépendamment de l'âge, de l'inoculum et du médecin.

Ces résultats confirment qu'une part des faux négatifs du TDR est liée aux faibles inocula bactériens et que ces faibles inocula sont plus fréquents en cas de portage sain qu'en cas d'angine. Cependant, le TDR semble se comporter différemment chez les

enfants ayant une angine et chez les porteurs sains et cela nous empêche d'estimer précisément la part des faux négatifs du TDR pouvant être expliquée par des porteurs ayant une angine virale intercurrente.

## 3. Conclusion

Certains experts jugent insuffisante la sensibilité du TDR et recommandent la confirmation systématique des résultats négatifs par une culture. Pourtant, la sensibilité de la culture au laboratoire de bactériologie n'est que de 90-95 % [7] et la sensibilité de la culture réalisée au cabinet du médecin, surtout pratiquée aux États-Unis, est inférieure ou égale à celle du TDR (50 % à 89 %) [8]. Il semble raisonnable de ne pas cibler une sensibilité de 100 % pour les stratégies diagnostiques des angines de l'enfant en raison du portage sain. Si on accepte une sensibilité de l'ordre de 90 %, la stratégie française consistant à réaliser un TDR sans contrôle systématique des résultats négatifs est acceptable.

Bien que l'étude ACTIV ait été menée au sein d'un réseau de pédiatres formés à la réalisation du TDR, il existait d'importantes variations de la sensibilité du TDR en fonction du médecin ayant réalisé le test. Les pédiatres qui décident de baser leur décision thérapeutique uniquement sur le résultat d'un TDR devraient d'abord s'assurer de la qualité de leur technique de réalisation du test. Ceux qui obtiennent les moins bons résultats devraient avoir accès à des formations ciblées leur permettant d'améliorer leur technique et leurs performances diagnostiques.

## Références

- [1] Cohen J, Levy C, Bidet P, et al. Test de diagnostic rapide dans les angines de l'enfant : effet et biais de spectre. Arch Pediatr, hors série Congrès des Sociétés Médico-Chirurgicales de Pédiatrie 2012;19:147-1.
- [2] Bell SM, Smith DD. Quantitative throat-swab culture in the diagnosis of streptococcal pharyngitis in children. Lancet 1976;2:62-3.
- [3] Shaikh N, Leonard E, Martin JM. Prevalence of streptococcal pharyngitis and streptococcal carriage in children: a meta-analysis. Pediatrics 2010;126:e557-64.
- [4] Gerber MA, Randolph MF, Chanatry J, et al. Antigen detection test for streptococcal pharyngitis: evaluation of sensitivity with respect to true infections. J Pediatr 1986;108:654-8.
- [5] Johnson DR, Kurlan R, Leckman J, et al. The human immune response to streptococcal extracellular antigens: clinical, diagnostic, and potential pathogenetic implications. Clin Infect Dis 2010;50:481-90.
- [6] Gupta R, Kalia A, Rattan A, et al. Comparative evaluation of two indigenously developed tests for rapid detection of group-A streptococci directly from throat swabs. Indian J Med Res 1997;105:200-5.
- [7] Kellogg JA. Suitability of throat culture procedures for detection of group A streptococci and as reference standards for evaluation of streptococcal antigen detection kits. J Clin Microbiol 1990;28:165-9.
- [8] Tanz RR, Gerber MA, Kabat W, et al. Performance of a rapid antigen-detection test and throat culture in community pediatric offices: implications for management of pharyngitis. Pediatrics 2009;123:437-44.

## Faux positifs du test de diagnostic rapide du streptocoque du groupe A dans les angines de l'enfant

J.-F. Cohen<sup>a,\*</sup>, R. Cohen<sup>b,c</sup>, P. Bidet<sup>d</sup>, C. Levy<sup>b</sup>, P. Deberdt<sup>b</sup>, C. d'Humieres<sup>d</sup>, S. Liguori<sup>d</sup>, F. Corrad<sup>b</sup>, F. Thollot<sup>e</sup>, P. Mariani-Kurkdjian<sup>d</sup>, M. Chalumeau<sup>a,f</sup>, E. Bingen<sup>d</sup>

<sup>a</sup>Inserm U953, Recherche épidémiologique en santé périnatale et santé des femmes et des enfants, Hôpital Saint-Vincent-de-Paul, Paris, France

<sup>b</sup>Association Clinique et Thérapeutique du Val-de-Marne (ACTIV), Saint-Maur-des-Fossés, France

<sup>c</sup>Laboratoire de Bactériologie, Centre Hospitalier Intercommunal, Créteil, France

<sup>d</sup>Laboratoire de Bactériologie, Hôpital Robert Debré, AP-HP, Faculté de médecine Paris Diderot, Sorbonne Paris Cité EA 3105, F-75505 Paris, France

<sup>e</sup>Association Française de Pédiatrie Ambulatoire (AFPA), Essey-lès-Nancy, France

<sup>f</sup>Service de Pédiatrie Générale, Hôpital Necker-Enfants Malades, AP-HP, Université Paris Descartes, Paris, France

La spécificité du test de diagnostic rapide (TDR) du streptocoque du groupe A (SGA) dans les angines de l'enfant a été très peu étudiée. Alors qu'elle est de 100 % *in vitro*, la spécificité des TDR est en moyenne de 95 % en pratique clinique [1]. Ainsi, certains patients ont un TDR positif et une culture de prélèvement de gorge négative pour le SGA, on parle alors de faux positifs (FP) du TDR. Même s'ils sont rares, ces FP ne sont pas clairement expliqués dans la littérature et peuvent représenter une source d'antibiothérapie inutile, entraînant des effets indésirables et participant à l'émergence de résistances bactériennes.

### 1. Principales hypothèses expliquant les faux positifs du TDR

Différentes hypothèses ont été avancées dans la littérature pour expliquer les FP du TDR.

#### 1.1. Réaction antigénique croisée

La plupart des TDR actuellement disponibles réagissent avec le polysaccharide C du groupe A. Le polysaccharide C est l'antigène de paroi des streptocoques qui est à la base de la classification en groupes de Lancefield. *Streptococcus pyogenes* est ainsi également connu sous le nom de streptocoque du groupe A. En théorie, le polysaccharide C du groupe A est spécifique de *Streptococcus pyogenes*, mais il a été montré que cet antigène pouvait également être porté par les streptocoques du groupe milleri (*Streptococcus intermedius*, *anginosus* ou *constellatus*) [2,3]. Ce phénomène semble rare puisqu'une étude systématique sur

77 FP du TDR chez des adultes atteints d'angine avait mis en évidence un seul *Streptococcus milleri* porteur d'un antigène du groupe A [2]. Les *S. milleri* sont des germes commensaux des sphères oro-pharyngées, uro-génitales et du tractus digestif. Leur rôle pathogène chez l'homme est connu, notamment en tant qu'agent d'appendicite ou de complications suppurées post-appendicectomie, mais leur rôle en tant qu'agent d'angines est discuté.

#### 1.2. Réaction avec une bactérie non-viable ou morte

Le TDR est un test antigénique et il peut donc réagir en présence de l'antigène cible, même si la bactérie est non-viable ou morte. A l'inverse, la croissance du SGA en culture nécessite que la bactérie soit viable. Ainsi, certains auteurs ont émis l'hypothèse que les FP du TDR pourraient être liés à des bactéries devenues non-viables ou mortes [2] par différents mécanismes :

- soit les bactéries sont déjà mortes dans la gorge du patient, par exemple en raison de la prise récente d'antibiotiques ;
- soit les bactéries sont mortes durant le transport du prélèvement entre le cabinet du pédiatre et le laboratoire de bactériologie, par exemple parce que le milieu de transport était inadapté ou la durée du transport trop longue ;
- soit les bactéries sont mortes dans la gorge du patient, durant le transport du prélèvement ou dans le milieu de culture en raison d'un phénomène d'interférence bactérienne ou de croissance compétitive avec une bactérie de la flore oro-pharyngée, comme les streptocoques  $\alpha$ -hémolytiques ou *Staphylococcus aureus*.

L'interférence bactérienne désigne différents mécanismes par lesquels les bactéries interagissent entre elles de manière antagoniste ou compétitive, par exemple par la production de substances inhibitrices (bactériocines). La mise en évidence de bactéries non-viables peut se faire par technique de *polymerase*

\*Auteur correspondant.  
e-mail : jeremie.cohen@inserm.fr

*chain reaction* (PCR), car cette technique très spécifique ne nécessite pas que les bactéries soient viables mais simplement que l'ADN bactérien correspondant soit présent.

### 1.3. Perte de la $\beta$ -hémolyse

La plupart des SGA sont producteurs de  $\beta$ -hémolyse, ce qui entraîne une lyse des hématies en culture sur gélose au sang. Cette propriété est à la base de l'identification en routine des SGA en culture dans les laboratoires de bactériologie. Face à une colonie  $\beta$ -hémolytique à partir d'un prélèvement de gorge, la présence de SGA est confirmée par un test au latex : la colonie suspecte est mise en contact avec des billes de latex portant des anticorps spécifiques du SGA et le test est positif si on observe une réaction d'agglutination. La propriété de  $\beta$ -hémolyse est due à la production de streptolysine S, une puissante cytotoxine streptococcique, également capable de lyser les leucocytes. Cependant, certaines souches de SGA ayant une propriété d'hémolyse incomplète ( $\alpha$ -hémolyse) ou ayant perdu totalement leur capacité d'hémolyse ont été identifiées et pourraient expliquer certains FP du TDR [4,5]. En effet, ces souches de SGA réagissent de manière normale avec le TDR mais en raison de leur défaut de  $\beta$ -hémolyse en milieu de culture habituel, ces souches ne sont pas identifiées comme étant du SGA par les méthodes de culture de routine.

### 1.4. Variants nutritionnels

Une autre atypie du SGA peut concerner ses besoins nutritionnels. Les conditions habituelles de mise en culture du SGA au laboratoire de bactériologie sont un gel d'agar classique de type trypticase-soja additionné de sang de cheval ou de mouton, en milieu anaérobie et à température de 37 °C. Cependant, certaines souches de SGA atypiques ne sont pas adaptées à ce type de milieu et nécessitent un milieu particulier, par exemple un milieu enrichi de type tryptone-glucose-levure (TGY) ou un bouillon de Todd-Hewitt, ou encore une atmosphère enrichie en CO<sub>2</sub> [2].

## 2. Étude ACTIV

Une étude ayant pour but d'explorer de manière systématique les FP du TDR en testant les différentes hypothèses identifiées dans la littérature a été mise en place par l'Association Clinique et Thérapeutique du Val-de-Marne (ACTIV) et le Centre National de Référence des infections à SGA chez l'enfant (CNR-Strep). Parmi 1482 enfants atteints d'angine [6], des échantillons consécutifs de FP du TDR et de témoins vrais négatifs (VN) et vrais positifs (VP) ont été analysés par reprise sur un milieu de culture enrichi de type TGY, PCR spécifique du SGA (PCR-SGA), mise en culture sur gélose au mannitol spécifique de *S. aureus* et identification biochimique en cas de réaction d' $\alpha$ -hémolyse sur gélose au sang. Le délai moyen entre le prélèvement de gorge et la mise en culture a également été comparé entre les cas et

les témoins. Au total, 51 cas FP, 58 témoins VN et 117 témoins VP ont été étudiés. Aucun variant nutritionnel n'a été identifié après reprise sur milieu enrichi TGY. Les PCR-SGA étaient plus fréquemment positives chez les FP que chez les témoins VN (74 % contre 2 %,  $p < 0,001$ ). Nous n'avons pas mis en évidence de différence de délai moyen entre prélèvement et mise en culture entre les FP et les témoins VP (2,5 contre 2,6 jours,  $p = 0,5$ ). La fréquence de *S. aureus* en culture était plus élevée chez les cas FP que chez les témoins VP (74 % contre 14 %,  $p < 0,001$ ) et 65 % des FP avaient à la fois une PCR-SGA positive et une culture positive pour *S. aureus*. Aucune souche de *S. milleri* n'a été identifiée chez les FP ni chez les témoins VN et VP. Ces résultats sont compatibles avec ceux d'une étude ayant montré que *S. aureus* pouvait produire une toxine conduisant à des altérations sévères de la structures du SGA sans altération de l'ADN [7].

## 3. Conclusion

Les résultats FP du TDR sont rares en pratique courante et les hypothèses permettant de les expliquer sont nombreuses dans la littérature. Les résultats de l'étude systématique ACTIV/CNR-Strep montrent que la plupart des résultats FP du TDR sont expliqués par un phénomène d'interférence bactérienne entre *S. aureus* et le SGA qui empêche la croissance du SGA en culture. Ainsi, les FP du TDR surviennent chez des enfants ayant réellement du SGA au niveau pharyngé et cela confirme que la spécificité du TDR peut être considérée comme étant proche de 100 % en pratique clinique. Le poids des FP du TDR dans le nombre des antibiothérapies inutiles prescrites dans le cadre des angines de l'enfant est donc minime.

## Références

- [1] Gerber MA, Shulman ST. Rapid diagnosis of pharyngitis caused by group A streptococci. Clin Microbiol Rev 2004;17:571-80.
- [2] Johnson DR, Kaplan EL. False-positive rapid antigen detection test results: reduced specificity in the absence of group A streptococci in the upper respiratory tract. J Infect Dis 2001;183:1135-7.
- [3] Rubin LG, Kahn RA, Vellozzi EM, et al. False positive detection of group A Streptococcus antigen resulting from cross-reacting Streptococcus intermedius (Streptococcus milleri group). Pediatr Infect Dis J 1996;15:715-7.
- [4] Rubin LG, Mirkin GS. Apparent false positive detection of group a Streptococcus antigen resulting from pharyngeal infection with a nonhemolytic Streptococcus pyogenes. Pediatr Infect Dis J 2000;19:672-4.
- [5] Schroeder S, Procop GW. False positive strep A antigen test. Pediatr Infect Dis J 2000;19:1114-5.
- [6] Cohen J, Levy C, Bidet P, et al. Sensibilité du test de diagnostic rapide du streptocoque du groupe A chez les porteurs sains et les enfants ayant une angine. Arch Pédiatr, hors série Congrès des Sociétés Médico-Chirurgicales de Pédiatrie 2012;19:143-7.
- [7] Clawson ML, Dajani AS. Effect of Bactericidal Substance from Staphylococcus aureus on Group A Streptococci II. Structural Alterations. Infect Immun 1970;1:491-8.

## Test de diagnostic rapide du streptocoque du groupe A dans les angines de l'enfant : effet et biais de spectre

J.-F. Cohen<sup>a,\*</sup>, C. Levy<sup>b</sup>, P. Bidet<sup>c</sup>, F. Thollot<sup>d</sup>, A. Wollner<sup>b</sup>, P. Mariani-Kurkdjian<sup>c</sup>,  
M. Chalumeau<sup>a,e</sup>, E. Bingen<sup>c</sup>, R. Cohen<sup>b,f</sup>

<sup>a</sup>Inserm U953, Recherche épidémiologique en santé périnatale et santé des femmes  
et des enfants, Hôpital Saint-Vincent-de-Paul, Paris, France

<sup>b</sup>Association Clinique et Thérapeutique du Val-de-Marne (ACTIV), Saint-Maur-des-Fossés, France

<sup>c</sup>Laboratoire de Bactériologie, Hôpital Robert Debré, AP-HP, Faculté de médecine Paris Diderot,  
Sorbonne Paris Cité EA 3105, F-75505 Paris, France

<sup>d</sup>Association Française de Pédiatrie Ambulatoire (AFPA), Essey-lès-Nancy, France

<sup>e</sup>Service de Pédiatrie Générale, Hôpital Necker-Enfants Malades, AP-HP, Université Paris  
Descartes, Paris, France

<sup>f</sup>Laboratoire de Bactériologie, Centre Hospitalier Intercommunal, Créteil, France

Les tests de diagnostic rapide (TDR) permettent de faire en quelques minutes le diagnostic d'angine à streptocoque  $\beta$ -hémolytique du groupe A (SGA). Leur utilisation est recommandée en France chez tous les enfants âgés de 3 à 15 ans ayant une angine afin de limiter l'antibiothérapie aux cas confirmés d'infection à SGA. Parmi les critiques faites aux TDR, les plus fréquentes sont leurs sensibilités proches seulement de 85 % [1] et la variabilité de leurs performances diagnostiques en fonction de facteurs liés au patient (effet et biais de spectre) [2-4] et de la quantité de bactéries présentes dans la gorge du patient (effet de l'inoculum bactérien) [5].

### 1. Effet et biais de spectre

Contrairement à une conception largement admise, la sensibilité et la spécificité d'un test diagnostique ne sont pas des valeurs fixes mais peuvent varier d'un groupe de patients à l'autre, entre différentes études ou au sein d'une même étude. Ce phénomène est appelé effet ou biais de spectre [6]. Le terme « spectre » évoque d'une part le spectre clinique de la maladie, c'est-à-dire principalement les différents degrés de gravité de la maladie, et d'autre part, le spectre de la population d'étude. Le terme d'« effet de spectre » semble plus adapté que le terme de « biais de spectre » car les variations de performances diagnostiques observées en fonction du spectre de la maladie ou de la population d'étude ne sont pas forcément dues à un biais au sens épidémiologique du terme, à savoir une erreur systématique.

La mise en évidence d'un effet de spectre pose la question de la généralisabilité des performances diagnostiques observées

dans une étude. S'il est par exemple montré que la sensibilité d'un test diagnostique augmente avec la gravité de la maladie, cela implique que la sensibilité observée dans une étude est influencée par la gravité des patients inclus dans l'étude.

### 2. Effet de spectre du TDR pour le diagnostic des angines à SGA

Le spectre des angines à SGA est habituellement mesuré par la présence de certains signes cliniques et/ou avec le niveau du score de McIsaac [2-4]. Ce score clinique peut varier de 0 à 5, chacun des items suivants étant coté 1 point : âge entre 3 et 15 ans, absence de toux, fièvre, adénopathies cervicales antérieures sensibles, tuméfaction ou exsudat des amygdales [7]. Il a été montré qu'il existait une assez bonne corrélation entre l'augmentation de ce score et la probabilité d'isoler un SGA en culture de prélèvement de gorge [7]. Dans les 3 études pédiatriques publiées sur l'effet de spectre du TDR pour le diagnostic des angines à SGA, il a été montré que la sensibilité du TDR augmentait avec le score de McIsaac. Par exemple, dans l'étude d'Edmonson et al. portant sur 1184 patients nord-américains âgés de moins de 24 ans, la probabilité pré-test de SGA augmentait de 9 % à 50 % et la sensibilité du TDR de 67 % à 88 % chez les enfants ayant un score de McIsaac à 1 et 4, respectivement [2]. Les auteurs de cette étude concluent que le TDR n'est pas assez fiable chez les enfants à bas risque de SGA et que les TDR négatifs devraient être confirmés par une culture. Toutefois, la portée de ces 3 études est limitée car elles présentaient toutes des défauts méthodologiques. Par exemple, dans l'étude d'Edmonson et al., les TDR n'étaient pas réalisés de manière systématique en cas d'angine et cela a probablement conduit à une sélection des patients les plus graves cliniquement (biais d'indication) [2]. Dans l'étude de Hall et al., seuls les TDR négatifs étaient contrôlés par culture, ce qui introduit un biais de vérification partielle ayant ten-

\*Auteur correspondant :  
e-mail : jeremie.cohen@inserm.fr

dance à surévaluer la prévalence du SGA et la sensibilité du TDR [3]. Dans l'étude de Tanz *et al.*, les auteurs concluaient à un effet de spectre sur la sensibilité du TDR sur la base d'une sensibilité plus élevée chez les patients ayant un score de Mclsaac > 2 mais cela n'était pas testé statistiquement [4]. Aucune de ces études n'a recherché d'éventuelles variations simultanées de sensibilité, spécificité, rapports de vraisemblance et valeurs prédictives, ni étudié de manière concomitante d'éventuelles variations de sensibilité du TDR en fonction de l'inoculum bactérien.

### 3. Étude ACTIV

En raison des limites des études antérieures, une évaluation de l'effet du spectre et de l'effet de l'inoculum bactérien a été menée à partir des données d'une étude prospective multicentrique évaluant les performances diagnostiques d'un TDR (StreptAtest, Dectrapharm, France) chez l'enfant de 3 à 15 ans. Les performances diagnostiques du TDR ont été comparées entre différents sous-groupes de patients définis par le score de Mclsaac où chacune de ses composantes cliniques et la sensibilité du TDR a été étudiée en fonction de l'inoculum bactérien, évalué en fonction de la densité des colonies de SGA en culture. Parmi les 785 enfants atteints d'angine inclus dans l'étude, la prévalence du SGA était de 36 %, dont 85 % d'inoculum fort. La sensibilité et la spécificité moyenne du TDR étaient de 87 % et de 95 %, respectivement. Nous avons mis en évidence des variations significatives de sensibilité, de spécificité, de rapports de vraisemblance et de valeur prédictive positive en fonction du score de Mclsaac et de ses composantes cliniques. Cependant, nous n'avons pas mis en évidence de variations de la valeur prédictive négative (VPN) du TDR, très stable autour de 93 %, indépendamment des caractéristiques des patients, c'est-à-dire indépendamment du spectre clinique de la maladie et de la population d'étude. La sensibilité du TDR était plus élevée chez les patients ayant un inoculum fort que chez les patients ayant un inoculum faible (95 % contre 40 %,  $p < 0,001$ ). En analyse multivariée, l'effet de spectre sur la sensibilité du TDR n'était significatif que chez les patients ayant un inoculum faible. En d'autres termes, la sensibilité du TDR était stable chez les patients ayant un inoculum fort.

### 4. Conclusion

L'étude ACTIV montre que l'effet de spectre touche la sensibilité et la spécificité du TDR. La VPN est stable par un mécanisme de compensation entre des variations de sensibilité, de spécificité et de prévalence du SGA. L'effet de spectre sur la sensibilité du TDR ne concerne que les patients ayant un faible inoculum bactérien,

c'est-à-dire un groupe de patients probablement constitué en grande partie de porteurs de SGA plutôt que de patients ayant une vraie angine à SGA. Ainsi, la sensibilité du TDR est élevée et stable en cas de fort inoculum bactérien.

Il est fréquemment avancé que le raisonnement diagnostique ne devrait pas être basé sur les valeurs prédictives d'un test rapportées dans une étude diagnostique car les valeurs prédictives varient avec la prévalence de la maladie dans la population d'étude. Pourtant, les valeurs prédictives sont capitales pour le clinicien. En effet, en pratique clinique, le médecin dispose du résultat du TDR et il cherche à savoir dans quelle mesure il peut s'y fier pour baser sa décision de traiter ou pas par antibiotique son patient. Les résultats de l'étude ACTIV montrent que la valeur prédictive négative du TDR est stable, ce qui permet d'interpréter les résultats négatifs du test indépendamment du profil clinique du patient et de son score de Mclsaac. Ainsi, bien que les performances diagnostiques du TDR soient soumises à un effet de spectre, ces variations n'affectent pas la VPN, et un résultat négatif du TDR rend très improbable le diagnostic d'angine à SGA de manière fiable et stable. La confirmation systématique des TDR négatifs par une mise en culture de prélèvement de gorge n'est donc pas nécessaire.

### Références

- [1] Gerber MA, Shulman ST. Rapid diagnosis of pharyngitis caused by group A streptococci. *Clin Microbiol Rev* 2004;17:571-80.
- [2] Edmonson MB, Farwell KR. Relationship between the clinical likelihood of group a streptococcal pharyngitis and the sensitivity of a rapid antigen-detection test in a pediatric practice. *Pediatrics* 2005;115:280-5.
- [3] Hall MC, Kieke B, Gonzales R, *et al.* Spectrum bias of a rapid antigen detection test for group A beta-hemolytic streptococcal pharyngitis in a pediatric population. *Pediatrics* 2004;114:182-6.
- [4] Tanz RR, Gerber MA, Kabat W, *et al.* Performance of a rapid antigen-detection test and throat culture in community pediatric offices: implications for management of pharyngitis. *Pediatrics* 2009;123:437-44.
- [5] Kurtz B, Kurtz M, Roe M, *et al.* Importance of inoculum size and sampling effect in rapid antigen detection for diagnosis of *Streptococcus pyogenes* pharyngitis. *J Clin Microbiol* 2000;38:279-81.
- [6] Mulherin SA, Miller WC. Spectrum bias or spectrum effect? Subgroup variation in diagnostic test evaluation. *Ann Intern Med* 2002;137:598-602.
- [7] Mclsaac WJ, White D, Tannenbaum D, *et al.* A clinical score to reduce unnecessary antibiotic use in patients with sore throat. *CMAJ* 1998;158:75-83.

