



**Bristol, North Somerset
and South Gloucestershire**
Clinical Commissioning Group

Understanding the distribution of health needs using Population Segmentation

An applied comparison of methods



@BNSSGAnalytics

Richard Wood
Ben Murch
Adrian Pratt

Population Health Management

1.4. To respond to these challenges, improve care for patients and reduce pressure on staff, this plan means that:

- **more joined-up** care institutions, people with long health service as
- **more proactive** with the NHS on practice, A&E or into action. But **health management** screening, case to avoid illness con
- **more different** to make further to the diverse pe and location of use of 'multicha of all ages can a mental wellbeing determinants, a tailored preventi with the people

This chapter the to bring this ab

1. We will **boost** primary and c
2. The NHS will
3. People will ge when they ne
4. **Digitally-ena**
5. Local NHS org partnerships v Systems (ICSS

Here's how:

2.6. As described in Chapter One, our new integrated care systems (ICSS) will help deliver these programmes as the NHS **continues to move** from reactive care towards a model

embodying a and care syst foundations f agenda of pr teams across network com Long Term Pl

2.7. The role deterioration **the NHS cor bring home action to do improve the healthy life ex**

Smoking

2.8. Smokin lost than any Smokers see nearly half a of women in

Figure 10

5.24. **Technology will enable the NHS to redesign clinical pathways.** Easy to referral decision trees, referral templates and direct access to investigations t evidence-based best practice and universal access to 'one click away' specialist guidance for GPs, will avoid many patients from requiring referral for an appoint (and potentially completing) some specialist referrals such as in dermatology wi questionnaires will allow some patients to be managed entirely digitally. Virtual escalation to face-to-face appointments where needed – such as the virtual frag in Bradford and renal care in Tower Hamlets – can replace follow-up appointme conditions, as set out in Chapter One.

5.25. **By 2022, technology will better support clinicians to improve the and reduce the health risks faced by children and adults.** An integrated cl system will replace dozens of legacy systems and we will deliver a screening and solution that is worthy of the NHS' world leading services.

4. Improving population health

5.26. **During 2019, we will deploy population health management solu support ICSSs to understand the areas of greatest health need and match to meet them.** Over the coming years these solutions will become increasingly in identifying those groups of people who are at risk of adverse health outcome which individuals are most likely to benefit from different health and care inter as shining a light on health inequalities. We will be able to routinely identify m of pathways of care for individuals and ensure that those gaps are filled. This w greater transparency of health and social care data on population health outco organisational performance.

5.27. **The use of de-personalised data extracted from local records, in li information governance safeguards, will enable more sophisticated pop management approaches and support world-leading research.** We will m APIs available to industry and the developer community to stimulate innova integration with other products. We will enable the NHS to work with suppliers user journeys, supplemented with data and insights, that help clinicians to do t effectively and more efficiently. The initial API and workflow integration initiati towards full integration with smart home and wearable devices.

Source: Roy with other w

TheKingsFund

Ideas that change health care

A vision for population health Towards a

David Buck
Alex Baylis
Durka Dougall
Ruth Robertson

November 2018

nuffieldtrust

Evidence for better health care

Our priorities

Research

Health & social care explained

News & comment

Vive la re health m

In a guest blog, Ar why population he systems all over th approaches we ha

Blog post

Published: 19/06/2019

Please note that views exp the views of the Nuffield Tr

Population health man trend du jour, not just in Long Term Plan, but as health systems all over t helps system leaders bet populations and deliver for, and proactively inte risks and outcomes.

This may sound somev health management is d we have taken before in at promoting, protecti offered to the whole pop identified groups. Often diagnosis.

Deloitte.



The transition integrated c Population he management

March 2019

HSJ

QUALITY AND PERFORMANCE

Pinpoint, predict, prevent: how population health management can make a difference

By Amanda Doyle | 11 July 2019



Population health management is causing us to change the way we work so that we intervene at a point at which we can influence what happens to people, writes Amanda Doyle

Sponsored by
NHS England

A lot of people think that Population Health Management is all about data – crunching it, monitoring it and governing it. Yet that's only part of the equation. What the data illuminates is where the real focus should be.

From helping to pinpoint people at risk of ill health, to predicting those in danger of exacerbating, to preventing ill health altogether, population health management provides a way of bringing together various sources of intelligence to systematically deliver personalised care.

What it's really about is using data-driven insights to make intelligent decisions about how to deliver care differently. Data may be the jump-off point, but it's by no means the whole story.

Put resources to best use

In Lancashire and South Cumbria, where inequalities in health are particularly stark, it's helping us to get under the skin of the causes of ill health and how we can use our resources to best effect.

NHS

Bristol, North Somerset
and South Gloucestershire
Clinical Commissioning Group

What is PHM?

