Objective Monitoring of the Obesogenic Behaviour and the Role of the Local Environment. New methods for collecting evidence

Anastasios DELOPOULOS

Associate Professor

Multimedia Understanding Group, Aristotle University of Thessaloniki

Greece

Abstract. The way we eat and what we eat, the way we move and the way we sleep significantly impact the risk of becoming obese. These three aspects of behavior decompose into a long list of personal behavioural elements including our food choices, eating place preferences, transportation choices, sleeping periods and duration etc. Most of these elements are highly correlated in a causal way with the conditions of our local urban, social, regulatory and economic environment. In this presentation we will examine technological solutions that have been developed in order to (a) objectively monitor a matrix of obesogenic behavioural elements, (b) acquire information related to the local environment conditions. The first rely mostly on signals captured by very simple wearable devices (accelerometers, gyroscopes, GPS) embedded in smart phones and smart watches while the latter resort to public sources, like maps, and to the data of statistical authorities. Based on this infrastructure, we are ready to link the obesogenic behaviours with the local environment; this creates a wealth of evidence for medical experts, epidemiologists, urban designers, obesity experts and public health authorities. Apart of the acquisition technologies that will be presented, we will investigate (i) The scalability of population engagement approaches that rely on the citizen scientist concept in order to collect behavioural and use of urban resources data in an efficient and still accurate and representative way. (ii) How these data (in raw and processed form) can be standardized and be communicated to interested scientist (statisticians, epidemiologists, urban designers, etc) and community stakeholders without breaching privacy. (iii) How sensory data and their derivatives can be converted and/or combined with non aggregated microdata; benefits and constraints of this approach.

1. Introduction

On average, obesity affects one in every three children aged six to nine years in Europe [1], with a particularly high rate in children of families with low socioeconomic status [2]. The problem with childhood obesity is clear: children who are obese are more likely to stay obese into adulthood, which puts them at increased risk for non-communicable diseases (NCDs). Moreover, the slow but continuous raise in the prevalence in the last forty years [3] jeopardizes the sustainability of our health systems.

The World Health Organization's (WHO) Commission on Ending Childhood Obesity has recently released a comprehensive report [4] outlining a high-level set of recommendations to tackle the childhood obesity epidemic, grouped into 6 broad categories: (i) promoting healthy foods, (ii) promoting physical activity (PA), (iii) preconception and pregnancy care, (iv) early childhood, (v) school-aged children, and (vi) weight management for overweight and obese children. Cross-cutting through all recommendations is the need for "robust monitoring and accountability systems", which "are vital in providing data for policy development and in offering evidence of the impact and effectiveness of interventions". Furthermore, the report recognizes that successful measures should address the entire obesogenic environment. There are several difficulties towards implementing these recommendations. Individuals and their behavioural choices are situated within and influenced by their broader social and environmental context [5], which consists of a complex array of local external factors [6], like community, demographic and socioeconomic characteristics. Measures that combine multiple strategies that modify the obesogenic environment may improve the dietary and sleeping habits, increase physical activity and reduce sedentariness. Such interventions can be successful [7], if they are evidence-based and context-specific [8].

2. The concept

The primary goal of the examined methodology, which is currently being developed within the H2020 project "BigO: Big Data Against Childhood Obesity" (bigoprogram.eu) is to create new sources of evidence together with exploration tools for the Public Health Authorities assisting their effort to tackle childhood obesity. The methodology includes the following functionalities:

1. The collection of Big Data (e.g., accelerometry, geolocation), using different technologies (smartphone, wristband) about the user's behavioural patterns. The BigO data pool will be analyzed, combined with various online data (e.g., maps, registries and GIS) referring to the Local Extrinsic Conditions of the local community (LECs).