JOURNÉES PARISIENNES DE PÉDIATRIE

*Comité d'organisation :*

Y. AUJARD, P. BOUGNÈRES, B. CHABROL, G. CHÉRON,  
G. DESCHENES, D. DEVICTOR, A. FISCHER, J. GAUDELUS,  
D. GENDREL, E. GRIMPREL, P. LABRUNE, F. LECLERC  
G. LEVERGER, M. TARDIEU

*Journées Parisiennes  
de Pédiatrie 2012*

Vendredi 5 et samedi 6 octobre 2012

## TEST DE DIAGNOSTIC RAPIDE DU STREPTOCOQUE DU GROUPE A DANS LES ANGINES DE L'ENFANT

par

J. F. COHEN, C. LEVY, P. MARIANI-KURKDJIAN, M. CHALUMEAU,  
E. BINGEN, R. COHEN

### INTRODUCTION

L'enjeu clinique principal des angines de l'enfant est la distinction entre les infections virales et les 20 à 40 % d'infections à streptocoque  $\beta$ -hémolytique du groupe A (SGA) qui justifient d'une antibiothérapie dont les objectifs sont de diminuer l'incidence des complications suppuratives et non-suppuratives liées aux infections à SGA, la durée des symptômes et la dissémination du germe à l'entourage [1]. Chez l'enfant, il est recommandé dans la plupart des pays de baser l'identification du SGA sur un prélèvement de gorge couplé soit au test de référence qui est la mise en culture sur une gélose au sang durant 48 heures dans un laboratoire de bactériologie, soit à un test de diagnostic rapide (TDR) [2].

La place du TDR dans les recommandations de prise en charge des angines de l'enfant est variable entre les pays occidentaux [2]. Les recommandations françaises sont de réaliser un TDR chez tous les enfants ayant une angine entre 3 et 15 ans [3], alors que l'usage du TDR est clairement non recommandé dans certains pays (Belgique, Pays-Bas, Canada) [2]. D'autres auteurs proposent des stratégies d'utilisation sélective du TDR sur la base de critères cliniques [4]. Ces différences à l'échelle internationale peuvent être expliquées en partie par une appréciation variable des avantages et des limites du TDR.

Afin d'apporter des connaissances nouvelles sur le diagnostic des angines de l'enfant et notamment la place du TDR, une étude a été réalisée par les pédiatres du réseau ACTIV chez 1482 enfants ayant une angine et 294 enfants asymptomatiques entre octobre 2009 et juin 2011 [5]. Pour chaque enfant, une fiche de renseignements cliniques a été remplie et un double prélèvement de gorge a été réalisé. Le premier écouvillon servait à réaliser un TDR au cabinet du pédiatre et le deuxième écouvillon était envoyé au laboratoire de bactériologie de l'Hôpital Robert-Debré pour mise en culture selon une technique de référence, à l'aveugle des éléments cliniques et des résultats du TDR.

## AVANTAGES DU TDR

### Un test simple, peu coûteux, au résultat immédiat

La plupart des TDR reposent sur la détection du polysaccharide C du groupe A, un antigène de paroi spécifique du SGA. Les bandelettes immuno-chromatographiques sont les TDR les plus répandus en pratique clinique car ils présentent le meilleur rapport performances-praticabilité [6]. La réalisation du TDR, contrairement à celle de la culture ou des techniques de PCR, ne nécessite pas de matériel ni de compétence technique particulière. Le résultat est disponible en quelques minutes alors que le résultat d'une culture au laboratoire nécessite par définition 24 à 48 heures. Un autre avantage du TDR est son coût faible. Depuis 2002, la Caisse Nationale d'Assurance Maladie le met gratuitement à disposition des médecins libéraux et des services hospitaliers. Le coût facturé par le fabricant est de 0,79 €/test, à comparer au coût d'une antibiothérapie comme l'amoxicilline pendant 6 jours, de l'ordre de 3 €.

### Des performances diagnostiques globales élevées

Les performances diagnostiques globales des TDR immuno-chromatographiques sont très élevées. Leur rapport de vraisemblance positif est habituellement  $\geq 15$  et leur rapport de vraisemblance négatif  $\leq 0,15$  [7,8]. En pratique, cela signifie qu'en cas de TDR positif, la probabilité post-test de SGA (c'est-à-dire une fois le résultat du test connu) dépasse les 90 % alors qu'en cas de TDR négatif, la probabilité post-test de SGA est autour de 5 % (Figure 1) [7]. A l'inverse, aucun signe clinique ne permet de faire la distinction entre angine virale et angine à SGA avec une performance correcte [9]. De même, dans l'étude conduite par les pédiatres d'ACTIV, la règle de décision clinique composite basée sur le score clinique de McIsaac n'avait une spécificité que de 52 % (Figure 2) [8].

### Une spécificité proche de 100 %

La spécificité des TDR est habituellement de l'ordre de 95 %, il y a donc 5 % de faux-positifs [10]. Dans l'étude d'ACTIV, il a été montré que 76 % des faux-positifs du TDR avaient une PCR SGA positive. Les enfants avec un résultat faux-positif du TDR étaient donc très majoritairement colonisés ou infectés par le SGA. De plus, 75 % des enfants avec un résultat faux-positif du TDR avaient également une culture positive à *Staphylococcus aureus*. Une hypothèse explicative de ces faux-positifs du TDR serait la présence inhibitrice de *S.aureus* via des mécanismes d'interférences bactériennes. La spécificité rapportée du TDR est donc sous-estimée et peut être considérée en pratique comme proche de 100 % [11].

## LIMITES DU TDR

Une des limites du TDR fréquemment alléguée est que ses performances diagnostiques peuvent varier en fonction de différents facteurs : présentation clinique des patients, charge bactérienne présente dans la gorge, portage pharyngé et médecin réalisant le test.

### Effet de spectre du TDR

Contrairement à ce qui est classiquement enseigné, la sensibilité et la spécificité d'un test diagnostique ne sont pas fixes, mais peuvent varier en fonction du spectre clinique de la maladie. On parle alors d'effet de spectre [12]. Il a été montré que la sensibilité du TDR variait en fonction de la présentation clinique des patients ayant une angine, la sensibilité du TDR augmentant avec le score de McIsaac [13]. Cela a été confirmé dans l'étude d'ACTIV : la sensibilité du TDR augmentait de 75 % à 90 % en fonction du score de McIsaac [8]. Cependant, cette étude a montré pour la première fois que l'effet de spectre n'affectait pas la valeur prédictive négative du TDR, qui restait stable, autour de 92 %, quelle que soit la présentation clinique des patients. Ainsi, un résultat négatif du TDR rendait très improbable le diagnostic d'angine à SGA indépendamment du spectre clinique, confirmant que la mise en culture systématique en cas de TDR négatif ne semblait pas nécessaire [3,14].

### Effets de l'inoculum et du portage

L'inoculum bactérien est la quantité de bactéries présente dans l'écouvillon ayant servi à réaliser le prélèvement de gorge. On considère que le nombre de colonies  $\beta$ -hémolytiques présentes sur la boîte de culture est un reflet de l'inoculum. Le seuil de positivité *in vitro* du TDR correspond à un inoculum bactérien de 10<sup>5</sup>-10<sup>7</sup> unités formant colonie/mL [6]. Plusieurs études ont montré que la sensibilité du TDR augmentait avec l'inoculum bactérien [15,16]. Cela est parfois considéré comme un défaut risquant d'entraîner un résultat faux-négatif du TDR chez les patients réellement infectés par le SGA mais avec un faible inoculum [15]. D'autres auteurs considèrent que les patients ayant de faibles inocula bactériens seraient des porteurs de SGA ayant une angine virale concomitante plutôt que des enfants réellement infectés par le SGA, minimisant le poids de ces faux-négatifs du TDR [17].

Dans l'étude réalisée par les pédiatres d'ACTIV, 11 % des enfants asymptomatiques étaient des porteurs sains de SGA [5]. La sensibilité du TDR augmentait avec l'inoculum bactérien et était plus élevée dans le groupe angine que chez les porteurs sains : 89 % contre 41 % respectivement. Cela était en partie expliqué par le fait que les inocula faibles étaient plus fréquents en cas de portage sain que d'angine : 66 % contre 14 %. Cependant après ajustement sur l'inoculum bactérien par analyse multivariée, nous avons montré que la sensibilité du TDR était toujours plus élevée en cas d'angine que de portage sain. Cela nous a empêché d'estimer précisément dans quelle mesure les résultats faux-négatifs du TDR seraient expliqués par des enfants ayant un portage de SGA et une angine virale concomitante. Notre principale hypothèse est que cette différence de sensibilité pourrait être associée d'une part à des modifications locorégionales liées à l'infection et à l'inflammation et d'autre part à des différences dans la distribution des génotypes *emm* de SGA présents chez les enfants ayant une angine et les porteurs sains.

### Effet médecin

Une autre limite du TDR est que sa sensibilité varie en fonction de la personne réalisant le test, ce qu'on appelle un « effet médecin ». Cela a été montré dans plusieurs études et l'hypothèse principale est que ces variations sont dues à une différence de technique d'écouvillonnage probablement liée à l'expérience du médecin et conduisant à des différences

d'inoculum bactérien [18]. Dans l'étude réalisée par les pédiatres d'ACTIV, nous avons confirmé qu'il existait des variations « inter-investigateurs » importantes de sensibilité du TDR, allant de 56 % à 96 %, même dans un groupe de professionnels motivés et très entraînés [5]. Grâce à une modélisation multivariée multiniveau, nous avons montré que ces variations étaient indépendantes de variables liées aux patients (et donc indirectement du recrutement des investigateurs) comme l'inoculum bactérien et le statut clinique des patients (angine ou porteurs sains) mais au contraire très bien expliquées par des variables liées au médecin (type de pratique clinique, connaissances et croyances sur l'angine à SGA).

### UNE SENSIBILITÉ DE 90 % : AVANTAGE OU LIMITÉ ?

La sensibilité des TDR immuno-chromatographiques est située entre 85 % et 90 % [10]. Certains auteurs estiment que cette sensibilité est insuffisante et que les résultats négatifs du TDR devraient être contrôlés par une mise en culture de prélèvement de gorge [2]. Pourtant, il semble raisonnable de ne pas avoir une cible de 100 % de sensibilité pour le diagnostic des angines de l'enfant. En effet, en raison du portage pharyngé asymptomatique de SGA, qui concerne en moyenne 12 % des enfants d'âge scolaire [19], il est certain qu'un certain nombre d'enfants avec une angine et une culture positive à SGA sont en fait des porteurs avec une angine virale concomitante. Un TDR ayant 100 % de sensibilité (en comparaison à la culture) conduirait inéluctablement à traiter inutilement des porteurs sains avec une angine virale. De plus, la plus redoutée des complications non-suppuratives des angines à SGA, le rhumatisme articulaire aigu, qui était à l'origine du dogme d'avoir une sensibilité de 100 %, est devenu très rare en Europe de l'Ouest ( $\approx 10$  cas/an en France) [20]. Ces différents éléments sont à l'origine, entre autre, des dernières recommandations européennes de se fier aux résultats négatifs du TDR sans nécessité de confirmation systématique par une mise en culture [14].

### LE TDR EN PRATIQUE

Diminuer la consommation d'antibiotiques est un objectif majeur pour les années à venir en raison de l'augmentation des résistances bactériennes. Le TDR est un test diagnostique simple et très performant qui permet de sélectionner les enfants pouvant réellement bénéficier d'une antibiothérapie ciblée. Les résultats des études réalisées par les pédiatres d'ACTIV permettent de tirer les conclusions suivantes :

- il est nécessaire de réaliser un TDR chez tous les enfants présentant une angine, même pauci-symptomatique, car l'examen clinique n'est pas suffisamment discriminant ;
- les cliniciens peuvent se fier aux résultats positifs du TDR car la spécificité du TDR est proche de 100 % ;
- les cliniciens peuvent se fier aux résultats négatifs du TDR (sans confirmation par la culture) indépendamment de la présentation clinique des patients s'ils acceptent un taux de faux-négatifs de l'ordre de 10 % ;
- la sensibilité du TDR dépend de manière importante de la personne réalisant le test et il serait intéressant de pouvoir proposer aux cliniciens des stratégies d'auto-évaluation et d'amélioration des pratiques professionnelles ciblant la réalisation du TDR.

## FINANCEMENTS

- Sources internes : l'étude a été financée par ACTIV.
- Sources externes : travail financé partiellement par Dectrapharm, fabricant d'un TDR SGA ; JFC était financé par l'ARS Ile-de-France et la bourse Guigoz SFP - Groupe de Pédiatrie Générale - Groupe de Recherches Epidémiologiques en Pédiatrie. Les financeurs externes n'ont participé à aucune des étapes suivantes : conception, réalisation, analyses, rédaction des manuscrits.

## REMERCIEMENTS

A : M Boucherat, F de La Rocque, I Ramay, D Menguy, S Tortorelli, M Fernandes, M Benani, F Corrad, P Deberdt, A Elbez, M Goldrey, J Gosselin, M Koskas, P Martin, AS Michot, N Panis, D Qutob, C Romain, O Romain, C Schlemmer, F Thollot, A Wollner (ACTIV) ; A Liboz, C d'Humières, S Liguori, P Bidet (Service de microbiologie, hôpital Robert-Debré, Paris).

### AUTEURS :

Jérémie F. Cohen<sup>1,2</sup>, Corinne Levy<sup>3</sup>, Patricia Mariani-Kurkdjian<sup>4</sup>, Martin Chalumeau<sup>1,2</sup>, Edouard Bingen<sup>4</sup>, Robert Cohen<sup>3,5</sup>

<sup>1</sup>Inserm U953, Inserm, Paris

<sup>2</sup>Service de Pédiatrie Générale, Hôpital Necker-Enfants-Malades, AP-HP, Université Paris Descartes, PRES Sorbonne Paris Cité, Paris

<sup>3</sup>ACTIV, Association Clinique et Thérapeutique Infantile du Val-de-Marne, Saint-Maur-des-Fossés

<sup>4</sup>AP-HP, Laboratoire de Microbiologie Hôpital Robert Debré, F-75935 Paris, France, Univ Paris Diderot, PRES Sorbonne Paris Cité, F-75505 Paris, France

<sup>5</sup>Service de Microbiologie, Centre Hospitalier Intercommunal de Créteil, Créteil

### AUTEUR CORRESPONDANT :

Martin Chalumeau - martin.chalumeau@nck.aphp.fr

FIGURES

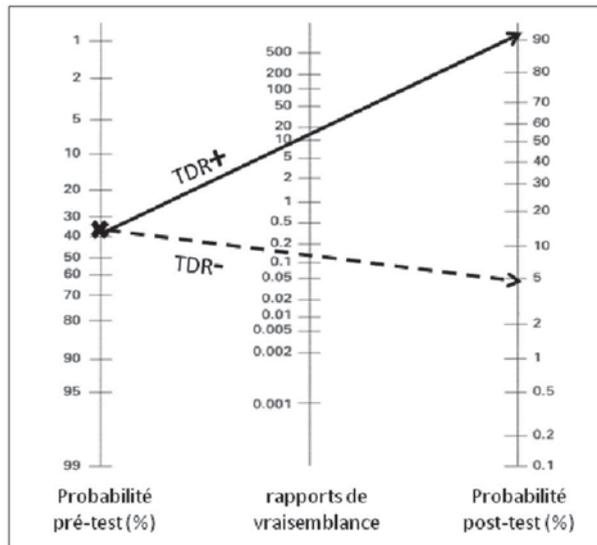


Figure 1. Nomogramme de Fagan montrant la relation entre la probabilité pré-test (prévalence) et les probabilités post-test de SGA en fonction du résultat du test de diagnostic rapide (TDR).

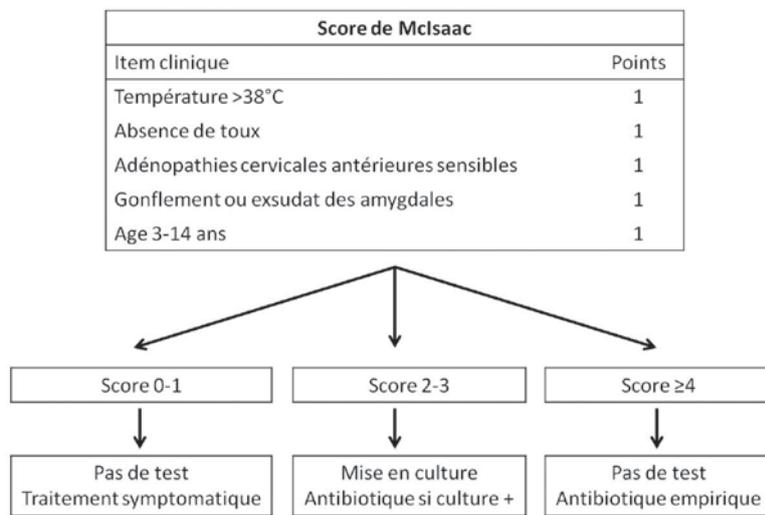


Figure 2. Score de McIsaac et règle de décision clinique qui en découle

## RÉFÉRENCES

- [1] Del Mar CB, Glasziou PP, Spinks AB. Antibiotics for sore throat. *Cochrane Database Syst Rev* 2006;CD000023.
- [2] Mathys J, De Meyere M, van Driel ML, et al. Differences among international pharyngitis guidelines: not just academic. *Ann Fam Med* 2007;5:436-43.
- [3] SFP-SPLIF-GPIP. Antibiothérapie par voie générale en pratique courante dans les infections respiratoires hautes de l'adulte et de l'enfant [en ligne]. 2011. Consulté le 15/05/2012. Disponible sur: <http://www.infectiologie.com>.
- [4] McIsaac WJ, Kellner JD, Aufricht P, et al. Empirical validation of guidelines for the management of pharyngitis in children and adults. *JAMA* 2004;291:1587-95.
- [5] Cohen JF, Levy C, Bidet P, et al. Sensibilité du test de diagnostic rapide du streptocoque du groupe A chez les porteurs sains et les enfants ayant une angine. *Arch Pediatr* 2012;19 Hors-série 2:143-4.
- [6] Charlier-Bret N, Boucher B, Poyart C, et al. Tests de diagnostic rapide des angines à streptocoque du groupe A : étude comparative de la sensibilité et de la praticabilité de 16 dispositifs médicaux de diagnostic in vitro dans le cadre d'un contrôle du marché effectuée par l'Afssaps en juillet 2002. *Pathol Biol (Paris)* 2004;52:438-43.
- [7] Cohen R, Martinot A. Angine à streptocoque du groupe A : enjeux nationaux et internationaux. *Arch Pediatr* 2012;19 Hors-série 2:141-2.
- [8] Cohen JF, Chalumeau M, Levy C, et al. Spectrum and inoculum size effect of a rapid antigen detection test for group A streptococcus in children with pharyngitis. *PLoS One* 2012;7:e39085.
- [9] Shaikh N, Swaminathan N, Hooper EG. Accuracy and precision of the signs and symptoms of streptococcal pharyngitis in children: a systematic review. *J Pediatr* 2012;160:487-93 e3.
- [10] Gerber MA, Shulman ST. Rapid diagnosis of pharyngitis caused by group A streptococci. *Clin Microbiol Rev* 2004;17:571-80.
- [11] Cohen JF, Cohen R, Bidet P, et al. Faux positifs du test de diagnostic rapide du streptocoque du groupe A dans les angines de l'enfant. *Arch Pediatr* 2012;19 Hors-série 2:145-6.
- [12] Mulherin SA, Miller WC. Spectrum bias or spectrum effect? Subgroup variation in diagnostic test evaluation. *Ann Intern Med* 2002;137:598-602.
- [13] Edmonson MB, Farwell KR. Relationship between the clinical likelihood of group A streptococcal pharyngitis and the sensitivity of a rapid antigen-detection test in a pediatric practice. *Pediatrics* 2005;115:280-5.
- [14] Pelucchi C, Grigoryan L, Galeone C, et al. Guideline for the management of acute sore throat. *Clin Microbiol Infect* 2012;18 Suppl 1:1-28.
- [15] Gerber MA, Randolph MF, Chanatry J, et al. Antigen detection test for streptococcal pharyngitis: evaluation of sensitivity with respect to true infections. *J Pediatr* 1986;108:654-8.
- [16] Kurtz B, Kurtz M, Roe M, et al. Importance of inoculum size and sampling effect in rapid antigen detection for diagnosis of *Streptococcus pyogenes* pharyngitis. *J Clin Microbiol* 2000;38:279-81.
- [17] Roddey OF, Jr. Quantitation of group A beta-hemolytic streptococci in throat culture. *J Clin Microbiol* 1991;29:1279-80.
- [18] Fox JW, Cohen DM, Marcon MJ, et al. Performance of rapid streptococcal antigen testing varies by personnel. *J Clin Microbiol* 2006;44:3918-22.
- [19] Shaikh N, Leonard E, Martin JM. Prevalence of streptococcal pharyngitis and streptococcal carriage in children: a meta-analysis. *Pediatrics* 2010;126:e557-64.
- [20] Olivier C, Portier H, Cohen R, et al. Rhumatisme articulaire aigu : résultats d'une enquête nationale (1995-1997). *Bulletin Epidémiologique Hebdomadaire* 1999;12:45-7.

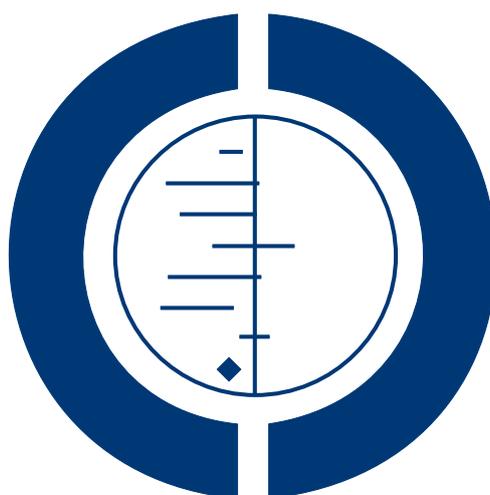


---

ANNEXE 5. REVUE SYSTEMATIQUE ET META-ANALYSE DES  
PERFORMANCES DIAGNOSTIQUES DU TDR (PROTOCOLE)

# Rapid antigen detection test for group A streptococcus in children with pharyngitis (Protocol)

Cohen JF, Cohen R, Chalumeau M



**THE COCHRANE  
COLLABORATION®**

This is a reprint of a Cochrane protocol, prepared and maintained by The Cochrane Collaboration and published in *The Cochrane Library* 2013, Issue 4

<http://www.thecochranelibrary.com>



---

Rapid antigen detection test for group A streptococcus in children with pharyngitis (Protocol)  
Copyright © 2013 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

## TABLE OF CONTENTS

HEADER . . . . .	1
ABSTRACT . . . . .	1
BACKGROUND . . . . .	2
OBJECTIVES . . . . .	4
METHODS . . . . .	4
ACKNOWLEDGEMENTS . . . . .	6
REFERENCES . . . . .	7
ADDITIONAL TABLES . . . . .	9
APPENDICES . . . . .	12
CONTRIBUTIONS OF AUTHORS . . . . .	13
DECLARATIONS OF INTEREST . . . . .	13
SOURCES OF SUPPORT . . . . .	13

[Diagnostic Test Accuracy Protocol]

# Rapid antigen detection test for group A streptococcus in children with pharyngitis

Jérémie F Cohen<sup>1</sup>, Robert Cohen<sup>2,3</sup>, Martin Chalumeau<sup>1</sup>

<sup>1</sup>INSERM U953 and Department of Pediatrics, Necker Hospital, AP-HP and Paris Descartes University, Paris, France. <sup>2</sup>Association Clinique et Thérapeutique du Val-de-Marne (ACTIV), Saint Maur, France. <sup>3</sup>Department of Microbiology, Centre Hospitalier Inter-communal de Créteil (CHIC), Créteil, France

Contact address: Martin Chalumeau, INSERM U953 and Department of Pediatrics, Necker Hospital, AP-HP and Paris Descartes University, Paris, France. [martin.chalumeau@gmail.com](mailto:martin.chalumeau@gmail.com). [martin.chalumeau@nck.aphp.fr](mailto:martin.chalumeau@nck.aphp.fr).