*“PHM is the concept of gathering data and insights about population health and wellbeing across multiple care and service settings, with a view to identifying the **main health care needs** of the community and adapting services accordingly”*

Deloitte

*“Population Health Management focuses on **key outcomes for identified groups**. Often these groups share more specific common characteristics, not just a disease diagnosis”*

Nuffield Trust

*“During 2019, we will deploy PHM solutions to support ICSs to understand the areas of greatest health need and match NHS services to meet them. Over the coming years these solutions will become increasingly sophisticated in identifying those **groups of people** who are at risk of adverse health outcomes”*

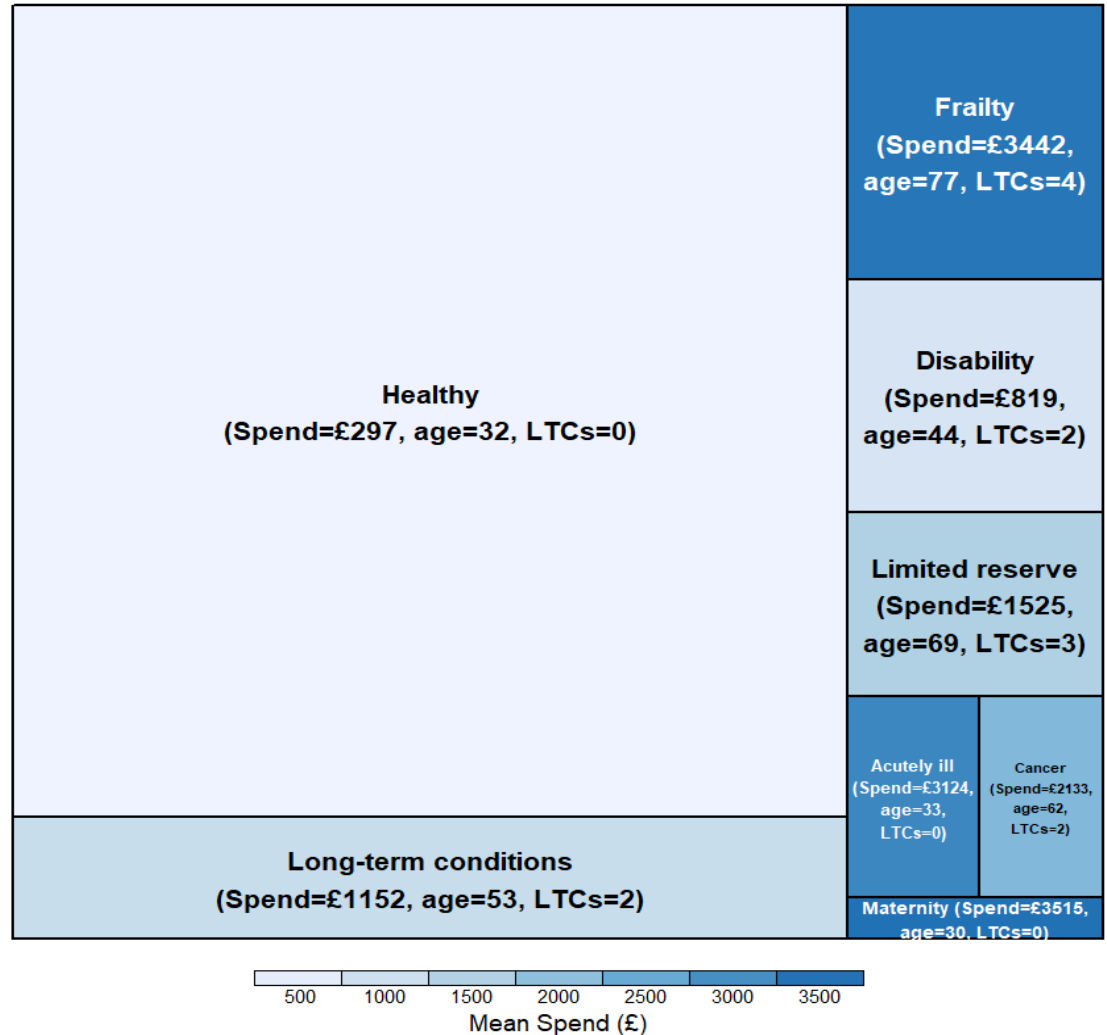
NHS Long Term Plan



Bristol, North Somerset
and South Gloucestershire
Clinical Commissioning Group

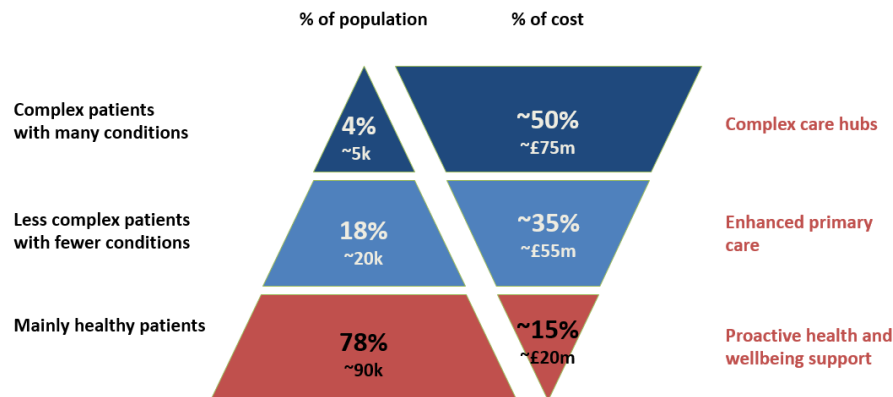
Example segmentation

- **Block size** represents segment population
- **Colour depth** represents spend differential



Why segment?

