

Ozgen, Heval, et al. "Morphological features in children with autism spectrum disorders: a matched case-control study." *Journal of autism and developmental disorders* 41.1 (2011): 23-31.

COMMENTAIRE

Cet article a été publié dans la revue *Journal of Autism and Developmental Disorders*, publication qui a un impact factor de 3.6 en 2015.

Son titre est quelque peu surprenant dans la mesure où une approche anthropométrique semble peu pertinente dans le cadre d'un trouble mental. La justification d'une telle démarche, ainsi que la notion de corrélation, voire de lien de cause à effet, ne sont pas évidentes.

Les auteurs viennent majoritairement des Pays Bas (9/10), et huit d'entre eux travaillent à Utrecht. En dehors de leurs affiliations nous n'apprenons rien sur ces personnes, ni leurs fonctions, ni leurs titres, ni leurs contributions à l'étude, et il n'y a pas de section spécifique sur ce que chacun a pu apporter. La liste des publications de l'auteur principal nous apprend deux informations intéressantes. D'une part dans les trois ans avant la publication de cet article elle a publié deux articles sur les anomalies morphologiques et le cancer chez l'enfant avant de s'intéresser aux anomalies morphologiques chez l'enfant atteint d'hypothyroïdisme. D'autre part, cette étude a engendré un deuxième article publié en 2013 intitulé '*Predictive value of morphological features in patients with autism versus normal controls*'. Il semblerait que ces articles sont les derniers qu'elle ait publiés. Le titre du deuxième article nous renseigne peut-être sur l'objectif réel de l'étude qui est donc d'établir un lien corrélatif entre certaines anomalies morphologiques et l'autisme.

Ce lien est confirmé dans l'abstract très concis qui s'attarde surtout sur la conception de l'étude. L'abstract termine par une ouverture concernant l'hypothèse que comprendre les mécanismes génétiques de cette association pourrait aider à mieux comprendre l'autisme. On ne nous dit pas exactement comment, ni quel aspect de l'autisme serait concerné.

L'introduction s'ouvre sur une définition des TSA, ce qui semble étrange dans une revue portant sur l'autisme et où on s'attendrait donc que les lecteurs connaissent déjà la définition. Il est étonnant que les *reviewers* n'aient pas demandé sa révision. Il devient vite clair que les auteurs partent du principe que l'autisme est surtout dû à des facteurs génétiques, et que l'objectif ultime est d'avoir un outil de diagnostic.

Une phrase de cette introduction est particulièrement surprenante. En effet, les auteurs évoquent les preuves solides ('robust evidence') tout en rappelant de nombreuses défaillances méthodologiques à ces études. N'y a-t-il pas une incompatibilité? Ils relèvent par ailleurs les problèmes avec l'échelle de Waldrop qui ne comporte que 18 items et qui a un niveau de sensibilité trop bas, produisant ainsi trop d'erreurs de type II (faux négatifs). Ils proposent plutôt d'employer un outil comportant pas moins de 683 items. Au nom d'une recherche d'exhaustivité on semble passer d'un extrême à l'autre. Il est donc possible que leur outil gagne en sensibilité mais perd en spécificité.

La section des méthodes est construite de façon tout à fait classique en 3 parties ; participants, recueil des données puis traitement de celles-ci. On y trouve les critères d'inclusion et d'exclusion, et il est intéressant de noter que les auteurs ne donnent pas d'explication concernant l'exclusion des personnes présentant une déficience mentale, ce qui n'est pas une évidence pour le lecteur étant donné que l'étude

porte avant tout sur des mesures physiques. Pas moins de trois références sont fournies pour expliquer le choix de la zone géographique d'Haarlem comme source des contrôles. Les procédures de consentement et d'approbation éthiques ont été respectées.

L'examen physique des participants a été pratiqué par une personne expérimentée, mais nous ne savons pas exactement la nature de ses fonctions ou de ses qualifications, les auteurs se contenant d'employer une abréviation sans explication (HMO) qui a un sens bien spécifique aux Etats Unis, mais qui n'est pas utilisé en Europe, lieu de l'étude. L'inclusion des mesures 'qualitatives' à côté des mesures quantitatives est pour le moins surprenant. Par ailleurs, après avoir plus ou moins discrédité l'échelle de Waldrop, les auteurs nous assurent que les items de cette échelle figuraient bien parmi les mesures relevées. Certaines mesures n'ont pas été prises, mais les auteurs ne donnent pas les raisons, à moins que ce ne soit tout simplement que les seuls équipements utilisés étaient une règle et un mètre. L'emploi de l'IMC prend bien en compte un ajustement d'interprétation pour l'âge des participants.