**Editorial group:** Cochrane Acute Respiratory Infections Group.

**Publication status and date:** New, published in Issue 4, 2013.

**Citation:** Cohen JF, Cohen R, Chalumeau M. Rapid antigen detection test for group A streptococcus in children with pharyngitis. *Cochrane Database of Systematic Reviews* 2013, Issue 4. Art. No.: CD010502. DOI: 10.1002/14651858.CD010502.

Copyright © 2013 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

## ABSTRACT

This is the protocol for a review and there is no abstract. The objectives are as follows:

The primary objective is to determine the diagnostic accuracy of RADTs for GAS in children with pharyngitis.

A secondary objective is to assess the relative diagnostic accuracy of EIA and OIA tests by indirect and direct comparison.

## BACKGROUND

### Target condition being diagnosed

Pharyngitis is defined as an acute inflammation of the pharynx, tonsils or both. Viruses are the most common cause of pharyngitis but the bacterium most frequently identified during acute pharyngitis is *Streptococcus pyogenes* (*S. pyogenes*), also known as group A  $\beta$ -haemolytic streptococcus (GAS). GAS is estimated to account for 20% to 40% of cases of pharyngitis in children and 5% to 15% in adults (Shaikh 2010; Wessels 2011). The estimated number of cases of GAS pharyngitis in children is 450 million/year worldwide (Carapetis 2005a). GAS pharyngitis is ubiquitous but is more frequent in low-income countries. Most cases are benign and self limiting within a week but suppurative complications (cervical lymphadenitis, retropharyngeal abscess, peritonsillar cellulitis or abscess (quinsy), sinusitis, acute otitis media and mastoiditis) or non-suppurative post-streptococcal diseases (acute rheumatic fever and rheumatic heart disease, acute glomerulonephritis, Sydenham's chorea, scarlet fever, streptococcal toxic shock syndrome and paediatric autoimmune neuropsychiatric disorder associated with group A streptococci) can occur (Gerber 2005).

In low-income countries, rheumatic heart disease remains the most commonly acquired heart disease in children, adolescents and young adults: a recent estimate of the number of deaths from rheumatic heart disease is 233,000 per year worldwide (Carapetis 2005a). In high-income countries acute rheumatic fever and rheumatic heart disease are rare (e.g.,  $\leq 10$  cases/year/100,000 children for acute rheumatic fever) (Carapetis 2005b; Seckeler 2011) because of improvements in living conditions, hygiene, increased antibiotic usage, increased access to primary care providers and changes in GAS epidemiology (Carapetis 2007). As a result, the public health goal is shifting from preventing rare GAS complications to minimising inappropriate use of antibiotics.

Many experts recommend the prescription of antibiotics for children with GAS-suspected or GAS-proven pharyngitis (Matthys 2007). The goal of antibiotic treatment is to reduce the individual risk of suppurative or non-suppurative complications, the duration of symptoms and the spread of the condition (Spinks 2011). Correct identification of GAS ensures against not missing GAS-positive cases that can lead to complications. The correct exclusion of GAS ensures against unnecessary use of antibiotics (thus reducing the incidence of adverse drug reactions, antibiotic resistance and associated costs).

There is a lack of consensus on the most suitable diagnostic method for GAS in children with pharyngitis and the 'standard' diagnostic practice varies greatly amongst countries. Because the signs and symptoms of GAS and viral pharyngitis overlap broadly (Shaikh 2011), most guidelines that recommend antibiotic treatment of GAS also recommend confirmation of the presence of GAS on the basis of a throat swab (Matthys 2007). However, throat swabs are

explicitly not recommended in some countries (e.g., United Kingdom, Belgium and The Netherlands) (Matthys 2007). International discrepancies might be explained by academic reasons and 'clinical traditions', different targets of sensitivity and specificity because of local epidemiological differences (i.e., rheumatic fever and rheumatic heart disease prevalence), international differences in health systems and policies, and the sparseness of recent data on the incidence of GAS complications and the efficacy of antibiotic treatment for their prevention.

The standard criterion for the diagnosis of GAS in children with pharyngitis is throat culture on a blood agar plate in a microbiology laboratory (AAP 2009). The major advantage of laboratory throat culture is its detection of GAS from swabs with a very low number of bacteria, but the major limitation is the 48-hour delay in obtaining results. In addition, throat cultures cannot distinguish true GAS infection from GAS carriage with intercurrent viral pharyngitis. Asymptomatic pharyngeal GAS carriage is usually defined as positive throat culture results for GAS without a GAS-specific immune response (anti-streptolysin O and anti-DNase B antibodies) (Tanz 2007). Asymptomatic GAS carriage occurs in 10% to 15% of healthy children (Shaikh 2010) and does not require antibiotic treatment (Tanz 2007).

Agreement is lacking on the most suitable culture technique for diagnosing GAS in children with pharyngitis. Several parameters are likely to affect the sensitivity of the test (culture medium, atmosphere of incubation, duration of incubation, group A identification technique and the number of plates inoculated) (Kellogg 1990; Tanz 1997). These variables affect the diagnostic accuracy of the throat culture and thus the diagnostic accuracy of rapid antigen detection tests (RADTs) as compared to throat culture.

### Index test(s)

Simple RADTs were developed in the 1980s to provide an immediate indication for the clinician about the presence or absence of GAS in children with pharyngitis. RADTs do not require any special equipment and can be performed at the point of care with a throat swab (Cohen 2004). They can provide immediate results and are calibrated to produce binary results (positive or negative). In children, the reported sensitivity of RADTs is about 85% (Gerber 2004) but varies greatly amongst studies (from 66% (Van Limbergen 2006) to 99% (Harbeck 1993)) and the specificity is high and stable, about 95% (Gerber 2004). Because of this high specificity, most experts agree on prescribing antibiotics with positive RADT results, even if RADTs cannot differentiate GAS true infection from GAS carriage. However, the consequences of a negative RADT result depend on national guidelines. North American guidelines recommend backing up negative RADT results with throat culture to avoid not treating RADT false-negative cases (Bisno 2002; Gerber 2009) but most recent European guidelines recommend relying on negative RADT results without culture confirmation (Pelucchi 2012). In low-income countries, the clini-

cal consequences of RADT results might be the same as in high-income countries (treat RADT-positive cases only) but resources for testing might be limited and practices may vary from generalised empiric antibiotic treatment to selective antibiotic treatment or selective rapid testing based on clinical scoring systems (Joachim 2010; Steinhoff 2005; WHO 1995).

All available RADTs involve the detection of the Lancefield group A carbohydrate, a GAS-specific cell-wall antigen. Different immunologic techniques are available for carbohydrate detection (Gerber 2004), from older to most recent:

- Latex agglutination (LA) assay: the sample is placed in the presence of latex beads coupled with GAS-specific antibodies; the result is determined by observing the agglutination of the beads if they are related to the specific antigen in the sample. These first-generation tests are no longer used in clinical practice and will not be considered in this review.

- Enzyme immunoassay (EIA): the sample is placed at the end of a nitrocellulose strip and then migrates to an area where it forms an antigen-antibody complex. These second-generation tests are also known as immunochromatographic, sandwich or lateral-flow assays. They are the most widespread and most used RADTs in clinical practice.

- Optical immunoassay (OIA): the sample is placed on a silicon membrane in the presence of the reagent. The result is based on the change in optical properties of the inert membrane in the presence of an antigen-antibody complex. These third-generation tests seem to be more sensitive than EIAs but their use is limited because of their high cost.

## Clinical pathway

### Alternative test(s)

#### Office culture

Another test for the diagnosis of GAS in children with pharyngitis is a throat culture performed in the physician's office (office culture). Office culture has the same disadvantage as a laboratory culture (a 48-hour delay in obtaining results), with the major limitation being insufficient sensitivity (from 50% to 85%) (Battle 1971; Mondzac 1967; Rosenstein 1970; Tanz 2009; Wegner 1992). Office culture is almost completely abandoned and will not be considered in this review.

#### Streptococcal antibody tests

Assessment of GAS-specific antibodies is the traditional reference test to differentiate true GAS infection and GAS carriage. The most commonly used GAS-specific antibody assays tests are for anti-streptolysin O and anti-DNase B antibodies. Increased antibody titre assessment diagnoses true GAS infection better than

a single absolute titre assessment (Gerber 1986; Johnson 2010). Streptococcal antibody tests are not used for the diagnosis of GAS in children with pharyngitis because of the need for repeat blood samples. Moreover, the information about the kinetics of the immune response to GAS in children with pharyngitis is very limited and the most recent data show that the interpretation of streptococcal antibody test results is not straightforward (Johnson 2010). Therefore, their use is usually limited to documenting recent GAS infection in patients suspected of having GAS non-suppurative complications or to epidemiologic studies (Gerber 1986; Johnson 2010).

#### Clinical scoring systems

Clinical scoring systems have been developed to diagnose GAS on the basis of clinical grounds. The most popular of these scores are the Centor score (Centor 1981) and the McIsaac score (McIsaac 1998). The scores are based on assessing simple clinical criteria (history of fever, cough, tonsillar swelling or exudate, tender cervical adenopathy and age). Their use is recommended in adults but might be inappropriate in children; several authors have reported a lack of diagnostic accuracy in this population (Cohen 2012; Fischer Walker 2006; Shaikh 2011). Clinical scoring systems will not be considered in this review.

#### Rapid molecular biology assays

Rapid molecular biology assays for GAS in children with pharyngitis have been recently developed (Group A Streptococcus Direct Test; GenProbe Inc., San Diego, CA; and LightCycler Strep-A assay; Roche Applied Science, Indianapolis, IN) (Chapin 2002; Heelan 1996; Pokorski 1994; Uhl 2003). These techniques, based on DNA-rRNA hybridisation or PCR, are highly sensitive but are not currently used widely because of their cost, the need for highly specialised equipment and personnel and the two-hour delay in results (Gerber 2004). Molecular assays are not antigen-detection tests and will not be considered in this review.

## Rationale

Childhood pharyngitis is a significant public health problem with, on the one hand, suppurative and non-suppurative complications of GAS pharyngitis (especially acute rheumatic fever and rheumatic heart disease) and, on the other, costly diagnostic tests and unnecessary antibiotics. RADTs for GAS are now widely available and their use in children with pharyngitis might increase accurate diagnosis and reduce antibiotic consumption.

According to local clinical guidelines, RADTs may be used as stand-alone diagnostic tests in replacement of throat culture (e.g., in contexts where throat culture is unavailable or not used), or as triage tests, with negative results being supported by a throat culture. These international discrepancies might be explained in part

by persistent gaps in knowledge regarding the diagnostic accuracy of RADTs:

- What is the accuracy of RADTs for GAS in children with pharyngitis compared to the most consensual reference test (throat culture on a blood agar plate)?
- Are there significant differences in diagnostic accuracy between EIAs and OIAs?
- Which study-level factors could explain variations in diagnostic accuracy across clinical studies?

The questions of whether RADTs should be performed in all patients presenting with signs and symptoms of pharyngitis or only in selected patients on the basis of a clinical score (selective testing strategies), and whether clinical protocols that incorporate RADTs are sufficient to reduce antibiotic prescription, will not be addressed in this review. We aim to provide information to help clinicians and public health decision makers better define the precise role of RADTs in the diagnosis of GAS in children with pharyngitis on the basis of unbiased evidence.

## OBJECTIVES

The primary objective is to determine the diagnostic accuracy of RADTs for GAS in children with pharyngitis.

### Secondary objectives

A secondary objective is to assess the relative diagnostic accuracy of EIA and OIA tests by indirect and direct comparison.

## METHODS

### Criteria for considering studies for this review

#### Types of studies

We will include reports of cross-sectional studies reporting the diagnostic accuracy of one or more RADTs for the diagnosis of GAS in children with pharyngitis, with laboratory throat culture as the standard test. Reports of randomised controlled trials (RCTs) will also be included if we can extract 2 x 2 tables for children. Reports of studies involving the RADT result as part of the reference test will be included but we will investigate the effect of such studies by using this study characteristic (i.e., RADT as part of the reference test) as a potential criterion to carry out sensitivity analyses.

### Participants

We will include reports of studies of children or young adults (age  $\leq 21$  years, according to the upper limit chosen by the American Academy of Pediatrics) seeking ambulatory medical care because of a sore throat or with a diagnosis of pharyngitis, who provided a throat swab for a RADT and laboratory throat culture. Ambulatory care settings will include private physicians' offices, walk-in clinics, hospital outpatients, emergency departments and family medicine centres.

We will also include reports of studies with only a subgroup of participants eligible for inclusion in the review, provided that we can extract relevant data specific to that subgroup. Reports of studies will not be excluded on the basis of whether studies were performed in high-income or low-income countries because no data exist to support variations in the accuracy of RADTs according to this criterion.

### Index tests

We will include only studies of EIA or OIA RADTs for GAS in children with pharyngitis, including those no longer marketed.

### Target conditions

GAS in children with pharyngitis.

### Reference standards

Studies are required to diagnose GAS with throat culture on a blood agar plate in a microbiology laboratory used as the reference test. Several parameters may affect the accuracy of throat culture. For studies involving more than one throat culture technique (different medium, duration or atmosphere of incubation), we a priori chose to extract data related to the culture technique recommended by a panel of North American content experts, i.e., simple blood agar plate (versus selective or enriched media), incubation 48 hours total (versus 18 to 24 hours only), aerobic atmosphere (versus other) (Shulman 2000), in order to avoid data-driven approaches.

### Search methods for identification of studies

#### Electronic searches

We will search MEDLINE using the search strategy described in Appendix 1. The search strategy was developed in consultation with a Medical Librarian and the Trials Search Co-ordinator for the Acute Respiratory Infections Group and will be adapted to search EMBASE (Ovid) and Web of Science (which incorporates the Science Citation Index). We will not use any filter related to age because many RADT studies enrol adults and children and could provide extractable data for children. We will not use

methodological filters to identify diagnostic studies because such filters may result in omission of relevant studies (Leefflang 2006; Whiting 2011b). The searches will be run from 1980 onwards because RADTs were not available prior to this date. We will search the Cochrane Register of Diagnostic Test Accuracy Studies for relevant studies.

We will search the following databases to identify potentially relevant studies referenced in reviews and guidelines:

- the Cochrane Database of Systematic Reviews;
- DARE (Database of Abstracts of Reviews of Effects);
- the MEDION database (for Systematic Reviews of

Diagnostic Studies); and

- TRIP (Turning Research Into Practice).

The search strategy will also include names of some commercial kits from the most common manufacturers: Test Pack Strep A (Abbott), ICON Strep A (Beckman Coulter), Link 2 Strep A Rapid Test (Becton Dickinson), Acceava Strep A (Inverness Medical), OSOM Strep A (Genzyme), Poly Stat Strep A (Polymedco) and QuickVue Strep A (Quidel).

### Searching other resources

We will handsearch reference lists of included articles and any relevant review articles identified through the search and the 'related articles' function in PubMed (20 first related articles of each included article) for eligible articles. We will use Science Citation Index and Google Scholar to search for reports that cite included articles.

We will electronically search the following databases for 'grey' literature (unpublished studies, including conference proceedings and reports):

- OpenSIGLE database; and
- OAISTER Database.

We will contact manufacturers of the most common RADTs to seek additional or unpublished studies. Manufacturers include Abbott, Beckman Coulter, Becton Dickinson, Genzyme, Inverness Medical, Polymedco and Quidel.

## Data collection and analysis

### Selection of studies

We will consider studies published in any language. One review author (JFC) will exclude studies that are not related to pharyngitis or RADT on the basis of the titles and abstracts identified by the search strategy. Two review authors (JFC, MC) will retrieve the full text of relevant articles and independently evaluate them for inclusion by using a *pro forma* as a guide. One review author (RC) will act as arbiter in case of discrepancies between two review authors (JFC, MC) who will discuss the inclusion of the studies.

We will select the most recent or most complete report in cases of multiple reports for a given study or when we cannot exclude the possibility of overlapping populations. The study selection process will involve use of ReSyWeb, a web service developed at the French Cochrane Centre and INSERM U872. The tool handles the importation of references from multiple databases, the semi-automatic deletion of duplicate records of the same report, the selection of studies by independent reviewers, consensus procedures for disagreements and the linkage of multiple reports of the same study. Finally, the web service automatically produces a flowchart to report the search process. We will record and report reasons for excluding studies but we will not report their references.

### Data extraction and management

Two authors (JFC, MC) will independently extract a standard set of data from each study by using a pre-specified data extraction form. We will extract the number of true positives, true negatives, false positives and false negatives for each index test evaluated in each study to construct 2 x 2 tables. If such data are not provided by the trial authors, we will attempt to contact them to construct the 2 x 2 table for the study population or the pre-specified subgroups. Otherwise, we will calculate the number of true positives, true negatives, false positives and false negatives from the summary estimates of sensitivity and specificity of the index test, if available. If reported, we will extract data on the number of undetermined or uninterpretable RADT results but they will be excluded from the analysis. For studies for which only a subgroup of patients will be included in the review, we will extract, analyse and present data for this subgroup only. If some data are unclear or missing, we will attempt to contact study authors to obtain additional data. The data extraction will be cross-checked and we will resolve discrepancies by discussion until a consensus is reached or we will consult the third review author (RC). See Table 1 for a description of which data will be extracted for each study.

### Assessment of methodological quality

Two review authors (JFC, MC) will independently assess the methodological quality of each study using a four-domain tool adapted from QUADAS-2 (Whiting 2011a). We tailored the quality assessment tool to our review question. We developed review-specific guidance on how to assess each signalling question and how to use this information to judge the risk of bias and applicability. We refined the tool until satisfactory inter-rater agreement was achieved. We will summarise the methodological quality assessment in tables for each study. See Table 2.

### Statistical analysis and data synthesis

We will enter data of the 2 x 2 tables into RevMan 2012 and we will plot estimates of sensitivity and specificity on forest plots and in the receiver-operating characteristic (ROC) space to represent the variability in diagnostic test accuracy within and between studies.

The hierarchical bivariate model described by Reitsma (Reitsma 2005) will be fitted by use of Stata/SE, which will allow for calculating summary estimates of sensitivity and specificity and the associated 95% confidence intervals (CIs). We will also report the estimate of correlation between sensitivity and specificity ( $\rho$ ). We will put the results from the bivariate model into RevMan 2012 to provide plots of the estimated curve(s) or summary point(s) and confidence region(s), superimposed on the study-specific estimates of sensitivity and specificity in the ROC space.

We will include the same study in the same meta-analysis more than once if needed, i.e., if one study reports different index tests. We will not present results in groups according to commercial test name because RADTs can be considered the same diagnostic test (all RADTs detect the same antigen using immunology-based techniques) and because there might be too many different commercial kits (> 50).

We will present results according to test type (EIA versus OIA) and will also attempt to directly compare the accuracy of EIA versus OIA RADTs using only studies that directly compare the two techniques.

### Investigations of heterogeneity

We will initially visually inspect the forest plots and ROC space to check for heterogeneity between study results. To investigate sources of heterogeneity, we will incorporate covariates in the bivariate model, i.e., meta-regression. We will assess the significance of the difference in covariate by likelihood ratio test comparing the bivariate model with and without the covariate. A P value of less than 0.05 will be used to denote statistical significance. With a significant test result, we will assess effects of covariates on sensitivity and specificity separately by testing the significance of the change in -2 log-likelihood of the model with or without corresponding terms. If the bivariate model does not converge or produces unstable parameter estimates, we will simplify the bivariate model by assuming fixed-effect estimates.

If sufficient studies are available, we will address the five following sources of heterogeneity by adding variables to the meta-analysis model:

#### a. Effect of test type

Some authors suggested that OIA may be more sensitive than EIA tests (Gerber 2004). Therefore, we will try to indirectly compare the RADT tests by using test type as a categorical covariate in the models (EIA versus OIA).

#### b. Effect of the reference standard

In this review, the reference standard will be throat culture on a blood agar plate. However, the following parameters may affect the accuracy of throat culture on blood agar and therefore the accuracy of RADTs as compared with throat culture: culture medium (standard versus inhibitory), atmosphere of incubation (aerobic versus

other), duration of incubation (18 to 24 hours versus 48 hours total), use of an enrichment broth before plating (yes/no), group A identification technique (latex agglutination versus bacitracin testing) and number of plates inoculated (one versus more than one). We will assess the effect of such parameters on the accuracy of the index test by adding categorical covariates in the models.

#### c. Effect of age

The sensitivity of RADTs is known to be higher in younger children than in older ones (Cohen 2012; Edmonson 2005). This might be explained by higher GAS prevalence in school-age children with pharyngitis than in older children. Therefore, we will try to explore age as a potential source of heterogeneity by using the mean age of patients in the study as a categorised covariate in the model (e.g., 0 to 4; 5 to 9; 10 to 14; 15 to 20 years).

#### d. Effect of disease severity

Spectrum effect has been demonstrated for RADTs, with increasing sensitivity with increasing disease severity, usually assessed by the McIsaac score (Cohen 2012; Edmonson 2005; Hall 2004; Tanz 2009). Therefore, disease severity might be a relevant source of heterogeneity to explore, for example by using the % of patients with a McIsaac score > two as a numerical covariate in the model.

#### e. Effect of GAS prevalence

Diagnostic accuracy may vary with disease prevalence (Leefflang 2009), usually with better performances in a population with higher disease prevalence. We will consider GAS prevalence as a dichotomised covariate to define low-risk versus high-risk study populations (i.e., below or above median of GAS prevalence across studies).

### Sensitivity analyses

We will carry out the following sensitivity analyses to explore the robustness of the results:

- include studies for which patient selection was avoided;
- include studies for which patients were excluded on the basis of antibiotics use within seven days before inclusion;
- include studies for which GAS antibody response was used as the reference test;
- include only studies of high quality according to QUADAS-2.

### Assessment of reporting bias

We will not try to assess reporting bias (Macaskill 2010).

## ACKNOWLEDGEMENTS

We thank Philippe Ravaud and Ludovic Trinquart (French Cochrane Centre, Université Paris Descartes, Paris, France) and the members of the Cochrane ARI Group and the Cochrane DTA Group for their comments and support. We also thank the following people for commenting on the draft protocol: Noorin Bhimani, Samileh Noorbakhsh, Saleh Altamimi, Conor Teljeur and Jenny Doust.

## REFERENCES

### Additional references

#### AAP 2009

American Academy of Pediatrics. Group A streptococcal infections. In: Pickering L, Baker C, Long S, McMillan J editor(s). *Red Book: 2009 Report of the Committee on Infectious Disease*. 28th Edition. Elk Grove Village, IL: American Academy of Pediatrics, 2009:616–28.

#### Battle 1971

Battle CU, Glasgow LA. Reliability of bacteriologic identification of beta-hemolytic streptococci in private offices. *American Journal of Diseases of Children* 1971;**122**(2):134–6.

#### Bisno 2002

Bisno AL, Gerber MA, Gwaltney JM, Jr, Kaplan EL, Schwartz RH. Practice guidelines for the diagnosis and management of group A streptococcal pharyngitis. Infectious Diseases Society of America. *Clinical Infectious Diseases* 2002;**35**(2):113–25.

#### Carapetis 2005a

Carapetis JR, Steer AC, Mulholland EK, Weber M. The global burden of group A streptococcal diseases. *Lancet* 2005;**5**(11):685–94.