2. The creation of comprehensive models of the obesity prevalence dependence matrix, through the association of the LECs, the community behavioural patterns, and the local Obesity prevalence. The targeted models will be used to

    a. Identify the most important obesogenic factors of the local environment. Although it is in general known what are the main conditions of the urban, the social, the regulatory and legal environment that negatively affect the obesogenic behavior, the examined methodology will be able to identify those that are prominent at a local level.

b. Simulate the effect of interventions to the obesogenic behaviours. Local authorities will have indication of the effectiveness of their counter-obesity measures before their actual implementation.

3. The visualisation of the acquired data and their relations is also part of the functionalities that facilitate the exploration of populations behaviours vs environment conditions.

3. Information Model

Information in BigO can be grouped into three layers which are (i) the raw data from sensors and external sources, (ii) behavioural indicators and LECs and (iii) models and analytics.

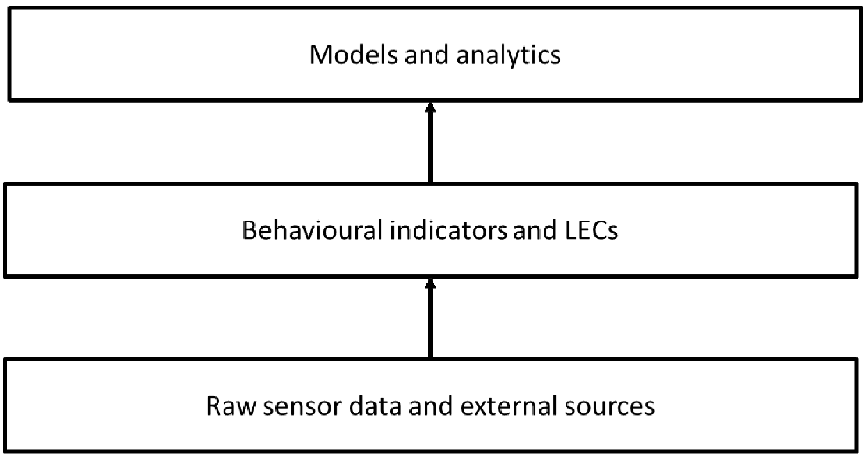


Figure 1: BigO layers of information

Raw data sources include mobile phone and smartwatch sensors (accelerometer, GPS, etc), map Points of Interest (PoI) and GIS data, statistical authority data, self-reports.

Behavioural indicators and Local Environment Conditions (or *Local Extrinsic Conditions* - LECs) are measurements derived from processing of the original raw data and provide information which is interesting with respect to childhood obesity. Specifically, behavioural indicators are measures that describe an individual's behaviour on diet (what you eat), eating behaviour (how you eat), physical activity (how you move) and sleep. Similarly, LECs quantify the characteristics of the environment which can affect individual behaviour. These include LECs related to the urban landscape, school programs and policies, socioeconomic factors as well as food marketing.

Behavioural indicator and LEC data are processed to produce *models and analytics* for data-driven decision support by public health authority experts and policy makers. Models and analytics include:

1. Models providing quantified relations between LECs and behavioural indicators, as well as identification of LECs and behavioural indicators which are associated with obesogenic behaviours in each location
2. Models providing predictions of the impact of interventions (changes of measured behaviours resulting from changes in LECs)
3. Visualisations of the collected data to provide further intuition regarding local factors which contribute to the development of childhood obesity

3.1. Raw sensor data

The first source of raw data used in BigO is Personal Sensory Data acquired by smartphones and commercial smartwatches. This includes Inertial Measurement Unit (IMU) Data, (3D Accelerometer, 3D Gyroscope), Location Data (mainly GPS longitude and latitude) and Photographs. IMU data are sampled at 5 to 100 Hz while Location is recorded every minute. Photographs are captured whenever users decide to submit pictures of their meals or food related advertisements. In our implementation raw data from the smartphone or from Bluetooth paired smartwatches are first buffered in the Smartphone and when wifi link is available they are transmitted to a central server.

Acquisition of the raw data is constrained by battery consumption limitations which currently prohibit continuous acquisition at high sampling frequency.