- Cut through the complexity of large and unwieldy datasets in making sense of the **key patient-related attributes** that drive the most significant differences in some targeted measure of interest
- **Activity or spend** can be targeted in representing **healthcare need**
- Identified segments thereafter **amenable to tailored interventions**
 - E.g. Complex Care Hub for the most seriously ill patients



From Somerset
Symphony project

Aim

- To review a number of Population Segmentation methods **side-by-side in a consistent manner** using a common dataset...
- ...that identify meaningful and interpretable population cohorts which are *heterogeneous between* and *homogeneous within*
- To open up **a range of options** allowing clinicians and managers an informed choice on which approach to use for their situation



Outline of introduction

1. Setting and data
2. Methods
3. Comparison
4. Discussion

Data requirements

- **Attributes**

Demographic – *age, sex, ethnicity*

Clinical – *listed chronic conditions, obesity, frailty scores*

Other – *deprivation, smoking status, housebound*



- **Activity**

Primary Care contacts – *GP and nurse appointments*

Secondary Care contacts – *outpatient, inpatient, A+E*

Prescriptions – *medicines*



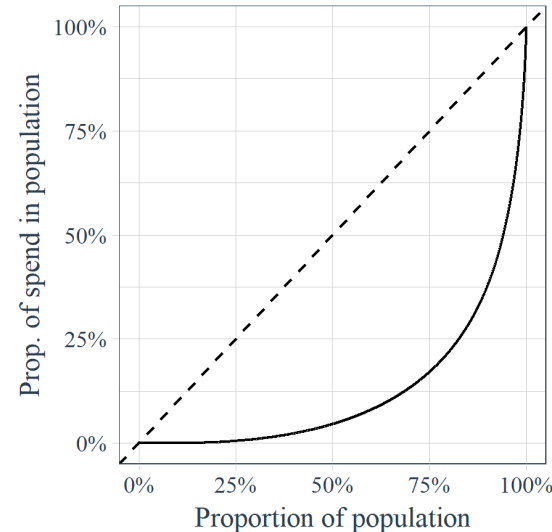
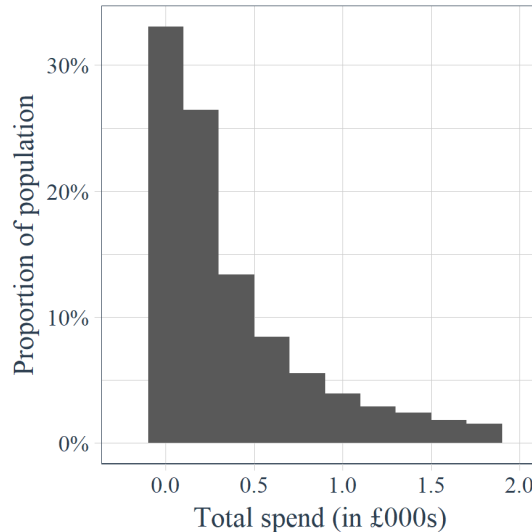
Sample data at a glance (n=51,072)

Highly skewed

Mean £853

Median £240

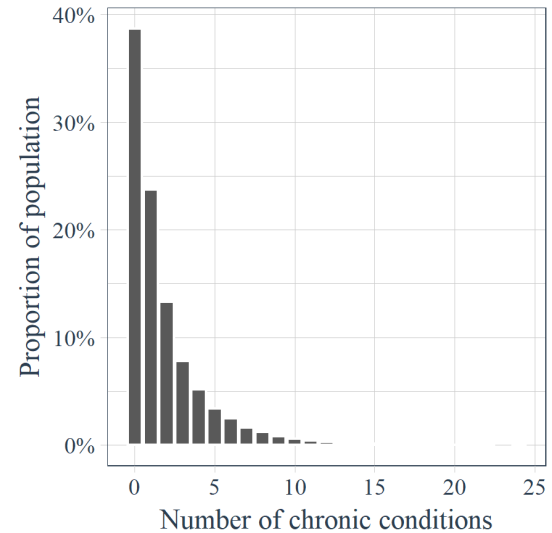
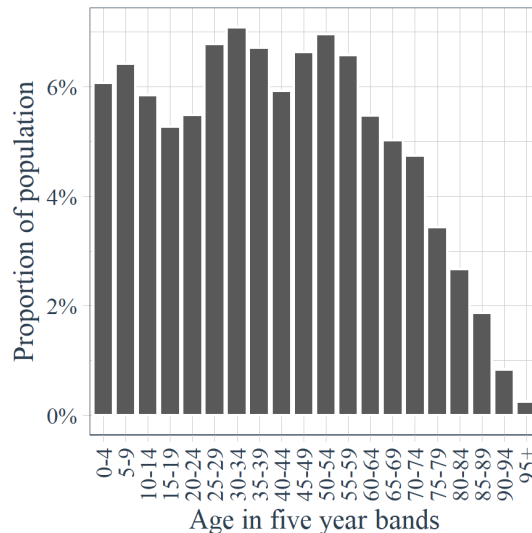
Multi-modal
population in
Bristol area



