

Supporting information

Population socioeconomics predicted using wastewater

Phil M. Choi,¹ Jake W. O'Brien,¹ Ben J. Tschärke,¹ Jochen F. Mueller,¹ Kevin V. Thomas,¹
Saer Samanipour^{1,2,3*}

¹ Queensland Alliance for Environmental Health Sciences, The University of Queensland,
Brisbane, QLD 4102 Australia

² Norwegian Institute for Water Research, 0349 Oslo, Norway

³ Van 't Hoff Institute for Molecular Sciences, University of Amsterdam, The Netherlands

*Corresponding author, s.samanipour@uva.nl

Table S1. Socioeconomic index for area (SEIFA) features used in the present study. Details regarding variable composition can be found elsewhere¹. Adapted with changes from Choi, P. M. et al. Social, demographic, and economic correlates of food and chemical consumption measured by wastewater-based epidemiology. PNAS 116, 21864-21873, (2019). Copyright © 2019 the Author(s). Published by PNAS.

Features	IRSAD weighting	Definition
ATSCHOOL	NA	% People aged 15 years and over who are still attending secondary school
ATUNI	0.36	% People aged 15 years and over at a university or other tertiary institution
CERTIFICATE	-0.36	% People aged 15 years and over whose highest level of education is a certificate III or IV (mid-high level vocational) qualification
DEGREE	NA	% People aged 15 years and over whose highest level of education is a bachelor degree or higher
DIPLOMA	0.5	% People aged 15 years and over whose highest level of education is an advanced diploma or diploma
NOEDU	-0.34	% People aged 15 years and over who have no educational attainment
NOYEAR12ORHIGHER	-0.85	% People aged 15 years and over whose highest level of education is year (grade) 11 or lower
CHILDJOBLESS	-0.76	% Families with children under 15 years of age and jobless parents
DISABILTYU70	-0.69	% People aged under 70 who need assistance with core activities
ENGLISHPOOR	NA	% People who do not speak English well or not at all
FEWBED	NA	% Classifiable occupied private dwellings with one or no bedrooms
HIGHBED	0.44	% Occupied private dwellings with four or more bedrooms
GROUP	NA	% Occupied private dwellings that are group occupied private dwellings
LONE	NA	% Occupied private dwellings that are lone person occupied private dwellings
HIGHCAR	NA	% Occupied private dwellings with three or more cars
NOCAR	-0.33	% Occupied private dwellings with no cars
HIGHMORTGAGE	0.72	% Occupied private dwellings paying more than \$2,800 per month in mortgage
HIGHRENT	0.47	% Occupied private dwellings paying more than \$470 per week in rent
LOWRENT	-0.64	% Occupied private dwellings paying less than \$215 per week in rent (excluding \$0 per week)
INC_LOW	-0.89	% People with stated annual household equalized income between \$1 and \$25,999 (approx. 1st and 2nd deciles)
INC_HIGH	0.83	% People with stated annual household equalized income greater than \$78,000 (approx. 9th and 10th deciles)
NONET	-0.78	% Occupied private dwellings with no internet connection
OCC_DRIVERS	-0.62	% Employed people classified as machinery operators and drivers
OCC_LABOUR	-0.79	% Employed people classified as laborers
OCC_MANAGER	0.47	% Employed people classified as managers
OCC_PROF	0.71	% Employed people classified as professionals
OCC_SKILL5	NA	Employed people working in a Skill Level 5 occupation (commensurate with compulsory secondary education, vocational Certificate I, or a short period of on-the-job training)
OCC_SKILL4	NA	Employed people working in a Skill Level 4 occupation (commensurate with vocational Certificate II or III or at least one year of relevant experience)
OCC_SKILL2	NA	Employed people working in a Skill Level 2 occupation (commensurate with AQF Associate Degree, Advanced Diploma or Diploma, or at least three years of relevant experience)
OCC_SKILL1	NA	Employed people working in a Skill Level 1 occupation (commensurate with bachelor degree or higher qualification, or at least five years of relevant experience)
ONEPARENT	-0.65	% Families that are one parent families with dependent offspring only
OWNING	NA	% Occupied private dwellings owning the dwelling they occupy without a mortgage
MORTGAGE	NA	% Occupied private dwellings owning the dwelling they occupy with a mortgage
SEPDIVORCED	-0.6	% People aged 15 and over who are separated or divorced
UNEMPLOYED	-0.66	% People (in the labor force) who are unemployed
UNEMPLOYED1	NA	% People aged 15 years and over who are unemployed
UNINCORP	NA	% Owner of an unincorporated enterprise

Table S2. Biomarkers used in the present study.