#### Carapetis 2005b

Carapetis JR, McDonald M, Wilson NJ. Acute rheumatic fever. *Lancet* 2005;**366**(9480):155–68.

#### Carapetis 2007

Carapetis JR. Rheumatic heart disease in developing countries. *New England Journal of Medicine* 2007;**357**(5):439–41.

#### Centor 1981

Centor RM, Witherspoon JM, Dalton HP, Brody CE, Link K. The diagnosis of strep throat in adults in the emergency room. *Medical Decision Making* 1981;**1**(3):239–46.

#### Chapin 2002

Chapin KC, Blake P, Wilson CD. Performance characteristics and utilization of rapid antigen test, DNA probe, and culture for detection of group A streptococci in an acute care clinic. *Journal of Clinical Microbiology* 2002;**40**(11):4207–10.

#### Cohen 2004

Cohen R, Levy C, Ovetchkine P, Boucherat M, Weil-Olivier C, Gaudelus J, et al. Evaluation of streptococcal clinical scores, rapid antigen detection tests and cultures for

childhood pharyngitis. *European Journal of Pediatrics* 2004;**163**(4-5):281–2.

#### Cohen 2012

Cohen JF, Chalumeau M, Levy C, Bidet P, Thollot F, Wollner A, et al. Spectrum and inoculum size effect of a rapid antigen detection test for group A streptococcus pharyngitis in children. *PLoS One* 2012;**7**(6):e39085.

#### Edmonson 2005

Edmonson MB, Farwell KR. Relationship between the clinical likelihood of group A streptococcal pharyngitis and the sensitivity of a rapid antigen-detection test in a pediatric practice. *Pediatrics* 2005;**115**(2):280–5.

#### Fischer Walker 2006

Fischer Walker CL, Rimoin AW, Hamza HS, Steinhoff MC. Comparison of clinical prediction rules for management of pharyngitis in settings with limited resources. *Journal of Pediatrics* 2006;**149**(1):64–71.

#### Gerber 1986

Gerber MA, Randolph ME, Chanatry J, Wright LL, DeMeo KK, Anderson LR. Antigen detection test for streptococcal pharyngitis: evaluation of sensitivity with respect to true infections. *Journal of Pediatrics* 1986;**108**(5 Pt 1):654–8.

#### Gerber 2004

Gerber MA, Shulman ST. Rapid diagnosis of pharyngitis caused by group A streptococci. *Clinical Microbiology Reviews* 2004;**17**(3):571–80.

#### Gerber 2005

Gerber MA. Diagnosis and treatment of pharyngitis in children. *Pediatric Clinics of North America* 2005;**52**(3):729–47, vi.

#### Gerber 2009

Gerber MA, Baltimore RS, Eaton CB, Gewitz M, Rowley AH, Shulman ST, et al. Prevention of rheumatic fever and diagnosis and treatment of acute streptococcal pharyngitis: a scientific statement from the American Heart Association Rheumatic Fever, Endocarditis, and Kawasaki Disease Committee of the Council on Cardiovascular Disease in the Young, the Interdisciplinary Council on Functional Genomics and Translational Biology, and the Interdisciplinary Council on Quality of Care and Outcomes Research: endorsed by the American Academy of Pediatrics. *Circulation* 2009;**119**(11):1541–51.

**Hall 2004**

Hall MC, Kieke B, Gonzales R, Belongia EA. Spectrum bias of a rapid antigen detection test for group A beta-hemolytic streptococcal pharyngitis in a pediatric population. *Pediatrics* 2004;**114**(1):182–6.

**Harbeck 1993**

Harbeck RJ, Teague J, Crossen GR, Maul DM, Childers PL. Novel, rapid optical immunoassay technique for detection of group A streptococci from pharyngeal specimens: comparison with standard culture methods. *Journal of Clinical Microbiology* 1993;**31**(4):839–44.

**Heelan 1996**

Heelan JS, Wilbur S, Depetris G, Letourneau C. Rapid antigen testing for group A Streptococcus by DNA probe. *Diagnostic Microbiology and Infectious Disease* 1996;**24**(2):65–9.

**Joachim 2010**

Joachim L, Campos D Jr, Smeesters PR. Pragmatic scoring system for pharyngitis in low-resource settings. *Pediatrics* 2010;**126**(3):e608–14.

**Johnson 2010**

Johnson DR, Kurlan R, Leckman J, Kaplan EL. The human immune response to streptococcal extracellular antigens: clinical, diagnostic, and potential pathogenetic implications. *Clinical Infectious Diseases* 2010;**50**(4):481–90.

**Kellogg 1990**

Kellogg JA. Suitability of throat culture procedures for detection of group A streptococci and as reference standards for evaluation of streptococcal antigen detection kits. *Journal of Clinical Microbiology* 1990;**28**(2):165–9.

**Leeflang 2006**

Leeflang MM, Scholten RJ, Rutjes AW, Reitsma JB, Bossuyt PM. Use of methodological search filters to identify diagnostic accuracy studies can lead to the omission of relevant studies. *Journal of Clinical Epidemiology* 2006;**59**(3):234–40.

**Leeflang 2009**

Leeflang MM, Bossuyt PM, Irwig L. Diagnostic test accuracy may vary with prevalence: implications for evidence-based diagnosis. *Journal of Clinical Epidemiology* 2009;**62**(1):5–12.

**Macaskill 2010**

Macaskill P, Gatsonis C, Deeks JJ, Harbord RM, Takwoingi Y. Chapter 10: Analysing and Presenting Results. In: Deeks JJ, Bossuyt PM, Gatsonis C editor(s). *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy*. Chichester, UK: The Cochrane Collaboration, 2010.

**Matthys 2007**

Matthys J, De Meyere M, van Driel ML, De Sutter A. Differences among international pharyngitis guidelines: not just academic. *Annals of Family Medicine* 2007;**5**(5):436–43.

**McIsaac 1998**

McIsaac WJ, White D, Tannenbaum D, Low DE. A clinical score to reduce unnecessary antibiotic use in patients with

sore throat. *Canadian Medical Association Journal* 1998;**158**(1):75–83.

**Mondzac 1967**

Mondzac AM. Throat culture processing in the office - a warning. *JAMA* 1967;**200**(12):1132–3.

**Pelucchi 2012**

Pelucchi C, Grigoryan L, Galeone C, Esposito S, Huovinen P, Little P, et al. ESCMID Guideline for the management of acute sore throat. *Clinical Microbiology and Infection* 2012;**18**(Suppl 1):1–28.

**Pokorski 1994**

Pokorski SJ, Vetter EA, Wollan PC, Cockerill FR, 3rd. Comparison of Gen-Probe Group A streptococcus Direct Test with culture for diagnosing streptococcal pharyngitis. *Journal of Clinical Microbiology* 1994;**32**(6):1440–3.

**Reitsma 2005**

Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *Journal of Clinical Epidemiology* 2005;**58**(10):982–90.

**RevMan 2012**

The Nordic Cochrane Centre, The Cochrane Collaboration. Review Manager (RevMan). 5.2. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration, 2012.

**Rosenstein 1970**

Rosenstein BJ, Markowitz M, Gordis L. Accuracy of throat cultures processed in physicians' offices. *Journal of Pediatrics* 1970;**76**(4):606–9.

**Seckeler 2011**

Seckeler MD, Hoke TR. The worldwide epidemiology of acute rheumatic fever and rheumatic heart disease. *Clinical Epidemiology* 2011;**3**:67–84.

**Shaikh 2010**

Shaikh N, Leonard E, Martin JM. Prevalence of streptococcal pharyngitis and streptococcal carriage in children: a meta-analysis. *Pediatrics* 2010;**126**(3):e557–64.

**Shaikh 2011**

Shaikh N, Swaminathan N, Hooper EG. Accuracy and precision of the signs and symptoms of streptococcal pharyngitis in children: a systematic review. *Journal of Pediatrics*. 2011/11/04 2011 Nov 1 [Epub ahead of print].

**Shulman 2000**

Shulman ST, Tanz RR, Gerber MA. Streptococcal pharyngitis. In: Stevens DL, Kaplan EL editor(s). *Streptococcal Infections: Clinical Aspects, Microbiology, and Molecular Pathogenesis*. New York: Oxford University Press, 2000:76–101.

**Spinks 2011**

Spinks A, Glasziou PP, Del Mar CB. Antibiotics for sore throat. *Cochrane Database of Systematic Reviews* 2011, Issue 9. [DOI: 10.1002/14651858.CD000023.pub3]

**Steinhoff 2005**

Steinhoff MC, Walker CF, Rimo AW, Hamza HS. A clinical decision rule for management of streptococcal

- pharyngitis in low-resource settings. *Acta Paediatrica* 2005;**94**(8):1038–42.
- Tanz 1997**  
Tanz RR, Gerber MA, Shulman ST. What is a throat culture?. *Advances in Experimental Medicine and Biology* 1997;**418**:29–33.
- Tanz 2007**  
Tanz RR, Shulman ST. Chronic pharyngeal carriage of group A streptococci. *Pediatric Infectious Disease Journal* 2007;**26**(2):175–6.
- Tanz 2009**  
Tanz RR, Gerber MA, Kabat W, Rippe J, Seshadri R, Shulman ST. Performance of a rapid antigen-detection test and throat culture in community pediatric offices: implications for management of pharyngitis. *Pediatrics* 2009;**123**(2):437–44.
- Uhl 2003**  
Uhl JR, Adamson SC, Vetter EA, Schleck CD, Harmsen WS, Iverson LK, et al. Comparison of LightCycler PCR, rapid antigen immunoassay, and culture for detection of group A streptococci from throat swabs. *Journal of Clinical Microbiology* 2003;**41**(1):242–9.
- Van Limbergen 2006**  
Van Limbergen J, Kalima P, Taheri S, Beattie TF. Streptococcus A in paediatric accident and emergency: are rapid streptococcal tests and clinical examination of any help?. *Emergency Medicine Journal* 2006;**23**(1):32–4.
- Wegner 1992**  
Wegner DL, Witte DL, Schrantz RD. Insensitivity of rapid antigen detection methods and single blood agar plate culture for diagnosing streptococcal pharyngitis. *JAMA* 1992;**267**(5):695–7.
- Wessels 2011**  
Wessels MR. Clinical practice. Streptococcal pharyngitis. *New England Journal of Medicine* 2011;**364**(7):648–55.
- Whiting 2011a**  
Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine* 2011;**155**(8):529–36.
- Whiting 2011b**  
Whiting P, Westwood M, Beynon R, Burke M, Sterne JA, Glanville J. Inclusion of methodological filters in searches for diagnostic test accuracy studies misses relevant studies. *Journal of Clinical Epidemiology* 2011;**64**(6):602–7.
- WHO 1995**  
World Health Organization. *The Management of Acute Respiratory Infections in Children: Practical Guidelines for Outpatient Care*. Geneva: World Health Organization, 1995.
- \* Indicates the major publication for the study

## ADDITIONAL TABLES

Table 1. Data to be extracted from each study

<b>Study ID</b>	First author, year of publication
<b>Type of study</b>	Journal article or unpublished study
<b>Clinical features and settings</b>	Presenting signs and symptoms
	Clinical selection of patients (none, clinical score, explicit criteria but not a score, implicit criteria)
	Exclusion if antibiotics use within 7 days before inclusion (yes/no)
	Clinical setting (office-based, emergency dept., walk-in clinic, mixed, other)
	Single- or multi-centre study
	Age range for inclusion
<b>Participants</b>	Sample size (n)
	Age (distribution)

**Table 1. Data to be extracted from each study** (Continued)

	GAS prevalence according to culture (with 95% confidence interval)
	Country of study
	Sex (% of girls)
	Clinical severity assessment (Centor score, McIsaac score, other, none)
<b>Study design</b>	Cross-sectional study or RCT
	Retrospective or prospective design
	Sample (consecutive, random or unclear)
	Direct comparison of different RADTs (yes/no)
	Direct comparison of several throat culture techniques (yes/no)
	Throat swab (1 single, 1 double, 2 different)
	Person performing the throat sample (physician, nurse, laboratory personnel, other)
<b>Reference standard(s)</b>	Throat culture medium (standard, enrichment, inhibitory)
	Atmosphere of incubation (aerobic, aerobic with CO <sub>2</sub> enrichment, anaerobic)
	Duration of incubation ( $\leq 24$ , 24 to 48, $\geq 48$ hours)
	GAS confirmation (bacitracin disk, latex test, other, none)
	Number of plates inoculated (n)
	Assessment of GAS antibody response (yes/no)
	Relevant details
<b>Index tests</b>	Commercial name of the RADT
	Type of RADT (EIA, OIA)
<b>Data</b>	Number of true positives, false positives, true negatives, false negatives and undetermined/uninterpretable results
<b>Notes</b>	Source of funding (whether any of the authors is affiliated with the manufacturer of the RADT, the study was directly funded by the manufacturer, authors reported conflicts of interests related to the manufacturer or other funding sources)

**Table 1. Data to be extracted from each study** (Continued)

	Anything else of relevance
--	----------------------------

**Table 2. Methodological quality assessment table for each study**

<b>Domain 1: Patient selection</b>	
Was a consecutive or random sample of patients enrolled?	Yes, No or Unclear
Was it a cross-sectional study or a RCT?	Yes, No or Unclear
Were selection criteria clearly described (at least presenting signs and symptoms and age limits for inclusion)?	Yes, No or Unclear
Were patients seen in an ambulatory care setting?	Yes, No or Unclear
Was clinical selection of patients avoided?	Yes, No or Unclear
<b>Could the selection of patients have introduced bias?</b>	<b>Risk: Low, High or Unclear</b>
<b>Is there concern that the included patients do not match the review question?</b>	<b>Concern: Low, High or Unclear</b>
<b>Domain 2: RADT (index test)</b>	
Were RADTs conducted during consultation time?	Yes, No or Unclear
Were the RADT results interpreted with blinding of the results of culture?	Yes, No or Unclear
Was the type of the RADT mentioned (EIA or OIA)?	Yes, No or Unclear
<b>Could the conduct or interpretation of the RADT have introduced bias?</b>	<b>Risk: Low, High or Unclear</b>
<b>Is there concern that the RADT, its conduct or interpretation differ from the review question?</b>	<b>Concern: Low, High or Unclear</b>
<b>Domain 3: Throat culture (reference standard)</b>	
Were culture results interpreted with blinding of the results of the RADT?	Yes, No or Unclear
Is the throat culture method likely to correctly identify GAS (laboratory culture on a blood agar plate during $\geq 48$ hr)?	Yes, No or Unclear
Were the culture medium, atmosphere, duration of incubation and GAS-confirmation technique described?	Yes, No or Unclear

Table 2. Methodological quality assessment table for each study (Continued)

Could the throat culture, its conduct, or its interpretation have introduced bias?	Risk: Low, High or Unclear
Is there concern that the target condition as defined by the reference standard does not match the review question?	Concern: Low, High or Unclear
<b>Domain 4: Flow and timing</b>	
Was the delay between the performance of the RADT and throat culture plating $\leq$ 48 hours?	Yes, No or Unclear
Did all patients receive a throat culture?	Yes, No or Unclear
Did patients receive the same throat culture method?	Yes, No or Unclear
Were undetermined/uninterpretable results reported?	Yes, No or Unclear
Were withdrawals from the study explained?	Yes, No or Unclear
Could the patient flow have introduced bias?	Risk: Low, High or Unclear

## APPENDICES

### Appendix I. MEDLINE (Ovid) search strategy

- 1 Pharyngitis/
- 2 pharyngitis.tw
- 3 Tonsillitis/
- 4 tonsillitis.tw
- 5 (tonsillopharyngitis or pharyngotonsillitis).tw
- 6 sore throat\*.tw
- 7 ((throat\* or pharyn\* or tonsil\*) adj5 (infect\* or inflam\*)).tw
- 8 Pharynx/mi [Microbiology]
- 9 Streptococcal Infections/
- 10 (strep\* adj5 (throat\* or pharyn\* or tonsil\*)).tw
- 11 ("group a" adj5 streptococc\*).tw
- 12 gabhs.tw
- 13 (beta-hemoly\* or beta-haemoly\*).tw
- 14 lancefield group a.tw
- 15 Streptococcus pyogenes/
- 16 (streptococcus pyogenes or "s. pyogenes" or "s.pyogenes").tw
- 17 or/1-16
- 18 Immunoassay/

19 exp Immunoenzyme Techniques/  
20 (enzyme adj2 (immunoassay\* or immuno-assay\* or immunosorbent)).tw.  
21 Immunochromatography/  
22 immunochromatograph\*.tw.  
23 Immunosorbent Techniques/  
24 exp Enzyme-Linked Immunosorbent Assay/  
25 (elisa or elisas or eia or eias).tw.  
26 (sandwich\* adj2 assay\*).tw.  
27 (lateral flow adj2 assay).tw.  
28 (optical adj2 (immunoassay\* or immuno-assay\*)).tw.  
29 (oia or oias).tw.  
30 Antigens, Bacterial/  
31 Reagent Kits, Diagnostic/  
32 Point-of-Care Systems/  
33 ((rapid or "point of care" or "near patient" or poc or poct or bedside) adj5 (test or tests or testing or detect\* or diagnos\* or screen\* or kit or kits or assay\*)).tw.  
34 (radt or radts or rdt or rdts).tw.  
35 (antigen\* adj3 detect\*).tw.  
36 or/18-35  
37 17 and 36  
38 exp animals/ not humans/  
39 37 not 38

## **CONTRIBUTIONS OF AUTHORS**

MC and JFC had the original idea for the review and wrote first draft of the protocol.

RC edited the protocol.

## **DECLARATIONS OF INTEREST**

All authors declared no financial conflicts of interest.

All authors have been involved in studies that might be included in the review.

## **SOURCES OF SUPPORT**

### **Internal sources**

- No sources of support supplied

## External sources

- Laboratoires Guigoz - Société Française de Pédiatrie - Groupe de Pédiatrie Générale - Groupe de Recherches Epidémiologiques en Pédiatrie, France.

Educational Grant to JFC (2010)

- Agence Régionale de Santé d'Ile-de-France, France.

Educational Grant to JFC (2011)

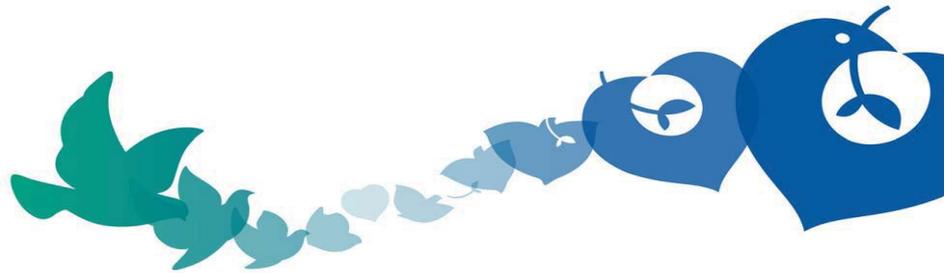


---

ANNEXE 6. CONSTRUCTION D'UN NOUVEL OUTIL PREDICTIF

(PROTOCOLE)

(projet financé par le PHRC régional n° AOR 12089)



*Unité de Recherche Clinique Cochin-Necker*

**PHRC2012 - APPEL A PROJETS REGIONAL**

**Construction et validation interne d'un nouvel outil informatisé de calcul séquentiel de risque individuel pour le diagnostic des angines à Streptocoque  $\beta$ -hémolytique du groupe A chez l'enfant**

***Investigateur coordonnateur***

**Pr Martin Chalumeau**

Service de Pédiatrie Générale  
Hôpital Necker  
Paris

16 Novembre2011

## PHRC REGIONAL 2012

## I. Résumé et liste des investigateurs

**Investigateur coordonnateur :**

Prénom et Nom : Pr Martin Chalumeau  
Fonction (PU-PH, PH, CCA, ...) : PU-PH

**Responsable Scientifique (si nécessaire) :**

Prénom et Nom :  
Fonction (PU-PH, PH, CCA, ...) :

**Titre du projet :**

**Construction et validation interne d'un nouvel outil informatisé de calcul séquentiel de risque individuel pour le diagnostic des angines à Streptocoque  $\beta$ -hémolytique du groupe A chez l'enfant**

**Hôpital :** Hôpital Necker Enfants Malades**Service :** Pédiatrie Générale**Chef de Service :** Pr Dominique Gendrel**Téléphone :** 01 44 49 48 82 **télécopie :** 01 44 38 17 50**Promoteur souhaité :**  APHP autre institutionnel, préciser lequel : .....**Pôle :** Pédiatrie générale et spécialisée**Chef de Pôle :** Pr Jacques de Blic**e-mail :** martin.chalumeau@nck.aphp.fr**URC impliquée :** URC/CIC Cochin Necker

- pour la conception méthodologique du projet :  non  oui
- pour l'élaboration du budget :  non  oui
- pour la mise en œuvre et le suivi, préciser l'URC si elle différente de celle impliquée pour la conception :

**Liste des centres participants (hôpital, service) et nom de l'investigateur principal pour chaque centre :**

Nom de l'investigateur	Service	Hôpital	Ville
Dr R. Cohen	Bactériologie	CHIC	Créteil
Pr E. Bingen	Bactériologie	Robert Debré	Paris
Pr M. Chalumeau	Pédiatrie Générale	Necker	Paris
Pédiatres du réseau ACTIV (N=20)	-	-	cf liste en annexe 3

**Axe thématique du PHRC Régional (un seul axe à indiquer en clair) :** Pédiatrie et adolescence Axe 3**Si votre projet utilise des médicaments, indiquez s'ils correspondent aux critères ci-après :**

- besoin d'une forme pharmaceutique ou d'un dosage spécifique (ex : pédiatrie) :  oui  non
- médicament sans AMM :  oui  non
- importation d'un pays tiers :  oui  non
- placebo fourni par un industriel :  oui  non

**Le projet a déjà été présenté à un appel à projets géré par l'AP-HP :**  oui  non

si oui, ① - préciser :  PHRC régional  PHRC national  CRC  autre, préciser :  
année : 2011 code du projet : CRC11052

② - joindre les réponses argumentées aux expertises

**Demandeur**

nom

Martin Chalumeau

signature


**Responsable de l'URC**

Jean Marc Tréluyer

date 16/11/2011

**RESUME DU PROJET (495 mots)****Introduction**

Le streptocoque du groupe A (SGA) est retrouvé dans 35% des angines de l'enfant, le reste des cas étant d'origine virale. Seules les angines à SGA justifient un traitement antibiotique dans le but de diminuer la durée des symptômes, l'incidence des complications suppuratives et non suppuratives ainsi que la dissémination du SGA. Les pratiques observées en France et aux USA concernant le traitement des angines de l'enfant sont sub-optimales puisque les taux de prescription d'antibiotiques sont de 70% au lieu des 35% attendus. Ces antibiothérapies inutiles ont d'importantes conséquences individuelles (effets secondaires des antibiotiques) et collectives (résistances bactériennes, coût). Afin de rationaliser la prescription d'antibiotiques, plusieurs recommandations basées sur l'utilisation de scores cliniques et/ou d'examen de prélèvements bactériologiques de gorge ont été émises. Les recommandations sont très hétérogènes car elles répondent à des rationnels de soins différents selon le pays. Par ailleurs, les recommandations basées sur des scores présentent des limites car leur méthodologie de construction et de validation est faible au regard des standards actuels, ce qui peut affecter leurs performances diagnostiques et la confiance des utilisateurs. Enfin, aucune de ces stratégies ne permet un calcul de risque individuel.

**Hypothèses**

Les discordances entre recommandations et pratiques dans le diagnostic étiologique des angines de l'enfant pourraient être liées en partie aux limites des recommandations actuelles : hétérogénéité, défauts méthodologiques et à absence de possibilité de calcul de risque individuel.

**Objectif principal**

Construction d'un nouvel outil informatisé de calcul séquentiel de risque individuel adaptable aux exigences locales de diagnostic en fonction des balances sensibilité/spécificité et bénéfiques/risques ciblées, en suivant les recommandations méthodologiques de l'Evidence Based Medicine Working Group et les recommandations sur la construction de modèles multivariés.

