EDITORIAL

Smart city lifestyle sensing, big data, geo-analytics and intelligence for smarter public health decision-making in overweight, obesity and type 2 diabetes prevention: the research we should be doing

Maged N. Kamel Boulos[1*] and Keumseok Koh[2]

Abstract

The public health burden caused by overweight, obesity (OO) and type-2 diabetes (T2D) is very significant and continues to rise worldwide. The causation of OO and T2D is complex and highly multifactorial rather than a mere energy intake (food) and expenditure (exercise) imbalance. But previous research into food and physical activity (PA) neighbourhood environments has mainly focused on associating body mass index (BMI) with proximity to stores selling fresh fruits and vegetables or fast food restaurants and takeaways, or with neighbourhood walkability factors and access to green spaces or public gym facilities, making largely naive, crude and inconsistent assumptions and conclusions that are far from the spirit of 'precision and accuracy public health'. Different people and population groups respond differently to the same food and PA environments, due to a myriad of unique individual and population group factors (genetic/epigenetic, metabolic, dietary and lifestyle habits, health literacy profiles, screen viewing times, stress levels, sleep patterns, environmental air and noise pollution levels, etc.) and their complex interplays with each other and with local food and PA settings. Furthermore, the same food store or fast food outlet can often sell or serve both healthy and non-healthy options/portions, so a simple binary classification into 'good' or 'bad' store/outlet should be avoided. Moreover, appropriate physical exercise, whilst essential for good health and disease prevention, is not very effective for weight maintenance or loss (especially when solely relied upon), and cannot offset the effects of a bad diet. The research we should be doing in the third decade of the twenty-first century should use a systems thinking approach, helped by recent advances in sensors, big data and related technologies, to investigate and consider all these factors in our quest to design better targeted and more effective public health interventions for OO and T2D control and prevention.

Keywords: Smart city lifestyle sensing, Big data, Geo-analytics, Systems science, Overweight, Obesity, Type 2 diabetes

*Correspondence: mnkboulos@mail.sysu.edu.cn
[1] School of Information Management, Sun Yat-Sen University, East Campus, Guangzhou 510006, Guangdong, China
Full list of author information is available at the end of the article

Introduction

The ongoing pandemic of coronavirus disease (COVID-19) has heavily scarred the healthcare capacities and systems across the world during the whole year of 2020 and until now (Q1 2021). At the time of writing, more than 1 per cent of the world population (82 million) were infected (diagnosed/confirmed

cases only), and over 1.8 million deaths were claimed by the pandemic [1]. Despite the recent arrivals of vaccines and medications, COVID-19 is likely to pose an additional burden to public health for the next few years [2]. While an effective control and prevention of COVID-19 would be high on the list of public health interventions worldwide, combating the existing global epidemic of non-communicable diseases (NCDs), attributable each year for 41 million deaths or 71 per cent of all death globally, should not be neglected [3].

Among NCDs, overweight and obesity (OO) and type 2 diabetes (T2D) are unique because they are both medical conditions (OO: ICD-10 code E66; T2D: ICD-10 code E11) and leading risk factors for other comorbidities, including many NCDs and infectious diseases such as COVID-19 [4–7]. The impacts of OO and T2D on public health are much devastating. A global study estimated that about 9.7 per cent of the world population (711.4 million) were obese and 4.0 million deaths were attributable for obesity in 2015 [8]. T2D affected 463 million adults aged 20 to 79 years worldwide and caused 4.2 million deaths in 2019 [9]. Scholars predicted no decline in OO and T2D prevalence any time soon if the current increasing trend continues [10, 11].

The fact that OO and T2D are unevenly distributed across different socioeconomic statuses (SES), demographic groups and geographies make researchers and policymakers more puzzled to design effective public health interventions [12, 13]. Low-SES individuals, specific racial/ethnic groups (e.g., OO: non-Hispanic blacks in the United States and Western Pacific Islanders; T2D: non-Hispanic black, Hispanic-American and Native American population in the United States, and South Asian and black African/Caribbean population in the United Kingdom), and rural populations were found more vulnerable to OO and/or T2D [14–17]. Notably, China and India, the two most populous and emerging economies in the world, are among the world's largest and fastest-growing populations with OO and/or T2D [18, 19]. Designing effective prevention and control policies for OO and T2D is of importance not only to decrease their direct epidemiological costs, but also to ensure successful achievement of the United Nation's Sustainable Development Goals (SDGs) [20].

To get our act together for successful future strategies and interventions, this article provides a brief overview of the latest OO and T2D geospatial research trends, especially focusing on novel approaches using precision and accuracy medicine/public health, smart technologies, big data and systems science.

Innovations in research methods during the last decade

OO and T2D have been conventionally and rather naively attributed to a long-term energy imbalance between calories intake from food and beverages and expenditure via physical activity (PA) and behaviour, but the reality of their pathogenesis is far more complex than that. Researchers have extensively investigated the multifactorial and deeply interrelated interactions among multiple biological, sociodemographic and contextual factors of OO and T2D. Clinic-based studies have identified more than 50 genes associated with OO and 36 genes related to T2D that play specific roles in appetite and food intake, insulin action, cholesterol and fatty acid synthesis, and family heredity [21, 22]. Population-based studies have also found various risk factors, including, but not limited to, age, race/ethnicity, SES, lifestyle choices and behaviours, sugar and high-fructose corn syrup-laden food products (not all calories are equal in their health effect, even though each of them carries the same amount of energy), and different environments [23].

Geographic Information Systems (GIS) have contributed uniquely to OO and T2D literature. Researchers used GIS to investigate the built environment and its associations with OO and T2D. Since 2002 (and up to December 2020), more than 130 articles using GIS have been published in *International Journal of Health Geographics* (IJHG) on the topic of OO (n=77) and T2D (n=64). Examples of OO and T2D articles published in IJHG include studies about youth's obesity [24, 25], older people's body weight [26], the built environment [27, 28] and new methodologies [29, 30]. With advances in related technologies, traditional GIS research approaches have become smarter. Geo-analytics and GeoAI are two rapidly growing GIS research areas with promising applications in the formulation and monitoring of OO and T2D public health interventions. Geo-analytics combines location-based (big) data with advanced analysis functionalities on a website or application (desktop or mobile), enabling interactive data interrogation, analysis and visualisation. GeoAI covers a range of technologies for extracting meaningful information and knowledge from geospatial big data using artificial intelligence and data science methods, such as data mining, machine learning/deep learning and high-performance computing [31, 32].

