BMJ Open

Model-based recursive partitioning to identify risk clusters for metabolic syndrome and its components: findings from the International Mobility in Aging Study

Catherine M Pirkle,[1] Yan Yan Wu,[1] Maria-Victoria Zunzunegui,[2] José Fernando Gómez[3]

[1]Office of Public Health Studies, University of Hawai'i at Mānoa, Honolulu, Hawaii, USA
[2]Institut de recherche en santé publique, Université de Montréal, Montréal, Canada
[3]Facultad de Ciencias para la Salud, Universidad de Caldas, Manizales, Colombia

Correspondence to
Dr Catherine M Pirkle;
cmpirkle@hawaii.edu

ABSTRACT

Objective Conceptual models underpinning much epidemiological research on ageing acknowledge that environmental, social and biological systems interact to influence health outcomes. Recursive partitioning is a data-driven approach that allows for concurrent exploration of distinct mixtures, or clusters, of individuals that have a particular outcome. Our aim is to use recursive partitioning to examine risk clusters for metabolic syndrome (MetS) and its components, in order to identify vulnerable populations.

Study design Cross-sectional analysis of baseline data from a prospective longitudinal cohort called the International Mobility in Aging Study (IMIAS).

Setting IMIAS includes sites from three middle-income countries—Tirana (Albania), Natal (Brazil) and Manizales (Colombia)—and two from Canada—Kingston (Ontario) and Saint-Hyacinthe (Quebec).

Participants Community-dwelling male and female adults, aged 64–75 years (n=2002).

Primary and secondary outcome measures We apply recursive partitioning to investigate social and behavioural risk factors for MetS and its components. Model-based recursive partitioning (MOB) was used to cluster participants into age-adjusted risk groups based on variabilities in: study site, sex, education, living arrangements, childhood adversities, adult occupation, current employment status, income, perceived income sufficiency, smoking status and weekly minutes of physical activity.

Results 43% of participants had MetS. Using MOB, the primary partitioning variable was participant sex. Among women from middle-incomes sites, the predicted proportion with MetS ranged from 58% to 68%. Canadian women with limited physical activity had elevated predicted proportions of MetS (49%, 95% CI 39% to 58%). Among men, MetS ranged from 26% to 41% depending on childhood social adversity and education. Clustering for MetS components differed from the syndrome and across components. Study site was a primary partitioning variable for all components except HDL cholesterol. Sex was important for most components.

Strengths and limitations of this study

► Explores social and behavioural risk clustering for metabolic syndrome among community-dwelling older adults from five diverse global settings.
► Applies model-based recursive partitioning (MOB), which is more intuitive and computationally efficient than classification and regression tree, to identify risk clusters.
► Provides an example of how MOB can be used in a modestly sized sample for hypothesis generation about complex admixtures of risk factors.
► Lacks data on participant diet, which likely clusters with many of the social and behavioural factors examined.
► Strong contextual influences may have masked variance attributable to individual behaviours.

Conclusion MOB is a promising technique for identifying disease risk clusters (eg, vulnerable populations) in modestly sized samples.

INTRODUCTION

With ageing, life's hazards and rewards amass and become embodied in ways that diminish or protect health. Differences in health trajectories are the product of cumulative risk and protective factors that are programmed into biobehavioural regulatory systems.[1] The cardiometabolic pathologies commonly observed in older adults (partially) reflect the collective burden exacted on their bodies as they adapt to life's challenges.[2] The types and magnitude of challenges that bodies are exposed to vary across societies and time, as these reflect underlying social orders with regard to the distribution of economic, political and social resources.[2] Research that purposefully compares populations of older adults across heterogeneous

societies may inform our understanding of modifiable social and behavioural factors that influence the dysregulation of biological systems. Of importance, social norms and patterning are capable of creating toxic or protective clusters that manifest among identifiable subgroups.[3] Such information is useful for directing public health interventions and for considering how contextual conditions render groups particularly vulnerable.

Metabolic syndrome (MetS) is a highly prevalent health condition among older adults; it confers an approximate twofold increased risk of cardiovascular disease and fivefold increased risk of diabetes.[4] It entails a constellation of components including obesity, impaired glucose metabolism, hypertension and atherogenic dyslipidaemia.[4] In older adults, MetS prevalence varies considerably across populations. Among older adults in the USA and Europe, MetS prevalence is estimated at 30%[4 5]; in urban China researchers estimate a prevalence of 60%.[6] In Canada, MetS prevalence increases with chronological age; approximately 40% of adults aged 60–79 years were estimated to have the syndrome according to the 2007–2009 Canadian Health Measures Survey.[7] Studies of older adults frequently document greater prevalence in women compared with men,[6–10] but report inconsistent associations with income and education. Some document linear associations between education and MetS prevalence, with the lowest educated at highest risk.[10] Others document non-linear associations, in which the highest risk groups are lower-middle income and high school graduates, while those in the lowest income group and without a high school education are slightly protected.[7] Yet, others observe strong associations with education, but not income[10] or associations with education and income in one sex over the other.[10 11] Heterogeneity in MetS prevalence likely reflects complex, contextually specific risk admixtures.

Epidemiological research on ageing explicitly acknowledges that environmental, social, psychological and biological systems interact to influence health outcomes.[13] The well-known ecological model posits that patterns of health are affected by a dynamic interplay among these factors across the life course.[2 12] A challenge for epidemiologists, especially with modest sample sizes, is to operationalise models that assume the joint effects of multiple risk factors on health conditions, such as MetS.[13]

Recursive partitioning is a technique that allows for exploration of distinct mixtures, or clusters, of individuals that have a particular outcome. Based on a set of candidate independent variables, it can produce classification trees with a series of binary splits highlighting subgroups with relatively similar risk profiles for a given outcome.[14 15] The classification trees depict the joint effects of multiple risk factors.[15] It is a data-driven approach with the potential to identify complex interactions worthy of future investigation.[15] Researchers have applied partitioning techniques to identify high-risk subgroups for cardiovascular disease, diabetes and falls in population-based studies.[16–18] Most of this work examines clinical or genetic risk factors, but the same technique can be expanded to examine social and behavioural risk factors.

