A Review of Population Segmentation Tools for Population Health Management: Applicability for Ontario Health Teams

Mudathira Kadu
Xiaomeng Ma
Lief Pagalan
Laura C Rosella
Walter P Wodchis

Population Health Management Series: Part 2

February 2021

HSPN Health System Performance Network

The opinions, results and conclusions included in this report are those of the authors and are independent from the funding sources.

About Us

The Health System Performance Network (HSPN) is a collaborative network of investigators, visiting scholars, post-doctoral fellows, graduate students and research staff working with health system leaders, and policymakers to improve the management and performance of our health system. Building on Ontario's established record of performance measurement created by the 1998 ground-breaking Hospital Report Research Collaborative, the HSPN was established in 2009 and has built a track record in performance measurement, research, evaluation and improvement in Ontario with expertise in multiple domains of health system performance including perspectives of patients, providers, population health, and cost. The HSPN receives funding from the Ontario Ministry of Health.

Contact Information

Health System Performance Network
155 College Street, Suite 425
Toronto ON M5T 3M6
Telephone: +1 (416) 946-5023
Email: hspn@utoronto.ca

Financial Support

This research was supported by a grant from the Ontario MOH to the HSPN. The funders had no role in data analysis, decision to publish, or preparation of the report.

Suggested citation

Kadu M, Ma X, Pagalan L, Rosella LC, Wodchis WP. A Review of Population Segmentation Tools for Population Health Management: Applicability for Ontario Health Teams. Toronto, ON: Health System Performance Network. 2021.

ISBN 978-1-7774610-3-4 (Online)

This document is available at hspn.ca

Executive Summary

Background

In April 2019, following the enactment of the Connecting Care Act, 2019, the Ontario Ministry of Health (MOH) introduced OHTs as a new way of organizing and delivering care that is more connected to patients in their local communities. In their first year, OHT candidates were asked to select target population segments with the greatest potential for significant impact. As OHTs mature, they will be responsible for their whole attributed population while emphasizing high-risk and specific sub-populations. The emphasis on population segments or sub-groups requires that OHTs segment their attributed populations into population sub-groups with shared needs. This rapid review summarizes population segmentation tools aimed at supporting population health management to inform the choice of segmentation tools that may be considered by Ontario Health Teams.

Purpose

This report provides a review of population segmentation tools for population health management as they apply to Ontario Health Teams. In this report we:

1. describe the different approaches to segmentation and the contexts in which segmentation has been applied,
2. capture the data sources, features and methodologies used to develop population segments,
3. assess the actionability, and target audience for each segmentation approach.

Methods

We searched PubMed, CINAHL, EMBASE, SCOPUS and PROSPERO databases from November 2015 to November 2019. In our search strategy we included key terms capturing the concepts of “population”, “segmentation”, “tools” and “population management” to identify the articles conducting the segmentation is for population management. Articles were included if: (a) they were full-text original studies that segment a population for population management and (b) the segmentation was undertaken through a data-driven approach using empirical data, use of clinical judgement or risk assessments, or a combination of the approaches. Three reviewers completed data extraction from the selected articles. Data items were abstracted into three domains including: (a) basic characteristics of the study; (b) population health application; and (c) data content of the segmentation tool.

Findings

Three major approaches for population segmentation have emerged over the years. Expert-driven approaches defined or validated segments with concordance of expert judgements prior to or after the tools were completed. Predictive models using algorithms aim to create groups of patients with similar health care needs, primarily proxied by current and/or future healthcare use. Data-driven approaches employed a post-hoc statistical analysis such as clustering techniques or latent class analysis on empirical data to segment a population.

- Of the 17 identified tools, 7 were completely developed from data sources, while 10 were synthesized with expert review. Studies segmented populations at various levels, including at the whole population (macro) level, organization or sub-population (meso) level, and at the patient (micro) level to identify high-risk individuals. More than half of the studies focused on segmentation at the population level and gave explicit descriptions of the population they segmented (N = 15).

- The most common segmentation approach used Aggregated Clinical Groups (ACGs) (N = 6), followed by Minnesota Tiering (N = 2), and Clinical Risk Groups (CRGs) (N = 2). Among the modelling approaches, K-mean aggregating was commonly used for population latent feature clustering (N = 3).
- The data sources for health-related features used in the segmentation tools were derived from health administrative databases (N=12), clinical electronic health records (N = 9), or through primary data collection using health care needs assessment tools (N=1). The majority of tools focused on predicting future health care utilization (e.g., ED visits, admissions, readmissions, hospitalizations; (N = 14) and costs (N = 9).
- The end users of the segmentation tools ranged from clinical practitioners, policymakers, healthcare providers, financial experts, and managers of care coordination programs, aligning to the purpose of tool development
- Across the range of studies and applications, the number of segments ranged widely from 3 to 534. The segments could be defined based on the health status, health resource utilization, demographics, geography, the severity of a designated illness, or case-mix clusters defined by risk scores from hybrid models; equity dimensions were considered in only a rare few cases.

Key Learnings

We recommend that OHTs examine both a tool based on an existing validated algorithm and a tool derived from a needs-oriented perspective to assess the relative merits of each in the Ontario population and the potential for transition to a needs-based approach to care. We recommend that each be applied using OHT attributed populations and built upon Ontario's existing population health administrative databases supplemented as possible by other data sources such as surveys or functional clinical assessments. Future exploration of data-driven approaches with machine learning or related statistical techniques may be considered although these tend to lack the important consideration of the linkage between patient groups and appropriate health care services.

Conclusion

There is a substantial literature in population segmentation tools, although the application of these tools in practice is less well described. Tools that derived from population administrative or electronic health databases and coded through existing validated algorithms have the greatest opportunity to enable inclusion of an entire population. We recommend that OHTs examine both a tool based on an existing validated algorithm and a tool derived from a needs-oriented perspective to assess the relative merits of each in the Ontario population and the potential for translation as a needs-based approach to care. We recommend that each be applied using OHT attributed populations and built upon Ontario's existing population health administrative databases supplemented as possible by other data sources such as surveys or functional clinical assessments.

Table of Contents

Introduction

Context: Ontario Health Teams

In 2019, the Ontario Ministry of Health and Long-Term Care invited health service providers across the full continuum of care to come together and to demonstrate their readiness to become or join an Ontario Health Team (OHT). OHTs are groups of providers and organizations that are clinically and fiscally accountable for delivering a full and coordinated continuum of care to a defined attributed population (Ontario Ministry of Health and Long-Term Care 2019). As well as being accountable for an attributed population, OHTs are responsible for optimizing their population's experience of care and health outcomes while also managing per capita costs. The goal of this population health management approach is to shift the whole population curve toward better health and decrease health inequities [1]. OHTs can achieve this goal through preventative and coordinated care, by offering services to patients seeking care, and by proactively connecting with those who have not had adequate access to the health system[1].

In their first year, OHT candidates were asked to select target population segments with the greatest potential for significant impact. As OHTs mature, they will be responsible for their whole attributed population while emphasizing high-risk and specific sub-populations. The emphasis on population segments or sub-groups requires that OHTs segment their attributed populations into population sub-groups with shared needs (e.g. common health conditions or risk factors, utilization patterns, sociodemographic factors, etc.)[1]. This rapid review summarizes population segmentation tools aimed at supporting population health management to inform the choice of segmentation tools that may be considered by Ontario Health Teams.

Segmentation for Population Health Management: Macro, Meso and Micro Application

This rapid review focuses on population segmentation tools to support population health management strategies. Population segmentation involves dividing patient populations into distinct groups based on common health characteristics among individuals. Population segmentation is closely related to risk-stratification, and tools are often described synonymously because some segments of the population may have a higher risk for adverse and expensive health care utilization. In this report, we focus on population segmentation as a tool to support the optimization of a health system by matching services to healthcare needs. Segmentation tools are often based on the existing patterns of healthcare, which may not meet the current needs of the population.

In population health management, segmentation provides efficient ways to develop integrated health care strategies and allocate resources tailored to the needs of distinct population segments [2]. Population segmentation can also be used to support detailed information on sub-populations to help develop and calculate capitated budgets [3]. For whole populations, segmentation ensures that the health care needs of all population groups are considered and that targeted programs are delivered based on individual characteristics and need [3]. At a policy level, organizing integrated care around population segments shifts the focus towards patient-centred care, allowing all relevant stakeholders across the health system to be involved, including social and community care [3].

Population segmentation can be described at the macro (whole populations), meso (sub-populations), and micro (high-risk individuals) level [2,3] . Macro-populations can be defined by a geographic region, catchment area, or membership to a health plan or health system responsible for payment and delivery of health services [4]. At the macro-level, whole populations are assessed to understand population health, system-level health care needs, and inform population health management. Population health management involves improving healthcare quality, optimizing healthcare spending, and designing interventions to maintain and improve people's health across a large population with varying levels of health [4]. At the macro-level, population segmentation identifies how care needs vary across the whole population and informs how to tailor policies and budgets for homogeneous patient groups [3].

