

Mémoire de Maîtrise en médecine No 3320

Evaluation d'une approche de mise à jour de SYNDROC,

un programme d'aide au diagnostic des syndromes dysmorphiques

(Evaluation of an updating approach of SYNDROC, a computer-aided diagnosis of human malformations)

Etudiant

Vahldieck Lukas

Tuteur

Prof. Schorderet Daniel Institut de Recherche en Ophtalmologie

Expert

Prof. Jean-Marc Joseph Dpt médico-chirurgical de pédiatrie

Lausanne, 10.01.2017

TABLE DES MATIERES

1.	RESUME		1
2.	INTRODU	JCTION	2
	2.1.	SYNDROC	3
	2.1.1.	Introduction	3
	2.1.2.	En pratique	3
	2.1.3.	L'algorithme descriptif	
	2.1.4. 2.1.5.	L'algorithme pseudo-bayésien Performance du programme	
3.	METHOD	OLOGIE	7
٠.		Etude 1 : Evaluation des performances de SYNDROC	
	3.1.1.	La sélection des articles dans la littérature médicale spécialisée	
	3.1.2.	L'extraction des case reports des articles sources	
	3.1.3.	Le choix des signes introduits dans SYNDROC	
	3.1.4.	La génération des premiers résultats	9
	3.2.	Etude 2 : Impact de la mise à jour de SYNDROC	9
	3.2.1.	La vérification de l'échantillonnage des groupes test (A) et contrôle (B)	
	3.2.2.	La mise à jour de SYNDROC	
	3.2.3.	La génération des deuxièmes résultats	10
4.	RESULTA	ATS	11
	4.1.	Etude 1 : Evaluation des performances de SYNDROC	11
	4.1.1.	La reconnaissance des signes et symptômes	
	4.1.2.	L'algorithme descriptif	
	4.1.3.	L'algorithme pseudo-bayésien	12
	4.2.	Etude 2 : Impact de la mise à jour de SYNDROC	12
	4.2.1.	La reconnaissance des signes et symptômes	
	4.2.2. 4.2.3.	L'algorithme descriptifL'algorithme pseudo-bayésien	
	4.2.3.	L algorithme pseudo-bayesien	10
5.		SION	
	5.1.	Etude 1 : Evaluation des performances de SYNDROC	
	5.1.1. 5.1.2.	Les facteurs externes à SYNDROC	
	0=.		
	5.2.	Etude 2 : Impact de la mise à jour de SYNDROC	
	5.2.1.	L'analyse des résultats après la mise à jour	21
6.	CONCLU	SION	22
7.	BIBLIOG	RAPHIE	23
8	ANNEXE	s	24

1. RESUME

Le diagnostic des maladies héréditaires rares de l'homme (prévalence inférieure à 5/10'000) (1) a toujours été une tâche difficile de par leurs caractéristiques. En effet, celles-ci sont rares, très nombreuses et leur nombre en constante augmentation. (2,3) Il en résulte inévitablement un manque d'expérience des médecins à les reconnaitre.

Pour faire face à ce manque d'expérience, les premiers outils informatiques d'aide au diagnostic ont été élaborés il y a 30 ans déjà. SYNDROC a été le premier. (2) Ces programmes ont pour objectif d'assister les médecins dans le diagnostic différentiel des situations complexes (4). Ils disposent d'une base de données de syndromes et de signes ainsi que divers algorithmes permettant d'établir une liste de diagnostics différentiels en fonction des signes et symptômes introduits par son utilisateur. (2) (4)

L'étude que j'ai menée a permis d'évaluer un processus de mise à jour des bases de données de SYNDROC dans le but d'améliorer ses performances. Un ensemble de 100 case reports extraits de la littérature spécialisée, regroupant 50 syndromes différents a été utilisé.

De manière générale sur les 100 case reports avant la mise à jour, les résultats obtenus n'atteignent pas la même magnitude que ceux rapportés dans les études précédentes. Une concordance de diagnostic de 71% (contre 91% rapporté en 1987) parmi les 3 premières propositions de l'algorithme descriptif de SYNDROC a été atteinte. Cette perte de précision s'explique par la sélection différente des case reports utilisés et l'évolution de SYNDROC.

La mise à jour effectuée de SYNDROC à l'aide de la moitié des case reports a entrainé une amélioration considérable de SYNDROC à proposer les diagnostics concordants de l'autre moitié des case reports. Elle a permis la reconnaissance de tous les syndromes testés en utilisant l'algorithme descriptif et a amené à une réduction du nombre de case reports non reconnus de l'algorithme pseudo-bayésien de plus de la moitié. Parmi les syndromes identifiés avant la mise à jour, elle a entrainé une augmentation de 24% du nombre de diagnostics concordants dans les 3 premières propositions de l'algorithme descriptif et une augmentation de 18% dans les 5 premières propositions de l'algorithme pseudo-bayésien.

Le processus de mise à jour s'est montré efficace et peut être poursuivi. Pour augmenter la spécificité de la description de SYNDROC en ce qui concerne l'algorithme descriptif l'ajout de signes non discriminatifs à la base de données devrait néanmoins être évité.

Key words: CAD, algorithm, databases

2. INTRODUCTION

Le diagnostic des maladies héréditaires rares de l'homme (prévalence inférieure à 5/10'000) (1) a toujours été une tâche difficile de par leurs caractéristiques. En effet, celles-ci sont rares, très nombreuses et leur nombre en constante augmentation. (2,3) Il en résulte inévitablement un manque d'expérience des médecins à les reconnaitre. Pouvoir poser le bon diagnostic amène cependant dans la plupart des cas des bénéfices de prise en charge pour le patient, sa famille et le clinicien. (5)

Pour faire face à ce manque d'expérience, les premiers outils informatiques d'aide aux diagnostics ont été élaborés il y a 30 ans déjà. SYNDROC a été le premier. (2) Ces programmes ont pour objectif d'assister les médecins dans le diagnostic différentiel des situations complexes (4). Ils disposent d'une base de données de syndromes et de signes ainsi que divers algorithmes permettant d'établir une liste de diagnostics différentiels en fonction des signes et symptômes introduits par son utilisateur. (2) (4)

L'utilité et l'efficacité de tel programme d'aide au diagnostic différentiel ainsi que la consultation de bases de données informatisées en générale ont été démontrées il y a plus de 20 ans déjà. (4) De nos jours, les ordinateurs sont devenus une partie intégrante de la recherche scientifique et de la pratique clinique quotidienne (6,7). Les outils informatiques permettant un diagnostic médical sont devenus de plus en plus nombreux, sophistiqué et spécifiques, mais souvent très onéreux (8).