Inequal
distribution

Gini 0.75

80/20 rule
holds



61.3% of
people
have at
least one
chronic
condition



Outline of introduction

1. Setting and data

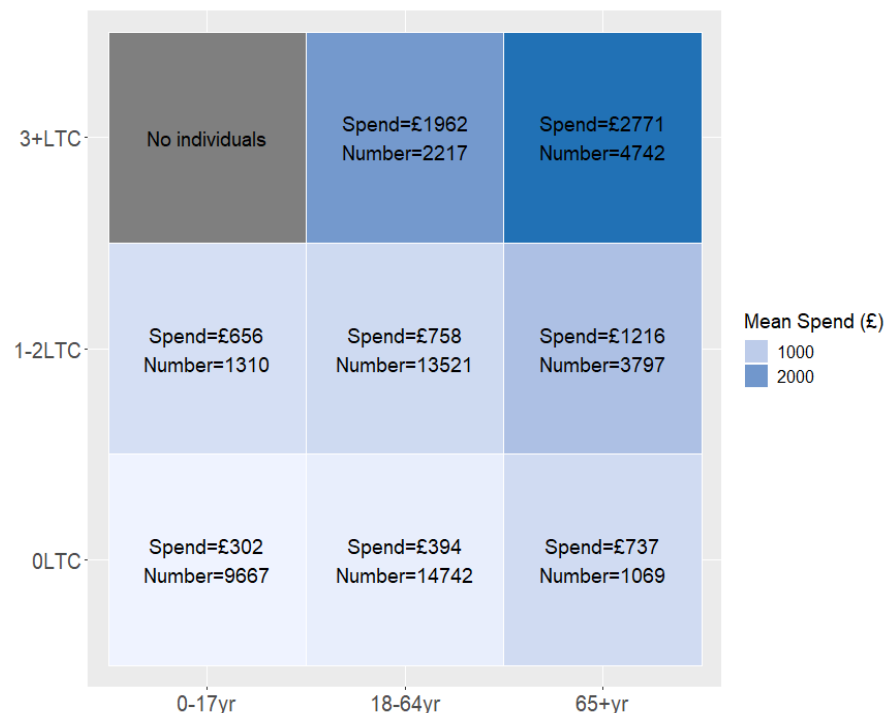
2. **Methods**

3. Comparison

4. Discussion

Judgemental splits

- Most basic approach
- Simply obtain segments by **making arbitrary splits** on explanatory variables (patient attributes)
- E.g. by age and chronic condition count



1. Age (0-12, 13-17, 18-49, 50-74, 75+)
2. Sex (male/female)
3. Chronic conditions (0-4, 5-9, 10-14, 15+)
4. Age and chronic conditions (0-17, 18-64, 65+ by CCs 0-4, 5-10, 11+)

Prescribed binning criteria

- Bin patients according to **preset rules** defining segment membership
- No guarantee that rules will be:
 - Discriminative
 - Relevant to geography/setting
 - Appropriate to targeted measure
- Some rules **not well-defined**

TABLE 1
Population Segments with Typical Patient Examples

Population	Patient and Services
1. Healthy	Mr. Smith, a 37-year-old carpenter, usually books an appointment with his primary care physician each year around his birthday for an annual checkup and necessary screenings. He also may contact his physician's office for acute, self-limiting problems such as a sore throat.
2. Maternal and infant health	Mrs. Brown, a 26-year-old waitress, had regular contact with her gynecologist for contraception and general health monitoring until deciding to become pregnant. A year later, she sought fertility treatment and had monitoring through normal pregnancy and delivery. Her newborn's checkups and immunizations follow national guidelines.
3. Acutely ill	Tom Jones, an 18-year-old high school student, broke his femur while playing football. An ambulance promptly transported him to the local emergency room. Following an uneventful surgical procedure, Tom received physical therapy to rehabilitate his leg and maintain his body strength. He returned as the team quarterback eight weeks later.
4. Chronic conditions, normal function	Mrs. Gomez, a 49-year-old teacher, has hypertension and diabetes. While she has taken classes to learn how to reduce her risks and control these conditions, she still finds that both are occasionally out of control and then makes an appointment with her physician, whose office sends her reminders for immunizations, regular checkups, and monitoring for possible complications.
5. Stable but serious disability	Mr. White, a 56-year-old telemarketer, also is a former paratrooper who is quadriplegic from a gunshot wound to the neck. He lives with his brother in an extensively adapted apartment and has a paid aide for personal care. He has a motorized wheelchair and transportation for shopping and outings. He has been suicidal at various times and often has urinary tract infections. He uses a medical home team for continuity and comprehensive coordination of services, and he and the team work from a negotiated plan of care.

