Analyzing the interplay between societal trends and socio-demographic variables with local pattern mining: Discovering exceptional trends in adolescent alcohol use in the Netherlands – Supplementary material

Rianne M. Schouten[1] (✉), Gonneke W.J.M. Stevens[2], Saskia A.F.M. van Dorsselaer[3], Elisa L. Duinhof[3], Karin Monshouwer[3], Mykola Pechenizkiy[1], and Wouter Duivesteijn[1]

[1] Eindhoven University of Technology, Data Mining Group, Eindhoven, the Netherlands, {r.m.schouten,m.pechenizkiy,w.duivesteijn}@tue.nl
[2] Utrecht University, Department of Interdisciplinary Social Science, Utrecht, the Netherlands, g.w.j.m.stevens@uu.nl
[3] National Institute of Mental Health and Addiction, Utrecht, the Netherlands, {sdorsselaer,eduinhof,kmonshouwer}@trimbos.nl

Abstract. Over the last two decades, alcohol use has been in decline among Dutch adolescents. However, the declining trend has been flatlining: prevalence of monthly alcohol use among Dutch 12-to-16-year-olds decreased from 54% in 2003 to 26% in 2013, but merely to 23% in 2019. Dutch governmental policy makers aim to decrease this prevalence further. To do so effectively, it would benefit them to know whether social group memberships correspond to exceptional alcohol use trends. With traditional statistical approaches, it is challenging to analyze such a relation between societal trends and social group memberships: only a few socio-demographic variables can be included, subgroups must be pre-defined, and linearity assumptions are required. We resolve these issues and automatically identify social subgroups of the Dutch adolescent population by deploying Exceptional Model Mining for Repeated Cross-Sectional data (EMM-RCS) on data that interleaves two quadrennial studies: the Health Behaviour in School-Aged Children study (HBSC), and the Dutch National School Survey on Substance Use (DNSSSU). Our findings confirm existing knowledge that age, educational level, and migration background are important descriptors of monthly alcohol use, and provide further insights into the existence of an interplay effect with life satisfaction, urbanization degree, and truancy.

Keywords: Adolescent Alcohol Use · Exceptional Model Mining · Intersectionality · Local Pattern Mining · Validation of Subgroups · Trend Analysis

Table 1: Demographics and descriptive statistics of all descriptive attributes.

