

Table des matières

Résumé.....	III
Abstract.....	IV
Table des matières	V
Liste des figures et illustrations	VI
Liste des abréviations	VII
Avant-Propos	VIII
Introduction générale	1
Chapitre 1. Contexte de l'étude	3
1 Les outils d'évaluation alimentaire et de dépistage	3
1.1 <i>Le rappel de 24h</i>	3
1.2 <i>Le questionnaire de fréquence alimentaire</i>	4
1.3 <i>Les outils de dépistage</i>	4
2 Le développement d'outils de dépistage	6
2.1 <i>Les arbres de décision</i>	6
2.2 <i>Les modèles de régression logistique</i>	8
2.3 <i>Évaluation de la performance des outils de dépistage</i>	10
3 Conclusion partielle	13
4 Objectifs et hypothèses	14
Chapitre 2. Étude portant sur le développement et la validation d'un outil de dépistage d'une consommation adéquate en glucides chez des athlètes d'endurance.....	15
Résumé	16
Abstract.....	18
Conclusion générale.....	36
Bibliographie.....	41

Liste des figures et illustrations

Figure 1	Structure typique d'un arbre de classification.....	6
Figure 2	Représentation visuelle de différentes courbes ROC.....	11

Liste des abréviations

AUC : aire sous la courbe, pour *area under the curve*

FFQ : questionnaire de fréquence alimentaire

GLU : glucides

g : grammes

IMC : indice de masse corporelle

INAF : Institut sur la nutrition et les aliments fonctionnels

kg : kilogrammes

MC : masse corporelle

NPV : valeur prédictive négative, pour *negative predictive value*

NS : nutritionniste du sport

PPV : valeur prédictive positive, pour *positive predictive value*

ROC : *Receiver Operating Characteristics* (caractéristique de fonctionnement du récepteur)

Avant-Propos

Durant mes études graduées à la maîtrise en nutrition, réalisées à l'Institut sur la nutrition et les aliments fonctionnels (INAF) de l'Université Laval, j'ai eu l'immense privilège de réaliser deux projets de recherche différents, intitulés «*Développement et validation d'un court outil de dépistage ciblant la consommation en glucides des athlètes d'endurance*» et «*La qualité globale de l'alimentation d'athlètes non-élite impliqués dans des épreuves multi-sports d'endurance est élevée*». Cependant, le présent mémoire abordera seulement les résultats de l'étude visant à développer et valider un outil de dépistage pour sportifs.

Je me considère chanceuse d'avoir pu occuper un rôle central lors de la réalisation de ce projet, de la conception à l'analyse des résultats en passant par la rédaction d'un article scientifique, à titre de première auteure. Cet article, révisé par mon directeur de recherche et mes coauteurs, Élise Carbonneau, Denis Talbot et Simone Lemieux a été soumis à la revue *Journal of the International Society of Sports Nutrition* le 28 février 2018. Plusieurs chemins différents ont été empruntés pour arriver au but final et la tâche n'a pas toujours été facile, mais la satisfaction et les acquis que j'en retire valent plus que toutes les difficultés rencontrées durant cette dernière année et demie.

Malgré mon omniprésence dans la réalisation de ce projet, il m'aurait été impossible de réussir sans l'aide de plusieurs individus clés. Tout d'abord, j'aimerais remercier mon directeur de recherche, Dr. Benoît Lamarche, de m'avoir accueillie dans son équipe, mais aussi de m'avoir offert des opportunités d'apprentissage uniques qui ont dépassé mes attentes. C'est aussi grâce à sa passion apparente pour la recherche, qu'il a su me transmettre, que j'ai entamé, depuis l'automne dernier, mon doctorat sous sa direction. Ce projet n'aurait pas été possible sans l'aide de Denis Talbot, biostatisticien qui m'a grandement assistée lors de la réalisation de mes analyses. Il est pour moi important d'également remercier Simone Lemieux puisque le projet aurait été difficilement

réalisable sans les données récoltées dans le cadre du projet PREDISE, présentement en cours au sein de son équipe.

Je tiens également à remercier tous/toutes mes collègues de l'INAF qui ont su me soutenir et répondre à mes interrogations, mais surtout qui m'ont fait sentir à ma place parmi eux. Je n'aurais pu choisir un meilleur milieu de travail. Mes derniers remerciements sont réservés à ma famille et mes amis, qui m'ont autant encouragée à persévérer qu'à décrocher de la maîtrise, une fois de temps en temps. Je vous dois une partie de mon succès, sachez-le.

Introduction générale

Un athlète d'endurance de haut calibre se doit de porter une attention particulière aux composantes de son alimentation.¹ Via des apports adéquats pour ses besoins, l'athlète s'assure d'avoir tous les éléments nécessaires pour tirer le maximum de bénéfices de sa forme physique lors d'entraînements ou de compétitions. En effet, plusieurs macro- et micronutriments sont des alliés non-négligeables à l'optimisation de la performance et de la récupération chez un athlète, tels les protéines, le fer, le calcium et, surtout, les glucides (GLU) [1]. Ceux-ci sont en fait la pierre angulaire de l'alimentation d'un athlète et doivent occuper une place de choix dans l'assiette de celui-ci. Leur pertinence a d'ailleurs été prouvée par de multiples études et la conclusion est claire : les GLU ont leur place avant, pendant et après l'entraînement [1]. Bien sûr, la quantité et la fréquence de consommation de ceux-ci varient selon plusieurs facteurs comme la durée ou le type d'entraînement, le sport pratiqué et le poids/sexe de l'athlète. Dans la majorité des cas, la recommandation pour un athlète d'endurance se trouve entre 6 et 10 grammes (g) de GLU par kilogramme (kg) de masse corporelle (MC) [2].

Que ce soit causé par la popularité grandissante des protéines ou encore le manque de connaissances en nutrition, la majorité des athlètes d'endurance, élite ou non, ne consomme pas suffisamment de GLU pour leurs besoins. Ce constat a malheureusement été atteint par plusieurs équipes de recherche différentes, et ce dans diverses populations cibles [3-6]. Ces résultats sont inquiétants, puisqu'une consommation insuffisante de GLU peut avoir de nombreuses conséquences à court ou long terme sur la santé de l'athlète, comme une augmentation de la perception de l'effort et même un affaiblissement du système immunitaire [1, 7].

Toutefois, au quotidien, il peut être difficile, long et coûteux de mesurer les apports alimentaires, surtout chez les athlètes, considérant l'horaire chargé de ceux-ci. De

¹ Par sport d'endurance, nous entendons, par exemple, la course sur de longues distances, le cyclisme, le triathlon, etc.

plus, les ressources en nutrition sportive sont souvent limitées, ce qui complique l'évaluation des apports alimentaires de la totalité des sportifs afin de s'assurer que ceux-ci aient une alimentation adéquate [8]. D'ailleurs, il arrive fréquemment que les athlètes utilisent leur entraîneur ou leurs collègues comme source d'information en nutrition puisqu'un/une nutritionniste ne leur est pas disponible [8, 9]. Ceux-ci n'ont toutefois pas les outils ni la formation nécessaire pour adéquatement évaluer les apports alimentaires et peuvent donc difficilement fournir les conseils personnalisés nécessaires à l'atteinte des recommandations chez l'athlète. Nous avons donc cherché à développer et valider un court outil de dépistage ciblant la consommation en GLU chez les athlètes d'endurance dans le but de combler cette lacune dans la pratique en nutrition sportive. L'outil aurait comme but ultime d'identifier les athlètes ayant un apport suffisant en GLU, ce qui permettrait de diriger ceux ayant un apport jugé insuffisant vers les ressources adéquates pour optimiser leur alimentation.

Le présent mémoire sera divisé en deux chapitres, en plus de l'introduction et de la conclusion. Le premier chapitre portera sur le contexte de l'étude réalisée, c'est-à-dire une brève présentation des outils d'évaluation alimentaire présentement disponibles ainsi que les méthodes de développement d'outil applicables à notre objectif principal. Le chapitre 2 présente un article scientifique, rédigé en langue anglaise, soumis pour publication. Finalement, la conclusion générale vise à résumer les résultats de l'étude et à présenter les perspectives futures découlant de ceux-ci.

Chapitre 1. Contexte de l'étude

1 Les outils d'évaluation alimentaire et de dépistage

Dans le domaine de la nutrition, il est souvent pertinent et même nécessaire d'effectuer une évaluation primaire de l'alimentation d'un patient ou d'un client avant même de planifier une intervention. En effet, l'estimation des apports alimentaires permet de cibler le problème principal pour mieux cadrer la dite intervention. En recherche, l'évaluation alimentaire prend une toute autre dimension, c'est-à-dire qu'elle fournit une infinité d'information à manipuler pour obtenir réponses à nos diverses questions. De nombreux outils d'évaluation alimentaire sont à notre disposition, mais ne se prêtent pas nécessairement à nos besoins spécifiques. La présente section portera majoritairement sur deux outils couramment utilisés en recherche, soit le rappel de 24h et le questionnaire de fréquence alimentaire, ainsi que les outils de dépistages. Le journal alimentaire, qui est également un outil de choix dans plusieurs branches de la recherche en nutrition, ne sera pas abordé, puisqu'il est moins souvent utilisé dans la recherche en nutrition sportive.