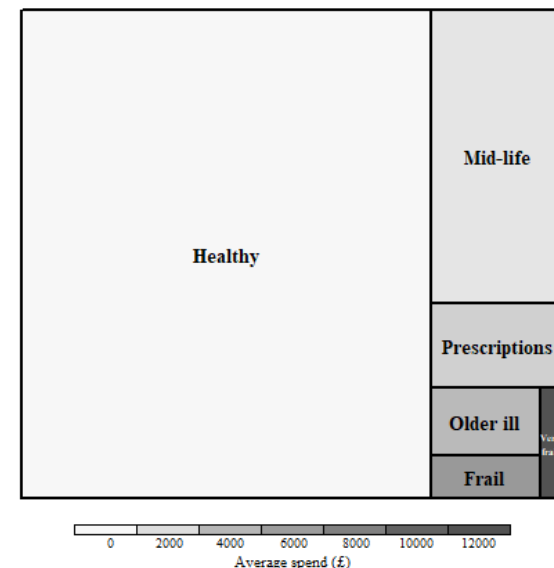
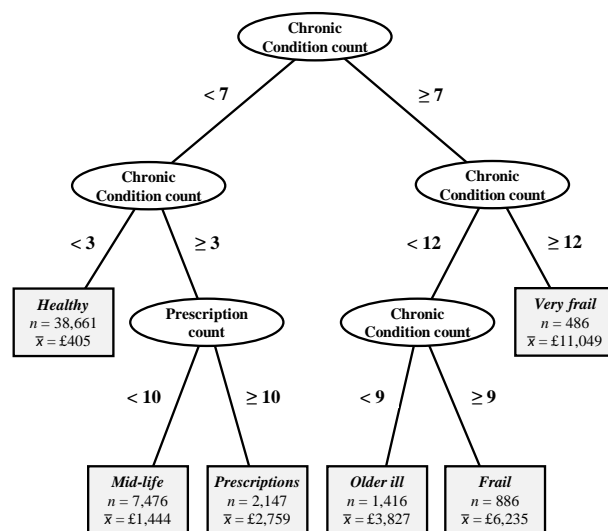
(Continued)

Bridges to Health

- Lynn et al, 2017 "Bridges to Health"
- Low et al, 2017 (Singapore)
- Joynt et al, 2017 (USA Medicare)
- Electronic Frailty Index (UK, four levels)

Decision trees

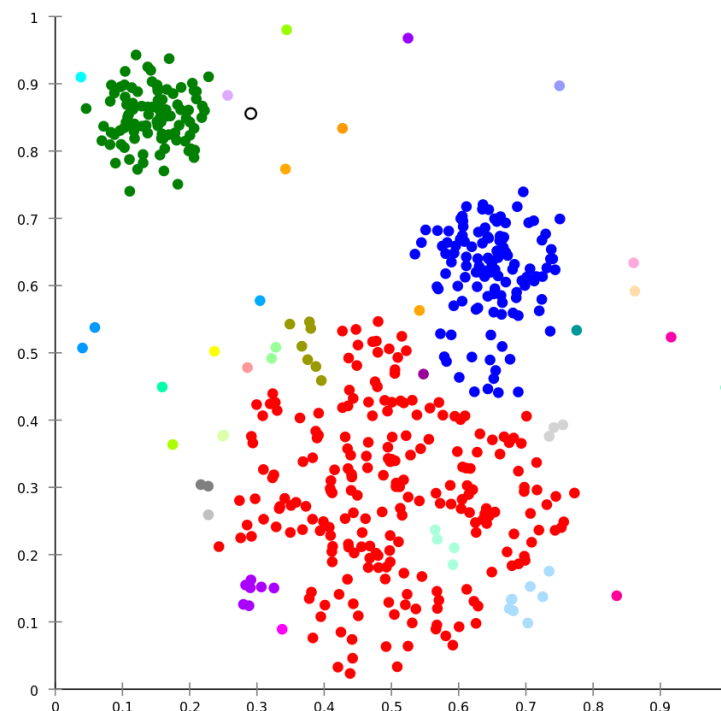
- A “**derived**” approach
- Objectively **seeks discrimination** in target variable (spend)
- Segment names and definitions may be less intuitive



9. CART (Breiman method)
10. Conditional inference trees
11. C5.0 (information gain)
12. CHAID (Chi-square significance testing)