Survey	DNSSSU	HBSC	DNSSSU	HBSC	DNSSSU	HBSC	DNSSSU	HBSC	DNSSSU
Year	2003	2005	2007	2009	2011	2013	2015	2017	2019
Total *n(%)*	6791 (12.8)	5272 (10.0)	6234 (11.8)	5490 (10.4)	6374 (12.1)	5421 (10.2)	6232 (11.8)	6060 (11.5)	5022 (09.5)
Gender *n(%)*									
Boy	3165 (46.6)	2609 (49.5)	3049 (48.9)	2686 (48.9)	3250 (51.0)	2668 (49.2)	3172 (50.9)	2950 (48.7)	2603 (51.8)
Girl	3626 (53.4)	2663 (50.5)	3185 (51.1)	2804 (51.1)	3124 (49.0)	2753 (50.8)	3060 (49.1)	3110 (51.3)	2415 (48.1)
Missing	0 (00.0)	0 (00.0)	0 (00.0)	0 (00.0)	0 (00.0)	0 (00.0)	0 (00.0)	0 (00.0)	4 (00.1)
Ethnic group *n(%)*									
Dutch	5370 (79.1)	4041 (76.7)	5036 (80.8)	4478 (81.6)	5079 (79.7)	4266 (78.7)	4963 (79.6)	4751 (78.4)	3706 (73.8)
Non-Western	1001 (14.7)	911 (17.3)	796 (12.8)	770 (14.0)	891 (14.0)	838 (15.5)	924 (14.8)	1001 (16.5)	933 (18.6)
Western	416 (06.1)	300 (05.7)	376 (06.0)	239 (04.4)	388 (06.1)	315 (05.8)	337 (05.4)	307 (05.1)	368 (07.3)
Missing	4 (00.1)	20 (00.4)	26 (00.4)	3 (00.1)	16 (00.3)	2 (00.0)	8 (00.1)	1 (00.0)	15 (00.3)
Does father have a job *n(%)*									
Yes	6142 (90.4)	4554 (86.4)	5715 (91.7)	4731 (86.2)	5768 (90.5)	4384 (80.9)	5615 (90.1)	5321 (87.8)	4566 (90.9)
No	423 (06.2)	324 (06.1)	292 (04.7)	262 (04.8)	320 (05.0)	355 (06.5)	369 (05.9)	278 (04.6)	237 (04.7)
Don't know/no contact	145 (02.1)	282 (05.3)	156 (02.5)	224 (04.1)	238 (03.7)	289 (05.3)	237 (03.8)	291 (04.8)	209 (04.2)
Missing	81 (01.2)	112 (02.1)	71 (01.1)	273 (05.0)	48 (00.8)	393 (07.2)	11 (00.2)	170 (02.8)	10 (00.2)
Does mother have a job *n(%)*									
Yes	4887 (72.0)	3723 (70.6)	4763 (76.4)	4039 (73.6)	5768 (90.5)	3970 (73.2)	4988 (80.0)	4783 (78.9)	4175 (83.1)
No	1696 (25.0)	1327 (25.2)	1350 (21.7)	1086 (19.8)	320 (05.0)	952 (17.6)	1177 (18.9)	1006 (16.6)	770 (15.3)
Don't know/no contact	46 (00.7)	103 (02.0)	42 (00.7)	83 (01.5)	238 (03.7)	94 (01.7)	63 (01.0)	100 (01.7)	70 (01.4)
Missing	162 (02.4)	119 (02.3)	79 (01.3)	282 (05.1)	48 (00.8)	405 (07.5)	4 (00.1)	171 (02.8)	7 (00.1)
Lives with both parents *n(%)*									
Yes	5354 (78.8)	4121 (78.2)	4922 (79.0)	4397 (80.1)	4756 (74.6)	3998 (73.8)	4676 (75.0)	4611 (76.1)	3669 (73.1)
No	1326 (19.5)	1093 (20.7)	1264 (20.3)	1089 (19.8)	1600 (25.1)	1423 (26.2)	1553 (24.9)	1442 (23.8)	1346 (26.8)
Missing	111 (01.6)	58 (01.1)	48 (00.8)	4 (00.1)	18 (00.3)	0 (00.0)	3 (00.0)	7 (00.1)	7 (00.1)
School level *n(%)*									
VMBO-p/t	1927 (28.4)	1384 (26.3)	1532 (24.6)	995 (18.1)	1543 (24.2)	1337 (24.7)	1370 (22.0)	1116 (18.4)	1082 (21.5)
VMBO-t/HAVO	2130 (31.4)	1653 (31.4)	2059 (33.0)	1982 (36.1)	2081 (32.6)	1486 (27.4)	2147 (34.5)	1973 (32.6)	1628 (32.4)
HAVO/VWO	1542 (22.7)	1271 (24.1)	1474 (23.6)	1247 (22.7)	1294 (20.3)	1418 (26.2)	1566 (25.1)	1402 (23.1)	1244 (24.8)
VWO	1086 (16.0)	939 (17.8)	1169 (18.8)	1266 (23.1)	1436 (22.5)	1180 (21.8)	1149 (18.4)	1569 (25.9)	1068 (21.3)
Missing	106 (01.6)	25 (00.5)	0 (00.0)	0 (00.0)	20 (00.3)	0 (00.0)	0 (00.0)	0 (00.0)	4 (00.1)
Skipped classes *n(%)*									
0 hours	5817 (85.7)	4470 (84.8)	5395 (86.5)	4645 (84.6)	5716 (89.7)	4649 (85.8)	5716 (91.7)	5213 (86.0)	4433 (88.3)
1 hour	285 (04.2)	281 (05.3)	271 (04.3)	236 (04.3)	242 (03.8)	213 (03.9)	204 (03.3)	298 (04.9)	209 (04.2)
2 hours	196 (02.9)	176 (03.3)	206 (03.3)	170 (03.1)	143 (02.2)	110 (02.0)	101 (01.6)	174 (02.9)	131 (02.6)
3-4 hours	163 (02.4)	131 (02.5)	153 (02.5)	101 (01.8)	111 (01.7)	64 (01.2)	84 (01.3)	112 (01.8)	106 (02.1)
5-6 hours	73 (01.1)	48 (00.9)	79 (01.3)	28 (00.5)	51 (00.8)	26 (00.5)	41 (00.7)	49 (00.8)	50 (01.0)
7 or more hours	149 (02.2)	86 (01.6)	97 (01.6)	64 (01.2)	68 (01.1)	49 (00.9)	78 (01.3)	92 (01.5)	89 (01.8)
Missing	108 (01.6)	80 (01.5)	33 (00.5)	246 (04.5)	43 (00.7)	310 (05.7)	8 (00.1)	122 (02.0)	4 (00.1)
Urbanity degree *n(%)*									
Very high	1021 (15.0)	729 (13.8)	1033 (16.6)	731 (13.3)	1234 (19.4)	1070 (19.7)	1143 (18.3)	957 (15.8)	931 (18.5)
High	1912 (28.2)	1304 (24.7)	1778 (28.5)	1351 (24.6)	1881 (29.5)	1422 (26.2)	1842 (29.6)	2049 (33.8)	1587 (31.6)
Moderate	1219 (18.0)	1081 (20.5)	1234 (19.8)	1089 (19.8)	1048 (16.4)	801 (14.8)	821 (13.2)	978 (16.1)	706 (14.1)
Low	1604 (23.6)	1047 (19.9)	1278 (20.5)	1070 (19.5)	1495 (23.5)	1480 (27.3)	1669 (26.8)	1547 (25.5)	1289 (25.7)
None	1035 (15.2)	1111 (21.1)	911 (14.6)	1249 (22.8)	716 (11.2)	648 (12.0)	757 (12.1)	529 (08.7)	509 (10.1)
Missing *n(%)*	0 (00.0)	0 (00.0)	0 (00.0)	0 (00.0)	0 (00.0)	0 (00.0)	0 (00.0)	0 (00.0)	0 (00.0)
Age (12-16)									
Observed *M(SD)*	13.8 (1.2)	13.8 (1.2)	13.9 (1.2)	13.8 (1.3)	13.8 (1.2)	13.8 (1.2)	13.8 (1.2)	13.7 (1.3)	13.7 (1.2)
Missing *n(%)*	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Life satisfaction (1-10)									
Observed *M(SD)*	7.7 (1.6)	7.8 (1.5)	7.7 (1.5)	7.9 (1.4)	7.7 (1.5)	7.6 (1.6)	7.8 (1.5)	7.6 (1.6)	7.6 (1.6)
Missing *n(%)*	39 (0.6)	33 (0.6)	52 (0.8)	94 (1.7)	31 (0.5)	205 (3.8)	0 (0.0)	6 (0.1)	1 (0.0)

