



The Unit Commensurability Problem in the Nutri-Score Algorithm: a Construct Validity Analysis

José A. Martínez^{1*}

¹ *Department of Business Management. Technical University of Cartagena - Member of European University of Technology. Cartagena, Spain.*

Abstract

Background: The Nutri-Score labelling system, based on the Food Standards Agency modified Nutrient Profiling System (FSAm-NPS) algorithm, assigns penalty points to four food components — energy, total sugars, saturated fatty acids (SFAs), and sodium — calculated per 100 g. Prospective cohort studies have validated the FSAm-NPS dietary index as a mortality predictor.

Aims: To examine whether the per-100-g reference unit of the FSAm-NPS algorithm is commensurable with the epidemiologically relevant exposure unit in dose-response nutrition research, and to evaluate implications for construct validity.

Methods: Analytical review of the FSAm-NPS scoring architecture, Messick's construct validity framework, and the exposure units reported in dose-response meta-analyses for the four penalised components.

Results: Dose-response evidence for all four FSAm-NPS components is expressed in daily intake units (mg/day or g/day), not per 100 g of food. Converting these estimates to per-100-g units requires the assumption of a fixed portion size, rendering component weights arbitrary from the perspective of individual health risk. Prospective cohort validations do not resolve this problem: they validate a composite dietary quality index, not individual component weights.

Conclusion: The FSAm-NPS algorithm contains an unresolved construct validity gap arising from unit incommensurability. Future iterations should either derive component weights from portion-corrected dose-response functions or explicitly declare these weights as normative conventions subject to periodic revision.

Keywords: Construct Validity, Dose-response Relationship, Food Labelling, Nutrient Profiling, Nutri-Score.

INTRODUCTION

Nutri-Score is the front-of-pack nutritional labelling system with the widest regulatory uptake in Europe, currently endorsed or in use in France, Belgium, Spain, Germany, Luxembourg, the Netherlands, Portugal, and Switzerland. Its underlying algorithm, the FSAm-NPS (modified French version of the UK Food Standards Agency nutrient profiling score), assigns up to 10 unfavourable points to each of four components—energy, total sugars, saturated fatty acids (SFAs), and sodium, calculated per 100 g or per 100 ml of food [1]. Favourable points are assigned for fruit, vegetables, legumes, nuts, fibre, and protein. The final score falls within a five-letter, colour-coded grade from A (most favourable) to E (least favourable).

The scientific case for Nutri-Score rests substantially on a body of prospective cohort evidence. The EPIC study, involving 521,324 participants across 10 European countries, showed that a higher FSAm-NPS dietary index score was associated with significantly greater all-cause and cancer-specific mortality. Analogous findings have been reported in the Spanish SUN cohort and in the ENRICA nationwide study [2-4].

These associations are methodologically robust and have appropriately elevated the system's public health credibility. However, epidemiological validation of a composite dietary quality index does not necessarily establish the validity of the individual component weights that constitute the index. The present paper argues that the FSAm-NPS algorithm contains an unresolved construct validity gap arising from unit incommensurability: the unit upon which the algorithm operates (nutrient concentration per 100 g of food) is structurally incompatible with the epidemiologically relevant unit of exposure (nutrient intake in mg or g per day). This incompatibility has substantive consequences for the relative penalisation of food components and, by extension, for the health inferences that

consumers, policymakers, and food manufacturers are expected to draw from Nutri-Score grades.

METHODS

Analytical approach

This study is a conceptual and analytical review, not an empirical investigation. The analysis proceeds in three steps. First, we examine the FSAm-NPS scoring architecture with reference to the original model documentation by Rayner et al. [1] to characterise the rationale and constraints of the per-100-g unit choice. Second, we apply Messick's construct validity framework [5]—the standard psychometric reference for instrument validity—to assess whether the unit of measurement is coherent with the construct being measured. Third, we review the exposure units reported in landmark dose-response meta-analyses for the four penalised components (energy, sugars, SFA, sodium) to determine whether any existing epidemiological evidence supports calibration of the FSAm-NPS weights in per-100-g units.