Biomarker	Category	Metabolite/indicator of
Hydroxycotinine	licit drug	cotinine
Cotinine	licit drug	tobacco use
Nicotine	licit drug	tobacco use
Paraxanthine	licit drug	caffeine
Caffeine	licit drug	itself
Amphetamine	illicit drug	amphetamine and methamphetamine consumption
Methamphetamine	illicit drug	itself
Methadone	pharmaceutical	itself
EDDP	pharmaceutical	methadone
Codeine	pharmaceutical	itself
Morphine	pharmaceutical	morphine and heroin consumption
Ethanol	licit drug	alcohol consumption
Oxycodone	pharmaceutical	itself
Noroxycodone	pharmaceutical	oxycodone
Tramadol	pharmaceutical	itself
Desvenlafaxine	pharmaceutical	venlafaxine and desvenlafaxine consumption
Venlafaxine	pharmaceutical	itself
Citalopram	pharmaceutical	itself
Mirtazapine	pharmaceutical	itself
Amitriptyline	pharmaceutical	itself
Carbamazepine	pharmaceutical	itself
Gabapentin	pharmaceutical	itself
Pregabalin	pharmaceutical	itself
Ibuprofen	pharmaceutical	itself
Naproxen	pharmaceutical	itself
Fexofenadine	pharmaceutical	itself
Cetirizine	pharmaceutical	itself
Atenolol	pharmaceutical	itself
Hydrochlorthiazide	pharmaceutical	itself
Cephalexin	pharmaceutical	itself
Sulphamethoxazole	pharmaceutical	itself
Trimethoprim	pharmaceutical	itself
Acesulfame	artificial sweetener	itself
Saccharin	artificial sweetener	itself
2PY	dietary metabolite	vitamin B3 consumption
4PY	dietary metabolite	vitamin B3 consumption
4-Pyridoxic acid	dietary metabolite	vitamin B6 consumption
Proline betaine	dietary metabolite	citrus
Enterodiol	dietary metabolite	dietary fiber consumption
Enterolactone	dietary metabolite	dietary fiber consumption

EDDP: 2-ethylidene-1,5-dimethyl-3,3-diphenylpyrrolidine. 2PY: N1-methyl-2-pyridone-5-carboxamide. 4PY: N1-methyl-4-pyridone-3-carboxamide.

Table S3. Sampling characteristics for the WWTPs in the present study.

Dataset	WWTP ¹	Sample composite basis	Composite frequency (min)	Composite frequency (ML)	Compositing events per 24h ²
Test	1	Time		0.04	85
Test	2	NP			
Test	3	Time	15		96
Test	4	Time	15		96
Test	5	Flow		0.01	2000
Test	6	"intermittent"			
Test	7	Time	15		96
Test	8	Time	15		96
Test	9	Time	15		96
Test	10	Time	60		24
Test	11	Time	15		96
Test	12	Time	60		24
Test	13	Time	15		96
Test	14	Flow		2	62
Test	15	Time	15		96
Test	16	Time	60		24
Test	17	Flow		NP	
Test	18	Time	15		96
Test	19	Flow		NP	
Test	20	Time	15		96
Test	21	Time	60		
Test	22	Time	60		
Test	23	Time	60		24
Training	24	Time	15		96
Training	25	Time	NP		
Training	26	Time	15		96
Training	27	Flow		0.05	98
Training	28	Time	15		96
Training	29	Time	60		24
Training	30	Time	15		
Training	31	Flow		0.05	67
Training	32	Time	60		24

NP: not provided by WWTP operators.

¹ For privacy reasons, these identifiers have been designated randomly and do not reflect those in Tables S4 and S5 (Excel file).

² For flow based composite samples, the average 24h influent volume was used to calculate this metric.

Please refer to the associated Excel file for Tables S4 and S5.

Table S6. Socioeconomic variables with entries in the training dataset as zero values or values between zero and one percent. No other socioeconomic features in the present study had entries below one percent. Refer to Table S1 for feature definitions.