**Objectif(s) secondaire(s)**

- 1/ Validation interne du nouvel outil prédictif ;
- 2/ Dépôt d'un brevet ;
- 3/ Développement d'une application informatique (plateforme web et *smartphone*).

**Critère d'évaluation principal**

Discrimination du modèle : aire sous la courbe ROC du modèle.

**Critère(s) d'évaluation secondaire(s)**

- Calibration du modèle : test d'adéquation d'Hosmer-Lemeshow, *calibration-plot*.
- Comparaison des performances du nouvel outil prédictif (aire sous la courbe ROC) à celles des différentes recommandations existantes.

**Méthodologie :**

Analyse secondaire (étude ancillaire) des données d'une étude prospective multicentrique d'évaluation des performances d'un dispositif médical de diagnostic in vitro.

**Nombre de sujets nécessaires**

Le calcul initial du nombre de sujets nécessaires pour l'étude princeps était N=714. Le nombre de sujets inclus est de 749 (recrutement terminé).

**Critères d'inclusion et principaux critères de non inclusion**

Enfants âgés de 3 à 15 ans ; vus en ambulatoire par leur pédiatre de ville ; ayant une angine ; n'ayant pas reçu d'antibiotique depuis 1 semaine.

**Durée totale de l'étude :** recrutement terminé, 3 mois d'analyses

**Période d'inclusion :** octobre 2010 - juin 2011 (recrutement terminé)

**Durée de participation pour un patient :** étude transversale, pas de suivi des patients

**Nombre de centres participants :** 20 pédiatres de ville faisant partie d'un réseau professionnel (ACTIV)

**Nombre moyen d'inclusions par mois par centre :** 10

**Montant total demandé :** 33000euros

## II. Projet Scientifique

### Justification scientifique

#### Justificatif

Le streptocoque  $\beta$ -hémolytique du groupe du groupe A (SGA) est retrouvé dans 35% des angines de l'enfant, le reste des cas étant d'origine virale [1]. On compte 450 millions de cas d'angines à SGA chez l'enfant par an dans le monde [2], ce qui en fait un véritable problème de santé publique. Seules les angines à SGA justifient un traitement antibiotique afin de diminuer la durée des symptômes, l'incidence des complications suppuratives et non suppuratives ainsi que la dissémination du SGA à l'entourage [3]. Les pratiques observées en France [4] et aux Etats-Unis [5] concernant le traitement des angines de l'enfant sont sub-optimales puisque les taux de prescription d'antibiotiques sont de 70% au lieu des 35% attendus. Ces antibiothérapies inutiles ont d'importantes conséquences individuelles (effets secondaires des antibiotiques) et collectives (résistances bactériennes, coût). Afin de rationaliser la prescription d'antibiotiques, plusieurs recommandations basées sur l'utilisation de scores cliniques [6, 7] et/ou de prélèvements bactériologiques de gorge (Test Diagnostique Rapide - TDR, culture au laboratoire durant 48h - gold standard, culture durant 24h au cabinet) ont été émises [8-12]. Ces scores et recommandations répondent à des rationnels de soins différents selon le pays (en particulier sur la sensibilité et la spécificité ciblées) et il n'y a donc pas de consensus international [13].

Recommandation	France (Afssaps)	USA (AAP/IDSA)*	USA (AAFP)*	Royaume-Uni (NICE)*
Score clinique	non	non	Maclsaac**	Centor**
Test rapide (TDR)	pour tous	selon clinique	selon score	non
Culture de gorge	non	si TDR -	si TDR -	non
Antibiotique	si TDR +	si test +	si test +	si score $\geq 3$

Tableau 1. Angines de l'enfant : résumé des principales recommandations internationales

\* AAP : American Academy of Pediatrics; IDSA : Infectious Disease Society of America; AAFP : American Academy of Family Physicians; NICE : National Institute for health and Clinical Excellence.

\*\* voir scores en Annexe 1

Par ailleurs, ces scores et recommandations présentent des limites méthodologiques au regard des standards de l'Evidence Based Medicine Working Group publiés dans le JAMA en 1997 sur les règles de décision clinique (Annexe 2) [14-16] et des recommandations concernant la construction de modèles prédictifs multivariés [17-21].

- Limites des recommandations françaises (Afssaps) [8] :
  - basées uniquement sur des avis d'experts (faible niveau de preuve) ;
  - approche non-sélective de pratique du TDR (pour tous les enfants à partir de 3 ans), ce qui peut limiter son application ;
  - sensibilité  $\approx 85\%$ , jugée insuffisante par certains auteurs [9, 22] ;

- Limites des recommandations de l'American Academy of Pediatrics et de l'Infectious Disease Society of America (TDR sur arguments cliniques, culture si TDR négatif) [9, 10] :
  - basées uniquement sur des avis d'experts (faible niveau de preuve) ;
  - validation impossible car les arguments devant faire suspecter un SGA et réaliser un test diagnostique ne sont pas explicités dans les recommandations.
- Les limites des recommandations de l'American Academy of Family Physicians [11] correspondent aux limites du score de McIsaac par rapport aux recommandations sur la construction de modèles multivariés [17-21] :
  - score construit sur une population adulte et pédiatrique alors que la problématique y est différente (fréquence du SGA, signes cliniques, portage asymptomatique de SGA chez l'enfant) ;
  - score construit par régression logistique sans recherche d'interactions ;
  - pas d'ajustement sur l'âge dans le modèle multivarié ;
  - pas de mesure d'adéquation (calibration) du modèle logistique ;
  - transformation des variables continues en variables discrètes (exemple : température) ;
  - validation externe défavorable du score : l'utilisation du score conduit chez l'enfant à une augmentation du taux d'antibiothérapies inutiles (+18%) [23] ;
- Les limites des recommandations britanniques (National Institute for Health and Clinical Excellence) correspondent aux limites du score de Centor [6, 12]:
  - score identique au score de McIsaac sauf qu'il ne prend pas en compte l'âge ;
  - pas de validation interne ni externe chez l'enfant.

Le principe des scores intégrés à ces recommandations est d'établir des groupes de risque puis de proposer une décision clinique en fonction du groupe de risque dans lequel se situe le patient. Aucune stratégie de calcul de risque individuel n'a été développée dans le diagnostic des angines alors que cela est envisageable depuis quelques années grâce à la généralisation de l'accès à l'informatique (comme cela est fait en oncologie par exemple) [24, 25].

Enfin, nous avons montré dans un précédent travail que les performances diagnostiques du TDR pouvaient varier en fonction du spectre clinique des patients (article soumis, voir abstract en annexe 9) et il semble intéressant de proposer un outil permettant au clinicien de tenir compte de ces variations lorsqu'il interprète le TDR [26].

### **Hypothèse principale**

Les discordances entre recommandations et pratiques pourraient être liées en partie à :

- l'absence de consensus international ;
- des défauts méthodologiques dans l'élaboration des recommandations ;
- l'absence actuelle d'outils de calcul de risque individuel.

L'utilisation d'un nouvel outil informatisé de calcul séquentiel de risque individuel permettrait au clinicien de discuter avec les parents du patient du risque individuel d'étiologie bactérienne de l'angine, de la nécessité d'un test diagnostique, et du rapport bénéfices/risques d'un traitement antibiotique.

## **Objectifs de l'étude**

1. Construction d'un nouvel outil prédictif séquentiel de l'étiologie des angines de l'enfant permettant un calcul de risque individuel, en suivant les recommandations méthodologiques de l'Evidence Based Medicine Working Group [14-16] et les recommandations sur la construction de modèles multivariés [17-21] ;
2. Validation interne de l'outil prédictif ;
3. Développement et validation d'une application informatique destinée aux praticiens.

<h3><b>Principes de l'outil de calcul séquentiel de risque individuel</b></h3>
--

Le nouvel outil prédictif sera séquentiel :

- étape 1 : en consultation, si l'enfant présente une angine, le clinicien pourra calculer un risque individuel de SGA en fonction des variables relevées à l'examen clinique ;
- (+/-) étape 2 : si le clinicien réalise un TDR, il pourra calculer un risque individuel de SGA post-test en fonction des variables utilisées à l'étape 1 et du résultat du TDR.

A chaque étape, le calcul de risque individuel sera réalisé à l'aide d'une application informatique (cf Développement et validation de l'application informatique).

A chaque étape, la probabilité individuelle prédite constituera une aide à la décision :

- étape 1 : en fonction de la probabilité individuelle prédite par le modèle 1, le médecin pourra, en discussion avec les parents, soit décider d'emblée de traiter ou de ne pas traiter par antibiotique, soit décider de compléter son examen clinique par un TDR ;
- étape 2 : si le médecin décide de réaliser un TDR, il pourra entrer le résultat de celui-ci et en fonction de la probabilité individuelle prédite par le modèle 2, le médecin pourra, en discussion avec les parents, soit décider de traiter ou de ne pas traiter par antibiotique, soit décider de compléter l'investigation étiologique par une mise en culture bactérienne standard de prélèvement de gorge (cf Figure 1 page suivante).

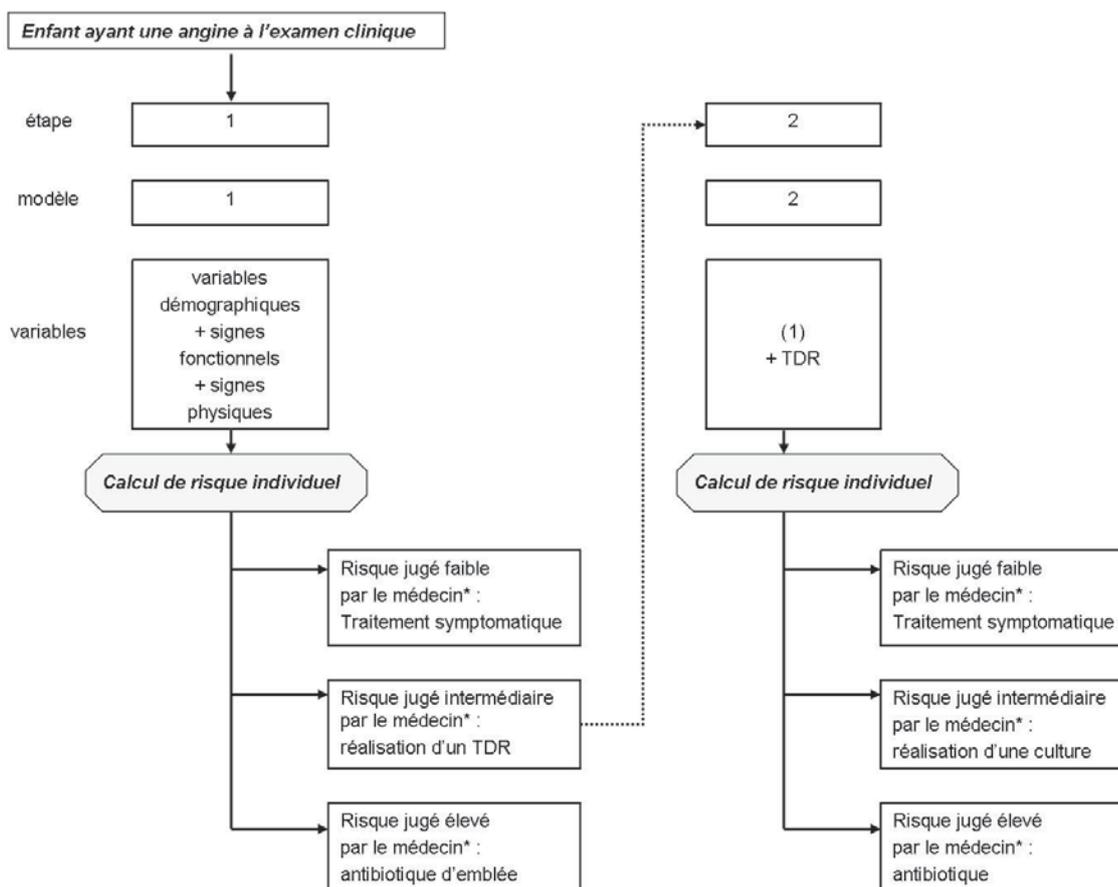


Figure 1. Principe de l'outil de calcul séquentiel de risque individuel

\* les seuils de risque orientant la décision peuvent être adaptés de manière contextuelle (souhaits des parents, exigences diagnostiques et balance bénéfiques/risques locales).

## Conception et déroulement de la recherche

### Description générale

- Etude épidémiologique ancillaire non interventionnelle ;
- Analyse secondaire de données existantes (étude prospective multicentrique d'enfants atteints d'angine).

### Critères d'inclusion et de non-inclusion

- Enfants vus en ambulatoire ;
- N'ayant pas reçu d'antibiotiques depuis 1 semaine ;
- N'ayant pas été déjà inclus dans l'étude depuis 1 mois ;
- Agés de 3 à 15 ans ;
- Ayant une angine définie cliniquement par : amygdales et/ou pharynx d'aspect érythémateux ou érythémato-pultacé (correspondant au terme anglo-saxon pharyngitis).

## Centres participants, investigateurs

### Centres participants

- Centre Hospitalier Intercommunal de Créteil

Service de Microbiologie - Dr Robert Cohen

40, avenue de Verdun

94010 Créteil Cedex

Tél : 01 45 17 53 54 Fax : 01 45 17 53 49

- ACTIV

Dr Corinne Lévy, Dr Robert Cohen

27 rue d'Inkerman

94100 St Maur des Fossés

Tél : 01 48 85 04 04 Fax : 01 48 85 18 58

- Hôpital Robert Debré

Service de Microbiologie - Pr Edouard Bingen

48 Bd Serrurier

75019 Paris

Tél : 01 40 03 23 40 Fax : 01 40 03 24 50

- Hôpital Necker Enfants Malades

Service de Pédiatrie Générale - Pr Martin Chalumeau

149 rue de Sèvres

75015 Paris

Tél: 01 44 49 48 82 Fax : 01 44 38 17 50

### Investigateurs

Le présent projet de recherche est une analyse secondaire des données d'une étude portant sur les performances du Test Diagnostique Rapide (TDR) des angines menée par 20 pédiatres du réseau ACTIV (Association Clinique et Thérapeutique du Val-de-Marne). Ce réseau de pédiatrie ambulatoire dédié à la recherche et à l'enseignement est coordonné par le Dr Robert Cohen, Praticien Hospitalier du service de Microbiologie du Centre Hospitalier Intercommunal de Créteil. La liste des pédiatres investigateurs participant à cette étude est détaillée en Annexe 3.

## Données recueillies

### **Données cliniques**

Pour tous les enfants, 25 variables sont recueillies sur un formulaire standardisé (CRF) :

- variables démographiques : âge et sexe de l'enfant ;
- signes fonctionnels : contagé familial de pathologie à SGA, début brutal, odynophagie, toux, rhinorrhée, enrouement, conjonctivite, céphalées, douleurs abdominales, nausées, vomissements, diarrhée ;
- signes physiques : température maximale observée depuis le début de l'épisode, œdème pharyngé, érythème pharyngé, pharynx érythémato-pultacé, purpura du voile du palais, présence d'adénopathies sous-angulo-maxillaires, taille des adénopathies sous-angulo-maxillaires si présentes, adénopathies sous-angulo-maxillaires sensibles si présentes, éruption cutanée, éruption scarlatiniforme si éruption présente (cf CRF en Annexe 4).

Ces variables ont été choisies car elles sont classiquement discutées dans la littérature [27] et/ou qu'elles rentrent dans les scores cliniques existants [6, 7]. Les variables « âge » et « température maximale rapportée au cours de l'épisode infectieux » sont des variables continues, la variable « taille des adénopathies sous-angulo-maxillaires » est une variable discrète en 3 classes et les autres variables cliniques sont dichotomiques. Ces variables seront les prédicteurs cliniques potentiels dont la performance sera testée dans la construction des modèles de prédiction.

### **Données bactériologiques**

Tous les enfants inclus ont un prélèvement de gorge par double écouvillon (MAG-ETP2, CML, France) :



Figure 2. Double écouvillon MAG-ETP2 (CML, France)

- Le premier écouvillon sert à réaliser le TDR au cabinet du pédiatre (StreptAtest, Dectrapharm, France). Le StreptAtest est une bandelette réactive immunochromatographique, spécifique d'un antigène de paroi spécifique du SGA : le carbohydrate C. Il s'agit d'un des TDR recommandé par l'Agence Française de Sécurité Sanitaire des Produits de Santé (AFSSAPS) depuis 2002 [28] car il présentait le meilleur couple performances/praticabilité et il est fourni gratuitement aux médecins généralistes, pédiatres et oto-rhino-laryngologistes par la Caisse Nationale d'Assurance Maladie.
- Le deuxième écouvillon sert à réaliser la mise en culture (gold standard). Le prélèvement est acheminé par coursier au laboratoire de microbiologie de l'Hôpital Robert Debré, Centre National de Référence (CNR) du SGA chez l'enfant [29], dans un milieu de transport de type AMIES gélosé, accompagné d'un bon de bactériologie anonymisé. Les prélèvements sont mis en culture sur un milieu gélosé trypticase-caséine-soja additionné de 5% de sang de cheval ou de mouton et les boîtes

de Petri sont incubées à 37°C en anaérobiose pendant 48 heures (technique de référence). Le groupage des colonies  $\beta$ -hémolytiques est effectué par un test d'agglutination sur particules de latex sensibilisées (Prolex, Pro-Lab Diagnostics, Canada). Les souches de SGA isolées sont congelées à -80°C dans un bouillon cœur-cervelle additionné de 15% de glycérol et conservées au CNR. Toutes les analyses microbiologiques effectuées au CNR sont faites à l'aveugle des données cliniques individuelles et du résultat du TDR.

Les résultats du TDR et de la culture sont rendus de manière binaire (positif/négatif).

### **Critères de jugement**

#### **Critère de jugement principal**

Notre critère de jugement principal sera le pouvoir discriminant du modèle, c'est-à-dire la mesure d'aire sous la courbe ROC (ASC ROC) des 2 modèles prédictifs séquentiels (modèles "clinique" et "clinique+TDR").

En effet, l'objectif principal est de pouvoir calculer un risque individuel de présence du SGA en maximisant le nombre de sujets bien classés, ce qui revient à maximiser l'ASC ROC des modèles de prédiction.

#### **Critères de jugement secondaires**

La calibration des modèles sera appréciée par le test d'Hosmer-Lemeshow et représentée graphiquement par un graphe de calibration.

Les performances diagnostiques des nouveaux modèles prédictifs seront comparées aux performances des recommandations internationales existantes dans les différents pays appliquées à notre population par comparaison d'ASC ROC.

### **Statistiques : construction et validation des modèles prédictifs**

#### **Nombre de sujets**

Le calcul initial du nombre de sujets nécessaires pour l'étude princeps était de  $N=714$ .

Le recrutement est terminé et le nombre de sujets nécessaires initialement prévu a été légèrement dépassé ( $N=749$ ).

En se basant sur une fréquence de SGA attendue de 35% [1], cela nous permet d'attendre  $(749 \times 0.35) = 262$  événements d'intérêt (angines à SGA). Ceci fournira assez de puissance pour construire des modèles multivariés contenant jusqu'à 25 prédicteurs (variables prédictives et termes d'interaction) [17].

#### **Variable prédite, variables prédictives**

Les modèles seront construits dans le but de prédire le risque individuel de SGA en fonction de variables cliniques et paracliniques (TDR) simples.

La variable prédite par le modèle est le résultat de la culture de prélèvement de gorge pour la recherche de SGA (gold standard) considéré de manière binaire (positif/négatif).

Les variables prédictives potentielles sont les 25 variables cliniques recueillies sur le CRF et le résultat du TDR (Annexe 4).

## **Stratégie d'analyse**

### **Plan d'analyse**

- description des caractéristiques générales de la population ;
- construction de 2 modèles prédictifs permettant un calcul séquentiel de risque individuel de SGA en 2 étapes :
  - modèle 1 : modèle destiné aux cliniciens face à un enfant présentant une angine clinique ;
  - modèle 2 : modèle destiné aux cliniciens face à un enfant présentant une angine et chez qui un TDR a été réalisé;
- validation interne des nouveaux modèles prédictifs ;
- comparaison des performances des modèles entre eux ;
- comparaison des performances des nouveaux modèles prédictifs à celles des recommandations nationales et internationales existantes.

### **Contraintes de construction des nouveaux modèles prédictifs**

Nous nous baserons sur les recommandations de l'Evidence Based Medicine Working Group sur les règles de décision clinique [14-16] et les recommandations sur l'élaboration et la validation des modèles multivariés de prédiction [17-21].

Les 2 modèles devront permettre de calculer un risque d'angine à SGA de manière continue, contrairement aux scores actuellement disponibles qui situent le patient dans un groupe à risque (risque faible, intermédiaire ou élevé).

Les modèles devront :

- utiliser des variables cliniques relevées à l'interrogatoire et à l'examen physique ;
- se baser sur un nombre restreint de variables ( $\leq 10$ ) ;
- être construits sur l'ensemble de la population d'enfants présentant une angine.

### **Construction des modèles prédictifs**

#### **Analyses univariées**

Les prédicteurs cliniques potentiels seront testés en analyse univariée par un test du  $\chi^2$  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.

#### **Analyses multivariées**

Nous utiliserons un modèle de régression logistique multivarié afin d'identifier les variables associées de manière indépendante à la présence du SGA après ajustement sur les co-variables d'intérêt.

Les modèles finaux seront de la forme :  $\text{score} = \alpha + \sum \beta_i X_i$  ;

La probabilité prédite par chaque modèle sera alors donnée par :

$$P(\text{SGA}) = 1/[1+e^{(-\text{score})}].$$

### Linéarité

En cas d'écart significatif à la linéarité, les variables continues seront traitées en continu dans le modèle multivarié en utilisant si nécessaire une transformation en polynômes fractionnaires d'ordre le plus bas ( $\leq 3$ ).

### Additivité, interactions

Les interactions ayant un substratum physiopathologique (par exemple entre fièvre et âge) seront testées par un test de Mantel-Haenszel sur les odds-ratio pour les variables catégorielles ou directement dans un modèle logistique contenant les 2 variables d'intérêt et leur terme d'interaction.

### 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 pourront être forcées dans le modèle en raison de leur pertinence clinique ou bibliographique et/ou de leur notoriété.

### Prise en compte des données manquantes

Si les données manquantes concernant les prédicteurs conduisent à la perte de plus de 5% des observations, une technique d'imputations multiples sera utilisée afin de conserver ces observations dans le modèle [30].

### Performances des modèles

Le pouvoir discriminant des modèles finaux sera apprécié par l'ASC ROC.

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 validation croisée de type *jackknife* (10-fold cross-validation). Il s'agit d'une technique de ré-échantillonnage dans laquelle l'échantillon est coupé en 10 sous-échantillons de taille égale. Un modèle est construit par régression logistique sur les 9/10 de la population puis la RDC est appliquée sur le 1/10 restant. Cette procédure est appliquée 10 fois en prenant tour à tour 1/10 différent comme population test [21, 31, 32].

Cette procédure permet :

- d'obtenir les performances du modèle de manière non-biaisée (sensibilité, spécificité et ASC ROC) ;

- d'utiliser tout l'échantillon pour construire le modèle en maximisant l'information utilisée (contrairement aux procédures de type split-sample réalisant une partition en un effectif de construction de 2/3 et de validation de 1/3 par exemple) ;
- de vérifier de manière fiable la stabilité du modèle logistique (stabilité des variables retenues dans le modèle, des coefficients de régression et de l'ASC ROC).

### **Comparaison des performances des nouveaux modèles prédictifs**

- comparaison des modèles entre eux par comparaison d'ASC ROC par un test de DeLong [33] ;
- comparaison de l'ASC ROC des nouveaux modèles à celle des recommandations existantes par un test de DeLong [33].