Les programmes d'aide aux diagnostic différentiels basés sur la description clinique, tel que SYNDROC dont il est sujet dans cette étude, restent un moyen simple, rapide et peu onéreux d'orienter le diagnostic et les suites de prise en charge pour un clinicien expérimenté.

Cette étude a pour but d'évaluer l'impact d'une mise à jour à petite échelle de SYNDROC sur sa performance d'établir une liste de diagnostics différentiels et de décider de la faisabilité d'une extension de ce processus de mise à jour.

Dans un premier temps, les performances actuelles de SYNDROC à proposer le diagnostic concordant de 100 case reports extraits de la littérature spécialisée, regroupant 50 syndromes différents, ont été mesurées et comparés aux résultats rapportés dans les 3 études précédentes sur SYNDROC (étude 1).

Dans un deuxième temps, la moitié des cases reports est utilisé pour effectuer une mise à jour des bases de données de SYNDROC. L'impact de la mise à jour est évalué à l'aide des résultats obtenus par la deuxième moitié des case reports (étude 2).

2.1. SYNDROC

2.1.1. Introduction

SYNDROC est un programme informatique d'aide au diagnostic différentiel des malformations humaines rares. Ecrit en 1985 et continuellement mis à jour, il permet l'établissement d'un diagnostic différentiel basé sur les signes et symptômes introduits par son utilisateur. Sa base de données initiale comptait 700 signes et 300 syndromes distincts. (2) La base de données actuelle regroupe plus de 1100 syndromes différents et 814 signes.

Afin d'établir une liste de diagnostics différentiels, ce programme a recourt à deux algorithmes : descriptif et pseudo-bayésien. (9) Lors de sa création en 1985, le programme disposait encore d'un troisième algorithme dit heuristique permettant le raffinement du diagnostic, en questionnant le clinicien sur la présence ou l'absence de signes pivot. (2)

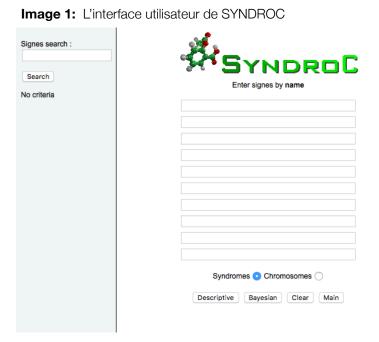
Une version de SYNDROC était accessible en ligne librement sous l'adresse "www.SYNDROC.ch". Les coûts importants d'hébergement du programme ont malheureusement provoqué récemment la fermeture de son site internet en attendant d'autres options.

2.1.2. En pratique

L'utilisation de SYNDROC ne présuppose aucune connaissance informatique, un examen clinique détaillé du patient ainsi qu'une recherche approfondie de ses antécédents médicaux sont les seuls prérequis à la bonne utilisation du programme.

A l'ouverture du programme l'utilisateur de SYNDROC a le choix d'effectuer une recherche de syndromes à partir de signes et de symptômes - fonctionnalité du programme dont il est sujet dans cette étude - ou d'afficher les signes d'un syndrome.

Lors de la recherche de syndromes à partir de signes et de symptômes, l'utilisateur dispose d'un outil de recherche permettant de trouver l'équivalent SYNDROC des signes et symptômes que présente son patient. (cf. Image 1)



Par recherche de syndrome, SYNDROC permet l'introduction de 10 signes. L'utilisateur a la possibilité de raffiner le diagnostic différentiel en recherchant dans la base de données des syndromes ou des anomalies chromosomiques de SYNDROC. Cette dernière fonctionnalité n'a pas pu être prise en considération dans cette étude en raison des difficultés rencontrées dans la recherche d'articles source.

Une fois tous les signes introduit dans SYNDROC, l'utilisateur a le choix entre les deux algorithmes. Dans les deux cas, SYNDROC affiche une liste de syndromes avec leur identifiant MIM McKusick (Mendelian Inheritance in Man, identifiant internationale des maladies génétiques) groupés selon le nombre de signes reconnus (algorithme descriptif) ou le coefficient de certitude (algorithme pseudo-bayésien) et trier par ordre alphabétique au sein d'un groupe. Dans ses deux publications, l'auteur de SYNDROC parle de "fenêtre de diagnostics différentiels", dans cette étude le mot "position" est préféré pour désigner un même groupe.

Pour chaque syndrome, il y a également la possibilité d'en afficher la description complète avec une mise en évidence des signes et symptômes reconnus par SYNDROC dans la description introduite. Dans cette vue, un code couleur indique également l'importance du signe au sein de la description de SYNDROC du syndrome.

2.1.3. L'algorithme descriptif

La description clinique des syndromes est au cœur de l'algorithme descriptif. Il permet de lister les syndromes selon le nombre de signes qu'ils ont en commun avec la description introduite dans SYNDROC. Des syndromes partageant la même position dans le résultat de cet algorithme partagent tous le même nombre de signe qu'ils ont en commun avec la description introduite, cependant leur combinaison peut varier au sein d'une position. (2)

Ainsi, la performance de l'algorithme descriptif à identifier les syndromes augmente avec chaque description nouvelle d'un syndrome (auto-apprentissage). Les relations des différents signes et symptômes entre eux sont sans importance dans cette démarche diagnostic, leur association doit seulement être suffisamment stable. (2)

2.1.4. L'algorithme pseudo-bayésien

L'algorithme pseudo-bayésien repose sur le théorème de Bayes permettant de calculer la probabilité conditionnelle P(M|C), c'est à dire la probabilité d'une maladie (M) connaissant la présentation clinique (C). (cf. Image 2)

P(M|C) est égale à la probabilité P(C|M) de retrouver cette présentation clinique (C) si la maladie (M) est présente, multipliée par la fréquence de cette maladie P(M) dans cette population et divisée par l'incidence de la présentation clinique (C) dans la population générale. (2)

Image 2:

Equation bayésienne de la probabilité conditionnelle P(M | C) sur laquelle repose l'algorithme pseudo-bayésien de SYNDROC.