To better manage population health at the meso-level, individuals are assessed and segmented into sub-populations based on their risk for adverse health events, health needs, or characteristics and behaviours (e.g. healthy, chronically ill, maternal and infant health, frail seniors, etc.) [4]. Segmentation at the meso-level helps identify priority subgroups and provides a better understanding of the target population [3]. As well, segmentation can inform targeted interventions, care delivery, and tailored policies for priority populations with specific health conditions, such as improved care coordination for patients with multiple long-term conditions and effective diagnosis and treatment for people with no long-term care needs [3,4]. Segmentation at the meso-level can also be used to set target budgets and capitation payment levels.

At the micro-level, the focus is on individuals who are at high risk of specific outcomes. Segmentation at the micro-level can be used for risk-stratification to deliver appropriate care management interventions [3].

Purpose

This report was prepared for OHTs interested in applying population segmentation tools for population health management. The report focuses on the tools and their characteristics and describes how they were actioned in practice.. A scoping review approach was taken to enable a wide array of segmentation approaches to be included. In particular, we sought to:

1. identify the different purposes and approaches to segmentation and the contexts in which segmentation had been applied;
2. capture the data sources, features and methodologies used to develop population segments;
3. assess the actionability, and target audience for each segmentation approach.

Methods

Data sources and search strategy

We searched PubMed, CINAHL, EMBASE, SCOPUS and PROSPERO databases from November 2015 to November 2019. In our search strategy we included key terms capturing the concepts of “population”, “segmentation”, “tools” and “population management” to identify the articles conducting the segmentation is for population management. Key terms linked to concepts can be found in appendix (Table S1). The articles found by three previous literature reviews [4-6] of a similar topic and including search results early than 2015/11/01, were added as supplement. Electronic records from databases were imported to and managed using the Mendeley reference management software.

Table 1. Search key terms matching with concepts

Concept	Search term
• population	communit* population*
• segmentation	typolog* stratification* segmentation* classification* categorization* categorisation*
• population health management	health management population management population health care management community engagement facility management Health Care Quality, Access, and Evaluation Delivery of Health Care Population Health Management public health administration public health practice public policy
• tool	model* tool guideline* method* approach* technique* formula* strateg*

Study eligibility criteria

Articles were included if A) they were full-text original studies that segment a population for population management and B) the segmentation was undertaken through a data-driven approach using empirical data, use of clinical judgement or risk assessments, or a combination of the approaches. We excluded 1) non-English studies, 2) reviews, study protocols or conference proceedings, , 3) studies that focused on the effectiveness of care management programs without any description of subgroup segmentation methods, 5) studies with an objective of segmentation only for cost/financing purposes. Inclusion and exclusion criteria were decided a-priori and maintained throughout the literature screening.

Study selection

The literature search was conducted by XM. Two authors (XM and MK) independently screened the titles and abstracts of the citation records and then eligible articles with full text. Ten percent of full-text papers identified by XM and MK were randomly selected and reviewed by two researchers (DW and JH) to confirm the eligibility of inclusion. Disagreements in articles and validation on preliminary eligible criteria were resolved through group discussion.

Data extraction and synthesis

Three reviewers (LP, XM and MK) completed data extraction from the selected articles. Data items were abstracted into three domains including: (a) basic characteristics of the study; (b) population health application; and (c) data content of the segmentation tool. These aspects were selected to compare the original context to Ontario Health Team context, determine evidence regarding the applicability for the purposes of population health management, and finally assess whether the data elements required are available at a population level in Ontario.

Basic characteristics included the year and country of publication, study population, data source, name of the segmentation tool, purpose of the tool, end user and ability to capture the change. Operational characteristics denoted feature and derivation of segments, including number and description of segments, primary outcome, the purpose of segmentation, method of delivery of segmentation information to user and actionability of segmentation results. Data content was categorized as health status, sociodemographic characteristics, healthcare utilization, basis for classification (diagnosis (groups), clinical procedures), other health-related risk factors, health equity adjustment and other factors. A collaborative online tool was applied for data entry. Non-conformity of data classification across studies was addressed through group discussion. Data from three domains were comprised into tables with matched illustrations expanded in the text. Summary figures were created to synthesize the tabular results.

Results

Our search resulted in 1,837 citation records from PubMed, CINAHL, Embase, Scopus, and PROSPERO (N = 1,166 after removing the duplications). After screening based on the inclusion and exclusion criteria, 93 articles in total published from November 2015 to November 2019 were included for full-text screening. Combined with the articles identified by Jeffery and colleagues (2019) (N = 35), Chong and colleagues (2019) (N = 16), and Yan and colleague (2018) (N = 216), we obtained 360 full texts [4-6]. After reviewing the full text, we included 25 articles for data abstraction.

A. Basic characteristics of the studies

Most studies were undertaken in North America [7-22] (USA: N = 13, Canada: N = 3). Six studies were conducted in Europe [3,23-27](Spain: N = 4, UK = 1, Netherland = 1) and 3 were conducted in Singapore [28,29].

Overall approaches to segmentation

Two major approaches for population segmentation have emerged over the years. Expert-driven approaches defined or validated segments with concordance of expert judgements prior to or after the tools were completed, while data-driven approaches employed a post-hoc statistical analysis such as clustering analysis or latent class analysis on empirical data to segment a population. Out of the 17 identified tools, 7 were completely developed from data sources, while 10 were synthesized with expert review [7,11,12,14,15,17,18,26,29,30]. An example of a data-driven approach was developed by Low and colleagues (2017). The authors first assessed whether the clustering analysis was able to generate segments of patients with unique healthcare utilization patterns and disease profiles in a public healthcare organization in Singapore. Secondly, they examined the validity of their cluster-driven segments on their discriminative properties on 4-year healthcare utilization, mortality and association with clinical chronic diseases [31]. In an example of a hybrid approach, Zhou and colleagues (2014) examined the concordance of the Senior Segmentation Algorithm (SSA) with physician clinical assessed segmentation. Six primary care physicians assigned members of their patients who were older than 65 years to a care group. Physicians were aware of the SSA-assigned care group, and the physician assigned group was identical to that of the SSA in 85% of the senior panel members [7].

Segmentation population

Studies segmented populations at various levels, including at the whole population (macro) level, organization or sub-population (meso) level, and at the patient (micro) level to identify high-risk individuals. More than half of the studies focused on segmentation at the population level and gave explicit descriptions of the population they segmented (N = 15). Using a whole-level population approach, Hanley and colleagues (2010) examined the predictive validity of the adjusted clinical groups system (ACGs) and the Charlson index and reported the correlation between the predicted and observed health expenditures amongst all residents of British Columbia, Canada [19]. There was no consistent approach for examining segmentation at the sub-population or meso level. For example, some studies segmented the populations based on age criteria or type of health service utilized. Armstrong and colleagues (2012) explored the heterogeneity of all home care clients who used rehabilitation services in Ontario and examined previously unidentified clinical characteristics patterns by creating client profiles to identify different subgroups [22]. In another study, the Senior Segmentation Algorithm used administrative and clinical data from electronic medical records to identify older adults aged 65 and older with similar needs; those without chronic conditions, with one or more chronic conditions, with advanced illness or end-organ failure or with extreme frailty or nearing end of life [7]. At the micro-level, Juncosa and colleagues (1999) used data from a random sample of 1,467 patients from 13 voluntary doctor and nurse teams to classify patients using the ACGs system at the clinical level in a primary care setting in Spain [26].

The segmentation tools identified predominantly defined their populations as health plan enrollees (N = 3), utilization-defined population (N = 8), or used an entire geographical region (N = 7). García-Goñi & Ibern (2008) applied the Clinical Risk Groups (CRGs) classification system to 87,691 individualized belonging to the Serveis de Salut Integrats Baix Emporada (SSIBE), an integrated healthcare delivery organization in Catalonia, Spain. The CRGs classification system allows the segmentation of individuals in mutually exclusive categories using information from encounters between the health system and the patient and allocates a severity level to each patient[27]. Other segmentation approaches identified individuals within a geographical boundary as their population of interest. For example, Orueta and colleagues (2018) developed the FINGER (Forming and Identifying New Groups of Expected Risk) approach to a cross-sectional population of all individuals covered by the Basque public health system in Spain. Their focus was on the design of a risk stratification model based on the presence of chronic conditions and to identify individuals with the highest healthcare needs [25].