In an international, multisite cohort of community-dwelling older adults, we apply recursive partitioning to investigate social and behavioural risk factors for MetS and its components. Our objectives are to assess if there are social/behavioural risks clusters for those with MetS or the components of the syndrome, and whether these risk clusters vary across societies. To date, we know of no other studies employing recursive-partitioning techniques to investigate predictors of MetS that are informed by a social epidemiological perspective.

METHODS

Data source and study populations

This is an analysis of 2012 baseline data from the International Mobility in Aging Study (IMIAS). IMIAS is composed of community-dwelling older adults, 65–74 years of age. This study comprises three sites in middle-income countries—Tirana (Albania), Natal (Brazil) and Manizales (Colombia)—and two from a high-income country—Kingston (Ontario, Canada) and Saint-Hyacinthe (Quebec, Canada). These cities represent diverse ways of living in distinct societies, providing a wide range of risk factors and outcomes. For example, Tirana is the capital of an ex-communist country in rapid transition to capitalism, while Manizales is in the Andean coffee-growing region, of Catholic tradition and relatively affluent. Approximately 200 men and women, each, were recruited per site for a sample size of 2002. A detailed description of the study sites and cohort is available elsewhere.[19]

Recruitment

In Tirana, Manizales and Natal, we recruited participants through their neighbourhood primary care centres by selecting a random sample of older adults registered at each.[19] The response rate was over 90%.[19] Ethics' committees in Canada prohibited researchers directly contacting potential participants. Invitations to participate in the project were therefore sent indirectly via family physicians.[19] Thirty per cent of people receiving a letter of invitation from their doctor in Kingston and St. Hyacinthe contacted the IMIAS research team; 95% agreed to participate.[19] Comparison with 2006 Canadian census data suggests participants in Kingston were more educated than the general population of that city, while participants from St. Hyacinthe had similar educational levels to inhabitants of that city. Otherwise, characteristics between those recruited and the sampling frame were very similar.[19] At all sites, over 80% of older adults were registered at a health centre or had a primary care physician[19]; it is unlikely that our recruitment strategy systematically excluded a large segment of older adult population.

Exclusion criterion

Those with four or more errors on the orientation scale of the Leganes Cognitive Test[20]; low scores indicate inability to complete study procedures.

Measurements

Study procedures were carried out at the participant's home, in the local language, by a trained interviewer. Detailed descriptions of data collection procedures are provided elsewhere.[19]

Metabolic syndrome

Except for the measure of insulin resistance, we defined MetS according to the Adult Panel Treatment III criteria.[21] IMIAS did not collect fasting glucose and the corresponding glycosylated haemoglobin (HbA1c) value was used instead.[22] Thus, MetS was defined as the presence of three or more of the following: abdominal obesity measured by waist circumference (women >88 cm; men >102 cm); elevated triglycerides (≥150 mg/dL); low high-density lipoprotein cholesterol (HDL-C men <40 mg/dL; HDL-C women <50 mg/dL); elevated HbA1c (≥5.7%); and high blood pressure (≥135 mm Hg systolic and/or ≥85 mm Hg diastolic).

Socioeconomic and demographic characteristics

We categorised education as: less than secondary school and/or illiterate, some secondary school to completed secondary school and postsecondary education. A participant's living arrangement was determined with the following questions: Do you live alone (yes/no)? (If no) Who do you live with? Responses were then categorised as: alone, spouse only and multiple people.

We determined exposure to childhood social and economic adversity with a scale that varied from 0 to 3 including parental alcohol or drug abuse, witnessing family physical violence and having been physically abused (childhood social adversity); poor economic status, hunger and parental unemployment (childhood economic adversity).[23] Occupation was grouped into five categories: non-manual, service, agriculture, manual and housewife, according to self-reported longest held occupation (based on International Labour Organization categories). We enquired about current annual income levels. Based on the annual minimum salary for each site, individuals were categorised as poor, middle, upper middle and high income. For example, in Canada, the minimum salary is $C19 680 per year. Thus, we categorised Canadian participants as poor if they earned less than $C20 000 per year. Those who earned more than the minimum salary but less than twice it ($C20 000–39 999) were classified as middle income, while those that earned twice or higher the minimum salary, but less than three times it, were classified as upper middle ($C40 000–59 999) and those that earned three times or more the minimum salary (≥$C60 000) were classified as upper income. This was done for each site based on the site-specific minimum salary. Income sufficiency was assessed accordingly: to what extent does your income allow you to meet your needs? Responses were categorised into: very sufficient, sufficient and not (at all) sufficient. We asked participants about their work history in the past 2 weeks and categorised them as: worked with remuneration; worked without remuneration; had a job, but did not work; retired or pensioned; and did not work. We also asked participants if they currently smoke. Responses were categorised as regular, occasional, used to be a smoker and never. Finally, we assessed minutes of physical activity per week with a validated computer-animated assessment tool.[24]

Statistical analysis

Descriptive statistics summarise overall sample characteristics. Because the distributions were positively skewed for some measures, we report the median, first and third quartiles for all continuous variables. We performed X^2 tests to investigate the associations between MetS and categorical independent variables and carried out two-sample t-tests to examine mean differences in biomarkers. Random forest method was used to impute missing physical activities (n=59).[25]