$$P(M \mid C) = \frac{P(C \mid M) \times P(M)}{P(C)}$$

Appliqué à la médecine ces variables restent pour la plupart inconnues et ainsi il est plus facile d'utiliser une estimation. (2,9) La probabilité d'un signe clinique utilisé dans cet algorithme ne provient pas de la population, mais est fonction du nombre de lien que celui-ci a dans la base de données. L'auteur de SYNDROC parle pour cette raison d'un algorithme pseudo-bayésien. (9)

Dans SYNDROC, chaque signe est lié à un diagnostic par trois variables ; M1 l'importance du signe pour le diagnostic, M2 la fréquence du signe (calculée), et M3 la fréquence du diagnostic dans la population générale. (2)

Avec l'augmentation du nombre de syndromes dans la base de données de SYNDROC l'estimation correcte de P(C|M) est devenue de plus en plus difficile et a amené en 1987 au développement de l'algorithme descriptif. (9)

La liste finale de diagnostics différentiels proposée par SYNDROC est fonction d'un coefficient de certitude (CC), 0 diagnostic exclu et 1 diagnostic retenu, et non d'une probabilité réelle. (2,9) SYNDROC affiche tous les syndromes ayant un coefficient de certitude plus grand que la moitié du coefficient de certitude le plus élevé trouvé.

La principale différence entre l'approche descriptive et pseudo-bayésienne est le nombre de diagnostics proposés. En effet, celui-ci est plus important pour l'approche descriptive. De ce fait l'algorithme pseudo-bayésien est plus performant dans la distinction des syndromes. (9)

2.1.5. Performance du programme

La performance de SYNDROC à établir une liste de diagnostics différentiels a été évaluée dans 2 études différentes réalisées par l'auteur de SYNDROC respectivement en 1985 et 1987. En 1985, sur 100 descriptions de cas comptant 40 syndromes différents, une concordance de 95% a été mesurée entre les 2 premiers diagnostics proposés par SYNDROC et le diagnostic de généticiens expérimentés. (2) En 1987, sur 100 descriptions de cas comptant 100 syndromes différents, une concordance de 91% pour les trois premiers diagnostics proposés par SYNDROC a été observée. (9)

Une troisième étude réalisée en 1996 par un groupe indépendant a comparé les performances des programmes LDDB (London Dysmorphology database) (10), POSSUM (Pictures of Standard Syndromes and Undiagnosed Malformations) (11) et SYNDROC. Dans cette étude SYNDROC s'est montré être le programme d'aide au diagnostic le plus performant des trois avec un taux de concordance du premier diagnostic proposé par SYNDROC et celui des experts de 78%, en utilisant une approche d'experts. (4)

De cette même étude, une approche de novice – l'introduction de tous les symptômes cliniques pour le diagnostic - a entrainé une diminution de l'efficacité de SYNDROC à différencier les syndromes, mais a augmenté cependant la probabilité de trouver le bon diagnostic en première position. (4) En revanche, une approche d'expert – l'introduction uniquement des symptômes d'importance cliniques pour le diagnostic - a entrainé une augmentation l'efficacité de SYNDROC. (4)

3. METHODOLOGIE

3.1. Etude 1 : Evaluation des performances de SYNDROC

Dans le cadre de cette étude, 100 descriptions de cas cliniques, en anglais "case reports", de 50 syndromes héréditaires rares différents ont été récoltés dans deux journaux médicaux spécialisés, l'"American Journal of Medical Genetics" et le journal "Pediatric Dermatology", accessibles en ligne via le site internet "www.onlinelibrary.wiley.com".

3.1.1. La sélection des articles dans la littérature médicale spécialisée

Pour être retenus dans la recherche de littérature, les articles sources devaient remplir trois critères distincts: 1. Décrire un syndrome connu par SYNDROC, 2. Contenir un ou plusieurs case reports avec une description clinique suffisante - c'est à dire une description clinique composée de ≥ 2 signes SYNDROC autres que les signes non discriminatifs ou subjectifs tels que "developmental delay", "hypotonia" ou "short philtrum", 3. Etre accessible électroniquement avec les accès de l'université de Lausanne - c'est à dire publiés après 1995.

3.1.2. L'extraction des case reports des articles sources

Tous les signes cliniques des case reports des articles sources ont été extraits directement des articles ou de leurs tableaux récapitulatifs. Ils ont été ensuite introduits dans un tableau excel permettant la recherche de leur équivalent SYNDROC. Dans le cas où aucun équivalent SYNDROC a été trouvé pour le signe utilisé dans l'article source ou que le signe relevait de la normalité (périmètre crânien dans les normes, taille dans les normes, etc.), le signe n'a pas été retenu.

66 articles ont rempli les 3 critères et 341 case reports regroupant 50 syndromes différents ont été récoltés. Pour cette étude, seul 2 case reports par syndrome ont été retenus pour avoir un total de 100 descriptions cliniques de 50 syndromes différents (cf. Tableau 1). Les case reports d'un même syndrome ont été choisis et attribués de manière aléatoire au groupe test (A) et groupe contrôle (B) à l'aide un script informatique utilisant la fonction random du langage de programmation "Applescript" sur macintosh. Le groupe test (A) et le groupe contrôle (B) contiennent ainsi chacun 50 case reports regroupant 50 syndromes différents mais identiques dans les 2 groupes.

Tableau 1: Nom des 50 syndromes de 100 case reports récoltés dans la littérature spécialisées. 2 case reports par syndrome, 1 case report d'un même syndrome par groupe test (A) et contrôle (B).

Syndrome	Syndrome
Acrorenoocular	Kenny-Caffey
Alagille	Keutel
Alstrom	Knobloch
Arthrodentoosteo dysplasia	Malpuech
Asphyxiating thoracic dysplasia	Menkes
Barber-say	Multiple lentigines
Brachmann-de lange	Mutchinick
Cantu	Myhre
Cardio-facio-cutaneous	Neu-laxova
Cerebral Gigantism	Noonan
Coffin-siris	Opitz-kaveggia
Costello	Osteoglophonic
Cranio-cerebello-cardiac	Perlman
Cranioectodermal dysplasia	Pitt-Rogers-Danks
Distal arthrogryposis	Prader-Willi
EEC	Robinow
Fibrodysplasia ossificans congenita	Rubinstein-Taybi
Floating-Harbor	Russell-Silver
Gapo	Seckel
Gillespie	Shprintzen-Goldberg
Holt-Oram	Simpson-Golabi-Behmel
Johnson-mcmillin	Smith-Fineman-Myers
Kabuki make-up	Spondyloepiphyseal dysplasia congenita
Karsch-Neugebauer	Toriello-carey
Kaufman oculo-cerebro-facial	Williams

3.1.3. Le choix des signes introduits dans SYNDROC

En raison de la limitation de SYNDROC à 10 signes par recherche de diagnostic différentiel, une sélection des signes des case reports avant leur introduction dans SYNDROC a été nécessaire.

