

Citizen generated data – an opportunity for public health?

August 2019

The expanding availability of digital technologies is generating large amounts of detailed health-related data outside formal healthcare settings. Interest in this citizen generated data (CGD) has to date focused on how the information it provides about an individual can be used to improve the health of that same individual. This policy briefing explores the potential CGD has to improve the health of populations.

Key points

- Citizen generated data has potential to inform and improve public health by filling in gaps in pooled health data, facilitating behaviour change and providing feedback on and to healthcare services
- Public health professionals should be made aware of the potential CGD has to improve population health
- It is early days for the use of CGD in healthcare and, while the aggregation of CGD could have benefits at the population level, the interpretation and actionability of CGD appears limited, unless it is used in conjunction with data from established sources
- Acceptance, adoption and sustained use of CGD for population health by the public, patients and professionals will depend on demonstrating clinical benefits
- Regulation of devices that generate and collect CGD and how that data can be analysed and used is unclear

POLICY BRIEFING

In order to prevent disease, promote health and prolong life, public health requires datasets to understand what the key health issues are and how they are distributed across society.

Where are the opportunities?

Interest in citizen generated data has up to now focused on how data about an individual can be used to improve the health of that person. In principle, the aggregation of CGD may also have benefits for populations by providing opportunities to:

- Enrich existing health datasets
- Obtain new sources of data
- Collect useful health data outside traditional healthcare settings
- Gain a more comprehensive view of individual and population health

CGD may also be useful in informing, evaluating, and even delivering innovative public health interventions.

Monitoring the population's health and their environment

Current gaps in data used to monitor population health include data on conditions known to be considerably underreported (e.g. mental health conditions) and conditions commonly managed without seeing a doctor (e.g. hayfever). Furthermore, most health data is currently gathered within a healthcare setting, but a diverse range of factors that influence health arise in the home and other environments. These are rarely captured, but potentially very valuable.

Examples – measurement of air pollution

- Citizen science initiatives have enabled communities to measure air quality. A project in London mapped nitrogen dioxide levels through citizens placing and collecting a series of diffusion tubes. Results were fed back to the communities in meetings and several suggestions were listed to be put forward to the Mayor and Local Authority, contributing to Putney Transport for London's decision to use hybrid buses on the high street¹.
- Technology is now enabling air pollution sensors to be attached to phones, potentially directly assessing personal exposure. A recent study found several limitations that may hinder its use in continuous monitoring of pollution, or measuring individual exposure levels, but it may be useful for comparative (rather than absolute) measures and assessments².

1. mappingforchange.org Citizen science used to map community air quality

2. Nyarku M, Mazaheri M, Jayaratne R, et al. [Mobile phones as monitors of personal exposure to air pollution: Is this the future?](#) PLoS ONE 13(2): e0193150. (2018)

POLICY BRIEFING

Citizen generated data could help to address these data gaps. For example, smartphone associated sensor technology could help to quantify determinants such as pollution (see example box) or internet search data could be used to monitor population-level health symptoms, such as the spread of flu. For some conditions there is evidence that health data generated from citizen sources correlates well with health data gathered via more established methods (such as clinical consultations). However, lack of information about, for example, the person searching online and why they are searching, greatly limits the scope for accurately interpreting and taking action on CGD unless it is in conjunction with data from established sources.

The need for accurate, consistent data over time is also limiting the current utility of CGD for health surveillance, although there is potential for providing targeted qualitative information at a local level to better profile the health of a community. Such insights may prove useful in informing public health initiatives.

Facilitating behaviour change

Promoting healthy lifestyle choices and behavioural changes is a significant public health challenge. The tools that generate CGD could be used to prompt and target behaviour change, for example through a combination of mobile phone applications and wearable technology to monitor and encourage a person's physical activity.

However current evidence indicates mixed results as to whether there has been any impact. Where change has been reported, it was generally short term with a high rate of drop off over time. When used to prompt behaviour change, there is better evidence of impact when targeting to individuals who could benefit most from them, but the challenge to sustain use remains.

Evaluating healthcare services

Patients are using the internet to describe their experiences of healthcare, providing a readily available source of information to inform healthcare improvement and ultimately the quality of healthcare services. The use of internet captured views is currently being assessed by groups such as HDRUK and the Health Foundation to determine whether CGD can be a source of viable information.

Early days for CGD powered public health

The utility of CGD to public health is just beginning to be explored and understanding how best to use this data and approaches for doing so still need to be developed. Actual use of CGD in public health has tended to be geographically localised, and limited to small scale studies. Scope for widescale adoption is low, constrained by general infrastructure challenges as well as the quality of available studies.

Whilst CGD may support surveillance and inform current public health initiatives, it will not replace them for the foreseeable future. Developing the use of CGD to improve public health will depend on finding answers to important questions:

Is the data of sufficient quality and manageable? There are various dimensions to consider, including:

- **Validity** – data are produced which may not be representative of the general population due to differences between groups within the population
- **Accuracy** – people tend to modify their behaviour when they know they are being observed, which would undermine the accuracy of the data generated

POLICY BRIEFING

- **Volume and range** – the volume of health related data generated by individuals is rapidly expanding as are the formats in which it is available creating challenges for implementation and utilisation
- **Integration with healthcare data** – the variety of types of data generated makes developing appropriate standards and infrastructure to manage data curation and ensure interoperability particularly challenging
- **Access** – many CGD related digital technologies have been developed by commercial organisations, and any proprietary rights over the resulting data could have an impact on the availability of data for public health applications

Acceptance, adoption and sustained use by the public, patients and professionals will depend on demonstrating how data is of clinical benefit. To date studies have not generally focused on the impact on clinical outcomes of this nascent development.

How can citizens be encouraged to share data?

Public trust is crucial if the health system is to have access to data generated by citizens where the primary intent was not to infer conclusions about their personal health. The regulation around the devices that generate and collect CGD and how data can be analysed and used is uncertain.

Will it widen health inequalities?

The population group producing and using certain types of CGD (e.g. data from health wearables) are generally healthier – differential usage and acceptance will need to be factored into design and development for its use to be targeted for public health purposes. Furthermore, there may be differences in the demographics of those who are willing to share data.

Considerations for policymakers

While there are many challenges, the changing ways in which health data is generated and the scope of CGD means its potential benefits for public health should not be dismissed.

Public health professionals should be aware of its potential and what is necessary to progress work in this field. To realise the potential that CGD may offer public health as a support to established data sources, end-users must be involved in the design of data gathering processes and devices. Collaboration between technology companies and public health providers will also be essential.

Authors: Helen Green and Sobia Raza

Contact: intelligence@phgfoundation.org

PHG Foundation is a health policy think tank with a special focus on how genomics and other emerging health technologies can provide more effective, personalised healthcare