3.2. Extrenal online sources

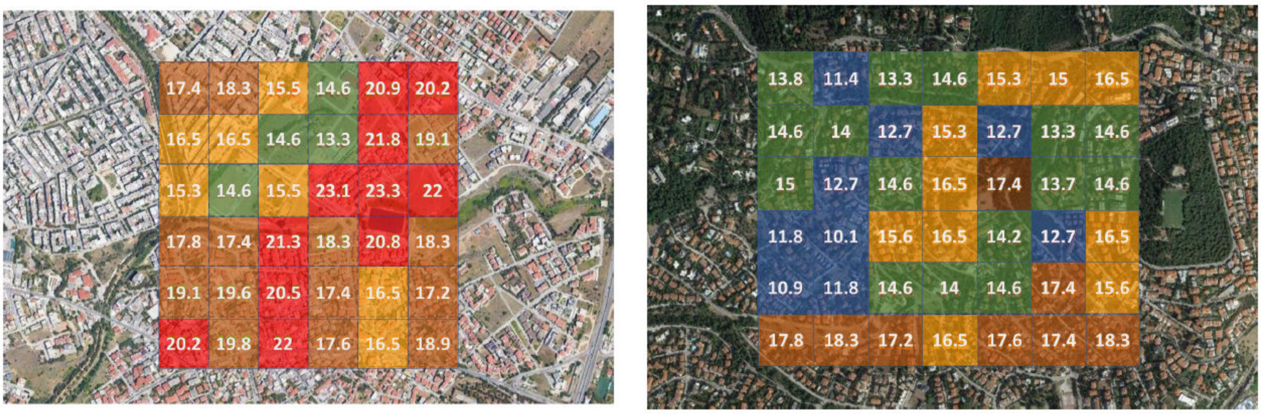
3.2.1. Geoaligned Points of Interest

A number of online providers, such as the Google Maps and Foursquare yield access to the metadata of Public Points of Interest. In BigO we acquire information that refers to (a) food related places (restaurants, food service facilities, grocery stores, etc), (b) gyms and recreation facilities, (c) transportation related POIs (bus stops, metro stations, etc), (d) public facilities like schools and parks. POIs are accompanied by their location metadata, as well as a code corresponding to their main role. Additional metadata may characterize the functionality of the POIs (e.g., timetables, transportation routes, etc). In order to guarantee uniformity we have adopted an internal coding scheme and certain heuristic rules are used to map the characterizations of each provider to our internal taxonomy.

3.2.2. Demographic, Social and Financial Statistical Data

Socioeconomic indices such as the Average Income, Unemployment Rates, Type of Employment, Educational Level etc are collected from published archives of the Eurostat and the National Statistical Authorities. This type of information is directly related to the Local Extrinsic Conditions and heavily influence our aetiology models. Two important limitations of these sources are (a) the relatively coarse spatial resolution of the statistical data (NUTS 1, 2 and rarely 3) and (b) the relatively low temporal resolution limited by the census periodicity. Two complementary sources of similar information are being currently explored:

1. The microdata repositories that are kept by the statistical authorities.
2. Inference of statistics of interest from the analysis of publicly available data. A very promising illustration of this option is our work in [9] that attempts to predict Unemployment rate at a very fine resolution by applying machine learning and image processing techniques to Google street view images.



(a) Pylaia (b) Panorama

Figure 2: Unemployment rate in blocks of two areas, inside the same Greek municipality. Orange and red values indicate high, while blue and green values indicate low unemployment rates as estimated by the analysis of car images appearing on Google Street View [9].

3.2.3. Isolated Repositories of Statistical Data

Many other sources of information regarding the behavior/habits/life style of the population, the prevalence of Obesity, etc have been collected and are provided mostly for free under an open access approach from the World Health Organizations and initiatives like COSI stemming from WHO. BigO intents to integrate this data too.

