NIHR Policy Research Programme Reviews Facility
To support national policy development and implementation

Precision public health
A critical review of the opportunities and obstacles

Dylan Kneale, Theo Lorenc, Alison O'Mara-Eves, Quan nha Hong
Katy Sutcliffe, Amanda Sowden, James Thomas

September 2020

The NIHR Policy Research Programme Reviews Facility is a collaboration between the following:

UCL Institute of Education
EPPI CENTRE
LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE
UNIVERSITY of York

Precision public health
A critical review of the opportunities and obstacles

Dylan Kneale, Theo Lorenc, Alison O'Mara-Eves, Quan nha Hong
Katy Sutcliffe, Amanda Sowden, James Thomas

September 2020

The authors of this report are:
Dylan Kneale[1], Theo Lorenc[2], Alison O'Mara-Eves[1], Quan nha Hong[1],
Katy Sutcliffe[1], Amanda Sowden[2], James Thomas[1]

Editorial and cover design by: Lionel Openshaw[1]

[1] EPPI-Centre, Social Science Research Unit, UCL Social Research Institute, University College London.
[2] Centre for Reviews and Dissemination, University of York.

Acknowledgements: We would like to thank the thoughtful input of our peer reviewers in helping to strengthen this work.

Funding
This review was commissioned by the National Institute for Health Research (NIHR) Policy Research Programme (PRP) for the Department of Health and Social Care (DHSC). It was funded through the NIHR PRP contract with the EPPI Centre at UCL (Reviews facility to support national policy development and implementation). The views expressed in this publication are those of the authors and not necessarily those of the NHS, the NIHR or the DHSC.

Funder involvement
This work is part of an ongoing programme of work funded by DHSC. Throughout the review, DHSC and Public Health England policy team members were consulted to understand the context of the issue under study, and to identify the focus and scope of the analysis.

Conflicts of interest
There were no conflicts of interest in the writing of this report.

Contributions
The opinions expressed in this publication are not necessarily those of the EPPI-Centre or the funders. Responsibility for the views expressed remains solely with the authors.

This report should be cited as:
Kneale D, Lorenc T, O'Mara-Eves A, Hong QN, Sutcliffe K, Sowden A, Thomas J (2020)
Precision public health – A critical review of the opportunities and obstacles. London: EPPI-Centre, Social Science Research Unit, UCL Institute of Education, University College London.

ISBN: 978-1-911605-17-1

Cover photo by Alina Grubnyak on Unsplash

Contents

Table of figures

Table of tables

Key Messages

Background

The term ‘precision public health’ (PPH) refers to a new approach in public health which involves the use of novel data sources and/or computer science-driven methods of data analysis to predict risk or outcomes, in order to improve how interventions are targeted or tailored, with the aim of making them more individualised and therefore more effective and cost-effective than methods currently in use. These data may include, for example, information from social media or devices, genomic or clinical data, and information from healthcare services.

Aims and methods

In this critical review, which was conducted between March and October 2019, we outline key assumptions underpinning the PPH approach and identify potential challenges in its application. We adopted a pragmatic, non-systematic review methodology to examine: (i) the general principles underlying PPH; (ii) the validity of claims made about PPH in empirical studies and commentaries; and (iii) the potential opportunities and challenges of adopting a PPH approach through examining two case studies: health checks and community-based interventions. Non-empirical studies (commentaries and think-pieces) were included in this review because PPH represents an emerging approach and many of the ideas around the potential of PPH are only described in such studies. There remains a need to develop an empirical evidence base around PPH.

Principles of PPH

PPH as an approach rests on a number of assumptions. These include:

- The data required will be available and reliable
- Novel data (e.g. genomic data) and computer science-driven analysis methods (e.g. machine learning) provide better estimates of individual risk
- Better estimates of risk will translate to more effective interventions
- Data can be collected and used in an ethically responsible way

Assessing the evidence on PPH

There may be merit in the greater use of novel data and analysis methods to improve the health of populations, although there is limited direct empirical evidence showing PPH to be effective, and the arguments in its favour are often not well supported by evidence. Analysis of commentaries suggests that (i) the PPH field may be highly influenced by commentary and (non-systematic) review pieces that lack transparent methods but make claims about the potential of PPH; (ii) commentators on PPH often attempt to provide evidence for the claims they make, but the link between the evidence and the claim is rarely substantiated; and (iii) many of the assumptions underlying PPH have no underpinning evidence. This suggests a need for a measured approach to the adoption of PPH, alongside a programme of evaluation measuring implementation processes and effectiveness of PPH approaches. In addition, PPH seems to be largely based on individual determinants of health, particularly information

provision, although the evidence suggests this is not a promising approach. This focus contrasts with the increasing emphasis within public health practice on improving the wider social determinants of health.

Potential applications of PPH

Case study 1: Health checks

Health checks involve screening populations for clinical or behavioural risk factors and use the findings to improve risk management and support behaviour change. Although evidence on the application of PPH approaches to health checks is largely speculative at this stage and subject to further research, PPH approaches could potentially be incorporated into health checks in a number of ways, including:

- A digital health check, with only higher-risk patients being referred for clinical testing
- Tailored interventions to support risk management after the health check
- Low-cost interventions such as apps to support behaviour change

Case study 2: Community-based interventions

Community-based interventions aim to create multilevel change by identifying and building on resources in the community itself. Although speculative at this stage and subject to further research, PPH approaches could be applied to community-based interventions in several ways including:

- Using PPH principles, new data sources and analysis methods (e.g. social media, social network analysis, sentiment analysis) could help to identify communities of need, to pinpoint subgroups within communities, and to better describe how social norms are propagated within communities.
- Using new data to elicit community views to tailor interventions
- Involving communities in designing and delivering interventions

What could come next?

Defining PPH is contentious and our findings reflect the difficulty in assessing and operationalising the broad ambition of using emerging data and technologies to better understand profiles, predict risk and outcomes, and act upon this evidence. Future work in this area should seek to introduce more focus around the concept of PPH, including being clearer about the goals and breaking down the concept into a series of components that can each be evaluated.

The bulk of the work presented here took place between March and October 2019. There is scope for further analysis to understand the potential of PPH in the future, as the number of studies adopting a PPH approach grows. This larger pool of studies may also lend itself to more systematic approaches to reviewing the evidence, particularly if there is an interest in evaluating a particular component or principle of PPH. In addition, the evidence examined in this report predates the COVID-19 global pandemic, and many of the measures taken to mitigate the spread of the pandemic may provide a further source of evidence and data to understand the potential role of PPH in public health decision-making.

Conclusions

- There are few empirical evaluations of PPH as an approach.
- Theoretically, PPH could represent a shift towards a more individualised approach to public health decision-making which is at odds with a focus on the wider (social) determinants of health. There may be merit in further consideration of how the approaches included within PPH can simultaneously provide insight on the multilevel factors and social determinants that influence individual behaviours. However, this did not appear to be a strong priority in several of the applications or theoretical arguments put forward around PPH.
- Arguments for PPH often rest on assumptions which are not supported by empirical evidence. There are few studies evaluating implementation processes and effectiveness.
- However, PPH could inform incremental improvements to a range of public health interventions, and while the impacts may be modest, they may nevertheless be of value at a population level.
- There are unanswered questions around the ethical and social implications of PPH.

Summary

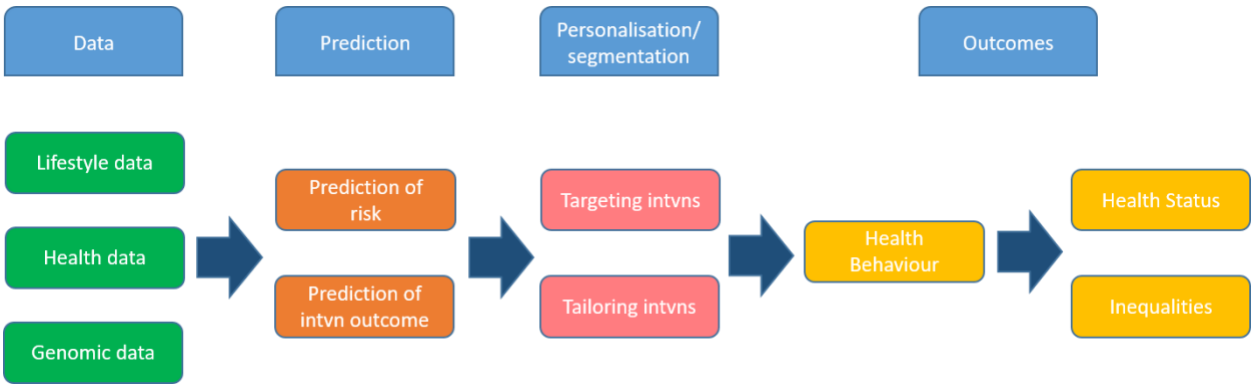
Introduction

The term ‘precision public health’ (PPH) refers to a range of new approaches for use in public health. These new approaches involve the use of novel data sources and/or computer science-driven methods of data analysis to predict risk or outcomes, in order to improve how interventions are targeted or tailored, with the aim of making them more individualised and therefore more effective and cost-effective. These data include information gathered through social media or devices, genomic or clinical data, and information from healthcare services.

The aim of this critical review is to outline key assumptions underpinning the PPH approach and to identify potential challenges in its application. Between March 2019 and October 2019, we adopted a pragmatic, non-systematic review methodology. Non-empirical studies (commentaries and think-pieces) were included in this review because PPH represents an emerging approach and many of the ideas about its potential of PPH are only described in these think pieces. There remains a need to develop an empirical evidence to support the adoption of PPH. We examine the characteristics of studies that represent examples of PPH. We then appraise the evidence claims and the lines of argumentation emerging around PPH using a structured approach (Toulmin’s Model of Argument). Finally, we look at the general principles behind PPH, and then consider its application to two case studies: health checks and community-based interventions.

Principles of PPH

PPH involves the use of *data* to enable intervention providers to *target* interventions to the right people and/or *tailor* them to suit their needs. Underpinning the success of these approaches is the assumption that it is possible to use these new data sources to *predict health risk* to target interventions, and to *predict outcomes* to determine which tailored interventions will have the largest impact on health *outcomes* via changes in behaviours. PPH theory thus rests on a series of dependencies, shown schematically below with a focus on changing health behaviours as a means of improving health status and reducing inequalities. Our definition of an intervention is broad in scope and we include any planned ‘set of actions with a coherent objective to bring about change or produce identifiable outcomes’ within this (Rychetnik, Frommer et al. 2002). Public health interventions are those that ‘promote or protect health or prevent ill health in communities or populations’ (Rychetnik, Frommer et al. 2002), as opposed to clinical interventions which have the same ambition albeit at an individual level. Interventions focussed on changing health behaviours may follow more individualised models, for example through health checks (see case study 1), or may target health behaviours through mobilising community-level changes (see case study 2) or through making environmental or system-wide changes, or through a combination of these strategies. Public health interventions therefore take place within a number of community settings, including primary care and clinical settings, as well as, for example, in schools, libraries or through social media.



Summary figure 1. Dependencies in the implementation of PPH.

Hence, PPH involves a series of assumptions, and as yet there is no empirical evidence in support of several of these:

- *Data are available, usable, and applicable to the broader population*. While new forms of data are promising, there are practical and ethical barriers to their use in many cases, and there may be biases in the data which limit their applicability.

- *New data*[1] *(e.g. genomic data) and computer science-driven methods of analysis (e.g. machine learning) provide significant benefits in terms of predicting risks and outcomes as against existing methods*. While work is ongoing in this area, current results generally show only incremental improvements in accuracy. Claims that new methods can result in accurate forecasting of individual risks and outcomes are largely unfounded.

- *Tailored and targeted interventions are more likely to be effective and cost-effective*. While there is some evidence in favour of tailored information provision, there is currently limited direct evidence for the benefits of targeted approaches. There is also a compelling theoretical case and empirical evidence for the benefits of some universal interventions.

- *New data sources offer significant benefits for tailoring or targeting as against existing data*. While new data may offer more precise targeting of individuals as against existing (e.g. demographic or clinical) data, the likely improvements in effectiveness are probably limited. This said, more research on specific strategies is needed.

- *Individual behaviour change interventions are an effective approach to creating sustained changes and improving public health in general*. PPH approaches tend to focus on individual determinants of behaviour, implying a focus on interventions which

[1] Note, novel data sources may be used alongside existing or traditional data sources, although the defining feature of a PPH approach is that there exist new data sources and/or new (AI-driven) ways of analysing large data sources.

seek to provide individuals with information and/or to change the factors influencing decisions at an individual level, as opposed to broader environmental or community-based intervention. Current evidence finds that the benefits of many individual approaches may be relatively modest.

Assessing the evidence on PPH

With the exception of genomic studies, existing empirical PPH studies mostly offer evidence of greater precision in terms of ecological-level data (for example using new approaches to better estimate the burden/risk of disease on an area level), potentially allowing for areas to be targeted more efficiently. There are fewer studies that incorporate micro-level data or that use data to enable the more precise tailoring of interventions.

Commentary studies emphasise that precision can be achieved in targeting interventions towards narrow social profiles of people through the incorporation of data reflecting micro-level day-to-day insights into the lives of individuals.

Structured analysis of commentary studies shows that (i) the PPH field may be highly influenced by commentary and non-systematic review pieces that lack transparent methods but make claims about the potential of PPH; (ii) commentators on PPH often attempt to provide evidence for claims but the link between the evidence and the claim is often unsubstantiated when critically examined; and (iii) many of the assumptions underlying PPH are not supported by empirical evidence suggesting that there needs to be a measured approach to adopting PPH approaches. Claims around the effectiveness of PPH and around PPH being an advance on current public health approaches tended not to be supported by empirical evidence.

Potential Applications

Case study 1: health checks

Health checks, such as the NHS Health Check programme, involve screening populations for clinical or behavioural risk factors, and using this information to (a) inform clinical risk assessment and provide more timely intervention and (b) facilitate counselling on lifestyle risks and support behaviour change if required. The evidence on the effectiveness and cost-effectiveness of health checks is mixed. PPH approaches could be employed at several points in the health check pathway.

Population targeting: The evidence suggests that a more targeted approach would be more likely to be cost-effective, but not necessarily more effective. This could take the form of a digital health check offered to everyone, with face-to-face checks for those at higher risk.

Tailoring of interventions: A range of interventions may be implemented after the health check (lifestyle advice, preventive medication, referral to specialist services). PPH approaches could help to inform which interventions are offered to individuals, although the likely improvements in effectiveness are probably modest.

Ongoing support: PPH approaches could provide support after the health check, for example using apps or wearable devices to support behaviour change.

With the possible exception of population targeting, the impacts of these PPH approaches on effectiveness and cost-effectiveness are likely to be modest, but they are generally low-cost and merit further exploration.

Case study 2: community-based interventions

Community-based interventions create multilevel change by identifying and building on resources in the community itself. They target a group of people united by some common characteristic, which may be geographical (area-based interventions) or may be based on shared values, identities or behaviours. We identify how PPH approaches could be employed at distinct stages in the intervention pathway.

Clarifying focus: Community-based interventions see communities as the target, agents or resources of change. Using PPH principles, new data sources and analysis methods (e.g. social media, social network analysis, sentiment analysis) could help to identify communities of need, to pinpoint subgroups within communities, and to better describe how social norms are propagated within communities. The combination of data on online interactions with geo-location may be particularly promising. However, while such methods have been fairly widely used to characterise and identify communities, it is unclear how far this information translates into better interventions.

Identifying need: Community-based interventions are based on supporting communities to identify their own needs. PPH approaches facilitated by new technology could help with this process, particularly for disadvantaged populations, although there may be issues with reliability.

Community involvement: Community-based interventions aim to involve community members as much as possible in their design and delivery. PPH approaches such as social networks could help to deepen involvement partnerships and stimulate community empowerment, although there are relatively few examples of this in practice.

Overall, while some PPH-informed approaches may be valuable as supplementary processes for planning and implementing community-based interventions, they are not a substitute, and there is limited evidence or theory supporting their use. Adoption of such approaches could divert attention from structural or wider determinants of health and towards individual level behaviours.

What could come next?

Defining PPH is contentious and our findings reflect the difficulty in assessing and operationalising a broad ambition of using emerging data and technologies to better understand profiles, predict risk and outcomes, and act upon this evidence.

Two factors may be useful to consider as part of future developments. First, there may be utility in avoiding short ambiguous definitions of PPH in favour of establishing a set of principles by which PPH could be operationalised. Second, in order to evaluate the utility of PPH, it might be appropriate to link PPH to a particular form of data (e.g. data from social media interactions or digital apps), and then to regard PPH as a particular model of intervention that utilises these data. This could mean that PPH is viewed as a way of harnessing particular forms of data, often

reflecting the individual level, that is used alongside evidence around social determinants and wider health influences, to inform public health decision-making. This would be instead of seeing PPH as a completely separate approach and as a 'new way' of conducting public health practice.

The bulk of the work presented here took place between March and October 2019. There is scope for further analysis to understand the potential of PPH in the future, as the number of studies adopting a PPH approach grows. The term 'Precision Public Health' only emerged within the past decade and the self-defined PPH empirical studies were published only very recently, suggesting that this is a rapidly expanding area of interest. A larger pool of studies in the future may also lend itself to more systematic approaches to reviewing the evidence, particularly if there is an interest in evaluating a particular component or principle of PPH. Future systematic reviews in this area may also benefit from greater patient public involvement (PPI) in shaping the specific types of questions and concerns that should be accounted for, including issues around equity. For example, greater PPI involvement in developing review questions and sub-questions may steer reviewers to consider a greater range of potential ethical considerations and equity considerations.

In addition, the evidence examined in this report predates the COVID-19 global pandemic, and many of the measures taken to mitigate the spread of the pandemic may provide a further source of evidence and data to understand the potential role of PPH in public health decision-making.

Future work in this area should seek to introduce more focus around the concept of PPH including being clearer about the goals, and seek to break down what is currently an expansive definition into a series of components that can each be evaluated.

Key messages

As a relatively new concept, there is limited direct empirical evidence showing PPH to be effective, and the theoretical arguments in its favour are often not well supported by evidence. The more ambitious claims made for PPH in the literature often rest on questionable readings of the evidence – for example, citing the possibility of identifying subgroups of the population through better targeting as though this automatically promises greater effectiveness among interventions targeting those subgroups.

In practice, it seems that PPH is less a radically new paradigm and more a range of incremental improvements to public health interventions. The case studies outlined above indicate several ways in which new data or tools could be productively used to inform the design and implementation of public health interventions. Current evidence suggests the impact of these is likely to be fairly modest, although further focused research (e.g. exploring the utility of strategies for targeting or involving communities using PPH) may merit further exploration and evaluation.

The discourse around PPH is arguably counter-productive in that it focuses attention on individual characteristics and behaviours (although practical PPH approaches may be applicable at a broader level). The development of PPH should not detract from policies

addressing the social and structural determinants of health, and there should be greater focus on how PPH approaches can target and tailor action on the wider determinants of health.

The ethical and social implications of PPH, particularly when genomic data are used, are potentially challenging and need to be explored further.

Main report

1. Introduction

The potential for 'artificial intelligence' and new sources of data to enable interventions to be precisely targeted and tailored to the needs of specific individuals has been gaining increasing interest across a range of disciplines. Within clinical fields, Precision Medicine is described as 'an emerging approach for disease treatment and prevention that takes into account individual variability in genes, environment, and lifestyle for each person' (Prasad and Groop 2019, p40). This endeavour is also known as personalised medicine, P4 medicine (personalised, predictive, preventive, and participatory), and individualised medicine. Implementation of a precision medicine approach should see patients receiving the right treatment(s) with the right dose in the right sequence at the right time, with minimum harm and maximum effectiveness (Mirnezami, Nicholson et al. 2012). Accompanying the rise of 'precision medicine' there has been an active debate as to whether the same approaches can be applied in public health, and where the possible limitations may lie. Here we report on our work that set out to critically assess the claims and counter-claims made about precision public health.

The term 'precision public health' (PPH) was first coined by Public Health Practitioners in 2013 from the Health Department of Western Australia to complement parallel developments in precision medicine (Weeramanthri, Dawkins et al. 2018). This term was developed through forging partnerships across policy, practice and academia and was specifically intended to reflect developments in genomics, spatial technology in health, and data linkage. Since then, the term has bloomed across the health literature, with a search on PubMed[2] showing that some of the first academic publications only emerged in 2016 (four records), rising to 16 records and 25 records in 2017 and 2018 respectively, with 26 papers referencing PPH published to date by early October in 2019. The term has sparked interest among public health policy-makers globally, with recent developments including the designation of an Office of Genomics and Precision Public Health by the US Centers for Disease Control (CDC) (Khoury 2019). This rapidly increasing interest is a motivator for exploring the term and the emerging evidence base further.

PPH appears to involve the ambition of using new sources of data, technologies and computer science-driven methods (particularly Artificial Intelligence and Machine Learning) to predict risk and susceptibility to health conditions more granularly than is currently the case, and to use this information to generate more precise risk profiles for the design and personalising of interventions (including digital interventions) (Dolley 2018, Newton, Epke et al. 2018) and to design and target surveillance, health protection and health improvement programmes (Dowell, Blazes et al. 2016) in response. The definition of PPH is somewhat contested in the literature, and while we provide a working definition of PPH in the glossary below, we focus here on the assumptions that need to be met in order for new data and technologies to contribute to public

[2] A database of references and abstracts on medical, health and life sciences topics holding close to 30 million records.

health decision-making. This thinking is shaped both by considering PPH as well as an allied term, 'Predictive Prevention' (Newton, Epke et al. 2018), which was developed by Public Health England to describe the use of hitherto unused data to enable intervention providers to target appropriate interventions to the right people and tailor them to suit their needs, with a particular focus on data collected through digital apps. These assumptions underpin the remainder of the report and we explore the extent to which these assumptions are met, both conceptually and through exploring applied examples of PPH, using a number of different approaches.

<table>
<tr><td colspan="3">Public Health: The science and art of preventing disease, prolonging life and promoting health, through the organised efforts and informed choices of society, organisations, public and private, communities and individuals (Acheson 1998)</td></tr>
<tr><td colspan="3">Precision approaches for informing, designing and implementing interventions</td></tr>
<tr><td colspan="3">Precision Public Health (PPH): <u>In this report</u> we view PPH as a movement towards refining public health practice across a range of functions through (i) the greater use of new data sources on the genetic and behavioural profiles of individuals and populations; and (ii) the use of new methods of analysing structured and unstructured data, specifically using artificial intelligence</td></tr>
<tr><td colspan="2">Predictive Prevention: A movement towards targeting and tailoring of interventions based on hitherto unused data to enable intervention providers to target appropriate interventions to the right people and tailor them to suit their needs</td><td>Public Health Genomics: The use of genomic data to deliver preventive care and disease treatments with better specificity, tailored to the genetic makeup of each patient (see Khoury, Engelgau et al. 2018)</td></tr>
<tr><td colspan="2">Infodemiology: involving scanning the internet for user-contributed health related content, with the ultimate goal to improve public health (see Eysenbach 2009)</td><td></td></tr>
<tr><td>Ecological Momentary Interventions: Interventions provided to people during their everyday lives (i.e. in real time) and in natural settings (i.e. real world).</td><td>Ecological Momentary Assessments: Assessments of people's behaviour during their everyday lives (i.e. in real time) and in natural settings (i.e. real world).</td><td></td></tr>
</table>

Data sources for Precision Public health	
Big data: "Extensive datasets primarily in the characteristics of volume, variety, velocity, and/or variability – that require a scalable architecture for efficient storage, manipulation, and analysis" (Chang and Grady 2015, p5)	**Genomic data**: Data on the structure, function, evolution, mapping, and editing of genomes
Digital/data/health avatar: "a virtual representation of a person with all their associated health information" (Prosperi, Min et al. 2018)	**Polygenic risk scores:** Scores based on variation in multiple genetic loci and their associated weights for predicting particular conditions; can be incorporated into polygenic screening (Natarajan, Young et al. 2017)
Internet of things: "a global infrastructure for the information society, enabling advanced services by interconnecting (physical and virtual) things based on existing and evolving interoperable information and communication technologies" (Wortmann and Flüchter 2015)	**Other Omics data:** Including epigenomics (data measuring changes in gene expression that do not occur through changes the DNA sequence), transcriptomics (measures of the expression of RNA transcripts), metabolomics (data reflecting metabolites, the products of metabolism), and proteomics (data reflecting protein structures). Omics data can form the basis of various screening tests (Prosperi, Min et al. 2018)
Analysis tools for Precision Public Health	
Artificial Intelligence: Human intelligence processes undertaken by computer systems	
Machine learning: the use of algorithms and statistical models used by computer systems to perform a specific task without using explicit instructions, relying on inference and data patterns	
Ensemble modelling: "A process where multiple diverse models are created to predict an outcome"(Arora, Chandna et al. 2020 p6)	

2. Research objectives

The objectives of this report are to:

1. Develop an understanding of the parameters of precision public health (PPH) through focussing on the implicit assumptions made when adopting a PPH approach

2. Examine some of the challenges and opportunities discussed in the literature around PPH and how these may undermine or facilitate meeting the underpinning assumptions
3. Assess the validity of claims that are included within commentary studies around the potential application of PPH
4. Critically assess the features of empirical PPH studies to further understand the extent to which the assumptions made are verified
5. Present two case studies to understand the potential gains and challenges of adopting a PPH approach. This includes:
    a) Assessment of the extent to which a PPH approach is compatible with current intervention theory
    b) Assessment of the extent to which a PPH approach could improve on current practice

3. Approach and Methods

We undertook a critical review which provided a framework for critically evaluating the literature and developing new conceptual understandings surrounding PPH (Grant and Booth 2009). This involved synthesising evidence from a diverse set of sources and identification of models or hypotheses for further exploration (Grant and Booth 2009). We aimed to explore the underlying assumptions of PPH and whether these are considered in studies. Commentaries and think-pieces were included in this review because PPH represents an emerging approach and many of the ideas around the potential of PPH are only described in these studies, and there remains a need to develop an empirical evidence base around PPH. Key features of a critical review are the focus on interrogating a smaller pool of studies at the expense of systematicity (the identification of all studies on a given topic) (Grant and Booth 2009). Here, we worked with different types of literature – including commentaries or discussion pieces, which focused on the utility of PPH, and empirical studies identified as examples of PPH. This review was undertaken between March and October 2019 through:

(i) Establishing and understanding the underlying assumptions and considerations (involving targeted searches of the literature; a workshop; and follow-up research and meetings to discuss initial findings) (section 5)
(ii) Critically appraising the line of argument of commentary studies (drawing on Toulmin's Model of Argumentation) (section 6)
(iii) Critically appraising empirical PPH studies (section7)
Developing case studies of potential applications of PPH (sections 8 and 9)

The methods used to complete steps are outlined in full in Appendix 1.

4. Key assumptions underpinning Precision Public Health (PPH)

As described in the introduction, PPH can refer to a range of approaches. In this section we focus on constructing a coherent intervention theory to describe how the targeting and tailoring of interventions, and the use of novel types of data and methods of analysing data, are intended to improve health outcomes. Other public health functions including surveillance, health protection and the reduction of health inequalities are also encompassed. However, in line with current areas of interest of the DHSC/PHE, as expressed in recent policy announcements (DHSC 2019), our assumptions are focussed on the potential for PPH to target preventative interventions, change behaviours and improve health. The underpinning assumption is that new data sources can better predict health risks and intervention outcomes, which in turn can improve the targeting of interventions, as well as better identify which forms of intervention (tailoring) will have the largest impact on health status, often via preventative changes in health behaviours. It is possible for a PPH approach to be applied discretely to one link in the chain of dependencies as depicted in Figure 1, e.g. in the prediction of risk, although in this report we assess the promise of PPH against the widespread assumption that PPH approaches have the capacity to improve the overall health of populations (for example, Baynam, Bauskis et al. 2017, Dolley 2018). Similarly, while a PPH approach is viewed as offering improved prediction of health risk through 'applying emerging methods and technologies for measuring disease,

pathogens, exposures, behaviours, and susceptibility in populations', a PPH approach also requires that improved precision in the prediction of risk is actionable and is followed by 'developing policies and targeted implementation programs to improve health' (Khoury and Galea 2016, p1358).

The overall success of PPH thus depends on a chain of dependencies that begins with the availability of suitable data, requires that those data can be used to predict which interventions might be appropriate, and ends with the requirement that people's behaviour can be changed, and health outcomes improved, through the provision of targeted and tailored interventions (see Figure 1). There is research available that evaluates each of these dependencies to different extents. We outline here each of the key assumptions on which PPH depends and consider some of the arguments that support or question the validity of each assumption.

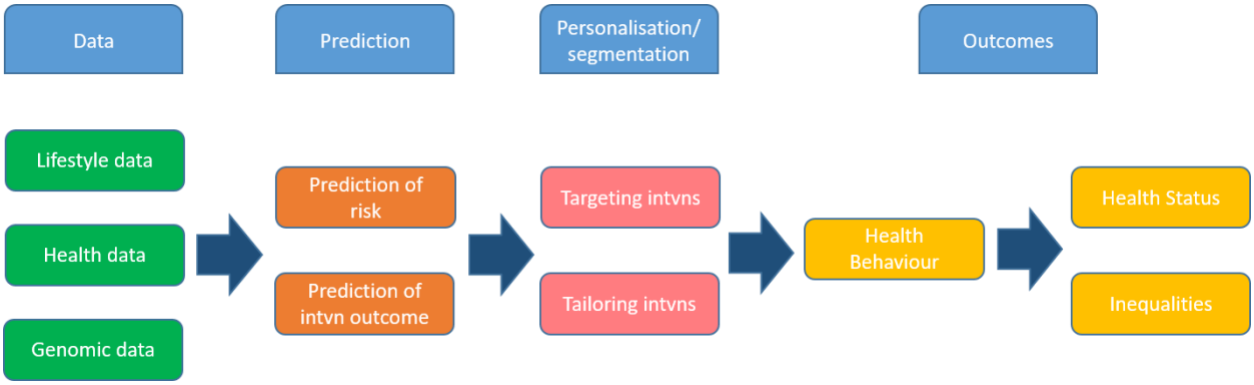


Figure 1: Dependencies in the implementation of PPH.

