ISSN 2282-0930 • Epidemiology Biostatistics and Public Health - XIII[th] SISMEC Congress - Vol. 20 Suppl. 1

Identifying and Characterizing Shared and Ethnic Background Site-Specific Dietary Patterns by Hispanic/Latino Background and Site: The Use of Bayesian Multi-Study Factor Analysis in the Hispanic Community Health Study/Study of Latinos

De Vito Roberta[1], Stephenson Briana[2], Sotres-Alvarez Daniela[3], Siega-Riz Anna-Maria[4], Mattei Josiemer[5], Parpinel Maria[6], Peters Brandilyn[7], Bainter Sierra A.[8], Daviglus Martha L.[9], Van Horn Linda[10], Edefonti Valeria[11]

(1) Department of Biostatistics and Data Science Initiative, Brown University, Providence, RI, USA
(2) Department of Biostatistics, Harvard TH Chan School of Public Health, Boston, MA, USA
(3) Department of Biostatistics, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
(4) Departments of Nutrition and Biostatistics and Epidemiology, School of Public Health and Health Sciences, University of Massachusetts, Amherst, MA, USA
(5) Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA, USA
(6) Department of Medicine - DAME, Università degli Studi di Udine, Udine, Italy
(7) Department of Epidemiology & Population Health, Albert Einstein College of Medicine, Bronx, NY, USA
(8) Department of Psychology, University of Miami, Coral Gables, FL, USA
(9) Institute for Minority Health Research, University of Illinois at Chicago, Chicago, IL, USA
(10) Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA
(11) Branch of Medical Statistics, Biometry and Epidemiology "G. A. Maccacaro", Department of Clinical Sciences and Community Health – Dipartimento di Eccellenza 2023-2027, Università degli Studi di Milano, Milano, Italy
Equally contributed to the paper

CORRESPONDING AUTHOR: *Edefonti Valeria, valeria.edefonti@unimi.it*

INTRODUCTION

Dietary patterns (DPs) are combinations of dietary components intended to summarize key aspects of diet, while taking advantage of synergies between single components. A posteriori DPs are defined from the application of multivariate statistics, including principal component and factor analyses. New statistical methods like multi-study factor analysis have been recently used to distinguish subpopulation-specific DPs (i.e., study/country-specific features within an international consortium or subpopulation-specific features within a single study), as well as those shared among all groups in a population [1].

The Hispanic Community Health Study/Study of Latinos (HCHS/SOL), the most extensive and ongoing community-based cohort of Hispanic/Latino adults from 4 US sites to date, provides a unique opportunity to identify shared and subpopulation-specific a posteriori DPs.

AIMS

The present work aims to: 1. identify shared and ethnic background-site (EBS)-specific (nutrient-based) DPs within the HCHS/SOL study and 2. characterize the identified DPs in terms of food-group consumption, an overall measure of diet quality, socio-demographic and lifestyle characteristics.

METHODS

THE HCHS/SOL

The HCHS/SOL is a population-based cohort study designed to identify disease prevalence rates and risk factors of Hispanic/Latino populations residing within 4 urban US communities (Bronx, Chicago, Miami, and San Diego) and representing individuals with 7 ethnicity backgrounds (Cuban, Dominican, Mexican, Puerto Rican, Central and South American, and mixed). Participants were selected using a probability sampling design [2].

Dietary habits at baseline (16,415 subjects from 2008 to 2011) were assessed using two 24-hr recalls, the first conducted in person and the second via telephone <=30 days after. The Nutrition Data System for Research software allowed for nutrient intake estimation [3].

DOI: 10.54103/2282-0930/29374

SELECTION OF SUBJECTS AND VARIABLES

We excluded Hispanics and Latinos from other/mixed backgrounds, with unreliable dietary recalls, or providing extreme energy intake. We also excluded subpopulations <200 participants after previous exclusions. This gave a final sample size of 15,021 participants.

We selected 42 nutrients that well represent the overall diet for Hispanics/Latinos. For each participant, nutrient intakes were derived from either one available reliable recall or the mean of the two available reliable recalls.

STATISTICAL ANALYSIS

Bayesian multi-study factor analysis (BMSFA) was carried out on the correlation matrices of the log-transformed nutrient intakes. The total number of factors to retain was selected using the spectral decomposition of the factors. After the singular value decomposition method used in the BMSFA for identifiability, the varimax rotation was applied to the shared factor-loading matrix to achieve a better-defined loading structure [4]. Characterization of DPs against selected food groups, a measure of diet quality, selected socio-demographic and lifestyle factors was based on survey-weighted regression models. Calculations were carried out using the R software [5].

RESULTS

The selected model included 4 shared (62.5% total variance explained) and 12 EBS-specific DPs (variance around 10%), one for each of the 12 EBS combinations (Figure 1). Among shared DPs, the first, named **Plant-based foods**, loaded highly on vegetable protein, several minerals, vitamin B1, niacin, natural folate, soluble and insoluble fiber, the second, named **Dairy products**, loaded highly on short- and medium-chain saturated fatty acids and calcium, vitamins B2, B12, D, and retinol; the third shared factor, named **Seafood**, loaded highly on EPA, DPA, and DHA and the fourth, named **Processed foods**, loaded highly on several fats, including long-chain saturated and monounsaturated fatty acids, linoleic and linolenic acids, total trans fatty acids, and natural alpha-tocopherol. Most EBS-specific DPs were further grouped into overarching profiles: Animal vs. vegetable source, Animal source only, and Poultry vs. dairy products, to capture nuances within animal-based DPs. Puerto Rican background participants from Chicago expressed a strikingly different DP.