Feature	NOEDU	ENGLISHPOOR	GROUP	HIGHRENT	LOWRENT
Entries as 0%	18%	8%	none	5%	none
Entries between 0 and 1%	68%	50%	5%	35%	8%

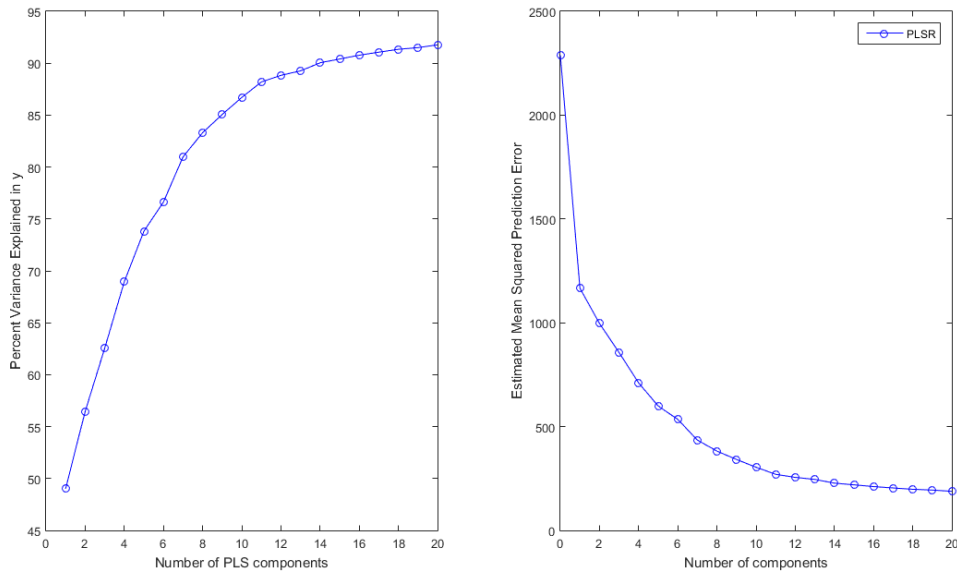


Figure S1. Cumulative percentage of variance in socioeconomics and estimated mean squared prediction error derived from partial least squared model components.

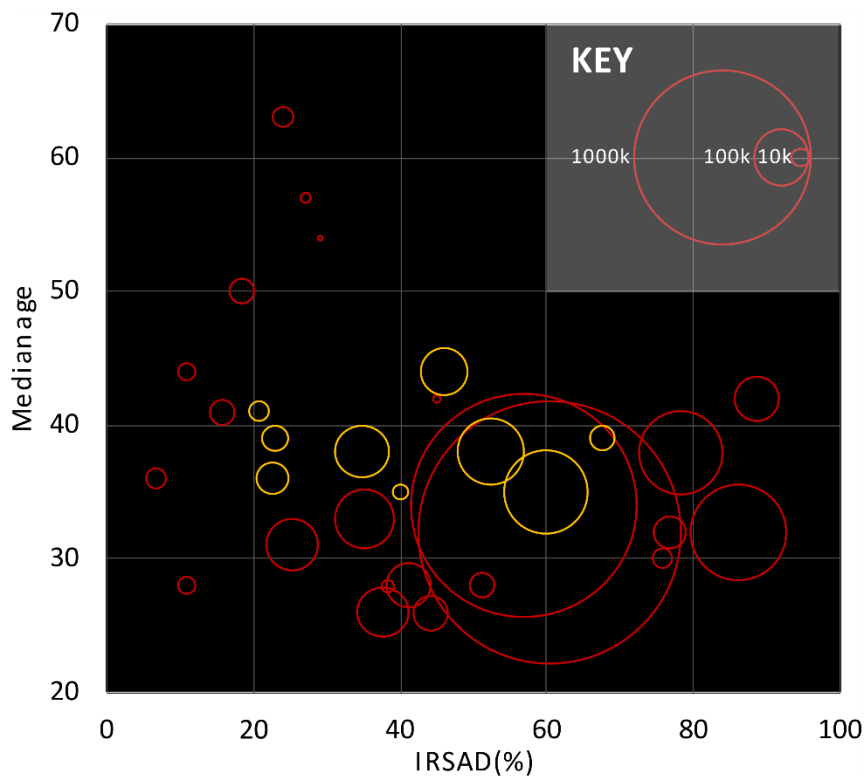


Figure S2. Wastewater treatment plant (WWTP) populations featured in this study, expressed in terms of IRSAD (index of relative socioeconomic advantage and disadvantage), median age and population size (circle area). Populations in the training dataset are shown in red. Populations from the test dataset are shown in yellow. Adapted with changes from Choi, P. M. et al. Social, demographic, and economic correlates of food and chemical consumption measured by wastewater-based epidemiology. *PNAS* 116, 21864-21873, (2019). Copyright © 2019 the Author(s). Published by PNAS.

References

- 1 Australian Bureau of Statistics (ABS). Census of Population and Housing: Socio-Economic Indexes for Areas (SEIFA), Australia, 2016. Report No. 2033.0.55.001, (© Commonwealth of Australia, 2018).