Assumption A: New data sets can provide usable data for public health interventions

The focus on (relatively) new data sources is key to the PPH approach[3]. The argument is that most existing data sets, such as surveys or electronic patient records, enable only a vague and general characterisation of factors, particularly behavioural factors, which can impact on health outcomes. Non-traditional sources of information from outside the health sector, sometimes called Big Data – such as wearables or social media analytics – could potentially provide highly granular data about individual behaviours, and about social relationships and environments. In some cases, they may also promise greater validity. For example, self-report data on health behaviours are notoriously unreliable for some health outcomes, and tracking behaviours directly might produce more accurate data, although the validity of such data over long time periods is not well understood.

These much more detailed data could then be used to undertake more sensitive analyses of the interactions between individual behaviours, social and environmental factors and health

[3] Note, novel data sources may be used alongside existing or traditional data sources, although the defining feature of a PPH approach is that there exist new data sources or new (AI-driven) ways of analysing large data sources. Without incorporating new data sources or new AI-driven analysis frameworks, there is little to distinguish a PPH approach from existing Public Health approaches.

outcomes. Such analyses could investigate the micro-level of individuals' day-to-day lives and interactions with other people and the environment, as well as the macro-level of population variables. However, this assumes that such data can be utilised in practice. Although considerable progress has been made in this area, there are numerous practical challenges involved in collecting or acquiring such data, and in linking different data sets so as to make them usable (Prosperi, Min et al. 2018).

In some cases, the data themselves may be skewed in ways which limit their applicability to the broader population. Concerns about genomic data have particularly focused on the underrepresentation of minority groups (see Dankwa-Mullan, Bull et al. 2015; Chowkwanyun, Bayer et al. 2018; Ramaswami, Bayer et al. 2018). Evidence which is based on unrepresentative samples will hinder understanding of the relationships between different exposures (including genetic factors) and a given health outcome in populations which have not been included. This may lead to erroneous associations being detected (and others to be overlooked) and is likely to provide insufficient evidence quantifying the impact of risk factors (including genetic variants) on disease in diverse populations (Landry, Ali et al. 2018). In turn, a lack of visibility and representation may mean that services are not linked to the specific needs of a population or group, unless specific mechanisms are put into place. While none of these issues are insurmountable, the collection or processing of usable data needs to be evaluated in terms of the opportunity costs.

Assumption B: New methods of predicting risk will enable prediction of individual outcomes

The PPH approach rests on the idea that new data sources will result in substantially more accurate predictions of risks and health outcomes than existing data (behavioural, clinical, psychological, social and/or demographic). Two types of data have been of particular interest: genomic data and big data. Recent advances in genomic data analysis such as polygenic risk scores can provide estimates of individuals' susceptibility to disease which are more accurate than those using existing data (Torkamani, Wineinger et al. 2018). Work on using genomic data to predict health behaviours is less advanced, although some indicative findings are emerging (Minicã, Mbarek et al. 2016). Thus, the use of genomic data, in conjunction with existing sources of information about individuals, could improve identification of individuals who are at risk. However, genomic datasets may lack sufficient statistical power to detect rare, but important, allelic variants, and in some cases genetic factors only help explain small amounts of disease risk, for example accounting for less than 10% of disease risk in type II diabetes (Arnett and Claas 2016). Hence, estimations of risk from genomic data remain probabilistic, and in many cases only offer incremental improvements as against other methods. Arguably, describing these as 'predictive' is misleading, as it suggests that they enable reliable forecasting of whether a given individual will develop a disease or not, which is not always the case.

Genomic studies have also faced challenges in prioritising candidate genes for exploration from the millions of potential variants in a patient genome to confirm a diagnosis (Baynam, Bowman et al. 2017). In some cases, studies have been conducted exploring outcomes in isolation that may not provide sufficient rationale for action in isolation (Meagher, McGowan et al. 2017). For example, researchers exploring risk stratification among bladder cancer patients found that

targeting smoking cessation programmes based on genetic risk would prevent cases of bladder cancer, although additional data on gene/smoking interactions for other smoking-related diseases, including lung cancer, would be needed before actions could be taken as the basis for exploring bladder cancer in isolation is unclear (Garcia-Closas, Rothman et al. 2013; Meagher, McGowan et al. 2017). Defining and selecting appropriate antecedents and outcomes in big datasets may be challenging for many areas of public health, where the nature of conditions are often hard to define (for example loneliness). Greater involvement of different stakeholders in the design of PPH and big data projects may be one way of ensuring the salience of the outcomes selected (Ioannidis and Khoury 2018), although there are few examples documenting this.

There has also been interest in using Big Data to improve the estimation of health risks, but this remains largely speculative; automated analysis of electronic health records has shown promise in identifying high-risk patients in healthcare contexts (Goldstein, Navar et al. 2016), but it is as yet unclear how such methods might transfer to non-traditional data sources (and they face considerable challenges even with standardised healthcare data). There may be an assumption that the datasets that can be gathered and linked contain all the correct variables necessary to account for the considerable variance in effects that is usually observed; this is likely to be problematic in many cases, although this is not to rule out the possibility for specific data sets to enable improvements in certain contexts.

Finally, new AI-driven technologies, including wearables, may be useful adjunct strategies in the detection of disease and understanding the optimal treatment. For example, AI is being incorporated in screening programmes for the detection of diabetic retinopathy, including through images taken with smartphones, with human and AI detection rates found to be similar (Bellemo, Lim et al. 2019) and AI analysis of smartphone imaging deemed to be a sensitive initial tool for mass retinal screening in people with diabetes (Rajalakshmi, Subashini et al. 2018). The potential for smart-watches, supported by machine learning, to detect irregular heart rhythms symptomatic of atrial fibrillation (AF), a condition which can lead to complications including cardiac arrest and stroke, has also been of interest, prompting the enrolment of over 400,000 participants in an observational study (Yan, Zhang et al. 2019). The results were suggestive of the ability of the technology to correctly identify atrial fibrillation in users whom it notified of irregular heart rhythms (Campion and Jarcho 2019; Perez, Mahaffey et al. 2019), although the evidence does remain suggestive given the young age profile of the majority of participants (unlike the majority of the population at risk of AF), the high levels of attrition, and potential issues over full compliance with the study protocol (Campion and Jarcho 2019). Nevertheless, both examples suggest that data collection via new technologies for a defined purpose, combined with AI-supported analyses of these data, may be promising avenues for PPH exploration.

Assumption C: New computer science-driven methods of analysis will enable substantially more accurate prediction of individual risk

The use of artificial intelligence and machine learning is seen as a promising strategy to make use of large data sets and enable better prediction. In the context of clinical healthcare, these methods have been used to improve diagnosis and early detection of disease, estimation of

risks and prognosis, and to inform treatment decisions; while results have been mixed (Christodoulou, Ma et al. 2019), some applications show considerable promise (Jiang, Jiang et al. 2017). The argument is, then, that such methods could be adapted for non-health data sets and non-clinical populations to enable better identification of those at risk. As yet, this idea remains largely conjecture, and so the identification of assumptions around increased accuracy of risk prediction using these data is speculative.

One potential issue is that the data sources used to populate machine learning models may be subject to sampling bias and other limitations. For example, data from social media sites or 'wearables' will be skewed towards heavy users, who are not necessarily representative of the general population (Hargittai 2015). Data from convenience samples alone can lead to selection bias and unreliable prediction models (Khoury, Iademarco et al. 2016), and several sources of big data could be considered convenience data based on self-selecting samples. In contrast, data used for conventional epidemiological studies have usually been purposefully collected to ensure a diversity of population characteristics and can allow for unbiased assessment of genetic and environmental factors (Khoury, Iademarco et al. 2016).

There also exists potential for different forms of bias to be carried through into machine learning/artificial intelligence algorithms that have been trained on inaccurate or selective data (Yu and Kohane 2019). Amassing more data and using more automation may even work to solidify different forms of automation bias (Ioannidis and Khoury 2018) and therefore it is unclear whether the results derived from these methods are in fact more accurate and cost-effective than established forms of data analysis (Ramaswami, Bayer et al. 2018). Methods which involve greater segmentation and repeated testing may lead to an increase in instances of erroneous inferences and associations being detected by chance (Manrai, Patel et al. 2018) and may lack adequate control of potential confounding variables (e.g. genetic associations based on race being confounded by poverty and inequalities in access) (Meagher, McGowan et al. 2017, Manrai, Patel et al. 2018).

On a more general level, such novel methods of analysis can be characterised as exploratory or data-driven as opposed to the inferential and hypothesis-driven nature of conventional statistical methods, and advocates argue that they do not require *a priori* theories, or the positing of mechanisms of causal effect, in order to generate meaningful results (Anderson 2008). In the context of public health interventions, the implication is that we do not need to know *how* an intervention works in order to predict who will benefit. Such an approach is contrary to most current thinking in public health, which emphasises the importance of theory and understanding of mechanisms of impact in evaluating interventions, particularly complex interventions, and in using the findings to inform policy (Bonell, Jamal et al. 2015; Moore, Audrey et al. 2015).

Many of the critiques around the role of theory are not unique to PPH. However, the hype surrounding PPH approaches raise the risk that big or novel data sources are viewed as having the capacity to produce insights, regardless of the complexity of the phenomenon, without even asking defined research questions of the data (Kitchin 2013). This raises the risk that spurious findings with low transferability starts to inform public health practice. Several of the limitations noted above could potentially be overcome through carefully designed epidemiological studies that draw on theory, but this represents a different methodological paradigm to much of that

found in the Big Data literature. In particular, it is unclear whether a PPH approach is compatible with a system-based approach that recognises interconnections between different layers, and is gaining prominence in public health (Rutter, Savona et al. 2017).

Assumption D: Tailored interventions are more likely to change behaviour

A key assumption in PPH is that behaviour change interventions which use information about individuals to tailor intervention content are more likely to be effective in changing behaviour.[4] There is some evidence for this assumption, in that the provision of tailored health information appears to be more effective than the provision of generic health information in bringing about behaviour change, although effect sizes are not large and we know little about medium- to long-term impacts (Krebs, Prochaska et al. 2010; Lustria, Noar et al. 2013; Wolfenden, Nathan et al. 2015). It is less clear how tailored health information might compare to other approaches not reliant on tailored information provision to individuals (e.g. food labelling or tobacco plain packaging).

Assumption E: Better targeting of interventions leads to greater effectiveness

PPH is based on the assumption that health interventions will be more effective and/or cost-effective if they are targeted at people who are most likely to benefit from the intervention. The argument is that conventional, population-based public health interventions produce a wide range of responses since many of the people who receive the intervention are unlikely to benefit. Using new data sources to identify those with greater potential to benefit, and preferentially delivering interventions to them while withholding interventions from those unlikely or less likely to benefit, should lead to better outcomes overall as well as more efficient use of resources (Arnett and Claas 2016; Torkamani, Wineinger et al. 2018). However, while there is now a substantial body of proof-of-concept studies on how data could be used to assess risk and prioritise preventive interventions such as cancer screening (Seibert, Fan et al. 2018) or statin therapy (Natarajan, Young et al. 2017), there is little evidence that such targeting leads to better health outcomes in reality.

Moreover, the argument for targeting does not align with two key findings in the public health literature. First, interventions that foster small improvements in health behaviours across the whole population produce greater benefit overall than those targeted towards achieving large improvements in health behaviours among high-risk individuals (Geoffrey Rose's 'paradox of prevention') (Chowkwanyun, Bayer et al. 2018). Second, community-level or environmental-level interventions are not concordant with targeting higher-risk individuals, but a strategy that targets both individual and wider determinants is more likely to be effective than one focused on individual determinants alone (Marteau, McGowan et al. 2018). Such interventions may be targeted at higher-risk communities or areas initially, and strategies for targeting such interventions according to need have been identified as a research priority (Egan, Kearns et al. 2016), and this is a potentially promising application of PPH principles which has not been

[4] In this review we distinguish between using data to *target* interventions (i.e. to provide or not provide interventions to individuals or groups based on the likelihood of the latter benefiting from the intervention) and using data to *tailor* interventions (i.e. to provide different intervention content to different individuals or groups).

widely explored. Our case study around community-based interventions seeks to explore this angle further.

Assumption F: Tailoring and targeting based on new types of data will be more effective than the use of existing data

The tailoring implemented in existing interventions is generally based on a small number of factors relating to individuals, such as demographics or clinical markers, or the results of questionnaires assessing self-efficacy or health behaviours at baseline. Thus, tailoring and targeting interventions, as currently practiced, mostly does not require very extensive or sophisticated data collection, or data from novel sources, but is based on broad-brush estimates of baseline risk. As described above, PPH aims to go beyond this by drawing on a greater range of data, including lifestyle, health and genomic data, in order to predict risk more accurately, and hence to tailor and target interventions more precisely. Thus, the assumption is that the use of more data, and more diverse data – and specifically of new types of data such as genomic or Big Data – will significantly increase the effectiveness of tailoring and/or targeting interventions relative to the use of existing information on individuals.

In relation to targeting interventions, the idea is that new data sources will enable more accurate prediction of who is likely to benefit from an intervention, and this in turn will enable more effective interventions (see previous section on Assumption E). In relation to tailoring, the idea is that taking a greater range of information into account will enable interventions to be better suited to individual needs. However, it is not entirely clear that there is a major need here, i.e. that the relatively crude data currently used to tailor interventions is insufficient. Indeed there is some evidence against the assumption that incorporation of more data is always an improvement; for example, providing information on genetic disease risk has been found to be largely ineffective in changing health behaviours (Hollands, French et al. 2016). This said, there is a need for more evidence about how tailoring works, which strategies are most promising, and which types of data are likely to be useful.

Assumption G: Information-based behaviour change interventions are a promising approach to public health in general

The practical applications of PPH tend to involve interventions focused on individuals. The argument for this flows naturally from those already given: if the key to better interventions is individual targeting and tailoring, then public health practice will need to focus mainly on interventions which can be targeted and tailored. This implies a focus on interventions which seek to provide individuals with information and/or to change the factors influencing decisions at an individual level. However, the evidence for many of these intervention types, such as ‘health check’ programmes (Si, Moss et al. 2014; Krogsbøll, Jørgensen et al. 2019) or incentives for health behaviour change (Giles, Robalino et al. 2014; Mantzari, Vogt et al. 2015) does not indicate that they have substantial lasting impacts, although there may be positive short-term impacts in some cases. There are also concerns about potential adverse effects (Jochelson 2007; Capewell, McCartney et al. 2015).

More broadly, some researchers have raised concerns about stigmatising at-risk groups (and thereby exacerbating inequalities), and conferring responsibilities to individuals without evidence for positive behaviour change (Meagher, McGowan et al. 2017). In the case of genomic data, there is generally low awareness among the public about diagnostic processes and support available after conferring information based on PPH/Precision Medicine (e.g. genetic counselling) (Ioannidis and Khoury 2018); similar critiques could be levelled at other forms of PPH if increasing precision in terms of risk is not linked to changes in care practices (Neff 2013).

Assumption H: PPH approaches and methods are ethically sound

As outlined above, there are some concerns that a more individualised approach to public health is out of step with established public health practice and overlooks the structural factors that shape health outcomes and health inequalities. Such omission is ethically questionable. Proponents of PPH have argued that, far from overlooking these factors, PPH emphasises the importance of social and structural determinants of health for communities that have been marginalised (Horton 2018). However many applications of PPH appear to take into account only individual level factors, where for example the utility of 'knowing the speed with which people metabolize nicotine, based on genetic and other factors, could lead to personalized smoking-cessation interventions' (Khoury, Iademarco et al. 2016).

The greater incorporation of data into decision-making raises concerns around the welfare of individuals (and communities) in terms of the potential for stigmatisation and exclusion, consent, privacy, anonymity and confidentiality. While stringent procedures have historically operated around the collection of data used for social and biomedical academic research, these have been more relaxed in terms of other forms of data and recent concerns indicate that individuals may not have given informed consent (Ioannidis 2013, Metcalf and Crawford 2016). Although the General Data Protection Regulations (GDPR) have greatly improved procedures around the collection and processing of individual data, the designation of data on some groups and categories as sensitive, and subject to additional conditions of use, could lead to further issues in the representation of minority groups (for example Kneale, French et al. 2019) in applications of PPH. Moreover, implications in terms of the validity of data, such as gaps in representation, also have important ethical dimensions.

In the next section of this report, we examine the utility of a PPH approach through an assessment of the claims made in commentaries on PPH.

5. Assessing the line of argument used in PPH commentaries

In this section we focus on examining how PPH commentaries and think-pieces draw on evidence, and consider the extent to which evidence directly supports the arguments made.

The influence of non-empirical studies in shaping debate

Commentaries, reviews, letters and editorials on PPH have the capacity to describe complex ideas about PPH in a format that can be understood across a range of different readers and

audiences. These outputs may be influential because they outline and explain new ideas and trends emerging within a discipline and expound on the potential benefits; some may adopt a different perspective and may seek to encourage scepticism about a particular issue among readers.

Measuring the impact of these studies is challenging. Academic measures of impact, such as citation counts, are unlikely to be suitable as the studies themselves tend to have only been published within the past two years. However, using alternative measures of impact (Altmetrics (Adie and Roe 2013)) we find that commentary studies have the capacity to be of substantial interest across a wide range of audiences (see table 2), attracting a good deal of attention on social media and the mainstream media, as well as a number of readers on referencing software Mendeley. A piece by Dowell, Blazes et al. (2016) was particularly influential, being in the top 5% of all research outputs scored by Altmetric, and being the only example to have been cited within policy (Centers for Disease Control policy outputs on emerging infectious diseases).

Table 1: Altmetric scores and breakdown of score among commentary, editorial and non-systematic review studies (measured August 2019)

	Altmetric score	Cited by policies	Cited by media	Tweeted	Readers on Mendeley	Included on Facebook walls
(Chowkwanyun, Bayer et al. 2018)	326	0	6	443	78	5
(Davey and Deribe 2017)	33	0	0	51	29	4
(Dolley 2018)	413	0	45	67	91	0
(Dowell, Blazes et al. 2016)	524	1	3	687	99	21
(Dunn, Mandl et al. 2018)	59	0	0	84	37	0
(Horton 2018)	144	0	0	242	24	1
(Khoury, Engelgau et al. 2018)	27	0	0	36	8	0
(Kuo, Summers et al. 2019)	5	0	0	10	8	0
(Lyles, Lunn et al. 2018)	36	0	0	43	19	0
(Newnham, Kemp et al. 2017)	11	0	0	18	56	1
(Prosperi, Min et al. 2018)	28	0	0	48	83	0
(Riddle 2017)	6	0	0	9	8	0
(Taylor-Robinson and Kee 2018)	54	0	0	98	40	0

The inclusion of evidence within commentary pieces is intended to justify the argument. Evidence is offered as a means of supporting conclusions or recommendations to act (Upshur and Colak 2003). However, as described above, in making their arguments, it is possible that some studies may overstate the claims being made because of the paucity of evidence (in either direction). Commentary studies draw on a number of different types of evidence to support the argument that PPH is a worthwhile endeavour including empirical research studies. However, not all studies may fit within a definition of utilising new sources of data and applying new methods.

Furthermore, evidence may be used in a persuasive rather than factual way. For example, a study by Holmberg, Chaplin John et al. (2016) drew on information from Instagram, a photo sharing social network, to understand which foods were presented and how by adolescents in Sweden. The study concluded that adolescents may focus on the aesthetic presentation of food, mirroring advertising practices, but that few of the images actually showed adolescents directly consuming food and were unlikely to mirror eating habits (Holmberg, Chaplin John et al. 2016). The conclusions appear at odds from the reporting of the findings elsewhere, which suggested that the study was an example of where Instagram could be used in lieu of a food diary or

dietary intake questionnaire (Prosperi, Min et al. 2018)[5]. In commentary studies there is a strong element of persuasion in the use of evidence, necessitating a need to interrogate an argument and the way in which evidence is used to support that argument, to discover its weakness or the basis of its validity. To aid us to understand how evidence is used within the commentary studies listed in Table 2, we adapted Toulmin's Model of argumentation approach for breaking the argument down into its constituent parts (Toulmin 2003).

We explore the extent to which a claim is supported by robust grounds. A claim in this case is a statement either in support or in opposition of PPH while the grounds are the basis for making that claim. While different grounds for an argument can be used, for example logic or emotion, our focus here is on the appropriate use of evidence, therefore we confine our interest to claims that use cited studies as grounds. We then need to make a link between the grounds and a claim in the form of a warrant. The warrant can be regarded as the bridge between the grounds and the claim, and involves considering how and if the grounds can support the claim. Backing to the warrant is, in turn, examined to understand how the warrant can (or cannot) support the leap between the grounds and the claim. We use this model to disentangle claims about PPH to understand which are supported adequately by evidence and where further research is needed to establish these claims (see further details in Appendix 2). We restricted our focus to claims that could be used to support or undermine the assumptions underlying a PPH approach (see section 5).

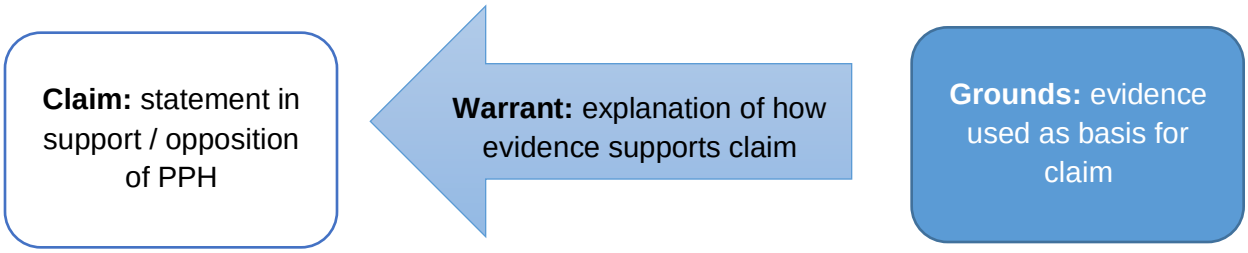


Figure 2: Schematic of line of argument.

Initially we started with a pool of 15 commentaries, two were excluded because they made no clear claims related to the underpinning assumptions of interest (this was related to the type of output, an editorial and a response to commentary respectively (Weeramanthri, Dawkins et al. 2018; Chowkwanyun, Bayer et al. 2019)). The claims are disaggregated, with further evidence found in Appendix 3, and summarised below. A further study does not feature in this table, as the grounds for each of the claims of interest were not supported by cited evidence, but represented other types including reasoning or logic (Chowkwanyun, Bayer et al. 2018)

[5] However, the same commentary also alerted us to the use of Instagram and machine learning methods to correctly identify predictive markers of depression, highlighting the potential of the approach in future as a mental health screening tool Reece, A., G. and C. Danforth, M. (2017). "Instagram photos reveal predictive markers of depression." EPJ Data Science **6**(1): 15.

Which claims about Precision Public Health are supported by evidence and where is further research needed?

A number of claims about PPH are not adequately supported by the evidence sources provided within commentary studies to sufficiently justify the claim. We observed that the interpretation of a number of evidence sources used by commentary study authors to justify a claim did not provide sufficient backing to support the nature of the claim itself. In these instances, evidence was used in a way to persuade the reader that the claim had been validated by others in the field, although the basis for using the evidence in this way was weak or ostensibly absent. Almost half of the claims (15/28 claims) assessed were deemed to fall within this category in that there was not a sufficient link made between the cited study and the claim made by the commentary authors, or that the claim was only partially supported. This is not tantamount to concluding that the claim itself was false or illogical, and in many cases, the claim itself may ostensibly be logically sound. Nevertheless, claims that are not supported by evidence are reliant on (at best) plausibility rather than certainty.

As shown in table 2, a quarter of claims made about PPH were made about the availability of new data sources and their use for targeting or tailoring public health (7/28), with a similar number made around the capacity of new methods to enable more accurate prediction of risk (6/28). Far fewer claims, supported by cited evidence, were explicitly made around the capacity of PPH to more effectively change behaviours, or that new sources of data would outperform existing sources of data. This is not to say that the commentary studies did not make these claims, but they did not draw on cited evidence to provide any grounds for making these claims. The claims (table 2) are mapped onto the assumptions underpinning PPH; further information can be found in Appendix 3).

Table 2: Summary of results from using Toulmin's Line of Argument approach for claims made in studies (claims are grouped according to assumptions underpinning PPH)

	Is the claim substantiated by the evidence presented? *Strength of evidence* 			
	No	In part	Yes, support provided	Total claims
Assumption A: New data sets can provide usable data for public health interventions	3	2	2	7
Assumption B: New methods of predicting risk will enable prediction of individual outcomes	2	1	1	4
Assumption C: New analysis methods will enable substantially more accurate prediction of individual risk	3	1	2	6
Assumption D: Tailored interventions are more likely to change behaviour	1		3*	4
Assumption E: Better targeting of interventions leads to greater effectiveness	1		2	3

Assumption F: Tailoring and targeting based on new types of data will be more effective than the use of existing data				
Assumption G: Information-based behaviour change interventions are a promising approach to public health in general				
Assumption H: PPH approaches and methods are ethically sound	1		3	4
Total claims	11	4	13	28

*One claim deemed to be weakly supported (unlike claims supported in part, the whole claim is substantiated albeit with weak evidence)

Of the seven claims around the assumption that new data can provide usable data for public health interventions (Assumption A), three were not substantiated by the evidence cited. In most of these cases, the claims were not substantiated because the cited evidence did not draw on data considered to be newly available. This reflected ambiguities around what could be considered newly available and hitherto unused sources of data or sources of big data. For example, a claim that big data can be used to predict risk in order implement preventative interventions was supported by a study that included 1,300 participants of a randomised controlled study, and not what would be considered a large volume of data requiring specialist software or skill for the analysis (big data) (Dolley 2018). Some claims in opposition of PPH related to this assumption were also deemed to be unsupported or only partially supported by the cited evidence. For example, a claim that we know little about the potential unintended consequences of PPH approaches was supported by a warrant; the warrant suggested that bias and lack of representativeness were inherent properties of Electronic Health Record data. Although such a warrant may have some grounds in the broader literature, this was deemed to be unsupported by the evidence provided in the commentary study.

Four claims related to the assumption that PPH approaches can improve the prediction of risk (Assumption B). Only one such claim was fully substantiated where the authors provided a warrant focussed on genomic studies, with the cited evidence showing that more precise estimation of risk and response to treatment can be calculated through incorporating genomic data. In contrast, a claim that adopting a PPH approach can effectively lower risk of harmful outcomes (in this case pre-term birth) was not directly supported by the cited evidence, as the cited studies were not applications of PPH approaches (Newnham, Kemp et al. 2017). Claims related to the assumption that PPH approaches can improve the prediction of risk, but which highlighted the potential drawbacks of PPH approach (negative claims), were also not deemed to be fully supported by evidence. This highlights that confirming or refuting the promise of PPH approaches in improving the prediction of risk is challenging. In general, few commentary studies drew on appropriate evidence, a likely reflection of the breadth of the evidence base, and the broad assumption that PPH approaches can improve the prediction of risk appears to be weakly supported in the literature.

We found that although fewer authors focussed on claims related to the assumption that better targeting (Assumption E) and tailoring (Assumption D) of interventions is a more effective approach, this smaller pool of claims were more likely to be substantiated (five out of seven claims across both assumptions). For example, there was deemed to be sufficient backing for warrants that machine learning applied to big data can help to identify subpopulations with

unique health needs, and that this can be integrated into decision support tools (the backing in this case involving a study on Familial Hypercholesterolemia (Khoury, Engelgau et al. 2018)). However, there were very few claims supported by evidence that made the explicit link between PPH approaches to targeting and tailoring of interventions being more effective than conventional practice and could provide direct evidence as grounds to this assumption. Similarly, there was little empirical evidence used to support claims that PPH approaches can improve public health outcomes through effectively changing behaviours. In fact, as highlighted in one commentary study (Taylor-Robinson and Kee 2018), there is little reason for optimism in thinking that providing tailored information, in this case based on genomic profiles, can change individuals' behaviours; this was a claim that was found to be substantiated by the evidence.

Finally, not only were claims around the components and effectiveness of PPH approaches found to be unsubstantiated in several cases, so was a claim around PPH being ethically sound (Assumption H). A broad claim about the harmonisation and linking of data being a means of maximising social justice were not directly substantiated (Lyles, Lunn et al. 2018). Claims made in another commentary that PPH approaches could be used to reduce health disparities were supported, through offering evidence where PPH approaches had enabled greater understanding of geographic health inequalities, although the evidence that this information was actioned was not directly presented (Horton 2018). In addition, ethical issues were discussed in several commentaries, but often not evidenced, suggesting that there was little systematic investigation into this dimension of PPH.