1.1 Le rappel de 24h

Le rappel de 24h est un outil subjectif et rétrospectif qui peut être administré par un interviewer ou encore auto-administré sur le web. Comme son nom l'indique, le sujet est questionné par rapport à son alimentation la journée précédant l'administration [10]. Les informations fournies par le répondant doivent inclure le type de nourriture consommé ainsi que ses caractéristiques et ce avec le plus de précision possible (quantité consommée, méthode de préparation, marque du produit, etc.) [11]. Pour que cette méthode soit efficace, il est nécessaire de fournir au sujet des exemples visuels d'aliments et de grosseur de portions afin de minimiser la sous-estimation des apports [11]. D'ailleurs, la sous-estimation est un des limites du rappel de 24h, surtout auprès des enfants et des personnes âgées [10, 11]. Notons également la nécessité de multiples rappels de 24h pour bien

estimer l'apport régulier d'un individu et la compilation des données fastidieuse requérant un ou une nutritionniste expérimenté(e) [11]. Néanmoins, cet outil est valide, peut être utilisé auprès de population avec un faible niveau de littératie et offre une bonne répétabilité d'un intervenant à l'autre [10]. L'administration d'un unique rappel de 24h peut prendre entre 30 et 45 minutes [10].

1.2 Le questionnaire de fréquence alimentaire

Le questionnaire de fréquence alimentaire (FFQ) est également un outil rétrospectif questionnant les sujets à propos de leur alimentation usuelle durant une période de temps fixe (semaine, mois, année). Il est normalement composé de 100 à 150 questions/aliments et peut être auto-administré ou administré par entrevue. Il est composé d'une sélection d'aliments concise et structurée pour laquelle chacun des aliments est détaillé en terme de fréquence de consommation de portions usuelles [12]. Le FFQ est avantageux dans le cadre de la recherche, surtout pour des projets de grande envergure, puisqu'il est modifiable afin de cibler un nutriment en particulier et est peu coûteux pour de grands échantillons [10]. Cependant, il s'agit d'une méthode qui peut prendre du temps à administrer et à analyser et qui tend à surestimer les apports [10]. De plus, la fréquence et la grosseur de portion préétablies par le questionnaire ne reflèteront pas nécessairement l'apport usuel du répondant. L'utilisation du FFQ est plus précise lorsque combinée à un rappel de 24 heures ou encore une mesure biochimique [12]. L'administration d'un FFQ peut prendre de 30 à 60 minutes [10].

1.3 Les outils de dépistage

Malgré la qualité des outils d'évaluation alimentaire précédemment décrits, il arrive parfois que ceux-ci ne se prêtent pas au contexte de recherche ou au travail sur le terrain [13]. C'est dans cet optique que les outils de dépistage entrent en jeu, servant normalement à étudier un ou deux nutriments ou groupes alimentaires précis. La majorité du temps, ils sont issus de modifications apportées à des outils existants, comme le FFQ, ou encore composés d'une courte série de questions

spécifiques à une habitude alimentaire, outre la fréquence de consommation d'aliments [10]. Ce type d'outil est grandement utile lorsqu'il n'est pas nécessaire d'avoir un détail précis de l'alimentation globale d'un sujet. Par exemple, un outil de dépistage pourrait être utilisé sur une large cohorte afin de cibler les gens plus à risque/qui bénéficieraient le plus d'une intervention nutritionnelle, surtout dans un milieu possédant des ressources limitées en professionnels de la santé [10]. Il est aussi possible de les utiliser pour surveiller les habitudes alimentaires et l'adhésion à certaines recommandations nutritionnelles ou *patterns* alimentaires, comme la diète méditerranéenne, et ce à grande échelle [13]. Bien que leur rapidité et leur facilité d'administration soient des avantages non-négligeables, les outils de dépistages ne sont pas sans failles. En effet, leur estimation des apports usuels d'un sujet est moins précise que celle d'outils plus complets et compte souvent des erreurs de mesure [10]. Dans ce sens, avant d'utiliser un tel outil, il est recommandé de le valider auprès de la population cible, en le comparant à un outil d'évaluation alimentaire de référence, comme un rappel de 24h [10, 13].

Dans le monde du sport, il n'existe que très peu d'études visant à développer et/ou à valider des outils de dépistage. De plus, à notre connaissance, aucune de ces études ne concerne l'alimentation des athlètes, ce qui est plutôt surprenant considérant l'importance de celle-ci dans l'atteinte d'une performance optimale. En ce sens, le présent projet de maîtrise était non seulement grandement nécessaire pour la pratique en nutrition sportive, mais également innovant dans son approche. Les méthodes de développement d'outils de dépistage envisagées seront d'ailleurs présentées dans la prochaine section de ce mémoire.

2 Le développement d'outils de dépistage

Il va sans dire qu'il existe de nombreuses méthodes et approches pour développer des outils de dépistage en nutrition. Pour les besoins du présent projet de maîtrise, deux méthodes avaient été ciblées *a priori*, soit les arbres de décision et les modèles de régression logistique. De plus amples informations sur l'utilisation de chaque approche et le processus décisionnel quant au choix de la méthode retenue se trouve à même l'article scientifique présenté au chapitre 2 de ce mémoire.

2.1 Les arbres de décision

Les arbres de décision (régression et classification) sont une méthode utilisée pour construire des modèles prédictifs à partir de données d'intérêt [14]. Ces derniers sont extrêmement efficaces lorsqu'il est question d'exploration de données² [15]. Ils sont d'ailleurs appliqués dans plusieurs domaines, surtout pour leur facilité d'utilisation et d'interprétation; ils sont aisément compréhensibles par des non-statisticiens [15, 16]. Cette méthode peut être utilisée autant pour prédire une variable dépendante catégorique (arbre de classification) que continue (arbre de régression) [17]. Elle permet, entre autres, de discerner les variables à inclure dans un modèle prédictif, déterminer l'importance d'une variable d'intérêt, gérer les données manquantes, prédire un dénouement précis ou encore tout simplement manipuler des données, même si elles sont biaisées ou incomplètes [15].

² Traduction libre de *data mining* (procédé utilisé pour extraire des informations utiles de larges bases de données et pour les exposer de manière visuelle et facile à interpréter).

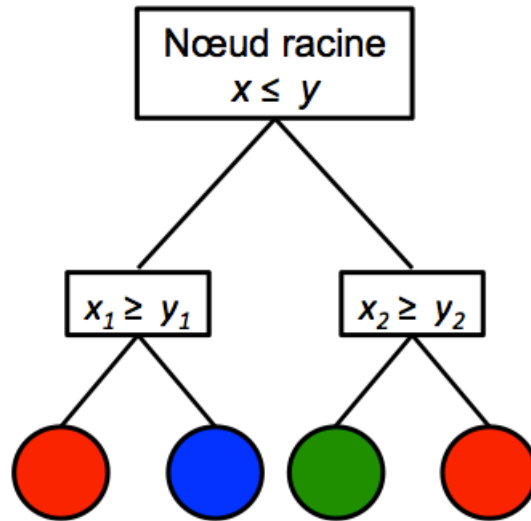


Figure 1 : Structure typique d'un arbre de classification, tirée et adaptée de Loh, 2011 [14]

La construction de l'arbre se fait par analyse discriminante, c'est-à-dire que l'algorithme teste plusieurs combinaisons de variables afin d'obtenir celle ayant le meilleur pouvoir prédictif [14]. La décision de classification de l'arbre est représentée visuellement par chaque chemin tracé du nœud racine (son premier nœud) à un des nœuds terminaux. Ceux-ci, représentés en cercles de différentes couleurs sur la figure 1, catégorisent le sujet dans un dénouement précis (souvent de type oui/non) [15].

Les arbres de décision sont souvent utilisés pour la recherche en médecine, servant majoritairement à prédire le développement d'une maladie grâce à plusieurs facteurs de risque. Par exemple, Miller et *coll.* (2014) ont développé un modèle grâce aux arbres de classification afin de détecter de manière prématurée le syndrome métabolique chez une population de jeunes adultes [16]. L'objectif derrière cette étude était de créer un outil de dépistage permettant l'identification de personnes à risque de développer une maladie cardio-métabolique chronique à l'âge adulte. Batterham et *coll.* (2009), quant à eux, ont utilisé les arbres afin d'identifier les facteurs de risques les plus associés au développement de troubles dépressifs majeurs chez les Australiens [18]. Parmi les facteurs de risque étudiés se trouvaient les symptômes dépressifs, l'usage d'alcool, de drogues et de tabac,

l'indice de masse corporelle (IMC), la présence d'hypertension et la pratique d'activité physique [18].

Brièvement, les arbres de décision présentent des avantages non-négligeables pour la création de modèles prédictifs, principalement leur facilité d'interprétation. Néanmoins, il s'agit d'une méthode présentement très peu explorée dans le domaine de la nutrition, il est donc difficile de savoir si son utilisation est appropriée pour la prédiction du risque d'apport suffisant d'un nutriment chez un individu.