During the past few years, new technologies have been applied in OO and T2D research. Firstly, utilising lifestyle sensors (LS) has gained popularity among researchers. LS are devices that detect and collect changes in the environment or behaviours of an object and relay the collected data to other data processing devices or controllers (e.g., smartphones and/or cloud computing platforms). Types

and levels (intensity and duration) of physical activity, food intake, mobile personal indirect calorimetry, heart rate (variability), blood sugar, blood lipids and blood pressure levels are among the measures that OO and T2D studies have frequently used [33] or can use (see, for example, slide 30 in Additional file 1). Secondly, big data have proved their potential to enhance our understanding of OO and T2D. Data aggregates from restaurant chains, supermarkets and their loyalty card schemes/apps (e.g., about sales and customers' purchasing habits/patterns), on-board vehicle sensors, weight management programmes (e.g., clinical and sociodemographic data from clients), geospatial built environment (e.g., layouts of restaurants and the different types of food they serve, remote sensing data, etc.), social media, and smartphones and wearables (e.g., fitness apps and smartwatches) have all been (or can be) collected, crosslinked and utilised for advanced analyses, e.g., [34], subject to appropriate individual privacy safeguards. And finally, systems science—an interdisciplinary field of science to investigate highly complex structures and interactions among components within systems—have helped researchers adopt a holistic understanding of OO and T2D, especially focusing on lifestyle, food environment, social networks (offline and online), and healthcare [35–39].

What we need to do in future OO and T2D studies: a feasibility demonstrator roadmap for a large metropolis

This demonstrator (see Additional file 1) features Guangzhou, the capital and most populous city (total population: about 15 million in 2019) of Guangdong Province in China, as the study area, but the same roadmap can be adopted for other cities/regions in China and other countries. As of 2020, Guangzhou is the 4th largest regional economy by Gross Domestic Product in China (1.7 trillion Chinese yuan, equivalent to 262 billion US dollars). With its very rapid economic development during the last few decades, China is now confronting the dual epidemics of OO and T2D. It has the largest overweight population in the world, and its population figures for childhood obesity and adult obesity ranked 1st and 2nd, respectively, across the world in 2014/15 [8, 40]. Over one out of ten Chinese adults (10.9%) were estimated to have diabetes in 2013 [41]. Immediate, evidence-based public health interventions focusing both on locale- and population-specific evidence will be vital to help release China from the tightening grip of OO and T2D.

Previous research into food and PA neighbourhood environments has mainly attempted to associate body mass index (BMI) with proximity to stores selling fresh fruits and vegetables or fast food restaurants and takeaways, or with urbanisation, neighbourhood walkability factors and access to green spaces or public gym facilities, making largely naive, blanket assumptions and crude, incomplete and often inconsistent (across similar studies) conclusions that are far from the spirit and requirements of twenty-first century precision and accuracy public health.

Different people and population groups respond differently to the same food and PA environments, due to a myriad of unique individual and population group factors and their complex interplays with each other and with food and PA elements (genetic/epigenetic factors, metabolic factors, gut bacteria profiles, gut hormones profiles, health literacy profiles, dietary and lifestyle habits, screen viewing times, stress levels, sleep patterns, SES, local cuisine and food industry standards and regulations [e.g., food processing levels, food labelling practices, etc.], environmental air and noise pollution levels, activity spaces [associating people with a single address/postcode is not always ideal in food and PA environment studies], etc.).

Furthermore, the same food store or fast food outlet can often sell or serve both healthy and non-healthy options/portions and use both good and bad food processing and cooking methods for different products, so a simple binary classification into 'good' or 'bad' store/outlet should be avoided. Population dietary behaviours, including amounts consumed per individual snack/meal/day or per family are also important, as even the healthiest options can prove unhealthy when overconsumed. Moreover, appropriate physical exercise, whilst essential for good health and disease prevention, is not very effective for weight maintenance or loss (especially when solely relied upon), and cannot offset the effects of a bad diet [42]. In fact, the "wrong" type of physical exercise might sometimes result in weight gain in certain individuals, and research has shown that some individuals avoid or 'hate' exercise because of their genetic makeup, even when living in close proximity to green spaces and public gyms. As far as urbanisation is concerned, research also shows the gap of BMI between urban and rural is closing, mostly by an unprecedented increase in rural BMI around the world in recent years, especially in low- and middle-income regions (so it is not urbanisation that is to blame as such or alone) [43].

The research we should be doing in the third decade of the twenty-first century should use a systems thinking approach, helped by recent advances in sensors, big data and related technologies, to cater for this myriad of interconnected factors in our quest to design better targeted and more effective public health interventions for OO and T2D control and prevention. The remainder of this section briefly describes the high-level task clusters that are necessary to successfully execute the feasibility

demonstrator roadmap and build prototype smart dashboards for future precision and accuracy public health OO and T2D interventions (Fig. 1).

Task cluster 1: Agree a comprehensive list of key population lifestyle data

To design and develop effective evidence-based public health interventions for OO and T2D, there should be an agreement on a comprehensive list of population lifestyle data relevant to the study area and study population. Besides standard population variables, such as age, gender, race/ethnicity, SES and family status, a set of important lifestyle risk factors for OO and T2D studies may include health literacy levels (per population group/neighbourhood derived from up-to-date, representative community-wide surveys), food shopping and dietary habits, behavioural characteristics (e.g., sedentary lifestyle, physical activity, screen viewing, sleep, smoking, mental health, etc.), relevant health conditions and exposure to environmental pollutants (e.g., noise, light pollution and fine particulate matter).