Summary of line of argument analysis

Commentary and review studies on PPH have the capacity to be highly influential, and several are among the most influential recorded using alternative metrics of impact. They also provide useful shorthand summaries of the main arguments surrounding PPH and outline opportunities and critiques of the ideas. They offer a glimpse of which of these arguments are evidenced with empirical studies and where the arguments of PPH are based on logical extensions of the evidence and emotive arguments. We mapped these arguments according to the assumptions that inherently underlie the premise of PPH. Here, our focus was on empirical evidence, and the absence of commentaries offering such evidence on the assumptions surrounding the differential effectiveness of PPH approaches compared to established public health approaches is illuminating. We also observed that where evidence is used, in many cases the inferences made go beyond the claims made in the source data. In this case, evidence is not being used to substantiate facts, but as part of persuasion dialogue (as opposed to precision dialogue). We do not directly assess whether the claims made, and the inferences made based on the cited evidence, are plausible. However, the analyses here emphasises a number of points. Firstly, that the PPH field may be highly influenced by commentary and non-systematic review pieces. Secondly, that commentators on PPH often attempt to provide evidence for claims, but the link between the evidence and the claim is often unsubstantiated when critically examined. Thirdly, that many of the assumptions underlying PPH have not been evidenced, suggesting that there needs to be a measured approach to adopting PPH approaches and greater investment in understanding and evaluating the added value of the approach. PPH represents an emerging concept, making it challenging to evidence different facets; however the analyses here suggests that there needs to be a much clearer communication of where direct evidence exists and where

the parameters of these studies lie. If PPH is to become a reality in public health decision-making, the type and use of supporting evidence needs also to become much more precise both by advocates and detractors.

6. Assessing the characteristics of empirical studies on PPH

We set out to identify and further understand the features of PPH studies through examining two sets of empirical studies, one set that could be regarded as self-described PPH studies (n=14) and a smaller set of Ecological Momentary Intervention (EMI) studies or Ecological Momentary Assessment (EMA) studies (n=3). EMA and EMI studies use new wireless enabled technology and/or social media to collect real-time data and use this to target interventions or tailor intervention content. EMI and EMA studies were of interest as they may share some of the same goals as PPH and this is explored further here.

This purposive approach to identifying studies was used because: (i) the interest in examining how the language of PPH is being used and interpreted in empirical literature; and (ii) a much more extensive searching and screening approach would have been needed to identify studies which examine PPH but don't use the term, which was not commensurate with the aims and timescale of the review. We used a structured tool to help understand the characteristics of these studies (see Appendix 4).

Empirical studies using the term PPH

Characteristics

Starting with a simple search on PubMed for 'Precision Public Health', we identified 14 empirical studies that use this term and that undertook empirical research described by the authors as either explicitly involving or reflecting the principles of PPH. Here we provide a narrative synthesis of the features of these studies to help to further understand the features of PPH, as well as how they meet our underlying assumptions. These were empirical studies that contained novel analyses drawing upon different methods, including case studies based on the direct experiences of the study authors, which explicitly or implicitly identified as contributions to the PPH literature.

Studies which explicitly self-identified as PPH studies were those that reported on applying a PPH approach to the analysis (Cummings, Tokarz et al. 2019) or reported on implementing a PPH strategy in public health practice (Baynam, Bowman et al. 2017). Studies which implicitly identified as PPH studies tended to first provide an outline of the principles of PPH, and then aligned the purposes of the study with these goals.

The majority of the studies were based in low and middle income countries (LMIC; n=10), including a well-known example focussed on child mortality across Africa (Golding, Burstein et al. 2017). There was a particular focus on infectious disease (n=5), including a cluster of four studies authored by overlapping teams exploring PPH approaches to understanding malaria transmission patterns in a specific district of Mozambique, Chimoio (Ferrao, Mendes et al. 2016; Ferrao, Mendes et al. 2017a; Ferrao, Mendes et al. 2017b; Ferrao, Niquisse et al. 2018).

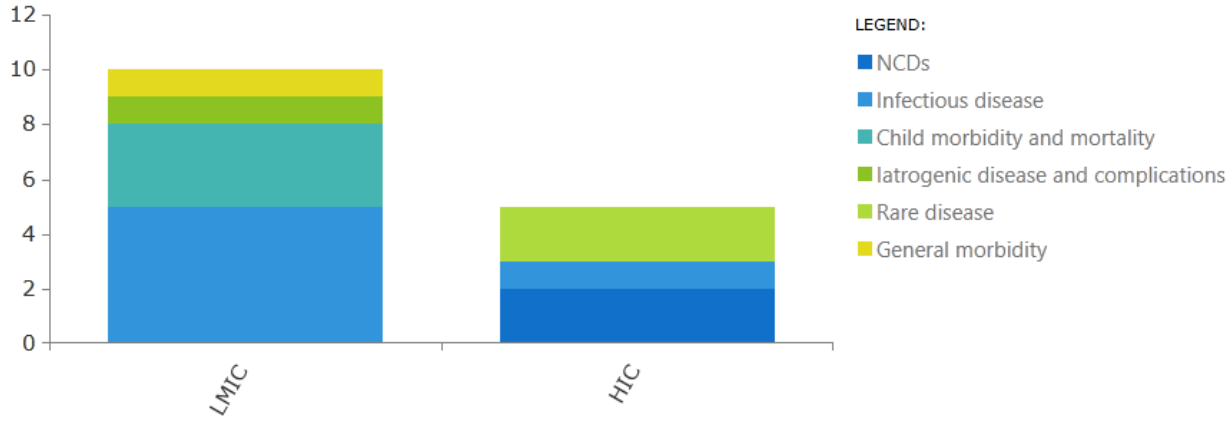


Figure 3: Empirical PPH studies by area of focus

Only one study focussed solely on utilising genomic data through examining genetic susceptibility, based on ethnic origin, to iatrogenic disease and complications across a number of treatments within two areas of Colombia (Nagar, Moreno et al. 2019). A further study combined genomic data with administrative data to understand patterns of severe acute respiratory illness (Cummings, Tokarz et al. 2019).

Relationship with definitions of PPH

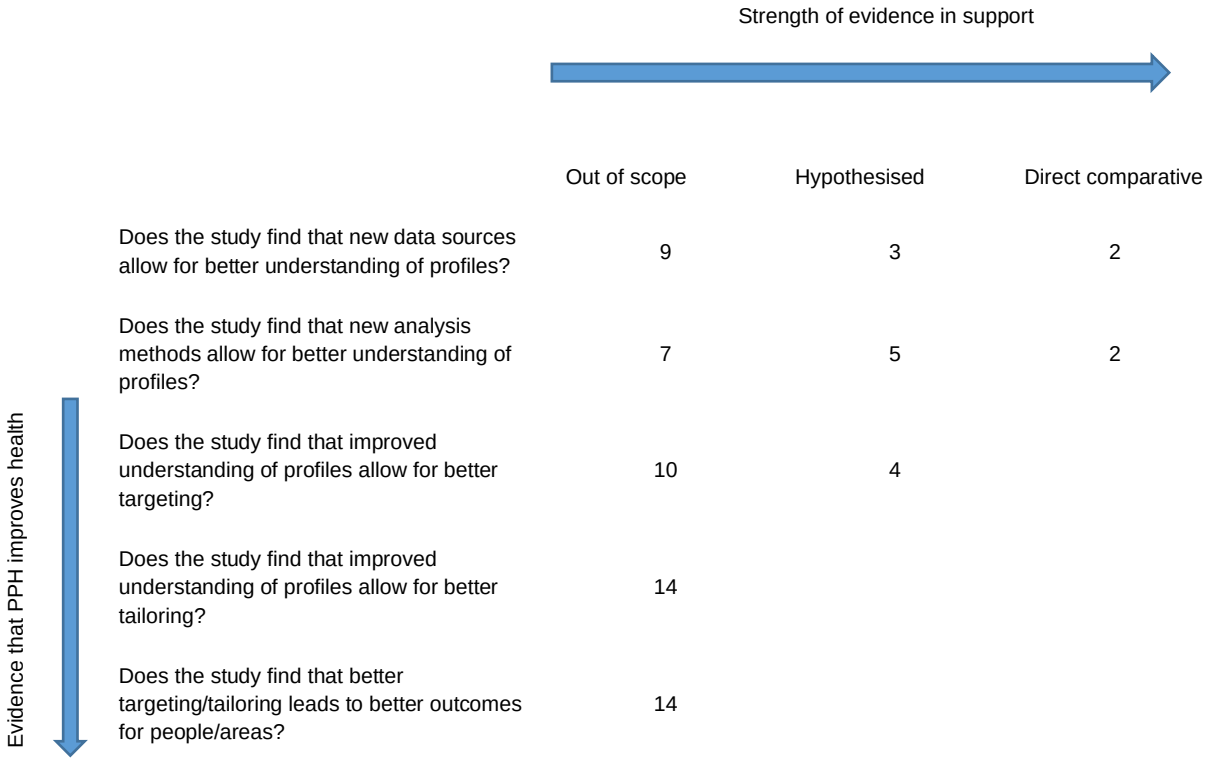
Examination of the 14 included studies highlighted that several of these appeared to challenge two principles which underpinned our understanding of PPH; namely that PPH should involve *new sources of data* and/or that PPH should involve *emerging analytical techniques* and specifically artificial intelligence and machine learning techniques. Four studies did not appear to meet these criteria; these were studies that either drew upon electronic health records to develop greater precision in terms of risk profiles of individuals (n=1 (Franchini, Pieroni et al. 2018)) and/or exploring geographic or temporal precision including incorporating Geographic Information System (GIS) approaches to better estimate the burden/risk of disease (n=3) (Ferrao, Mendes et al. 2016; Freeman, Boylan et al. 2017; Rieger, Trommlerová et al. 2019). These studies in particular highlight the problematic nature of treating PPH as a concept with crisp boundaries. For example, although electronic health records do not constitute a new source of data, across several settings, including high income settings, they are underutilised as a source for decision-making (Kneale, Khatwa et al. 2016). For those studies that took place within low and middle income countries, the strengthening of some forms of administrative data to provide sufficient quality for analysis does indeed constitute new and hitherto unused sources of data. As such, this small group of studies helps to illuminate the risk of focussing on a definition of PPH, as opposed to honing in on steps that can be taken to increase precision.

Relationship with underpinning assumptions of PPH

Among the select group of empirical PPH studies (n=14) we found little direct evidence that adopting PPH methods to refine targeting and tailoring of interventions led to an improvement in

health outcomes for people or areas. Examining health outcomes was out of the scope of the included studies; most reported on intermediate outcomes such as the impact of using new data or analysis methods to develop a better risk profiling. There was little direct evidence on how targeting had been conducted and none of the included studies reported on how interventions had been tailored using PPH methods. As reported above, four of the studies did not appear to draw on new analysis methods or sources of data.

Table 3: Strength of evidence for the impact of Precision Public Health



With reference to using new sources of data to better understand the profile of people/areas, three studies hypothesised added improvement – where they described the existing literature or status quo and described the added improvement of utilising new sources of data. These included reports by Baynam, Bowman et al. (2017) on how the implementation of new screening programmes based on genomic testing were improving the detection of rare and undiagnosed diseases; how the incorporation of newly available data from satellite imaging into geostatistical models allowed for greater precision in understanding the geographic profile of malaria cases (Ferrao, Niquisse et al. 2018); and how genomic data allowed for better profiling of areas and groups in terms of their susceptibility of adverse impacts of medicines (Nagar, Moreno et al. 2019). An example of a study providing direct evidence of the added value of a PPH approach comes from the work of Cummings, Tokarz et al. (2019). They report on the use of genomic and administrative data to understand the epidemiology of unexplained severe acute respiratory infections. Use of genomic information allowed for the characterisation of previously unidentified

illness, with a cluster of measles-related infections identified, and a number of vaccine preventable illnesses identified. One study drew on big data sources harvested through online interactions; this study also directly compared the results with previous approaches within the study to provide direct evidence on the added value of PPH approaches (Lu, Hattab et al. 2019). In examining precision forecasting of influenza cases at a US-state level, Lu, Hattab et al. (2019) drew on data collected from influenza-related Google search frequencies (Google Flu Trends), electronic health records, and historical flu trends within each state, and analysed their data with new methods. Combining data from different sources helped to address some of the deficiencies of relying on Google Flu Trends data alone for prediction, and represented a substantial increase in the variance explained in models. This same study also provided direct evidence that improved analytical techniques and ensemble methods also more accurately tracked reports of influence-like illness produced by the Centers for Disease Control and Prevention. A further study also provided direct evidence that PPH analytical approaches were more effective, with Ferrao, Mendes et al. (2017) finding that their machine learning derived models provided a better fit than standard modelling approaches.

Few studies provided evidence that greater understanding of profiles of people or areas through PPH approaches had been, or could be, implemented into targeting strategies. Many of the studies represented retrospective precision surveillance studies. Baynam, Bowman et al. (2017) present a case study of their approaches in implementing a programme for detecting rare and undiagnosed diseases, and while they hypothesise that the benefits to individuals of such targeted approaches are clear, supporting data are not presented. Another three studies hypothesised that predictive risk models developed were an advance on current benchmarks and targeting strategies through comparing their findings with previous studies (Ferrao, Mendes et al. 2017; Ferrao, Niquisse et al. 2018; Lu, Hattab et al. 2019).

Theoretical underpinnings and ethics

Data used in studies were collected through active interaction with researchers (i.e. completing a survey) and through regular passive interaction (i.e. ecological level counts of diseases and electronic health records). Only one study used data through continuous passive interactions (i.e. based on public internet use and search trends) although these data were not linked to individual profiles within the study. Few of the studies reported that stakeholders had been involved in the design of the study. In one study clinicians had been involved, although this was to verify decisions around the coding of data and not to ensure that patient and practitioner views were represented (Franchini, Pieroni et al. 2018).

In terms of theoretical underpinnings, none of the studies were identified as drawing on established public health theory (e.g. social ecological models of health) although this is perhaps reflective of the studies predominantly being focussed on surveillance and predicting risk. There may be scope among this pool of studies to become more aligned with socioecological models of health (Dahlgren and Whitehead 1991) through incorporating data from different levels of influence, although no examples were found.

Summary of findings from empirical studies using the term PPH

All fourteen studies, regardless of whether they utilised new sources of data or employed new methods of analysis, reported that the findings represented an advance on existing public health knowledge through gaining more precision about ill-health, disease and mortality in terms of genomic, geographic or temporal precision. For example, Franchini, Pieroni et al. (2018) developed a rules-based algorithm for detecting the risk of heart failure, and found that the model was more efficient the early detection of people either at risk or in early stages of heart failure and enable their referral to specialists who can further optimize their cardiovascular care. Their approach may not draw on machine learning methods or unused data, although does represent an advance in terms of the concept of precision from the perspective of decision-makers.

The included studies also highlighted a discrepancy between some of the content from commentary studies (see previous section) and the focus of empirical PPH studies (those meeting the definition of PPH outlined in the glossary). The former studies emphasise that precision can be achieved to target interventions towards narrow social profiles through the incorporation of data reflecting micro-level of individuals' day-to-day lives. The latter empirical examples, with the exception of genomic studies, tend to offer evidence of greater precision predominantly using ecological level data, allowing for areas and groups to be targeted more efficiently. Overall, based on these empirical studies, we know little about how PPH approaches can be implemented to (i) tailor interventions; and (ii) whether this has an impact on outcomes.

Ecological Momentary Intervention and Ecological Momentary Assessment Studies

In addition to examining self-defined PPH studies, we also examined a small group of interventions that appear to share some of the goals of PPH. The studies appeared to use newly available data sources to either develop targeting strategies and/or tailor interventions. Ecological Momentary Interventions (EMI) are interventions provided to people during their everyday lives (real time) and outside clinical settings (real world) and draw on new technologies in order to tailor interventions. Ecological Momentary Assessments aim to capture data in real time through repeated sampling of individual behaviours and experiences in real time and in the real world. Although the concept of capturing data from individuals through repeated measures and outside the laboratory is not new (Stone and Shiffman 1994), EMI and EMA studies that draw on new technologies could meet some of the underlying principles that are associated with PPH.

The origin of EMA and EMI studies appears to lie in the psychological literature, although recent studies have explored their application within public health. Within the literature we encountered little cross-over between EMA/EMI studies and those considered as PPH studies, including among commentary studies. We examine three applications below of EMA studies with a focus on two aspects: (i) how closely do the studies align with the principles of PPH; and (iii) can the studies reveal additional conceptual and ethical considerations of undertaking PPH studies? The studies were purposefully selected as examples of EMI/EMA studies that focussed on a public health issue. One study focussed on reducing alcohol consumption among young people (Wright, Dietze et al. 2018); one study examined the use of new data and technology to help support weight loss (Martin, Miller et al. 2015, Martin, Gilmore et al. 2016), while a further

EMI/EMA study was conducted to examine reductions in the risk of smoking cessation (Businelle, Ma et al. 2016, Hébert, Stevens et al. 2018).

All of the studies claimed to undertake EMA activities as a preliminary to delivering a personalised intervention. In the case of Wright, Dietze et al. (2018) and (Businelle, Ma et al. 2016; Hébert, Stevens et al. 2018), participants completed short surveys through mobile phones before receiving text messages (the intervention) that were tailored to the survey responses submitted, intended to reduce the number of drinks consumed during drinking events and reduce the risk of smoking relapse respectively. In both interventions, while the data were submitted in real time presenting an advantage to the timing and tailoring of the intervention content, there was some evidence that real-time data could share the same methodological challenges as conventional survey approaches, with a 58% response rate recorded in one study (Wright, Dietze et al. 2018), although a higher level was obtained in a second (87% (Businelle, Ma et al. 2016)). A third study drew on internet-enabled weighing scales among participants of a weight loss intervention. Participants weighed themselves every day and the weight was automatically and wirelessly transmitted to a website that was accessible by a weight loss counsellor. The recorded weight was used to determine progress and tailor advice or positive reinforcement messages alongside other examples of new technology including activity monitors (Martin, Miller et al. 2015; Martin, Gilmore et al. 2016). While all the studies used new *technologies* to collect data, particularly in the case of data collected through smartphone surveys it may be more questionable as to whether survey-based data collected through new technology do constitute 'new' sources of *data*. However, they do offer additional precision in collecting data in real-time from natural settings. None of the studies employed new methods to analyse the data, and all the studies included relatively modest numbers of participants (the largest included 90 young people randomised to receive an intervention).

The two EMA studies that collected data through smartphone surveys and delivered text messages for the intervention showed mixed evidence of effectiveness. Young people receiving a tailored alcohol reduction intervention showed no difference in the number of drinks consumed during drinking episodes compared to a control group (Wright, Dietze et al. 2018). However, in a study of a smoking cessation intervention, receipt of messages that were tailored to address urges to smoke, stress or access to cigarettes respectively were more effective in reducing these smoking triggers than non-tailored messages (Hébert, Stevens et al. 2018); 20% of the participants were abstinent at twelve weeks (Businelle, Ma et al. 2016). A final study that used internet enabled weighing scales as the basis for tailoring weight loss information provided by counsellors found that the intervention was effective in supporting weight loss compared to a health education control group, albeit based on a small sample of participants (Martin, Miller et al. 2015).

All three intervention studies align with the aims of PPH of utilising hitherto unused data, although the focus here is on data actively collected and submitted through new technology, as opposed to drawing on (big) data sources where data is collected incidentally or passively (Prosperi, Min et al. 2018). However, the goals of EMI/EMA studies may align much more closely with the concept of Predictive Prevention, which as outlined by Public Health England, involves 'combining person-generated data with existing health data can help us predict poor health in the future and create an opportunity to prevent it with more personalised advice and

services' (Newton 2019). Conceptually, these types of studies appear to be compatible with transtheoretical (stages of change) models of health, where interventions are tailored towards an individual's readiness to change and adopt healthier behaviours (DiClemente and Prochaska 1998). In this case, an individual's readiness to change, and potential for relapse, may be detected through real-time assessment. However, the model and the approach overall do have some recognised limitations, including limited to account for social context, and there remain questions around the effectiveness of the approach across different health behaviours. Finally, all three studies described that issues of consent had been accounted for. However, the experience from Wright, Dietze et al. (2018) also suggest that obtaining informed consent can be challenging for this type of study given that there are more complex flows of information. Wright, Dietze et al. (2018) explained that they 'found it difficult to clearly and concisely present all information relating to the study [to participants] in a format that also fulfilled the requirements of the ethics committee. Despite our use of diagrams, the description of procedures seemed overly complex, and we found that when we had the chance to explain the study verbally (in telephone calls to remind/follow-up participants), participants were more inclined to consent'.

7. The application of a Precision Public Health approach to Health Checks

This section reports on the first of two case studies in which we sought to examine the potential of PPH when applied to specific use scenarios. This example focuses on an individual level intervention, Health Checks.

Intervention theory

Health check programmes involve inviting people in the general population, without diagnosed disease or specific symptoms, to an appointment with a health professional where screening tests are carried out; these tests may include clinical tests (e.g. blood pressure measurement) and/or questionnaires (e.g. about health behaviours or family history). The results of these tests are then used to inform ongoing clinical care. The health check visit may also include conversation about risk factors and/or a range of behavioural interventions.

Health check programmes aim to achieve several different goals. They aim to improve the assessment of disease risk, and identify individuals at risk of developing chronic disease before serious problems emerge. This should enable both more timely clinical treatments (e.g. statins or diabetes medications), and also (in conjunction with further support and referrals to relevant services) help people to make positive changes in their health behaviours. Finally, they may reassure individuals who are at low risk but are concerned about future health problems. An outline logic model for the intervention is shown in Figure 4.

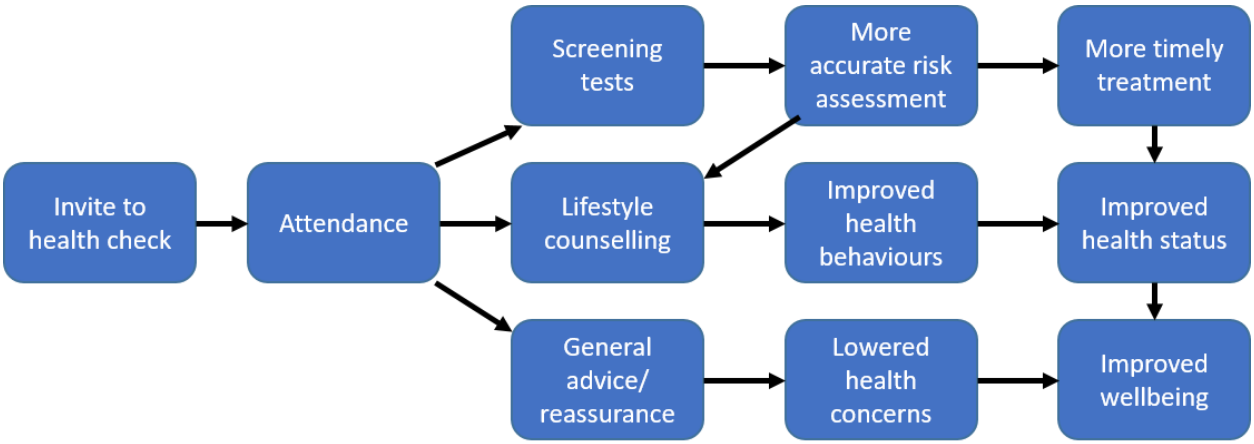


Figure 4: Outline logic model for health checks

The NHS Health Check was introduced in 2009. It targets 40-74-year-olds without pre-existing conditions, a total of 15 million people in England, who are invited to attend a health check every five years (Public Health England 2019). The programme aims to reduce the incidence of highly prevalent chronic non-communicable diseases, with a particular focus on reducing cardiovascular risk. The NHS Health Check includes a risk assessment which incorporates both self-report measures of demographics, family history and health behaviours (smoking, alcohol use, physical activity), body mass index, and clinical screening tests (cholesterol, blood pressure and, depending on risk score, diabetes).

Potential unintended consequences

Health checks may also have a range of unintended consequences (Goodyear-Smith 2013; Capewell, McCartney et al. 2015; Stol, Schermer et al. 2016; Krogsbøll, Jørgensen et al. 2019). Offering screening tests to healthy people runs the risk of overdiagnosis, in that the people identified as at risk by the test may not always face a risk which is clinically significant. This may lead to harms including overtreatment, and the associated unnecessary costs. It could also cause worry and mental stress through the 'medicalisation' of healthy people; statistical risks are often not well understood (Zipkin, Umscheid et al. 2014). Conversely, receiving reassurance that one is at low risk could reduce motivation to engage in beneficial health behaviour change. Any extensive health check programme will also be costly to implement, and it is legitimate to ask whether this is the best use of resources.

It is also possible that the health check could widen health inequalities at a population level because of differential uptake. If people who are at higher risk, due to behaviours and/or demographic or environmental factors, are less likely to respond to the invitation to the health check – and if the intervention has on average a beneficial effect for individuals – then inequalities in health outcomes may become wider as a result. Figure 5 shows an outline 'dark logic' model (Bonell, Jamal et al. 2015) illustrating potential unintended consequences.

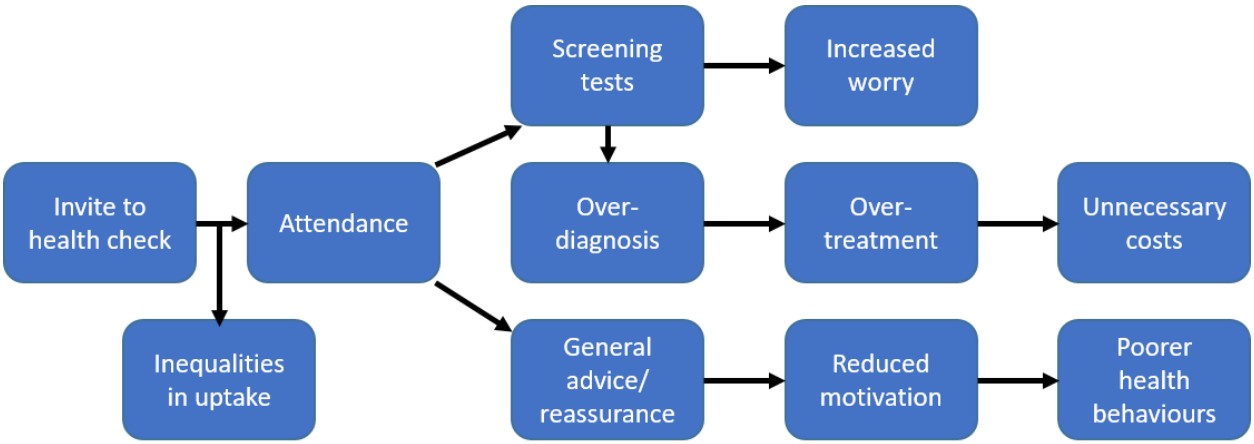


Figure 5: Outline 'dark logic' model for health checks

Effectiveness and implementation

Evidence from systematic reviews

A systematic review and meta-analysis of health checks for the Cochrane Collaboration finds that while they increase the number of new diagnoses, they are not effective for health status outcomes (overall mortality or incidence of heart disease or stroke) (Krogsbøll, Jørgensen et al. 2019).[6] However, another review argues that health checks carried out in general practice settings rather than community or workplace settings (like the NHS programme) are effective for clinical measures of risk such as blood pressure and BMI, although not for smoking behaviour (Si, Moss et al. 2014). Thus, the evidence indicates that health checks can improve clinical indicators, but do not improve health behaviours or health status outcomes. It should be noted that the evidence base is relatively small ((n=19) primary studies across the two reviews) and most studies are several decades old, so the applicability of these data is debatable.

A systematic review focusing on uptake finds evidence for inequalities in attendance, with men less likely to attend than women and lower-SES people less likely than high-SES; people with higher risk factors and/or less healthy lifestyles are also less likely to attend (Dryden, Williams et al. 2012). Thus, there is some evidence to support the concern that health checks may worsen health inequalities. A systematic review of cost-effectiveness reports conflicting findings, with some studies showing health checks to be very far from cost-effective, and others showing them to be cost-saving (Lee, Lawson et al. 2017). The uncertainty appears to be due to a range of factors, including the effectiveness of health checks for clinical risk indicators, the extrapolation of the latter to health status outcomes, and implementation factors including the rate of uptake.

Evidence on the NHS Health Check programme

Studies and routine data on the NHS Health Check programme show that it is effective for increasing new diagnoses (Robson, Dostal et al. 2016). Observational data indicate that the

[6] Krogsbøll and colleagues' analysis informed the Danish government's decision not to implement a health check programme (Krogsbøll, L. T., K. J. Jørgensen and P. C. Gøtzsche (2013). "Universal health checks should be abandoned." BMJ : British Medical Journal **347**: f5227.).

programme has some effect on some clinical indicators, although effect sizes are small (Artac, Dalton et al. 2013; Chang, Lee et al. 2016); modelling studies suggest these translate into improved health outcomes, although again the benefits are relatively modest (Mytton, Jackson et al. 2018).

Uptake of the programme has generally risen over time, and the most recent routine data (Q1 2019) show an uptake rate of 49.1% nationally, although this varies widely by area.[7] Studies reach conflicting conclusions on equity in uptake, with some studies finding uptake to be lower in lower-SES areas (Attwood, Morton et al. 2016) and others no difference (Robson, Dostal et al. 2016); it is unclear how individual-level SES impacts on uptake.

A systematic review of evidence of patients' views of the programme shows generally high levels of satisfaction, but some desire for more detailed information and better follow-up (Usher-Smith, Harte et al. 2017). Studies of clinicians' views suggest that most are broadly supportive, but reveal a range of concerns, particularly uncertainty about implementing the scheme and concerns about workload and administrative burden (Baker, Loughren et al. 2015; Krska, du Plessis et al. 2016).

Evidence on unintended consequences

There is limited evidence on the unintended consequences of health checks. The NHS programme is costly to implement, with one estimate putting the total annual cost at £450 million (Capewell, McCartney et al. 2015); however, as noted above, it is as yet unclear whether the programme is cost-effective overall. The evidence on inequalities in uptake is also somewhat conflicting (above). The limited data on anxiety and worry following health checks do not indicate that this is a significant harm (Krogsbøll, Jørgensen et al. 2019). It is unclear how far health checks lead to overdiagnosis and unnecessary treatment; however, studies of screening interventions show that these harms are potentially important (Nelson, Pappas et al. 2016), and this may also be true of health checks.