Pour les case reports composés de ≤ 10 signes SYNDROC, l'ensemble des signes a été retenu. Pour les case reports composés de > 10 signes SYNDROC, les signes les plus objectifs tel que "micro/macrocephaly", "polydactyly" ou "short stature" ont été privilégiés. Les signes non discriminatifs ou subjectifs tels que "developpmental delay", "hypotonia" ou "short philtrum" n'ont pas été retenu en faveur d'autres signes plus objectifs. (cf. Annexe 1)

3.1.4. La génération des premiers résultats

Les ≤ 10 signes choisis pour chaque case report ont été introduits de manière automatisée dans SYNDROC. Pour chaque case report, les listes de diagnostics différentiels ont été générés d'abord en utilisant l'algorithme descriptif de SYNDROC puis en utilisant l'algorithme pseudo-bayésien. Par la suite, ils ont été reportés dans un tableau excel (étude 1).

3.2. Etude 2 : Impact de la mise à jour de SYNDROC

3.2.1. La vérification de l'échantillonnage des groupes test (A) et contrôle (B)

Deux tests de Mann-Whitney ont été réalisés à l'aide du programme d'analyses statistiques SPSS 23 pour identifier une éventuelle différence entre les résultats SYNDROC obtenus dans les deux groupes en utilisant l'algorithme descriptif et ceux obtenus dans les deux groupes en utilisant l'algorithme pseudo-bayésien. Ces deux tests obtiennent une P-value > 0.05, la différence intergroupe des résultats obtenus en utilisant les deux algorithmes dans les deux groupes est donc statistiquement non significative.

Par la suite, une série de t-tests a été utilisée pour déceler d'autres éventuelles différences intergroupes statistiquement significatives.

Ils ont été comparés sur plusieurs déterminants : le nombre de signes trouvés dans les articles sources par case report, le nombre de signes introduits dans SYNDROC par case report et le nombre de signes reconnus par SYNDROC comme appartenant au syndrome concordant avec celui des articles sources.

Tous les t-tests réalisés à l'aide de Microsoft excel 2016 pour Macintosh ont montré une P-value > 0.05. H0 : groupe test (A) = groupe contrôle (B) sur les déterminants listés ci-dessus n'a donc pas pu être rejetée. Le groupe contrôle (B) a été utilisé correctement comme groupe contrôle par la suite.

3.2.2. La mise à jour de SYNDROC

Une fois les premiers résultats des 100 case reports générés, les 50 case reports du groupe test (A) ont été utilisés pour procédé à une mise à jour de la description que possède SYNDROC du syndrome concordant. Cette mise à jour a été faite en trois étapes distinctes.

Dans une première étape, tous les signes SYNDROC utilisés d'un case report mais ne figurant pas dans la description SYNDROC du syndrome concordant ont été rajoutés à la description de SYNDROC du syndrome. Le 100% des signes SYNDROC utilisés dans les case reports du groupe A a été ainsi reconnu par SYNDROC comme appartenant au syndrome concordant avec celui des articles sources.

Dans une deuxième étape, une pondération par défaut de 5, 10 ou 25 en fonction de l'objectivité du signe et de son pouvoir de discrimination, a été attribuée à chaque nouveau signe dans la description SYNDROC du syndrome et cela indépendamment du syndrome en question.

Dans une troisième étape, les pondérations de tous les signes de la description SYNDROC du syndrome concordant ont été modifiées jusqu'à ce qu'il apparaisse dans les 5 premières positions du diagnostic différentiel pseudo-bayésien. Aucune modification des pondérations des signes n'a été apportée si le syndrome concordant apparaissait déjà en première position.

3.2.3. La génération des deuxièmes résultats

Pour la génération des résultats après modification des descriptions SYNDROC des syndromes, les mêmes signes SYNDROC des case reports du groupe test (A) et contrôle (B) ont été réintroduits dans SYNDROC de la même manière que lors de la génération des premiers résultats. Seul les résultats obtenus pour le groupe contrôle (B) ont été analysés par la suite pour juger de l'impact de la mise à jour effectuée (étude 2).

4. RESULTATS

4.1. Etude 1 : Evaluation des performances de SYNDROC

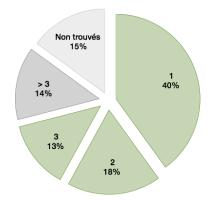
4.1.1. La reconnaissance des signes et symptômes

L'ensemble des 100 case reports (groupe A et B) extraits de la littérature spécialisée est composé de 1146 signes SYNDROC dont 837 ont pu être introduits dans SYNDROC pour effectuer les recherches de diagnostics différentiels. Au total, 372 signes ont été reconnus comme appartenant au diagnostic décrit dans les articles sources. Ainsi, en moyenne 3,72 signes (SD: +/- 1,92) par recherche ont été correctement reconnus soit 44,4% des signes introduits.

4.1.2. L'algorithme descriptif

En utilisant l'algorithme descriptif de SYNDROC, le diagnostic concordant, celui proposé dans l'article source du case report, a été listé 71 fois parmi les 3 premières propositions de SYNDROC. Dans 40 cas, il a été le premier diagnostic proposé par SYNDROC. 18 fois, il a été listé en position 2, 13 fois en position 3 et 14 fois à une position > 3. Dans 15 cas, le diagnostic concordant n'a pas été trouvé par SYNDROC.

Tableau 2: Position et nombre des diagnostics concordants proposés par l'algorithme descriptif de SYNDROC, groupe test (A) et contrôle (B): 100 case reports.