Population dietary habits

While studies have often requested participants to write food diaries (usually resulting in inaccurate and/or incomplete data collection), later studies have used dedicated smartphone apps to collect and analyse meal records and food barcodes (packaged products) and pictures (using machine learning and computer vision, sometimes coupled with smartphone spectroscopy and other sensors [44]). Aggregates of individual-level dietary habit data from specialised smartphone apps (see, for example, slide 30 in Additional file 1) may be further combined with territory-wide databases of food outlets, food/restaurant guide and review websites, food sales/payment transaction records (segmented by food class/type), food retailers consumer loyalty card/app data aggregates (can provide detailed geodemographic classifications of households and consumers, offering unique insights about food spending patterns/trends among different population groups), purchasable food options/availability, and/or other relevant food databases at various neighbourhood levels.
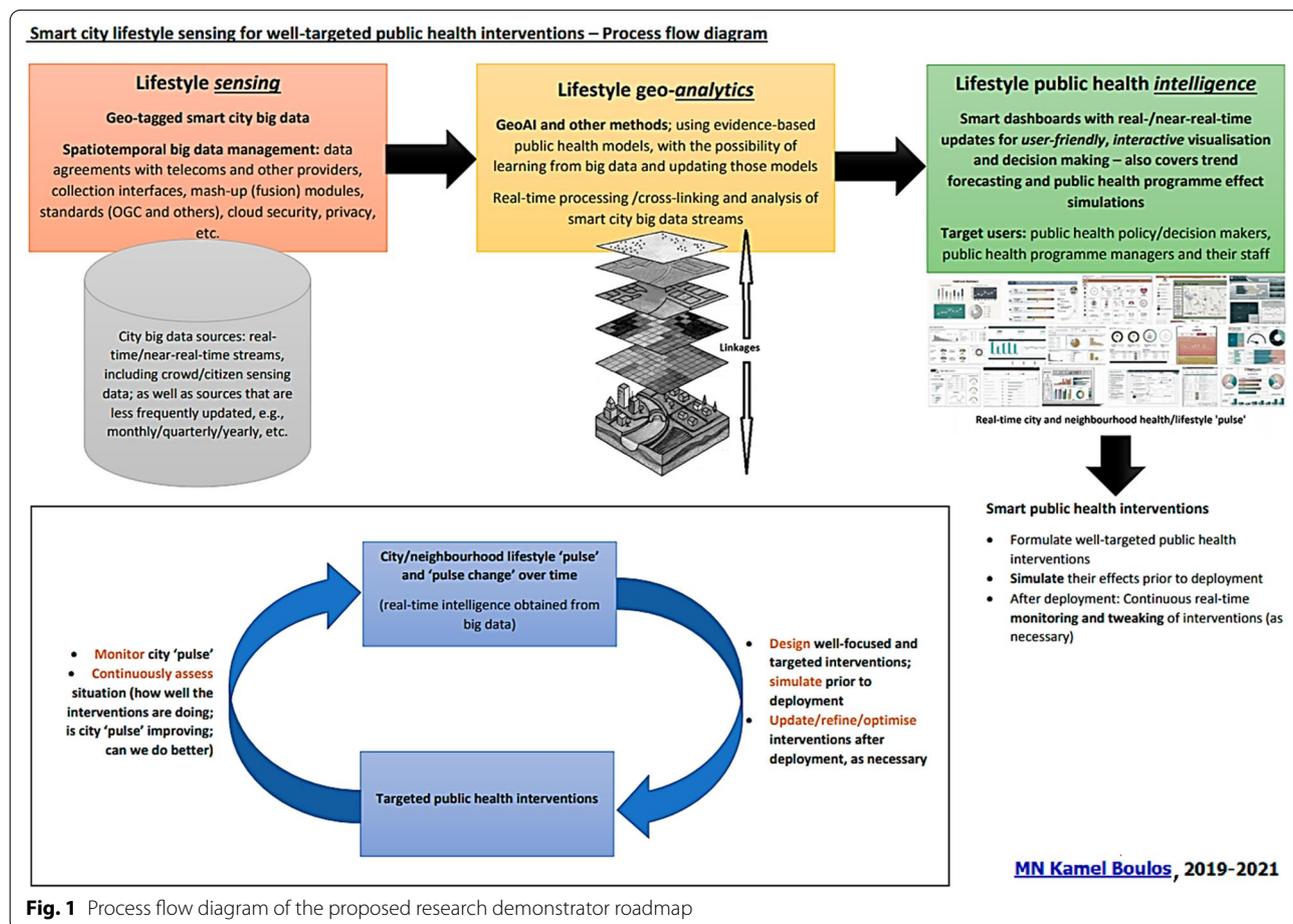

Fig. 1 Process flow diagram of the proposed research demonstrator roadmap

Population PA and behavioural characteristics

Measures of PA and estimates of energy expenditure levels, e.g., by computing daily walking steps, are commonly used in OO and T2D research. Some studies measured calories spent on PA using LS, GPS/location-based applications or exergaming devices [45, 46]. Fitness wearables and apps also exist that can perform some form of PA segmentation by type and intensity, which when combined with specialised LS for mobile personal indirect calorimetry, can further calibrate and enhance the accuracy of energy expenditure estimates [42]. Big data aggregates from these wearables, gadgets and apps can reveal interesting trends among different population groups.

Other measurable lifestyle factors associated with PA in OO and T2D literature include the use of television and other screen devices (can affect sleep quality), sugar-sweetened beverage intake (often marketed using football and sport themes, which can be deceiving), smoking (reduces physical endurance) and sleep quality [47, 48]. Smartphone screen time trends, sedentary behaviour/walking steps and sleep quality/patterns in the community can be readily extracted from user population big data aggregates derived from smartphones (e.g., Google Digital Wellbeing platform on Android) and fitness/PA apps and gadgets, providing invaluable 'lifestyle intelligence' for designing better public health interventions.