Model-based recursive partitioning method (MOB) was applied to cluster individuals into subgroups with similar response values.[26–28] MOB is reminiscent of the classification and regression tree (CART) algorithms, which split the datasets into subsets based on independent (partitioning) variables, of which the distributions of the response values are most different.[13] Whereas CARTs have constant fits in the terminal nodes, MOB trees have parametric models with one or more predictor variables controlled in each step of the partitioning, and in their terminal nodes. For instance, age is controlled in the MOB analysis of MetS using logistic regression models and the MOB algorithm cycles iteratively through the following steps: (1) fit the logistic regression with MetS as response variable and age as control variable, (2) test for parameter instability over a set of partitioning variables (socioeconomic and demographic characteristics) while controlling for age, (3) if there is some overall parameter instability, split the data set with respect to the variable associated with the highest instability (ie, the smallest P value) and (4) repeat the procedure in each of the resulting subsamples with different risk of MetS. The process is termed recursive because each subpopulation may be split a number of times until a particular stopping criterion is reached. Our stopping criteria were: 5% level of significance and minimum sample size of 100 at terminal nodes. For continuous partitioning variables, MOB tests and selects an optimal cut-off point and split subjects into two subgroups.[26–28]

We performed MOB for MetS, systolic blood pressure (SBP), diastolic blood pressure (DBP), waist circumference, log transformed triglyceride, HDL-C and HbA1c, controlling for age. The partitioning variables included social and behavioural risk characteristics described previously. Statistical software R (V.3.2.2) and the R package 'party' were used.

RESULTS

Complete data on all variables were available for 1628 (81%) participants. Table 1 presents the frequency of

Table 1 Descriptive characteristics of the participants and frequency of MetS

	Overall (n=1628)	MetS Yes (42.7%)	No (57.3%)	
	n (%)	n (%)	n (%)	P value*
Site				
Kingston	289 (17.8%)	92 (31.8%)	197 (68.2%)	<0.0001
St. Hyacinthe	310 (19.0%)	96 (31.0%)	214 (69.0%)	
Tirana	344 (21.1%)	171 (49.7%)	173 (50.3%)	
Manizales	374 (23.0%)	172 (46.0%)	202 (54.0%)	
Natal	311 (19.1%)	164 (52.7%)	147 (47.3%)	
Sex				
Female	844 (51.8%)	424 (50.2%)	420 (49.8%)	<0.0001
Male	784 (48.2%)	271 (34.6%)	513 (65.4%)	
Educational attainment				
Primary/illiterate	760 (46.7%)	366 (48.2%)	394 (51.8%)	0.0001
Secondary	217 (13.3%)	91 (41.9%)	126 (58.1%)	
Postsecondary	651 (40.0%)	238 (36.6%)	413 (63.4%)	
Current living arrangements				
Alone	254 (15.6%)	100 (39.4%)	154 (60.6%)	0.0001
Spouse only	609 (37.4%)	225 (36.9%)	384 (63.1%)	
Other	765 (47.0%)	370 (48.4%)	395 (51.6%)	
Adult occupation				
Non-manual	593 (36.4%)	211 (35.6%)	382 (64.4%)	<0.0001
Service	160 (9.8%)	67 (41.9%)	93 (58.1%)	
Agriculture	94 (5.8%)	41 (43.6%)	53 (56.4%)	
Manual	607 (37.3%)	274 (45.1%)	333 (54.9%)	
Housewife	174 (10.7%)	102 (58.6%)	72 (41.4%)	
Current income				
Poor	417 (25.6%)	189 (45.3%)	228 (54.7%)	0.001
Middle	704 (43.2%)	315 (44.7%)	389 (55.3%)	
Upper middle	346 (21.3%)	145 (41.9%)	201 (58.1%)	
High	161 (9.9%)	46 (28.6%)	115 (71.4%)	
Perceived income sufficiency				
Very sufficient	339 (20.8%)	102 (30.1%)	237 (69.9%)	<0.0001
Sufficient	537 (33.0%)	215 (40.0%)	322 (60.0%)	
Not (at all) sufficient	752 (46.2%)	378 (50.3%)	374 (49.7%)	
Childhood economic adversity				
No adversities	858 (52.7%)	351 (40.9%)	507 (59.1%)	0.0291
One adversity event	453 (27.8%)	186 (41.1%)	267 (58.9%)	
Two adversity events	212 (13.0%)	102 (48.1%)	110 (51.9%)	
Three adversity events	105 (6.4%)	56 (53.3%)	49 (46.7%)	
Childhood social adversity				
No adversities	1231 (75.6%)	505 (41.0%)	726 (59.0%)	0.0792
One adversity event	245 (15.0%)	113 (46.1%)	132 (53.9%)	
Two adversity events	119 (7.3%)	59 (49.6%)	60 (50.4%)	
Three adversity events	33 (2.0%)	18 (54.5%)	15 (45.5%)	
Smoker				

Table 1 Continued

	Overall (n=1628)	MetS Yes (42.7%)	No (57.3%)	
	n (%)	n (%)	n (%)	P value*
Regular	138 (8.5%)	64 (46.4%)	74 (53.6%)	0.109
Occasional	37 (2.3%)	18 (48.6%)	19 (51.4%)	
Used to be	676 (41.5%)	265 (39.2%)	411 (60.8%)	
Never	777 (47.7%)	348 (44.8%)	429 (55.2%)	
Current employment status				
Worked with remuneration	185 (11.4%)	66 (35.7%)	119 (64.3%)	0.011
Worked without remuneration	201 (12.3%)	105 (52.2%)	96 (47.8%)	
Had a job, but did not work	25 (1.5%)	9 (36.0%)	16 (64.0%)	
Retired or pensioned	1119 (68.7%)	468 (41.8%)	651 (58.2%)	
Did not work	98 (6.0%)	47 (48.0%)	51 (52.0%)	
	Median(Q1, Q3)	Median(Q1, Q3)	Median(Q1, Q3)	
Age (years)	69 (67, 72)	69 (67, 72)	69 (67, 71)	0.0451
Physical activity (min/week)	20 (9, 39)	15 (6, 32)	24 (10, 43)	<0.0001
SBP (mm Hg)	138 (126, 152)	143 (134, 157)	132 (122, 146)	<0.0001
DBP (mm Hg)	79 (71, 86)	81 (74, 88)	77 (70, 84)	<0.0001
Waist (cm)	96 (88, 104)	101 (94, 108)	92 (85, 100)	<0.0001
HbA1c (%)	5.8 (5.5, 6.2)	6.1 (5.8, 6.6)	5.7 (5.4, 6.0)	<0.0001
Triglyceride (mg/dL)	126 (91, 172)	165 (125, 211)	105 (78, 134)	<0.0001
HDL (mg/dL)	50 (43, 60)	45 (39, 53)	55 (47, 63)	<0.0001