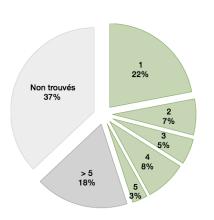


Position	n	%
1	40	40%
2	18	18%
3	13	13%
> 3	14	14%
Non trouvés	15	15%
Total	100	100%

4.1.3. L'algorithme pseudo-bayésien

En utilisant l'algorithme pseudo-bayésien de SYNDROC, le diagnostic concordant, celui proposé dans l'article source du case report, a été listé 45 fois parmi les 5 premières propositions de SYNDROC. Dans 22 cas, il a été le premier diagnostic proposé par SYNDROC. 7 fois, il a été listé en position 2, 5 fois en position 3, 8 fois en position 4, 3 fois en position 5 et 18 fois à une position > 5. Dans 37 cas, le bon diagnostic a eu un coefficient de certitude (CC) 50% inférieur à celui du meilleur diagnostic proposé et a par conséquent non été listé par SYNDROC.

Tableau 3: Position et nombre des diagnostics concordants proposés par l'algorithme pseudo-bayésien de SYNDROC, groupe test (A) et contrôle (B): 100 case reports.



Position	n	%
1	22	22%
2	7	7%
3	5	5%
4	8	8%
5	3	3%
> 5	18	18%
Non trouvés	37	37%
Total	100	100%

4.2. Etude 2 : Impact de la mise à jour de SYNDROC

Le processus de mise à jour de SYNDROC basé sur les descriptions des cases reports du groupe test (A) a induit une amélioration statistiquement significative (Wilcoxon signed ranks test, p-value < 0.01) de SYNDROC à reconnaître les différents syndromes au sein du groupe contrôle (B), pour les deux algorithmes de SYNDROC.

4.2.1. La reconnaissance des signes et symptômes

Quel que soit l'algorithme utilisé de SYNDROC, sur 421 signes introduits, 268 ont été reconnus comme appartenant au diagnostic décrit dans les articles sources, soit 91 signes nouvellement reconnus (cf. Tableau 4).

En moyenne, le nombre de signes reconnus par SYNDROC a augmenté de 1,82 signes par recherche. Ceci équivaut à une reconnaissance de 63,6% des signes introduits et ainsi à une augmentation de 21,6% par rapport aux résultats obtenus pour ce groupe contrôle (B) avant la mise à jour.

Pour 5 case reports, le 100% des signes introduits par recherche a été reconnu soit 2 cas de plus qu'avant la mise à jour.

Tableau 4: Reconnaissance des signes et symptômes des diagnostics concordants au sein du groupe contrôle (B) après la mise à jour.

		AVANT	APRES			
	Signes introduits	Signes reconnus	Signes reconnus	Signes nouveaux		
Total	421	177	268	91		
Moyenne	8.42	3.54	5.36	1.82		
Déviation standard	2.04	1.94	2.15	1.77		
%	-	42.04%	63.66%	21.62%		

4.2.2. L'algorithme descriptif

En ce qui concerne l'algorithme descriptif de SYNDROC, tous les diagnostics concordants ont été trouvés par SYNDROC après la mise à jour (cf. Tableau 5).

Dans 90% des cas (45 case reports), le diagnostic concordant a été listé parmi les 3 premières propositions de SYNDROC, soit une augmentation de 24% (12 cas reports) par rapport aux résultats obtenu avant la mise à jour.

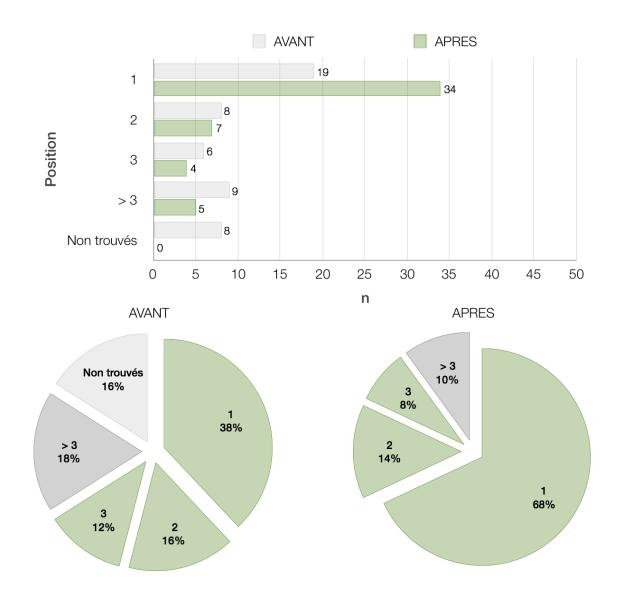
Pour 8 case reports dont le diagnostic n'a pu être identifié auparavant par SYNDROC au sein du groupe contrôle (B), le diagnostic concordant a été trouvé après la mise à jour et cela en moyenne en position 3. Pour 18 case reports le diagnostic concordant a gagné au moins une position (moyenne : +1,57 positions, max : 4 positions).

Pour 21 case reports, dont 17 ou le diagnostic concordant figurait déjà en première position, les résultats restent inchangés.

Pour 3 cases reports (Noonan, Russell-Silver, Cerebral gigantism), dont 2 figuraient en première position avant la mise à jour, le diagnostic concordant a perdu une position après la mise à jour.

Tableau 5: Position et nombre des diagnostics concordants proposés par l'algorithme descriptif de SYNDROC au sein du groupe contrôle (B) avant et après la mise à jour.

	AVANT		APR	S	
Position	n	%	n	%	
1	19	38%	34	68%	
2	8	16%	7	14%	
3	6	12%	4	8%	
> 3	9	18%	5	10%	
Non trouvés	8	16%	0	0%	
Total	50	100%	50	100%	



4.2.3. L'algorithme pseudo-bayésien

En ce qui concerne l'algorithme pseudo-bayésien de SYNDROC, le diagnostic concordant a été listé parmi les 5 premiers diagnostics proposés par SYNDROC dans 62% des cas (31 case reports) après la mise à jour, soit une augmentation de 18% (9 case reports) par rapport aux premiers résultats. Le diagnostic concordant a été trouvé 16 fois en première position, 8 fois en position 2, 5 fois en position 3, 2 fois en position 4 et 10 fois en une position > 5 (cf. Tableau 6).