Table 2: Proportion and descriptions of top-17 validated subgroups of adolescents with exceptional deviations of the prevalence of monthly alcohol use. Subgroup descriptions are conjunctions of conditions 1, 2 and 3 (in that order). The prevalence trends can be clustered into five main trend groups.

TG	SG	Cov	Description		
			condition 1	condition 2	condition 3
1	1	0.11	age: 12	skipped classes: 0	urbanity: at least moderate
	2	0.15	age: 12	life satisf: 7-10	skipped classes: 0
	3	0.14	age: 12	life satisf: 7-10	
	10	0.09	age: 12	skipped classes: 0	
2	4	0.35	age: 12-13	skipped classes: 0	life satisf: 7-10
	5	0.37	age: 12-13	skipped classes: 0	life satisf: 6-10
	7	0.40	age: 12-13	skipped classes: 0	
	9	0.25	age: 12-13	life satisf: 6-10	urbanity: at least moderate
	12	0.37	age: 12-13	life satisf: 7-10	
	14	0.40	age: 12-13	life satisf: 6-10	
	16	0.41	age: 12-13	skipped classes: 0-1	
	17	0.43	age: 12-13		
3	6	0.26	age: 15-16	ethnicity: dutch, western	
	8	0.24	age: 15-16	ethnicity: dutch	
4	11	0.48	age: 14-16	ethnicity: dutch, western	
	15	0.44	age: 14-16	ethnicity: dutch	
5	13	0.32	age: 15-16		

Table 3: Proportion and descriptions of top-14 validated subgroups of adolescents with exceptional deviations in the course of the trend in monthly alcohol use (i.e. exceptional slope deviations). Subgroup descriptions are conjunctions of conditions 1 and 2 (in that order); there are no subgroups with 3 conditions. The trends can be clustered into five main trend groups.

TG	SG	Cov	Description		
			condition 1	condition 2	condition 3
1	1	0.19	age: 12		
	4	0.15	age: 12-13	life satisf: 9-10	
	5	0.29	age: 12-13	life satisf: 8-10	
	8	0.37	age: 12-13	life satisf: 7-10	
	10	0.54	age: 12-14	ethnicity: dutch	
	14	0.53	age: 12-14	mother job: yes	
2	2	0.1	urbanity: very high	age: 14-16	
	7	0.14	urbanity: very high	life satisf: 0-9	
	13	0.14	urbanity: very high	age: 13-16	
3	3	0.26	age: 15-16	school level: at least vmbo-p/t	
	9	0.32	age: 15-16		
	12	0.11	skipped classes: ≥ 1		
4	6	0.23	school level: at least havo	urbanity: at most moderate	
5	11	0.13	school level: vmbo-p/t - havo	ethnicity: western, non-western	

Table 4: Proportion and descriptions of top-10 validated subgroups of adolescents with exceptionally horizontal trends in monthly alcohol use. Subgroup descriptions are conjunctions of conditions 1, 2 and 3 (in that order). The subgroups are clustered into four main trend groups.

TG	**SG**	**Cov**	**Description**		
			condition 1	condition 2	condition 3
1	1	0.09	ethnicity: non-western	life satisf: 0-8	skipped classes: 0-2
	3	0.09	ethnicity: non-western	life satisf: 0-8	skipped classes: 0-4
	4	0.08	ethnicity: non-western	life satisf: 0-8	school level: $\leq$ havo/vwo
	6	0.08	ethnicity: non-western	age: 13-16	school level: $\geq$ vmbo-t
2	2	0.08	ethnicity: non-western	age: 14-16	urbanity: $\geq$ moderate
	9	0.09	ethnicity: non-western	age: 14-16	urbanity: $\geq$ little
	10	0.08	ethnicity: non-western	age: 14-16	skipped classes: 0-2
3	5	0.10	ethnicity: non-western	complete family: yes	skipped classes: 0-4
	8	0.10	ethnicity: non-western	complete family: yes	father job: yes,no
4	7	0.11	school level: $\geq$ havo/vwo	ethnicity: (non)-western	life satisf: 0-8