Potential applications of PP/PPH approach

A PP/PPH approach could potentially be used in a number of ways within the NHS Health Check programme. Below we briefly consider some examples. Figure 6 shows the potential applications superimposed on a partial version of the logic model.

[7] www.healthcheck.nhs.uk/commissioners-and-providers/data/

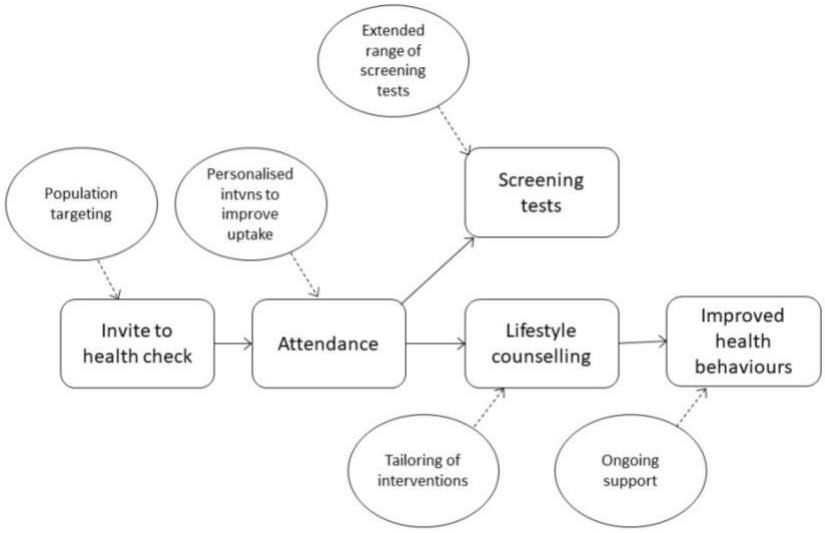


Figure 6: Potential applications of PP/PPH approach

1. Population targeting and health checks

Rather than including all people in the given age range, the health check programme could be targeted only at those individuals or communities at higher risk. Economic modelling studies indicate that this would make the programme markedly more cost-effective and reduce health inequalities (Kypridemos, Collins et al. 2018). Individual targeting can be based on clinical judgement of higher risk, as for example in the Australian health check programme, which focuses on 45-49-year-olds at elevated risk of cardiovascular disease (Australian Government Department of Health 2014). However, a broader range of data could potentially be used to inform an algorithm for deciding which patients to invite, including information from electronic patient records. This could take the form of a two-stage process in which a digital health check is offered to everyone in the population, with face-to-face checks for those identified as at higher risk (Department of Health and Social Care 2019). Targeting could also take into account demographic and social characteristics, addressing some of the limitations of the current 'one size fits all' approach; for example, some ethnic minority populations have earlier onset for cardiovascular disease than the white population, suggesting that people from these groups should be invited to the health check at an earlier age (Khunti, Walker et al. 2011). This being said, it seems unlikely that any highly sophisticated data analysis would be required for population targeting, and relatively straightforward approaches based on a brief questionnaire and/or clinical judgement would probably work just as well.

2. Personalised interventions to improve uptake of health checks

As already noted, while the uptake of NHS Health Checks has steadily increased over the lifetime of the programme, it still remains substantially below target levels, and there are concerns about the potential for differential uptake to worsen health inequalities. Interventions to improve uptake could thus improve the effectiveness, cost-effectiveness and equity impact of the programme.

Current guidance already recommends targeting interventions to increase uptake, e.g. telephone reminders, to high-risk groups (Public Health England 2016). Public health commissioners have implemented targeted outreach initiatives to improve uptake in disadvantaged communities (Riley, Coghill et al. 2015). While such programmes are promising, they are much more costly than generic reminders. PPH-informed approaches such as targeted text reminders or personalised advertising are less resource-intensive and could play a part in increasing uptake, although the potential gains are probably limited.

3. Extended range of screening tests as part of health checks

The NHS Health Check currently includes testing for blood pressure, cholesterol, and, for those at high risk, diabetes. In theory, a wide range of testing could be included in a health check programme. Some older programmes included very broad testing regimens (including, for example, chest X-rays and electrocardiograms), but these are no longer recommended as they frequently find abnormalities which are not clinically significant, and offering these tests to healthy people is likely to do more harm than good (Krogsbøll, Jørgensen et al. 2019).

This argument is likely to apply to newer forms of testing, including genetic testing. The main value of genetic testing to date has been in personalised medicine, and in testing for genetic disorders in people with known family risk. Testing healthy people with no known family history of genetic disorders would represent a very different kind of intervention, with no clear clinical value and a real potential for serious harm (Caulfield, Evans et al. 2013; van El, Cornel et al. 2013). While some have suggested that communicating genetic risk could help to promote healthy behaviour change, the evidence indicates that this is not the case (Hollands, French et al. 2016).

4. Tailoring of health checks

The NHS Health Check pathway includes a range of interventions including general lifestyle advice, preventive medication (e.g. statins) and referral to specialist services such as smoking cessation, alcohol services and weight management (Public Health England 2017). The existing pathway is already tailored in a sense, in that progress through the pathway is guided by risk scores and algorithms based on the data collected in the health check, as well as by the clinical judgement of the GP. In its current form this tailoring is relatively straightforward, and in principle could be made more sensitive by drawing on a wider range of data or more sophisticated methods of analysis, with the goal of making the identification of risk factors more accurate, and more clearly identifying interventions of value to individuals.

It is unclear how much scope there is for tailoring to improve outcomes related to behavioural risks (as distinct from genetic risks). It is not clear that new sources of data would significantly improve the accuracy of measurement of risk factors, and utilising personalised risk within behaviour change interventions does not appear to be effective (French, Cameron et al. 2017). However, the approaches evaluated to date have been relatively limited, and there may be scope for further improvement.

For example, the use of incentives to support behaviour change has received considerable interest. While incentives appear to be effective for many people (Giles, Robalino et al. 2014; Mantzari, Vogt et al. 2015; Finkelstein, Bilger et al. 2019), their effectiveness may depend on

both the type and amount of incentive offered and on individual characteristics such as socio-economic status (Finkelstein, Bilger et al. 2019). This suggests that personalised approaches to incentives could be a promising focus of future research.

5. Ongoing support with health checks

There is scope for PPH approaches to inform the ongoing support offered after the health check, particularly using tools such as apps or wearables to assist in making or maintaining healthy behaviour change. Such approaches are much less costly than in-person services, and could be made more widely available to people who are not at high risk but would still benefit from ongoing support. They could include personalised reminders and feedback to address specific needs identified at the health check.

There is some evidence that such approaches are effective in improving health behaviours, although engagement often drops off after a few weeks (Schoeppe, Alley et al. 2016). Given that they are low cost and unlikely to have negative impacts, there is potential to explore their use further, provided that they do not displace more focused support for those at high risk.

Summary and discussion of applications of PPH approaches to health checks

PPH approaches could be implemented within a health check programme in a range of ways, particularly for improving the targeting of the programme overall, increasing uptake, and offering ongoing support after the health check. With the possible exception of population targeting, the likely improvements in outcomes and/or reductions in costs would probably not be dramatic, but are worth investigating. In particular, further work would be valuable to explore the potential of PPH approaches to increase uptake of the health check in disadvantaged groups, and to facilitate maintenance of healthy behaviours beyond the health check itself.

The health check programme could also be seen as a framework for rolling out PPH approaches across the general population, particularly personalised technologies such as apps. As noted, this is of particular interest as a means of engaging low- and medium-risk groups who do not meet the criteria for more intensive intervention. However, more evidence is needed on the effectiveness of such strategies.

Health check programmes focus on the individual determinants of health outcomes, and arguably promote a narrative of individual responsibility for those outcomes which detracts from the social and structural determinants of health (Capewell, McCartney et al. 2015, Stol, Schermer et al. 2016). One modelling study indicates that population-wide interventions addressing the latter would significantly improve effectiveness, cost-effectiveness and equity relative to the health check programme alone, even on very optimistic assumptions about the implementation of the programme (Kypridemos, Collins et al. 2018). Given that the impacts of the programme as currently implemented appear to be modest, and the likely additional improvements from PPH approaches to be incremental, there is an argument that public health resource would be better spent on structural and policy interventions addressing the determinants of population health outcomes (Marteau, White et al. 2019).

8. The application of a Precision Public Health approach to Community-based interventions

This section reports on the second of two case studies in which we sought to examine the potential of PPH when applied to specific use scenarios. Here we explore how a PPH approach could add value to community-based interventions.

Intervention theory

We begin by examining community-based interventions as a generic intervention model (see additional information in Appendix 5), and consider a more focussed example of school-based smoking interventions in Appendix 6. Community-based interventions are defined here as interventions that target 'a group of people united by at least one but perhaps more than one common characteristic, including geography, ethnicity, shared interests, values, experience or traditions' (O'Mara-Eves, Brunton et al. 2013). Communities may be self-defined, or communities may be defined by people outside the community. Area-based interventions are aligned with community-based interventions although are focussed exclusively on certain geographies, which may or may not also align with a 'community'. Community-based interventions draw on a number of different intervention modes and foci, although share a common set of principles.

The epidemiological foundation of community-based interventions is grounded in the work of Geoffrey Rose, who demonstrated the principle that at a population level, more cases of a disease could arise from large numbers of people exposed to a low risk than a smaller number of individuals at high risk (Rose 2001; Merzel and D'Afflitti 2003). In turn the logic follows that, interventions focused on the entire population aiming to shift the risk distribution for all result in the most effective improvement in population health. Community-based interventions draw on these principles, in that the community may be deemed to be at higher risk of a particular outcome or health determinant, but the individuals may or may not be at high risk (Katikireddi and Valles 2015).

As the logic model below emphasises (Figure 7), community-based interventions involve a number of different approaches and creation of multilevel systemic change can follow from a number of stages:

1. Focus – communities are viewed as the target, agents or resources of change, as opposed to merely the intervention setting
2. Identification of need – public health needs and intervention foci are identified by the community
3. Involvement – Communities members are involved in the design or delivery of an intervention
4. Processes – Key processes are set in place to involve the community including the development of underlying theory

Where interventions incorporate these stages (shaded green below), they tend to lead to positive change. However, where these processes are not observed, there is a risk that desired

outcomes are not reached, and that potentially unintended consequences occur (shaded blue below) such as greater levels of stigmatisation and the failure to stimulate systemic change. Further details of these processes are provided in Appendix 5.

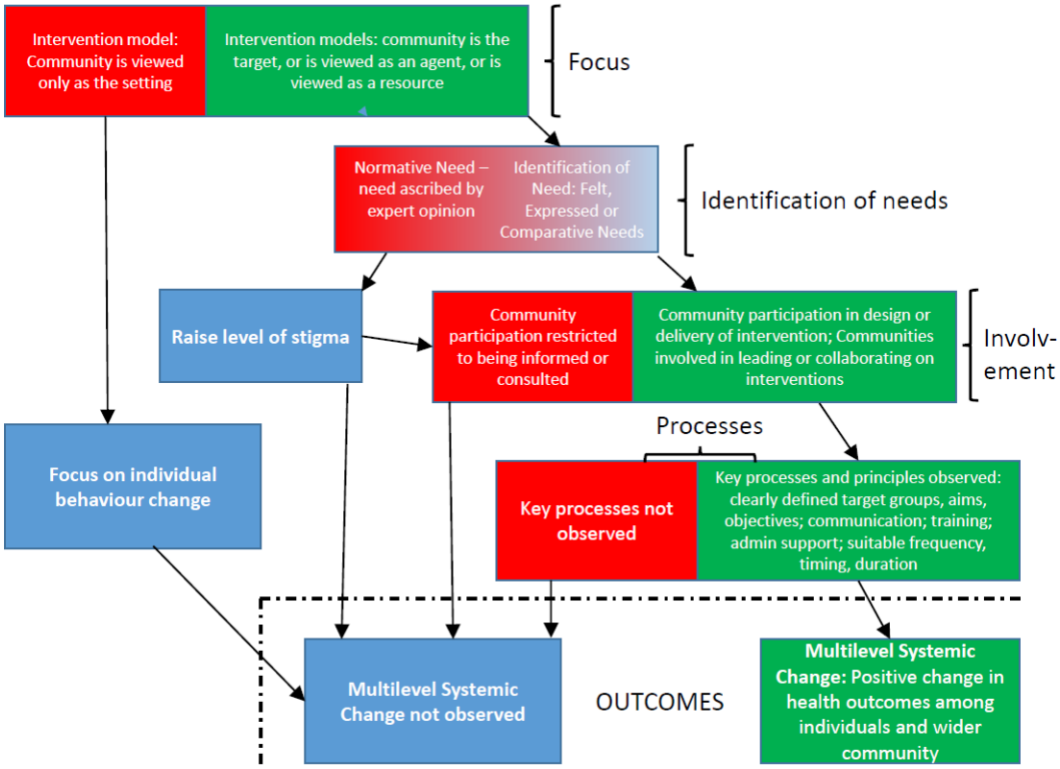


Figure 7: Logic model - general principles of community based interventions

Impacts of community-based interventions of public health outcomes

Higher levels of participation, empowerment, coproduction, delegation of power and control to communities are associated with a greater impact on outcomes, while low intensity forms of engagement that do not seek to build community capacity such as consulting and informing do not have a substantial impact on outcomes (O'Mara-Eves, Brunton et al. 2013, Brunton, Caird et al. 2015). Consequently, interventions that include higher levels of engagement are associated with greater effectiveness (Brunton, Caird et al. 2015).

Overall however, community-based interventions are generally characterised by moderate levels of effectiveness (Wandersman and Florin 2003; Brunton, Caird et al. 2015; O'Mara-Eves, Brunton et al. 2015). This may reflect issues around the limitations in the theories used and in the intervention design, with community-based interventions generally less intensive on an individual level. An example where community-based interventions have been an effective approach include HIV prevention approaches (Merzel and D'Afflitti 2003) due to their impact in successfully changing of community norms around safe sex among men who have sex with men. HIV prevention strategies sought to modify the social context in which risk behaviours

occurred and employed strategies such as identifying role models (e.g. within bars and clubs) and strengthening whole-community capability in engagement with risk reducing behaviours.

Potential applications of PP/PPH approach

Below in Figure 8, we outline where a PPH approach could potentially strengthen the delivery of community-based interventions. In Appendix 6, we consider the more specific example of school-based smoking prevention intervention and potential applications of a PPH approach for community-based interventions.

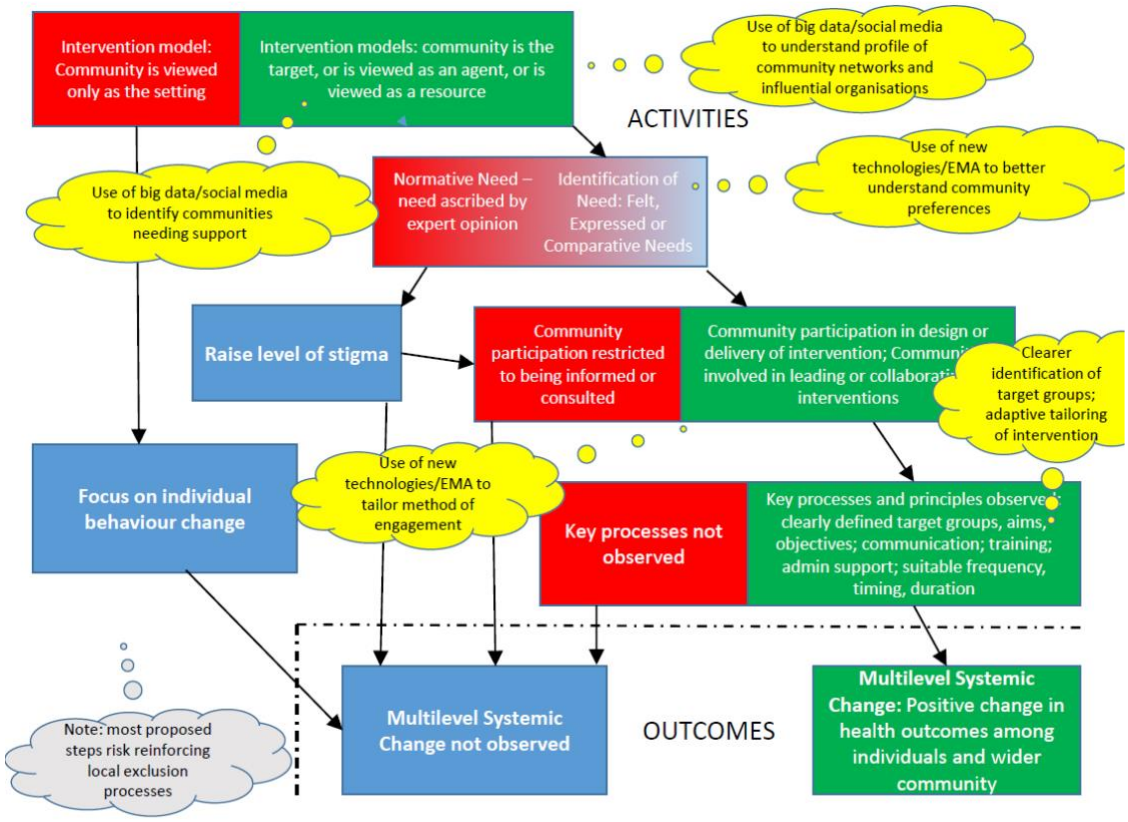


Figure 8: Logic model - general principles of community based interventions and potential of PPH to adapt the model

Focus

Identifying communities on the basis of new data and methods

(i) Targeting communities of need through social media: New sources of data, and new means of analysing these data, could offer opportunities for better targeting of communities through identifying and establishing need on the basis of a wider range of factors and with more up-to-

date observations. Online social networks offer round-the-clock access to vast numbers of individuals whose data could be used to identify communities and networks through which social norms, social influence, and social support may be analysed in real time (Cobb, Graham et al. 2011).

Online communities of need may be identified and targeted through similar algorithms as have been used successfully in marketing campaigns and in recruitment strategies for public health interventions. Among younger people for example, Facebook has been used to target younger people on the basis of users' profile information including age, gender, location, language and keywords (Park and Calamaro 2013); Facebook and other forms of social networking have also been used to reach those who have been difficult to reach through other means (including email, phone calls and school visits) (Park and Calamaro 2013; Cahill, Wertz et al. 2019). Using Twitter and machine learning methods, Chu, Colditz et al. (2019) examined the way in which sentiment analysis could be used to identify a narrow group of users who posted positive and negative tweets about hookah tobacco smoking. Communities of users who express ambivalence towards harmful health behaviours may be of interest to public health practitioners in the future as they may be more receptive towards targeted campaigns to cease or prevent adoption of harmful behaviours. Other examples where subgroups and networks have been identified using social media include the use of Reddit to monitor sentiments around eating disorders to identify online communities who expressed pro-eating disorder comments and those who expressed 'thinspiration' comments (McCaig, Elliott et al. 2019); and the use of Grindr (an online dating app) to identify specific subgroups of men who have sex with men based on location and access to medical treatment (Hampel, Kusejko et al. 2017). These studies can be considered as examples of 'infodemiology', involving the study of the determinants and distribution of information in an electronic medium (Eysenbach 2009); similarly 'infoveillance' involves the use of 'infodemiology' methods to monitor health trends (Eysenbach 2009). In an approach that blended online and geographic data, Liu, Chen et al. (2019) combined Twitter information with geo-located data and survey data to observe that the volume of tweets on physical activity across areas was correlated with other survey-based indicators of physical activity; in the future such information could provide a more timely assessment for targeting areas/communities based on low levels of physical activity.

While these studies exemplify the use of social media to classify individuals and networks, their utility in (i) identifying 'communities', with precision; and (ii) in producing insights that can be used to inform community-based interventions is more questionable. The notion of community usually connotes a set of relationships based on a common tie, usually a sense of identity (Scott and Marshall 2009); the use of a particular form of social media and expressing common sentiments may transcend this notion of community. Furthermore, using social media to target communities does raise methodological challenges. This includes the capacity of machine learning methods to detect and classify minority classes (or communities) due to conventional algorithms being biased towards identifying large groupings in an effort to optimise error rates (Chu, Colditz et al. 2019). Ethical issues may also arise from researchers 'lurking'

online and harvesting data, including monitoring the extent to which communities are aware of the way in which their data are being collected and whether they provide informed consent (for example Griffiths and Whitty 2010). The promise of precision approaches to identifying communities in need on the basis of social media data alone is not well established and such approaches may conflate terms such as community, networks and users (of technology). While it is clear that there are a number of uses of social media data for developing public health campaigns (Moorhead, Hazlett et al. 2013), there has been less around using these data to identify communities at risk. Nevertheless, these data may be useful in understanding community dynamics (see below).

(ii) Targeting areas through geo-located data: A persistent issue in researching area and neighbourhood effects on health has been the failure of administrative area boundaries to adequately reflect communities and neighbourhoods on the ground (Van Ham, Manley et al. 2012). New sources of geo-spatial data may offer new opportunities to understand the features of neighbourhoods that influence public health with a greater degree of nimbleness than has been the case with conventional survey-based sources of data. Webcams, crowdsourcing and social media all offer opportunities to understand the dynamic features of neighbourhoods that potentially can be used to target areas of greatest need (Schootman, Nelson et al. 2016). Google Street View, for example, which is a source of big data that offers interactive high resolution panoramas of public thoroughfares has been used in health research to assess the condition of the built environment, to assess compliance with health policy (e.g. the location of traffic calming features and the presence/absence of smoke-free signage), and for study site selection (Rzotkiewicz, Pearson et al. 2018). Google Street View data were found in one study to have a good level of reliability compared to physical observations, albeit with some caveats around how current these data were (Less, McKee et al. 2015; Jones-Webb, McKee et al. 2018). Similarly, machine learning has been applied to geo-located photos uploaded to Flickr to extract information on geographic features of areas (Sengstock and Gertz 2012); by extension such methods could be suitable for identifying public health features of interest such as particular brands of fast food outlets. The representativeness of such data haven't been explored however; it may be assumed that the areas with the highest levels of deprivation (that are less picturesque) may have the weakest profiles on such media and less extensive data.

There may be more promise in identifying areas of need (as opposed to communities per se) using novel computer science-driven methods of analysis (machine learning, artificial intelligence) to analyse established data sources, rather than those sources that are 'new' per se. For example, Goin, Rudolph et al. (2018) used machine learning approaches to identify a subset of 18 of the most predictive ecological characteristics of firearm violence in the USA, from a potential 340 characteristics under consideration. The methods employed helped to overcome common analytic issues usually encountered in analysing data of these types including multicollinearity (Goin, Rudolph et al. 2018). Similarly, in one of the few self-defined PPH studies encountered, Freeman, Boylan et al. (2017) explored the feasibility of producing cancer incidence rates for geographic units that transcended administrative (census)

boundaries, and redefined boundaries to reflect health service boundaries more closely aligned with public health decision-making needs. There may be greater potential in applying PPH approaches in identifying and targeting *areas* for public health interventions, as opposed to targeting *communities*.

Using new sources of data to understand how communities interact and their influences:

While the potential of PPH for identifying communities in need (as opposed to areas) may be unclear, there may be greater scope for using PPH for understanding community dynamics and community influencers once target communities have been identified.

For a number of years, public health researchers have been interested in examining online behaviours and expressions as a way of understanding health trends, for example in analysing content from within online chat groups frequented by men who have sex with men to inform HIV/AIDS prevention programmes (Rhodes 2004), or researching online cannabis communities to detect new forms of cannabis ingestion (Meacham, Roh et al. 2019). Recent applications have also used machine learning in order to understand characteristics and behaviours of self-defined online communities, such as survivors of suicide (Ambalavan, Moulahi et al. 2019). These analyses have also extended to offer insight into how users react to one another as opposed to solely being based on the self-presentation of users. For example, in examining alcohol discussions in an online social network for smoking cessation Cohn, Amato et al. (2019) examined how interactions with other users patterned alcohol discussions. The researchers speculated that the findings could be used to shape the development of interventions that were customised or tailored in real-time, based on users' connectivity and the content of their sentiments. However, while there are greenshoots around the potential for using these data to understand some aspects of how online communities interact, a number of these applications have been somewhat limited in offering additional understanding of community dynamics as they typically include little additional information on the sociodemographic characteristics of users.

Social media data may be useful to better understand community influencers and how communities network. For example, there have been a number of studies based on QuitNet, a large virtual community for smoking cessation which has over 100,000 new registrants annually (Myneni, Cobb et al. 2013). Data extracted from QuitNet, and analysed through machine learning (Natural Language Processing), has been used to identify sub-communities of users and to identify opinion leaders within those communities who may change social norms and accelerate behaviour change (Myneni, Cobb et al. 2013). PPH approaches utilising social media data are likely to be able to identify influencers across different platforms, and while there may be some advantages in identifying and potentially involving online opinion leaders terms of gaining access and legitimacy among harder to reach participants (Heldman, Schindelar et al. 2013; Myneni, Cobb et al. 2013); involvement of influencers in public health campaigns can

pose some risks in terms of loss of message control (Heldman, Schindelar et al. 2013). Furthermore, as was the case above, while it may be possible to detect opinion leaders based on relative measures, whether these individuals do in fact influence behaviour is far from certain, and likely to be moderated by other characteristics including age and educational level which are often measured imprecisely (or not at all) on social media.

On an ecological level, a number of apps support users to add further contextual features of areas that can illuminate how people engage with areas, or the condition of areas such as the provision of open spaces for physical activity (Hoffimann, Campelo et al. 2018) or obesogenic environment (Vandevijvere, Williams et al. 2017). Crowdmapping involves citizens becoming active agents in data collection, and can be useful in helping to understand how communities experience public health challenges. For example, crowdmapping has been used in the North East of England to enable mothers to explore and contribute to a map which 'describes how supportive the local community and services are toward women who breastfeed' (Balaam, Comber et al. 2015, p5). This type of data could enable public health practitioners to better understand wider community-level determinants of breastfeeding. As well as providing data on ecological features, new sources of data and methods of data collection could also allow for insights into networks and engagement in specific areas or among physically connected communities. For example Stopczynski, Sekara et al. (2014) collected data through questionnaires, but also through sensors on smartphones that collected data including GPS, WiFi, Bluetooth, calls, SMS, battery, and application usage to build a better knowledge of how 1,000 university students interacted in networks in Copenhagen.

PPH approaches may be able to establish and monitor the emergence and nature of some community norms, and how they are shaped by opinion leaders, in a way that traditional sources have not been able to. However, there are clear caveats around their use and scalability. Similarly some of the most ostensibly promising approaches for understanding how geographic communities operate, using smartphones and sensor data, have only been trialled with narrow digitally literate populations. In terms of online communities, as was discussed above, the value of online sentiments in predicting actual health behaviours is unclear and the validity of ascribing the label of 'community' onto users of different social media platform who share some sentiments is contestable. Issues of consent are not addressed fully in some studies examining online communities. For example, in their paper on content analysis of QuitNet, Myneni, Cobb et al. (2013) include no information on ethical approvals or the ethical dimensions of the work they undertook. In contrast in developing their study capturing sensor data, Stopczynski, Pietri et al. (2014) propose the concept of 'living informed consent' and their study included an 'Authorisation Dashboard' that allowed participants involved to monitor and provide (or withdraw) consent at different points in the study (Stopczynski, Sekara et al. 2014).

Identifying need

Using new sources of data to identify community preferences regarding intervention targeting and tailoring

Community-based interventions that are based on supporting communities to directly identify their own needs (felt need) are likely to be most aligned with intervention models that view communities agents or resources of change (O'Mara-Eves, Brunton et al. 2013; Brunton, Caird et al. 2015). PPH approaches, utilising data collected online and through apps, may help to understand felt need through:

(i) Enabling public health researchers to 'listen in' to conversations taking place to identify preferences, which could be considered a form of identifying implicit felt needs. Such studies, as described above, could involve researchers increasing understanding of community preferences through analysing data collected passively or actively submitted to forums, sometimes using artificial intelligence/machine learning. Examples of similar applications are described above.

(ii) Enabling public health researchers to measure the preferences of hard to reach populations using survey-based methods delivered online, for example measuring preferences regarding HIV self-testing among young men (Merchant, Clark et al. 2017). These methods could be considered a form of identification of explicit felt need, with data usually collected from specific and pre-defined community populations similar to conventional survey methods.

(iii) Crowdsourcing preferences from communities, which could be considered a form of identification of explicit felt need. Crowdsourcing involves a group of non-experts and experts working together to solve a problem, through the input of large numbers of people on a particular task or project. Crowdsourcing may be aligned with community-based intervention theory through providing a mechanism through which communities can share their ideas and preferences through existing social networks, thereby harnessing existing community resources (Tang, Ritchwood et al. 2019). In a narrative review of sexual health interventions that compared crowdsourced materials with materials produced by researchers, studies that drew on crowdsourced materials were found to be of similar or greater effectiveness than standard materials, and were less costly (Tang, Ritchwood et al. 2019). However, the term crowdsourcing is being used widely in this field, and is sometimes applied to strategies that simply involve the recruitment of individuals using social media or specific platforms such as Mechanical Turk, and their subsequent participation in standard (pre-determined) interventions delivered online (Naslund, Aschbrenner et al. 2015; Cunningham, Godinho et al. 2019); such strategies are not aligned with communities identifying a 'felt need', and seem to stray from some of the underlying principles of crowdsourcing.