Segmentation techniques and approaches

The rapid review identified a total of 20 unique adult population health segmentation tools from 25 studies. The most common segmentation approach used Aggregated Clinical Groups (ACGs) (N = 6) [8,11,19,20,24,26], followed by Minnesota Tiering (N = 3) [9,10,15], and Clinical Risk Groups (CRGs) (N = 2) . Among the modelling approaches, K-mean aggregating was commonly used for population latent feature clustering (N = 3) [3,22,29]. Armstrong and colleagues (2012) used K-means cluster analysis to examine the heterogeneity of a complex geriatric population of rehabilitation service users in the home health care system of Ontario, Canada [22]. Other studies used machine learning approaches such as gradient boosting machine models. For example, in their hospitalization prediction model, Takahashi and colleagues (2015) considered variables such as self-reported employment, education status and availability of psychosocial health and social support in their gradient boosting machines models to predict the risk of hospitalization [9].

Data sources and objectives of the segmentation tools

The data sources for health-related features used in the segmentation tools were derived from health administrative databases (N=12 , Table 3), clinical electronic health records (N = 9), or through primary data collection using health care needs assessment tools (N=1) [23]. The majority of tools focused on predicting future health care utilization (e.g., ED visits, admissions, readmissions, hospitalizations; (N = 14, Table 2) and costs (N = 9). For example, in a 4-year longitudinal study of healthcare utilization and mortality, Lian Leng Low and colleagues (2018) generated segments of patients with unique healthcare utilization patterns within the largest public health care organization in Singapore [29]. Other tools focused on classifying patients based on illness or health needs. Eissens van der Laan and colleagues (2014) developed a tool that stratified older adults 65-101 years living in the northern part of the Netherlands based on the difficulties experienced in biopsychosocial functioning [23]. Using a person-centred approach, they note that the goal was to develop segments of older adults based on the elderly population's needs, frailty and functioning and the extent in which their needs were fulfilled, rather than on diseases, impairments and disabilities [3]. To segment the older adults according to their unfulfilled bio-psychosocial needs, the authors required variables from these different functional domains. The variables were selected from validated instruments: the Groningen Frailty indivator and the INTERMED. They selected the biological and psychosocial subscales to measure a person's felt needs in the physical and psychological domains of human functioning [23]. In other cases, the goal was to segment populations to tailor the implementation of new care intervention programs. Segmenting Medicare fee-for-service beneficiaries into major healthcare spending categories, Joynt and colleagues (2017) sought to characterize their spending by type (inpatient, outpatient, post-acute, etc.) in order to determine if there were areas that might be particularly critical for intervention that could differ by group [14].

Analytical techniques

Several studies applied existing algorithms such as the Adjusted Clinical Groups (ACGs) Chronic Disease Index (CDI), or Clinical Risk Groups (CRGs), with the first of these used most frequently (N = 6). Rather than develop new approaches to segmentation, some studies compared the predictive validity of existing algorithms. For example, Wahls and colleagues (2004) compared the predictive validity of the Adjusted Clinical Groups (ACGs) and Chronic Disease Index (CDI) in a retrospective cohort study of 31,212 primary care patients in a Veterans Health Administration network. The ACG® system is a widely used approach that uses diagnoses from healthcare claims to segment populations, while the CDI is an alternative nonproprietary approach using computerized medication data as a proxy for chronic illness [8].

Among the modeling approaches, K-means clustering was commonly used (N = 3) [3,22,29] such as in Armstrong and colleagues (2012) analysis to examine the heterogeneity of home rehabilitation service users in Ontario, Canada. One study applied a factor mixture model in which a confirmatory factor analysis and a latent class analysis were combined to investigate the common content of observed scores of items measuring the unfulfilled needs in various domains [23]. Takahashi and colleagues (2015) considered variables such as self-reported employment, education status, and availability of psychosocial health and social support in their gradient boosting machine learning model to predict hospitalization risk [9].

End-Users

The end users of the segmentation tools ranged from clinical practitioners, policymakers, healthcare providers, financial experts, and managers of care coordination programs, aligning to the purpose of tool development (Table 1). For example, Rezaeiahari and colleagues (2017) proposed a segmenting strategy to be embedded in the electronic health information system for a local community hospital. An added feature of the tool is that it could update the patient risk level at the point of care in real-time [12]. The risk-stratification tool was designed by a group of physicians at an Upstate New York hospital, which could have applications for clinicians and care managers interested in identifying high risk patients for care coordination programs [12]. Similarly, Hong and colleagues (2015) developed a predictive tool for estimated physician-identified complexity called ePDC. The purpose of the tool was to assist primary care physicians with a qualitative assessment of their patient's complexity by identifying patients with suboptimal clinical quality and high acute care utilization [18]. Some tools were developed with health policy and other decision-makers in mind. For example, Powell and colleagues (2017) developed an approach to analyzing and visualizing indicators of health service use that accounted for the temporal progression of an individual disease course. Ultimately, the goal was to help health policymakers identify and interpret patterns within the health system and facilitate data-driven and evidence-based decision-making [13].

B. Operational Characteristics

Across the range of studies and applications, the number of segments ranged widely from 3 to 534, though explicit segments were not reported in 11 of 25 studies. The segments could be defined based on the health status, health resource utilization, demographics, geography, the severity of a designated illness, or case-mix clusters defined by risk scores from hybrid models (Table 2).
Furthermore, other than a few exceptions (N = 5), most of the segmentation approaches were in static nature, failing to capture the population changes [12,15,16,20,28]. In order to capture population changes, models generally use cross-sections of data at a specific point in time and must be repeated frequently in order to capture changes over time. Powel and colleagues (2017) modelled a given individual's movement between discrete latent states of health service use, as defined by five dimensions of service uses (GPs, specialists, ED, hospitalization, or no use) [13]. Health service use data were obtained from the public health insurance provider in Quebec, Canada and applied a tool, PopHR, which uses an open cohort of approximately 1 million people created by taking a 25% random sample of the population of the Montreal census metropolitan area. PopHR is a semantic web application that helps measure and monitor population health and health system performance, integrating administrative data on health service use with data on behavioral health determinants from surveys and other sources [13].

C. Data Content

The segments were defined based on the demographic characteristics, health status, healthcare utilization, diagnoses or case-mix clusters defined by risk scores from hybrid models, additional risk factors such as the severity of a designated illness, other measures of function or need, and in a rare few cases equity dimensions (Table 2). Almost all studies reported basic demographic measures of age and sex (N = 22), previous health care utilization (N = 21) and diagnoses (N = 18). Other studies adjusted for physical function, cognitive function, morbidity needs, emotional status, mental health status, health behaviors, and geographical location. Factors that we considered related to equity and socioeconomic factors including race, education, employment, insurance status, social support eligibility, and area deprivation were uncommon.

Equity Measurement

The use of equity measures in segmentation and risk models must be carefully considered because it is important not to have measures that adjust for characteristics such as deprivation or race that may entrench inequities in access if these factors are associated with under-use (which could entrench inequities and potentially lower payments). On the other hand, if measures of equity align with particular population segments, they could be used to increase equity-oriented intervention component. Equity measures can be used as inputs when developing the segmentation models but can also be used to stratify or be reported within population segments. Measures of equity such as living in a marginalized neighbourhood – as measured by ethnic concentration, residential instability, material deprivation and income can be used as ecological measures in developing population segmentation [32]. Ecological analyses are an efficient approach to analyzing the socioeconomic status of a population and capture community-level factors that may affect access and health outcomes. Where individual data are not easily or often not routinely collected, multidimentional socio-economic deprivation indices, derived from census data and population based surveys can be used to support such ecological analyses [32]. Point-of-care collection of race, socioeconomic and related equity measures is also recommended to better understand inequity in need and service planning and provision.

Discussion

Approximately half of the tools and approaches identified in this report used data available through health administrative claims data, while nearly as many used data collected from electronic medical records. An electronic system that allows the linkage of patient variables from various settings and sources (e.g. primary care, acute care, home care and patient-reported variables, etc) is essential for comprehensive patient population segments [2]. Enhancements from deeper clinical understanding of severity may enable greater clinical specificity of segments, needs and associated intensity of interventions. However, because this type of data is less available at the population level, it is less likely to be used as a population health planning tool [33].

It is important to develop segments that can be actionable, which means that the quality and quantity of information collected across the health care system should be meaningful and interpretable by providers and system administrators who are ultimately responsible for deciding on intervention, care and resource allocation. There is also a need to minimize potential burdens of data collection to clinicians who might already be spending a significant time on electronic medical record data entry as part of their routine clinical practice. Decision-makers and clinician managers can tailor data entry fields in electronic medical records, which have the dual benefit of enabling population segmentation, while reducing clinical workload [2].