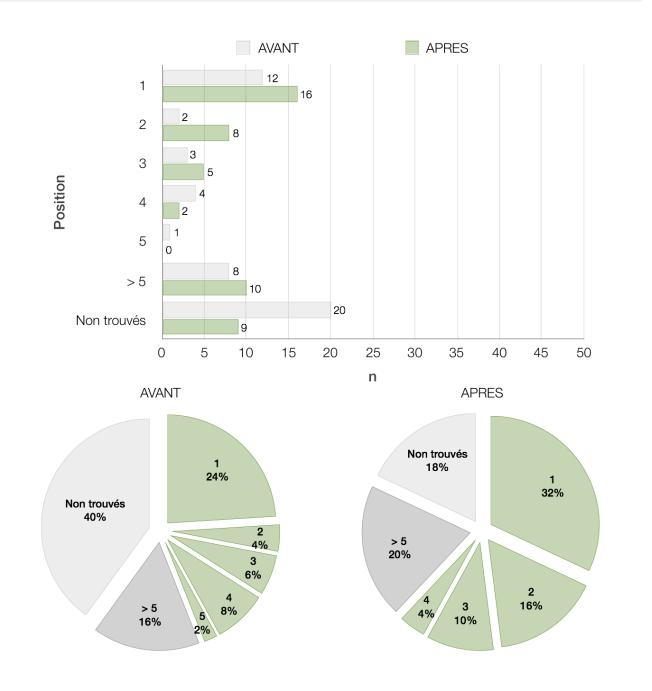
Pour 11 case reports pour lesquels le diagnostic concordant n'a initialement pas été listé par SYNDROC (coefficient de certitude < 50% du coefficient de certitude max par recherche), il a été trouvé après la mise à jour. Ceci correspond à une réduction des syndromes non listés par SYNDROC de 55% (20 cases reports avant contre 9 case reports non reconnus après la mise à jour). Aucun diagnostic concordant préalablement listé a disparu du diagnostic différentiel après la mise à jour.

Le coefficient de certitude (CC) de 24 diagnostics concordants a augmenté après la mise à jour et cela en moyenne de 0,078 points. 13 fois cette augmentation du CC a entrainé une amélioration de la position du diagnostic concordant dans le diagnostic différentiel proposé par SYNDROC, 11 fois la position a été maintenue. Aucune diminution de position n'a été observée lorsque le CC a augmenté.

Pour le diagnostic de 6 case reports aucun changement du CC a été observé, 5 fois la position du diagnostic concordant a été maintenue et une fois elle a diminué de 3 positions (Floatingharbor syndrome). Aucune amélioration de position n'a été observée lors du maintien du CC.

Tableau 6: Position et nombre des diagnostics concordants proposés par l'algorithme pseudo-bayésien de SYNDROC au sein du groupe contrôle (B) avant et après la mise à jour.

	AVANT		API	RES
Position	n	%	n	%
1	12	24%	16	32%
2	2	4%	8	16%
3	3	6%	5	10%
4	4	8%	2	4%
5	1	2%	0	0%
> 5	8	16%	10	20%
Non trouvés	20	40%	9	18%
Total	50	100%	50	100%



5. DISCUSSION

5.1. Etude 1 : Evaluation des performances de SYNDROC

L'obtention d'un diagnostic différentiel basé sur seulement 10 signes et symptômes cliniques est une tâche difficile. SYNDROC a su proposer le diagnostic concordant dans 40% des cas en première position en utilisant l'algorithme descriptif et ceci indépendamment de la répétition des syndromes (21 position 1 dans le groupe A contre 19 position 1 dans le groupe B). Cette proportion s'élève à 46% en combinant les résultats obtenus des deux algorithmes.

De manière générale, les résultats obtenus du groupe test (A) et contrôle (B) avant la mise à jour sont moins bons que ceux obtenus lors des précédentes études. (2,4,9). Pour cette raison la comparaison avec ces études est faite avec les résultats du groupe contrôle (B) obtenus après la mise à jour.

Un certain nombre de facteurs externes et internes à SYNDROC - détaillés plus bas - ont influencé les résultats.

De manière générale, l'écart considérable entre le classement des syndromes proposé par les deux algorithmes laisse suspecter que les signes introduits dans SYNDROC n'ont pas été les plus importants dans la description SYNDROC des syndromes. Le grand nombre de diagnostics proposés dans les 3 premières positions (46.72 diagnostics) de l'algorithme descriptif ainsi que 37 case reports sans diagnostic en utilisant l'algorithme pseudo-bayésien renforcent cette idée.

5.1.1. Les facteurs externes à SYNDROC

Bien qu'aucune connaissance préalable soit présupposée à l'utilisation de SYNDROC, l'utilisateur maintient une influence non négligeable sur la qualité des résultats obtenus. Trois facteurs externes à SYNDROC ayant contribué aux résultats obtenus sont discutés cidessous.

5.1.1.1. L'échantillonnage des case reports

L'extension phénotypique d'un syndrome est le tendon d'Achille des programmes informatiques d'aide au diagnostic différentiel.

Contrairement aux trois autres études qui ont utilisé des registres de patients examinés en Suisse, aux Etats-Unis ou en Allemagne, cette étude repose sur des case reports extraits de la littérature spécialisée de 1996 à 2016 et ceci n'a pas été un avantage. En effet, la sélection

des syndromes des case reports dans cette étude n'est pas le reflet de leur importance ni de leur incidence dans la population générale, mais le résultat d'un biais de publication.

Pour être publié dans un journal médical spécialisé un case report doit amener une information nouvelle et pertinente. Le plus souvent les articles sources décrivent ainsi une nouvelle présentation d'un syndrome rare, une extension phénotypique du syndrome après analyse génétique ou une présentation atypique d'un syndrome bien connu.

La reconnaissance de tels case reports est une tâche difficile pour un programme d'aide au diagnostic différentiel qui évalue la similitude de la description clinique introduite avec celle d'une description précédente d'un syndrome.

5.1.1.2. La qualité des descriptions

La qualité de la description clinique introduite est un élément au fondement même du bon fonctionnement d'un programme d'aide au diagnostic différentiel basé sur la description clinique. Comme tous les articles source ne proviennent pas des mêmes domaines de recherche, il en résulte un aperçu très détaillé mais partiel de l'expression phénotypique du syndrome qu'on pourrait appeler dans le contexte de cette étude le biais de la description. Des descriptions partielles, précises et imprécises à la fois représentent, en raison d'un vocabulaire distincts, une difficulté supplémentaire pour SYNDROC à proposer des diagnostics.