3.3. Behavioural Indicators and LECs

Behavioural indicators and Local Environment Conditions (or Local Extrinsic Conditions - LECs) are measurements derived from processing of the original raw data and provide information which is interesting with respect to childhood obesity. Specifically, behavioural indicators are measures that describe an individual's behaviour on diet (what you eat), eating behaviour (how you eat), physical activity (how you move) and sleep. Similarly, LECs quantify the characteristics of the environment which can affect individual behaviour. These include LECs related to the urban landscape, school programs and policies, socioeconomic factors as well as food marketing.

In the proposed pipeline, behavioural indicators and LEC data – rather than raw data - are processed to produce models and analytics for data-driven decision support by public health authority experts, policy makers and clinicians. Based on previous studies and scientific findings in the field of obesity we have compiled a long list of indicators (Table 1) and LECs (Table 2).

Table 1: Behavioural Indicators

ID	Name	Units	Sensors
	Diet Indicators		
D1	Eating fast food	Occurrence	L, P, U
D2	Fast-food eating frequency	Times/week	L, P, U
D3	Eating dinner outside of the home?	Occurrence	L, P, U
D4	Eating at home	Occurrence	L, P, U
D5	Food type	Categorical	U, P
D6	Meal type (breakfast, lunch, dinner, snack)	Categorical	L, P, U
D7	Meal frequency (e.g., breakfast)	Occurrence	U,P
D8	Soda or fizzy drinks	Occurrence	U, P
D9	Diet soda	Occurrence	U, P
D10	Juice	Occurrence	U, P
D11	Drinking water	Occurrence	U, P
D12	Drinking milk	Occurrence	U, P
	Physical Activity Indicators		
P1	Energy expenditure (at minute intervals)	MET (based on	A
P2	Activity type (minute)	Categorical	A
P3	Activity intensity	Categorical	A
P4	Activity level	Categorical	A
P5	Activity counts	Counts/minute	A
P6	Walking/cycling to/from school	Times/week	A, L
P7	Mins of active commute to school	Mins/day	A, L
P8	Exercise frequency	Times/week	A
P9	Mins of Sedentary behaviours after school	Times/week	A
P10	Distribution of physical activity at school	Mins per activity	A
P11	Distribution of Physical activity after school	Mins per activity	U, A
P12	Frequency of 10 min bouts of consecutive mod-vig	Times/week	A
	Sleep Indicators		
S1	Hours of sleep per night	Hours	A
S2	Sleep/wake-up times per night	Timestamp	A
S3	Interruptions of sleep	Number	A
S4	Duration of each interruption	Minutes	A
S5	Movement during sleep	Categorical	A
T1	Mins smartphone screen is active per day	Minutes	Other

[1] **L**: Location-related sensors, such as GPS, magnetometer. Either on the mobile phone or in wristband/smartwatch, **A**: Activity-related indicators, such as accelerometer, gyroscope. Either on the mobile phone or in wristband/smartwatch, **P**: Smartphone camera, **U**: User self-reports, **O**: Media monitoring reports, **E**: External sources (e.g. Google maps)

Table 2: Local Extrinsic Conditions

ID	Name	Units	Sensors
	Urban Environment		
U1	Availability of supermarkets and grocery stores	Yes/No, count & locations	E, L
U2	Availability of restaurants and food outlets	Yes/No, count & locations	E, L
U3	Availability of Take-away restaurants	Yes/No, count & locations	E, L
U4	Availability of Café/bars	Yes/No, count & locations	E, L
U5	Availability of wine/liquor stores	Yes/No, count & locations	E, L
U6	Availability of public parks	Yes/No, count & locations	E, L
U7	Availability of indoor recreational facilities	Yes/No, count & locations	E, L
U8	Availability of outdoor recreational facilities	Yes/No, count & locations	E, L
U9	Open spaces in neighbourhood	Percentage/Categorical	E, L
U10	Density of food outlets	Number per square km	E, L
U11	Number of food outlets within 400 m of school	Number	E, L
U12	Number of food outlets within 400 m of home	Number	E, L
U13	Number of food outlets along school commute	Number	E, L
U14	Proximity of recreational facilities from home	Kilometres	E, L
U15	Proximity of recreational facilities from school	Kilometres	E, L
U16	Density of recreational facilities	Number per square km	E, L
U17	Distribution of recreational facility type	Categorical?	E,L
U18	Time spent in recreational facilities	Mins per week	L,E
U19	Additional SPOTLIGHT neighbourhood typology	-	-
	School Environment		
S1	School exercise programs	Times per week, duration	E
S2	School meals/breaks	Nr of instances & duration	E
S3	School hours	Start/end timestamps	E
	Socioeconomic Environment		
E1	Average income in neighbourhood	EUR(SEK)/person/year	E
E2	Education level statistics	Education level distribution	E
E3	Unemployment rates	%	E
	Food Marketing		
M1	Exposure to food advertising from TV	Categorical or nr ads/day	O
M2	Exposure to food advertising from the internet	Categorical or nr ads/day	O
M3	Exposure to food advertising in urban environment	Number of food ads in area	O
M4	Food advertising at specific times	Series of timestamps	O