Further population-level PA indicators to consider include population heart rate variability levels, also derived from fitness/PA wearable and app data aggregates; and, if a platform such as Google Fit is used, average 'Move Minutes' and 'Heart Points' per person/per day in different age groups and country regions, cities and city neighbourhoods; percentages of poor, average and high performers in target populations; also cumulative (total) population walking steps, 'Move Minutes' and 'Heart Points' per region, city and neighbourhood for a given period. These indicators can be monitored and compared every few months to determine population progress and trends over time, or at different times of the year/seasons, or following specific PA community interventions. Indicators can also be normalised by population number to compare different regions, cities and neighbourhoods. (Location-tagging of app data aggregates, e.g., by neighbourhood or city, should be straightforward and readily available, as most PA and fitness mobile apps require 'location permission' to log user's home base location, fetch corresponding weather data, and map walking, running and cycling routes.) Further interesting patterns might emerge from user data aggregates that can help city planners better understand and optimise the PA landscape of their city and its population; for example, data might reveal that residents of some neighbourhoods are doing most of their daily exercises in some walkable streets, parks or public gym facilities in a different neighbourhood [42].

Population health status/conditions and stress levels

Population clinical data (aggregates and trends stratified by population group and neighbourhood/region), such as BMI and body fat percentage (BFP, a better measure than just body weight or BMI), blood sugar and blood pressure levels, cholesterol/blood lipid profiles, data about relevant long-term conditions, etc., can be combined with other population lifestyle factors/data, since OO and T2D are related to, or often coexist with, other chronic diseases, e.g., cardiovascular conditions.

Population aggregates from platforms such as Amazon's Halo wearable and health monitoring service (subscription) can provide unique insights about the prevailing moods and stress levels among different population groups over time. Powered by machine learning, speech processing and computer vision, Amazon Halo infers mood and stress levels from the wearer's voice tone, and also measures BFP [49].

Activity spaces, exposure to environmental pollutants, etc.

The rapid advances and improvements in mobile technologies and their affordability over the past two decades have enabled researchers to use smart devices and wearables more actively in their research. For example, a Dutch study obtained a GPS-tracked mobility dataset from its participants for seven consecutive days [50]. Such a dataset, if collected for a longer period of months or years, could be extremely useful for researchers to measure the precise and accurate levels of exposure to environmental factors, rather than using a home, school, or office address as a proxy for exposure locations, (in addition to providing more accurate and highly valuable details about participants' *actual* food and PA environments). Citizen noise pollution monitoring is also possible using smartphones [51]. The analysis of relevant photos and videos of neighbourhood environments and street scenes can provide additional insights.

It should be noted that the aforementioned lifestyle variables are not an exhaustive list of all the factors one should consider. Researchers also need to prioritise the acquisition of these variables according to their relative importance, data availability and reliability, and ease and cost of collection, by carefully examining previous public health interventions (their level of success and types of data feeding into them), the decision support needs of public health professionals and policymakers, and the available human and financial resources.

Task cluster 2: identify the most efficient ways of data collection

Identifying the most efficient, appropriate and reliable ways of data collection is as important as selecting the relevant variables. Technological feasibility, as well as traditional issues, such as budget and staffing, should all be examined when planning data collection. There are several important questions that researchers need to carefully consider; for example, are the data purchasable from vendors or need to be collected? Will the data be collected by a research team, by some outsourcing contract, or by crowdsourcing? Which method or instrument would be best for collecting each dataset depending on its nature, real- or near-real-time datasets and those that need to be collected less frequently; for example, using data streams from LS/wearables, smartphone apps, or traditional survey methods (e.g., by telephone, post, websites, or social media)? What levels of detail granularity will be set for the research and data (e.g., by some specified population groups, neighbourhood, or city)? How long will it take to collect enough data for the system to reason with, and what will be the ideal update frequency/interval, where applicable (e.g., (every) 24 h, one week, one month, yearly, or before, during and after some public health intervention, etc.)? For data that are going to be sampled rather than extracted from population aggregates of routine service/app users, we need to determine how many participants should be recruited (and their characteristics) to secure a representative sample of the target population, and what is the most appropriate method to recruit them (e.g., snowball sampling, convenience sampling, stratified sampling, etc.)? (Big) data quality and reliability dimensions must also be carefully considered for each dataset, e.g., validity, accuracy, consistency, completeness, timeliness, etc.

Task cluster 3: data warehousing and management

This task cluster focuses on big data governance and management, covering a number of interrelated issues that are critical to the success, scalability and sustainability of the demonstrator, including compliance with relevant data and metadata standards, e.g., OGC (Open Geospatial Consortium) standards; compliance with data protection regulations, ensuring individual data privacy is always preserved, especially when non-aggregate data are collected and processed; establishing adequate data sharing agreements with providers; and the adoption of best practices in ontologies, big data warehousing and cloud security.

Specifically, data should be protected from any unauthorised access and potential corruption at all times. Individuals' privacy and confidentiality protection should comply with up-to-date regulations for data storage

and management, all while enabling the demonstrator's seamless, secure access to the data for analysis. Data taxonomies and ontologies are particularly essential for managing and reasoning with big data, since the majority of big data are often collected in unstructured form [52, 53].

Task cluster 4: data fusion, geo-analytics and GeoAI

Data fusion refers to the process of integrating data from different datasets and multiple sources, including managing data uncertainties and incompleteness/missing data points across the fused datasets. For example, a geo-tagged social media dataset aggregated at neighbourhood levels can be 'fused' with other aggregate neighbourhood data, such as governmental census or survey data, to paint a more accurate and detailed picture of the target population. Geo-analytics and GeoAI tools are becoming increasingly essential for crosslinking and reasoning with ('making sense of') the growing amounts of geo-tagged population big data that are continually generated through mobile health (mHealth) devices/apps and precision medicine practices [32].

Task cluster 5: design smart public health dashboards with mechanisms for intervention simulations and near-real-time monitoring and optimisation of interventions.