5.1.1.3. La sélection des signes introduits

Dans la troisième étude s'intéressant à SYNDROC publié en 1996 par un groupe indépendant (4), les auteurs avaient déjà mis en évidence une influence non négligeable des signes introduits sur les résultats finaux de l'algorithme descriptif de SYNDROC. Une concordance de diagnostics de 78% a été obtenu en utilisant l'approche d'expert - sélection des signes d'importance clinique - contre 68% en utilisant l'approche de novice - utilisation de tous les signes disponibles - sur un échantillon de 102 de case reports regroupant 102 syndromes différents. L'algorithme pseudo-bayésien - capable de choisir de manière autonome l'importance qu'il accorde à chaque signe introduit - n'a pas été influencé par ces différentes approches (58% de concordance contre 57% respectivement). (4)

Les deux approches de sélection des signes utilisées dans cette étude, novice – introduction de tous les signes disponibles si 10 - et semi-expert - utilisation de 10 signes en rejetant les moins objectifs - ont ainsi très probablement influencé la performance de SYNDROC à proposer des diagnostics en utilisant l'algorithme descriptif de SYNDROC.

5.1.2. Les facteurs internes à SYNDROC

Plusieurs modifications de SYNDROC depuis les dernières publications à son sujet ont influencé les résultats obtenus. Deux des facteurs internes à SYNDROC sont discutés cidessous.

5.1.2.1. L'extension de la base de données

L'augmentation du nombre de syndromes dans la base de données de SYNDROC s'est déjà montrée dans le passé comme étant un des principaux responsables de la diminution de précision diagnostic de SYNDROC. (4) (9) De plus, l'augmentation considérable du nombre de syndromes dans la base de donnés associée à un nombre de signes et symptômes relativement constant a inévitablement contribué à l'augmentation du nombre de diagnostics proposés par recherche en utilisant l'algorithme descriptif.

Au cours des 30 dernières années SYNDROC a bénéficié de mises à jour considérables. Comptant initialement 300 syndromes différents et 700 signes, SYNDROC a rapidement évolué. Seulement 2 ans après sa publication SYNDROC a vu sa base de données triplée, regroupant plus de 1000 syndromes différents. En 1996, le nombre de syndromes s'élevait à plus de 1100.

5.1.2.2. L'adaptation des fonctionnalités de SYNDROC.

Le nombre de signes pouvant être introduits par recherche de diagnostic, initialement illimité, puis limité à 50, a été réduit à 10 signes par recherche au cours des années. Ceci pourrait expliquer en partie les performances abaissées de l'algorithme pseudo-bayésien, capable de faire le tri des signes d'importance, plus que celles de l'algorithme descriptif.

5.2. Etude 2 : Impact de la mise à jour de SYNDROC

Lors de la mise à jour de SYNDROC 217 signes regroupant 116 signes distincts ont été rajoutés à l'ensemble des 50 descriptions de syndromes dans SYNDROC. Ceci équivaut à 4.34 signes en moyenne par description. (cf. Tableau 7).

Sur l'ensemble des 100 case reports, 166 signes distincts ont été non reconnus par SYNDROC (absent de la description SYNDROC du syndrome concordant), 111 de ces signes ont été utilisés 1 à 2 fois.

Lors de la mise à jour, 69 signes des 116 signes distincts non reconnus ont été ajoutés une seule fois sur 31 descriptions de syndromes différents.

Tableau 7: Fréquences des signes selon les catégories indiquées (totaux des colonnes: nombre de signes différents, **somme des signes**).

Fréquences combinées des signes du groupe test (A) et du groupe contrôle (B)									Fréquence des signes rajoutés lors de la mise à jour (Signes non reconnus du groupe contrôle A)		
Littérature	n	Introduits	n	Non introduits	n	Reconnus	n	Non reconnus	n	Rajoutés	n
Mental deficiency	60	Mental deficiency	36	Mental deficiency	24	Mental deficiency	29	Ear anomaly	27	Ear anomaly	11
Hypertelorism	39	Hypertelorism	35	Hypotony	14	Hypertelorism	25	Dental anomaly	13	Polyhydramnios	6
Short stature	36	Short stature	32	Joint hypermobility	10	Short stature	20	Micrognathia	13	Broad forehead	6
Ear anomaly	34	Ear anomaly	31	Long philtrum	9	Microcephaly	14	Short stature	12	Dental anomaly	6
Hypotony	26	Microcephaly	24	Low-set ear	8	Prenatal onset growth deficiency	13	Microcephaly	10	Prenatal onset growth deficiency	5
Microcephaly	25	Micrognathia	21	Low nasal bridge	8	Clinodactyly	9	Hypertelorism	10	Short stature	5
Micrognathia	23	Prenatal onset growth deficiency	19	Immunity deficiency	7	Macrocephaly	9	Broad forehead	9	Strabismus	5
Prenatal onset growth deficiency	20	Congenital heart defect	13	Small nose	7	Micrognathia	8	Polyhydramnios	8	Micrognathia	5
Low-set ear	18	Deafness	13	Anteverted nare	7	Deafness	7	Congenital heart defect	7	Microcephaly	5
Deafness	16	Macrocephaly	13	Broad forehead	6	Frontal bossing	7	Mental deficiency	7	Ptosis of eyelid	4
215	1146	213	837	131	309	119	372	166	465	116	21

"Mental deficiency" a été le signe le plus fréquemment retrouvé dans la description des case reports. En raison de la méthodologie appliquée dans cette étude - introduction de 10 signes si la description originale présentait > 10 signes, il a été non seulement le signe le plus fréquemment introduit mais également le plus fréquemment rejeté (non introduit).

Une attention particulière est à donner au signe "ear anomaly" rencontré 34 fois dans la littérature, le plus fréquemment sous la forme de "anteverted ears" mais restant non reconnu 27 fois. Il devient ainsi le signe le plus fréquemment rajouté lors de la mise à jour.

Parmi les signes les plus fréquemment rejetés (non introduits) on retrouve d'autre signes non discriminatifs ou subjectifs tel que "hypotony", "joint hypermobility", "long philtrum" et "low-set ear" qui démontrent une bonne sélection des signes introduits, si sélection il y a eu.

Si les signes les plus fréquemment rajoutés et ainsi non reconnus par SYNDROC avant la mise à jour permettent d'augmenter la sensibilité de SYNDROC à reconnaitre l'ensemble des syndromes testés, ce sont les signes les moins rajoutés qui augmentent la spécificité des descriptions des syndromes.

5.2.1. L'analyse des résultats après la mise à jour

La mise à jour de SYNDROC a induit une amélioration des résultats des deux algorithmes tout en réduisant le nombre de diagnostics proposés dans les 3 premières positions de l'algorithme descriptif de 45,21% en moyenne (47,6 diagnostics avant, 26.08 diagnostics après) et en ne modifiant pas significativement la fréquence à laquelle les syndromes testés sont proposés par les deux algorithmes.