*P values obtained from χ^2 tests of association between MetS (yes/no) and categorical explanatory variables, and t-tests for difference in MetS (yes/no) for continuous variable.

DBP, diastolic blood pressure; HbA1c, glycosylated haemoglobin; HDL, high-density lipoprotein; MetS, metabolic syndrome; SBP, systolic blood pressure.

MetS according to the participant characteristics. MetS was observed in 43% participants, 50% of women and 35% of men. For most variables, there were important differences in the proportion of participants with MetS. It concentrated among those of lower socioeconomic status: those with lower education, lower incomes, manual workers and housewives. More MetS was observed among those reporting childhood adversities. Those with MetS reported had higher mean blood pressure, waist circumference, HbA1c, triglyceride, low HDL measures and walked less on average than those without it.

Table 2 presents participant characteristics by study site and shows a notably higher frequency of MetS among participants from the middle-income sites (46%–53%) compared with those from the Canadian sties (≈30%).

Figure 1 depicts the MOB tree for MetS, adjusting for participant age. The highest estimates of MetS prevalence were observed in clusters of women from the middle-income study sites (Tirana, Manizales and Natal). In these clusters, the predicted proportion with MetS varied from 58% to 68%, depending on education. Better-educated women from these sites had more MetS. Among women from the Canadian sites, less walking time per week distinguished the higher from lower probability cluster. The lowest predicted proportion (26%) of MetS was observed in men with postsecondary education reporting no childhood social adversities, and in women from the Canadian sites who had more walking time per week. The graphs under each node in figure 1 depict the estimated prevalence according to age and demonstrate that for certain nodes (e.g., 7), there is a strong association between increasing age and higher estimated MetS prevalence.

The online supplementary files contain the MOB trees for MetS components. Each tree depicts unique clusters that do not correspond with those observed for MetS as a whole. For example, the highest estimated mean SBP (154 mm Hg) was observed among participants from Tirana and Natal, with income insufficiency, and who smoked regularly or used to. For this outcome, in contrast to MetS as a whole, sex was not a partitioning variable. Overall, in all models except HDL, study site was the primary partitioning variable and in some cases (triglycerides), the only one. Typically, participants from Natal and Tirana had unfavourable estimates for MetS components; however, participants from Manizales had the highest estimated triglyceride concentrations (145 mg/dL). Participant sex was a key partitioning variable for DBP, waist circumference and

Table 2 Descriptive characteristics of the participants by study site

	Kingston	St. Hyacinthe	Tirana	Manizales	Natal
	289 (17.8%)	**310 (19%)**	**344 (21.1%)**	**374 (23%)**	**311 (19.1%)**
	n (%)	**n (%)**	**n (%)**	**n (%)**	**n (%)**
MetS					
Yes	92 (31.8%)	96 (31.0%)	171 (49.7%)	172 (46.0%)	164 (52.7%)
No	197 (68.2%)	214 (69.0%)	173 (50.3%)	202 (54.0%)	147 (47.3%)
Sex					
Female	152 (52.6%)	166 (53.5%)	178 (51.7%)	190 (50.8%)	158 (50.8%)
Male	137 (47.4%)	144 (46.5%)	166 (48.3%)	184 (49.2%)	153 (49.2%)
Educational attainment					
Primary/illiterate	29 (10.0%)	85 (27.4%)	54 (15.7%)	310 (82.9%)	282 (90.7%)
Secondary	37 (12.8%)	67 (21.6%)	79 (23.0%)	19 (5.1%)	15 (4.8%)
Postsecondary	223 (77.2%)	158 (51.0%)	211 (61.3%)	45 (12.0%)	14 (4.5%)
Current living arrangements					
Alone	88 (30.4%)	72 (23.2%)	31 (9.0%)	47 (12.6%)	16 (5.1%)
Spouse only	127 (43.9%)	201 (64.8%)	151 (43.9%)	68 (18.2%)	62 (19.9%)
Other	74 (25.6%)	37 (11.9%)	162 (47.1%)	259 (69.3%)	233 (74.9%)
Adult occupation					
Non-manual	226 (78.2%)	158 (51.0%)	116 (33.7%)	63 (16.8%)	30 (9.6%)
Service	24 (8.3%)	40 (12.9%)	23 (6.7%)	22 (5.9%)	51 (16.4%)
Agriculture	2 (0.7%)	18 (5.8%)	5 (1.5%)	37 (9.9%)	32 (10.3%)
Manual	27 (9.3%)	78 (25.2%)	199 (57.8%)	161 (43.0%)	142 (45.7%)
Housewife	10 (3.5%)	16 (5.2%)	1 (0.3%)	91 (24.3%)	56 (18.0%)
Current income					
Poor	51 (17.6%)	109 (35.2%)	36 (10.5%)	191 (51.1%)	30 (9.6%)
Middle	97 (33.6%)	122 (39.4%)	204 (59.3%)	112 (29.9%)	169 (54.3%)
Upper middle	61 (21.1%)	62 (20.0%)	82 (23.8%)	50 (13.4%)	91 (29.3%)
High	80 (27.7%)	17 (5.5%)	22 (6.4%)	21 (5.6%)	21 (6.8%)
Perceived income sufficiency					
Very sufficient	174 (60.2%)	132 (42.6%)	5 (1.5%)	19 (5.1%)	9 (2.9%)
Sufficient	101 (34.9%)	157 (50.6%)	121 (35.2%)	89 (23.8%)	69 (22.2%)
Not (at all) sufficient	14 (4.8%)	21 (6.8%)	218 (63.4%)	266 (71.1%)	233 (74.9%)
Childhood economic adversity					
No adversities	190 (65.7%)	201 (64.8%)	147 (42.7%)	216 (57.8%)	104 (33.4%)
One adversity event	73 (25.3%)	92 (29.7%)	89 (25.9%)	108 (28.9%)	91 (29.3%)
Two adversity events	24 (8.3%)	15 (4.8%)	75 (21.8%)	28 (7.5%)	70 (22.5%)
Three adversity events	2 (0.7%)	2 (0.6%)	33 (9.6%)	22 (5.9%)	46 (14.8%)
Childhood social adversity					
No adversities	210 (72.7%)	229 (73.9%)	281 (81.7%)	287 (76.7%)	224 (72.0%)
One adversity event	42 (14.5%)	54 (17.4%)	27 (7.8%)	63 (16.8%)	59 (19.0%)
Two adversity events	25 (8.7%)	22 (7.1%)	30 (8.7%)	18 (4.8%)	24 (7.7%)
Three adversity events	12 (4.2%)	5 (1.6%)	6 (1.7%)	6 (1.6%)	4 (1.3%)
Smoker					
Regular	14 (4.8%)	19 (6.1%)	43 (12.5%)	37 (9.9%)	25 (8.0%)
Occasional	3 (1.0%)	5 (1.6%)	12 (3.5%)	13 (3.5%)	4 (1.3%)
Used to be	141 (48.8%)	169 (54.5%)	81 (23.5%)	147 (39.3%)	138 (44.4%)