The smart public health dashboards should (1) provide public health decision makers and programme planners with a user-friendly one-stop portal for intelligent data analysis, interactive data querying and visualisation in different ways, and distributed team collaboration; and (2) offer lay summaries/views of the dashboards to keep members of the general public informed and engaged. Public health planners can use the dashboards and the generated big data intelligence to interactively design and prioritise new public health interventions, identify their potential target population groups and areas, and perform simulation modelling of intervention costs and impacts under different settings or 'what if' scenarios prior to deployment [42]. Various GIS and simulation modelling techniques are available for this purpose (see one example in [37]). Once an intervention is agreed and deployed in the real world, the smart dashboards can then serve as a 'situation room'. Fed with fresh, post-intervention-deployment data streams, the dashboards can be used to dynamically monitor the intervention in near real-time and tweak it as necessary. The ideal dashboards should be intuitive and user-error-tolerant. They should help decision makers uncover subtle, unfolding changes of concern among their populations before they grow into bigger problems. To achieve this goal, the dashboards should serve intelligent, contextual alerts and

reminders to its users in a timely manner, inferred from the combined 'patterns of change' of multiple interrelated data streams, (going beyond conventional, simple threshold-based triggers derived from single data sources, which often fail to provide timely 'early warnings').

Task cluster 6: evaluate dashboards and other components
Public health dashboards serve critical functions as information and intelligence communication tools and in decision support. Adequate user involvement in all stages of their development (via a representative sample of all stakeholders and end users' roles) is key to their lasting success and ultimate user acceptance [54]. Development should be conducted iteratively and incrementally to incorporate user feedback. A comprehensive, multifaceted evaluation, covering usability, utility, accuracy/reliability, etc., must be carried out for all the interfaces, tools and other components created in task clusters 3, 4 and 5 above, individually and in their integrated form as one whole system/service. Dashboard aspects that should be evaluated include end user interfaces (ease of use, customisability, feedback to users, human error tolerance, etc.), knowledge discovery, public health information delivery and communication, system security, and integration and interoperability with other relevant/existing public health systems [55].

Task cluster 7: research and technological development (R&D) coordination and management
Designing effective public health dashboards is a highly multidisciplinary and interdisciplinary undertaking that must address a number of essential criteria and tasks concurrently. These tasks cover (1) research integrity, quality assurance and system/evidence updates, (2) multidisciplinary collaboration, (3) risk management, and (4) exploitation of the demonstrator's results.

Firstly, a dashboard should achieve scientific rigour. Data, analytics methods, visualisation and underlying research design, as well as the underpinning clinical/public health evidence must all be robust and sound, avoiding the well-documented failings and traps of big data [56]. Dashboards must be designed and developed in such a way that accommodates and streamlines future maintenance and expansion updates of software components (functionality, security, etc.), consumed datasets, and underpinning clinical/public health evidence and guidelines (as science progresses). Updates (and the mechanisms for implementing them) ensure the quality of the dashboards can be maintained and improved over time. Standard best practices in research ethics and individual privacy preservation must be adhered to.

Secondly, it is vital to build and maintain strong partnerships, close collaboration and mutual understanding between the multidisciplinary members of the demonstrator team (with their different but complementary professional backgrounds and areas of specialism), target users and other stakeholders. The research literature offers some excellent discussions about participatory practices and platforms in dashboard building that should prove helpful in fostering and facilitating these collaborative partnerships [57, 58].

Thirdly, risks assessment and mitigation plans ('plan B') must be put in place at the very start of the work on the demonstrator. Lastly, outputs dissemination activities and exploitation plans should be developed to keep the wider stakeholders and managers engaged and supportive of the work, secure continuing funding, and ensure future viability and expansions of the demonstrator, both in functionality and to other cities/regions and populations.

Looking forward to research in 2025 and beyond
Big data approaches in genomics, epigenomics and bioinformatics, coupled with geographic information (systems), are unveiling the complex interplay of environment, genes and other factors in health and disease [59]. Genes confer potential protections and predispositions. But it is the lifestyle/environmental (exposome) modulation – up and down, on and off – of their expression (gene expression) via the epigenome (epigenetics) that determines their ultimate effects, i.e., can increase/enhance or decrease/mute the negative (and positive) effects of lifestyle and environmental exposures on the individual, depending on the unique interplay between a person's genetic profile and her/his lifestyle and environmental exposures [60]. Some of the epigenome-tagged genome DNA and histones can even be heritable. Environmental pollutants (obesogens), gut microbiota modifications and unbalanced food intake can induce, through epigenetic mechanisms, OO and altered metabolic consequences [61].

But that is not the end of the story; we can still 'tame' (tone down/turn off the expression of) unfavourable genes with suitable external factor modifications and interventions, e.g., by introducing targeted gamified exercise interventions (exergames/mobile exergame apps) for added exercise 'appeal' and PA behavioural sustainability in those individuals and population groups that 'hate' exercise because of their genetic makeup [62].

Indeed, future public health interventions will evolve to precisely identify, and address the specific characteristics and needs of, target population groups or areas, inasmuch the same way as personalised precision medicine is evolving in managing individual patients. In the future, we will know more about our target populations, thanks to population genome data banks,

where profiles of large local population samples can be mined and analysed for the presence specific relevant genes, gene variants/mutations and dysfunctions, such as those genes controlling appetite, or implicated in overweight and obesity predisposition, or determining our base exercise behaviour, e.g., fat mass and obesity-associated (FTO), melanocortin-4 receptor (MC4R), leptin (LEP), etc. [63]. We will then use this intelligence to inform the design of optimised individual- and population-group-specific interventions that can make our genes work best for us and/or offset their negative effects (predispositions).

Similarly, we can have population gut microbiome data banks of large local population samples (with regularly updated profiles, say every 6 months, to monitor change). A growing number of consumer-oriented labs are already offering 'gut bacteria profiling' services today at affordable prices. Gut bacteria types and diversity/ratios (some gut bacteria are 'bad' in relation to obesity, e.g., *Firmicutes*, while others are 'good', e.g., *Bacteroidetes* [64]) can be modified or modulated/influenced, as necessary, through both individual- and population-group-specific interventions to make our gut microbiomes work best for us, e.g., through targeted interventions involving specific dietary and lifestyle modifications/mass health education about diet, such as encouraging the consumption of high-fibre diet, sauerkraut (cabbage), yoghurt and kimchi, etc., all of which are known to promote good gut bacteria.