5.2.1.1. L'algorithme descriptif

En ce qui concerne l'algorithme descriptif, les résultats bien que meilleurs que ceux obtenus avant la mise à jour restent inférieurs à ceux obtenus lors des études précédentes, pour les raisons détaillées plus haut (facteurs internes et externes à SYNDROC), à l'exception de 2 points.

Les résultats du groupe B après la mise à jour atteignent presque la même magnitude que ceux obtenus par l'auteur de SYNDROC en 1987 (9) où une concordance de diagnostic de 91% pour les positions 1 à 3 a été rapporté en utilisant l'algorithme descriptif contre 90% obtenus au sein du groupe B après la mise à jour.

Au sein du même groupe, on observe également une concordance de diagnostics de 68% avec la position 1 de l'algorithme descriptif, résultat identique à celui obtenu dans l'étude de 1996 par un groupe indépendant en utilisant une approche de novice. (4)

Le nombre de diagnostics proposés dans les 3 premières positions, bien que nettement réduit à 26.08 après la mise à jour reste toujours supérieur à 18,78, valeur rapportée dans l'étude de l'auteur de SYNDROC en 1987.

5.2.1.2. L'algorithme pseudo-bayésien

Les résultats de l'algorithme pseudo-bayésien restent sensiblement inférieurs à ceux rapportés par les 3 études pour les raisons détaillées plus haut (facteurs internes et externes à SYNDROC).

Les résultats qui se rapproche le plus des précédents résultats sont ceux qui concernent la concordance de diagnostics parmi l'ensemble des diagnostics proposés par l'algorithme pseudo-bayésien, 92% de concordance dans l'étude réalisée en 1987 par l'auteur de SYNDROC (9) contre 82% de concordance obtenu au sein du groupe B après mise à jour.

6. CONCLUSION

Les résultats obtenus pour les 100 case reports n'atteignent pas les magnitudes rapportées des études précédentes. Cette perte de précision observée dans les résultats de SYNDROC peut être expliquée en partie par les caractéristiques des case reports utilisés, plus favorable au processus de mise à jour qu'à l'évaluation de la performance d'un programme d'aide au diagnostic, à l'augmentation constante du nombre de syndromes dans la base de données de SYNDROC et à l'évolution au cours du temps des fonctionnalités de SYNDROC.

Cependant, en tenant compte des situations dans lesquelles l'utilisateur peut faire appel à un programme d'aide au diagnostic différentiel, de la longueur des diagnostics différentiels respectifs proposée par les deux algorithmes de SYNDROC ainsi que du fait que les résultats proviennent de l'analyse de ≤ 10 signes et symptômes cliniques, une concordance de diagnostics de plus de 70% parmi les 3 premières propositions de l'algorithme descriptif et 45% parmi les 5 premières propositions de l'algorithme pseudo-bayésien est plus que respectable.

La mise à jour effectuée de SYNDROC dans cette étude a entrainé une amélioration considérable de SYNDROC. Elle a permis la reconnaissance de tous les syndromes testés en utilisant l'algorithme descriptif et a amené à une réduction du nombre de diagnostics concordants non listés de l'algorithme pseudo-bayésien de plus de la moitié. Parmi les syndromes identifiés avant la mise à jour, elle a entrainé une augmentation de 24% du nombre de diagnostics concordants dans les 3 premières propositions de l'algorithme descriptif et une augmentation de 18% dans les 5 premières propositions de l'algorithme pseudo-bayésien.

Le processus de mise à jour s'est montré efficace et peut être poursuivi. Pour augmenter la spécificité de la description de SYNDROC en ce qui concerne l'algorithme descriptif l'ajout de signes non discriminatifs à la base de données devrait néanmoins être évité.

Le programme d'aide au diagnostic SYNDROC reste un outil précieux tout particulièrement dans les mains d'un médecin expérimenté pour lequel il a été développé. En effet, SYNDROC et son utilisateur forme encore de nos jour un binôme efficace, dans la démarche diagnostic des maladies héréditaires syndromiques rares.

7. BIBLIOGRAPHIE

- 1. Quevauvilliers J. Dictionnaire médical. 2007.
- 2. Schorderet D, Aebischer P. SYNDROC: microcomputer based differential diagnosis of malformation patterns. Arch Dis Child. BMJ Group; 1985 Mar;60(3):248–51.
- 3. OFSP. Office fédéral de la santé publique OFSP. Concept national maladies rares. 2014.
- 4. Pelz J, Arendt V, Kunze J. Computer assisted diagnosis of malformation syndromes: An evaluation of three databases (LDDB, POSSUM, and SYNDROC). American Journal of Medical Genetics. Wiley Subscription Services, Inc., A Wiley Company; 1996 May 3;63(1):257–67.
- 5. Reardon W, Donnai D. Dysmorphology demystified. Archives of Disease in Childhood Fetal and Neonatal Edition. 2007 May 1;92(3):F225–9.
- 6. Siemens. Soarian Gaining Momentum [Internet]. 2010 [cited 2017 Jan 7]. pp. 1–6. Available from: http://www.usa.siemens.com/entry/features/us_healthcare/us/en/pdf/ACHE_Trifold.pdf
- 7. Mickan S, Atherton H, Roberts NW, Heneghan C, Tilson JK. Use of handheld computers in clinical practice: a systematic review. BMC Med Inform Decis Mak. BioMed Central; 2014 Jul 6;14(1):67–10.
- 8. van Nimwegen KJM, van Soest RA, Veltman JA, Nelen MR, van der Wilt GJ, Vissers LELM, et al. Is the \$1000 Genome as Near as We Think? A Cost Analysis of Next-Generation Sequencing. Clinical Chemistry. Clinical Chemistry; 2016 Nov;62(11):1458–64.
- 9. Schorderet DF. Diagnosing human malformation patterns with a microcomputer: evaluation of two different algorithms. Wiley Subscription Services, Inc., A Wiley Company; 1987 Oct;28(2):337–44.
- 10. FDNA. London Dysmorphology database.
- 11. Murdoch Childrens Research Institute. POSSUM web. POSSUM web.

8. ANNEXES

Cf. document PDF « Annexes_TM3320_Lukas_Vahldieck.pdf »