2.2 Les modèles de régression logistique

La régression logistique permet de créer des modèles évaluant la relation entre :

- Une variable dépendante et dichotomique (régression logistique binomiale) ou encore une variable ayant plus de deux valeurs possibles (régression logistique multinomiale) et
- Une ou plusieurs variables explicatives indépendantes, qu'elles soient qualitatives ou quantitatives, continues ou dichotomiques. [19]

L'utilisation de cette méthode peut être motivée par deux objectifs, soit l'aspect prédictif d'un modèle de régression logistique ou encore son côté explicatif [19]. Concrètement, une régression logistique modélisera les chances qu'un certain dénouement se produise en se basant sur plusieurs caractéristiques individuelles [20, 21].

Pour développer un tel modèle, il est important de déterminer quelles variables y seront incluses [19, 20]. Pour ce faire, il est recommandé d'étudier *a priori* la relation entre les différentes variables indépendantes à considérer et la variable réponse [19]. Ces analyses peuvent être réalisées, entre autres, à partir de régressions logistiques univariées, qui étudieront le lien entre une seule variable explicative et sa variable réponse. Le modèle initial devrait inclure les variables statistiquement significatives issues de ces régressions univariées, en plus des paramètres cliniquement pertinent au dénouement étudié [19].

Cette approche a plusieurs avantages, le principal étant sa versatilité. En fait, les régressions logistiques ne nécessitent pas une distribution normale ou encore une égalité au niveau des variances, elles peuvent donc être appliquées à une multitude de situations [21]. En nutrition, la régression logistique est souvent choisie pour observer les liens entre une maladie et les habitudes alimentaires d'une population ciblée. Par exemple, Wang et *coll.* (2012) ont développé un outil de prédiction qui permet d'identifier les individus à risque élevé de dyslipidémie dans un contexte rural en utilisant cette méthode [22]. Plus précisément, ils ont utilisé la dyslipidémie (oui/non) comme variable réponse et les facteurs de risque de celle-ci comme variables explicatives. Je, Kim et Park (2017), quant à eux, se sont servi du modèle de régression logistique pour développer un outil auto-administrable cherchant à mesurer le risque de syndrome métabolique chez des adultes Coréens non-obèses [23]. Concrètement, cette équipe a utilisé comme variable réponse la présence ou non d'un syndrome métabolique et comme variables explicatives toutes celles pouvant potentiellement être associées au développement d'un syndrome métabolique. Afin de garder leur outil le plus efficient possible, Je, Kim et Park ont décidé de considérer uniquement les paramètres facilement disponibles, c'est-à-dire qui ne nécessitaient aucune mesure en laboratoire. Parmi ces paramètres se trouvaient, entre autres, l'âge, l'IMC, l'activité physique, le statut fumeur et les habitudes alimentaires par rapport au sodium.

En somme, la régression logistique multiple est une approche de choix lorsqu'il est question d'observer la relation entre une variable réponse et plusieurs variables explicatives. Comme l'apport en GLU d'un individu est affecté par la consommation de plusieurs aliments, nous trouvons intéressant de considérer la régression logistique dans le cadre des analyses liées à ce projet de maîtrise. La littérature actuelle présente peu d'études utilisant la régression logistique multiple pour prédire l'apport en nutriments d'un individu. De surcroît, il ne semble pas y avoir

d'études présentement disponibles sur l'utilisation de cette approche pour prédire l'apport en GLU d'un individu.

2.3 Évaluation de la performance des outils de dépistage

Plusieurs méthodes peuvent être utilisées pour déterminer la qualité d'un outil, qu'il soit développé par arbre de décision ou par régression logistique. La qualité de l'outil développé durant mon projet de maîtrise a été évaluée à partir des valeurs de sensibilité et spécificité, des valeurs prédictives positive et négative ainsi que l'analyse de courbes ROC (*Receiver Operating Characteristic*), ce sont donc ces mesures qui seront expliquées dans la présente section. De plus, celles-ci sont souvent utilisées lorsqu'il est question d'un modèle basé sur une variable réponse/dépendante de type oui/non [24].

2.3.1 Sensibilité, spécificité et valeurs prédictives positive et négative

D'abord, la sensibilité et la spécificité sont des valeurs complémentaires. Prenons l'exemple d'un outil qui tenterait de prédire si un individu est atteint d'une maladie; il pourrait classer un sujet comme «malade» (test positif) ou «non-malade» (test négatif). Dans cette optique, la sensibilité de l'outil reflète sa capacité à classer correctement les sujets malades, alors que sa spécificité reflète sa capacité à classer comme non-malades les sujets étant réellement des non-malades [25]. La valeur prédictive positive (PPV, pour *positive predictive value*) représente, quant à elle, la probabilité qu'un sujet classé comme malade soit réellement malade. Sa valeur complémentaire, la valeur prédictive négative (NPV, pour *negative predictive value*), représente la probabilité qu'un sujet classé comme non-malade soit réellement non-malade [26]. Tout dépendant de l'outil développé, différentes caractéristiques peuvent être recherchées. Par exemple, un outil qui cherche à identifier les sujets potentiellement atteints d'une maladie grave aura avantage à être plus sensible et à avoir une PPV élevée.

2.3.2 Les courbes ROC

L'analyse d'une courbe ROC permet d'évaluer spécifiquement la précision d'un outil de prédiction [25]. Concrètement, la courbe ROC correspond à $1 -$ la spécificité en axe des x , contre la sensibilité, en axe des y [25]. C'est en fait l'aire sous la courbe (AUC, pour *area under the curve*) de la courbe ROC, ou la statistique C, qui représente la précision de l'outil. Celle-ci est facile à interpréter et est propice à la comparaison; plus la statistique C est élevée, plus l'outil est précis [27]. À titre d'exemple, dans la figure 2, la ligne de couleur bleue représente une AUC de 0.5, ce qui correspond en fait au hasard. La courbe de couleur rouge pointillée correspond à une courbe ROC standard, avec une précision d'environ 85% (AUC \sim 0.85), tandis que la courbe de couleur verte, avec son AUC de 1.0, représente un outil parfait ou le *gold standard*. Typiquement, une courbe ROC se retrouvera entre celle du hasard et le *gold standard*, c'est-à-dire qu'il aura une précision pouvant varier entre 50 et 100% [27].

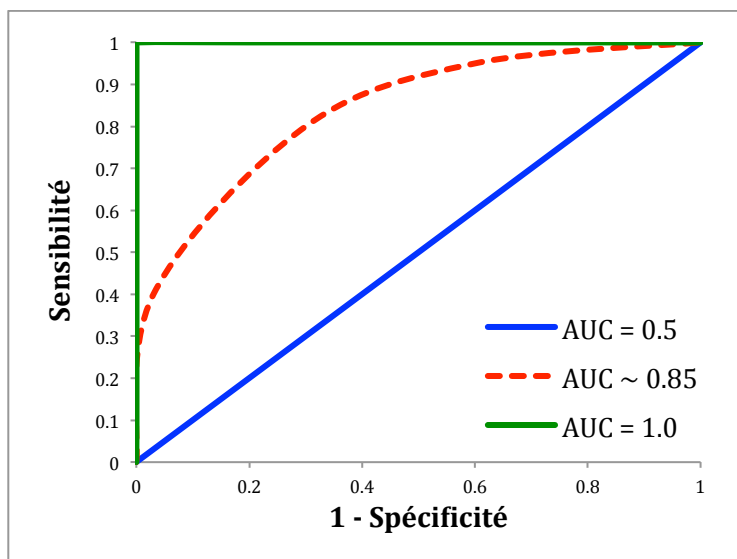


Figure 2 : Représentation visuelle de différentes courbes ROC

Dans le cadre de ce projet visant à prédire les apports en GLU chez une population d'athlètes, nous cherchions à obtenir un outil ayant une sensibilité et

une PPV élevées. En fait, comme l'objectif ultime du projet était d'identifier les athlètes ayant un apport suffisant en GLU, il était essentiel pour nous de s'assurer que l'outil final soit en mesure de correctement classer les sujets avec des apports suffisants. Pour ce qui est de la statistique C, il est évident que la plus haute valeur possible était recherchée, dans le but d'avoir l'outil le plus précis possible. Dans cet ordre d'idée, la décision quant au choix de l'outil final était basée sur une série de caractéristiques recherchées et non un seul trait désirable; nous cherchions à avoir l'outil le plus sensible et précis, sans toutefois être au détriment de la spécificité.

Rapport-Gratuit.com

3 Conclusion partielle

À la lumière des informations présentées, il est clair que l'article associé au présent projet de maîtrise saura contribuer de manière innovante à la recherche dans le domaine de la nutrition sportive. En effet, la majorité des manuscrits présentement publiés dans ce champ de recherche concernent les aides ergogéniques visant l'amélioration de la performance, comme la supplémentation en caféine ou en beta-alanine, ou encore les habitudes alimentaires des athlètes selon leur discipline sportive. Ce dernier type d'étude permet certes de dresser un portrait des lacunes présentes dans l'alimentation des sportifs, mais reste toutefois dans un cadre très théorique. Il est pertinent de savoir que les athlètes consomment, pour la majorité, trop peu de GLU, mais que faire avec cette information? Puisque nous savons que les ressources en nutrition sportive sont souvent limitées, il était pertinent pour nous de chercher à développer un nouvel outil permettant de cibler rapidement les individus ayant un apport suffisant en GLU et, par le fait même, identifier les athlètes ayant un risque d'apport sous-optimal en ce même nutriment, afin de les diriger vers un/une nutritionniste.