The types and quantity of data elements collected depend on whether the stakeholder is interested in conducting health service assessment and/or planning at the micro-, meso- and macro-level. For example, in Ontario, the interRAI-HC, a routinely administered clinical assessment tool in home care and long-term care is intended for individual-level care planning. Segmentation at the clinical or micro-level requires granular information such as an individual's ability to bathe oneself [2]. Health assessment data can also be aggregated to meso-level measures of a home care population's ability to perform activities of daily living, which can enable planning for the overall level or allocation of home care supports [34]. Contrary to other systematic reviews which found that most population segmentation was undertaken at the meso-level, many of the segmentation tools identified in this review identified approaches at the macro-level and segmented whole populations. For example, clustering patients by ACGs provides a summative characteristic of the population based on routinely collected health administrative data, requiring less time consuming and intensive data collection. On the other hand, macro-level segments cannot fully meet the needs of clinicians developing services at the individual level, and therefore cannot be completely substituted for meso- or micro-level variables [2].

This report recommends that clinical, social and functional factors could be included in segmentation approaches (as available) to determine need for medical and other support services. Additionally, equity measures must be included to identify subgroups where equity-oriented interventions may be needed. Understanding the population health care needs in relation to availability of services and their effectiveness to meet these needs can promote resource redistribution to optimize population health. Well-matched services with needs will be associated with improved efficiency and equity in a health care system. Failure to meet the health care needs of patients can lead to worse clinical outcomes and may potentially increase health service utilization in the long-term.

Population data that includes health needs assessment from sources other than health care claims allows segmentation to consider an entire population including under-serviced individuals without realized access to health care. Segmentation of the population facilitates the efficient development of services that are based on common sets of needs associated with each segment. Understanding the distinct needs of population segments may help ease the identification of unmet needs by comparing services recommended to individuals in a segment with the typical service packages received by patients in that segment. A large proportion of segmentation tools were developed and validated against their ability to predict future health care utilization. However, it is unclear whether such utilization is a measure of need, or the outcome of unmet need in for lower intensity, lower cost healthcare services. By segmenting a population-based on health service needs, a more upstream approach can be adopted.

Tracking individual and population transitions and changes over time also provide an important clinical and system planning tool. In this scenario, interventions can be designed across segments and the outcomes of the interventions can be evaluated based on their rate of progression to higher needs population segments. Unfortunately, most of the segmentation tools identified in this review were not dynamic, in that they did not capture changes in the population over time. Population segmentation can also be used when evaluating interventions to determine whether there are subgroups that experience greater or less effectiveness. Given the greater calls for new means of evaluating healthcare interventions' ability to meet health care needs, it is imperative for new segmentation tools to be able to quantify intervention effectiveness as a reduction in the probability of progression to higher or more severe population segments [2]. This would improve evaluation over the common practice of measuring outcomes of health care interventions through variables such as incidence of hospitalization and disease biomarkers [2].

The focus of segmentation for macro-level populations in our review is of interest as it suggests that through this approach, different care needs can be identified, and policies and budgets can be tailored for homogenous patient groups. Examples from our review include Kaiser Permanente [7] and SigHealth [31], which aimed to segment all the population covered by these integrated care organizations. Meso-level integrated care models can use segmentation to choose specific subpopulations for their interventions. Segmentation tools such as the Senior Segmentation Algorithm help prioritize care coordination for care groups with specific needs or characteristics [7]. Micro-level integration, which often occurs at the clinical or patient-level requires high-risk patients to be identified at the point of care. With a common data platform, information from population segmentation can inform tools used in clinical practice at the point-of-care, but most importantly, they can be used to transform clinical care delivery. For clinicians, population segmentation tools can be used to guide approaches to population health management at the meso or macro level. The segmentation tool developed by Rezaeiahari and colleagues was embedded in the electronic health information system of a local community hospital, with the added feature of updating the patient risk level at the point of care at the real-time [12]. Such tools can enable risk-stratification at the meso- or clinical-level and could have applications for clinicians and care managers interested in identifying high risk patients for care coordination programs [3].

Limitations

This rapid review did not collect data from the grey literature such as government or institutional reports regarding their development or use of segmentation tools, or promotional material from developers or distributors of segmentation tools. While this exclusion may have missed useful tools, it also ensured greater disclosure of the content of underlying contributing factors and uses. On the other hand, a number of included peer-reviewed articles put more focus on explaining the technical logistics of the segmentation tools but were ambiguous about the users at the front line can adopt these tools. We also found the included studies did not clearly distinguish the purposes for tool development (i.e., population management or clinical use).

Conclusion

There is a substantial literature in population segmentation tools, although the application of these tools in practice is less well described. Tools developed through data-driven approaches appear to have relatively low interpretability. Tools that derived from population administrative or electronic health databases and coded through existing validated algorithms have the greatest opportunity to enable inclusion of an entire population (or at least the population that have encountered a health or social care service provider that contributes this information to a common linked database). The clinical utility of such approaches however is less certain as the focus for aggregating conditions may be too broad or abstract to be meaningful to potential users. Clinically-derived tools have the opportunity to be useful in practice but may not include sufficient information to attribute an entire population. Need-based measurement and segmentation are an optimal resolution but face a challenge of sharing information back to front line providers. While our review did not clearly identify a tool developed from a health needs assessment, a number of tools did appear to have a need-oriented clinical applicability.

Recommendation

We recommend that OHTs examine both a tool based on an existing validated algorithm and a tool derived from a needs-oriented perspective to assess the relative merits of each in the Ontario population and the potential for transition to a needs-based approach to care. We recommend that each be applied using OHT attributed populations and built upon Ontario's existing population health administrative databases supplemented as possible by other data sources such as surveys or functional clinical assessments. The most common tool reported in this review is the ACG® system which has been deployed in some Canadian jurisdictions including Ontario. Although this is a proprietary tool from a US-based organization and may have incremental costs for use, an ACG® or equivalent system would be an important consideration. There is one example of a new, possibly equivalent, segmentation tool developed by the Canadian Institute for Health Information that has provided some early indication of predictive validity [35]. A few tools in our review have a stated purpose of having groups to match services to needs and have segment descriptions that are aligned with service offerings such as high-intensity treatment clinics or patients with social or mobility needs. These would serve as a useful needs-oriented example.