A few additional notes

It should be noted that the demonstrator discussed above is all about making sense of 'crude' population data aggregates for public health purposes. Therefore, it is not about collecting ultra-precise individual patient data or the precise clinical management of individual patients. Moreover, there will be other research opportunities related to this demonstrator that were not fully developed or described in this article. For example, task clusters 4 and 5 can additionally help generate new hypotheses for further clinical and epidemiological research beyond the scope of this demonstrator, as an extra advantage of having such big population lifestyle data at hand. An open platform/API (application programming interface) or a 'data cooperative' can be offered as one of the demonstrator's outputs to enable other research groups, nationally and internationally, to interrogate and interact with select sets of the demonstrator's data and analytics, subject to governing data regulations and adequate data privacy and security safeguards. The companion 'Additional file 1' provides further important details and pointers to the research literature that complement the material presented in this article.

Conclusions

Geo-tagged big data from smartphones, wearables and other sensors are enabling researchers to conduct innovative studies in OO and T2D. (Smartphones are not just useful for data collection; they can also be used to deliver some location-based targeted public health interventions and campaigns.) Public health professionals can greatly benefit from well-conceived big data dashboards and related technologies in unveiling and acting upon the multifaceted challenges of OO and T2D in their target populations. The roadmap presented in this article with its list of high-level task clusters should provide a good start for teams willing to develop these dashboards for smarter public health decision-making in OO and T2D control and prevention in their locales. Dashboards must always be designed and developed in such a way that accommodates and streamlines future updates of not just the software components, but also the consumed datasets (e.g., adding new population genetic/epigenetic datasets and population gut microbiome profiles/trends when they become available) and the underpinning clinical/public health evidence and guidelines (as science progresses). It is hoped this article will initiate and stimulate further fruitful discussions among public health communities worldwide, and inspire many future groundbreaking studies about food and PA environments and population factors in OO and T2D.

Disclaimer

Reference in the manuscript to any specific commercial product, process or service by trade name, trademark, manufacturer or otherwise does not necessarily constitute or imply its endorsement, recommendation or favouring by the authors or the entities they are affiliated to, and shall not be used for commercial advertising or product endorsement purposes.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12942-021-00266-0.

> **Additional file 1.** A feasibility demonstrator roadmap (supplementary slide set).

Author details
[1] School of Information Management, Sun Yat-Sen University, East Campus, Guangzhou 510006, Guangdong, China. [2] Department of Geography, The University of Hong Kong, Pokfulam RD, Hong Kong, China.