Pour nous rassurer, les auteurs évoquent le recours à une deuxième, et puis une troisième personne pour vérifier les mesures. Si on nous précise que l'une est pédiatre, aucune explication n'est donnée des abréviations (FAB et RCH). Ils ont en revanche éliminé un biais possible, en s'assurant que le statut diagnostique des patients était inconnu de ces personnes.

L'analyse statistique est très rigoureux, d'autant plus qu'avec un nombre de mesures aussi élevé le risque de faux positifs (erreurs de type 1) est très élevé. Pour pallier ce problème le recours à un test *post hoc*, la méthode Holm-Bonferroni, permet de réduire ces erreurs.

En ce qui concerne les résultats, une section qui est relativement courte, le lecteur se pose un certain nombre de questions. Tout d'abord concernant les participants on apprend que l'âge moyen se situe, après élimination d'environ la moitié selon certains critères, à 10,6 ans. Sachant que l'objectif des auteurs est de se servir des résultats pour poser le diagnostic d'un TSA, et la logique voudrait que ce soit pour des diagnostics plus précoces, on peut légitimement remettre en question son utilité réelle, d'autant plus qu'il n'y a pas de données antérieures pour les patients provenant de leurs dossiers. Seule une étude longitudinale commençant à l'âge où on souhaiterait se servir des mesures comme indicateur pourrait vraiment démontrer l'utilité en tant qu'outil de diagnostic.

Par ailleurs on apprend qu'il y a eu 31 exclusions de cas qui avait un diagnostic, mais qui ont été exclus par les évaluateurs qui considéraient que ces personnes n'avait pas vraiment de TSA. Représentant 7% de l'ensemble, cela peut être interprété de deux façons : soit leurs confrères sont incompetents, soit les TSA sont évolutifs et peuvent être 'guéris', chose qui est exclu dans leur article de 2013 qui est également basé sur cette étude. Ces exclusions ne sont pas commentées par les auteurs, ni dans cette section, ni dans la discussion.

Il est difficile d'évaluer le nombre réel de mesures utilisées. En effet, par endroits on évoque 683 mesures et dans le tableau de résultats il y a finalement moins de mesures que dans l'échelle de Waldrop. En outre certaines de ces mesures pourraient être imputées à autre chose, comme par exemple l'IMC et la croissance influencés par des troubles alimentaires. On n'évoque nullement les mesures qualitatives dont il est question dans la section 'méthodes'.

Si le tableau 1 et la figure 1 sont tous les deux compréhensibles, le tableau 2 est plus difficile à lire, et le choix des auteurs de trier par sévérité de l'anomalie ne facilite pas la tâche. Un tri par OR aurait peut-être facilité la lecture. Dans ce tableau on voit des différences statistiquement significatives, mais la

fréquence des différentes anomalies n'est peut-être pas assez élevée chez les sujets TSA pour en faire un outil de diagnostique.

Dans la discussion les auteurs reviennent sur les chiffres qu'ils qualifient de 'strikingly high', et soulignent que cela confirme ce que l'on trouve dans la littérature, notamment dans une méta-analyse réalisée par les auteurs eux-mêmes! Par la suite ils listent les points fort de cette étude, dont sa puissance statistique. Quatre observations que les auteurs considèrent intéressantes sont ensuite énumérées; une fréquence élevée d'anomalies, un taux plus élevé chez les garçons, une variation non-systématique (valeurs plus ou moins élevées que les normes) et des résultats similaires à une étude sur la schizophrénie.

Les limites de l'étude sont nombreuses, et les auteurs tentent d'en faire un recensement complet. Ces limites font apparaître des perspectives possibles pour des études plus poussées ainsi que pour les implications cliniques, en termes de classification des TSA en fonction des anomalies constatées, mais aussi en termes de dépistage précoce.

Malheureusement, les auteurs ne reviennent pas sur l'âge moyen des participants qui est plutôt élevé par rapport à l'âge habituel de diagnostic des TSA, et qui reste donc, pour moi, le frein principal à son utilité clinique.