Reference

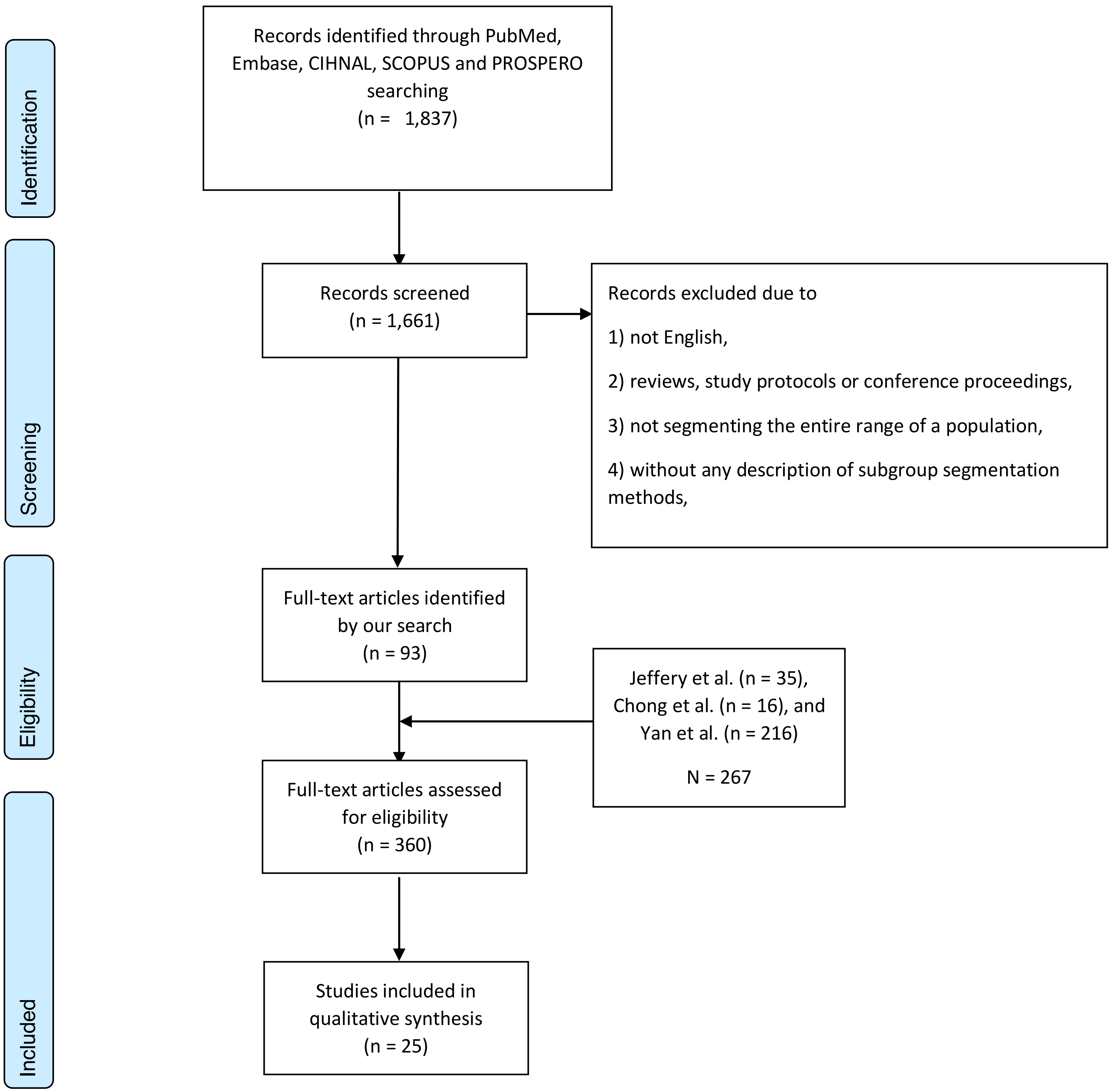
1. Team OH. RISE brief 6: Population-health management. https://www.mcmasterforum.org/docs/default-source/rise-docs/rise-briefs/rb6_population-health-management.pdf?sfvrsn=8. Published 2019. Accessed December 17, 2020.
2. Chong JL, Matchar DB. Benefits of population segmentation analysis for developing health policy to promote patient-centred care. *Ann Acad Med Singapore.* 2017;46(7):287-289.
3. Vuik SI, Mayer EK, Darzi A. Patient segmentation analysis offers significant benefits for integrated care and support. *Health affairs.* 2016;35(5):769-775.
4. Jeffery AD, Hewner S, Pruinelli L, et al. Risk prediction and segmentation models used in the United States for assessing risk in whole populations: a critical literature review with implications for nurses' role in population health management. *JAMIA open.* 2019;2(1):205-214.
5. Yan S, Kwan YH, Tan CS, Thumboo J, Low LL. A systematic review of the clinical application of data-driven population segmentation analysis. *BMC medical research methodology.* 2018;18(1):121.
6. Chong JL, Lim KK, Matchar DB. Population segmentation based on healthcare needs: a systematic review. *Systematic reviews.* 2019;8(1):202.
7. Zhou YY, Wong W, Li H. Improving care for older adults: a model to segment the senior population. *The Permanente Journal.* 2014;18(3):18.
8. Wahls TL, Barnett MJ, Rosenthal GE. Predicting resource utilization in a veterans health administration primary care population: comparison of methods based on diagnoses and medications. *Medical Care.* 2004:123-128.
9. Takahashi PY, Ryu E, Olson JE, et al. Health behaviors and quality of life predictors for risk of hospitalization in an electronic health record-linked biobank. *International journal of general medicine.* 2015;8:247.
10. Takahashi PY, Heien HC, Sangaralingham LR, Shah ND, Naessens JM. Enhanced risk prediction model for emergency department use and hospitalizations in patients in a primary care medical home. *Am J Manag Care.* 2016;22(7):475-483.
11. Starfield B, Weiner J, Mumford L, Steinwachs D. Ambulatory care groups: a categorization of diagnoses for research and management. *Health services research.* 1991;26(1):53.
12. Rezaeiahari M, Poranki S, Khasawneh M. Predicting emergency department visits among patients receiving primary care using risk scores. Paper presented at: IIE Annual Conference. Proceedings2017.
13. Powell G, Luo YT, Verma A, Stephens DA, Buckeridge DL. Multidimensional and Longitudinal Indicators in Population Health. Paper presented at: AAAI Workshops2017.
14. Joynt KE, Figueroa JF, Beaulieu N, Wild RC, Orav EJ, Jha AK. Segmenting high-cost Medicare patients into potentially actionable cohorts. Paper presented at: Healthcare2017.
15. Johnson TL, Brewer D, Estacio R, et al. Augmenting predictive modeling tools with clinical insights for care coordination program design and implementation. *eGEMs.* 2015;3(1).
16. Johnson TJ, Jones A, Lulias C, Perry A. Practice innovation, health care utilization and costs in a network of federally qualified health centers and hospitals for Medicaid enrollees. *Population Health Management.* 2018;21(3):196-201.

17. Hughes JS, Averill RF, Eisenhandler J, et al. Clinical Risk Groups (CRGs): a classification system for risk-adjusted capitation-based payment and health care management. *Medical care.* 2004:81-90.
18. Hong CS, Atlas SJ, Ashburner JM, et al. Evaluating a model to predict primary care physician-defined complexity in a large academic primary care practice-based research network. *Journal of general internal medicine.* 2015;30(12):1741-1747.
19. Hanley GE, Morgan S, Reid RJ. Explaining prescription drug use and expenditures using the adjusted clinical groups case-mix system in the population of British Columbia, Canada. *Medical care.* 2010:402-408.
20. Haas LR, Takahashi PY, Shah ND, et al. Risk-stratification methods for identifying patients for care coordination. *American Journal of Managed Care.* 2013;19(9):725-732.
21. Bertsimas D, Bjarnadóttir MV, Kane MA, et al. Algorithmic prediction of health-care costs. *Operations Research.* 2008;56(6):1382-1392.
22. Armstrong JJ, Zhu M, Hirdes JP, Stolee P. K-means cluster analysis of rehabilitation service users in the home health care system of Ontario: Examining the heterogeneity of a complex geriatric population. *Archives of physical medicine and rehabilitation.* 2012;93(12):2198-2205.
23. van der Laan ME, van Offenbeek M, Broekhuis H, Slaets J. A person-centred segmentation study in elderly care: towards efficient demand-driven care. *Social science & medicine.* 2014;113:68-76.
24. Sicras-Mainar A, Velasco-Velasco S, Navarro-Artieda R, et al. Adaptation Of The Categories Johns-Hopkins ACG Case-Mix System In Primary Healthcare: A Longitudinal, Retrospective Claim Database Study. 2011.
25. Orueta JF, García-Alvarez A, Aurrekoetxea JJ, García-Goñi M. FINGER (Forming and Identifying New Groups of Expected Risks): developing and validating a new predictive model to identify patients with high healthcare cost and at risk of admission. *BMJ open.* 2018;8(5).
26. Juncosa S, Bolibar B, Roset M, Toma R. Performance of an ambulatory casemix measurement system in primary care in Spain: ambulatory care groups (ACGs). *The European Journal of Public Health.* 1999;9(1):27-35.
27. García‐Goñi M, Ibern P. Predictability of drug expenditures: an application using morbidity data. *Health Economics.* 2008;17(1):119-126.
28. Saxena N, You AX, Zhu Z, et al. Singapore's regional health systems—a data‐driven perspective on frequent admitters and cross utilization of healthcare services in three systems. *The International journal of health planning and management.* 2017;32(1):36-49.
29. Low LL, Yan S, Kwan YH, Tan CS, Thumboo J. Assessing the validity of a data driven segmentation approach: A 4 year longitudinal study of healthcare utilization and mortality. *PloS one.* 2018;13(4):e0195243.
30. Yan S, Seng BJJ, Kwan YH, et al. Identifying heterogeneous health profiles of primary care utilizers and their differential healthcare utilization and mortality–a retrospective cohort study. *BMC family practice.* 2019;20(1):54.
31. Low LL, Kwan YH, Liu N, Jing X, Low ECT, Thumboo J. Evaluation of a practical expert defined approach to patient population segmentation: a case study in Singapore. *BMC health services research.* 2017;17(1):771.
32. Anderson G, Frank JW, Naylor CD, Wodchis W, Feng P. Using socioeconomics to counter health disparities arising from the covid-19 pandemic. *BMJ.* 2020;369.

33. Healthcare OB. Segmentation for Outcomes - Segmentation approaches for outcomes based commissioning. https://outcomesbasedhealthcare.com/wp-content/uploads/2018/12/OBH-Part-1-Segmentation-for-Outcomes.pdf. Published 2016. Accessed December 17, 2020.
34. Mello JDA, Hermans K, Van Audenhove C, Macq J, Declercq A. Evaluations of home care interventions for frail older persons using the interRAI home care instrument: a systematic review of the literature. *Journal of the American Medical Directors Association.* 2015;16(2):173. e171-173. e110.
35. Li Y, Weir S, Steffler M, Shaikh S, Wright JG, Kantarevic J. Using Diagnoses to Estimate Health Care Cost Risk in Canada. *Medical care.* 2019;57(11):875.