Cluster analysis

- Also a derived approach
- Finds groups where observations are **most similar**
- **Non-objective**: similarity driven by attributes and not target
- Need to define “k” upfront



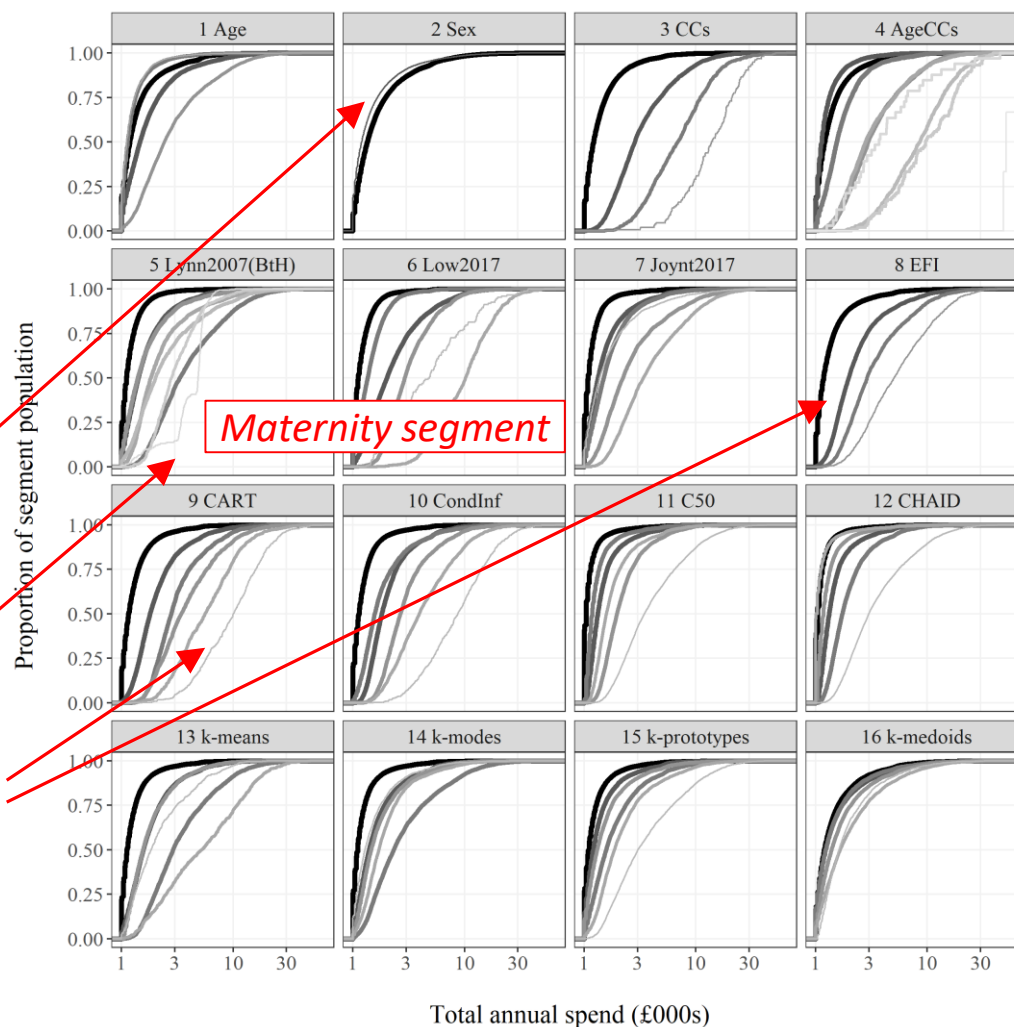
13. k-means (mean Euclidean distance)
14. k-modes (Hamming distance)
15. k-prototypes (numeric + categorical data)
16. k-medoids (sum of square distance)