Dans un autre ordre d'idée, le développement d'outils de dépistage permettant de prédire un apport adéquat en un nutriment précis est encore peu documenté dans la littérature, que ce soit dans le monde du sport ou tout simplement dans le domaine de la nutrition. Ce projet a donc le potentiel d'ouvrir la voie au développement de plusieurs instruments du même genre. Au final, de tels outils permettraient d'augmenter l'efficacité de la pratique autant en nutrition sportive qu'en clinique. De plus, l'utilisation des arbres de classification comme méthode de développement d'outil de dépistage est actuellement très peu documentée parmi les manuscrits publiés en nutrition, encore moins en nutrition sportive. La publication de l'article associé au présent projet permettra donc de combler un manque dans la littérature en nutrition.

4 Objectifs et hypothèses

En considérant que les outils d'évaluation alimentaire présentement disponibles nécessitent un certain temps à administrer et analyser, que les ressources en nutrition sont limitées dans le monde du sport et que le but final de ce projet était d'identifier les athlètes ayant un apport adéquat en un seul nutriment, soit les GLU, nous avons établi les **objectifs** suivant :

- Développer un outil de dépistage d'une consommation adéquate en GLU, en utilisant les arbres de classification et/ou les modèles de régression logistique multiple.
- Valider l'outil développé auprès de la population cible, soit les athlètes d'endurance.

L'**hypothèse** émise au début de nos travaux était la suivante :

- Il est possible de développer un outil de dépistage valide qui sera rapide et facile à utiliser dans le domaine des sports d'endurance.

Le prochain chapitre présente, sous forme d'article scientifique, les différentes étapes de l'étude ainsi que les résultats encourus, accompagnés de leur discussion. L'article a été rédigé en langue anglaise et a été soumis à la revue *Journal of the International Society of Sports Nutrition* au mois de février 2018.

Chapitre 2. Étude portant sur le développement et la validation d'un outil de dépistage d'une consommation adéquate en glucides chez des athlètes d'endurance

DEVELOPMENT AND VALIDATION OF A DIETARY SCREENER FOR CARBOHYDRATE INTAKE IN ENDURANCE ATHLETES

Stéphanie Harrison, Élise Carbonneau, Denis Talbot, Simone Lemieux, Benoît Lamarche.

Affiliations :

Institute of Nutrition and Functional Foods (INAF), Laval University, Quebec, Canada G1V 0A6 (SH, ÉC, SL, BL)

School of Nutrition, Laval University, Quebec, Canada G1V 0A6 (SH, ÉC, SL, BL)

Department of Social and Preventive Medicine, Laval University, Quebec, Canada G1V 0A6 (DT)

CHU de Quebec Research Center, Laval University, Quebec, Canada G1V 0A6 (DT)

Résumé

Objectif : Plusieurs études ont démontré qu'une majorité d'athlètes d'endurance n'atteint pas la recommandation en glucides (GLU) de 6g/kg de masse corporelle, ce qui peut engendrer de nombreux impacts négatifs sur la récupération et la performance. L'objectif de cette étude était de développer et valider un outil de dépistage rapide et facile à utiliser qui permet d'identifier les athlètes d'endurance à risque de ne pas avoir une consommation suffisante en glucides.

Méthode : L'outil de dépistage a été développé par une approche de régression logistique multiple dans une cohorte de 1571 adultes non-athlètes (826 femmes et 745 hommes, âge moyen de 44.8 ± 14.2 ans). Les apports alimentaires de ces sujets avaient été préalablement déterminés par un questionnaire de fréquence alimentaire web (web-FFQ) validé. Trois modèles ont été développés à partir de la consommation auto-rapporté des aliments, utilisant les 5, 10 et 15 variables les plus significatives dans la prédiction de l'apport en GLU. Ces trois modèles ont ensuite été validés auprès d'un échantillon d'athlètes d'endurance non-élites ayant participé à des épreuves multisports ($n=175$, 64 femmes et 111 hommes, âge moyen de 37.1 ± 11.3 ans).

Résultats: Le modèle basé sur 15 variables était significativement plus performant pour prédire un apport adéquat en GLU chez les athlètes en endurance (statistique-c=0.94) que les modèles de 10 et 5 variables (statistique-c= 0.90 et 0.71, respectivement). Le modèle basé sur 15 variables prédit l'atteinte de la

recommandation en GLU avec une sensibilité de 89.5%, une spécificité de 87.3% et des valeurs prédictives positive et négative de 77.4% et 94.5%, respectivement.

Conclusion: L'outil de dépistage de la consommation en GLU permet d'identifier avec précision et rapidité les athlètes à risque de ne pas rencontrer les recommandations pour les sports d'endurance. L'utilisation de cet outil de dépistage a le potentiel de réduire la prévalence élevée de consommation sous-optimale en GLU dans le monde du sport d'endurance.

Mots clés: Glucides, outil d'évaluation alimentaire, athlètes d'endurance

Abstract

Background: Studies have shown that the majority of endurance athletes do not achieve the minimal recommended carbohydrate (CHO) intake of 6 g/kg of body weight, with potentially negative impacts on recovery and performance. The purpose of this study was to develop and validate a rapid and easy to use dietary screener to identify athletes who do and do not achieve the minimal recommendation for CHO intake for endurance sports during moderate-intense training periods.

Methods: The dietary screener was developed using multiple logistic regression modeling of data from a sample of 1571 non-athlete adults (826 women and 745 men, mean age 44.75 ± 14.2 years) among whom dietary intake was assessed using a validated web-based food frequency questionnaire (web-FFQ). Three models were developed based on whole food intake using the 5, 10 and 15 most significant variables predicting CHO intake. The three models were then validated in a target population of non-elite endurance athletes having taken part in multisport events ($n=175$, 64 women and 111 men, mean age 37.1 ± 11.3 years) and compared using sensitivity, specificity, positive and negative predictive values (PPV and NPV, respectively) and c-statistics.

Results: The 15-variables model provided significantly better accuracy in predicting CHO intake adequacy in non-elite endurance athletes (c-statistic=0.94) compared with the 10-and 5-variables model (c-statistic=0.90 and 0.71 respectively). The 15-variables model predicts CHO intake adequacy in the target population with a sensitivity of 89.5%, a specificity of 87.3% and PPV and NPV of 77.3% and 94.5%, respectively.

Conclusion: We have successfully developed a short and valid dietary screener that identifies endurance athletes at risk of not achieving adequate CHO intake.

Use of this rapid screener may help alleviate the highly prevalent issue of suboptimal CHO consumption in the endurance sports realm.

Keywords : Carbohydrates, dietary screener, endurance athletes

Background

Carbohydrates (CHOs) are a crucial component of an athlete's diet, especially in endurance sports. Dietary CHOs contribute to restoring muscle and liver glycogen between training sessions and increase performance when their availability is maintained during the effort [1, 2]. Recommendations on dietary CHOs are specific to sport, training regimen and competition schedule [1]. It is generally recommended that athletes consume between 6 and 10g of CHO per kg of body weight (BW) per day when involved in an endurance program comprising moderate-to-high intensity trainings [2].

Multiple studies have shown that a large proportion of endurance athletes do not meet these recommendations. For instance, 45% of non-elite men and women participating in endurance multisport events such as IRONMAN triathlons were below the targeted 6g CHO/kg of BW [3]. The prevalence of inadequate CHO intake was 80% among elite endurance athletes [4]. Studies have also shown that average CHO consumption among young pentathlon athletes and female collegiate athletes was below the recommendation [5, 6]. This is an important concern because inadequate CHO consumption may lead to decreased work rates, impaired skills and concentration and increased effort perception, all of which are partly caused by fatigue [1]. Weakening of the immune system and increased risks of complications due to over-training have also been associated with low CHO intake in endurance athletes [28].

The difficulty in rapidly measuring food and nutrient intakes on the field certainly represents one of the most significant barriers to more optimal management of diet among athletes. Dietary assessment is costly and time-consuming and this is particularly troublesome in environments that generally rely on limited resources towards nutrition support. Having access to a rapid and cost-effective screening tool that identifies athletes at risk of not achieving dietary CHO recommendations will prove to be extremely useful to assign the limited nutrition support resources to those who need it the most. We hypothesized that it is possible to develop such a

simple validated tool for use in the endurance sports realm. Therefore, the purpose of this study was to develop and validate an easy to use and rapid dietary screener to identify athletes who do and do not achieve the minimal CHO intake recommendation for endurance sports during training periods.