Reference meta-analyses and systematic reviews were identified from existing citations within the FSAm-NPS validation literature and supplemented by targeted PubMed searches for dose-response studies published before December 2024. For sodium, the primary reference is the systematic review and meta-analysis by Wang et al. [6], and for SFA, the Cochrane review by Hooper et al. [7].

RESULTS

The per-100-g unit choice: rationale and inherent constraints

The decision to score food quality per 100 g rather than per serving is explicitly documented by Rayner et al. [1]. The model uses the per-100-g measure on the grounds that it is designed to assess the nutritional quality of the food independently of how it is consumed. The authors acknowledge that a per-serving approach would have been feasible in principle, but that portion sizes and

consumption patterns are individual and cannot be standardised. Penalty thresholds were derived from proportions of guideline daily amounts (GDA) and from analyses of the distribution of nutrient concentrations in the food supply, not from quantitative dose-response functions relating food nutrient concentration to individual health risk.

This design choice has a formal corollary that has not been systematically addressed in the subsequent validation literature: because the scoring is expressed in mg per 100 g, it is incommensurable with the unit of exposure in dose-response nutritional epidemiology, which is individual intake in mg per day. A food may receive a fixed Nutri-Score penalty regardless of whether it is consumed in a 10 g portion or a 200 g portion. A food with 600 mg sodium per 100 g delivers 60 mg when consumed in a 10 g condiment serving and 1,200 mg when consumed in a 200 g main-course serving; these contributions differ by a factor of 20, yet the FSAm-NPS assigns identical points in both cases.

Construct validity: the unit coherence requirement

Messick's construct validity framework requires that an instrument's unit of measurement be coherent with the unit of the construct it purports to measure [5]. Nutri-Score purports to inform consumers about the health risk that a food represents. Such health risk is a function of ingested dose; not of nutrient concentration independent of the amount consumed. The bridge between concentration per 100 g and actual intake is the habitual consumption pattern—precisely the variable that the FSAm-NPS model declared impossible to standardise and therefore excluded from its measurements.

This exclusion produces a construct validity gap: the instrument measures a property of the food (nutrient density) rather than a property of the exposure (nutrient intake). These are related but

not identical, and treating them as equivalent requires an implicit assumption, that all food items in the scoring domain are consumed in amounts proportional to their concentration per 100 g, that is both empirically false and explicitly rejected by the model developers themselves.

Dose-response evidence and the calibration problem

A consequence of unit incommensurability is that the relative weights assigned to different components within the FSAm-NPS cannot be validated against dose-response evidence. All available meta-analyses quantify risk in terms of daily intake. The systematic review by Wang et al. on dietary sodium and cardiovascular risk reports hazard ratios per increment of sodium intake expressed in mg/day [6]. The Cochrane review by Hooper et al. on SFA reduction expresses benefits per gram-per-day reduction in SFA intake [7]. Converting these effect estimates to 100-g food units requires that a fixed reference portion size be specified, which reintroduces the very assumption the original model declined to make.

In practice, this means that the scoring equation, in which one additional point for sodium and one additional point for SFA are treated as equivalent, rests on an implicit claim that the population-level risk associated with a unit increase in sodium concentration per 100 g of food is equal to the population-level risk associated with a unit increase in SFA concentration per 100 g. No dose-response literature supports this equivalence, because no available epidemiological evidence reports risk in those units. The parity of component weights is therefore a normative convention, not an empirical calibration.

For example, extra-virgin olive oil (EVOO), widely recognised as a core component of the Mediterranean diet with demonstrated cardiovascular benefits, received Nutri-Score grades of C or D primarily on account of its energy and SFA density per 100 g. At typical culinary

portions of 10–15 g, its contribution to daily SFA intake is negligible (approximately 1–2 g) and its sodium contribution is effectively zero. The poor Nutri-Score grade does not reflect adverse health outcomes associated with EVOO consumption; rather, it reflects a high nutrient concentration measured independently of realistic portion size. The 2023 revision of the FSAm-NPS algorithm introduced partial corrections for fats and sugars and improved the treatment of EVOO and nuts. However, it preserved the per-100-g base unit and did not address the underlying calibration gap [8,9]. This is consistent with the observation that incremental parameter adjustments cannot resolve a structural unit mismatch; they can mitigate specific anomalies, but the formal validity problem remains unaddressed.