While there is potential to harness new sources of data to understand community preferences and identify felt needs to design community-based interventions, there are some clear issues

around the quality of these data. There is no evidence that data collected through these means do provide a more accurate depiction of community preferences than data collected through conventional sources, although they may be less costly to collect/harvest. Many sources of big data lack principles for sampling on particular user characteristics and the resulting data are likely to be highly unrepresentative with regards to sociodemographic characteristics. This impedes the generalisability of the findings, to the extent that the findings generated from one social media platform may not be generalisable to another (Lazer and Radford 2017). Uneven expressions of felt need could exacerbate social exclusion and health inequalities within communities, with felt need being solicited by those who are more socially engaged or included. There are also additional considerations around the validity of preferences expressed online.

Involvement

Using new approaches for involving communities in leading, designing and delivering community-based interventions

Community-based interventions that empower communities in leading, designing and delivering interventions are those that achieve greater effect sizes. These interventions may view enhancing community capacity and empowerment as an outcome in itself. There are a number of examples in the literature where community participation and online approaches have been integrated:

- (i) community input on the design of an (online) intervention (e.g. Muessig, Baltierra et al. 2014);
- (ii) to aid in the delivery of an intervention, for example through peer delivery (Hwang, Ottenbacher et al. 2013);
- and (iii) understand empowerment processes within an online communities (e.g. Verberne, Batenburg et al. 2019).

There are fewer examples where online approaches have been used to directly stimulate community empowerment. One potential exception is the ‘Deadly Choices’ (mainly) online intervention, focussed on improving the health choices of Indigenous Australians in South East Queensland, with ‘deadly’ reflecting the Aboriginal English use of deadly to signify good or fantastic (McPhail-Bell, Appo et al. 2017). The ‘Deadly Choices’ intervention expanded online guided by ongoing interactions between a team of mainly Indigenous Australian researchers, the Indigenous community and public health practitioners, and managed to achieve a large following. The online presence was tailored to reflect Indigenous Australian community-based interactions and concepts of health, and five interconnected principles for the collaborative use of social networks for Indigenous health promotion were enacted to ensure empowerment including the creation of a dialogue; building community online and offline; incentivisation of

healthy online engagement; celebrations of Indigenous identity and culture; and the prioritisation of community partnerships (McPhail-Bell, Appo et al. 2017). While ethnographic research methods highlight high levels of engagement and reports of changing health risk behaviours, other comparative evaluation evidence was not found. The study authors noted the absence of other exemplar studies in the literature that employed 'community-oriented, dialogical uses of social network sites rather than as one-way tools for health education' (McPhail-Bell, Appo et al. 2017). Similarly, a systematic review of online community engagement interventions found that while leadership and collaboration processes were recorded in all eleven included studies, there were generally lower levels of engagement compared to face-to-face interventions (Stokes, Richardson et al. 2015).

In the examples above, while online methods were used to involve communities within interventions, there is less evidence that the *data* collected online (or computer science-driven methods of data analysis) was instrumental to this process. PPH approaches may be relatively limited with respect to involving communities in leading, designing and delivering community-based health interventions.

Summary and discussion of applications of PPH approaches to community-based interventions

Many existing community-based interventions have limited ecological reach and are too focussed on individuals (Merzel and D'Afflitti 2003). Similarly, several existing community-based interventions fail to take a systems-based approach to understanding connections between individuals and their networks. This critique is unlikely to be redressed by most of the applications of PPH described in the literature. A risk of adopting a PPH approach is continued under-theorising of how individuals interact with, and form, communities. Critiques about the potential unfair and exclusionary stigmatising impact of targeting communities of people (as opposed to communities of behaviours), have been levelled at existing community-based approaches (Katikireddi and Valles 2015). Data-driven approaches characterised by a PPH approach could exacerbate this issue, and a more granular level of risk prediction could be erroneously interpreted as a more granular understanding of individual causal forces. These critiques are not necessarily unique to PPH, although it is unclear how adopting a PPH approach could help to resolve these.

Nevertheless, we have identified some areas where PPH approaches could aid in understanding the features and dynamics of areas and communities. The availability of new sources of data, and the ability to crowdsource data about community characteristics, dynamics, needs and wishes may facilitate community engagement within target areas. The availability of these data are also likely to be useful for improved understanding of selective forms of community interactions. While not explicitly discussed in the literature, new methodological techniques may also open the window to more complex statistical modelling techniques that better capture system-level relationships and dynamics. Community engagement interventions have been associated with relatively modest effects (Popay, Whitehead et al. 2015), although there are examples of well-known community-based interventions that have been highly

effective in improving health outcomes (Merzel and D'Afflitti 2003). The extent to which PPH approaches can help to revolutionise the design and delivery of community-based interventions is unclear, and uncritical application could in fact be harmful in some cases. However, it is likely that the greater availability of data sources to understand communities could represent a useful adjunct, provided that issues around representativeness and validity of constructs and data are investigated before use.

9. Summary and conclusions

What is precision public health?

The term precision public health (PPH) is used differentially by authors and appears to encompass:

(i) An ambition towards achieving greater precision in understanding the profiles of areas and people

(ii) An approach which seeks to mobilise new sources of data and new methods of analysis involving Artificial Intelligence and Machine Learning for the design and implementation of interventions

Each aspect is somewhat problematic when trying to understand what PPH is and assessing the underlying evidence around the concept.

PPH can be seen as a process in which various methods and data are marshalled in order to add greater precision to understanding risk profiles (and predicting outcomes), usually with improved precision at a geographic or temporal level, or at an individual level. All self-defined PPH studies focus on increasing precision, and usually claim (whether substantiated or not) that the methods and data lead to an improvement in objective accuracy of predicting risk compared to previous research on the same topic. The underlying assumption is simply that there is an advance on current understandings of the health and social profile of people or areas. However, this is not a clearly distinguishable approach or process as all public health research strives towards greater granularity, or precision, as a way of increasing knowledge of profiles, risks and predicted outcomes. In addition, an uncritical move towards greater precision risks overlooking a key question which is whether targeted interventions are actually more effective.

We regarded studies that used either new data or new AI-driven analytical methods as being examples of PPH studies. In applying these definitions, we encountered challenges. Firstly, a number of self-defined PPH studies did not meet the parameters of this definition, despite offering an improvement on 'precision'. Secondly, there is potential for studies to be classed as examples of PPH, for instance through using data collected through online technologies, despite not offering an advance in 'precision'. For example, information collected from social media user profiles on characteristics such as age, gender, ethnicity and location, is unlikely to offer any advance on conventional survey data (and could have additional flaws). Conversely, data on these baseline characteristics are often missing from studies that analyse online interactions, despite being critical as determinants of health. Thirdly, we observed little direct empirical evidence of any advance in precision in targeting and particularly tailoring, and little empirical evidence that targeting and tailoring had an impact on improving outcomes. Therefore, if PPH is

a distinct concept, using these parameters, we are only at the starting blocks of understanding why we need to adopt PPH approaches and what benefits could be observed.

We also recognise that in examining PPH as being separate from traditional public health practice, we may be creating a false dichotomy. PPH and traditional public health practice are unlikely to be mutually exclusive in practice and could be used in tandem. However, as discussed above, without this dichotomy, it becomes unclear as to why PPH should be regarded as a distinct concept or approach and not just a suite of different data types and analysis methods. Given the lack of empirical evidence around PPH as an approach, it may even be advantageous to regard PPH in this way as simply a suite of new data sources and computer science-driven analysis approaches, each needing to be evaluated distinctly with regards to the added value to public health practice.

What is the evidence base around precision public health?

Analyses of commentary studies found that (i) the PPH field may be highly influenced by commentary and non-systematic review pieces; (ii) that commentators on PPH often attempt to provide evidence for claims but the link between the evidence and the claim is often unsubstantiated; (iii) that many of the assumptions underlying PPH have not been evidenced, suggesting that there needs to be a measured strategy to adopting PPH approaches. Claims about the effectiveness of PPH and of PPH being an advance on current public health approaches tended not to be supported by empirical evidence (see table 3).

Empirical studies that use the term PPH highlight a discrepancy between the claims made in commentaries about the potential of PPH, and the focus of empirical examples of PPH. The former emphasise that precision can be achieved through targeting and tailoring interventions towards narrow social profiles using data reflecting the micro-level of individuals' day-to-day lives. The latter, with the exception of genomic studies, offer evidence about greater precision predominantly using ecological level data, allowing for areas and groups to be targeted more efficiently. A different set of studies, Ecological Momentary Intervention studies and Ecological Momentary Assessment studies appear to share the ambitions expressed earlier by the Department of Health and Social Care and Public Health England in 'combining person-generated data with existing health data' to help 'predict poor health in the future and create an opportunity to prevent it with more personalised advice and services' (Newton 2019). However, selected studies provided mixed evidence about effectiveness, and did not demonstrate how passively collected data through real-world or online interactions could be incorporated into interventions. Conceptually, the interventions were aligned with some elements of behaviour change theory, and in particular the transtheoretical (stages of change) theory; and therefore shared the same limitations and are open to the same critiques about being too individualised in focus.

What can the case studies tell us about the potential benefits of PPH?

In order to theorise the possible benefits of PPH we explored using two case studies: one involving an individual level and one a community level intervention. We examined how these

interventions are used presently using a logic model approach, and theorised the potential benefits a PPH approach could bring.

Individual level: starting with health checks, PPH approaches offer potential for improving the targeting of the programme overall, increasing uptake, and providing ongoing support after health checks. With the possible exception of population targeting, the likely improvements in outcomes and/or reductions in costs would probably not be dramatic, but are worth investigating. Further work would be valuable to explore the potential of PPH approaches to increase uptake of health checks in disadvantaged groups, and to facilitate maintenance of healthy behaviours beyond health checks themselves.

Community level: similarly, there is potential for PPH approaches in the design of community-based interventions to further understanding of the features and dynamics of areas and communities. The availability of new sources of data, and the ability to crowdsource data on community characteristics, dynamics, needs and wishes may facilitate community engagement within target areas. This may be particularly relevant among populations that are already digitally well connected, as discussed in the example of school-based smoking cessation interventions (see Appendix 6). These data may also increase understanding of community dynamics through the analysis of online interactions and new statistical modelling techniques may better capture system-level relationships and dynamics. However, while it is likely that the greater availability of data will enable better understanding of communities, the techniques are more likely to represent a useful adjunct rather than to revolutionise current approaches.

Both case studies suggest that PPH approaches could make a marginal difference in improving outcomes, which may have important population level consequences, albeit based on logical extensions to current practice. In each case, issues of equity and the potential obscuring of structural factors impacting communities and individuals were identified as concerns.

What could come next?

Defining PPH is contentious and our findings reflect the difficulty in assessing and operationalising a broad ambition of using emerging data and technologies to better understand profiles, predict risk and outcomes, and act upon this evidence. Two factors may be useful to consider in future developments.

First there may be utility in avoiding short ambiguous definitions of PPH in favour of establishing a set of principles by which PPH could be operationalised. Principles could include, for example, a commitment to establishing improved precision through the use of novel sources of data on behaviours or social determinants of health; a focus on using new data and methods to reduce health inequalities; and the use of specific metrics to demonstrate improvements in predictive value through the utilisation of these new sources of data and/or methods. Adding this comparative/evaluation component would mean that the field develops and that many of the existing studies that claim to be examples of PPH would instead be regarded as preliminary feasibility studies. Regardless, these principles should reflect that PPH is not a single object or entity, but a collection of actions that can be undertaken to harness novel sources of data and

analytical methods in order to act upon the social determinants of health and reduce health inequalities, among other goals.

Second it might be appropriate to establish parameters around the definition of PPH and to align it more specifically with a particular type of data. This type of thinking may mean that PPH is regarded as a particular type of intervention, and not a different approach of undertaking all interventions.

The bulk of the work presented here took place between March and October 2019. There is scope for further analysis to understand the potential of PPH in the future, as the number of studies adopting a PPH approach grows. The term 'Precision Public Health' only emerged within the past decade and the self-defined PPH empirical studies were published only very recently, emphasising that this is a rapidly expanding area of interest. A larger pool of studies in the future may also lend itself to more systematic approaches to reviewing the evidence, particularly if there is an interest in evaluating a particular component or principle of PPH. Adjunct methods to further contextualise the arguments and evidence, such as conducting key informant interviews, may also be useful in exploring PPH in the future.

Future work in this area may also benefit from greater patient public involvement (PPI) in shaping the specific types of questions and concerns that should be accounted for. For example, greater PPI in developing review questions and sub-questions may provide a more extensive framework for reviewers to consider whether public concerns around equity, data and ethics were addressed in PPH studies. These concerns could reflect issues such as privacy, the commodification and ownership of personal data, implications of use and misuse of data needed to undertake PPH, the potential intrusiveness of personalised public health and medicine, and broader ethical issues of trust, including stigmatisation, coercion, and data ownership. Furthermore, although the research itself was conducted in 2019, the critiques around equity and representation of ethnic minority groups in PPH approaches are particularly salient at the time of publication of this report (summer 2020), which coincides with the COVID-19 pandemic and the Black Lives Matter movement, both of which have cast a spotlight on the health implications of unchecked systemic discrimination. Issues of equity and representation are only touched upon briefly in many of the existing studies and commentaries, although these issues should occupy a more central position in appraising the utility of PPH in future. Any PPI involvement supporting future evaluations of PPH should also ensure adequate representation of a diversity of perspectives. It is critical that, if PPH is deemed to be a promising approach overall, that it does not serve to perpetuate systemic biases that disadvantage the health and wellbeing of minoritised people.

Some of the aspects of PPH that we expected to find, but appeared largely absent included: the (i) use of new methods to re-define what we mean by community and the identification of new profiles of networks and communities, (ii) the greater utilisation of data that can help to understand real-time behaviours (e.g. the use of sales data to target areas with high levels of negative health behaviours); (iii) the combination of different data types to simulate systems of influence that generate health outcomes; and (iv) greater focus on allied terminology, for example infodemiology, within PPH studies. As discussed, the evidence examined in this report predates the COVID-19 global pandemic, and many of the measures taken to mitigate the

spread of the pandemic may provide a further source of evidence and data to understand the potential role of PPH in public health decision-making. For example, a number of the tracing applications that are in development may fit within a definition of PPH and may provide evidence of the potential utility and drawbacks of PPH approaches within a defined use scenario. Emerging evidence on the efficacy and implementation of such approaches suggests that real-world applications of PPH do face challenges (for example Calvo, Deterding et al. 2020).

Overall, the results here suggest that many opportunities for examining PPH have not been considered or fully explored. For example, there may be further scope to align ecological momentary interventions with the goals of health decision-makers, particularly where these interventions are able to also incorporate assessments of environmental and structural factors. Our case studies in particular suggest that PPH approaches as they currently stand could represent an important adjunct element to current public health practice. However, where opportunities have been explored, our results suggest that a more measured, and potentially systematic approach, may be needed to fully understand and assess the opportunities that PPH can bring. This would be facilitated through introducing more focus around the concept of PPH, being clearer about the goal of PPH, and breaking down what is currently an expansive definition into a series of components that can each be evaluated.

References

Acheson, D. (1998). Independent Inquiry into Inequalities in Health Report. London, The Stationery Office.

Adie, E. and W. Roe (2013). "Altmetric: enriching scholarly content with article-level discussion and metrics." Learned Publishing **26**(1): 11-17.

Ambalavan, A. K., B. Moulahi, J. Azé and S. Bringay (2019). "Unveiling online suicide behavior: what can we learn about mental health from suicide survivors of Reddit?" Studies in Health Technology and Informatics **264**: 50-54.

Anderson, C. (2008). "The end of theory: The data deluge makes the scientific method obsolete." Wired **16**(7).

Arnett, D. K. and S. A. Claas (2016). "Precision medicine, genomics, and public health." Diabetes Care **39**(11): 1870-1873.

Arora, T., R. Chandna, S. Conant, B. Sadler and R. Slater (2020). "Demand Forecasting In Wholesale Alcohol Distribution: An Ensemble Approach." SMU Data Science Review **3**(1): 7.

Artac, M., A. R. H. Dalton, A. Majeed, J. Car and C. Millett (2013). "Effectiveness of a national cardiovascular disease risk assessment program (NHS Health Check): Results after one year." Preventive Medicine **57**(2): 129-134.

Attwood, S., K. Morton and S. Sutton (2016). "Exploring equity in uptake of the NHS Health Check and a nested physical activity intervention trial." Journal of Public Health **38**(3): 560-568.

Australian Government Department of Health. (2014). "Health assessment for people aged 45 to 49 years who are at risk of developing chronic disease: Fact Sheet." Retrieved 4 July 2019, from https://www.health.gov.au/internet/main/publishing.nsf/Content/mbsprimarycare_mbsitem701_703_705_707.

Baker, C., E. A. Loughren, D. Crone and N. Kallfa (2015). "Perceptions of health professionals involved in a NHS Health Check care pathway." Practice Nursing **26**(12): 608-612.

Balaam, M., R. Comber, E. Jenkins, S. Sutton and A. Garbett (2015). FeedFinder: a location-mapping mobile application for breastfeeding women. Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, ACM.

Baynam, G., A. Bauskis, N. Pachter, L. Schofield, H. Verhoef, R. L. Palmer, S. Kung, P. Helmholz, M. Ridout, C. E. Walker, A. Hawkins, J. Goldblatt, T. S. Weeramanthri, H. J. S. Dawkins and C. M. Molster (2017). "3-Dimensional Facial Analysis-Facing Precision Public Health." Front Public Health **5**: 31.

Baynam, G., F. Bowman, K. Lister, C. E. Walker, N. Pachter, J. Goldblatt, K. M. Boycott, W. A. Gahl, K. Kosaki, T. Adachi, K. Ishii, T. Mahede, F. McKenzie, S. Townshend, J. Slee, C. Kiraly-Borri, A. Vasudevan, A. Hawkins, S. Broley, L. Schofield, H. Verhoef, T. Groza, A. Zankl, P. N. Robinson, M. Haendel, M. Brudno, J. S. Mattick, M. E. Dinger, T. Roscioli, M. J. Cowley, A. Olry, M. Hanauer, F. S. Alkuraya, D. Taruscio, P. de la, P. M, H. Lochmuller, K. Bushby, R.

Thompson, V. Hedley, P. Lasko, K. Mina, J. Beilby, C. Tifft, M. Davis, N. G. Laing, D. Julkowska, C. Le, S. F. Terry, P. Kaufmann, I. Eerola, I. Norstedt, A. Rath, M. Suematsu, S. C. Groft, C. P. Austin, R. Draghia-Akli, T. S. Weeramanthri, C. Molster and H. J. S. Dawkins (2017). Improved Diagnosis and Care for Rare Diseases through Implementation of Precision Public Health Framework. Rare Diseases Epidemiology: Update and Overview, 2nd Edition. M. P. DeLaPaz, D. Taruscio and S. C. Groft. Cham, Springer International Publishing Ag. **1031:** 55-94.

Baynam, G., F. Bowman, K. Lister, C. E. Walker, N. Pachter, J. Goldblatt, K. M. Boycott, W. A. Gahl, K. Kosaki, T. Adachi, K. Ishii, T. Mahede, F. McKenzie, S. Townshend, J. Slee, C. Kiraly-Borri, A. Vasudevan, A. Hawkins, S. Broley, L. Schofield, H. Verhoef, T. Groza, A. Zankl, P. N. Robinson, M. Haendel, M. Brudno, J. S. Mattick, M. E. Dinger, T. Roscioli, M. J. Cowley, A. Olry, M. Hanauer, F. S. Alkuraya, D. Taruscio, M. Posada de la Paz, H. Lochmuller, K. Bushby, R. Thompson, V. Hedley, P. Lasko, K. Mina, J. Beilby, C. Tifft, M. Davis, N. G. Laing, D. Julkowska, Y. Le Cam, S. F. Terry, P. Kaufmann, I. Eerola, I. Norstedt, A. Rath, M. Suematsu, S. C. Groft, C. P. Austin, R. Draghia-Akli, T. S. Weeramanthri, C. Molster and H. J. S. Dawkins (2017). "Improved Diagnosis and Care for Rare Diseases through Implementation of Precision Public Health Framework." Adv Exp Med Biol **1031**: 55-94.

Bellemo, V., Z. W. Lim, G. Lim, Q. D. Nguyen, Y. Xie, M. Y. Yip, H. Hamzah, J. Ho, X. Q. Lee and W. Hsu (2019). "Artificial intelligence using deep learning to screen for referable and vision-threatening diabetic retinopathy in Africa: a clinical validation study." The Lancet Digital Health **1**(1): e35-e44.

Bonell, C., F. Jamal, G. J. Melendez-Torres and S. Cummins (2015). "'Dark logic': theorising the harmful consequences of public health interventions." Journal of Epidemiology & Community Health **69**(1): 95-98.

Bonell, C., F. Jamal, G. J. Melendez-Torres and S. Cummins (2015). "'Dark logic': theorising the harmful consequences of public health interventions." Journal of Epidemiology and Community Health **69**(1): 95-98.

Brunton, G., J. Caird, D. Kneale, J. Thomas and M. Richardson (2015). Review 2: Community engagement for health via coalitions, collaborations and partnerships: a systematic review and meta-analysis. London., UCL Institute of Education.

Brunton, G., J. Caird, G. Stokes, C. Stansfield, D. Kneale, M. Richardson and J. Thomas (2015). Review 1: Community engagement for health via coalitions, collaborations and partnerships: a systematic review. London, EPPI-Centre, UCL Institute of Education.

Businelle, M. S., P. Ma, D. E. Kendzor, S. G. Frank, D. J. Vidrine and D. W. Wetter (2016). "An ecological momentary intervention for smoking cessation: evaluation of feasibility and effectiveness." Journal of Medical Internet Research **18**(12): e321.

Cahill, T. J., B. Wertz, Q. Zhong, A. Parlato, J. Donegan, R. Forman, S. Manot, T. Wu, Y. Xu and J. Cummings (2019). "The Search for Consumers of Web-Based Raw DNA Interpretation Services: Using Social Media to Target Hard-to-Reach Populations." Journal of Medical Internet Research **21**(7): e12980.

Cakar, T., L. Harrison-Barry, M. Pukallus, S. Kazoullis and W. Seow (2018). "Caries experience of children in primary schools with long-term tooth brushing programs: A pilot Australian study." International Journal of Dental Hygiene **16**(2): 233-240.

Calvo, R. A., S. Deterding and R. M. Ryan (2020). Health surveillance during covid-19 pandemic, British Medical Journal Publishing Group.

Campion, E. W. and J. A. Jarcho (2019). "Watched by Apple." N Engl J Med.

Capewell, S., M. McCartney and W. Holland (2015). "Invited debate: NHS Health Checks—a naked emperor?" Journal of Public Health **37**(2): 187-192.

Caulfield, T., J. Evans, A. McGuire, C. McCabe, T. Bubela, R. Cook-Deegan, J. Fishman, S. Hogarth, F. A. Miller, V. Ravitsky, B. Biesecker, P. Borry, M. K. Cho, J. C. Carroll, H. Etchegary, Y. Joly, K. Kato, S. S.-J. Lee, K. Rothenberg, P. Sankar, M. J. Szego, P. Ossorio, D. Pullman, F. Rousseau, W. J. Ungar and B. Wilson (2013). "Reflections on the Cost of "Low-Cost" Whole Genome Sequencing: Framing the Health Policy Debate." PLOS Biology **11**(11): e1001699.

Chang, K. C.-M., J. T. Lee, E. P. Vamos, M. Soljak, D. Johnston, K. Khunti, A. Majeed and C. Millett (2016). "Impact of the National Health Service Health Check on cardiovascular disease risk: a difference-in-differences matching analysis." Canadian Medical Association Journal **188**(10): E228-E238.

Chang, W. L. and N. Grady (2015). NIST Big Data Interoperability Framework: Volume 1, Big Data Definitions.

Chowkwanyun, M., R. Bayer and S. Galea (2018). ""Precision" Public Health - Between Novelty and Hype." N Engl J Med **379**(15): 1398-1400.

Chowkwanyun, M., R. Bayer and S. Galea (2019). "Precision public health: pitfalls and promises." The Lancet **393**(10183): 1801.

Christodoulou, E., J. Ma, G. S. Collins, E. W. Steyerberg, J. Y. Verbakel and B. Van Calster (2019). "A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models." Journal of Clinical Epidemiology **110**: 12-22.

Chu, K.-H., J. Colditz, M. Malik, T. Yates and B. Primack (2019). "Identifying Key Target Audiences for Public Health Campaigns: Leveraging Machine Learning in the Case of Hookah Tobacco Smoking." Journal of Medical Internet Research **21**(7): e12443.

Cobb, N. K., A. L. Graham, M. J. Byron and D. B. Abrams (2011). "Online social networks and smoking cessation: a scientific research agenda." Journal of Medical Internet Research **13**(4): e119.

Cohn, A. M., M. S. Amato, K. Zhao, X. Wang, S. Cha, J. L. Pearson, G. D. Papandonatos and A. L. Graham (2019). "Discussions of alcohol use in an online social network for smoking cessation: analysis of topics, sentiment, and social network centrality." Alcoholism: Clinical and Experimental Research **43**(1): 108-114.

Cummings, M. J., R. Tokarz, B. Bakamutumaho, J. Kayiwa, T. Byaruhanga, N. Owor, B. Namagambo, A. Wolf, B. Mathema, J. J. Lutwama, N. W. Schluger, W. I. Lipkin and M. R. O'Donnell (2019). "Precision Surveillance for Viral Respiratory Pathogens: Virome Capture Sequencing for the Detection and Genomic Characterization of Severe Acute Respiratory Infection in Uganda." Clinical infectious diseases : an official publication of the Infectious Diseases Society of America **68**(7): 1118-1125.

Cunningham, J. A., A. Godinho and N. Bertholet (2019). "Outcomes of two randomized controlled trials, employing participants recruited through Mechanical Turk, of Internet interventions targeting unhealthy alcohol use." BMC Medical Research Methodology **19**(1): 124.

Dahlgren, G. and M. Whitehead (1991). Policies and strategies to promote social equity in health. Stockholm Institute for Future Studies.

Dankwa-Mullan, I., J. Bull and F. Sy (2015). "Precision medicine and health disparities: advancing the science of individualizing patient care." Am J Public Health **105 Suppl 3**: S368.

Davey, G. and K. Deribe (2017). "Precision public health: mapping child mortality in Africa." The Lancet **390**(10108): 2126-2128.

Department of Health and Social Care (2019). Advancing our health: Prevention in the 2020s. London, HMSO.

DHSC (2019). Advancing our health: prevention in the 2020s – consultation document. London, Department of Health and Social Care.

DiClemente, C. C. and J. O. Prochaska (1998). Toward a comprehensive, transtheoretical model of change: Stages of change and addictive behaviors. Applied Clinical Psychology: Treating Addictive Behaviors. W. R. Miller and N. Heather. New York, Plenum Press.

Dolley, S. (2018). "Big data's role in precision public health." Frontiers in Public Health **6**: 68.

Dowell, S. F., D. Blazes and S. J. N. N. Desmond-Hellmann (2016). "Four steps to precision public health." Frontiers in Public Health **540**(7632): 189.

Dryden, R., B. Williams, C. McCowan and M. Themessl-Huber (2012). "What do we know about who does and does not attend general health checks? Findings from a narrative scoping review." BMC Public Health **12**(1): 723.

Dunn, A. G., K. D. Mandl and E. Coiera (2018). "Social media interventions for precision public health: promises and risks." NPJ Digital Medicine **1**(1): 1-4.

Egan, M., A. Kearns, S. V. Katikireddi, A. Curl, K. Lawson and C. Tannahill (2016). "Proportionate universalism in practice? A quasi-experimental study (GoWell) of a UK neighbourhood renewal programme's impact on health inequalities." Social Science & Medicine **152**: 41-49.

Evers, K. E., A. L. Paiva, J. L. Johnson, C. O. Cummins, J. O. Prochaska, J. M. Prochaska, J. Padula and N. S. Gökbayrak (2012). "Results of a transtheoretical model-based alcohol, tobacco and other drug intervention in middle schools." Addictive Behaviors **37**(9): 1009-1018.

Eysenbach, G. (2009). "Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet." Journal of Medical Internet Research **11**(1): e11.

Ferrao, J. L., J. M. Mendes, M. Painho and S. Z. Joao (2016). "Spatio-temporal variation and socio-demographic characters of malaria in Chimoio municipality, Mozambique." Malaria Journal **15**: 329.

Ferrao, J. L., J. M. Mendes and M. Painho (2017a). "Modelling the influence of climate on malaria occurrence in Chimoio Municipality, Mozambique." Parasites & Vectors **10**(1): 260.

Ferrao, J. L., J. M. Mendes, M. Painho and S. Zacarias (2017b). "Malaria mortality characterization and the relationship between malaria mortality and climate in Chimoio, Mozambique." Malaria Journal **16**(1): 212.

Ferrao, J. L., S. Niquisse, J. M. Mendes and M. Painho (2018). "Mapping and Modelling Malaria Risk Areas Using Climate, Socio-Demographic and Clinical Variables in Chimoio, Mozambique." International Journal of Environmental Research and Public Health **15**(4).

Finkelstein, E. A., M. Bilger and D. Baid (2019). "Effectiveness and cost-effectiveness of incentives as a tool for prevention of non-communicable diseases: A systematic review." Social Science & Medicine **232**: 340-350.

Flay, B. R. (2009). "The promise of long-term effectiveness of school-based smoking prevention programs: a critical review of reviews." Tobacco Induced Diseases **5**(1): 7.

Franchini, M., S. Pieroni, C. Passino, M. Emdin and S. Molinaro (2018). "The CARPEDIEM Algorithm: A Rule-Based System for Identifying Heart Failure Phenotype with a Precision Public Health Approach." Frontiers in Public Health **6**: 6.

Freeman, V. L., E. E. Boylan, O. Pugach, S. L. Mclafferty, K. Y. Tossas-Milligan, K. S. Watson and R. A. Winn (2017). "A geographic information system-based method for estimating cancer rates in non-census defined geographical areas." Cancer Causes & Control **28**(10): 1095-1104.