Table 2 Continued

	Kingston	St. Hyacinthe	Tirana	Manizales	Natal
	289 (17.8%)	**310 (19%)**	**344 (21.1%)**	**374 (23%)**	**311 (19.1%)**
	n (%)	n (%)	n (%)	n (%)	n (%)
Never	131 (45.3%)	117 (37.7%)	208 (60.5%)	177 (47.3%)	144 (46.3%)
Current employment status					
Worked with remuneration	43 (14.9%)	39 (12.6%)	8 (2.3%)	58 (15.5%)	37 (11.9%)
Worked without remuneration	8 (2.8%)	18 (5.8%)	8 (2.3%)	99 (26.5%)	68 (21.9%)
Had a job, but did not work	8 (2.8%)	2 (0.6%)	4 (1.2%)	8 (2.1%)	3 (1.0%)
Retired or pensioned	228 (78.9%)	246 (79.4%)	323 (93.9%)	133 (35.6%)	189 (60.8%)
Did not work	2 (0.7%)	5 (1.6%)	1 (0.3%)	76 (20.3%)	14 (4.5%)
	Median(Q1, Q3)	Median (Q1, Q3)	Median(Q1, Q3)	Median(Q1, Q3)	Median(Q1, Q3)
Age (year)	69(67, 71)	68(67, 71)	70(66, 72)	69(67, 72)	69 (67, 71)
Physical activity (min/week)	26(9, 51)	24(9, 39)	27(11, 48)	17(9, 39)	10(4, 23)
SBP (mm Hg)	135 (124, 145)	134 (125, 143)	144 (130, 161)	133 (123, 146)	146 (130, 161)
DBP (mm Hg)	77(71, 83)	75 (68, 81)	84(75, 91)	79(72, 86)	78(71, 86)
Waist (cm)	96(88, 107)	94(87, 102)	102 (95, 107)	90(83, 96)	100 (93, 106)
HbA1c (%)	5.8 (5.5, 6.0)	5.8 (5.5, 6.1)	5.6 (5.3, 6.3)	5.9 (5.7, 6.2)	6.0 (5.6, 6.5)
Triglyceride (mg/dL)	88(60, 136)	122 (91, 159)	130 (106, 165)	140 (106, 201)	135 (97, 187)
HDL (mg/dL)	54(45, 66)	56(46, 67)	47 (41, 52)	48(39, 57)	52 (47, 61)

DBP, diastolic blood pressure; HbA1c, glycosylated haemoglobin; HDL, high-density lipoprotein; MetS, metabolic syndrome; SBP, systolic blood pressure.

HDL concentration. Other partitioning variables for one or two of the MetS components included: weekly walking time, current employment status, perceived income sufficiency, living arrangements, smoking, adult occupation and current income.

DISCUSSION

The MOB technique identified distinct clusters of individuals with differential probabilities of MetS and its components, according to multiple social and behavioural risk factors. For the syndrome as a whole, in clusters of women from middle-income sites, the predicted proportion with MetS was quite high (58% or 68% depending on the cluster). In clusters of men, the predicted proportion with MetS was lower (26%, 38% or 41% depending on the cluster) and highest among men reporting childhood social adversities (41%). MetS in women from the Canadian sites varied considerably based on average walking time per week. Women from Kingston and St. Hyacinthe who walked minimally (>11 min/week) had predicted probabilities of MetS identical to men with postsecondary education and no childhood social adversities. This work demonstrates the potential of using MOB to identify joint effects in a moderately sized sample of individuals. It raises questions for future investigation, especially related to the concentration of risk(s) in certain subgroups.