References

1. The World Health Organization. WHO Coronavirus Disease (COVID-19) Dashboard. https://covid19.who.int. Accessed 3 Jan 2021.
2. The Lancet Microbe. COVID-19 vaccines: the pandemic will not end overnight. Lancet Microbe. 2020. https://doi.org/10.1016/S2666-5247(20)30226-3.
3. The World Health Organization. Non communicable diseases (Fact sheet, 1 June 2018). https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases. Accessed 6 Jan 2021
4. Gribsholt SB, Pedersen L, Richelsen B, Thomsen RW. Validity of ICD-10 diagnoses of overweight and obesity in Danish hospitals. Clin Epidemiol. 2019;11:845–54. https://doi.org/10.2147/CLEP.S214909.
5. ICD-10 Version:2016. https://icd.who.int/browse10/2016/en#/E10. Accessed 6 Jan 2021
6. Alberca RW, De Oliveira LM, Branco ACCC, Pereira NZ, Sato MN. Obesity as a risk factor for COVID-19: an overview. Crit Rev Food Sci Nutr. 2020. https://doi.org/10.1080/10408398.2020.1775546.
7. Vas P, Hopkins D, Feher M, Rubino F, Whyte MB. Diabetes, obesity and COVID-19: a complex interplay. Diabetes Obes Metab. 2020;22(10):1892–6. https://doi.org/10.1111/dom.14134.
8. The GBD 2015 Obesity Collaborators. Health Effects of Overweight and Obesity in 195 Countries over 25 Years. *N Engl J Med*. 2017;377(1):13-27. Doi: https://doi.org/10.1056/NEJMoa1614362
9. International Diabetes Federation. Diabetes facts & figures (2020. https://www.idf.org/aboutdiabetes/what-is-diabetes/facts-figures.html). Accessed 6 Jan 2021
10. Ampofo AG, Boateng EB. Beyond 2020: Modelling obesity and diabetes prevalence. Diabetes Res Clin Pract. 2020;167:108362. https://doi.org/10.1016/j.diabres.2020.108362.
11. Saeedi P, Petersohn I, Salpea P, et al. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9th edition. Diabetes Res Clin Pract. 2019;157:107843. https://doi.org/10.1016/j.diabres.2019.107843.
12. El-Sayed AM, Scarborough P, Galea S. Unevenly distributed: a systematic review of the health literature about socioeconomic inequalities in adult obesity in the United Kingdom. BMC Public Health. 2012;12(1):18. https://doi.org/10.1186/1471-2458-12-18.
13. Chaufan C, Yeh J, Ross L, Fox P. You can't walk or bike yourself out of the health effects of poverty: active school transport, child obesity, and blind spots in the public health literature. Crit Public Health. 2015;25(1):32–47. https://doi.org/10.1080/09581596.2014.920078.
14. Jørgensen ME, Christensen DL. Ethnicity and obesity: why are some people more vulnerable? Int Diabetes Monit. 2008;20(5):9.
15. Mathur R, Farmer RE, Eastwood SV, Chaturvedi N, Douglas I, Smeeth L. Ethnic disparities in initiation and intensification of diabetes treatment in adults with type 2 diabetes in the UK, 1990–2017: A cohort study. PLOS Med. 2020;17(5):e1003106. https://doi.org/10.1371/journal.pmed.1003106.
16. Golden SH, Yajnik C, Phatak S, Hanson RL, Knowler WC. Racial/ethnic differences in the burden of type 2 diabetes over the life course: a focus on the USA and India. Diabetologia. 2019;62(10):1751–60. https://doi.org/10.1007/s00125-019-4968-0.
17. Massey CN, Appel SJ, Buchanan KL, Cherrington AL. Improving diabetes care in rural communities: an overview of current initiatives and a call for renewed efforts. Clin Diabetes. 2010;28(1):20–7. https://doi.org/10.2337/diaclin.28.1.20.
18. International Diabetes Federation. Demographic and geographic outline (IDF Diabetes Atlas, 9th edition 2019). https://diabetesatlas.org/en/sections/demographic-and-geographic-outline.html. Accessed 6 Jan 2021
19. Luhar S, Timæus IM, Jones R, et al. Forecasting the prevalence of overweight and obesity in India to 2040. PLoS ONE. 2020;15(2):e0229438. https://doi.org/10.1371/journal.pone.0229438.
20. The World Obesity Federation (World Obesity). BLOG | Obesity and the SDGs: an opportunity hidden in plain sight. https://www.worldobesity.org/news/blog-obesity-and-the-sdgs-an-opportunity-hidden-in-plain-sight. Accessed 7 Jan 2021
21. US CDC. Genes and obesity (17 May 2013). https://www.cdc.gov/genomics/resources/diseases/obesity/obesedit.htm. Accessed 8 Jan 2021
22. Herder C, Roden M. Genetics of type 2 diabetes: pathophysiologic and clinical relevance. Eur J Clin Invest. 2011;41(6):679–92. https://doi.org/10.1111/j.1365-2362.2010.02454.x.
23. Mayo Clinic. Obesity - Symptoms and causes (18 November 2020. https://www.mayoclinic.org/diseases-conditions/obesity/symptoms-causes/syc-20375742). Accessed 8 Jan 2021
24. Zhang X, Christoffel KK, Mason M, Liu L. Identification of contrastive and comparable school neighborhoods for childhood obesity and physical activity research. Int J Health Geogr. 2006;5(1):14. https://doi.org/10.1186/1476-072X-5-14.
25. Duncan DT, Castro MC, Gortmaker SL, Aldstadt J, Melly SJ, Bennett GG. Racial differences in the built environment—body mass index relationship? A geospatial analysis of adolescents in urban neighborhoods. Int J Health Geogr. 2012;11(1):11. https://doi.org/10.1186/1476-072X-11-11.
26. Okuyama K, Abe T, Hamano T, et al. Hilly neighborhoods are associated with increased risk of weight gain among older adults in rural Japan: a 3-years follow-up study. Int J Health Geogr. 2019;18(1):10. https://doi.org/10.1186/s12942-019-0174-z.
27. Cebrecos A, Díez J, Gullón P, Bilal U, Franco M, Escobar F. Characterizing physical activity and food urban environments: a GIS-based multicomponent proposal. Int J Health Geogr. 2016;15(1):35. https://doi.org/10.1186/s12942-016-0065-5.
28. Hanibuchi T, Kondo K, Nakaya T, et al. Neighborhood food environment and body mass index among Japanese older adults: results from the Aichi Gerontological Evaluation Study (AGES). Int J Health Geogr. 2011;10(1):43. https://doi.org/10.1186/1476-072X-10-43.
29. Forsyth A, Van Riper D, Larson N, Wall M, Neumark-Sztainer D. Creating a replicable, valid cross-platform buffering technique: The sausage network buffer for measuring food and physical activity built environments. Int J Health Geogr. 2012;11(1):14. https://doi.org/10.1186/1476-072X-11-14.
30. Owens PM, Titus-Ernstoff L, Gibson L, Beach ML, Beauregard S, Dalton MA. Smart density: a more accurate method of measuring rural residential density for health-related research. Int J Health Geogr. 2010;9(1):8. https://doi.org/10.1186/1476-072X-9-8.
31. VoPham T, Hart JE, Laden F, Chiang Y-Y. Emerging trends in geospatial artificial intelligence (geoAI): potential applications for environmental epidemiology. Environ Health. 2018. https://doi.org/10.1186/s12940-018-0386-x.
32. Kamel Boulos MN, Peng G, VoPham T. An overview of GeoAI applications in health and healthcare. Int J Health Geogr. 2019;18(1):7. https://doi.org/10.1186/s12942-019-0171-2.
33. McGrath MJ, Scanaill CN. Wellness, fitness, and lifestyle sensing applications. In: McGrath MJ, Scanaill CN, editors. Sensor technologies: healthcare, wellness, and environmental applications. New York: Apress; 2013. p. 217–48. https://doi.org/10.1007/978-1-4302-6014-1_10.
34. Timmins KA, Green MA, Radley D, Morris MA, Pearce J. How has big data contributed to obesity research? A review of the literature. Int J Obes. 2018;42(12):1951–62. https://doi.org/10.1038/s41366-018-0153-7.

35. Christakis NA, Fowler JH. The spread of obesity in a large social network over 32 years. N Engl J Med. 2007;357(4):370–9. https://doi.org/10.1056/NEJMsa066082.

36. Hieronymi A. Understanding systems science: a visual and integrative approach. Syst Res Behav Sci. 2013;30(5):580–95. https://doi.org/10.1002/sres.2215.

37. Koh K, Reno R, Hyder A. Examining disparities in food accessibility among households in Columbus, Ohio: an agent-based model. Food Secur. 2019;11(2):317–31. https://doi.org/10.1007/s12571-019-00900-7.

38. Vojnovic I, Ligmann-Zielinska A, LeDoux TF. The dynamics of food shopping behaviour: Exploring travel patterns in low-income Detroit neighborhoods experiencing extreme disinvestment using agent-based modeling. PLoS ONE. 2020;15(12):e0243501. https://doi.org/10.1371/journal.pone.0243501.

39. Hirsch GB, Homer J. System Dynamics Applications to Health Care in the United States. In: Dangerfield B, editor. System dynamics: theory and applications. Encyclopedia of complexity and systems science series. New York: Springer US; 2020. p. 209–27. https://doi.org/10.1007/978-1-4939-8790-0_270.