Participants and methods

Study participants

A database of adult non-athlete subjects from previous projects conducted at the Institute of Nutrition and Functional Foods (INAF) in Quebec City was used to develop the screener (DEV sample). Multiple projects, in which subjects were all healthy, were included in the database. All participants provided consent in written form to have their data included in a database for use in research other than the main project to which they participated. The validity of the screener in the targeted population was assessed in a sample of non-elite endurance athletes (VALID sample). These athletes competed in Ironman triathlons (IM), Ironman 70.3 triathlons (IM 70.3), winter pentathlon (tandem or solo category) (9-15 km of cycling, 3.6-5.5 km of running, 4.9-8 km of cross-country skiing, 5-8.4 km of ice-skating, and 3.4-5.1 km of snowshoeing) or winter triathlon (5 km of snowshoeing, 12 km of ice-skating, and 8 km of cross-country skiing). Non-elite athletes provided consent through an online system.

Dietary data collection

Participants in both the DEV and the VALID samples completed a validated web-based food frequency questionnaire (web-FFQ) [29]. This questionnaire contains 136 questions split into eight different sections: dairy products, fruits, vegetables, meat and alternatives, cereals and grain products, beverages, 'other foods' and dietary supplements. The web-FFQ inquires about food intake during the month prior to questionnaire completion. The Nutrition Data System for Research (software version 4.03, Food and Nutrient Database 31, Minneapolis, MN, USA) (Schakel et al., 1988) and the Canadian Nutrient File (CNF, version 2007b, Ottawa, ON, Canada) (Health Canada, 2007) were used to obtain nutrient intakes based on the answers provided in the web-FFQ. Food items from a broad category were grouped as one variable to simplify application and use of the final screener. For example, answers pertaining to brown and white rice consumption were added together to form only one category (frequency and amount of rice consumption).

Model development

Logistic regression modeling was used to develop the CHO-specific screener using data from the DEV sample. Analyses were undertaken in SAS (University Edition) unless stated otherwise. *P-values* less than 0.05 were considered statistically significant. Food servings per day, derived from the frequency and portion size of food categories in the web-FFQ, and participants' sex were considered in the primary phases of model development. Intake of individual nutrients, such as fat or protein, was not considered because athletes generally do not know how much of any specific nutrients they are consuming.

Spearman's correlations between food intake (in servings/d) and CHO consumption (in g/kg of BW) were first calculated. The 25 foods showing the strongest univariate correlation with CHO consumption were retained. Then, for each of these 25 foods, cut-off points that best correlated with CHO intake above or below 6g/kg BW were identified using logistic regression in R (version 3.3.0), in order to create dichotomic variables that can be answered simply by yes/no. These cut-off points were further adjusted to best reflect plausible daily or weekly servings. For example, the cut-off point for bread that best predicted a CHO intake > 6g/kg of BW in univariate logistic regression was 2.67 servings/d. This value was rounded up to 3 servings per day in order to facilitate the answering of the question by athletes. Next, multiple stepwise logistic regression models were constructed in SAS based on the 5, 10 and 15 variables that best predicted a CHO consumption > 6g /kg of BW. Contingency tables with derived sensitivity, specificity, positive and negative predictive values (PPV and NPV, respectively) as well as receiver operating characteristic (ROC) curves and derived c-statistic were used to compare the performance of the three models. A model based on 20 variables or more was considered, but data indicated that model performance was no longer increased beyond 15 variables (not shown).

Model validation

Model validation was ascertained in the VALID sample of non-elite endurance athletes. The 5, 10 and 15-variables models derived from the development phase

were compared for performance again using statistics from contingency tables as well as ROC curves.

Results

Participants characteristics

The DEV sample included 1571 participants (826 women and 745 men). Mean age was 44.8 years (SD = 14.3), mean body mass index (BMI) was 28.1 kg/m² (SD = 5.8) and mean CHO consumption was 3.75 g/kg of BW (SD = 1.5). Only 7.2% of participants in this sample had a CHO consumption > 6 g/kg of BW (Table 1).

The VALID cohort included 175 athletes (64 women and 111 men). Mean age was 37.1 years (SD = 11.3), mean BMI was 23.3 kg/m² (SD = 2.6) and mean CHO consumption was 5.4 g/kg BW (SD = 2.5). A total of 32.6% consumed more than 6g of CHO/kg of BW (Table 2).

Model development

The 15-variables model showed the highest c-statistic (0.89, $p < 0.004$ vs other models, Figure 1) with a sensitivity of 73.5%, a specificity of 86.7%, a NPV of 97.7% and a PPV of 30.0% (Table 3).

Model validation

Table 4 shows the characteristics of the 5, 10 and 15-variables models when applied to the target population of endurance athletes (VALID cohort), using the predetermined cut-offs for each food in the model. Consistent with data from the DEV sample, the 15-variables model performed significantly better than the 5 and 10-variables models in predicting CHO intake (Figure 2). The 15-variables model identified athletes achieving the minimal CHO recommendation (>6g/kg of BW) with a sensitivity of 89.5% and a specificity of 87.3%. NPV and PPV were 94.5% and 77.3%, respectively. Table 5 presents the final screener based on the 15-variables model, presenting each food retained in the model with its corresponding cut-off (formulated as a question) and their associated multivariate β derived from the multivariate logistic model. These β are used to define the predictive model that will be deployed to predict one's risk of not achieving the CHO recommendations for endurance sports.

Discussion

Rapid and cost-efficient assessment of CHO consumption among endurance athletes is challenging on the field. Although multiple dietary assessment tools, such as FFQs, 24hr recalls and dietary journals, are available to calculate an athlete's CHO intake, these tools usually take a lot of time to complete and require the experience of a trained professional for analysis. This, combined with the fact that large proportions of endurance athletes do not meet the recommended CHO intake, is a concerning issue. Here, we have developed a CHO-specific dietary screener that allows rapid detection of endurance athletes at risk of not achieving a CHO intake of 6g/kg of BW or more. To our knowledge, this is the first validated tool that screens for adequate CHO intake among athletes.

The final model upon which the screener is based has both a high sensitivity and specificity in the target population (89.5% and 87.3%, respectively), which are desired traits [25]. Such statistics indicate that the screener is as accurate in adequately identifying athletes who meet and those who do not meet the recommendation for CHO intake. The high AUC of the ROC curve (or c-statistic) yielded by the 15-variables model (0.94 on a range from 0.5 to 1.0) is also reflective of a dietary screener that has excellent accuracy[25]. Furthermore, the model's NPV was considerably higher than its PPV (94.5% vs. 77.3%), indicating that the screener is slightly more accurate in identifying endurance athletes who do not achieve adequate CHO intake than those who do. Such characteristic is highly desirable in the context of this research, as the ultimate goal of the CHO screener is to target athletes who would benefit from nutritional counseling, i.e. those with inadequate CHO intakes.

Very few studies have used an approach similar to ours to develop predictive models of adequate/inadequate dietary intakes, which makes comparison difficult. Most attempts were undertaken with a health rather than sports perspective. In those previous studies, predictive models and tools often achieved either a high sensitivity or a high specificity, but rarely both. For instance, Cook *et al.* built single-

question and five-question screeners to rapidly assess fruits and vegetables intake among non-athletes. Sensitivity values ranged from 35.7% to 45.5% while specificity values ranged from 81.8% to 84.9% among all five-question screeners developed by the research team. Using a single-question approach yielded high sensitivity but low specificity in the same population [30]. In most of these studies, the AUC of the ROC curves were fairly low, indicative of poor accuracy.

In an attempt to develop the simplest and yet most accurate CHO screener possible, we gave important considerations to limitations specific to the sports work environment. First, we had access to numerous dietary variables for the development of the model, such as energy, vitamin and protein intake, which may have contributed to a better prediction accuracy. However, such information is not readily available to either the respondent or the resource responsible for the screening test. It was therefore decided *a priori* to exclude such information. All anthropometric measures were also *a priori* excluded as they are too-closely related to the outcome measure to predict, which is based on BW. Similarly, cut-off values for each predictive food in the model were rounded to full daily or weekly servings to facilitate screener administration.

Several methods can be used to develop the predictive model of an outcome. Here, a multifaceted approach was used, but ultimately a stepwise logistic regression modeling approach yielded the final model. A classification tree (CT) approach was also considered to develop the screener. This method uses discriminant analysis to test various combinations of variables in order to maximize the CT's predictive power [14]. Different algorithms can be used to build CTs ; the CART algorithm was the chosen method for our purpose. What characterizes the CART algorithm is that it builds a very large CT and then prunes it to a smaller size to minimize classification errors [14]. A 10-fold cross-validation is used to prune the initial CT. The use of this method would have been beneficial for this particular research since ideal cut-off points are calculated directly in the CT algorithm.

Unfortunately, this method yielded underwhelming results, with unacceptably high values for false negatives (approximately 30%) when applied to the athletes sample (the VALID cohort). We hypothesize that the sample of non-athletes used to develop the CT may have comprised too few individuals with a CHO intake > 6g/kg of BW, thereby reducing the data usable by the algorithm to maximize the CT's predictive power.