3.4. Use of urban resources

In addition to indicators of behaviour and LECs that describe the environment, we can produce measurements which are in-between the two, i.e., measurements on the use of urban resources. The following tables provide examples of such measurements from two different viewpoints, (a) from the individual viewpoint (in what urban context does an individual perform his/her activities? - Table 3) and (b) from the location viewpoint (what behaviours are observed in this location? - Table 4).

Table 3: Examples of individual-level measurements resulting from a combination of LECs and behaviour, without necessarily revealing the location of individuals

ID	Name	Units
H1	Time spent in open spaces in neighbourhood	Mins/week
H2	Number of times per week user is within 400 m of food outlet	Frequency

H3	Time spent in recreational facilities	Mins/week
H4	Distance from school to home	Meters

Table 4: LECs based on aggregate children statistics

ID	Name	Units
A1	Number of children visited geohash[2] or PoI	Number
A2	Number of meals registered at geohash or PoI	Number
A3	Average PA of children at geohash or PoI at specific time segments	METs
A4	Distribution of means of transportation for geohash or PoI	Vector corresponding to the distribution
A5	Most common path from (geohash or public PoI) A to B	Series of locations or geohashes defining the path

4. Privacy preservation

Collecting and using dense behavioural data, including location recordings, must be accompanied by strong guarantees that personal identifiers will be removed from this data. Pseudonymization is necessary but not sufficient; the reason is that combining location traces of even anonymous users may reveal their ID.