This study corroborates previous findings that the prevalence of MetS varies according to age, sex and socioeconomic status.[4 10 29] Consistent with other studies, overall, we observed a concentration of MetS in participants of lower socioeconomic status[10 30 31]; although, among women from the middle-income sites, MetS was more prevalent among those with postsecondary education. Better educated women from the middle-income sites likely had/have more money to afford obesogenic, westernised foods and over their lifetimes, may have engaged in less exercise, as educational attainment may have allowed them to 'escape' physically strenuous jobs. Consistent with other research,[32] early life adversity was associated to higher prevalence estimates of MetS; however, this association was only observed in men, whereas it has been observed in both men and women elsewhere.[32] The strong context specificity of our findings highlights the utility of using MOB to identify unique admixtures that might have been overlooked with tradiational statistical techniques and/or would have been impossible to identify without a very large sample size.

When applying MOB, we observed distinct risk clusters according to study site and participant sex, education and childhood adversity. These findings support a dynamic interplay between contextual and social risk factors and the concentration of risks in certain subgroups, which is consistent with the notion of vulnerable populations proposed by Frohlich and Potvin.[3] Accordingly, vulnerable populations are defined by shared social characteristics that put them at 'higher risk of risks' (Frohlich and
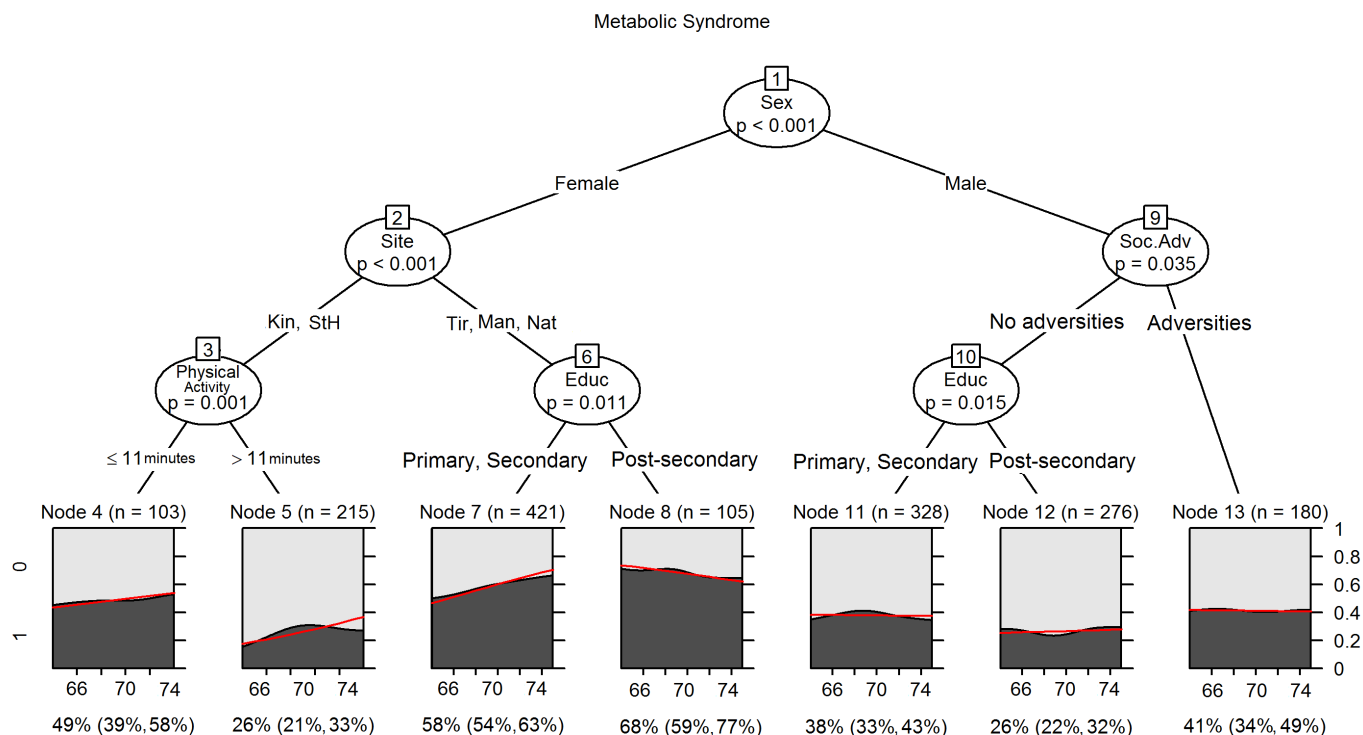
Figure 1 Model based recursive partitioning for MetS controlling for age. The horizontal axis of the terminal plots is age (64–75 years), and the vertical axis shows the predicted mean proportions of MetS obtained from logistic regression models by age. The predicted mean proportion of MetS and 95% CI for each terminal node are listed under the plots. MetS, metabolic syndrome.

Potvin, p218).[3] These risks and their accumulation across the life course relate to fundamental causes linked to one's social position within the predominant social structure.[3] This may explain why study site and sex were key partitioning variables. Study site proxies societal opportunities for education, occupation and income and expectations surrounding behaviours and diet. The clustering by site supports research underscoring the importance of context in patterning the risk exposures of individuals.[3 33] Sex/gender likely underpins access to resources such as money, knowledge and power affecting health outcomes through multiple risk factors.[34]