French, D. P., E. Cameron, J. S. Benton, C. Deaton and M. Harvie (2017). "Can Communicating Personalised Disease Risk Promote Healthy Behaviour Change? A Systematic Review of Systematic Reviews." Annals of Behavioral Medicine **51**(5): 718-729.

Garcia-Closas, M., N. Rothman, J. D. Figueroa, L. Prokunina-Olsson, S. S. Han, D. Baris, E. J. Jacobs, N. Malats, I. De Vivo and D. Albanes (2013). "Common genetic polymorphisms modify the effect of smoking on absolute risk of bladder cancer." Cancer Research **73**(7): 2211-2220.

Garrison, M. M., D. A. Christakis, B. E. Ebel, S. E. Wiehe and F. P. Rivara (2003). "Smoking cessation interventions for adolescents: a systematic review." American Journal of Preventive Medicine **25**(4): 363-367.

Giles, E. L., S. Robalino, E. McColl, F. F. Sniehotta and J. Adams (2014). "The effectiveness of financial incentives for health behaviour change: Systematic review and meta-analysis." PLOS ONE **9**(3): e90347.

Goin, D. E., K. E. Rudolph and J. Ahern (2018). "Predictors of firearm violence in urban communities: a machine-learning approach." Health & Place **51**: 61-67.

Golding, N., R. Burstein, J. Longbottom, J. Browne Annie, N. Fullman, A. Osgood-Zimmerman, L. Earl, S. Bhatt, E. Cameron and C. Casey Daniel (2017). "Mapping under-5 and neonatal mortality in Africa, 2000–15: a baseline analysis for the Sustainable Development Goals." The Lancet **390**(10108): 2171-2182.

Goldstein, B. A., A. M. Navar, M. J. Pencina and J. P. A. Ioannidis (2016). "Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review." Journal of the American Medical Informatics Association **24**(1): 198-208.

Goodyear-Smith, F. (2013). "Government's plans for universal health checks for people aged 40-75." BMJ **347**: f4788.

Grant, M. J. and A. Booth (2009). "A typology of reviews: an analysis of 14 review types and associated methodologies." Health Information Libraries Journal **26**(2): 91-108.

Griffiths, M. and M. Whitty (2010). "Online behavioural tracking in Internet gambling research: Ethical and methodological issues." International Journal of Internet Research Ethics **3**(1): 104-117.

Hampel, B., K. Kusejko, D. Braun, J. Harrison-Quintana, R. Kouyos and J. Fehr (2017). "Assessing the need for a pre-exposure prophylaxis programme using the social media app Grindr®." HIV Medicine **18**(10): 772-776.

Hargittai, E. (2015). "Is bigger always better? Potential biases of big data derived from social network sites." Annals of the American Academy of Political and Social Science **659**(1): 63-76.

Hébert, E. T., E. M. Stevens, S. G. Frank, D. E. Kendzor, D. W. Wetter, M. J. Zvolensky, J. D. Buckner and M. S. Businelle (2018). "An ecological momentary intervention for smoking cessation: the associations of just-in-time, tailored messages with lapse risk factors." Addictive Behaviors **78**: 30-35.

Heldman, A. B., J. Schindelar and J. B. Weaver (2013). "Social media engagement and public health communication: implications for public health organizations being truly “social”." Public Health Reviews **35**(1): 13.

Hodder, R. K., M. Freund, L. Wolfenden, J. Bowman, S. Nepal, J. Dray, M. Kingsland, S. L. Yoong and J. Wiggers (2017). "Systematic review of universal school-based ‘resilience’interventions targeting adolescent tobacco, alcohol or illicit substance use: a meta-analysis." Preventive Medicine **100**: 248-268.

Hoffimann, E., D. Campelo, P. Hooper, H. Barros and A. I. Ribeiro (2018). "Development of a smartphone app to evaluate the quality of public open space for physical activity. An instrument for health researchers and urban planners." Landscape and Urban Planning **177**: 191-195.

Hollands, G. J., D. P. French, S. J. Griffin, A. T. Prevost, S. Sutton, S. King and T. M. Marteau (2016). "The impact of communicating genetic risks of disease on risk-reducing health behaviour: systematic review with meta-analysis." BMJ **352**: i1102.

Holmberg, C., E. Chaplin John, T. Hillman and C. Berg (2016). "Adolescents' presentation of food in social media: An explorative study." Appetite **99**: 121-129.

Horton, R. (2018). "Offline: In defence of precision public health." The Lancet **392**(10157): 1504.

Huang, G. C., J. B. Unger, D. Soto, K. Fujimoto, M. A. Pentz, M. Jordan-Marsh and T. W. Valente (2014). "Peer influences: the impact of online and offline friendship networks on adolescent smoking and alcohol use." Journal of Adolescent Health **54**(5): 508-514.

Hwang, K. O., A. J. Ottenbacher, A. L. Graham, E. J. Thomas, R. L. Street Jr and S. W. Vernon (2013). "Online narratives and peer support for colorectal cancer screening: a pilot randomized trial." American Journal of Preventive Medicine **45**(1): 98-107.

Ioannidis, J. P. A. (2013). "Informed consent, big data, and the oxymoron of research that is not research." The American Journal of Bioethics **13**(4): 40-42.

Ioannidis, J. P. A. and M. J. Khoury (2018). "Evidence-based medicine and big genomic data." Hum Mol Genet **27**(R1): R2-r7.

Jiang, F., Y. Jiang, H. Zhi, Y. Dong, H. Li, S. Ma, Y. Wang, Q. Dong, H. Shen and Y. Wang (2017). "Artificial intelligence in healthcare: past, present and future." Stroke and Vascular Neurology **2**(4): 230-243.

Jochelson, K. (2007). Paying the patient: Improving health using financial incentives. London, King's Fund.

Jones-Webb, R., P. McKee, S. Joshi, D. Erickson, T. Toomey and T. Nelson (2018). "Is restricting sales of malt liquor beverages effective in reducing crime in urban areas?" Journal of Studies on Alcohol and Drugs **79**(6): 826-834.

Katikireddi, S. V. and S. A. Valles (2015). "Coupled ethical-epistemic analysis of public health research and practice: Categorizing variables to improve population health and equity." American Journal of Public Health **105**(1): e36-e42.

Khoury, M. J. (2019). "Introducing the CDC Office of Genomics and Precision Public Health: What's in a Name?" Genomics and Public Health https://blogs.cdc.gov/genomics/2019/09/30/introducing/ 2019.

Khoury, M. J., M. Engelgau, D. A. Chambers and G. A. Mensah (2018). "Beyond Public Health Genomics: Can Big Data and Predictive Analytics Deliver Precision Public Health?" Public health genomics **21**(5-6): 244-250.

Khoury, M. J. and S. Galea (2016). "Will precision medicine improve population health?" JAMA **316**(13): 1357-1358.

Khoury, M. J., M. F. Iademarco and W. T. Riley (2016). "Precision Public Health for the Era of Precision Medicine." Am J Prev Med **50**(3): 398-401.

Khoury, M. J., M. F. Iademarco and W. T. Riley (2016). "Precision public health for the era of precision medicine." American Journal of Preventive Medicine **50**(3): 398.

Khunti, K., N. Walker, N. Sattar and M. Davies (2011). "Unanswered questions over NHS health checks." BMJ **342**: c6312.

Kitchin, R. (2013). "Big data and human geography: Opportunities, challenges and risks." Dialogues in Human Geography **3**(3): 262-267.

Kneale, D., R. French, H. Spandler, I. Young, C. Purcell, Z. Boden, S. D. Brown, D. Callwood, S. Carr, A. Dymock, R. Eastham, J. Gabb, J. Henley, C. Jones, E. McDermott, N. Mkhwanazi, J. P. Ravenhill, P. Reavey, R. Scott, C. Smith, M. Smith, J. Thomas and K. Tingay (2019). "Conducting sexualities research: an outline of emergent issues and case studies from ten Wellcome-funded projects." Wellcome Open Research **4**(137): 137.

Kneale, D., M. Khatwa and J. Thomas (2016). Identifying and appraising promising sources of UK clinical, health and social care data for use by NICE. EPPI-Centre, UCL Institute of Education, University College London.

Krebs, P., J. O. Prochaska and J. S. Rossi (2010). "A meta-analysis of computer-tailored interventions for health behavior change." Preventive Medicine **51**(3): 214-221.

Krogsbøll, L. T., K. J. Jørgensen and P. C. Gøtzsche (2013). "Universal health checks should be abandoned." BMJ: British Medical Journal **347**: f5227.

Krogsbøll, L. T., K. J. Jørgensen and P. C. Gøtzsche (2019). "General health checks in adults for reducing morbidity and mortality from disease." Cochrane Database of Systematic Reviews(1).

Krogsbøll, L. T., K. J. Jørgensen, C. Grønhøj Larsen and P. C. Gøtzsche (2019). "General health checks in adults for reducing morbidity and mortality from disease." Cochrane Database of Systematic Reviews(1): CD009009.

Krska, J., R. du Plessis and H. Chellaswamy (2016). "Views of practice managers and general practitioners on implementing NHS Health Checks." Primary Health Care Research & Development **17**(2): 198-205.

Kuo, A. K., N. M. Summers, S. Vohra, R. S. Kahn and K. Bibbins-Domingo (2019). "The Promise of Precision Population Health: Reducing Health Disparities Through a Community Partnership Framework." Advances in Pediatrics **66**: 1-13.

Kypridemos, C., B. Collins, P. McHale, H. Bromley, P. Parvulescu, S. Capewell and M. O'Flaherty (2018). "Future cost-effectiveness and equity of the NHS Health Check cardiovascular disease prevention programme: Microsimulation modelling using data from Liverpool, UK." PLOS Medicine **15**(5): e1002573.

Landry, L. G., N. Ali, D. R. Williams, H. L. Rehm and V. L. Bonham (2018). "Lack of diversity in genomic databases is a barrier to translating precision medicine research into practice." Health Affairs **37**(5): 780-785.

Langford, R., C. Bonell, H. Jones, T. Pouliou, S. Murphy, E. Waters, K. Komro, L. Gibbs, D. Magnus and R. Campbell (2015). "The World Health Organization's Health Promoting Schools framework: a Cochrane systematic review and meta-analysis." BMC public health **15**(1): 130.

Lazer, D. and J. Radford (2017). "Data ex machina: introduction to big data." Annual Review of Sociology **43**: 19-39.

Lee, J. T., K. D. Lawson, Y. Wan, A. Majeed, S. Morris, M. Soljak and C. Millett (2017). "Are cardiovascular disease risk assessment and management programmes cost effective? A systematic review of the evidence." Preventive Medicine **99**: 49-57.

Less, E. L., P. McKee, T. Toomey, T. Nelson, D. Erickson, S. Xiong, R. J. E. Jones-Webb and p. planning (2015). "Matching study areas using Google Street View: A new application for an emerging technology." Eval Program Plann. 2015; **53**: 72-79.

Liu, S., B. Chen and A. Kuo (2019). "Monitoring Physical Activity Levels Using Twitter Data: Infodemiology Study." Journal of Medical Internet Research **21**(6): e12394.

Lu, F. S., M. W. Hattab, C. L. Clemente, M. Biggerstaff and M. Santillana (2019). "Improved state-level influenza nowcasting in the United States leveraging Internet-based data and network approaches." Nature Communications **10**(1): 147.

Lupton, R. and S. Thomson (2015). The coalition's record on schools: Policy, spending and outcomes 2010-2015. Social Policy in a Cold Climate Working Paper. **13:** 5-68. London, London School of Economics and Political Science.

Lustria, M. L., S. M. Noar, J. Cortese, S. K. Van Stee, R. L. Glueckauf and J. Lee (2013). "A meta-analysis of web-delivered tailored health behavior change interventions." Journal of Health Communication **18**(9): 1039-1069.

Lutkenhaus, R. O., J. Jansz and M. P. Bouman (2019). "Tailoring in the digital era: Stimulating dialogues on health topics in collaboration with social media influencers." Digital health **5**: 2055207618821521.

Lyles, C. R., M. R. Lunn, J. Obedin-Maliver and K. Bibbins-Domingo (2018). "The new era of precision population health: insights for the All of Us Research Program and beyond." Journal of Translational Medicine **16**(1): 211.

Manrai, A. K., C. J. Patel and J. P. A. Ioannidis (2018). "In the Era of Precision Medicine and Big Data, Who Is Normal?" Jama **319**(19): 1981-1982.

Mantzari, E., F. Vogt, I. Shemilt, Y. Wei, J. P. T. Higgins and T. M. Marteau (2015). "Personal financial incentives for changing habitual health-related behaviors: A systematic review and meta-analysis." Preventive Medicine **75**: 75-85.

Markham, W. A. and P. Aveyard (2003). "A new theory of health promoting schools based on human functioning, school organisation and pedagogic practice." Social Science & Medicine **56**(6): 1209-1220.

Marteau, T. M., J. G. McGowan, M. Petticrew, H. Rutter, J. Pearson-Stuttard and M. White (2018). Changing behaviour for a healthier population. Annual Report of the Chief Medical Officer, 2018. Health 2040 - Better Health Within Reach. S. C. Davies. London, Department of Health & Social Care.

Marteau, T. M., M. White, H. Rutter, M. Petticrew, O. T. Mytton, J. G. McGowan and R. W. Aldridge (2019). "Increasing healthy life expectancy equitably in England by 5 years by 2035: could it be achieved?" The Lancet **393**(10191): 2571-2573.

Martin, C. K., L. A. Gilmore, J. W. Apolzan, C. A. Myers, D. M. Thomas and L. M. Redman (2016). "Smartloss: a personalized mobile health intervention for weight management and health promotion." JMIR mHealth and uHealth **4**(1): e18.

Martin, C. K., A. C. Miller, D. M. Thomas, C. M. Champagne, H. Han and T. Church (2015). "Efficacy of SmartLossSM, a smartphone-based weight loss intervention: Results from a randomized controlled trial." Obesity **23**(5): 935-942.

McCaig, D., M. T. Elliott, C. S. Siew, L. Walasek and C. Meyer (2019). "Profiling Commenters on Mental Health–Related Online Forums: A Methodological Example Focusing on Eating Disorder–Related Commenters." JMIR mental health **6**(4): e12555.

McLeroy, K. R., B. L. Norton, M. C. Kegler, J. N. Burdine and C. V. Sumaya (2003). "Community-based interventions." American Journal of Public Health **93**(4): 529-533.

McPhail-Bell, K., N. Appo, A. Haymes, C. Bond, M. Brough and B. Fredericks (2017). "Deadly Choices empowering Indigenous Australians through social networking sites." Health Promotion International **33**(5): 770-780.

Meacham, M. C., S. Roh, J. S. Chang and D. E. Ramo (2019). "Frequently asked questions about dabbing concentrates in online cannabis community discussion forums." International Journal of Drug Policy **74**: 11-17.

Meagher, K. M., M. L. McGowan, R. A. Settersten, Jr., J. R. Fishman and E. T. Juengst (2017). "Precisely Where Are We Going? Charting the New Terrain of Precision Prevention." Annu Rev Genomics Hum Genet **18**: 369-387.

Merchant, R., M. Clark, T. Liu, J. G. Rosenberger, J. Romanoff, J. Bauermeister and K. Mayer (2017). "Preferences for oral fluid rapid HIV self-testing among social media-using young black, Hispanic, and white men-who-have-sex-with-men (YMSM): implications for future interventions." Public Health **145**: 7-19.

Merzel, C. and J. D'Afflitti (2003). "Reconsidering community-based health promotion: promise, performance, and potential." American Journal of Public Health **93**(4): 557-574.

Metcalf, J. and K. Crawford (2016). "Where are human subjects in Big Data research? The emerging ethics divide." Big Data & Society **3**(1): 2053951716650211.

Minicã, C. C., H. Mbarek, R. Pool, C. V. Dolan, D. I. Boomsma and J. M. Vink (2016). "Pathways to smoking behaviours: biological insights from the Tobacco and Genetics Consortium meta-analysis." Molecular Psychiatry **22**: 82.

Mirnezami, R., J. Nicholson and A. Darzi (2012). "Preparing for precision medicine." New England Journal of Medicine **366**(6): 489-491.

Moore, G. F., S. Audrey, M. Barker, L. Bond, C. Bonell, W. Hardeman, L. Moore, A. O'Cathain, T. Tinati, D. Wight and J. Baird (2015). "Process evaluation of complex interventions: Medical Research Council guidance." BMJ **350**: h1258.

Moorhead, S. A., D. E. Hazlett, L. Harrison, J. K. Carroll, A. Irwin and C. Hoving (2013). "A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication." Journal of Medical Internet Research **15**(4): e85.

Muessig, K. E., N. B. Baltierra, E. C. Pike, S. LeGrand and L. B. Hightow-Weidman (2014). "Achieving HIV risk reduction through HealthMpowerment. org, a user-driven eHealth intervention for young Black men who have sex with men and transgender women who have sex with men." Digital Culture & Education **6**(3): 164.

Myneni, S., N. K. Cobb and T. Cohen (2013). Finding meaning in social media: content-based social network analysis of QuitNet to identify new opportunities for health promotion. MedInfo.

Mytton, O. T., C. Jackson, A. Steinacher, A. Goodman, C. Langenberg, S. Griffin, N. Wareham and J. Woodcock (2018). "The current and potential health benefits of the National Health Service Health Check cardiovascular disease prevention programme in England: A microsimulation study." PLOS Medicine **15**(3): e1002517.

Nagar, S. D., A. M. Moreno, E. T. Norris, L. Rishishwar, A. B. Conley, K. L. O'Neal, S. Velez-Gomez, C. Montes-Rodriguez, W. V. Jaraba-Alvarez, I. Torres, M. A. Medina-Rivas, A. Valderrama-Aguirre, I. K. Jordan and J. E. Gallo (2019). "Population Pharmacogenomics for Precision Public Health in Colombia." Frontiers in Genetics **10**: 241.

Naslund, J. A., K. A. Aschbrenner, L. A. Marsch, G. J. McHugo and S. J. Bartels (2015). "Crowdsourcing for conducting randomized trials of internet delivered interventions in people with serious mental illness: a systematic review." Contemporary Clinical Trials **44**: 77-88.

Natarajan, P., R. Young, N. O. Stitziel, S. Padmanabhan, U. Baber, R. Mehran, S. Sartori, V. Fuster, D. Reilly, F. , A. Butterworth, D. J. Rader, I. Ford, N. Sattar and S. Kathiresan (2017). "Polygenic risk score identifies subgroup with higher burden of atherosclerosis and greater relative benefit from statin therapy in the primary prevention setting." Circulation **135**(22): 2091-2101.

Neff, G. (2013). "Why big data won't cure us." Big Data **1**(3): 117-123.

Nelson, H. D., M. Pappas, A. Cantor, J. Griffin, M. Daeges and L. Humphrey (2016). "Harms of Breast Cancer Screening: Systematic Review to Update the 2009 U.S. Preventive Services Task Force RecommendationHarms of Breast Cancer Screening." Annals of Internal Medicine **164**(4): 256-267.

Newnham, J. P., M. W. Kemp, S. W. White, C. A. Arrese, R. J. Hart and J. A. Keelan (2017). "Applying precision public health to prevent preterm birth." Frontiers in Public Health **5**: 66.

Newton, J. (2019). "Predictive Prevention: Harnessing the power of data and tech to improve health." https://publichealthmatters.blog.gov.uk/2019/07/24/predictive-prevention-harnessing-the-power-of-data-and-tech-to-improve-health/ 2019.

Newton, J., M. Epke and P. Bradley (2018). "Predictive prevention and the drive for precision public health." https://publichealthmatters.blog.gov.uk/2018/11/20/predictive-prevention-and-the-drive-for-precision-public-health/ 2019.

NICE (2016). Community engagement: improving health and wellbeing and reducing health inequalities (NICE guideline [NG44]). London, National Institute of Health and Care Excellence.

O'Mara-Eves, A., G. Brunton, G. McDaid, S. Oliver, J. Kavanagh, F. Jamal, T. Matosevic, A. Harden and J. Thomas (2013). "Community engagement to reduce inequalities in health: a systematic review, meta-analysis and economic analysis." Public Health Research **1**(4).

O'Mara-Eves, A., G. Brunton, S. Oliver, J. Kavanagh, F. Jamal and J. Thomas (2015). "The effectiveness of community engagement in public health interventions for disadvantaged groups: a meta-analysis." BMC Public Health **15**(1): 129.

Park, B. K. and C. Calamaro (2013). "A systematic review of social networking sites: Innovative platforms for health research targeting adolescents and young adults." Journal of Nursing Scholarship **45**(3): 256-264.

Pearce, N. (2000). "The ecological fallacy strikes back." Journal of Epidemiology & Community Health **54**(5): 326-327.

Perez, M. V., K. W. Mahaffey, H. Hedlin, J. S. Rumsfeld, A. Garcia, T. Ferris, V. Balasubramanian, A. M. Russo, A. Rajmane and L. Cheung (2019). "Large-scale assessment of a smartwatch to identify atrial fibrillation." New England Journal of Medicine **381**(20): 1909-1917.

Peterson, A. J., M. Donze, E. Allen and C. Bonell (2019). "Effects of Interventions Addressing School Environments or Educational Assets on Adolescent Sexual Health: Systematic Review and Meta-analysis." Perspectives on Sexual and Reproductive Health Matters.

Popay, J., M. Whitehead, R. Carr-Hill, C. Dibben, P. Dixon, E. Halliday, J. Nazroo, E. Peart, S. Povall and M. Stafford (2015). "The impact on health inequalities of approaches to community engagement in the New Deal for Communities regeneration initiative: a mixed-methods evaluation." Public Health Res 2015;3(12)

Prasad, R. B. and L. Groop (2019). "Precision medicine in type 2 diabetes." Journal of Internal Medicine **285**(1): 40-48.

Prokhorov, A. V., S. H. Kelder, R. Shegog, J. L. Conroy, N. Murray, R. Peters, P. M. Cinciripini, C. De Moor, K. S. Hudmon and K. H. Ford (2010). "Project ASPIRE: an Interactive, Multimedia

Smoking Prevention and Cessation curriculum for culturally diverse high school students." Substance Use & Misuse **45**(6): 983-1006.

Prosperi, M., J. S. Min, J. Bian and F. Modave (2018). "Big data hurdles in precision medicine and precision public health." BMC Medical Informatics and Decision Making **18**(1): 139.

Public Health England (2016). Top tips for increasing the uptake of NHS Health Checks. London, Public Health England.

Public Health England (2017). NHS Health Check: Best practice guidance. London, Public Health England.

Public Health England. (2019). "NHS Health Checks: applying All Our Health." from https://www.gov.uk/government/publications/nhs-health-checks-applying-all-our-health/nhs-health-checks-applying-all-our-health.

Rajalakshmi, R., R. Subashini, R. M. Anjana and V. Mohan (2018). "Automated diabetic retinopathy detection in smartphone-based fundus photography using artificial intelligence." Eye **32**(6): 1138.

Ramaswami, R., R. Bayer and S. Galea (2018). "Precision Medicine from a Public Health Perspective." Annu Rev Public Health **39**: 153-168.

Reece, A., G. and C. Danforth, M. (2017). "Instagram photos reveal predictive markers of depression." EPJ Data Science **6**(1): 15.

Rhodes, S. D. (2004). "Hookups or health promotion? An exploratory study of a chat room-based HIV prevention intervention for men who have sex with men." AIDS education and prevention **16**(4): 315-327.

Riddle, M. S. (2017). "Big data meet precision public health: the modeling of acute gastroenteritis, norovirus and its present and future utility." Journal of Infectious Diseases **216,**.

Rieger, M., S. K. Trommlerová, R. Ban, K. Jeffers and M. Hutmacher (2019). "Temporal stability of child growth associations in Demographic and Health Surveys in 25 countries." SSM-population health **7**: 100352.

Riley, R., N. Coghill, A. Montgomery, G. Feder and J. Horwood (2015). "The provision of NHS health checks in a community setting: an ethnographic account." BMC Health Services Research **15**(1): 546.

Robson, J., I. Dostal, A. Sheikh, S. Eldridge, V. Madurasinghe, C. Griffiths, C. Coupland and J. Hippisley-Cox (2016). "The NHS Health Check in England: an evaluation of the first 4 years." BMJ Open **6**(1): e008840.

Rose, G. (2001). "Sick individuals and sick populations." International Journal of Epidemiology **30**(3): 427-432.

Rutter, H., N. Savona, K. Glonti, J. Bibby, S. Cummins, D. T. Finegood, F. Greaves, L. Harper, P. Hawe and L. Moore (2017). "The need for a complex systems model of evidence for public health." The Lancet **390**(10112): 2602-2604.

Rychetnik, L., M. Frommer, P. Hawe and A. Shiell (2002). "Criteria for evaluating evidence on public health interventions." Journal of Epidemiology and Community Health **56**(2): 119-127.

Rzotkiewicz, A., A. L. Pearson, B. V. Dougherty, A. Shortridge and N. Wilson (2018). "Systematic review of the use of Google Street View in health research: major themes, strengths, weaknesses and possibilities for future research." Health & Place **52**: 240-246.

Schoeppe, S., S. Alley, W. Van Lippevelde, N. A. Bray, S. L. Williams, M. J. Duncan and C. Vandelanotte (2016). "Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: a systematic review." International Journal of Behavioral Nutrition and Physical Activity **13**(1): 127.

Schootman, M., E. Nelson, K. Werner, E. Shacham, M. Elliott, K. Ratnapradipa, M. Lian and A. McVay (2016). "Emerging technologies to measure neighborhood conditions in public health: implications for interventions and next steps." International Journal of Health Geographics **15**(1): 20.

Schreuders, M., P. A. Nuyts, B. van den Putte and A. E. Kunst (2017). "Understanding the impact of school tobacco policies on adolescent smoking behaviour: A realist review." Social Science & Medicine **183**: 19-27.

Scott, J. and G. Marshall (2009). A Dictionary of Sociology (3 rev. ed.) Oxford, Oxford University Press.

Seibert, T. M., C. C. Fan, Y. Wang, V. Zuber, R. Karunamuni, J. K. Parsons, R. A. Eeles, D. F. Easton, Z. Kote-Jarai, A. A. Al Olama, S. B. Garcia, K. Muir, H. Grönberg, F. Wiklund, M. Aly, J. Schleutker, C. Sipeky, T. L. J. Tammela, B. G. Nordestgaard, S. F. Nielsen, M. Weischer, R. Bisbjerg, M. A. Røder, P. Iversen, T. J. Key, R. C. Travis, D. E. Neal, J. L. Donovan, F. C. Hamdy, P. Pharoah, N. Pashayan, K.-T. Khaw, C. Maier, W. Vogel, M. Luedeke, K. Herkommer, A. S. Kibel, C. Cybulski, D. Wokolorczyk, W. Kluzniak, L. Cannon-Albright, H. Brenner, K. Cuk, K.-U. Saum, J. Y. Park, T. A. Sellers, C. Slavov, R. Kaneva, V. Mitev, J. Batra, J. A. Clements, A. Spurdle, M. R. Teixeira, P. Paulo, S. Maia, H. Pandha, A. Michael, A. Kierzek, D. S. Karow, I. G. Mills, O. A. Andreassen and A. M. Dale (2018). "Polygenic hazard score to guide screening for aggressive prostate cancer: development and validation in large scale cohorts." BMJ **360**: j5757.

Sengstock, C. and M. Gertz (2012). Latent geographic feature extraction from social media. Proceedings of the 20th International Conference on Advances in Geographic Information Systems, ACM.

Si, S., J. R. Moss, T. R. Sullivan, S. S. Newton and N. P. Stocks (2014). "Effectiveness of general practice-based health checks: a systematic review and meta-analysis." Br J Gen Pract **64**(618): e47-e53.

Si, S., J. R. Moss, T. R. Sullivan, S. S. Newton and N. P. Stocks (2014). "Effectiveness of general practice-based health checks: a systematic review and meta-analysis." British Journal of General Practice **64**(618): e47-e53.

South, J. and G. Phillips (2014). "Evaluating community engagement as part of the public health system." J Epidemiol Community Health **68**(7): 692-696.

Stokes, G., M. Richardson, G. Brunton, M. Khatwa and J. Thomas (2015). Review 3: Community engagement for health via coalitions, collaborations and partnerships (on-line social media and social networks)–a systematic review and meta-analysis. London UK, Social Science Research Unit, UCL Institute of Education, University College London.

Stol, Y. H., M. H. N. Schermer and E. C. A. Asscher (2016). "Omnipresent Health Checks May Result in Over-responsibilization." Public Health Ethics **10**(1): 35-48.

Stone, A. A. and S. Shiffman (1994). "Ecological momentary assessment (EMA) in behavorial medicine." Annals of Behavioral Medicine.

Stopczynski, A., R. Pietri, A. Pentland, D. Lazer and S. Lehmann (2014). "Privacy in sensor-driven human data collection: A guide for practitioners." arXiv **preprint arXiv:1403.5299**.

Stopczynski, A., V. Sekara, P. Sapiezynski, A. Cuttone, M. M. Madsen, J. E. Larsen and S. Lehmann (2014). "Measuring large-scale social networks with high resolution." PloS one **9**(4): e95978.

Tang, W., T. D. Ritchwood, D. Wu, J. J. Ong, C. Wei, J. Iwelunmor and J. D. Tucker (2019). "Crowdsourcing to Improve HIV and Sexual Health Outcomes: a Scoping Review." Current HIV/AIDS Reports: 1-9.

Taylor-Robinson, D. and F. Kee (2018). "Precision public health—the Emperor's new clothes." International Journal of Epidemiology **48**(1): 1-6.