Figures and Tables

Figure S1. PRISMA diagram



Search Term Sample Input

CINAHL

MW communit* OR MW population*
AND
MW typolog* OR MW stratification* OR MW segmentation* OR MW class* OR MW categorization* OR MW categorisation* OR MW group* OR MW risk* OR MW profil* OR MW cluster* OR MW cluster analysis OR MW pattern*
AND
MW health management OR MW population management OR MW population health OR MW care management OR MW community engagement OR MW facility management OR MW health care quality OR MW health care evaluation OR MW health care delivery OR MW health policy OR MW health administration OR MW health practice
AND
MW model* OR MW tool* OR MW guideline* OR MW method* OR MW approach* OR MW technique* OR MW algorithm* OR MW implement* OR MW device* OR MW instrument* OR MW strateg* OR MW application*
Limiters - Published Date: 20161101-

Pubmed

(((communit*[Title/Abstract] OR population*[Title/Abstract]) AND (typolog*[Title/Abstract] OR stratification*[Title/Abstract] OR segmentation*[Title/Abstract] OR classification*[Title/Abstract] OR categorization*[Title/Abstract] OR “risk management”[Title/Abstract] OR “risk grouping”[Title/Abstract] OR “risk profiling”[Title/Abstract] OR cluster*[Title/Abstract] OR “risk pattern”[Title/Abstract]) AND ("Population Health Management"[MeSH Terms] OR "public Health administration"[MeSH Terms] OR "public health practice"[MeSH Terms] OR "public policy"[MeSH Terms] OR “health management”[Title/Abstract] OR “population management”[Title/Abstract] OR “care management”[Title/Abstract] OR “population health”[Title/Abstract] OR “community engagement”[Title/Abstract] OR “facility management”[Title/Abstract]) AND (model*[Title/Abstract] OR tool*[Title/Abstract] OR guideline*[Title/Abstract] OR approach*[Title/Abstract] OR technique*[Title/Abstract] OR algorithm*[Title/Abstract] OR implement*[Title/Abstract] OR device*[Title/Abstract] OR instrument*[Title/Abstract] OR application*[Title/Abstract] OR strateg*[Title/Abstract]) AND data[Title/Abstract])) AND ("2017/01/01"[Date - Publication] : "3000"[Date - Publication])

EMBASE

(communit* or population*).mp. [mp=title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword, floating subheading word, candidate term word]
AND
(typolog* or stratification* or segmentation* or classification* or categorization* or categorisation* or (risk management) or (risk grouping) OR (risk profiling) OR cluster* OR (risk pattern)).mp. [mp=title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword, floating subheading word, candidate term word]
AND
((health adj3 management) or (population adj3 management) or (population adj3 health) or (care adj3 management) or (community adj3 engagement) or (facility adj3 management) or (health care adj3 quality adj3 evaluation) or (deliver* adj3 healthcare) or (population adj3

health adj3 management) or (health adj3 administration) or (health adj3 practice) or (health adj3 policy)).mp. [mp=title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword, floating subheading word, candidate term word]

AND

(model* or tool* or method* or approach* or technique* or application* or strateg* or algorithm* or implement* or instrument*).mp. [mp=title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword, floating subheading word, candidate term word]

AND

(data).mp. [mp=title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword, floating subheading word, candidate term word]

limit to yr="2017"

Scopus

(TITLE-ABS-KEY (typolog*) OR TITLE-ABS-KEY (stratification*) OR TITLE-ABS-KEY (segmentation*) OR TITLE-ABS-KEY (classification*) OR TITLE-ABS-KEY (categorization*) OR TITLE-ABS-KEY (categorisation*) OR TITLE-ABS-KEY (risk management) OR TITLE-ABS-KEY (risk grouping) OR TITLE-ABS-KEY (risk profiling) OR TITLE-ABS-KEY (cluster) OR TITLE-ABS-KEY (risk pattern))

AND

(TITLE-ABS-KEY (model*) OR TITLE-ABS-KEY (tool*) OR TITLE-ABS-KEY (guideline*) OR TITLE-ABS-KEY (method*) OR TITLE-ABS-KEY (approach*) OR TITLE-ABS-KEY (technique*) OR TITLE-ABS-KEY (*application* *) OR TITLE-ABS-KEY (strateg*) OR TITLE-ABS-KEY (algorithm*) OR TITLE-ABS-KEY (instrument*) TITLE-ABS-KEY (device*) OR TITLE-ABS-KEY (implement*))

AND

(TITLE-ABS-KEY ("health management") OR TITLE-ABS-KEY ("population management") OR TITLE-ABS-KEY ("population health") OR TITLE-ABS-KEY ("care management") OR TITLE-ABS-KEY ("community engagement") OR TITLE-ABS-KEY ("facility management") OR TITLE-ABS-KEY (("health care" AND ("administration" OR "access" OR "evaluation" OR "delivery"))) OR TITLE-ABS-KEY ("population health management") OR TITLE-ABS-KEY ("public health practice") OR TITLE-ABS-KEY ("health policy"))

AND

(TITLE-ABS-KEY (communit*) OR TITLE-ABS-KEY (population*))

AND

TITLE-ABS-KEY (data)

PUBYEAR > 2016

Table 1. Study characteristics

No.	Year	Country	Study Population	Data sources	Name/ technique of the tool	Purpose of the tool	End-users	Ability to capture changes in population
16	2018	USA	All Medicaid enrollees who were members of Medical Home Network	Claims-based data from the Illinois Department of Healthcare and Family Services	Not specified risk score	Calculate the risk-adjusted utilization and cost of Medicaid enrollees	Researchers, analysts	Risk scores were used to adjust utilization and cost measures for both cohorts and years
8	2004	USA	Patients in a Veterans Health Administration (VA) network who received outpatient medication prescriptions or had VA utilization	Pharmacy Benefits Management database (the Patient Treatment File, and the Outpatient Care File). Outpatient prescriptions filled in VA pharmacies.	Comparison of ACG and CDI	Predict patient utilization	Physician, policy makers for resource allocation	Not specified
29	2018	Singapore	All adult patients who utilized services in SingHealth	Regional administrative health data	Hierarchical clustering analysis (Ward's linkage) and K-means cluster analysis	Segment a regional health system patient population	Policy makers	Not specified
22	2012	Canada	Clients who received rehabilitation services within the first 3 months of their initial home care assessment	Not specified	K-mean clustering	Examine the heterogeneity of home care clients who use rehabilitation services	Front-line providers and data available to researchers	Not specified
7	2014	USA	Kaizer Permanente (KP) North West and KP Hawaii	Electronic health record within KP health connect	Senior Segmentation Algorithm	Identify and address distinct health care profiles and priorities of different groups comprising it	Clinical decision-makers, healthcare system manager	Not specified
21	2020	USA	Individuals from a commercially insured population across the country	Medical and pharmaceutical claims data	Data-Mining Methods with classification tree and clustering	Provide prediction of health care costs	Health care managers and decision makers	Not specified
23	2014	Netherland	Older adults living in the Northern part of the Netherlands	Patient sampling data from 25 diverse healthcare, welfare organizations and elderly associations	Factor Mixture Model	Utilize the elderly segmentation as a first triage step	Care providers, decision makers and policy makers	Not specified
14	2017	USA	All high cost patients in Medicare fee-for-service population	Medicare claims files, including the Medicare Beneficiary Denominator and Enrollment Database	Not specified	Identify all high-cost patients in Medicare fee-for-service population and define six mutually exclusive subpopulations, characterize their spending by type in order to determine if there were areas that might be particularly critical for intervention	Decision makers and policy makers	Not specified

18	2014	USA	Adult patients receiving primary care within the Massachusetts General Hospital	Cohort data from an electronic data repository containing demographic, clinical, appointment, and billing data	Predictive model to estimate physician-defined complexity (ePDC)	Evaluate physician-defined complexity prediction model against outpatient Charlson score and a commercial risk predictor	Health system: risk prediction; predict suboptimal clinical quality outcomes and future acute care utilization	Not specified
25	2017	Spain	All individuals covered by the Basque public health system	Primary care electronic health records, hospital discharge reports, electronic records from day hospitals and from visits to emergency departments and specialized care	FINGER (Forming and Identifying New Groups of Expected Risks)	Characterize patients by chronic disease groups	Assessment scale for clinicians	Not specified
15	2015	USA	Adult patients who receive primary care at a Denver Health primary care clinic	Billing claims data, Chronic Illness and Disability Payment System	Denver Health's 21st Century Care Project	Conduct risk and financial stratification for tiered care coordination. The numeric "risk score" calculated as individual risk in relation to the average risk of future spending.	Multidisciplinary team including clinical directors, health services researchers, clinical operations staff, finance experts, primary care providers, and quality improvement experts.	The dynamics is defined through monthly runs of a population attribution and risk tiering algorithm
17	2004	USA	Patients from Medicare, Medicaid, and a privately insured population	Claims for inpatient care, hospital-based outpatient care, hospice care, skilled nursing facility, physician office, and ancillary services	Clinical Risk Groups (CRGs)	Claims-based classification system for risk adjustment to predict future use of healthcare resources.	Health planners interested in risk adjustment for capitated payment systems and management systems that support care pathways	Not specified
24	2012	Spain	Primary healthcare patients at 13 Catalonia health centers	Computerized medical records	Adjusted Clinical Groups (ACG) system	Examine the clusters in relation to the burden of disease to consumption of resources and the cost of care.	Health managers who support mechanism to allocate health resources	Not specified
19	2010	Canada	British Columbia residents who registered for benefits under the province's public health insurance plan	Population administrative databases describing prescription drug use, demographics, household income, and diagnostic information	Adjusted Clinical Group (ACG) system; Charlson index	Explained and predicted prospective expenditures on prescription	Health system planners	Not specified
9	2015	USA	Mayo Clinic patients enrolled in Employee and Community Health program	Mayo Clinic electronic health record, self-administered questionnaires at enrollment	Gradient boosting machine (GBM) model	Determine if quality of life or health behaviors captured in an EHR-linked biobank can predict future risk of hospitalization	Clinical practitioners and medical providers	Not specified
11	1990	USA	Patients insured with Columbia Medical Plan in Maryland	Database from Columbia Medical Plan in Maryland	Ambulatory care groups (ACGs)	Measure and compare burden of illness of patients overtime in different ambulatory care facilities; predict utilization and charges	As "case mix" tool for health providers; as management tool for health planners	Not specified