### **Déroulement des analyses**

Les analyses statistiques seront réalisées au sein de l'unité INSERM U953 sous la responsabilité du Pr Martin Chalumeau, pédiatre et épidémiologiste, qui encadrera M Jérémie Cohen, interne en DES de pédiatrie qui sera en disponibilité pour une thèse de Science d'Epidémiologie.

Les analyses statistiques seront faites sur le logiciel Stata/SE 11 (StataCorp LP, College Station, Tx, USA).

## **Développement et validation de l'application informatique**

### **Développement de l'application informatique**

#### **Principes**

Une fois les différents modèles prédictifs construits et validés, une phase de développement informatique sera réalisée en partenariat avec une entreprise de développement de logiciels informatiques dans le but d'élaborer une application informatique permettant de calculer facilement le risque individuel d'angine à SGA.

L'utilisateur sera invité à rentrer une à une les variables cliniques retenues dans le modèle final (exemple : sexe, âge...) puis l'application informatique utilisera les coefficients du modèle pour calculer une probabilité individuelle d'angine à SGA exprimée en pourcentage et son intervalle de confiance à 95%.

#### **Cahier des charges**

- développement d'une interface web ;
- développement d'une interface smartphones pour l'application de calcul de risque individuel (compatible iPhone©, Blackberry© et Android©) ;
- développement en langue française et anglaise (dans un premier temps).

### **Validation de l'application informatique**

L'application sera soumise à une évaluation auprès des pédiatres du réseau ACTIV et d'un groupe de médecins généralistes afin d'évaluer les paramètres suivants :

- ergonomie ;
- facilité d'utilisation ;
- compréhension des risques fournis par l'application ;
- satisfaction globale.

Cette phase de validation d'une durée de 6 semaines permettra de repérer d'éventuels points sur lesquels l'outil pourrait être amélioré et conduira à une mise à jour de l'application.

### **Taches respectives**

- le Dr Robert Cohen, pédiatre infectiologue, directeur scientifique d'ACTIV et praticien hospitalier dans le service de microbiologie du Centre Hospitalier Intercommunal de Créteil supervisera ce travail en tant qu'expert infectiologue ;
- les pédiatres du réseau ACTIV (cf liste en Annexe 3) seront responsables de l'inclusion des patients, du recueil des données cliniques sur les CRF et de leur transmission à ACTIV et de la réalisation et de l'envoi du prélèvement de gorge au CNR du SGA de l'Hôpital Robert Debré à Paris. Ils participeront également à la phase d'évaluation de l'outil informatique ;
- le Pr Edouard Bingen, chef du service de Bactériologie de l'Hôpital Robert Debré et responsable du CNR du SGA sera responsable des analyses microbiologiques entrant dans le cadre de cette étude ;
- le Pr Martin Chalumeau, pédiatre dans le service de Pédiatrie Générale, Hôpital Necker et épidémiologiste dans l'unité INSERM 953 (Paris) sera responsable de l'analyse des données et de leur intégrité ;
- Jérémie Cohen, interne en DES de Pédiatrie et mis en disponibilité de l'internat jusqu'en novembre 2013 sera recruté pour l'analyse statistique des données.

### **Retombées potentielles et perspectives**

#### **Retombées potentielles**

La construction et la validation d'un nouvel outil informatisé de calcul de risque individuel satisfaisant aux critères méthodologiques actuels permettrait de proposer aux médecins un outil universel d'aide à la décision dans le cadre du diagnostic et du traitement des angines à SGA de l'enfant. Cet outil aura l'avantage d'être basé sur les preuves et adaptable aux exigences locales de diagnostic en fonction des balances sensibilité/spécificité et bénéfices/risques ciblées ainsi qu'aux souhaits des parents.

#### **Perspectives**

En cas de succès de la construction et de validation interne de ce nouvel outil prédictif, nous envisageons :

- le dépôt d'un brevet pour les modèles et l'application informatisée associée ;
- une phase de validation externe afin de vérifier la stabilité du modèle et de ses performances (discrimination, calibration) ;

- une étude d'impact par la mise en place d'un essai quasi-expérimental de type "avant-après et ici-ailleurs" afin d'évaluer la praticabilité de l'outil et la réduction des antibiothérapies inutiles liées à l'utilisation de l'application.

## **Faisabilité**

### **Disponibilité des données**

Le recueil des données est terminé, les fichiers sont saisis et nettoyés. ACTIV est responsable du monitoring de la base de données.

### **Application informatique**

Un contact a été pris avec une société spécialisée dans le développement d'applications pour smartphones et plateformes web afin d'évaluer la faisabilité du projet (budget, temps de développement).

## **Aspects éthiques**

### **CPP**

L'étude originale était une étude de comparaison des performances diagnostiques du test rapide chez des enfants ayant une angine et chez des enfants bien portants. Le protocole concernant les enfants bien portants a reçu un avis favorable du CPP Ile-de-France XI (décision n°09016, annexe 7) et ce CPP a indiqué que la partie concernant les enfants ayant une angine ne nécessitait pas d'être examinée car il s'agissait d'une pratique courante. Toutefois, pour la soumission d'un article analysant le biais de spectre du TDR chez les 785 premiers enfants inclus avec angine (voir abstract en annexe 9), nous avons demandé une confirmation de cette indication donnée par le CPP Ile-de-France XI au comité d'éthique de l'hôpital Necker (rattachement hospitalier du demandeur du CIRC) qui a donné un avis favorable (voir annexe 8 du protocole).

### **Autorisation parentale**

La non-opposition du parent accompagnateur à l'utilisation des données de routine du patient a été recueillie (Annexe 6).

### **CNIL**

Pour l'étude princeps, une déclaration simple (chap.X) à la CNIL a été faite, en accord avec les articles 62 et suivants de la loi du 6 janvier 1978 modifiée (notification d'autorisation reçue le 20/08/2009, n°1354254).

## **Gestion des données**

### **Cahier d'Observation**

Les cahiers d'observation (ou Case Report Form - CRF) ont été complétés pour chaque patient (Annexe 4). Les CRF ont été faxés à ACTIV où les données ont été saisies. Aucune donnée nominative ne figure dans la base de données.

### **Traitement des données**

Les CRF ont été adressés à ACTIV pour validation et double saisie en temps réel.

Les données cliniques et microbiologiques ont été saisies sur une base de données (logiciel 4D). Des demandes de correction seront faites par fax en cas de donnée manquante, aberrante ou en cas d'incohérence de double saisie. Les réponses des investigateurs datées et signées ont alors été saisies sur la base, une fois refaxées à ACTIV.

### **Monitoring**

Une base de données spécifique à l'étude a été créée, testée et validée avant le début de la saisie. Seules les personnes habilitées ont le droit d'accès. Cette base est située sur un poste informatique au siège d'ACTIV, sous la responsabilité d'une personne chargée de la conception et de la mise en place de cette base, de la saisie des données initiales, de la réception des questionnaires médicaux, de leur saisie dans la base et de la transmission régulière des résultats aux différentes personnes concernées. La base de données est sauvegardée de façon quotidienne.

### **Archivage**

ACTIV conservera toute la documentation liée à l'étude pendant 10 ans.

Le médecin conservera les documents liés à l'étude jusqu'à la fin de la période d'analyse des données ou l'interruption prématurée de l'étude. Il est rappelé que pendant et à l'issue de l'étude, la tenue des dossiers médicaux est soumise à la législation en vigueur.

## **Conseil de l'ordre**

Les médecins ne seront pas rémunérés pour cette étude qui ne sera pas soumise au Conseil National de l'Ordre des Médecins.

## **Financement et assurance**

### **Financement**

L'étude princeps a été auto-financée par le groupe ACTIV (recueil des données, nettoyage de la base). Le budget de l'étude princeps ne prévoyait pas les analyses nécessaires à cette étude ancillaire et le budget nécessaire à cette étude ancillaire n'a pas été obtenu par ailleurs.

**Assurance**

Un contrat de responsabilité civile a été souscrit en date du 29/01/2009 auprès de la compagnie HDI-GERLING INDUSTRIES, 111 Rue de Longchamp 75116 Paris, agissant en qualité d'assureur. N° de contrat (1680)90712.

**Calendrier prévisionnel**

- novembre 2012 - janvier 2013 : construction et validation interne des 2 modèles de prédiction ;
- février - avril 2013 : développement des applications informatiques ;
- mai - juin 2013 : phase d'évaluation et de mise à jour de l'outil informatique ;
- juillet 2013 : rédaction des manuscrits.

**Propriété des résultats, règles relatives à la publication**

Les analyses ancillaires réalisées sur cette base de données au sein de l'unité Inserm 953 seront la propriété conjointe des différents partenaires (ACTIV, CNR SGA, Inserm U953). Aucune communication ou publication ne sera réalisée sans discussion et accord avec les différents partenaires (ACTIV, CNR SGA, Inserm U953). Le rôle des différents partenaires devra être mentionné dans la (les) publication(s). Seront signataires des publications, les personnes ayant réellement participé à l'élaboration du protocole et son déroulement, à l'analyse statistique ainsi qu'à la rédaction des manuscrits.

## **Bibliographie**

1. Shaikh N, Leonard E, Martin JM. Prevalence of streptococcal pharyngitis and streptococcal carriage in children: a meta-analysis. *Pediatrics*. 2010;126:e557-564.
2. Carapetis JR, Steer AC, Mulholland EK, Weber M. The global burden of group A streptococcal diseases. *Lancet Infect Dis*. 2005;5:685-694.
3. Del Mar CB, Glasziou PP, Spinks AB. Antibiotics for sore throat. *Cochrane Database Syst Rev*. 2006;CD000023.
4. Allemand H. Etat des lieux de la consommation d'antibiotiques en France. Résistances aux antibiotiques : Quels enjeux aujourd'hui pour la médecine de ville?; Table ronde de la CNAM, 19 mai 2010; Table ronde de la CNAM, 19 mai 2010.
5. McCaig LF, Besser RE, Hughes JM. Trends in antimicrobial prescribing rates for children and adolescents. *JAMA*. 2002;287:3096-3102.
6. Centor RM, Witherspoon JM, Dalton HP, Brody CE, Link K. The diagnosis of strep throat in adults in the emergency room. *Med Decis Making*. 1981;1:239-246.
7. McIsaac WJ, White D, Tannenbaum D, Low DE. A clinical score to reduce unnecessary antibiotic use in patients with sore throat. *CMAJ*. 1998;158:75-83.
8. AFSSAPS. Antibiothérapie par voie générale en pratique courante au cours des infections respiratoires hautes de l'adulte et de l'enfant. Octobre 2005 [cited Mai 2010]; Available from: <http://www.afssaps.fr>
9. American Academy of Pediatrics. Group A Streptococcal Infections. In: Pickering L, Baker C, Long S, McMillan J, eds. *Red Book: 2009 Report of the Committee on Infectious Disease*. 28th ed. Elk Grove Village, IL: American Academy of Pediatrics 2009:616-628.
10. Bisno AL, Gerber MA, Gwaltney JM, Jr., Kaplan EL, Schwartz RH. Practice guidelines for the diagnosis and management of group A streptococcal pharyngitis. *Infectious Diseases Society of America. Clin Infect Dis*. 2002;35:113-125.
11. Choby BA. Diagnosis and treatment of streptococcal pharyngitis. *Am Fam Physician*. 2009;79:383-390.
12. NICE. Respiratory tract infections – antibiotic prescribing. July 2008 [cited January 2011]; Available from: <http://www.nice.org.uk/CG69>
13. Matthys J, De Meyere M, van Driel ML, De Sutter A. Differences among international pharyngitis guidelines: not just academic. *Ann Fam Med*. 2007;5:436-443.
14. Laupacis A, Sekar N, Stiell IG. Clinical prediction rules. A review and suggested modifications of methodological standards. *JAMA*. 1997;277:488-494.
15. McGinn TG, Guyatt GH, Wyer PC, Naylor CD, Stiell IG, Richardson WS. How to use articles about clinical decision rules. Evidence-Based Medicine Working Group. *JAMA*. 2000;284:79-84.
16. Wasson JH, Sox HC, Neff RK, Goldman L. Clinical prediction rules. Applications and methodological standards. *The New England journal of medicine*. 1985;313:793-799.
17. Harrell FE, Jr., Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*. 1996;15:361-387.

18. Royston P, Moons KG, Altman DG, Vergouwe Y. Prognosis and prognostic research: Developing a prognostic model. *BMJ (Clinical research ed)*. 2009;338:b604.
19. Sauerbrei W, Royston P, Binder H. Selection of important variables and determination of functional form for continuous predictors in multivariable model building. *Stat Med*. 2007;26:5512-5528.
20. Royston P, Ambler G, Sauerbrei W. The use of fractional polynomials to model continuous risk variables in epidemiology. *Int J Epidemiol*. 1999;28:964-974.
21. Steyerberg EW. *Clinical Prediction Models*. 1st ed: Springer 2009.
22. Mirza A, Wludyka P, Chiu TT, Rathore MH. Throat culture is necessary after negative rapid antigen detection tests. *Clin Pediatr (Phila)*. 2007;46:241-246.
23. McIsaac WJ, Goel V, To T, Low DE. The validity of a sore throat score in family practice. *CMAJ*. 2000;163:811-815.
24. Hamburg MA, Collins FS. The path to personalized medicine. *The New England journal of medicine*. 2010;363:301-304.
25. Ravdin PM, Siminoff LA, Davis GJ, Mercer MB, Hewlett J, Gerson N, *et al*. Computer program to assist in making decisions about adjuvant therapy for women with early breast cancer. *J Clin Oncol*. 2001;19:980-991.
26. Cohen J, Chalumeau M, Levy C, Thollot F, Bingen E, Cohen R. Spectrum and inoculum size effect of a rapid antigen detection test for group A streptococcus pharyngitis in children. (Submitted).
27. Ebell MH, Smith MA, Barry HC, Ives K, Carey M. The rational clinical examination. Does this patient have strep throat? *JAMA*. 2000;284:2912-2918.
28. Charlier-Bret N, Boucher B, Poyart C, Quesne G, Bingen E, Doit C, *et al*. Tests de diagnostic rapide des angines a streptocoque du groupe A : étude comparative de la sensibilité et de la praticabilité de 16 dispositifs médicaux de diagnostic in vitro dans le cadre d'un contrôle du marché effectuée par l'Afssaps en juillet 2002. *Pathol Biol (Paris)*. 2004;52:438-443.
29. Bidet P, Plainvert C, Doit C, Mariani-Kurkdjian P, Bonacorsi S, Lepoutre A, *et al*. Infections a Streptococcus pyogenes ou streptocoque du groupe A chez l'enfant : donnees du Centre national de reference (CNR). *Arch Pediatr*. 2010;17:201-208.
30. Rubin D. *Multiple Imputation for Nonresponse in Surveys*. Hoboken, NJ: J Wiley 1987.
31. Efron B, Gong G. A leisurely look at the bootstrap, the jackknife and cross validation. *Am Stat*. 1983;37:36-48.
32. Molinaro AM, Simon R, Pfeiffer RM. Prediction error estimation: a comparison of resampling methods. *Bioinformatics*. 2005;21:3301-3307.
33. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988;44:837-845.

---

ANNEXE 7. VALIDATION DU TEST  $Q$  DE COCHRAN POUR EXPLORER

L'HETEROGENEITE DE RAPPORTS DE VRAISEMBLANCE

(article accepté pour publication dans le *Journal of Clinical Epidemiology*)

**Cochran's  $Q$  test can be used to assess heterogeneity in likelihood ratios in studies of diagnostic accuracy**

Jérémie F. Cohen<sup>a,b,c</sup>, Martin Chalumeau<sup>a,b</sup>, Robert Cohen<sup>d,e</sup>, Daniël A. Korevaar<sup>c</sup>, Babak Khoshnood<sup>a</sup>, Patrick M. M. Bossuyt<sup>c</sup>

*<sup>a</sup>INSERM U1153, Obstetrical, Perinatal and Pediatric Epidemiology Research Team, Research Center for Epidemiology and Biostatistics Sorbonne Paris Cité (CRESS), Paris Descartes University, Paris, France;*

*<sup>b</sup>Department of Pediatrics, Necker-Enfants Malades Hospital, Assistance Publique-Hôpitaux de Paris, Paris Descartes University, Paris, France;*

*<sup>c</sup>Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Academic Medical Center, University of Amsterdam, Amsterdam, the Netherlands;*

*<sup>d</sup>Association Clinique et Thérapeutique Infantile du Val-de-Marne (ACTIV), Saint-Maur-des-Fossés, France;*

*<sup>e</sup>Unité Court Séjour Nourrisson, Centre Hospitalier Intercommunal de Créteil, Créteil, France.*

**CORRESPONDING AUTHOR:**

Jérémie F. Cohen, MD, MSc

INSERM U1153, Obstetrical, Perinatal and Pediatric Epidemiology Research Team

Maternité de Port Royal

53 avenue de l'Observatoire 75014 PARIS

Phone: +33(0)142345570

Fax: +33(0)143268979

Email: jeremie.cohen@inserm.fr

**WORD COUNT: 2379**

## **ABBREVIATIONS**

LR, likelihood ratio

PLR, positive likelihood ratio

NLR, negative likelihood ratio

95CI, 95% confidence interval

TP, true-positive result

FP, false-positive result

FN, false-negative result

TN, true-negative result

## **COMPETING INTERESTS**

We declare no conflicts of interest.

## **ACKNOWLEDGMENTS**

We thank Professor Alfredo Morabia, MD, PhD (Barry Commoner Center for the Biology of Natural Systems, Queens College, City University of New York, NY, USA) for providing *R* code needed to apply Goehring's method. J.F.C. is a PhD student supported by educational grants from Agence Régionale de Santé d'Ile-de-France, Laboratoires Guigoz - Société Française de Pédiatrie - Groupe de Pédiatrie Générale - Groupe de Recherches Épidémiologiques en Pédiatrie and research grants from the French Ministry of Health (PHRC Régional 2012, AOR 12089) and Assistance Publique-Hôpitaux de Paris (Année Médaille de l'Internat 2013). The funders had any role in study design, data collection and analysis, decision to publish, or preparation of the article.

## **ABSTRACT**

**Objective** – Empirical evaluations have demonstrated that diagnostic accuracy frequently shows significant heterogeneity between subgroups of patients within a study. We propose to use Cochran’s  $Q$  test to assess heterogeneity in diagnostic likelihood ratios (LRs).

**Study Design and Setting** – We reanalysed published data of six articles that showed within-study heterogeneity in diagnostic accuracy. We used the  $Q$  test to assess heterogeneity in LRs and compared the results of the  $Q$  test to those obtained using another method for stratified analysis of LRs, based on subgroup confidence intervals. We also studied the behaviour of the  $Q$  test using hypothetical data.

**Results** – The  $Q$  test detected significant heterogeneity in LRs in all six example datasets. The  $Q$  test detected significant heterogeneity in LRs more frequently than the confidence interval approach (38% vs. 20%). When applied to hypothetical data, the  $Q$  test would be able to detect relatively small variations in LRs, of about a twofold increase, in a study including 300 participants.

**Conclusion** – Reanalysis of published data using the  $Q$  test can be easily performed to assess heterogeneity in diagnostic LRs between subgroups of patients, potentially providing important information to clinicians who base their decisions on published LRs.

**KEYWORDS** – Bayes theorem; sensitivity and specificity; diagnostic techniques and procedures/statistics & numerical data; data interpretation, statistical; likelihood functions; predictive value of tests

**RUNNING TITLE** – Heterogeneity in likelihood ratios

**WORD COUNT** – 195 words

## WHAT IS NEW?

- There is no simple way to assess heterogeneity in diagnostic likelihood ratio on the basis of published subgroup data.
- This study provides empirical evidence that Cochran's  $Q$  test can be used to explore within-study heterogeneity in likelihood ratios, similar to what can be done to explore between-study heterogeneity in meta-analysis.
- The  $Q$  test detected significant heterogeneity in likelihood ratios in all example datasets and was more powered than a previously published confidence interval approach.
- Readers of primary studies can reanalyse published data using the  $Q$  test to assess heterogeneity in likelihood ratios between subgroups. The analysis can be easily performed using any spreadsheet program.

## INTRODUCTION

In clinical practice, physicians need to know whether the result of a diagnostic test provides sufficient information to rule out or to confirm a diagnosis, or if another test is warranted to further reduce diagnostic uncertainty. Diagnostic likelihood ratios (LRs) became increasingly popular in the past decades. They allow clinicians to use Bayes' theorem in estimating post-test probabilities of disease, with use of published estimations of LRs, estimated pre-test probabilities and the result for a given patient of the diagnostic test they used [1].

Empirical data have shown that diagnostic accuracy can vary substantially within a primary study [2-8]. If there are significant variations in LRs between subgroups of patients, defined by patient characteristics, the overall summary estimates of LRs from a given study might lead to invalid estimates of post-test probabilities in a given patient, and thus to inappropriate clinical decision-making. When clinicians evaluate a report of diagnostic accuracy, they should be able to assess whether they can rely on the overall LRs of the study to take medical decisions.

In diagnostic research, several modelling strategies have been proposed to evaluate whether variables, such as age or sex, affect diagnostic test results through significant variations in LRs, mostly through logistic regression [9-12]. These models can be used to explore heterogeneity in LRs and to estimate subgroup-specific LRs that take the risk profiles of patients into account. They allow for an assessment of the conditional effect of variables through multivariable analysis.

Modelling requires access to individual patient data. If the authors of a study of diagnostic accuracy did not provide the results of such a regression analysis, readers may still want to know whether there is significant heterogeneity in LRs between subgroups. To our knowledge, only one method has been proposed to perform a stratified analysis on the basis of aggregated subgroup-specific data from published studies. Goehring *et al.* proposed to

estimate the ratio between the LR observed in a subgroup of patients ( $LR_{sub}$ ) over the LR observed in the overall population ( $LR$ ) [13]. They suggested that significant heterogeneity in LRs is present if  $LR_{sub}/LR$  differs from unity. As a test, they proposed to evaluate whether the 95% confidence interval (95CI) of the ratio contains unity. The main drawback of this method is the complex calculation of the 95CI for the ratio of LRs, especially if the subgroup variable has more than two categories. Moreover, this method does not take the effect of the size of each subgroup on the estimation of the overall LR into account: there is no weighting.

The issue of variation in LRs between subgroups within a study can be viewed as a classic question of heterogeneity. In that sense, it is similar to the issue of between-study heterogeneity in meta-analysis. A simple way to assess heterogeneity in meta-analysis is to perform Cochran's  $Q$  test. The  $Q$  test is based on the estimation of a weighted average of effect in the overall population and the evaluation of whether stratum-specific effects significantly depart from the common effect. Here, we present how Cochran's  $Q$  test can be used to detect significant variations in LRs between subgroups of patients within a study.

## COCHRAN'S $Q$ FOR HETEROGENEITY IN LIKELIHOOD RATIOS

We consider a sample of individuals in which the diagnostic accuracy of an index test is estimated, by comparing test results against the classification from a clinical reference standard. Let our sample of individuals be stratified in  $k$  subgroups according to a clinical variable such as age or sex. We can build 2x2 contingency tables and calculate corresponding LRs for each subgroup  $i$ . First, a summary estimate of LRs is calculated as a pooled average of the subgroup LRs using the Mantel-Haenszel method [14]. Then, Cochran's  $Q$  statistic of heterogeneity is given by:

$$Q = \sum_i w_i (\theta_i - \theta_{MH})^2 \sim \chi_{k-1}^2$$

where  $\theta_i$  is the estimate of the LR for each subgroup,  $w_i$  is the inverse variance weight given specific to each subgroup, and  $\theta_{MH}$  is the pooled summary estimate of LRs using Mantel-Haenszel weighting (on a logarithmic scale). Statistical details about the  $Q$  test are presented in appendix 1.

## METHODS

### *Application of the $Q$ test to published examples and comparison with Goehring's method*

As an example, we applied the  $Q$  test to six published studies. We selected the three articles already chosen by Goehring *et al.* to illustrate their stratified analysis approach [3, 4, 6], the studies used by Janssens *et al.* and Elie *et al.* to illustrate their LRs modelling framework [11, 12], and a study in which some of us participated [5]. These studies were chosen because they explicitly aimed to demonstrate the existence of heterogeneity in accuracy between subgroups of patients and provided data that allowed reanalysis.