Outline of introduction

1. Setting and data
2. Methods
3. Comparison
4. Discussion

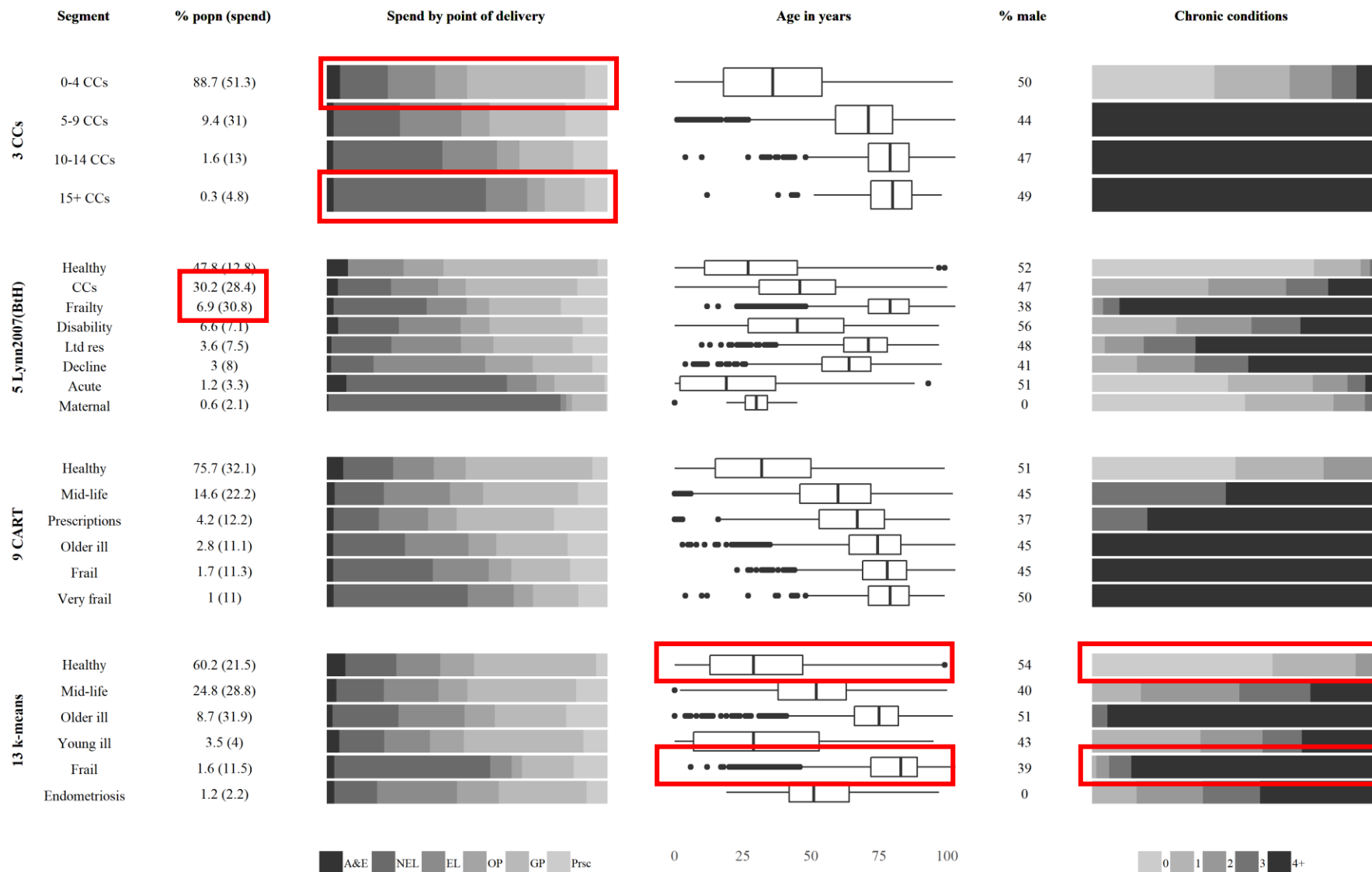
Segment-level spend distributions

- Can understand:
 - Number of segments
 - Spend profile
 - Number in segment
- Rules of thumb:
 - Closer lines = less discrimination
 - More volatility = less homogeneity
 - Consistent progression in line shading = larger less expensive to smaller more costly



3. COMPARISON

Spend and attributes differential



Reduction in variance (spend)

- To be useful for PHM, methods must be **discriminative**

- Assessed through:

$$\sigma_W^2 = \frac{\sum_{i=1}^N n_i \sigma_i^2}{\sum_{i=1}^N n_i}$$

- A **mixed bag** of results
- Some methods “cheat”

Category	Method	Variance	Reduction
<i>Baseline</i>	<i>Baseline</i>	5,433,321	-
Judgemental	Age	4,912,301	10%
	Sex	5,418,924	0%
	Chronic conditions	3,516,112	35%
	Age and chronic conditions	3,505,768	35%
Prescribed binning	Bridges to Health	4,275,849	21%*
	Low2017	3,145,663	42%*
	Joynt2017	4,376,765	19%
	Elec Frailty Index	4,319,655	20%
Decision trees	CART	3,352,816	38%
	Conditional inference	3,403,747	37%
	C5.0	4,017,665	26%
	CHAID	4,180,762	23%
Clustering	K-means	4,088,182	25%
	K-modes	4,647,659	14%
	k-prototypes	4,564,590	16%
	K-medoids	5,331,346	2%

* methods in which activity is used as an explanatory variable

Outline of introduction

1. Setting and data
2. Methods
3. Comparison
4. Discussion

Considerations for practical use

- **Discrimination vs segment interpretability**
 - CHAID achieves just 4% better discrimination than Joynt2017 method
 - Really worth this given that Joynt has very comprehensible binning rules?
- **Implementation complexity**
 - Performing a judgemental split is trivial
 - Performing FAMD based *k*-medoids requires specialist skillset
- **Availability of explanatory variables**
 - Different methods require use of different variables
 - Derived methods should have lots of candidate variables to choose from
 - Also raises computational issues