When interpreted in terms of food groups, the identified DPs confirmed the names based on nutrients. Higher overall diet quality was observed with increasing categories of **Plant-based foods**, **Seafood**, and the "Puerto Rican background–Chicago" EBS-specific DP, whereas increasing categories of **Dairy products**, **Processed foods**, and the remaining EBS-specific DPs were related to lower diet quality. Compared to non-US-born participants, US-born individuals exhibited lower adherence to the **Plant-based foods** and **Dairy products** DPs but higher adherence to **Processed foods**, **Seafood**, and 6 EBS-specific DPs.

CONCLUSIONS

In its first application in nutritional epidemiology, BMSFA succeeded in simultaneously estimating well-interpretable shared and EBS-specific DPs within 12 combinations of background and site.

REFERENCES

1. De Vito R., Lee Y. C. A., Parpinel M., et al. Shared and study-specific dietary patterns and head and neck cancer risk in an international consortium. Epidemiology, 2019;30(1):93-102.
2. Lavange L. M., Kalsbeek W. D., Sorlie P. D., et al. Sample design and cohort selection in the Hispanic Community Health Study/Study of Latinos. Ann Epidemiol., 2010;20:642–649.
3. Sorlie P. D., Aviles-Santa L. M., Wassertheil-Smoller S., et al. Design and implementation of the Hispanic Community Health Study/Study of Latinos. Ann Epidemiol., 2010;20:629–641.
4. De Vito, R., Bellio, R., Trippa, L., and Parmigiani, G. Bayesian multi-study factor analysis for high-throughput biological data. The Annals of Applied Statistics 2021, 15:1723-1741.
5. R Core Team, R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2025.

DOI: 10.54103/2282-0930/29374

ISSN 2282-0930 • Epidemiology Biostatistics and Public Health - XIIIth SISMEC Congress - Vol. 20 Suppl. 1

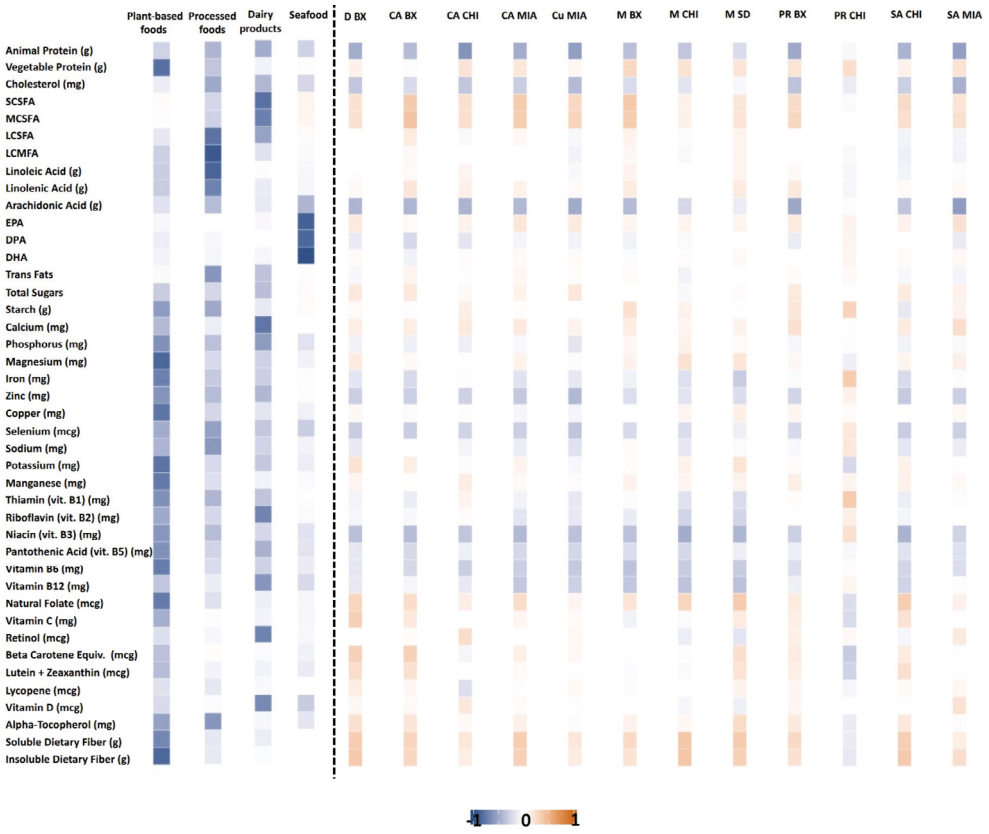

Figure 1. Heatmap of the estimated factor loadings for the shared (left) and background/site-specific (right) dietary patterns. Hispanic Community Health Study/Study of Latinos – baseline examination (2008-2011). Abbreviations: BMSFA: Bayesian multi-study factor analysis; BX: Bronx; CA: Central American; Cu: Cuban; CHI: Chicago; D: Dominican; DHA: docosahexaenoic acid; DPA: docosapentaenoic acid; EPA: eicosapentaenoic acid; M: Mexican; MCSFA: medium-chain saturated fatty acids; LCMFA: long-chain monounsaturated fatty acids; LCSFA: long-chain saturated fatty acids; MIA: Miami; PR: Puerto Rican; SA: South American; SCSFA: short-chain saturated fatty acids; SD: San Diego

DOI: 10.54103/2282-0930/29374