Thomas, J., D. Kneale, J. E. McKenzie, S. E. Brennan and S. Bhaumik (2019). Determining the scope of the review and the questions it will address. Cochrane Handbook for Systematic Reviews of Interventions. J. Thomas and J. P. T. Higgins. Hoboken, NJ, USA, John Wiley & Sons**:** 13-31.

Thomas, R. E., J. McLellan and R. Perera (2013). "School-based programmes for preventing smoking." Evidence-Based Child Health: A Cochrane Review Journal **8**(5): 1616-2040.

Thomas, R. E., J. McLellan and R. Perera (2015). "Effectiveness of school-based smoking prevention curricula: systematic review and meta-analysis." BMJ open **5**(3): e006976.

Torkamani, A., N. E. Wineinger and E. J. Topol (2018). "The personal and clinical utility of polygenic risk scores." Nature Reviews Genetics **19**(9): 581-590.

Toulmin, S. E. (2003). The uses of argument. Cambridge, Cambridge University Press.

Upshur, R. E. and E. Colak (2003). "Argumentation and evidence." Theoretical Medicine and Bioethics **24**(4): 283-299.

Usher-Smith, J. A., E. Harte, C. MacLure, A. Martin, C. L. Saunders, C. Meads, F. M. Walter, S. J. Griffin and J. Mant (2017). "Patient experience of NHS health checks: a systematic review and qualitative synthesis." BMJ Open **7**(8): e017169.

van El, C. G., M. C. Cornel, P. Borry, R. J. Hastings, F. Fellmann, S. V. Hodgson, H. C. Howard, A. Cambon-Thomsen, B. M. Knoppers, H. Meijers-Heijboer, H. Scheffer, L. Tranebjaerg, W. Dondorp and G. M. W. R. de Wert (2013). "Whole-genome sequencing in health care." European Journal Of Human Genetics **21**: 580.

Van Ham, M., D. Manley, N. Bailey, L. Simpson and D. Maclennan (2012). Neighbourhood effects research: New perspectives. Neighbourhood effects research: New perspectives, Springer**:** 1-21.

Vandevijvere, S., R. Williams, E. Tawfiq and B. Swinburn (2017). "A food environments feedback system (FoodBack) for empowering citizens and change agents to create healthier community food places." Health Promotion International **34**(2): 277-290.

Verberne, S., A. Batenburg, R. Sanders, M. van Eenbergen, E. Das and M. S. Lambooij (2019). "Analyzing Empowerment Processes Among Cancer Patients in an Online Community: A Text Mining Approach." JMIR Cancer **5**(1): e9887.

Wandersman, A. and P. Florin (2003). "Community interventions and effective prevention." American Psychologist **58**(6-7): 441.

Weeramanthri, T. S., H. J. Dawkins, G. Baynam, M. Bellgard, O. Gudes and J. B. Semmens (2018). "Precision public health." Frontiers in Public Health **6**: 121.

Wiehe, S. E., M. M. Garrison, D. A. Christakis, B. E. Ebel and F. P. Rivara (2005). "A systematic review of school-based smoking prevention trials with long-term follow-up." Journal of Adolescent Health **36**(3): 162-169.

Wolfenden, L., N. Nathan and C. M. Williams (2015). "Computer-tailored interventions to facilitate health behavioural change." British Journal of Sports Medicine **49**(22): 1478-1479.

Wortmann, F. and K. Flüchter (2015). "Internet of things." Business & Information Systems Engineering **57**(3): 221-224.

Wright, C., P. M. Dietze, P. A. Agius, E. Kuntsche, M. Livingston, O. C. Black, R. Room, M. Hellard and M. S. J. J. m. Lim (2018). "Mobile phone-based ecological momentary intervention to reduce young adults' alcohol use in the event: a three-armed randomized controlled trial." JMIR mHealth and uHealth **6**(7): e149.

Yan, Y., J.-W. Zhang, G.-Y. Zang and J. Pu (2019). "The primary use of artificial intelligence in cardiovascular diseases: what kind of potential role does artificial intelligence play in future medicine?" Journal of Geriatric Cardiology: JGC **16**(8): 585.

Yu, K. H. and I. S. Kohane (2019). "Framing the challenges of artificial intelligence in medicine." BMJ Qual Saf **28**(3): 238-241.

Zipkin, D. A., C. A. Umscheid, N. L. Keating, E. Allen, K. Aung, R. Beyth, S. Kaatz, D. M. Mann, J. B. Sussman, D. Korenstein, C. Schardt, A. Nagi, R. Sloane and D. A. Feldstein (2014). "Evidence-based risk communication: a systematic review." Ann Intern Med **161**(4): 270-280.

Appendices

Appendix 1 – Methods used to complete the review

1. Establishing and understanding the underlying assumptions and considerations

The assumptions underpinning PPH were identified and explored by undertaking:

(i) targeted searches of the literature;
(ii) a workshop;
(iii) follow-up research and meetings to discuss initial findings.

Initial searches were undertaken on Scopus, PubMed and Web of Science using a simple search string to identify studies and commentaries that focussed on PPH. From these, nine commentaries, reviews and editorials were prioritised to present a snapshot of the emerging debates and issues; these were purposefully selected on the basis of their breadth, impact and we also selected a balance of studies that were ostensibly supportive of PPH as well as those that were critical. These were supplemented with information from an additional nine studies. A pro forma for data extraction was created and extracted information formed the basis of discussions in a workshop involving all team members. The purpose of this workshop was to identify the key issues arising from the literature and to generate a shared understanding of gaps in knowledge around PPH. From this workshop, the focus honed in on (i) establishing and understanding the underpinning assumptions surrounding PPH; and (ii) describing the additional considerations that were discussed within the literature. A further workshop helped to define future directions for the review.

2. Critically Appraising the Line of Argument of Commentary studies

We applied a critical appraisal tool to consider the way in which evidence is used by advocates and detractors of PPH to substantiate lines of argument as documented in commentaries and reviews. To understand the validity of the arguments, we used an adaptation to Toulmin's Model of argumentation approach for breaking the argument down into its constituent parts (Toulmin 2003), and this process is described below.

(i)	Identify claims - We identified positive and negative claims (i.e. a statement made about PPH either in support or in opposition). Because commentaries often contain several claims, we examined those that could help us to further understand evidence the underpinning assumption of PPH (e.g. the availability of new methods and the effectiveness of tailoring interventions – these are outlined in full in the first section of the findings)
(ii)	Identify the grounds for the claim – Following from (i), we considered the grounds for the claim; i.e. the basis for the claim. We were solely interested in grounds that were based in

	evidence and considered those claims that drew on specific cited study to support their argument.
(iii)	Examine qualifiers and rebuttals – Once the claim and its grounds were identified, we considered any qualifiers and any rebuttals to the claim. Qualifiers are indicators of the strength of the leap from the data to the claim and may limit the universality of the claim. Rebuttals are made when an author anticipates potential counterarguments to their claim and outlines why potential counterarguments may not be valid. This provided a sense of whether the author endorsed the claim and under what conditions.
(iv)	Identify warrants – The link between the grounds and the claim is established through a warrant. The warrant should explain how the grounds support the claim.
(v)	Establish the backing for the warrant – Finally we identified whether the backing provided in the cited study supported the claim being made. From the outset it was anticipated that in most cases, both the warrant/s and their backing would be implicit rather than stated.
(vi)	Assess the extent to which the claim is upheld – Using the information collected above, we assessed whether the authors claims were upheld by the evidence that they cited.

Each commentary study was assessed and a pro forma completed by a reviewer. A second reviewer also examined the information extracted to ensure consistency and robustness and the review team met frequently to discuss the emerging findings. Commentaries with a specific focus on PPH were identified through a simple search on PPH via PubMed at the end of July 2019 and 20 studies were identified; of these 15 were purposefully selected as they approached PPH from a broad perspective (i.e. were focussed on both big data and genomic data sources) or superseded earlier commentaries (where lead authors published successive commentaries the latest was selected). Commentaries on dental PPH were excluded.

3. Critically Appraising Empirical PPH studies

We set out to identify and further understand the features of PPH studies through examining two sets of empirical studies:

(i) Empirical studies (n=14) that use the term ‘PPH’, identified via a simple PubMed search for ‘Precision Public Health’.

(ii) Public health-focussed empirical studies that were described as being Ecological Momentary Intervention (EMI) studies or Ecological Momentary Assessment (EMA) studies. This third set helped us to identify whether there was overlap between the goals of EMI studies and PPH studies.

This purposive approach to identifying studies was used because: (i) the interest in examining how the language of PPH is being used and interpreted in empirical literature; and (ii) a much more extensive searching and screening approach would have been needed to identify studies

which examine PPH but don't use the term, which was not commensurate with the aims and timescale of the review.

We developed a tool for assessing the features of these studies, and how the studies substantiated or refuted the underlying assumptions of PPH. We extracted key features of each study including (i) the provenance of the study (how it was identified) and whether it met the definition of a PPH study; (ii) the extent to which the underpinning assumptions of PPH are examined and whether they are refuted/upheld; (iii) if additional methodological considerations are examined and if they are refuted/upheld; (iv) if additional conceptual considerations are examined and if they are refuted/upheld; and (v) if additional ethical considerations are examined and if they are refuted/upheld. The tool was piloted and refined after feedback, and applied to three sets of studies (an outline of the tool is provided in the appendix 2).

4. Developing case studies of potential applications

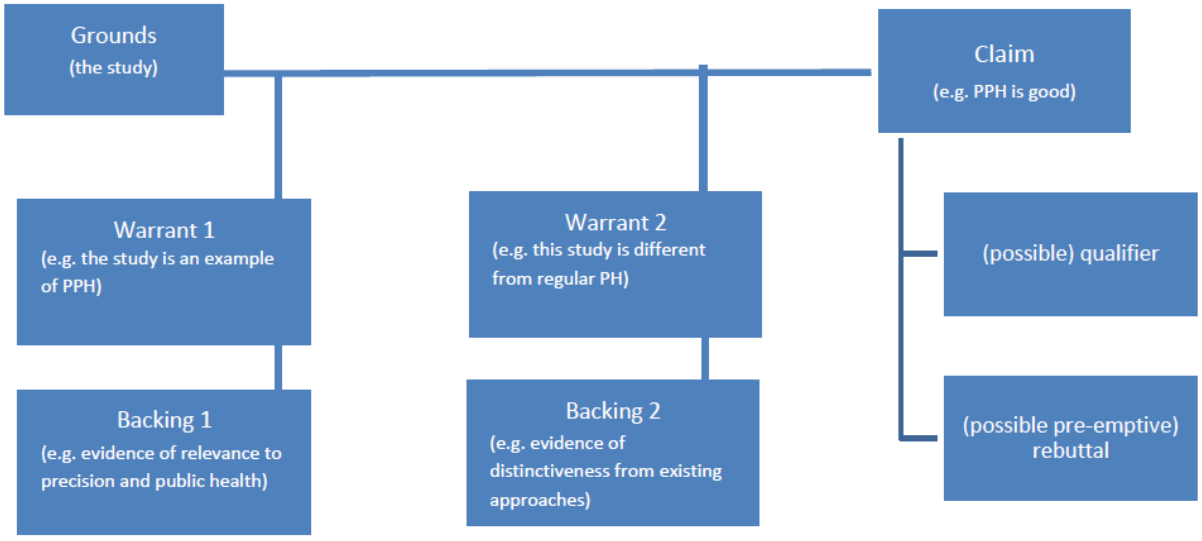
A case study approach was adopted to further understand the extent to which different intervention strategies offer specific opportunities and challenges when considering a PPH approach. In particular, this exercise intended to explore the applicability of PPH approaches to social, environmental or policy-level interventions (and further understand challenges regarding conceptual issues). The areas of interest identified were Health Checks and Community-based interventions. Here we aimed to: (i) develop an understanding of these interventions and how they are intended to 'work' and understand how this changes when adopting a PPH by developing a logic model (Thomas, Kneale et al. 2019); (ii) explore if the assumptions surrounding PP/PPH (e.g. around the availability of data and the effectiveness of tailored interventions) appear to be supported using Health Checks and Community based interventions as case studies; and (iii) identify key evaluation criteria and questions for future research. Key systematic and critical reviews on the intervention were identified and information on how the intervention was intended to work was synthesised to create a logic model; we then returned to the assumptions underpinning a PPH approach to help consider how PPH methods could change the way an intervention was designed or implemented; where possible we sought additional information to understand whether the suggested modifications were feasible and effective; finally the case studies were discussed and refined during a further all-day workshop meeting involving all team members.

Appendix 2: Further Details of adaptation to Toulmin's Model of Argument approach for use with Commentary and Review Studies

To understand the validity of the arguments, we used an adaptation to Toulmin's Model of argumentation tool for breaking the argument down into its constituent parts (Toulmin 2003). We identified claims made about PPH, both in defence and denigrating aspects of PPH, that were in relation to the assumptions identified earlier. Starting from the claim, we considered any qualifiers (indicators of the strength of the leap from the data to the claim, they may limit the universality of the claim) and any rebuttals to the claim (focus here on pre-emptive counter-arguments). This helped to give us a sense of the extent to which the author endorses the claim and under what conditions. We then considered the grounds for the claim. Some points will be

grounded in emotion, ideals, or logic. We were most interested in grounds for which the author provides some “data” or “fact”; in this case, the fact is a specific study that is cited as a supporting example. The link between the grounds (the study) and the claim (for example that PPH is worth pursuing) is established through warrants. We examined whether the backing provided was substantiated. A pictorial representation is displayed below.

Figure 9: Pictorial representation of line of argument



·What claim is being made? (describe)	
·What is the claim being made in author's words (if applicable)	
·Which assumption does the claim align with	·A1: New sources of data exist and are accessible
	·A2: New methods and data allow for better understanding of health and social profile of people and areas
	·A3: We can use improved profiles to tailor interventions
	·A4: We can use improved profiles to target people
	·A5: Better targeting/tailoring leads to improved outcomes
·Are there any qualifiers to this claim?	·“Promising”, “potential”, or "may"
	·For certain health issues
	·For certain populations
	·Other (specify)

	·No qualifiers
·Have the authors pre-empted any rebuttals to this claim?	
·What are the grounds for the claim?	·Specific empirical evidence cited
	·Indirect empirical evidence cited
	·Reasoning/ logic/ theory
	·Idealistic/emotive appeal
	·Other (specify)
	·Unclear?
·Warrant 1: specify	
·Does the citing document explicitly state this warrant?	·Yes – explicitly says
	·No – implicit
	·No, not at all
·Does the citing document explicitly provide appropriate backing for this warrant? (i.e., evidence of relevance of study to precision and public health)	·Yes – explicit and appropriate (the study is PPH)
	·No – explicit but inappropriate (the study is not PPH)
	·No – implicit but appropriate (the study is PPH)
	·No – implicit but inappropriate (the study is not PPH)
	·No, not at all

Appendix 3: Table of results from Line of Argument Analysis

Broad claim category	Claim in author's words	Claim made in	Grounds	Warrant (link between the grounds and the claim)	Backing	Limitations of the claim considered?	Is the claim substantiated by the evidence presented?
ASSUMPTION A: NEW DATA SETS CAN PROVIDE USABLE DATA FOR PUBLIC HEALTH INTERVENTIONS							
PPH allows targeting/tailoring interventions	Big data are used for targeting treatment interventions.	Dolley (2018)	Cited studies for various diseases and settings	**Inferred:** There are several examples of application of big data for targeting intervention.	**Explicit and partly appropriate:** A number of studies are cited, some of which are observational and do not involve interventions. In addition, some are questionable as to whether they fall within the remit of PPH using neither 'big data' (e.g., instead using traditional survey data) or new analytical methods.	Yes	**No.** The arguments made about targeting of interventions are not substantiated and the backing involves traditional data sources and analytical methods. The author pinpointed several potential gaps of using big data for targeting interventions such as the lack of showing treatment effectiveness via electronic health records, heightened privacy risks since subjects can be uniquely identified, and lack of plan on how to implement a treatment to a high-risk area.
PPH allows targeting/tailoring interventions	Big data are used for predicting risk that can lead to implement preventive interventions.	Dolley (2018)	Cited studies for various diseases	**Inferred:** There are several examples of application of big data for risk prediction.	**Explicit and partly appropriate:** The studies cited are on risk prediction but some do not use big data. For example, one cited study collected data on 1303 participants to study the risk of gestational diabetes amongst obese women (White et al. 2016).	Yes	**No.** The evidence presented mainly focus on risk prediction but is not necessarily on big data nor on the implementation of preventive interventions. The author pinpointed several potential gaps of using big data for predicting risks such as lack of clinical data, missing novel determinants, lack of data for some populations (such as children and healthy subjects), and lack of geographical precision.
PPH is an improvement on traditional public health	"Changes in the way people live and communicate have made it possible to access new forms of data."	Dunn (2018)	Several papers cited	**Explicit:** Population-level studies have demonstrated the capacity to model spatial variations in several fields. Individual-level studies have demonstrated the capacity to predict attitudes, behaviours, and health outcomes.	**Explicit and appropriate:** A number of the cited studies provide evidence on the predictive power of social media in detecting health outcomes, attitudes, and behaviours.	Yes	**Yes**, although the cited studies provide evidence, they do contain a number of caveats which are recognized in the claim that PPH can be a useful adjunct to PH surveillance. For example, the authors discussed the risks and unintended consequences of using of social media data such as erosion of privacy, social manipulation, driving unhealthy behaviours, and backlash.

Broad claim category	Claim in author's words	Claim made in	Grounds	Warrant (link between the grounds and the claim)	Backing	Limitations of the claim considered?	Is the claim substantiated by the evidence presented?
Big data improves precision	Increasingly, a large volume of health- and non-health related data from multiple sources is becoming available that has the potential to drive precision implementation.	Khoury (2018)	Dolley, 2018	**Explicit:** There are new sources of data and the term big data captures large datasets requiring new methods for our analysis, Collecting these data to inform evidence and ensuring uptake of this new evidence will improve health.	**Explicit and appropriate:** The claim pertains to the increasing availability of sources which are useful for PPH. The cited document is a review of different sources of big data for PPH.	Yes	**Yes**. While our knowledge of different sources is improving, the case that these data are usable isn't explicitly made. Conditional limitations are highlighted.
Big data allows a wider range of variables to be in scope	Tools, combined with social networks platforms provide a window into the behavioural and social domains of health, data-rich environments that need to be considered in the context of precision medicine and precision public health, to create a 'digital phenotype' of disease.	Prosperi (2018)	Jain 2015; Holmberg 2016; Reece 2017; Dwyer-Lindgren 2017; Lazer 2014; Butler 2013	**Inferred:** There are studies that highlight the potential contribution of patient-generated data to individual and public health decision-making.	**Yes, explicit and appropriate:** This claim is backed up by a systematic review that examined areas of need for collecting and using patient generated health data from the perspective of patients and providers; the review finds synergies and divergence. Other studies are cited on the use of social media to identify health behaviour and risk factors.	Yes	**In part.** The systematic review quotes focusses on the perspectives of patients and providers. Both sets of stakeholders can see the value, with caveats, of these data. However, the aspect of the claim that the 'clinical domain can be significantly enhanced' is untested because no consideration is given to the validity or reliability of the data. A later limitation (that seems to apply here also) is that research using non-traditional health-related data from these domains have been conducted with some success as well as with some controversy.
Lack of knowledge and potential unintended consequences of PPH approaches *Claim is a critique of PPH*	"Researchers must be wary of the 'big data hubris' or "that big data are a substitute for, rather than a supplement to, traditional data collection and analysis"".	Prosperi (2018)	Lazer 2014; Rajkomar 2018; Shickel 2017; Choi 2016; Miotto 2016	**Implicit**: Big data suffer from bias but this isn't always recognised. Because people overlook the fallacies of big data, analytical methods used to analyse big data present flawed results.	**Implicit and inappropriate:** Specific examples are provided outlining where EHR and ML has been used to gain advances in prediction. One review paper cautions against having too much faith in big data unquestioningly. However, the link between EHR data and lack of representativeness in big data is made on the basis of a logical extension.	No	**In part.** The claim is upheld and the grounds are suitable. However. some further qualifiers are needed around the warrant to make it clear that the studies focused on Electronic Health Records do not make the link with bias.
Big data improves precision	"The use of study designs that rely on more readily accessible clinical data from large	Riddle (2017)	Verstraeten 2017	**Implicit**: A case study shows that electronic health records (EHRs) can be used to estimate norovirus	**Explicit but inappropriate:** The cited study does not self-refer as an example precision public health study. The source data are Electronic health records and while	No	**No.** While the claim itself may be plausible, and the source study does offer a new granularity of evidence; there isn't a huge movement beyond

Broad claim category	Claim in author's words	Claim made in	Grounds	Warrant (link between the grounds and the claim)	Backing	Limitations of the claim considered?	Is the claim substantiated by the evidence presented?
	repositories (big data), which meet unique constraints based on a particular disease and population of interest, holds promise to allow for important public health implications beyond traditional surveillance and burden of disease estimation—precision public health."			attributable disease at the population level.	the methods are relatively complex – Poisson regression with bootstrapping – these are methods that have been available for some time. The added level of precision is monthly estimated norovirus cases.		traditional surveillance and burden of disease estimation.
ASSUMPTION B: NEW METHODS OF PREDICTING RISK WILL ENABLE PREDICTION OF INDIVIDUAL OUTCOMES							
PPH can improve risk and outcome prediction	In characterising gaps and disparities in implementation and outcomes, personal characteristics of patients, providers, and policy-makers can be further refined beyond the use of traditional indicators such as age, gender, and race/ethnicity.	Khoury (2018)	Collins 2018; Knowles 2017; Yurgelun 2018	**Explicit**: Genomic studies show that more precise estimation of risk and response to treatment can be calculated. Examples on Familial Hypercholestorolemia and Lynch syndrome show the integration of genomic data into screening programmes, leading to different treatment regimes.	**Explicit and appropriate:** The two examples do provide evidence that knowledge of genomic profiles can be useful for detecting people's outcomes (e.g. Bowel cancer in the case of Lynch syndrome).	No	**Yes**, this claim is substantiated by the evidence. The claim is interpreted rather narrowly through the warrant and draws on two well-known examples from genomic studies. Limitations not provided (or needed).
PPH can improve risk and outcome prediction	By applying precision analysis using factors such as genetic predisposition, prior history, and lifestyle variables, we may better understand which pregnancies can safely be left until after 39 weeks' gestation and which cases require earlier intervention.	Newnham (2017)	Glavind 2013; Nicholson 2016	**Inferred:** PPH approaches can help to shape interventions around avoiding non-medically indicated late preterm and early-term birth. Precision data are needed to determine where interventions should be directed. **Explicit:** PPH can help lower Pre-term Birth in areas with high levels but may be less effective in other areas.	**No, explicit but inappropriate:** The specific sources are not examples of PPH. These examples are presented with multiple caveats around their precision and potential. The summarising sentence is theorised application but not based on evidence.	Yes	**No**, the warrants provided are themselves surrounded by caveats and nuance; but the claim is not supported. Limitations are noted: "continuing success and translation into other environments where the baseline PTB rates are already lower may be more challenging".

Broad claim category	Claim in author's words	Claim made in	Grounds	Warrant (link between the grounds and the claim)	Backing	Limitations of the claim considered?	Is the claim substantiated by the evidence presented?
Providing better predictions may not change behaviour *Claim is a critique of PPH*	"In spite of the potentially higher accuracy in predicting disease diagnoses and health outcomes, many machine learning methods are usually regarded as non-transparent to the end user and labelled as black-boxes."	Prosperi (2018)	Krause 2016	**Inferred:** PPH black-box models lower levels of user confidence. **Explicit:** Although black-box models may provide a very precise calculation of the probability of a target event or outcome, they are often regarded with scepticism due to the lack of consideration for causal pathways.	**No, not at all.** Although the claim is situated in a section about the interpretability of models, there is no backing provided that machine learning methods are regarded as non-transparent.	Yes	**No,** there are very few studies that provide comparative evidence. The claim is supported by a counterargument with backing that studies show that confidence can be strengthened when they are integrated into electronic health records to support decision-making. However, the counterargument describes enhancing confidence of users, but the study used to support this perspective explored the views of data scientists and not clinicians, who would be the users of interest.
No evidence PPH prediction is more effective *Claim is a critique of PPH*	"Although there are undoubted public health benefits to harnessing population-level 'big data' to inform policy and evaluation, as epitomized by data linkage in the Nordic countries, risk stratification approaches are already well developed in clinical practice and public health on the basis of established risk factors."	Taylor-Robinson (2018)	Ioannidis 2009	**Implicit**: The warrant to this claim is that there is no evidence of an added predictive value of genetic markers.	**No, explicit but inappropriate:** The point being made about risk prediction is supported by the literature cited but the potential of PPH to improve on this is not addressed. The cited review used as backing for the warrant is almost ten years old.	Yes	**In part**. The authors do present a case that risk prediction is fallible but the warrant is not directly related to the claim and the backing is dated (although may be the best available evidence).
ASSUMPTION C: NEW ANALYSIS METHODS WILL ENABLE SUBSTANTIALLY MORE ACCURATE PREDICTION OF INDIVIDUAL RISK							
PPH allows policy making and targeting resources	"Level of precision is important for equitable policy making and efficient targeting of resources".	Davey (2017)	Golding 2017	**Explicit:** Higher level of precision can help to reduce the rate of child mortality.	**Explicit and partly appropriate**: Golding et al. (2017) do present a greater level of precision for quantifying child mortality rates, but the use in decision-making is not substantiated.	Yes	**No**. We can conclude that PPH approaches can add precision to estimates of mortality but it is unclear how/if these estimates are used in decision-making. The authors acknowledged two limitations regarding the scarcity of data in some regions due to the lack of strong registration systems as well as important potential covariates not taken into account in the study.

Broad claim category	Claim in author's words	Claim made in	Grounds	Warrant (link between the grounds and the claim)	Backing	Limitations of the claim considered?	Is the claim substantiated by the evidence presented?
PPH is an improvement on traditional public health	"Improved analysis, data visualization and machine learning have expanded our ability to use disparate data sources to decide what to do."	Dowell (2016)	Bhatt 2015	**Inferred**: Precise modelling methods can be useful to target effective interventions.	**Explicit and appropriate:** The study cited is about the use of model-based geostatistics to identify the intervention that had the most impact on reducing the incidence of malaria.	Yes	**Yes**. They presented one example of application of new methods to target effective intervention that can reduce disease incidence. However, the authors also underlined some limits of using new methods such as data that are unavailable, unreliable and of poor quality.
Big data improves precision	The data collected through digital devices give a picture of how the interventions have been implemented and the outcomes generated with much greater precision.	Khoury (2018)	Topol 2015	**Explicit:** use of personal devices such as sensors, smartphones, and other digital devices can provide measurement of variability over time for various health indicators such as nutrition, physical activity, and blood pressure.	**Implicit but appropriate:** The backing for the warrant supported by a commentary outlining the potential of digital devices. The commentary outlines that these devices have greater coverage but is more ambivalent about the quality of the data regarding precision.	No	**No**. Yes digital devices can give more continuous data flow and may have good levels of coverage in a number of high income settings, but their use for greater precision, particularly with reference to outcomes, isn't explicitly evidenced. Ability to measure indicators is not commensurate with collecting and analysing data. Limited qualifiers are provided in terms of language (e.g. big data *may* improve…).
PPH can improve risk and outcome prediction	A nuanced, multilevel approach to controlling tobacco exposure can help to prevent preterm birth (PTB) via precision public health.	Newnham (2017)	Several papers cited	**Explicit**: Preventive programs based around national public education and legislative control are clearly warranted and, when well-executed, demonstrate marked reductions in population-level tobacco use coincident with significant reductions in population PTB rates. Alongside national programmes, assessment of a number of maternal and community factors should be considered, including ethnicity and education, which constitute a PPH approach to preventing tobacco associated pre-term birth.	**No, explicit but inappropriate:** Many of the specific sources are not examples of PPH but better community level engagement. Many of the examples also provide epidemiological evidence but not evidence on how to implement within screening programmes	Yes	**No**, there is no ground or backing to the claim that a PPH approach has been developed to control tobacco exposure here. A qualifier/limitation is that it is recognised that countries employing comprehensive, multi-faceted tobacco control measures at a national level remain in the minority.
PPH can improve risk and outcome prediction	"Omic" approaches may be suitable for both primary screening, as well as precision refinement in women previously identified as being at	Newnham (2017)	Bergen 2013	**Explicit**: Studies show that it is possible to target/tailor an intervention. One example is that women who are genetically more likely to succeed in smoking cessation with nicotine replacement therapy may be offered this intervention, while those genetically	**Yes, explicit and appropriate:** This is a case of a logical extension. Women with a particular genomic profile were more/less likely to quit smoking on eight different interventions.	No	**In part.** Women with a particular genomic profile have been found to have differential quit rates on interventions. However, the case of extending this and providing an alternative and more effective intervention is not made. The only qualifiers include the language that this is a

Broad claim category	Claim in author's words	Claim made in	Grounds	Warrant (link between the grounds and the claim)	Backing	Limitations of the claim considered?	Is the claim substantiated by the evidence presented?
	risk of PTB by other screening modalities. This may be used on a larger scale than is currently employed in order to direct interventions in those who screen at increased risk of PTB.			likely to fail may avoid the potential adverse outcomes of this therapy.			potential application at large scale not an observed application.
PPH can reduce health disparities	"The application of precision medicine to the social determinants of health is emerging as a potential method to reduce health disparities."	Kuo 2019	Khoury 2018; Beck 2014 Warrant is based on Beck.	**Implicit**: The warrant given is based on a case study on asthma admission rates, researchers have been able to link asthma "hot spots" to specific public housing properties, identifying exacerbating environmental conditions, such as extensive mould and cockroach infestations.	**Yes, explicit and appropriate:** An example is provided looking at correlations between housing code violations and asthma. There are novel links between health records and housing violations; the latter could be considered a version of big data.	No	**Yes**. However, the extent to which data on housing violations in a census tract area is adding useful precision, and whether this is 'big data' is debatable. There are no limitations made around this specific claim. Note: They did not cite a reference for the study they described for the warrant. We assume it is the same paper that is used for the backing in the following paragraph.
ASSUMPTION D: TAILORED INTERVENTIONS ARE MORE LIKELY TO CHANGE BEHAVIOUR							
Big data allows for targeting of interventions	Big-data-driven public health assessment studies provide directions about how to enhance implementation in subpopulations and can drive implementation studies that tailor interventions by place.	Khoury (2018)	Steinhubl 2015; Engelgau 2019	**Explicit:** Others discuss how PPH can be applied to understand implementation with greater precision. Precision is considered at the level of place.	**Implicit and appropriate:** The backing is provided by another commentary piece although in turn this draws on empirical examples that fit within a PPH framework.	No	**Yes it is (weakly) supported**. There is potential for implementation to be more precisely understood at a geographic level. Qualifiers are provided by language with an emphasis that there are new approaches on offer but that these may not be established.
Big data allows for targeting of interventions	Big-data-driven public health assessment studies provide directions about how to enhance implementation in subpopulations and	Khoury (2018)	Banda 2019	**Explicit:** Machine learning applied to big data can help to identify subpopulations with unique health needs.	**Explicit and appropriate:** The cited study shows that the classifier is effective and has a high Positive Predictive Value.	No	**Yes**, it is supported. The backing is limited to one health condition so it may have limitations that are not raised.