40. NCD Risk Factor Collaboration (NCD-RisC). Trends in adult body-mass index in 200 countries from 1975 to 2014: a pooled analysis of 1698 population-based measurement studies with 19·2 million participants. Lancet. 2016;387(10026):1377-1396. https://doi.org/10.1016/S0140-6736(16)30054-X

41. Wang L, Gao P, Zhang M, et al. Prevalence and ethnic pattern of diabetes and prediabetes in China in 2013. JAMA. 2017;317(24):2515. https://doi.org/10.1001/jama.2017.7596.

42. Kamel Boulos MN, Yang SP. Mobile physical activity planning and tracking: a brief overview of current options and desiderata for future solutions. *mHealth*. 2021;7:13. https://doi.org/10.21037/mhealth.2020.01.01 https://mhealth.amegroups.com/article/view/38687/html

43. NCD Risk Factor Collaboration (NCD-RisC). Rising rural body-mass index is the main driver of the global obesity epidemic in adults. *Nature*. 2019;569:260–264. https://www.nature.com/articles/s41586-019-1171-x

44. Rateni G, Dario P, Cavallo F. Smartphone-based food diagnostic technologies: a review. Sensors (Basel). 2017;17(6):1453. https://doi.org/10.3390/s17061453.

45. Pendergast FJ, Ridgers ND, Worsley A, McNaughton SA. Evaluation of a smartphone food diary application using objectively measured energy expenditure. Int J Behav Nutr Phys Act. 2017;14(1):30. https://doi.org/10.1186/s12966-017-0488-9.

46. Teixeira V, Voci SM, Mendes-Netto RS, da Silva DG. The relative validity of a food record using the smartphone application MyFitnessPal. Nutr Diet. 2018;75(2):219–25. https://doi.org/10.1111/1747-0080.12401.

47. Kenney EL, Gortmaker SL. United States adolescents' television, computer, videogame, smartphone, and tablet use: associations with sugary drinks, sleep, physical activity, and obesity. J Pediatr. 2017;182:144–9. https://doi.org/10.1016/j.jpeds.2016.11.015.

48. Patja K, Jousilahti P, Hu G, Valle T, Qiao Q, Tuomilehto J. Effects of smoking, obesity and physical activity on the risk of type 2 diabetes in middle-aged Finnish men and women. J Intern Med. 2005;258(4):356–62. https://doi.org/10.1111/j.1365-2796.2005.01545.x.

49. Amazon Halo: A better measure of health. https://www.aboutamazon.com/news/devices/a-better-measure-of-health. Accessed 29 Dec 2020

50. Klous G, Smit LAM, Borlée F, et al. Mobility assessment of a rural population in the Netherlands using GPS measurements. Int J Health Geogr. 2017;16(1):30. https://doi.org/10.1186/s12942-017-0103-y.

51. Stevens M, D'Hondt E. Crowdsourcing of Pollution Data using Smartphones. In: Proceedings of UbiComp '10, Copenhagen, Denmark, 26–29 September 2010. http://soft.vub.ac.be/Publications/2010/vub-tr-soft-10-15.pdf

52. Hurwitz J, Nugent A, Halper F, Kaufman M. Unstructured Data in a Big Data Environment (dummies). https://www.dummies.com/programming/big-data/engineering/unstructured-data-in-a-big-data-environment/. Accessed 16 Jan 2021

53. Kamel Boulos MN, Yassine A, Shirmohammadi S, Namahoot CS, Brückner M. Towards an "Internet of Food": food ontologies for the internet of things. Future Internet. 2015;7(4):372–92. https://doi.org/10.3390/fi7040372.

54. Bano M, Zowghi D. A systematic review on the relationship between user involvement and system success. Inf Softw Technol. 2015;58:148–69. https://doi.org/10.1016/j.infsof.2014.06.011.

55. Zhuang M, Concannon D, Manley E. A framework for evaluating dashboards in healthcare. ArXiv (Preprint, 10 September 2020). https://arxiv.org/abs/2009.04792

56. Lazer D, Kennedy R, King G, Vespignani A. Big data. The parable of Google Flu: traps in big data analysis. Science. 2014;343(6176):1203–5. https://doi.org/10.1126/science.1248506.

57. Lock O, Bednarz T, Leao SZ, Pettit C. A review and reframing of participatory urban dashboards. City Cult Soc. 2020;20:100294. https://doi.org/10.1016/j.ccs.2019.100294.

58. Ma Z, Chen M, Yue S, et al. Activity-based process construction for participatory geo-analysis. GIScience Remote Sens. 2020. https://doi.org/10.1080/15481603.2020.1868211.

59. O'Donnell E. Zip Code vs. Genetic Code. *Harvard Magazine* (online). 2019 (Jul-Aug). https://harvardmagazine.com/2019/07/nature-nurture-genetics. Accessed 29 Dec 2020

60. Kanherkar RR, Bhatia-Dey N, Csoka AB. Epigenetics across the human lifespan. Front Cell Dev Biol. 2014. https://doi.org/10.3389/fcell.2014.00049.

61. Lopomo A, Burgio E, Migliore L. Epigenetics of obesity. Prog Mol Biol Transl Sci. 2016;140:151–84. https://doi.org/10.1016/bs.pmbts.2016.02.002.

62. Schutte NM, Nederend I, Hudziak JJ, Bartels M, de Geus EJC. Heritability of the affective response to exercise and its correlation to exercise behavior. Psychol Sport Exerc. 2017;31:139–48. https://doi.org/10.1016/j.psychsport.2016.12.001.

63. Crovesy L, Rosado EL. Interaction between genes involved in energy intake regulation and diet in obesity. Nutrition. 2019;67–68:110547. https://doi.org/10.1016/j.nut.2019.06.027.

64. Davis CD. The gut microbiome and its role in obesity. Nutr Today. 2016;51(4):167–74. https://doi.org/10.1097/NT.00000000000001.

Publisher's Note