Although this is the first study to develop a CHO-specific dietary screener for endurance athletes, limitations should be noted. First and foremost, the sample used to build the screener for application among athletes comprised non-athletes. This may have been a very significant shortcoming, considering that the diets of non-athletes and of endurance athletes are quite different. Second, a small proportion of the sample of individuals used to develop the screener achieved an intake of CHO greater than 6g/kg of BW, which may have hindered our ability to accurately predict this nutritional outcome. Ideally, the development of this CHO-specific screener would have been based on data from a large cohort of endurance athletes, but this was not possible. Nevertheless, the accuracy and hence external validity of the CHO-specific screener is considered to be excellent, despite these limitations. Lastly, exploring different approaches for model development is a strength considering that very few studies in the field of nutrition have used CTs to create predictive models.

Conclusion

This study succeeded in developing a simple, 15-questions dietary screening tool that predicts with accuracy the risk for an athlete of not achieving adequate CHO intake for endurance sports. Since the screener was validated in non-elite endurance athletes, further research should be conducted to test the accuracy of the screening tool in elite athletes. This easy-to-use screening tool will be a great asset to field work in sports nutrition as it rapidly identifies athletes who may benefit the most from receiving dietary counseling to optimize their diet.

List of abbreviations

AUC: area under the curve

BW: body weight

CHO: carbohydrate

CNF: Canadian nutrient file

CT: classification tree

IM: IRONMAN triathlon

IM70.3: IRONMAN 70.3 triathlon

INAF: Institute of Nutrition and Functional Foods

NPV: negative predictive value

PPV: positive predictive value

ROC: Receiver operating characteristic

Web-FFQ: web-based food-frequency questionnaire

References

1. Burke LM, Hawley JA, Wong SH, et al.: Carbohydrates for training and competition. *J Sports Sci* 2011, 29 Suppl 1:S17-27.
2. Thomas DT, Erdman KA, and Burke LM: Position of the academy of nutrition and dietetics, dietitians of canada, and the american college of sports medicine: Nutrition and athletic performance. *J Acad Nutr Diet* 2016, 116:501-28.
3. Masson G and Lamarche B: Many non-elite multisport endurance athletes do not meet sports nutrition recommendations for carbohydrates. *Appl Physiol Nutr Metab* 2016, 41:728-34.
4. Baranauskas M, Stukas R, Tubelis L, et al.: Nutritional habits among high-performance endurance athletes. *Medicina (Kaunas)* 2015, 51:351-62.
5. Shriver LH, Betts NM, and Wollenberg G: Dietary intakes and eating habits of college athletes: Are female college athletes following the current sports nutrition standards? *J Am Coll Health* 2013, 61:10-6.
6. Coutinho LA, Porto CP, and Pierucci AP: Critical evaluation of food intake and energy balance in young modern pentathlon athletes: A cross-sectional study. *J Int Soc Sports Nutr* 2016, 13:15.
7. Gunzer W, Konrad M, and Pail E: Exercise-induced immunodepression in endurance athletes and nutritional intervention with carbohydrate, protein and fat-what is possible, what is not? *Nutrients* 2012, 4:1187-212.
8. Labonte ME, Cyr A, Baril-Gravel L, et al.: Validity and reproducibility of a web-based, self-administered food frequency questionnaire. *Eur J Clin Nutr* 2012, 66:166-73.
9. Zou KH, O'Malley AJ, and Mauri L: Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. *Circulation* 2007, 115:654-7.
10. Cook A, Roberts K, O'Leary F, et al.: Comparison of single questions and brief questionnaire with longer validated food frequency questionnaire to assess adequate fruit and vegetable intake. *Nutrition* 2015, 31:941-7.
11. Loh W-Y: Classification and regression trees. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 2011, 1:14-23.

Table 1 Characteristics of subjects in the DEV sample (n=1571)

Women, %	52.6%
Age, y	44.8 ± 14.3 ¹
Body weight, kg	79.9 ± 18.9
BMI, kg/m ²	28.1 ± 5.8
Carbohydrates consumption, g/kg of body weight	3.75 ± 1.5
Subjects consuming >6g CHO/kg of body weight, (%)	7.2%

¹ Mean ± SD (all such values) unless stated otherwise.

Table 2 Characteristics of subjects in the VALID sample (n=175)

Women, %	36.6%
Age, y	37.1 ± 11.3
Body weight, kg	69.1 ± 11.1
BMI ² , kg/m ²	23.3 ± 2.6
Carbohydrates consumption, g/kg of body weight	5.4 ± 2.5
Subjects consuming >6g CHO/kg of body weight, %	32.6%

¹ Mean ± SD (all such values) unless stated otherwise.

² n=147 because of 28 missing height values

Table 3 Characteristics of the multiple logistic regression models in the DEV sample

Model	Sensitivity ¹	Specificity	False positives	False negatives	PPV	NPV	c statistic
5 variables	63.7	83.8	70.0	2.3	23.4	96.8	0.78
10 variables	64.6	87.7	71.0	3.0	29.0	97.0	0.85
15 variables	73.5	86.7	70.0	2.3	30.0	97.7	0.89

¹ % (all such values)

PPV : Positive predictive value

NPV : Negative predictive value

Table 4 Characteristics of the multiple logistic regression models in VALID sampler

model	Sensitivity ¹	Specificity	False positives	False negatives	PPV	NPV	c statistic
5 variables	52.6	82.2	12.0	15.4	58.8	78.2	0.71
10 variables	75.4	86.4	9.1	8.0	72.9	87.9	0.90
15 variables	89.5	87.3	8.6	3.4	77.3	94.5	0.94

¹ % (all such values)

PPV : Positive predictive value

NPV : Negative predictive value

Table 5 Final dietary screener

Questions of the final dietary screener ¹	β^2
Do you consume melons (watermelon, honeydew or cantaloup) on a daily basis?	0.5287
Do you consume pancakes twice a week?	1.9666
Do you consume avocado twice a week?	-0.0433
Do you consume cereal bars 6 times a week?	2.0899
Do you consume rice 5 times a week?	2.0401
Do you drink chocolate milk 5 times a week?	2.3249
Do you consume chocolate (white, milk or dark) every week?	1.8776
Do you consume corn on a daily basis?	0.7994
Do you consume milk, soy milk or silk tofu based desserts 3 times a week?	5.3373
Do you consume cold breakfast cereals on a daily basis?	3.0771
Do you consume pasta on a daily basis?	0.8276
Do you consume jam, maple by-products, hazelnut spread, jelly or chocolate syrup twice a day?	2.3477
Do you consume salad, lettuce or spinach twice a day?	2.7638
Do you drink soft drinks 3 times a day?	10.3662
Are you a woman ?	0.3734

¹ Final questions are based on optimal cut-off points calculated by R (version 3.3.0) that were further adjusted to best fit a daily or weekly number of servings. Cut-off points represent the number of servings of each specific food that best predicted a CHO consumption >6g/kg of BW.

² β from the multivariate logistic regression model for each dichotomic variable (yes/no) in the final dietary screener. All β are significant ($P < 0.05$).

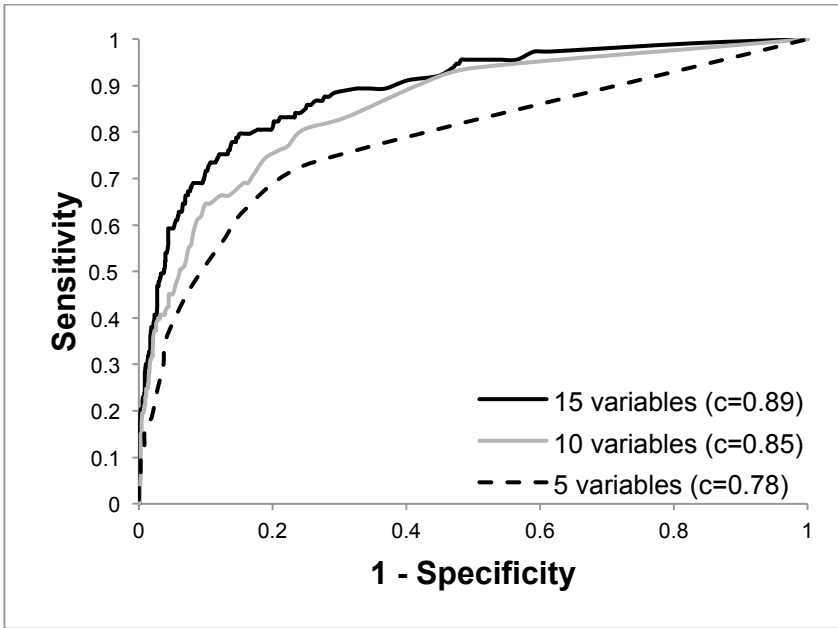


Figure 1 ROC curves comparison of multiple logistic regression models in DEV sample. (c represents the c statistic on a scale of 0.5 to 1.0).