In studies with large sample sizes (>10 000 participants), complex joint effects have been observed with traditional regression analysis techniques. For example, using data from representative samples of US adults aged 25 and older, Loucks et al reported that overall, the prevalence of MetS was similar in women and men, both low education and poverty were associated with MetS, and the social gradient of the prevalence of MetS was more pronounced in women than in men.[11] These US findings differ from ours since in IMIAS, sex was the first stratifying variable, with an overall higher frequency of MetS in women than in men. However, for both men and women in our study, education was an important predictor of MetS at most sites; although, the direction of the association between education and MetS was not consistent. Among women from the middle-income sites, greater educational attainment

was associated with a higher predicted prevalence of MetS. Among men with no childhood social adversity, low educational attainment was associated to a higher predicted prevalence of MetS. In the study by Loucks et al, education was generally not associated to MetS in men; although, their study did not consider childhood adversity experiences,[10 11] in contrast to our own work. Interestingly, in the Loucks et al study, among men, low education was associated with the MetS components of abdominal obesity, hypertension and hyperglycaemia.[11]

When MOB was applied to the syndrome components, only study site and participant sex were consistent partitioning variables for most components. In general, when measures of socioeconomic status were partitioning variables, lower status was associated with poorer outcomes. For example, income insufficiency predicted higher mean DBP among Tirana participants. Measures of socioeconomic status appeared as partitioning variables more frequently than risk behaviours. This is consistent with recent work analysing individual-level data from >1.7 million people in which low socioeconomic status was associated with premature mortality across multiple disease categories and ranked third in population attributable fraction for mortality among a large list of risk factors (physical inactivity ranked second).[35] In our study, higher mean weekly physical activity predicted lower waist circumferences and higher HDL concentrations, consistent with the literature.[36–38]

This study has strengths. First is the use of the MOB technique. Traditional CART methods have the vulnerability of over fitting, selection bias and no concept of statistical significance. Thus pruning and cross-validation methods are used to avoid the over-fitting problems characteristic of CART. MOB is implemented via hypothesis tests, which leads to regression models whose predictive performance is equivalent to optimally pruned trees, therefore offering an intuitive and computationally efficient solution to the over-fitting problem, and the resulting models are easier to communicate to practitioners.[27 28 39] Finally, this is one of the very few studies that apply MOB to examine social and behavioural risk clusters for disease. Most research applying recursive partitioning focuses on identifying patient subgroups within a clinical setting[26] and/or how to better define components that constitute syndromes such as MetS.[39]

This study has limitations. Study site and sex are structuring variables that may mask more proximal risk factors for disease. Individual risk behaviours, such as smoking, may be ubiquitous within certain subgroups,[33] rendering it difficult to detect the influence of these behaviours on MetS using the MOB technique. Another limitation is that we used a single waist circumference cut-off value for all populations. Arguments exist for country and/or ethnicity-specific waist circumference cut-offs, but more work is required to optimally determine these.[40] Finally, we did not collect individual dietary data or data on the early nutritional environment.[32]

CONCLUSION

We applied a recursive partitioning technique to investigate risk clustering for MetS in an international, multisite study of community-dwelling older adults and observed unique risk clusters according to mostly contextual and socioeconomic characteristics. The main partitioning variables in our results were study site and sex, which for most people are not easily modifiable. However, the policies and opportunities afforded to residents of different communities and to men versus women can be modified and do vary dramatically on a global scale, which likely helps to explain the large variations in MetS prevalence across communities and even the inconsistencies observed across the literature in which women sometimes, but not always, have higher MetS prevalence. By identifying risk clusters with techniques such as MOB, we can generate novel hypotheses about both contributing and protective factors that might have been missed with traditional regression techniques, as relatively few studies have sufficient resources to recruit large enough samples for multiple order joint effects. MOB may also prove particularly informative in studies with much larger samples, such as the Health and Retirement Study, where it can be used to generate new hypotheses about risk clustering and then more traditional deterministic techniques can be applied to the same sample in order to corroborate or contradict these hypotheses. Finally, with regard to both clinical practice and health promotion activities, identifying risk clusters is important for targeting purposes, as the intensity and type of programmes may differ according to subgroups.[15]

REFERENCES

1. Halfon N, Hochstein M. Life course health development: an integrated framework for developing health, policy, and research. *Milbank Q* 2002;80:433–79.
2. Satariano WA. *Epidemiology of aging: an ecological approach*. Sudbury, MA: Jones and Bartlett Publishers, 2006.
3. Frohlich KL, Potvin L. Transcending the known in public health practice: the inequality paradox: the population approach and vulnerable populations. *Am J Public Health* 2008;98:216–21.
4. Samson SL, Garber AJ. Metabolic syndrome. *Endocrinol Metab Clin North Am* 2014;43:1–23.
5. Scuteri A, Najjar SS, Morrell CH, *et al*. The metabolic syndrome in older individuals: prevalence and prediction of cardiovascular events: the Cardiovascular Health Study. *Diabetes Care* 2005;28:882–7.
6. Liu M, Wang J, Jiang B, *et al*. Increasing Prevalence of Metabolic Syndrome in a Chinese Elderly Population: 2001-2010. *PLoS One* 2013;8:e66233.
7. Riediger ND, Clara I. Prevalence of metabolic syndrome in the Canadian adult population. *CMAJ* 2011;183:E1127–34.
8. Davila EP, Quintero MA, Orrego ML, *et al*. Prevalence and risk factors for metabolic syndrome in Medellin and surrounding municipalities, Colombia, 2008-2010. *Prev Med* 2013;56:30–4.
9. Moreira GC, Cipullo JP, Ciorlia LA, *et al*. Prevalence of metabolic syndrome: association with risk factors and cardiovascular complications in an urban population. *PLoS One* 2014;9:e105056.
10. Wu HF, Tam T, Jin L, *et al*. Age, gender, and socioeconomic gradients in metabolic syndrome: biomarker evidence from a large sample in Taiwan, 2005-2013. *Ann Epidemiol* 2017;27:315–22.
11. Loucks EB, Rehkopf DH, Thurston RC, *et al*. Socioeconomic disparities in metabolic syndrome differ by gender: evidence from NHANES III. *Ann Epidemiol* 2007;17:19–26.
12. Bronfenbrenner U. *Ecological models of human development. International encyclopedia of education*. Elsevier: Oxford, 1994.
13. Breiman L, Friedman J, Alshen R, *et al*. *CART: Classification and Regression Trees*. Belmont, CA: Wadsworth, 1984.
14. Nelson LM, Bloch DA, Longstreth WT, *et al*. Recursive partitioning for the identification of disease risk subgroups: a case-control study of subarachnoid hemorrhage. *J Clin Epidemiol* 1998;51:199–209.