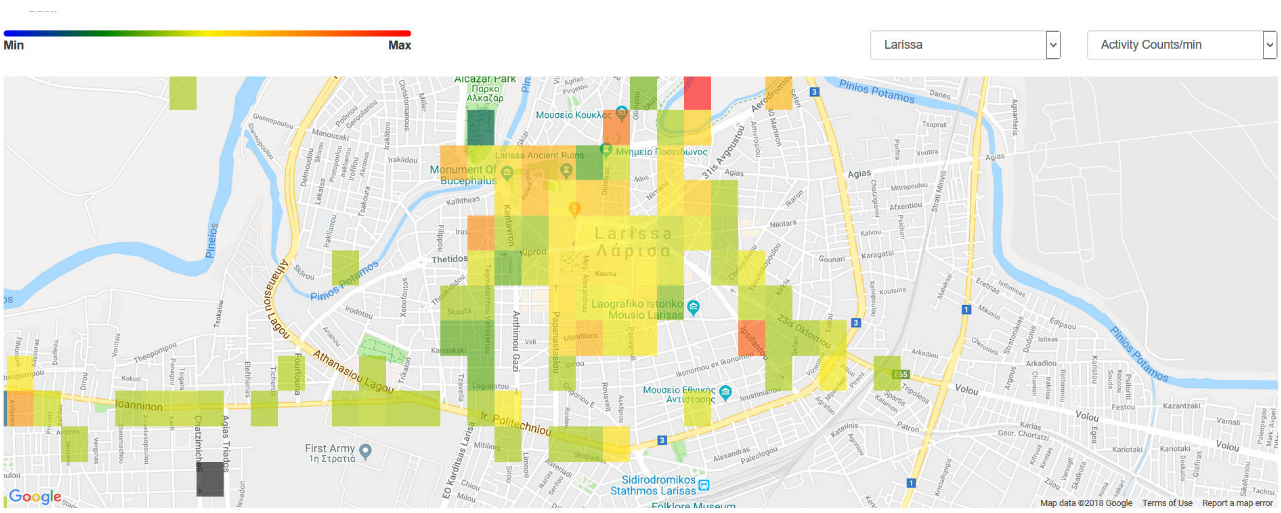


Figure 3: Activity counts (indicator P5) in the area of Larissa, Greece. Monthly (June 18) average normalized to an arbitrary min-max scale. Quantification has been performed at the level of geohashes; Geohashes with few votes appear blank

Our solution to this problem adopts an arbitrary partition of locations according the geohash geocoding. The adopted geohash resolution is fine enough for the purpose of describing the local obesogenic environment (e.g., an area of a few city blocks, a neighborhood). At the same time it should be coarse enough to support k-anonymity of the involved users. Let L be the set of adopted geohashes.

---

[2] Geohashes are alphanumeric strings following a geocoding mechanism which defines a hierarchical grid of rectangular regions around the globe. Each geohash is a rectangular region. The longer the length of the geohash, the smaller the region. In the context of this document Geohashes refer also to the region itself.

The next element of our solution is the quantization of behavioural indicators. With reference to Table 1, let B_k be a partition of the k-th behavioural indicator. Partitioning of categorical indicators is straightforward. For the numerical ones (either scalar or multidimensional) some binning is being assumed.

After the aforementioned quantization of the spatial domain and each of the behavioural indicators, our approach is to avoid storing behavioural data associated with even pseudonyms of the users; instead we implement an anonymous voting scheme where for each time slot (say once every 30 minutes) each user casts a vote to one element of the Cartesian product $L \times B_k$, for each indicator k. Voting is automatically controlled by the mobile application software without any user intervention. The internal intelligence of the software decides after each time slot (a) which geohash is to be voted; the decision relies on the duration of time spent to each geohash during the specific time slot, (b) which of the bins of behavioural indicator B_k is the most representative during the specific slot. K-anonymity constraints are enforced by the automatic teller resulting to appropriate increase of the geohash size (temporary reduction of spatial resolution), spreading the votes to more than one geohash areas or even aborting casted votes.

The outcome of this voting procedure is a map of distributions for each behavioural indicator B_k:

$$C(l,b,k), \quad l \in L, b \in B_k$$

Subsequent steps of deriving correlations between obesogenic behaviours and the environment resort to this type of aggregate data. Following a similar voting approach, two other types of privacy preserving aggregations are also under consideration. In the first geohashes are replaced by public Point of Interests while in the second the geographic elements are routes within a city graph.

5. Data exchange and standardization

Aggregated data that have been sanitized from traceable personal information become available to the local authorities, public health experts and the scientific community. In this perspective, in BigO we act as intermediate processors that collect information from users – under the citizen scientists paradigm – and other providers as well, aggregate and combine the collected big data input and produce detailed geoaligned statistics. This can be useful only under two conditions: (a) The derived statistics are of satisfactory quality and (b) their interpretation –together with the interpretation of the original data– is unambiguous. Both requirements are challenging.

5.1. Quality

The first difficulty stems from the multiplicity of commercial devices used to produce raw data in the wild, in an uninstructed way. Raw data may thus suffer from calibration problems and also from noise contamination. For example, a smartwatch firmly fastened produces different accelerometry data compared to a loose one. Robust signal processing and machine learning modules are used to abate these effects. The second source of quality concerns is the sparseness of user data; users deliberately start and stop recording or even decide to carry/wear their smartphone/watch. This uncontrolled behavior is the reason of missing data. Extraction of behavioural indicators from such data is not trivial, it requires strong imputation algorithms and long term recordings.