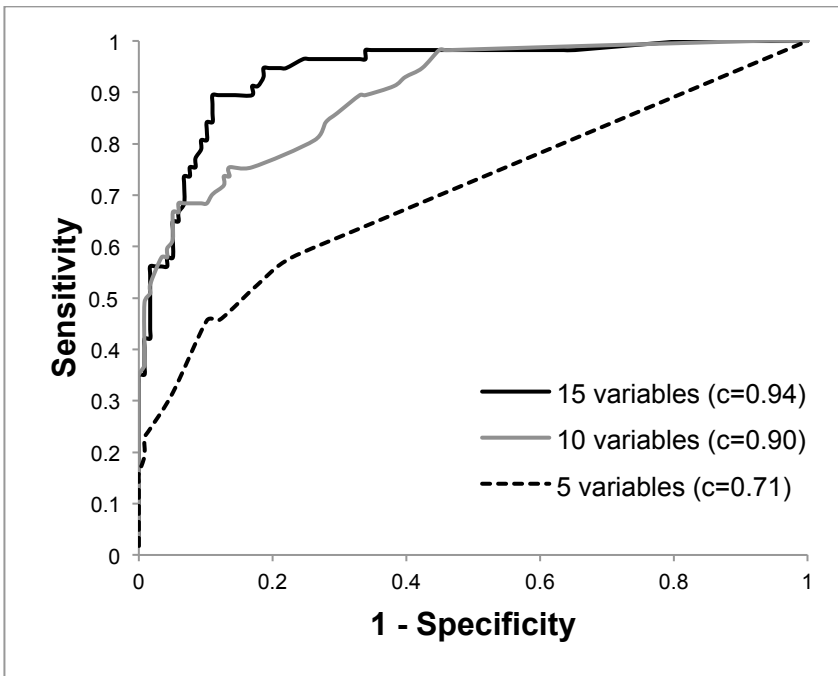


Figure 2 ROC curves comparison of multiple logistic regression models in the VALID sample. (c represents the c statistic on a scale of 0.5 to 1.0).

Conclusion générale

Le présent projet de maîtrise avait pour objectifs principaux de développer un outil de dépistage permettant d'identifier de manière simple et rapide les athlètes ayant un apport suffisant en GLU selon les recommandations minimales applicables aux sportifs impliqués dans des disciplines d'endurance et de le valider dans cette même population. Une cohorte différente a été utilisée pour la réalisation de chaque objectif. Des hommes et des femmes non-athlètes composaient la cohorte utilisée pour le développement de l'outil, alors que celle utilisée pour la validation comprenait 175 athlètes, hommes et femmes. Les résultats obtenus ont confirmé l'hypothèse émise initialement, soit qu'il est possible de prédire les apports en GLU à l'aide d'un outil de dépistage dans une population d'athlètes d'endurance. Les résultats présentés montrent que l'outil final développé, soit celui de 15 questions courtes de type oui/non, est non seulement rapide à administrer, mais extrêmement précis; l'outil de dépistage, avec sa précision de 94%, prend environ de 2 à 3 minutes à administrer à un seul sujet. Il a été choisi puisqu'il prédisait avec la plus grande précision l'apport en GLU des athlètes. Celui-ci pouvait être qualifié de suffisant suite à l'administration complète de l'outil; il est impossible, avec ce genre d'outil prédictif, de quantifier le nombre de questions à répondre à l'affirmative pour avoir un apport qualifié comme suffisant, puisque chaque aliment est associé à une probabilité différente d'avoir un apport en GLU plus grand que 6g/kg MC.

Il est important de rappeler que peu d'études dans le domaine de la nutrition ont fait appel aux arbres de décision. Dans le cadre de ce projet, ils ont été choisis comme première approche pour leur facilité d'interprétation et leur capacité à prédire un dénouement selon plusieurs facteurs distincts. Le fait de pouvoir contrôler de multiples paramètres de l'arbre à même un seul algorithme était également un attrait de choix pour nous. En effet, nous voulions d'abord et avant tout que notre outil soit rapide et efficace et les arbres de classification nous permettaient de limiter le nombre de questions dans l'outil final, sans toutefois

compromettre sa précision. De plus, les arbres engendrent des questions de type oui/non, ce qui est un facteur facilitant dans l'administration rapide de l'outil final. Cependant, comme nous n'avions aucun comparatif quant à l'utilisation de cette approche pour un objectif similaire au nôtre, ce choix représentait un risque, mais aussi une potentielle avancée pour la recherche en nutrition. Ultimement, le dénouement que nous cherchions à prédire, soit un apport suffisant en GLU, combiné à la base de données utilisée pour le faire, ne s'apparentaient pas aux arbres de décision. En fait, nous pensons que l'utilisation de la base de données de non-athlètes, qui contenait trop peu de sujets ayant un apport en GLU correspondant aux recommandations pour sportifs, a nuit à la réussite du développement de notre outil par arbre de classification. Selon nos résultats, il est impossible de dire si la présente approche est inadéquate pour prédire l'apport en un nutriment précis, mais il semblerait que la base de données utilisée doive contenir une plus grande proportion de sujets atteignant le dénouement étudié pour que l'outil développé soit performant et précis. Néanmoins, il est intéressant de noter qu'une autre étude présentement en cours à l'INAF fait usage des arbres de classification et, jusqu'à maintenant, ceux-ci semblent se prêter aux besoins du projet et offre des résultats prometteurs. Ceci étant dit, nous nous devons de trouver une alternative à cette méthode initialement ciblée pour réaliser nos objectifs et les modèles de régression logistique, de par leur aspect prédictif, ont semblé la meilleure option selon nos besoins[19].

Plusieurs limites de la présente étude étaient, au final, hors de notre contrôle. En effet, l'utilisation d'une cohorte non-athlète pour le développement de l'outil était certes sous-optimale, mais le recrutement d'un nombre assez élevé d'athlètes d'endurance dans un contexte de recherche représente un défi de taille qui aurait été irréaliste dans le contexte d'un projet de maîtrise. Considérant ceci, il était intéressant pour nous de se tourner vers une base de données rassemblant plusieurs anciens projets de notre centre de recherche puisque le nombre élevé de sujet dans cette cohorte ($n=1571$) nous permettait d'obtenir un meilleur pouvoir prédictif [31]. De surcroît, peu d'études se sont tournées vers les différences

présentes entre l'alimentation des athlètes et celle des non-athlètes. Néanmoins, une étude réalisée auprès de jeunes italiennes a noté comme seule différence entre ces deux groupes un pourcentage des apports quotidiens plus élevé en GLU et plus faible en lipides chez les athlètes, une différence caractérisée par une consommation plus élevée en sucres raffinés chez ces derniers [32]. Sedek et Yih ont observé, quant à eux, un score des habitudes alimentaires plus élevé chez des adolescents malaisiens non-athlètes, lorsque comparés à leur confrères athlètes [33]. Ces résultats, obtenus chez des adolescents, nous empêchent de déterminer à quel point l'alimentation des athlètes d'âge adulte diffère de celle des non-athlètes. Par le fait même, il est difficile de savoir l'effet que cette différence aurait pu avoir sur nos résultats. Malgré tout, il est logique de penser qu'un outil de dépistage comme le nôtre, mais développé dans une cohorte d'athlètes, inclurait des aliments plus spécifiques à ceux-ci, comme les gels ou bonbons pour sportifs. Nous considérons d'ailleurs cet élément comme une limite de l'étude réalisée. Il était toutefois impossible pour nous d'inclure de tels aliments dans nos analyses puisqu'ils n'étaient pas représentés dans le web-FFQ utilisé pour la compilation des données nutritionnelles de notre cohorte.

Dans un autre ordre d'idée, il serait pertinent de valider le présent outil dans une cohorte d'athlètes d'endurance d'élite, puisqu'il a été élaboré précisément pour cette population, car elle présente un grand nombre d'athlètes ayant des apports en GLU sous-optimaux. Cependant, la cohorte de validation utilisée dans le présent projet comprenait majoritairement des sportifs participant à des épreuves d'ultra-endurance extrêmement exigeantes (*IRONMAN* et *demi-IRONMAN*, entre autres), nous croyons donc que leurs besoins glucidiques se rapprochent de ceux d'athlètes d'élite. En ce sens, nous jugeons que la précision de notre outil sera très peu affectée lorsqu'il sera validé auprès d'athlètes d'endurance d'élite. Au contraire, Il serait intéressant de voir si notre outil, développé pour des sports d'endurance, serait aussi valide pour des athlètes de d'autres disciplines, comme

les sports de puissance.³ Aussi, puisque l'outil a été développé et validé dans une cohorte canadienne-française, il serait intéressant de le tester chez des sujets de différentes provinces canadiennes. Cette validation à plus grande échelle serait un atout si l'outil venait à être intégré dans la pratique en nutrition sportive au Canada. En fait, elle permettrait non seulement d'adapter l'outil à une population anglophone, mais aussi de s'assurer de sa validité auprès d'athlètes provenant de d'autres provinces, qui pourraient avoir des habitudes alimentaires légèrement différentes de celles des Québécois.