First, we extracted the data needed to build the 2x2 tables for the whole study population and for each subgroup. We reanalysed the data to calculate LRs and corresponding 95CIs in each study and in each subgroup using the method described by Simel *et al.* [15]. Second, we applied the  $Q$  test for LRs as described above. Third, we applied Goehring's method to each dataset. We then compared the results from the  $Q$  test with those obtained using Goehring's method by comparing the proportion of significant variations in LRs detected by each method. Goehring's method assesses the effect of subgroup variables using a 95CI approach in each of the two subgroups defined by the subgroup variable, which increases type I error to 10%. To keep type I error equal to 5% using both methods, variations in LRs were considered significant using Goehring's method only if 95CI did not contain 1 in both subgroups. We arbitrarily chose to restrict the analysis to 20 comparable stratified analyses of PLRs and NLRs, for a total of 40 stratified analyses. Statistical analysis involved use of *R* (R Foundation for Statistical Computing, Vienna, Austria).

### *Investigation of the statistical power of the $Q$ test*

We additionally studied the behaviour of the  $Q$  test under different conditions of variations in LRs to estimate the minimum difference in LRs that the  $Q$  test could detect using data from

hypothetical studies of diagnostic accuracy. We arbitrarily considered the situation of a highly accurate diagnostic test, especially to rule in the disease. For that purpose, we set the parameters of the experiment as follows: two subgroups of patients of equal size, constant disease prevalence across subgroups 0.40,  $PLR_{common}$  6.0, specificity 0.90, and constant global accuracy across subgroups (i.e., diagnostic odds ratio between 10 and 20 across subgroups), when possible. We arbitrarily set the PLR to 3.0 in the first subgroup and considered different situations in which the PLR would be higher in the second subgroup, following a progressive increase. The number of participants was first set to 150, which seems to be typical for a study of diagnostic accuracy [16, 17], and then to 300. Setting these parameters allowed us to calculate the numbers in each cell of the 2x2 table for each subgroup in each situation. We used the diagnostic test accuracy calculator implemented in Revman 5.2 to calculate the number of true-positives (TP), true-negatives (TN), false-positives (FP) and false-negatives (FN) that would match our parameters as much as possible. Then, we applied the  $Q$  test in each situation using  $R$ .

## RESULTS

### *Application of the Q test to published examples and comparison with Goehring's method*

Table 1 shows the main characteristics of the six example studies and overall summary estimates of sensitivity, specificity, and LRs. Table 2 shows the results of applying the *Q* test and Goehring's method to the six datasets we reanalysed. For example, in Lachs' study, the *Q* test showed a significantly lower NLR in patients with high clinical probability of urinary tract infection (UTI) as compared to patients with low clinical probability of UTI (0.18 [0.07–0.49] vs. 0.57 [0.34–0.96],  $NLR_{MH}=0.39$  [0.25–0.62],  $p=0.038$ ). In comparison, Goehring's method showed discrepant results between subgroups, therefore nonsignificant, according to the clinical probability of UTI (0.76 [0.41–1.44] and 2.42 [1.73–3.40] for ratios of NLRs). In total, the *Q* test found significant heterogeneity in LRs in all six example studies. The *Q* test found heterogeneity in LRs more frequently than Goehring's method (38% [15/40] cases vs. 20% [8/40] cases).

### *Investigation of the statistical power of the Q test*

Table 3 shows the results of the *Q* test applied to data from two hypothetical diagnostic accuracy studies. In the case of a study with 150 participants, the *Q* test was unable to detect statistical significance when assessing variability in PLRs, even in the case of a tenfold increase of PLR between subgroups (3.00 [1.79–5.03] vs. 30.42 [1.86–496.62],  $p=0.061$ ). In contrast, the *Q* test detected statistical significance when assessing heterogeneity in NLRs when the NLR followed a 2.3-fold increase between subgroups: 0.27 [0.13–0.57] versus 0.62 [0.46–0.83],  $p=0.019$ ). In the case of a study with 300 participants, the *Q* test could detect statistical significance for about a twofold increase in PLRs and NLRs (3.00 [2.08–4.32] vs. 6.00 [3.26–11.06],  $p=0.049$ ; 0.27 [0.16–0.46] vs. 0.50 [0.38–0.66],  $p=0.032$ , respectively).

## DISCUSSION

In this article, we have described the use of Cochran's  $Q$  as a valid but yet very simple test to assess heterogeneity in LRs between patient subgroups. In a series of six published example studies, the  $Q$  test found statistically significant variations in all studies and was statistically more powered than Goehring's method to detect significant heterogeneity in LRs. When evaluating the behaviour of the  $Q$  test on hypothetical data, we found that the  $Q$  test might suffer from low power in case of a small number of participants in the study.

We believe our method has advantages over the method previously proposed by Goehring *et al.* to assess heterogeneity in LRs on the basis of published subgroup data [13]. First, our approach based on the investigation of heterogeneity is steadily understandable by clinicians because meta-analysis has become widely accepted by the health sciences community over the past 20 years. The  $Q$  test approach can also be considered a classical weighted least squares test procedure analogous to homogeneity tests of risk ratios [7]. The  $Q$  test for heterogeneity of LRs that we propose can be easily performed using any spreadsheet editor or any free statistical software dedicated to meta-analysis of diagnostic accuracy studies (appendix 2). In contrast, Goehring's method needs the use of advanced statistical software such as SAS or R, which may not be accessible to all clinicians. Another advantage of the  $Q$  test-approach is that it allows for the comparison of LRs even when the subgroup variable has more than two levels (i.e., non-dichotomic variables). Goehring's method estimates a test statistic and a corresponding 95CI for each level of the subgroup variable, increasing the likelihood of a type I error, whereas the  $Q$  test provides a unique p-value for the overall effect of the subgroup variable on LRs.

Our method has a number of limitations. It needs the extraction of subgroup specific data to build 2x2 tables. Because authors sometimes only report rounded estimates of diagnostic accuracy in each subgroup, we cannot be sure that the numbers of TP, FP, FN and

TN we extracted corresponded exactly to those from the original studies. Backcalculation of numbers in contingency tables can lead to errors, especially when the total number of participants in the study is large and when the estimates of diagnostic accuracy are not precisely reported.

The  $Q$  test might suffer from insufficient power to detect true heterogeneity between subgroups of patients, especially when the number of patients and subgroups is small, which is quite common in studies of diagnostic accuracy. Nevertheless, the  $Q$  test was able to detect several significant variations both in NLRs and PLRs when applied to published examples. When applying the  $Q$  test to hypothetical datasets, we observed that it detected smaller variations in NLRs than in PLRs. Studies of diagnostic accuracy usually include more non-diseased than diseased patients.

A third limitation of our method lies in its univariate approach, which ignores the paired nature of test accuracy data. We propose to analyse PLRs and NLRs separately, whereas it might be more rigorous to analyse PLRs and NLRs jointly. Zwinderman *et al.* have shown that separate fixed- or random-effects meta-analysis of PLR and NLR can lead to inadmissible pairs of sensitivity and specificity (i.e., point estimates out of [0;1]) [18]. Since both the PLR and NLR are a function of sensitivity and specificity, it may be more sensible to use a single test to evaluate homogeneity of test accuracy between subgroups.

It is increasingly recognized that variability in diagnostic accuracy is the rule rather than the exception. A test's sensitivity, specificity, PLR, and NLR are not fixed characteristics, but properties that describe the behaviour of a test in a specific clinical setting, in a well-defined patient population. Despite encouragement to report on variability, as in the STARD recommendations, and the availability of statistical methods to do so, explicit explorations of heterogeneity are not yet frequent in reports of test accuracy evaluations [19, 20].

We have proposed a way to perform stratified analysis on LRs and such a method cannot assess the effect of a given variable on LRs conditional on other variables. We conducted these analyses to address issues that clinicians might encounter when reading the results from a study of diagnostic accuracy. The  $Q$  test should be used to assess heterogeneity in LRs within primary studies only when one does not have access to the full dataset. The  $Q$  test is not to replace modelling, and we strongly advise that authors who want to assess variations in LRs and have access to individual patient data use multivariable regression strategies. However, the  $Q$  test might be useful as a first step when exploring heterogeneity in LRs and selecting candidate variables that might be relevant for multivariable analysis. In such a situation, a p-value  $<0.20$  could be used, considering that the  $Q$  test has relatively low power.

We believe the  $Q$  test can help clinicians who base their decisions on published LRs, for whom statistically significant heterogeneity in LRs could discourage the use of overall summary estimates of LRs. We advise that clinicians who use the Bayes' theorem for decision making should be cautious when using published LRs and should first evaluate whether significant heterogeneity in LRs exists.

## TABLES

Table 1. Characteristics of the six published studies used as example datasets with corresponding overall measures of diagnostic accuracy

Author	Target condition	Index test	Reference standard	N	Disease prevalence	Sensitivity (95CI)	Specificity (95CI)	PLR (95CI)	NLR (95CI)
Lachs [3]	Urinary tract infection	Urinary dipstick	Urine culture	366	0.20	0.83 (0.75–0.92)	0.71 (0.66–0.77)	2.92 (2.37–3.59)	0.23 (0.14–0.39)
Moons [4]	Coronary artery disease	Exercise test	Coronary angiography	295	0.70	0.57 (0.51–0.64)	0.92 (0.86–0.98)	7.23 (3.52–14.85)	0.46 (0.39–0.55)
Levy [6]	Left ventricular hypertrophy	ECG	Echocardiography	4684	0.16	0.07 (0.05–0.09)	0.99 (0.98–0.99)	5.76 (3.91–8.48)	0.94 (0.92–0.96)
Janssens [11]	Renal artery stenosis	Serum creatinine	Renal angiography	426	0.23	0.74 (0.66–0.83)	0.55 (0.50–0.61)	1.66 (1.41–1.96)	0.46 (0.33–0.66)
Elie [12]	Cervical cancer	Cervical smear test <sup>a</sup>	Colposcopy <sup>b</sup>	1781	0.20	0.88 (0.85–0.92)	0.90 (0.89–0.92)	8.95 (7.61–10.51)	0.13 (0.10–0.17)
Cohen [5]	Streptococcal pharyngitis	Rapid antigen detection test	Throat culture	785	0.36	0.87 (0.82–0.90)	0.95 (0.92–0.96)	16.1 (11.1–23.2)	0.14 (0.10–0.19)

PLR, positive likelihood ratio; NLR, negative likelihood ratio.

Disease prevalence, sensitivity, specificity, likelihood ratios and corresponding 95% confidence intervals calculated from extracted data.

<sup>a</sup>Clinical reading, <sup>b</sup>followed by biopsy if abnormalities were detected.

Table 2. Application of the  $Q$  test to six published example datasets and comparison with Goehring's quantitative method

Author	Subgroups	N	Test accuracy		$Q$ test		Goehring's method			
			PLR (95CI)	NLR (95CI)	PLR <sub>MH</sub> (95CI)	$p$	NLR <sub>MH</sub> (95CI)	$p$	Ratio of PLR (95CI)	Ratio of NLR (95CI)
Lachs	Clinical probability of urinary tract infection <sup>a</sup>									
	High clinical probability	103	1.59 (1.24–2.04)	0.18 (0.07–0.49)	1.78 (1.43–2.22)	0.092	0.39 (0.25–0.62)	<b>0.038</b>	0.54 (0.41–0.71)	0.76 (0.41–1.44)
	Low clinical probability	259	2.53 (1.57–4.07)	0.57 (0.34–0.96)	–	–	–	–	0.86 (0.61–1.20)	2.42 (1.73–3.40)
Moons	Age (years)									
	28–50	117	6.64 (2.55–17.30)	0.46 (0.35–0.62)	7.32 (3.52–15.19)	0.796	0.46 (0.39–0.55)	0.966	0.92 (0.48–1.75)	1.00 (0.20–5.10)
	51–70	178	8.03 (2.67–24.12)	0.46 (0.37–0.57)	–	–	–	1.11 (0.49–2.52)	0.99 (0.19–5.34)	
Sex	Sex									
	Male	222	5.51 (2.57–11.79)	0.41 (0.33–0.51)	6.02 (2.97–12.21)	0.542	0.52 (0.44–0.60)	<b>&lt;0.001</b>	0.76 (0.54–1.07)	0.89 (0.19–4.12)
	Female	73	10.70 (1.46–78.69)	0.72 (0.58–0.90)	–	–	–	1.48 (0.24–9.03)	1.57 (0.27–9.16)	
Symptoms <sup>b</sup>	Symptoms <sup>b</sup>									
	Nonspecific	55	6.17 (1.90–20.05)	0.54 (0.34–0.87)	6.51 (3.02–14.02)	0.866	0.50 (0.41–0.61)	0.582	n/a	n/a
	Atypical	89	8.66 (2.20–34.04)	0.52 (0.39–0.69)	–	–	–	–	n/a	n/a
	Typical	151	5.20 (1.41–19.25)	0.44 (0.33–0.58)	–	–	–	–	n/a	n/a
Diabetes mellitus <sup>ac</sup>	Diabetes mellitus <sup>ac</sup>									
	Yes	34	7.40 (0.51–106.72)	0.42 (0.25–0.70)	6.24 (3.11–12.51)	0.894	0.48 (0.40–0.57)	0.587	1.18 (0.09–15.48)	0.88 (0.10–7.43)
	No	249	6.14 (2.99–12.60)	0.49 (0.41–0.58)	–	–	–	0.98 (0.81–1.18)	1.02 (0.58–1.79)	
Smoking	Smoking									
	Yes	145	9.38 (2.44–36.14)	0.42 (0.33–0.54)	6.57 (3.14–13.72)	0.451	0.49 (0.41–0.58)	0.164	1.43 (0.47–4.37)	0.89 (0.16–5.03)
	No	137	5.17 (2.20–12.17)	0.54 (0.42–0.68)	–	–	–	0.79 (0.49–1.28)	1.13 (0.28–4.54)	
Beta-blocker use	Beta-blocker use									
	Yes	168	6.49 (2.52–16.75)	0.46 (0.37–0.58)	7.18 (3.50–14.73)	0.755	0.46 (0.39–0.55)	0.968	0.90 (0.48–1.69)	1.01 (0.20–5.18)
	No	127	8.19 (2.71–24.78)	0.46 (0.36–0.60)	–	–	–	1.13 (0.50–2.59)	1.00 (0.19–5.27)	
Systolic blood pressure, baseline (mmHg) <sup>b</sup>	Systolic blood pressure, baseline (mmHg) <sup>b</sup>									
	100–140	131	16.79 (4.27–65.97)	0.37 (0.27–0.50)	6.89 (3.45–13.74)	0.059	0.45 (0.37–0.54)	<b>0.034</b>	2.32 (0.76–7.12)	0.80 (0.19–3.41)
	141–240	164	3.83 (1.67–8.77)	0.54 (0.43–0.68)	–	–	–	–	0.53 (0.32–0.88)	1.18 (0.19–7.46)
Levy	Sex									
	Women	2642	9.36 (4.85–18.08)	0.95 (0.93–0.97)	6.15 (4.16–9.09)	0.092	0.94 (0.92–0.96)	0.283	1.63 (0.96–2.75)	1.01 (0.17–6.10)
	Men	2042	4.62 (2.82–7.58)	0.93 (0.90–0.96)	–	–	–	0.80 (0.58–1.11)	0.99 (0.13–7.69)	
BMI (women) <sup>b</sup>	BMI (women) <sup>b</sup>									
	Quartile I	661	29.05 (6.21–135.87)	0.87 (0.74–1.03)	7.60 (4.03–14.35)	<b>0.044</b>	0.95 (0.92–0.97)	0.135	n/a	n/a
	Quartile II	659	26.40 (5.24–133.04)	0.92 (0.84–0.99)	–	–	–	–	n/a	n/a
	Quartile III	663	12.71 (3.42–47.17)	0.94 (0.89–0.98)	–	–	–	–	n/a	n/a
	Quartile IV	657	3.32 (1.15–9.60)	0.97 (0.95–1.00)	–	–	–	–	n/a	n/a
BMI (women)	BMI (women)									
	Quartile I&II	1320	25.13 (8.42–75.05)	0.90 (0.84–0.97)	8.26 (4.28–15.93)	<b>0.032</b>	0.95 (0.92–0.97)	0.103	<b>2.52 (1.04–6.13)</b>	0.95 (0.01–117.24)
	Quartile III&IV	1320	5.95 (2.61–13.57)	0.96 (0.94–0.98)	–	–	–	–	<b>0.60 (0.36–0.98)</b>	1.01 (0.01–203.44)

<sup>a</sup>Values were missing in the report of the original study; <sup>b</sup>Not included in the comparison of the  $Q$  test vs. Goehring's method; <sup>c</sup>Added 0.5 to all cells because of a zero cell. Statistically significant results indicated in bold.

Table 2. Application of the  $Q$  test to six published example datasets and comparison with Goehring's quantitative method (continued)

Author	Subgroups	N	Test accuracy			Q test			Goehring's method			
			PLR (95CI)		NLR (95CI)	PLR <sub>MH</sub> (95CI)		NLR <sub>MH</sub> (95CI)	Ratio of PLR (95CI)		Ratio of NLR (95CI)	
			N	PLR (95CI)	NLR (95CI)	PLR <sub>MH</sub> (95CI)	NLR <sub>MH</sub> (95CI)	p	p	Ratio of PLR (95CI)	Ratio of NLR (95CI)	
Janssens	Gender											
	Women	192	2.52 (1.71–3.72)	0.55 (0.39–0.78)	1.71 (1.47–2.00)	<b>0.003</b>	0.45 (0.32–0.63)	0.086	1.52 (1.17–1.98)	1.19 (0.96–1.47)		
	Men	234	1.45 (1.25–1.67)	0.26 (0.11–0.61)	–	–	–	0.87 (0.75–1.01)	0.56 (0.42–0.76)			
	Age (years)											
	≥50	242	1.63 (1.35–1.97)	0.38 (0.23–0.63)	1.59 (1.35–1.89)	0.657	0.49 (0.35–0.69)	0.077	0.98 (0.86–1.13)	<b>0.81 (0.68–0.96)</b>		
	<50	184	1.49 (1.04–2.14)	0.68 (0.43–1.07)	–	–	–	0.90 (0.69–1.17)	<b>1.48 (1.11–1.96)</b>			
Elic	HPV test											
	Positive	537	4.02 (3.22–5.03)	0.07 (0.04–0.11)	4.72 (3.92–5.68)	< <b>0.001</b>	0.19 (0.15–0.24)	< <b>0.001</b>	0.45 (0.37–0.55)	<b>0.54 (0.34–0.88)</b>		
	Negative	1244	9.31 (7.10–12.20)	0.38 (0.28–0.52)	–	–	–	1.04 (0.86–1.25)	<b>2.99 (2.33–3.82)</b>			
	Study setting											
	Screening	1320	12.32 (9.38–16.19)	0.32 (0.22–0.46)	3.14 (2.69–3.66)	< <b>0.001</b>	0.22 (0.16–0.28)	<b>0.001</b>	<b>1.38 (1.14–1.67)</b>	2.48 (1.98–3.11)		
	Referral clinic	461	2.34 (1.94–2.81)	0.12 (0.08–0.18)	–	–	–	<b>0.26 (0.22–0.32)</b>	0.91 (0.44–1.90)			
	Age (years)											
	<35	981	11.66 (9.14–14.87)	0.12 (0.07–0.18)	8.47 (7.22–9.94)	<b>0.001</b>	0.13 (0.10–0.17)	0.538	<b>1.30 (1.09–1.55)</b>	0.91 (0.67–1.24)		
	≥35	800	6.79 (5.48–8.41)	0.14 (0.10–0.20)	–	–	–	<b>0.76 (0.65–0.88)</b>	1.09 (0.76–1.57)			
	Current smoking <sup>a</sup>											
	Yes	602	7.22 (5.61–9.30)	0.11 (0.07–0.18)	8.56 (7.27–10.07)	0.051	0.13 (0.10–0.17)	0.494	0.81 (0.66–0.99)	0.89 (0.60–1.33)		
	No	1161	9.99 (8.09–12.32)	0.14 (0.10–0.20)	–	–	–	1.12 (0.98–1.27)	1.10 (0.84–1.44)			
	European origin											
	Yes	1623	9.71 (8.14–11.58)	0.13 (0.10–0.18)	8.68 (7.40–10.18)	<b>0.001</b>	0.13 (0.10–0.17)	0.803	<b>1.09 (1.02–1.16)</b>	1.02 (0.90–1.16)		
	No	158	4.74 (3.22–6.99)	0.12 (0.05–0.29)	–	–	–	<b>0.53 (0.37–0.77)</b>	0.90 (0.49–1.66)			
	Educational level <sup>a</sup>											
	< Higher	1063	7.15 (5.93–8.62)	0.14 (0.10–0.19)	8.80 (7.49–10.33)	< <b>0.001</b>	0.13 (0.10–0.17)	0.625	<b>0.79 (0.72–0.88)</b>	1.06 (0.78–1.44)		
	≥ Higher	709	14.37 (10.42–19.84)	0.12 (0.07–0.19)	–	–	–	<b>1.59 (1.22–2.09)</b>	0.91 (0.63–1.32)			
	Contraception <sup>b</sup>											
	None	640	8.30 (6.42–10.73)	0.12 (0.08–0.20)	8.95 (7.61–10.51)	0.769	0.13 (0.10–0.17)	0.756	n/a	n/a		
	Combined oral pill	905	9.26 (7.35–11.67)	0.14 (0.10–0.21)	–	–	–	n/a	n/a			
	Other	236	9.68 (6.11–15.34)	0.10 (0.04–0.23)	–	–	–	n/a	n/a			
Cohen	Age (years)											
	3–8	663	15.13 (10.24–22.36)	0.12 (0.08–0.17)	15.53 (10.73–22.48)	0.624	0.15 (0.11–0.20)	< <b>0.001</b>	0.94 (0.82–1.08)	0.83 (0.43–1.58)		
	9–14	122	20.44 (6.53–64.01)	0.35 (0.21–0.57)	–	–	–	1.27 (0.44–3.70)	2.45 (0.62–9.72)			
	Lack of cough											
	Yes	428	15.30 (9.18–25.49)	0.11 (0.07–0.16)	15.59 (10.67–22.79)	0.887	0.15 (0.11–0.20)	<b>0.022</b>	0.95 (0.67–1.36)	0.75 (0.25–2.24)		
	No	357	16.12 (9.40–27.65)	0.21 (0.14–0.32)	–	–	–	1.01 (0.68–1.48)	1.49 (0.53–4.23)			
	Anterior cervical adenopathy											
	No	407	36.79 (16.62–81.43)	0.17 (0.12–0.25)	15.57 (10.82–22.41)	<b>0.003</b>	0.14 (0.11–0.19)	0.118	<b>2.29 (1.15–4.55)</b>	1.24 (0.45–3.39)		
	Yes	378	10.06 (6.67–15.18)	0.11 (0.07–0.18)	–	–	–	<b>0.63 (0.50–0.78)</b>	0.76 (0.25–2.31)			

<sup>a</sup>Values were missing in the report of the original study; <sup>b</sup>Not included in the comparison of the  $Q$  test vs. Goehring's method. Statistically significant results indicated in bold.