5.2. Standardization

The strong requirement that raw behavioural recordings are never passing the internal processing boundary of BigO has the direct consequence that only the second level of our information model will be accessible from external stakeholders. Behavioural indicators must thus be uniquely interpretable. Standardizing a well defined list of indicators with associated measurement units and ranges is our answer to this.

6. Conclusions

BigO develops tools that allow for the monitoring of obesogenic behaviours, of users that voluntarily offer their data according to the citizen scientist paradigm. Raw data originating from worn IMU sensors, GPS, pictures captured by the users and user responses to questionnaires are being aggregated in order to produce behavioural indicators. Collection of Local Extrinsic Conditions is also supported. Extracted behavioural indicators are being correlated to LECs as a means to identify the local factors that cause (childhood) obesity. BigO adopts strict privacy preservation mechanisms including an innovative aggregation method based on voting. Anonymized aggregate data are also exported for use by external stakeholders.

7. References

[1] Wijnhoven TM, van Raaij JM, Sjöberg A, Eldin N, Yngve A, Kunešová M, Starc G, Rito AI, Duleva V, Hassapidou M, Martos E1, Pudule I, Petrauskiene A, Sant'Angelo VF, Hovengen R, Breda J. WHO European Childhood Obesity Surveillance Initiative: school nutrition environment and body mass index in primary schools. International journal of environmental research and public health. 2014 Oct 30;11(11):11261-85.

[2] Bammann K, Gwozdz W, Lanfer A, Barba G, De Henauw S, Eiben G, Fernandez-Alvira JM, Kovács E, Lissner L, Moreno LA, Tornaritis M, Veidebaum T, Pigeot I; IDEFICS Consortium. Socioeconomic factors and childhood overweight in Europe: results from the multi‐centre IDEFICS study. Pediatric Obesity. 2013 Feb 1;8(1):1-12.

[3] EU Action Plan on Childhood Obesity 2014-2020: https://ec.europa.eu/health/sites/health/files/nutrition_physical_activity/docs/childhoodobesity_actionplan_2014_2020_en.pdf

[4] WHO: Report of the commission on ending childhood obesity: http://apps.who.int/iris/bitstream/10665/204176/1/9789241510066_eng.pdf

[5] Lyn R, Aytur S, Davis TA, Eyler AA, Evenson KR, Chriqui JF, Cradock AL, Goins KV, Litt J, Brownson RC. Policy, systems, and environmental approaches for obesity prevention: a framework to inform local and state action. Journal of public health management and practice: JPHMP. 2013 May;19(3 Suppl 1):S23-33.

[6] Lobstein T, Jackson-Leach R, Moodie ML, Hall KD, Gortmaker SL, Swinburn BA, James WP, Wang Y, McPherson K. Child and adolescent obesity: part of a bigger picture. The Lancet. 2015 Jun 26;385(9986):2510-20.

[7] Wang Y, Wu Y, Wilson RF, Bleich S, Cheskin L, Weston C, Showell N, Fawole O, Lau B, Segal J. Childhood obesity prevention programs: comparative effectiveness review and meta-analysis.

[8] De Bourdeaudhuij I, Verbestel V, De Henauw S, Maes L, Huybrechts I, Mårild S, Eiben G, Moreno LA, Barba G, Kovács É, Konstabel K, Tornaritis M, Gallois K, Hebestreit A, Pigeot I; IDEFICS consortium. Behavioural effects of a community ‐ oriented setting ‐ based intervention for prevention of childhood obesity in eight European countries. Main results from the IDEFICS study. Obesity reviews. 2015 Dec 1;16(Suppl 2):30-40.

[9] Diou, C.; Lelekas, P.; Delopoulos, A. Image-Based Surrogates of Socio-Economic Status in Urban Neighborhoods Using Deep Multiple Instance Learning. Preprints 2018, 2018080154 (doi: 10.20944/preprints201808.0154.v1).