DISCUSSION

The central finding of this analysis is that the FSAm-NPS algorithm operates on a measurement unit (nutrient concentration per 100 g of food) that is structurally incommensurable with the epidemiologically relevant exposure unit (nutrient intake per day) used in dose-response nutrition research. This is not a minor technical limitation but a formal constraint that affects the interpretive validity of the component weights and, by extension, the health inferences that Nutri-Score grades are intended to convey.

An important clarification is warranted regarding the role of prospective validation evidence. The EPIC, SUN, and ENRICA studies provide robust evidence that a higher FSAm-NPS dietary index is associated with worse health outcomes [2-4]. These studies validate the system as a classifier of overall dietary quality: individuals who habitually consume foods with higher FSAm-NPS scores tend to have less healthy diets and worse health outcomes. This is a meaningful and practically valuable finding. However, it does not validate the individual component weights, because the validation methodology aggregates scores across thousands of food items, allowing asymmetries in

component weights to cancel out in the composite index. Thus, the associated validity is demonstrated at the level of the overall dietary pattern, not at the level of individual food scoring or component weights.

This distinction maps onto a well-recognised problem in psychometrics: the predictive validity of a composite score does not imply the validity of each subscale. An analogy from clinical measurement illustrates this point: a composite frailty index that reliably predicts mortality does not guarantee that each of its components contributes equally to mortality risk, even if the index assigns them equal weights. The composite's predictive validity and the components' relative weights are logically independent properties that require independent validation.

The implications for regulatory policy are significant. Nutri-Score grades are increasingly embedded in procurement policies, fiscal incentives (e.g., differential VAT rates under discussion in several EU member states), and reformulation targets for the food industry. If the component weights embedded in the algorithm are normative conventions rather than empirically calibrated risk estimates, regulatory decisions based on comparisons between individual foods' profiles, rather than comparisons between dietary patterns, may be epistemologically unsupported. A food manufacturer may reformulate to shift from grade D to grade C by reducing sodium, but the health benefit of such change cannot be quantified in per-100-g terms, because the dose-response literature does not provide risk estimates in those units.

A further implication concerns communication transparency. The Rayner et al. model documentation is explicit about the normative character of the penalty thresholds [1]. However, Nutri-Score's public communication and, in some cases, its academic commentary, conflates the two levels of evidence. The prospective cohort validation of the dietary index has been presented

as validation of the scoring system as a whole, when in fact it validates only the system's ability to discriminate between dietary patterns, not its calibration at the food level. Distinguishing these levels of validity is not pedantry; it is a prerequisite for appropriate policy use.

Two methodological pathways could address the identified gap. The first, applicable in the longer term, requires the generation of dose-response evidence expressed in serving-size-corrected units. This would involve defining standard reference portions by food category, and then deriving risk estimates per unit of nutrient concentration at that portion size. Such data would allow the FSAm-NPS weights to be anchored empirically rather than conventionally. The second pathway, implementable immediately, involves transparency: future versions of the algorithm and their accompanying documentation should explicitly characterise the component weights as normative conventions calibrated to the distribution of nutrient concentrations across food products, not as expressions of relative epidemiological risk. This would not diminish the utility of the system as a consumer information tool but would prevent its misapplication as a component-level risk quantification instrument.

This analysis is conceptual rather than empirical and does not involve primary data collection. It does not claim that Nutri-Score is ineffective as a dietary quality classifier (the prospective evidence to the contrary is substantial), but argues that its construct validity at the component weight level remains undemonstrated. The analysis is also limited to the four penalised components of the FSAm-NPS; the validity of the favourable-component scoring (fibre, protein, fruits, and vegetables) raises related but distinct questions that are not addressed here.