Table 3. Power of the  $Q$  test to detect variations in LRs according to the number of participants in the study (disease prevalence 0.40)

Subgroups	N	Test accuracy				$Q$ test			
		Sensitivity (95CI)	Specificity (95CI)	PLR (95CI)	NLR (95CI)	PLR <sub>MH</sub> (95CI)	$p$	NLR <sub>MH</sub> (95CI)	$p$
<b>N=150</b>									
Total study sample	150	0.60 (0.48-0.72)	0.90 (0.84-0.96)	6.00 (3.12-11.53)	0.44 (0.32-0.61)	-	-	-	-
Subgroup 1	75	0.80 (0.61-0.92)	0.73 (0.58-0.85)	3.00 (1.79-5.03)	0.27 (0.13-0.57)	3.43 (2.31-5.09)	0.479	0.26 (0.16-0.44)	0.869
Subgroup 2	75	0.80 (0.61-0.92)	0.80 (0.65-0.90)	4.00 (2.17-7.37)	0.25 (0.12-0.52)	-	-	-	-
Subgroup 1	75	0.80 (0.61-0.92)	0.73 (0.58-0.85)	3.00 (1.79-5.03)	0.27 (0.13-0.57)	3.67 (2.37-5.68)	0.278	0.33 (0.22-0.51)	0.448
Subgroup 2	75	0.67 (0.47-0.83)	0.87 (0.73-0.95)	5.00 (2.28-10.98)	0.39 (0.23-0.65)	-	-	-	-
Subgroup 1	75	0.80 (0.61-0.92)	0.73 (0.58-0.85)	3.00 (1.79-5.03)	0.27 (0.13-0.57)	3.88 (2.48-6.09)	0.165	0.33 (0.21-0.51)	0.481
Subgroup 2	75	0.67 (0.47-0.83)	0.89 (0.76-0.96)	6.00 (2.53-14.24)	0.38 (0.22-0.63)	-	-	-	-
Subgroup 1	75	0.80 (0.61-0.92)	0.73 (0.58-0.85)	3.00 (1.79-5.03)	0.27 (0.13-0.57)	4.00 (2.47-6.48)	0.107	0.40 (0.28-0.58)	0.130
Subgroup 2	75	0.53 (0.34-0.72)	0.93 (0.82-0.99)	8.00 (2.55-25.10)	0.50 (0.34-0.74)	-	-	-	-
Subgroup 1	75	0.80 (0.61-0.92)	0.73 (0.58-0.85)	3.00 (1.79-5.03)	0.27 (0.13-0.57)	4.06 (2.46-6.71)	0.088	0.41 (0.29-0.59)	0.088
Subgroup 2	70 <sup>a</sup>	0.50 (0.31-0.69)	0.95 (0.83-0.99)	10.00 (2.47-40.45)	0.53 (0.37-0.76)	-	-	-	-
Subgroup 1	75	0.80 (0.61-0.92)	0.73 (0.58-0.85)	3.00 (1.79-5.03)	0.27 (0.13-0.57)	4.07 (2.42-6.84)	0.075	0.46 (0.34-0.64)	<b>0.019</b>
Subgroup 2	70 <sup>a</sup>	0.40 (0.23-0.59)	0.98 (0.87-1.00)	16.00 (2.20-116.39)	0.62 (0.46-0.83)	-	-	-	-
Subgroup 1	75	0.80 (0.61-0.92)	0.73 (0.58-0.85)	3.00 (1.79-5.03)	0.27 (0.13-0.57)	4.18 (2.46-7.09)	0.061	0.48 (0.36-0.65)	<b>0.012</b>
Subgroup 2	70 <sup>a</sup>	0.37 (0.20-0.56)	1.00 (0.91-1.00)	30.42 (1.86-496.62)	0.64 (0.49-0.84)	-	-	-	-
<b>N=300</b>									
Total study sample	300	0.60 (0.51-0.69)	0.90 (0.86-0.94)	6.00 (3.78-9.52)	0.44 (0.36-0.56)	-	-	-	-
Subgroup 1	150	0.80 (0.68-0.89)	0.73 (0.63-0.82)	3.00 (2.08-4.32)	0.27 (0.16-0.46)	3.43 (2.59-4.53)	0.317	0.26 (0.18-0.38)	0.816
Subgroup 2	150	0.80 (0.68-0.89)	0.80 (0.70-0.88)	4.00 (2.60-6.16)	0.25 (0.15-0.42)	-	-	-	-
Subgroup 1	150	0.80 (0.68-0.89)	0.73 (0.63-0.82)	3.00 (2.08-4.32)	0.27 (0.16-0.46)	3.67 (2.69-4.99)	0.125	0.33 (0.25-0.45)	0.283
Subgroup 2	150	0.67 (0.53-0.78)	0.87 (0.78-0.93)	5.00 (2.87-8.72)	0.39 (0.27-0.56)	-	-	-	-
Subgroup 1	150	0.80 (0.68-0.89)	0.73 (0.63-0.82)	3.00 (2.08-4.32)	0.27 (0.16-0.46)	3.88 (2.82-5.34)	<b>0.049</b>	0.33 (0.24-0.45)	0.319
Subgroup 2	150	0.67 (0.53-0.78)	0.89 (0.81-0.95)	6.00 (3.26-11.06)	0.38 (0.26-0.54)	-	-	-	-
Subgroup 1	150	0.80 (0.68-0.89)	0.73 (0.63-0.82)	3.00 (2.08-4.32)	0.27 (0.16-0.46)	4.00 (2.84-5.63)	<b>0.022</b>	0.40 (0.31-0.52)	<b>0.032</b>
Subgroup 2	150	0.53 (0.40-0.66)	0.93 (0.86-0.98)	8.00 (3.56-17.96)	0.50 (0.38-0.66)	-	-	-	-
Subgroup 1	150	0.80 (0.68-0.89)	0.73 (0.63-0.82)	3.00 (2.08-4.32)	0.27 (0.16-0.46)	4.06 (2.85-5.79)	<b>0.016</b>	0.41 (0.32-0.53)	<b>0.016</b>
Subgroup 2	140 <sup>a</sup>	0.50 (0.37-0.63)	0.95 (0.88-0.99)	10.00 (3.72-26.86)	0.53 (0.41-0.68)	-	-	-	-
Subgroup 1	150	0.80 (0.68-0.89)	0.73 (0.63-0.82)	3.00 (2.08-4.32)	0.27 (0.16-0.46)	4.07 (2.82-5.87)	<b>0.012</b>	0.46 (0.37-0.58)	<b>0.001</b>
Subgroup 2	140 <sup>a</sup>	0.40 (0.28-0.54)	0.98 (0.91-1.00)	16.00 (3.93-65.09)	0.62 (0.50-0.76)	-	-	-	-
Subgroup 1	150	0.80 (0.68-0.89)	0.73 (0.63-0.82)	3.00 (2.08-4.32)	0.27 (0.16-0.46)	4.20 (2.90-6.08)	<b>0.006</b>	0.48 (0.39-0.60)	<b>&lt;0.001</b>
Subgroup 2	150	0.37 (0.25-0.50)	0.99 (0.94-1.00)	33.00 (4.57-238.35)	0.64 (0.53-0.78)	-	-	-	-

<sup>a</sup>We had to reduce the number of patients in subgroup 2, for a disease prevalence of 0.43, in order to set PLR to 10.00, 16.00, and 30.42, respectively. Statistically significant results indicated in bold.

## REFERENCES

1. Deeks JJ, Altman DG. Diagnostic tests 4: likelihood ratios. *BMJ* 2004;329:168-9.
2. Ransohoff DF, Feinstein AR. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. *N Engl J Med* 1978;299:926-30.
3. Lachs MS, Nachamkin I, Edelstein PH, Goldman J, Feinstein AR, Schwartz JS. Spectrum bias in the evaluation of diagnostic tests: lessons from the rapid dipstick test for urinary tract infection. *Ann Intern Med* 1992;117:135-40.
4. Moons KG, van Es GA, Deckers JW, Habbema JD, Grobbee DE. Limitations of sensitivity, specificity, likelihood ratio, and Bayes' theorem in assessing diagnostic probabilities: a clinical example. *Epidemiology* 1997;8:12-7.
5. Cohen JF, Chalumeau M, Levy C, Bidet P, Thollot F, Wollner A, et al. Spectrum and inoculum size effect of a rapid antigen detection test for group A streptococcus in children with pharyngitis. *PLoS One* 2012;7:e39085.
6. Levy D, Labib SB, Anderson KM, Christiansen JC, Kannel WB, Castelli WP. Determinants of sensitivity and specificity of electrocardiographic criteria for left ventricular hypertrophy. *Circulation* 1990;81:815-20.
7. Mulherin SA, Miller WC. Spectrum bias or spectrum effect? Subgroup variation in diagnostic test evaluation. *Ann Intern Med* 2002;137:598-602.
8. Willis BH. Spectrum bias--why clinicians need to be cautious when applying diagnostic test studies. *Fam Pract* 2008;25:390-6.
9. Knottnerus JA. Application of logistic regression to the analysis of diagnostic data: exact modeling of a probability tree of multiple binary variables. *Med Decis Making* 1992;12:93-108.
10. Leisenring W, Pepe MS. Regression modelling of diagnostic likelihood ratios for the evaluation of medical diagnostic tests. *Biometrics* 1998;54:444-52.
11. Janssens AC, Deng Y, Borsboom GJ, Eijkemans MJ, Habbema JD, Steyerberg EW. A new logistic regression approach for the evaluation of diagnostic test results. *Med Decis Making* 2005;25:168-77.
12. Elie C, Coste J. A methodological framework to distinguish spectrum effects from spectrum biases and to assess diagnostic and screening test accuracy for patient populations: application to the Papanicolaou cervical cancer smear test. *BMC Med Res Methodol* 2008;8:7.
13. Goehring C, Perrier A, Morabia A. Spectrum bias: a quantitative and graphical analysis of the variability of medical diagnostic test performance. *Stat Med* 2004;23:125-35.
14. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst* 1959;22:719-48.
15. Simel DL, Samsa GP, Matchar DB. Likelihood ratios with confidence: sample size estimation for diagnostic test studies. *J Clin Epidemiol* 1991;44:763-70.
16. Bachmann LM, Puhan MA, ter Riet G, Bossuyt PM. Sample sizes of studies on diagnostic accuracy: literature survey. *BMJ* 2006;332:1127-9.
17. Bochmann F, Johnson Z, Azuara-Blanco A. Sample size in studies on diagnostic accuracy in ophthalmology: a literature survey. *Br J Ophthalmol* 2007;91:898-900.
18. Zwinderman AH, Bossuyt PM. We should not pool diagnostic likelihood ratios in systematic reviews. *Stat Med* 2008;27:687-97.
19. Irwig L, Bossuyt P, Glasziou P, Gatsonis C, Lijmer J. Designing studies to ensure that estimates of test accuracy are transferable. *BMJ* 2002;324:669-71.
20. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. Standards for Reporting of Diagnostic Accuracy. *Clin Chem* 2003;49:1-6.

**Cochran's  $Q$  test can be used to assess heterogeneity in likelihood ratios in studies of diagnostic accuracy**

Jérémie F. Cohen, Martin Chalumeau, Robert Cohen, Daniël A. Korevaar, Babak Khoshnood,  
Patrick M. M. Bossuyt

**COMPETING INTERESTS**

We declare no conflicts of interest.

**FINANCIAL DISCLOSURE**

J.F.C. is a PhD student supported by educational grants from Agence Régionale de Santé d'Ile-de-France, Laboratoires Guigoz - Société Française de Pédiatrie - Groupe de Pédiatrie Générale - Groupe de Recherches Epidémiologiques en Pédiatrie and research grants from the French Ministry of Health (PHRC Régional 2012, AOR 12089) and Assistance Publique-Hôpitaux de Paris (Année Médaille de l'Internat 2013). The funders had any role in study design, data collection and analysis, decision to publish, or preparation of the article.

**WEB-ONLY APPENDIX 1**

**Cochran’s Q For heterogeneity in likelihood ratios**

*Testing framework*

We consider a sample of  $N$  individuals in which the diagnostic accuracy of an index test is estimated, by comparing test results against the classification from a clinical reference standard. The reference standard defines disease status; diseased individuals being noted  $D$  and non-diseased individuals being noted  $ND$ . By definition, the positive LR (PLR) and negative LR (NLR) can be expressed as:

$$PLR = sensitivity/(1-specificity)$$

$$NLR = (1-sensitivity)/specificity$$

Let  $a_i, b_i, c_i$  and  $d_i$  be the number of true-positive (TP), false-positive (FP), false-negative (FN) and true-negative (TN) test results in our sample of  $N$  patients, stratified by a variable  $X$ , where  $i$  ( $i=1, \dots, k$ ) indicates the level of the subgroup variable; with  $N_i$  the total number of patients in level  $i$  being the sum of diseased individuals ( $D_i$ ) and non-diseased individuals ( $ND_i$ ) in level  $i$ .

We can build contingency tables and calculate corresponding  $PLR_i$  and  $NLR_i$  for each level of the subgroup variable (eTable 1). If we assume there is a common PLR across subgroups of patients defined by subgroup variable levels, and let  $PLR_{common}$  denote that common PLR, then our hypotheses of interest are:

$$H_0: PLR_1 = PLR_2 = \dots = PLR_k = PLR_{common}$$

$$H_1: PLR_i \neq PLR_j \text{ for at least two subgroups } (i, j = 1, \dots, k)$$

A pooled summary estimate of  $PLR_{common}$  can be calculated as a weighted average of the subgroup PLR estimates using the Mantel-Haenszel method (fixed-effect assumption) [1] by:

$$LR_{MH} = \frac{\sum_i w_i^{MH} LR_i}{\sum_i w_i^{MH}}$$

with the Mantel-Haenszel weights for PLRs being:

$$w_i^{MH} = \frac{b_i D_i}{N_i}$$

Similarly for NLRs with Mantel-Haenszel weights being:

$$w_i^{MH} = \frac{d_i D_i}{N_i} .$$

### *Test statistics*

The test statistics, which are also based on binary response data, formally test for the heterogeneity of the PLR and NLR, respectively. Because LRs are ratios of two proportions we can assume a normal approximation for the distribution of the logarithm of LRs [2]. Further analysis of heterogeneity of LRs will be performed after a logarithmic transformation. Cochran's  $Q$  statistic can be calculated as the weighted sum of squared differences between subgroup variable level estimates of LRs and the pooled Mantel-Haenszel estimate of the common LR across subgroups on a logarithmic scale:

$$Q = \sum_i w_i (LnLR_i - LnLR_{MH})^2$$

where  $LnLR_i$  is the estimate of the logarithm of the LR for each level of the subgroup variable,  $w_i$  is the inverse variance weight given specific to each subgroup variable level and  $LnLR_{MH}$  is the logarithm of the pooled summary estimate of LRs using Mantel-Haenszel weighting.

Test statistics are calculated separately for PLRs and NLRs with, for PLRs:

$$w_i^+ = \frac{1}{\text{var}(\text{LnPLR}_i)}$$

$$\text{var}(\text{LnPLR}_i) = \frac{1}{a_i} + \frac{1}{b_i} - \frac{1}{D_i} - \frac{1}{ND_i}$$

and similarly for NLRs:

$$w_i^- = \frac{1}{\text{var}(\text{LnNLR}_i)}$$

$$\text{var}(\text{LnNLR}_i) = \frac{1}{c_i} + \frac{1}{d_i} - \frac{1}{D_i} - \frac{1}{ND_i}$$

If any subgroup contingency table has a zero cell, we suggest adding 0.5 to all cells in the table.

Under the null hypothesis of homogeneity of LR<sub>s</sub>, these statistics follow a  $\chi^2$  distribution with  $k-1$  degrees of freedom [3].

### *Running the Q test in a spreadsheet*

The  $Q$  test for LR<sub>s</sub> can be easily run in any spreadsheet editor. An example Microsoft Excel spreadsheet with the calculations for two subgroups of patients is provided. In Excel, enter data for the number of TP, FP, FN and TN for each subgroup. The spreadsheet then calculates the LR<sub>s</sub> in each subgroup and the Mantel-Haenszel weighted common LR<sub>s</sub>, and performs the  $Q$  test as described above. We manually checked that our Excel spreadsheet gave consistent results with the  $Q$  test run using R (R Foundation for Statistical Computing, Vienna, Austria).

**eTable 1. Contingency table for each level of the subgroup variable**

Index test	Reference standard		Sum
	Positive	Negative	
Positive	$a_i$	$b_i$	$a_i + b_i$
Negative	$c_i$	$d_i$	$c_i + d_i$
Sum	$a_i + c_i = D_i$	$b_i + d_i = ND_i$	$a_i + c_i + b_i + d_i = N_i$

## References

1. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst* 1959;22:719-48.
2. Simel DL, Samsa GP, Matchar DB. Likelihood ratios with confidence: sample size estimation for diagnostic test studies. *J Clin Epidemiol* 1991;44:763-70.
3. Cochran WG. The combination of estimates from different experiments. *Biometrics* 1954;10:101-29.

---

ANNEXE 8. FICHES DE RECUEIL DE DONNEES

**Enquête sur les TDR dans les infections possiblement streptococciques de l'enfant  
Groupe Angine chez l'enfant de 3 à 15 ans**

Date de consultation

N° de patient

Date de naissance

Sexe  M  F

Initiales du patient :

	/	
1ère lettre du nom		1ère lettre du prénom

**Critères d'inclusion et de non-inclusion (toutes les cases doivent être cochées)**

- |   |  |
|---|--|
| <input type="checkbox"/> Angine fébrile et/ou douloureuse | <input type="checkbox"/> Pas d'antibiothérapie depuis ≤ 7 jours      |
| <input type="checkbox"/> Enfants âgés de 3 à 15 ans       | <input type="checkbox"/> Enfant n'ayant pas été inclus depuis 1 mois |

**Examen clinique**

**Température** maximale observée depuis le début de cet épisode ,  C°

*La notion de fièvre a été sollicitée par l'interrogatoire des parents ?*  Oui  Non

**Dysphagie ou douleur pharyngée**

absente  modérée  intense  inconnue

si présente

*spontanément décrite par les parents*  ou *sollicitée par l'interrogatoire des parents*

**Toux**

absente  modérée  intense  inconnue

si présente

*spontanément décrite par les parents*  ou *sollicitée par l'interrogatoire des parents*

**Rhinorrhée**  Oui  Non

**Enrouement**  Oui  Non

**Conjonctivite**  Oui  Non

**Aspect du pharynx**

Oedématisé (augmentation du volume des amygdales)  Oui  Non

Erythémateux  Oui  Non

Erythémato-pultacé (exsudat)  Oui  Non

**Autres signes**

**Adénopathies sous-angulo-maxillaires**

Absentes

Présentes

taille

<1 cm

1 à 2 cm

> 2 cm

sensibles  Oui  Non

**Eruption**

Oui

Non

évocatrice de Scarlatine

Oui  Non

**Lecture du test rapide**

Positif

Faiblement positif

Négatif

Ininterprétable

**Quelque soit le résultat du test**, adresser le 2ème écouvillon à Robert Debré avec le bon de bactériologie complété dans le conteneur

Faxer ce formulaire d'inclusion à ACTIV (01 48 85 18 58) à la fin de la consultation

Tampon et signature  
de l'investigateur

Enquête sur les TDR dans les infections possiblement streptococciques de l'enfant  
Groupe Angine chez l'enfant de 3 à 15 ans

Date de consultation       2 0 1

N° de patient

Date de naissance          Sexe  M  F

Initiales du patient :

 /   
 1ère lettre du nom / 1ère lettre du prénom

**Critères d'inclusion et de non-inclusion** (toutes les cases doivent être cochées)

- Angine et/ou Pharyngite et/ou mal de gorge  Pas d'antibiothérapie depuis  $\leq$  7 jours
- Enfants âgés de 3 à 15 ans  Enfant n'ayant pas été inclus depuis 1 mois

**Examen clinique**

- Contage familial de pathologie à streptocoque du groupe A**  Oui  Non
- Début brutal**  Oui  Non
- Température**  
maximale observée depuis le début de cet épisode    C°
- La notion de fièvre a été sollicitée par l'interrogatoire des parents ?*  Oui  Non
- Dysphagie ou douleur pharyngée**  absente  inconnue  
 modérée  intense
- si présente a t-elle été spontanément décrite par les parents*  Oui  Non
- si présente a t-elle été sollicitée par l'interrogatoire des parents*  Oui  Non
- Toux**  absente  inconnue  
 modérée  intense
- si présente a t-elle été spontanément décrite par les parents*  Oui  Non
- si présente a t-elle été sollicitée par l'interrogatoire des parents*  Oui  Non
- Rhinorrhée**  Oui  Non
- Enrouement**  Oui  Non
- Conjonctivite**  Oui  Non
- Céphalées**  Oui  Non

**Aspect du pharynx**

- Oedématié (augmentation du volume des amygdales)  Oui  Non
- Erythémateux  Oui  Non
- Erythémato-pultacé (exsudat)  Oui  Non
- Purpura du voile  Oui  Non

**Signes digestifs**

- Nausées  Oui  Non
- Douleurs abdominales  Oui  Non
- Vomissements  Oui  Non
- Diarrhée  Oui  Non

**Adénopathies sous-angulo-maxillaires**

- Absentes  Présentes
- si présentes, taille  <1 cm  1 à 2 cm  > 2 cm
- si présentes, sensibles  Oui  Non

**Eruption**  Oui  Non

Si oui évocatrice de Scarlatine  Oui  Non

**Lecture du test rapide**

- Positif  Faiblement positif  Négatif  Ininterprétable

**Quelque soit le résultat du test**, adresser le 2ème écouvillon à Robert Debré avec le bon de bactériologie complété dans le conteneur

Faxer ce formulaire d'inclusion à ACTIV (01 48 85 18 58) à la fin de la consultation

Tampon et signature  
de l'investigateur

## TDR SGA : étude ancillaire

1. En quelle année avez-vous soutenu votre thèse de médecine ?

2. Quel est votre nombre d'heures moyen d'activité clinique au cabinet /semaine ?

3. Combien êtes-vous de pédiatres dans le cabinet où vous exercez ?

4. Quel est votre âge (années) ?

5. Pour vous, le TDR SGA est un test diagnostique ...

- très mauvais
- mauvais
- bon
- très bon

6. Avez-vous accès à internet au cabinet ?

- oui
- non

7. D'après vous, quelle est l'incidence en nombre de cas/an du rhumatisme articulaire aigu en France chez les enfants âgés de 3 à 15 ans ?

## TDR SGA : étude ancillaire

8. Quelle est la part de vos patients bénéficiant de la Couverture Maladie Universelle (CMU) ?

[entre 0% et 100%]

9. Avez-vous une activité clinique hospitalière régulière ?

oui

non

10. Durant l'étude SGA, qui faisait la plupart du temps le prélèvement de gorge servant à réaliser le TDR ?

vous-même

votre assistante

l'un ou l'autre

11. Durant l'étude SGA, qui lisait les TDR de vos patients la plupart du temps ?

vous-même

votre assistante

l'un ou l'autre

12. Utilisez-vous couramment un score clinique de risque de Streptocoque du groupe A pour le diagnostic étiologique des angines (score de McIsaac, de Centor ou de Wald) ?

oui

non

## TDR SGA : étude ancillaire

13. Vous arrive-t-il de prescrire un antibiotique alors que le TDR SGA est négatif ?  
(Promis c'est anonyme !)

- jamais
- rarement
- souvent

14. De mémoire, combien de gouttes de chaque réactif faut-il utiliser pour le TDR utilisé dans l'étude SGA (StreptAtest) ?

- 2
- 3
- 4
- 5
- Ne sait pas

15. Utilisez-vous un moyen précis de mesure des temps recommandés par le fabricant du StreptAtest (1 minute pour l'étape d'extraction de l'écouvillon puis 5 minutes pour l'étape d'immersion de la bandelette) ?

- oui
- non

16. Utilisez-vous l'un de ces autres tests de diagnostic rapide en routine au cabinet ?

- TDR grippe
- micro-CRP
- non

17. De mémoire, quelle est la fréquence des souches de SGA résistant à la pénicilline ?

- 0%
- 0-20%
- >20%
- Ne sait pas

## TDR SGA : étude ancillaire

18. Comment écouvillonnez-vous le plus souvent la gorge du patient pour réaliser le TDR SGA ?

- 1 amygdale
- 2 amygdales
- 1 amygdale + voile ou piliers
- 2 amygdales + voile ou piliers

19. Quand vous réalisez un TDR SGA, faites-vous attention à ne pas toucher la langue et la face interne des joues ?

- jamais
- autant que possible
- toujours

20. De mémoire, quelle est la fréquence moyenne du SGA dans les angines de l'enfant ?

- <20%
- 20-40%
- >40%