Broad claim category	Claim in author's words	Claim made in	Grounds	Warrant (link between the grounds and the claim)	Backing	Limitations of the claim considered?	Is the claim substantiated by the evidence presented?
	can drive implementation studies that tailor interventions by a person.			The example of Familial Hypercholestorolamia shows that this is possible.			
Big data allows for targeting of interventions	Big-data-driven public health assessment studies provide directions about how to enhance implementation in subpopulations and can drive implementation studies that tailor interventions over time.	Khoury (2018)	Dunn 2018	**Explicit:** Apps could collect data for tailoring interventions. There are examples where smartphone apps have been used to improve outcomes among a group of patients with hypertension. The potential for smartphones to be used have been reviewed elsewhere.	**Implicit and inappropriate:** The paper by Dunn is described as a review although is based on a 'perspective' article. There is likely to be a misdirected citation – there is explicit reference to an RCT on hypertension but it's not included in the references.	No	**No**, this is not supported. There is a misdirected citation and also the warrant states that there has been a review undertaken but this does not appear to be the case.
Providing better predictions might not change behaviour *Claim is a critique of PPH*	There is little reason for optimism around the claim that providing genetic information to individuals has negligible impact on behaviour	Taylor-Robinson (2018)	Hollands 2016	**Implicit**: providing genetic information to individuals has negligible impact on health behaviours.	**Yes, explicit and appropriate:** The authors point to evidence (a systematic review) that supports this	No	**Yes**. The claim is established. Limitations are not discussed (not needed).
ASSUMPTION E: BETTER TARGETING OF INTERVENTIONS LEADS TO GREATER EFFECTIVENESS							
PPH can improve health outcomes	"Precision public health can save lives".	Dowell (2016)	Maïnassara 2015; Ferguson 2015; Kraemer 2015	**Explicit:** More precise disease surveillance can trigger more quickly effective interventions. More precise disease surveillance can illuminate causes of disease and spark opportunities for prevention.	**Explicit and appropriate:** The studies cited are on the use of regular surveillance to trigger more quickly mass vaccination campaigns, on the use of disease surveillance to target treatment (penicillin) to a needed population (women with risky pregnancies) to prevent diseases (group B streptococci infections in newborns), and on geospatial modelling to target intervention to	Yes	**No**. The warrants do not directly support the general claim (that precision public health can save lives). The cited studies did not use new data sets or new methods (they used databases and mathematical modelling). Despite the potential benefits of precision public health, the authors discussed several needs to achieve precision such as the need to improve access to demographic data, to improve surveillance by implementing infrastructure and systems for data collection and analysis, to provide more accurate data by incorporating laboratory analysis, and to better train public health professionals.

Broad claim category	Claim in author's words	Claim made in	Grounds	Warrant (link between the grounds and the claim)	Backing	Limitations of the claim considered?	Is the claim substantiated by the evidence presented?
				More precise disease surveillance allows for more efficient use of resources and allows for more people to receive interventions.	at risk areas. The backing supports the warrants but not the claim.		
PPH can reduce health disparities	The use of big data sources could allow a more in- depth analysis of disease burden and implementation gaps and disparities in health care systems and population subgroups.	Khoury (2018)	Golding 2017; Kind 2018	**Explicit**: We can draw on examples where using small-area analysis, we might be able to uncover pockets of disparities in the implementation of health interventions that are often masked in analyses performed on larger areas.	**Explicit and appropriate:** The warrant is unproblematic and we can use methods for small area analysis with increasing precision.	No	**Yes**, the claim is substantiated by the evidence and both of the examples show that greater geographic precision is perfectly possible and can be used to highlight pockets of deprivation not usually spotted otherwise.
Methodological issues need to be overcome before PPH is implemented *Claim is a critique of PPH*	It is misguided to try to create a false dichotomy between 'personalized' and 'population-based' approaches to prevention.	Taylor-Robinson (2018)	Kypridemos 2016; Beheshti 2017	**Implicit**: Omitting the population-level perspective leads to poorer quality results. There are examples highlighting that when information on different layers of influence is optimal.	**Yes, explicit and appropriate:** The backing is two specific studies highlighting these points.	Yes	**Yes**. The claim is established with qualifiers around the language and they also provide a rebuttal that while this point is not novel, it is pervasive and relevant.
ASSUMPTION H: ETHICS OF PPH APPROACHES AND METHODS							
PPH is still concerned with SDH	"…precision public health is about using the best available data to target more effectively and efficiently Interventions of all kinds to those most in need. Nothing in this definition excludes the traditional concerns of public health. On the contrary, precision public health emphasises the importance of those determinants for	Horton (2018)	Golding 2017; Wagner 2018	**Explicit:** For Golding et al. example, "As Nick Golding and his colleagues wrote, their work "provides key information for decision makers to target interventions at populations in the greatest need". For Wagner et al. example, "To those who worry that precision public health strips politics out of public health, these findings argue exactly the opposite." **Inferred:** These studies show that SDH can still be emphasised in a PPH approach.	**Explicit and appropriate:** The warrants, especially for Golding, are stretched, but the studies do use PPH to explore SDH and so the backing is acceptable.	No	**Yes**. Whilst the warrants are overstated and not quite appropriate, the main point that these studies used big data to identify/ explore health inequalities supports the broader claim that SDH is not necessarily inconsistent with a PPH approach.

Broad claim category	Claim in author's words	Claim made in	Grounds	Warrant (link between the grounds and the claim)	Backing	Limitations of the claim considered?	Is the claim substantiated by the evidence presented?
	communities that have been invisibilised."						
PPH Approaches can maximise social justice	Methodologic deficiencies such as systematic bias in prediction models and non-representative studies, along with limited or differential access in sub-populations, can contribute to widening of health disparities, especially for racial and ethnic minority populations.	Khoury (2018)	Martin 2013	**Explicit**: Recent studies have consistently shown that the accuracy of genetic risk prediction models based on genome-wide association studies is reduced among non-European populations compared to European populations.	**Explicit and appropriate:** The warrant is appropriate and there is sufficient backing – plenty of evidence in the source document that genomic studies are disproportionately overrepresented by people of European ancestry.	No	**Yes**, the claim is substantiated by the evidence with an appropriate warrant and backing.
PPH approaches can maximise social justice	"In order to maximize social justice, we need to break down silos between medicine, public health, and 'omics' science as well as institutional barriers between local/regional/national government departments, academia, healthcare systems, and industry."	Lyles (2018)	Beck 2013	**Explicit:** "This example demonstrates that precision approaches can act as both a root-cause analysis to discover trends, but also as an intervention strategy combining evidence-based medicine/public health and social policy." **Inferred:** The example shows that multi-layered data sources when used together (presumably an example of PPH) can improve outcomes; the reader has to make a big inference that certain applications can enhance social justice.	**Explicit and inappropriate:** This is the conclusion from the cited study: "In a single year, asthma admission rates varied 88-fold across neighborhood quintiles in one county; a reduction of the county-wide admission rate to that of the bottom quintile would decrease annual admissions from 862 to 34. A rate of zero was present in 15 neighborhoods, which is evidence of what may be attainable" Beck et al. 2013).	Yes	**No**. Inferred limitation is that this will only work if the 'silos' 'break down'. The backing indicates that data can be joined up for surveillance/prediction, but Lyles et al. make a leap to say that it "prevent[s] future admissions" or that it is an intervention, and any link to social justice is hypothetical. The grounds are perhaps more related to the assumption that "New methods and data allow for better understanding of health and social profile of people and areas".
PPH can be perceived as a threat to privacy *Claim is a critique of PPH*	"The collation of non-standard data, e.g. momentary ecological assessment via Twitter, Facebook, or smartphone GPS monitoring, is prone to serious privacy and security concerns."	Prosperi (2018)	Dwork 2014	**Inferred:** There are studies that show that, despite some measures put into place to preserve anonymity, there remain issues and greater availability of data introduces privacy concerns. **Explicit:** While differential privacy has facilitated data sharing, it remains challenging to safely anonymize data	**Explicit and appropriate:** The claim is backed up by evidence around differential privacy. Differential privacy can provide some interference to hackers, although there remain limitations and fallibilities in data security.	Yes	**Yes,** this claim is upheld. The authors acknowledge that some steps have been taken but there remains some risk. Counter arguments/limitations are presented around steps taken to minimise privacy concerns – backing is provided for these.

Broad claim category	Claim in author's words	Claim made in	Grounds	Warrant (link between the grounds and the claim)	Backing	Limitations of the claim considered?	Is the claim substantiated by the evidence presented?
				while preserving all their multivariate statistical properties.			

Appendix 4: Critical Appraisal Tool for use with Empirical Studies

Category	Items
A. Is this really PPH (or related)?	1. Is it public health? 2. Is the study a (self-) defined PPH study? 1. What's "precision" about it?
B. Study provenance	Add description of provenance
C. Assumptions	1. Does the study use new or novel data sources? 2. Does the study find that new data (or linkages) allow for better understanding of profiles? 3. Does the study find that new methods allow for better understanding of profiles? 4. Does the study find that improved understanding of profiles allow for better targeting? 5. Does the study find that improved understanding of profiles allow for better tailoring? 6. Does the study find that better targeting/tailoring leads to better outcomes for people/areas?
D. Challenges and Properties	1. Ethics and public trust 2. Conceptual considerations 3. Targeting and Reach 4. Data quality and coverage 5. Outcomes
E. Contribution and Added Value	1. How is the study described as an advance on existing PH techniques 2. Scope and applicability – Health 3. Scope and applicability – Geography 4. Scope and applicability – Population
F. Treatment in source data (where applicable)	Are the claims made in the source consistent with the findings of this primary study?
G. Additional reviewer reflections/notes	

Appendix 5: Supporting information on community-based interventions: background of intervention theory

Focus

Community-based interventions can be categorised into four different models dependent on the way in 'community' is viewed as the focus to bring about changes in health (McLeroy, Norton et al. 2003; South and Phillips 2014). The four models include:

- (i) Intervention models that focus on community as a setting, which tend to primarily target changes in individuals' behaviours through standardised intervention programmes; these may still be delivered by peers or lay people, but are distinguishable through the absence of communities in the design of interventions and in that the target of change remains on changing the behaviours of individuals (O'Mara-Eves, Brunton et al. 2013; South and Phillips 2014).
- (ii) Intervention models that regard community as a target attempt to change environmental and wider area-based antecedents of health, although engagement with individuals may remain limited.
- (iii) Intervention models that view communities as 'agents' view a high degree of community ownership of the intervention as being essential to success; while (iv) models of intervention that view communities as 'resources' similarly place emphasis on fostering the adaptive, supportive, and developmental capacities of communities (McLeroy, Norton et al. 2003).

Interventions that view communities as agents or resources have been of substantial interest in recent years to systematic reviewers and guideline developers (O'Mara-Eves, Brunton et al. 2013; Brunton, Caird et al. 2015, NICE 2016). They often incorporate community capacity-building as a central organising principle and are focussed developing community capital to tackle public health issues through processes such as community engagement, participation and empowerment (South and Phillips 2014). Among interventions that incorporate community engagement, the greatest public health benefits are theorised to arise from instances where the community is in full control of the intervention (O'Mara-Eves, Brunton et al. 2013). All models of community-based interventions implicitly recognise that individuals' behaviours are shaped by dynamic interactions between the individual and their social and physical environment including interpersonal, organisational, community and policy levels, and seek to change these interactions in a positive way (Merzel and D'Afflitti 2003). In doing so they draw on social-ecological models of health (Dahlgren and Whitehead 1991) and seek to promote individual and collective change.

Identification of needs

The health challenges or needs identified may be (i) directly identified by community members themselves (a felt need); or (ii) may be identified through observation, of a community's use of services, for example (expressed need); or (iii) may be identified through comparisons with other communities (comparative need), or (iv) may be understood from comparing a community's needs or service use with societal norms or standards (normative need) (O'Mara-Eves, Brunton et al. 2013).

Where needs are exclusively ascribed by those from outside the community, this increases the risk of stigmatising communities. Katikireddi and Valles (2015) use men who have sex with men (MSM) as an example of a community deemed to be at high behavioural risk of contracting HIV/AIDS, even though some individuals within this 'community' may be at very low risk. Use of MSM as a means of identifying risk inadvertently leads to sexuality being understood to be the risk factor for HIV/AIDS, and not unprotected sex between HIV negative and HIV positive partners. Such processes lead to the stigmatisation of communities, raising both ethical issues and epistemic issues around the way in which data may be collected from these communities in future. Area-based (as opposed to community-based) interventions also often tend to involve

ascribing risk/needs by those outside the community. Ascription of individual risk based on area-based profiles raises the risk of ecological fallacy occurring, in that factors associated with communities may not be associated with the lives or actions of individuals (Pearce 2000). Targeting on a community level can mean that small numbers of individuals with high risk receive an intervention, which can be viewed as disadvantageous by policymakers. However, in population health, failure to recognise that areas and communities are often determinants and modifiers of risk can be an equally pernicious (and individualistic) fallacy (Pearce 2000).

Community intervention processes

In seeking to promote individual and collective change, community-based interventions may use multiple intervention approaches targeting multiple layers of influence on individual and collective health behaviours (Merzel and D'Afflitti 2003). They may draw on a multitude of individual-level strategies, mass media strategies and other population level approaches, and policy and environmental interventions to create environments that optimise health (Merzel and D'Afflitti 2003). O'Mara-Eves, Brunton et al. (2013) recognise three models that broadly capture the multitude of ways in which community-based interventions are delivered. Classical peer or lay-delivered interventions have limited levels of community control, with community needs ascribed by others and with beneficiaries usually reflecting individuals rather than communities. A second model includes processes to engage communities in intervention design and implementation, with programme theories including explicit engagement with the community of interest.

A third model centres on the concept of empowerment and ownership (Merzel and D'Afflitti 2003), where the intervention is identified by the community itself and the community occupies a central role in designing and delivering the intervention. These latter models may view community empowerment and enhancing community capacity both as a mediator and as a desired outcome in itself. Such initiatives may focus on other determinants of health and multiple causes of disadvantage, and consequently such approaches may be more closely aligned with public health theories that incorporate information about systems and system perspectives (Rutter, Savona et al. 2017). Intervention processes that enable the development of community capacity include the co-design of an intervention theory that clearly defines target groups, objectives, interventions and intervention components; a focus on developing communication and relationships within the community and beyond; the development of skills among community members to ensure sustainability; high(er) degrees of collective decision-making; and the provision of administrative and financial support through the lifetime of the intervention (O'Mara-Eves, Brunton et al. 2013).

Appendix 6: A focus on school-based smoking prevention interventions

Intervention theory

School-based smoking prevention interventions take different forms, have been described as adhering to five main models ((Wiehe, Garrison et al. 2005; Thomas, McLellan et al. 2013; Thomas, McLellan et al. 2015) represented as 'curriculum ethos' in figure 10 below), which can

also shape other design features including the deliverer and additional components that may be present.

- Information-only interventions are designed to provide accurate facts on smoking to young people;
- Social competence (or affective information) interventions are intended to reduce children's personal susceptibility through developing skills such as self-esteem, self-control, and cognitive skills for resisting interpersonal or media influences;
- Social influence interventions teach children about social influences that support substance use and the skills to refuse tobacco and build resistance against peer pressure;
- Combined social competence and social influence approaches;
- Multimodal interventions which combine curricular approaches with wider initiatives within and beyond the school, including the involvement of those external to the school (e.g. parents) and wider environmental and policy changes.

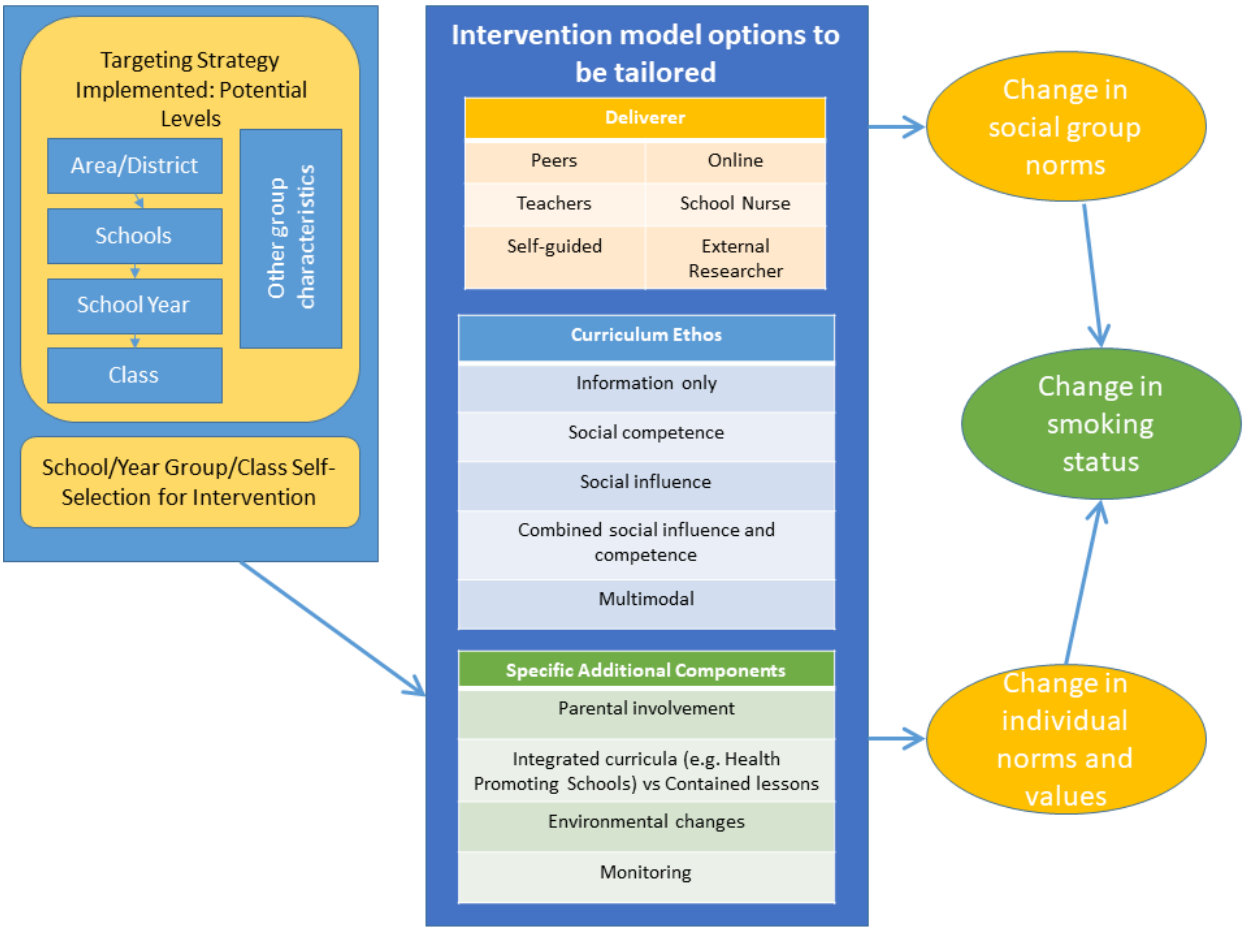
In charting progress of school-based smoking prevention interventions, Wiehe, Garrison et al. (2005) note progress away from information-only interventions, which were the mainstay of interventions during the 1960s, and followed an information deficit model where it was believed that information about harmful impacts of smoking would be enough to change behaviour. Social competence interventions, first developed during the 1970s, and social influence interventions, first developed during the mid-70s, recognised the influence of individuals' self-worth and peer group influences respectively (Wiehe, Garrison et al. 2005). More recent moves towards multimodal school-based smoking prevention interventions have sought to accompany a focus on individual change with wider changes in children's social networks and environments (Thomas, McLellan et al. 2013; Thomas, McLellan et al. 2015), mirroring thinking in public health that seeks to change the systems that generate inequalities in health, such as the adoption of smoking. Multimodal strategies have also spurred the development of new theoretical bases underpinning interventions, such as the theory of 'health promoting schools', which states that schools can improve students' levels of practical reasoning, social affiliations and autonomy necessary to avoid risky health behaviours such as smoking through, among other strategies, enhancing relationships between students and staff, student involvement in the running of schools, and aligning the values of schools with their local communities (Markham and Aveyard 2003; Peterson, Donze et al. 2019).

School-based smoking prevention interventions can involve some form of targeting of schools where children are thought to be at particular risk. For example in their study measuring the effectiveness of the ASPIRE intervention (A Smoking Prevention Interactive Experience) specifically targeted schools that were 'ethnically diverse and located in socioeconomically challenged neighbourhoods' (Prokhorov, Kelder et al. 2010). In addition, school-based smoking prevention interventions can be delivered through a variety of different means (deliverers) and can involve a number of different additional components or variants that reflect different curricula but also go beyond this distinction, some of which are listed in figure 10.

Unintended consequences

Few unintended consequences are reported, although there are concerns that school-based smoking prevention interventions could inadvertently strengthen smoking cultures among those at risk and marginalised in the school system. Theories of Health Promoting Schools, for example, suggest that where a disproportionate emphasis is placed on academic achievement and where students feel excluded from decision-making in the school, then students may have weaker capacity for practical reasoning and may feel estranged from the school culture (Markham and Aveyard 2003), and may seek out counter (pro-smoking) cultures. Adverse consequences of school-based smoking prevention interventions include the possibilities that the interventions can cause adolescents to find alternative places to smoke, develop alternative values around smoking and the attractiveness of smoking peer groups, and may encourage young people to view smoking as an expression of personal autonomy (Schreuders, Nuyts et al. 2017). However, many of these potential mechanisms are not quantified in the literature.

Figure 10: School-based smoking intervention logic model



Intervention effectiveness

Critical reviews suggest that school-based smoking interventions can be effective when supported by specific combinations of components and processes including being interactive and incorporating social influence and social competence elements, involvement of a number of reinforcing sessions, and that change the social meaning of being a smoker (Flay 2009). Recent evidence has suggested school-based resilience-only or information-only interventions are unlikely to be successful in preventing tobacco use (Thomas, McLellan et al. 2013; Thomas, McLellan et al. 2015; Hodder, Freund et al. 2017), but that dual component (social influence and social competence) and health promoting schools (multicomponent) interventions are successful in reducing tobacco use (Langford, Bonell et al. 2015, Thomas, McLellan et al. 2015). Realist systematic reviews on school tobacco policies have identified four mechanisms of action that include (i) a threat of sanction to be faced if found smoking; (ii) reduction in social pressure to smoke; (iii) internalisation of anti-smoking messages; and (iv) the practical difficulty of being able to smoke during school hours making it easier for children to maintain their resolve around not smoking (Schreuders, Nuyts et al. 2017). However, historically the field has been characterised by studies with weak designs (Wiehe, Garrison et al. 2005), particularly with regards to long-term follow-up, and interventions are yet to specifically address vaping. Alongside prevention programmes, school-based interventions aimed at helping young people to quit smoking are also found to be effective in some reviews (Garrison, Christakis et al. 2003).

Potential applications of a PPH approach

Targeting

PPH approaches could be implemented to better target schools at risk of high numbers of pupils smoking. Current approaches can involve some targeting strategies based on socioeconomic and demographic factors (Prokhorov, Kelder et al. 2010; Langford, Bonell et al. 2015), although the selection and recruitment of schools into smoking prevention interventions are generally not described in great detail. School-based interventions focussed on other health issues have used other characteristics to target schools; for example interventions aimed at reducing dental caries have targeted schools by level of disadvantage and fluoridation of water supply (Cakar, Harrison-Barry et al. 2018), while within the education literature there exist a number of approaches for targeting schools for various education and school improvement interventions (Lupton and Thomson 2015). Approaches to targeting school-based smoking prevention interventions could involve identifying schools with high levels of smoking among young people, and potentially situated in areas with high levels of smoking across all ages, although it is questionable whether this does add a ‘sufficient’ degree of precision, or meets a definition of PPH.

A theory driven PPH approach to identifying schools may involve identifying those that do not adhere to the health promoting schools framework through having a curriculum that does not integrate health education into general education; that has low involvement of pupils in school-level decision-making; and where there is only a weak alignment between the values of the

school and the wider environment. There may be some possibilities around the use of data collected from online social networks, although in general it is difficult to contemplate how 'new' sources of data could help to identify and target schools according to these criteria without also relying on traditional approaches such as surveys. There may be a role for PPH approaches to play an adjunct role in identifying schools with weak student involvement in creating the school environment or where health is not well integrated into day-to-day decision-making. For example, some machine learning applications may be useful in identifying schools where health is not frequently discussed in educational meetings or where there is little pupil involvement, based on meeting records; such an approach would clearly be experimental. However, the merits of targeting schools for school smoking prevention interventions should be critically questioned. Given the public health benefits of smoking prevention and that the evidence suggests that schools can be effective sites for smoking prevention, an effect that is generalizable across a wide range of schools and student demographic characteristics, the case for anything other than universal approaches to this type of intervention is questionable.

Tailoring the intervention

The evidence around the effectiveness of school-based smoking prevention interventions is strongly suggestive that interventions that involve combined social competence and social influence curricula (Thomas, McLellan et al. 2015), or adopt a health promoting schools framework (Langford, Bonell et al. 2015), are those that lead to higher effect size. These align with current public health thinking that interventions that focus on individual behaviour change and that aim to disrupt systems across individuals' social networks and broader environments, are most effective. There is little within the evidence to suggest that other alternative approaches should be selected dependent on student or school composition.

Within a school however, PPH approaches could add depth to decisions around how to tailor an intervention to reflect the sentiments and behaviours of intervention recipients (Lutkenhaus, Jansz et al. 2019). For example, the language around schools, social relationships and smoking could be modified within an intervention to reflect the values and language of young people within a particular school. PPH approaches could be used to understand the risk of adopting smoking across different social networks and peer groups, mirroring applications described earlier around the use of PPH to understand community dynamics. This is particularly pertinent for school-based smoking prevention interventions, given that students who are exposed to pictures of risk-taking behaviours on online social networks are more to smoke themselves (Huang, Unger et al. 2014). Similarly, PPH approaches using crowdsourcing to collect data on preferences and norms, and potentially forms of AI/machine learning to understand patterns within the data in order to shape interventions, could align with theories around health promoting schools which seek to maximise student involvement in designing the school environment.

It is also likely that future interventions can incorporate greater tailoring towards individual preferences and needs. For example, some online interventions delivered within schools for smoking cessation have tailored content according to the trans-theoretical (stages of change) model to adapt the intervention content to reflect individual students' needs (Evers, Paiva et al. 2012). However, based on the evidence from the systematic reviews above, such individualised content is likely to be most successful when the intervention also addresses social network

norms and children's wider relationships with schools; again this suggests that a PPH approach may be a useful adjunct but unlikely to form the mainstay of a school-based intervention.

The NIHR Policy Research Programme Reviews Facility aims to put the evidence into development and implementation of health policy through:

- Undertaking policy-relevant systematic reviews of health and social care research
- Developing capacity for undertaking and using reviews
- Producing new and improved methods for undertaking reviews
- Promoting global awareness and use of systematic reviews in decision-making

The Reviews Facility is a collaboration between the following centres:
EPPI-Centre (Evidence for Policy and Practice Information and Co-ordinating Centre), UCL Institute of Education, University College London;
CRD (Centre for Reviews and Dissemination), University of York;
and the London School of Hygiene and Tropical Medicine.

The NIHR Policy Research Programme Reviews Facility collaboration has grown out of a previous 'reviews facility' in Health Promotion and Public Health based at the EPPI-Centre, and has been funded by the Department of Health and Social Care since 1995.

First produced in 2020 by:

Evidence for Policy and Practice Information and Co-ordinating Centre
(EPPI-Centre)
Social Science Research Unit
UCL Institute of Education, University College London
18 Woburn Square
London WC1H 0NR

http://eppi.ioe.ac.uk
http://www.ucl.ac.uk/ioe

ISBN: 978-1-911605-17-1

This document is available in a range of accessible formats including large print. Please contact the Social Science Research Unit for assistance.

Email: ioe.ssru@ucl.ac.uk
Telephone: +44 (0)20 7331 5263