10	2016	USA	Patients from an academic medical center in Rochester, Minnesota	Electronically abstracted from the electronic medical record and administrative databases	enhanced Minnesota tiering model	Identify population at highest risk of hospitalization and/or ED visit	Care coordination programs	Not specified
20	2013	USA	Patients over 18 enrolled in Employee and Community Health program	Clinic EHRs and administrative databases within Mayo Clinic	A Hybrid Model with Adjusted Clinical Groups (ACG), Minnesota Tiering, Hierarchical Condition Categories (HCC), Chronic Condition Count, Elder Risk Assessment	Predict future healthcare utilization	healthcare coordinators and managers of care coordination programs	Risk to a specific outcome of an individual patient will be updated every time the patient visit the clinic
27	2008	Spain	Individuals belonging to a integrated healthcare delivery organization	Pharmaceutical consumption (outpatient, specialist, and hospital) data and the morbidity profiles in a publicly funded healthcare system	Clinical Risk Groups (CRGs)	Control the drug expenditure and provide incentives for efficiency in the use of pharmaceutical benefits	policy makers and healthcare managers, Healthcare delivery organization	Not specified
13	2017	Canada	A random sample of the population from Montreal census metropolitan area	An administrative claims data for patients with a previous diagnosis of chronic obstructive pulmonary disease in Montreal	A hidden Markov model	Facilitate decision makers to identify the areas of concern, predict future disease burden, and implement appropriate policies.	regional decision maker	Not specified
12	2017	USA	All adult patients receiving primary care at a community hospital	Clinical EHRs from a community hospital	ED visit risk score	provide an easy-to-implement and effective tool to identify patients based on additional factors other than demographics and comorbidity	managers of care coordination program, physicians	The proposed tool is a real-time updated algorithm at the point of care
26	1999	Spain	The study population was based on 13 voluntary doctor and nurse teams	Hospital EHRs from 14 primary care settings	Ambulatory Care Groups (ACGs)	Analyze the case-mix groups obtained, describe the explanatory power vis-il-vis variability in the measurement of resource availability and use and compare the results obtained with the authors' own initial results from application in an ambulatory care setting.	Decision maker of primary care systems	Not specified
30	2019	Singapore	Adult Singapore citizens or permanent residents who utilized Singapore Health Services	Singapore Health Services Regional Health System administrative database	Latent class analysis (LCA) model	Segmented the heterogeneous population of primary care utilizers into six patient classes with distinct disease patterns, demonstrated the derived classes have predicative ability on mortality and long term healthcare utilization.	health policy makers, community-based service providers	Not specified

28	2015	Singapore	Patients from three Reginal Health Service over 6 years who have used any of the nine public primary care	Unspecified linked patient administrative databases and disease registry data	a data-driven tool on frequent admitters and cross utilization of healthcare services	Deliver information in regard to population health management. Each regional health service needs to understand its population, prevalence of risk factors and drivers of high utilization and cost.	population health managers, interventions programs designers	A transition matrix was constructed using an annual cohort of inpatients to determine the annual number of readmissions or death.
3	2016	UK	The study used a random sample of 300,000 patients to reflect a general local population	Linked databeses including primary care records in the Clinical Practice Research Datalink, acute care information	utilization-based cluster analysis with k-mean	Segment the patient population into distinct groups with unique care priorities, provide a quantitative evidence base to improve population health. Segments lower-needs populations to inform preventive interventions. The identification of different care user types provides insight into needs.	population health managers, policy makers	Not specified

Table 2. Operational characteristics

No.	Number of segments	Segment description	Primary outcomes	Purpose for segmentation	Method of delivery of segmentation information to user	Actionability of segmentation results
16	Not specified	Derived risk scores	Health care utilization and costs	Determined health care utilization and costs	Electronic medical records accessible to providers	Not actioned in the study.
8	104	1 of 104 mutually exclusive ACG groups	Healthcare resource utilization rate	Predictive outpatient clinic visits and days of hospital care	Electronic medical records and computerized formats	Not actioned in the study.
29	3	1) "Young, healthy", 2) "Middle age, healthy", 3) "Complicated chronic disease"	Population segments	Organize health services around segments of patients with similar healthcare needs	Not specified	Results can be used to facilitate the policy makers' development of population health policy strategies and design of targeted healthcare service packages that meet each segment's specific needs.
22	7	7 clusters containing home care client population generated from K-mean	Utilization rate	Predict resource utilization	Not specified	Researchers can use cluster analysis within large administrative databases to focus on pattern discovery in both a general fashion or in a more targeted fashion focusing on specific domains of interest.
7	3	1) patients in the absence of illness, 2) patients need more emphasis on disease management services, 3) patients were more complex and require approaches beyond disease management	Utilization rate	Predict resource utilization	Not specified	Segmentation results provide a foundation for individualized assessment for patient-centered care. The tool is intended to ensure that individualized needs of all patients in each group are met by informing clinical decision making.
21	5	1) low, 2) emerging, 3) moderate, 4) high, 5) and very high risk of medical complications	Utilization rate	Predict resource utilization	Not specified	Not actioned in study.
23	5	1) physical needs, 2) psychological needs, 3) social needs, 4) mobility needs, and 5) cognition needs.	Healthcare utilization per elderly segments	1) identification and description of robust, person-centered groups, 2) evaluation of the alignment between the resulting segments' experience	The information can be accessed through triage systems	Providers access broad patient needs. In a second step, comprehensive assessment is conducted focusing on the typical difficulties experienced for a specific segment.

14	Not specified	1) presence of end-stage renal disease or disability, 2) presence of at least two conditions, 3) beneficiaries >65 with chronic illness 4) all other relatively health beneficiaries	Clinically meaningful subgroups with different spending profiles	Identify high-cost patients, define mutually exclusive high-cost and non-high-cost subpopulations	Not specified	Resource allocation
18	Not specified	Risk predictor incorporating level of complexity of the disease	Utilization rates of primary care and emergency department	Categorize population and compare results between physician defined complexity	Not specified	Not actioned in study.
25	Not specified	Individual risk score	Patients clusters based on diagnoses of long-term health problems	Stratify the patient population to identify high-risk patients based on clinical criteria	Not specified	Not actioned in study, but suggest tool be used to identify patients who may be candidates for specific interventions. The tool could be used to describe the burden of morbidity and of certain health problems.
15	4	1) high intensity treatment clinics, 2) complex case management, 3) chronic Disease management, 4) panel management	Assess individual risk in relation to the average risk of future spending	Segment populations into tiers to match services based on individuals needs within tiers	Not specified	Offer all patients text message reminders about appointments and recommended preventive services.
17	534	534-episode diagnostic categories.	Use of healthcare resources	Categorize each individual to risk group to optimize the healthcare resource allocation	Not specified	Not actioned in study but suggest categorizing patients according to their risk of debility and expected future resource use for risk management.
24	106	ACG tool derived population segments	Cost of care	Assess poorly performing and highly variable ACGs according to the cost of care	Not specified	Not actioned in study.
19	Not specified	ACG tool derived population segments	Predict pharmaceutical expenditures	Cluster medical condition and cost-defined groups, predict the use of medical utilization	Not specified	Not actioned in study.
9	3	Segment by self-perceived health status 1) fair or poor, 2) good, 3) very good	Hospitalization during the next 12 months and related cost	Predict hospital utilization	Not specified	Not actioned in study, but suggest adding self-rated health to EHRs, which could be assessed at each patient encounter and trigger EHR alert system to medical providers
11	34	Thirty-four ambulatory diagnostic groups	Expected persistence or recurrence of the condition over time	Measure the burden of illness of patients overtime in different ambulatory care facilities, predict utilization and charges	Not specified	Not actioned in study, but suggest assessing medical practice variations and utilization levels by ACGs; setting rates for capitation