CONCLUSION

The Nutri-Score front-of-pack labelling system contains a construct validity gap at the food scoring level that has not been addressed by the

prospective epidemiological validation of the FSAm-NPS as a dietary quality index. This gap arises from unit incommensurability: the algorithm scores nutrient concentration per 100 g of food, while health risk is a function of nutrient intake per day. Any meaningful conversion between these two systems requires fixing a portion size, which the original model explicitly declined to do. As a result, the relative weights assigned to the four penalised components: energy, sugars, SFA, and sodium, cannot be derived from or validated against the dose-response literature in nutrition, because such literature does not report risk estimates in per-100-g-of-food units. Prospective cohort studies validate Nutri-Score as a dietary pattern classifier, not as a component-level risk quantification tool. We recommend two complementary lines of action: (1) investment in dose-response research that reports nutrient risk per serving-standardised concentration unit, which would provide an empirical basis for component weight calibration; and (2) explicit declaration, in future algorithm documentation and public communication, that the FSAm-NPS component weights are normative conventions subject to periodic revision, not empirically calibrated risk estimates. Addressing this validity gap is essential to ensure that Nutri-Score-based policy instruments, including procurement rules, fiscal measures, and reformulation targets, are epistemologically grounded

ABBREVIATIONS

ENRICA: Study on Nutrition and Cardiovascular Risk in Spain

EPIC: European Prospective Investigation into Cancer and Nutrition

EVOO: extra-virgin olive oil

FSAm-NPS: Food Standards Agency nutrient profiling score (modified French version)

GDA: guideline daily amounts

SFA/SFAs: saturated fatty acids

SUN: Seguimiento Universidad de Navarra cohort

ACKNOWLEDGEMENT

Generative AI tools were employed exclusively to assist with the language, organisation, and formatting of the manuscript. The author has thoroughly reviewed, verified, and edited all content and assumes full responsibility for the accuracy and integrity of the work. No confidential or identifiable information was input into the AI tools.

REFERENCES

1. Rayner M, Scarborough P, Boxer A, Stockley L. Nutrient profiles: Development of final model. London: Food Standards Agency; 2005. Available from: <https://www.ndph.ox.ac.uk/food-ncd/files/about/uk-ofcom-nutrient-profile-model.pdf>
2. Deschasaux-Tanguy M, Huybrechts I, Julia C, Hercberg S, Egnell M, Srouf B, et al. Association between nutritional profiles of foods underlying Nutri-Score front-of-pack labels and mortality: EPIC cohort study in 10 European countries. *BMJ*. 2020;370:m3173. doi:10.1136/bmj.m3173
3. Gómez-Donoso C, Martínez-González MA, Perez-Cornago A, Sayón-Orea C, Martínez JA, Bes-Rastrollo M. Association between the nutrient profile system underpinning the Nutri-Score front-of-pack nutrition label and mortality in the SUN project. *Clin Nutr*. 2021;40(3):1085-94. doi:10.1016/j.clnu.2020.07.008
4. Donat-Vargas C, Sandoval-Insausti H, Rey-García J, Ramón Banegas J, Rodríguez-Artalejo F, Guallar-Castillón P. Five-color Nutri-Score labeling and mortality risk in a nationwide, population-based cohort in Spain: the Study on Nutrition and Cardiovascular Risk in Spain (ENRICA). *Am J Clin Nutr*. 2021 May 8;113(5):1301-1311. doi:10.1093/ajcn/nqaa389
5. Messick S. Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *Am Psychol*. 1995;50(9):741-9. doi:10.1037/0003-066X.50.9.741
6. Wang YJ, Yeh TL, Shih MC, Tu YK, Chien KL. Dietary sodium intake and risk of cardiovascular disease: a systematic review and dose-response meta-analysis. *Nutrients*. 2020 Sep 25;12(10):2934. doi:10.3390/nu12102934
7. Hooper L, Martin N, Jimoh OF, Kirk C, Foster E, Abdelhamid AS. Reduction in saturated fat intake for cardiovascular disease. *Cochrane Database Syst Rev*. 2020;(5):CD011737. doi:10.1002/14651858.CD011737.pub3
8. Scientific Committee of the Nutri-Score. Update of the Nutri-Score algorithm for beverages. Second update report from the scientific committee of the Nutri-Score. 2023. Available from: <https://www.santepubliquefrance.fr/en/nutri-score>.
9. Deschasaux-Tanguy M, Huybrechts I, Julia C, Hercberg S, Sarda B, Fialon M, et al. Nutritional quality of diet characterized by the Nutri-Score profiling system and cardiovascular disease risk: a prospective study in 7 European countries. *Lancet Reg Health Eur*. 2024 Sep 10;46:101006. doi:10.1016/j.lanepe.2024.101006.