En 2016, Hull et ses collaborateurs se sont attardés à l'impact qu'ont les nutritionnistes du sport (NS) sur l'alimentation des athlètes [34]. L'étude, réalisée chez des athlètes au collégial, a identifié plusieurs différences entre l'alimentation des athlètes ayant accès à une nutritionniste et ceux n'en ayant pas. Entre autres, les collations post-entraînement des athlètes utilisant les services d'une NS étaient plus adéquates et ces athlètes avaient moins tendance à consommer des repas de restauration rapide avant/après un entraînement ou une compétition. Au final, les NS aidaient les athlètes à mieux appliquer les principes de base en nutrition à leur propre alimentation. Un constat similaire a été réalisé dans une équipe féminine de volleyball par Valliant *et coll.* (2012) [35]. Dans le cadre de cette étude, les athlètes avaient une rencontre en nutrition à un rythme mensuel, pendant une période de 4 mois, ce qui a engendré au final une différence significative quant à leur apport en énergie et en macronutriments et leur niveau de connaissance en nutrition. Parmi les macronutriments, on remarque un apport moyen en GLU ayant augmenté de manière significative suite à l'intervention. Ces résultats offrent une perspective intéressante à l'application future de notre outil. En effet, comme le but ultime de celui-ci est d'identifier les athlètes ayant un apport insuffisant en GLU dans le but de les diriger vers une NS qui pourra leur offrir des conseils personnalisés afin d'optimiser leur alimentation, il est intéressant de voir que l'intervention d'une nutritionniste a un impact concret sur l'alimentation des athlètes. Il faut toutefois

³ Par sport de puissance, nous entendons, par exemple, la course sur de moyennes distances, la natation, l'aviron, le canoë/kayak, etc.

garder en tête que les études présentées ont été réalisées spécifiquement auprès d'athlètes américains au collégial. Néanmoins, nous aspirons à présenter une version numérique de l'outil à des NS et des organisations sportives d'ici la fin de la présente année afin d'encourager son inclusion à la pratique courante en nutrition sportive. Par exemple, l'outil pourrait être présenté à des équipes du Rouge et Or ou, à plus grande échelle, à l'Institut National du Sport du Québec, qui offre un service de nutrition à des athlètes d'élite du Québec/Canada.

En terminant, l'outil développé se concentre spécifiquement sur l'apport en GLU chez des athlètes d'endurance. Toutefois, dans cette même population, la méthode employée pourrait être utilisée pour cibler d'autres nutriments pertinents à l'alimentation d'un athlète, comme les protéines ou le fer. À plus grande échelle, un outil similaire spécifique à un nutriment différent pourrait être développé dans d'autres populations.

Bibliographie

1. Thomas, D.T., K.A. Erdman, and L.M. Burke, *Position of the Academy of Nutrition and Dietetics, Dietitians of Canada, and the American College of Sports Medicine: Nutrition and Athletic Performance*. J Acad Nutr Diet, 2016. **116**(3): p. 501-28.
2. Burke, L.M., et al., *Carbohydrates for training and competition*. J Sports Sci, 2011. **29 Suppl 1**: p. S17-27.
3. Masson, G. and B. Lamarche, *Many non-elite multisport endurance athletes do not meet sports nutrition recommendations for carbohydrates*. Appl Physiol Nutr Metab, 2016. **41**(7): p. 728-34.
4. Baranauskas, M., et al., *Nutritional habits among high-performance endurance athletes*. Medicina (Kaunas), 2015. **51**(6): p. 351-62.
5. Coutinho, L.A., C.P. Porto, and A.P. Pierucci, *Critical evaluation of food intake and energy balance in young modern pentathlon athletes: a cross-sectional study*. J Int Soc Sports Nutr, 2016. **13**: p. 15.
6. Shriver, L.H., N.M. Betts, and G. Wollenberg, *Dietary intakes and eating habits of college athletes: are female college athletes following the current sports nutrition standards?* J Am Coll Health, 2013. **61**(1): p. 10-6.
7. Hawley, J.A., et al., *Nutritional modulation of training-induced skeletal muscle adaptations*. J Appl Physiol (1985), 2011. **110**(3): p. 834-45.
8. Torres-McGehee, T.M., et al., *Sports nutrition knowledge among collegiate athletes, coaches, athletic trainers, and strength and conditioning specialists*. J Athl Train, 2012. **47**(2): p. 205-11.
9. Couture, S., et al., *Evaluation of Sports Nutrition Knowledge and Recommendations Among High School Coaches*. Int J Sport Nutr Exerc Metab, 2015. **25**(4): p. 326-34.
10. Thompson, F.E. and A.F. Subar, *Chapter 1 - Dietary Assessment Methodology A2 - Coulston, Ann M, in Nutrition in the Prevention and Treatment of Disease (Third Edition)*, C.J. Boushey and M.G. Ferruzzi, Editors. 2013, Academic Press. p. 5-46.
11. Salvador Castell, G., L. Serra-Majem, and L. Ribas-Barba, *What and how much do we eat? 24-hour dietary recall method*. Nutr Hosp, 2015. **31 Suppl 3**: p. 46-8.
12. Perez Rodrigo, C., et al., *Food frequency questionnaires*. Nutr Hosp, 2015. **31 Suppl 3**: p. 49-56.
13. Perez Rodrigo, C., et al., *Screeners and brief assessment methods*. Nutr Hosp, 2015. **31 Suppl 3**: p. 91-8.
14. Loh, W.-Y., *Classification and regression trees*. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2011. **1**(1): p. 14-23.
15. Song, Y.Y. and Y. Lu, *Decision tree methods: applications for classification and prediction*. Shanghai Arch Psychiatry, 2015. **27**(2): p. 130-5.
16. Miller, B., et al., *Use of CHAID decision trees to formulate pathways for the early detection of metabolic syndrome in young adults*. Comput Math Methods Med, 2014. **2014**: p. 242717.
17. Speybroeck, N., *Classification and regression trees*. Int J Public Health, 2012. **57**(1): p. 243-6.

18. Batterham, P.J., H. Christensen, and A.J. Mackinnon, *Modifiable risk factors predicting major depressive disorder at four year follow-up: a decision tree approach*. BMC Psychiatry, 2009. **9**: p. 75.
19. Dominguez-Almendros, S., N. Benitez-Parejo, and A.R. Gonzalez-Ramirez, *Logistic regression models*. Allergol Immunopathol (Madr), 2011. **39**(5): p. 295-305.
20. Sperandei, S., *Understanding logistic regression analysis*. Biochem Med (Zagreb), 2014. **24**(1): p. 12-8.
21. Campbell, M., *Multiple logistic regression models--what are they?* Midwifery, 2004. **20**(3): p. 236-9.
22. Wang, C.J., et al., *Development and evaluation of a simple and effective prediction approach for identifying those at high risk of dyslipidemia in rural adult residents*. PLoS One, 2012. **7**(8): p. e43834.
23. Je, Y., Y. Kim, and T. Park, *Development of a self-assessment score for metabolic syndrome risk in non-obese Korean adults*. Asia Pac J Clin Nutr, 2017. **26**(2): p. 220-226.
24. Moisen, G.G., *Classification and Regression Trees*. 2008(1): p. 582-588.
25. Zou, K.H., A.J. O'Malley, and L. Mauri, *Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models*. Circulation, 2007. **115**(5): p. 654-7.
26. Lalkhen, A.G. and A. McCluskey, *Clinical tests: sensitivity and specificity: Fig 1*. Continuing Education in Anaesthesia, Critical Care & Pain, 2008. **8**(6): p. 221-223.
27. Fawcett, T., *An introduction to ROC analysis*. Pattern Recognition Letters, 2006. **27**(8): p. 861-874.
28. Gunzer, W., M. Konrad, and E. Pail, *Exercise-induced immunodepression in endurance athletes and nutritional intervention with carbohydrate, protein and fat-what is possible, what is not?* Nutrients, 2012. **4**(9): p. 1187-212.
29. Labonte, M.E., et al., *Validity and reproducibility of a web-based, self-administered food frequency questionnaire*. Eur J Clin Nutr, 2012. **66**(2): p. 166-73.
30. Cook, A., et al., *Comparison of single questions and brief questionnaire with longer validated food frequency questionnaire to assess adequate fruit and vegetable intake*. Nutrition, 2015. **31**(7-8): p. 941-7.
31. Everitt, B.S. and A. Skrondal, *The Cambridge Dictionary of Statistics*. 2010: Cambridge University Press.
32. Cupisti, A., et al., *Nutrition knowledge and dietary composition in Italian adolescent female athletes and non-athletes*. Int J Sport Nutr Exerc Metab, 2002. **12**(2): p. 207-19.
33. Sedek, R. and T.Y. Yih, *Dietary Habits and Nutrition Knowledge among Athletes and Non-Athletes in National University of Malaysia (UKM)*. Pakistan Journal of Nutrition, 2014. **13**(12): p. 752-759.
34. Hull, M.V., et al., *Gender differences and access to a sports dietitian influence dietary habits of collegiate athletes*. J Int Soc Sports Nutr, 2016. **13**: p. 38.

35. Valliant, M.W., et al., *Nutrition education by a registered dietitian improves dietary intake and nutrition knowledge of a NCAA female volleyball team.* *Nutrients*, 2012. **4**(6): p. 506-16.