10	Not specified	Individual risk scores	Combined binary outcome of hospitalization and/or ED visit	Improve the organization of care delivery and management of complex patients	Not specified	Not actioned in study.
20	Not specified	Hybrid clusters	Inpatient visits, ED visits, 30-day readmission, high-cost user	Develop an optimal method to guide the implementation of highly cost-effective care coordination programs	Risk scores will be generated at the time of patient encounter	Not actioned in study, but this tool offers valuable information for providers and health plans who undertake case management
27	Not specified	Codified health encounters groups	Drug expenditure	Propose risk adjustment tool in setting the premiums for pharmaceutical benefits	Investigator present and interpreted the results to policymakers and healthcare managers	Not actioned in study.
13	57	57 local community service clusters	Aggregates of yearly health service use	Help decisionmakers to identify and interpret patterns within the health system and facilitate decision making	Analyze and visualize of health indicators to provide a foundation for information displays	Not actioned in study, but this study identifies regional clusters based on patterns across four health service indicators
12	Not specified	Risk strata was defined by the risk scores	ED visits	Validate the performance of the risk scores obtained from the developed tool in identifying the ED visits	This tool captures the level of risk of patient who are already in the contact by the system, and the predictive results in terms of ED visits will be shared with care coordinators	Not actioned in study.
26	51	51 ACG derived case mix clusters	The number of visits, the number of episodes, PC cost and total cost	Depict the resource use characteristics to inform of future situation	Not specified	Not actioned in study, the performance of the groups was evaluated in relation to their ability to explain the variability of several measures of resource use in the multivariant models
30	6	1) Relatively healthy, 2) Stable metabolic disease, 3) Metabolic disease with vascular complications, 4) High respiratory burden, 5) High metabolic disease without complication, 6) Metabolic disease with end-organ failure	Healthcare utilizations and mortality	Report the disease patterns, assess the predictive ability of class membership on one-year follow up healthcare utilizations and mortality	The utilization and outcome differentiation could be explained on a class basis with medical sense.	Action through a six-step process involving needs assessment, definition of proximal program objective matrices, selection of theory-based methods and practical strategy, production of program components and design, program adoption and implementation plan, and finally evaluation plan.

28	Not specified	Population segments were generated on the basis of frequent admitters for 4 hospital departments	One-year readmission and mortality	Quantify health service utilization, predict patient readmission or death	The paper provides the reader with a thorough understanding of the assimilated regional health service data with a special focus on the high resource utilizers, the frequent admitters.	Not actioned in study, but the comprehensive analytical results of the population health service utilization data were presented using figures and tables in this study as a report format.
3	8	Eight clusters were generated on the basis of care utilization, demographic and chronic diseases	ED visit, mortality	Explore the potential value of using utilization-based cluster analysis to segment a general patient population	The segmentation results were visualized in a diagram, clearly demonstrating demographics, the cost, care utilization features of population in each segment	Policymakers can adopt the analytical results to develop population health strategy that considers both care and prevention to delivers interventions tailored to the segments' needs.

Table 3. Data content

No.	Health status	Socio-demo characteristics	Healthcare utilization	Medication/ diagnosis/ clinical procedure	Identified risk factors	Health equity adjustment	Other factors
16	Patients with any conditions	Age, sex	Inpatient visits, emergency department visits, pharmacy costs, primary care visits, readmission	Not specified	Medical and pharmacy risk score	No	Not specified
8	Patients with any conditions	Age, sex	Outpatient visits, primary care, specialty care, and ancillary care; emergency department visits; inpatient visits and; length of stay	Not specified	Risk scores from model output	No	Not specified
29	Chronic disease status	Not specified	Inpatient visits, specialist outpatient visits, emergency department visits, primary care visits	Not specified	Not specified	No	Mortality
22	Changes in Health, end-stage disease, symptoms, and signs of disease	Age	Not specified	Not specified	Not specified	No	Physical function (instrumental activities of daily living)
7	Chronic disease status	Not specified	Hospital utilization	Not specified	Not specified	No	Not specified
21	Patients with any conditions	Age and sex	Medical services utilizations	ICD-9-CM diagnostic codes, diagnostic clusters (DRG), procedures, supplies, products and services codes (HCPCS, CPT4)	Not specified	No	Not specified
23	Patients with any conditions	Not specified	Adjust for chronicity, diagnostic dilemma, severity of illness, diagnosis	Not specified	Not specified	Adjust for difficulties in psycho-social coping	Mobility needs, able to do activities independently, cognitive function (Memory complaints)
14	presence of end-stage renal disease or disability, frailty, chronic illness	Age	Inpatient visits, ambulatory visits, emergency department visits, durable medical equipment, post-acute/rehabilitative/long-term/hospice care, and pharmaceutical spending	Not specified	Not specified	No	Not specified
18	complex patient	Age, marital status, insurance	Clinic visits, no-show appointments, urgent care visits, primary care visits,	multiple complex diagnoses, Comorbidity score, diabetes, COPD, number of e-prescribed medications	last hemoglobin A1c >9, Warfarin prescribed, opiate prescribed, computed tomography scan procedure, magnetic resonance imaging procedure	patient needs across medical, social, behavioral, and environmental dimensions	Not specified

HSPN

25	chronic conditions	Age, sex	Resource utilization, emergency department visits, prolonged hospital stay, mortality	ICD-9-CM diagnostic codes	Not specified	No	Not specified
15	chronic illness and disability	Age, sex	Financial stratification	diagnostic clusters from Clinical Risk Groups (CRGs)	Not specified	No	Not specified
17	Chronic, acute, and manifestations of chronic disease	Yes, but not specified	Medical expenditures, claims for inpatient care, hospital-based outpatient care, hospice care, skilled nursing facility, physician office, and ancillary services	ICD-9-CM diagnostic codes	Not specified	No	Not specified
24	Patients with any conditions	Age, sex	Mean number of episodes, cost of primary healthcare, care provider, clinical service	ICD-9-CM diagnostic codes, comorbidity index	risk scores from model output	No	Not specified
19	Patients with any conditions	Age, sex	Hospital separations, and physician paid claim records	ICD-9-CM and ICD-10 diagnostic codes	Not specified	No	Geographic local of residence
9	Patients with any conditions	Education, Employment,	General dental checkups	Diagnostic cluster from Minnesota tiering	smoking status, alcohol consumption, vegetable intake, physical activity score	Education, Employment, social support measure	Emotional/spiritual health, optimism/pessimism score, physical health, physical shape
11	Patients with any conditions	Age, sex	Measures of resource consumption	ICD-8 and ICD-9-CM diagnostic codes; major ambulatory categories	Not specified	No	Not specified
10	Comorbid mental health and medical conditions	Age, gender, marital status, insurance, language	Emergency department visits, inpatient visits, specialty provider visits	ICD-9-CM diagnostic codes, medication use	High-risk medication (warfarin, insulin, narcotics), BMI	No	Quality of life, health behaviors
20	Presence of chronic conditions	Age, sex	Not specified	ICD-9-CM diagnostic codes	risk scores from aggregated clinical risk group	No	Not specified
27	Health levels, history of severe disease (acute, chronic, catastrophic conditions)	Age, sex	Not specified	ICD-9-CM diagnostic codes, procedure codes	risk scores from aggregated clinical risk group	No	Not specified
13	Patients with COPD	Age, sex	General practitioners, specialist visits, emergency department visits and inpatient visits	ICD-9-CM diagnostic codes	Not specified	Estimates were aggregated by the regions associated with the local community service centers	Not specified

12	Chronic illness	Age, sex, marital status, insurance status	Inpatient admissions, outpatient visits	Comorbidity conditions, medication use	risk scores derived from the features fitted in	No	Mental health status
26	Patients with any conditions	Age, sex	Not specified	Diagnosis codes	ambulatory diagnostic groups (ADGs) derived risk clusters	No	Not specified
30	Patients with any conditions	Age, sex, race	Primary care visits, specialist visits, hospital admissions, emergency department visits	Comorbidities	Not specified	Adjusted for public rental housing	Not specified
28	Patients with any conditions	Age, sex, race, medical fund/ public assistance status	Physician or technical consult at polyclinic, surgery visit, a specialist outpatient consult, emergency visit or inpatient admission	Chronic condition clusters, complications	Not specified	No	Not specified
3	General population with or without medical conditions	Age, sex	Inpatient admissions, outpatient visits, general practitioner visits	Chronic condition clusters, medication prescriptions	Not specified	Townsend Deprivation Index	Not specified