15. Nayak S, Hubbard A, Sidney S, *et al*. PP66 Self-rated health in the CARDIA Study: a recursive partitioning approach to contextualising health determinants. *J Epidemiol Community Health* 2014;68:A73.2–A74.
16. Gomez F, Wu YY, Auais M, *et al*. A simple algorithm to predict falls in primary care patients aged 65 to 74 Years: The International Mobility in Aging Study. *J Am Med Dir Assoc* 2017;18:774–9.
17. Costello TJ, Swartz MD, Sabripour M, *et al*. Use of tree-based models to identify subgroups and increase power to detect linkage to cardiovascular disease traits. *BMC Genet* 2003;4(Suppl 1):S66.
18. Stern SE, Williams K, Ferrannini E, *et al*. Identification of individuals with insulin resistance using routine clinical measurements. *Diabetes* 2005;54:333–9.
19. Zunzunegui MV, Alvarado BE, Guerra R, *et al*. The mobility gap between older men and women: the embodiment of gender. *Arch Gerontol Geriatr* 2015;61:140–8.
20. De Yébenes MJ, Otero A, Zunzunegui MV, *et al*. Validation of a short cognitive tool for the screening of dementia in elderly people with low educational level. *Int J Geriatr Psychiatry* 2003;18:925–36.
21. National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III). Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) final report. *Circulation* 2002;106:3143–421.
22. American Diabetes Association. Diagnosis and classification of diabetes mellitus. *Diabetes Care* 2010;33(Suppl 1):S62–9.
23. Sousa AC, Guerra RO, Thanh Tu M, *et al*. Lifecourse adversity and physical performance across countries among men and women aged 65-74. *PLoS One* 2014;9:e102299.
24. Marsh AP, Janssen JA, Ip EH, Eh I, *et al*. Assessing walking activity in older adults: development and validation of a novel computer-animated assessment tool. *J Gerontol A Biol Sci Med Sci* 2015;70:1555–61.
25. Stekhoven DJ, Bühlmann P. MissForest--non-parametric missing value imputation for mixed-type data. *Bioinformatics* 2012;28:112–8.
26. Seibold H, Zeileis A, Hothorn T. Model-based recursive partitioning for subgroup analyses. *Int J Biostat* 2016;12:45–63.
27. Zeileis A, Hothorn T, Hornik K. Model-based recursive partitioning. *J Comput Graph Stat* 2008;17:492–514.
28. Hothorn T, Hornik K, Zeileis A. Unbiased recursive partitioning: a conditional inference framework. *J Comput Graph Stat* 2006;15:651–74.
29. Eckel RH, Grundy SM, Zimmet PZ. The metabolic syndrome. *Lancet* 2005;365:1415–28.
30. Loucks EB, Magnusson KT, Cook S, *et al*. Socioeconomic position and the metabolic syndrome in early, middle, and late life: evidence from NHANES 1999-2002. *Ann Epidemiol* 2007;17:782–90.
31. Wamala SP, Lynch J, Horsten M, *et al*. Education and the metabolic syndrome in women. *Diabetes Care* 1999;22:1999–2003.
32. Delpierre C, Fantin R, Barboza-Solis C, *et al*. The early life nutritional environment and early life stress as potential pathways towards the metabolic syndrome in mid-life? A lifecourse analysis using the 1958 British Birth cohort. *BMC Public Health* 2016;16:815.
33. Rose G. Sick individuals and sick populations. *Int J Epidemiol* 2001;30:427–32.
34. Phelan JC, Link BG, Tehranifar P. Social conditions as fundamental causes of health inequalities: theory, evidence, and policy implications. *J Health Soc Behav* 2010;51:S28–40.
35. Stringhini S, Carmeli C, Jokela M, *et al*. Socioeconomic status and the 25 × 25 risk factors as determinants of premature mortality: a multicohort study and meta-analysis of 1·7 million men and women. *Lancet* 2017;389:1229–37.
36. Kokkinos PF, Fernhall B. Physical activity and high density lipoprotein cholesterol levels: what is the relationship? *Sports Med* 1999;28:307–14.
37. Skoumas J, Pitsavos C, Panagiotakos DB, *et al*. Physical activity, high density lipoprotein cholesterol and other lipids levels, in men and women from the ATTICA study. *Lipids Health Dis* 2003;2:3.
38. Recio-Rodríguez JI, Sanchez-Aguadero N, Rodríguez-Sánchez E, *et al*. Physical Activity and Adiposity Among Older Adults of the EVIDENT Study. *J Aging Phys Act* 2017;25:254–60.
39. Girman CJ, Dekker JM, Rhodes T, *et al*. An exploratory analysis of criteria for the metabolic syndrome and its prediction of long-term cardiovascular outcomes: the Hoorn study. *Am J Epidemiol* 2005;162:438–47.
40. Alberti KG, Eckel RH, Grundy SM, *et al*. Harmonizing the metabolic syndrome: a joint interim statement of the International Diabetes Federation Task Force on Epidemiology and Prevention; National Heart, Lung, and Blood Institute; American Heart Association; World Heart Federation; International Atherosclerosis Society; and International Association for the Study of Obesity. *Circulation* 2009;120:1640–5.