Conclusions

- There is **not necessarily a right or wrong method** for descriptive population segmentation – many factors to weigh w.r.t. discrimination and practicality
 - **Cluster analysis** on-the-whole unsuitable – computationally expensive, lacking discrimination, laborious pre-processing
 - **Prescribed binning** e.g. Bridges to Health: unlikely to achieve high levels of discrimination, but may be more interpretable and useful for benchmarking
 - **Decision trees** arguably preferred option since they offer a sound conceptual fit to the problem and promote good, data-derived discrimination
 - Otherwise, if insufficient data or expertise, then **judgemental splits** focusing on the number of chronic conditions should be favoured



Contents lists available at ScienceDirect

Operations Research for Health Care

journal homepage: www.elsevier.com/locate/orhc

A comparison of population segmentation methods

R.M. Wood^{a,b,*}, B.J. Murch^b, R.C. Betteridge^{b,c}^aSchool of Management, University of Bath, Claverton Down, Bath, BA2 7AY, United Kingdom^bModelling and Analytics, UK National Health Service (NHS) CCG, South Plaza, Marlborough St, Bristol, BS1 3NQ, United Kingdom^cData Science, Genentech, 27 North Wharf Road, London, W2 1NF, United Kingdom

HIGHLIGHTS

- Addresses a gap in the literature regarding population segmentation methods.
- Provides a side-by-side comparison of methods used in practice and by researchers.
- Results show that locally-calibrated decision trees offer the best discrimination.
- Findings provide useful advice to healthcare managers on the ground.

ARTICLE INFO

Article history:

Received 6 February 2019

Received in revised form 11 June 2019

Accepted 21 June 2019

Available online 26 June 2019

Keywords:

Population health

Population segmentation

Decision trees

Cluster analysis

Healthcare utilisation

ABSTRACT

This paper presents the first comparison of descriptive segmentation methods for population health management. The aim of descriptive segmentation is to identify heterogeneous segments according to some target observed measure. In healthcare it can be used to understand how utilisation is distributed among a population, and to identify the patient attributes which explain the greatest differences (knowledge of which can help shape segment-tailored services). In reviewing a number of segmentation methods that are both employed on the ground and explored more experimentally within the academic literature, this paper aims to open up a range of options allowing clinicians and managers an informed choice on which approach to use for their situation. Results support the recommendation that decision tree approaches are on-the-whole most suitable, being configurable to local data and providing the best inter-segment discrimination. More basic judgemental splits on patient attributes can be powerful, with the cost of chronic conditions being a key variable. Prescribed binning methods such as Bridges to Health are unlikely to achieve high levels of discrimination but do have easily interpretable segments and could be useful for benchmarking. Clustering methods are found to lack discriminative power, which can be attributed to a lack of conceptual appropriateness to the problem.

© 2019 Published by Elsevier Ltd.

1. Introduction

Population health has been defined as “the health outcomes of a group of individuals, including the distribution of such outcomes within the group” [1]. Interest in this field has grown in recent years driven by the combination of rising costs of care with increasingly polarised health needs leading to greater inequality in per capita spend and clinical outcome [2]. Facilitated by the rise of big data and the availability of associated analytical methods, this has led to the growth of population health analytics as a discipline concerned with quantitatively approaching matters of population health.

One of the principal investigative areas within the field of population health management – and the subject of this paper – is population segmentation. This involves using information about individuals, such as age and sex, to partition a population into similar groups. Ultimately, the aim is to identify meaningful and interpretable population cohorts which are heterogeneous between and homogeneous within. It is important to have such discrimination since it allows the greatest differences to be uncovered.

Population segmentation is an important tool in healthcare since it allows managers and clinicians to cut through the complexity of large and unwieldy datasets in making sense of the key patient-related attributes that drive the most significant differences in some targeted measure of interest – clinical outcome, waiting time, or utilisation as measured through activity or spend. These insights can help determine the nature and scale of intervention that may be required for cohorts of the population

* Corresponding author at: Modelling and Analytics, UK National Health Service (NHS) CCG, South Plaza, Marlborough St, Bristol, BS1 3NQ, United Kingdom.

E-mail address: richardwood16@nhs.net (R.M. Wood).

<https://doi.org/10.1016/j.orhc.2019.100192>

2213-6824/© 2019 Published by Elsevier Ltd.

Wood, R. M., Murch, B. J., & Betteridge, R. C. (2019). A comparison of population segmentation methods. *Operations Research for Health Care*, 100192.



@BNSSGAnalytics

richard.wood16@nhs.net



Bristol, North Somerset
and South Gloucestershire
Clinical Commissioning Group

Outline of today's workshop

1. Understanding the data
2. Basic segmentation by age
3. Inclusion of long-term conditions alongside age
4. Decision tree segmentation
5. Bridges to Health segmentation
6. Report generation via Rmarkdown
7. Further work on R-based PHM suite